

NIEPEWNOŚĆ W PROGNOZOWANIU HYDROLOGICZNYM



MAURYCY EMILIAN CIUPAK



NIEPEWNOŚĆ W PROGNOZOWANIU HYDROLOGICZNYM

Maurycy Emilian Ciupak

Instytut Meteorologii i Gospodarki Wodnej
Państwowy Instytut Badawczy
Warszawa 2021



SERIA PUBLIKACJI NAUKOWO-BADAWCZYCH IMGW-PIB

NIEPEWNOŚĆ W PROGNOZOWANIU HYDROLOGICZNYM

Autor

Maurycy Emilian Ciupak
IMGW-PIB/Centrum Badań i Rozwoju
Zakład Hydrologii i Inżynierii Zasobów Wodnych

Recenzenci

dr hab. Wojciech Jakubowski
prof. dr hab. inż. Renata Julita Romanowicz

Opracowanie redakcyjne

Rafał Stepnowski

Opracowanie graficzne i techniczne

Michał Seredin

ISBN: 978-83-64979-40-8

Wydawca

Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy
01-673 Warszawa, ul. Podleśna 61, E. content@imgw.pl

Redaktor Wydawniczy/Publishing Editor

Rafał Stepnowski, E. rafal.stepnowski@imgw.pl

PRZEDMOWA

W prezentowanej pracy przedstawiono zbiór istniejących oraz autorskich metod i modeli probabilistycznych, mających zastosowanie w praktyce hydrologicznej, które mogą być wykorzystane do poprawy symulacji i prognozy hydrologicznej. Pod uwagę wzięto dwa rodzaje złożonych modeli hydrologicznych składających się z preprocesora danych wejściowych do modeli regresyjnych i szwedzkiego modelu typu opad–odpływ HBV (Hydrologiska Byråns Vattenbalansavdelning), przy czym należy zaznaczyć, że preprocesor danych wejściowych jest także pewnym modelem probabilistycznym. W odniesieniu do modeli regresyjnych preprocesor danych wejściowych identyfikuje zdarzenia hydrologiczne będące przedmiotem modelowania matematycznego, wyznacza ilościowe charakterystyki zmiennych prognozowanych i prognozujących oraz wykonuje analizę statystyczną zmiennych będących wejściem do modeli za pomocą metod statystycznych. Natomiast preprocesor danych do modelu HBV przetwarza satelitarne obserwacje opadu będące zmiennymi wymuszającymi model hydrologiczny oraz asymiluje satelitarne obserwacje wilgotności gleby. Modelowanie wezbrania deszczowo-roztopowego wykonano za pomocą statycznych liniowych modeli Bayesowskich (SLMB) z parametrami estymowanymi metodą Bayesowską i techniką odwrotnego modelowania, używając wielokrotnej i wielowymiarowej wielokrotnej liniowej regresji (WLR-B, WWLR-B, WLR-InvB, WWLR-InvB). Modelowanie całorocznego hydrogramu przepływów i krótkoterminową prognozę obliczono za pomocą dynamicznych liniowych modeli Bayesowskich (DLMB), tj. Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR). Uzyskane wyniki symulacji porównano z wielokrotną liniową regresją (WLR) i sztucznymi sieciami neuronowymi (NN).

Do usuwania obciążenia z satelitarnych obserwacji meteorologicznych (opadu i wilgotności gleby) zastosowano algorytm transformaty funkcji dystrybuanty (BC) wykorzystujący trzy trójparametrowe rozkłady prawdopodobieństwa, tj. uogólniony wykładniczy (GE), gamma (GA) i Weibulla (WE). Powyższy algorytm zastosowano w procedurze asymilacji satelitarnych obserwacji wilgotności gleby do generowania wolnej od obciążenia wiązki prognozowanych zmiennych stanu modelu HBV w filtrach Kalmana (EnKF-BC i EnSRF-BC) oraz w połączeniu z filtrem wariacyjnym (4DVAR) w modelach hybrydowych (EnKF/4DVAR i EnSRF/4DVAR). Niepewność prognozy hydrologicznej we wszystkich zastosowanych modelach określono ilościowo w formie rozkładu prawdopodobieństwa za pomocą prognozy przedziałowej lub prognozy w skali błędu.

Zaproponowane w pracy metody i modele nie mają na celu zastąpienie używanych modeli deterministycznych, lecz utworzenie zbioru matematycznych alternatyw lub metod komplementarnych do istniejących rozwiązań zbudowanych na podstawach fizyki. Zastosowane prawdopodobieństwo i statystyka poszukują prawidłowości, praw przyrody, wyodrębiają maksymalne informacje z zaobserwowanych zjawisk w przeszłości w celu wnioskowania o tym, czego należy spodziewać się w przyszłości.

SPIS TREŚCI

WYKAZ ZMIENNYCH, STAŁYCH, OZNACZEŃ I SKRÓTÓW	7
SPIS UŻYTYCH POJĘĆ I STOSOWANE NAZEWNICTWO	21
1. WPROWADZENIE	25
1.1. Niepewność prognozy hydrologicznej a praktyka inżynierska	27
1.2. Cel i zakres pracy	28
1.3. Układ pracy.....	30
2. DANE	32
2.1. Zlewnie testowe do modelowania probabilistycznego w prognozowaniu hydrologicznym.....	32
2.2. Metodyka analizy statystycznej danych wejściowych do modeli.....	32
2.3. Dane wejściowe do statycznych liniowych modeli Bayesowskich (SLMB)	35
2.3.1. Opis hydrologii zlewni Sokołdy.....	35
2.3.2. Opracowanie danych wejściowych do modeli SLMB.....	36
2.3.3. Analiza statystyczna danych wejściowych do modeli SLMB	38
2.4. Dane wejściowe do dynamicznych liniowych modeli Bayesowskich (DLMB)	42
2.4.1. Opis hydrologii zlewni Górnej Narwi.....	42
2.4.2. Opracowanie danych wejściowych do modeli DLMB	43
2.4.3. Analiza statystyczna danych wejściowych do modeli DLMB	44
2.5. Dane wejściowe do usuwania błędu systematycznego (obciążenia) z meteorologicznych obserwacji satelitarnych i asymilacji satelitarnych obserwacji wilgotności gleby	51
2.5.1. Opis hydrologii zlewni Soły	51
2.5.2. Opracowanie meteorologicznych obserwacji z naziemnej sieci pomiarowo-obserwacyjnej i obserwacji satelitarnych.....	52
2.5.3. Analiza statystyczna danych wejściowych do korekcji obciążenia i metod asymilacji	54
3. METODY I MODELE	68
3.1. Metody i modele użyte do analizy niepewności w prognozowaniu hydrologicznym.....	68
3.2. Wielokrotna liniowa regresja (WLR) i wielowymiarowa wielokrotna liniowa regresja (WWLR).....	69
3.3. Statyczne liniowe modele Bayesowskie (SLMB)	71
3.3.1. Wnioskowanie i reguła Bayesowska	71
3.3.2. Modelowanie Bayesowskie modelu wielokrotnej liniowej regresji	74
3.3.3. Modelowanie Bayesowskie wielowymiarowego modelu wielokrotnej liniowej regresji	78
3.3.4. Zastosowanie modeli SLMB i warunki stosowalności	81
3.4. Sieci neuronowe.....	82
3.5. Dynamiczne liniowe modele Bayesowskie (DLMB).....	83

3.5.1. Dynamiczny liniowy model Bayesowski: Varying coefficient Regression (VCR)	84
3.5.2. Dynamiczny liniowy model Bayesowski: Discount Weighted Regression (DWR).....	87
3.5.3. Zastosowanie modeli DLMB i warunki stosowalności.....	89
3.6. Konceptualny hydrologiczny model typu opad-odpływ HBV	90
3.7. Metody usuwania błędu systematycznego (obciążenia)	90
3.7.1. Korekcja obciążenia metodą transformaty dystrybuanty (BC)	92
3.7.2. Zastosowanie metody do usuwania błędu systematycznego i warunki stosowalności	93
3.8. Metody asymilacji danych	94
3.8.1. Asymilacja satelitarnych obserwacji wilgotności gleby przy użyciu filtrów wiążkowych z uwzględnieniem metody transformaty dystrybuanty	96
3.8.1.1. Błędy modelu, obserwacji oraz zaburzanie zmiennych stanu modelu hydrologicznego	96
3.8.1.2. Nowa procedura wiążkowego filtra Kalmana w połączeniu ze schematem korekcji obciążenia przy użyciu transformaty dystrybuanty	99
3.8.2. Filtr wariacyjny (VAR) i metody hybrydowe	100
3.8.2.1. Asymilacja obserwacji satelitarnych przy użyciu filtra wariacyjnego (4DVAR).....	100
3.8.2.2. Asymilacja obserwacji satelitarnych metodami hybrydowymi	102
3.8.3. Zastosowanie metod asymilacji satelitarnych obserwacji wilgotności gleby i warunki stosowalności	103
3.9. Klasyfikacja modeli	104
3.10. Ocena modeli	105

4. PROGNOZA WEZBRANIA DESZCZOWO-ROZTOPOWEGO ZA POMOCĄ STATYCZNYCH LINIOWYCH MODELI BAYESOWSKICH (SLMB)	107
4.1. Metodyka opracowania statycznych liniowych modeli Bayesowskich.....	107
4.2. Przykład zastosowania statycznych liniowych modeli Bayesowskich do prognozowania wezbrania deszczowo-roztopowego	109
4.2.1. Wyniki i dyskusja.....	109
4.2.2. Podsumowanie i wnioski.....	117

5. KRÓTKOTERMINOWA PROGNOZA PRZEPŁYWU RZECZNEGO ZA POMOCĄ DYNAMICZNYCH LINIOWYCH MODELI BAYESOWSKICH (DLMB)	118
5.1. Metodyka opracowania modeli do symulacji rocznego hydrogramu i krótkoterminowej prognozy hydrologicznej.....	118
5.2. Przykład zastosowania modeli Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR) do symulacji rocznego hydrogramu oraz krótkoterminowej prognozy hydrologicznej.....	120
5.2.1. Wyniki i dyskusja.....	120
5.2.1.1. Ocena modeli WLR, NN, VCR i DWR do symulacji rocznych hydrogramów	123
5.2.1.2. Ocena modeli WLR, NN, VCR i DWR do prognozowania z wyprzedzeniem 1-, 2- i 3-dobowym	129

5.2.2. Podsumowanie i wnioski.....	133
6. GENEROWANIE WIĄZKOWYCH KRÓTKOTERMINOWYCH PROGNOZ PRZEPIŹYU RZECZNEGO Z WYKORZYSTANIEM KOREKCCI I ASYMLACJI OBSERWACJI SATELITARNYCH	134
6.1. Metodyka opracowania preprocesora danych wejściowych do modelu HBV.....	134
6.2. Przetwarzanie naziemnych i satelitarnych obserwacji meteorologicznych.....	136
6.2.1. Kalibracja deterministycznego modelu hydrologicznego HBV.....	136
6.2.2. Interpolacja i korekcja obciążenia (BC) obserwacji meteorologicznych z naziemnej sieci pomiarowo-obserwacyjnej i danych satelitarnych.....	137
6.2.3. Schemat wiązkowego filtra Kalmana (KF) użytego w procedurze asymilacji..	138
6.3. Przykład korekcji obciążenia i asymilacji obserwacji satelitarnych do generowania wiązkowych prognoz hydrologicznych.....	139
6.3.1. Symulacja przepływu przy użyciu preprocesora HBV.....	140
6.3.1.1. Ocena wpływu korekcji obciążenia BC na dokładność modelu i wybór skorygowanych obserwacji satelitarnych do procedury updatingu...	140
6.3.1.2. Algorytm symulacji przepływu przy użyciu preprocesora modelu HBV.....	142
6.3.1.3. Symulacja przepływu z korekcją obciążenia satelitarnego opadu bez asymilacji.....	144
6.3.1.4. Symulacja przepływu z użyciem dwóch metod updatingu, tj.: zastąpienie zmiennej stanu modelu HBV skorygowaną satelitarną wilgotnością gleby i asymilacja satelitarnej wilgotności gleby.....	144
6.3.2. Przykład updatingu i prognozy hydrologicznej modelu HBV w zlewni Soły w profilu wodowskazowym Żywiec.....	147
6.3.3. Podsumowanie i wnioski.....	150
7. PODSUMOWANIE I WNIOSKI KOŃCOWE	152
BIBLIOGRAFIA	157
SUMMARY	166
ZAŁĄCZNIKI	
Dodatek A. Brzegowe rozkłady prawdopodobieństwa zastosowane w metodzie usuwania obciążenia.....	167
Dodatek B. Metoda symulacji Monte Carlo (MCMC) do uzyskiwania rozkładów <i>a posteriori</i>	167
Dodatek C. Wielowymiarowe i macierzowe rozkłady prawdopodobieństwa występujące w statycznych i dynamicznych modelach Bayesowskich.....	168
Dodatek D. Równanie do updatingu zmiennych stanu w filtrze – ensemble Kalman filter (EnKF).....	170
Dodatek E. Równanie do updatingu zmiennych stanu w filtrze – ensemble square root filter (EnSRF).....	171
Dodatek F. Metoda wyboru zbioru zmiennych prognozujących, dla którego średnia wartość średniego błędu prognozy przyjmuje minimalną wartość.....	172

WYKAZ ZMIENNYCH, STAŁYCH, OZNACZEŃ I SKRÓTÓW

Rozdział 1

Zmienne i stałe:

$Q_{s,\tau}$	symulowany przepływ w chwili τ
δ_τ	proces resztowy w chwili τ
Q_o	przepływ obserwowany
τ_0	chwila początkowa prognozy

Oznaczenia i skróty:

bias	błąd systematyczny (obciążenie)
SLMB	statyczny liniowy model Bayesowski
DLMB	dynamiczny liniowy model Bayesowski
HBV	szwedzki konceptualny, częściowo rozłożony, model hydrologiczny typu opad-odpływ, Hydrologiska Byråns Vattenbalansavdelning
WLR	wielokrotna liniowa regresja
WWLR	wielowymiarowy model wielokrotnej liniowej regresji
WLR-B	WLR z parametrami estymowanymi metodą Bayesowską
WWLR-B	WWLR z parametrami estymowanymi metodą Bayesowską
WLR-InvB	WLR z parametrami estymowanymi techniką odwrotnego modelowania (Bayesian inverse modelling, InvB)
WWLR-InvB	WWLR z parametrami estymowanymi techniką odwrotnego modelowania (Bayesian inverse modelling, InvB)
VCR	dynamiczny liniowy model Bayesowski – Varying Coefficient Regression
DWR	dynamiczny liniowy model Bayesowski – Discount Weighted Regression
BC	statystyczna metoda usuwania obciążenia (bias correction), np. za pomocą transformaty dystrybuanty (distribution derived transformation)
EnKF	wiązkowy filtr Kalmana (ensemble Kalman filter)
EnSRF	wiązkowy filtr (ensemble square root filter)
EnKF-BC	wiązkowy filtr Kalmana z postprocesową procedurą usuwania obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu modelu HBV za pomocą metody BC (ensemble Kalman filter-bias correction)
EnSRF-BC	wiązkowy filtr SRF z postprocesową procedurą usuwania obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu modelu HBV za pomocą metody BC (ensemble square root filter-bias correction)
4DVAR	czterowymiarowy filtr wariacyjny (4D variational filter)
EnSRF-BC/4DVAR	hybrydowy filtr złożony z filtru SRF z procedurą BC i filtru wariacyjnego 4DVAR
EUMETSAT	Europejska Organizacja Eksploatacji Satelitów Meteorologicznych (European Organization for the Exploitation of Meteorological Satellites)
GE	uogólniony rozkład wykładniczy
GA	rozkład gamma
WE	rozkład Weibulla
GE-GE, GA-GA, WE-WE, GE-GA, GE-WE, GA-WE	transformacja dystrybuanty typu uogólniony wykładniczy-uogólniony wykładniczy, gamma-gamma, Weibull-Weibull, uogólniony wykładniczy-gamma, uogólniony wykładniczy-Weibull, gamma-Weibull
NN	sieci neuronowe (neural networks)

Rozdział 2

Zmienne i stałe:

d	wymiar przestrzeni danych występujący w definicji uogólnionego rozkładu metody wektorów nośnych do wykrywania elementów odstających, SVM
τ	zmienna niezależna czas
n	liczebność próby losowej
X_1, X_2, \dots, X_n	próba losowa składająca się z n niezależnych zmiennych losowych
x_1, x_2, \dots, x_n	realizacja zmiennej losowej X
Q_τ	przepływ rzeczny w chwili τ (m^3s^{-1})
Q_{\max}	przepływ kulminacyjny wezbrania (m^3s^{-1})
V_w	objętość wezbrania (10^6m^3)
T_w	czas trwania wezbrania (doba)
ES_5	wodny ekwiwalent śniegu z uwzględnieniem sumy opadu z 5 dób poprzedzających roztopy (10^6m^3)
AT_5	średnia dobową temperaturę powietrza w ciągu 5 dób przed roztopami ($^{\circ}\text{C}$)
Q_1	przepływ w momencie początku roztopów (m^3s^{-1})
GT_1	średnia dobową temperaturę gleby na głębokości 0,5 m na początku roztopów ($^{\circ}\text{C}$)
AT_{SRE}	5-dniowa prognoza średniej dobowej temperatury powietrza opracowana (postawiona) w dniu początku roztopów ($^{\circ}\text{C}$); w opracowaniu reprezentuje 5-dniową zaobserwowaną średnią dobową temperaturę powietrza
$T_{\text{SRE}}^{\text{pow}}$	średnia temperatura powietrza ($^{\circ}\text{C}$)
$\tau_p, \tau_{k1}, \tau_{p1}, \tau_k$	odpowiednio początek i koniec wezbrań cząstkowych
τ_{dws}	moment wejścia do modelu danych zaobserwowanych: ES_5, AT_5, Q_1, GT_1
τ_{dw}	początek roztopów zbieżny z momentem, w którym średnia dobową temperaturę powietrza osiągnęła lub przekroczyła 0°C i moment wejścia do modelu danych prognozowanych, tj. AT_{SRE}
τ_p	początek okresu prognostycznego
τ_k	zakończenie okresu prognostycznego
$\Delta \tau_{\text{dw}}$	różnica czasu między terminem stawiania prognozy, τ_{dw} , a momentem, w którym wyznaczone są zaobserwowane dane wejściowe do modelu, τ_{dws}
$\Delta \tau_w$	okres wyprzedzenia (uprzedzenia), tj. czas jaki upływa od momentu stawiania prognozy, τ_{dw} , do momentu, w którym zaobserwowany przepływ przekroczy graniczną wartość $Q_{g,w}$, tj. do początku okresu prognostycznego, τ_p
$\Delta \tau_{pk}$	okres prognostyczny wyznaczany jako różnica między zakończeniem okresu prognostycznego, τ_k , tzn. w dniu w którym zaobserwowany przepływ będzie niższy od wartości granicznej $Q_{g,w}$, a początkiem okresu prognostycznego, τ_p
$\Delta \tau_h$	horyzont prognozy obejmujący czas od momentu wejścia danych zaobserwowanych do modelu, τ_{dws} , do końca okresu prognostycznego, τ_k
$\ln Q_{\max}, \ln V_w, \ln T_w,$ $\ln ES_5, \ln Q_1$	zlogarytmowane zmienne losowe: Q_{\max}, V_w, T_w, ES_5 , i Q_1
Q_τ^{WIZ}	codzienny przepływ w zlewni Górnej Narwi w Wiźnie (m^3s^{-1})
$Q_{\tau-2}^{\text{STR}}, Q_{\tau-2}^{\text{BUR}}$	przepływy w profilach wodowskazowych Strękowa Góra i Burzyn zaobserwowane 2 doby wcześniej w stosunku do symulowanego hydrogramu w Wiźnie (m^3s^{-1})
$P_{\tau-2}^{\text{AREA}}$	średnia wysokość warstwy opadu w zlewni Górnej Narwi powyżej wodowskazu Wizna, zaobserwowana 2 doby wcześniej (mm)
$T_{\tau-2}^{\text{MAX}}, T_{\tau-2}^{\text{MIN}}$	średnie dobowe wartości maksymalnej i minimalnej temperatury powietrza zaobserwowane 2 doby wcześniej na obszarze modelowanej zlewni ($^{\circ}\text{C}$)
x, y	odpowiednio zmienna niezależna i zależna
p	prawdopodobieństwo
P_{ann}	opad roczny w zlewni Soły (mm)

P_{03}	obserwacja satelitarna natężenia opadu otrzymana w ramach projektu H-SAF, reprezentująca średnie natężenie opadu mierzone w cyklach 15-minutowych czterokrotnie w ciągu godziny (mm h^{-1}). Obserwacje P_{03} są związane z rozdzielczością siatki pomiarowej – od 3,5 km horyzontalnie do 5,9 km wertykalnie
$P_{04}, P_{05_{24}}$	obserwacje satelitarne opadu otrzymane w ramach projektu H-SAF, P_{04} do opisu aktualnego stanu pogody i jego rozwoju w ciągu 6 godzin i $P_{05_{24}}$ reprezentujący 24-godzinną sumę opadu (mm). Obserwacje satelitarne związane są z rozdzielczością siatki pomiarowej – odpowiednio od 5,2 km horyzontalnie do 8,5 km wertykalnie oraz od 3,5 km do 5,9 km
P_{OBS} H_{14}	sumy dobowe opadów zaobserwowane na stacjach meteorologicznych (mm) satelitarna obserwacja wilgotności gleby (%). Obserwacje związane są z rozdzielczością siatki pomiarowej – od 16,29 km horyzontalnie do 25,07 km wertykalnie
P_d, P_h	odpowiednio dobowe i godzinne obserwowane sumy opadu z naziemnej sieci pomiarowo-obszaryjnej IMGW-PIB (mm)
I_{rz0} $H_{14}^{1,00}, H_{14}^{2,89}$	wskaźnik wilgotności gleby w strefie korzeniowej satelitarna wilgotność gleby w pionowym profilu, obliczona jako średnia ważona z trzech i czterech głębokości, tj. odpowiednio (0-1,00 m, $H_{14}^{1,00}$ i 0-2,89 m, $H_{14}^{2,89}$) (%)
θ_{HBV} Fc θ	wilgotność gleby wygenerowana przez model HBV (mm) połowa pojemność wodna gleby, zmienna stanu modelu hydrologicznego HBV przeliczona asymilowana (w ramach preprocesora danych wejściowych do modelu HBV) prognozowana zmienna stanu wilgotności gleby (%)
$\theta_{HBV-ZIMA}, \theta_{HBV-LATO},$ $H_{14}^{1,00-ZIMA}, H_{14}^{1,00-LATO},$ $H_{14}^{2,89-ZIMA}, H_{14}^{2,89-LATO}$	zmiennie losowe: $\theta_{HBV}, H_{14}^{1,00}$ i $H_{14}^{2,89}$ pogrupowane według sezonu zimowego i letniego

Oznaczenia i skróty:

SVM	metoda wektorów nośnych do wykrywania elementów odstających (support vector machine)
$Q \in \mathbb{R}^d$	uogólniony kwantyl rozkładu d -wymiarowej przestrzeni danych \mathbb{R}
M-K	test Manna-Kendalla do wykrywania i badania istotności tendencji w ciągach pomiarowych
ADF	rozszerzony test pierwiastka jednostkowego Dickey'a-Fullera do sprawdzenia stacjonarności zależnej zmiennej losowej
KPSS	Test Kwiatkowskiego, Phillipsa, Schmidta i Shina, uzupełniający sprawdzanie stacjonarności zależnej zmiennej losowej
$E(X)$	wartość oczekiwana zmiennej losowej X
\bar{X}	wartość średnia zmiennej losowej X
\bar{x}	wartość średnia z próby losowej
$\text{Var}(X)$	wariancja zmiennej losowej X
s^2	estymator wariancji w próbie losowej
$\text{Cov}(X, Y)$	kowariancja zmiennych losowych X i Y
SKE	współczynnik skośności
K	współczynnik spłaszczenia
$Q_{g,w}$	wartość przepływu granicznego wg kryterium hydrologicznego
NWW	Najniższa Wysoka Woda
NWQXI-IV, 1966-2002	Najniższy Wysoki Przepływ zaobserwowany w sezonie zimowym w latach 1966-2002
$Q_{r,w}$	kryterium rozdziału sąsiadujących wezbrań
SSQXI-IV, 1966-2002	średni ze średnich przepływów zaobserwowany w sezonie zimowym w latach 1966-2002
α, λ	parametry skali i kształtu rozkładów prawdopodobieństwa GE, GA i WE

χ^2	test dobroci dopasowania χ^2 -Pearsona
$\chi^2_{kr,\alpha}$	wartość krytyczna testu χ^2 -Pearsona przy poziomie istotności α
H_0	hipoteza zerowa
ε	oszacowane dolne ograniczenie rozkładu prawdopodobieństwa
$\ln(X - \varepsilon)$	funkcja logarymiczna nieprzekształconej zmiennej losowej $X - \varepsilon$
λ	statystyka Kołmogorowa
λ_{kr}	wartość krytyczna statystyki Kołmogorowa
NPN	Narwiański Park Narodowy
Lag	parametr opóźnienia występujący w teście współczynnika autokorelacji
Beta	znormalizowane współczynniki regresji
Błąd std. Beta	błąd standardowy znormalizowanych współczynników regresji
β	współczynniki regresji
Błąd std. β	błąd standardowy estymacji β
B_0, β_1	odpowiednio wyraz wolny i współczynnik kierunkowy
F	test Fishera-Snedecora
r^2, r	odpowiednio współczynnik determinacji i korelacji liniowej Pearsona
p -wartość	prawdopodobieństwo testowe (p -value)
$ADF_{kr,\alpha=0,05}$	wartość krytyczna statystyki ADF
$KPSS_{kr,\alpha=0,05}$	wartość krytyczna statystyki KPSS
H-SAF	Infrastruktura Zastosowań Satelitarnych na rzecz Wsparcia Hydrologii Operacyjnej i Gospodarki Wodnej (Support to Operational Hydrology and Water Management)
IMGW-PIB	Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy
IR	obrazy w podczerwieni z operacyjnych satelitów geostacjonarnych GEO (Geostationary Earth Orbit)
MW	detektor mikrofalowy operacyjnych satelitów LEO (Low Earth Orbit)
SEVIRI IFOV	optyczny przyrząd radiometru zapewniający obraz w czterech kanałach widzialnych i w bliskiej podczerwieni VNIR oraz ośmiu kanałach podczerwieni IR
QPF	ilościowa prognoza opadów
NWP	numeryczny model pogody (Numerical Weather Prediction)
ECMWF IFS	Europejskie Centrum Prognoz Średnioterminowych w Zintegrowanym Systemie Prognozowania danych (European Centre for Medium Range Weather Forecast Integrated Forecasting System)
SM-DAS-2 (H ₁₄)	wskaźnik wilgotności gleby $I_{r2\theta}$ w strefie korzeniowej otrzymywany w ECMWF IFS
Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land	procedura asymilacji danych hydrologicznych
$f(\dots)$	notacja zależności funkcyjnej, np. $f(p_{OBS})$ oznacza funkcję gęstości realizacji zmiennej losowej P_{OBS}
K-S	statystyka testu zgodności Kołmogorowa-Smirnowa
x_p	funkcja kwantylowa
$t_p(\lambda)$	kwantyl standaryzowany rozkładu gamma o prawdopodobieństwie przewyższenia p
$\Gamma(\lambda)$	funkcja gamma Eulera

Rozdział 3

Zmienne i stałe:

Wielokrotna liniowa regresja (WLR)

Y	zmienna losowa prognozowana
\mathbf{X}	wektor zmiennych losowych prognozujących
y	realizacja losowej zmiennej prognozowanej Y (w postaci skalara)

x_1, x_2, \dots, x_d	d zmiennych prognozujących
$E(y x_1, x_2, \dots, x_d)$	oczekiwana wartość zmiennej prognozowanej Y dla WLR
y_p	realizacja zmiennej losowej Y dla $p = 1, \dots, l$ w wielorównaniowej zależności regresji WLR
\mathbf{x}	realizacja d wymiarowego wektora zmiennej losowej prognozującej \mathbf{X}
\mathbf{y}	realizacja l wymiarowego wektora zmiennej losowej prognozowanej \mathbf{Y}
y_1, y_2, \dots, y_l	l zmiennych prognozowanych
\mathbf{Y}	n wymiarowy wektor z zaobserwowanymi wartościami zmiennej prognozowanej w modelu wielokrotnej regresji, WLR
\mathbf{X}	$n \times d$ wymiarowa macierz obserwacji zmiennych prognozujących
$\boldsymbol{\beta}$	d wymiarowy wektor współczynników kierunkowych w modelu wielokrotnej regresji WLR
$\boldsymbol{\epsilon}$	n wymiarowy wektor błędów modelu i niepewności parametrów o wielowymiarowym rozkładzie normalnym, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{E})$, gdzie \mathbf{E} jest $n \times n$ wymiarową macierzą jednostkową
\mathbf{b}	d wymiarowy wektor przykładowych wartości współczynników kierunkowych dla $\boldsymbol{\beta}$
$S(\mathbf{b})$	funkcja minimalizująca sumę kwadratów odchyłań w metodzie najmniejszych kwadratów, MNK
b_0, b_1, \dots, b_d	d nieznanych współczynników regresji

Wielowymiarowa wielokrotna liniowa regresja (WWLR)

$E(\mathbf{y} \mathbf{x})$	oczekiwana wartość wektorowej zmiennej prognozowanej $\mathbf{y}: y_1, y_2, \dots, y_l$ zależnej od wektora zmiennych prognozujących $\mathbf{x}: x_1, x_2, \dots, x_d$ dla WWLR
\mathbf{Y}	$n \times l$ wymiarowa macierz z obserwacjami zmiennych prognozowanych w wielowymiarowym modelu wielokrotnej regresji, WWLR
\mathbf{E}	$n \times n$ wymiarowa macierz jednostkowa
$\boldsymbol{\beta}_{(p)}$	d wymiarowy wektor współczynników kierunkowych dla p -tej zmiennej prognozowanej, gdzie $p = 1, \dots, l$
$\boldsymbol{\epsilon}_{(p)}$	n wymiarowy wektor błędów modelu dla p -tej zmiennej prognozowanej
$\hat{\boldsymbol{\beta}}_{(p)}$	estymowany d wymiarowy wektor współczynników kierunkowych dla p -tej zmiennej prognozowanej metodą najmniejszych kwadratów w modelu WWLR

Statyczne liniowe modele Bayesowskie (SLMB)

$\boldsymbol{\theta}$	wektor parametrów modelu matematycznego
$\boldsymbol{\theta} = (\theta_1, \dots, \theta_{k_j})$	zestaw parametrów modelu, gdzie k_j określa liczbę parametrów θ_i w hipotezie (modelu) H_j przy danym stanie wiedzy
$H_j = H_j(\boldsymbol{\theta})$	j -ta hipoteza dotycząca wybranego do analizy modelu matematycznego przy danym stanie wiedzy dla $j = 1, \dots, m$, gdzie m jest liczbą możliwych hipotez
σ^2	wariancja błędu dopasowania modelu w wielokrotnej regresji
d	w algorytmie Gibbsa liczba zmiennych jednowymiarowego warunkowego rozkładu (liczba parametrów),
$\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \dots, \theta_d^0)$ $\theta^1, \dots, \theta^M$	zbiór początkowych wartości łańcucha Markowa w generatorze Gibbsa wylosowana pełna próba parametrów modelu regresji z wielowymiarowego warunkowego rozkładu o rozmiarze M
$\hat{\boldsymbol{\beta}}$	estymator metody najmniejszych kwadratów (MNK wektora $\boldsymbol{\beta}$)
\mathbf{X}^T	$d \times n$ wymiarowa macierz transponowana zmiennych prognozujących
$\mathbf{V}_{\boldsymbol{\beta}}$	$d \times d$ wymiarowa macierz obliczana jako $(\mathbf{X}^T \mathbf{X})^{-1}$
$\hat{\sigma}^2$	estymator wariancji σ^2 , tzw. współczynnik skali
\mathbf{Y}_0	l wymiarowy wektor zmiennych prognozowanych
\mathbf{X}_0	$n \times d$ wymiarowa macierz zmiennych prognozujących w równaniu prognostycznym
$\hat{\boldsymbol{\beta}}_0$	średnia wartość parametrów wektora $\boldsymbol{\beta}$

\mathbf{V}_0	średnia wartość macierzy kowariancji rozkładu <i>a priori</i>
\mathbf{Y}	$n \times l$ wymiarowa macierz z obserwacjami zmiennych prognozowanych w wielowymiarowym modelu wielokrotnej regresji WWLR
$\boldsymbol{\beta}$	$d \times l$ wymiarowa macierz współczynników kierunkowych modelu WWLR
$\boldsymbol{\epsilon}$	$n \times l$ wymiarowa macierz błędów modelu WWLR
$\boldsymbol{\Sigma}$	$l \times l$ wymiarowa macierz kowariancji modelu WWLR
Sieci neuronowe	
$Y_{\tau k}$	prognozowane k -te wyjście w chwili τ
d	liczba wejściowych warstw neuronów (liczba zmiennych prognozujących)
m	liczba ukrytych warstw neuronów
w_{j0}	waga niezwiązana z żadnym sygnałem dla j -tego ukrytego neuronu
$F_{\tau i}^{w_{ji}}$	waga, która łączy i -tą zmienną prognozującą w warstwie wejścia z j -tym neuronem w ukrytej warstwie
w_{k0}	waga niezwiązana z żadnym sygnałem dla k -tego neuronu wyjścia
w_{kj}	waga, która łączy j -ty neuron w ukrytej warstwie i k -ty neuron warstwy wyjściowej
$F_{\tau i}$	i -ta zmienna prognozująca w chwili τ
Dynamiczne liniowe modele Bayesowskie (DLMB)	
\mathbf{Y}	$n \times l$ wymiarowa macierz z zaobserwowanymi wartościami zmiennych prognozowanych
\mathbf{Y}_{τ}	l wymiarowy wektor z zaobserwowanymi wartościami zmiennych prognozowanych w chwili τ
\mathbf{F}_{τ}^T	$n \times d$ wymiarowa transponowana macierz zaobserwowanych zmiennych prognozujących w chwili τ
$\boldsymbol{\theta}_{\tau}$	$d \times l$ wymiarowa nieobserwowana macierz zmiennych stanu
\mathbf{v}_{τ}	l wymiarowy wektor zaobserwowanych błędów w równaniu obserwacji
$\boldsymbol{\Sigma}$	$l \times l$ wymiarowa macierz kowariancji macierzy
V_{τ}	obserwacyjny czynnik skali wariancji
\mathbf{G}_{τ}	$d \times d$ wymiarowa macierz przejścia z jednego stanu do drugiego
$\boldsymbol{\Omega}_{\tau}$	$d \times l$ wymiarowa macierz, w której kolumny są błędami transformacji zmiennych w równaniu stanu
\mathbf{W}_{τ}	$d \times d$ wymiarowa macierz kowariancji macierzy losowej $\boldsymbol{\Omega}_{\tau}$ w równaniu przejścia
\mathbf{m}_{τ}	$d \times l$ wymiarowa macierz zawierająca współczynniki kierunkowe dla każdej zmiennej prognozowanej
\mathbf{C}_{τ}	$d \times d$ wymiarowa macierz kowariancji
\mathbf{S}_{τ}	$l \times l$ wymiarowa zaobserwowana <i>a priori</i> macierz wariancji
n_{τ}	<i>a priori</i> stopień swobody
\mathbf{a}_{τ}	$d \times l$ wymiarowa macierz będąca estymatorem współczynników kierunkowych
\mathbf{R}_{τ}	$d \times d$ wymiarowa macierz ze zmodyfikowanymi wartościami kowariancji \mathbf{C}_{τ}
D_0	zbiór początkowej informacji (tj.: $\mathbf{m}_0, \mathbf{C}_0, \mathbf{S}_0, \mathbf{G}_0, \mathbf{F}_0$) w chwili $\tau = 0$, dla $\tau > 0$ $D_{\tau} = \{\mathbf{Y}_{\tau}, D_{\tau-1}\}$
f_{τ}	l wymiarowy wektor z prognozą wektora \mathbf{Y}_{τ} na podstawie informacji $D_{\tau-1}$
Q_{τ}	skala wariancji
$\mathbf{A}_{\tau}, \mathbf{e}_{\tau}$	macierz i wektor o wymiarach odpowiednio $d \times n$ i l
\mathbf{m}_0	początkowa (inicjująca) $d \times l$ wymiarowa macierz zawierająca współczynniki kierunkowe dla każdej zmiennej prognozowanej
\mathbf{C}_0	początkowa (inicjująca) $d \times d$ wymiarowa diagonalna macierz kowariancji
$\boldsymbol{\theta}_0$	początkowa (inicjująca) $d \times l$ wymiarowa nieobserwowana macierz zmiennych stanu
δ	współczynnik dyskontowy

Konceptualny hydrologiczny model typu opad–odpływ HBV

<i>MELT</i>	topniejący śnieg (mm)
<i>T</i>	temperatura powietrza (°C)
<i>WC</i>	zawartość topniejącej wody w pokrywie śniegu (mm)
<i>SP</i>	pokrywa śniegu (mm)
<i>DRAIN</i>	drenaż ze śniegu (mm)
<i>IN</i>	infiltracja (mm)
<i>R</i>	odpływ (mm)
<i>SM</i>	wilgotność gleby (mm)
<i>EA</i>	rzeczywista ewapotranspiracja (mm)
<i>EP</i>	potencjalna ewapotranspiracja (mm)
<i>UR</i>	górne ograniczenie zawartości zbiornika wody gruntowej (mm)
<i>LR</i>	dolne ograniczenie zawartości zbiornika wody gruntowej (mm)
<i>Q0, Q1, Q4</i>	komponenty odpływu (mm doba ⁻¹)
<i>TT</i>	progowa temperatura powietrza (°C)
<i>CFMAX</i>	współczynnik topnienia śniegu w metodzie stopień-dzień (mm °C ⁻¹ doba ⁻¹)
<i>LP</i>	progowa część pojemności połowej
<i>BETA</i>	współczynnik zasilania strefy saturacji
<i>K0</i>	maksymalny współczynnik recesji górnego zbiornika wody gruntowej (doba ⁻¹)
<i>K1</i>	współczynnik recesji górnego zbiornika wody gruntowej (doba ⁻¹)
<i>UZL0</i>	wartość progowa współczynnika recesji K0 (mm)
<i>PERC</i>	pojemność perkolacji z górnego do dolnego zbiornika (mm doba ⁻¹)
<i>K2</i>	współczynnik recesji dolnego zbiornika wody gruntowej (doba ⁻¹)
<i>MAXBAS</i>	parametr funkcji wagowej transformującej odpływ (doba)

Metody usuwania błędu systematycznego (obciążenia)

X_{MOD}	zmienna losowa modelowana
X_{OBS}	zmiennie losowe obserwowana
X_{COR}	zmienna losowa skorygowana
\hat{X}_{OBS}	estymator zmiennej losowej obserwowanej X_{OBS}
$a, b, c, \varepsilon, \tau$	parametry funkcji transformacji parametrycznej podlegające estymacji
ε	oszacowane dolne ograniczenie rozkładu prawdopodobieństwa
P_{MOD}	opad modelowany (np. satelitarny)
P_{MOD}^{BC}	skorygowany metodą BC opad modelowany

Metody asymilacji danych

G	stała – parametr relaksacji występujący w metodzie nudgingu
Q	przepływ rzeczny (m ³ s ⁻¹)
θ'	zaburzona zmienna stanu wilgotności gleby
UR', LR', P'	zaburzone zmienne stanu górnego i dolnego ograniczenia zawartości zbiornika wody gruntowej i opadu
δ_j	średnie obciążenia zaburzonych zmiennych stanu w n elementowej wiązce w chwili j
θ	prognozowany stan wilgotność gleby przez model hydrologiczny
$\hat{\theta}_{j,i}$	skorygowany i -ty element wiązki (o liczebności n) prognozowanych zmiennych stanu w modelu hydrologicznym w chwili j
$\theta'_{j-1,i}$	zaburzony i -ty element wiązki w chwili τ
θ_j, θ_{j-1}	wilgotność gleby obliczona przez model hydrologiczny w chwili j i $j-1$
$\mathbf{\theta}, \mathbf{\theta}'$	odpowiednio wektor niezaburzonej i macierz zaburzonej prognozy wilgotności gleby
$UR'_{j,i}$	zaburzona zmienna stanu UR_j , tj. górne ograniczenie zawartości zbiornika wody gruntowej w fazie odpowiedzi w j -tym kroku symulacji dla i -tego elementu wiązki, gdzie: $i = 1, 2, \dots, n$

$UR_{out,j}$	odpływ z górnego zbiornika wody gruntowej w fazie odpowiedzi (mm) obliczony w j -tym kroku symulacji
khq_j	zmienna stanu modelu HBV użyta do obliczania współczynnika recesji górnego zbiornika wody gruntowej w j -tym kroku symulacji
α	miara nieliniowości generowanego odpływu, który jest funkcją odpowiedzi przekształcającą nadmiar wody ze strefy wilgotności gleby w odpływ zaburzona zmienna stanu LR_j , tj. dolne ograniczenie zawartości zbiornika wody gruntowej w fazie odpowiedzi w j -tym kroku symulacji dla i -tego elementu wiązki, gdzie: $i = 1, 2, \dots, n$
$LR'_{j,i}$	
$LR_{out,j}$	odpływ z dolnego zbiornika wody gruntowej w fazie odpowiedzi (mm) obliczony w j -tym kroku symulacji
k_{4j}	współczynnik recesji dolnego zbiornika wody gruntowej w fazie odpowiedzi (doba ⁻¹)
m	liczba zmiennych stanu modelu, liczba zmiennych tła stanu modelu, tj. zmiennych stanu obliczonych przez model hydrologiczny
H	$l \times m$ wymiarowa macierz operatora, który przekształca stan modelu do przestrzeni obserwacji
K	$m \times l$ wymiarowa macierz zysku Kalmana
X^a	$m \times n$ wymiarowa macierz uaktualnionych zmiennych stanu modelu
X^b	$m \times n$ wymiarowa macierz tła zmiennych stanu modelu (zmiennych obliczonych przez model)
X₁^b, ..., X_n^b	wektory wszystkich zmiennych stanu dla każdego n elementu wiązki przed updatingiem
Z	$l \times n$ wymiarowa macierz obserwacji <i>in situ</i>
Z₁, ..., Z_n	wektory obserwacji <i>in situ</i> dla każdego n elementu wiązki
P^b	$m \times m$ wymiarowa macierz kowariancji błędu modelu, która jest obliczana <i>a priori</i> bezpośrednio z anomalii (odchylenia) wiązki
X^{b'} = X^b - X̄^b	macierz odchylenia zmiennych stanu modelu od średnich wartości wiązki
R	$l \times l$ wymiarowa macierz kowariancji błędu obserwacji <i>in situ</i>
Z' = Z - Z̄	macierz odchylenia obserwacji <i>in situ</i> od średnich wartości wiązki
K̄	$m \times l$ wymiarowa macierz zysku Kalmana użyta do aktualizacji odchylenia (anomalii) od średniej wartości wiązki zmiennych stanu
X̄^a	$m \times l$ wymiarowa macierz średnich wartości wiązki uaktualnionych zmiennych stanu modelu
X̄^b	$m \times l$ wymiarowa macierz średnich wartości wiązki tła zmiennych stanu modelu
X^{a'}	$m \times l$ wymiarowa macierz odchyłeń macierzy tła zmiennych stanu (po aktualizacji) od macierzy średnich wartości wiązki uaktualnionych zmiennych stanu modelu, tj. $X_i^{a'} = X_i^a - \bar{X}_i^a$
X^{b'}	$l \times n$ wymiarowa macierz odchyłeń wszystkich elementów wiązki tła zmiennych stanu modelu (przed aktualizacją) od macierzy średnich wartości wiązki tła zmiennych stanu modelu, tj. $X_i^{b'} = X_i^b - \bar{X}_i^b$
Z'	$l \times n$ wymiarowa macierz odchyłeń każdego elementu wiązki obserwacji <i>in situ</i> od średniej wartości wiązki obserwacji, tj. $Z_i' = Z_i - \bar{Z}_i$
Filtr wariacyjny (VAR) i metody hybrydowe	
w	długość okna asymilacji w filtrze 4DVAR, równa długości updatingu
l	wymiar wektora obserwacji <i>in situ</i> , liczba zaburzanych obserwacji zmiennych stanu w oknie asymilacji o długości w
LAG	długość updatingu
X_i^b	tło stanu modelu w i -tej chwili wewnątrz okna asymilacji o długości w
X_{i=0}^b	tło stanu modelu na początku okna asymilacji dla $i = 0$ (warunki inicjujące)

\mathbf{X}^a	$m \times w$ wymiarowa macierz z uaktualnionymi zmiennymi stanu modelu
$\mathbf{X}_{i=0}^a$	uaktualnione tło stanu modelu na początku okna asymilacji
\mathbf{P}^s	$m \times m$ wymiarowa macierz kowariancji błędu modelu skojarzona z macierzą kowariancji błędu modelu \mathbf{P}^b w metodzie filtru Kalmana
\mathbf{R}	$l \times l$ wymiarowa macierz kowariancji błędu obserwacji <i>in situ</i>
\mathbf{Z}	$l \times w$ wymiarowa macierz obserwacji <i>in situ</i>
\mathbf{H}	$l \times w$ wymiarowa macierz przyporządkowująca stany obliczone przez model do stanów zaobserwowanych
$\mathbf{Z}_i, \mathbf{H}_i, \mathbf{R}_i$	macierze opisujące zależności między obserwacjami \mathbf{Z}_i a tłem stanu modelu \mathbf{X}_i^b w i -tej chwili okna asymilacji
$k = 1, \dots, m$	indeks m zmiennych stanu zmagazynowanych wewnątrz macierzy $\mathbf{X}_{i=0,k}^b$ i $\mathbf{X}_{i=0,k}^a$ na początku okna asymilacji $i = 0$
f	moment, w którym obliczana jest prognoza hydrologiczna
$\mathbf{X}_{i=0,k,f-w}^b$	macierz zaobserwowanego tła stanu modelu w chwili $f-w$ na początku okna asymilacji $i = 0$
$\mathbf{X}_{i=0,k,f-w}^a$	macierz z uaktualnionym stanem modelu wraz ze zmiennymi stanu obliczonymi przez model hydrologiczny w chwili $f-w$ na początku okna asymilacji $i = 0$
\mathbf{X}_{f-w}^b	tło stanu modelu w chwili $f-w$
$\mathbf{X}_p, \mathbf{X}_{PE}$	w wymiarowe wektory z multiplikatywnymi czynnikami korygującymi opad i ewapotranspirację, minimalizowanymi w celu odszukania takiej macierzy zmiennych stanu modelu \mathbf{X}_{f-w}^a , która minimalizuje funkcję J_f wewnątrz okna asymilacji w w kroku $i = f-w+1, \dots, f$ dla $i = i+1$
$\mathbf{X}_k^{\text{b,min}}, \mathbf{X}_k^{\text{b,max}}$	m wymiarowe wektory zawierające odpowiednio dolne i górne ograniczenie zmienności k -tej zmiennej stanu modelu
β	współczynnik mieszania, waga do szacowania kowariancji błędów dla każdego elementu wiązki

Klasyfikacja i ocena modeli

a	stała w równaniu zasady superpozycji
$x_i(\tau)$	i -ta zmienna stanu w chwili τ
$y_{0,i}, y_{m,i}$	odpowiednio zmienne obserwowane i modelowane dla $i = 0, \dots, n$, gdzie n jest liczbą zmiennych modelowanych
\bar{y}_0	średnia wartość obserwowana
\bar{y}_m	średnia wartość modelowana

Oznaczenia i skróty:

MCMC	algorytm Monte Carlo wykorzystujące łańcuchy Markowa (Markov Chain Monte Carlo)
LN	rozkład log-normalny
LGA	rozkład logarytmiczny gamma – log-Pearson typu III
$p(\boldsymbol{\theta} \mathbf{Y}, H_j)$	prawdopodobieństwo <i>a posteriori</i> hipotez H_j uwarunkowanych zbiorem zmiennych prognozowanych \mathbf{Y}
$p(\mathbf{Y} H_j)$	wiarygodność globalna porównująca różne hipotezy H_j z zaobserwowanym zbiorem parametrów $\boldsymbol{\theta}$
$p(H_j)$	prawdopodobieństwo <i>a priori</i> wszystkich możliwych hipotez H_j
$p(\boldsymbol{\theta} H_j)$	prawdopodobieństwo hipotezy <i>a priori</i> wiedzy o zbiorze parametrów $\boldsymbol{\theta}$ uwarunkowane hipotezą H_j
$p(\mathbf{Y} \boldsymbol{\theta}, H_j)$	prawdopodobieństwo otrzymania danych \mathbf{Y} pod warunkiem słuszności hipotezy H_j , tzw. Bayesowska funkcja wiarygodności
$p(\boldsymbol{\theta} \mathbf{Y})$	rozkład <i>a posteriori</i> , czyli rozkład parametrów $\boldsymbol{\theta}$ odzwierciedlający końcową wiedzę badacza o parametrach modelu
H_{j_0}	przestrzeń hipotez dotycząca wyłącznie parametrów modelu H_j dla $j = j_0$

$p(\mathbf{Y}, \boldsymbol{\theta})$	model Bayesowski, łączny rozkład wektora obserwacji \mathbf{Y} i wektora parametrów $\boldsymbol{\theta}$
$p(\mathbf{Y})$	brzegowy rozkład prawdopodobieństwa obserwacji
$p(\mathbf{Y} \boldsymbol{\theta})$	model próbkowy, czyli warunkowy przy ustalonym wektorze parametrów $\boldsymbol{\theta}$ rozkład obserwacji
$p(\boldsymbol{\theta})$	rozkład <i>a priori</i> , czyli rozkład parametrów przed wglądem w dane
$F(\boldsymbol{\theta})$	dystrybuanta próbkowanego rozkładu parametrów modelu regresji
$p(\mathbf{Y} \boldsymbol{\beta}, \sigma^2, \mathbf{X})$	model próbkowy opisujący zależność między obserwacjami i parametrami modelu liniowej regresji
$N(\mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{E})$	wielowymiarowy rozkład normalny ze średnią $\mathbf{X} \boldsymbol{\beta}$ i kowariancją $\sigma^2 \mathbf{E}$
$\text{Cov}(\dots)$	kowariancja
$\mathbf{I}(\boldsymbol{\beta})$	informacja Fishera w modelu wielokrotnej regresji
$p(\boldsymbol{\theta}) \propto \mathbf{I}(\boldsymbol{\theta}) ^{0.5}$	reguła Jeffreysa
$p(\boldsymbol{\beta}, \sigma^2 \mathbf{Y})$	rozkład <i>a posteriori</i> w modelu liniowej regresji
$N(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\hat{\boldsymbol{\beta}}}\sigma^2)$	wielowymiarowy rozkład normalny ze średnią $\hat{\boldsymbol{\beta}}$ i kowariancją $\mathbf{V}_{\hat{\boldsymbol{\beta}}}\sigma^2$
MNK	metoda estymacji nieznanymi parametrów najmniejszych kwadratów
$(\chi^2)^{-1}$	odwrotny rozkład χ^2
γ^{-1}	odwrotny rozkład gamma
$p(\sigma^2 \mathbf{Y})$	brzegowy rozkład <i>a posteriori</i> σ^2 uwarunkowany wektorem zmiennych prognozowanych \mathbf{Y}
$p(\boldsymbol{\beta} \mathbf{Y})$	brzegowy rozkład <i>a posteriori</i> wektora współczynników kierunkowych $\boldsymbol{\beta}$ uwarunkowany wektorem \mathbf{Y}
$T_{n-d}(\hat{\boldsymbol{\beta}}, \mathbf{s}^2)$	wielowymiarowy rozkład <i>t</i> -Studenta
$E(\beta_i \mathbf{Y}, \mathbf{X})$	wartość oczekiwana brzegowych rozkładów <i>a posteriori</i> poszczególnych elementów wektora $\boldsymbol{\beta}$
$\text{Var}(\beta_i \mathbf{Y}, \mathbf{X})$	wariancja brzegowych rozkładów <i>a posteriori</i> poszczególnych elementów wektora $\boldsymbol{\beta}$
$D(\beta_i \mathbf{Y})$	odchylenie standardowe brzegowych rozkładów <i>a posteriori</i> poszczególnych elementów wektora $\boldsymbol{\beta}$
$E(\mathbf{Y}_0 \mathbf{Y})$	wartość oczekiwana rozkładu \mathbf{Y}_0
$\text{Var}(\mathbf{Y}_0 \sigma^2, \mathbf{Y})$	wariancja prognozy brzegowego rozkładu prawdopodobieństwa \mathbf{Y}_0 <i>a priori</i>
$p(\mathbf{Y}_0 \mathbf{Y})$	brzegowy rozkład <i>a posteriori</i> wektora \mathbf{Y}_0 uwarunkowany wektorem danych zaobserwowanych \mathbf{Y}
$T_{n-d}(\mathbf{X}_0 \hat{\boldsymbol{\beta}}, (\mathbf{E} + \mathbf{X}_0^T + 1 \mathbf{V}_{\hat{\boldsymbol{\beta}}}) \hat{\sigma}^2)$	wielowymiarowy rozkład <i>t</i> -Studenta
$\mathbf{Y} = f(\boldsymbol{\beta}, \mathbf{X}) + \boldsymbol{\varepsilon}$	równanie ogólnej postaci nieliniowego modelu
$\log(f_p)$	logarytm funkcji rozkładu <i>a posteriori</i>
$\log(f_0)$	logarytm funkcji wiarygodności
$\log(f_i)$	logarytm funkcji rozkładu <i>a priori</i>
$E(\boldsymbol{\varepsilon})$	wartość oczekiwana w wielowymiarowym modelu wielokrotnej regresji WWLR
$\text{Cov}(\boldsymbol{\varepsilon}_{(p)}, \boldsymbol{\varepsilon}_{(q)})$	kowariancja modelu WWLR, gdzie p i $q = 1, \dots, l$
$p(\mathbf{Y} \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X})$	model próbkowy w wielowymiarowym modelu wielokrotnej regresji WWLR
$\boldsymbol{\Sigma} \otimes \mathbf{E}$	iloczyn Kroneckera, czyli macierz blokowa o wymiarach $ln \times ln$
$N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} \otimes \mathbf{E})$	macierzowy rozkład normalny zmiennych prognozowanych \mathbf{Y} uwarunkowany parametrami $\boldsymbol{\beta}$ i $\boldsymbol{\Sigma}$
$\mathbf{I}(\boldsymbol{\beta})$	informacja Fishera w wielowymiarowym modelu wielokrotnej regresji
$p(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mathbf{Y})$	rozkład <i>a posteriori</i> w wielowymiarowym modelu wielokrotnej regresji
$p(\boldsymbol{\beta} \boldsymbol{\Sigma}, \mathbf{Y})$	brzegowy rozkład <i>a posteriori</i> dla macierzy współczynników kierunkowych $\boldsymbol{\beta}$
$p(\boldsymbol{\Sigma} \mathbf{Y})$	brzegowy rozkład <i>a posteriori</i> dla macierzy wariancji błędów
$p(\boldsymbol{\Sigma} \mathbf{Y}) \sim \text{IW}(\mathbf{S}^{-1}, n)$	rozkład <i>a posteriori</i> macierzy wariancji podlegający odwrotnemu rozkładowi Wisharta
$-\log(L(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mathbf{Y}))$	ujemny logarytm funkcji wiarygodności w wielowymiarowym modelu wielokrotnej regresji
$\text{tr}(\cdot)$	śląd macierzy

$F_{l,n-d-l}(\alpha)$	rozkład prawdopodobieństwa F -Snedecora o l i $n-d-l$ stopniach swobody dla przyjętego w obliczeniach poziomu istotności α
ANN	sztuczne sieci neuronowe (artificial neuron networks)
MLP	perceptron wielowarstwowy (multilayer perceptron)
FFBPNN	sieć neuronowa wstecznej propagacji błędów (feed forward backpropagation neural network)
LM	algorytm wstecznej propagacji błędów (Levenberg-Marquardt)
MSE	średniokwadratowy błąd
f_0	funkcja aktywacji dla neuronu wyjściowego
f_n	funkcja aktywacji dla neuronu ukrytego
$N(\mathbf{0}, V_\tau \Sigma)$	wielowymiarowy rozkład normalny o wartości oczekiwanej $\mathbf{0}$ i macierzy kowariancji Σ z uwzględnieniem czynnika skali wariancji V_τ
$N(\mathbf{0}, \mathbf{W}_\tau \otimes \Sigma)$	$d \times l$ macierzowy rozkład normalny o macierzy wartości oczekiwanej $\mathbf{0}$ oraz lewej i prawej macierzy wariancji kowariancji \mathbf{W}_τ i Σ
$NIW_{n_{\tau-1}}(\mathbf{m}_{\tau-1}, \mathbf{C}_{\tau-1}, \mathbf{S}_{\tau-1})$	macierzowy rozkład normalny/odwrotny rozkład Wisharta macierzy $\boldsymbol{\theta}_{\tau-1}$, $(\boldsymbol{\theta}_{\tau-1}, \Sigma D_{\tau-1})$
$N(\mathbf{f}_\tau, Q_\tau \Sigma)$	wielowymiarowy rozkład normalny prognozy wektora \mathbf{Y}_τ , $(\mathbf{Y}_\tau \Sigma, D_{\tau-1})$
$T_{n_{\tau-1}}(\mathbf{f}_\tau, Q_\tau \mathbf{S}_{\tau-1})$	wielowymiarowy rozkład t -Studenta brzegowego rozkładu, $(\mathbf{Y}_\tau D_{\tau-1})$
$p(\mathbf{Y}_{\tau+k} D_\tau)$	k -rozkładów brzegowych naprzód do prognozowania w chwili τ
$T_{n_\tau}(\mathbf{f}_\tau(k), \boldsymbol{\theta}_\tau(k) \mathbf{S}_\tau)$	wielowymiarowy rozkład t -Studenta prognozy wektora $\mathbf{Y}_{\tau+k}$, $(\mathbf{Y}_{\tau+k} D_\tau)$
$N(\mathbf{m}, \mathbf{C} \otimes \Sigma)$	macierzowy rozkład normalny nieobserwowanej macierzy zmiennych stanu $\boldsymbol{\theta}$, $(\boldsymbol{\theta} \Sigma)$
IW_τ	odwrotny rozkład Wisharta
$N(\mathbf{m}_{\tau-1}, \mathbf{C}_{\tau-1} \otimes \mathbf{S}_{\tau-1})$	macierzowy rozkład normalny macierzy nieobserwowanych współczynników kierunkowych $\boldsymbol{\theta}_\tau$ na podstawie informacji $D_{\tau-1}$, $(\boldsymbol{\theta}_{\tau-1}, D_{\tau-1} \Sigma = \mathbf{S}_\tau)$
$f_{\mu, \Sigma}(\mathbf{X})$	funkcja gęstości n wymiarowego rozkładu normalnego
\otimes	iloczyn Kroneckera
$f_{\mu, \Sigma, \nu}(\mathbf{X})$	funkcja gęstości n wymiarowego rozkładu t -Studenta o ν stopniach swobody
MNR	macierzowy liniowy model dynamiczny (Matrix Normal Regression)
MDWR	macierzowy liniowy model dynamiczny (Multivariate Discount Weighted Regression)
F_{MOD}	dystrybuanta zmiennej losowej modelowanej X_{MOD}
F_{OBS}^{-1}	funkcja odwrotna dystrybuanty (lub funkcji kwantylowej) zmiennej losowej obserwowanej X_{OBS}
WLRR	wielokrotna liniowa regresja z randomizacją
B-WE	transformacja dystrybuanty typu Bernoulliego-Weibulla
B-LN	transformacja dystrybuanty typu Bernoulliego-lognormalny
B-E	transformacja dystrybuanty typu Bernoulliego-wykładniczy
MNW	estymacji nieznanymi parametrami metodą największej wiarygodności
GCM	Model Ogólnej Cyrkulacji (Global Climate Model)
F_{MOD}	dystrybuanta zmiennej losowej obserwowanej P_{OBS}
Q-Q	wykres zależności Kwanty teoretyczny-Kwantyl empiryczny
DA	metoda asymilacji danych (data assimilation)
ASCAT nudging	zaawansowana skaterometria atmosfery (Advanced SCATTERometer)
EKF	metoda asymilacji danych (Newtonian relaxation)
SWAT	rozszerzony liniowy filtr Kalmana
PF	model transportu materiału zawieszzonego (Soil and Water Assessment Tool)
REnKF	filtr cząstkowy (Particle Filter)
K	rekurencyjny wiązkowy filtr Kalman (recursive ensemble Kalman filter)
$N(0, \sigma_0^2)$	zysk Kalmana
a, b	rozkład normalny o średniej równej zero i wariancji σ_0^2
$g(f)_{(a, b)}$	ograniczenia przedziałowe obustronnie uciętego rozkładu normalnego
U	funkcja gęstości obustronnie uciętego rozkładu normalnego
	rozkład jednostajny

ξ_{θ}	błąd obserwacji wilgotności gleby θ (mm)
ξ_P	błąd obserwacji opadu P (mm)
ξ_{UR}	błąd obserwacji górnego ograniczenia zbiornika wody gruntowej UR (mm)
ξ_{LR}	błąd obserwacji dolnego ograniczenia zbiornika wody gruntowej LR (mm)
F_{θ}	dystrybuanta niezaburzonej zmiennej stanu wilgotności gleby
$F_{\theta'_1}, \dots, F_{\theta'_n}$	dystrybuanty zaburzonych stanów wilgotności gleby n elementowej wiązki
F_{θ}^{-1}	funkcja odwrotna do dystrybuanty niezaburzonej zmiennej stanu
SAC-SMA	model hydrologiczny (Sacramento Soil Moisture Accounting Model)
VAR	filtr wariacyjny (variational filter)
KF	filtr Kalmana
\tilde{K}	zmodyfikowany zysk Kalmana
J_{τ}	minimalizowana funkcja celu w filtrze wariacyjnym 4DVAR
$J_f(\mathbf{X}_{f-w}^b, \mathbf{X}_P, \mathbf{X}_{PE})$	funkcja celu w momencie stawiania prognozy hydrologicznej f
$M_{\tau_i, \tau}$	nieliniowy dynamiczny model, który oblicza $m \times w$ wymiarową macierz stanu systemu \mathbf{X}^b od momentu τ_i do τ
J_f	minimalizowana funkcja celu w filtrze wariacyjnym 4DVAR w chwili f
σ_Q	odchylenie standardowe błędu obserwacji przepływu rzecznoego (m^3s^{-1})
$\sigma_{\theta}, \sigma_P, \sigma_{PE}$	odpowiednio odchylenia standardowe błędu obserwacji <i>in-situ</i> satelitarnej wilgotności gleby, pomiaru opadu i ewapotranspiracji (mm)
L	operator funkcji liniowego modelu
MAD	maksymalny błąd absolutny (maximum absolute deviation)
MAE	średni błąd absolutny (mean absolute deviation)
R_0M	stosunek wartości średnich (ratio of the means)
RMSE	pierwiastek kwadratowy średniego błędu kwadratowego (root mean square error)
R^2	współczynnik determinacji (coefficient of determination)
E_{NS}	współczynnik skuteczności Nasha i Sutcliffea
$\bar{K} \in (\delta_{Q_{max}}, \delta_{V_w}, \delta_{T_w})$	kryterium przyjęte dla funkcji celu opisane w 3D przestrzeni akceptacji wyznaczonej sumą błędów dla zmiennych Q_{max}, V_w i T_w
δ_j	suma błędów dla j -tej zmiennej prognozowanej, gdzie $j = 1, \dots, l$
$\omega = \prod_{j=1}^l \delta_j$	błąd modelu l -wymiarowej zmiennej prognozowanej

Rozdział 4

Zmienne i stałe:

$\mathbf{X}_{OBS}(\tau)$	wektor zaobserwowanych zmiennych prognozujących w chwili stawiania prognozy τ
$\mathbf{X}_{PROG}(\tau)$	wektor prognozowanych zmiennych prognozujących w chwili stawiania prognozy τ

Oznaczenia i skróty:

U_0	błąd średni prognozy
R_0	współczynnik korelacji wielokrotnej
$p(\beta_{ES_5} Y), p(\beta_{Q_1} Y)$	brzegowe rozkłady <i>a posteriori</i> zmiennych losowych: ES_5, Q_1 , i AT_5
$p(\beta_{AT_5} Y)$	
MLR	multiple linear regression (wielokrotna liniowa regresja, WLR)
MMLR	multivariate multiple linear regression (wielowymiarowy model wielokrotnej liniowej regresji, WWLR)
MLR-B, MMLR-B	modele regresyjne MLR i MMLR z parametrami estymowanymi metodą Bayesowską, B
MLR-InvB, MMLR-InvB	modele regresyjne MLR i MMLR z parametrami estymowanymi techniką odwrotnego modelowania, InvB

σ_w

warunkowe odchylenie standardowe modeli regresyjnych

Rozdział 5

Zmienne i stałe:

d_l, d_g

odpowiednio dolna i górna granica przedziału wyników statystyki Durбина-Watsona

$Q_{\tau+2}^{WIZ}$

prognozowany przepływ w zlewni Górnej Narwi w Wiźnie z wyprzedzeniem 2-dobowym (m^3s^{-1})

$Q_{DWR+1}, Q_{VCR+1},$

Q_{WLR+1}, Q_{NN+1}

prognozowany dobowy hydrogram w zlewni Górnej Narwi w przekroju wodowskazowym WIZNA z wyprzedzeniem 1-dobowym za pomocą modeli: DWR, VCR, WLR i NN

$Q_{DWR+2}, Q_{VCR+2},$

Q_{WLR+2}, Q_{NN+2}

prognozowany dobowy hydrogram w zlewni Górnej Narwi w przekroju wodowskazowym WIZNA z wyprzedzeniem 2-dobowym za pomocą modeli: DWR, VCR, WLR i NN

$Q_{DWR+3}, Q_{VCR+3},$

Q_{WLR+3}, Q_{NN+3}

prognozowany dobowy hydrogram w zlewni Górnej Narwi w przekroju wodowskazowym WIZNA z wyprzedzeniem 3-dobowym za pomocą modeli: DWR, VCR, WLR i NN

Oznaczenia i skróty:

I-H-O

struktura modelu NN, gdzie I – liczba neuronów wejściowych, H – liczba ukrytych neuronów w ukrytej warstwie, O – liczba zmiennych prognozowanych w warstwie wyjściowej

diag(.)

macierz diagonalna (zazwyczaj kwadratowa), której wszystkie współczynniki leżące poza główną przekątną są zerowe

WAV-DLMB

model hybrydowy łączący analizę falkową z dynamicznym liniowym modelem regresji

Rozdział 6

Zmienne i stałe:

P_{0524}^{BC}

opad satelitarny P_{0524} z usuniętym obciążeniem metodą BC

$H_{14}^{1,00-BC}, H_{14}^{2,89-BC}$

satelitarna wilgotność gleby $H_{14}^{1,00}, H_{14}^{2,89}$ z usuniętym obciążeniem metodą BC
średni opad, temperatura powietrza i przepływ rzeczny z naziemnej sieci pomiarowo-obszaryjnej

$P_{OBS}, T_{OBS}, Q_{OBS}$

P_{SAT}

reprezentuje satelitarne opady: P_{03}, P_{04}, P_{0524}

H_{14}

reprezentuje satelitarne obserwacje wilgotności gleby: $H_{14}^{1,00}, H_{14}^{2,89}$

Q_{HBV}

przepływ rzeczny obliczony przez model HBV z wejściem w postaci zaobserwowanego opadu P_{OBS} z naziemnej sieci pomiarowo-obszaryjnej IMGW-PIB

Q_{HBV-BC}

przepływ rzeczny obliczony przez model HBV z wejściem w postaci skorygowanego opadu satelitarnego bez asymilacji danych

$Q_{HBV-BC-BC}$

przepływ rzeczny obliczony przez model HBV z wejściem w postaci skorygowanego opadu satelitarnego i wilgotności gleby (bez asymilacji wilgotności gleby)

$Q_{HBV-BC-EnKF},$

$Q_{HBV-BC-EnKF-BC},$

$Q_{HBV-BC-EnSRF-BC}$

$Q_{\text{HBV-BC-EnSRF/4DVAR}}$	przepływy rzeczne obliczone przez model HBV ze skorygowanym opadem satelitarnym i asymilacją wilgotności gleby odpowiednio metodami: EnKF, EnKF-BC, EnSRF-BC i EnSRF/4DVAR
$P_{\text{OBS}+1}, \dots, P_{\text{OBS}+72},$ $T_{\text{OBS}+1}, \dots, T_{\text{OBS}+72}$	odpowiednio historyczny opad godzinny P (mm) i temperatura powietrza T ($^{\circ}\text{C}$) w zakresie +1 h, ..., +72 h
$Q_{\text{HBV}+1}, \dots, Q_{\text{HBV}+72}$	hydrogram uśrednionej wiązki przepływów

SPIS UŻYTYCH POJĘĆ I STOSOWANE NAZEWNICTWO

(w kolejności występowania w tekście)

1. **symulacja hydrologiczna** – odwzorowanie przebiegu zjawiska, np. przepływu rzecznego, za pomocą modelu matematycznego przy założeniu znajomości danych wymuszających, np. zaobserwowanego opadu, temperatury powietrza.
2. **hipoteza** – przypuszczenie dotyczące pewnych prawidłowości w świecie. Hipoteza badawcza stawiana jest, by sprawdzić, przetestować nową teorię. W przypadku rozpatrywania przypuszczeń dotyczących mechanizmu losowego generującego obserwacje mówimy o hipotezie statystycznej. Wówczas weryfikacji poddawane jest przypuszczenie badacza na temat rozkładu cechy w populacji. Hipotezy statystyczne dzielą się na parametryczne (dotyczące jednego z parametrów rozkładu cechy) oraz nieparametryczne (dotyczące postaci funkcyjnej rozkładu cechy). Przykładem hipotezy statystycznej jest przypuszczenie dotyczące modeli matematycznych wybieranych do analizy Bayesowskiej wraz ze zbiorami parametrów, które mogą zostać użyte do modelowania danych.
3. **kalibracja modelu** – identyfikacja parametrów modelu połączona z ich oszacowaniem. Kalibracja zapewnia zgodność predyktywną modelu w innych warunkach niż te, dla których model został opracowany.
4. **zmienne stanu** (state variables) – z matematycznego punktu widzenia koncepcja zmiennych stanu umożliwia przekształcenie równania różniczkowego dowolnego rzędu w równoważny zbiór równań rzędu pierwszego. Z fizycznego punktu widzenia zmienne stanu systemu określone wartościami liczbowymi w danym momencie zawierają wszystkie istotne informacje o systemie, niezbędne do przewidywania jego zachowania w przyszłości. Koncepcja zmiennej stanu jest przydatna do analizy systemów dynamicznych. Zależności zmiennych stanu z innymi ważnymi zmiennymi występują w modelu matematycznym w postaci równania stanu. Na przykład w modelach hydrologicznych typu opad–odpływ zmiennymi stanu mogą być: pokrywa śniegu, wilgotność gleby, zbiorniki wody itp. W modelach operacyjnych stan modelu na końcu obliczeń można zapisać, aby umożliwić rozpoczęcie nowego cyklu obliczeń ze stanem modelu z poprzedniego cyklu.
5. **updating** – aktualizacja zmiennych stanu modelu matematycznego, bieżąca korekta prognozy.
6. **obserwacje *in situ*** – obserwacje i pomiary prowadzone w miejscu, którego dotyczą (w odróżnieniu od pomiarów zdalnych, tj. teledetekcyjnych).
7. **asymilacja danych** (data assimilation), DA [rozdział 3.8, str. 94] – przetwarzanie np. meteorologicznych obserwacji i pomiarów do postaci pozwalającej na ich bezpośrednie użycie w matematycznym modelu, np. w modelu hydrologicznym. Dane wejściowe do asymilacji mogą pochodzić z naziemnej sieci pomiarowo-obserwacyjnej, obserwacji satelitarnych, radarowych lub z matematycznych modeli meteorologicznych.
8. **metoda Monte Carlo**, MC [dodatek B, str. 167] – opiera się na numerycznych obliczeniach z wykorzystaniem zmiennych losowych w celu zastąpienia problemu numerycznego zadaniem z dziedziny prawdopodobieństwa o takim samym rozwiązaniu. Wykorzystuje ona liczby losowe o określonym rozkładzie w charakterze próbek opisujących analizowany proces.
9. **algorytm genetyczny** – algorytm poszukiwania oparty na mechanizmach doboru naturalnego oraz łączenia cech rozwiązań. Algorytm stosowany jest w celu wyjaśnienia procesów adaptacyjnych występujących w przyrodzie oraz w zadaniach optymalizacji i uczenia.
10. **dane gridowe** – w meteorologii utożsamiane są z wartościami elementów meteorologicznych występujących w punktach węzłowych regularnej siatki przestrzennej o określonej rozdzielczości.
11. **konceptualny model hydrologiczny HBV** [rozdział 3.6, str. 90] – szwedzki model konceptualny typu opad–odpływ (szw. Hydrologiska Byråns Vattenbalansavdelning) generujący ciągły hydrogram odpływu. Do opisu procesów hydrologicznych stosowane są równania matematyczne oraz zależności empiryczne, a parametry modelu identyfikowane są metodą Monte Carlo lub metodą algorytmu genetycznego. Standardowym wejściem do modelu mogą być punktowe lub gridowe dane meteorologiczne opadu i temperatury powietrza.
12. **preprocesor danych wejściowych do modelu HBV** [rozdział 6.1, str. 134] – zbiór procedur przetwarzających dane meteorologiczne w celu dostosowania ich do wejścia do modelu hydrologicznego typu opad–odpływ HBV. W skład preprocesora danych wejściowych wchodzi metody usu-

wania błędu systematycznego (obciążenia) z satelitarnych obserwacji meteorologicznych oraz metody asymilujące satelitarne obserwacje wilgotności gleby. Obie procedury są zintegrowane z informatycznym środowiskiem modelu HBV.

13. **rozkład *a priori*** [rozdział 3.3, str. 71] – rozkład prawdopodobieństwa przypisany zdarzeniu przed wglądem w dane zaobserwowane, często subiektywny przy założeniu zasady obojętności.
14. **rozkład *a posteriori*** [rozdział 3.3, str. 71] – rozkład prawdopodobieństwa przypisany do zdarzenia na podstawie jego obserwowanej częstotliwości w próbie losowej i wyznaczony z rozkładu *a priori* zgodnie z teorią Bayesowską.
15. **szum Gaussa** – utożsamiany z szumem białym, dowolny stacjonarny proces stochastyczny o gaussovskim rozkładzie prawdopodobieństwa o wartości średniej $\mu = 0$ i skończonej wariancji σ^2 .
16. **filtr Kalmana** [rozdział 3.8, str. 94] – algorytm wyznaczony układem matematycznych równań do wydajnego rekursywnego estymowania stanu modelu, co minimalizuje błąd średniokwadratowy. Filtr oszacowuje przyszły stan modelu na podstawie znanego stanu poprzedniego oraz aktualnej obserwacji. Przyjmuje się, że proces estymacji obarczony jest błędem o rozkładzie Gaussowskim, tj. białym szumem.
17. **tło stanu modelu** – zbiór nieobserwowanych zmiennych stanu, obliczanych przez model hydrologiczny.
18. **filtr wariacyjny** [rozdział 3.8.2, str. 100] – stosowany jest do wariacyjnej wymiarowej asymilacji danych i polega na estymacji zbioru parametrów za pomocą optymalizacji dopasowania między modelem (np. hydrologicznym) a zbiorem obserwacji, które model ma prognozować. W filtrze wariacyjnym asymilacja danych związana jest z minimalizacją funkcji celu. Typową funkcję celu tworzy:
 - suma kwadratów odchyleń wartości symulowanych zmiennych stanu od obserwacji ważonych dokładnością obserwowanych zmiennych stanu,
 - suma kwadratów odchyleń tła stanu modelu w chwili uruchomienia metody od uaktualnionego tła stanu modelu ważonego macierzą błędów modelu skojarzonego z tłem stanu modelu.
19. **obserwacje *pseudo in situ*** – obserwacje i pomiary teledetekcyjne (np. wilgotności gleby) zastosowane w procesie asymilacji danych za pomocą filtrów Kalmana.
20. **metoda transformaty dystrybuanty** [rozdział 3.7.1, str. 92] – jest to metoda odwracania dystrybuanty. Jeżeli składowe wektora losowego U i X nie są skorelowane, to można je transformować niezależnie wykorzystując następującą tożsamość: $F_U(u) = F_X(x)$; czyli transformację zmiennej losowej X na zmienną U można zapisać jako $u = F_U^{-1}(F_X^{-1}(x))$, gdzie F_U jest dystrybuantą rozkładu prawdopodobieństwa zmiennej losowej U , a F_X jest dystrybuantą rozkładu prawdopodobieństwa zmiennej losowej X .
21. **metoda postprocesowa korekcji obciążenia** (bias correction), BC [rozdział 3.7.1, str. 92] – metoda usuwania błędu systematycznego (obciążenia) z zaburzonej wiązki zmiennych stanu w wiązkowym filtrze Kalmana. Algorytm wykonywany jest metodą transformaty dystrybuanty. Jest to metoda postprocesowa, czyli uruchamiana po zakończeniu przetwarzania danych przez model hydrologiczny.
22. **prognoza wiązkowa** – zbiór (wiązka) prognozowanych hydrogramów obliczonych za pomocą symulacyjnego modelu hydrologicznego uruchomionego w celu otrzymania prognozy. Wiązka hydrogramów jest obliczona na podstawie, uaktualnionej w ostatnim kroku, updatingu, macierzy zmiennych stanu.
23. **prognoza w skali błędu** [rozdział 6.2.3, str. 138] – prognoza odzwierciedlająca zmienność danych prognozowanych i wielkości błędów (**niepewność**). Błąd prognozy pochodzi z tej samej skali danych, zatem prognozy różnych szeregów można dokonać tylko wtedy, gdy szeregi są na tej samej skali.
24. **analiza statystyczna** [rozdział 2.2, str. 32] – obejmuje statystykę opisową (np. metody obserwacji statystycznej, sposoby opracowywania i prezentacji materiału statystycznego) oraz statystykę matematyczną (np. metody wnioskowania, tj. estymacji i weryfikacji hipotez, o całej zbiorowości generalnej na podstawie badania wybranej w sposób losowy pewnej jej części zwanej próbą losową).
25. **analiza niejednorodności** [rozdział 2.2, str. 32] – analiza zmienności zbioru czynników warunkujących wystąpienie i przebieg zjawiska oraz zmiany warunków przeprowadzania eksperymentu, tj. zmiany metody i przyrządu pomiarowego, miejsca i czasu wykonywania pomiaru, częstości próbkowania itp. Analiza wykonywana jest metodami genetycznymi i statystycznymi.

26. **łańcuch Markowa** [dodatek B, str. 167] – proces będący ciągiem zmiennych losowych X_0, X_1, \dots określonych na wspólnej przestrzeni probabilistycznej i spełniający warunek:

$$\bigwedge_n \bigwedge_{i_0, \dots, i_{n-1}, j \in \{0, 1, 2, \dots\}} P(X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) = P(X_n = j | X_{n-1} = i_{n-1}).$$
Dla łańcucha Markowa rozkład prawdopodobieństwa warunkowego położeń w n -tym kroku zależy tylko od prawdopodobieństwa warunkowego położenia w kroku poprzednim, a nie od wartości punktów trajektorii.
27. **symulacja Monte Carlo łańcuchami Markowa** (Markov Chain Monte Carlo), MCMC – algorytm próbkowania dowolnego rozkładu prawdopodobieństwa. Rozróżniamy próbkowanie, tzw. sampling, który może być wykonany np.: za pomocą metody odwracania dystrybuanty, metodą eliminacji, superpozycji rozkładów lub algorytm polegający na konstrukcji łańcucha Markowa o rozkładzie równoważnym szukanemu rozkładowi. Powyższa symulacja umożliwia wydajne próbkowanie złożonych rozkładów prawdopodobieństwa.
28. **nieinformatywny rozkład *a priori*** [rozdział 3.3.3, str. 78] – postać rozkładu *a priori*, która opisuje brak wiedzy na temat jego parametrów. Jest to sposób na rozwiązanie subiektywności rozkładu *a priori*.
29. **funkcja wiarygodności** – iloczyn prawdopodobieństwa *a posteriori* dla n dostępnych prób losowych: $L = \prod_{i=1}^n f(x_i, \theta)$, $\ln(L) = \sum_{i=1}^n \ln(f(x_i, \theta))$, gdzie: $f(x_i, \theta)$ jest funkcją gęstości prawdopodobieństwa; x_i są realizacjami zmiennej losowej X dla $i = 1, \dots, n$; θ oznacza parametr lub wektor parametrów; a n jest liczebnością próby losowej.
30. **informacja Fishera** [rozdział 3.3.2, str. 74] – niech $X = (X_1, \dots, X_n)$ jest próbą losową z rozkładu o różniczkowalnej, względem θ , gęstości f_θ , gdzie $\theta \in \Theta \subset \mathbb{R}$, to funkcję $I(\theta) = E_\theta \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right)^2$ nazywamy informacją Fishera o parametrze θ zawartej w próbie losowej X , gdzie wartość oczekiwana $E_\theta[\cdot]$ wzięta została w sensie rozkładu $p(x|\theta)$, tj. klasycznej funkcji wiarygodności.
31. **rozkład aprioryczny Jeffreysa** [rozdział 3.3.2, str. 74] – najpopularniejsza reguła znajdowania nieinformatywnych rozkładów *a priori*. Nieinformatywny rozkład *a priori* parametru θ powinien spełniać warunek $f(\theta) \propto |I(\theta)|^{-\frac{1}{2}}$, gdzie $I(\theta)$ jest informacją Fishera. Zaletą rozkładów *a priori* znajdujących *a priori* zgodnie z regułą Jeffreysa jest to, że rozkłady *a priori* i *a posteriori* są niezmiennicze ze względu na jednoznaczne przekształcenia parametrów.
32. **próbkowanie Gibbsa** [dodatek B, str. 167] – algorytm oparty o metodę łańcucha Markowa, który do generowania danych wykorzystuje zasadę symulacji Monte Carlo. Umożliwia generowanie danych pochodzących z rozkładu wielowymiarowego bezpośrednio lub za pomocą generowania danych z rozkładów warunkowych poszczególnych zmiennych brzegowych.
33. **metoda Metropolis-Hastingsa** [dodatek B, str. 167] – najważniejszy algorytm symulacji Monte Carlo łańcuchami Markowa. Przy założeniu znajomości generowania łańcucha Markowa, algorytm modyfikuje taki łańcuch, wprowadzając specjalnie dobraną regułę akceptacji w taki sposób, aby wymusić zbieżność do zadanego rozkładu.
34. **entropia krzyżowa** (cross-entropy) – w teorii informacji jest to sposób porównywania modeli statystycznych. Entropia krzyżowa między dwoma rozkładami prawdopodobieństwa p i q tego samego zdarzenia mierzy średnią liczbę bitów informacji potrzebnych do zidentyfikowania zdarzenia wylosowanego ze zbioru, jeżeli jest użyty schemat kodowania, który jest zoptymalizowany dla sztucznego rozkładu prawdopodobieństwa q , zamiast prawdziwego prawdopodobieństwa p .
35. **hiperparametry modelu** – parametry, od których zależą rozkłady aprioryczne, tj. rozkłady przyjęte z góry, nieoparte na doświadczeniu (obserwacjach i pomiarach).
36. **regularyzacja we wnioskowaniu Bayesowskim** – w przypadku wielowymiarowych zbiorów danych zastosowanie kryterium największej wiarygodności (maximum likelihood) do estymacji parametrów modelu prowadzi do obciążenia i dużej wariancji parametrów. W podejściu Bayesowskim stosowane jest penalizowane prawdopodobieństwo (penalized likelihood), które wprowadza człon kary w formie wiedzy *a priori* na temat rozkładów parametrów (np. średnia zero i odchylenie standardowe równe 1) lub też szacowane jest bezpośrednio z danych (poprzez hiperparametry, odchylenie standardowe szacowane jest bezpośrednio z danych). Oba rozwiązania prowadzą do powstania hierarchicznej struktury modelu.

37. **kwantyl zmiennej losowej** – kwantyl rzędu p rozkładu prawdopodobieństwa to taka liczba x_p , że $P(X \leq x_p) \geq p$ i $P(X \geq x_p) \geq 1 - p$.
38. **Bayesowski kryterium informacyjne Schwarza** (Bayesian Information Criterion), BIC wskaźnik dopasowania modelu, stanowiący metodę porównywania modeli dla zmiennej zależnej wyrażony za pomocą wzoru: $BIC = -2 \ln L(\hat{\theta}) + K \ln(n)$, gdzie: $\hat{\theta}$ jest estymatorem parametru lub wektora parametrów; $L(\hat{\theta})$ oznacza funkcję wiarygodności dla danego modelu; K liczbę parametrów oraz n – liczbę obserwacji.
39. **homoscedastyczność** – stała wariancja reszt, tj. stała różnica między wartością zaobserwowaną a modelowaną.
40. **współczynnik dyskontowy** [rozdział 3.5.2, str. 87] – czynnik stosowany w ekonomii. Jest pewną wielkością liczbową, która pozwala na określenie wartości obecnej przepływu pieniężnego. Jeżeli spodziewamy się pewnego przepływu pieniężnego w przyszłości, to znając jej obecną wartość i mnożąc ją przez odpowiedni współczynnik dyskontowy otrzymamy bieżącą wartość tego przepływu pieniężnego w przeszłości. Współczynnik dyskontowy stosowany jest także w dynamicznych liniowych modelach Bayesowskich i może odnosić się do przepływu rzecznoego. Różnica między wartością przepływu w przyszłości a wartością obecnie zaobserwowaną wynika ze zmienności przepływu rzecznoego w czasie (np. zmienności w cyklach rocznych lub sezonowych).
41. **stan początkowy** (initial state) – początkowy stan modelu, od którego rozpoczynają się obliczenia; zwykle jest to stan końcowy z poprzedniego uruchomienia modelu. Z reguły możliwe jest utworzenie standardowego stanu początkowego dla pierwszego uruchomienia dla nowej modelowanej zlewni.
42. **walidacja modelu** – ocena przydatności użytkowej modelu. Polega na porównaniu wyników obserwacji (np. zaobserwowanego przepływu rzecznoego) z wynikami symulacji.
43. **black box** – model matematyczny opisujący układ, o którym nie posiada się absolutnie żadnej informacji. W tego typu modelach należy wyznaczyć funkcję wiążącą wielkości występujące w układzie, jak i wartości liczbowe parametrów tych funkcji. W tym celu wykorzystywane są obserwacje i wielkości wejściowe i wyjściowe modelowanego układu.
44. **analiza falkowa** (wavelet analysis) – analiza zmiany częstotliwości zdarzenia, np. przepływu rzecznoego, w dziedzinie czasu. Analiza przeprowadzana jest w oparciu o transformatę falkową pozwalającą na przejście z układu czas–wartość (obserwacyjne ciągi przepływów) do układu czas–skala (częstotliwość).
45. **metoda hindcast** – to sposób testowania modelu matematycznego. Znane lub dokładnie oszacowane dane wejściowe (wymuszenie modelu) dla przeszłych zdarzeń są wprowadzane do modelu, aby zobaczyć, jak dobrze dane wyjściowe odpowiadają znanym wynikom. Prognoza metodą hindcast (tzw. prognoza wsteczna) nazywana jest również ponowną prognozą historyczną i używana się jej do sprawdzenia, czy zaktualizowany model faktycznie poprawia prognozy. Aktualizacją modelu mogą być niedostępne wcześniej dane wejściowe lub dane asymilowane w ramach preprocesora danych wejściowych. Prognozę metodą hindcast używa się również do badania przypadków ekstremalnych.
46. **błąd prognozy ex post** [rozdział 6.2.3, str. 138] – błąd, który jest obliczany po zapoznaniu się z rzeczywistymi wartościami zmiennej/zmiennych prognozowanych.
47. **błąd prognozy ex ante** [rozdział 6.2.3, str. 138] – błąd, który jest obliczany w chwili stawiania prognozy, czyli bez poznania rzeczywistych wartości zmiennej/zmiennych prognozowanych. Wielkość błędu wynika wyłącznie z informacji dostępnych w czasie prognozy, tj. z informacji dotyczących oczekiwanych średnich odchyleń realizacji prognozowanej zmiennej losowej.

1. WPROWADZENIE

Jednym z istotnych zagrożeń hydrologiczno-meteorologicznych jest występowanie powodzi, czyli wysokiego stanu wody w rzekach i jeziorach, które przewyższają pewną przeciętną miarę wielkich wód powodujących w mniejszym lub w większym stopniu straty powodziowe [Lambor 1962]. Tego rodzaju zdarzenia występują zarówno w obszarach górskich dorzeczy Odry i Wisły, jak i na nizinach i rzekach Przymorza. W ostatnich latach coraz częściej obserwowane są powodzie o gwałtownych przebiegach, których wielkość i zasięg oddziaływania są trudne do prognozowania. W hydrologii spotykamy się z wieloma modelami pozwalającymi oszacować wielkość odpływu zlewniowego – są to głównie modele deterministyczne, dostarczające punktowych wyników o przebiegu procesu bazujące na zależności opad–odpływ oraz uwzględniające efekt procesu transformacji fali wezbraniowej w korycie rzeczonym. Podstawą opracowania takich modeli są fizyczne analizy procesów hydrologicznych (np. ewapotranspiracji, infiltracji, spływu powierzchniowego, podpowierzchniowego, gruntowego itp.). Stosowane są też modele stochastyczne, czyli modele deterministyczne dostarczające wyniki z podaniem zmienności losowej, tj. określeniem prawdopodobieństwa, oraz modele konceptualne opisujące strukturę i działanie systemu zlewniowego w uniwersalnym modelu niezależnym od modelu implementacyjnego.

Ben [1988] zaproponował, aby metody probabilistyczne postrzegać jako metody komplementarne do modeli deterministycznych budowanych na podstawach fizyki. W modelowaniu hydrologicznym ważne jest łączenie symulacji deterministycznej [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 1] z niepewnością, zdefiniowaną jako odchylenie in plus albo in minus od oczekiwanego stanu lub nieoczekiwanego zdarzenia w przyszłości, które można określić za pomocą rachunku prawdopodobieństwa. W związku z powyższym, celem niniejszej pracy jest opracowanie zbioru metod probabilistycznych uwzględniających modelowanie niepewności. Zaproponowane rozwiązania mają stanowić alternatywę lub uzupełnienie istniejących modeli deterministycznych przeznaczonych do prognozowania hydrologicznego.

W hydrologii metody probabilistyczne służą modelowaniu procesów, które mają skojarzone z nimi komponenty losowe. Dziedzina ta proponuje metody (modele) matematyczne odpowiednie do rozwiązywanych problemów, sposoby szacowania ich parametrów oraz techniki symulacji procesów, a także umożliwia za pomocą tych metod wykonywanie prognoz hydrologicznych. Jednym z najważniejszych problemów, które mogą pomóc rozwiązać metody probabilistyczne to oszacowanie niepewności. Źródłem niepewności jest nie tylko losowość zdarzenia, ale także fakt wystąpienia wielu zdarzeń jednostkowych, o czym decydują wiele czynników. Tylko część z tych czynników jest znana, część stanowi grupę czynników niepowtarzalnych (tzn. nieznanych). Niepewność z reguły wynika z szacowania właściwości losowych badanej populacji na podstawie niedokładnych danych lub próby losowej o niedostatecznej liczebności. Są to główne przyczyny błędów i strat informacyjnych. W konsekwencji powstają różnice między badaną populacją zmiennych losowych a własnościami losowymi takiej próby. Szczególnie jest to ważne w prognozowaniu hydrologicznym w momencie stawiania prognozy. Różnica między hydrogramem symulowanym a obserwowanym jest podstawową przyczyną błędnej prognozy. Różnice między wynikami modelu hydrologicznego a obserwacjami (tzw. szumy) mogą wynikać z następujących przyczyn:

- 1) Zmienności zjawisk przyrody.
- 2) Przyjmowania założeń i hipotez [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 2] dotyczących nierozpoznanych w pełni powiązań przyczynowo-skutkowych zjawiska.
- 3) Wpływu czynników antropogenicznych zmieniających środowisko naturalne zlewni, tym samym zmieniających badaną populację generalną.
- 4) Niejednorodności i niedoskonałości danych wejściowych do modelu i parametrów systemów hydrologicznych. Parametry systemów hydrologicznych są przeważnie zmienne w przestrzeni i często w czasie oraz reprezentowane przez pojedyncze wartości dla całej zlewni, a także uzyskiwane z interpolacji z kilku miejsc, w których takie pomiary są prowadzone.
- 5) Niezgodności skali wyjścia modeli matematycznych w porównaniu do pomiarów i obserwacji, które w rzeczywistości reprezentują wartości średnie dla bardzo małych objętości, praktycznie punktów.
- 6) Dyskretyzacji ciągłych zmiennych losowych na ich dyskretne odpowiedniki.
- 7) Błędów obserwacji (np. zmiany temperatury i ciśnienia atmosferycznego podczas obserwacji poziomu zwierciadła wody).

- 8) Losowości pomiaru, błędów obliczeniowych obserwacji.
- 9) Błędów systematycznych, tzw. obciążenia (bias) przyrządów pomiarowych i niewłaściwej kalibracji modelu [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 3].
- 10) Błędów warunków brzegowych, warunków początkowych i danych wejściowych do modeli hydrologicznych (np. początkowe wartości zmiennych stanu modelu [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 4] nie zawsze są dokładnie znane, zatem oszacowywane są z błędem).
- 11) Błędów modelu lub systemu, które są uproszczonymi wersjami rzeczywistości i tym samym nie mogą zawierać wszystkich skomplikowanych mechanizmów i interakcji, które występują w naturalnych systemach przyrodniczych. Z reguły modelujemy cząstkowe, elementarne procesy występujące w złożonych systemach hydrologicznych w sposób przybliżony – tym samym fizyka modelu różni się od fizyki rzeczywistości. Powoduje to dodatkowy błąd w wynikach modelu.

Z jednej strony procesy przyrodnicze zachodzą w określonej kolejności, podlegają pewnym prawidłowościom i naturalnym ograniczeniom w rzeczywistości fizycznej, co jest głównym warunkiem możliwości ich prognozowania. Z drugiej strony sporadycznie pojawiają się zjawiska i zdarzenia przypadkowe, świadczące o tym, że świat przyrody nie jest w pełni przewidywalny. Powyższe procesy są w większości przypadków złożone i nieliniowe, co wpływa na niestabilność układu i jest źródłem niepewności w modelowaniu hydrologicznym. Niestabilność układu ma jeszcze jedną ważną praktyczną konsekwencję. Przebieg zjawiska z reguły staje się bardzo wrażliwy na wpływ najrozmaitszych, przypadkowych, zewnętrznych czynników. W Polsce przejściowość klimatu sprawia, że czynniki kształtujące ilościową strukturę bilansu wodnego charakteryzują się naturalną niestabilnością. Rzeki są bardzo wrażliwe na fluktuacje różnych zmiennych definiujących klimat. Większy efekt wywołują okresowe nadwyżki i niedobory opadu niż podobne wahania terenowe związane ze zmianami średniej temperatury [Bartnik, Jokiel 1997]. Dodatkowo niepewność modelu w trybie prognozowania wzmacniana jest niedokładnością danych wejściowych pozyskiwanych z meteorologicznych modeli mezoskalowych.

Funkcjonowanie zlewni charakteryzuje duża zmienność i niejednorodność procesów fizycznych, które dominują w funkcji odpowiedzi hydrologicznej, np. intercepcja, infiltracja, procesy zachodzące w strefie saturacji, generowanie odpływu bezpośredniego. W ocenie ilościowej tych procesów istotną rolę odgrywają prawidłowe obserwacje, pomiary oraz przeniesienie pomiarów z małej skali, tj. na podstawie pomiarów punktowych, do pomiarów w skali zlewni. Błąd pomiaru opadu atmosferycznego w stanie ciekłym może dochodzić do 40% objętości, natomiast błędy >20% związane są z pomiarem opadu w stanie stałym, tj. śniegu, gradu [Uhlenbrook 2006]. Niereprezentatywność pomiarowo-obszaryjna opadów stałych prowadzi do nieregularnego odwzorowania takich parametrów śniegu, jak wodny ekwiwalent śniegu czy też dynamika topnienia pokrywy śniegu.

Przy założeniu, że ilościowe pomiary opadu są poprawne, pojawia się kolejny problem związany z niedokładnym szacunkiem intercepcji. Z jednej strony znaczne ilości wody pochodzącej z opadu ulegają wyparowaniu podczas zdarzeń związanych z występowaniem ruchu mas powietrza, z drugiej strony zwiększona retencja wody wynika ze struktury gruntu, np. więcej wody zmagazynują zlewnie zalesione charakteryzujące się grubymi warstwami organicznymi występującymi powyżej gleby mineralnej. Na obszarach pokrytych roślinnością wytwarzana jest niejednorodna struktura gleby, co bezpośrednio skutkuje dużą zmiennością retencji w małej skali i pośrednio oddziałuje na procesy infiltracji, retencji, perkolacji, a także na odpływ i procesy erozji gleby. Dalszy wzrost niepewności pomiarowej związany jest z ilościową reprezentacją ewapotranspiracji, tj. szybkości procesu parowania. W tym przypadku źródłami błędów są sporadyczność wykonywania pomiarów oraz mała liczba pomiarów bezpośrednich.

Dużą rolę odgrywa skala przestrzenna analizy hydrologicznej, która zmienia się od małych obszarów w górnym biegu rzeki do coraz większych wzdłuż profilu podłużnego głównego cieku zlewni. Niedokładność oceny zmiany retencji zlewni wynika z niereprezentatywności rozkładu przestrzennego posterońków pomiarowych rozlokowanych na takich obszarach. Na przykład w zlewniach górskich posterońki opadowe w większości przypadków znajdują się w dolinach, natomiast w zlewniach nizinnych – w ujściowych odcinkach rzek. Istotnym elementem wpływającym na zmienność procesów zachodzących w zlewni hydrologicznej jest skala czasu reakcji rzek. Zmienia się ona od zdarzeń krótkotrwałych (sekundy, minuty, godziny), np. odpływ ze zlewni w wyniku opadów konwekcyjnych, frontalnych, do zdarzeń długotrwałych i trendów (dekady, wieki), np. zjawisko erozji, trendy w zmianach odpływu rzeczno-ego itp.

Błędy pomiaru przepływu rzeczno-ści są kojarzone głównie z pomiarami zdarzeń maksymalnych, tj. wezbrań i niżówek. W okresach, w których rzeki zasilane są wodami gruntowymi (niżówki), pojawia się dodatkowy problem związany z nieprecyzyjnym zdefiniowaniem rozgraniczenia działu wody powierzchniowej i podziemnej. Szczególnie mało informacji dostarcza nam proces przepływu podpowierzchniowego, który w wielu zlewniach jest istotną składową odpływu bezpośredniego ze zlewni. Wiele problemów wiąże się z pomiarem maksymalnych przepływów powodziowych, dla których błędy pomiarowe >20% nie są rzadkością. W takich sytuacjach pomiary prędkości i strumienia objętości wody są najczęściej realizowane przy użyciu mechanicznych młynków hydrometrycznych (ze względu na bezpieczeństwo wykonywanego pomiaru), mimo że współczesna hydrometria oparta jest na technikach akustycznych. Na zmniejszenie dokładności pomiarów hydrometrycznych w warunkach powodziowych ma wpływ turbulentny charakter ruchu wody, który charakteryzuje się pulsacją prędkości, w której cząstki wody poruszają się po torach nierównoległych, często kolistych, określanymi jako wiry – wykonują one ruchy postępowe, obrotowe, śrubowe, a nawet wsteczne.

Z punktu widzenia modelu matematycznego procesy zachodzące w zlewni, dotyczące większości składników cyklu hydrologicznego, tj. wody atmosferycznej, wody powierzchniowej i gruntowej, wilgotności gleby, śniegu, lodu oraz retencji jeziornej, mogą być obiektami podlegającymi modelowaniu. Zadaniem modelu matematycznego jest imitowanie wybranego obiektu lub jego wybranych cech. Matematyczny model obiektu wraz ze zbiorem wszystkich możliwych stanów tworzy układ dynamiczny lub statyczny, w zależności od tego czy stany modelu zmieniają się w czasie, czy też nie. Układy, w których niewielka zmiana warunków początkowych wywołuje niewielką zmianę rozwiązania, są stabilne. Układy, w których niewielka zmiana warunków początkowych wywołuje znaczne zmiany rozwiązania, określane są jako niestabilne. Układy dynamiczne i statyczne mogą być liniowe, czyli takie, w których niewiadome w równaniach występują wyłącznie w pierwszej potęgze, lub nieliniowe. Niestabilne mogą być tylko układy nieliniowe.

1.1. Niepewność prognozy hydrologicznej a praktyka inżynierska

W praktyce inżynierskiej stosuje się dwa sposoby postępowania z błędami w wynikach modelu matematycznego. W modelach deterministycznych błędy bierze się pod uwagę już w procesie kalibracji modeli. Poszukuje się wartości zmiennych stanu w taki sposób, aby zminimalizować reszty, czyli różnice między wynikami symulacji a obserwacjami. Po zakończeniu kalibracji, w kolejnych etapach modelowania, błędy nie są już uwzględniane. Drugim, alternatywnym sposobem jest stosowanie metod probabilistycznych, które są używane nie tylko do prognozowania hydrologicznego, ale także do ilościowej oceny błędu (błędów) modelu. Zazwyczaj na podstawie próby losowej pomiarów hydrologicznych o stosunkowo niewielkiej liczebności możemy określić rozkład prawdopodobieństwa błędów, czyli uzyskać wiedzę, jakie jest prawdopodobieństwo, że błąd ma określoną zmienność. W rzeczywistości użytkownik modelu nie zna dokładnych wartości błędów prognozy. Jeżeli dysponowalibyśmy informacją o błędach prognozy, to nasze przewidywania byłyby stuprocentowo dokładne.

Świadomość niedoskonałości modelu wymusza potrzebę określenia spodziewanego błędu prognozy (niepewności). W przypadku prognozy punktowej wynik powinien zawierać wartości prognozowane i miarę dokładności. Innym sposobem prezentacji niepewności jest prognoza przedziałowa, która opiera się zarówno na prognozie punktowej, jak i błędzie prognozy. Prognoza przedziałowa polega na wyznaczeniu przedziału liczbowego, w którym znajdzie się rzeczywista wartość zmiennej prognozowanej w okresie prognostycznym. Możliwe jest także oszacowanie przedziału liczbowego w taki sposób, że przedział będzie zawierać z określonym prawdopodobieństwem przyszłą realizację zmiennej losowej. Prognozy wyznaczone na podstawie modelu hydrologicznego mogą odbiegać od wartości rzeczywistych zmiennej prognozowanej. Wynika to przede wszystkim z istnienia składnika losowego, tzw. reszty modelu. Odchylenie między zaobserwowanym przepływem w i -tym kroku czasowym a przepływem symulowanym jest podstawową miarą błędu prognozy hydrologicznej. Szerokość przedziału dla konkretnego prawdopodobieństwa zależy od błędu prognozy, zatem chcąc poprawiać prognozę hydrologiczną należy dążyć do minimalizacji błędów.

Błędy prognozy eliminuje się w ramach procedury korekty prognoz hydrologicznych, tzw. updatingu [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 5; Kruszewski, Żelaziński 1985]. Najprostszą metodą bieżącej korekty prognozy jest dodawanie błędu symulacji obliczonego w momencie stawiania prognozy do prognoz obliczonych przez model w kolejnych krokach symulacji, tj. $Q_{s,t} + \delta_s$,

gdzie: $Q_{s,\tau}$ jest przepływem symulowanym w chwili τ ; δ_τ jest procesem resztowym obliczanym jako $\delta_\tau = Q_{s,\tau} - Q_o$; Q_o jest przepływem obserwowanym. Metoda opiera się na założeniu, że błąd symulacji nie zmienia się w czasie. Natomiast błąd prognozy w chwili początkowej τ_0 , tj. w ostatniej chwili, dla której znane (określone) są wartości prognozowanego procesu, wyraźnie wpływa na błędy w kolejnych etapach przetwarzania. Błędy prognoz rosną w miarę wydłużania okresu wyprzedzenia prognozy. Powyższe założenie czasami prowadzi do ujemnych wartości prognozowanego przepływu, kiedy ujemny błąd symulacji δ_τ ma wyższą wartość od przepływu symulowanego $Q_{s,\tau}$.

Często stosowaną metodą jest korekta nie stanu początkowego modelu symulacyjnego. Proces polega na doprowadzeniu do pełnej zgodności między wynikami symulacji a obserwacjami w momencie stawiania prognozy. Zakładając, że błędy obserwowanych danych wymuszających (tj. wejścia do modelu) są jedynym źródłem błędów symulacji, updating uzyskiwany jest w wyniku zmiany wartości danych wejściowych, np. opadu w sezonie letnim czy też temperatury powietrza zimą (temperatura determinuje tajanie pokrywy śniegu). Innym sposobem jest korygowanie zmiennych stanu modelu hydrologicznego. Najczęściej korekta dotyczy wilgotności gleby, pokrywy śniegu, zawartości wody w śniegu oraz nieobserwowanych (niemierzonych) zmiennych, np. wskaźnika infiltracji gleby, górnego i dolnego ograniczenia zawartości zbiornika wody gruntowej itp. Zmiany wartości tych zmiennych stanu prowadzą do poprawy zgodności symulacji i obserwacji w momencie stawiania prognozy. Jeśli wziąć pod uwagę punktowy pomiar powyższych wielkości fizycznych, a w konsekwencji wysoki poziom niepewności w ocenie obszarowej pomiarów naziemnych, to wydaje się, że lepszym rozwiązaniem jest wykorzystanie danych satelitarnych potraktowanych w tym przypadku jako potencjalne obserwacje *in situ* [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 6]. Ponieważ rozkład przestrzenny i czasowy danych satelitarnych może znacznie różnić się od naziemnych danych obserwowanych [Kalbarczyk, Kalbarczyk 2004; Sidur 2018], pomiary powyższych zmiennych powinny być asymilowane przez symulacyjny model hydrologiczny lub odpowiednio skonstruowany preprocesor, a nie być bezpośrednim wejściem do modelu deterministycznego. W związku z powyższym, w większości hydrologicznych modeli symulacyjnych procedura asymilacji danych [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 7] powinna być niezbędną procedurą (preprocesorem) łączącą dane satelitarne z parametrami stanu modelu.

1.2. Cel i zakres pracy

Celem niniejszego opracowania było przedstawienie zbioru istniejących oraz autorskich metod i modeli probabilistycznych uwzględniających niepewność w modelowaniu i zastosowanie ich w praktyce hydrologicznej do poprawy symulacji i prognozy hydrologicznej. Niedostateczna precyzja opisu modelowanego zjawiska wynika często z uwzględniania zbyt wielu zmiennych, przyjmowania innych niż liniowa postaci analitycznych równań i stosowania zbyt wielu założeń o niespójnej wewnętrznej zgodności. Zatem umiejętność opisu i redukcji niepewności w prognozowaniu hydrologicznym, niezależnie od jej pochodzenia, jest bardzo ważnym składnikiem procesu modelowania. Jednym ze sposobów zredukowania niedokładności modeli stochastycznych jest wykorzystanie dodatkowej pozastatystycznej informacji, zastosowanie bardziej wyrafinowanych metod modelowania oraz wykorzystanie szerszego zestawu funkcji konstruowanych modeli.

W pracy omówiono dwa złożone modele hydrologiczne składające się z dwóch członów, tj. preprocesora danych wejściowych oraz modelu matematycznego, przy czym preprocesor danych jest również pewnym modelem probabilistycznym. W pierwszym modelu preprocesor danych wejściowych jest stosowany do dwóch modeli regresyjnych, tj. statycznego [Howson, Urbach 2005] i dynamicznego [Harvey 1986] liniowego modelu Bayesowskiego (odpowiednio SLMB i DLMB), natomiast w drugim – do korekcji satelitarnych obserwacji wysokości opadu i wilgotności gleby oraz do autorskiego rozwiązania asymilacji satelitarnych obserwacji wilgotności gleby, będących wejściem do szwedzkiego deterministycznego modelu hydrologicznego typu opad–odpływ HBV [Bergström 1976, 1992, 1995; Lindström i in. 1997; Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 11, 12].

Preprocesor danych wejściowych do modeli regresyjnych identyfikuje zdarzenia hydrologiczne będące przedmiotem modelowania matematycznego, wyznacza ilościowe charakterystyki zmiennych prognozowanych i prognozujących oraz wykonuje analizę statystyczną zmiennych będących wejściem

do modeli za pomocą metod statystycznych. W ramach analizy preprocesor danych sprawdza niezależność i stacjonarność danych wejściowych; w zależności od wyników testów statystycznych ciągi pomiarowe są klasyfikowane jako odpowiednie wejście do statycznych lub dynamicznych modeli regresji.

Statyczne liniowe modele Bayesowskie, przeznaczone do prognozowania wezbrania deszczowo-roztopowego, opracowano w oparciu na dwóch wielowymiarowych modelach liniowych w postaci wielokrotnej regresji I rodzaju (WLR) ze skalarną zmienną prognozowaną i wielowymiarowej wielokrotnej regresji (WWLR), w której zmienna prognozowana jest wektorem. Parametry w obu statycznych modelach estymowano metodą Bayesowską (WLR-B, WWLR-B) i techniką odwrotnego modelowania (inverse modelling) (WLR-InvB, WWLR-InvB). Natomiast modele dynamiczne przeznaczone do krótkoterminowej prognozy przepływu utworzono na podstawie dwóch teoretycznych liniowych modeli Bayesowskich, tj. Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR).

W klasie modeli Bayesowskich przyjęto założenie, że zmienne prognozujące określają zmienną prognozowaną przy stałych parametrach modelu. Oznacza to, że parametry strukturalne modelu uznaje się za wielkości ustalone, lecz nieznanne, które należy oszacować. Dopuszcza się, by parametry strukturalne modelu były różne. W tym przypadku przyjmuje się, że o wartościach parametrów decyduje hydrolog dysponujący pewną wiedzą *a priori*, tzn. wiedzą znaną przed uzyskaniem informacji z próby. Uznaje się ponadto, że wiedza ta może być wyrażona w postaci probabilistycznej, tzn. że parametry modelu są zmiennymi losowymi, a hydrolog jest w stanie przyporządkować tym parametrom znany rozkład, zwany *a priori* [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 13].

Nowością w pracy jest zaprogramowanie SLMB i DLMB pod kątem wymagań dotyczących hydrologicznego prognozowania operacyjnego poprzez uwzględnienie:

- 1) Updatingu modelu.
- 2) Zdolności do zapisywania zmiennych stanu modelu w każdym etapie przetwarzania, a następnie uruchomienia algorytmów z zapisanymi w ostatnim kroku zmiennymi stanu.
- 3) Opracowania prognozy punktowej i przedziałowej.

Drugi, przedstawiony w niniejszy pracy, złożony model hydrologiczny ukierunkowany jest na poprawę krótkoterminowej prognozy hydrologicznej poprzez zminimalizowanie rozbieżności między symulowanym a obserwowanym przepływem rzeczny w momencie stawiania prognozy. Model składa się z preprocesora danych wejściowych do modelu hydrologicznego typu opad–odpływ (HBV). Preprocesor danych przetwarza satelitarne obserwacje opadu będące zmiennymi wymuszającymi model hydrologiczny oraz asymiluje satelitarne obserwacje wilgotności gleby. Preprocesor w pierwszej części obliczeń analizuje statystycznie dane wejściowe oraz wybiera najlepiej dopasowane funkcje rozkładu prawdopodobieństwa dla opadu i wilgotności gleby do następujących metod:

- 1) Usuwanie błędu systematycznego metodą korekcji obciążenia (bias correction, BC) z satelitarnych obserwacji opadów [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 21].
- 2) Usuwanie błędu systematycznego z satelitarnych obserwacji wilgotności gleby, w celu bezpośredniego zastąpienia odpowiednich prognozowanych zmiennych stanu wilgotności gleby modelem hydrologicznym obserwacjami satelitarnymi z usuniętym obciążeniem.
- 3) Asymilacji nieskorygowanej satelitarnej wilgotności gleby, z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV za pomocą zmodyfikowanych filtrów Kalmana [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 16]: EnKF-BC i EnSRF-BC.
- 4) Asymilacji nieskorygowanej satelitarnej wilgotności gleby, z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV za pomocą hybrydowego filtra złożonego z filtra EnSRF i filtra wariacyjnego 4DVAR [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 18], tj. EnSRF-BC/4DVAR.

Metodę BC oparto na algorytmie transformaty dystrybucyjnej [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 20] i zastosowano do obserwowanych dobowych sum opadów i obserwacji wilgotności gleby monitorowanych przez satelity Europejskiej Organizacji Eksploatacji Satelitów Meteorologicznych (European Organization for the Exploitation of Meteorological Satellites, EUMETSAT). Procedura BC wykorzystuje trzy teoretyczne, trójparametrowe rozkłady prawdopodobieństwa: uogólniony wykładniczy (GE), gamma (GA) i Weibull (WE), do generowania wolnych od obciążenia obserwacji satelitarnych. Transformacja statystyczna wykonywana jest dla takich samych rozkładów

prawdopodobieństwa (np. GE-GE, GA-GA, WE-WE) lub dla różnych (np. GE-GA, GE-WE, GA-WE, itd.). Nowością w budowie preprocesora danych wejściowych do modelu HBV było wykorzystanie algorytmu BC w procedurze asymilacji satelitarnych obserwacji wilgotności gleby do usuwania obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu hydrologicznego za pomocą zmodyfikowanego w pracy algorytmu wiązkowych filtrów Kalmana i filtru hybrydowego.

1.3. Układ pracy

Rozdział 1

Przedstawiono cel pracy wraz z uzasadnieniem wyboru tematu oraz wprowadzono do zagadnień związanych z metodami probabilistycznymi stosowanymi w prognozowaniu hydrologicznym. Opisano zasady i problemy modelowania zjawiska hydrologicznego w odniesieniu do procesów przyrodniczych występujących w zlewni, modelowania matematycznego oraz zasad praktyki inżynierskiej dotyczącej minimalizowania niepewności prognozy hydrologicznej, czyli sposobów postępowania z błędami modelu matematycznego.

Rozdział 2

Dotyczy danych wejściowych do modeli probabilistycznych zastosowanych w prognozowaniu hydrologicznym. Przedstawiono opis zlewni testowych, metodykę opracowania danych wejściowych i analizę statystyczną w ramach preprocesora wejściowego do modeli regresyjnych SLMB, DLMB i deterministycznego modelu HBV. Analiza statystyczna obejmowała badanie jednorodności oraz sprawdzenie własności losowych badanych zmiennych, które uwzględniono w wyznaczaniu charakterystyk rozkładu prawdopodobieństwa. Zależność, niezależność, stacjonarność i niestacjonarność zmiennych losowych były podstawą, kwalifikującą, badaną zmienną losową jako wejście do odpowiedniego modelu hydrologicznego.

Rozdział 3

Wprowadza czytelnika w zagadnienia związane z modelowaniem matematycznym, uwzględniającym w obliczeniach niepewności wynikające z przyczyn wymienionych w rozdziale 1. Prezentuje metody i modele zastosowane do analizy niepewności w prognozowaniu hydrologicznym oraz do modelowania wezbrania deszczowo-roztopowego, całorocznego hydrogramu przepływów i poprawy krótkoterminowej prognozy hydrologicznej. W rozdziale przedstawiono algorytmy modeli regresyjnych, tj. wielokrotnej liniowej regresji (WLR), wielowymiarowego modelu wielokrotnej regresji (WWLR), oraz algorytmy Bayesowskich modeli regresyjnych, tj. SLMB i DLMB, umożliwiające modelowanie niepewności występującej w procesie modelowania wezbrania hydrologicznego i przepływu rzecznoego. Prezentację modeli, opierających swoje algorytmy na danych, uzupełniono o sieci neuronowe (NN). W drugiej części rozdziału przedstawiono deterministyczny model typu opad-odpływ HBV oraz metody zastosowane w preprocesorze danych wejściowych do modelu HBV. Zastosowane metody dotyczą usuwania błędów systematycznego (obciążenia) i asymilacji danych. Przedstawiono algorytm usuwania obciążenia z satelitarnych obserwacji opadu na podstawie danych z naziemnej sieci pomiarowo-obszaryjnej i wilgotności gleby w oparciu na ciągach wilgotności gleby wygenerowanych przez model hydrologiczny HBV. Opracowano algorytmy asymilujące satelitarne obserwacje wilgotności gleby za pomocą, zmodyfikowanych w pracy, wiązkowych filtrów Kalmana i metody hybrydowej w połączeniu z filtrem wariacyjnym.

Rozdział 4

Przedstawiono przykład zastosowania statycznych liniowych modeli Bayesowskich do symulacji hydrologicznej i średnioterminowej prognozy wezbrania deszczowo-roztopowego. Opracowana metodyka uwzględniła procedury preprocesora wejścia danych, czyli identyfikację wezbrań deszczowo-roztopowych, analizę statystyczną danych wejściowych do modeli SLMB, wybór optymalnego zbioru zmiennych prognozowanych (w postaci skalarnej i wektorowej) oraz zbioru zmiennych prognozujących do wielokrotnej i wielowymiarowej wielokrotnej liniowej regresji z parametrami estymowanymi metodami B i InvB. W drugiej części rozdziału wykonano analizę graficzną modeli SLMB z uwzględnieniem przedziału niepewności.

Rozdział 5

W całości dotyczy dynamicznych liniowych modeli Bayesowskich (DLMB). Zagadnienia merytoryczne zilustrowano przykładem zastosowania dwóch liniowych modeli Bayesowskich – Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR) – do symulacji rocznego hydrogramu oraz krótkoterminowej prognozy hydrologicznej przepływu rzecznoego. Uzyskane wyniki symulacji i prognozy na podstawie DLMB porównano z wielokrotną liniową regresją (WLR) i sieciami neuronowymi (NN). Opracowana metodyka uwzględniała preprocesor danych wejściowych do modeli (tj. identyfikację modelowanego zjawiska hydrologicznego, analizę statystyczną danych wejściowych), kalibrację modeli DLMB, WLR i NN oraz analizę porównawczą czterech metod obliczeniowych zastosowanych do rocznych symulacji hydrologicznych i prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym.

Rozdział 6

Skoncentrowano się na zastosowaniu metod probabilistycznych do budowy preprocesora danych wejściowych do modelu hydrologicznego HBV, w celu poprawy krótkoterminowej prognozy hydrologicznej. W preprocesorze danych zastosowano procedury analizy statystycznej danych wejściowych, badania jednorodności i wyboru najlepiej dopasowanej funkcji rozkładu prawdopodobieństwa dla opadu i wilgotności gleby do korekcji obciążenia i asymilacji meteorologicznych obserwacji satelitarnych. Preprocesor danych asymiluje satelitarne obserwacje wilgotności gleby za pomocą zmodyfikowanych w pracy wiązkowych filtrów Kalmana i metody hybrydowej w połączeniu z filtrem wariacyjnym. Poprawione prognozy hydrologiczne przedstawiono w postaci prognozy wiązkowej i w skali błędów [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 22, 23].

Rozdział 7.

Całość pracy kończy podsumowanie poruszonych zagadnień oraz wnioski końcowe, wskazujące na możliwość praktycznego wykorzystania metod probabilistycznych zawartych w preprocesorze danych wejściowych, w połączeniu z modelami regresyjnymi i modelem deterministycznym, w operacyjnym prognozowaniu hydrologicznym.

We wszystkich opracowanych w pracy metodach (modelach) niepewność prognozy hydrologicznej określono ilościowo, w formie rozkładu prawdopodobieństwa, za pomocą prognozy przedziałowej lub prognozy w skali błędu. Dla podkreślenia skuteczności i uniwersalności zaproponowanej w pracy analizy probabilistycznej w prognozowaniu hydrologicznym, przedstawiono przykłady obejmujące dużą i małą zlewnię nizinną oraz zlewnię górską. Ważnym elementem pracy jest implementacja wszystkich zaproponowanych metod w taki sposób, aby była możliwość ich uruchomienia w trybie pracy operacyjnej, tj. dla potrzeb prognozy hydrologicznej

2. DANE

2.1. Zlewnie testowe do modelowania probabilistycznego w prognozowaniu hydrologicznym

Analizę niepewności w prognozowaniu hydrologicznym przedstawiono na przykładzie modelowania trzech różnych zlewni hydrologicznych. Do prognozowania wezbrania deszczowo-roztopowego wybrano małą niziną zlewnię rzeki Sokółdy, charakteryzującą się występowaniem stałej pokrywy śniegu w sezonie zimowym, tj. od listopada do kwietnia. Na skutek gwałtownego ocieplenia, topniejąca pokrywa śniegu, często wspomagana opadami deszczu, wyzwała wzrost stanów wody. Zlewnia Sokółdy umożliwiła identyfikację wezbrań deszczowo-roztopowych niezakłóconych zjawiskami lodowymi czy też pracą urządzeń hydrotechnicznych. Prognozowanie codziennego przepływu rzecznoego wykonano dla dużej zlewni nizinnej Górnej Narwi w profilu wodowskazowym Wizna. Całoroczny hydrogram przepływu rzecznoego obrazuje zmienność odpływu ze zlewni. Przepływy gwałtownie wzrastają wraz z pojawiającym się spływem powierzchniowym, który jest efektem opadów lub/i topnienia śniegu. Natomiast okresy bezdeszczowe i okresy z dominującym zasilaniem wodami podziemnymi są przyczyną występowania niżówek. Reżim hydrologiczny uzupełniają spiętrzenia zwierciadła wody w wyniku nasilenia zjawisk lodowych, do których zalicza się np. intensywne tworzenie się sryżu lub lodu dennego, powodujących zmniejszenie przekroju przepływu lub spiętrzenie spływającej kry. Ostatnią modelowaną zlewnią jest górską zlewnia rzeki Soły – mocno urzeźbiona, z dużym udziałem procentowym spadków (20-30%), co ma wpływ na dokładność pomiarów meteorologicznych – zwłaszcza opadu [Lipski i in. 2005]. Liczba i intensywność oraz rozkład opadów mają wpływ na wilgotność gleby, która jest jedną z najważniejszych zmiennych stanu modelu hydrologicznego typu opad–odpływ. Na przykładzie zlewni Soły przedstawiono wpływ asymilacji satelitarnych obserwacji meteorologicznych na niepewność prognozy codziennego przepływu w ujściowym profilu wodowskazowym w Żywcu.

2.2. Metodyka analizy statystycznej danych wejściowych do modeli

Analiza statystyczna [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 24] danych wejściowych do modeli składa się z dwóch etapów, tj.:

- I. Badania jednorodności statystycznej zmiennych losowych.
- II. Wyznaczania charakterystyk rozkładu prawdopodobieństwa.

I. BADANIE JEDNORODNOŚCI STATYSTYCZNEJ ZMIENNYCH LOSOWYCH

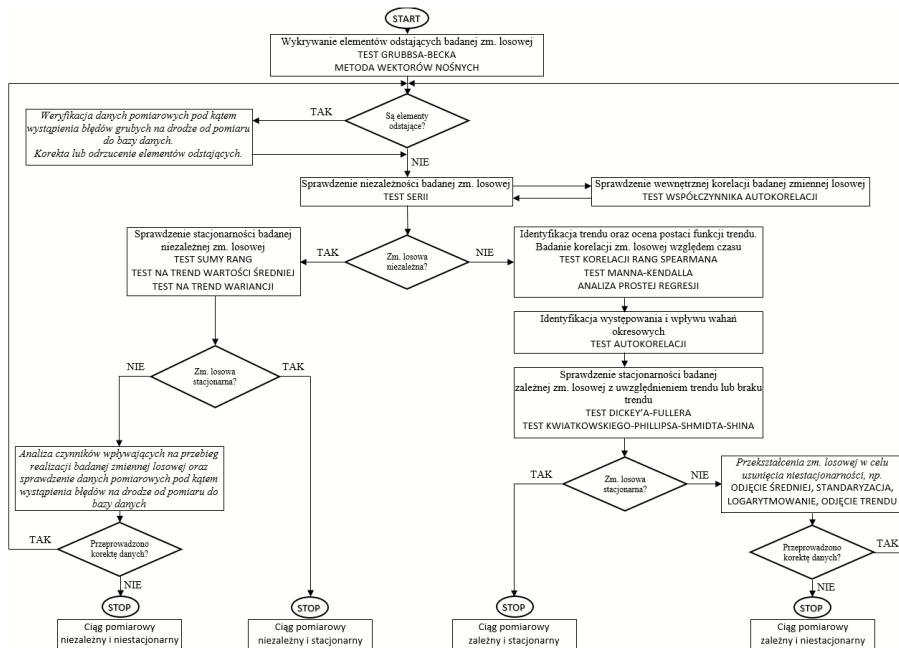
Analizę niejednorodności [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 25] ciągów pomiarowych metodami statystycznymi przeprowadzono zgodnie z procedurą zilustrowaną na rysunku 2.1. W pierwszym etapie wykonano wykrywanie elementów odstających badanej zmiennej losowej za pomocą testu Grubbsa-Becka [Grubbs, Beck 1972] lub metodą wektorów nośnych SVM [Hawkins 1980; Ben-Hur i in. 2001]. Test Grubbsa-Becka [Cohn i in. 2013] służy do odpowiedzi na pytanie, czy wartości maksymalna lub minimalna w próbie są elementami odstającymi. Powyższa metoda zalecana jest do wykrywania elementów odstających o niskich wartościach i w próbach losowych o liczebności $10 \leq n \leq 150$. Do analizy prób losowych o większej liczebności zastosowano metodę SVM. Pozwala ona na wyznaczenie uogólnionego rozkładu generującego dane z analizowanego zbioru [Trzęsiok 2007]. Uogólniony kwantyl rozkładu to taki obszar $Q \in \mathbb{R}^d$ d wymiarowej przestrzeni danych, który spełnia warunek, że niemal wszystkie obserwacje wygenerowane z rozkładu należą do Q , z drugiej strony niemal wszystkie obiekty nie pochodzące z rozkładu generującego dane należą do dopełnienia zbioru Q . Powyższą metodę zastosowano dla przypadku jednowymiarowego, tj. dla $d = 1$.

Kluczowym elementem procedury analizy niejednorodności (rys. 2.1) jest sprawdzenie niezależności i wewnętrznego skorelowania badanej zmiennej losowej – odpowiednio przy wykorzystaniu testu serii i testu współczynnika autokorelacji [Wald, Wolfowitz 1940; Anderson 1941]. Zadaniem obu testów jest sprawdzenie, czy obserwowana realizacja badanej zmiennej losowej ma w rzeczywistości charakter losowy, czy też jej wybór podlega tendencji lub cykliczności. W teście współczynnika autokore-

lacji przyjęto arbitralnie maksymalne opóźnienie $Lag = 8$ dób. Zatem wyznaczane są autokorelacje między pierwszą a kolejną, tj. drugą, trzecią, ..., ósmą obserwacją, a sprawdzian testu dla odpowiednich przesunięć jest porównywany z wartościami dolnej i górnej granicy przedziału ufności. W przypadku, gdy sprawdzian testu nie mieści się w przedziale ufności, hipotezę H_0 o braku autokorelacji między elementami ciągu pomiarowego należy odrzucić. Brak autokorelacji między kolejnymi elementami z maksymalnym przesunięciem 8 dób powinien wskazać na brak wewnętrznej współzależności (skorelowania) i tym samym na losowy charakter badanej zmiennej. W zależności od uzyskanego wyniku testu procedura rozgałęzia się na część związaną z testowaniem stacjonarności niezależnej zmiennej losowej (lewa strona na rys. 2.1) lub zależnej (prawa strona na rys. 2.1).

Charakterystykami statystycznymi ciągów pomiarowych najczęściej ulegającymi zmianom są wartość średnia i wariancja. Zmiany w wartościach tych charakterystyk mają charakter skokowy lub występują w postaci trendów. Zbadanie stacjonarności, tj. niezmienności czasowej powyższych parametrów statystycznych dla niezależnej zmiennej losowej (lewa strona na rys. 2.1), wykonano za pomocą testu sumy rang Kruskala-Wallisza oraz testu współczynnika Spearmana korelacji rangowej na trend wartości średniej i trend wariancji [Kruskal, Wallis 1952; Lehmann 1959; Yevjevich, Jeng 1969; Pruchnicki 1976].

W przypadku zależnej zmiennej losowej (prawa strona na rys. 2.1) sprawdzenie stacjonarności poprzedza procedura identyfikacji trendu, oceny funkcji trendu oraz identyfikacji występowania i wpływu wahań okresowych. Korelacje między badanymi zmiennymi losowymi określone są za pomocą nieparametrycznego testu korelacji rang Spearmana [Spearman 1940]. Do wykrywania i zbadania istotności tendencji zmian w ciągach pomiarowych zastosowano test Manna-Kendalla M-K [Hamed 2008] oraz analizę prostej regresji [Kala 2002]. Test M-K służy do analizy danych w dziedzinie czasu pod kątem stale rosnących lub malejących trendów. Jest to test nieparametryczny, zatem może zostać zastosowany dla dowolnego rozkładu, tzn. badany ciąg nie musi spełniać założenia normalności, jednak nie powinna występować w nim autokorelacja. Funkcja autokorelacji potwierdziła lub wykluczyła występowanie wahań okresowych w badanym ciągu pomiarowym i obliczana jest na danych oryginalnych bez eliminowania ewentualnego trendu. Ocena funkcji trendu wykonuje się za pomocą analizy prostej regresji, gdzie zmienną zależną jest badana zmienna losowa, a zmienną niezależną czas, τ .



Rys. 2.1. Procedura analizy niejednorodności niezależnych i zależnych ciągów pomiarowych.
Fig. 2.1. Analysis of the homogeneity of independent and dependent data.

Ostatnim elementem procedury analizy niejednorodności zależnych ciągów pomiarowych jest sprawdzenie stacjonarności badanej zmiennej losowej z uwzględnieniem wykrytego trendu lub braku trendu. Ciąg pomiarowy jest stacjonarny, jeżeli ma skończoną wariancję oraz jeśli kowariancja między obserwacjami nie zależy od czasu, a jedynie od odległości między obserwacjami. Podstawowym testem do badania stacjonarności jest rozszerzony test pierwiastka jednostkowego Dickey'a-Fullera ADF [Dickey, Fuller 1979], który uwzględnia możliwość występowania autokorelacji składnika losowego [Charemza, Deadman 1997]. Testem uzupełniającym badanie w całym zakresie pomiarowym jest test Kwiatkowskiego, Phillipsa, Schmidta i Shina KPSS [Kwiatkowski i in. 1992].

Wyjściem z procedury (rys. 2.1) jest informacja o jednorodności badanych ciągów pomiarowych. Badany ciąg może zostać zakwalifikowany jako niezależny stacjonarny lub niezależny niestacjonarny albo zależny stacjonarny lub zależny niestacjonarny. W zależności od uzyskanej informacji, badane ciągi będą przetwarzane za pomocą odpowiednich modeli matematycznych.

II. WYZNACZANIE CHARAKTERYSTYK ROZKŁADU PRAWDOPODOBIENSTWA

Procedura wyznaczania charakterystyk rozkładu prawdopodobieństwa uwzględnia następujące miary rozkładu:

- 1) miarę asymetrii rozkładu:
 - współczynnik skośności;
- 2) miarę koncentracji rozkładu:
 - współczynnik spłaszczenia (kurtoza);
- 3) miarę rozproszenia rozkładu:
 - wariancję;
- 4) miary położenia rozkładu:
 - średnią arytmetyczną;
 - medianę.

Zakładając, że próba losowa składa się z n niezależnych zmiennych losowych X_1, X_2, \dots, X_n , wartość oczekiwaną wyznacza się za pomocą wzoru [Kala 2002]:

$$E(X) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad (2.1)$$

natomiast wariancję przy użyciu formuły [Kala 2002]:

$$\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var} X_i = \frac{1}{n} (X_i - \bar{X})^2 \quad (2.2)$$

Zatem wariancję w próbie losowej s^2 obliczamy jako średnią z odchyleń od średniej:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.3)$$

gdzie \bar{x} jest wartością średnią z próby.

W przypadku zależnych ciągów pomiarowych wariancja sumy n zmiennych losowych X_i , $i = 1, \dots, n$ wynosi [Zhang i in. 2012]:

$$\text{Var}(\sum_{i=1}^n X_i) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var} X_i + \sum_{i \neq j}^n \text{Cov}(X_i, X_j) \quad (2.4)$$

natomiast wariancję w próbie losowej obliczamy za pomocą wzoru [Zhang i in. 2012]:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{1}{n-1} \sum_{i \neq j}^n (x_i - \bar{x})(x_j - \bar{x}) \quad (2.5)$$

Rozróżnienie sposobu wyznaczania wariancji dla zmiennych niezależnych i zależnych wpływa także na obliczanie wyższych momentów z próby, tj. współczynnika skośności (2.6) i współczynnika spłaszczenia (kurtozy) (2.7) [Górka 2007]:

$$SKE = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)s^3} \quad (2.6)$$

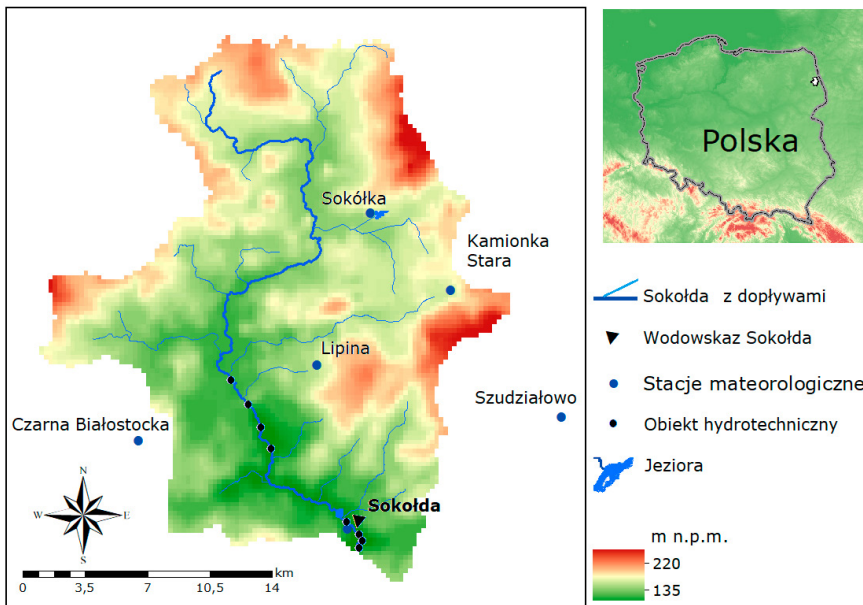
$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3 \quad (2.7)$$

gdzie powyższe statystyki są obciążonymi estymatorami odpowiednio współczynnika skośności i spłaszczenia; \bar{x} jest średnią z próby, a s odchyleniem standardowym.

2.3. Dane wejściowe do statycznych liniowych modeli Bayesowskich (SLMB)

2.3.1. Opis hydrologii zlewni Sokółdy

Zlewnia Sokółdy, o powierzchni 464 km², leży w północno-wschodnim regionie Polski (dł. geogr. 23,06° – 23,61°E, szer. geogr. 53,25° – 53,72°N; rys. 2.2). Źródła rzeki znajdują się na wysokości 206 m n.p.m., a ujście na wysokości 125,5 m n.p.m. Na całej jej długości (47,2 km) znajdują się liczne przepusty wodne, młyny i jazy służące do nawadniania i pracujące w okresie wegetacyjnym. W czasie odpływu zimowo-wiosennego Sokółda jest wolna od zakłóceń wynikających z pracy obiektów hydrotechnicznych. Ciek w zlewni Sokółdy charakteryzują się wysoką inercją i umiarkowanym reżimem hydrologicznym. Wezbrania roztopowe i deszczowo-roztopowe dominują zimą i wczesną wiosną. Poprzedzone są występowaniem stałej pokrywy śnieżnej. W okresach wysokich przepływów, szczególnie od marca do kwietnia, Sokółda wylewa na słabo zaludnione terasy zalewowe, co nie stanowi poważnego zagrożenia dla ludzi lub infrastruktury.



Rys. 2.2. Mapa hydrograficzna zlewni Sokółdy, przedstawiająca lokalizację posterunku wodowskazowego Sokółda w ujściowym odcinku rzeki.

Fig. 2.2. Hydrographic map of the Sokółda River watershed including the location of the Sokółda gauging site situated at the watershed outlet.

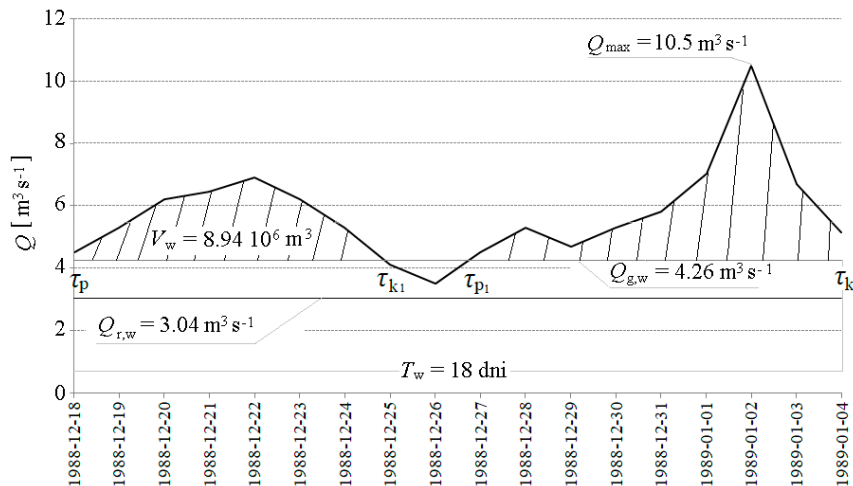
W górnym biegu rzeki dolina Sokołdy charakteryzuje się pagórkowatym terenem i glebami piaszczystymi, podczas gdy dno doliny jest głównie gliniaste. Liczne ciekły odwadniają rozległe obniżenia wypełnione torfem. W środkowej części zlewni Sokołdy i jej dopływy odwadniają obszary sandrowe Puszczy Knyszyńskiej i strefy moren czołowych o wysokości do 186 m n.p.m. Południową część tworzą piaszczyste gleby przykryte gęstą warstwą lasów (rys. 2.2). Zlewnia użytkowana jest głównie rolniczo.

2.3.2. Opracowanie danych wejściowych do modeli SLMB

IDENTYFIKACJA WEZBRAŃ DESZCZOWO-ROZTOPOWYCH W ZLEWNI SOKOŁDY

Do wyodrębnienia wezbrań deszczowo-roztopowych w zlewni w przekroju wodowskazowym Sokołda arbitralnie przyjęto wartość przepływu granicznego wg kryterium hydrologicznego $Q_{g,w}$ (rys. 2.3) korespondującego do NWW, tj. najniższego stanu wysokiego zaobserwowanego w ujściowym odcinku rzeki Sokołdy w latach 1966-2002. $Q_{g,w}$ został wyrażony w postaci $Q_{g,w} = NWQ_{XI-IV, 1966-2002}$, gdzie $NWQ_{XI-IV, 1966-2002}$ jest najniższym wysokim przepływem odpowiadającym NWW. Wezbranie zdefiniowano jako okres, w którym przepływy są równe lub wyższe od przyjętego przepływu granicznego $Q_t \geq Q_{g,w}$ i trwa do dnia, w którym zaobserwowany przepływ obniżył się poniżej wartości granicy $Q_{g,w}$. Poza tym wymagana jest stałość czynników wywołujących występowanie w analizowanym przekroju wodowskazowym wezbrania deszczowo-roztopowego.

Dodatkowo przyjęto kryterium rozdziału sąsiadujących wezbrań. Z punktu widzenia hydrologicznego ustalono, że wezbrania przedzielone okresami o przepływach niższych od NWQ i wyższych od $Q_{g,w} \geq SSQ_{XI-IV, 1966-2002}$, gdzie $SSQ_{XI-IV, 1966-2002}$ jest średnim ze średnich przepływów, oraz trwających krócej niż 5 dni traktowane są jako jedno wezbranie. Okresy historycznych wezbrań różnią się między sobą, natomiast ich geneza powinna być taka sama. Powyższa definicja wezbrania umożliwiła zestawienie zaobserwowanych (historycznych) ciągów pomiarowych, na podstawie których opracowano charakterystyki ilościowe wezbrań deszczowo-roztopowych, będące zbiorem zmiennych zależnych (prognozowanych). Odpowiadające im fizyczne uwarunkowania wezbrań, utworzyły zbiór potencjalnych zmiennych niezależnych (prognozujących). W kolejnym etapie opracowania sformułowano model statystyczny uwzględniający zależności występujące między oboma zbiorami.



Rys. 2.3. Definiowanie wezbrania w zlewni Sokołdy w przekroju wodowskazowym Sokołda na przykładzie wezbrania roztopowego w grudniu 1988 roku.

Fig. 2.3. Defining the flood in the Sokołda watershed at Sokołda on the example of the snow-melt flood in December 1988.

Fizyczne uwarunkowania występowania jednorodnych wezbrań deszczowo-roztopowych analizowano, biorąc pod uwagę następujące czynniki meteorologiczne:

- 1) Dobowe i miesięczne sumy opadów deszczu i śniegu.

- 2) Średnią i minimalną dobową temperaturę powietrza.
- 3) Średnią dobową temperaturę powietrza na poziomie gruntu.
- 4) Zmienność temperatury powietrza w ciągu doby.
- 5) Występowanie pokrywy śniegu.
- 6) Stan gruntu, tj. informację o tym czy grunt jest rozmarznięty, czy zamrożony oraz średnią dobową temperaturę na głębokości 0,50 m, która jest bardzo ważnym czynnikiem sprzyjającym wezbraniom roztopowym w małych zlewniach.

Źródłem powyższych informacji były pomiary i obserwacje wykonane na stacjach meteorologicznych w Czarnej Białostockiej, Sokółce, Lipinie, Szudziałowie, Sokółdzie i Supraślu (rys. 2.2).

Przeprowadzona analiza zidentyfikowała trzydzieści osiem wezbrań deszczowo-roztopowych w zlewni Sokołdy w profilu wodowskazowym Sokołda w latach 1966-2002 i umożliwiła wyodrębnienie więcej niż jednego wezbrania w jednym roku hydrologicznym (dwa wezbrania w 1971) oraz lata hydrologiczne bez wezbrania (np. w 1969). Wezbrania z lat 1966-1990 wykorzystano do kalibracji modeli, do walidacji wybrano okres 1991-2002. Największe wezbrania poprzedzone były występowaniem na obszarze zlewni pokrywy śnieżnej o grubości przekraczającej 0,5 m (1970, 1979, 1980, 1988, 1991, 1996, 1999 i 2002) i zamrożonego gruntu (1978, 1979, 1981, 1982, 1989, 1994, 1996, 1999, 2000 i 2002), co przyczyniło się do szybkiego spływu powierzchniowego i podpowierzchniowego, a w efekcie do uformowania się wysokiej fali wezbraniowej.

CHARAKTERYSTYKI ILOŚCIOWE WEZBRAŃ DESZCZOWO-ROZTOPOWYCH

Wezbrania deszczowo-roztopowe opisano trzema charakterystykami: przepływem kulminacyjnym Q_{\max} ($m^3 \cdot s^{-1}$), tj. maksymalnym natężeniem przepływu zaobserwowanym w okresie, w którym przepływy są równe lub wyższe od $Q_{g,w}$, objętością wezbrania V_w ($10^6 m^3$) i czasem trwania wezbrania T_w (doby). Zgodnie z przyjętą definicją wezbrania, charakterystyki ilościowe (rys. 2.3) określono dla przepływu powyżej $Q_{g,w}$:

1) czas trwania:

$$T_w = \begin{cases} \tau_k - \tau_p + 1 & Q_\tau > Q_{g,w} \quad \tau \in \{\tau_p, \dots, \tau_k\} \\ (\tau_{k1} - \tau_p) + (\tau_k - \tau_{p1}) + 2 & \text{gd}y \quad Q_\tau > Q_{g,w} \quad \tau \in \{\tau_p, \dots, \tau_{k1}\} \cup \tau \in \{\tau_{p1}, \dots, \tau_k\} \\ SSQ \leq Q_\tau < Q_{g,w} \cap \tau \in \{\tau_{k1}, \dots, \tau_{p1}\} \cap \tau_{p1} - \tau_{k1} + 1 < 5 \text{ d}ó\text{b} \end{cases} \quad (2.8)$$

2) objętość wezbrania:

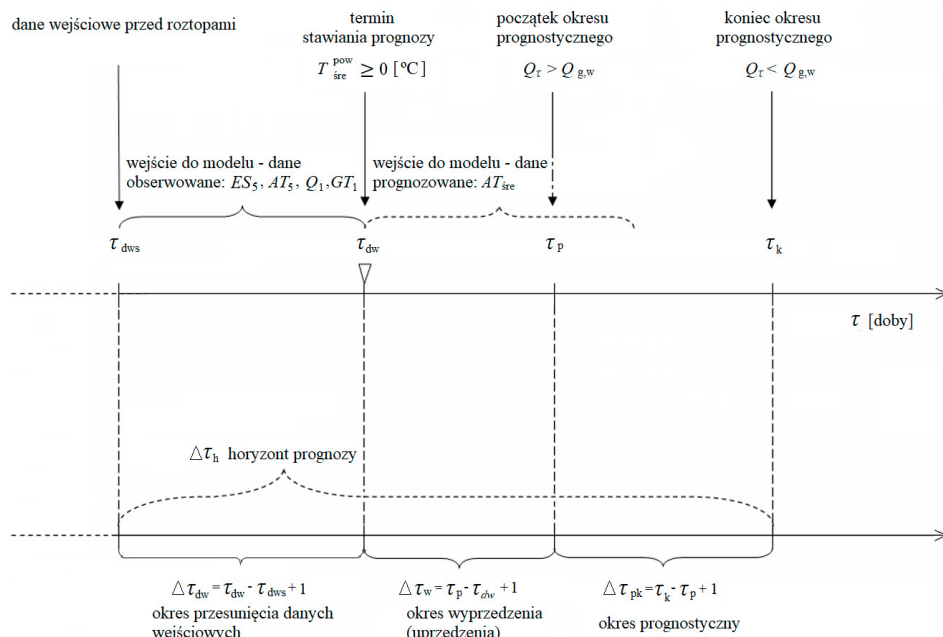
$$V_w = \begin{cases} \int_{\tau_p}^{\tau_k} (Q_\tau - Q_{g,w}) d\tau & Q_\tau > Q_{g,w} \quad \tau \in \{\tau_p, \dots, \tau_k\} \\ \int_{\tau_p}^{\tau_{k1}} (Q_\tau - Q_{g,w}) d\tau + \int_{\tau_{p1}}^{\tau_k} (Q_\tau - Q_{g,w}) d\tau & \text{gd}y \quad Q_\tau > Q_{g,w} \quad \tau \in \{\tau_p, \dots, \tau_{k1}\} \cup \tau \in \{\tau_{p1}, \dots, \tau_k\} \\ SSQ \leq Q_\tau < Q_{g,w} \cap \tau \in \{\tau_{k1}, \dots, \tau_{p1}\} \cap \tau_{p1} - \tau_{k1} + 1 < 5 \text{ d}ó\text{b} \end{cases} \quad (2.9)$$

gdzie: τ_p , τ_{k1} , τ_{p1} , τ_k – odpowiednio początek i koniec wezbrań cząstkowych; Q_τ – przepływ w chwili τ .

ANALIZA ZBIORÓW ZMIENNYCH PROGNOZOWANYCH I PROGNOZUJĄCYCH

Modele wezbrania deszczowo-roztopowego prognozują Q_{\max} , V_w (reprezentującą całą objętość wezbrania, gdy $Q_\tau > Q_{g,w}$) oraz czas trwania wezbrania T_w , wyznaczonymi zgodnie ze wzorami (2.8-2.9) uwzględniającymi kryterium rozdziału sąsiadujących wezbrań. Na rysunku 2.4 przedstawiono schemat elementów średnioterminowej prognozy hydrologicznej wezbrania deszczowo-roztopowego w zlewni Sokołdy w profilu wodowskazowym Sokołda. Początek roztopów, τ_{dw} , jest zbieżny z momentem, w którym średnia dobową temperatura powietrza osiągnęła lub przekroczyła $0^\circ C$. Okres przesunięcia danych wejściowych, $\Delta\tau_{dw}$, jest różnicą między terminem stawiania prognozy, τ_{dws} , a momentem, w którym wyznaczane są zaobserwowane dane wejściowe do modelu, tj. τ_{dws} . W opisywanym modelu $\Delta\tau_{dw} = 5$ dób (przyjęty na podstawie wykonanych eksperymentów obliczeniowych). Okres wyprzedzenia prognozy (uprzedzenia), $\Delta\tau_w$, to czas jaki upływa od momentu stawiania prognozy, τ_{dws} , do momentu, w którym zaobserwowany przepływ przekroczy graniczną wartość $Q_{g,w}$, τ_p . Powyższy okres zależy od początkowej retencji koryta rzeki i dynamiki roztopów (głównie warunków termicznych, opadów deszczu, siły

wiatru) i zmienia się od jednej do kilku dób. Następnym elementem występującym w schemacie średnioterminowej prognozy hydrologicznej (rys. 2.4) jest okres prognostyczny, $\Delta\tau_{pk}$, wyznaczany jako różnica między zakończeniem okresu prognostycznego, τ_k , tj. w dniu w którym zaobserwowany przepływ będzie niższy od wartości granicznej $Q_{g,w}$, a początkiem okresu prognostycznego, τ_p . Okres prognostyczny odpowiada czasowi trwania wezbrania i podlega prognozie. Wartość okresu prognostycznego jest uzależniona od czynników wywołujących wezbranie i jego przebieg. Ostatnim elementem schematu jest horyzont prognozy, $\Delta\tau_h$, obejmujący czas od momentu wejścia danych zaobserwowanych do modelu, τ_{dws} , do końca okresu prognostycznego, τ_k .



Rys. 2.4. Schemat przedstawiający elementy średnioterminowej prognozy hydrologicznej wezbrania deszczowo-roztopowego w zlewni Sokołdy w profilu wodowskazowym Sokołda.

Fig. 2.4. Diagram of the elements of the snowmelt flood medium-term forecast in the Sokołda River watershed as monitored at the Sokołda gauging site.

Wejściem do modeli SLMB jest zbiór zmiennych prognozujących w postaci bieżących lub prognozowanych warunków hydrologicznych i meteorologicznych w zlewni, dla której przygotowywana jest prognoza (rys. 2.4). Wejście do modeli podzielono na dwie części:

- 1) Pierwsza część składa się z następujących danych: wodnego ekwiwalentu śniegu z uwzględnieniem sumy opadów z 5 dób poprzedzających roztopy – ES_5 (10^6 m³); średniej dobowej temperatury powietrza w ciągu 5 dób przed roztopami – AT_5 (°C); zaobserwowanego przepływu w momencie początku roztopów – Q_1 (m³s⁻¹); średniej dobowej temperatury gruntu na głębokości 0,50 m na początku roztopów – GT_1 (°C); oszacowanych w czasie bezpośrednio poprzedzającym roztopy.
- 2) Teoretycznie druga część składa się z danych prognozowanych, czyli 5-dniowej prognozy średniej dobowej temperatury powietrza opracowanej w dniu początku roztopów – AT_{sre} (°C). W opracowanym przykładzie wejściem do modeli zamiast meteorologicznych danych prognozowanych były zaobserwowane dane historyczne.

2.3.3. Analiza statystyczna danych wejściowych do modeli SLMB

Analiza statystyczna danych wejściowych do modeli SLMB obejmuje:

- I. Badanie jednorodności statystycznej zmiennych prognozowanych i prognozujących.

II. Wyznaczanie charakterystyk rozkładu prawdopodobieństwa.

III. Sprawdzanie normalności rozkładu zmiennych prognozowanych i prognozujących.

I. BADANIE JEDNORODNOŚCI STATYSTYCZNEJ ZMIENNYCH PROGNOZOWANYCH I PROGNOZUJĄCYCH

Zbadano jednorodność genetyczną, czasową i pomiarową. Poddano wizualnej ocenie przebiegi ciągów pomiarowych w celu wykrycia potencjalnych punktów niejednorodności. Na podstawie informacji zawartych w rocznych opracowaniach stanów wody i przepływów oraz po wizualnej ocenie realizacji badanego ciągu pomiarowego nie stwierdzono istotnych czynników mogących wpłynąć na jednorodność genetyczną badanych ciągów w latach 1966–2002. Badając jednorodność genetyczną nie jesteśmy w stanie uwzględnić wszystkich czynników mogących mieć wpływ na jednorodność ciągu pomiarowego. W związku z tym do analizy wykorzystano nieparametryczne testy istotności przydatne do badania jednorodności hydrologicznych ciągów pomiarowych podlegających rozkładowi asymetrycznym [Ozga-Zieliński 1999; Zasady... 2001; Guidelines... 2005]. Analizę niejednorodności wykrywanej metodami statystycznymi wykonano zgodnie ze schematem procedury dla zmiennych niezależnych (lewa strona na rys. 2.1).

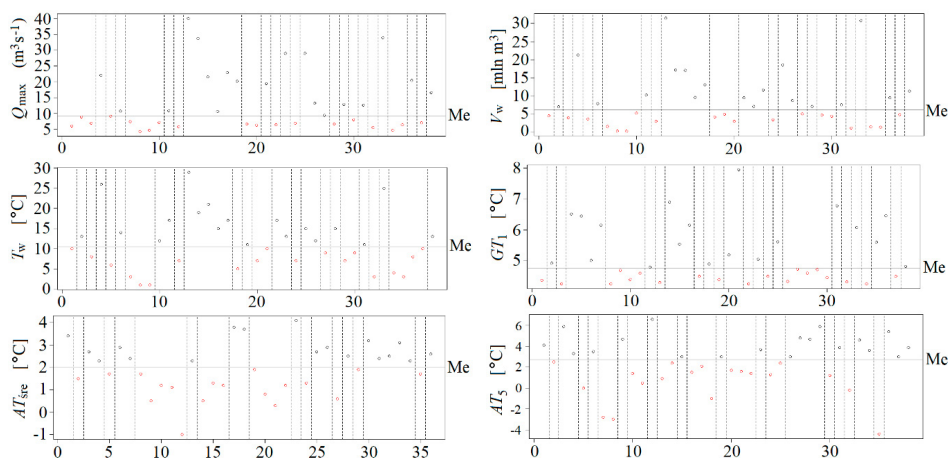
Tabela 2.1. Analiza niejednorodności ciągów pomiarowych w zlewni Sokołdy w profilu wodowskazowym Sokołda przy poziomie istotności $\alpha = 0,05$.

Table 2.1. Nonhomogeneity analysis of the variables in Sokołda River watershed at Sokołda site at significance level $\alpha = 0.05$.

	Q_{max} (m ³ s ⁻¹)	V_w (10 ⁶ m ³)	T_w (doby)	ES_5 (10 ⁶ m ³)	Q_1 (m ³ s ⁻¹)	GT_1 (°C)	AT_{5re} (°C)	AT_5 (°C)
Wykrywanie elementów odstających TEST GRUBBSA I BECKA								
Występują (+)	(-)	(+)	(+)	(-)	(-)	(-)	(-)	(-)
Elementy odstające		[8]~0,37 [9]~0,41	[8]~1,0 [9]~1,0					
Sprawdzenie niezależności TEST SERII I WSPÓŁCZYNNIK AUTOKOREKACJA (max. opóźnienie Lag = 8 dób)								
Liczba obserwacji powyżej/poniżej Me	19/19	19/19	19/19	19/19	19/19	19/19	18/18	19/19
p-wartość	0,188	0,188	0,188	0,048	0,188	0,188	0,102	0,324
Ocena zmiennej	niezależna	niezależna	niezależna	niezależna	niezależna	niezależna	niezależna	niezależna
Sprawdzenie stacjonarności niezależnej zm. losowej TEST SUMY RANG KRUSKALA-WALLISA								
p-wartość	0,977	0,945	0,456	0,872	0,907	0,189	0,447	0,313
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST NA TREND WARTOŚCI ŚREDNIEJ								
p-wartość	0,637	0,772	0,610	0,819	0,915	0,156	0,347	0,433
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST NA TREND WARIANCJI								
p-wartość	0,792	0,418	0,359	0,860	0,823	0,296	0,660	0,839
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna

Wyniki analizy niejednorodności zmiennych prognozowanych i prognozujących przedstawiono w tabeli 2.1. Dla V_w dwa elementy (nr 8 = $0,37 \times 10^6$ m³ i nr 9 = $0,41 \times 10^6$ m³) oraz dla T_w dwa elementy (nr 8 = 1 doba i nr 9 = 1 doba) okazały się być wartościami odstającymi. Zmienne losowe V_w i T_w podlegają rozkładowi gamma; np. parametry rozkładu gamma dla V_w są równe $\alpha = 4,448$ i $\lambda = 1,492$. Do zweryfikowania hipotezy o zgodności empirycznego rozkładu prawdopodobieństwa z hipotetycznym zastosowano test dobroci dopasowania χ^2 -Pearsona przy poziomie istotności $\alpha = 0,05$. Obliczona wartość testu $\chi^2 = 2,975$ była niższa od wartości krytycznej $\chi_{kr,\alpha}^2 = 7,815$. Zgodnie z twierdzeniem

Neymana i Scotta [1971], dowodzącym że rodziny rozkładów log-normalnego, gamma i Weibulla są całkowicie podatne na elementy odstające w próbie losowej, wartości elementów odstających należy uznać za całkowicie naturalne i takie elementy nie powinny być odrzucane z ciągu pomiarowego. Na rysunku 2.5 przedstawiono przykłady wykresów ciągów pomiarowych wraz z linią referencyjną równą medianie. Liczba zaobserwowanych realizacji badanych zmiennych losowych jest jednakowa zarówno powyżej, jak i poniżej mediany, co jest cechą charakterystyczną niezależności ciągów. W teście serii prawdopodobieństwo testowe (p -wartość) było wyższe od przyjętego w badaniu poziomu istotności, $\alpha = 0,05$, zatem nie stwierdzono podstaw do odrzucenia hipotezy zerowej, H_0 , o niezależności badanej zmiennej losowej. Wszystkie analizowane próby zostały uznane za niezależne i stacjonarne, nie wykryto skoku wartości średniej, trendu wartości średniej i wariancji. Prawdopodobieństwa testowe (p -wartość) występujące w procedurze sprawdzania stacjonarności zmiennych niezależnych (tab. 2.1) były wyższe od przyjętego w badaniu poziomu istotności, $\alpha = 0,05$. Nie było powodu, aby odrzucać H_0 w odniesieniu do jednorodności danych.



Rys. 2.5. Wykres weryfikujący liczbę obserwacji powyżej i poniżej mediany, jak również liczbę przebiegów w górę i w dół dla zmiennych: Q_{\max} , V_w , T_w , GT_1 , AT_{sre} i AT_5 za pomocą testu serii.

Fig. 2.5. The chart that verifies the runs above and below the median as well the runs up and down for variables: Q_{\max} , V_w , T_w , GT_1 , AT_{sre} and AT_5 using runs test.

II. WYZNACZANIE CHARAKTERYSTYK ROZKŁADU PRAWDOPODOBIEŃSTWA

W kolejnym etapie analizy danych wejściowych do modeli SLMB [Ciupak 2004] zestawiono próbę losową zmiennych: Q_{\max} , V_w , T_w , ES_5 , Q_1 , GT_1 , AT_{sre} i AT_5 o liczebności $n = 38$. Dla każdej zmiennej losowej określono charakterystyki rozkładu [Węglarczyk 2010]: asymetrii, koncentracji, rozproszenia i położenia (tab. 2.2). Obliczanie miar rozproszenia rozkładu i tym samym współczynnika skośności oraz spłaszczenia uwzględniały stosowaną standardowo formułę dla niezależnych zmiennych losowych (2.3). Dodatkowo wartości współczynników skośności zmiennych: Q_{\max} , V_w , T_w , ES_5 , Q_1 i GT_1 oznaczają, że rozkłady tych zmiennych są prawostronnie asymetryczne (prawy ogon rozkładów tych zmiennych jest grubszy od lewego), natomiast AT_{sre} i AT_5 z ujemnymi skośnościami cechuje asymetria lewostronna. Warto zauważyć, że skośność w zakresie $\pm 1,5$ wskazuje, że rozkłady są zbliżone do rozkładu normalnego, z nieco wydłużonym odpowiednio prawym lub lewym ramieniem rozkładu. Skośność zmiennej losowej ($V_w + 1,573$) w niewielkim stopniu przewyższa wartość graniczną $+1,5$, można zatem uznać, że jej rozkład jest zbliżony do rozkładu normalnego. Niewielkie dodatnie lub ujemne wartości współczynnika spłaszczenia, mieszczące się w przedziale $<-3,0; 3,0>$, zaobserwowane dla wszystkich badanych zmiennych losowych, świadczą o bliskości ich rozkładów w stosunku do rozkładu normalnego, choć są one bardziej smukłe.

Tabela 2.2. Charakterystyki rozkładu prawdopodobieństwa zmiennych prognozowanych i prognozujących w modelowanej zlewni Sokółdy w przekroju wodowskazowym Sokółda.

Table 2.2. Statistical measures of distribution probability for the predictands and predictors for the modeled Sokółda River watershed at Sokółda site.

	Q_{\max} (m^3s^{-1})	V_w (10^6m^3)	T_w (doby)	ES_5 (10^6m^3)	Q_1 (m^3s^{-1})	GT_1 ($^{\circ}\text{C}$)	$AT_{\text{śre}}$ ($^{\circ}\text{C}$)	AT_5 ($^{\circ}\text{C}$)
Miara asymetrii rozkładu								
Współczynnik skośności	1,172	1,573	0,674	1,303	1,110	1,760	-0,311	-0,747
Miara koncentracji rozkładu								
Współczynnik spłaszczenia	0,281	2,155	0,033	1,800	0,306	2,893	-0,025	0,361
Miary rozproszenia rozkładu								
Wariancja	91,041 [(m^3s^{-1}) ²]	55,704 [(10^6m^3) ²]	45,326 [(doby) ²]	203,591 [(10^6m^3) ²]	0,878 [(m^3s^{-1}) ²]	0,489 [($^{\circ}\text{C}$) ²]	1,168 [($^{\circ}\text{C}$) ²]	6,090 [($^{\circ}\text{C}$) ²]
Miary położenia rozkładu								
Średnia arytmetyczna	13,514 (m^3s^{-1})	8,414 (10^6m^3)	11,395 (doby)	27,421 (10^6m^3)	5,143 (m^3s^{-1})	0,329 ($^{\circ}\text{C}$)	1,979 ($^{\circ}\text{C}$)	2,371 ($^{\circ}\text{C}$)
Mediana	9,235 (m^3s^{-1})	6,175 (10^6m^3)	10,500 (doby)	24,695 (10^6m^3)	4,770 (m^3s^{-1})	0,150 ($^{\circ}\text{C}$)	2,000 ($^{\circ}\text{C}$)	2,750 ($^{\circ}\text{C}$)

III. SPRAWDZANIE NORMALNOŚCI ROZKŁADU ZMIENNYCH PROGNOZOWANYCH I PROGNOZUJĄCYCH

Dla zmiennych Q_{\max} , V_w , T_w , ES_5 i Q_1 wykonano estymację dolnego ograniczenia rozkładu prawdopodobieństwa, tj. parametru ε . Do obliczeń wykorzystano zależność wynikającą z metody kwantyli [Strupczewski 1967; Ozga-Zielińska, Brzeziński 1997]. Normalizację zmiennych Q_{\max} , V_w , T_w , ES_5 i Q_1 przeprowadzono wykorzystując funkcję logarytmiczną [$\ln(X - \varepsilon)$], gdzie X jest nieprzekształconą zmienną losową. Następnie dla wszystkich parametrów modelu, w tym nieznormalizowanych i znormalizowanych, przeprowadzono test dobroci dopasowania teoretycznego rozkładu normalnego z rozkładem empirycznym opartym na statystyce Kolmogorowa, λ [Marsaglia i in. 2003]. W teście przyjęto wartość krytyczną statystyki $\lambda_{\text{kr}} = 0,1809$, przy poziomie istotności $\alpha = 0,05$, którą obliczono za pomocą metody bootstrap [Amal 2006]. Najlepsze dopasowanie rozkładu empirycznego do teoretycznego (tab. 2.3) uzyskano dla znormalizowanych zmiennych $\ln(Q_{\max} - 4,202)$, $\ln(V_w)$, $\ln(T_w)$, $\ln(ES_5 - 8,858)$ i $\ln(Q_1 - 4,231)$. W celu uproszczenia zapisu, znormalizowane zmienne oznaczono jako $\ln Q_{\max}$, $\ln V_w$, $\ln T_w$, $\ln ES_5$ i $\ln Q_1$, a nieznormalizowane jako GT_1 , AT_5 i $AT_{\text{śre}}$. Zmienną losową V_w cechowała stosunkowo wysoka wartość współczynnika skośności (+1,573) i spłaszczenia (+2,155) (tab. 2.2), jednak uzyskane wartości mieściły się w granicach wskazujących na ich bliskość z rozkładem normalnym, odpowiednio +1,5 i +3,0. Powyższe wnioskowanie potwierdziło przeprowadzone badanie zgodności z rozkładem normalnym (tab. 2.3). Uzyskana wartość testowa dla V_w , $\lambda = 0,1647$, była niższa od wartości krytycznej $\lambda_{\text{kr}}^{\alpha = 0,05} = 0,1809$.

Mimo, że uzyskano zgodność brzegowych rozkładów prawdopodobieństwa nieznormalizowanych zmiennych losowych V_w , T_w , ES_5 i Q_1 z rozkładem normalnym (tab. 2.3), to do dalszej analizy wybrano ich znormalizowane formy, tj. $\ln V_w$, $\ln T_w$, $\ln ES_5$ i $\ln Q_1$. Wykonano to w celu umożliwienia ich wzajemnego porównywania w kolejnych krokach analizy za pomocą statycznych liniowych modeli Bayesowskich. Na podstawie uzyskanych wyników sprawdzania normalności (tab. 2.3), zbiór danych wejściowych uzupełniono o zmienną $\ln Q_{\max}$.

Tabela 2.3. Badanie zgodności brzegowych rozkładów prawdopodobieństwa zmiennych losowych i zmiennych, znormalizowanych z parametrem ε (dolnym ograniczeniem rozkładu prawdopodobieństwa), z rozkładem normalnym z zastosowaniem testu λ -Kolmogorowa z wartością krytyczną przy poziomie istotności $\alpha = 0,05$.

Table 2.3. Verification of normality of marginal distribution for unnormalized and normalized variables, with ε parameter estimated (lower limit in the probability distribution), using Kolmogorov goodness-of-fit test with the critical value at significance level $\alpha = 0.05$.

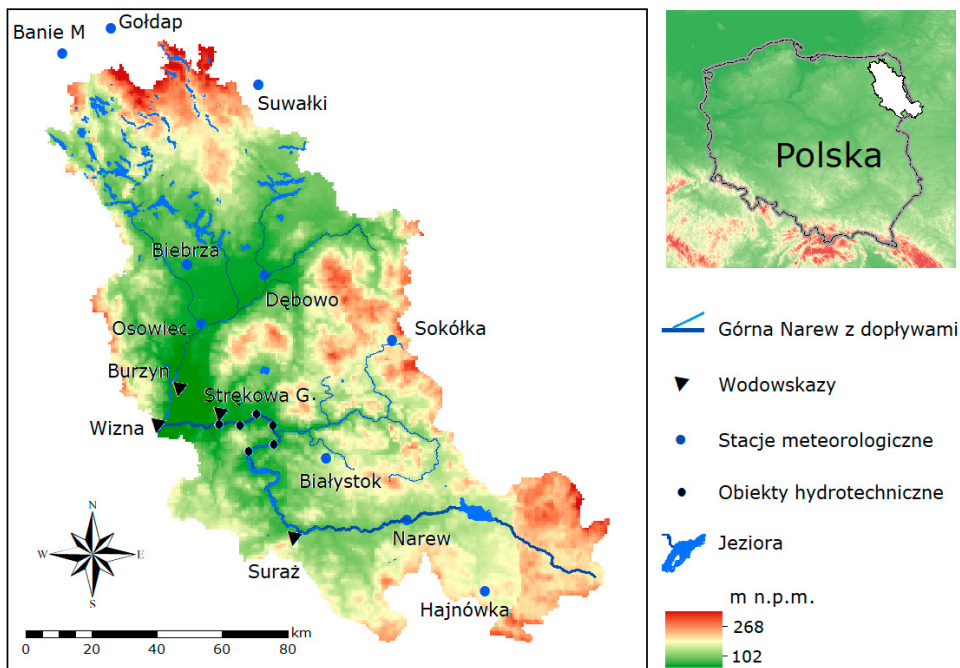
	Q_{\max}	$\ln(Q_{\max} - \varepsilon)$	V_w	$\ln(V_w - \varepsilon)$	T_w	$\ln(T_w - \varepsilon)$	
ε		4,202		0,0		0,0	
$\lambda_{kr.}^{\alpha=0,05} = 0,1809$	0,2558	0,1241	0,1647	0,1512	0,1661	0,0830	
	ES_5	$\ln(ES_5 - \varepsilon)$	Q_1	$\ln(Q_1 - \varepsilon)$	GT_1	AT_{5se}	AT_5
ε	0,0	8,858		4,231			
$\lambda_{kr.}^{\alpha=0,05} = 0,1809$	0,1473	0,1220	0,2051	0,1093	0,2314	0,0934	0,0921

2.4. Dane wejściowe do dynamicznych liniowych modeli Bayesowskich (DLMB)

2.4.1. Opis hydrologii zlewni Górnej Narwi

Zlewnia Górnej Narwi leży na Wysoczyźnie Podlasko-Białoruskiej, w północno-wschodnim rejonie Polski (rys. 2.6). Źródła rzeki znajdują się około 30 km poza granicami Polski, na wschodnim skraju Puszczy Białowieskiej na terenie Białorusi. Jest to obszar w znacznym stopniu zmeliorowany o powierzchni 605,8 km². Na rzece znajduje się zbiornik wodny Siemianówka wybudowany w latach 1977-1988. Zlewnia charakteryzuje się rozległą doliną, w części północnej podmokłą, a w południowej całkowicie zalesioną. Między Surażem a Wizną rzekę charakteryzuje silnie rozbudowany system koryt. Na tym odcinku występują liczne starorzecza i rowy, rzeka płynie wieloma ramionami. Dno doliny pokrywają torfy. W 1985 roku utworzono Narwiański Park Krajobrazowy, który w 1996 r. przekształcono w Narwiański Park Narodowy (NPN). W latach 1990-2012 zrealizowano projekt renaturalizacji strefy buforowej NPN, który udroźnił zarośnięte starorzecza w strefie buforowej, połączył je z siecią rzeczną oraz uwodnił przesuszone tereny. Zlewnia od ujścia Biebrzy do profilu wodowskazowego Wizna obejmuje szerokie, zatorfione i podmokłe dno doliny Narwi, tzw. Bagno Wizna oraz część Wysoczyzny Kolneńskiej porożcinanej licznymi dolinami. Powierzchnia zlewni Górnej Narwi do wodowskazu Wizna bez dorzecza Biebrzy wynosi 7243,5 km²; po uwzględnieniu Biebrzy powierzchnia zwiększa się do 14308,0 km².

Do wodowskazu Wizna zlewnia charakteryzuje się wyraźną asymetrią przyrostu dorzecza, z bardziej rozwiniętą siecią prawostronną (rys. 2.6). Ze względu na charakter doliny Górnej Narwi, w szczególności na hydrologię, hydraulikę i morfologię koryta rzeki, prognozowanie hydrologiczne jest istotnym problemem. Odcinek Narwi od Rzędzian do Strękowej Góry jest uregulowany, znajduje się tam sześć jazów: Rzędziany, Babino, Złotoria, Góra, Tykocin i Strękowa Góra. Główny zadaniem budowli hydrotechnicznych jest nawadnianie w okresie wegetacyjnym (rys. 2.6). Morfologia gleby w węzle hydrologicznym Wizna, Strękowa Góra i Burzyn komplikuje system połączenia wód powierzchniowych i gruntowych. W strefie depozycji Biebrzy do Narwi na terenie Wizna ujęcie wody składa się z obszarów torfowych, podmokłych i bagien. W czasie marcowo-kwietniowych wezbrań występują na tym obszarze naturalne rozlewiska, co zwiększa ryzyko wystąpienia powodzi w pobliskich wsiach, głównie w rejonie węzła hydrologicznego (wlot Biebrzy do Narwi) w Wiźnie. Ze względu na brak znajomości hydromorfologii zlewni, zastosowanie modeli hydrologicznych typu czarna skrzynka (black box) może być obiecującym rozwiązaniem, szczególnie w zastosowaniach prognostycznych.



Rys. 2.6. Mapa hydrograficzna zlewni Górnej Narwi do profilu wodowskazowego Wizna.
 Fig. 2.6. Hydrographic map of the Upper Narew River watershed at Wizna site.

2.4.2. Opracowanie danych wejściowych do modeli DLMB

IDENTYFIKACJA DANYCH WEJŚCIOWYCH DO MODELI

Wejściami do modeli dynamicznych były, prowadzone w regularnych odstępach czasu (na przykład raz na dobę, co godzinę), obserwacje lub pomiary hydrologiczno-meteorologiczne wpływające bezpośrednio lub pośrednio na hydrogram przepływu w zlewni Górnej Narwi w profilu wodowskazowym Wizna. Potencjalny zbiór zmiennych prognozujących utworzyły zaobserwowane hydrogramy w Strękowej Górze i Wiźnie na rzece Narew oraz w Burzynie i Osowcu na Biebrzy. Ponieważ czas dobiegu fali wezbraniowej ze Strękowej Góry i Burzyna do wodowskazu Wizna trwa około dwóch dób, a od wodowskazu w Osowcu około trzech dób, wzięto pod uwagę obserwacje dwie i trzy doby przed terminem postawienia prognozy. Analizowano następujące charakterystyki zmiennych meteorologicznych:

- 1) Sumę dobową opadu, rodzaj opadu (deszcz, śnieg).
- 2) Grubość śniegu, ekwiwalent wodny śniegu.
- 3) Średnią, maksymalną i minimalną dobową temperaturę powietrza, które wpływają na dynamikę drenażu zlewni w skali roku hydrologicznego.

Dane meteorologiczne pozyskano z następujących posterunków meteorologicznych: Banie M, Gołdap, Suwałki, Osowiec, Biebrza, Dębowo, Białystok, i Narew. Lokalizacje tych posterunków przedstawiono na rysunku 2.6.

Codzienny przepływ w zlewni Górnej Narwi w Wiźnie, Q_t^{WIZ} (m^3s^{-1}) (rys. 2.6), uzależniono od wektora zmiennych prognozujących (Q_{t-2}^{STR} , Q_{t-2}^{BUR} , P_{t-2}^{AREA} , T_{t-2}^{MAX} , T_{t-2}^{MIN}), gdzie: Q_{t-2}^{STR} i Q_{t-2}^{BUR} (m^3s^{-1}) – przepływy w profilach wodowskazowych Strękowa Góra i Burzyn zaobserwowane dwie doby wcześniej w stosunku do odtwarzanego hydrogramu w Wiźnie; P_{t-2}^{AREA} ($10^6 m^6$) – średnia wysokość warstwy opadu w zlewni Górnej Narwi powyżej wodowskazu Wizna zaobserwowana dwie doby wcześniej i wy-

znaczona metodą wielokątów równego zadeszczenia; $T_{\tau-2}^{\text{MAX}}, T_{\tau-2}^{\text{MIN}}$ ($^{\circ}\text{C}$) – średnie dobowe wartości maksymalnej i minimalnej temperatury powietrza zaobserwowane dwie doby wcześniej na obszarze modelowanej zlewni.

Wszystkie dane pomiarowe obejmowały okres od 1 listopada 1989 r. do 31 października 1996 r. i zostały uznane za potencjalne dane wejściowe do dynamicznych liniowych modeli Bayesowskich (DLBM) oraz matematycznych modeli wielokrotnej liniowej regresji (WLR) i sieci neuronowych (NN) zastosowanych do porównania efektywności DLMB. O ostatecznym składzie i zakresie zbiorów wejściowych do modelowania matematycznego decyduje analiza statystyczna potencjalnych danych wejściowych do modeli DLMB.

2.4.3. Analiza statystyczna danych wejściowych do modeli DLMB

Analiza statystyczna danych wejściowych do modeli DLMB obejmuje:

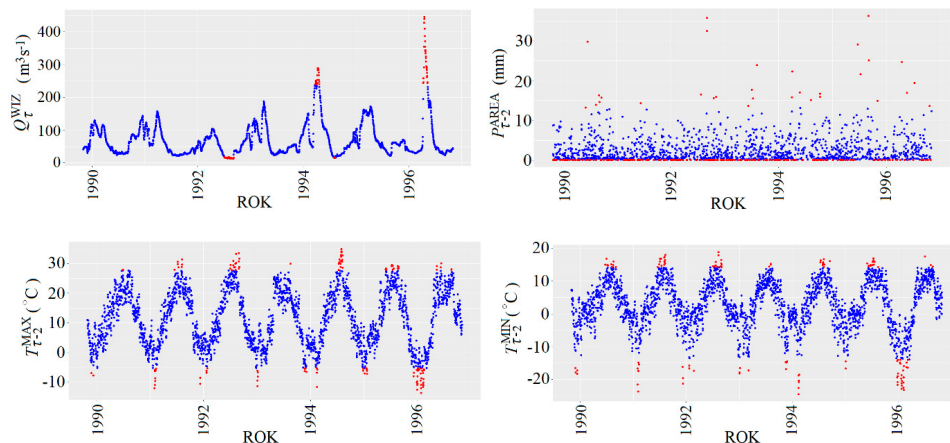
- I. Badanie jednorodności statystycznej danych wejściowych do modeli DLMB.
- II. Wyznaczanie charakterystyk rozkładu prawdopodobieństwa.

I. BADANIE JEDNORODNOŚCI STATYSTYCZNEJ DANYCH WEJŚCIOWYCH DO MODELI DLMB

Analizę niejednorodności wykrywanej metodami statystycznymi wykonano zgodnie ze schematem procedury (rys. 2.1). Wyniki analizy niejednorodności danych wejściowych do modeli DLMB przedstawiono w tabeli 2.4.

Identyfikacja obserwacji odstających na podstawie metody SVM

Obserwacje odstające mogą w istotny sposób zaburzać informację o rodzaju i sile występowania trendu. Takie obserwacje mogą być wynikiem błędnego pomiaru lub pomyłki przy wprowadzaniu pomiaru lub obserwacji do bazy danych. Do identyfikacji obserwacji odstających w ciągach pomiarowych $Q_{\tau}^{\text{WIZ}}, Q_{\tau-2}^{\text{STR}}, Q_{\tau-2}^{\text{BUR}}, P_{\tau-2}^{\text{AREA}}, T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$, każdy o liczebności $n = 2557$, zastosowano metodę wektorów nośnych [Hawkins 1980; Ben-Hur i in. 2001]. Metoda SVM nie nakłada niemal żadnych założeń na badany ciąg pomiarowy oraz wykrywa głównie pojedyncze obserwacje odstające. Przykład wykrytych obserwacji odstających dla $Q_{\tau}^{\text{WIZ}}, P_{\tau-2}^{\text{AREA}}, T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$ pokazano na rysunku 2.7.



Rys. 2.7. Identyfikacja obserwacji odstających w ciągach pomiarowych $Q_{\tau}^{\text{WIZ}}, P_{\tau-2}^{\text{AREA}}, T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$ za pomocą metody SVM.

Fig. 2.7. Identification of outliers in measurement data $Q_{\tau}^{\text{WIZ}}, P_{\tau-2}^{\text{AREA}}, T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$ using SVM method.

Tabela 2.4. Analiza niejednorodności ciągów pomiarowych w zlewni górnej Narwi w profilu wodowskazowym Wizna przy poziomie istotności $\alpha = 0,05$. Pogrubione wartości wskazują na zależność lub/i niestacjonarność próby losowej oraz na istotność trendu.

Table 2.4. Nonhomogeneity analysis of the variables in Upper Narew River watershed at Wizna site at significance level $\alpha = 0.05$. The bold values indicate on the dependance or/and non-stationarity of the random sample and the significance of the trend.

	Q_{τ}^{WIZ} (m^3s^{-1})	$Q_{\tau-2}^{STR}$ (m^3s^{-1})	$Q_{\tau-2}^{BUR}$ (m^3s^{-1})	$p_{\tau-2}^{AREA}$ ($10^6 m^3$)	$T_{\tau-2}^{MAX}$ ($^{\circ}C$)	$T_{\tau-2}^{MIN}$ ($^{\circ}C$)
Wykrywanie elementów odstających METODA WEKTORÓW NOŚNYCH, SVM						
Występują (+)	(+)	(+)	(+)	(+)	(+)	(+)
Odstające/nieodstające	130/2427	118/2439	127/2430	49/2508	133/2424	125/2432
Sprawdzenie niezależności TEST SERII I WSPÓŁCZYNNIK AUTOKORELACJI (max. opóźnienie Lag = 8 dób)						
Liczba obserwacji powyżej/poniżej M_e	1277/1278	1274/1278	1278/1269	1201/1266	1278/1274	1270/1271
p -wartość	0,000	0,000	0,000	0,000	0,000	0,000
Ocena zmiennej	zależna	zależna	zależna	zależna	zależna	zależna
Identyfikacja trendu oraz ocena funkcji trendu TEST WSPÓŁCZYNNIKA KORELACJI RANG SPEARMANA						
	-0,0023	0,0359	-0,0212	-0,0481	0,0485	0,0388
TEST MANNA-KENDALLA						
Wartość testu	-0,0021	0,0212	-0,0110	-0,0337	0,0314	0,0257
p -wartość	0,872	0,108	0,405	0,016	0,017	0,052
Identyfikacja występowania i wpływu wahań okresowych TEST AUTOKORELACJI (opóźnienie Lag = 365 dób)						
Występują (+)	(+)	(+)	(+)	(+)	(+)	(+)
Sprawdzenie stacjonarności z uwzględnieniem trendu lub braku trendu TEST DICKEY'A-FULLERA, ADF						
Statystyka ADF $ADF_{kr,\alpha=0,05} = -3,41$ (z trendem) $ADF_{kr,\alpha=0,05} = -2,89$ (bez trendu)	-5,003	-6,011	-5,225	-11,920	-3,258	-4,090
Lag	13	13	13	13	13	13
p -wartość	0,010	0,010	0,010	0,010	0,076	0,010
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	niestacjonarna	stacjonarna
TEST KWIATKOWSKIEGO-PHILLIPSA-SHMIDTA-SHINA, KPSS						
Statystyka KPSS $KPSS_{kr,\alpha=0,05} = 0,463$	0,622	0,718	0,672	0,884	0,193	0,469
Lag	8	8	8	8	8	8
p -wartość	0,021	0,012	0,016	0,010	0,100	0,010
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	niestacjonarna	stacjonarna

Na wykresach (rys. 2.7) kolorem czerwonym zaznaczono obserwacje uznane przez metodę SVM jako odstające. W celu wyeliminowania lub pozostawienia w ciągu liczbowym zidentyfikowanych obserwacji odstających, poddano wizualnej ocenie przebiegi ciągów pomiarowych. Na podstawie informacji zawartych w opracowaniach rocznikowych (rocznikach meteorologicznych, rocznikach hydrologicznych wód powierzchniowych, wód podziemnych, wynikach pomiarów hydrometrycznych, rocznikach opadów atmosferycznych, i innych opracowaniach takich jak: biuletyny hydro-meteorologiczne i operaty hydrologiczne zawierające materiał potrzebny do identyfikacji jednorodnych zjawisk w badanym profilu wodowskazowym) nie stwierdzono istotnych czynników mogących wpłynąć na jednorodność badanych ciągów w latach 1990-1996. Zidentyfikowane obserwacje odstające codziennych przepływów w profilu wodowskazowym Wizna na Narwi dotyczyły zdarzeń związanych z wezbraniem deszczowo-roztopowymi w latach 1994 i 1996 oraz niżówką latem 1992 roku. Fala wezbraniowa w badanym profilu wodowskazowym przemieszczała się wzdłuż Narwi i Biebrzy od marca 1994 do kwietnia 1994 r. oraz od kwietnia 1996 do początku maja 1996 roku. W analizowanym okresie, tj. w latach 1990-1996, maksymalne przepływy w Wiźnie ($Q_{\tau}^{WIZ} = 446 \text{ m}^3\text{s}^{-1}$) zarejestrowano w dniach 16-17 kwietnia 1996 roku. Natomiast 2 września 1992 r. wystąpił jeden z najniższych przepływów ($Q_{\tau}^{WIZ} = 12,9 \text{ m}^3\text{s}^{-1}$) w całej historii prowadzenia obserwacji hydrometrycznych w tym profilu, kończący okres niedoboru opadów w zlewni Górnej Narwi trwający od maja do września 1992 roku.

Metoda SVM zakwalifikowała brak opadów jako obserwacje odstające. Po przeanalizowaniu roczników meteorologicznych uznano, że zarejestrowane braki opadów są uzasadnione zaistniałą sytuacją synoptyczną na obszarze badanej zlewni. Maksymalne sumy dobowe opadu w zlewni Górnej Narwi w latach 1990-1996 wystąpiły w półroczu letnim, tj. od maja do października, co było związane z niestabilnością konwekcyjną występującą w atmosferze. Maksymalny dobowy opad uśredniony na powierzchnię zlewni, $P_{\tau-2}^{AREA} = 36,3 \text{ mm}$, zanotowano 3 września 1995 roku. Najwyższą średnią dobową temperaturę powietrza na obszarze zlewni Górnej Narwi, $T_{\tau-2}^{MAX} = 34,9^{\circ}\text{C}$, zarejestrowano 1 sierpnia 1994 r., a najniższą, $T_{\tau-2}^{MIN} = -24,5^{\circ}\text{C}$, 15 lutego 1994 roku. Najniższe temperatury powietrza obserwowano w 1991 r. od stycznia do lutego, w latach 1992 i 1993 na przełomie grudnia i stycznia oraz w 1994 r. w lutym, ze spadkami temperatury odpowiedni do -21°C , -18°C i $-24,5^{\circ}\text{C}$.

Zidentyfikowane na podstawie metody SVM obserwacje odstające pozostawiono w ciągu. Zaobserwowane dane dotyczyły maksymalnych i minimalnych wartości charakterystyk wezbrań roztopowych, deszczowo-roztopowych, niżówek oraz zjawisk meteorologicznych. Stwierdzono, że istnieje hydrologiczne i meteorologiczne uzasadnienie ich wystąpienia w zlewni Górnej Narwi w latach 1990-1996.

Sprawdzanie niezależności badanych zmiennych losowych za pomocą testu serii i testu współczynnika autokorelacji

Sprawdzenie niezależności zmiennych losowych Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$, $T_{\tau-2}^{MAX}$ i $T_{\tau-2}^{MIN}$ przeprowadzono przy użyciu testu serii [Wald, Wolfowitz 1940] i testu współczynnika autokorelacji dla maksymalnego opóźnienia $Lag = 8$ dób, które posłużyły do sprawdzenia wewnętrznej korelacji badanych zmiennych losowych. Test serii weryfikuje liczbę obserwacji powyżej i poniżej wartości mediany oraz wyznacza prawdopodobieństwo testowe, p -wartość. Dla każdej analizowanej zmiennej liczba obserwacji powyżej i poniżej mediany była różna, a wartość testu była mniejsza od przyjętego w badaniu poziomu istotności $\alpha = 0,05$ (tab. 2.4). W takiej sytuacji uznano, że weryfikowaną hipotezę zerową o niezależności badanej zmiennej losowej należy odrzucić, a realizacje wszystkich badanych zmiennych losowych nie mają charakteru losowego – ich dobór do próby podlega tendencji lub cykliczności, co oznacza, że badane zmienne losowe są zależne. Na podstawie tego wniosku wykonano dalszą analizę niejednorodności przy użyciu testów statystycznych dla zależnych zmiennych losowych (prawa strona na rys. 2.1).

Identyfikacja trendu

Analizowano korelację między badanymi zmiennymi losowymi, tj. Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$, $T_{\tau-2}^{MAX}$ i $T_{\tau-2}^{MIN}$, a zmienną określającą czas, τ . W tym celu skorzystano z nieparametrycznego testu korelacji rang Spearmana [Spearman 1904] (tab. 2.4). Ujemne i dodatnie wartości wskazują na występowanie trendu odpowiednio malejącego lub rosnącego. Współczynnik pozwala również określić siłę występującego trendu. Im wartości są bliższe -1 lub $+1$, tym związek badanej zmiennej losowej ze zmienną

określającą czas, τ , jest silniejszy. Współczynnik Spearmana określa dowolną monotoniczną zależność, także nieliniową.

W tabeli 2.4 pogrubione wartości wskazują na istotne korelacje z $p < 0,05$. Istotność wykrytej tendencji zmian w badanych ciągach pomiarowych określono za pomocą nieparametrycznego testu statystycznego Manna-Kendalla [Hamed 2008], polegającego na weryfikacji hipotezy o braku trendu na podstawie nieparametrycznego współczynnika korelacji. Wartość testu, tzw. współczynnik tau Kendalla, opiera się na różnicy między liczbą zgodnych (w tym samym porządku) i niezgodnych par w obrębie obserwowanych danych. Wyniki testu wokół wartości zerowej tego współczynnika świadczą na rzecz hipotezy o braku trendu, duże wartości dodatnie współczynnika przemawiają na korzyść występowania trendu rosnącego, zaś duże wartości ujemne oznaczają występowanie trendu malejącego. Na podstawie obliczonej p -wartości można precyzyjnie określić poziom istotności wykrytego trendu. Test Manna-Kendalla (tab. 2.4) potwierdził istotność występowania trendu malejącego dla zmiennej $p_{\tau-2}^{AREA}$ i rosnącego dla $T_{\tau-2}^{MAX}$, odpowiednio na poziomie istotności 0,016 i 0,017 (p -wartość).

Tabela 2.5. Wyniki analizy regresji prostej, gdzie zmienną zależną jest zmienna prognozowana lub zmienne prognozujące w modelowanej zlewni Górnej Narwi, a zmienną niezależną czas (τ) dla $n = 2557$. Pogrubione wartości wskazują na istotność parametrów regresji, czyli na istnienie istotnego wpływu zmiennej losowej τ na badaną zmienną zależną, z $p < 0,05$.

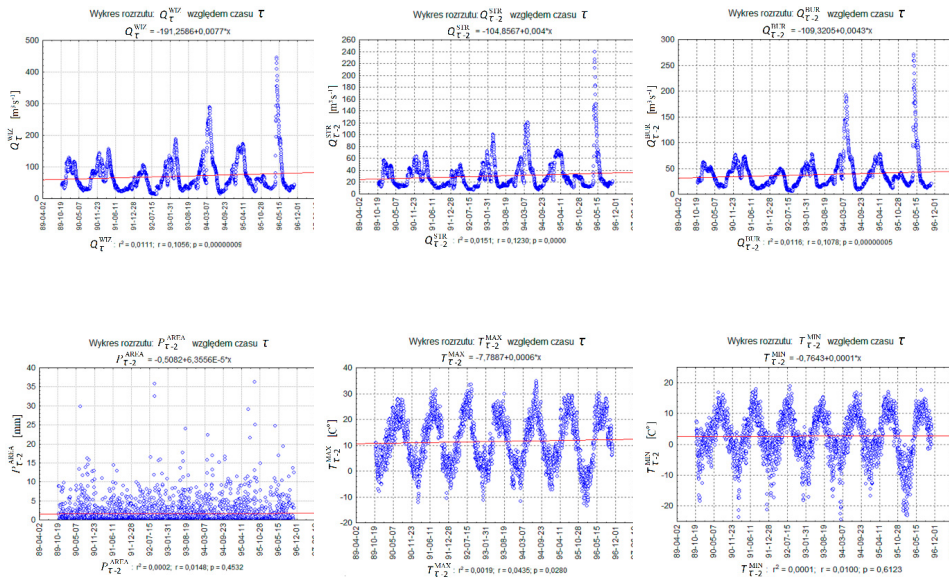
Table 2.5. The results of simple regression analysis, where the dependent variable is the predictand or predictors for the modeled Upper Narew River watershed, and the independent variable is time (τ) for $n = 2557$. Bold values indicate the significance of the regression parameters, *i.e.* the existence of a significant effect of the random variable τ on the dependent variable under study, with $p < 0,05$.

	Beta	Błąd standardowy Beta	β	Błąd standardowy β	$t(n-2)$	Poziom p
Q_{τ}^{WIZ}	R = 0,1056, R² = 0,0111, F(1,2555) = 28,8010 Błąd standardowy estymacji = 53,397					
Wyraz wolny			60,6557	2,1126	28,7119	0,000
τ	0,1056	0,0197	0,0077	0,0014	5,3667	0,000
$Q_{\tau-2}^{STR}$	R = 0,1230, R² = 0,0151, F(1,2555) = 39,256 Błąd standardowy estymacji = 23,652					
Wyraz wolny			25,4151	0,9358	27,1598	0,000
τ	0,1230	0,0196	0,0039	0,0006	6,2654	0,000
$Q_{\tau-2}^{BUR}$	R = 0,1078, R² = 0,0116, F(1,2555) = 30,051 Błąd standardowy estymacji = 29,382					
Wyraz wolny			32,2747	1,1625	27,7640	0,000
τ	0,1078	0,0197	0,0043	0,0008	5,4819	0,000
$p_{\tau-2}^{AREA}$	R = 0,0148, R² = 0,0002, F(1,2555) = 0,5629 Błąd standardowy estymacji = 3,1619					
Wyraz wolny			1,5771	0,1251	12,6074	0,000
τ	0,0148	0,0198	0,0001	0,0001	0,7503	0,453
$T_{\tau-2}^{MAX}$	R = 0,0435, R² = 0,0019, F(1,2555) = 4,8336 Błąd standardowy estymacji = 9,5937					
Wyraz wolny			10,7529	0,3796	28,3303	0,000
τ	0,0434	0,0198	0,0006	0,0003	2,1986	0,028
$T_{\tau-2}^{MIN}$	R = 0,0100, R² = 0,0001, F(1, 2555) = 0,2568 Błąd standardowy estymacji = 7,3767					
Wyraz wolny			2,5220	0,2918	8,6418	0,000
τ	0,0100	0,0198	0,0001	0,0002	0,5068	0,612

Ocena postaci funkcji trendu

Ocenę postaci funkcji trendu sprawdzono za pomocą analizy prostej regresji (tab. 2.5), gdzie zmienną zależną jest badana zmienna losowa, a zmienną niezależną czas, τ . W tabeli 2.5 przedstawiono następujące wyniki regresji: znormalizowane współczynniki regresji (Beta), błąd standardowy znormalizowanych współczynników regresji (Błąd std. Beta), współczynniki regresji (β), błąd standardowy estymacji β (Błąd std. β), test t testujący istotność każdego współczynnika regresji wyrażony w postaci ilorazu $\frac{\beta}{\text{Błąd std.}\beta}$, poziom istotności przyjęty w modelu (p), współczynnik korelacji Pearsona (r) [Pearson 1931], współczynnik determinacji (r^2), błąd standardowy estymacji, globalny test F (test Fishera-Snedecora) [Box 1953] oraz liczebność próby losowej (n).

Globalny test F testuje trzy równoważne hipotezy zerowe: $H_0: \beta_1 = 0$ – istotność współczynnika kierunkowego; $H_0: r^2 = 0$ – istotność współczynnika determinacji; $H_0: \beta_1 x + \beta_0 = y$ – istotność liniowego związku między analizowanymi zmiennymi; gdzie: β_1 jest współczynnikiem kierunkowym, β_0 jest wyrazem wolnym; x i y to odpowiednio zmienna niezależna i zależna. Weryfikowana jest hipoteza zerowa, że zmienna niezależna x (w tab. 2.5 zmienną niezależną jest czas τ) nie wywiera wpływu na badaną zmienną zależną y (w tab. 2.5 zmiennymi zależnymi są: $Q_{\tau-2}^{WIZ}$, $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$, $T_{\tau-2}^{MAX}$ i $T_{\tau-2}^{MIN}$). Jeżeli w trakcie weryfikacji odrzucana jest hipoteza zerowa, to ocena współczynnika regresji jest istotna. Oznacza to, że τ istotnie wpływa na badaną zmienną zależną. W tabeli 2.5 pogrubione wartości wskazują na istnienie istotnego wpływu czasu, τ , na badaną zmienną losową. Trend wykryto dla $Q_{\tau-2}^{WIZ}$, $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$ i $T_{\tau-2}^{MAX}$. Estymowane współczynniki nachylenia prostej regresji β różniły się od zera. Prawdopodobieństwo testowe p , wyznaczone na podstawie wyliczonych współczynników regresji, było niższe od przyjętego w badaniu poziomu istotności $\alpha = 0,05$.



Rys. 2.8. Wykres rozrzutu zależnych zmiennych losowych $Q_{\tau-2}^{WIZ}$, $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$, $T_{\tau-2}^{MAX}$ i $T_{\tau-2}^{MIN}$ względem zmiennej niezależnej czas ($x = \tau$), z zaznaczonym równaniem prostej regresji, współczynnika determinacji (r^2), współczynnika korelacji liniowej Pearsona (r) oraz prawdopodobieństwem testowym (p) porównywanym z przyjętym poziomem istotności $p < 0,05$.

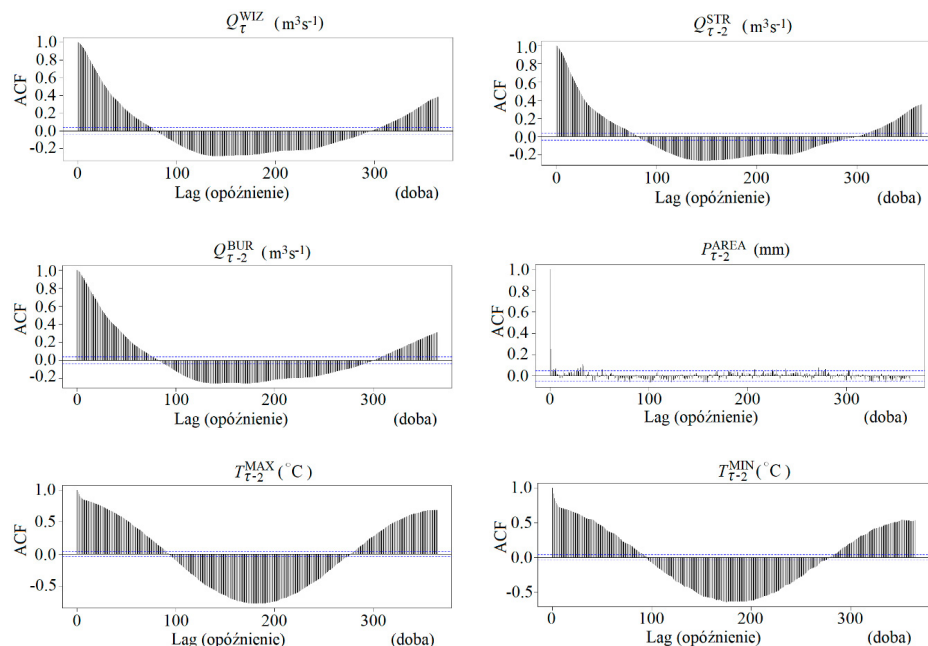
Fig. 2.8. Scatter chart of dependent random variables: $Q_{\tau-2}^{WIZ}$, $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$, $T_{\tau-2}^{MAX}$ and $T_{\tau-2}^{MIN}$ relative to the independent variable time ($x = \tau$), with the equation of the regression line, coefficient of determination (r^2), Pearson's linear coefficient (r) and test probability (p) compared with assumed significance level $p < 0.05$.

Graficzną analizę prostej regresji przedstawiono na rysunku 2.8. Wykresy pokazują niewielką liniową tendencję rosnącą w odniesieniu do $Q_{\tau-2}^{WIZ}$, $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$ i $T_{\tau-2}^{MAX}$ z wyjątkiem zmiennej $P_{\tau-2}^{AREA}$

i $T_{\tau-2}^{\text{MIN}}$, dla których współczynniki kierunkowe są bliskie 0, a prawdopodobieństwo testowe ($p_{\tau-2}^{\text{AREA}}$, $p = 0,453$) i ($T_{\tau-2}^{\text{MIN}}$, $p = 0,612$) wyraźnie przewyższa przyjęty w badaniu poziom istotności $\alpha = 0,05$. Przeprowadzana za pomocą testów nieparametrycznych i analizy prostej regresji identyfikacja trendu, potwierdziła występowanie we wszystkich badanych zmiennych losowych niewielkich tendencji rosnących lub malejących, z różnym poziomem istotności wykrytego trendu, o charakterze liniowy.

Identyfikacja występowania i wpływu wahań okresowych

Zbadano autokorelację ciągów zmiennych losowych Q_{τ}^{WIZ} , $Q_{\tau-2}^{\text{STR}}$, $Q_{\tau-2}^{\text{BUR}}$, $p_{\tau-2}^{\text{AREA}}$, $T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$. Obserwacje w analizowanych ciągach pomiarowych uszeregowano w cyklu hydrologicznym, tj. od 1 listopada do 31 października. Analizę wykonano dla opóźnienia 365 dób w celu zidentyfikowania występowania cykliczności rocznej z wyróżnieniem wahań sezonowych (letnich i zimowych). Uzyskane wyniki pokazano na rysunku 2.9. Na wykresach wyraźnie widoczny jest bardzo silny związek między bieżącą realizacją badanej zmiennej losowej (z wyjątkiem $p_{\tau-2}^{\text{AREA}}$) a jej wartością, z opóźnieniem do kilku dób (współczynnik autokorelacji powyżej 0,90). Tak duża wartość daje nadzieję, że wykorzystanie w modelu obserwacji opóźnionych o dwie doby, $\tau - 2$, w stosunku do Q_{τ}^{WIZ} będzie sprzyjać poprawnej prognozie hydrologicznej.



Rys. 2.9. Funkcja autokorelacji zmiennych losowych Q_{τ}^{WIZ} , $Q_{\tau-2}^{\text{STR}}$, $Q_{\tau-2}^{\text{BUR}}$, $p_{\tau-2}^{\text{AREA}}$, $T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$ dla opóźnienia 365 dób, z zaznaczonymi współczynnikami autokorelacji i poziomem ufności p (przerywana linia).

Fig. 2.9. Autocorrelation function of random variables Q_{τ}^{WIZ} , $Q_{\tau-2}^{\text{STR}}$, $Q_{\tau-2}^{\text{BUR}}$, $p_{\tau-2}^{\text{AREA}}$, $T_{\tau-2}^{\text{MAX}}$ and $T_{\tau-2}^{\text{MIN}}$ for a 365 days delay, with autocorrelation coefficients and confidence level p (dotted line).

W funkcji autokorelacji zaobserwowano stały cykl o rocznym okresie, z uwzględnieniem dwóch sezonów, tj. zimowego i letniego. Bardzo wyraźna sezonowość zmiennych losowych występuje dla $T_{\tau-2}^{\text{MAX}}$ i $T_{\tau-2}^{\text{MIN}}$ – współczynniki autokorelacji przekraczają 0,70. Z funkcji autokorelacji wynika, że cykl roczny rozpoczyna się pod koniec października (Lag = 0), a kończy na początku listopada (Lag = 365). Sezon letni dla przepływów rzecznych Q_{τ}^{WIZ} , $Q_{\tau-2}^{\text{STR}}$ i $Q_{\tau-2}^{\text{BUR}}$ ma swój początek na przełomie maja i czerwca (w przybliżeniu Lag = 210). Powyższy wniosek potwierdza długość corocznej fali wezbraniowej obserwowanej na Narwi w przekrojach wodowskazowych Wizna, Strękowa Góra i na Biebrzy

w Burzynie, będącej konsekwencją wezbrań roztopowo-deszczowych. Czas przemieszczania się fali często przekracza 30-50 dni. Początek termicznego lata wg funkcji autokorelacji przypada na początku maja (Lag = 180), koniec zaś na ostatnie dni października (Lag = 365).

Sprawdzenie stacjonarności zależnej zmiennej losowej z uwzględnieniem trendu lub braku trendu
 Sprawdzenie stacjonarności zależnej zmiennej losowej wykonano z zastosowaniem dwóch testów: rozszerzonego testu Dickey'a-Fullera ADF [Charemza, Deadman 1997], dla którego hipoteza zerowa zakłada niestacjonarność badanego ciągu, oraz testu Kwiatkowskiego, Phillipsa, Schmidta i Shina KPSS [Kwiatkowski i in. 1992], w którym hipoteza zerowa zakłada stacjonarność badanej zmiennej losowej. Test ADF uwzględnia istnienie trendu w badanym ciągu pomiarowym po przez odpowiedni dobór wartości krytycznej testu dla przyjętego w badaniu poziomu istotności $\alpha = 0,05$ [Dickey, Fuller 1979]. Wyniki wskazują (tab. 2.4), że statystyka testu ADF dla badanych zmiennych losowych jest niższa od wartości krytycznej, a p -wartość jest niższa od przyjętego w badaniu testu istotności. W związku z tym należy odrzucić H_0 o występowaniu w tych ciągach pierwiastka jednostkowego na rzecz hipotezy alternatywnej o stacjonarności badanych zmiennych losowych. Wyjątkiem jest zmienna $T_{\tau-2}^{MAX}$, dla której uzyskana statystyka jest wyższa od wartości krytycznej (statystyka ADF = -3,258, $ADF_{kr,\alpha=0,05} = -3,41$), a p -wartość = 0,0762 jest wyższa od przyjętego w badaniu poziomu istotności $\alpha = 0,05$. W związku z tym nie ma podstaw do odrzucenia H_0 , czyli zmienna jest niestacjonarna. Sprawdzenie stacjonarności ciągów pomiarowych wykonano również za pomocą testu KPSS. Statystyka testowa była wyższa od przyjętej w badaniu wartości krytycznej, zatem nie było podstaw do odrzucenia H_0 o stacjonarności badanych zmiennych losowych, z wyjątkiem zmiennej $T_{\tau-2}^{MAX}$ (statystyka KPSS = 0,193, $KPSS_{kr,\alpha=0,05} = 0,463$; p -wartość = 0,1, $\alpha = 0,05$). Oba testy wykazały, że badane zmienne losowe Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$ i $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$ i $T_{\tau-2}^{MIN}$ są stacjonarne, natomiast zmienna, $T_{\tau-2}^{MAX}$ jest niestacjonarna.

Analiza niejednorodności ciągów pomiarowych, zaobserwowanych w zlewni Górnej Narwi w okresie od 1990 do 1996 r., nie wykryła elementów odstających we wszystkich analizowanych ciągach. Zidentyfikowane na podstawie metody SVM elementy odstające, uznano za wiarygodne, gdyż dotyczyły maksymalnych zdarzeń hydrologiczno-meteorologicznych występujących naturalnie na obszarze zlewni Górnej Narwi. Sprawdzenie niezależności za pomocą testu serii i współczynnika autokorelacji wykazało, że wszystkie badane zmienne są zależne. W konsekwencji kolejne etapy badania niejednorodności wykonano za pomocą testów przeznaczonych do badania zależnych zmiennych losowych. Przeprowadzone analizy statystyczne, tj. identyfikacja trendu, ocena postaci funkcji trendu, identyfikacja występowania i wpływu wahań okresowych oraz sprawdzenie stacjonarności zależnych zmiennych losowych, wykazało że Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, i $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$ i $T_{\tau-2}^{MIN}$ są zmiennymi zależnymi i stacjonarnymi, natomiast $T_{\tau-2}^{MAX}$ jest zmienną zależną i niestacjonarną.

Analizowane zmienne losowe stanowią dane wejściowe do modeli DLMB. W związku z tym, że w algorytmach DLMB uwzględniany jest czynnik sezonowy, reprezentujący funkcję cykliczności prognozy oraz występowanie trendu liniowego [West, Harrison 1999], zdecydowano się na pozostawienie $T_{\tau-2}^{MAX}$ bez usuwania niestacjonarności.

II. WYZNACZENIE CHARAKTERYSTYK ROZKŁADU PRAWDOPODOBIEŃSTWA

W celu wyznaczenia charakterystyk rozkładu prawdopodobieństwa zestawiono oddzielne ciągi liczbowe badanych zmiennych losowych, tj. Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$, $T_{\tau-2}^{MAX}$, każdy o liczebności $n = 2557$. Dla każdej zmiennej losowej określono następujące charakterystyki rozkładu prawdopodobieństwa: asymetrii, koncentracji, rozproszenia i położenia (tab. 2.6). Analiza niejednorodności (tab. 2.4) wykazała, że wszystkie zmienne losowe są zależne. Zatem do obliczenia wariancji zmiennych zależnych zastosowano wzór (2.5), który jest uwzględniany w wyznaczaniu wyższych momentów, tj. współczynnika skośności (2.6) i spłaszczenia (2.7).

Dodatnia (powyżej +1,5) skośność uzyskana dla zmiennych Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$ (tab. 2.6) wskazuje na wyraźną prawostronną asymetrię, co oznacza, że prawy ogon rozkładu jest grubszy od lewego. Zmienne losowe, ($T_{\tau-2}^{MAX}$, -0,029) i ($T_{\tau-2}^{MIN}$, -0,547), wykazały większą lewostronną asymetrię zaobserwowanych wyników. Skośność w zakresie $\pm 1,5$ wskazuje, że ich rozkłady są zbliżone do rozkładu normalnego, z nieco wydłużonym lewym ramieniem rozkładu (lewostronna asymetria). Dodatnie współczynniki spłaszczenia dla zmiennych losowych (Q_{τ}^{WIZ} , +11,233), ($Q_{\tau-2}^{STR}$, +15,573), ($Q_{\tau-2}^{BUR}$, +16,669), ($P_{\tau-2}^{AREA}$, +29,237) i ($T_{\tau-2}^{MIN}$, +3,190) oznaczają grubsze ogony ich rozkładów prawdopodobieństw.

stwa w porównaniu do rozkładu normalnego. Powyższe rozkłady charakteryzują się mniejszym spłaszczeniem niż rozkład normalny, co świadczy o większej koncentracji zmiennej losowej wokół wartości średniej. Wartość współczynnika spłaszczenia mieszcząca się w przedziale $\pm 3,0$, np. ($T_{\tau-2}^{MAX}$, +2,039), wskazują że rozkład zmiennej jest zbliżony do rozkładu normalnego, co potwierdziło porównanie wartości średniej i mediany dla $T_{\tau-2}^{MAX}$ (odpowiednio 11,477°C i 11,645°C).

Tabela 2.6. Charakterystyki rozkładu prawdopodobieństwa zmiennych prognozowanych i prognozujących w modelowanej zlewni Górnej Narwi w przekroju wodowskazowym Wizna.

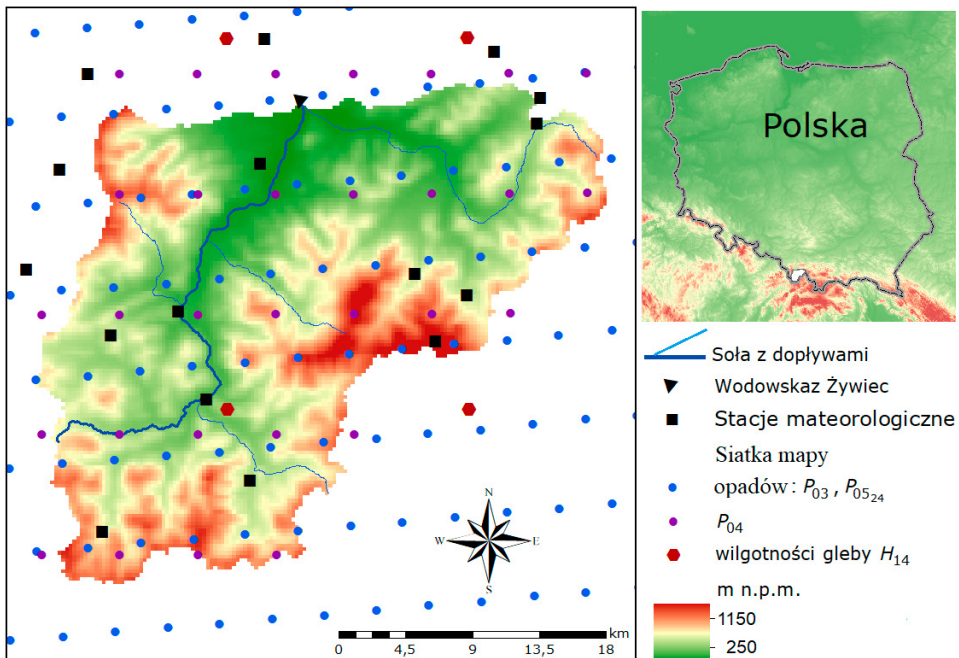
Table 2.6. Statistical measures of distribution probability for the predictand and predictors for the modeled Upper Narew River watershed at Wizna site.

	Q_{τ}^{WIZ} (m^3s^{-1})	$Q_{\tau-2}^{STR}$ (m^3s^{-1})	$Q_{\tau-2}^{BUR}$ (m^3s^{-1})	$P_{\tau-2}^{AREA}$ ($10^6 m^3$)	$T_{\tau-2}^{MAX}$ (°C)	$T_{\tau-2}^{MIN}$ (°C)
Miara asymetrii rozkładu						
Współczynnik skośności	2,353	2,757	2,988	4,072	-0,029	-0,547
Miara koncentracji rozkładu						
Współczynnik spłaszczenia	11,232	15,573	16,669	29,237	2,039	3,190
Miary rozproszenia rozkładu						
Wariancja	2880,016 [(m^3s^{-1}) ²]	567,363 [(m^3s^{-1}) ²]	872,462 [(m^3s^{-1}) ²]	9,999 [($10^6 m^3$) ²]	92,102 [(°C) ²]	54,369 [(°C) ²]
Miary położenia rozkładu						
Średnia arytmetyczna	70,475 (m^3s^{-1})	30,493 (m^3s^{-1})	37,794 (m^3s^{-1})	1,658 ($10^6 m^3$)	11,477 (°C)	2,650 (°C)
Mediana	54,200 (m^3s^{-1})	23,000 (m^3s^{-1})	31,300 (m^3s^{-1})	0,263 ($10^6 m^3$)	11,645 (°C)	2,804 (°C)

2.5. Dane wejściowe do usuwania błędu systematycznego (obciążenia) z meteorologicznych obserwacji satelitarnych i asymilacji satelitarnych obserwacji wilgotności gleby

2.5.1. Opis hydrologii zlewni Soły

Pierwszym większym prawostronnym dopływem Wisły jest rzeka Soła, o długości 88 km i powierzchni zlewni do przekroju wodowskazowego Żywiec równej 1400 km² (dł. geogr. 19,23°E – 22,68°E, szer. geogr. 50,05°N – 53,71°N; rys. 2.10). Rzeka bierze swój początek z połączenia kilku górskich potoków w miejscowości Rajcza, skąd typowy górski potok przekształca się w ciek o charakterze rzeczny. W górnej części zlewni Soły panuje klimat górski, w dolnej klimat podgórski, natomiast część środkowa leży w klimacie Pogórza Karpackiego. Opisany obszar charakteryzuje się dużą zmiennością miesięcznych i rocznych sum opadów. Średnia roczna suma opadów jest zależna od rzeźby terenu i wysokości nad poziomem morza. Na zmienność opadów ma także wpływ występowanie cieni opadowych w dolinach i na zewnętrznych zboczach. W zlewni Soły występuje wyraźne przesunięcie strefy opadów w stronę hydrologicznego półrocza letniego. Około dwóch trzecich rocznego opadu (P_{ann}), którego wartość zmienia się od 750 mm rok⁻¹ do 1300 mm rok⁻¹, przypada na okres między majem a październikiem. 64% opadów występuje w letnim półroczu hydrologicznym, 36% – w zimowym. W ciągu trzech miesięcy letnich (czerwiec – sierpień) obserwuje się 39% P_{ann} , w porównaniu do 17% w okresie od grudnia do lutego. Wysokie poziomy wody w dolinie Soły występują głównie w marcu i kwietniu (wezbrania roztopowe) oraz w lipcu i sierpniu, a ich przyczyną są nawalne lub długotrwałe deszcze o znacznym zasięgu. Letnie wezbrania (opadowe) charakteryzują się wysokimi kulminacjami, gwałtownością przebiegu i krótkim czasem występowania.



Rys. 2.10. Mapa hydrograficzna zlewni Soły w przekroju wodowskazowym Żywiec, przedstawiająca siatkę mapy opadów z obserwacji satelitarnych P_{03} , P_{0524} , P_{04} H-SAF i opadów zaobserwowanych na stacjach meteorologicznych, P_{OBS} , oraz obserwacji wilgotności gleby, H_{14} .

Fig. 2.10. Hydrographic map of the Soła basin in the Żywiec water gauge with the grid of the satellite precipitation observations P_{03} , P_{0524} , P_{04} H-SAF and precipitation observed at meteorological stations, P_{OBS} , and soil moisture observations, H_{14} .

2.5.2. Opracowanie meteorologicznych obserwacji z naziemnej sieci pomiarowo-obszaryjnej i obserwacji satelitarnych

IDENTYFIKACJA DANYCH WEJŚCIOWYCH DO MODELI

Przetwarzanie satelitarnych obserwacji opadu i wilgotności gleby wykonano w odniesieniu do obserwowanych danych z naziemnej sieci pomiarowo-obszaryjnej IMGW-PIB. Dobbowe i godzinne dane opadowe (odpowiednio zbiór P_{OBS} , P_d i P_h), jak również obserwacje satelitarne (P_{03} , P_{04} , P_{0524}), otrzymano z okresu od 1 maja 2013 do 30 kwietnia 2015 roku. Lokalizacja naziemnych stacji meteorologicznych IMGW-PIB (czarne kwadraty, rys. 2.10) oraz odczytów opadów satelitarnych H-SAF, P_{03} i P_{0524} (niebieskie koła, rys. 2.10) oraz P_{04} (fioletowe koła, rys. 2.10), podkreślają różnice w siatce opadów P_{03} i P_{0524} vs. P_{04} vs. w zlewni rzeki Soły. Rozdzielczość P_{03} i P_{0524} na obszarze zlewni Soły zmienia się od 3,5 km horyzontalnie do 5,9 km wertykalnie, natomiast rozdzielczość P_{04} od 5,2 km horyzontalnie do 8,5 km wertykalnie.

Dane P_{OBS} otrzymano z różnych systemów pomiarowych (w tym analogowych i cyfrowych) pracujących w innych warunkach pomiarowych na stacjach meteorologicznych. Obserwacje P_{03} (mm h^{-1}) pochodziły z map opadowych generowanych przez obrazy w podczerwieni (IR) z operacyjnych satelitów geostacjonarnych (Geostationary Earth Orbit, GEO), które kalibrowano wykorzystując opady pozyskane z detektora mikrofalowego (MW) z operacyjnych satelitów (Low Earth Orbit, LEO) operujących na orbitach heliosynchronicznych [EUMETSAT 2015a]. Mapy opadowe P_{03} reprezentują średnie natężenie opadu, mierzone w cyklach 15-minutowych czterokrotnie w ciągu godziny. Na obszarze Polski mapy opadowe P_{03} mają charakter jedynie jakościowy, gdyż nie jest wykonywana operacyjnie we-

ryfikacja tego produktu. Ocena dokładności wymaga porównania P_{03} z pomiarami naziemnymi i radarem meteorologicznym. Dokładność P_{03} dla opadów o natężeniu $>10 \text{ mmh}^{-1}$ jest szacowana na 40-80%, natomiast dla natężenia opadu w przedziale ($1-10 \text{ mmh}^{-1}$) na 80-160% [EUMETSAT 2015a]. Powyższy produkt jest przeznaczony głównie do analizy jakościowej opadów konwekcyjnych. Zastosowanie P_{03} w niniejszej pracy do matematycznego modelu hydrologicznego ma charakter badawczy i jest poprzedzone przeliczeniem natężenia opadu do sumy godzinnej (agregacji zgodnej z wejściowym krokiem modelu matematycznego HBV) oraz usunięciem błędu systematycznego za pomocą metody transformaty dystrybucyjnej (BC) [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 20]. P_{04} (mm) jest wysokością opadu pochodzącą z map opadowych tworzonych za pomocą obrazów z detektora mikrofalowego (MW), pozyskiwanych z operacyjnych satelitów LEO i interpolowanych w czasie, z wykorzystaniem dynamicznych informacji o opadzie dostępnych za pośrednictwem obrazów w podczerwieni (IR) z satelitów GEO. Zastosowanie powyższej techniki mieszania źródeł informacji umożliwiło ekstrapolację pola opadu do kilku kroków naprzód. Konsekwencją była z jednej strony mniejsza dokładność pola opadu, z drugiej poprawa wartości opadu przeznaczonego do opisu aktualnego stanu pogody i jej rozwoju w ciągu najbliższych kilku godzin (najczęściej do 6 godzin), czyli do tzw. nowcastingu [EUMETSAT 2015b]. Opad satelitarny P_{0524} (mm) pozyskano z map opadowych generowanych przez scalanie obrazów pochodzących z detektora mikrofalowego (MW), otrzymywanych z operacyjnych satelitów LEO operujących na orbitach heliosynchronicznych i z obrazów w podczerwieni (IR) z operacyjnych satelitów geostacjonarnych GEO i zintegrowanych w zakresie 3, 6, 12 i 24 godzin. Próbkowanie przeprowadzane jest za pomocą przyrządu optycznego (SEVIRI IFOV), radiometru skanującego linia po linii, zapewniającego dane obrazu w czterech kanałach widzialnych i bliskiej podczerwieni (VNIR) oraz ośmiu kanałach podczerwieni (IR). W celu zminimalizowania błędu systematycznego i błędów losowych, wyznaczone satelitarne pole opadowe zintegrowano z informacją o intensywności opadów z naziemnej sieci posterunków opadowych i z ilościową prognozą opadów (QPF) pochodzącą z modelu prognozowania numerycznego (Numerical Weather Prediction, NWP). Wybrany do analizy parametr P_{0524} reprezentuje 24-godzinną sumę opadu [EUMETSAT 2015c].

Na potrzeby analizy statystycznej danych wejściowych (2.5.3), opady P_{OBS} , P_{03} , P_{04} zagregowano do sumy dobowej. Dobowe sumy P_{OBS} , P_{03} , P_{04} i P_{0524} uwzględniały zarówno zdarzenia z opadem, jak i bez opadu. Natomiast wejściem do metody usuwania obciążenia (błędu systematycznego) były tylko te zdarzenia opadowe, których sumy dobowe były większe od zera. Usuwanie obciążenia z obserwacji satelitarnych wykonano metodą transformaty dystrybucyjnej, za pomocą trzech trójparametrowych teoretycznych dystrybucyj, z parametrem dolnego ograniczenia $\varepsilon > 0$.

Obserwacje wilgotności gleby, SM-DAS-2 (H_{14}) w strefie korzeniowej, przedstawiono za pomocą wskaźnika wilgotności gleby $I_{rz\theta}$, który jest opracowywany w Europejskim Centrum Prognoz Średnioterminowych w Zintegrowanym Systemie Prognozowania danych (ECMWF's IFS) w wyniku asymilacji wielkopowierzchniowej wilgotności gleby (dane ASCAT), czyli produktu skaterometrii SM-OBS-1 Meteorologicznych Satelitów Operacyjnych (MetOp) [EUMETSAT 2015d]. Rozdzielczość H_{14} (czerwone sześciokąty, rys. 2.10) na obszarze zlewni Soły zmienia się od 16,29 km horyzontalnie do 25,07 km wertykalnie. Obserwacje satelitarne pozyskano z okresu od 1 maja 2012 do 30 kwietnia 2015 roku.

Wilgotność gleby H_{14} (%) uzyskiwana jest w wyniku analizy powierzchni Ziemi na poziomie 2 m na bazie pomiarów pozyskiwanych z obserwacji synoptycznych i analizy wilgotności gleby opartej na uproszczonym filtrze Kalmana (Extend Kalman Filter, EKF). W procedurze asymilacji danych hydrologicznych (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land) satelitarna wilgotność gleby H_{14} (%) reprezentuje pionowy profil w zakresie czterech głębokości gleby: 0-0,07 m, 0,07-0,28 m, 0,28-1,00 m i 1,00-2,89 m. Zgodnie ze strukturą modelu hydrologicznego HBV (możliwość uwzględniania w modelu stref wegetacyjnych gleby), parametr H_{14} został uproszczony poprzez obliczenie średniej ważonej z trzech i czterech głębokości, tj. odpowiednio 0-1,00 m – $H_{14}^{1,00}$ i 0-2,89 m – $H_{14}^{2,89}$.

W celu usunięcia obciążenia z satelitarnych obserwacji H_{14} (%) lub asymilacji obserwacji satelitarnych przy użyciu metody EnKF-BC i EnSRF-BC, wygenerowano za pomocą symulacji HBV ciąg wartości wilgotności gleby θ_{HBV} (mm) w okresie odpowiadającym obserwacjom danych satelitarnych. W metodzie usuwania obciążenia wilgotność gleby, θ_{HBV} , musi być wyrażona w procentach, podobnie jak H_{14} . Dlatego wilgotność gleby θ_{HBV} została przeliczona wg wzoru (2.10):

$$\theta_{\text{HBV}} = \left[\frac{\theta_{\text{HBV}}}{F_c} \right] 100\% \quad (2.10)$$

gdzie F_c (mm) jest połową pojemnością wodną gleby, będącą zmienną stanu modelu hydrologicznego HBV, której wartość jest ustalana w czasie kalibracji modelu HBV. Po wykonaniu asymilacji satelitarnej wilgotności gleby, za pomocą metody EnKF-BC lub EnSRF-BC, należy wprowadzić asymilowane wartości prognozowanych zmiennych stanu wilgotności gleby z powrotem do modelu HBV. W tym celu θ (%) przeliczana jest wg wzoru (2.11):

$$\theta = \left[\frac{F_c \theta}{100\%} \right] \quad (2.11)$$

gdzie θ to przeliczona prognozowana zmienna stanu wilgotności gleby wyrażona w (mm).

2.5.3. Analiza statystyczna danych wejściowych do korekcji obciążenia i metod asymilacji

Analiza statystyczna danych wejściowych do korekcji obciążenia i metod asymilacji obejmuje:

- I. Badanie jednorodności statystycznej danych wejściowych do korekcji obciążenia i metod asymilacji.
- II. Przekształcenie ciągów pomiarowych w celu usunięcia zależności i niestacjonarności.
- III. Wyznaczanie charakterystyk rozkładu prawdopodobieństwa.
- IV. Wybór najlepiej dopasowanej funkcji rozkładu prawdopodobieństwa dla opadu P i wilgotności gleby θ na podstawie testu Kolmogorowa-Smirnowa.

I. BADANIE JEDNORODNOŚCI STATYSTYCZNEJ DANYCH WEJŚCIOWYCH DO KOREKCJI OBCIĄŻENIA I METOD ASYMILACJI

Analizę niejednorodności wykrywanej metodami statystycznymi wykonano zgodnie ze schematem procedury (rys. 2.1). Na potrzeby badania jednorodności danych wejściowych, opady P_{OBS} , P_{03} , P_{04} zagregowano do sumy dobowej. Dobowe sumy uwzględniały zdarzenia bez opadu. Wyniki analizy niejednorodności danych wejściowych do korekcji obciążenia i metod asymilacji przedstawiono w tabeli 2.7.

Identyfikacja obserwacji odstających na podstawie metody SVM

Do identyfikacji obserwacji odstających zastosowano metodę wektorów nośnych [Hawkins 1980; Ben-Hur i in. 2001]. Zidentyfikowane obserwacje odstające (czerwone punkty) pokazano na rysunku 2.11. W celu wyeliminowania lub pozostawienia w ciągu liczbowym obserwacji odstających, poddano wizualnej ocenie przebiegi ciągów pomiarowych. W całym przedziale obserwacji (tj. od 1 maja 2013 do 30 kwietnia 2015 r.) sumy dobowe opadów nie przekroczyły 50 mm, z wyjątkiem P_{04} (9 sierpnia 2013 r.: $P_{04} = 73,2$ mm, $P_{\text{OBS}} = 61,5$ mm, $P_{03} = 61,1$ mm i $P_{0524} = 56,4$ mm). Średnia arytmetyczna dobowych opadów w analizowanym okresie kształtowała się od $P_{\text{OBS}} = 2,862$ mm do $P_{0524} = 1,950$ mm. Najwyższa ilościowa i jakościowa rozbieżność między naziemnymi a satelitarnymi obserwacjami występuje latem (od lipca do sierpnia – niestabilność konwekcyjna w atmosferze) i zimą w okresie występowania ujemnych temperatur powietrza i stałej pokrywy śniegu. Najmniejsze różnice zaobserwowano jesienią, tj. od drugiej połowy września do listopada. Metoda SVM za obserwacje odstające dla P_{04} i P_{0524} (rys. 2.11) uznała zdarzenia bezopadowe, które na podstawie dostępnej informacji meteorologicznej (np. roczniki opadów atmosferycznych) uznano za wiarygodne.

Wilgotność gleby analizowano w okresie od 1 maja 2012 do 30 kwietnia 2015 roku. Maksymalna wilgotności gleby dla wszystkich zmiennych losowych, tj. θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$, nie przekroczyła 80%. Większe różnice zaobserwowano dla wartości minimalnej ($\theta_{\text{HBV}} = 22,9\%$, $H_{14}^{1,00} = 49,1\%$ i $H_{14}^{2,89} = 19,9\%$). Wartości odstające wystąpiły w miesiącach jesienno-zimowych. Jakość oszacowania wilgotności gleby z ASCAT [EUMETSAT 2015d] jest niska, gdy występuje pokrywa śniegu. $H_{14}/\text{SM-DAS-2}$ jako produkt ECMWF jest walidowany na podstawie gruntowych pomiarów wilgotności gleby *in situ*. W Polsce stacją pomiarową przeznaczoną do walidacji satelitarnej wilgotności gleby jest nizinna stacja w Łukowie (zachodnia część Niziny Południowopodlaskiej), działająca w ramach Systemu Monitoringu Suszy Rolniczej. Dane pomiarowe *in situ* są porównywane z $H_{14}/\text{SM-DAS-2}$ za pomocą średniego odchylenia, odchylenia standardowego, współczynnika korelacji i średniej kwadratowej różnicy. W Polsce

stwierdzono niski współczynnik korelacji ($r = 0,61$), stąd produkty satelitarne w okresie późnego lata i zimy są dodatkowo filtrowane, gdy temperatura powietrza spada poniżej $+3^{\circ}\text{C}$.

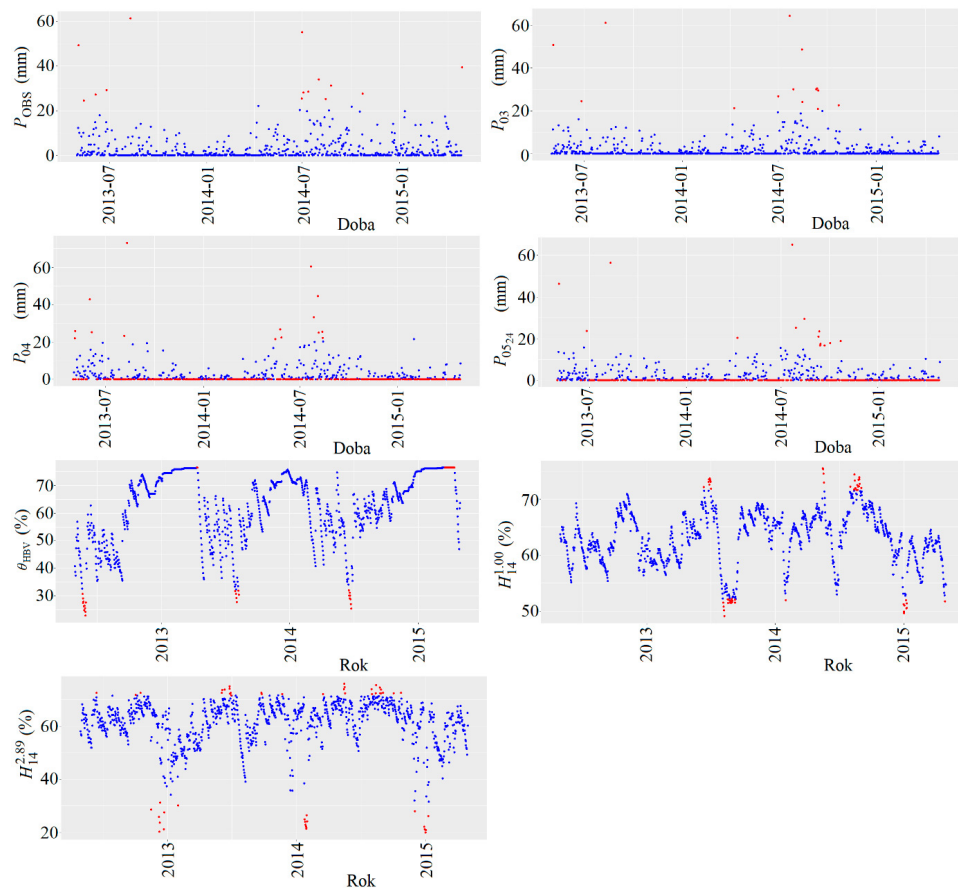
Tabela 2.7. Analiza niejednorodności ciągów pomiarowych w zlewni Soły przy poziomie istotności $\alpha = 0,05$. Pogrubione wartości wskazują na zależność lub/i niestacjonarność próby losowej oraz na istotność trendu.

Table 2.7. Nonhomogeneity analysis of the variables in Sola River watershed at significance level $\alpha = 0.05$. The bold values indicate on the dependence or/and non-stationarity of the random sample and the significance of the trend.

	P_{OBS} (mm)	P_{03} (mm)	P_{04} (mm)	P_{0524} (mm)	θ_{HBV} (%)	$H_{14}^{1,00}$ (%)	$H_{14}^{2,89}$ (%)
Wykrywanie elementów odstających METODA WEKTORÓW NOŚNYCH, SVM							
Występują (+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
Odstające/nieodstające	54/675	14/715	38/691	41/688	54/1041	55/1040	54/1041
Sprawdzenie niezależności TEST SERII I WSPÓLCZYNNIK AUTOKOREKACJA (max. opóźnienie Lag = 8 dób)							
Liczba obserwacji powyżej/poniżej Me	361/358	347/336	347/336	347/336	545/549	545/549	546/547
p -wartość	0,012	0,000	0,000	0,000	0,000	0,000	0,000
Ocena zmiennej	zależna	zależna	zależna	zależna	zależna	zależna	zależna
Identyfikacja trendu oraz ocena funkcji trendu TEST WSPÓLCZYNNIKA KORELACJI RANG SPEARMANA							
	0,0456	-0,0644	-0,1357	-0,1154	0,3350	0,0797	0,0487
TEST MANNA-KENDALLA							
Wartość testu	0,033	-0,046	-0,091	-0,084	0,257	0,045	0,036
p -wartość	0,221	0,085	0,0003	0,002	0,000	0,026	0,078
Identyfikacja występowania i wpływu wahań okresowych TEST AUTOKORELACJI (opóźnienie Lag = 365 dób)							
Występują (+)	(-)	(-)	(-)	(-)	(+)	(+)	(+)
Sprawdzenie stacjonarności z uwzględnieniem trendu lub braku trendu TEST DICKEY'A-FULLERA, ADF							
Statystyka ADF $ADF_{kr,\alpha=0,05} = -3,41$ (z trendem) $ADF_{kr,\alpha=0,05} = -2,89$ (bez trendu)	-7,499	-7,349	-5,827	-7,636	-3,971	-4,231	-5,705
Lag	8	8	8	8	10	10	10
p -wartość	0,010	0,010	0,010	0,010	0,010	0,010	0,010
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST KWIATKOWSKIEGO-PHILLIPSA-SHMIDTA-SHINA, KPSS							
Statystyka KPSS $KPSS_{kr,\alpha=0,05} = 0,463$	0,464	0,465	0,532	0,476	0,321	0,473	0,738
Lag	5	6	6	6	7	7	7
p -wartość	0,049	0,042	0,034	0,045	0,100	0,046	0,010
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	niestacjonarna	stacjonarna	stacjonarna

Przeprowadzona analiza występowania obserwacji odstających nie wykluczyła żadnej obserwacji opadów i wilgotności gleby otrzymanych zarówno z pomiarów z naziemnej sieci pomiarowo-obszaryjnej, jak i produktów satelitarnych. Największe różnice ilościowe i jakościowe zaobserwowano w stosunku do θ_{HBV} . Ciąg θ_{HBV} jest generowany przez symulacyjny model hydrologiczny HBV, zatem na

jego wartości ma wpływ kalibracja modelu matematycznego. θ_{HBV} w rzeczywistości jest zmienną stanu modelu HBV i w kolejnych etapach obliczeń będzie korygowana w ramach asimilacji satelitarnych obserwacji wilgotności gleby: $H_{14}^{1,00}$ i $H_{14}^{2,89}$.



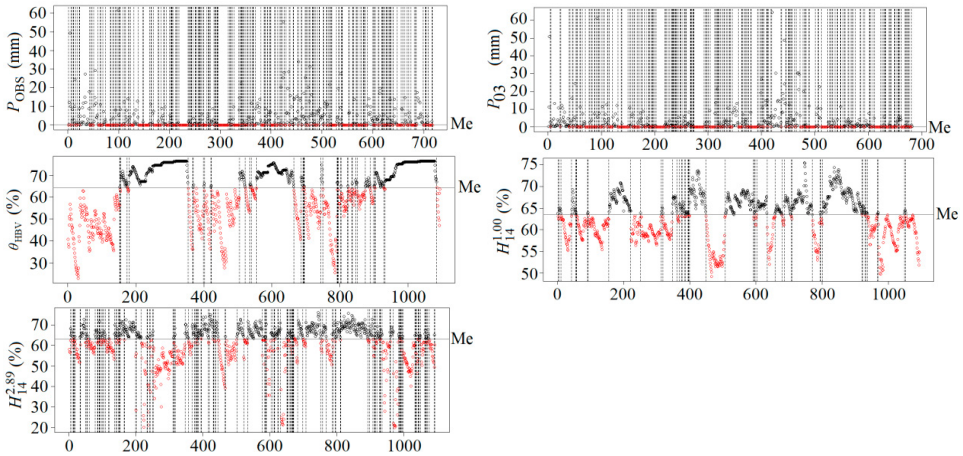
Rys. 2.11. Identyfikacja obserwacji odstających w ciągach pomiarowych P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ z użyciem metody SVM.

Fig. 2.11. Identification of outliers in measurement data P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ and $H_{14}^{2,89}$ using SVM method.

Sprawdzanie niezależności badanych zmiennych losowych za pomocą testu serii i testu współczynnika autokorelacji

Sprawdzenie niezależności zmiennych losowych P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ przeprowadzono przy użyciu testu serii [Wald, Wolfowitz 1940] i testu współczynnika autokorelacji dla maksymalnego opóźnienia Lag = 8 dób służącego do sprawdzenia wewnętrznej korelacji badanych zmiennych losowych. Przykłady wykresów ciągów pomiarowych wraz z linią referencyjną równą medianie pokazano na rysunku 2.12. Wykresy weryfikują liczbę obserwacji powyżej i poniżej mediany, jak również liczbę przebiegów w górę i w dół. W tabeli 2.7 zaznaczono liczbę obserwacji powyżej i poniżej mediany oraz prawdopodobieństwo testowe, p -wartość. Ciągi niezależne powinny charakteryzować się jednakową liczbą obserwacji w obu przedziałach. Do obliczeń przyjęto poziom istotności $\alpha = 0,05$. W przypadku p -wartości $< 0,05$ hipotezę zerową o niezależności badanej zmiennej losowej należy odrzucić. Na podstawie wyników testów statystycznych (tab. 2.7) uznano, że realizacje wszystkich badanych

zmiennych losowych nie mają charakteru losowego, a ich dobór do próby podlega tendencji lub cykliczności, tzn. badane zmienne losowe są zależne. Na podstawie tego wniosku dalszą analizę niejednorodności wykonano za pomocą testów statystycznych dla zależnych zmiennych losowych (prawa strona na rys. 2.1).



Rys. 2.12. Wykres weryfikujący liczbę obserwacji powyżej i poniżej mediany, jak również liczbę przebiegów w górę i w dół dla zmiennych P_{OBS} , P_{03} , θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ za pomocą testu serii.

Fig. 2.12. The chart that verifies the runs above and below the median as well the runs up and down for variables P_{OBS} , P_{03} , θ_{HBV} , $H_{14}^{1,00}$ and $H_{14}^{2,89}$ using runs test.

Identyfikacja trendu

W tabeli 2.7 przedstawiono korelacje między badanymi zmiennymi losowymi, tj. P_{OBS} , P_{03} , P_{04} , $P_{05,24}$, θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ a zmienną reprezentującą czas, τ . Pogrubione wartości współczynników korelacji wskazują na istotne zależności z $p < 0,050$. Występowanie dowolnego trendu monotonicznego, tj. wg współczynnika korelacji rang Spearmana [Spearman 1904], zaobserwowano dla zmiennych: (P_{04} , $-0,1357$), ($P_{05,24}$, $-0,1154$), (θ_{HBV} , $0,3350$) i ($H_{14}^{1,00}$, $0,0797$). Najsilniejszy związek między badaną zmienną a czasem, τ , zaobserwowano dla θ_{HBV} . Powyższa własność statystyczna wynika z faktu, że θ_{HBV} nie jest obserwacją naturalną, tj. prowadzoną w warunkach nielaboratoryjnych, lecz została wygenerowana przy użyciu symulacji skalibrowanego modelu hydrologicznego HBV.

Istotność wykrytej tendencji zmian w badanych ciągach pomiarowych określono za pomocą nieparametrycznego testu statystycznego Manna-Kendalla [Hamed 2008], polegającego na weryfikacji hipotezy o braku trendu na podstawie nieparametrycznego współczynnika korelacji. Test Manna-Kendalla (tab. 2.7) potwierdził istotność występowania trendu malejącego dla zmiennej $P_{05,24}$ na poziomie istotności p -wartość = 0,002 i rosnącego dla P_{04} , θ_{HBV} i $H_{14}^{1,00}$, odpowiednio na poziomie istotności 0,0003, 0,000 i 0,026 (p -wartość).

Ocena postaci funkcji trendu

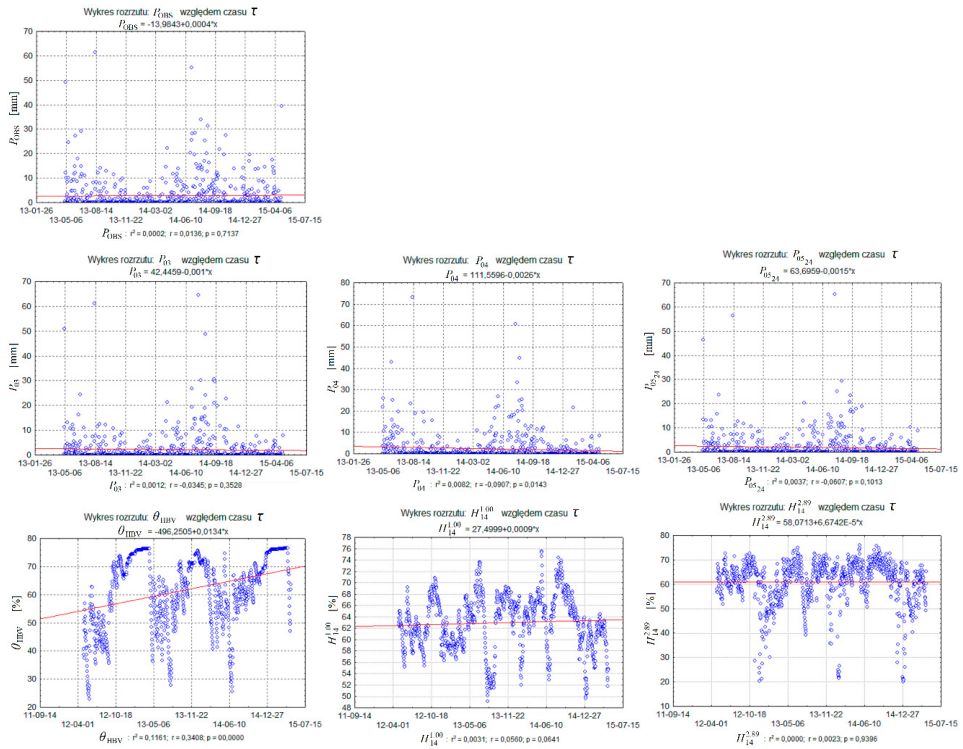
Wyniki analizy prostej regresji w dużym stopniu potwierdzają analizę współczynników korelacji. W tabeli 2.8 pogrubione wartości wskazują na istnienie istotnego wpływu czasu, τ , na badaną zmienną losową. Prawdopodobieństwo testowe p , wyznaczone na podstawie otrzymanych współczynników regresji, analizowano przy przyjętym w badaniu poziomie istotności $\alpha = 0,05$. Współczynnik nachylenia prostej jest istotnie różny od zera dla P_{04} ($\beta = -0,0026$, $p = 0,014$) i θ_{HBV} ($\beta = 0,013$, $p = 0,000$). Dla pozostałych badanych zmiennych losowych, tj. P_{OBS} , P_{03} , $P_{05,24}$, $H_{14}^{1,00}$ i $H_{14}^{2,89}$, analiza prostej regresji nie wykryła istotnego statystycznie trendu.

Z wykresów rozrzutu badanych zmiennych losowych względem zmiennej czasu τ (rys. 2.13) wynika, że wyraźną tendencję wzrostu zaobserwowano tylko dla θ_{HBV} oraz niewielki trend malejący dla P_{04} . Dla pozostałych badanych zmiennych losowych współczynnik nachylenia prostej regresji był nieistotnie różny od zera z prawdopodobieństwem testowym $p > 0,05$.

Tabela 2.8. Wyniki analizy regresji prostej, gdzie zmiennymi zależnymi są: P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$, a zmienną niezależną jest czas (τ). Pogrubione wartości wskazują na istotność parametrów regresji, czyli na istnienie istotnego wpływu zmiennej losowej τ na badaną zmienną zależną, z $p < 0,05$.

Table 2.8. The results of simple regression analysis, where the dependent variable are: P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ and $H_{14}^{2,89}$, and the independent variable is time (τ). Bold values indicate the significance of the regression parameters, i.e. the existence of a significant effect of the random variable τ on the dependent variable under study, with $p < 0.05$.

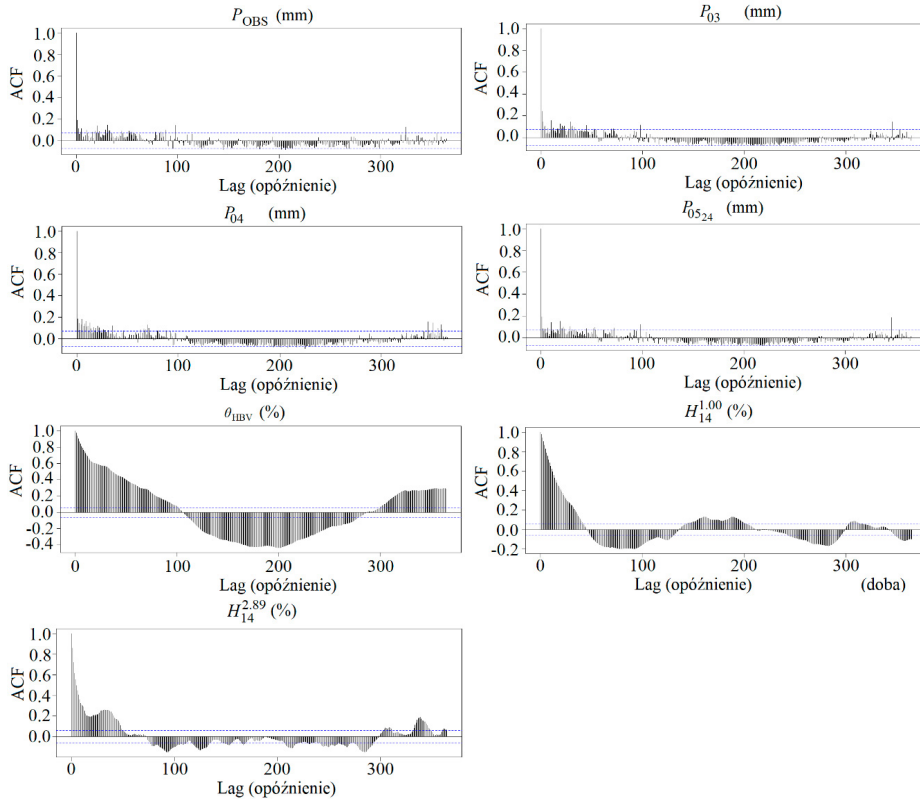
	Beta	Błąd standardowy Beta	β	Błąd standardowy β	t(n-2)	Poziom p
P_{OBS}	R = 0,0136, R ² = 0,0002, F(1,727) = 0,1347 Błąd standardowy estymacji = 6,2452					
Wyraz wolny			2,7152	0,4631	5,8633	0,000
τ	0,0136	0,0371	0,0004	0,0011	0,3670	0,714
P_{03}	R = 0,0345, R ² = 0,0012, F(1,727) = 0,8645 Błąd standardowy estymacji = 5,8890					
Wyraz wolny			2,5552	0,4367	5,8514	0,000
τ	-0,0345	0,0371	-0,0009	0,0010	-0,9298	0,353
P_{04}	R = 0,0907 , R ² = 0,0082 , F(1,727) = 6,0341 Błąd standardowy estymacji = 6,0521					
Wyraz wolny			3,2548	0,4488	7,2529	0,000
τ	-0,0907	0,0369	-0,0026	0,0011	-2,4565	0,014
P_{0524}	R = 0,0672, R ² = 0,0037, F(1,727) = 2,6911 Błąd standardowy estymacji = 5,1214					
Wyraz wolny			2,4900	0,3797	6,5569	0,000
τ	-0,0607	0,0370	-0,0015	0,0009	-1,6405	0,101
θ_{HBV}	R = 0,3408 , R ² = 0,1161 , F(1,1093) = 143,62 Błąd standardowy estymacji = 11,7150					
Wyraz wolny			54,4546	0,7085	76,8526	0,000
τ	0,3408	0,0284	0,0134	0,0011	11,9840	0,000
$H_{14}^{1,00}$	R = 0,0559, R ² = 0,0031, F(1,1093) = 3,4361 Błąd standardowy estymacji = 4,8148					
Wyraz wolny			62,5083	0,2912	214,6526	0,000
τ	0,0559	0,0302	0,0008	0,0005	1,8537	0,064
$H_{14}^{2,89}$	R = 0,0022, R ² = 0,0000005, F(1,1093) = 0,0057 Błąd standardowy estymacji = 9,2074					
Wyraz wolny			60,8096	0,5569	109,1986	0,000
τ	0,0023	0,0302	0,00007	0,0008	0,0758	0,940



Rys. 2.13. Wykres rozrzutu zależnych zmiennych losowych P_{OBS} , P_{03} , P_{04} , P_{0524} (skumulowanych do doby z uwzględnieniem zdarzeń bezopadowych), θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$, względem zmiennej niezależnej czas ($x = \tau$) z zaznaczonym równaniem prostej regresji, współczynnika determinacji (r^2), współczynnika korelacji liniowej Pearsona (r) oraz prawdopodobieństwem testowym (p) porównywanym z przyjętym poziomem istotności $p < 0,05$.
Fig. 2.13. Scatter chart of dependent random variables: P_{OBS} , P_{03} , P_{04} , P_{0524} (accumulated for day, including precipitation-free events), θ_{HBV} , $H_{14}^{1,00}$ and $H_{14}^{2,89}$, relative to the independent variable time ($x = \tau$) with the equation of the regression line, coefficient of determination (r^2), Pearson's linear coefficient (r) and test probability (p) compared with assumed significance level $p < 0.05$.

Identyfikacja występowania i wpływu wahań okresowych

Analizę autokorelacji ciągów zmiennych P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ wykonano dla opóźnienia 365 dób (rys. 2.14). Na wykresach autokorelacji zmiennych opadowych związek między bieżącym opadem a jego poprzednimi wartościami jest słaby, współczynniki korelacji mieszczą się w przedziale $(\pm 0,15)$. W funkcji autokorelacji wilgotności gleby zaobserwowano stały cykl o rocznym okresie z uwzględnieniem dwóch sezonów, tj. zimowego i letniego. Wahanía okresowe są wyraźne dla θ_{HBV} (współczynniki korelacji mieszczą się w przedziale $\pm 0,25$) i dużo słabsze dla $H_{14}^{1,00}$ i $H_{14}^{2,89}$ (współczynniki korelacji w przedziale $\pm 0,15$). Obserwacje wilgotności gleby uszeregowano w układzie rocznym od 1 maja do 30 kwietnia, wynikającym z dostępności do danych historycznych i satelitarnych. Z wykresów funkcji autokorelacji dla θ_{HBV} wynika, że cykl roczny dla wilgotności gleby rozpoczyna się na przełomie października i listopada (Lag = 200). Sezon letni ma swój początek na przełomie marca i kwietnia (LAG = 0) i jest to związane z wiosennym zasilaniem rozmarznętego gruntu wodą roztopowo-opadową.



Rys. 2.14. Funkcja autokorelacji zmiennych losowych P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ dla opóźnienia 365 dób z zaznaczonymi współczynnikami autokorelacji p (przerywana linia).

Fig. 2.14. Autocorrelation function of random variables P_{OBS} , P_{03} , P_{04} , P_{0524} , θ_{HBV} , $H_{14}^{1,00}$ and $H_{14}^{2,89}$ for a 365 days delay with autocorrelation coefficients and confidence level p (dotted line).

Sprawdzenie stacjonarności zależnej zmiennej losowej z uwzględnieniem trendu lub braku trendu

Sprawdzenie stacjonarności zależnej zmiennej losowej wykonano z zastosowaniem dwóch testów, tj. rozszerzonego testu Dickey'a-Fullera ADF [Charemza Deadman 1997] oraz testu Kwiatkowskiego, Phillipsa, Schmidta i Shina KPSS [Kwiatkowski i in. 1992]. Oba testy wykazały, że nie ma podstaw do odrzucenia hipotezy o niestacjonarności zmiennych losowych P_{OBS} , P_{03} , P_{04} , P_{0524} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$. W przypadku zmiennej θ_{HBV} wartość statystyki testu KPSS była niższa od wartości krytycznej, a p -wartość była wyższa od przyjętego w badaniu poziomu istotności (statystyka KPSS = 0,321, $KPSS_{kr,\alpha=0,05} = 0,463$; p -wartość = 0,100, $\alpha = 0,05$). W związku z tym odrzucono H_0 o stacjonarności zmiennej losowej, θ_{HBV} .

Analiza niejednorodności ciągów pomiarowych, zaobserwowanych w zlewni Soły w okresie od 1 maja 2012 do 30 kwietnia 2015 r., nie wykryła elementów odstających we wszystkich analizowanych ciągach. Zidentyfikowane na podstawie metody SVM elementy odstające uznano za wiarygodne, gdyż dotyczyły maksymalnych zdarzeń hydrologiczno-meteorologicznych występujących naturalnie na obszarze zlewni Soły. Sprawdzenie niezależności za pomocą testu serii i współczynnika autokorelacji wykazało, że wszystkie badane zmienne są zależne oraz wykazują liniowy trend na różnym poziomie istotności. W przypadku zmiennej θ_{HBV} można zidentyfikować roczną cykliczność z wyróżnieniem wahań sezonowych (letnich i zimowych).

Analizowane ciągi pomiarowe stanowią wejście do metody korekcji obciążenia, obliczanej za pomocą metody transformaty dystrybucyjnej (BC). W takim przypadku wymagane jest zbadanie ich zgodności z jednym z rozpatrywanych w tej metodzie rozkładów prawdopodobieństwa, tj. uogólnionym rozkładem wykładniczym, gamma oraz Weibulla. Testowanie takiej zgodności należy wykonać przy założeniu niezależności i stacjonarności próby losowej. W związku z tym wszystkie badane zmienne losowe przed kolejnym etapem badań zostały przekształcone, w celu usunięcia zależności i niestacjonarności.

II. PRZEKSZTAŁCENIE CIĄGÓW POMIAROWYCH W CELU USUNIĘCIA ZALEŻNOŚCI I NIESTACJONARNOŚCI

W przypadku danych opadowych najprostszym sposobem przekształcenia ciągów pomiarowych P_{OBS} , P_{03} , P_{04} , P_{0524} było usunięcie zdarzeń bezopadowych. Uzasadnieniem takiej operacji jest, występujące w metodzie transformaty dystrybucyjnej, ograniczenie zmienności sumy opadu. Do metody brane są pod uwagę sumy opadów $P > 0,1$ mm. Wyniki przeprowadzonej analizy niejednorodności dla dobowych sum opadów, bez uwzględniania zdarzeń bezopadowych, przedstawiono w tabeli 2.9.

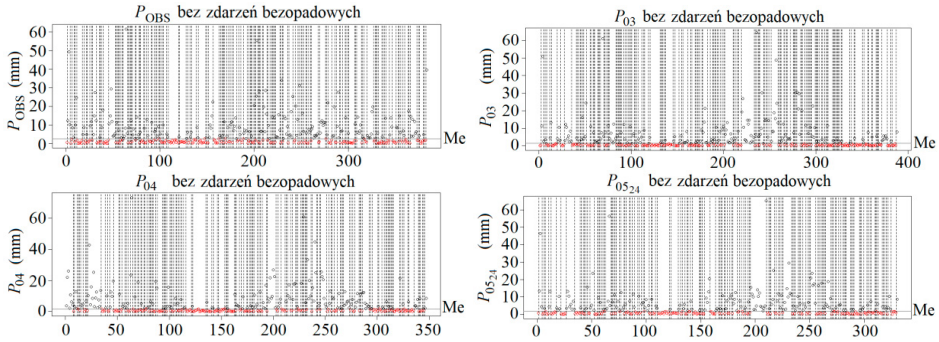
Tabela 2.9. Analiza niejednorodności opadowych ciągów pomiarowych (bez zdarzeń bezopadowych) do korekcji obciążenia metodą transformaty dystrybucyjnej (BC) w zlewni Soły przy poziomie istotności $\alpha = 0,05$.

Table 2.9. Nonhomogeneity analysis of the precipitation variables (without precipitation-free events) to bias correction using the distribution derived transformation method in Soła River watershed at significance level $\alpha = 0.05$.

	P_{OBS} (mm)	P_{03} (mm)	P_{04} (mm)	P_{0524} (mm)
Sprawdzenie niezależności TEST SERII I WSPÓŁCZYNNIK AUTOKOREKACJA (max. opóźnienie Lag = 8 dób)				
Liczba obserwacji powyżej/poniżej Me	192/191	192/196	174/173	169/169
p -wartość	0,061	0,052	0,060	0,078
Ocena zmiennej	niezależna	niezależna	niezależna	niezależna
Sprawdzenie stacjonarności niezależnej zm. losowej TEST SUMY RANG KRUSKALA-WALLISA				
p -wartość	0,391	0,921	0,578	0,241
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST NA TREND WARTOŚCI ŚREDNIEJ				
p -wartość	0,161	0,253	0,532	0,337
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST NA TREND WARIANCJI				
p -wartość	0,173	0,304	0,481	0,196
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna

Sprawdzanie niezależności badanych zmiennych losowych oraz stacjonarności

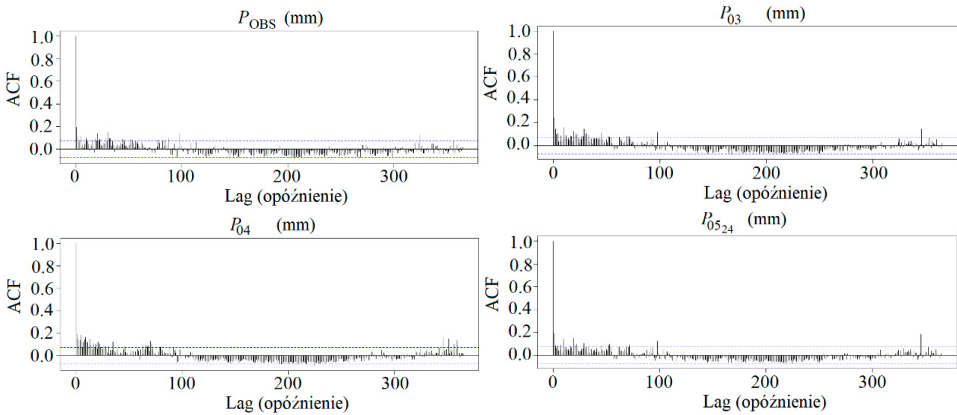
Sprawdzenie niezależności przekształconych zmiennych losowych P_{OBS} , P_{03} , P_{04} , P_{0524} (bez zdarzeń bezopadowych), z użyciem testu serii [Wald, Wolfowitz 1940] i testu współczynnika autokorelacji dla maksymalnego opóźnienia Lag = 8 dób, wykazało że nie ma podstaw do odrzucenia weryfikowanej hipotezy zerowej o niezależności badanej zmiennej losowej. Zweryfikowano liczbę obserwacji powyżej i poniżej wartości mediany oraz wyznaczono prawdopodobieństwo testowe, p -wartość. Dla każdej analizowanej zmiennej losowej, liczby obserwacji powyżej i poniżej mediany były zbliżone (rys. 2.15), a wartość testu była większa od przyjętego w badaniu poziomu istotności $\alpha = 0,05$ (tab. 2.9). Uznano, że realizacje wszystkich badanych zmiennych losowych mają charakter losowy, a ich dobór do próby nie podlega tendencji lub cykliczności, tzn. badane zmienne losowe są niezależne.



Rys. 2.15. Wykres weryfikujący liczbę obserwacji powyżej i poniżej mediany, jak również liczbę przebiegów w górę i w dół dla zmiennych P_{OBS} , P_{03} , i $P_{05_{24}}$ (bez zdarzeń bezopadowych) za pomocą testu serii.

Fig. 2.15. The chart that verifies the runs above and below the median as well the runs up and down for variables P_{OBS} , P_{03} and $P_{05_{24}}$ (without precipitation-free events) using runs test.

Na podstawie tego wniosku dalsza analiza niejednorodności została wykonana za pomocą testów statystycznych dla niezależnych zmiennych losowych (lewa strona na rys. 2.1). Zbadano autokorelację ciągów zmiennych losowych P_{OBS} , P_{03} , P_{04} , $P_{05_{24}}$ (bez zdarzeń bezopadowych). Obserwacje w analizowanych ciągach pomiarowych uszeregowano od 1 maja do 30 kwietnia. Powyższy dobór zmiennych losowych był uzależniony od dostępności obserwacji satelitarnych. Analizę wykonano dla opóźnienia 365 dób w celu zidentyfikowania występowania cykliczności rocznej z wyróżnieniem wahań sezonowych (letnich i zimowych). Współczynniki autokorelacji w większości przypadków znajdowały się wewnątrz przedziału ufności dla przyjętego w badaniu poziomu istotności $\alpha = 0,05$. Uzyskane wyniki pokazano na rysunku 2.16. W funkcji autokorelacji nie zaobserwowano stałego cyklu o rocznym okresie. Wszystkie analizowane próby uznano za niezależne i stacjonarne, nie wykryto skoku wartości średniej, trendu wartości średniej i wariancji. Prawdopodobieństwa testowe (p -wartość) były wyższe od przyjętego w badaniu poziomu istotności $\alpha = 0,05$ (tab. 2.9). Nie było powodu, aby odrzucać H_0 w odniesieniu do jednorodności danych opadowych po usunięciu z ciągów zdarzeń bezopadowych.



Rys. 2.16. Funkcja autokorelacji zmiennych losowych P_{OBS} , P_{03} , P_{04} , $P_{05_{24}}$ (bez zdarzeń bezopadowych) dla opóźnienia 365 dób, z zaznaczonymi współczynnikami autokorelacji i poziomem ufności p (przerywana linia).

Fig. 2.16. Autocorrelation function of random variables P_{OBS} , P_{03} , P_{04} , $P_{05_{24}}$ (without precipitation-free events) for a 365 days delay, with autocorrelation coefficients and confidence level p (dotted line).

Pogrupowanie obserwacji wilgotności gleby,
tj. θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$, według sezonu zimowego i letniego

Obserwacje wilgotności gleby wygenerowane przez model HBV oraz obserwacje satelitarne, tj. θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$, pogrupowano według sezonu. Dla każdej zmiennej losowej, metodą odpowiedniego uporządkowania, utworzono dwa oddzielne ciągi reprezentujące obserwacje zimą i latem. Na podstawie uzyskanych wyników identyfikacji występowania i wpływu wahań okresowych dla zmiennej θ_{HBV} (rys. 2.14) ustalono sezon letni od maja do października i zimowy od listopada do kwietnia. Powyższy podział jest zgodny z cyklem hydrologicznym w Polsce. Klimatyczna zima rozpoczyna się w jednym z dwóch ostatnich miesięcy roku, a odpływ wód deszczowych utrudniony zimą z powodu mrozów ma miejsce w sezonach wiosennych następnego roku kalendarzowego. Pogrupowanie zmiennych losowych, polegające na zbudowaniu nowych ciągów zmiennych losowych na podstawie ciągów już istniejących, wykonano w celu przekształcenia szeregu czasowego obserwacji w ciąg obserwacji mający cechy próby prostej, tj. w ciąg niezależny i stacjonarny.

Pogrupowane ciągi zmiennych losowych wg sezonu (zima, lato), tj. θ_{HBV} -ZIMA, θ_{HBV} -LATO, $H_{14}^{1,00}$ -ZIMA, $H_{14}^{1,00}$ -LATO, $H_{14}^{2,89}$ -ZIMA i $H_{14}^{2,89}$ -LATO, poddano analizie niejednorodności zgodnie algorytmem na rysunku 2.1. Wyniki przedstawiono w tabeli 2.10.

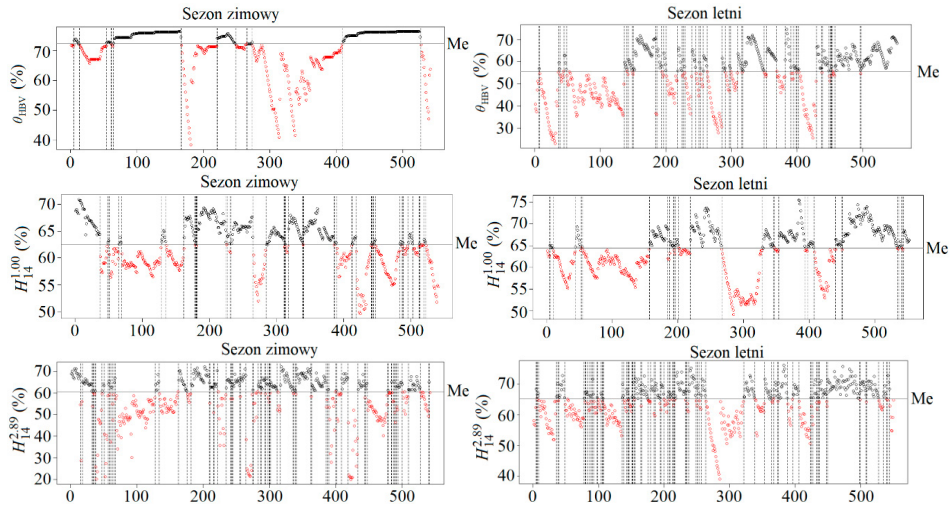
Tabela 2.10. Analiza niejednorodności ciągów pomiarowych wilgotności gleby w sezonie zimowym i letnim do korekcji obciążenia metodą transformaty dystrybuanty (BC) w zlewni Soły przy poziomie istotności $\alpha = 0,05$. Pogrubione wartości wskazują na zależność lub/i niestacjonarność próby losowej.

Table 2.10. Nonhomogeneity analysis of the soil moisture variables in winter and summer season with bias correction using the distribution derived transformation method in Soła River watershed at significance level $\alpha = 0.05$. The bold values indicate on the dependence or/and non-stationarity of the random sample.

	θ_{HBV} -ZIMA (%)	θ_{HBV} -LATO (%)	$H_{14}^{1,00}$ -ZIMA (%)	$H_{14}^{1,00}$ -LATO (%)	$H_{14}^{2,89}$ -ZIMA (%)	$H_{14}^{2,89}$ -LATO (%)
Sprawdzenie niezależności TEST SERII I WSPÓLCZYNNIK AUTOKOREKACJA (max. opóźnienie Lag = 8 dób)						
Liczba obserwacji powyżej/poniżej Me	270/271	276/276	271/270	276/276	271/271	275/275
<i>p</i> -wartość	0,018	0,028	0,063	0,069	0,050	0,052
Ocena zmiennej	zależna	zależna	niezależna	niezależna	niezależna	niezależna
Sprawdzenie stacjonarności niezależnej zm. losowej TEST SUMY RANG KRUSKALA-WALLISA						
<i>p</i> -wartość	0,019	0,408	0,131	0,192	0,271	0,447
Ocena zmiennej	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST NA TREND WARTOŚCI ŚREDNIEJ						
<i>p</i> -wartość	0,027	0,016	0,065	0,077	0,092	0,105
Ocena zmiennej	niestacjonarna	niestacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna
TEST NA TREND WARIANCJI						
<i>p</i> -wartość	0,030	0,013	0,052	0,082	0,103	0,123
Ocena zmiennej	niestacjonarna	niestacjonarna	stacjonarna	stacjonarna	stacjonarna	stacjonarna

Wykonane sprawdzenie niezależności zmiennych losowych θ_{HBV} -ZIMA, θ_{HBV} -LATO, $H_{14}^{1,00}$ -ZIMA, $H_{14}^{1,00}$ -LATO, $H_{14}^{2,89}$ -ZIMA i $H_{14}^{2,89}$ -LATO, z użyciem testu serii [Wald, Wolfowitz 1940] i testu współczynnika autokorelacji dla maksymalnego opóźnienia Lag = 8 dób, dało negatywne wyniki w odniesieniu do zmiennych θ_{HBV} -ZIMA i θ_{HBV} -LATO. Obliczone prawdopodobieństwo testowe, *p*-wartość, okazało się niższe od przyjętego w badaniu poziomu istotności $\alpha = 0,05$, co wskazuje na zależność badanych zmiennych losowych. W związku z tym, że zmienne θ_{HBV} -ZIMA i θ_{HBV} -LATO nie są obser-

wacjami naturalnymi, lecz wygenerowanymi przy użyciu symulacji skalibrowanego modelu hydrologicznego HBV, postanowiono nie odrzucać hipotezy zerowej o niezależności tych dwóch zmiennych losowych przy jednoczesnym obniżeniu poziomu istotności $\alpha = 0,01$. Wykresy weryfikujące liczbę obserwacji powyżej i poniżej mediany, jak również liczbę przebiegów w górę i w dół, przedstawiono na rysunku 2.17. Analizowane ciągi charakteryzują się jednakową liczbą obserwacji w obu przedziałach (zaobserwowano różnicę jednego przebiegu dla zmiennej θ_{HBV} -ZIMA i $H_{14}^{1,00}$ -ZIMA).



Rys. 2.17. Wykres weryfikujący liczbę obserwacji powyżej i poniżej mediany, jak również liczbę przebiegów w górę i w dół, dla pogrupowanych (wg sezonu: zima i lato) zmiennych θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ za pomocą testu serii.
Fig. 2.17. The chart that verifies the runs above and below the median as well the runs up and down for grouped (by season: winter and summer) variables θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ using runs test.

Zbadanie stacjonarności, tj. niezmienności czasowej wartości średniej i wariancji dla niezależnej zmiennej losowej (algorytm – lewa strona na rys. 2.1), wykonano za pomocą testu sumy rang Kruskala-Wallisa oraz testu współczynnika Spearmana korelacji rangowej na trend wartości średniej i trend wariancji [Lehmann 1959; Yevjevich, Jeng 1969; Pruchnicki 1976]. Zmienne $H_{14}^{1,00}$ -ZIMA, $H_{14}^{1,00}$ -LATO, $H_{14}^{2,89}$ -ZIMA i $H_{14}^{2,89}$ -LATO uznano za stacjonarne, nie wykryto skoku wartości średniej, trendu wartości średniej i wariancji. Nie było powodu, aby odrzucać H_0 w odniesieniu do jednorodności tych danych. Problemem jest występowanie trendu wartości średniej i wariancji w ciągu liczbowym θ_{HBV} -ZIMA i θ_{HBV} -LATO. W związku z tym, że oba ciągi są niezbędnym wejściem do korekcji obciążenia za pomocą metody transformaty dystrybuanty oraz reprezentują ciągi wygenerowane przez model hydrologiczny, postanowiono uznać oba ciągi za stacjonarne przy obniżonym poziomie istotności $\alpha = 0,01$.

II. WYZNACZENIE CHARAKTERYSTYK ROZKŁADU PRAWDOPODOBIEŃSTWA

W celu wyznaczenia charakterystyk rozkładu prawdopodobieństwa zestawiono oddzielne ciągi liczbowe badanych zmiennych losowych, tj.: sumy dobowe opadów P_{OBS} , P_{03} , P_{04} , P_{0524} (bez zdarzeń bezopadowych), oraz ciągi wilgotności gleby pogrupowane wg sezonu (zima, lato), tj. θ_{HBV} -ZIMA, θ_{HBV} -LATO, $H_{14}^{1,00}$ -ZIMA, $H_{14}^{1,00}$ -LATO, $H_{14}^{2,89}$ -ZIMA i $H_{14}^{2,89}$ -LATO. Dla każdej zmiennej losowej określono następujące charakterystyki rozkładu prawdopodobieństwa: asymetrii, koncentracji, rozproszenia i położenia (odpowiednio tab. 2.11 i 2.12). Analiza niejednorodności tych zmiennych (tab. 2.9 i 2.10) wykazała, że wszystkie analizowane zmienne losowe są niezależne (θ_{HBV} -ZIMA i θ_{HBV} -LATO przy poziomie istotności $\alpha = 0,01$). Zatem obliczanie miar rozproszenia rozkładu i tym samym współczynnika skośności i spłaszczenia uwzględniła stosowaną standardowo formułę dla niezależnych zmiennych losowych (2.3). Obliczone charakterystyki prawdopodobieństwa, tj. współczynnik skośności, współczynnik spłaszczenia, wariancja, wartość średnia i mediana, wspierają dobór 1D teoretycznych rozkładów

prawdopodobieństwa w procedurze korekcji obciążenia za pomocą transformaty dystrybucyjnej. Powyższa analiza jest przydatna w wyborze (na podstawie zaobserwowanych wartości zmiennych losowych) zarówno dokładnych funkcji gęstości: $f(P_{OBS})$, $f(P_{03})$, $f(P_{04})$, $f(P_{0524})$, $f(\theta_{HBV-ZIMA})$, $f(\theta_{HBV-LATO})$, $f(h_{14}^{1,00}-ZIMA)$, $f(h_{14}^{1,00}-LATO)$, $f(h_{14}^{2,89}-ZIMA)$, $f(h_{14}^{2,89}-LATO)$, jak i dokładnych ich dystrybucyj.

Tabela 2.11. Charakterystyki rozkładu prawdopodobieństwa zaobserwowanego opadu P_{OBS} i opadów satelitarnych P_{03} , P_{04} i P_{0524} (bez zdarzeń bezopadowych) do korekcji obciążenia metodą transformaty dystrybucyjnej (BC) w modelowanej zlewni Soły (obliczenia uwzględniają niezależność zmiennych losowych).

Table 2.11. Statistical measures of probability distribution for observed precipitation P_{OBS} and satellite precipitation P_{03} , P_{04} and P_{0524} (without precipitation-free events), for the modeled Soła River watershed (the calculations take into account the independence of random variables).

	P_{OBS} (mm)	P_{03} (mm)	P_{04} (mm)	P_{0524} (mm)
Miara asymetrii rozkładu				
Współczynnik skośności	3,241	4,379	4,055	4,647
Miara koncentracji rozkładu				
Współczynnik spłaszczenia	14,891	25,342	23,617	30,729
Miary rozproszenia rozkładu				
Wariancja	59,900 [(mm) ²]	25,342 [(mm) ²]	64,369 [(mm) ²]	47,893 [(mm) ²]
Miary położenia rozkładu				
Średnia arytmetyczna	5,420 (mm)	4,087 (mm)	4,723 (mm)	4,295 (mm)
Mediana	2,500 (mm)	1,400 (mm)	1,700 (mm)	1,900 (mm)

Tabela 2.12. Charakterystyki rozkładu prawdopodobieństwa zmiennych pogrupowanych wg sezonu, tj. wilgotności gleby θ_{HBV} obliczonej przez model HBV oraz satelitarnej wilgotności gleby $H_{14}^{1,00}$ i $H_{14}^{2,89}$, w modelowanej zlewni Soły (obliczenia uwzględniają niezależność zmiennych losowych).

Table 2.12. Statistical measures of probability distribution for grouped (by season) variables, i.e. moisture computed by HBV, θ_{HBV} , and satellite soil moisture $H_{14}^{1,00}$ and $H_{14}^{2,89}$, for the modeled Soła River watershed (the calculations take into account the independence of random variables).

	$\theta_{HBV-ZIMA}$ (%)	$\theta_{HBV-LATO}$ (%)	$H_{14}^{1,00}-ZIMA$ (%)	$H_{14}^{1,00}-LATO$ (%)	$H_{14}^{2,89}-ZIMA$ (%)	$H_{14}^{2,89}-LATO$ (%)
Miara asymetrii rozkładu						
Współczynnik skośności	-1,737	-0,575	-0,393	-0,488	-1,502	-0,925
Miara koncentracji rozkładu						
Współczynnik spłaszczenia	2,706	-0,295	-0,165	-0,274	2,369	1,262
Miary rozproszenia rozkładu						
Wariancja	62,098	67,997	26,775	28,546	35,000	34,489
Miary położenia rozkładu						
Średnia arytmetyczna	69,932	53,821	62,231	63,708	57,604	64,036
Mediana	72,300	55,680	62,470	64,445	60,500	65,110

Zmienne opadowe (tab. 2.11) uzyskały dodatnią skośność, co wskazuje, że ich rozkłady są prawostronnie asymetryczne. Oznacza to, że prawy ogon rozkładów tych zmiennych jest grubszy od lewego. Duże wartości współczynnika spłaszczenia ($>+1,5$) informują nas o dużym rozrzucie uzyskanych wyników, natomiast dodatnie wartości świadczą o tym, że ich rozkład jest bardziej wysmukły (większa koncentracja realizacji zmiennych losowych) w porównaniu do rozkładu normalnego. Powyższy wniosek potwierdza duża różnica między wartością średnią i medianą.

Zmienne wilgotności gleby mają ujemny współczynnik skośności mieszczący się w przedziale $<-1,5; 1,5>$, z wyjątkiem zmiennej $\theta_{HBV-ZIMA}$ ($-1,737$). Oznacza to, że ich rozkłady są asymetryczne, ale zbliżone do normalnego, z nieco wydłużonym lewym ramieniem rozkładu. Niewielki ujemny współczynnik spłaszczenia mieszczący się w przedziale $<-3; 3,0>$ świadczy o mniejszej koncentracji uzyskanych wyników wokół wartości średniej. Rozkład jest lekko spłaszczony w porównaniu do rozkładu normalnego, jednocześnie zbliżony do niego, co potwierdziły małe różnice między wartością średnią a medianą.

III. WYBÓR NAJLEPIEJ DOPASOWANEJ FUNKCJI ROZKŁADU PRAWDOPODOBIEŃSTWA DLA OPADU P I WILGOTNOŚCI GLEBY θ NA PODSTAWIE TESTU KOŁMOGOROWA-SMIRNOWA

Wejściem do procedury były dobowe sumy opadów P_{OBS} , P_{03} , P_{04} , $P_{05,24}$ (bez zdarzeń bezopadowych) oraz ciągi wilgotności gleby pogrupowane wg sezonu (zima, lato), tj. $\theta_{HBV-ZIMA}$, $\theta_{HBV-LATO}$, $H_{14}^{1,00-ZIMA}$, $H_{14}^{1,00-LATO}$, $H_{14}^{2,89-ZIMA}$ i $H_{14}^{2,89-LATO}$. Analiza niejednorodności wykazała, że zarówno opady, jak i obserwacje wilgotności gleby można uznać za niezależne i stacjonarne (w przypadku zmiennej $\theta_{HBV-ZIMA}$ i $\theta_{HBV-LATO}$ hipotezę H_0 w odniesieniu do jednorodności tych danych przyjęto przy obniżonym poziomie istotności $\alpha = 0,01$). Dla uproszczenia zapisu pogrupowane zmienne losowe wilgotności gleby będą oznaczane jako θ_{HBV} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$ z zaznaczeniem sezonu występowania, np. w sezonie zimowym lub letnim.

Dokonano wyboru najlepiej dopasowanego rozkładu do badanej zmiennej losowej na podstawie testu zgodności Kołmogorowa-Smirnowa [Genest i in. 2009]. Na podstawie uzyskanych charakterystyk rozkładu prawdopodobieństwa (rozkłady asymetryczne) postanowiono do obliczeń zastosować trójparametrowe $[\alpha, \lambda, i \varepsilon]$, równania (1-6) w Dodatku A] rozkłady prawdopodobieństwa GE, GA i WE. Wartości testu porównano z wartościami krytycznymi statystyki (tab. 2.13). Podkreślone wartości obliczonej statystyki wskazują rozkłady, dla których statystyka Kołmogorowa-Smirnowa odrzuciła hipotezę zerową na poziomie $\alpha = 0,05$. Rozkładem prawdopodobieństwa, który najlepiej dopasował się do P_{OBS} był uogólniony rozkład wykładniczy GE, a do opadów P_{03} , P_{04} i $P_{05,24}$ rozkład Weibulla WE. W przypadku wilgotności gleby dopasowanie rozkładów prawdopodobieństwa podzielono na dwa sezony. Zimą najlepiej dopasowanym rozkładem teoretycznym do θ_{HBV} był rozkład Weibulla, WE, do $H_{14}^{1,00}$ – gamma, GA, do $H_{14}^{2,89}$ – Weibulla, WE. Latem najlepiej dopasowanym rozkładem do θ_{HBV} był rozkład Weibulla, WE, do $H_{14}^{1,00}$ – gamma, GA, do $H_{14}^{2,89}$ – rozkład gamma, GA.

Wybrane rozkłady prawdopodobieństwa wykorzystano w kolejnych etapach przetwarzania, tj. do walidacji, korekcy obciążenia, modelowania i asymilacji za pomocą wiązkowych filtrów Kalmana.

Tabela 2.13. Wybór rozkładów prawdopodobieństwa użytych do korekcji obciążenia danych H-SAF, tj. opadu i wilgotności gleby P_{03}, P_{04}, P_{0524} (bez zdarzeń bezopadowych) i $H_{14}^{1,00}, H_{14}^{2,89}$ (pogrupowanych wg sezonu: ZIMA i LATO), gdzie: WE – rozkład Weibulla; GA – rozkład gamma; GE – rozkład uogólniony wykładniczy. Pogrubione wartości reprezentują najlepiej dopasowane rozkłady prawdopodobieństwa. Podkreślone wartości wskazują na odrzucenie hipotezy o zgodności rozkładu empirycznego wg statystyki Kołmogorowa-Smirnowa, (K-S) przy poziomie istotności $\alpha = 0,05$.

Table 2.13. Choice of the distribution used in the bias correction of precipitation and soil moisture H-SAF products: P_{03}, P_{04}, P_{0524} (without precipitation-free events) and $H_{14}^{1,00}, H_{14}^{2,89}$ (grouped by season: WINTER and SUMMER), where: WE – Weibull; GA – gamma; GE – generalized exponential distribution. Bolded values represent the best fitted distributions, while the underlined features indicate distributions for which the Kolmogorov-Smirnov statistics, (K-S) rejected the null hypothesis at $\alpha = 0.05$.

Dane	Rozkład prawdopodobieństwa				Test Kołmogorowa-Smirnowa λ	
	Typ	Współczynniki				
		α	λ	ε		
Opad					$\lambda_{\alpha=0,05}^{\text{kryt.}} = 0,1398$	
P_{OBS}	WE	0,0091	0,3389	0,0990	0,1305	
	GA	0,9979	0,2313	0,0000	0,1240	
	GE	1,2140	4,9784	0,0000	0,1233	
P_{03}	WE	0,0070	0,3451	0,0990	0,0877	
	GA	0,8010	0,2400	0,0280	0,1081	
	GE	1,2511	5,3353	0,0000	0,0878	
P_{04}	WE	0,0059	0,3523	0,0990	0,0704	
	GA	0,9143	0,1911	0,0250	0,0858	
	GE	1,5346	6,7433	0,0000	0,0734	
P_{0524}	WE	0,0130	0,3239	0,0990	0,0965	
	GA	0,7193	0,1824	0,0540	0,1039	
	GE	1,8952	8,2915	0,0000	0,0977	
Wilgotność gleby					$\lambda_{\alpha=0,05}^{\text{kryt.}} = 0,0721$	
Sezon zimowy	θ_{HBV}	WE	34,4556	5,6428	38,00	0,0342
		GA	69,1987	1,0107	0,00	<u>0,0914</u>
		GE	11,3523	0,0885	36,00	<u>0,0935</u>
	$H_{14}^{1,00}$	WE	20,6861	4,0103	42,40	<u>0,0896</u>
		GA	50,8409	0,8123	20,00	0,0399
		GE	12,1816	0,1415	40,00	<u>0,1638</u>
	$H_{14}^{2,89}$	WE	40,6269	3,8497	19,80	0,0556
		GA	19,8836	2,8677	0,00	<u>0,1441</u>
		GE	5,9666	0,0631	18,60	<u>0,1975</u>
					$\lambda_{\alpha=0,05}^{\text{kryt.}} = 0,0729$	
Sezon letni	θ_{HBV}	WE	33,2265	2,9743	22,80	0,0507
		GA	20,0311	2,6243	0,00	<u>0,1015</u>
		GE	9,6627	0,0820	18,00	<u>0,1171</u>
	$H_{14}^{1,00}$	WE	17,8183	2,9249	46,50	<u>0,0847</u>
		GA	61,1260	0,7925	14,00	0,0451
		GE	8,2893	0,1502	44,50	<u>0,1180</u>
	$H_{14}^{2,89}$	WE	26,5781	4,6010	38,80	<u>0,0966</u>
		GA	66,0066	0,8109	9,60	0,0367
		GE	28,7413	0,1320	34,00	<u>0,1423</u>

3. METODY I MODELE

3.1. Metody i modele użyte do analizy niepewności w prognozowaniu hydrologicznym

W prognozowaniu hydrologicznym szczególną rolę odgrywa umiejętność uwzględniania w obliczeniach niepewności. Ten kluczowy problem prognostyczny można rozwiązać za pomocą:

- i) metod i modeli probabilistycznych;
- ii) poprawy symulacji modelu hydrologicznego z uwzględnieniem: satelitarnych obserwacji opadu jako danych wymuszających, i updatingu zmiennych stanu modelu hydrologicznego z wykorzystaniem metod asymilacji satelitarnych obserwacji meteorologicznych.

W pierwszym przypadku zastosowanie metod i modeli probabilistycznych uwzględnia niepewność prognozy hydrologicznej, zawierając ją ilościowo, np. w prognozie przedziałowej opierającej się zarówno o prognozę punktową, jak i błąd prognozy. Prognoza przedziałowa jest przedziałem liczbowym, w którym z zadaniem prawdopodobieństwem zawarta jest nieznaną wartość zmiennej prognozowanej w okresie prognostycznym, czyli w czasie objętym przewidywaniem, zaś prognoza punktowa jest liczbą, która przyjmowana jest za najlepszą ocenę zmiennej prognozowanej.

W modelach probabilistycznych prawdopodobieństwo zdarzenia jest miarą (oceną) możliwości zajścia tego zdarzenia w określonych warunkach. Zaproponowane w niniejszej pracy metody i modele matematyczne powinny uwzględniać błędy i straty informacyjne wynikające głównie z niedostatecznej wiedzy hydrologa o prognozowanym zjawisku. Powyższy problem można rozwiązać za pomocą podejścia klasycznego, które opiera się na wnioskowaniu statystycznym opartym wyłącznie na próbie losowej, na podstawie której odbywa się wnioskowanie na temat całej badanej populacji (np. wezbrań, niszówek hydrologicznych).

Jednak stosując metody i modele matematyczne w prognozowaniu hydrologicznym, powinniśmy mieć możliwość skorzystania nie tylko z wyników zaobserwowanych w próbie, ale także z informacji *a priori* dotyczących rozważanego problemu. Takie podejście do problemu probabilistyki występuje we wnioskowaniu Bayesowskim, stąd w niniejszej pracy – zarówno w modelach statycznych, jak i dynamicznych – równania prognostyczne powinny zawierać prawdopodobieństwo Bayesowskie.

Drugim rozwiązaniem uwzględniającym niepewność w prognozie hydrologicznej jest korekta rozbieżności między hydrogramem symulowanym a obserwowanym w momencie stawiania prognozy, co stanowi podstawową przyczynę błędnej prognozy. Powyższy problem można rozwiązać, wprowadzając do modelu hydrologicznego satelitarne obserwacje opadu jako dane wymuszające i asymilując w modelu hydrologicznym satelitarne obserwacje wilgotności gleby. Z jednej strony większa rozdzielczość regularnej siatki obserwacji satelitarnych (od 1 km do 10 km w przypadku satelitów geostacjonarnych) zapewnia dokładniejszy rozkład przestrzenny opadu atmosferycznego w porównaniu do rzadkiej i nieregularnej naziemnej sieci pomiarowo-obserwacyjnej (rozdzielczość od 10 km do 30 km) [Sidur 2018]. Z drugiej strony wiele produktów satelitarnych cechuje słaba ilościowa jakość i błąd systematyczny, co niekorzystnie wpływa na przetwarzanie w modelach hydrologicznych.

Matematyczne metody i modele, zaproponowane w niniejszej pracy do symulacji hydrologicznej i prognozy, powinny aktualizować zmienne stanu modelu hydrologicznego w ramach updatingu wykorzystując obserwacje satelitarne. Problem polega na tym, że aktualizowane zmienne stanu w modelach hydrologicznych albo nie mają swoich odpowiedników w naziemnych obserwacjach i pomiarach, albo wyznaczane są na podstawie charakterystyk fizyczno-geograficznych zlewni, bilansów wodnych itp., co wpływa na niedokładność ich oszacowania. Procedura updatingu zaproponowana w pracy powinna wykorzystywać asymilację obserwacji satelitarnych. Z takim przetwarzaniem związane są problemy nieliniowości zjawisk hydrologicznych, obciążenia symulowanej wiązki zaburzonej szumem Gaussa, występowania błędu losowego, z którymi zaproponowane w pracy metody i modele powinny sobie poradzić.

Przedstawione w niniejszym rozdziale metody i modele probabilistyczne powinny wspomóc hydrologa w opracowaniu różnych typów prognozy hydrologicznej, np.:

- 1) Prognozy średnioterminowej charakterystyk ilościowych wezbrania o czasie wyprzedzenia od dwóch do dziesięć dób [Byczkowski 1999]. Prognoza powinna być obliczona na podstawie pomierzonych lub zaobserwowanych charakterystyk meteorologicznych, takich jak: opady, pokrywa śniegu, wodny ekwiwalent śniegu, temperatura powietrza itp., stanowiących bezpośrednią przyczynę prognozowanego zjawiska hydrologicznego. Równania prognostyczne powinny wynikać ze związków zachodzących między zaobserwowanym stanem zlewni a przewidywanym odpływem.
- 2) Prognozy krótkoterminowej przepływu rzecznoego z wyprzedzeniem 1-, 2- i 3-dobowym, odpowiednio z krokiem 1-dobowym i 1-godzinnym. Prognozowany przepływ rzeczny powinien być uzależniony od obserwacji z poprzedniego kroku obliczeniowego i opisany zależnością przyczynowo-skutkową, tzn. zmienne prognozujące powinny być przyczyną zmiennej prognozowanej.
- 3) Poprawie krótkoterminowej prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym, poprzez korektę satelitarnego opadu potraktowanego jako wymuszenie modelu hydrologicznego i asymilację satelitarnych obserwacji wilgotności gleby w ramach procedury updatingu zmiennych stanu modelu hydrologicznego.

3.2. Wielokrotna liniowa regresja (WLR) i wielowymiarowa wielokrotna liniowa regresja (WWLR)

Jednym z celów matematycznego modelowania jest dopasowanie modelu do obserwowanych danych. Często stosowanym rozwiązaniem są modele regresyjne, które formalnie opisują stochastyczne zależności różnego rodzaju zjawisk od czynników je kształtujących i wyrażone są w formie odpowiedniego równania matematycznego. Przykładem są modele prostej regresji, wielokrotnej liniowej regresji (WLR) lub wielowymiarowej wielokrotnej liniowej regresji (WWLR) – w tym ostatnim zmienną prognozowaną opisano nie tylko jako skalar, ale również jako wektor. Równania regresji opisują zjawisko przyrodnicze w taki sposób, że można wyjaśnić je za pomocą zależności przyczynowo-skutkowej (np. równania regresji mogą opisywać związek między stanami wody, przepływami, odpływem i opadem). Do opisu powyższych relacji najczęściej używano funkcji wielomianowych pierwszego stopnia. Mashayekhi i Mahjoub [1991] zastosowali WLR do prognozowania objętości odpływu rzek w celu zarządzania zasobami zbiornika wielozadaniowego Kardj w Iranie. Objętość odpływu z topniejącej pokrywy śniegu uzależniono od sumy opadów, przepływu bazowego, wodnego ekwiwalentu śniegu, średniej miesięcznej temperatury powietrza i prawdopodobnego opadu. Regresja liniowa jest również często wykorzystywana do opisu nieliniowej zależności z uwzględnieniem odpowiedniej transformacji zmiennych prognozujących i prognozowanych [Lefebvre 2003]. Transformacja oparta jest na zlogarytmizowaniu zmiennych obu stron równania regresji.

Wiele zastosowań metod regresji można znaleźć w literaturze hydrologicznej [Holder 1985; Adamowski i in. 2012], w szczególności w odniesieniu do średnio- i długoterminowych prognoz. Mahabir i in. [2006] wykorzystał wielokrotną liniową regresję (WLR) do prognozowania maksymalnych poziomów wody podczas wezbrania wiosennego. Gelfan i Motovilov [2009] zastosowali metodę indeksu, wykorzystującą liniowe i nieliniowe modele regresji zgodnie z regresją opartą na graficznej metodzie krzywych, do opisanego wpływu długoterminowej akumulacji śniegu na warunki odpływu powierzchniowego. Analiza WLR i WWLR koncentruje się głównie na dwóch ważnych kwestiach metodologicznych: wybraniu optymalnego zbioru zmiennych prognozowanych i prognozujących i oszacowaniu parametrów regresji. Wielu autorów wykorzystało podejście najmniejszych kwadratów do oszacowania parametrów [Garen 1992].

ESTYMACJA PARAMETRÓW MODELI WLR I WWLR

Funkcja regresji WLR uzależnia zmienną prognozowaną Y od zmiennych prognozujących X i wyrażona jest jako [Ozga-Zielińska, Brzeziński 1997]:

$$y = f(x_1, x_2, \dots, x_d) = E(y|x_1, x_2, \dots, x_d) \quad (3.1)$$

gdzie: ostatnie wyrażenie oznacza oczekiwaną wartość zmiennej prognozowanej Y uwarunkowaną zmiennymi x_1, x_2, \dots, x_d , a d – liczbę zmiennych prognozujących.

Kiedy zjawisko hydrologiczne opisane jest przez więcej niż jedną zmienną prognozowaną, stosowany jest układ równań zależności, tzn. dla każdej zmiennej prognozowanej wyznacza się oddzielne równanie [Darlington 1990]:

$$y_p = f(y_p | x_1, x_2, \dots, x_d) \text{ dla } p = 1, \dots, l \quad (3.2)$$

gdzie l jest liczbą zmiennych prognozowanych.

W przypadku jednoczesnego prognozowania kilku zmiennych prognozowanych należy zastosować funkcję WWLR w postaci hiperpowierzchni regresji:

$$\mathbf{y} = f(\mathbf{x}) = E(\mathbf{y}|\mathbf{x}) = E(y_1, y_2, \dots, y_l | x_1, x_2, \dots, x_d) \quad (3.3)$$

gdzie: \mathbf{x} jest wektorem realizacji wielowymiarowej zmiennej prognozującej x_1, x_2, \dots, x_d , \mathbf{y} jest wektorem realizacji wielowymiarowej zmiennej prognozowanej y_1, y_2, \dots, y_l .

Jeżeli założymy, że przestrzeń obserwacji modelu WLR tworzy n -wymiarowy wektor z zaobserwowanymi wartościami zmiennej prognozowanej \mathbf{Y} i $n \times d$ wymiarową macierz zmiennych prognozujących \mathbf{X} , gdzie n jest liczbą obserwacji, to model regresji WLR ma następującą postać:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3.4)$$

gdzie $\boldsymbol{\beta}$ jest d wymiarowym wektorem współczynników kierunkowych, a $\boldsymbol{\epsilon}$ jest n wymiarowym wektorem błędów modelu WLR.

Niech \mathbf{b} będzie d wymiarowym wektorem przykładowych wartości dla $\boldsymbol{\beta}$. Metoda najmniejszych kwadratów wybiera takie \mathbf{b} , które minimalizuje sumę kwadratów odchyłeń, tj.:

$$S(\mathbf{b}) = \sum_j^n (y_j - b_0 - b_1 x_{j1} - \dots - b_d x_{jd})^2 = (\mathbf{Y} - \mathbf{X}\mathbf{b})^T (\mathbf{Y} - \mathbf{X}\mathbf{b}) \quad (3.5)$$

gdzie: $S(\mathbf{b})$ jest funkcją celu; y_j są realizacjami zmiennej losowej \mathbf{Y} dla $j = 1, \dots, n$; x_{ji} są realizacjami zmiennej losowej \mathbf{X} dla $i = 1, \dots, d$; a $(\mathbf{Y} - \mathbf{X}\mathbf{b})^T$ oznacza macierz transponowaną.

W modelu WWLR przestrzeń obserwacji modelu tworzą obserwacje historyczne obejmujące $n \times l$ wymiarową macierz z obserwacjami zmiennych prognozowanych \mathbf{Y} , gdzie: n jest liczbą obserwacji; l jest liczbą zmiennych prognozowanych i $n \times d$ wymiarową macierz z obserwacjami dla zmiennych prognozujących \mathbf{X} . Model dla p -tej zmiennej prognozowanej $\mathbf{Y}_{(p)}$ spełnia model regresji:

$$\mathbf{Y}_{(p)} = \mathbf{X}\boldsymbol{\beta}_{(p)} + \boldsymbol{\epsilon}_{(p)} \text{ dla } p = 1, 2, \dots, l \quad (3.6)$$

$$\text{z } \text{Cov}(\boldsymbol{\epsilon}_{(p)}) = \sigma_{pp} \mathbf{E}$$

gdzie: $\boldsymbol{\beta}_{(p)}$ jest d wymiarowym wektorem współczynników kierunkowych dla p -tej zmiennej prognozowanej; $\boldsymbol{\epsilon}_{(p)}$ jest n wymiarowym wektorem błędów modelu dla p -tej zmiennej prognozowanej; \mathbf{E} jest $n \times n$ wymiarową macierz jednostkową.

Błędy dla różnych zmiennych prognozowanych w tej samej próbie losowej mogą być skorelowane. Biorąc pod uwagę wyniki macierzy \mathbf{Y} i wartości macierzy zmiennych prognozujących \mathbf{X} , można ustalić estymator najmniejszych kwadratów $\hat{\boldsymbol{\beta}}_{(p)}$ dla obserwacji $\mathbf{Y}_{(p)}$:

$$\hat{\boldsymbol{\beta}}_{(p)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_{(p)} \quad (3.7)$$

gdzie \mathbf{X}^T oznacza macierz transponowaną.

3.3. Statyczne liniowe modele Bayesowskie (SLMB)

W wyborze modelu hydrologicznego do symulacji, a zwłaszcza prognozy hydrologicznej, ważnym elementem jest umiejętność opisu i prezentowania niepewności poprzez:

- 1) Przydzielanie niepewności regułom wnioskowania.
- 2) Łączenie niepewnych informacji pochodzących z różnych źródeł.
- 3) Uwzględnianie propagacji niepewności w trakcie procesu wnioskowania [Carpenter 2003; Howson, Urbach 2005; Sikorska i in. 2012].

Rozwiązaniem powyższych problemów są modele z losowymi parametrami, tj. modele Bayesowskie, które umożliwiają modelowanie niepewności. Teoria Bayesowska została zastosowana w procesach Gaussowskich w połączeniu z wielowymiarowym modelem regresji [Neal 1997]. Na podstawie rozkładów *a posteriori*, oszacowanych na bazie informacji uzyskanej *a priori*, modele wielowymiarowe utworzyły rozkłady brzegowe zmiennych prognozowanych i prognozujących. Jednym z takich modeli jest statyczny liniowy model Bayesowski (SLMB) [Carpenter 2003; Howson, Urbach 2005], wykorzystujący wielokrotną liniową regresję (WLR) z parametrami estymowanymi metodą Bayesowską (WLR-B) lub techniką odwrotnego modelowania (Bayesian inverse modelling) (WLR-InvB). Estymacja parametrów za pomocą techniki odwrotnego modelowania (WLR-InvB) jest częściej praktykowana w modelach dynamicznych [Dowd, Meyer 2003; Scharnagl i in. 2011], których ewolucja wyznaczona jest jednoznacznie za pomocą równania stanu, a zmienne stanu zmieniają się z upływem czasu, czyli są funkcjami czasu.

Rozwinięcie wielokrotnej liniowej regresji do postaci wielowymiarowej może również uwzględnić wnioskowanie Bayesowskie. W tym celu wielokrotna liniowa regresja WLR została zmodyfikowana do postaci umożliwiającej estymację wielowymiarowej zmiennej prognozowanej metodami analizy Bayesa, tj.: WWLR-B i WWLR-InvB [Rosen, Thomson 2009; Royle i in. 2014]. Parametry większości z tych modeli oszacowywane są za pomocą klasycznych metod estymacji, tj. metodą najmniejszych kwadratów lub metodą największej wiarygodności. Natomiast zastosowanie podejścia Bayesowskiego (w sensie Bayesowskiego kryterium informacyjnego [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 38]) zwiększyło efektywność predykcijną modeli dzięki możliwości wybierania zbiorów zmiennych o najwyższej informacyjności.

SLMB stosowane są zarówno dla zmiennych losowych o rozkładzie normalnym, jak i dla zmiennych o innych rozkładach prawdopodobieństwa. Pełna analiza Bayesowska rozkładów *a posteriori* jest trudna obliczeniowo. Wykorzystywane są różne metody do przybliżania wartości parametrów. Przykładem takich rozwiązań są linearyzacja funkcji rozkładu w pobliżu optymalnych oszacowań parametrów lub metoda bootstrapowa pozwalająca obliczyć przybliżoną macierz kowariancji parametrów rozkładu.

Do estymacji parametrów modeli Bayesowskich stosowane są algorytmy Monte Carlo wykorzystujące łańcuchy Markowa (Markov Chain Monte Carlo, MCMC) [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 27]. Techniki bezpośredniej symulacji są szczególnie skuteczne przy wyznaczeniu rozkładu *a priori* w przypadku niepewności. W takich sytuacjach używane są rozkłady nieinformacyjne, np. rozkład aprioryczny Jeffreysa [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 31] lub rozkład jednostajny. Algorytmy MCMC stosowane są również do estymacji parametrów najprostszych modeli Bayesowskich. W takim przypadku algorytmy MCMC są elementem próbkowania Gibbsa [Walsh 2004] i metody Metropolisa-Hastingsa [Chib, Greenberg 1995; Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 32, 33]. Badanie Bayesowskich metod MCMC do oceny rozkładów *a posteriori* kwantyli prawdopodobieństwa przepływów kulminacyjnych wezbrań, ryzyka powodzi, parametrów takich rozkładów jak log-normalny (LN), log-Pearsona 3 typu (LGA) [Reis i in. 2005] wykazało, że metody MCMC pozwalają na bogatszą i pełniejszą prezentację informacji o zaobserwowanych wezbraniach.

3.3.1. Wnioskowanie i reguła Bayesowska

Ze względu na niepowtarzalny charakter zdarzeń, wezbrania są trudne do probabilistycznego zdefiniowania na bazie klasycznej (tzn. częstościowej) definicji prawdopodobieństwa. W ujęciu klasycznym (niebayesowskim) prawdopodobieństwo zdarzenia losowego w praktyce jest reprezentowane za pomocą częstości występowania tego zdarzenia w bardzo dużej liczbie identycznych prób [Rohatgi

1976]. Wnioskowanie częstościowe na podstawie informacji pochodzących z tych prób pozwala jedynie oszacować nieznanne wartości parametrów rozkładu. Natomiast wnioskowanie Bayesowskie wykorzystuje również wstępną wiedzę badacza o parametrach modelu. W tym przypadku stosuje się prawdopodobieństwo bezwarunkowe, tzw. *a priori*, wystąpienia zdarzenia losowego (np. wezbrania), które jest niczym innym, jak miarą racjonalnego przekonania, że dane zdarzenie wystąpi. Jeśli chcemy zmienić (zmodyfikować, wzbogacić) nasze przekonania, powinniśmy zidentyfikować dodatkowe wezbrania, wykonać pomiary i opracować ich charakterystyki ilościowe. Wyniki badań przekształcają prawdopodobieństwo *a priori* (wstępne oczekiwania) w tzw. prawdopodobieństwo *a posteriori* (prawdopodobieństwo wynikowe, czyli miarę racjonalnego oczekiwania wystąpienia zdarzenia po uzyskaniu wyników badań). W opisanym rozumowaniu prawdopodobieństwo częstościowe może być pomocne w określeniu wiarygodności Bayesowskiej lub prawdopodobieństw *a priorycznych* (wyjściowych).

Do modelowania wezbrań wybrano wnioskowanie Bayesowskie z następujących względów:

- 1) Ma zastosowanie zarówno do zjawisk powtarzalnych jak i niepowtarzalnych.
- 2) Prawdopodobieństwo staje się opisem niepewności niezależnie od jej pochodzenia. Na równych prawach traktowane są błędy statystyczne, tzn. niepewność wynikająca z błędów pomiarowych, czyli ze skończonej dokładności przyrządów oraz błędy systematyczne, czyli niepewność związana z naszą niewiedzą odnośnie procesów deterministycznych.
- 3) Używa dwóch źródeł informacji, tj. *a priori* i *a posteriori*.
- 4) Wnioskowanie Bayesowskie pokazuje, jak nowe obserwacje modyfikują wstępne oczekiwania, tj. prawdopodobieństwo *a priori*.
- 5) Statystyka Bayesowska odnosi się tylko do danych, które faktycznie zostały zaobserwowane, natomiast statystyka częstościowa do rozkładu wyników, które są możliwe, ale nie zostały zaobserwowane.
- 6) Wnioskowanie Bayesowskie pozwala na bezpośrednie podanie prawdopodobieństwa parametrów.

Wnioskowanie Bayesowskie tworzy związek między zbiorem parametrów θ a zbiorem zaobserwowanych zmiennych prognozowanych \mathbf{Y} poprzez wyznaczenie:

- 1) Funkcji rozkładu prawdopodobieństwa parametrów θ uzależnionego od stanu zmiennych prognozowanych.
- 2) Przewidywanego rozkładu prawdopodobieństwa danych niezaobserwowanych.

Parametry θ mają pewną ustaloną wartość, lecz obarczone są niepewnością, natomiast rozkład prawdopodobieństwa jest uwarunkowany posiadaną informacją o rozpatrywanym zjawisku oraz danymi obserwacyjnymi i pomiarowymi. We wnioskowaniu Bayesowskim funkcja rozkładu prawdopodobieństwa jest sposobem uwzględnienia niepewności konkurujących hipotez H_j (modeli) związanych z zestawem parametrów modelu $\theta = (\theta_1, \dots, \theta_{k_j})$, gdzie k_j określa liczbę parametrów θ w hipotezie (modelu) H_j przy danym stanie wiedzy. Zakładamy, że postawione zostały hipotezy (modele) H_j ze zbiorem parametrów θ , które zostaną użyte do modelowania danych \mathbf{Y} , tj. $H_j = H_j(\theta)$, dla $j = 1, \dots, m$, gdzie m jest liczbą możliwych hipotez. Hipotezy mogą dotyczyć spodziewanych zależności (relacji) między jakimś zjawiskiem, np. między wielkością wezbrania a bieżącymi lub prognozowanymi warunkami hydrologicznymi i meteorologicznymi w zlewni.

Matematyczny model Bayesowski zawiera dwa poziomy wnioskowania [Gibbs 1997]:

- 1) Obliczenie prawdopodobieństwa *a posteriori* hipotez H_j uwarunkowanych zbiorem zmiennych prognozowanych \mathbf{Y} , $p(\theta|\mathbf{Y}, H_j)$.
- 2) Wyznaczenie wiarygodności globalnej w celu porównania różnych hipotez H_j z zaobserwowanym zbiorem parametrów θ , $p(\mathbf{Y}|H_j)$.

W teorii Bayesa do realizacji obu poziomów wnioskowania wykorzystuje się rozkłady prawdopodobieństwa. Rozumowanie rozpoczyna się od wyrażenia naszych wcześniejszych oczekiwań o tym, które hipotezy są *a priori* odpowiednie w odniesieniu do danych, w sensie rozkładu prawdopodobieństwa wszystkich możliwych hipotez $p(H_j)$. Zawsze posiadamy *a priori* pewną wiedzę (informację) odnośnie badanego zjawiska i na jej podstawie formułujemy hipotezy H_j . Wiedzę tę wyraża prawdopodobieństwo hipotezy *a priori* $p(\theta|H_j)$, które jest miarą naszych wstępnych oczekiwań co do słuszności

hipotezy H_j zanim jeszcze zobaczymy dane. Wiedza *a priori* o parametrach θ jest zależna od ich funkcji w modelu, a zatem ten rozkład jest uwarunkowany hipotezą H_j .

Na pierwszym poziomie wnioskujemy o wartościach parametrów θ na podstawie danych \mathbf{Y} . Dla każdej hipotezy H_j tworzona jest prognoza dotycząca prawdopodobieństwa, że dane zostały wygenerowane przez ten model o parametrach θ . Powyższe przewidywania są zawarte w rozkładzie $p(\mathbf{Y}|\theta, H_j)$, który oznacza prawdopodobieństwo otrzymania danych \mathbf{Y} pod warunkiem słuszności hipotezy H_j (tzw. Bayesowska funkcja wiarygodności). Musimy wiedzieć, jak połączyć wiedzę *a priori* o zbiorze parametrów θ , $p(\theta|H_j)$, z wiedzą którą pozyskaliśmy z danych $p(\mathbf{Y}|\theta, H_j)$. W tym celu odwołujemy się do teorii Bayesa [Gibbs 1997]:

$$p(\theta|\mathbf{Y}, H_j) = \frac{p(\mathbf{Y}|\theta, H_j)p(\theta|H_j)}{p(\mathbf{Y}|H_j)} \quad (3.8)$$

Całkowita wiedza, jaką mamy o wartościach parametrów θ , która jest wyrażona za pomocą prawdopodobieństwa *a posteriori* $p(\theta|\mathbf{Y}, H_j)$, w świetle uzyskanych danych \mathbf{Y} jest iloczynem $p(\mathbf{Y}|\theta, H_j)$ i $p(\theta|H_j)$ z $p(\mathbf{Y}|H_j)$ potraktowanym jako czynnik normalizacyjny – niezależny od hipotez H_j , tzw. wiarygodność globalna, czyli ewidencja. W terminologii Bayesowskiej powszechnie stosowane są następujące określenia: $p(\theta|\mathbf{Y}, H_j)$ – ‘Posteriori’, $p(\mathbf{Y}|\theta, H_j)$ – ‘Likelihood’, $p(\theta|H_j)$ – ‘Prior’, i $p(\mathbf{Y}|H_j)$ – ‘Evidence’.

Na drugim poziomie wnioskowania Bayesowskiego hipoteza H_j jest porównywana z innymi hipotezami przy pomocy algorytmu rankingowego. W celu porównania różnych hipotez H_j z określonym zbiorem parametrów θ , należy oszacować wiarygodność globalną (3.9) wyrażoną jako [Gibbs 1997]:

$$p(\mathbf{Y}|H_j) = \int d\theta \cdot p(\mathbf{Y}|\theta, H_j) \cdot p(\theta|H_j) \quad (3.9)$$

Parametry θ różnych hipotez H_j utworzonych na podstawie danych zaobserwowanych na pierwszym poziomie wnioskowania są trudne do porównywania, ponieważ zarówno Bayesowska funkcja wiarygodności Likelihood, jak i prawdopodobieństwo *a priori* są na ogół nieliniową funkcją parametrów. Postać prawdopodobieństwa funkcji wiarygodności jest często oczywista. Występuje w postaci wykładniczej sumy kwadratów błędów dla regresji z szumem Gaussa i entropii krzyżowej (*cross-entropy*) [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 34] w ocenie poprawności predykcji (klasyfikacji). Biorąc pod uwagę trudności związane z całkowaniem równania (3.9), w praktyce całka jest przybliżana za pomocą algorytmu opartego na metodzie symulacji Monte Carlo [Chib, Greenberg 1995], tzw. MCMC [szczegóły w Załączniku: Dodatek A].

W celu lepszego poznania badanego zjawiska wykonywane są obserwacje i pomiary odzwierciedlające nasze dotychczasowe jego zrozumienie. Wnioskowanie Bayesowskie umożliwia aktualizację nowego prawdopodobieństwa *a posteriori* takiej hipotezy w świetle napływających danych (obserwacji) i jej dotychczasowego prawdopodobieństwa *a priori*. Po wykonaniu nowych badań i na podstawie uzyskanych wyników przechodzimy do drugiego poziomu wnioskowania Bayesowskiego. Biorąc pod uwagę fakt, że możemy obliczyć $p(\mathbf{Y}|H_j)$, tzn. globalną ewidencję (wiarygodność) z równania (3.9), ponownie stosujemy teorię Bayesa. Ewidencja $p(\mathbf{Y}|H_j)$ wyraża wiedzę o tym, czy i w jakim stopniu nowe obserwacje i pomiary zwiększają nasze poznanie badanego zjawiska.

Nie wszystkie parametry modelu Bayesowskiego są tego samego typu. Reprezentujący drugi poziom wnioskowania Bayesowskiego wyższy poziom parametrów lub hiperparametrów, np. szum wariancji czy stałe regularyzacji [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 35, 36], definiują ciągłą przestrzeń modelu. Eksploracja hipotetycznej przestrzeni polega na ciągłym poszukiwaniu hiperparametrów i dyskretnym poszukiwaniu parametrów architektonicznych modelu. Analizowanie ciągle jest korzystne, ponieważ automatyzacja powyższego procesu umożliwia stosowanie bardziej wyrafinowanych algorytmów gradientowych. Proces ten nazywany jest optymalizacją parametrów. Jeżeli wiarygodność $p(\mathbf{Y}|H_j)$ nie może być oszacowana analitycznie na każdym etapie procesu poszukiwania, to do optymalizacji parametrów θ stosowana jest większa symulacja Monte Carlo.

W wielu zastosowaniach wnioskowania Bayesowskiego mamy do czynienia z problemem, w którym pewien model jest z założenia słuszny. Dla ustalonego modelu (hipotezy) $H_j, j = j_0$ przestrzeń hipotez H_{j_0} dotyczy wyłącznie wartości parametrów tego modelu. Hipotezą jest wówczas wartość estymowanego wektora parametrów θ . W takiej sytuacji model Bayesowski (3.8) upraszcza się; łączny rozkład

prawdopodobieństwa parametrów θ i obserwacji \mathbf{Y} może zostać zdekomponowany do dwóch komponentów – $p(\mathbf{Y}|\theta)$ i $p(\theta)$. W związku z tym model próbkowy $p(\mathbf{Y}|\theta)$, czyli warunkowy przy ustalonym wektorze parametrów rozkładu obserwacji, łączymy z naszym oczekiwaniem *a priori* $p(\theta)$, czyli rozkładem parametrów przed wglądem w dane [Gibbs 1997]:

$$p(\theta|\mathbf{Y}) = \frac{p(\mathbf{Y},\theta)}{p(\mathbf{Y})} = \frac{p(\mathbf{Y}|\theta) \cdot p(\theta)}{p(\mathbf{Y})} \quad (3.10)$$

gdzie: $p(\theta|\mathbf{Y})$ oznacza rozkład *a posteriori*, który mierzy stopień poparcia danego wektorowi parametrów θ przez dane pomiarowo-obszawacyjne \mathbf{Y} , tzn. mierzy w jakim stopniu dane \mathbf{Y} rozróżniają wektor parametrów θ od innych alternatywnych wektorów. Rozkład *a posteriori* odzwierciedla naszą końcową wiedzę o parametrach modelu θ ; $p(\mathbf{Y}, \theta)$ oznacza model Bayesowski, czyli łączny rozkład wektora obserwacji i wektora parametrów; $p(\mathbf{Y})$ to brzegowy rozkład prawdopodobieństwa obserwacji – spełnia rolę współczynnika normalizacji w mianowniku równania (3.10); $p(\mathbf{Y}|\theta)$ to model próbkowy, czyli warunkowy przy ustalonym wektorze parametrów θ_i rozkład obserwacji; $p(\theta)$ to rozkład *a priori*, czyli rozkład parametrów przed wglądem w dane.

W powyższym przypadku schemat wnioskowania Bayesowskiego w praktyce sprowadza się do następujących kroków:

- 1) Zapisanie modelu próbkowego w postaci wyrażeń algebraicznych.
- 2) Wybór rozkładu *a priori*.
- 3) Zapisanie modelu Bayesowskiego w postaci wyrażeń algebraicznych.
- 4) Wyznaczenie rozkładu *a posteriori*.

3.3.2. Modelowanie Bayesowskie modelu wielokrotnej liniowej regresji

W SLMB parametry funkcjonalne i wiarygodność prognozy wyprowadzone są na podstawie danych zaobserwowanych [Walsh 2004]. Przestrzeń obserwacji modelu tworzą obserwacje historyczne i obejmują zmienne prognozowane \mathbf{Y} i prognozujące \mathbf{X} . W problemie regresji liniowej zakłada się liniową zależność pomiędzy zmienną prognozowaną \mathbf{Y} i zmienną prognozującą \mathbf{X} . Zatem równanie liniowej regresji Bayesowskiej wygląda w sposób następujący:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad (3.11)$$

gdzie: n – liczba obserwacji; l – liczba zmiennych prognozowanych, dla wielokrotnej liniowej regresji $l = 1$; d – liczba zmiennych prognozujących; \mathbf{Y} – n wymiarowy wektor z zaobserwowanymi wartościami zmiennej prognozowanej; \mathbf{X} – $n \times d$ wymiarowa macierz obserwacji zmiennych prognozujących; β – d wymiarowy wektor współczynników kierunkowych; ϵ – n wymiarowy wektor błędów modelu i niepewności parametrów o wartości oczekiwanej $E(\epsilon) = \mathbf{0}$ i kowariancji $\text{Cov}(\epsilon\epsilon^T) = \sigma^2\mathbf{E}$, tj. o wielowymiarowym rozkładzie normalnym, $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{E})$, gdzie \mathbf{E} oznacza $n \times n$ wymiarową macierz jednostkową.

Jeżeli stosujemy postać funkcji regresji z wartością stałą, jedna z kolumn macierzy \mathbf{X} musi być wypełniona wartością 1. Parametry regresji nie są podane explicite i ogólnie będą oznaczone symbolem θ . Szacowanie ich wartości odbywa się za pomocą analizy Bayesa. θ zawiera wektor współczynników kierunkowych β i wariancję błędu dopasowania modelu σ^2 , która jest traktowana jako precyzja składnika losowego ϵ , czyli $\theta = (\beta, \sigma^2)$.

MODEL PRÓBKOWY

Model próbkowy opisuje zależność między obserwacjami i parametrami modelu i zapisany jest jako [Carpenter 2003]:

$$p(\mathbf{Y}|\beta, \sigma^2, \mathbf{X}) \sim N(\mathbf{X}\beta, \sigma^2\mathbf{E}) \quad (3.12)$$

Oznacza to, że rozkład zmiennych prognozowanych \mathbf{Y} , uwarunkowany parametrami $\boldsymbol{\beta}$ i σ^2 oraz zmiennymi prognozującymi \mathbf{X} , podlega wielowymiarowemu rozkładowi normalnemu ze średnią równą $\mathbf{X}\boldsymbol{\beta}$ i macierzą kowariancji [szczegóły w Załączniku: Dodatek B] obliczaną jako $\text{Cov}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = \sigma^2\mathbf{E}$.

WYBÓR ROZKŁADU *a priori*

W problemie estymacji parametrów liniowej regresji zbiór historycznych obserwacji i pomiarów uwzględniony jest w postaci rozkładu prawdopodobieństwa *a priori*. W SLMB zastosowano rozkłady *a priori*, które mają na celu uchwycenie stanu wiedzy o testowanym modelu wezbrania deszczowo-roztopowego, tzw. rozkłady nieinformacyjne. Metodą konstrukcji takiego apriorycznego rozkładu jest tzw. reguła Jeffreys'a [Osiewalski 1991]:

$$p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{-\frac{1}{2}} \quad (3.13)$$

gdzie \mathbf{I} jest informacją Fishera [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 30] liczoną na podstawie wartości oczekiwanej $E[.]$ w sensie rozkładu $p(\boldsymbol{\theta})$ według wzoru:

$$\mathbf{I}(\boldsymbol{\theta}) = -E\left(\frac{\partial^2 \log(p(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^2}\right) \quad (3.14)$$

W przypadku nieinformacyjnego rozkładu *a priori* dla nieznanego wektora parametrów $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ rozkład Jeffreys'a dla wiarygodności Gaussowskiej przyjmuje postać $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}$. Wyrażenie oznacza, że łączny rozkład $\boldsymbol{\beta}$ i σ^2 jest płaską powierzchnią o stałym poziomie proporcjonalnym do $\frac{1}{\sigma^2}$.

WYZNACZANIE ROZKŁADU *a posteriori*

Podstawowym zadaniem analizy Bayesowskiej w zagadnieniach wielokrotnej regresji liniowej jest poszukiwanie rozkładów *a posteriori* i optymalizacja wektora parametrów $\boldsymbol{\beta}$ i wariancji błędów σ^2 . Obliczenia brzegowych rozkładów prawdopodobieństwa realizuje algorytm MCMC [szczegóły w Załączniku: Dodatek A], uwzględniający tworzenie łańcucha Markowa przy użyciu generatora Gibbsa [Cappé i in. 2005].

Rozkład *a posteriori* [Carpenter 2003; Walsh 2004] wyrażony jest jako:

$$p(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y}) = p(\boldsymbol{\beta} | \sigma^2, \mathbf{Y}) p(\sigma^2 | \mathbf{Y}) \quad (3.15)$$

gdzie brzegowy rozkład *a posteriori* współczynników kierunkowych $\boldsymbol{\beta}$ jest uwarunkowany wariancją błędów σ^2 i wektorem zmiennych prognozowanych \mathbf{Y} i podlega rozkładowi normalnemu ze średnią $\hat{\boldsymbol{\beta}}$ i kowariancją $\mathbf{V}_{\boldsymbol{\beta}}\sigma^2$. Obliczany jest zgodnie z wyrażeniem [Carpenter 2003; Walsh 2004]:

$$p(\boldsymbol{\beta} | \sigma^2, \mathbf{Y}) \sim N(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}} \sigma^2) \quad (3.16)$$

gdzie: $\hat{\boldsymbol{\beta}}$ jest estymatorem metody najmniejszych kwadratów (MNK) wektora $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$; $\mathbf{V}_{\boldsymbol{\beta}}$ jest $d \times d$ wymiarową macierzą obliczaną jako $\mathbf{V}_{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1}$; d jest liczbą zmiennych prognozujących, a \mathbf{X}^T oznacza macierz transponowaną.

ESTYMACJA PARAMETRÓW ROZKŁADU *a posteriori*

Do oszacowania brzegowego rozkładu *a posteriori* σ^2 stosuje się odwrotny rozkład prawdopodobieństwa $(\chi^2)^{-1}$ (3.17) [Gelman i in. 1995] lub odwrotny rozkład prawdopodobieństwa gamma γ^{-1} (3.19) [Gentle 2003]. Wybór rozkładu determinują własności rozkładu:

- 1) wszystkie wartości ≥ 0 ;
- 2) stosowany jest rozkład normalny, gdy liczebność próby losowej rośnie:

$$p(\sigma^2 | \mathbf{Y}) \sim [\chi^2(n - d, \hat{\sigma}^2)]^{-1} \quad (3.17)$$

gdzie: d – liczba zmiennych prognozujących; n – liczba obserwacji w macierzy \mathbf{X} ; $\hat{\sigma}^2$ – estymator wariancji σ^2 obliczany ze wzoru [Gentle 2003]:

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-d} \quad (3.18)$$

Losowanie odbywa się z populacji o rozkładzie $N(0, \sigma^2)$, gdzie parametr σ^2 jest estymowany zgodnie z (3.18). Z danych pochodzących z próby losowej obliczana jest statystyka rozkładu χ^2 . Z równania (3.17) widać, że rozkład prawdopodobieństwa σ^2 dla danego wektora \mathbf{Y} podlega odwrotnemu rozkładowi prawdopodobieństwa $(\chi^2)^{-1}$ [Gelman i in. 1995]. Ten rozkład jest w pełni zdefiniowany dwoma parametrami: $n-d$ stopniami swobody i współczynnikiem skali $\hat{\sigma}^2$ obliczanym za pomocą wzoru (3.18), w którym $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ jest wektorem reszt lub odchyleniem obserwacji od predykcji. Natomiast generator zmiennych losowych z odwrotnego rozkładu gamma $(\gamma)^{-1}$ może być wyrażony jako [Gentle 2003]:

$$p(\sigma^2 | \mathbf{Y}, \hat{\boldsymbol{\beta}}) \sim \gamma \left(a + \frac{n}{2}, b + \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{2} \right)^{-1} \quad (3.19)$$

gdzie a i b – są współczynnikami rozkładu $\gamma(a, b)$.

Odwrotny rozkład prawdopodobieństwa $(\chi^2)^{-1}$ jest szczególnym przypadkiem rozkładu $\gamma(a, b)$ [Gentle 2003], tj. $\chi^2(n) = \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. W opisywanym rozwiązaniu wybór rozkładu prawdopodobieństwa jest elementem kalibracji SLMB. Wybiera się ten rozkład, dla którego uzyskiwana jest wyższa ocena jakości modelu uruchomionego w trybie kalibracji.

Brzegowy rozkład *a posteriori* wektora współczynników kierunkowych $\boldsymbol{\beta}$, uwarunkowany wektorem \mathbf{Y} , podlega wielowymiarowemu rozkładowi *t*-Studenta [Kotz, Nadarajah 2004] zgodnie z wyrażeniem:

$$p(\boldsymbol{\beta} | \mathbf{Y}) \sim T_{n-d}(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}} \hat{\sigma}^2) \quad (3.20)$$

gdzie $n - d$ oznacza stopnie swobody wielowymiarowego rozkładu *t*-Studenta.

Wielowymiarowy rozkład *t*-Studenta [szczegóły w Załączniku: Dodatek B] ma trzy parametry: stopnie swobody równe $n - d$, średnią $\hat{\boldsymbol{\beta}}$ i kowariancję $\mathbf{V}_{\boldsymbol{\beta}} \sigma^2$. Rozkład ten jest obliczany przez całkowanie rozkładu *a posteriori* współczynników kierunkowych $\boldsymbol{\beta}$ i uwarunkowany kowariancją $\mathbf{V}_{\boldsymbol{\beta}} \sigma^2$ zgodnie z równaniem (3.16) za pomocą wzoru (3.17) lub (3.19). Przy rozpatrywaniu rozkładu brzegowego dla pojedynczego parametru wykorzystano własności wielowymiarowego rozkładu *t*-Studenta, tzn.: rozkład brzegowy dla pojedynczej zmiennej z wielowymiarowego rozkładu T_{n-d} to również (jednowymiarowy) rozkład *t*-Studenta.

Podstawowe charakterystyki brzegowych rozkładów *a posteriori* poszczególnych elementów wektora $\boldsymbol{\beta}$ to: wartość oczekiwana $E(\beta_i | \mathbf{Y}, \mathbf{X})$, wariancja $\text{Var}(\beta_i | \mathbf{Y}, \mathbf{X})$ i odchylenie standardowe $D(\beta_i | \mathbf{Y}) = \sqrt{\text{Var}(\beta_i | \mathbf{Y})}$ dla $i = 1, \dots, d$, gdzie d jest liczbą zmiennych prognozujących.

Dodatkowo analiza Bayesowska umożliwi obliczenie przewidywanego rozkładu zmiennych niezaobserwowanych. Zatem celem SLMB jest wyznaczenie zmiennej prognozowanej \mathbf{Y}_0 . Wartość oczekiwaną prognozowanego rozkładu \mathbf{Y}_0 dla macierzy zmiennych prognozujących \mathbf{X}_0 zapisujemy jako $E(\mathbf{Y}_0 | \mathbf{Y}) = \mathbf{X}_0 \hat{\boldsymbol{\beta}}$, a wariancję prognozy brzegowego rozkładu prawdopodobieństwa *a priori* [Carpenter 2003] jako:

$$\text{Var}(\mathbf{Y}_0 | \sigma^2, \mathbf{Y}) = \mathbf{E} \sigma^2 + \mathbf{X}_0 \mathbf{V}_{\boldsymbol{\beta}} \mathbf{X}_0^T \sigma^2 \quad (3.21)$$

gdzie $\mathbf{E} \sigma^2$ jest komponentem próbkowania wariancji nowych obserwacji zawartych w macierzy zmiennych prognozujących \mathbf{X}_0 , a $\mathbf{X}_0 \mathbf{V}_{\boldsymbol{\beta}} \mathbf{X}_0^T \sigma^2$ wyrażeniem służącym do obliczania niepewności wektora współczynników kierunkowych $\boldsymbol{\beta}$.

Brzegowy rozkład *a posteriori* wektora \mathbf{Y}_0 , uwarunkowany wektorem danych zaobserwowanych \mathbf{Y} , podlega wielowymiarowemu rozkładowi *t*-Studenta zgodnie z wyrażeniem [Carpenter 2003; Walsh 2004]:

$$p(\mathbf{Y}_0|\mathbf{Y}) \sim T_{n-d}(\mathbf{X}_0\hat{\boldsymbol{\beta}}, (\mathbf{E} + \mathbf{X}_0^T\mathbf{V}_\beta)\hat{\sigma}^2) \quad (3.22)$$

Rozkłady predykcyjne, określane dla danego zbioru danych, w dużym stopniu są uzależnione od macierzy kowariancji wyznaczonej ze wzoru $\mathbf{V}_\beta\sigma^2$ i parametrów. Istnieje tylko jedno ograniczenie macierzy kowariancji. Macierz kowariancji $(\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$ i macierz zmiennych prognozujących \mathbf{X} muszą być dodatnio określone dla dowolnej wartości n , gdzie n jest liczbą obserwacji w macierzy \mathbf{X} , która jest zbiorem zmiennych prognozujących, będącymi danymi wejściowymi do modelu regresji w celu zapewnienia, że rozkład $p(\mathbf{Y}|\mathbf{V}_\beta\sigma^2)$ jest normalizowalny.

W opisanym powyżej SLMB zastosowano stacjonarną kowariancję. Mimo, że wiele zbiorów danych może być skutecznie modelowanych za pomocą takiej macierzy, w niektórych przypadkach kowariancja musi wykazywać pewną postać niestacjonarności. Ponieważ trudno jest zbudować uniwersalną strukturę modelu, w ramach których generowane są obie kowariancje, to powyższy rozdział ograniczony jest do opisu użytecznej postaci stacjonarnej.

ESTYMACJA NIEZNANYCH PARAMETRÓW TECHNIKĄ ODWROTNEGO MODELOWANIA

Bayesowska metoda estymacji parametrów stosowana jest w modelach, w których zmienna prognozowana jest liniową funkcją zmiennych prognozujących. W przypadku analizy zjawisk nieliniowych do szacowania parametrów modelu na podstawie wielowymiarowych zbiorów danych, teoria Bayesa stosuje technikę odwrotnego modelowania (inverse modelling). Ogólną postać nieliniowego modelu opisuje równanie $\mathbf{Y} = f(\boldsymbol{\beta}, \mathbf{X}) + \epsilon$, gdzie f jest nieliniową funkcją wektora parametrów $\boldsymbol{\beta}$ i macierzy obserwacji zmiennej prognozującej \mathbf{X} , a ϵ wektorem reszt, czyli odchyleniem \mathbf{Y} od wartości prognozowanych \mathbf{Y}_0 . Zakłada się, że ϵ podlega rozkładowi normalnemu $\epsilon \sim N(\mathbf{0}, \sigma^2)$ [Gibbs 1997]. W takim przypadku ϵ można wyrazić w funkcji nieznanymi parametrów $\epsilon = \mathbf{Y} - f(\boldsymbol{\beta}, \mathbf{X})$. Nieznane wartości parametrów $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2\mathbf{E})$ można oszacować minimalizując ujemny logarytm funkcji wiarygodności zgodnie z zależnością:

$$-\log(f) = 0,5 [(\epsilon^T \epsilon / \sigma^2) + n \log(2\pi) + n \log(\sigma^2)] \quad (3.23)$$

gdzie n jest liczbą obserwacji.

Wiarygodność wszystkich reszt ϵ jest iloczynem wiarygodności poszczególnych reszt, zatem logarytm funkcji wiarygodności $\log(f)$ jest sumą ich logarytmów. W związku z tym, że jest to ujemny logarytm, minimalizacja tej funkcji w rzeczywistości będzie maksymalizacją wiarygodności.

W Bayesowskiej technice odwrotnego modelowania, do obliczenia rozkładu *a posteriori* parametrów $\boldsymbol{\theta}$, rozkład *a priori* wektora parametrów $\boldsymbol{\theta}$ jest łączony z danymi pozyskanymi z dodatkowych obserwacji lub pomiarów [Young 1977; Thacker, Long 1998]. W ten sposób parametry nieliniowego modelu szacowane są za pomocą minimalizacji sumy ujemnych logarytmów funkcji wiarygodności i rozkładu *a priori*:

$$-\log(f_p) = -\log(f_0) - \log(f_1) \quad (3.24)$$

gdzie f_p jest rozkładem *a posteriori* otrzymanym w wyniku połączenia rozkładu *a priori* f_0 i Bayesowskiej funkcji wiarygodności f_1 obliczonej na podstawie dodatkowych obserwacji lub pomiarów.

Dla średniej wartości parametrów $\boldsymbol{\beta}_0$ i macierzy kowariancji \mathbf{V}_0 rozkładu *a priori* i funkcji nieznanymi parametrów $\epsilon = \mathbf{Y} - f(\boldsymbol{\beta}, \mathbf{X})$ ujemny logarytm rozkładu *a priori* wyrażony jest jako:

$$-\log(f_0) = 0,5[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{V}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + d \log(2\pi) + \log(|\mathbf{V}_0|)] \quad (3.25)$$

gdzie β jest wektorem estymowanych parametrów modelu, a d jest liczbą zmiennych prognozujących. Natomiast ujemny logarytm funkcji wiarygodności występuje w postaci:

$$-\log(f_i) = 0,5 \left[\left(\frac{\epsilon_i^T \epsilon_i}{\sigma^2} \right) + n \log(2\pi - \sigma^2) \right] \quad (3.26)$$

Rozkład prawdopodobieństwa *a posteriori* parametrów modelu nie będzie przypominał takich rozkładów, jak wielowymiarowy rozkład normalny lub wielowymiarowy rozkład *t*-Studenta. Przybliżenie macierzy kowariancji parametrów V_0 można obliczyć różnymi metodami, takimi jak linearyzacja w pobliżu optymalnych oszacowań parametrów lub za pomocą metody bootstrap [Efron, Tibshirani 1994].

3.3.3. Modelowanie Bayesowskie wielowymiarowego modelu wielokrotnej liniowej regresji

Rozszerzając wielokrotną regresję liniową do przypadku wielowymiarowej wielokrotnej regresji należy uwzględnić więcej niż jedną zmienną prognozowaną, tj. $l > 1$. W takiej sytuacji przestrzeń obserwacji modelu tworzą obserwacje historyczne obejmujące macierze zmiennych prognozowanych Y i prognozujących X . Macierz zmiennych prognozowanych tworzą wektory $Y_{(p)}$ zawierające wartości p -tej zmiennej prognozowanej, gdzie $p = 1, 2, \dots, l$ wyznaczone dla n obserwacji. Podobnie każdy wektor błędów $\epsilon_{(p)}$ zawiera błędy losowe otrzymane dla n obserwacji dla p -tej zmiennej prognozowanej. Wektor $\beta_{(p)}$ złożony jest z nieznanymi współczynnikami kierunkowych modelu regresji otrzymanych dla p -tej zmiennej prognozowanej. Jedynie macierz zmiennych prognozujących X ma takie same wymiary jak w przypadku wielokrotnej regresji liniowej. Uogólniając model wielokrotnej regresji liniowej oddzielnie dla każdej zmiennej prognozowanej uzyskujemy następujące równanie regresji:

$$Y_{(p)} = X\beta_{(p)} + \epsilon_{(p)} \text{ dla } p = 1, 2, \dots, l \quad (3.27)$$

gdzie błąd wektora $\epsilon = (\epsilon_1, \epsilon_1, \dots, \epsilon_l)$ ma $E(\epsilon) = 0$ i $\text{Var}(\epsilon) = \Sigma$.

Poprzez włączenie do struktury macierzowej każdej pojedynczej odpowiedzi modelu (zmiennej prognozowanej) konstruuje się wielowymiarowy model wielokrotnej regresji liniowej zapisany w postaci:

$$Y = X\beta + \epsilon \quad (3.28)$$

gdzie: Y – $n \times l$ wymiarowa macierz z obserwacjami zmiennych prognozowanych; X – $n \times d$ wymiarowa macierz z obserwacjami zmiennych prognozujących; β – $d \times l$ wymiarowa macierz współczynników kierunkowych modelu; ϵ – $n \times l$ wymiarowa macierz błędów modelu.

Wielowymiarowy model złożony z oddzielnych l wielokrotnych regresji liniowych ma następujące własności losowe: $E(\epsilon_{(p)}) = 0$, $E(Y) = X\beta$ i $\text{Cov}(\epsilon_{(p)}) = \sigma_{pp}E$ dla p -tej zmiennej prognozowanej, gdzie $p = 1, 2, \dots, l$. Natomiast wielowymiarowy model wielokrotnej regresji liniowej ma następujące własności losowe: $E(\epsilon) = 0$ i $\text{Cov}(\epsilon_{(p)}, \epsilon_{(q)}) = \sigma_{pq}E$, gdzie p i $q = 1, 2, \dots, l$, a E jest $n \times n$ wymiarową macierzą jednostkową. Oznacza to, że błędy dla różnych zmiennych prognozowanych dla tej samej próby losowej mogą być skorelowane, czyli $\text{Cov}(\epsilon_{1p}, \epsilon_{1q}) \neq 0$, $\text{Cov}(\epsilon_{2p}, \epsilon_{2q}) \neq 0, \dots, \text{Cov}(\epsilon_{lp}, \epsilon_{lq}) \neq 0$ dla $p, q = 1, 2, \dots, l$.

W wielowymiarowym modelu wielokrotnej regresji liniowej nieznanymi parametrami są macierze Σ i β , gdzie Σ jest kowariancją w postaci $l \times l$ wymiarowej macierzy o strukturze $\Sigma = \{\sigma_{pq}\}$, dla $p, q = 1, \dots, l$. Różnice między wielokrotną regresją a wielowymiarowym modelem w zagadnieniach Bayesowskich to: wykorzystanie do obliczenia rozkładu próbkowego zamiast wielowymiarowego rozkładu prawdopodobieństwa, rozkładu macierzowego oraz wykorzystanie metody symulacji Monte Carlo do estymacji parametrów rozkładu *a posteriori*. Schemat wnioskowania dla obu modeli, tj. wielowymiarowego i wielokrotnego, jest taki sam. Analiza polega na utworzeniu modelu próbkowego, następnie wybraniu rozkładu *a priori* oraz wyznaczeniu rozkładu *a posteriori*.

MODEL PRÓBKOWY

$$p(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X}) \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}\otimes\mathbf{E}) \quad (3.29)$$

Oznacza to, że rozkład zmiennych prognozowanych \mathbf{Y} , uwarunkowany parametrami $\boldsymbol{\beta}$ i $\boldsymbol{\Sigma}$ oraz zmiennymi prognozującymi \mathbf{X} , podlega macierzowemu rozkładowi normalnemu ze średnią równą $\mathbf{X}\boldsymbol{\beta}$ i macierzą wariancji kowariancji obliczaną jako iloczyn *Kroneckera* [szczegóły w Załączniku: Dodatek B]. Do zapisu macierzowych rozkładów prawdopodobieństwa przyjęto notację zgodną z Krzyżko [2009]. Iloczyn Kroneckera $\boldsymbol{\Sigma}\otimes\mathbf{E}$ jest macierzą blokową o wymiarach $ln \times ln$, a $\mathbf{X}\boldsymbol{\beta}$ i \mathbf{Y} są macierzami o wymiarach $n \times l$.

WYBÓR ROZKŁADU *a priori*

W przypadku nieinformacyjnego rozkładu *a priori* dla wielowymiarowego modelu wielokrotnej regresji i nieznanego wektora $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$ rozkład Jeffreysa dla wiarygodności Gaussowskiej przyjmuje postać [Bain, Engelhardt 2000]:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto (\mathbf{I}(\boldsymbol{\theta}))^{\frac{1}{2}} \quad (3.30)$$

Zakłada się, że $\boldsymbol{\beta}$ i $\boldsymbol{\Sigma}$ są niezależne, zatem nieinformacyjny rozkład *a priori* wyrażony jest jako $f(\boldsymbol{\theta}) = f(\boldsymbol{\beta})f(\boldsymbol{\Sigma})$. Obliczenia rozkładu *a priori* $f(\boldsymbol{\theta})$ wykonuje się według następującego schematu:

1) nieinformacyjny rozkład *a priori* dla $\boldsymbol{\beta}$ według reguły Jeffreysa [Bain, Engelhardt 2000]:

$$\mathbf{I}(\boldsymbol{\beta}) = -E\left(\frac{\partial^2}{\partial \boldsymbol{\beta}^2} \log(p(\mathbf{Y}|\boldsymbol{\theta}))\right) = -E(-\boldsymbol{\Sigma}^{-1}) = \boldsymbol{\Sigma}^{-1} \quad (3.31)$$

$$f(\boldsymbol{\beta}) = \sqrt{|\mathbf{I}(\boldsymbol{\beta})|} \sqrt{\boldsymbol{\Sigma}^{-1}} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \quad (3.32)$$

2) dla l -wymiarowego modelu wielokrotnej regresji [Saputro i in. 2018]:

$$f(\boldsymbol{\beta}) = f(\boldsymbol{\beta}_1) \times f(\boldsymbol{\beta}_2) \times \dots \times f(\boldsymbol{\beta}_l) = \boldsymbol{\Sigma}^{-\frac{1}{2}} \times \boldsymbol{\Sigma}^{-\frac{1}{2}} \times \dots \times \boldsymbol{\Sigma}^{-\frac{1}{2}} = \boldsymbol{\Sigma}^{-\frac{l}{2}} \quad (3.33)$$

3) nieinformacyjny rozkład *a priori* dla $\boldsymbol{\Sigma}$ według reguły Jeffreysa [Saputro i in. 2018]:

$$\mathbf{I}(\boldsymbol{\Sigma}) = -E\left(\frac{\partial^2}{\partial \boldsymbol{\Sigma}^2} \log(p(\mathbf{Y}|\boldsymbol{\theta}))\right) = -E(-(\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-3} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})) = \boldsymbol{\Sigma}^{-2} \quad (3.34)$$

dla $k = 1, 2, \dots, n$

$$f(\boldsymbol{\Sigma}) = \sqrt{|\mathbf{I}(\boldsymbol{\Sigma})|} \sqrt{\boldsymbol{\Sigma}^{-2}} = \boldsymbol{\Sigma}^{-1} \quad (3.35)$$

4) nieinformacyjny rozkład *a priori* wg reguły Jeffreysa dla $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$ jest wyrażony jako [Saputro i in. 2018]:

$$f(\boldsymbol{\theta}) = f(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = f(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_l) f(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-\frac{l}{2}} \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-\frac{l+2}{2}} \quad (3.36)$$

WYZNACZANIE ROZKŁADU *a posteriori*

Rozkład *a posteriori* [Carpenter 2003; Walsh 2004] wyrażony jest jako:

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{Y}) = p(\boldsymbol{\beta}|\boldsymbol{\Sigma}, \mathbf{Y}) p(\boldsymbol{\Sigma}|\mathbf{Y}) \quad (3.37)$$

gdzie $p(\boldsymbol{\beta}|\boldsymbol{\Sigma}, \mathbf{Y})$ jest brzegowym rozkładem *a posteriori* dla macierzy współczynników kierunkowych $\boldsymbol{\beta}$, podlegającym wielokrotnemu rozkładowi normalnemu $p(\boldsymbol{\beta}|\boldsymbol{\Sigma}, \mathbf{Y}) \sim N(\bar{\mathbf{Y}}, \frac{\boldsymbol{\Sigma}}{n})$ z $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{k=1}^n \mathbf{Y}_k$, a $p(\boldsymbol{\Sigma}|\mathbf{Y})$

jest brzegowym rozkładem *a posteriori* macierzy wariancji błędów Σ , podlegającym odwrotnemu rozkładowi Wisharta $p(\Sigma|\mathbf{Y}) \sim IW(\mathbf{S}^{-1}, n)$, gdzie \mathbf{S} jest $l \times l$ wymiarową macierzą skali obliczaną jako [Saputro i in. 2018]:

$$\mathbf{S} = \sum_{k=1}^n (\mathbf{Y}_k - \bar{\mathbf{Y}})^T (\mathbf{Y}_k - \bar{\mathbf{Y}}) \quad (3.38)$$

a n jest stopniami swobody.

ESTYMACJA PARAMETRÓW ROZKŁADU *a posteriori*

Do estymacji parametrów wielowymiarowego modelu wielokrotnej regresji wykorzystano metodę Bayesowską polegającą na wyznaczeniu oczekiwanych wartości brzegowych rozkładów *a posteriori* $p(\boldsymbol{\beta} | \Sigma, \mathbf{Y})$ i $p(\Sigma | \mathbf{Y})$. Do oszacowania parametrów użyto metody symulacji Monte Carlo (MCMC) z algorytmem próbkowania Gibbsa [patrz ogólny algorytm próbkowania Gibbsa Dodatek B]. Algorytm Gibbsa dla wielowymiarowego modelu wielokrotnej regresji jest następujący [Saputro i in., 2018]:

1. Skonstruowanie wielowymiarowego modelu próbkowego dla wielowymiarowej wielokrotnej liniowej regresji do estymacji macierzy współczynników kierunkowych $\boldsymbol{\beta}$ i macierzy wariancji błędów Σ .
2. Zainicjowanie tworzenia łańcucha Markowa ze zbioru wartości początkowych estymowanych parametrów $\boldsymbol{\beta}^{(0)}$ i $\Sigma^{(0)}$ na podstawie informacji z kroku 1.
3. Obliczenie pierwszej realizacji rozkładu *a posteriori* na podstawie wielokrotnego rozkładu normalnego, tj. $\mathbf{Y}_k^{(1)} \sim N(\mathbf{X}_k \boldsymbol{\beta}^{(0)}, \Sigma^{(0)})$.
4. Wygenerowanie $\boldsymbol{\beta}^{(1)} | \mathbf{Y}_k^{(1)} \sim N(\bar{\mathbf{Y}}, \frac{\Sigma}{n} | \boldsymbol{\beta}^{(0)})$ dla $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{k=1}^n \mathbf{Y}_k$.
5. Wygenerowanie $\Sigma^{(1)} | \mathbf{Y}_k^{(1)} \sim IW(\mathbf{S}^{-1}, n | \Sigma^{(0)})$.
6. Powtórzenie kroków 3-5 g -krotnie dla $g = 1, 2, \dots, M$ do otrzymania:
 $\boldsymbol{\beta}^{(M)} | \mathbf{Y}_k^{(M)} \sim N(\bar{\mathbf{Y}}, \frac{\Sigma}{n} | \boldsymbol{\beta}^{(M-1)})$,
 $\Sigma^{(M)} | \mathbf{Y}_k^{(M)} \sim IW(\mathbf{S}^{-1}, n | \Sigma^{(M-1)})$.

W obliczeniach użyto metodę MCMC z algorytmem próbkowania Gibbsa zastosowanym do zmiennych losowych $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \dots, \boldsymbol{\beta}^{(M)})$ i $(\Sigma^{(1)}, \Sigma^{(2)}, \dots, \Sigma^{(M)})$. Estymowany parametr dla $\boldsymbol{\beta}$ jest równy $\hat{\boldsymbol{\beta}} = \frac{1}{M} \sum_{g=1}^M \boldsymbol{\beta}^{(g)}$, zaś dla Σ równy $\hat{\Sigma} = \frac{1}{M} \sum_{g=1}^M \Sigma^{(g)}$.

ESTYMACJA NIEZNANYCH PARAMETRÓW TECHNIKĄ ODWROTNEGO MODELOWANIA

W technice odwrotnego modelowania nieznane wartości parametrów $\boldsymbol{\theta} = (\boldsymbol{\beta}, \Sigma)$ można oszacować minimalizując ujemny logarytm funkcji wiarygodności zgodnie ze wzorem [Lin i in. 2018]:

$$-\log(L(\boldsymbol{\beta}, \Sigma | \mathbf{Y})) = 0,5 n l \log(2\pi) - 0,5 n \log(|\Sigma|) - 0,5 \mathbf{S} \quad (3.39)$$

gdzie \mathbf{S} jest $l \times l$ wymiarową macierzą skali obliczaną wg wzoru (3.38), a:

$$L(\boldsymbol{\beta}, \Sigma | \mathbf{Y}) = (2\pi)^{\frac{-nl}{2}} (|\Sigma|)^{\frac{-n}{2}} \exp\left\{-0,5 \text{tr}\left((\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\right)\right\} \quad (3.40)$$

jest funkcją gęstości funkcji wiarygodności wielowymiarowego modelu wielokrotnej liniowej regresji.

W przypadku logarytmicznej funkcji wiarygodności wyznacza się najlepsze estymowane parametry poprzez ustalenie pochodnej cząstkowej funkcji wiarygodności względem parametrów $\boldsymbol{\beta}$ i Σ . Obliczane są pochodne cząstkowe, dla których logarytmiczna funkcja wiarygodności przyjmuje wartość ekstremalną [w omawianym przypadku minimalną, zgodnie ze wzorem (3.39)], dla których pochodna funkcji logarytmicznej funkcji wiarygodności przyjmuje wartość zero, tj. $\frac{\partial \log(L)}{\partial \boldsymbol{\beta}} = 0$ i $\frac{\partial \log(L)}{\partial \Sigma} = 0$.

W związku z tym, że łączny rozkład *a posteriori* wielowymiarowego modelu wielokrotnej liniowej re-

gresji nie ma postaci analitycznej, to do estymacji odpowiednich parametrów stosuje się metody symulacji Monte Carlo, np. metodę Metropolisa-Hasingsa [Chib, Greenberg 1995] z połączeniem generatora Gibbsa [Sinay, Hsu 2014].

PROGNOZA PRZEDZIAŁOWA W WIELOWYMIAROWYM MODELU WIELOKROTNEJ LINIOWEJ REGRESJI
 Na podstawie znanej macierzy zmiennych prognozujących \mathbf{X} i estymowanych parametrów modelu $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ można obliczyć macierz zmiennych prognozowanych $\hat{\mathbf{Y}}$. Prognozy oblicza się za pomocą równania macierzowego $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Po obliczeniu macierzy $\hat{\mathbf{Y}}$ można wyznaczyć macierz prognozowanych błędów $\boldsymbol{\epsilon}$, tj. $\boldsymbol{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$.

Zatem l wymiarowy wektor zmiennych prognozowanych \mathbf{Y}_0 zostanie obliczony na podstawie zależności $\mathbf{X}_0\hat{\boldsymbol{\beta}}$, gdzie \mathbf{X}_0 jest d wymiarowym wektorem zmiennych prognozujących. W przypadku wielowymiarowego modelu wielokrotnej regresji $\hat{\mathbf{Y}}_0$ podlega wielowymiarowemu rozkładowi normalnemu, tj.:

$$\hat{\mathbf{Y}}_0 \sim N(\mathbf{X}_0\hat{\boldsymbol{\beta}}, (\mathbf{X}_0(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_0^{-1})\hat{\boldsymbol{\Sigma}}) \quad (3.41)$$

Na podstawie zależności (3.41) ustalony został $100(1 - \alpha)\%$ przedział jednoczesnej prognozy, obliczany osobno dla każdej zmiennej prognozowanej Y_{0p} dla $p = 1, 2, \dots = l$. Prognoza przedziałowa jest szacunkowym zakresem, w którym znajdują się prognozowane zmienne losowe, otrzymane na podstawie wektora zmiennych prognozujących wg wzoru:

$$\hat{\mathbf{Y}}_{0p} = \mathbf{X}_0\hat{\boldsymbol{\beta}}_{(p)} \sim N(\mathbf{X}_0\hat{\boldsymbol{\beta}}_{(p)}, (\mathbf{X}_0(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_0^{-1})\hat{\sigma}_{pp}^2) \quad (3.42)$$

Biorąc pod uwagę rozkład średniej wartości zmiennej prognozowanej dla p -tej zmiennej prognozowanej (3.42), określono rozkład różnicy odpowiedzi modelu i obserwacji. Wartość oczekiwana różnicy $E(\mathbf{Y}_{0p} - \mathbf{X}_0\hat{\boldsymbol{\beta}}_{(p)})$ jest równa 0. W przypadku wielowymiarowym powyższa różnica podlega wielokrotnemu rozkładowi normalnemu, tj.:

$$\mathbf{Y}_0 - \mathbf{X}_0\hat{\boldsymbol{\beta}} \sim N(\mathbf{0}, (1 + \mathbf{X}_0(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_0^{-1})\hat{\boldsymbol{\Sigma}}) \quad (3.43)$$

Wykorzystując rozkład różnicy odpowiedzi modelu i obserwacji zmiennych prognozowanych (3.43), utworzono przedział jednoczesnej prognozy l wymiarowego wektora zmiennych prognozowanych $\hat{\mathbf{Y}}_0$, obliczany oddzielnie dla p -tej zmiennej prognozowanej:

$$\mathbf{X}_0\hat{\boldsymbol{\beta}}_{(p)} \pm \sqrt{\left(\frac{l(n-d-1)}{n-d-l}\right) F_{l, n-d-l}(\alpha)} \sqrt{\left(1 + \mathbf{X}_0(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_0^T\right) \left(\frac{n}{n-d-1}\hat{\sigma}_{pp}\right)} \quad (3.44)$$

gdzie: $F_{l, n-d-l}(\alpha)$ jest rozkładem prawdopodobieństwa F -Snedecora [Phillips 1982] o l i $n - d - l$ stopniach swobody dla przyjętego w obliczeniach α poziomu istotności; $\hat{\boldsymbol{\beta}}_{(p)}$ jest d wymiarowym wektorem współczynników kierunkowych estymowanych dla p -tej zmiennej prognozowanej; $\hat{\sigma}_{pp}$ jest p -tym elementem diagonalnej macierzy $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}}$. Równanie (3.44) można zastosować dla jednowymiarowego modelu wielokrotnej regresji, tj. dla $l = 1$.

Przedziały jednoczesnej l wymiarowej prognozy umożliwiają określenie, gdzie (w jakim zakresie wartości) należy spodziewać się wystąpienia wektora zmiennej prognozowanej dla określonego wektora zmiennych prognozujących. Na przykład dla prognozy 80% można oczekiwać, że wektor zmiennej prognozowanej będzie mieścił się w przedziale wyznaczonym na podstawie 80% prób losowych.

3.3.4. Zastosowanie modeli SLMB i warunki stosowalności

Do modelowania wezbrania deszczowo-roztopowego zastosowano następujące statystyczne liniowe modele Bayesowskie:

- 1) Wielokrotną liniową regresję z parametrami oszacowanymi metodą Bayesowską (WLR-B).
- 2) Wielokrotną liniową regresję z parametrami oszacowanymi techniką odwrotnego modelowania (WLR-InvB).
- 3) 3D wielokrotną liniową regresję z parametrami oszacowanymi metodą Bayesowską (WWLR-B).
- 4) 3D wielokrotną liniową regresję z parametrami oszacowanymi techniką odwrotnego modelowania (WWLR-InvB).

Warunki stosowalności modeli SLMB:

- 1) Modele SLMB stosuje się do modelowania sytuacji występującej w rzece zdefiniowanej jako wezbranie.
- 2) Definicja wezbrania wymaga jednoznacznego określenia okresu, w którym płynie rzeką tak duża objętość wody, że ten okres jest uznawany za wezbranie.
- 3) Wezbranie opisywane jest za pomocą ilościowych charakterystyk (parametrów) wezbrania, tj. przepływem kulminacyjnym, objętością i czasem trwania wezbrania.
- 4) W modelach WLR-B, WLR-InvB oddzielnie modelowana jest każda ilościowa charakterystyka wezbrania – zmienna prognozowana (predyktand) jest skalarem.
- 5) W modelach WWLR-B, WWLR-InvB modelowane są ilościowe charakterystyki wezbrania łącznie – zmienna prognozowana jest wektorem.
- 6) Metoda Bayesowska zastosowana w SLMB umożliwia modelowanie zarówno zjawisk powtarzalnych (występujących okresowo, cyklicznie), jak i niepowtarzalnych.
- 7) Niepewność modelowanego zjawiska oraz niepewność związana z niewiedzą odnośnie procesów deterministycznych w modelach SLMB opisują występujące tam prawdopodobieństwo.
- 8) W analizie regresji zakłada się:
 - występowanie stałej wariancji reszt (tj. różnicy między wartością zaobserwowaną a modelowaną) dla poszczególnych wartości zmiennych prognozujących, tzw. homoscedastyczność [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 39];
 - reszty mają rozkład zbliżony do rozkładu normalnego, nie są skorelowane ze sobą (brak autokorelacji).
- 9) W SLMB są wymagania dotyczące wyboru optymalnego zbioru zmiennych prognozujących:
 - liczba zmiennych prognozujących (predyktorów) powinna być znacznie mniejsza niż liczebność próby losowej, na podstawie której ustalana jest zależność regresyjna;
 - do zastosowania analizy liniowej wymagana jest liniowa zależność, tj. liniowy związek między zmiennymi prognozującymi a zmienną prognozowaną;
 - w SLMB obowiązuje brak współliniowości zmiennych prognozujących, tzn. żaden z predyktorów (zmiennych prognozujących) nie może być kombinacją innych predyktorów, które są ze sobą silnie skorelowane.
- 10) Opracowanie modeli SLMB powinna poprzedzać statystyczna analiza danych wejściowych, w tym badanie jednorodności zmiennych prognozowanych i prognozujących za pomocą nieparametrycznych testów odpowiednich dla niezależnych zmiennych losowych.

3.4. Sieci neuronowe

Sztuczne sieci neuronowe (ANN) lub po prostu sieci neuronowe (NN) są szeroko stosowane w prognozowaniu hydrologicznym ze względu na ich zdolność do modelowania dowolnej zależności wejścia–wyjścia bez względu na stopień nieliniowości lub brak *a priori* wiedzy o własnościach fizycznych modelowanego systemu. W jednym z pierwszych zastosowań NN do prognozowania przepływu rzeki, Kang i in. [1993] wykorzystali te modele do prognozowania dobowego i godzinowego przepływu rzecznoego. Od tego czasu wiele badań potwierdziło użyteczność modeli NN w prognozowaniu przepływu rzeki [np.: Hsu i in. 2002; Tawfik 2003; Piotrowski i in. 2006; Bhadra i in. 2010; Pramanik i in. 2011; Artigue i in. 2012]. Kisi [2004] wykorzystał sieci neuronowe do prognozowania średniego miesięcznego przepływu w odniesieniu do operacji zbiornikowych, podczas gdy Kisi [2007] użył różnych algorytmów NN zarówno do krótkoterminowego, jak i długoterminowego prognozowania przepływu w rzece Norte Platte w Stanach Zjednoczonych. Powyższe eksperymenty przedstawiają możliwości zastosowania algorytmów NN jako narzędzi prognostycznych w hydrologii.

STRUKTURA SIECI NEURONOWEJ

Perceptron wielowarstwowy (Multilayer perceptron, MLP) jest typem sztucznych sieci neuronowych NN. Sieć tego typu składa się z warstwy wejściowej, jednej lub więcej warstw ukrytych oraz warstwy wyjściowej. Warstwę wyjściową mogą stanowić neurony liniowe występujące w zastosowaniach regresji lub neurony nieliniowe stosowane w zagadnieniach klasyfikacji. W metodzie użyto trójwarstwowej sieci neuronowej wstecznej propagacji błędów (Feed forward backpropagation neural network, FFBPNN) z jedną ukrytą warstwą połączoną jednym lub więcej ukrytymi węzłami.

ALGORYTM UCZENIA SIECI NEURONOWEJ

Podstawowy paradygmat modelowania NN wykorzystuje iteracyjny parametryczny schemat updatingu stanu modelu związany z procesem uczenia sieci neuronowej. Zatem ważnym aspektem tworzenia sieci FFBPNN jest wybór skutecznego algorytmu uczenia do zminimalizowania błędu sieci. W tym celu zastosowano algorytm wstecznej propagacji błędów Levenberga–Marquardta (LM), który jest wykorzystywany w procedurze backpropagation. Algorytm jest uruchamiany po to, aby zminimalizować błąd średniokwadratowy (MSE), tj. statystyczną miarę rozbieżności między obserwacją a wielkością modelowaną. Algorytm wstecznej propagacji błędów LM uważa się za skuteczny algorytm minimalizowania błędu sieci i jest on stosowany w licznych badaniach dotyczących prognozowania hydrologicznego [Babel, Shinde 2011; Campisi i in. 2012; Kim, Valdés 2003]. Adamowski i Karapatakis [2010] wykazali, że LM konsekwentnie przewyższa inne algorytmy uczenia, np. algorytm gradientów sprzężonych, quasi-Newtona. W FFBPNN wagi sieci w poszczególnych neuronach i wagi nie związane z żadnym z sygnałów są stale modyfikowane w taki sposób, aby minimalizować popełniany przez sieć błąd. Efekt uczenia sieci neuronowej jest dobry tylko wtedy, gdy w jego wyniku wszystkie wagi otrzymują wartości gwarantujące uzyskanie najmniejszej wartości błędu całej sieci. Powyższy proces odbywa się tak długo, dopóki błędy nie zostaną zmodyfikowane do odpowiedniego poziomu lub nie zostaną spełnione warunki zatrzymania algorytmu. Jest to w istocie proces minimalizacji funkcji błędu. W sieciach NN często pojawia się zjawisko przeuczenia spowodowane zbyt długim uczeniem, w wyniku którego sieć nadmiernie uzależnia swoje działanie od cech drugorzędnych, niedających podstaw do generalizacji. Parametryczna postać sieci NN oszacowana w procesie kalibracji nie przenosi generalizacji na nieobserwowane dane.

Relacja zaktualizowanych wag w poszczególnych neuronach i wag niezwiązanych z żadnym sygnałem może być powiązana z warstwą wejściową, warstwą ukrytą i wyjściami neuronów za pomocą równania:

$$Y_{\tau_k} = f_0 \left[\sum_{j=1}^m w_{kj} f_n \left(\sum_{i=1}^d w_{ji} F_{\tau_i} + w_{j0} \right) \right] + w_{k0} \quad (3.45)$$

gdzie: d – liczba wejściowych warstw neuronów (liczba zmiennych prognozujących); f_0 – funkcja aktywacji dla neuronu wyjściowego; f_n – funkcja aktywacji dla neuronu ukrytego; m – liczba ukrytych warstw neuronów; w_{j0} – waga niezwiązana z żadnym sygnałem dla j -tego ukrytego neuronu; $F_{\tau_i}^{w_{ji}}$ – waga, która łączy i -tą zmienną prognozującą w warstwie wejścia z j -tym neuronem w ukrytej warstwie; w_{k0} – waga niezwiązana z żadnym sygnałem dla k -tego neuronu wyjścia; w_{kj} – waga, która łączy j -ty neuron w ukrytej warstwie i k -ty neuron warstwy wyjściowej; F_{τ_i} – i -ta zmienna prognozująca w chwili τ ; Y_{τ_k} – prognozowane k -te wyjście w chwili τ [Kim, Valdés 2003].

3.5. Dynamiczne liniowe modele Bayesowskie (DLMB)

Przeprowadzone badania [Triantafyllopoulos, Pikoulas 2002; Ciupak 2007; Ciupak 2010a] wykazały przydatność dynamicznych liniowych modeli Bayesowskich (DLMB) w prognozowaniu hydrologicznym. DLMB potrafią odzwierciedlać zmieniającą się dynamikę poprzez liniowy updating zmieniającego się stanu w sposób zbliżony do dynamicznych zjawisk naturalnych będących cyklem hydrologicznym. Ponieważ składniki cyklu hydrologicznego występują jednocześnie na różnych poziomach złożoności, dokładne prognozowanie ich przy użyciu statycznych modeli jest trudne. Połączenie liniowych procesów z wnioskowaniem Bayesowskim w postaci DLMB [Ciupak 2010b; Ciupak i in. 2015] pomogło w wypełnieniu tej luki, umożliwiając tym samym skuteczne modelowanie pomiarowych ciągów hydrologicznych.

DLMB w postaci Varying Coefficient Regression (VCR) [Harvey 1986] zastosowano do modelowania wielowymiarowych ciągów pomiarowych z wykorzystaniem algorytmów dla przypadków jednowymiarowych, w tym do analizy międzynarodowych kursów wymiany walut. Rozkłady *a priori* i prawdopodobieństwo połączono za pomocą dynamicznej regresji Bayesa do postaci nowego rozkładu *a posteriori* w kolejnym kroku przetwarzania. Wynikiem przetwarzania był regularny updating warunkowej gęstości *a posteriori* parametrów regresji i predykcyjnego rozkładu prawdopodobieństwa. Triantafyllopoulos i Pikoulas [2002] opracowali model VCR i zastosowali go do rozwiązania problemu bezpieczeństwa sieci informatycznej, związanego z naruszeniami lub nieuprawnioną manipulacją informacjami w systemach komputerowych. W modelu nieznaną macierz kowariancji obserwacji pozostawiał nieokreślony, tzn. macierz kowariancji podlegała odwrotnemu rozkładowi Wisharta [O'Hagan, Forester 2004]. W ten sposób zostało zapewnione szybsze przetwarzanie i większa niezawodność prognozy. Salvador i Gargallo [2004] zaproponowali algorytm monitorowania i interwencji, który pozwolił na automatyczne porównywanie modeli i wybór rozwiązań w ramach wnioskowania Bayesowskiego.

Triantafyllopoulos [2006] zaproponował inne podejście do DLMB, w postaci modelu Discount Weighted Regression (DWR) przeznaczonego do prognozowania wielowymiarowych zmiennych losowych opisujących zmieniające się w czasie dane na Londyńskiej Giełdzie Metali. Prognoza dotyczyła terminowych danych giełdowych i przyszłych kontraktowych cen zamknięcia. Dyskontowe ważone modele regresji (DWR) zostały odpowiednio zmodyfikowane w celu uwzględnienia wielowymiarowego modelu na poziomie lokalnym, natomiast macierze kowariancji pozostały nieokreślone i szacowano je metodą największej wiarygodności.

Najbliższym prognozy hydrologicznej zagadnieniem było zastosowanie modelu DWR przez Solhjella [2009] do przewidywania odkształceń mostu nad rzeką Göta w Szwecji. W opracowanym rozwiązaniu wykorzystano wnioskowanie Bayesowskie opisane przez Westa i Harrisona [1999]. Ponieważ model uwzględniał temperaturę powietrza, prognozy wymagały zmiennych, które nie byłyby znane w czasie rzeczywistym. Kwestię tę rozwiązano poprzez wprowadzenie do modelu informacji o temperaturze powietrza za pomocą przygotowanych scenariuszy z danymi meteorologicznymi.

3.5.1. Dynamiczny liniowy model Bayesowski: Varying Coefficient Regression (VCR)

W kontekście hydrologii podstawowe założenia w modelach DLMB (VCR i DWR) mówią o tym że: i) przepływ w rzece jest obserwowany w regularnych odstępach czasu, ii) z każdą obserwacją skojarzony jest taki sam błąd pomiaru, iii) stan systemu z przeszłości wpływa na nową obserwację. Do zapisu wielokrotnych i macierzowych rozkładów prawdopodobieństwa przyjęto notację zgodną z Krzyżko [2009]. Opis modeli DLMB, tj. VCR i DWR, przedstawiono dla przypadku wielowymiarowego, tzn. dla wektora zmiennych prognozowanych o liczebności $l > 1$, natomiast do obliczeń wykorzystano modele jednowymiarowe, w których liczba zmiennych prognozowanych $l = 1$. Model VCR zdefiniowano za pomocą dwóch równań [West, Harrison 1999; Triantafyllopoulos 2002]:

1) równania obserwacji:

$$\mathbf{Y}_\tau^T = \mathbf{F}_\tau^T \boldsymbol{\theta}_\tau + \mathbf{v}_\tau \quad (\mathbf{v}_\tau | \boldsymbol{\Sigma}) \sim \mathbf{N}(\mathbf{0}, \mathbf{V}_\tau \boldsymbol{\Sigma}) \quad (3.46)$$

2) równania przejścia:

$$\boldsymbol{\theta}_\tau = \mathbf{G}_\tau \boldsymbol{\theta}_{\tau-1} + \boldsymbol{\Omega}_\tau \quad (\boldsymbol{\Omega}_\tau | \boldsymbol{\Sigma}) \sim \mathbf{N}(\mathbf{0}, \mathbf{W}_\tau \otimes \boldsymbol{\Sigma}) \quad (3.47)$$

gdzie: τ – moment obserwacji; d – liczba zmiennych prognozujących; l – liczba zmiennych prognozowanych; n – liczba obserwacji; \mathbf{Y} – $n \times l$ wymiarowa macierz z zaobserwowanymi wartościami zmiennych prognozowanych; \mathbf{Y}_τ – l wymiarowy wektor z zaobserwowanymi wartościami zmiennych prognozowanych w chwili τ ; \mathbf{F}_τ^T – $n \times d$ wymiarowa transponowana macierz zaobserwowanych zmiennych prognozujących; $\boldsymbol{\theta}_\tau$ – $d \times l$ wymiarowa nieobserwowana macierz zmiennych stanu \mathbf{v}_τ – l wymiarowy wektor zaobserwowanych błędów w równaniu obserwacji (3.46); $\boldsymbol{\Sigma}$ – $l \times l$ wymiarowa macierz kowariancji; \mathbf{V}_τ – obserwacyjny czynnik skali wariancji; \mathbf{G}_τ – $d \times d$ wymiarowa macierz przejścia z jednego

stanu do drugiego; $\Omega_\tau - d \times l$ wymiarowa macierz, w której kolumny są błędami transformacji zmiennych stanu w równaniu (3.47); $\mathbf{W}_\tau - d \times d$ wymiarowa macierz kowariancji macierzy losowej Ω_τ w równaniu przejścia (3.47); $N(\mathbf{0}, V_\tau, \Sigma)$ – wielowymiarowy rozkład normalny o wektorze wartości oczekiwanej $\mathbf{0}$ i macierzy kowariancji Σ z uwzględnieniem czynnika skali wariancji V_τ [patrz szczegóły w Dodatku B]; $N(\mathbf{0}, \mathbf{W}_\tau \otimes \Sigma) - d \times l$ macierzowy rozkład normalny o macierzy wartości oczekiwanej $\mathbf{0}$ i lewą i prawą macierzą kowariancji \mathbf{W}_τ i Σ [szczegóły w Załączniku: Dodatek B].

Wartości elementów wektora \mathbf{v}_τ opisują zmienność wektora \mathbf{Y}_τ nieopisaną przez zmienne stanu modelu, a Ω_τ opisują stratę w czasie informacji o wartości macierzy θ_τ w chwili $\tau - 1$ i τ . Z założenia wektor \mathbf{v}_τ i macierz Ω_τ są w każdym momencie τ wzajemnie niezależne i podlegają odpowiednio wielowymiarowemu i macierzowemu rozkładowi normalnemu ze średnią (wartością oczekiwaną) równą $\mathbf{0}$ [szczegóły w Załączniku: Dodatek B]. Jeżeli średnia wartość jest równa $\mathbf{0}$, to należy oczekiwać stałości zmiennych stanu macierzy θ_τ w całym okresie obserwacji. Zaobserwowane zmienne prognozujące (wejście do VCR) w postaci hydrologicznych i meteorologicznych obserwacji są zapisane w $d \times n$ macierzy \mathbf{F}_τ , podobnie jak zmienne prognozowane w postaci l wymiarowego wektora \mathbf{Y}_τ zawierającego zaobserwowane wartości w chwili τ dla l zmiennych prognozowanych. Dodatkowo model został uzupełniony o następujące zmienne stanu: $\mathbf{m}_\tau - d \times l$ wymiarowa macierz zawierająca współczynniki kierunkowe dla każdej zmiennej prognozowanej; $\mathbf{C}_\tau - d \times d$ wymiarowa macierz kowariancji; $\mathbf{S}_\tau - l \times l$ wymiarowa zaobserwowana *a priori* macierz wariancji; $n_\tau - a$ *a priori* stopień swobody.

W algorytmie pojawiają się następujące zmienne pomocnicze: $\mathbf{a}_\tau - d \times l$ wymiarowa macierz będąca estymatorem współczynników kierunkowych; $\mathbf{R}_\tau - d \times d$ wymiarowa macierz ze zmodyfikowanymi wartościami kowariancji \mathbf{C}_τ ; $D_0 -$ zbiór początkowej informacji (tj.: $\mathbf{m}_0, \mathbf{C}_0, \mathbf{S}_0, \mathbf{G}_0, \mathbf{F}_0$) w chwili $\tau = 0$, dla $\tau > 0$ $D_\tau = \{\mathbf{Y}_\tau, D_{\tau-1}\}$; $\mathbf{f}_\tau - l$ wymiarowy wektor z prognozą wektora \mathbf{Y}_τ na podstawie informacji $D_{\tau-1}$; $Q_\tau -$ skala wariancji; \mathbf{A}_τ i $\mathbf{e}_\tau -$ macierz i wektor o wymiarach odpowiednio $d \times n$ i l .

ALGORYTM PROCEDURY OBLICZENIOWEJ MODELU VARYING COEFFICIENT REGRESSION (VCR)

W pełni Bayesowski (zgodny z wnioskowaniem Bayesowskim) algorytm procedury obliczeniowej modelu VCR (rys. 3.1) zawiera pięć kroków [West, Harrison 1999]:

- 1) *a posteriori* w chwili $\tau - 1$;
- 2) *a priori* w chwili τ ;
- 3) pierwszy krok prognozy;
- 4) *a posteriori* w chwili τ (updating zgodnie z nowymi obserwacjami);
- 5) prognoza k -kroków naprzód.

Model może być uruchomiony w trybie kalibracji, walidacji i prognozy, tj. nie tylko jako model symulacyjny, ale również jako model do prognozowania hydrologicznego (rys. 3.1).

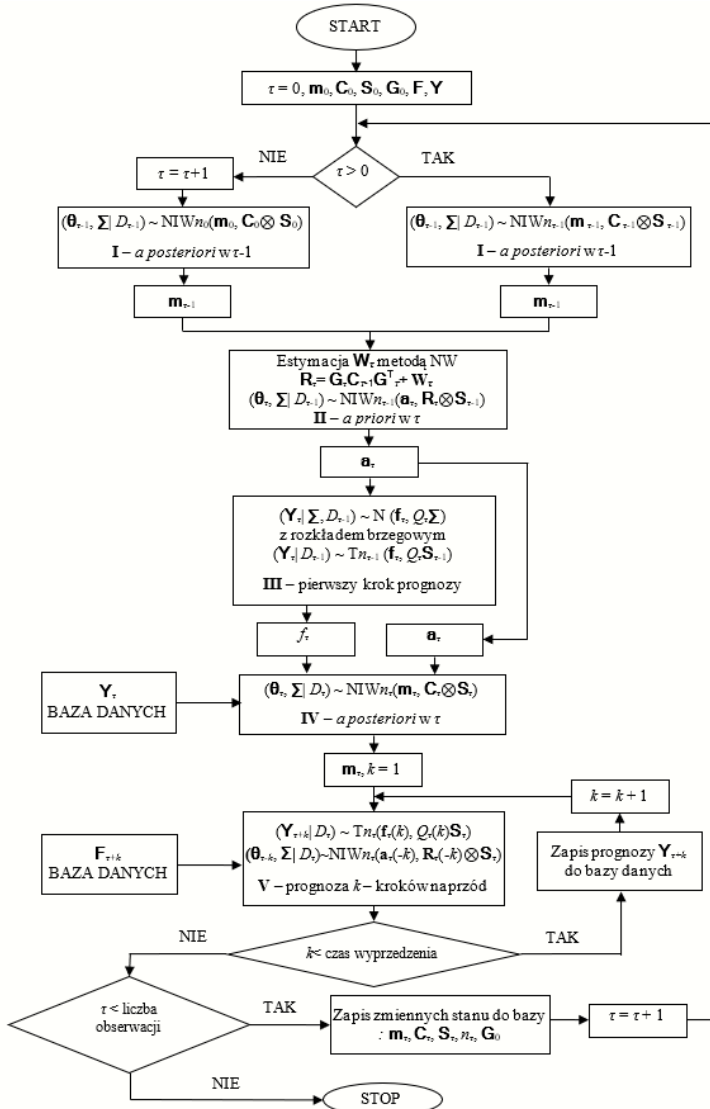
W trybie kalibracji w trakcie pierwszego uruchomienia modelu (dla $\tau = 0$) bardzo ważnym krokiem jest określenie początkowych (inicjujących) wartości opisanych zmiennych stanu, gdyż proces ten wyraża probabilistyczną reprezentację wiary synoptyka w te zmienne. Na przykład macierz \mathbf{m}_0 może być określona na podstawie historycznej informacji uzyskanej z eksperymentu, \mathbf{C}_0 może być zbiorem w postaci dużej diagonalnej macierzy odzwierciedlającej niską precyzję (lub wysoką niepewność) w określeniu wartości elementów macierzy θ_0 . Zatem zakres zmienności, w szczególności macierzy kowariancji \mathbf{C}_0 , jest obliczany na podstawie wielokrotnego uruchomienia modelu z różnymi wartościami \mathbf{C}_0 . W modelu VCR założono, że nieobserwowana macierz zmiennych stanu θ ma macierzowy rozkład normalny/odwrotny rozkład Wisharta ($\theta|\Sigma) \sim \text{NIW}_n(\mathbf{m}, \mathbf{C} \otimes \Sigma)$, czyli dodatkowo macierz kowariancji Σ podlega odwrotnemu rozkładowi Wisharta [West, Harrison 1999; O'Hagan, Forester 2004].

W pierwszym kroku algorytmu (rys. 3.1) model oblicza macierzowy rozkład normalny/odwrotny rozkład Wisharta macierzy $\theta_{\tau-1}$, tj. ($\theta_{\tau-1}, \Sigma|D_{\tau-1}) \sim \text{NIW}_{n_{\tau-1}}(\mathbf{m}_{\tau-1}, \mathbf{C}_{\tau-1} \otimes \Sigma_{\tau-1})$ ze zmiennymi stanu otrzymanymi w $\tau - 1$ [szczegóły algorytmu w Załączniku: Dodatek C]. W ten sposób algorytm przygotowuje updating (tj. aktualizację zmiennych stanu modelu) i równania prognostyczne.

W czasie drugiego kroku w chwili τ obliczany jest rozkład *a priori* macierzy θ_τ (tj. prognozę macierzy θ_τ na podstawie informacji $D_{\tau-1}$). Prognoza ma macierzowy rozkład normalny/odwrotny rozkład Wisharta z parametrami obliczonymi w następujący sposób: $\mathbf{a}_\tau = \mathbf{G}_\tau \mathbf{m}_{\tau-1}$ i $\mathbf{R}_\tau = \mathbf{G}_\tau \mathbf{C}_{\tau-1} \mathbf{G}_\tau^T + \mathbf{W}_\tau$. Macierz kowariancji \mathbf{W}_τ macierzy losowej Ω_τ jest oszacowywana za pomocą metody największej wiarygodności.

W trzecim kroku algorytmu obliczana jest prognoza wektora Y_τ w chwili τ na podstawie informacji $D_{\tau-1}$. Prognoza ma wielowymiarowy rozkład normalny $(Y_\tau | \Sigma, D_{\tau-1}) \sim N(f_\tau, Q; \Sigma)$ z rozkładem brzegowym $(Y_\tau | D_{\tau-1}) \sim Tn_{\tau-1}(f_\tau, Q; S_{\tau-1})$. Do obliczenia brzegowego rozkładu $(Y_\tau | D_{\tau-1})$ zastosowano wielowymiarowy rozkład t -Studenta z parametrami obliczonymi w następujący sposób: $f_\tau^T = F_\tau^T a_\tau$ i $Q_\tau = V_\tau + F_\tau^T R_\tau F_\tau$.

Symulowany hydrogram w chwili τ porównywany jest do obserwowanego. Normalnie obserwowane i symulowane hydrogramy różnią się znacząco, co sprawia, że konieczne jest zastosowanie procedury filtrowania. Różnica między obserwowanymi i symulowanymi wartościami zmiennej prognozowanej obliczana jest w czwartym kroku.



Rys. 3.1. Algorytm dynamicznego liniowego modelu Bayesowskiego – Varying Coefficient Regression (VCR).
Fig. 3.1. Algorithm of dynamic linear Bayesian model – Varying Coefficient Regression (VCR).

W czwartym kroku *a posteriori* w chwili τ za pomocą filtru Kalmana [Jaźwiński 1970] szacowana jest macierz $\boldsymbol{\theta}_\tau$ na podstawie informacji $D_\tau = \{\mathbf{Y}_\tau, D_{\tau-1}\}$. Szacowana macierz $\boldsymbol{\theta}_\tau$ ma macierzowy rozkład normalny odwrotny rozkład Wishart'a ($\boldsymbol{\theta}_\tau, \boldsymbol{\Sigma}|D_\tau \sim \text{NIW}n_\tau(\mathbf{m}_\tau, \mathbf{C}_\tau \otimes \mathbf{S}_\tau)$). Wejściem do procedury obliczającej parametry rozkładu macierzy $\boldsymbol{\theta}_\tau$ jest wektor zaobserwowanych zmiennych prognozowanych \mathbf{Y}_τ w chwili τ . Parametry modelu wyznaczone są według następujących równań: $\mathbf{m}_\tau = \mathbf{a}_\tau + \mathbf{A}_\tau \mathbf{e}_\tau^T$, $\mathbf{C}_\tau = \mathbf{R}_\tau - \mathbf{A}_\tau \mathbf{Q}_\tau \mathbf{A}_\tau^T$, $n_\tau = n_{\tau-1} + 1$, $\mathbf{S}_\tau = n_\tau^{-1}(n_{\tau-1} \mathbf{S}_{\tau-1} + \mathbf{e}_\tau \mathbf{e}_\tau^T \mathbf{Q}_\tau^{-1})$, gdzie: $\mathbf{A}_\tau = \mathbf{R}_\tau \mathbf{F}_\tau \mathbf{Q}_\tau^{-1}$ i $\mathbf{e}_\tau = \mathbf{Y}_\tau - \mathbf{f}_\tau$.

W trakcie obliczania macierzy wariancji \mathbf{S}_τ , jeżeli zachodzi taka konieczność, wprowadza się korektę i w ten sposób zmodyfikowana macierz \mathbf{S}_τ staje się wejściem do obliczania współczynników kierunkowych macierzy \mathbf{m}_τ . Jest to ważna procedura z punktu widzenia możliwości uruchomienia modelu do prognozowania, w którym należy brać pod uwagę możliwość zmiany wartości zmiennych stanu.

Do prognozowania w chwili τ , tj. w piątym kroku algorytmu, wymagana jest znajomość k -rozkładów brzegowych naprzód $p(\mathbf{Y}_{\tau+k}|D_\tau)$. W tym celu należy dostarczyć do modelu w każdym kolejnym kroku $\tau + k$ rzeczywistą wartość zaobserwowanej zmiennej modelowanej w chwili τ . W piątym kroku do updatingu rozkładu *a posteriori* (proces prognozowania) stosowany jest wielowymiarowy rozkład t -Studenta. Prognozę wektora $\mathbf{Y}_{\tau+k}$ i oszacowanie macierzy zmiennych stanu dla $1 < k \leq \tau$, tj. $\boldsymbol{\theta}_{\tau-k}$ opisują następujące równania: ($\mathbf{Y}_{\tau+k}|D_\tau \sim \text{T}n_\tau(\mathbf{f}_\tau(k), \boldsymbol{\theta}_\tau(k)\mathbf{S}_\tau)$) i ($\boldsymbol{\theta}_{\tau-k}, \boldsymbol{\Sigma}|D_\tau \sim \text{NIW}n_\tau(\mathbf{a}_\tau(-k), \mathbf{R}_\tau(-k) \otimes \mathbf{S}_\tau)$), gdzie: $\mathbf{a}_\tau(-k) = \mathbf{m}_{\tau-k} + \mathbf{C}_{\tau-k} \mathbf{G}_{\tau-k+1}^T \mathbf{R}_{\tau-k+1}^{-1} [\mathbf{a}_\tau(-k+1) - \mathbf{a}_{\tau-k+1}]$. Drugim wymogiem uruchomienia modelu symulacyjnego do prognozowania hydrologicznego jest umożliwienie zapisywania zmiennych stanu wraz z wynikami prognoz. Wyjściem z modelu VCR jest obliczony hydrogram w postaci prognozy punktowej i przedziałowej z określonym prawdopodobieństwem niepewności.

3.5.2. Dynamiczny liniowy model Bayesowski: Discount Weighted Regression (DWR)

Model DWR zdefiniowany jest za pomocą dwóch równań [Triantafyllopoulos 2006]:

1) równania obserwacji:

$$\mathbf{Y}_\tau^T = \mathbf{F}_\tau^T \boldsymbol{\theta}_\tau + \mathbf{v}_\tau \quad (\mathbf{v}_\tau|\boldsymbol{\Sigma}) \sim \text{N}(\mathbf{0}, V_\tau \boldsymbol{\Sigma}) \quad (3.48)$$

2) równania przejścia:

$$\boldsymbol{\theta}_\tau = \boldsymbol{\theta}_{\tau-1} + \boldsymbol{\Omega}_\tau \quad (\boldsymbol{\Omega}_\tau|\boldsymbol{\Sigma}) \sim \text{N}(\mathbf{0}, \mathbf{W}_\tau \otimes \boldsymbol{\Sigma}) \quad (3.49)$$

Wejście do modelu jest zdefiniowane w taki sam sposób, jak wejście do modelu VCR. Dodatkowo w modelu pojawia się stały współczynnik dyskontowy δ [Spis użytych pojęć i stosowanego nazywania, pkt. 40]. Macierz transformacji kowariancji \mathbf{W}_τ jest określona za pomocą przybliżenia dyskontowego wg wzoru:

$$\mathbf{W}_\tau = \frac{1-\delta}{\delta} \mathbf{C}_{\tau-1} \quad (3.50)$$

gdzie: $\delta \in (0,1]$, a $\mathbf{C}_{\tau-1}$ jest macierzą kowariancji obliczaną w momencie $\tau - 1$.

Podobnie jak w algorytmie modelu VCR pojawiają się zmienne pomocnicze: $\mathbf{a}_\tau - d \times l$ wymiarowa macierz będąca estymatorem współczynników kierunkowych; $\mathbf{R}_\tau - d \times d$ wymiarowa macierz z wartościami kowariancji \mathbf{C}_τ zmodyfikowanymi za pomocą czynnika dyskontowego δ .

ALGORYTM PROCEDURY OBLICZENIOWEJ MODELU DISCOUNT WEIGHTED REGRESSION (DWR)

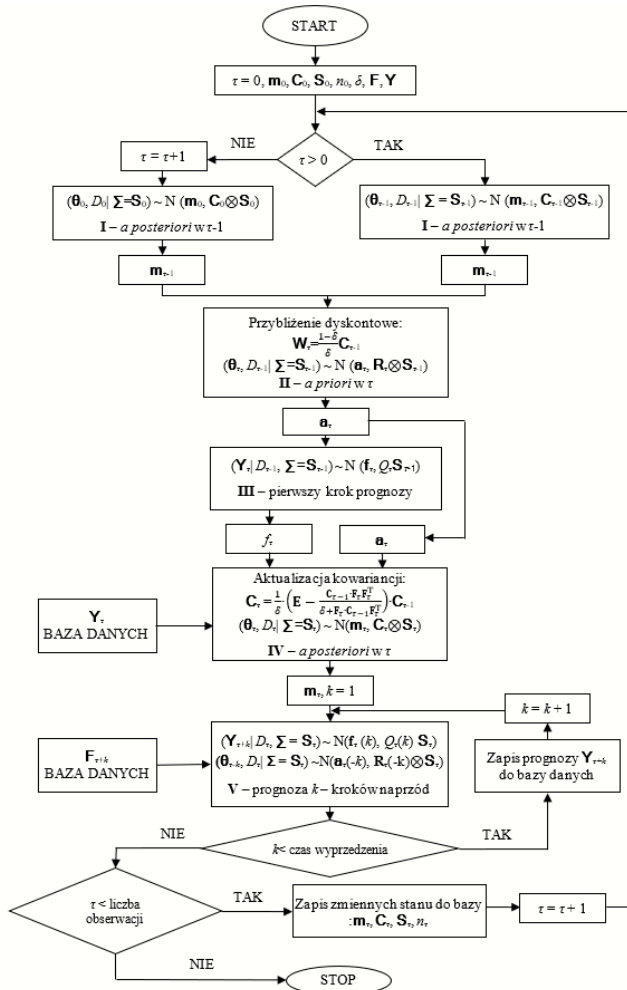
Algorytm procedury obliczeniowej DWR (rys. 3.2) jest podobny do algorytmu modelu VCR i składa się z tych samych pięciu kroków. Różnica między modelami polega na różnych wielowymiarowych lub macierzowych rozkładach prawdopodobieństwa zastosowanych w obu algorytmach. Zatem jedyna różnica pomiędzy DWR a VCR polega na zastosowanej metodzie modelowania macierzy kowariancji $\boldsymbol{\Sigma}$.

W modelu DWR założono, że odwrotny rozkład Wishart'a nieznannej macierzy kowariancji zastosowany w modelu VCR może być szczególnym przypadkiem dyskontowego komponentu w modelu DWR. Zakładając, że macierz przejścia \mathbf{G}_τ zostanie zastąpiona $d \times d$ wymiarową macierzą jednostkową

\mathbf{E} , obserwacyjny czynnik skali wariancji V_τ przyjmie wartość równą 1, a macierz kowariancji transformacji \mathbf{W}_τ będzie wyliczana zgodnie ze wzorem (3.50), zostanie otrzymany nowy dynamiczny liniowy model DWR.

Po uwzględnieniu powyższych podstawień ulegną zmianie równania następujących zmiennych w porównaniu do modelu VCR: $\mathbf{a}_\tau = \mathbf{G}_\tau \mathbf{m}_{\tau-1} = \mathbf{m}_{\tau-1}$, $\mathbf{f}_\tau^T = \mathbf{F}_\tau^T \mathbf{a}_\tau = \mathbf{F}_\tau^T \mathbf{m}_{\tau-1}$, $\mathbf{m}_\tau = \mathbf{a}_\tau + \mathbf{A}_\tau \mathbf{e}_\tau^T = \mathbf{m}_{\tau-1} + \mathbf{A}_\tau \mathbf{e}_\tau$, $\mathbf{e}_\tau = \mathbf{Y}_\tau - \mathbf{f}_\tau = \mathbf{Y}_\tau - \mathbf{F}_\tau^T \mathbf{m}_{\tau-1}$, $\mathbf{S}_\tau = n_\tau^{-1}(n_{\tau-1} \mathbf{S}_{\tau-1} + \mathbf{e}_\tau \mathbf{e}_\tau^T Q_\tau^{-1}) = n_\tau^{-1} [n_{\tau-1} \mathbf{S}_{\tau-1} + (\mathbf{Y}_\tau - \mathbf{m}_\tau \mathbf{F}_\tau) \mathbf{e}_\tau^T]$, a macierz kowariancji \mathbf{C}_τ jest aktualizowana z uwzględnieniem współczynnika dyskontowego:

$$\mathbf{C}_\tau = \frac{1}{\delta} \left(\mathbf{E} - \frac{\mathbf{C}_{\tau-1} \mathbf{F}_\tau \mathbf{F}_\tau^T}{\delta + \mathbf{F}_\tau \mathbf{C}_{\tau-1} \mathbf{F}_\tau^T} \right) \mathbf{C}_{\tau-1}.$$



Rys. 3.2. Algorytm dynamicznego liniowego modelu Bayesowskiego – Discount Weighted Regression (DWR).
Fig. 3.2. Algorithm of dynamic linear Bayesian model – Discount Weighted Regression (DWR).

W modelu DWR założono, że w dowolnej chwili $\tau > 0$ macierz nieobserwowanych współczynników kierunkowych θ_τ na podstawie informacji $D_{\tau-1}$ ma macierzowy rozkład normalny [Dawid 1981; Gupta, Nagar 1999], czyli $(\theta_{\tau-1}, D_{\tau-1} | \Sigma = \mathbf{S}_\tau) \sim N(\mathbf{m}_{\tau-1}, \mathbf{C}_{\tau-1} \otimes \mathbf{S}_{\tau-1})$, gdzie $D_{\tau-1}$ zawiera zbiór informacji w kroku $\tau - 1$ i wektor obserwacji \mathbf{Y}_τ [patrz szczegóły macierzowego rozkładu normalnego w Dodatku

B]. W trzecim kroku do obliczania rozkładu brzegowego ($\mathbf{Y}_\tau | D_{\tau-1}$) zastosowano wielowymiarowy rozkład normalny, tj. $(\mathbf{Y}_\tau | D_{\tau-1}, \boldsymbol{\Sigma} = \mathbf{S}_{\tau-1}) \sim N(\mathbf{f}_\tau, Q_\tau \mathbf{S}_{\tau-1})$, natomiast w piątym kroku do updatingu rozkładu *a posteriori* (proces prognozowania), zastosowano wielowymiarowy i macierzowy rozkład normalny zgodnie z następującymi równaniami: $(\mathbf{Y}_{\tau+k} | D_\tau, \boldsymbol{\Sigma} = \mathbf{S}_\tau) \sim N(\mathbf{f}_\tau(k), Q_\tau(k) \mathbf{S}_\tau)$ i $(\boldsymbol{\theta}_{\tau-k}, D_\tau | \boldsymbol{\Sigma} = \mathbf{S}_\tau) \sim N(\mathbf{a}_\tau(-k), \mathbf{R}_\tau(-k) \otimes \mathbf{S}_\tau)$. Rozkłady prawdopodobieństwa obliczane są z parametrami wyznaczonymi na podstawie równań opisanych w modelu VCR i ich modyfikacji w modelu DWR. Zarówno w algorytmie modelu VCR, jak i DWR zastosowano mechanizm uczenia w postaci filtru Kalmana. Oba modele rozpoczynają obliczenia ze zmiennymi stanu zapisanymi w poprzednich krokach. Powyższe wymagania są niezbędne do uruchomienia modeli w trybie operacyjnym (prognozy).

3.5.3. Zastosowanie modeli DLMB i warunki stosowalności

Dynamiczne liniowe modele Bayesowskie (DLMB) zastosowane do modelowania całorocznego hydrogramu przepływów:

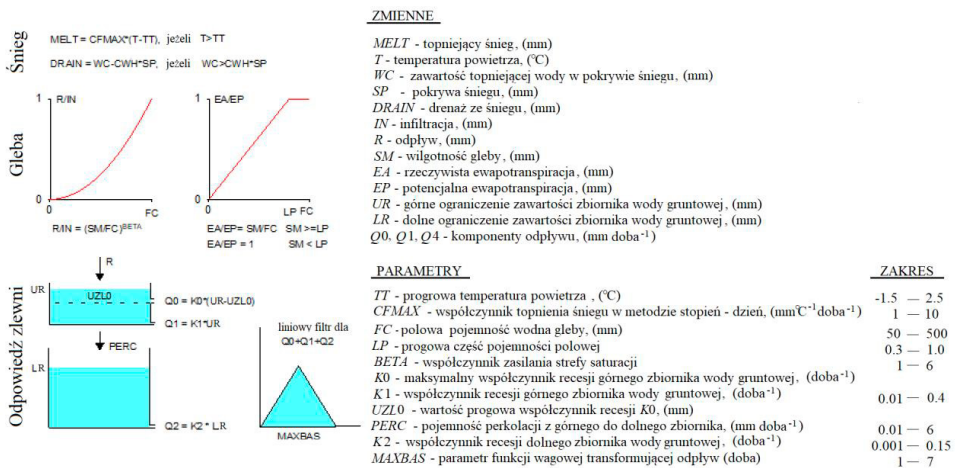
- 1) Dynamiczny liniowy model Bayesowski – Varying Coefficient Regression (VCR).
- 2) Dynamiczny liniowy model Bayesowski – Discount Weighted Regression (DWR).

Warunki stosowalności modeli DLMB:

- 1) Modele DLMB stosuje się do symulacji rocznych hydrogramów i krótkoterminowych prognoz przepływu rzecznej z wyprzedzeniem 1-, 2- i 3-dobowym.
- 2) Podstawowym warunkiem stosowania DLMB w symulacji i prognozie hydrologicznej jest:
 - obserwowanie przepływu rzecznej w regularnych odstępach czasu;
 - skojarzenie z każdą obserwacją takiego samego rozkładu błędu pomiaru;
 - uzależnienie nowej obserwacji przepływu od przepływu z poprzedniego kroku.
- 3) Modelowany przepływ rzeczny należy opisać zależnością przyczynowo-skutkową, tzn. zmienne prognozujące powinny być przyczyną zmiennej prognozowanej.
- 4) Ciągi pomiarowe będące elementem cyklu hydrologicznego są przetwarzane w DLMB dzięki połączeniu liniowych procesów regresji z wnioskowaniem Bayesowskim.
- 5) W modelach liniowej regresji zależność nieliniową można uwzględnić poprzez odpowiednią transformację zmiennych prognozowanych polegającą na zlogarytmowaniu obu stron równania regresji.
- 6) W modelach DLMB prognoza probabilistyczna wyrażana jest w postaci funkcji rozkładu prawdopodobieństwa *a posteriori*.
- 7) Modele DLMB mogą być stosowane w operacyjnych systemach hydrologicznych, gdyż zawarty w modelu mechanizm filtru Kalmana umożliwia updating zmiennych stanu oraz zapis zaktualizowanych zmiennych stanu w każdym kroku przetwarzania.
- 8) Symulacja i prognoza hydrologiczna obliczana jest za pomocą 1-wymiarowego modelu Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR), natomiast do 3-wymiarowej symulacji i symulacji w trybie prognozy stosowany jest odpowiednio Matrix Normal Regression (MNR) i Multivariate Discount Weighted Regression (MDWR).
- 9) Model 1-wymiarowy tworzy skalarna zmienna prognozowana, zaś model 3-wymiarowy 3-elementowy wektor zmiennych prognozowanych.
- 10) W ciągach wejściowych do DLMB może być obserwowany:
 - trend i wahania okresowe;
 - trend bez wahań okresowych;
 - wahania okresowe bez trendu.
- 11) Opracowanie DLMB wymaga przeprowadzenia analizy statystycznej pomiarowych ciągów wejściowych zawierającej:
 - badanie jednorodności statystycznej, tj. identyfikacji obserwacji odstających, sprawdzenia niezależności zmiennych losowych, identyfikacji trendu, identyfikacji występowania i wpływu wahań okresowych, sprawdzenia stacjonarności z uwzględnieniem trendu i braku trendu;
 - wyznaczanie charakterystyk rozkładu prawdopodobieństwa.

3.6. Konceptualny hydrologiczny model typu opad-odpływ HBV

Konceptualny model hydrologiczny Hydrologiska Byråns Vattenbalansavdelning, HBV (wersja HBV 7.3) [Bergström 1976, 1992, 1995; Lindström i in. 1997] jest częściowo rozłożonym modelem typu opad-odpływ zawierającym procedury służące do obliczania akumulacji śniegu, topnienia śniegu, wilgotności gleby, generowania odpływu i prostej transformacji wody w korycie rzeki. Wykorzystuje proste równania ciągłości i inne uzupełniające zależności służące do oszacowania niektórych elementów cyklu hydrologicznego. Strukturę modelu cechuje nieduża liczba parametrów (około 11), które wystarczają do kalibracji modelu w różnych warunkach hydrologicznych i klimatycznych (rys. 3.3). Lista wszystkich parametrów modelu, ich domyślne wartości i skala korekcji zawarte są w pracach Johnel i in. [2011] oraz Johansson [2013]. W asymilacji meteorologicznych satelitarnych obserwacji kluczową rolę odgrywa modelowanie złożonej dynamiki zmian wilgotności gleby. W modelu HBV modelowany jest tylko wpływ wilgotności gleby na generowanie odpływu w skali całej zlewni. Algorytm opiera się na modyfikacji procedury sortowania kubełkowego, w której zakłada się statystyczny rozkład potencjału magazynowania wody w zlewni [Lindström i in. 1997]. Największy odpływ jest osiągnięty, gdy wszystkie zbiorniki modelu są pełne i tym samym przyczyniają się do zwiększenia objętości odpływu. W takich sytuacjach równania transformacji wilgotności gleby prowadzą do maksymalnego przechowywania wilgoci w glebie. Precyzyjne oszacowanie tej wielkości fizycznej ma duży wpływ na symulację przepływu zwłaszcza podczas letnich wzebrań. HBV jako model częściowo-rozłożony umożliwia podział zlewni na kilka części i każdą z nich rozkłada zgodnie z podziałem na strefę wysokości i wegetacyjną.



Rys. 3.3. Schemat modułów obliczeniowych modelu hydrologicznego HBV, tj.: modułu pokrywy śnieżnej, modułu strefy aeracji, modułu generowania odpływu ze zlewni, równania matematyczne i zależności empiryczne oraz lista wybranych zmiennych stanu i parametrów do kalibracji modelu.

Fig. 3.3. Scheme of calculation modules of the HBV hydrological model, i.e.: snow routine, soil routine, response routine, mathematical equations and empirical relationships, and a list of selected state variables and parameters for model calibration.

3.7. Metody usuwania błędu systematycznego (obciążenia)

Obserwacje meteorologiczne dostarczane do modelu matematycznego, a pochodzące z różnych źródeł, np. z teledetekcji satelitarnej, naziemnej sieci pomiarowo-obszaryjnej, modeli ziemi integrujących meteorologiczne obserwacje wymuszające model hydrologiczny (opad, temperaturę), modeli klimatu, często wykazują różne wartości średnie i zmienność statystyczną (np. zakres obserwacji, rozstęp międzykwartylowy, wariancję i odchylenie standardowe). Problemem jest błąd systematyczny (obciąż-

żenie), który wpływa na przetwarzanie hydrologiczne w modelach matematycznych. W literaturze naukowej opisano wiele rozwiązań przeznaczonych do usuwania obciążenia z danych opadowych otrzymanych z modeli meteorologicznych i klimatycznych [Wilby i in. 1998; Lehner i in. 2006; Sharma i in. 2007; Piani i in. 2010; Sequí i in. 2010; Bennet i in. 2011; Ahmed i in. 2013, Osuch i in. 2016]. W dużo mniejszym stopniu opisane w powyższych publikacjach metody dotyczą korekty obciążenia obserwacji teledetekcji satelitarnej [Müller, Thomson 2013; Demirel i in. 2017]. Z punktu widzenia matematyki [Piani i in. 2010] usuwanie obciążenia jest transformacją statystyczną polegającą na znalezieniu optymalnej funkcji $f(\cdot)$, która odwzorowuje zmienną modelowaną X_{MOD} w taki sposób, że histogram intensywności skorygowanej zmiennej losowej X_{COR} jest zgodny z histogramem intensywności zmiennej obserwowanej X_{OBS} . Skorygowana zmienna losowa X_{COR} jest funkcją jej modelowanego odpowiednika X_{MOD} :

$$X_{COR} = f(X_{MOD}) \quad (3.51)$$

Metody korekcji stosowane do danych meteorologicznych podzielono na cztery grupy główne [Gudmundsson i in. 2012]:

- 1) Metody oparte na prostych zmianach [Lehner i in. 2006].
- 2) Transformacja parametryczna [Piani i in. 2010; Maraun 2013; Rojas i in. 2011].
- 3) Transformacja nieparametryczna [Wood i in. 2004; Bennet i in. 2011].
- 4) Transformacja błędu systematycznego metodą transformaty dystrybuanty (distribution-derived transformation algorithm) [Sharma i in. 2007; Salvi i in. 2011; Kurnik i in. 2012].

Przykładem metody opartej na prostych zmianach (i) jest metoda Delta [Lehner i in. 2006]. Ta podstawowa metoda korekcji obciążenia obejmuje korektę wartości średniej poprzez tymczasowe dodanie przesunięcia do wartości zaobserwowanej lub poprzez skojarzenie odpowiedniego współczynnika korekcji z danymi modelowanymi (korygowanymi). Dodana lub multiplikatywna stała kwantyfikuje średnie odchylenie między korygowanymi a zaobserwowanymi ciągami danych w okresie historycznym. Metodę Delta [Verzano 2009; Eisner i in. 2012; Hempel i in. 2013] zastosowano do statystycznej korekty obciążenia dobowego opadu na podstawie średnich miesięcznych sum opadu historycznego i danych otrzymanych z modelu meteorologicznego.

Drugą metodą usuwania obciążenia jest transformacja parametryczna (ii). Stosowana jest do wygładzania histogramu i do kwantylowego przyporządkowania (quantile mapping) opadu. Metoda została opisana szczegółowo przez Piani i in. [2010]. Zależność kwantyl-kwantyl może być modelowana bezpośrednio przy użyciu następujących transformacji parametrycznych:

$$X_{COR} = bX_{MOD} \quad (3.52)$$

$$X_{COR} = a + bX_{MOD} \quad (3.53)$$

$$X_{COR} = b(X_{MOD} - \varepsilon)^c \quad (3.54)$$

$$X_{COR} = (a + bX_{MOD}) \left(1 - \exp\left(-\frac{X_{MOD}-\varepsilon}{\tau}\right) \right) \quad (3.55)$$

gdzie: X_{MOD} – reprezentuje daną z modelu do korekcji; X_{COR} – reprezentuje najlepsze oszacowanie wartości X_{MOD} i a , b , c , ε i τ są parametrami, które podlegają estymacji. Wszystkie parametry funkcji transformacji są dopasowywane po przez zgodność histogramów dla $X_{MOD} > 0$. W powyższej metodzie w przypadku $X_{MOD} = 0$ (w okresach suchych, bezdopadowych) przyjmuje się, że $X_{COR} = \varepsilon$.

Transformacja nieparametryczna (iii) wyznaczana jest dwoma sposobami, tj. za pomocą kwantyli empirycznych i regresji nieparametrycznej. W metodzie kwantyli empirycznych przyjmuje się, że rozkład badanej zmiennej losowej jest nieznan, a transformacja jest zdefiniowana jako [Piani i in. 2010]:

$$X_{OBS} = F_{OBS}^{-1}(F_{MOD}(X_{MOD})) \quad (3.56)$$

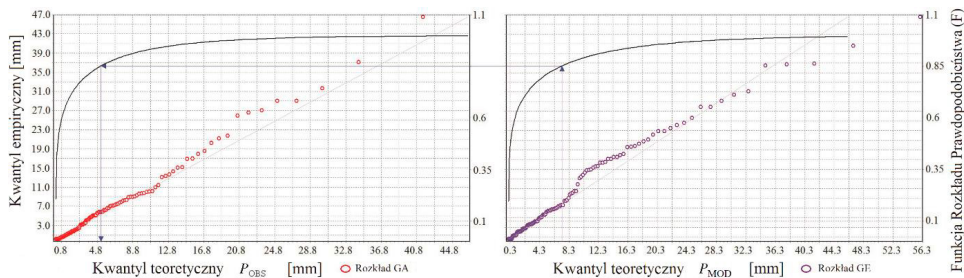
gdzie: F_{MOD} jest dystrybuantą zmiennej losowej modelowej, X_{MOD} i F_{OBS}^{-1} jest funkcją odwrotną dystrybuanty (lub funkcji kwantylowej) zmiennej losowej obserwowanej, X_{OBS} . Do rozwiązania powyższego równania zamiast założonych parametrycznych rozkładów prawdopodobieństwa wykorzystano rozkłady empiryczne zmiennych losowych zaobserwowanych i satelitarnych (modelowych) [Hashino i in. 2007; Themel i in. 2011]. Do rozwiązania równania (3.51) stosowana jest także nieparametryczna wielokrotna liniowa regresja (WLR) lub wielokrotna liniowa regresja z randomizacją (WLRR). W przypadku korygowania opadów atmosferycznych obliczenia wykonywane są dla dni z opadem, a w przypadku dni suchych dystrybuanta przyjmuje wartość zero [Themel i in. 2012].

Do usunięcia obciążenia z meteorologicznych obserwacji satelitarnych często używano metodę transformacji obciążenia za pomocą transformaty dystrybuanty (iv) [Reichle, Koster 2004; Alvarez-Garreton i in. 2014]. Transformacja statystyczna obliczana jest na podstawie równania (3.56) za pomocą teoretycznych rozkładów prawdopodobieństwa. W dotychczasowych opracowaniach do modelowania prawdopodobieństwa występowania opadów stosowano rozkład Bernoulliego (B), natomiast do modelowania intensywności opadu teoretyczny rozkład gamma (GA) [Cannon 2008]. Obliczenia wykonywano dla takich samych rozkładów zmiennych losowych X_{OBS} i X_{MOD} , np. Bernoulliego (B), Weibulla (WE), log-normalnego (LN), wykładniczego (E) lub dla różnych rozkładów zmiennych X_{OBS} i X_{MOD} , odpowiednio: Bernoulliego-Weibulla (B-WE), Bernoulliego-log-normalnego (B-LN) i Bernoulliego-Wykładniczy (B-E) [Cannon 2012]. Parametry tych rozkładów szacowano za pomocą metody największej wiarygodności (MNV) niezależnie dla obu brzegowych rozkładów prawdopodobieństwa, tj. dla X_{OBS} i X_{MOD} .

Powszechnie stosowana jest transformacja typu (GA-GA). Sharma i in. [2007] zastosował ten typ transformacji w celu zmniejszenia różnicy między wynikami modelu GCM (Global Climate Model) i opadem zaobserwowanym z naziemnej sieci pomiarowo-obserwacyjnej, wykorzystując w tym celu narzędzia GCM do usuwania błędu systematycznego z wyników modelu klimatycznego. W przypadku obserwacji i symulacji opadu mogą pojawić się zdarzenia zarówno z wysokimi opadami jak i dni bez opadu. W takich sytuacjach Piani i in. [2009] zaproponował odpowiednią funkcję przekształcania typu GA-GA w celu uzupełnienia frakcji ciągów bez opadów, a dopasowanie zostało wykonane minimalizując średni błąd kwadratowy. Zgodnie z Piani i in. [2010] wszystkie parametry transformacji były dopasowane do frakcji dystrybuanty odpowiadającej obserwacjom opadu ($X_{OBS} > 0$). Wartości modelowanego opadu odpowiadające suchej części dystrybuanty opadu zaobserwowanego były wyzerowane.

3.7.1. Korekcja obciążenia metodą transformaty dystrybuanty (BC)

W algorytmie korekcji obciążenia metodą transformaty dystrybuanty (BC) założono, że rozkład badanej zmiennej losowej jest znany i zdefiniowany za pomocą wzoru (3.56) [Piani i in. 2010]. Rysunek 3.4 ilustruje koncepcję transformacji obciążenia za pomocą teoretycznej dystrybuanty. Powyższa koncepcja polega na zastąpieniu opadu modelowanego, P_{MOD} (np. satelitarnego) wartością opadu zaobserwowanego, P_{OBS} z takim samym prawdopodobieństwem nieosiągnięcia, w tym przypadku z $p = 0,85$. Korekcja obciążenia jest realizowana poprzez obliczenie dystrybuanty F_{OBS} , a następnie powiązanie wartości P_{OBS} z każdą wartością P_{MOD} w taki sposób, że $F_{MOD}(P_{MOD}) = F_{OBS}(P_{OBS})$. W związku z tym w pierwszym kroku teoretyczne dystrybuanty F_{OBS} są dopasowywane do zaobserwowanej danej P_{OBS} (lewa strona rys. 3.4), jak również do danej modelowanej P_{MOD} (prawa strona rys.3.4). Na rys. 3.4 opad zaobserwowany z naziemnej sieci pomiarowo-obserwacyjnej P_{OBS} i modelowany P_{MOD} odnoszą się do średniego opadu obszarowego obliczonego dla podzlewni Rajcza zlewni rzeki Soły. Zgodnie z równaniem (3.56) dla danej wartości P_{MOD} obliczana jest dystrybuanta $F_{MOD}(P_{MOD})$ i korespondujący do obliczonej dystrybuanty F_{MOD} skorygowany opad modelowany P_{MOD}^{BC} , tj. $P_{MOD}^{BC} = F_{OBS}^{-1}(F_{MOD}(P_{MOD}))$.



Rys. 3.4. Ilustracja metody usuwania obciążenia za pomocą transformaty dystrybucyjnej typu GA–GE oraz wykres zależności Kwantyl teoretyczny–Kwantyl empiryczny (Q–Q) dla zaobserwowanego opadu P_{OBS} i opadu modelowanego (satelitarnego) P_{MOD} w podzlewni Rajcza w zlewni rzeki Soły. Czarna ciągła linia oznacza dystrybuantę rozkładu teoretycznego i empirycznego ze skalą na osi pomocniczej.

Fig. 3.4. Illustration of the distribution derived transformation method with transformation function type GA–GE and the Quantile theoretical–Quantile empirical (Q–Q) plot for the observed P_{OBS} and modelled (satellite) P_{MOD} at Rajcza in Sola catchment. The solid black line represents the theoretical and empirical distribution with a scale on the auxiliary axis.

Krytycznym momentem opisywanego algorytmu jest wybór optymalnego teoretycznego rozkładu brzegowego zarówno dla obserwacji historycznych, jak i modelowanych (satelitarnych). Zasada usuwania obciążenia opiera się na założeniu maksymalnego wykorzystania informacji o własnościach losowych ciągów obserwacyjnych w celu doboru wiarygodnego rozkładu prawdopodobieństwa. Powyższy efekt uzyskiwany jest poprzez wybór wiarygodnego rozkładu prawdopodobieństwa spośród możliwie licznych zbiorów potencjalnych typów funkcji rozkładów prawdopodobieństwa. Do analizy wybrano trzy trójparametrowe rozkłady, tj. gamma (GA), uogólniony rozkład wykładniczy (GE) i Weibulla (WE) [szczegóły w Załączniku: Dodatek D].

3.7.2. Zastosowanie metody do usuwania błędu systematycznego i warunki stosowalności

Zastosowane metody usuwania błędu systematycznego (obciążenia) z meteorologicznych obserwacji satelitarnych:

- 1) Usuwanie błędu systematycznego z satelitarnych obserwacji opadu metodą transformaty dystrybucyjnej w celu otrzymania zmiennej wymuszającej w modelu HBV.
- 2) Usuwanie błędu systematycznego z satelitarnych obserwacji wilgotności gleby metodą transformaty dystrybucyjnej do updatingu modelu HBV polegającego na bezpośrednim podstawieniu skorygowanych obserwacji satelitarnych pod odpowiednie zmienne stanu w modelu hydrologicznym.
- 3) Usuwanie obciążenia z zaburzonej prognozy zmiennych stanu wilgotności gleby modelu HBV w celu utworzenia nieobciążonej wiązki zmiennych stanu za pomocą filtru Kalmana w połączeniu z metodą transformaty dystrybucyjnej, (EnKF-BC, EnSRF-BC).

Warunki stosowalności metody usuwania błędu systematycznego (obciążenia) za pomocą transformaty dystrybucyjnej:

- 1) Metodę usuwania błędu systematycznego (korekcję obciążenia, BC) stosuje się do obserwacji meteorologicznych, których źródłem jest:
 - teledetekcja satelitarna;
 - teledetekcja radarowa;
 - pomiary z naziemnej sieci pomiarowo-obserwacyjnej;
 - modele ziemi integrujące meteorologiczne obserwacje wymuszające model hydrologiczny (np. opad, temperatura), modele klimatu, modele hydrologiczne, itp.
- 2) Metoda korekcji (BC) usuwa rozbieżności w wartościach średnich i zmienność w czasie w korygowanych ciągach meteorologicznych na podstawie ciągu wzorcowego (obserwowanego).

- 3) Usuwanie obciążenia z satelitarnych obserwacji meteorologicznych (tj. z opadów i wilgotności gleby) odbywa się za pomocą transformacji statystycznej obserwacji satelitarnych na podstawie obserwacji opadu z naziemnej sieci pomiarowo-obszaryjnej i ciągów wilgotności gleby wygenerowanych przez model hydrologiczny HBV.
- 4) Usuwanie obciążenia z wiązki zaburzonej prognozy zmiennych stanu wilgotności gleby modelu HBV odbywa się za pomocą transformaty dystrybucyjnej każdego zaburzonego elementu wiązki na podstawie satelitarnych obserwacji wilgotności gleby.
- 5) Warunkiem zastosowania transformaty dystrybucyjnej do usuwania obciążenia jest znajomość rozkładu prawdopodobieństwa modelowanej zmiennej, jak i obserwowanej zmiennej.
- 6) Obliczanie korekty (BC) powinien poprzedzić wybór optymalnych teoretycznych rozkładów brzegowych zarówno zmiennych obserwowanych, jak i modelowanych.
- 7) Wybór teoretycznych rozkładów prawdopodobieństwa powinien poprzedzić wyznaczenia charakterystyk rozkładu prawdopodobieństwa oraz analiza niejednorodności wejściowych ciągów meteorologicznych za pomocą nieparametrycznych testów odpowiednich dla niezależnych zmiennych losowych.
- 8) Na podstawie wyników analizy statystycznej (trendu i wahań okresowych) należy w podziale ciągów pomiarowych do obliczeń uwzględnić zmienność cykliczną (sezonową), zwłaszcza do korekty obciążenia z wilgotności gleby.
- 9) Usuwanie obciążenia z meteorologicznych obserwacji, a zwłaszcza z opadu, najkorzystniejsze jest w okresach jesienno-wiosennych. Latem wzrasta ilościowa i jakościowa różnica między opadem zaobserwowanym z naziemnej sieci pomiarowo-obszaryjnej a np. obserwacjami satelitarnymi ze względu na występującą w tym okresie niestabilność konwekcyjną atmosfery.

3.8. Metody asymilacji danych

Metody asymilacji danych (data assimilation, DA) pozwalają na dokładne modelowanie zmiennych hydrologicznych [Refsgaard 1997; Yang, Michelle 2001; Xiong, O'Connor 2002; Wöhling i in. 2006; Kiczko i in. 2015] poprzez zapewnienie właściwych warunków gruntu w modelach (np. wilgotności gleby, θ (mm), pokrywy śniegu, wodnego ekwiwalentu śniegu). Powyższe zmienne stanu wraz z innymi, np. górnym ograniczeniem zawartości zbiornika wody gruntowej, UR (mm), dolnym ograniczeniem zawartości zbiornika wody gruntowej, LR (mm), występującymi w koncepcyjnym modelu hydrologicznym typu opad-odpływ (HBV) mogą być wykorzystane w procedurach aktualizacji zmiennych stanu.

W większości hydrologicznych modeli typu opad-odpływ początkowe wartości zmiennych stanu [Spis użytych pojęć i stosowane nazewnictwo, patrz pkt. 41] szacowane są na podstawie obserwacji z naziemnych sieci pomiarowo-obszaryjnych lub obliczane na podstawie charakterystyk fizyczno-geograficznych zlewni, bilansów wodnych itp. Wilgotność gleby jest kluczową zmienną w wymianie wody i energii na powierzchni Ziemi, a jej dokładne oszacowanie w dużym stopniu wpływa na symulację przepływu podczas wezbrania. Tradycyjne metody szacowania wilgotności gleby, takie jak analiza grawimetryczna [Reynolds 1970] nie są dokładne. Biorąc pod uwagę wysoki poziom niepewności występujący w pomiarach wilgotności gleby, obserwacje satelitarne mogą być wykorzystane jako dane wejściowe do modeli hydrologicznych.

Przestrzenny i czasowy rozkład obserwacji satelitarnych może jednak znacznie różnić się od rozkładu danych uzyskanych za pomocą symulacji modeli hydrologicznych. Dlatego w większości hydrologicznych modeli symulacyjnych procedura asymilacji danych (DA) powinna być niezbędną procedurą (preprocesorem) łączącą dane satelitarne z parametrami stanu modelu. Prawidłowo skonfigurowany symulacyjny model hydrologiczny powinien asymilować dane satelitarne, zamiast bezpośrednio wykorzystywać je jako dane wejściowe do modelu deterministycznego. Powyższy efekt jest możliwy do osiągnięcia poprzez zastosowanie procedury DA w odniesieniu do zmiennych stanu modelu matematycznego. Szacunki wilgotności gleby w ramach procedury DA z wykorzystaniem liniowych filtrów [Sun i in. 2015] są wrażliwe na obserwacje i wariancję błędu modelu. Oszacowane zmienne stanu z wysokimi błędami wejściowymi mogą okazać się znacznie gorsze niż szacunki modelu pozbawionego DA.

METODA NUDGINGU

Najprostszą metodą asymilacji jest nudging (Newtonian relaxation) [Telford i in. 2008]. Interpolacja danych wykonywana na siatce modelu i dla określonego kroku czasowego obliczana jest za pomocą prostej liniowej interpolacji. Analizy są wprowadzane do modelu poprzez dodanie wielkości nie mającej interpretacji fizycznej do równań modelu. Wielkość ta jest proporcjonalna do różnicy między zmienną modelowaną, X_{MOD} a zmienną obserwowaną, X_{OBS} . Skorygowana wartość zmiennej modelowanej, X_{COR} jest obliczana wg wzoru:

$$X_{COR} = X_{MOD} + G(X_{OBS} - X_{MOD}) \quad (3.57)$$

gdzie: X_{COR} reprezentuje oszacowanie skorygowanej wartości X_{MOD} , X_{MOD} jest daną z modelu do korekcji, G jest stałą zwaną parametrem relaksacji z zalecaną wartością równą $\left(\frac{1}{\text{częstotliwość danych}}\right)$.

Metodę nudgingu zastosowali między innymi Laiolo i in. [2015], którzy przetestowali wpływ asymilacji trzech różnych obserwacji satelitarnych wilgotności gleby (Advanced SCATTERometer, ASCAT) na dokładność symulacji rozproszonego fizycznego modelu hydrologicznego w trzech małych włoskich zlewniach. Obserwacje ASCAT zostały wstępnie przetworzone, aby być bezpośrednim wejściem do preprocesora modelu hydrologicznego jako zmienne stanu, a następnie asymilowano je za pomocą prostego nudgingu.

ROZSZERZONY FILTR KALMANA (EKF)

W modelowaniu systemów, których dokładna natura nie jest znana, do szacowania zmiennych stanu deterministycznych modeli hydrologicznych wykorzystywany jest filtr Kalmana [Kalman 1960] i jego liczne modyfikacje. Procedura ta łączy własności predykcyjne modelu z rzeczywistymi danymi wejściowymi [Drécourt 2003]. Najwcześniejszą metodą asymilacji danych stosowaną w hydrologii był Extended Kalman Filter (EKF), rozszerzenie liniowego filtru Kalmana. Taka próba została po raz pierwszy podjęta w badaniu, w którym EKF został połączony z modelem SWAT (Soil and Water Assessment Tool) do asymilacji przepływu rzecznoego Q w górnym biegu rzeki Senegal w Afryce Zachodniej [Sun i in. 2015]. Okazało się jednak, że takie rozwiązanie jest niestabilne w przypadku silnych nieliniowych modeli, co sugerowało, że EKF może okazać się nieskuteczny do rozwiązywania dużych problemów asymilacji w połączeniu z rozproszonymi modelami hydrologicznymi [Evensen 1992; Reichle i in. 2002].

ENSEMBLE KALMAN FILTER (ENKF)

Powszechnie stosowaną metodą DA w hydrologii jest ensemble Kalman filter (EnKF) [Xie, Zhang 2010; Xiong, Cao 2011; Samuel i in. 2014; Rasmussen i in. 2015; Thiboult, Anctil 2015]. Używany jest do korygowania błędu losowego i uważany jest za najbardziej wydajny schemat asymilacji danych (DA), ponieważ wymaga na wejściu tylko: (i) zmiennych systemowych z poprzedniego kroku czasowego i warunków wymuszenia, oraz (ii) obserwacji z bieżącego kroku czasowego [Kalman 1960]. Filtr Kalmana zapewnia potężne wsparcie procesu szacowania przeszłych, bieżących i przyszłych zmiennych stanu modelu za pomocą układu równań, które zapewniają wydajne środowisko obliczeniowe szacujące stan procesu przy jednoczesnym minimalizowaniu średniego błędu kwadratowego [Welch, Bishop 2006]. EnKF wyprowadzony jest ze standardowego filtru Kalmana [Evensen 2003]. Generuje wiązki *ipso facto*, rozszerzając zakres próbkowania w celu uwzględnienia w modelu niepewności wynikającej z procesów nieliniowych [szczegóły w Załączniku: Dodatek D]. W przypadku filtru EnKF, kiedy zdefiniowane są wszystkie źródła niepewności przyjmuje się, że błędy modelu i niepewności podlegają rozkładowi Gaussa.

Metoda ta stosowana jest jako preprocesor danych wejściowych do modeli hydrologicznych, który uruchamiany jest z określoną liczebnością wiązki. W tym przypadku wiązka reprezentuje jednako wiarygodne zbiory zmiennych stanu modelu zlewni [Clark i in. 2008]. Błąd modelu szacowany jest bezpośrednio z wiązki jako różnica między poszczególnymi realizacjami elementów wiązki a średnią realizacją. Filtr Kalmana stosowany jest do rozwiązywania problemów liniowych. Takie sytuacje (tzn. liniowość) w hydrologii zdarzają się stosunkowo rzadko. Stąd wykorzystywany jest filtr wiązkowy. W filtrze EnKF zmienne stanu modelu są syntetycznie zaburzane. Otrzymywana w ten sposób wiązka zmiennych stanu reprezentuje niepewność modeli hydrologicznych. Obciążenie symulowanej wiązki

zmiennych stanu wynika z nieliniowego charakteru naturalnych procesów zachodzących w modelach hydrologicznych. Konsekwencją tych procesów jest obciążenie prognozowanych zmiennych stanu modelu.

MODYFIKACJE WIĄZKOWEGO FILTRU KALMANA: ENSEMBLE SQUARE FILTER (EnSRF), PARTICLE FILTER (PF) I RECURSIVE ENSEMBLE KALMAN FILTER (REnKF)

Kolejne modyfikacje filtra Kalmana stosowane w modelowaniu hydrologicznym obejmują: Ensemble Square Root Filter (EnSRF), Particle Filter (PF) i Recursive Ensemble Kalman Filter (REnKF). W przeciwieństwie do EnKF metoda EnSRF nie wymaga zaburzeń obserwacji, ale raczej aktualizuje stan modelu poprzez updating średniej wartości wiązki i odchylenia każdego elementu wiązki od średniej. Algorytm EnSRF jest szybszy i dokładniejszy niż EnKF dla takiej samej liczebności wiązki. W metodzie PF [Arulampalam i in. 2002] zmienne stanu modelu nie są aktualizowane. Każdy element wiązki skrajony jest z prawdopodobieństwem lub wagą. Zatem błąd modelu nie podlega rozkładowi Gaussowskiemu. Jednak do oszacowania tego błędu wymagana jest znacznie większa liczebność wiązki niż w przypadku metody EnKF i EnSRF. W tym sensie PF uznano za optymalną metodę DA dla numerycznego szacowania błędów w nieliniowych i niegaussowskich modelach przestrzennych. Najtrudniejszą i czasochłonną metodą DA, zwłaszcza przy zastosowaniu do asymilacji danych przepływu rzeczno- Q w trybie prognozowania operacyjnego jest metoda REnKF [McMillan i in. 2013]. Jest to metoda iteracyjna, wykorzystująca podstawowy cykl przetwarzania filtra EnKF [Clark i in. 2008]. Powtarzany jest updating prognozowanych i bieżących zmiennych stanu modelu w celu poprawienia warunków początkowych modelu, a w konsekwencji prognozy Q .

3.8.1. Asymilacja satelitarnych obserwacji wilgotności gleby przy użyciu filtrów wiązkowych z uwzględnieniem metody transformaty dystrybuanty

W niniejszym opracowaniu do asymilacji satelitarnych obserwacji wilgotności gleby zastosowano dwa filtry wiązkowe, tj. Ensemble Kalman Filter (EnKF) i Ensemble Square Root Filter (EnSRF). Wilgotność gleby θ jest ważną zmienną stanu w procesie updatingu. Prawidłowe oszacowanie wartości tej zmiennej w ramach asymilacji danych ma wpływ na wariancję błędu modelu. Obliczony przepływ przez model hydrologiczny z dużymi błędami wejściowymi może okazać się znacznie gorszy niż szacunki modelu pozbawionego procedury asymilacji. Wiązkowy filtr może zostać użyty do nieliniowych symulacji i niegaussowskich zależności występujących między zmiennymi stanu modelu i obserwacjami, jednak w analizie należy uwzględnić możliwość występowania błędu systematycznego w oszacowaniach procesów nieliniowych.

Jedną z głównych słabości EnKF jest występowanie błędów wynikających z przetwarzania zaburzonych obserwacji. Wynikowy szum dodany do zaburzonych obserwacji może być skorelowany z błędem tła (błędem prognozowanych zmiennych stanu obliczonych przez model hydrologiczny). Konsekwencją tego źródła błędu jest systematyczne niedoszacowanie analizy wariancji błędu [Whitaker, Hamill 2002]. W celu zminimalizowania tego błędu zastosowano filtr EnSRF [patrz Dodatek E z algorytmem updatingu zmiennych stanu modelu]. EnSRF asymiluje zmienne stanu modelu wykorzystując w tym celu updating średniej wartości wiązki zmiennych stanu zapisanych w macierzy tła modelu.

3.8.1.1. Błędy modelu, obserwacji oraz zaburzanie zmiennych stanu modelu hydrologicznego

Błędy modelu mogą wynikać z niepewności danych wejściowych (tzw. danych wymuszających), struktury modelu i parametrów modelu [McMillan i in. 2011; Alvarez-Garreton i in. 2015]. W rzeczywistości błędy te reprezentują nieuporządkowany szum, który jest dodawany do danych wymuszających, zmiennych stanu modelu i/lub parametrów modelu. W filtrze EnKF zmienne stanu modelu są syntetycznie zaburzane. Konsekwencją zaburzania jest obciążenie prognozowanych zmiennych stanu modelu, które należy usunąć. Zbiór zmiennych stanu, który może podlegać zaburzaniu i być uwzględniany w obliczaniu zysku Kalmana, K , zawiera: wilgotność gleby θ [mm], górne UR [mm] i dolne LR [mm] ograniczenie zawartości zbiornika wody gruntowej. Wybór opadu P do zaburzania jest uzależniony od decyzji użytkownika. W praktyce opad powinien być zaburzany tylko wtedy, kiedy model

hydrologiczny generuje odpływ wyraźnie różniący się od odpływu obserwowanego i istnieje uzasadnienie, że główną przyczyną niezgodności symulowanego hydrogramu z obserwowanym są błędy pomiarowe opadu (np. mała reprezentatywność sieci pomiarowej, niejednorodność ciągu pomiarowego, błędy wynikające z konstrukcji przyrządu pomiarowego, warunków meteorologicznych podczas opadu lub po nim, sposobu wykonywania pomiaru, itp.).

Niezaburzoną prognozę zmiennej stanu wilgotności gleby oznaczono jako θ , natomiast zaburzony stan przedstawiono jako θ' . Podobnie oznaczono niezaburzone i zaburzone zmienne stanu górnego i dolnego ograniczenia zawartości zbiornika wody gruntowej oraz zaburzoną i niezaburzoną zmienną wymuszającą opad, odpowiednio: UR, UR', LR, LR' oraz P, P' .

Powyższy problem był rozważany przez Ryu i in. [2009] oraz Alvarez-Garretona i in. [2015]. Autorzy zastosowali technikę postprocesową polegającą na równoległym uruchomieniu modelu hydrologicznego bez zaburzania i z zaburzaniem zmiennych stanu. Średnie obciążenie zaburzonych zmiennych stanu n elementowej wiązki w chwili j (δ_j), wyznaczono odejmując niezaburzoną prognozę zmiennej stanu θ_j od średniej wartości wiązki zmiennych zaburzonych [Ryu i in. 2009]:

$$\delta_j = \frac{1}{n} \sum_{i=1}^n (\theta'_{j-1,i} - \theta_j) \quad (3.58)$$

gdzie δ_j jest średnim obciążeniem zaburzonych zmiennych stanu w n elementowej wiązce w chwili j .

W celu otrzymania wiązki zmiennych stanu wolnych od obciążenia ($\tilde{\theta}_{j,i}$) (tj. skorygowanej wiązki), obciążenie zaburzenia (δ_j) odejmowano od poszczególnych elementów wiązki ($\theta'_{j-1,i}$) [Ryu i in. 2009]:

$$\tilde{\theta}_{j,i} = \theta'_{j-1,i} - \delta_j \quad (3.59)$$

Ograniczeniem tego rozwiązania jest to, że metoda nie uwzględnia rozkładu obciążenia w całej wiązce, mimo, że obciążenie może być różne dla każdego elementu wiązki zaburzonych zmiennych stanu. Rozkład obciążenia w wiązce zaburzonych stanów może wpływać na jakość filtrowania danych w ramach EnKF [Ciupak i in. 2019].

W przypadku wilgotności gleby zakres zmienności występowania zawiera się w przedziale $\langle 0\%, 100\% \rangle$. Problem pojawia się w procesie asymilacji za pomocą filtru EnKF. W momencie zaburzania wilgotności gleby z zastosowaniem rozkładu Gaussa często ograniczenie zmienności jest przekraczane, co bezpośrednio wpływa na przetwarzanie modelu hydrologicznego. Baguis i Roulin [2017] do wygenerowania wariancji szumu $\xi_\theta \sim N(0, \sigma_\theta^2)$, gdzie σ_θ^2 jest odchyleniem standardowym rozkładu normalnego, i jednocześnie do wyeliminowania możliwości wylosowania zaburzonej zmiennej poza górną (100) i dolną (0) granicę zmienności wilgotności gleby zastosowali obustronnie ucięty rozkład normalny lub log-normalny. Dla zmiennej losowej, X , o funkcji gęstości, f , i dystrybuancie, F , ograniczenie przedziałowe $(a, b]$ opisane jest funkcją rozkładu $g(f)_{(a,b]}$ według wzoru [Johnson i in. 1994]:

$$g(f)_{(a,b]}(x) = \frac{f(x)}{F(b) - F(a)} \quad (3.60)$$

gdzie

$$f(x) = \begin{cases} f(x) & a < x \leq b \\ 0 & x \leq a \cup x > b \end{cases} \quad (3.61)$$

dla $a = 0\%$ i $b = 100\%$.

W niniejszej metodyce zastosowano multiplikatywny model błędu opracowany przez McMillana i in. [2011]. Błąd obserwacji opadu został wyrażony jako $\xi_p \sim \ln N(0, \sigma_p^2)$, gdzie σ_p^2 jest odchyleniem standardowym rozkładu lognormalnego. W przypadku wilgotności gleby zastosowano uproszczony model $\xi_\theta \sim N(0, \sigma_\theta^2)$. Zgodnie z metodyką filtru EnKF, asymilacja danych jest poprzedzona analizą wrażliwości wybranych zmiennych stanu. Zaburzenia są parametryzowane jako ułamkowe parametry błędów

i wyznaczane dla: wilgotności gleby ξ_θ (mm), górnego ξ_{UR} (mm) i dolnego ξ_{LR} (mm) ograniczenia zawartości zbiornika wody gruntowej oraz opadu ξ_P (mm). Powyższe błędy ilościowo określono na podstawie analizy wrażliwości przy użyciu dopuszczalnych przedziałów: $0,10 \leq \xi_\theta \leq 0,40$; $0,05 \leq \xi_{UR} \leq 0,50$; $0,05 \leq \xi_{LR} \leq 0,50$ i $0,10 \leq \xi_P \leq 0,40$, natomiast granice przedziałów ustalono na podstawie wyników pracy McMillan i in. [2013]. Metoda EnSRF nie wymaga kwantyfikacji niepewności obserwacji [Whitaker, Hamill 2002], dlatego obserwowane zmienne stanu modelu nie są zaburzane, natomiast prognozowane zmienne stanu są zaburzane w sposób podobny do metody EnKF.

Zakłada się, że błędy podlegają rozkładowi jednostajnemu $U[-\xi_i q_i; +\xi_i q_i]$, gdzie ξ_i jest parametrem błędu q_i .

Równanie do zaburzania wilgotność gleby θ zdefiniowano jako:

$$\theta'_{j,i} \sim \theta_j + U[-\xi_\theta(\theta_j - \theta_{j-1}); +\xi_\theta(\theta_j - \theta_{j-1})] \text{ dla } \xi_\theta = \{0,10, 0,20, \dots, 0,40\} \quad (3.62)$$

gdzie: $\theta'_{j,i}$ jest zaburzoną zmienną θ_j , tj. wilgotnością gleby dla i -tego elementu wiązki, dla $j = 1, \dots, N$, gdzie N jest liczbą kroków symulacji i dla $i = 1, 2, \dots, n$, gdzie n jest liczebnością wiązki założoną w procedurze; θ_j jest wilgotnością gleby w j -tym kroku symulacji; θ_{j-1} jest wilgotnością gleby obliczoną w poprzednim ($j - 1$) kroku; U jest rozkładem jednostajnym w zakresie $\pm\xi_\theta(\theta_j - \theta_{j-1})$.

Dla górnego ograniczenia zawartości zbiornika wody gruntowej, UR (mm) i dolnego ograniczenia zawartości zbiornika wody gruntowej, LR (mm) powyższe błędy obliczane są z równania odwrotnego wyznaczanego w następujący sposób. Dla UR jest to zdefiniowane jako:

$$UR'_{j,i} \sim \exp \left[\frac{1}{1+\alpha} \ln \left(\frac{UR_{out,j} + U[-\xi_{UR}UR_{out,j}; +\xi_{UR}UR_{out,j}]}{k_h q_j} \right) \right] \text{ dla } \xi_{UR} = \{0,05, 0,10, \dots, 0,50\} \quad (3.63)$$

gdzie: $k_h q_j$ jest zmienną stanu modelu HBV [Lindström i in. 1997] użytą do obliczania współczynnika recesji górnego zbiornika wody gruntowej w j -tym kroku symulacji; α jest miarą nieliniowości generowanego odpływu, który jest funkcją odpowiedzi przekształcającą nadmiar wody ze strefy wilgotności gleby w odpływ; U jest rozkładem jednostajnym w zakresie $\pm\xi_{UR}UR_{out,j}$; $UR'_{j,i}$ jest zaburzoną zmienną stanu UR_j , tj. górnego ograniczenia zawartości zbiornika wody gruntowej w fazie odpowiedzi w j -tym kroku symulacji dla i -tego elementu wiązki, gdzie: $i = 1, 2, \dots, n$; $UR_{out,j}$ jest odpływem z górnego zbiornika wody gruntowej w fazie odpowiedzi (mm) obliczonym w j -tym kroku symulacji.

Dla LR jest to zdefiniowane jako:

$$LR'_{j,i} \sim \frac{1}{k_{a_j}} \left[LR_{out,j} + U[-\xi_{LR}LR_{out,j}; +\xi_{LR}LR_{out,j}] \right] \text{ dla } \xi_{LR} = \{0,05, 0,10, \dots, 0,50\} \quad (3.64)$$

gdzie: k_{a_j} jest współczynnikiem recesji dolnego zbiornika wody gruntowej w fazie odpowiedzi (doba⁻¹) w j -tym kroku symulacji; $LR'_{j,i}$ jest zaburzoną zmienną stanu LR_j , tj. dolnego ograniczenia zawartości zbiornika wody gruntowej w fazie odpowiedzi w j -tym kroku symulacji dla i -tego elementu wiązki, gdzie: $i = 1, 2, \dots, n$; $LR_{out,j}$ jest odpływem z dolnego zbiornika wody gruntowej w fazie odpowiedzi (mm) obliczonym w j -tym kroku symulacji; U jest rozkładem jednostajnym w zakresie $\pm\xi_{LR}LR_{out,j}$.

Zaburzony opad P zdefiniowano jako:

$$P'_j \sim P_j + U[-\xi_P P_j; +\xi_P P_j], \text{ for } \xi_P = \{0,10, 0,20, \dots, 0,40\} \quad (3.65)$$

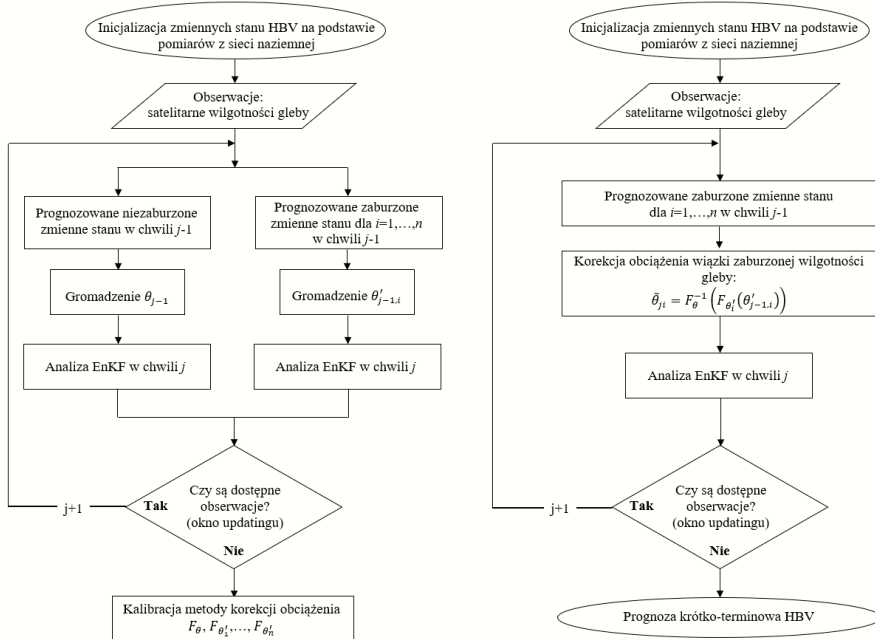
gdzie: P'_j jest zaburzonym opadem P_j w j -tym kroku symulacji; P_j jest uśrednionym powierzchniowo opadem w j -tym kroku symulacji; U jest rozkładem jednostajnym w zakresie $\pm\xi_P P_j$.

3.8.1.2. Nowa procedura wiązkowego filtru Kalmana w połączeniu ze schematem korekcji obciążenia przy użyciu transformaty dystrybuantry

W algorytmie wiązkowego filtru Kalmana użyto metody korekcji obciążenia (BC) (rozdział 3.7.1). Metoda BC, która służy do usuwania obciążenia z wiązki prognozowanych zmiennych stanu wilgotności gleby w modelu HBV jest techniką postprocesową (rys. 3.5). Procedura ta w przeciwieństwie do metody Ryu i in. [2009] bierze pod uwagę rozkład obciążenia zaburzonej n -elementowej wiązki zmiennych stanu w chwili j , a nie tylko jej średnie odchylenie, a do usuwania obciążenia wykorzystuje metodę transformaty dystrybuantry [Ciupak i in. 2019].

Procedura została podzielona na dwa etapy: dopasowanie rozkładu prawdopodobieństwa niezaburzonych i zaburzonych zmiennych stanu (lewa strona rys. 3.5) oraz korekcji obciążenia wiązki zaburzonych zmiennych stanu w algorytmie filtru EnKF-BC (prawa strona rys.3.5). Wejściem do procedury dopasowania rozkładów prawdopodobieństwa (schemat z lewej strony na rys.3.5) była próba losowa dwóch zmiennych: θ i Θ reprezentujących odpowiednio wektor niezaburzonej i macierz zaburzonej prognozy wilgotności gleby, gdzie $\theta = [\theta_j]$ i $\Theta = [\theta'_{j,i}]$, dla $j = 1, 2, \dots, N$ i $i = 1, 2, \dots, n$, gdzie N jest liczbą kroków symulacji, a n jest liczebnością wiązki. Powyższa procedura wymaga równoległej symulacji (uruchomienia modelu hydrologicznego) z i bez zaburzania zmiennych stanu modelu. W ramach dopasowania rozkładów prawdopodobieństwa w metodzie transformaty dystrybuantry [Sharma i in. 2007] wybrano najlepiej dopasowane (optymalne) teoretyczne rozkłady prawdopodobieństwa tj. F_θ i $F_{\theta'_1}, \dots, F_{\theta'_n}$ [patrz szczegóły w Dodatku D].

Transformacje statystyczne, podobnie jak każda technika statystyczna po cichu zakładają, że modelowane zależności będą zachowane w przypadku konfrontacji z nowymi danymi [Gudmundsson i in. 2012]. Powyższe założenie jest spełnione w kontekście zastosowanej w algorytmie techniki zaburzania wykorzystującej rozkład jednostajny. Transformata dystrybuantry wykazuje stosunkowo słabe wyniki w odniesieniu do skrajnych rejonów rozkładu prawdopodobieństwa [Gudmundsson i in. 2012]. Jednak teoretyczna postać równania transformacji (3.56) jest względnie łatwa do dopasowania rozkładów prawdopodobieństwa i wygodna do wdrożenia operacyjnego, co jest ważne w opisywanym zastosowaniu.



Rys. 3.5. Algorytm asymlacji danych z metodą korekcji obciążenia (EnKF-BC).
Fig. 3.5. Data assimilation algorithm with bias correction method (EnKF-BC).

Następnie dopasowane teoretyczne rozkłady prawdopodobieństwa w ramach metody BC (schemat z lewej strony na rys.3.5) zostały użyte do skorygowania wiązki prognozowanych zmiennych stanu modelu HBV θ_{ji} (schemat z prawej strony na rys.3.5) zgodnie ze wzorem:

$$\bar{\theta}_{j,i} = F_{\theta}^{-1} \left(F_{\theta'_i} \left(\theta'_{j-1,i} \right) \right) \quad (3.66)$$

gdzie: $F_{\theta}, F_{\theta'_1}, \dots, F_{\theta'_n}$ są dystrybuantami niezaburzonych i zaburzonych stanów wilgotności gleby, θ ; F_{θ}^{-1} jest funkcją odwrotną do dystrybuanty niezaburzonej zmiennej.

3.8.2. Filtr wariacyjny (VAR) i metody hybrydowe

Innym popularnym rozwiązaniem matematycznym używanym do asymilacji danych jest filtr wariacyjny (VAR) [Lee i in. 2011; Lee i in. 2015]. Jego główną ideą jest znalezienie trajektorii modelu w określonym przedziale czasowym, tak aby trajektoria przebiegała w pobliżu obserwacji tak blisko, jak to możliwe. W rzeczywistości podejście to ogranicza się do rozwiązania problemu odwrotnego do celu znalezienia warunków początkowych modelu i otrzymania trajektorii modelu uruchomionego od poszukiwanego punktu w pobliżu obserwacji w odniesieniu do przyjętej miary dokładności. Rozwiązanie problemu polega na minimalizacji przyjętej *a priori* funkcji celu. W celu udoskonalenia prognoz hydrologicznych modelu SAC-SMA (Sacramento Soil Moisture Accounting Model) w niekontrolowanej części zlewni rzeki Elk (Missouri, USA), Lee i in. [2015] zastosowali korekcję obciążenia i metodę VAR. Celem badania było zmniejszenie obciążenia w ilościowym oszacowaniu opadu pochodzącego z teledetekcji satelitarnej. Obydwie metody DA wyraźnie zredukowały obciążenie w analizowanym opadzie i zmniejszyły średni błąd kwadratowy (MSE) prognozy przepływu rzecznoego w ujściowych i górnych przekrojach wodowskazowych rzeki.

W ostatnich latach w literaturze można odnaleźć hybrydowe schematy DA łączące wiązkowe filtry Kalmana z metodami wariacyjnymi [Hamil, Snyder 2000; Zhang i in. 2009], np. EnKF/4DVAR, EnSRF/4DVAR. W tych rozwiązaniach poszukiwane jest minimum funkcji celu według przyjętej miary dokładności, którą reprezentuje wariancja określona na podstawie obserwacji zmiennych stochastycznych (filtrów EnKF i EnSRF).

3.8.2.1. Asymilacja obserwacji satelitarnych przy użyciu filtra wariacyjnego (4DVAR)

Metoda asymilacji danych za pomocą filtra wariacyjnego (4D VAR) polega na poszukiwaniu prawdopodobieństwa *a posteriori* metodą maksymalnej wiarygodności. Minimalizowana jest funkcja celu J , (równanie 3.67) zawierająca l zaburzanych obserwacji zmiennych stanu w oknie asymilacji o długości w i tło m zmiennych stanu modelu (obserwowanych i nieobserwowanych) w chwili τ . Funkcja celu filtra 4D VAR jest ogólnie zapisana jako [Zhang i in. 2009]:

$$J_{\tau}(\{\mathbf{X}_{i=0}^b\}) = \frac{1}{2} \sum_{i=1}^w \left[\mathbf{Z}_i - \mathbf{H}_i \left(\mathbf{M}_{\tau_i, \tau}(\mathbf{Z}_i^b) \right) \right]^T \mathbf{R}_i^{-1} \left[\mathbf{Z}_i - \mathbf{H}_i \left(\mathbf{M}_{\tau_i, \tau}(\mathbf{X}_i^b) \right) \right] + \frac{1}{2} (\mathbf{X}_{i=0}^b - \mathbf{X}_{i=0}^a)^T (\mathbf{P}_{i=0}^s)^{-1} (\mathbf{X}_{i=0}^b - \mathbf{X}_{i=0}^a) \quad (3.67)$$

gdzie: $\mathbf{M}_{i, \tau}$ – reprezentuje nieliniowy dynamiczny model, który oblicza $m \times w$ wymiarową macierz stanu systemu \mathbf{X}^b od momentu τ_i do τ ; \mathbf{X}_i^b – tło stanu modelu (prognozowane zmienne stanu przez model) w i -tej chwili wewnątrz okna asymilacji o długości w ; $\mathbf{X}_{i=0}^b$ – tło stanu modelu na początku okna asymilacji dla $i = 0$ (warunki inicjujące); \mathbf{X}^a – jest $m \times w$ wymiarową macierzą ze zmiennymi stanu modelu po updatingu; $\mathbf{X}_{i=0}^a$ – uaktualnione tło stanu modelu na początku okna asymilacji; \mathbf{P}^s – jest $m \times m$ wymiarową macierzą kowariancji błędu modelu skojarzoną z macierzą kowariancji błędu modelu w metodzie filtra Kalmana \mathbf{P}^b [patrz równanie (22) w Dodatku E]; \mathbf{R} – jest $l \times l$ wymiarową kowariancją błędu obserwacji *in situ*; \mathbf{Z} – jest $l \times w$ wymiarową macierzą obserwacji *in situ*; \mathbf{H} – jest $l \times w$ wymiarową macierzą przyporządkowująca stany obliczone przez model do stanów zaobserwowanych oraz

$\mathbf{Z}_i, \mathbf{H}_i, \mathbf{R}_i$ opisują zależności między obserwacjami \mathbf{Z}_i a tłem stanu modelu \mathbf{X}_i^b w i -tej chwili okna asymilacji.

W opracowaniu sformułowano problem asymilacji meteorologicznych obserwacji satelitarnych będących wejściem do częściowo-rozłożonego modelu konceptualnego typu opad-odpływ HBV za pomocą filtra wariacyjnego. Danymi wymuszającymi model HBV jest skorygowany opad satelitarny, a wejściem do procedury asymilacji są satelitarne wilgotności gleby traktowane jako potencjalne obserwacje *in situ*. W celu uproszczenia opisu metody założono, że f jest momentem, w którym obliczana jest prognoza hydrologiczna (zakończony jest updating), długość updatingu jest równy długości okna asymilacji w i obliczenia wykonywane są dla pojedynczej zlewni hydrologicznej.

RÓWNANIE FUNKCJI CELU FILTRU WARIACYJNEGO 4D VAR

Do rozwiązania problemu minimalizacji funkcji celu J_f (3.67) w chwili τ w oknie asymilacji o długości w należy wyznaczyć równanie stanu dla zmiennych stanu modelu HBV i równanie obserwacji dla wilgotności gleby θ , opadu P , ewapotranspiracji EP i przepływu rzecznoego Q . Tło zmiennych stanu modelu w chwili $f-w$, tj. na początku okna asymilacji (*a priori* lub przed updatingiem) reprezentują macierze zaobserwowanego tła stanu modelu $\mathbf{X}_{i=0,k,f-w}^b$ i uaktualnionego stanu modelu $\mathbf{X}_{i=0,k,f-w}^a$ wraz ze zmiennymi stanu modelu obliczonymi przez HBV, tj.: θ_{HBV} (mm) – wilgotność gleby, UR (mm) oraz LR (mm) – odpowiednio zawartość górnego i dolnego zbiornika wody gruntowej występującej w funkcji odpowiedzi modelu HBV. Zmienna i jest indeksem kroku wewnątrz okna asymilacji i przyjmuje wartości od $f-w+1$ do f , natomiast $k = 1, \dots, m$ jest indeksem m zmiennych stanu zmagazynowanych wewnątrz macierzy $\mathbf{X}_{i=0,k}^b$ i $\mathbf{X}_{i=0,k}^a$.

Równania (3.68-3.70) reprezentują funkcję celu J_f do zminimalizowania w chwili f :

$$J_f(\mathbf{X}_{f-w}^b, \mathbf{X}_P, \mathbf{X}_{PE}) = \left\{ \begin{array}{l} = \frac{1}{2} \times \sum_{i=f-w+1}^f [\mathbf{Z}_{Q,i} - \mathbf{H}_{Q,i}(\mathbf{X}_{f-w}^b; \mathbf{X}_P; \mathbf{X}_{PE})]^2 \times \sigma_Q^{-2} + \\ + \frac{1}{2} \times \sum_{i=f-w+1}^f [\mathbf{Z}_{\theta,i} - \mathbf{H}_{\theta,i}(\mathbf{X}_{f-w}^b)]^2 \times \sigma_{\theta}^{-2} + \\ + \frac{1}{2} \times \sum_{i=f-w+1}^f \mathbf{Z}_{P,i}^2 [1 - \mathbf{X}_P]^2 \times \sigma_P^{-2} + \\ + \frac{1}{2} \times \sum_{i=f-w+1}^f \mathbf{Z}_{PE,i}^2 [1 - \mathbf{X}_{PE}]^2 \times \sigma_{PE}^{-2} + \end{array} \right. \quad (3.68)$$

$$+ \frac{1}{2} \times \sum_{k=1}^l [\mathbf{X}_{i=0,k,f-w}^b - \mathbf{X}_{i=0,k,f-w}^a]^2 \times \mathbf{P}_{i=0,k}^{s-2} \quad (3.69)$$

Zmienne graniczne:

$$\left. \begin{array}{l} \mathbf{X}_i^b = \mathbf{M}(\mathbf{X}_{i-1}^b; \mathbf{X}_P; \mathbf{X}_{PE}); \quad i = f-w+1, \dots, f \\ \mathbf{X}_k^{\text{b,min}} \leq \mathbf{X}_{i,k}^b \leq \mathbf{X}_k^{\text{b,max}}; \quad i = f-w, \dots, f; \quad k = 1, \dots, m \end{array} \right\} \quad (3.70)$$

gdzie $J_f(\mathbf{X}_{f-w}^b, \mathbf{X}_P, \mathbf{X}_{PE})$ jest funkcją celu zależną od: \mathbf{X}_{f-w}^b – tło stanu modelu w chwili $f-w$, $\mathbf{X}_P, \mathbf{X}_{PE}$ – w wymiarowe wektory z multiplikatywnymi czynnikami korygującymi opad i ewapotranspirację, minimalizowanymi w celu odszukania takiej macierzy zmiennych stanu modelu \mathbf{X}_{f-w}^a , która minimalizuje funkcję J_f wewnątrz okna asymilacji o długości w , w kroku $i = f-w+1, \dots, f$ dla $i = i+1$.

Równanie (3.68) reprezentuje odległość symulowanych do obserwowanych zmiennych stanu w przedziale czasowym okna od chwili $f-w+1$ do f . Rozpatrywane są następujące zmienne: $\mathbf{Z}_{Q,i}$ – zaobserwowany i $\mathbf{H}_{Q,i}$ – symulowany przepływ rzeczny Q ; $\mathbf{Z}_{\theta,i}$ – satelitarne obserwacje *in situ* wilgotności gleby i $\mathbf{H}_{\theta,i}$ – obliczone przez model wartości wilgotności gleby θ . W równaniu również uwzględniono: $\mathbf{Z}_P, \mathbf{Z}_{PE}, \mathbf{X}_P, \mathbf{X}_{PE}$ – odpowiednio obserwacje i czynniki korygujące opad i ewapotranspirację oraz σ_Q – odchylenie standardowe błędu obserwacji przepływu rzecznoego; σ_{θ} – odchylenie standardowe błędu obserwacji *in-situ* satelitarnej wilgotności gleby; σ_P, σ_{PE} – odpowiednio odchylenia standardowe

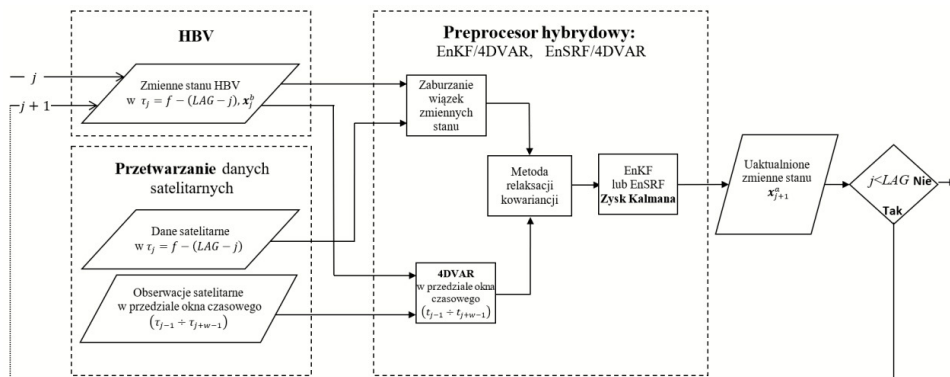
błędu pomiaru opadu i ewapotranspiracji. Równanie (3.69) pokazuje odległość tła stanu modelu w chwili inicjującej metodę $\mathbf{X}_{i=0,k,f-w}^b$ do uaktualnionego stanu modelu $\mathbf{X}_{0,k,f-w}^a$ z uwzględnieniem macierzy kowariancji błędu modelu $\mathbf{P}_{i=0}^s$ skojarzonego z tłem stanu modelu (tj. *a priori*). Obliczenia wykonywane są dla każdej k -tej zmiennej stanu z uwzględnieniem jednorodnych stref wilgotności gleby wyodrębnionych w modelu HBV. Macierz \mathbf{P}^s jest ekwiwalentem macierzy \mathbf{P}^b występującej w wiązkowych filtrach EnKF i EnSRF. W równaniu (3.70) $\mathbf{X}_k^{b,\min}$ i $\mathbf{X}_k^{b,\max}$ – oznaczają m wymiarowe wektory zawierające dolne i górne ograniczenie zmienności k -tej zmiennej stanu modelu.

3.8.2.2. Asymilacja obserwacji satelitarnych metodami hybrydowymi

W pracy utworzono algorytm łączący deterministyczną 4D wariacyjną metodę asymilacji z dwoma filtrami wiązkowymi, tj. Ensemble Kalman Filter (EnKF) i Ensemble Square Root Filter (EnSRF). Wynikiem tego połączenia są dwie metody hybrydowe: EnKF/4DVAR i EnSRF/4DVAR (rys. 3.6). Rozwiązanie hybrydowe korzysta z niepewności prognozowanych zmiennych stanu modelu dostarczanych przez metody wiązkowe przy jednoczesnym wykorzystaniu filtru 4D VAR do minimalizowania rozbieżności filtru wynikowego. Analiza 4D VAR oblicza prawdopodobieństwo *a posteriori* metodą największej wiarygodności poprzez minimalizację funkcji celu (3.67). W ramach tej procedury zaburzone wiązki prognozowanych zmiennych stanu modelu hydrologicznego mogą być propagowane do następnego cyklu asymilacji, tj. dopóki $j < LAG$, gdzie j jest kolejnym krokiem przetwarzania, a LAG długością updatingu (rys. 3.6), lub stanowić podstawowy zbiór zmiennych stanu modelu wykorzystany do generowania finalnego prognozowanego hydrogramu wiązkowego (przedziałowego), tj. gdy $j = LAG$.

W opracowaniu założono, że:

- i) Metody hybrydowe EnKF/4DVAR i EnSRF/4DVAR są zdolne do osiągnięcia lepszych wyników, tj. efektywniejszych zmiennych stanu modelu i w konsekwencji do dokładniejszego prognozowania hydrologicznego w porównaniu do asymilacji metodami EnKF i EnSRF.
- ii) Metoda hybrydowa w mniejszym stopniu uzależniona jest od liczebność wiązki (liczebność wiązki ma bezpośredni wpływ na czas przetwarzania w metodzie EnKF i EnSRF) oraz od długości okna asymilacji.



Rys. 3.6. Schemat modelu hybrydowego EnKF/4DVAR lub EnSRF/4DVAR.

Fig. 3.6. Schematic of the hybrid model EnKF/4DVAR or EnSRF/4DVAR.

W modelach hybrydowych: EnKF/4DVAR i EnSRF/4DVAR statyczna macierz kowariancji błędu modelu \mathbf{P}^s (3.67) występująca w standardowej metodzie 4DVAR zostaje zastąpiona macierzą kowariancji błędów \mathbf{P}^b [równanie (22), Załącznik: Dodatek E] występująca w metodzie KF z uwzględnieniem zależności [Hamil, Snyder 2000; Zhang i in. 2009]:

$$\mathbf{P}^b = \beta \mathbf{P}^b + (1 - \beta) - \mathbf{P}^s \quad (3.71)$$

gdzie: współczynnik mieszania β jest wagą przypisaną do szacowania kowariancji błędów dla każdego elementu wiązki. Dla $\beta = 0$ model hybrydowy EnKF/4DVAR lub EnSRF/4DVAR jest tożsamy z modelem 4DVAR, natomiast dla $\beta = 1$ model hybrydowy przybiera postać modelu EnKF lub EnSRF. Algorytm modelu hybrydowego jest połączeniem algorytmu filtru wiązkowego [szczegóły w Dodatku D lub E] z algorytmem filtru za pomocą metody relaksacji kowariancji (3.71).

3.8.3. Zastosowanie metod asymilacji satelitarnej obserwacji wilgotności gleby i warunki stosowalności

Zastosowane metody asymilacji satelitarnej wilgotności gleby do poprawiania symulacji i prognozowania hydrologicznego modelu typu opad-odpływ HBV:

- i) Wiązkowy filtr Kalmana – Ensemble Kalman Filter (EnKF).
- ii) Wiązkowy filtr Kalmana – Ensemble Kalman Filter z metodą usuwania obciążenia z wiązki zaburzonych prognoz zmiennych stanu wilgotność gleby (EnKF-BC).
- iii) Wiązkowy filtr – Ensemble Square Root Filter (EnSRF).
- iv) Wiązkowy filtr – Ensemble Square Root Filter z metodą usuwania obciążenia z wiązki zaburzonych prognoz zmiennych stanu wilgotności gleby (EnSRF-BC).
- v) Metoda hybrydowa wiązkowego filtru Kalmana z metodą wariacyjną (EnKF/4DVAR).
- vi) Metoda hybrydowa wiązkowego filtru Square Root Filter z metodą wariacyjną (EnSRF/4DVAR).

Warunki stosowalności metod asymilacji satelitarnej wilgotności gleby:

- i) Metody asymilacji danych (DA) stosuje się w procedurach updatingu zmiennych stanu modelu hydrologicznego.
- ii) Do updatingu hydrologicznych modeli typu opad-odpływ za pomocą procedury DA stosuje się najczęściej satelitarne obserwacje warunków gruntu, takich jak:
 - wilgotność gleby;
 - pokrywa śniegu;
 - wodny ekwiwalent śniegu.
- iii) Procedury asymilacji danych usuwają rozbieżność między przestrzennym i czasowym rozkładem obserwacji satelitarnych a obserwacjami z naziemnej sieci pomiarowo-obszaryjnej lub zmiennymi generowanymi przez modele hydrologiczne.
- iv) Wejściem do procedury asymilacji danych (DA) są syntetycznie zaburzane prognozowane zmienne stanu wilgotności gleby generowane przez model hydrologiczny (np. HBV) oraz satelitarne obserwacje wilgotności gleby traktowane jako potencjale obserwacje *in situ*.
- v) Przygotowanie danych wejściowych do procedury DA powinna poprzedzać analiza statystyczna ciągów obserwacji satelitarnych i wilgotności gleby wygenerowanych przez model hydrologiczny.
- vi) Analiza statystyczna powinna zawierać:
 - badanie jednorodności statystycznej danych wejściowych do korekcji obciążenia i metod asymilacji;
 - przekształcenie ciągów pomiarowych w celu usunięcia zależności i stacjonarności danych wejściowych do metody korekcji obciążenia, (np. grupowanie obserwacji opadu bez uwzględniania zdarzeń bezopadowych lub grupowanie obserwacji wilgotności gleby wg sezonu zimowego i letniego);
 - wyznaczenie charakterystyk rozkładu prawdopodobieństwa;
 - wybór najlepiej dopasowanej funkcji rozkładu prawdopodobieństwa do metody korekcji obciążenia.
- vii) Wynik analizy statystycznej jest podstawą do wyodrębnienia kalibracyjnych i walidacyjnych¹ ciągów obserwacyjnych i pomiarowych z uwzględnieniem wahań okresowych (np. sezonu zimowego i letniego).
- viii) Analiza niejednorodności meteorologicznych obserwacji satelitarnych i wilgotności gleby wygenerowanej przez model hydrologiczny do metody korekcji obciążenia powinna być wykonana za pomocą nieparametrycznych testów odpowiednich dla niezależnych zmiennych losowych.

¹ Spis użytych pojęć i stosowane nazewnictwo, pkt 42.

- ix) Procedury DA wymagają, żeby analiza danych i krok czasowy został przeniesiony do siatki modelu hydrologicznego.
- x) Interpolacja obserwacji satelitarnej wilgotności gleby wykonywana jest na siatce modelu hydrologicznego. W modelu HBV wykorzystano schemat skupiony (lumped scheme) do utworzenia pojedynczej zlewni z jedną strefą wegetacyjną.
- xi) Dane satelitarne z regularnej siatki należy przestrzennie interpolować do centroidu modelowanej zlewni, natomiast asymilacja obliczana jest tylko w kroku modelu, w którym obserwowana jest obserwacja satelitarna.
- xii) Ujemne temperatury powietrza i występowanie stałej pokrywy śniegu ma negatywny wpływ na asymilację wilgotności gleby, natomiast najbardziej korzystna sytuacja do asymilacji występuje w sezonie letnim.

3.9. Klasyfikacja modeli

Zastosowane w niniejszej pracy metody i modele matematyczne sklasyfikowano według [Ozga-Zielińska, Brzeziński 1997]. Podana klasyfikacja nie jest rozłączna, stąd wymienione rozwiązania mogą spełniać kilka kryteriów klasyfikacyjnych.

I. STRUKTURA MODELU I PRZEDMIOT MODELOWANIA

Opracowane modele dotyczą symulacji operatorów pojedynczych procesów:

- i) wybranych charakterystyk ilościowych wezbrania deszczowo-roztopowego, tj. przepływu kulminacyjnego wezbrania Q_{\max} (m^3s^{-1}), objętości V_w (10^6 m^3) i czasu trwania T_w (doby) – modele SLMB;
- ii) przepływu rzeczego w profilu wodowskazowym – modele DLMB oraz model HBV z procedurą DA.

II. STOSOWANE METODY BADAWCZE

Do przygotowania modeli matematycznych zastosowano statystyczne metody badawcze. Procedura modelowania składała się z trzech etapów:

- i) zgromadzenia długiego ciągu pomiarowego;
- ii) analizy danych wejściowych do modelu (np. analizy niejednorodności, wyznaczenia charakterystyk rozkładu prawdopodobieństwa, identyfikacji trendu, analizy autokorelacji, itp.);
- iii) syntezy, czyli opracowania modelu.

III. DOSTĘPNOŚĆ INFORMACJI O OBIEKCIE

Wszystkie opisane modele i metody należą do kategorii modeli tzw. czarnej skrzynki (black box)², którą charakteryzuje brak informacji o wewnętrznej budowie modelowanego obiektu. Ze względu na to, że opisywane metody (modele) przeznaczone są do celów prognozy hydrologicznej, zmienne występujące w zależnościach funkcyjnych nazywano zmiennymi prognozowanymi i prognozującymi.

IV. WARTOŚCI POZNAWCZE MODELU

Zastosowano modele przyczynowo-opisowe (SLMB i DLMB), których podstawą są prawa przyczynowo-skutkowe. W każdym modelu zbiór zmiennych prognozujących zapewnił merytoryczną poprawność wnioskowania i określenia wyjścia (skutku) z modelu, czyli wartości zmiennej/zmiennych prognozowanych.

V. ROLA CZYNNIKA CZASU

W zależności od roli czynnika czasu użyto modeli statycznych lub dynamicznych:

- i) W modelach statycznych (SLMB) opisano wezbrania deszczowo-roztopowe i powiązania z nimi zachodzące w pewnym momencie lub okresie przyjętym za jednostkę czasu. Opisano stan zjawiska, czyli relację między zmienną prognozowaną a zmiennymi prognozującymi dla pewnego przedziału czasu, lecz nie opisano zmiany stanu w czasie. Rozpatrywano natychmiastową reakcję zmiennej prognozowanej na zmianę zmiennej/zmiennych prognozujących. W modelu statycznym

² Spis użytych pojęć i stosowane nazewnictwo, pkt 43.

poszukiwano wartości parametrów strukturalnych występujących w równaniach opisujących zachowanie rzeczywistych wartości zmiennej/zmiennych prognozowanych po to, aby poznać relację między zmienną/zmiennymi prognozowanymi a zmienną/zmiennymi prognozującymi. Model statyczny został użyty do modelowania wezbrania deszczowo-roztopowego w zlewni Sokółdy w profilu wodowskazowym Sokółda i do opracowania prognozy średnioterminowej trzech ilościowych charakterystyk wezbrania, tj.: przepływu kulminacyjnego wezbrania Q_{\max} (m^3s^{-1}), objętości V_w (10^6 m^3) i czasu trwania T_w (doby).

- ii) W modelach dynamicznych (DLMB) opisano zjawiska i zachodzące między nimi powiązania w funkcji czasu. Stan modelowanego obiektu, jego wejścia i wyjścia są funkcjami czasu (przebieg przepływu opracowany na podstawie funkcji transformującej). Modele dynamiczne zastosowano do symulacji rocznego hydrogramu w zlewni Górnej Narwi w profilu wodowskazowym Wizna oraz do opracowania prognozy krótkoterminowej przepływu rzecznoego Q_t^{WIZ} (m^3s^{-1}).

VI. WŁASNOŚCI FUNKCJI PRZEKSZTAŁCAJĄCEJ

Funkcje przekształcające we wszystkich zastosowanych rozwiązaniach są liniowe w sensie działania opisywanych przez nie obiektów. W modelu liniowym obowiązuje zasada superpozycji $L[\sum_{i=1}^n a_i x_i(\tau)] = \sum_{i=1}^n L[a_i x_i(\tau)]$, gdzie: L jest operatorem funkcji $a = \text{const}$, $x_i(\tau)$ jest i -tą zmienną stanu w chwili τ oraz zachodzi możliwość wyjęcia przed nawias wartości stałej (współczynnika a) $L[ax(\tau)] = aL[x(\tau)]$. Z reguły obiekty hydrologiczne (np. odpowiedzi małych zlewni hydrologicznych) są nieliniowe. W takich sytuacjach zastosowano:

- 1) Przybliżenie obiektu modelem liniowym dla pewnych zakresów wejścia – modelowano jednorodne zdarzenia wezbrań deszczowo-roztopowych, np. w modelach SLMB.
- 2) Zmianę metody estymacji parametrów funkcji przekształcającej – w przypadku SLMB zastosowano dwie metody estymacji parametrów liniowej regresji: Bayesowską i technikę odwrotnego modelowania.
- 3) Zastosowanie metod wiązkowych i metod usuwających obciążenie z wiązki zaburzonych zmiennych stanu – użyto wiązkowego filtra Kalmana w połączeniu z metodą transformaty dystrybucyjnej.

Zastosowanie modeli liniowych wymusiło podział całego zaobserwowanego opadu na powierzchnię zlewni na jednostkowe opady, np. dobowe, godzinne, i transponowanie ich na jednostkowe odpływy, a następnie sumowanie odpływu.

VII. JEDNOZNACZNOŚĆ UZYSKIWANYCH WYNIKÓW

Opracowane rozwiązania należą do kategorii modeli probabilistycznych. Nie opisują zjawiska jednoznacznie. W każdym przypadku uzyskane wyniki przedstawiane są w przedziałach niepewności szacowanych zmiennych, tzw. prognozy przedziałowe.

3.10. Ocena modeli

W ocenie jakości modeli wykorzystano analizę reszt [Kala 2002]. Zastosowano podstawowe miary rozbieżności między obserwacją a wielkością modelowaną – maksymalny błąd absolutny $\text{MAD} = \max_{i=1, \dots, n} |e_i|$, gdzie $e_i = y_{o_i} - y_{m_i}$, y_{o_i} , y_{m_i} są i -tymi obserwowanymi i modelowanymi zmiennymi, a n jest liczbą zmiennych modelowanych (maximum absolute deviation) i średni błąd absolutny $\text{MAE} = \overline{|e|}$ (mean absolute deviation) [Węglarczyk 1998].

Do oceny jakości modeli zastosowano także:

- 1) Stosunek wartości średnich, R_0M (ratio of the means), pokazujący przeszacowanie – $R_0M > 1,0$, niedoszacowanie – $R_0M < 1,0$ lub perfekcję – $R_0M = 1,0$ modelu [Friedrich i in. 2008]:

$$R_0M = \frac{\frac{1}{n} \sum_{i=1}^n y_{m_i}}{\frac{1}{n} \sum_{i=1}^n y_{o_i}} = \frac{\overline{y_m}}{\overline{y_o}} \quad (3.72)$$

gdzie: n jest liczbą zmiennych modelowanych; y_{oi} jest i -tą wartością obserwowaną, \bar{y}_o jest średnią wartością obserwowaną, y_{mi} jest i -tą wartością modelowaną; \bar{y}_m jest średnią wartością modelowaną.

- 2) Pierwiastek kwadratowy średniego błędu kwadratowego, RMSE (root mean square error), zależy od skali dokładności pomiaru i jest stosowany do oceny zdolności różnych modeli do prognozowania pojedynczej zmiennej [Chai, Draxel 2014]. RMSE wyrażony jest w mianie danej wielkości, na przykład w milimetrach (mm) dla opadu atmosferycznego, m^3s^{-1} dla przepływu rzecznoego i $^{\circ}C$ dla temperatury:

$$RMSE = \left(\frac{\sum_{i=1}^n (e_i)^2}{1} \right)^{\frac{1}{2}} \quad (3.73)$$

gdzie $e_i = y_{oi} - y_{mi}$ jest i -tą resztą, czyli różnicą między i -tą zmienną obserwowaną a i -tą zmienną modelowaną.

- 3) Współczynnik determinacji R^2 (coefficient of determination) [Colin, Windmeijer 1997]:

$$R^2 = \frac{\sum_{i=1}^n (y_{mi} - \bar{y}_o)^2}{\sum_{i=1}^n (y_{oi} - \bar{y}_o)^2} \quad (3.74)$$

- 4) Współczynnik skuteczności Nasha i Sutcliffea [Nash, Sutcliffe 1970], gdzie $-\infty < E_{NS} \leq 1,0$ z $E_{NS} = 1,0$ reprezentujący perfekcyjne dopasowanie między obserwacją a wartością modelowaną i $E_{NS} \leq 0$ reprezentujący wartość modelowaną nie lepszą niż średnia z obserwacji. Wartości $E_{NS} \geq 0,5$ są uznawane za zadawalające wyniki modelu.

$$E_{NS} = 1 - \frac{\sum_{i=1}^n (y_{oi} - y_{mi})^2}{\sum_{i=1}^n (y_{oi} - \bar{y}_o)^2} \quad (3.75)$$

W modelowaniu wielowymiarowym interesuje nas stopień rozbieżności między i -tą zmienną prognozowaną a odpowiednią zmienną zaobserwowaną. Głównym źródłem rozbieżności między modelem a rzeczywistym procesem, np. wezbraniem roztopowym, jest istnienie parametrów koncepcyjnych, tj. współczynników występujących w każdym równaniu regresji otrzymanych w procesie kalibracji. W modelu wielowymiarowej wielokrotnej regresji zastosowano miary oparte na funkcjach kryterialnych – funkcji celu. W tym badaniu zastosowano kryterium w postaci wektora $\bar{K} \in (\delta_{Q_{max}}, \delta_{V_w}, \delta_{T_w})$, gdzie $\delta_{Q_{max}}$, δ_{V_w} i δ_{T_w} oznaczają sumy błędów odpowiednio trzech prognozowanych ilościowych charakterystyk wezbrania, tj. przepływu kulminacyjnego, objętości i czasu trwania wezbrania. Dla każdej zmiennej prognozowanej j obliczana jest suma błędów jako [Ciupak 2009; Ciupak i in. 2009]:

$$\delta_j = \sum_{i=0}^n \frac{y_{0,i} - y_{m,i}}{y_{0,i}} \quad (3.76)$$

gdzie $y_{0,i}$, $y_{m,i}$ są obserwowanymi i modelowanymi zmiennymi dla każdej zmiennej prognozowanej j .

Wektor \bar{K} opisany jest w 3D przestrzeni akceptacji. Błąd modelu wielowymiarowej zmiennej prognozowanej jest objętością ω przestrzeni $(\delta_{Q_{max}}, \delta_{V_w}, \delta_{T_w})$ i obliczany jako iloczyn sumy błędów wszystkich zmiennych prognozowanych [Ozga-Zielińska, Nawalany 1979; Ciupak 2009]:

$$\omega = \prod_{j=1}^l \delta_j \quad (3.77)$$

gdzie δ_j jest sumą błędów obliczoną dla każdej zmiennej prognozowanej j , gdzie $j = 1, \dots, l$, a l jest liczbą zmiennych prognozowanych.

4. PROGNOZA WEZBRANIA DESZCZOWO-ROZTOPOWEGO ZA POMOCĄ STATYCZNYCH LINIOWYCH MODELI BAYESOWSKICH (SLMB)

Jednym z celów tego rozdziału jest przedstawienie możliwości wykorzystania statycznych liniowych modeli Bayesowskich w odniesieniu do prognozy hydrologicznej, czyli numeryczne przetwarzanie informacji niepewnej, niepełnej i nieprecyzyjnej. Do estymacji parametrów wielowymiarowych zależności regresyjnych użyto metod i technik Bayesowskich. Modele Bayesowskie zastosowano do modelowania sytuacji występującej w rzece, zdefiniowanej jako wezbranie. Interpretacja stanu hydrologicznego rzeki za pomocą wezbrania wymaga jednoznacznego określenia okresu, w którym rzeką płynie tak duża objętość wody, że okres ten należy uznać za wezbranie. Metodę opracowania danych wejściowych do modelu oraz procedury analizy statystycznej, weryfikujące własności losowe badanych ciągów i wymagane przez modele Bayesowskie, opisano szczegółowo w rozdziale (2.3.2 i 2.3.3).

Zastosowane w niniejszym rozdziale modele SLMB składają się z wielokrotnej regresji liniowej (WLR) lub wielowymiarowej wielokrotnej regresji liniowej (WWLR), w którym to modelu zmienną prognozowaną opisano nie tylko jako skalar, ale również jako wektor. Metody Bayesowskie (B) i techniki odwrotnego modelowania (InvB) posłużyły do oszacowania parametrów regresji. Metody B i InvB zastosowano w statycznych modelach (WLR-B, WWLR-B) i (WLR-InvB, WWLR-InvB). Rozwiązanie posłużyło do dokładnego prognozowania wezbrania deszczowo-roztopowego, tj. przepływu kulminacyjnego Q_{\max} (m^3s^{-1}), objętości V_w (10^6m^3) i czasu trwania T_w (doby) w zlewni Sokołdy.

Struktura niniejszego rozdziału przedstawia się następująco:

- 1) Metodyka opracowania SLMB.
- 2) Zastosowanie SLMB do prognozowania wezbrania deszczowo-roztopowego.
- 3) Omówienie wyników.
- 4) Podsumowanie zalet i ograniczeń ujawnionych w trakcie zastosowania zaproponowanych metod.

4.1. Metodyka opracowania statycznych liniowych modeli Bayesowskich

Strukturalnie SLMB składa się z modelu WLR lub WWLR z parametrami szacowanymi metodą B lub InvB. Sekwencje przetwarzania w ramach zaprojektowanego SLMB przedstawiono na schemacie blokowym (rys. 4.1).

Modele wezbrania deszczowo-roztopowego przygotowano w następujący sposób:

- 1) Opracowanie danych wejściowych do modeli SLMB:
 - identyfikacja wezbrań deszczowo-roztopowych i wyznaczenie charakterystyk ilościowych [Ciupak 2004];
 - analiza statystyczna danych wejściowych do modeli SLMB:
 - badanie jednorodności statystycznej zmiennych prognozowanych i prognozujących [Ozga-Zieliński 1999; Zasady... 2001; Guidelines... 2005],
 - wyznaczenie charakterystyk rozkładu prawdopodobieństwa [Węglarczyk 2010],
 - sprawdzenie normalności zmiennych losowych [Strupczewski 1967].
- 2) Wybór zmiennych prognozowanych i prognozujących, statystyczne metody wyboru optymalnego zbioru zmiennych prognozujących do WLR i WWLR [Darlington 1990; Kaczmarek 1977].
- 3) Kalibracja modeli WLR i WWLR w okresie (1966-1990), estymacja parametrów metodami B i InvB [Carpenter 2003; Walsh 2004; Ciupak 2009].
- 4) Walidacja modeli WLR i WWLR z parametrami estymowanymi metodą B i InvB w okresie (1991-2002), oraz ocena prognozowanych charakterystyk ilościowych wezbrań deszczowo-roztopowych za pomocą miar ilościowych i graficznych [Adamowski i in. 2012].

Szczegóły procedur zastosowanych do przygotowania wejścia do modeli SLMB w ramach procesora danych wejściowych opisano w rozdziale (2.3.2), natomiast algorytmy do kalibracji modeli WLR i WWLR, estymacji parametrów metodami B i InvB i oceny prognozowanych hydrogramów w rozdziałach (3.2 i 3.3).

INPUT	Długie ciągi pomiarowe przepływów i dane meteorologiczne	
BUDOWANIE MODELI	identyfikacja wezbrań deszczowo-roztopowych	
	wyznaczenie ilościowych charakterystyk wezbrań	
	Analiza potencjalnych zmiennych prognozujących i prognozowanych	
	analiza niejednorodności	
	optymalizacja zbioru zmiennych prognozujących dla modeli 1D i 3D	
KALIBRACJA	WLR	Statystyczna synteza SLMB
	WWLR	
	$y_k = f(y_k x_1, x_2, \dots, x_d)$ for $k=1, \dots, l$	
	$y = f(x) = E(y x) = E(y_1, y_2, \dots, y_l x_1, x_2, \dots, x_d)$	
	Estymacja parametrów metodami: B i InvB	
WLR-B		WWLR-B
WLR-InvB		WWLR-InvB
Zapisanie parametrów do bazy danych		
WLR-B WLR-InvB		WWLR-B WWLR-InvB
WALIDACJA	Walidacja modeli na danych historycznych, niezależność danych pomiarowych	
	WLR-B	WWLR-B
	WLR-InvB	WWLR-InvB
OUTPUT	Prognozowane wartości modelowanych charakterystyk i ocena rozwiązania	
	$Q_{max}; K_w; T_w$ jako skalar	$\langle Q_{max}; K_w; T_w \rangle$ jako wektor

Rys. 4.1. Schemat obliczeniowy SLMB.

Fig. 4.1. Scheme calculation of the SLMB.

WYBÓR ZMIENNYCH PROGNOZOWANYCH I PROGNOZUJĄCYCH, STATYSTYCZNE METODY WYBORU OPTIMALNEGO ZBIORU ZMIENNYCH PROGNOZUJĄCYCH DO WLR I WWLR

Do wyboru optymalnego zbioru zmiennych prognozujących zastosowano trzyetapową procedurę opisaną przez Ozga-Zielińską i Brzezińskiego [1997]. Wybór uwzględnia konieczność spełnienia dwóch warunków formalnych:

- 1) Liczba zmiennych prognozujących musi być znacznie mniejsza niż liczebność próby losowej, na podstawie której ustalana jest zależność regresyjna.
- 2) Zmienne prognozujące powinny być wzajemnie niezależne.

Na podstawie zbioru jednorodnych ciągów pomiarowych ustalono optymalny zbiór zmiennych prognozowanych i prognozujących. Wybór zmiennych prognozujących, spośród zmiennych słabo wzajemnie skorelowanych, wykonano metodą krokową postępującą (forward stepwise regression) ze skalarną i wektorową zmienną prognozowaną, odpowiednio dla modeli WLR i WWLR [Darlington 1990]. Otrzymane zbiory zmiennych porównano z wynikami procedury, w której wybierane są zmienne prognozujące, dla których średnia wartość średniego błędu prognozy U_0 [szczegóły algorytmu w Załączniku: Dodatek F] osiąga wartość minimalną [Kaczmarek 1977; Sollis 2012]. Algorytm jest podobny do statystycznej metody *a priori*, w której analizowane są cząstkowe współczynniki korelacji. Na każdym etapie procedury wyboru generowana jest zmienna prognozująca, dla której wartość cząstkowa współczynnika korelacji jest najwyższa oraz wyznaczany jest współczynnik korelacji wielokrotnej R_0 [Ciupak 2009]. Kalibracja równań prognostycznych wykorzystuje wielowymiarowe warunkowe rozkłady praw-

dopodobieństwa w postaci modeli WLR i WWLR. W celu zidentyfikowania różnic między empirycznym a założonym w procedurze teoretycznym rozkładem normalnym, przeprowadzono analizę własności losowych rozkładów prawdopodobieństwa [szczegóły procedury w rozdziale 2.3.3].

KALIBRACJA MODELI WLR I WWLR

Do estymacji parametrów rozkładu *a posteriori* wielokrotnej liniowej regresji za pomocą metody Bayesowskiej (WLR-B) użyto równania (3.18), natomiast estymację nieznanych parametrów techniką odwrotnego modelowania (WLR-InvB) wykonano, minimalizując ujemny logarytm funkcji wiarygodności zgodnie ze wzorem (3.26). W przypadku wielowymiarowego modelu wielokrotnej regresji, estymacje parametrów modelu (WWLR-B) wykonano za pomocą wzoru (3.37), a modelu (WWLR-InvB) wg wzoru (3.39).

4.2. Przykład zastosowania statycznych liniowych modeli Bayesowskich do prognozowania wezbrania deszczowo-roztopowego

W tym rozdziale opracowano dwa statyczne liniowe modele Bayesowskie (SLMB) do średnioterminowej prognozy wezbrania deszczowo-roztopowego opisanego trzema zmiennymi prognozowanymi, tj. przepływem kulminacyjnym wezbrania Q_{\max} (m^3s^{-1}), objętością V_w (10^6 m^3) i czasem trwania T_w (doby) w zlewni Sokołdy w profilu wodowskazowym Sokołda. Do modelowania wezbrania zastosowano:

- 1) Wielokrotną liniową regresję z parametrami estymowanymi metodą Bayesowską (WLR-B) i techniką odwrotnego modelowania (WLR-InvB).
- 2) 3D wielokrotną liniową regresję z parametrami estymowanymi metodą Bayesowską (WWLR-B) i techniką odwrotnego modelowania (WWLR-InvB).

Średnioterminowe prognozy hydrologiczne wg Lambora [1962] odzwierciedlają dane wejściowe do modelu (tj. dominującą sytuację hydro-meteorologiczną zlewni) w momencie stawiania prognozy. Zakłada się, że wejście do modelu jest częściowo znane, a sytuacja determinująca prognozowane zjawisko będzie jeszcze trwać przez jakiś czas po postawieniu prognozy. W proponowanym rozwiązaniu nie określono czasu wyprzedzenia prognozy. Przedział czasu między momentem stawiania prognozy a początkiem okresu prognostycznego zależy od parametrów hydrologicznych i meteorologicznych właściwych dla badanej zlewni oraz dynamiki topnienia pokrywy śniegu specyficznej dla konkretnego wezbrania. Ponieważ prognozowane zjawisko hydrologiczne wiąże się bezpośrednio lub pośrednio z prognozami hydrologicznymi i meteorologicznymi, dane wejściowe do modelu mogą składać się z dwóch macierzy zmiennych prognozujących:

- i) Macierzy zmiennych zmierzonych i dostępnych przed stawianiem prognozy \mathbf{X}_{OBS} w chwili τ (np. topnienie pokrywy śniegu i opady deszczu przyczyniające się do roztopów).
- ii) Macierzy zmiennych wpływających na przebieg wezbrania \mathbf{X}_{PROG} w chwili τ (np. prognozowane opady i średnia temperatura powietrza pozyskane z prognostycznych modeli meteorologicznych lub od synoptyka meteorologa). W niniejszym badaniu wejściem do modelu, zamiast meteorologicznych danych prognozowanych, były zaobserwowane dane historyczne. Przyczyną był brak prognoz hydro-meteorologicznych dla zlewni Sokołdy w okresie kalibracji i walidacji modeli, tj. w latach 1966-2002.

4.2.1. Wyniki i dyskusja

Na podstawie szczegółowej analizy danych wejściowych do modeli SLMB (szczegóły w rozdziale 2.3.2) wyodrębniono $n = 38$ wezbrań deszczowo-roztopowych w zlewni rzeki Sokołdy w okresie 1966-2002 oraz zweryfikowano ich jednorodność. Wszystkie zmienne zostały uznane za niezależne i stacjonarne (szczegóły w rozdziale 2.3.3). Wejściem do modeli SLMB były zmienne znormalizowane oznaczone jako $\ln(Q_{\max})$, $\ln(V_w)$, $\ln(T_w)$, $\ln(ES_5)$, $\ln(Q_1)$ oraz nieznormalizowane GT_1 , $AT_{\text{śr}}e$ i AT_5 .

W tabeli 4.1 przedstawiono współczynniki korelacji, współczynniki korelacji wielokrotnej R_0 oraz średnie wartości średniego błędu prognozy U_0 wyznaczone dla optymalnego zbioru zmiennych prognozujących oddzielnie dla modelu WLR i WWLR. Wyznaczono trzy istotne liniowe modele dla

wielokrotnej regresji: $\langle \ln(Q_{\max}); \ln(ES_5); \ln(Q_1); GT_1 \rangle$, $\langle \ln(V_w); \ln(ES_5); GT_1; AT_{5re}; AT_5 \rangle$, i $\langle \ln(T_w); \ln(ES_5); \ln(Q_1); AT_{5re} \rangle$. Jednoczesną relację z trzema zmiennymi prognozowanymi uzależniono od wektora zmiennych prognozujących $\langle \ln(Q_{\max}); \ln(V_w); \ln(T_w); \ln(ES_5); \ln(Q_1); GT_1; AT_{5re}; AT_5 \rangle$.

Tabela 4.1. Współczynniki korelacji, współczynniki korelacji wielokrotnej R_0 , średnie wartości średniego błędu prognozy U_0 wyznaczone dla optymalnego zbioru zmiennych prognozujących modelu zlewni Sokołdy w Sokołdzie oddzielnie dla modelu WLR i WWLR.

Table 4.1. Correlation coefficients, multiple correlation coefficients R_0 , average values of the mean forecast error U_0 for the optimal set of predictors for the modeled Sokołda River watershed at Sokołda separately for the MLR and MMLR.

Metoda <i>a priori</i>	$\ln(ES_5)$	$\ln(Q_1)$	GT_1	AT_{5re}	AT_5	R_0	U_0
	Wielokrotna liniowa regresja						
$\ln(Q_{\max})$	0,9417	0,4987	-0,3316	-0,0948		0,9566	0,4308
Wybrano	√	√	√				
$\ln(V_w)$	0,8886		-0,5550	-0,4388	0,6054	0,9397	0,4041
Wybrano	√		√	√	√		
$\ln(T_w)$	0,7697	-0,2618		-0,3606		0,7946	0,5471
Wybrano	√	√		√			
Metoda <i>a priori</i>	Wielowymiarowa wielokrotna liniowa regresja						
$\ln(Q_{\max})$	0,7654	0,3130	0,3340	0,1203	0,1216	0,9675	0,3091
$\ln(V_w)$						0,9455	0,3698
$\ln(T_w)$						0,9006	0,3200
	√	√	√	√	√		

KALIBRACJA MODELI WLR I WWLR Z PARAMETRAMI ESTYMOWANYMI METODĄ B I INV B

Wszystkie modele SLMB skalibrowano (tj. określono współczynniki regresji) na bazie danych kalibracyjnych (1966-1990). Parametry WLR i WWLR oszacowano metodą B i InvB przy użyciu funkcji regresji ze stałą regresji. Współczynniki regresji i warunkowe odchylenie standardowe dla modeli WLR i WWLR zawarto w tabeli 4.2.

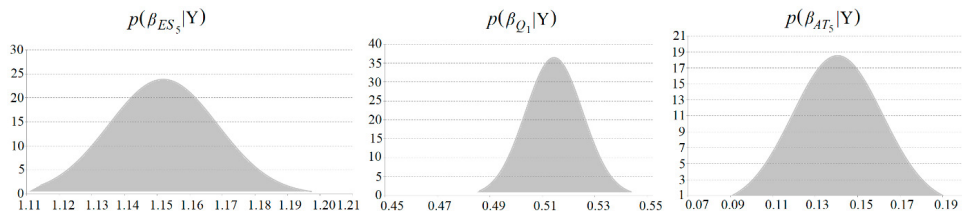
W podejściu Bayesowskim prawdopodobieństwo aprioryczne utożsamia wiedzę o badanej hipotezie lub modelu. W tym rozdziale przyjęto, że do modelowania wezbrań roztopowo-deszczowych został zastosowany statyczny liniowy model regresji. W związku z tym przestrzeń hipotez ograniczono wyłącznie do wartości parametrów modelu regresji $\theta = (\beta, \sigma^2)$ w przypadku wielokrotnej regresji (WLR) i $\theta = (\beta, \Sigma)$ dla wielowymiarowego modelu wielokrotnej regresji (WWLR). Zatem badacz konstruując SLMB powinien w pierwszej kolejności podjąć pewne wstępne założenia *a priori* dotyczące rozkładu wektora parametrów θ . Niekiedy inny stan posiadanej wiedzy (różne prawdopodobieństwa *a priori*) może doprowadzić do rozbieżnych wniosków. Silna zależność wyniku analizy, tzw. prawdopodobieństwa *a posteriori* od założeń apriorycznych wskazuje na fakt, że przeprowadzone obserwacje i/lub pomiary niewiele wniosły w wyjaśnienie analizowanego zjawiska. We wnioskowaniu Bayesowskim podstawową rolę odgrywają rozkłady *a priori*, które pozwalają danym mówić za siebie. Stąd w tym opracowaniu przyjęto, że nieinformacyjny rozkład *a priori* jest rozkładem apriorycznym Jeffreys'a opracowanym dla wiarygodności Gaussowskiej i dla przypadku, w którym wektor parametrów modelu θ jest nieznan (tzn. jest estymowany). Przyjęty w obliczeniach rozkład normalny Gaussa stanowi dobre przybliżenie wiarygodności Bayesowskiej, gdy szum odpowiedzialny za niepewność, czyli brak idealnej zgodności między wartościami zaobserwowanymi a prognozowanymi, jest złożeniem wielu niezależnych zmiennych losowych, tj. macierzy zmiennych prognozujących X . Parametry rozkładu *a priori* i finalowego *a posteriori* będą się różnić. W przypadku rozkładu Gaussa, gdy liczebność próby losowej n jest duża, wpływ założeń *a priori* maleje, podczas gdy n jest mała, średnia z wartości *a posteriori* jest zbliżona do średniej wartości *a priori*.

Tabela 4.2. Kalibracja modeli WLR i WWLR dla zmiennych prognozowanych $\ln(Q_{\max})$, $\ln(V_w)$ i $\ln(T_w)$ dla zlewni Sokoldy w profilu wodowskazowym Sokolda w okresie 1966-1990. Współczynniki kierunkowe i warunkowe odchylenie standardowe σ_w modelu WLR i WWLR z parametrami estymowanymi metodą B i InvB. Myślnik oznacza, że zmienna losowa nie należy do wektora zmiennych prognozujących dla zmiennej prognozowanej.

Table 4.2. Calibrator of the MLR and MMLR models for predictands $\ln(Q_{\max})$, $\ln(V_w)$ and $\ln(T_w)$ for the Sokolda River watershed at the Sokolda for the period 1966-1990. Slope coefficients and conditional standard deviation σ_w of the MLR and MMLR with parameters estimated by B and InvB methods. The dash indicates that variable does not belong to the vector of predictors for the predictand.

Statystyka zmiennej prognozującej	Model/Zmienna prognozowana/Metoda											
	WLR						WWLR					
	$\ln(Q_{\max})$ (m ³ s ⁻¹)		$\ln(V_w)$ (10 ⁶ m ³)		$\ln(T_w)$ (doby)		$\ln(Q_{\max})$ (m ³ s ⁻¹)		$\ln(V_w)$ (10 ⁶ m ³)		$\ln(T_w)$ (doby)	
	B	InvB	B	InvB	B	InvB	B	InvB	B	InvB	B	InvB
$\ln(ES_5)$	1,415	1,407	0,763	0,764	0,729	0,740	1,275	1,273	0,730	0,729	0,434	0,433
$\ln(Q_1)$	0,177	0,175	–	–	–0,063	–0,063	0,202	0,202	0,035	0,035	–0,011	–0,011
GT_1	–0,057	–0,057	–0,611	–0,617	–	–	–0,301	–0,302	–0,643	–0,643	–0,605	–0,604
AT_{5re}	–	–	–0,148	–0,149	–0,148	–0,149	–0,042	–0,042	–0,152	–0,152	0,161	0,161
AT_5	–	–	0,146	0,146	–	–	0,084	0,084	0,151	0,151	0,161	0,161
Const.	–1,960	–1,966	–0,085	–0,127	0,481	0,454	–1,591	–1,589	0,039	0,039	1,202	1,202
σ_w	0,255	0,255	0,258	0,258	0,278	0,278	0,257	0,071	0,258	0,076	0,255	0,066

Podstawą wyboru odpowiedniego rozkładu apriorycznego były obliczone charakterystyki rozkładu prawdopodobieństwa dla wszystkich zmiennych prognozowanych i prognozujących (szczegółowo w tab. 2.2 w rozdziale 2.3.3). Wartości współczynników skośności, mieszczące się w przedziale $<-1,5; 1,5>$, wskazują że rozkłady tych zmiennych są symetryczne i zbliżone do normalnego. Jedynie rozkłady prawdopodobieństwa zmiennych V_w i GT_1 mają prawostronną asymetrię (dłuższe prawe ramię rozkładu). Powyższe wnioski potwierdzają wartości współczynnika spłaszczenia mieszczące się w przedziale $<-3,0; 3,0>$, co świadczy o bliskości ich rozkładów w stosunku do rozkładu normalnego. Rozkłady asymetryczne wykazują większe prawdopodobieństwo uzyskania wartości odległej od przeciętnej niż w rozkładzie normalnym. Zastosowana w badaniu transformacja logarymiczna zmniejszyła w stosunku do zmiennej V_w anomalie, takie jak brak addytywności, normalności i heteroskedastyczności, co doprowadziło do zwiększenia efektywności oszacowań i tym samym do zmniejszenia błędów standardowych estymatorów.



Rys. 4.2. Wykres brzegowego rozkładu *a posteriori* (szary obszar) wybranych parametrów (współczynników kierunkowych) SLMB jako ilustracja wnioskowania Bayesowskiego.

Fig. 4.2. The chart of marginal *a posteriori* distribution (gray area) of SLMB selected parameters (slopes) as an example of Bayesian inference.

W przypadku estymacji parametrów regresji techniką odwrotnego modelowania InvB, problem minimalizacji ujemnego logarytmu funkcji wiarygodności (równanie 3.26 dla wielokrotnej regresji – WLR i równanie (3.39) dla wielowymiarowego modelu wielokrotnej regresji – WWLR) sprowadza się do redukcji czasu niezbędnego do optymalnego wygładzenia funkcji wiarygodności. W konsekwencji powyższej procedury prawdopodobieństwo *a posteriori* będzie miało w pełni rozkład Gaussowski i tym

samym będzie scharakteryzowane średnią i kowariancją. Filtr Kalmana zastosowany w metodzie InvB dywersyfikuje rozkład prawdopodobieństwa *a posteriori* w porównaniu do metody B.

Stosując wzory brzegowego rozkładu *a posteriori* (3.16-3.17) oraz gęstości wielowymiarowego rozkładu *t*-Studenta [Kotz, Nadarajah 2004], skonstruowano wykresy gęstości brzegowych rozkładów *a posteriori* wybranych współczynników kierunkowych β , tj. zmiennych losowych ES_5 , Q_1 , i AT_5 (rys. 4.2).

OCENA MODELI WLR Z PARAMETRAMI ESTYMOVANYMI METODAMI B I INV B

Dokładność modeli oceniono na podstawie pięciu statystycznych miar dopasowania, tj. MAD, MAE, RMSE, R^2 i E_{NS} (szczegółowe wzory w rozdziale 3.10). W celu uzyskania rzetelnych wyników modeli WLR i WWLR zmienne prognozowane (znormalizowane za pomocą funkcji logarytmicznej) przekształcano za pomocą funkcji wykładniczej zgodnie ze wzorem $(\exp^{\ln(\hat{x})} + \varepsilon)$, gdzie ε jest dolnym ograniczeniem rozkładu prawdopodobieństwa. W związku z tym wyniki modeli WLR i WWLR obliczono dla nielogarytmowanych wartości zmiennych prognozowanych.

Modele WLR szacujące parametry techniką InvB wygenerowały dokładniejsze średnio-terminowe prognozy hydrologiczne w porównaniu do modeli z parametrami szacowanymi metodą B (tab. 4.3).

Tabela 4.3. Prognozowane Q_{max} , V_w i T_w za pomocą modeli WLR i WWLR w zlewni rzeki Sokółdy w przekroju wodowskazowym Sokółda w okresie 1991-2002. Ocena metod B i InvB na podstawie statystyk dokładności modelu. Pogrubiona czcionka odnosi się do najlepszej metody dla każdej zmiennej prognozowanej. Myślnik oznacza, że dla modelu nie jest wyznaczana suma błędów δ i iloczyn błędu ω .

Table 4.3. Forecasted Q_{max} , V_w and T_w using the MLR and MMLR models for the Sokółda River watershed at the Sokółda gauging site in the period 1991-2002. Performance of the models with model accuracy statistics for B and InvB methods. Bold refers to the best method for each predictand. The dash means that for the model does not calculate the sum of errors δ and the product of errors ω .

Statystyka	Model/Zmienna prognozowana/Metoda											
	WLR						WWLR					
	Q_{max} (m^3s^{-1})		V_w ($10^6 m^3$)		T_w (doby)		Q_{max} (m^3s^{-1})		V_w ($10^6 m^3$)		T_w (doby)	
	B	InvB	B	InvB	B	InvB	B	InvB	B	InvB	B	InvB
MAD	4,642	3,479	5,420	6,407	6,318	6,007	11,445	11,292	5,432	4,557	7,946	7,903
MAE	1,361	1,239	1,759	1,652	2,573	2,572	2,403	2,378	1,700	1,793	3,114	3,094
RMSE	2,007	1,685	2,412	2,444	3,095	3,073	4,102	4,054	2,340	2,359	3,958	3,930
R^2	0,9778	0,977	0,919	0,920	0,752	0,753	0,965	0,965	0,924	0,923	0,829	0,829
E_{NS}	0,9428	0,959	0,908	0,909	0,741	0,745	0,761	0,766	0,913	0,912	0,577	0,583
δ	–	–	–	–	–	–	1,721	1,710	3,243	3,228	4,063	4,044
ω	–	–	–	–	–	–	22,676	22,322	22,676	22,322	22,676	22,322

W odniesieniu do prognozy Q_{max} wszystkie modele liczyły w podobny sposób (tab. 4.3). WLR-InvB nieznacznie przewyższył WLR-B, uzyskując wyższą E_{NS} (0,959 vs. 0,942) oraz niższe wartości RMSE (1,685 vs. 2,007), MAD (3,479 vs. 4,642) i MAE (1,239 vs. 1,361). W przypadku zlewni rzeki Sokółdy istnieje pewna nieliniowość, zwłaszcza biorąc pod uwagę niewielką powierzchnię zlewni (464 km²). Nieliniowość odpowiedzi systemu hydrologicznego ma większy wpływ na Q_{max} niż na V_w lub T_w [Kundzewicz, Napiórkowski 1986]. Innym czynnikiem zakłócającym ciągi pomiarowe w zlewni rzeki Sokółdy może być zdolność zlewni do retencjonowania dużych objętości wody w bagnistym terenie i torfowiskach.

W tym opracowaniu próbowano zminimalizować efekty uboczne modelowania nieliniowych systemów za pomocą modeli liniowej regresji, decydując się na użycie metody InvB do szacowania parametrów. W opisywanym rozwiązaniu przyjęto liniową postać modelu, redukując w ten sposób problem szacowania $p(\theta_i|Y)$ do zależnego od czasu problemu optymalnego wygładzania funkcji.

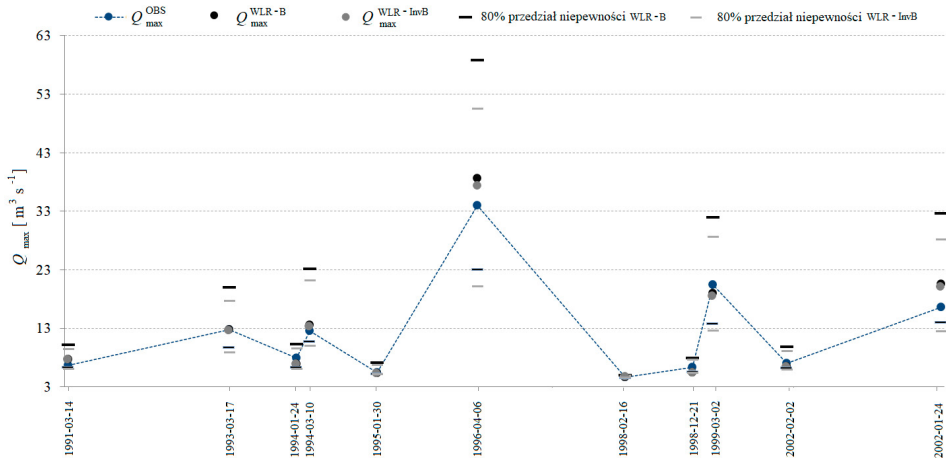
Większe różnice między obserwacjami a prognozami zaobserwowano podczas testowania modeli V_w i T_w . Stwierdzono, że model WLR-InvB dostarczył dokładniejsze prognozy V_w i T_w niż WLR-B.

OCENA MODELI WWLR Z PARAMETRAMI ESTYMOWANYMI METODAMI B I INV B

Wyniki prognozy wezbrania deszczowo-roztopowego opisanego wektorem zmiennych prognozowanych i prognozujących $\langle \ln(Q_{\max}); \ln(V_w); \ln(T_w); \ln(ES_s); \ln(Q_1); GT_1; AT_{\text{src}}; AT_s \rangle$ i uzyskane za pomocą modelu WWLR z parametrami oszacowanymi metodą B i InvB wskazują, że ta ostatnia metoda miała najwyższy R^2 i E_{NS} , jednocześnie z najniższym RMSE, MAD oraz MAE dla Q_{\max} i T_w (tab. 4.3). Do oceny modelu WWLR użyto miary opartej na funkcjach kryterialnych – funkcji celu. W tym badaniu funkcja celu przyjęła postać iloczynu i -tych wyjść modelu [Ozga-Zielińska i in. 1979], który tworzy przestrzeń błędów ω . Wyjściowy model błędów obliczono jako sumę błędów dla każdej z trzech zmiennych prognozowanych, tj. Q_{\max} , V_w i T_w . Procedura sprawdza, który z modeli, tj. WWLR-B lub WWLR-InvB, osiągnął najmniejszą przestrzeń błędów ω . W przypadku WWLR-InvB model uzyskał najmniejszą sumę błędów δ , dla każdej zmiennej prognozowanej, tj. (Q_{\max} , $\delta = 1,710$), (V_w , $\delta = 3,228$), (T_w , $\delta = 4,044$), a zatem uzyskał najmniejszy iloczyn błędów ω (22,322). Model WWLR-InvB uzyskał wyższą ocenę prognozy wektora $\langle Q_{\max}; V_w; T_w \rangle$ niż model WWLR-B (Q_{\max} , $\delta = 1,721$; V_w , $\delta = 3,243$; T_w , $\delta = 4,063$ i $\omega = 22,676$).

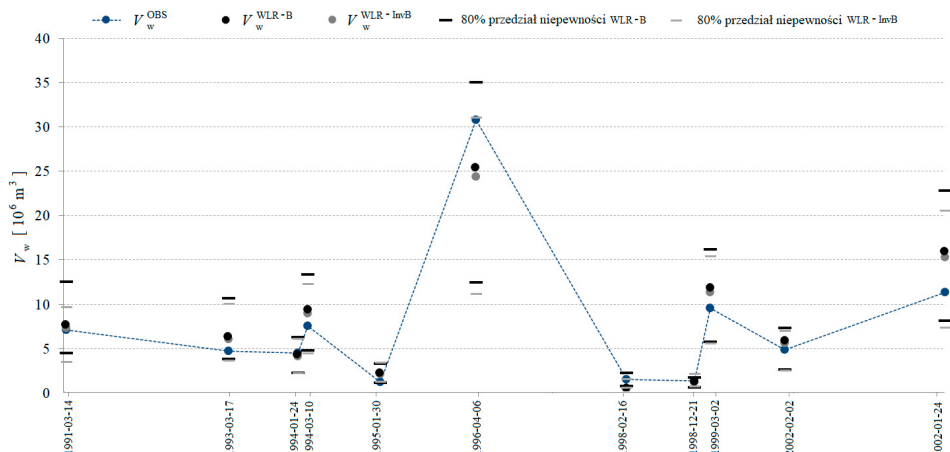
ANALIZA GRAFICZNA MODELI WLR I WWLR Z UWZGLĘDNIENIEM PRZEDZIAŁU NIEPEWNOŚCI

Porównano zaobserwowane wartości zmiennych Q_{\max} , V_w i T_w z wartościami prognozowanymi za pomocą modeli WLR-B i WLR-InvB (rys. 4.3-4.5) oraz prognozy otrzymane z najlepszego modelu WLR z wynikami najlepszego modelu WWLR (rys. 4.6-4.8). Wykresy zawierają prognozowane niepewności obliczone osobno dla każdego modelu, tj. WLR-B, WLR-InvB, WWLR-B i WWLR-InvB, wg wzorów (3.16 i 3.37). Zaznaczone na wykresach niebieskie oraz ciemno- i jasnoszare koła korespondują odpowiednio do obserwacji i prognoz punktowych Q_{\max} , V_w lub T_w , natomiast ciemno- i jasnoszare poziome linie wskazują 80% przedział niepewności prognozy wyznaczonej przy użyciu gęstości *a posteriori*. Wykresy na rysunkach 4.3-4.8 przedstawiają w skali czasu niezależne wezbrania deszczowo-roztopowe zaobserwowane w zlewni Sokołdy w latach 1991-2002.



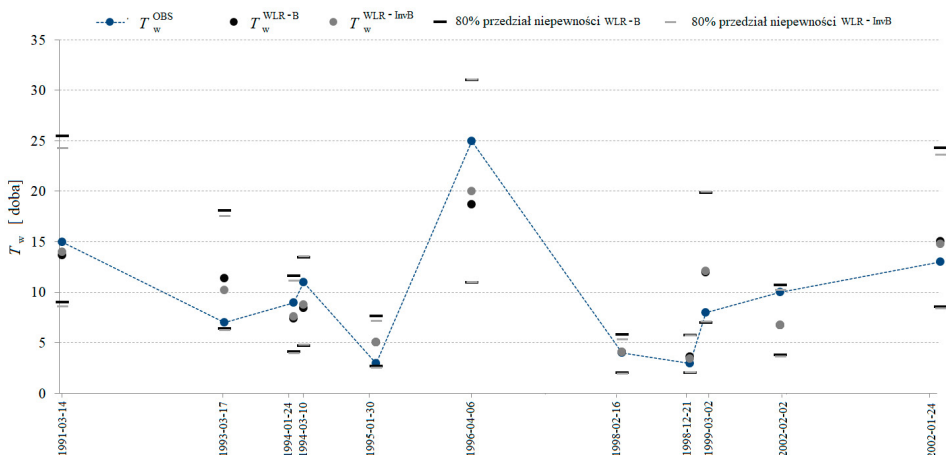
Rys. 4.3. Porównanie wartości obserwowanych (niebieskie koła) z prognozowanymi za pomocą modeli regresji WLR-B (ciemnoszare koła) i WLR-InvB (jasnoszare koła) dla Q_{\max} (przepływu kulminacyjnego) zidentyfikowanych jedenastu wezbrań deszczowo-roztopowych w ujściu zlewni Sokołdy (posterunek wodowskazowy Sokołda) w okresie 1991-2002. Ciemnoszare poziome linie przedstawiają 80% przedział niepewności modelu WLR-B, jasnoszare – modelu WLR-InvB.

Fig. 4.3. Comparison of observed values (blue circle points) with forecasted values from the regression models MLR-B (dark gray circle points) and MLR-InvB (light gray circle points) for Q_{\max} (peak of discharge) of the identified 11 snowmelt floods at the outlet of the Sokołda River watershed (Sokołda gauging station) over the period 1991-2002. Dark gray a horizontal lines depict the 80% uncertainty interval of MLR-B, light gray – MLR-InvB.



Rys. 4.4. Porównanie wartości obserwowanych (niebieskie koła) z prognozowanymi za pomocą modeli regresji WLR-B (ciemnoszare koła) i WLR-InvB (jasnoszare koła) dla V_w (objętości) jedenastu wezbrań deszczowo-roztopowych w ujściu zlewni Sokółdy (posterunek wodowskazowy Sokółda) w okresie 1991-2002. Ciemnoszare poziome linie przedstawiają 80% przedział niepewności modelu WLR-B, jasnoszare – modelu WLR-InvB.

Fig. 4.4. Comparison of observed values (blue circle points) with forecasted values from the regression models MLR-B (dark gray circle points) and MLR-InvB (light gray circle points) for V_w (the volume of snowmelt flood) 11 snowmelt floods at the outlet of the Sokółda River watershed (Sokółda gauging station) over the period 1991-2002. Dark gray a horizontal lines depict the 80% uncertainty interval of MLR-B, light gray – MLR-InvB.



Rys. 4.5. Porównanie wartości obserwowanych (niebieskie koła) z prognozowanymi za pomocą modeli regresji WLR-B (ciemnoszare koła) i WLR-InvB (jasnoszare koła) dla T_w (czasu trwania) jedenastu wezbrań deszczowo-roztopowych w ujściu zlewni Sokółdy (posterunek wodowskazowy Sokółda) w okresie 1991-2002. Ciemnoszare poziome linie przedstawiają 80% przedział niepewności modelu WLR-B, jasnoszare – modelu WLR-InvB.

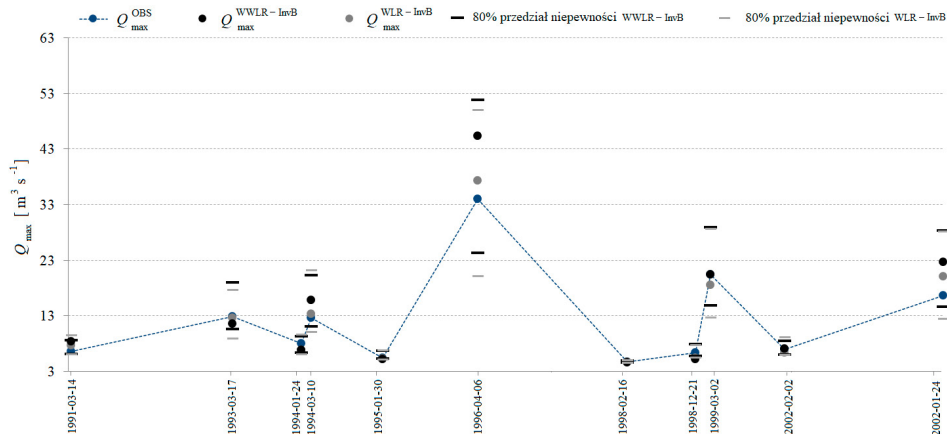
Fig. 4.5. Comparison of observed values (blue circle points) with forecasted values from the regression models MLR-B (dark gray circle points) and MLR-InvB (light gray circle points) methods for T_w (flood duration) 11 snowmelt floods at the outlet of the Sokółda River watershed (Sokółda gauging station) over the period 1991-2002. Dark gray a horizontal lines depict the 80% uncertainty interval of MLR-B, light gray – MLR-InvB.

Prognoza Q_{\max} (rys. 4.3) jednego z największych wezbrań deszczowo-roztopowych zanotowanych w historii zlewni (1996) została nieco zawyżona, natomiast pozostałe zdarzenia zostały przez obie metody dokładnie zaprognozowane. Wszystkie obserwacje znalazły się wewnątrz 80% przedziału niepewności prognozy. Na rysunkach 4.4-4.5 porównano wartości obserwowane i wartości prognozowane przez model WLR z parametrami szacowanymi metodą B i InvB, odpowiednio dla V_w i T_w . T_w różni się w zależności od warunków pogodowych. Na przykład powolny proces topnienia śniegu wywołany adwekcją ciepła, w porównaniu do gwałtownego topnienia śniegu na wiosnę, generuje inne objętości i czasy odpływu ze zlewni. Czynniki, które wpływają na niepewność danych wejściowych (wybór zmiennych prognozujących) oddziałują również na niepewność wyjściową modelu, tj. na symulację lub prognozę T_w ; tym samym wpływają na gorszą ocenę modelu T_w w porównaniu z modelami Q_{\max} i V_w .

Ostatnim elementem testu wydajności modeli była graficzna analiza porównawcza WLR i WWLR z parametrami szacowanymi metodami B i InvB, oddzielnie dla każdej zmiennej prognozowanej. Na rysunku 4.6 porównano: zaobserwowane wartości Q_{\max} , prognozowane wartości otrzymane z najlepszego modelu WLR, tj. WLR-InvB, najlepszego modelu WWLR, tj. WWLR-InvB, oraz 80% przedziały niepewności obliczone dla obu modeli. Wartości prognozowane dla Q_{\max} za pomocą modelu WLR-InvB wykazały, że RMSE jest o 140% dokładniejszy od modelu WWLR-InvB o parametrach szacowanych tą samą metodą.

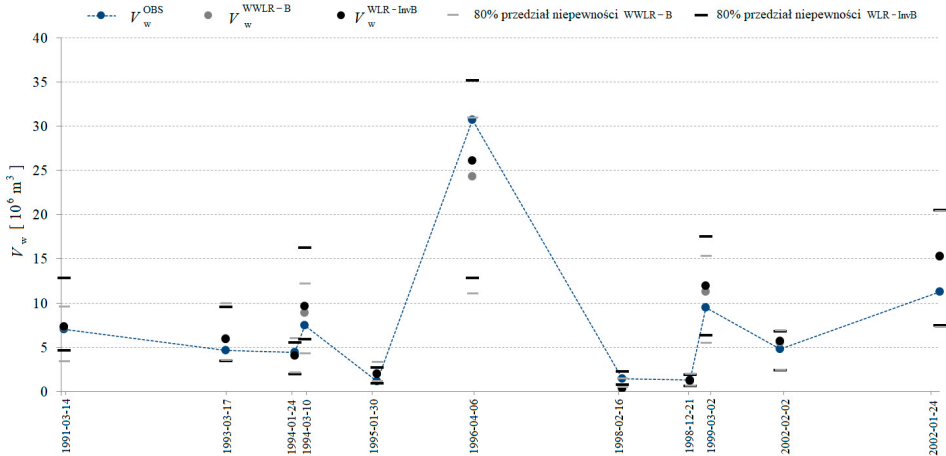
Hydrogramy porównujące zaobserwowane wartości V_w z prognozowanymi uzyskanymi z najlepszego modelu WLR-InvB, i najlepszym modelem WWLR-B wykazały, że ten ostatni ma RMSE o 4,3% mniej dokładny niż WLR-InvB (rys. 4.7). Objętość wezbrania V_w była nieco niedoszacowana w okresie wysokiego przepływu w 1996 r., natomiast była wyraźnie zawyżona w czasie marcowych wezbrań w latach 1993, 1994, 1999 oraz w styczniu 2002 roku.

Na rysunku 4.8 porównano hydrogramy zaobserwowanych wartości T_w z prognozowanymi otrzymanymi z najlepszego modelu WLR-InvB i najlepszym modelem WWLR-InvB. Uzyskane wyniki pokazują, że WLR-InvB lepiej prognozował mając o 27,9% niższy RMSE niż model WWLR-InvB. Model WWLR-InvB istotnie przeszacował T_w , zwłaszcza w marcu 1994, 1999 i w styczniu 2002 roku. Szacowanie parametrów SLMB metodą Bayesowską pozwala redukować wartości średniego warunkowego odchylenia standardowego, a co za tym idzie uzyskiwać dokładniejszą prognozę.



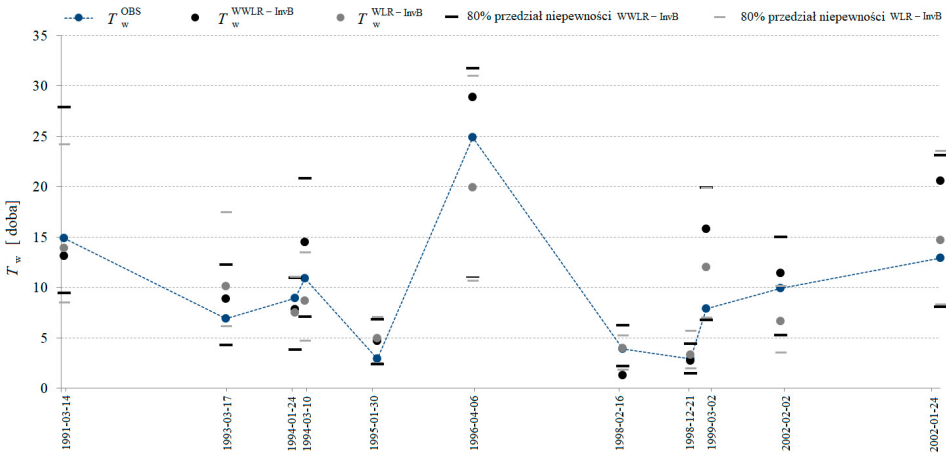
Rys. 4.6. Porównanie wartości obserwowanych (niebieskie koła) Q_{\max} (przepływu kulminacyjnego wezbrania) z prognozowanymi, za pomocą najlepszego modelu 3D wielokrotnej liniowej regresji z parametrami estymowanymi metodą InvB (WWLR-InvB) (ciemnoszare koła) i najlepszym modelem wielokrotnej liniowej regresji z parametrami estymowanymi metodą InvB (WLR-InvB) (jasnoszare koła) w okresie 1991-2002. Ciemnoszare poziome linie przedstawiają 80% przedział niepewności modelu WWLR-InvB, jasnoszare – modelu WLR-InvB.

Fig. 4.6. Comparison of observed values (blue circle points) of Q_{\max} (peak of discharge of snowmelt flood) with forecasted values from the best 3D multiple linear regression model with parameters estimated by InvB (MMLR-InvB) (dark gray circle points) and the best multiple linear regression model with parameters estimated by the InvB method (MLR-InvB) (light gray circle points) over the period 1991-2002. Dark gray a horizontal lines depict the 80% uncertainty interval of MMLR-InvB, light gray – MLR-InvB.



Rys. 4.7. Porównanie wartości obserwowanych (niebieskie koła) V_w (objętości wezbrania) z prognozowanymi, za pomocą najlepszego modelu 3D wielokrotnej liniowej regresji z parametrami estymowanymi metodą B (WWLR-B) (jasnoszare koła) i najlepszym modelem wielokrotnej liniowej regresji z parametrami estymowanymi metodą InvB (WLR-InvB) (ciemnoszare koła) w okresie 1991-2002. Jasnoszare poziome linie przedstawiają 80% przedział niepewności modelu WWLR-B, ciemnoszare – modelu WLR-InvB.

Fig. 4.7. Comparison of observed values (blue circle points) of V_w (the volume of snowmelt flood) with forecasted values from the best 3D multiple linear regression model with parameters estimated by B (MMLR-B) (light gray circle points) and the best multiple linear regression model with parameters estimated by the InvB method (MLR-InvB) (dark gray circle points) over the period 1991-2002. Light gray horizontal lines depict the 80% uncertainty interval of MMLR-B, dark gray – MLR-InvB.



Rys. 4.8. Porównanie wartości obserwowanych (niebieskie koła) T_w (czasu trwania wezbrania) z prognozowanymi, za pomocą najlepszego modelu 3D wielokrotnej liniowej regresji z parametrami estymowanymi metodą InvB (WWLR-InvB) (ciemnoszare koła) i najlepszym modelem wielokrotnej liniowej regresji z parametrami estymowanymi metodą InvB (WLR-InvB) (jasnoszare koła) w okresie 1991-2002. Ciemnoszare poziome linie przedstawiają 80% przedział niepewności modelu WWLR-InvB, jasnoszare – modelu WLR-InvB.

Fig. 4.8. Comparison of observed values (blue circle points) of T_w (flood duration) with forecasted values from the best 3D multiple linear regression model with parameters estimated by InvB (MMLR-InvB) (dark gray circle points), and the best multiple linear regression model with parameters estimated by the InvB (MLR-InvB) method (light gray circle points) over the period 1991-2002. Dark gray horizontal lines depict the 80% uncertainty interval of MMLR-InvB, light gray – MLR-InvB.

4.2.2. Podsumowanie i wnioski

W tym rozdziale do prognozowania ilościowych charakterystyk wezbrań deszczowo-roztopowych (tj. Q_{\max} , V_w , T_w) zastosowano modele WLR i WWLR z parametrami szacowanymi metodami B i InvB. Za pomocą modelu WWLR, w kontraście do WLR, otrzymano prognozy nie tylko skalarne, ale również wektorowe $\langle Q_{\max}; V_w; T_w \rangle$. Generalnie SLMB z parametrami estymowanymi metodą InvB dawał dokładniejsze prognozy w porównaniu do metody B. Zastosowane w modelach SLMB szacowanie parametrów regresji metodami Bayesowskimi, obniżyło wartości średniego warunkowego odchylenia standardowego i zwiększyło dokładność prognozy. Obserwacje Q_{\max} , V_w , T_w mieściły się wewnątrz 80% przedziału niepewności modelu. W modelach WLR i WWLR zależności między zmiennymi losowymi przyjmuje się jako liniowe, jednakże w praktyce to założenie często nie jest prawdziwe. Mała powierzchnia zlewni rzeki Sokoły (464 km²) sugeruje, że istnieje pewna nieliniowość. W tym przypadku warto byłoby zbadać możliwość wykorzystania modeli, które należą do kategorii nieliniowych i niegaussowskich. Niemniej zastosowane procedury WLR i WWLR tylko w niewielkim stopniu odstępają od powyższego założenia. W rozdziale założono liniową postać modelu InvB i dodatkowo niezależność błędów o rozkładzie Gaussowskim. Metody szacowania parametrów B i InvB (z Gaussowskim *a posteriori* rozkładem prawdopodobieństwa) są odpowiednie do modelowania zlewni o liniowej funkcji odpowiedzi systemu hydrologicznego. W przypadku modelowania zlewni z nieliniową odpowiedzią systemu hydrologicznego, technika modelowania InvB okazała się nieco lepszą metodą szacowania parametrów od metody Bayesowskiej B.

We wnioskowaniu Bayesowskim stosowane są założenia aprioryczne. Z jednej strony niewiedza o analizowanym zjawisku ukryta jest w prawdopodobieństwie *a priori*, które nie jest przetwarzane bezpośrednio, z drugiej strony podejście subiektywne jest silnie wsparte obecnością danych. W analizowanym problemie estymacji parametrów historyczne pomiary i obserwacje znalazły odbicie w postaci rozkładu *a priori*. Różne prawdopodobieństwa aprioryczne, wynikające z niejednoznacznego wstępnego stanu wiedzy, prowadzą do różnych konkluzji (prawdopodobieństw *a posteriori*). Z kolei silna zależność prawdopodobieństwa *a posteriori* od założeń apriorycznych świadczy o słabości przeprowadzonych obserwacji i pomiarów. Zatem zastosowanie wnioskowania Bayesowskiego do modelowania wezbrań deszczowo-roztopowych powinno być poprzedzone silną analizą statystyczną zmiennych prognozujących i prognozowanych, w tym niestacjonarności i nieliniowości odpowiedzi systemu hydrologicznego. Większą uwagę należy zwrócić na wyznaczenie optymalnego zbioru zmiennych prognozujących i prognozowanych, w połączeniu z analizą przyczynowo-skutkową modelowanego zjawiska.

5. KRÓTKOTERMINOWA PROGNOZA PRZEPŁYWU RZECZNEGO ZA POMOCĄ DYNAMICZNYCH LINIOWYCH MODELI BAYESOWSKICH (DLMB)

Celem było zbadanie przydatności wybranych dynamicznych liniowych modeli Bayesowskich, tj. Varying Coefficient Regression (VCR), czyli jednowymiarowej postaci modelu Matrix Normal Regression (MNR) [Harvey 1986] i Discount Weighted Regression (DWR) oraz jednowymiarowej postaci modelu Multivariate Discount Weighted Regression (MDWR) [Triantafyllopoulos 2006], do symulacji i prognozy hydrologicznej. Pomimo zastosowania w tym badaniu skalarnej zmiennej prognozowanej, opracowano algorytm obliczeniowy dla przypadku uogólnionego, tj. dla wektora zmiennych prognozowanych [szczegóły algorytmu w rozdziale 3.5]. Zaproponowane modele (VCR i DWR) zawierają mechanizm uczenia się w postaci filtru Kalmana, co pozwala na aktualizację (updating) zmiennych stanu, tzn.:

- 1) Uwzględnienie różnicy między obserwacjami i wartościami symulowanymi zmiennej prognozowanej przed uruchomieniem modelu symulacyjnego dla potrzeb prognozy.
- 2) Zapisanie zmiennych stanu modelu w każdym kroku przetwarzania, a następnie uruchomienie modelu ze zmiennymi stanu zapisanymi w poprzednim kroku.

Przedstawione zagadnienia zilustrowano przykładem zastosowania modeli VCR i DWR do:

- 1) Symulacji rocznego hydrogramu.
- 2) Krótkoterminowej prognozy przepływu rzeczne go z wyprzedzeniem 1-, 2- i 3-dobowym.

Eksperyment polegający na modelowaniu przepływu rzeczne go Q_r^{WIZ} (m^3s^{-1}) z jednodobowym krokiem przetwarzania przeprowadzono dla zlewni Górnej Narwi w profilu wodowskazowym Wizna. Oba modele, tj. VCR i DWR, zostały porównane z tradycyjną metodą prognozowania za pomocą wielokrotnej liniowej regresji (WLR) i nowszej metody, tj. sieci neuronowej (NN).

Struktura niniejszego rozdziału przedstawia się następująco:

- 1) Metodyka opracowania VCR, DWR, WLR i NN.
- 2) Zastosowanie VCR, DWR, WLR i NN do symulacji rocznego hydrogramu i krótkoterminowej prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym.
- 3) Omówienie wyników.
- 4) Podsumowanie przeprowadzonego eksperymentu i wnioski wynikające z porównania VCR, DWR, WLR i NN.

5.1. Metodyka opracowania modeli do symulacji rocznego hydrogramu i krótkoterminowej prognozy hydrologicznej

Do symulacji rocznego hydrogramu dane wejściowe podzielono na podzbiór do kalibracji (1 listopada 1989 – 31 października 1991) i do walidacji modeli (1 listopada 1991 – 31 października 1995). Symulowane i prognozowane hydrogramy obu dynamicznych modeli, tzn. VCR i DWR, porównano z analizą wielokrotnej liniowej regresji (WLR) i sieciami neuronowymi (NN). Do obliczenia krótkoterminowej prognozy hydrologicznej wybrano dane od 1 listopada 1995 do 31 października 1996 roku. Wyszczególnione czasy wyprzedzenia (1-, 2- i 3-dobowe) krótkoterminowej prognozy są standardowo stosowane w hydrologii operacyjnej w Biurach Prognoz Hydrologicznych na terenie Polski. Dane kalibracyjne i walidacyjne posłużyły jako dane wejściowe do każdego z czterech testowanych modeli, tj. WLR, NN, VCR i DWR. Do symulacji rocznego hydrogramu i do obliczenia krótkoterminowej prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym idealnym rozwiązaniem jest użycie takich samych równań przetwarzających wejście w wyjście, aby uniknąć niespójności modeli. Ustalono, że NN, VCR i DWR wymagały tylko jednego równania zarówno do symulacji rocznego hydrogramu, jak i do prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym, natomiast WLR wymagała oddzielnych równań do symulacji i prognoz w celu osiągnięcia najlepszych wyników.

Długie ciągi pomiarowe (dłuższe niż 2 lata) stosowane do kalibracji VCR i DWR znacząco wydłużają czas przetwarzania. Wynika to z następujących przyczyn:

- 1) Przetwarzania macierzowego angażującego procedury odwracania macierzy.
- 2) Używania algorytmów Monte Carlo.
- 3) Wzrostu liczby parametrów wraz z każdą nową przetwarzaną obserwacją.

W projektowaniu modelu hydrologicznego zbiory do kalibracji powinny być dłuższe od tych do walidacji. Wynika to z konieczności zapewnienia podstawowego paradygmatu modelowania, czyli większej liczby przykładów par (zmienne prognozujące i zmienne prognozowane) w celu efektywniejszej kalibracji modelu. Zastosowany podział danych na podzbiory do kalibracji i walidacji nie jest idealny dla WLR i NN. W pracy zdecydowano się użyć takich samych zbiorów danych wejściowych do wszystkich zastosowanych modeli, aby umożliwić bezpośrednie ich porównanie.

Ogólna metodyka opracowania VCR, DWR, WLR i NN zawiera następujące procedury:

- 1) Identyfikacja zjawisk hydrologicznych.
- 2) Analiza statystyczna zmiennych losowych będących wejściem do modeli za pomocą metod statystycznych.
- 3) Kalibracja i walidacja modeli VCR, DWR, WLR i NN.
- 4) Ocena rocznych symulacji hydrologicznych.
- 5) Ocena krótkoterminowych prognoz hydrologicznych z wyprzedzeniem 1-, 2- i 3-dobowym.

Szczególną uwagę należy zwrócić na przygotowanie danych wejściowych do modelu za pomocą metod analizy statystycznej [Helsel, Hirsch 2002; Ciupak 2004, 2009]. Zmienne prognozujące muszą być czynnikami, które bezpośrednio wpływają na wielkość i przebieg wezbrania. Powinny być wzajemnie niezależne (tj. nieskorelowane ze sobą), ponieważ silna korelacja oznacza nakładające się informacje, czyli zmienna prognozująca nie dostarcza dodatkowego źródła informacji do modelu o zmiennej prognozowanej. W wyborze optymalnego zbioru zmiennych prognozujących korzysta się metody analizy statystycznej, angażujące badanie zależności między potencjalnym zbiorem zmiennych prognozujących a zmienną prognozowaną. Częścią optymalizacji jest procedura wyboru zbioru zmiennych prognozujących, dla których średnia wartość średniego błędu prognozy osiąga wartość minimalną [Kaczmarek 1977; Sollis 2012] [szczegóły algorytmu w Załączniku: Dodatek F]. Następnie wykonywana jest analiza statystyczna danych wejściowych do modeli VCR i DWR [szczegóły w rozdziale 2.4.3] z zastosowaniem procedur odpowiednich do analizy zbioru zmiennych zależnych. Procedura wykonywana w ramach preprocesora danych wejściowych obejmuje badanie jednorodności statystycznej danych wejściowych do modeli VCR i DWR i wyznaczanie charakterystyk rozkładu prawdopodobieństwa. W ten sposób przygotowane ciągi pomiarowe służą jako wejście do modeli VCR, DWR, MLR i NN.

Szczegółowe informacje na temat przeprowadzonego opracowania danych wejściowych do modeli, tj. identyfikacji zjawiska hydrologicznego i analizy statystycznej zmiennych losowych, opisano w rozdziale 2.4.

OPRACOWANIE WIELOKROTNEJ REGRESJI LINIOWEJ (WLR)

Modele WLR opracowano oddzielnie do symulacji rocznego hydrogramu i prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dniowym, stosując sekwencyjne metody doboru zmiennych za pomocą metody regresji krokowej w przód. Parametry modeli wielokrotnej regresji liniowej oszacowano metodą najmniejszych kwadratów (MNK) [Rao, Toutenburg 1999; szczegóły algorytmu w rozdziale 3.2].

OPRACOWANIE SIECI NEURONOWYCH (NN)

W tym opracowaniu wszystkie modele NN ukierunkowano na potrzeby modelowania hydrologicznego oraz prognozowania przepływu rzeczne z wyprzedzeniem 1-, 2- i 3-dobowym. Wejście neuronu dla każdego modelu wybrano na podstawie zmiennych prognozujących w taki sposób, aby każda pojedyncza sieć miała pięć węzłów wejściowych, z których każdy reprezentował jedną ze zmiennych prognozujących. Liczba ukrytych węzłów zastosowanych w ukrytej warstwie wynosiła od 1 do 2 ($d + 1$) węzłów, gdzie d reprezentuje liczbę węzłów wejścia (zmiennych prognozujących) użytych w warstwie wejściowej [Mishra, Desai 2006]; w tej analizie $d = 5$. Dla każdej konfiguracji ukrytych neuronów utworzono $[2(d + 1) + 1]$ podstawowych modeli (po jednym dla każdej konfiguracji ukrytych neuronów) w celu określenia najlepszej kombinacji danych wejściowych i ukrytych neuronów reprezentujących optymalny model NN. Ponieważ funkcja wydajności modelu stara się minimalizować błąd średniokwadratowy MSE, wybrano taką architekturę modelu NN (konfigurację ukrytego neuronu), dla której MSE osiągnął najmniejszą wartość dla danych kalibracyjnych [szczegóły dotyczące struktury i algorytmu uczenia sieci NN w rozdziale 3.4]. Warstwa wyjścia pozostaje niezmienna w węzle wyjścia, który reprezentuje wyjście modelu określone poprzez odwzorowanie wysokiej rozdzielczości parametrów sieci

za pomocą zmiennych prognozujących. Wszystkie zmienne prognozujące i prognozowane zostały przeskalowane do przedziału $[-1, 1]$, w celu użycia neuronów tangensoidalnych i liniowych funkcji przekształcających w ukrytych warstwach i warstwie wyjściowej. W tym badaniu struktura modelu NN była reprezentowana jako NN (I-H-O), gdzie I reprezentuje liczbę neuronów wejściowych (zmienne prognozujące), H liczbę ukrytych neuronów w ukrytej warstwie, a O liczbę zmiennych prognozowanych w warstwie wyjściowej (która zawsze jest równa 1 dla modelu jednowymiarowego, tj. $p = 1$). W tym badaniu zarówno I, jak i O pozostały niezmiennione, odpowiednio na poziomie 5 i 1.

5.2. Przykład zastosowania modeli Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR) do symulacji rocznego hydrogramu oraz krótkoterminowej prognozy hydrologicznej

Dynamiczne liniowe modele Bayesowskie (VCR i DWR) w postaci jednowymiarowej (tj. obliczenia wykonywane są tylko dla jednej zmiennej prognozowanej) zastosowano w dwóch hydrologicznych kontekstach:

- 1) Symulacji rocznego hydrogramu.
- 2) Prognozy hydrologicznej z wyprzedzeniem 1-, 2-, i 3-dobowym.

5.2.1. Wyniki i dyskusja

ZBIÓR ZMIENNYCH PROGNOZUJĄCYCH, ANALIZA STATYSTYCZNA DANYCH WEJŚCIOWYCH DO MODELI VCR i DWR

Wejściem do procedury optymalizacji zbioru zmiennych prognozujących były hydrologiczne i meteorologiczne czynniki, które wpływają na hydrogram przepływu w zlewni Górnej Narwi w profilu wodowskazowym Wizna. Szczegółowe informacje dotyczące identyfikacji danych wejściowych do modelu zawarto w rozdziale 2.4.2. Optymalny zbiór zmiennych prognozujących (tab. 5.1) został określony metodą *a priori* i wg minimum średniej wartości średniego błędu prognozy [szczegóły algorytmu w Załączniku: Dodatek F]. W tabeli 5.1 zawarto współczynniki korelacji między wybranymi zmiennymi prognozującymi i zmienną prognozowaną, współczynnik korelacji wielokrotnej R_0 , średnią wartość średniego błędu prognozy U_0 i warunkowe odchylenie standardowe σ_w wyznaczone dla optymalnego zbioru zmiennych prognozujących dla modelu wielokrotnej liniowej regresji, WLR. Ostatecznie codzienny przepływ w zlewni Górnej Narwi w Wiźnie Q_{τ}^{WIZ} (m^3s^{-1}) uzależniono od wektora zmiennych prognozujących $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$ (m^3s^{-1}), $P_{\tau-2}^{AREA}$ (10^6m^3), $T_{\tau-2}^{MAX}$ i $T_{\tau-2}^{MIN}$ ($^{\circ}C$), tj. $\langle Q_{\tau}^{WIZ}; Q_{\tau-2}^{STR}; Q_{\tau-2}^{BUR}; P_{\tau-2}^{AREA}; T_{\tau-2}^{MAX}; T_{\tau-2}^{MIN} \rangle$, reprezentujących odpowiednio: i) przepływy w profilach wodowskazowych w Strękowej Górze i w Burzynie zaobserwowane dwie doby wcześniej w stosunku do odtwarzanego hydrogramu w Wiźnie, ii) średnią wysokość warstwy opadu w zlewni Górnej Narwi powyżej wodowskazu Wizna zaobserwowaną dwie doby wcześniej, iii) średnie dobowe wartości maksymalnej i minimalnej temperatury powietrza zaobserwowane dwie doby wcześniej na obszarze modelowanej zlewni.

Tabela 5.1. Współczynniki korelacji, współczynnik korelacji wielokrotnej R_0 , średnia wartość średniego błędu prognozy U_0 i warunkowe odchylenie standardowe σ_w wyznaczone dla optymalnego zbioru zmiennych prognozujących modelu zlewni Górnej Narwi w Wiźnie w modelu wielokrotnej regresji liniowej, WLR.

Table 5.1. Correlation coefficients, multiple correlation coefficient R_0 , average value of the mean forecast error U_0 and conditional standard deviation σ_w for the optimal set of predictors for the modeled Upper Narew River watershed at Wizna site in the multiple linear regression, MLR.

	$Q_{\tau-2}^{STR}$ (m^3s^{-1})	$Q_{\tau-2}^{BUR}$ (m^3s^{-1})	$P_{\tau-2}^{AREA}$ (10^6m^3)	$T_{\tau-2}^{MAX}$ ($^{\circ}C$)	$T_{\tau-2}^{MIN}$ ($^{\circ}C$)	R_0	U_0	σ_w (m^3s^{-1})
Q_{τ}^{WIZ} (m^3s^{-1})	0,971	0,956	-0,146	-0,568	-0,477	0,982	6,711	35,703

Przeprowadzona analiza statystyczna [patrz szczegóły w rozdziale 2.4.3] wykazała, że w ciągach pomiarowych nie wykryto elementów odstających. Po sprawdzeniu niezależności za pomocą testu serii i współczynnika autokorelacji okazało się, że wszystkie badane zmienne są zależne, wykryto natomiast

występowanie trendu malejącego dla zmiennej $P_{\tau-2}^{AREA}$ i rosnącego dla $T_{\tau-2}^{MAX}$, odpowiednio na poziomie istotności 0,016 i 0,017 (p -wartość w teście Manna-Kendalla). W funkcji autokorelacji zmiennych losowych Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $T_{\tau-2}^{MAX}$ i $T_{\tau-2}^{MIN}$ zaobserwowano stały cykl o rocznym okresie z uwzględnieniem dwóch sezonów, tj. zimowego i letniego. Sprawdzenie stacjonarności zależnych zmiennych losowych wykazało, że Q_{τ}^{WIZ} , $Q_{\tau-2}^{STR}$, $Q_{\tau-2}^{BUR}$, $P_{\tau-2}^{AREA}$ i $T_{\tau-2}^{MIN}$ są zmiennymi zależnymi i stacjonarnymi, natomiast $T_{\tau-2}^{MAX}$ jest zmienną zależną i niestacjonarną. Powyższe wnioski potwierdziły zasadność zastosowania modeli VCR i DWR do modelowania hydrogramu rzecznego [patrz rozdział 3.5.3].

KALIBRACJA MODELI VCR I DWR

Modele VCR i DWR użyto do symulacji rocznego hydrogramu, jak również do krótkoterminowej prognozy przepływu rzeczego z wyprzedzeniem 1-, 2- i 3-dobowym. Modele zostały skalibrowane na podstawie dwuletnich danych obserwacji hydrologiczno-meteorologicznych (1990-1991) i zwalidowane z wykorzystaniem danych walidacyjnych (1992-1995). Postępując zgodnie z procedurą opisaną w rozdziale (5.1), wejściem do obu modeli był ten sam zbiór zmiennych prognozujących (tab. 5.1) zarówno do symulacji rocznego hydrogramu, jak i krótkoterminowej prognozy przepływu rzeczego. VCR i DWR mają parametry stanu, których wartości zmieniają się w każdym kroku obliczeniowym. Dynamiczne modele Bayesowskie wymagają pierwszego uruchomienia, tj. uruchomienia w chwili $\tau = 0$ z domyślnymi wartościami zmiennych \mathbf{m}_0 , \mathbf{C}_0 , \mathbf{S}_0 (*a priori* wariancja) i δ . Wstępne przetwarzanie oszacowuje początkowe wartości zmiennych stanu. Oba modele były wielokrotnie uruchomiane ze zbiorem danych kalibracyjnych i z różnymi wartościami \mathbf{C}_0 . Jeżeli początkowe ustawienia przekątnej diagonalnej macierzy kowariancji \mathbf{C}_0 są odpowiednio duże, wówczas wybór \mathbf{m}_0 nie jest kluczowy, ponieważ parametry stanu będą szybko dostosowywać się do faktycznych danych. Pozostałe wartości macierzy kowariancji \mathbf{C}_0 są ustawione na 0. W tym badaniu zbiór wartości początkowych \mathbf{C}_0 mieścił się w zakresie określonym przez dwie skrajne wartości. Ustawienie diagonalnej przekątnej macierzy \mathbf{C}_0 elementami równymi 1,0, tj. diag(1,0, 1,0, ..., 1,0), oznacza że wstępna wartość (tj. *a priori*) jest całkiem dobrze uzasadniona, lecz ma niską precyzję, podczas gdy ustawienie głównej przekątnej elementami równymi 100,0, tj. diag(100,0, 100,0, ..., 100,0), wskazuje na wysoką niepewność posiadanej wiedzy.

W przypadku modelu statycznego wartość współczynnika dyskontowego jest równa $\delta = 1$ i elementy diagonalnej przekątnej macierzy kowariancji \mathbf{C}_0 wynoszą 1,0, natomiast dla modelu całkowicie niestabilnego $\delta = 0$ elementy diagonalnej przekątnej są równe $+\infty$. W praktyce należy unikać obu przypadków. Współczynnik dyskontowy δ powinien mieścić się w przedziale [0,1; 0,99]. Początkowe parametry powinny być starannie dobrane w celu zapewnienia nieinformacyjności wnioskowania Bayesowskiego. Powyższy stan odzwierciedlają wysokie wartości macierzy kowariancji \mathbf{C}_0 i niskie wartości stopni swobody n_{τ} . Optymalne parametry początkowe w badaniu ustalono na podstawie kilkukrotnego uruchomienia symulacji. Analizowano wrażliwość obu modeli na początkową wartość macierzy kowariancji \mathbf{C}_0 . Przykłady uzyskanych wyników dla przekątnej diagonalnej macierzy kowariancji równej {1,0; 5,1; 100,0} pokazano w tabeli 5.2.

Tabela 5.2. Ustalenie optymalnej początkowej wartości diagonalnej macierzy kowariancji \mathbf{C}_0 .

Table 5.2. Determining the optimal initial value of the diagonal matrix covariance \mathbf{C}_0 .

Model	\mathbf{C}_0	R ²	R ₀ M	RMSE	ENS
VCR	diag(1,0, ..., 1,0)	0,948	1,007	8,275	0,987
	diag(5,1, ..., 5,1)	0,965	0,992	6,754	0,992
	diag(100, ..., 100)	0,964	0,987	6,793	0,991
DWR	diag(1,0, ..., 1,0)	0,935	1,007	9,147	0,984
	diag(5,1, ..., 5,1)	0,979	0,989	5,263	0,994
	diag(100, ..., 100)	0,975	0,983	5,667	0,978

W celu otrzymania w modelach VCR i DWR stabilnego hydrogramu, na podstawie uzyskanych wyników analizy wrażliwości (tab. 5.2) określono następujące początkowe wartości zmiennych stanu: \mathbf{m}_0 – macierz z przekątną równą 0,15; \mathbf{C}_0 – macierz z przekątną równą 5,1; \mathbf{S}_0 – macierz z przekątną równą 0,001 i $\delta = 0,96$. Zmiany wartości współczynników kierunkowych i *a priori* wariancji \mathbf{S}_{τ} w kolejnych krokach przetwarzania (tj. od $\tau = 0$ do $\tau = 729$), w procesie kalibracji modeli VCR i DWR dla

okresu 1990-1991, przedstawiono w tabeli 5.3. W przypadku modeli VCR i DWR liczba zmiennych prognozowanych $l = 1$, zatem macierz wariancji \mathbf{S}_τ redukuje się do liczby rzeczywistej.

Tabela 5.3. Zmiany wartości współczynników kierunkowych i *a priori* wariancji \mathbf{S}_τ w kolejnych krokach przetwarzania, tj. od $\tau = 0$ do $\tau = 729$, dla modeli VCR i DWR kalibrowanych w okresie 1990-1991.

Table 5.3. Changes in the values of direct coefficients and prior variance \mathbf{S}_τ in subsequent processing steps, i.e. from $\tau = 0$ to $\tau = 729$, for VCR and DWR models calibration in the period 1990-1991.

	τ	$Q_{\tau-2}^{STR}$	$Q_{\tau-2}^{BUR}$	$p_{\tau-2}^{AREA}$	$T_{\tau-2}^{MAX}$	$T_{\tau-2}^{MIN}$	Stała	\mathbf{S}_τ
VCR	0	0,150	0,150	0,150	0,150	0,150	0,150	0,00001
	1	0,136	0,134	0,105	0,094	0,114	0,145	0,002
	2	0,167	0,167	-0,040	0,140	0,194	0,155	0,004

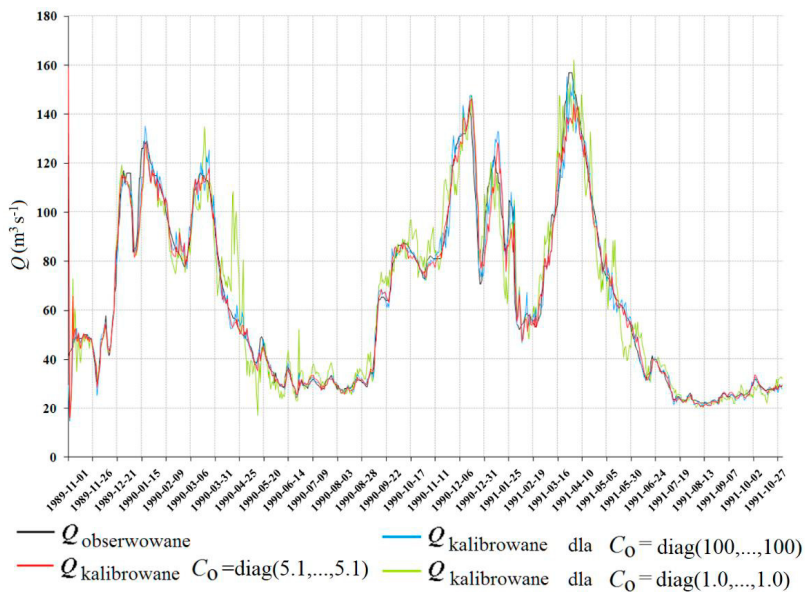
	728	0,775	-0,016	0,002	-0,004	0,000	1,343	0,034
	729	0,785	-0,018	0,002	-0,004	0,000	1,338	0,036
DWR	0	0,100	0,100	0,100	0,100	0,100	0,100	0,00001
	1	0,103	0,104	0,111	0,114	0,109	0,101	0,00082
	2	0,137	0,139	-0,046	0,164	0,196	0,113	0,003

	728	0,827	-0,015	-0,001	-0,001	0,000	1,232	0,034
	729	0,836	-0,016	-0,001	-0,001	0,001	1,212	0,035

Odpowiednie oszacowanie zbioru zmiennych stanu, inicjujących pracę obu DLMB (VCR DWR), pozwoliło na uzyskanie wysokiej stabilności modelowanych hydrogramów. W obu algorytmach zastosowano mechanizm filtru Kalmana. Filtr ten optymalizuje wartości *a priori* wariancji \mathbf{S}_τ w relatywnie krótkim czasie i stabilizuje modelowany hydrogram w kilku pierwszych krokach. Obserwowany i kalibrowany hydrogram modeli VCR i DWR pokazano na rysunkach 5.1 i 5.2. Kalibrację modeli osiąga się przy użyciu nieprecyzyjnej *a priori* macierzy kowariancji, czyli diagonalnej macierzy \mathbf{C}_0 z główną przekątną równą 100,0, 5,1 i 1,0. Takie ustawienia zapewniają szybszą konwergencję predykcyjnej wariancji (rys. 5.1 i 5.2). Powyższą własność VCR i DWR obserwuje się jedynie do początku kwietnia 1990 roku. Dla diagonalnej macierzy kowariancji \mathbf{C}_0 , z główną przekątną równą 5,1, wartości zmiennych uzyskane w ostatnim kroku kalibracji wykorzystano jako *a priori* dla pozostałej części zbioru danych, tj. w okresie walidacji. Jak zauważono (tab. 5.3), model DWR z *a priori* wariancją $\mathbf{S}_\tau = 0,035$ w chwili $\tau=729$ (ostatni krok kalibracji) wskazuje na nieco większą stabilność niż w modelu VCR z *a priori* wariancją $\mathbf{S}_\tau = 0,036$ (wartości *a priori* wariancji zmieniają się od $\mathbf{S}_0 = 0,00001$). W trakcie updatingu modeli VCR i DWR za pomocą filtru Kalmana obliczana jest wariancja \mathbf{S}_τ , która może zostać skorygowana [szczegóły algorytmu w rozdziale 3.5.1 i 3.5.2]. Zmodyfikowana wariancja \mathbf{S}_τ jest bezpośrednim wejściem do obliczania współczynników kierunkowych macierzy \mathbf{m}_τ , a zmiany ich wartości przyczyniają się do zwiększenia dokładności modelowanego hydrogramu, zwłaszcza w okresach wzbrań i niskich przepływów.

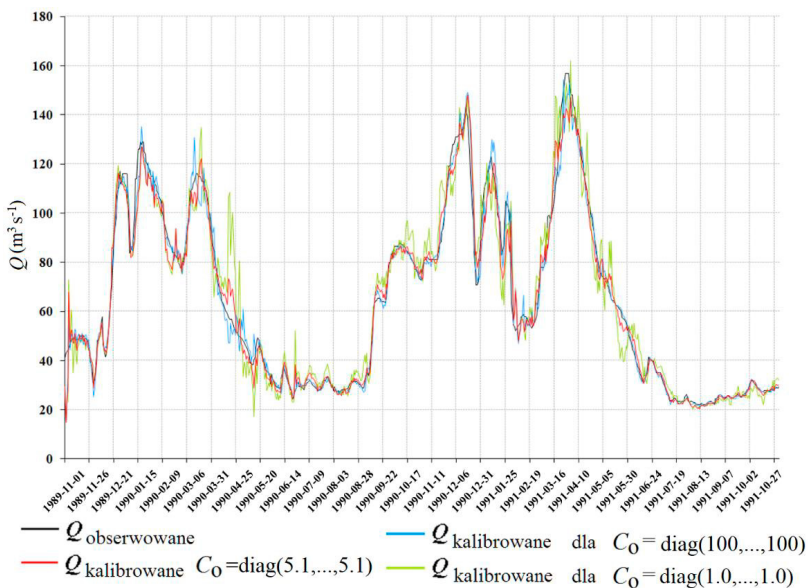
KALIBRACJA SIECI NEURONOWYCH NN

W przypadku sieci neuronowych (NN) zastosowano topologie, które używają zbioru zmiennych prognozujących w pełni połączonych z ukrytymi warstwami w stosunku 1:2 ($d + 1$) do ukrytych węzłów, gdzie d jest liczbą zmiennych prognozujących, a następnie testowano je pod kątem dokładnej symulacji rocznych hydrogramów przepływu. Spośród testowanych topologii sieci neuronowych, na podstawie uzyskanej optymalnej oceny wg błędu średniokwadratowego (MSE), do symulacji rocznego hydrogramu, a następnie walidacji, wybrano model NN o strukturze [5-1-1] [szczegóły w rozdziale 5.1]. Macierz zmiennych prognozujących skalibrowano pod kątem ukrytych połączeń neuronów (1-11). Optymalny model NN miał jeden ukryty neuron tangensoidalny w pełni połączony z wejściową i wyjściową warstwą sieci. W czasie kalibracji do krótkoterminowego prognozowania z wyprzedzeniem 1-, 2- i 3-dobowym użyto zoptymalizowany model sieci neuronowej, w celu uniknięcia niespójności w porównywaniu różnych modeli matematycznych.



Rys. 5.1. Kalibracja modelu VCR zlewni Górnej Narwi w profilu wodowskazowym Wizna w okresie 1990-1991 dla $\mathbf{C}_0 = \text{diag}(1,0, \dots, 1,0)$ lub $\text{diag}(5,1, \dots, 5,1)$ lub $\text{diag}(100, \dots, 100)$.

Fig. 5.1. Calibration of the VCR model for the Upper Narew River watershed at Wizna site using the data in calibration set 1990-1991 for $\mathbf{C}_0 = \text{diag}(1.0, \dots, 1.0)$ or $\text{diag}(5.1, \dots, 5.1)$ or $\text{diag}(100, \dots, 100)$.



Rys. 5.2. Kalibracja modelu DWR zlewni Górnej Narwi w profilu wodowskazowym Wizna w okresie 1990-1991 dla $\mathbf{C}_0 = \text{diag}(1,0, \dots, 1,0)$ lub $\text{diag}(5,1, \dots, 5,1)$ lub $\text{diag}(100, \dots, 100)$.

Fig. 5.2 Calibration of the DWR model for the Upper Narew River watershed at Wizna site using the data in calibration set 1990-1991 for $\mathbf{C}_0 = \text{diag}(1.0, \dots, 1.0)$ or $\text{diag}(5.1, \dots, 5.1)$ or $\text{diag}(100, \dots, 100)$.

5.2.1.1. Ocena modeli WLR, NN, VCR i DWR do symulacji rocznych hydrogramów

W tabeli 5.4 przedstawiono porównanie najlepszych modeli użytych do symulacji rocznego hydrogramu (podczas walidacji w okresie 1992-1995). Symulacja modelu DWR nieznacznie przewyższyła symulację modelu VCR. Wynika to z większej stabilności *a priori* wariancji S_τ modelu DWR w porównaniu do VCR. Modele WLR i NN [5-1-1] znalazły się odpowiednio na trzecim i czwartym miejscu. Mimo że WLR uzyskał dobrą ocenę symulacji, wielokrotna regresja liniowa nie ma możliwości uwzględniania nieliniowej dynamiki zmian przepływu, która wystąpiła w obserwowanym hydrogramie w okresie 1992-1995.

VCR i DWR skutecznie reprezentują *a priori* wariancję w ramach iteracyjnego updatingu zmieniających stanu modelu (czwarty krok *a posteriori* w chwili τ), istotnie aktualizują parametry stanu i tym samym zmniejszają niepewność związaną z ich prognozowaniem w każdym dyskretnym kroku przetwarzania. Natomiast modele NN i WLR wyszukują tylko taką postać parametryczną, która minimalizuje błąd kwadratowy. Przy każdej minimalizacji modeli NN istnieje niebezpieczeństwo uzyskania wyniku minimum lokalnego a nie globalnego.

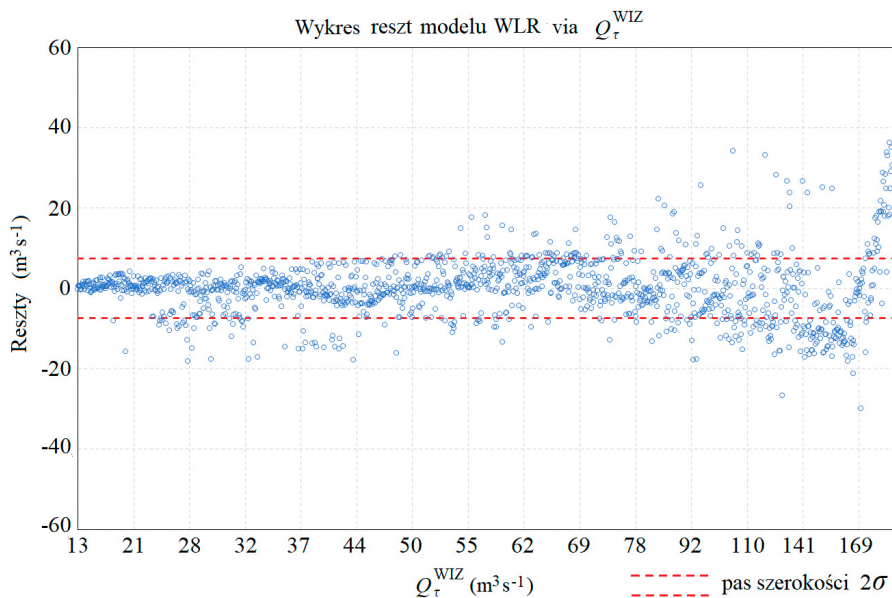
Tabela 5.4. Wyniki sprawdzenia poprawności walidacji hydrogramów modeli WLR, NN, VCR i DWR w okresie 1992-1995.

Table 5.4. Performance statistics for MLR, NN, VCR and DWR models for validation hydrographs in the period 1992-1995.

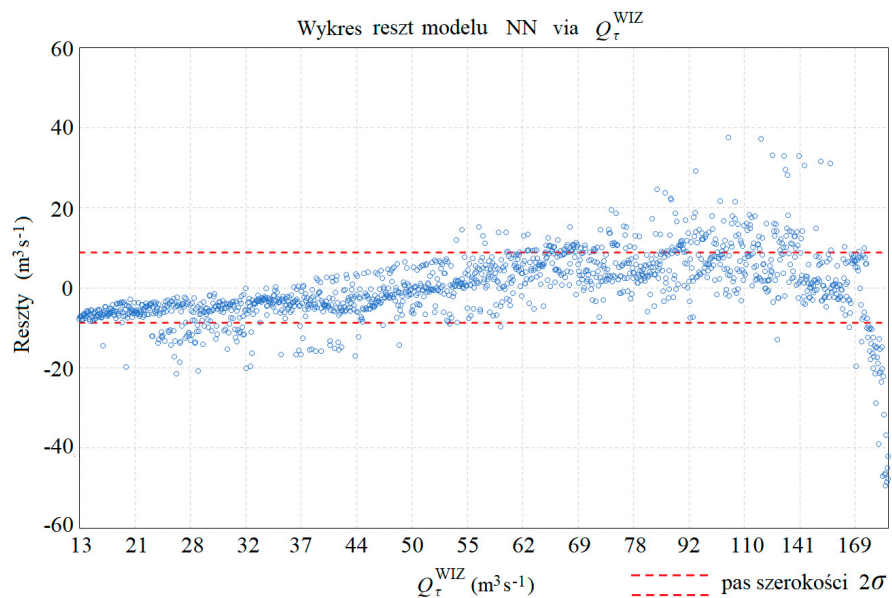
Lata		Modele	MAD	MAE	MSE	RoM	RMSE	ENS
1992-1995	walidacja	WLR	39,317	5,027	55,004	1,00	7,416	0,9780
		NN	80,596	6,415	128,962	0,97	11,356	0,9484
		VCR	60,493	3,531	37,481	1,00	6,122	0,9850
		DWR	55,066	2,586	24,001	1,00	4,899	0,9904

Na rysunkach 5.3-5.6 przedstawiono graficzną analizę reszt modeli WLR, NN, VCR i DWR. Wykresy wizualizują zależność między resztami modeli a wartościami obserwowanymi Q_τ^{WIZ} . Większość punktów (niebieskie koła) mieści się wewnątrz pasa (czerwona przerywana linia) o szerokości dwóch odchyłeń standardowych 2σ . Najwięcej reszt, e_τ , oddalonych o więcej niż 2σ zaobserwowano dla modelu NN (rys. 5.4). Na wykresie reszt modelu VCR (rys. 5.5) zaobserwowano duże zmiany ich wielkości. Jest to związane z dużymi skokami (zmianami dobowymi) przepływów, rzędu 42 i 65 m³s⁻¹, zaobserwowanymi w profilu wodowskazowym Wizna w okresie 1992-1995. W modelu VCR nie występuje stały współczynnik dyskontowy, δ , którego wartość wpływa na aktualizację macierzy kowariancji stanu C_τ i tym samym na dokładność symulacji przepływu. Wpływ współczynnika dyskontowego na dokładność symulacji i tym samym na analizę reszt jest szczególnie widoczny na wykresie reszt modelu DWR (rys. 5.6), gdzie reszty dla przepływów rzędu 42 i 65 m³s⁻¹ mieszczą się praktycznie wewnątrz pasa szerokości dwóch odchyłeń standardowych 2σ .

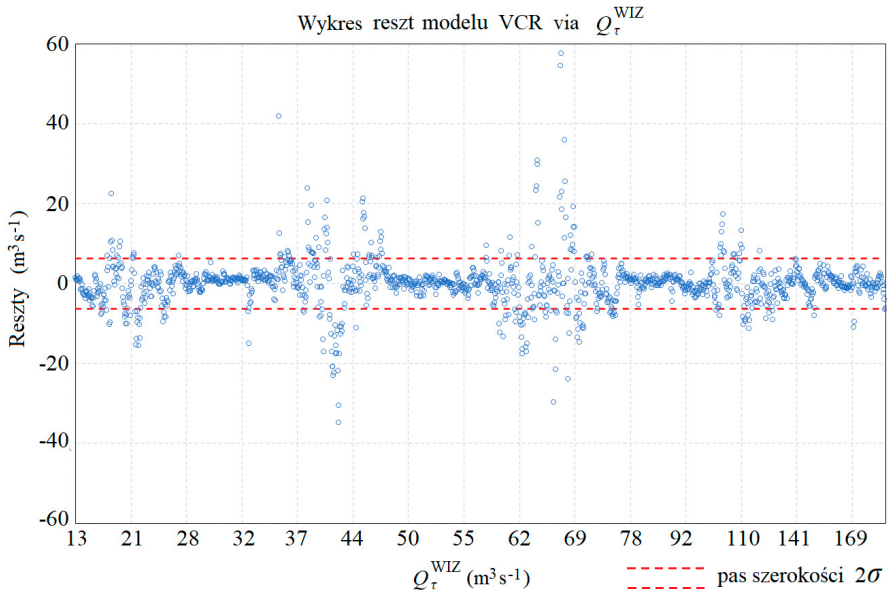
W kolejnym kroku analizy sprawdzono normalność rozkładu reszt modeli WLR, NN, VCR i DWR via Q_τ^{WIZ} za pomocą testu W Shapiro-Wilka [Royston 1992] oraz autokorelację reszt przy użyciu testu Durбина-Watsona [Durbin, Watson 1971]. Uzyskane wyniki przedstawiono w tabeli 5.5. Obliczone *p*-wartości testu W Shapiro-Wilka są wyższe od przyjętego w badaniu poziomu istotności $\alpha = 0,05$, co oznacza, że nie ma podstaw do odrzucenia hipotezy o normalności rozkładu reszt. Duże różnice między rozkładem reszt a rozkładem normalnym mogłyby zaburzać ocenę istotności współczynników pomiędzy zmiennymi modelu. Należy stwierdzić, że szereg czasowy reszt ma rozkład zbliżony do rozkładu normalnego i stanowi biały szum. Statystyki Durбина-Watsona porównano do dwóch wartości d_l i d_g , oznaczających przedział wyników (odpowiednio dolną i górną granicę), dla których nie można stwierdzić, czy zachodzi bądź nie zachodzi autokorelacja reszt. Ponieważ wartości statystyki testowej są niższe od dolnej granicy przedziału, d_l , należy stwierdzić, że reszty we wszystkich modelach są dodatnio skorelowane.



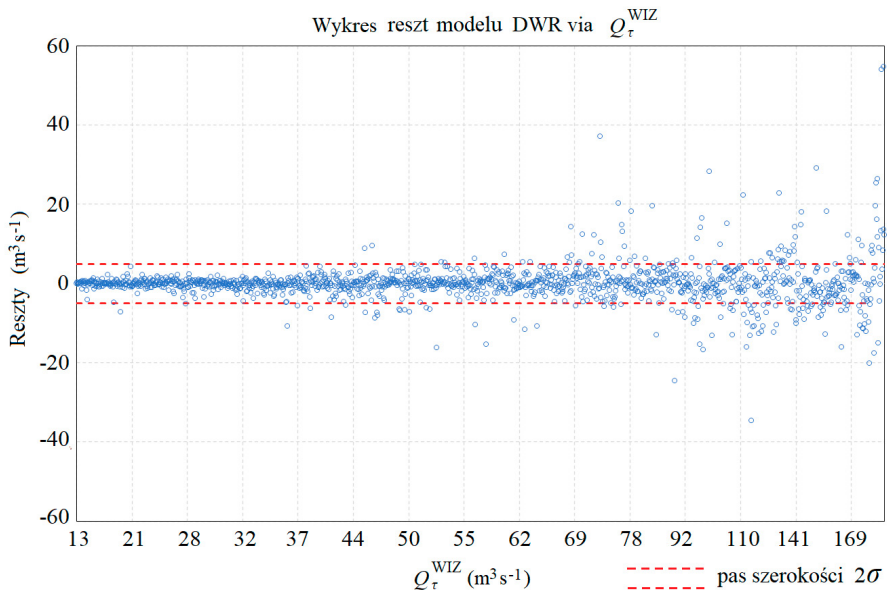
Rys. 5.3. Wykres reszt modelu WLR via dobowy przepływ Q_{τ}^{WIZ} w zlewni Górnej Narwi w profilu wodowskazowym Wizna dla danych walidacyjnych z okresu 1992-1995 z pasem szerokości dwóch odchyłeń standardowych.
Fig. 5.3. Chart of residuals of MLR model for daily discharge Q_{τ}^{WIZ} in the Upper Narew River watershed at Wizna site for validation data set 1992-1995 and the band with a width of two standard deviation.



Rys. 5.4. Wykres reszt modelu NN via dobowy przepływ Q_{τ}^{WIZ} w zlewni Górnej Narwi w profilu wodowskazowym Wizna dla danych walidacyjnych z okresu 1992-1995 z pasem szerokości dwóch odchyłeń standardowych.
Fig. 5.4. Chart of residuals of NN model for daily discharge Q_{τ}^{WIZ} in the Upper Narew River watershed at Wizna site for validation data set 1992-1995 and the band with a width of two standard deviation.



Rys. 5.5. Wykres reszt modelu VCR via dobowy przepływ Q_{τ}^{WIZ} w zlewni Górnej Narwi w profilu wodowskazowym Wizna dla danych walidacyjnych z okresu 1992-1995 z pasem szerokości dwóch odchyłeń standardowych.
Fig. 5.5. Chart of residuals of VCR model for daily discharge Q_{τ}^{WIZ} in the Upper Narew River watershed at Wizna site for validation data set 1992-1995 and the band with a width of two standard deviation.



Rys. 5.6. Wykres reszt modelu DWR via dobowy przepływ Q_{τ}^{WIZ} w zlewni Górnej Narwi w profilu wodowskazowym Wizna dla danych walidacyjnych z okresu 1992-1995 z pasem szerokości dwóch odchyłeń standardowych.
Fig. 5.6. Chart of residuals of VCR model for daily discharge Q_{τ}^{WIZ} in the Upper Narew River watershed at Wizna site for validation data set 1992-1995 and the band with a width of two standard deviation.

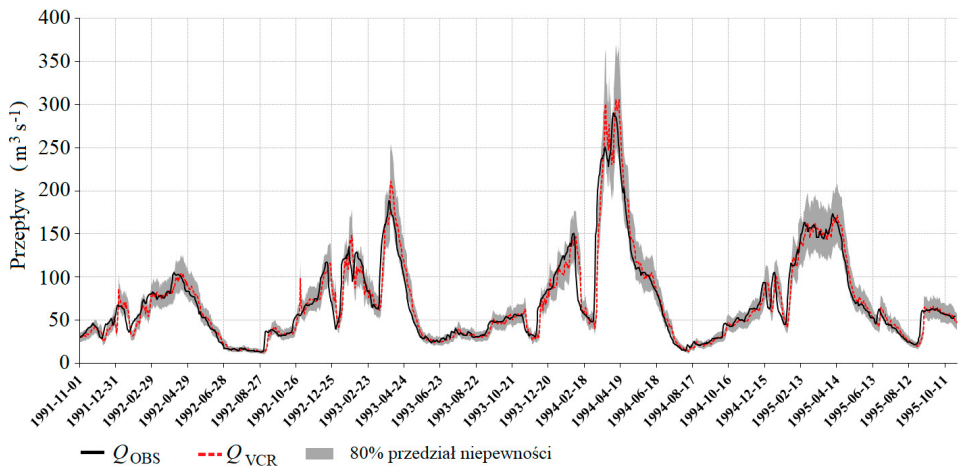
Tabela 5.5. Sprawdzenie normalności reszt modeli WLR, NN, VCR i DWR w okresie 1992-1995 za pomocą testu W Shapiro-Wilka, przy poziomie istotności $\alpha = 0,05$, oraz autokorelacji wśród reszt za pomocą testu Durbin-Watsona.

Table 5.5. Checking the normality of the rest of the models MLR, NN, VCR and DWR models for validation hydrographs in the period 1992-1995 using W Shapiro-Wilk test and autocorrelation of residuals using the Durbin-Watson test.

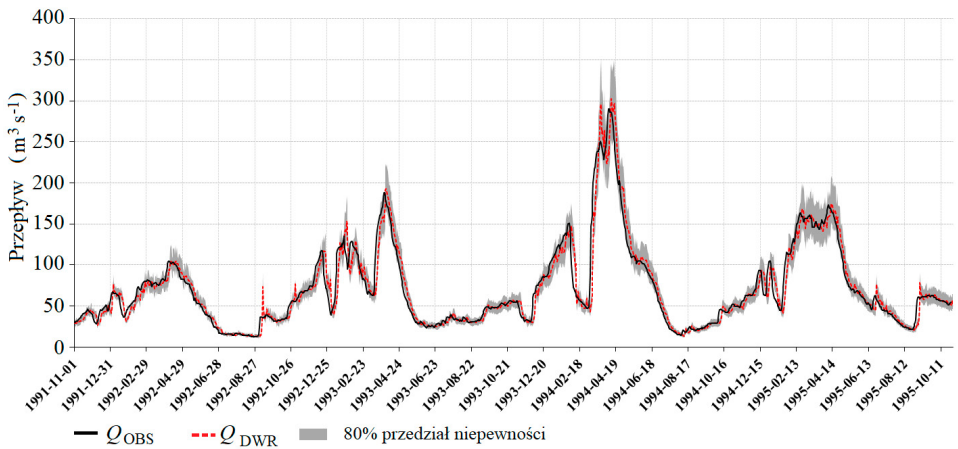
Lata		Modele	Statystyka W Shapiro-Wilka	p -wartość	Statystyka Durbin-Watsona $d_i = 1,91$ i $d_g = 1,92$ dla $n = 1462$
1992-1995	walidacja	WLR	0,8356	0,0521	0,5423
		NN	0,8146	0,0498	0,5312
		VCR	0,8478	0,0711	0,8713
		DWR	0,8427	0,0613	0,7642

Na rysunkach 5.7-5.8 porównano hydrogramy obserwowane (czarna ciągła linia) z symulowanymi (czerwona przerywana linia), odpowiednio dla modelu VCR i DWR, w okresie walidacji 1992-1995. Jasnoszare obszary wskazują 80% przedział niepewności symulacji wyznaczonej przy użyciu gęstości *a posteriori* w czwartym kroku algorytmu [szczegóły w paragrafie 3.5.1 i 3.5.2]. Większą dokładność modelowanych hydrogramów za pomocą DWR (w porównaniu do VCR) na etapie walidacji, szczególnie w okresie wezbrania deszczowo-roztopowego od kwietnia do maja i w okresie niskich przepływów od czerwca do września (odpowiednio rys. 5.7 i 5.8), można przypisać mniejszej wartości *a priori* wariancji obliczonej w ostatnim kroku kalibracji (1990-1991), gdzie wartości S_r dla modeli VCR i DWR były równe odpowiednio 0,036 i 0,035. W okresach niżówkowych codzienne przepływy Q_{OBS} są stabilne (brak gwałtownych zmian wartości przepływów, zbliżone amplitudy wahań dobowych), co wpływa na stabilizację wariancji S_r i na stabilność współczynników kierunkowych w obu modelach, tj. VCR i DWR, a to z kolei sprzyja uzyskiwaniu większej dokładności modelowanych hydrogramów. Inaczej wygląda sytuacja w okresach wezbraniowych, w których występują duże amplitudy codziennych przepływów. W modelu DWR, w porównaniu do VCR, pojawia się stały współczynnik dyskontowy δ [szczegóły w rozdziale 3.5.2], którego wartość wpływa na aktualizację macierzy kowariancji stanu C_r i tym samym na dokładność symulacji przepływu. West i Harrison [1999] sugerowali, że w modelach dynamicznych w przestrzeni stanu współczynnik dyskontowy powinien mieścić się w przedziale $\delta \in (0,8; 0,99]$, ponieważ w przeciwnym wypadku macierz transformacji kowariancji W_r (3,50) będzie miała bardzo duże wartości, dając niestabilny model z bardzo dużymi wartościami macierzy kowariancji stanu C_r . Z drugiej strony wysoki współczynnik dyskontowy δ nie pozwala na dokładną prognozę dużych skoków (zmian dobowych) przepływów. Według Triantafyllopoulosa [2006] dla różnych wartości δ różnice w C_r nie wpływają dramatycznie na macierz kowariancji i tym samym pozwalają na większą elastyczność modelowania. W przypadku modelowania całorocznego hydrogramu należy znaleźć optymalną wartość δ (w przytoczonym eksperymencie $\delta = 0,96$), natomiast do modelowania okresów wezbraniowych można byłoby zastosować niższe wartości δ .

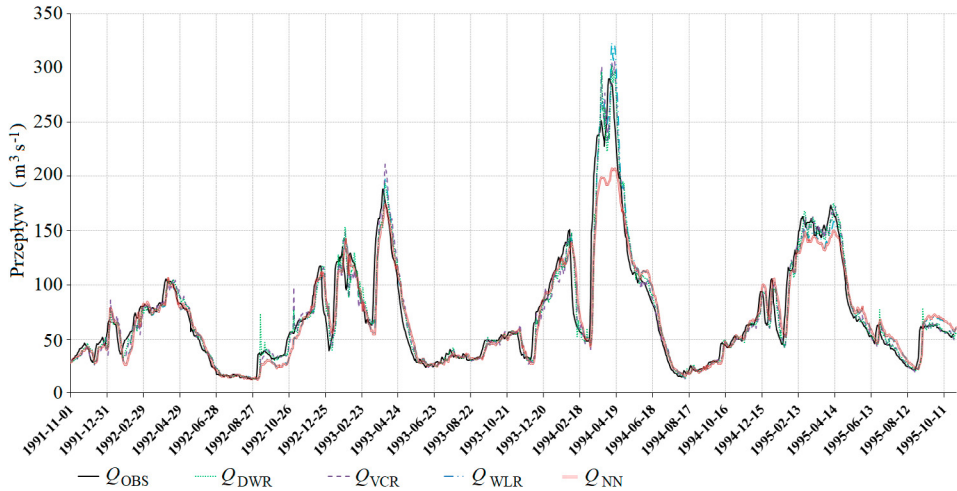
Z analizy ilościowej (tab. 5.4) i jakościowej (rys. 5.9) wynika, że najlepszymi modelami są (w kolejności od najlepszych do najgorszych): DWR > VCR > MLR > NN. Rozwiązania DWR i VCR okazały się najmniej zawodnymi modelami, tzn. uzyskane wyniki symulacji w najmniejszym stopniu odbiegały od zaobserwowanego hydrogramu w porównaniu do modeli WLR i NN. Skalibrowane parametry stanu modeli DWR i VCR lepiej odzwierciedlają dynamikę przepływu rzeczno, zwłaszcza podczas wezbrań. Model NN (czerwona ciągła linia) nie był w stanie osiągnąć zbliżonych do obserwacji wartości przepływów (czarna ciągła linia) z powodu przeszacowania danych kalibracyjnych i niedoszacowania danych walidacyjnych. Rozwiązania DWR i VCR również przewyższają symulacje uzyskane z najlepszego dopasowania zmiennej prognozowanej do skalibrowanej zmiennej prognozowanej za pomocą modelu WLR.



Rys. 5.7. Walidacja modelu VCR.
 Fig. 5.7. Validation of the VCR model.



Rys. 5.8. Walidacja modelu DWR.
 Fig. 5.8. Validation of the DWR model.



Rys. 5.9. Porównanie walidacji modeli DWR, VCR, WLR i NN.
Fig. 5.9. Comparison of the DWR, VCR, MLR and NN validation models.

5.2.1.2. Ocena modeli WLR, NN, VCR i DWR do prognozowania z wyprzedzeniem 1-, 2- i 3-dobowym

Zasadniczym celem modelu WLR jest poznanie zależności między zmiennymi prognozującymi a zmienną prognozowaną. Stosowanie WLR do celów prognostycznych zakłada, że:

- 1) dobrze skalibrowany model (zbiór statycznych współczynników regresji) pozostanie aktualny w przyszłości;
- 2) prognozowane zjawisko charakteryzuje się dużą bezwładnością i tym samym większe ilościowe zmiany nie wystąpią w krótkim odstępie czasu.

Tabela 5.6. Wyniki sprawdzenia poprawności krótkoterminowej prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym, obliczonej za pomocą modeli WLR, NN, VCR i DWR w 1996 roku.

Table 5.6. Performance statistics for MLR, NN, VCR and DWR models for 1-, 2- and 3- day lead time forecasting hydrographs in 1996.

Rok		Model	Prognoza	MAD	MAE	MSE	RMSE	E_{NS}
1996	prognozowanie	WLR	$Q + 1$	104,326	8,243	324,335	17,935	0,9571
			$Q + 2$	144,138	9,884	511,550	22,524	0,9323
			$Q + 3$	160,612	11,338	716,196	26,651	0,9052
		NN	$Q + 1$	160,154	11,626	523,124	22,872	0,9301
			$Q + 2$	197,611	12,410	667,510	25,836	0,9108
			$Q + 3$	215,611	13,288	839,944	28,982	0,8878
		VCR	$Q + 1$	84,951	3,178	70,980	8,390	0,9906
			$Q + 2$	92,211	4,115	120,050	10,911	0,9841
			$Q + 3$	126,383	5,249	212,128	14,504	0,9719
		DWR	$Q + 1$	72,462	3,915	71,811	8,439	0,9905
			$Q + 2$	87,647	4,611	109,461	10,419	0,9855
			$Q + 3$	105,135	5,238	152,801	12,310	0,9797

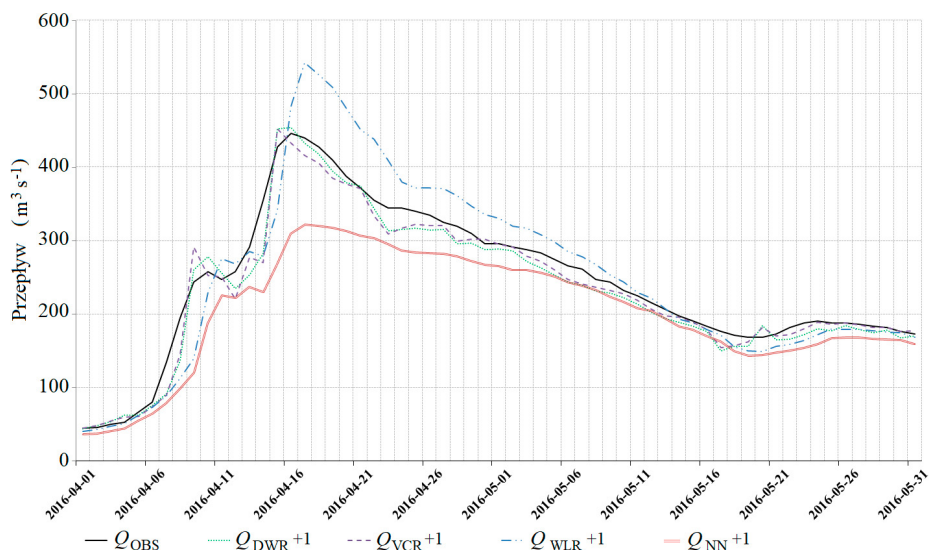
Jednak powyższe założenia są często nieprawdziwe, szczególnie w kontekście wzebrań (powodzi), czyli zjawisk o wysokiej dynamice zmian ich charakterystyk. Sieci neuronowe poszukują przestrzeni parametrów w celu odnalezienia rozwiązania, które minimalizuje błąd modelu. Rezultatem

minimalizacji nie zawsze są zbiory parametrów, które doskonale wyjaśniają modelowane zjawisko i nie uwzględniają zmian wariancji modelu w czasie. Natomiast modele VCR i DWR opierają się na dynamicznych zmianach stanu, a równanie regresji dynamicznie uwzględnia zbiór czynników wpływających na prognozowany hydrogram.

Modele WLR, NN, VCR i DWR zostały opracowane i użyte do prognozowania przepływu rzeki z wyprzedzeniem 1-, 2- i 3-dobowym. W algorytmach modeli symulacyjnych VCR i DWR uwzględniono proces prognozowania (tj. 5. krok opisany w rozdziale, odpowiednio 3.5.1 i 3.5.2 i zaznaczony na rys. 3.1 i 3.2), model WLR wykorzystał trzy oddzielne równania otrzymane w trakcie kalibracji do prognozy z wyprzedzeniem 1-, 2- i 3-dobowym, natomiast oszacowany model NN został użyty do prognozy przepływu rzecznego dla wszystkich czasów wyprzedzenia. Wyniki najlepszych modeli, na podstawie miar statystycznych MAD, MAE, MSE, RMSE i E_{NS} , przedstawiono w tabeli 5.6.

ANALIZA PORÓWNAWCZA CZTERECH METOD PROGNOSTYCZNYCH Z WYPRZEDZENIEM 1-DOBOWYM

Na rysunku 5.10 porównano obserwowany i prognozowany z wyprzedzeniem 1-dobowym hydrogramy WLR, NN [5-1-1], VCR i DWR w czasie 2 miesięcy, tj. od 1 kwietnia do 31 maja 1996 roku. Przyjęty do analizy zakres czasu pozwala na dokładną graficzną analizę porównawczą, szczególnie w czasie największego zaobserwowanego w tym okresie wezbrania w zlewni Górnej Narwi. Średnie i wysokie przepływy są bardzo dokładnie prognozowane przez modele VCR i DWR (najwyższe wartości E_{NS} uzyskane dla prognozy z wyprzedzeniem 1-dobowym, tab. 5.6). Zauważono, że zarówno VCR, jak i DWR nieznacznie przeszacowały przepływ w okresie do kulminacji, a następnie lekko niedoszacowały prognozowane hydrogramy w fazie opadania i stabilizacji (rys. 5.10).



Rys. 5.10. Porównanie obserwowanego i prognozowanego dobowego hydrogramu w zlewni Górnej Narwi w profilu wodowskazowym Wizna z 1-dobowym wyprzedzeniem w okresie od 1 kwietnia do 31 maja 1996 roku otrzymanym za pomocą modeli DWR, VCR, WLR i NN.

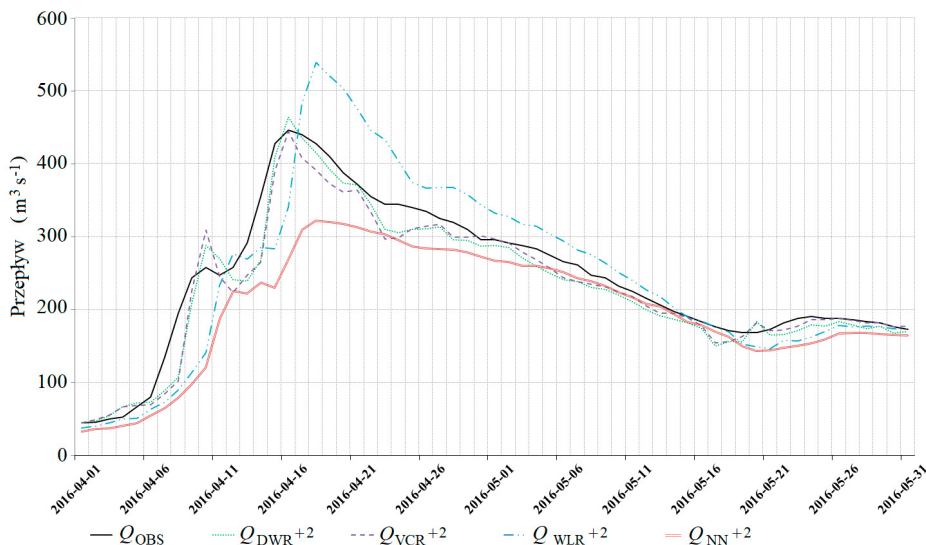
Fig. 5.10. Comparison of observed and forecasted daily hydrograph for the Upper Narew River watershed at Wizna site for 1-day lead time in the period from 1 April to 31 May 1996 using DWR, VCR, NN and MLR models.

Dokładne oszacowanie *a priori* wariancji S_r w modelach VCR i DWR zapewniło dokładność zmiennych stanu i znacznie zwiększyło wydajność modeli w okresie progностycznym, w którym zaobserwowano najwyższe przepływy i najsilniejszą hydrologiczną dynamikę w badanej zlewni. Mniejsza wartość S_r w DWR umożliwiła nieznacznie dokładniejsze wyjaśnienie wariancji modelu, tym samym DWR uzyskał najniższe ze wszystkich analizowanych modeli maksymalne odchylenie prognozy (MAD) (tab. 5.6).

Po porównaniu modeli WLR i NN okazało się, że dodatkowe parametry występujące w NN (w modelu NN są o dwa parametry więcej niż w WLR) mają tendencję do nadmiernej parametryzacji oszacowania procesu przepływu rzeki i generują dodatkowe szумы do prognozy, co prowadzi do pogorszenia wydajności w porównaniu do WLR. Sieci NN opierają swoje prognozy na założeniu posiadania wystarczającej liczby danych kalibracyjnych niezbędnych do poznania modelowanego zjawiska. Jeżeli podczas zdarzeń prognostycznych występuje dynamika nie opisana w danych kalibracyjnych, modele NN nie są w stanie wyodrębnić takich zdarzeń w prognozie. Lepsze dopasowanie obserwacji do symulacji modelu WLR w trybie prognozy wynika z uproszczonej formy parametrycznej WLR (6 parametrów) w porównaniu do modelu NN (8 parametrów).

ANALIZA PORÓWNAWCZA CZTERECH METOD PROGNOSTYCZNYCH Z WYPRZEDZENIEM 2-DOBOWYM

Wydajność wszystkich modeli uruchomionych w trybie prognozy z wyprzedzeniem 2-dobowym zmniejszyła się. Jeśli wziąć pod uwagę uzyskane miary statystyczne (tab. 5.6), najlepszym modelem prognostycznym okazał się model DWR, a następnie VCR, WLR i NN [5-1-1]. Kształty hydrogramów uzyskanych za pomocą VCR i DWR są bardzo podobne do siebie i zdecydowanie różnią się od hydrogramów WLR i NN (rys. 5.11). W modelach WLR i NN kulminacja wezbrania opóźnia się o dwa dni w porównaniu do rzeczywistej kulminacji, którą zaobserwowano 16 kwietnia 1996 roku. Co więcej, wszystkie modele wykazują przesunięcie w prawo wznoszącej gałęzi prognozowanego hydrogramu i przesunięcie w lewo gałęzi opadającej (tzn. zawężenie kulminacji hydrogramu). Uzyskane w VCR i DWR wyższe oceny symulacji, w porównaniu do modeli NN i WLR, pokazują korzyści wynikające z użycia tego paradygmatu modelowania. Wynika to z faktu, że *a priori* wariancja zastosowana do updatingu zmiennych stanu modelu VCR i DWR jest niezwykle istotna w obliczaniu prognoz, zwłaszcza w kontekście wydłużającego się horyzontu prognozy. Model DWR pokazuje, że większa dokładność w oszacowaniu *a priori* wariancji S_t może istotnie polepszyć dokładność prognozy w porównaniu do modelu VCR.



Rys. 5.11. Porównanie obserwowanego i prognozowanego dobowego hydrogramu w zlewni Górnej Narwi w profilu wodowskazowym Wizna, z 2-dobowym wyprzedzeniem w okresie od 1 kwietnia do 31 maja 1996 r., otrzymanym za pomocą modeli DWR, VCR, WLR i NN.

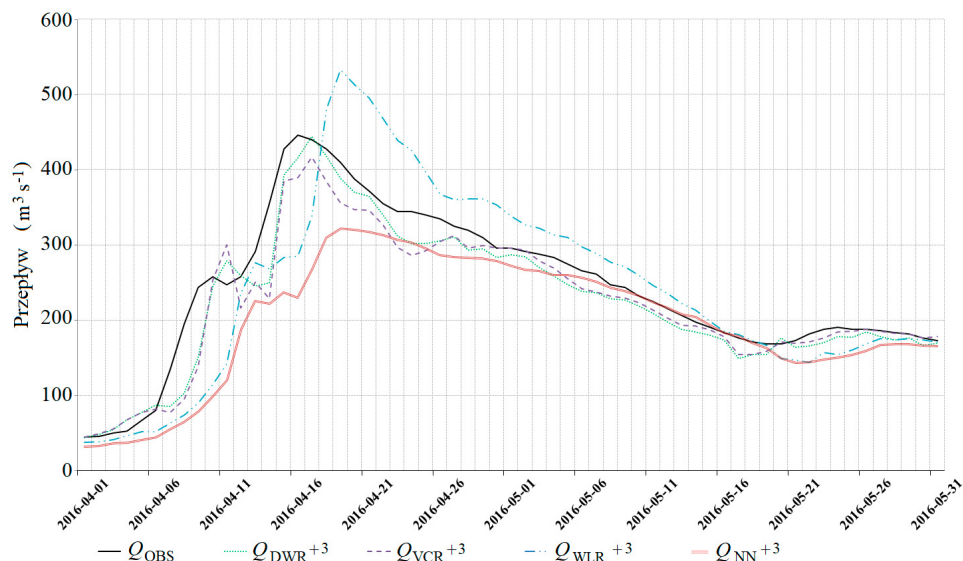
Fig. 5.11. Comparison of observed and forecasted daily hydrograph for the Upper Narew River watershed at Wizna site for 2-days lead time in the period from 1 April to 31 May 1996 using DWR, VCR, MLR and NN models.

Podobnie jak w przypadku prognozy z wyprzedzeniem 1-dobowym, model WLR wykazuje, w porównaniu do modelu NN, lepsze dopasowanie obserwacji do prognozy z wyprzedzeniem 2-dobowym. Wynika to z wyższej parametryzacji modelu NN, która nieliniowo odwzorowuje (w przestrzeni

wysokowymiarowej) i nadmiernie komplikuje liniową zależność występującą między zmiennymi prognozującymi a zmienną prognozowaną $Q_{\tau+2}^{WIZ}$.

ANALIZA PORÓWNAWCZA CZTERECH METOD PROGNOSTYCZNYCH Z WYPRZEDZENIEM 3-DOBOWYM

Zgodnie z oczekiwaniami, wraz ze wzrostem czasu wyprzedzenia prognozy z dwóch do trzech dób wydajność wszystkich modeli obniżyła się, niemniej DWL jest wciąż lepszy od VCR, WLR i NN [5-1-1]. Obserwowane i prognozowane hydrogramy pokazują, że prognozowany przepływ obliczony za pomocą NN, VCR i DWL był niedoszacowany w czasie średniego i niskiego przepływu, natomiast przepływ uzyskany z modelu WLR był przeszacowany w czasie średniego i wysokiego przepływu (rys. 5.12). Ponadto zauważalne jest przesunięcie w prawo rosnącej gałęzi hydrogramu i przesunięcie w lewo gałęzi opadającej (tzn. zawężenie kulminacji hydrogramu). W przypadku VCR i DWL kulminacja wezbrania opóźnia się o dwie doby, natomiast dla modeli WLR i NN opóźnienie w stosunku do rzeczywistej kulminacji (tj. 16 kwietnia 1996 r.) wynosi trzy doby. Objętość fali została wyraźnie przeszacowana przez model WLR i niedoszacowana przez NN. VCR i DWL pokazują, że dobre dopasowanie wartości *a priori* wariancji S_{τ} jest w stanie zapewnić dokładne parametry stanu i tym samym rzetelnie opisać dynamikę Q_{τ}^{WIZ} . Mniejsza *a priori* wariancja S_{τ} związana z modelem DWL zapewnia większą precyzję prognozowania w porównaniu z modelem VCR z większym oszacowaniem S_{τ} .



Rys. 5.12. Porównanie obserwowanego i prognozowanego dobowego hydrogramu w zlewni Górnej Narwi w profilu wodowskazowym Wizna, z 3-dobowym wyprzedzeniem w okresie od 1 kwietnia do 31 maja 1996 r., otrzymanym za pomocą modeli DWL, VCR, WLR i NN.

Fig. 5.12. Comparison of observed and forecasted daily hydrograph for the Upper Narew River watershed at Wizna site for 3-days lead time in the period from 1 April to 31 May 1996 using DWL, VCR, MLR and NN models.

Ogólnie rzecz biorąc, najlepszym modelem do symulacji rocznych hydrogramów przepływu rzecznej i do krótkoterminowej prognozy hydrologicznej z wyprzedzeniem 1-, 2- i 3-dobowym jest model DWL, a następnie VCR, WLR i NN. Model WLR miał tendencję do przeszacowania rocznego hydrogramu w obszarach, w których model NN niedoszacowywał hydrogramu, a powyższa charakterystyka obu modeli była również uwidoczniła w krótkoterminowych prognozach. W okresach niskiego przepływu modele WLR i NN reagują zgodnie z hydrologiczną dynamiką rzeki. Modele VCR i DWL efektywnie szacowały kształt i zmieniającą się dynamikę rocznego hydrogramu i krótkoterminowych prognoz i są najdokładniejszymi z badanych modeli w tej pracy.

5.2.2. Podsumowanie i wnioski

Zakładając, że przepływy rzeczne są obserwowane i rejestrowane w regularnych odstępach czasu i że ten sam błąd pomiaru jest związany z każdą obserwacją, każda nowa obserwacja może być wykorzystana do prognozowania tego zjawiska w przyszłości. W przypadku modelowania systemów hydrologicznych często brane są pod uwagę czynniki, które wpływają na procesy fizyczne opisane za pomocą pomiarów i obserwacji. W takich sytuacjach ważne jest, aby uwzględnić zmiany stanu modelu. W obu algorytmach dynamicznych modeli, tj. VCR i DWR, zastosowano mechanizm uczenia w postaci filtra Kalmana, który umożliwia updating zmiennych stanu. Algorytm pozwala na zapisywanie zmiennych stanu w każdym kroku przetwarzania, a to z kolei umożliwia uruchomienie modelu z parametrami obliczonymi w poprzednim kroku. Połączenie dynamicznych liniowych modeli z filtrem Kalmana odróżnia VCR i DWR od WLR i NN.

Porównując modele VCR i DWR z WLR i NN należy podkreślić, że VCR, DWR i NN potrzebują tylko jednego równania do symulacji rocznego hydrogramu i do prognozy z wyprzedzeniem 1-, 2- i 3-dobowym. Proces prognozowania w modelach VCR i DWR jest wyprowadzony jako rozkład prawdopodobieństwa w piątym kroku obu algorytmów w taki sposób, że tylko jedno równanie musi być aktualizowane w celu otrzymania walidowanego rocznego hydrogramu i prognozy.

Model NN używa parametrów skalibrowanego modelu i stosuje te parametry do walidacji zarówno rocznego hydrogramu, jak i prognozy z wyprzedzeniem 1-, 2- i 3-dobowym. W odróżnieniu do NN, technika WLR wymaga jednego równania do symulacji rocznego hydrogramu i dodatkowo trzech równań do prognozy z wyprzedzeniem 1-, 2- i 3-dobowym.

Głównymi różnicami między modelami VCR, DWR, WLR i NN jest to, że WLR i NN nie uwzględniają ewolucji wariancji połączonej z dynamiczną interakcją zmiennych prognozujących w czasie. Modele WLR i NN działają przy założeniu, że system fizyczny, który parametryzują (np. zlewnia hydrologiczna) cechuje wysoka bezwładność, tzn. że istotne zmiany modelowanych ciągów pomiarowo-obserwacyjnych wymagają długiego czasu. W związku z tym oba modele (WLR i NN) poszukują sparametryzowanej postaci, która minimalizuje średni błąd kwadratowy. W pracy założono, że oszacowane w czasie kalibracji parametry WLR i NN są optymalne do opisu modelowanego systemu w całym okresie analizy. Kluczowym problemem i przyczyną, dla której modele VCR i DWR osiągają lepsze wyniki od modeli WLR i NN jest to, że biorą pod uwagę dynamiczne zmiany współczynników kierunkowych zmiennych prognozujących i w ten sposób zapewniają adaptacyjność, niezawodność i dokładność modelu prognostycznego. Takie modele są w stanie uchwycić zmieniającą się dynamikę modelowanego zjawiska w czasie, co jest cechą nieuwzględnianą przez WLR i NN. W modelu DWR, w porównaniu do VCR, pojawia się dodatkowo współczynnik dyskontowy δ , który w zależności od przyjmowanej wartości umożliwia większą precyzję modelowania i prognozowania dużych skoków w obserwowanych ciągach pomiarowych.

Przeprowadzony eksperyment wykazał, że rozwiązania VCR i DWR zapewniły dokładną roczną symulację hydrogramów, a także prognozy z wyprzedzeniem 1-, 2- i 3-dobowym, szczególnie w czasie wezbrań i niskich przepływów. Techniki VCR i DWR są nowym krokiem w prognozowaniu hydrologicznym, w związku z tym jest wiele obszarów potencjalnego ich zastosowania. Na przykład w algorytmach VCR i DWR można uwzględnić czynnik sezonowy reprezentujący funkcję cykliczności prognozy [West, Harrison 1999]. W zastosowaniach hydrologicznych cykliczne (okresowe) zachowania są uwzględniane w wielu danych obserwacyjno-pomiarowych związanych z systemami hydrologicznymi [Belayneh i in. 2014]. Wzorce sezonowe mogą mieć zastosowanie w analizie zmienności odpływu bezpośredniego, odpływu wody gruntowej itp. Powyższe sezonowe wzorce łatwiej byłoby uwzględnić w analizie falkowej (wavelet analysis) [Spis użytych pojęć i stosowane nazewnictwo, pkt. 44], która umożliwia lokalizację i wydzielenie pewnych określonych zdarzeń częstotliwościowych, co mogłoby zwiększyć dokładność modelu hybrydowego WAV-DLMB.

W kolejnym etapie badań należy ocenić możliwość zastosowania modeli VCR i DWR do symulacji i prognozowania przepływów nie tylko w pojedynczym przekroju wodowskazowym, ale również w profilu wzdłużnym koryta rzeki i węźle hydrologicznym. W obliczu rosnącej niepewności, na przykład wynikającej ze zmian klimatycznych [Adamowski, Prokoph 2013], modele przedstawione w tym badaniu mogą ułatwić przejście do bardziej zintegrowanego i adaptacyjnego zarządzania zasobami wodnymi [Halbe i in. 2013, 2014].

6. GENEROWANIE WIĄZKOWYCH KRÓTKOTERMINOWYCH PROGNOZ PRZEPŁYWU RZECZNEGO Z WYKORZYSTANIEM KOREKcji I ASYMLACJI OBSERWACJI SATELITARNYCH

Jednym z głównych celów badania jest poprawienie krótkoterminowych prognoz hydrologicznych poprzez wzrost efektywności procedury updatingu. W tym celu skonstruowano preprocesor danych wejściowych do modelu HBV [Lindström i in. 1997]. Nowością w zaproponowanym podejściu jest:

- 1) Rozwój metody transformaty dystrybuanty (korekcji obciążenia, BC) z uwzględnieniem trzech trójparametrowych teoretycznych rozkładów prawdopodobieństwa, tj. uogólnionego rozkładu wykładniczego (GE), gamma (GA) i Weibulla (WE). Transformacja statystyczna obliczana jest dla takich samych rozkładów prawdopodobieństwa (np. GA-GA, GE-GE itd.) lub dla różnych (np. GA-GE, WE-GA itd.) w celu utworzenia wolnych od błędów systematycznego pomiarów satelitarnego opadu i wilgotności gleby.
- 2) Zastosowanie korekcji obciążenia (BC) do tworzenia wiązki zaburzonych prognozowanych zmiennych stanu modelu HBV nieobciążonych błędem systematycznym. Metodę BC użyto w zmodyfikowanych filtrach wiązkowych, tj. EnKF-BC i EnSRF-BC, zaimplementowanych w środowisku informatycznym modelu HBV.
- 3) Połączenie algorytmu deterministycznej 4D wariacyjnej metody asymilacji (4DVAR) z dwoma wiązkowymi filtrami (EnKF i EnSRF) w celu otrzymania metod hybrydowych, tj. EnKF/4DVAR i EnSRF/4DVAR.

Wejściem do zaprojektowanej procedury korekcji obciążenia (BC) i asymilacji danych były, przetworzone w ramach preprocesora danych wejściowych, obserwacje meteorologiczne:

- 1) Satelitarne obserwacje opadu (tj. skumulowane opady na powierzchni gruntu $P_{SAT}: P_{03}, P_{04}, P_{0524}$) uzyskane z geostacjonarnych satelitów Europejskiej Organizacji Eksploatacji Satelitów Meteorologicznych (EUMETSAT).
- 2) Satelitarne obserwacje wilgotności gleby ($H_{14}: H_{14}^{1,00}, H_{14}^{2,89}$) otrzymane w wyniku asymilacji produktów zaawansowanej skaterometrii atmosfery (ASCAT) w Europejskim Centrum Prognoz Średnioterminowych w Zintegrowanym Systemie Prognozowania (ECMWF's IFS).

Dane satelitarne opracowano w Biurze Prognoz Hydrologicznych IMGW-PIB w Krakowie. Wszystkie obliczenia i symulacje przedstawione w artykule dotyczą części zlewni górskiej rzeki Soły zamkniętej wodowskazem Żywiec.

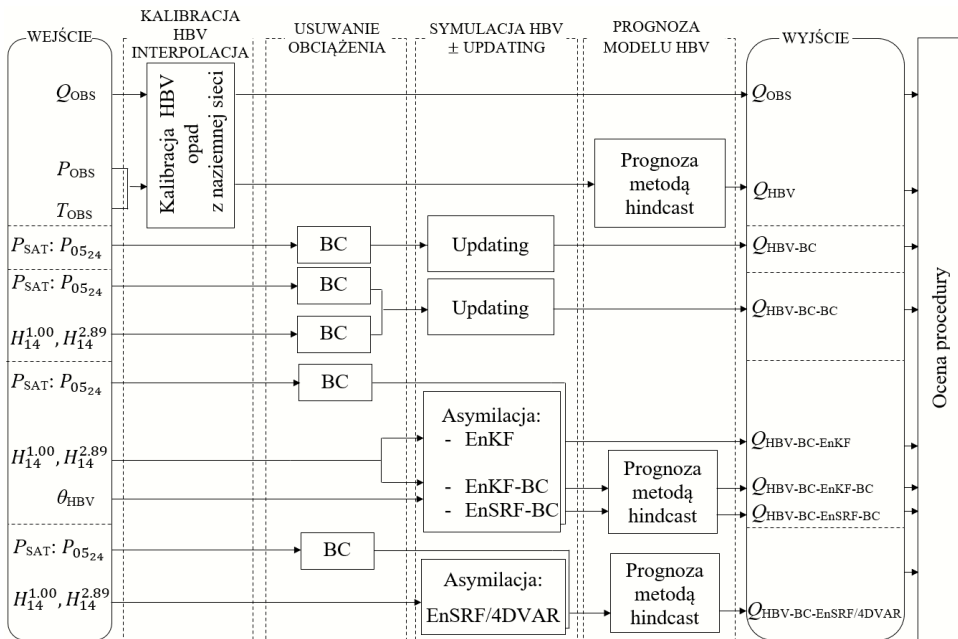
Struktura rozdziału przedstawia się następująco:

- 1) Przedstawienie metodyki preprocesora danych wejściowych do modelu HBV [Spis użytych pojęć i stosowane nazewnictwo, pkt. 40].
- 2) Przetwarzanie naziemnych i satelitarnych obserwacji meteorologicznych.
- 3) Przykład korekcji obciążenia i asymilacji obserwacji satelitarnych do generowania wiązkowych prognoz hydrologicznych.
- 4) Podsumowanie i omówienie uzyskanych wyników.

6.1. Metodyka opracowania preprocesora danych wejściowych do modelu HBV

W niniejszej metodyce wykorzystano: (i) metodę usuwania obciążenia (BC), (ii) asymilację satelitarnych obserwacji wilgotności gleby za pomocą wiązkowych filtrów z uwzględnieniem postprocesu BC (EnKF-BC i EnSRF-BC), (iii) asymilację metodą hybrydową EnSRF/4DVAR do nieobciążonego zaburzenia wiązki prognozowanych zmiennych stanu w ramach preprocesora modelu HBV (rys. 6.1). Zaproponowana metodyka (poziome kolumny na rys. 6.1) jest połączeniem czterech istotnych modułów z możliwością konfigurowania komponentów obliczeniowych zgodnie z wejściem, wyjściem i oceną procedury. Metodyka jest opisana w następujący sposób:

1. Kalibracja deterministycznego hydrologicznego modelu HBV na podstawie obserwacji i pomiarów z naziemnej sieci pomiarowo-obszaryjnej IMGW-PIB.



Rys. 6.1. Diagram blokowy preprocesora HBV, gdzie: P_{OBS} , T_{OBS} , Q_{OBS} reprezentują odpowiednio średni opad z naziemnej sieci pomiarowo-obszaryjnej, temperaturę powietrza i przepływ rzeczny; P_{SAT} reprezentuje satelitarne opady P_{05_24} ; $H_{14}^{1.00}$, $H_{14}^{2.89}$ to satelitarne obserwacje wilgotności gleby, a θ_{HBV} prognozowane zmienne stanu wilgotności gleby modelu HBV; Q_{HBV} jest produktem modelu HBV bez korekcji i asymilacji; Q_{HBV-BC} reprezentuje produkt modelu HBV z korekcją obciążenia opadów satelitarnych i bez asymilacji; $Q_{HBV-BC-BC}$ jest produktem modelu HBV z korekcją obciążenia opadów satelitarnych i korekcją obciążenia z satelitarnej wilgotności bez asymilacji; $Q_{HBV-BC-EnKF}$ jest produktem HBV z korekcją obciążenia opadu satelitarnego i z asymilacją satelitarnej wilgotności gleby za pomocą filtra EnKF; $Q_{HBV-BC-EnKF-BC}$, $Q_{HBV-BC-EnSRF-BC}$ reprezentują produkty HBV z korekcją obciążenia satelitarnych opadów i asymilacją wilgotności gleby z usuwaniem obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby HBV za pomocą filtrów EnKF-BC i EnSRF-BC; $Q_{HBV-BC-EnSRF/4DVAR}$ jest produktem HBV, który asymiluje satelitarne wilgotności gleby metodą hybrydową łączącą filtr EnSRF z filtrem wariacyjnym 4DVAR.

Fig. 6.1. The block diagram of the methodology, where: P_{OBS} , T_{OBS} and Q_{OBS} represent mean ground-based precipitation, temperature and river discharge, respectively; P_{SAT} represents the P_{05_24} ; $H_{14}^{1.00}$ and $H_{14}^{2.89}$ represent soil moisture observations and θ_{HBV} is the predicted soil moisture states of the HBV model; Q_{HBV} is the product of the HBV model without correction and assimilation procedures; Q_{HBV-BC} represents the product of the HBV model with bias-corrected satellite precipitation and without assimilation; $Q_{HBV-BC-BC}$ is the product of the HBV model with bias-corrected satellite precipitation and with bias-corrected satellite soil moisture data without assimilation; $Q_{HBV-BC-EnKF}$ is the product of the HBV model with bias-corrected satellite precipitation and with assimilation of the satellite soil moisture (EnKF); $Q_{HBV-BC-EnKF-BC}$, $Q_{HBV-BC-EnSRF-BC}$ represent the product of the HBV model with bias-corrected satellite precipitation and with assimilation of the satellite soil moisture data with bias correction of the perturbed soil moisture for the creation of an unbiased ensemble of model states (by EnKF-BC and EnSRF-BC); $Q_{HBV-BC-EnSRF/4DVAR}$ is the product of the HBV model, which assimilates the satellite soil moisture data using hybrid method combining EnSRF filter with variational method 4DVAR.

2. Usuwanie obciążenia z meteorologicznych obserwacji satelitarnych P_{SAT} : P_{03} , P_{04} i P_{05_24} i wilgotności gleby H_{14} : $H_{14}^{1.00}$, $H_{14}^{2.89}$ przy użyciu metody BC ze wszystkimi fazami obliczeniowymi, tj. dopasowaniem rozkładu prawdopodobieństwa, walidacją i korekcją, oraz interpolacją danych do centroidu zlewni.
3. Symulacja modelu HBV:
 - a) symulacja modelu HBV bez updatingu, tj. ze zmienną wymuszającą w postaci zaobserwowanego opadu z naziemnej sieci pomiarowo-obszaryjnej IMGW-PIB;

- b) symulacja modelu HBV ze zmienną wymuszającą w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} bez asymilacji wilgotności gleby;
- c) bezpośrednie zastąpienie zmiennych stanu modelu HBV skorygowanymi, tj. z usuniętym obciążeniem metodą BC, satelitarnymi obserwacjami wilgotności gleby $H_{14}^{1,00-BC}$, $H_{14}^{2,89-BC}$, bez asymilacji wilgotności gleby;
- d) asymilacja nieskorygowanej wilgotności gleby, tj. $H_{14}^{1,00}$, $H_{14}^{2,89}$, za pomocą filtru EnKF;
- e) asymilacja nieskorygowanej wilgotności gleby, tj. $H_{14}^{1,00}$, $H_{14}^{2,89}$, z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV, θ_{HBV} , za pomocą zmodyfikowanych filtrów EnKF-BC i EnSRF-BC;
- f) asymilacja nieskorygowanej wilgotności gleby, tj. $H_{14}^{1,00}$, $H_{14}^{2,89}$, z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV, θ_{HBV} , za pomocą filtru EnSRF/4DVAR.

W drugiej metodzie updatingu (c) ciągi wilgotności gleby wygenerowane przez model HBV, θ_{HBV} , traktowane są jako potencjalne obserwacje *in situ*. Natomiast w metodach asymilacji (d-f), jako potencjalne obserwacje *in situ* traktowane są obserwacje satelitarne. W powyższych pięciu metodach (b-f) zmienną wymuszającą jest skorygowany opad satelitarny P_{0524}^{BC} wybrany spośród dostępnych opadów satelitarnych w punkcie (2).

- 4. Prognozy modelu HBV w postaci wiązki, tj. prognozy przedziałowe, z użyciem metody hindcast [Spis użytych pojęć i stosowane nazewnictwo, pkt. 45], w której wejściem do modelu hydrologicznego (zmienne prognozujące) są, zamiast prognozy meteorologicznej, zaobserwowane (historyczne) opady i temperatura powietrza z naziemnej sieci obserwacyjno-pomiarowej IMGW-PIB (np. obserwowany opad z naziemnej sieci pomiarowo-obserwacyjnej P_{OBS} i temperatura powietrza T_{OBS}).

Wejściem do preprocesora (rys. 6.1) są: przepływ rzeczny (Q_{OBS}), zaobserwowane z naziemnej sieci pomiarowo-obserwacyjnej opady atmosferyczne (P_{OBS}), temperatura powietrza (T_{OBS}) i obserwacje satelitarne (tj. opady P_{SAT} : P_{03} , P_{04} i P_{0524} , wilgotność gleby H_{14} : $H_{14}^{1,00}$, $H_{14}^{2,89}$) oraz wilgotność gleby (θ_{HBV}) wygenerowana przez model HBV. W zależności od danych wejściowych, metodyka ocenia sześć różnych symulacji, w tym pięć symulacji z updatingiem (rys. 6.1). W pierwszym kroku jest zainicjowany model HBV, skalibrowany na podstawie historycznych obserwacji i pomiarów uzyskanych z naziemnej sieci pomiarowo-obserwacyjnej IMGW-PIB. Kroki użyte do kalibracji HBV opisano w kolejnych paragrafach niniejszego rozdziału. Odpowiednia konfiguracja preprocesora umożliwi wygenerowanie następujących wyjść z modelu HBV: Q_{HBV} , Q_{HBV-BC} , $Q_{HBV-BC-BC}$, $Q_{HBV-BC-EnKF}$, $Q_{HBV-BC-EnKF-BC}$ lub $Q_{HBV-BC-EnSRF-BC}$, oraz $Q_{HBV-BC-EnSRF/4DVAR}$.

6.2. Przetwarzania naziemnych i satelitarnych obserwacji meteorologicznych

6.2.1. Kalibracja deterministycznego modelu hydrologicznego HBV

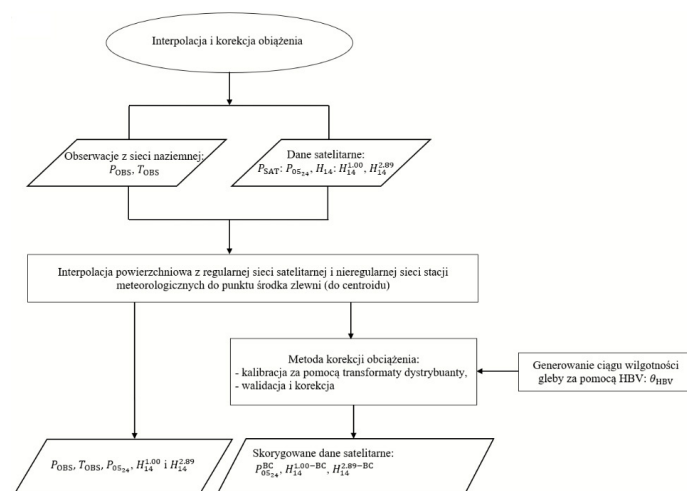
Do modelowania przepływu rzeczne w testowej zlewni Soły zastosowano konceptualny model hydrologiczny typu opad-odpływ (wersja HBV 7.3) [Bergström 1976, 1992, 1995; Lindström i in. 1997] [szczegóły w rozdziale 3.6]. W niniejszym eksperymencie wykonano analizę hydrologiczną testowej zlewni Soły bez podziału na podzlewnie. Wykorzystano schemat skupiony (lumped scheme) do utworzenia pojedynczej zlewni z jedną strefą wegetacyjną, tj. obszaru pola (field area) [Seibert, Vis 2012]. W konsekwencji takiego podziału obserwacje satelitarne z regularnej sieci satelitarnej (P_{03} , P_{04} , P_{0524} , $H_{14}^{1,00}$, $H_{14}^{2,89}$) i z nieregularnej naziemnej sieci pomiarowo-obserwacyjnej (P_{OBS} i T_{OBS}) zostały przestrzennie interpolowane do punktu środka zlewni (centroidu). Model HBV potraktowano wyłącznie jako dobrze skalibrowaną funkcję dynamiczną HBV z własnym zbiorem początkowych zmiennych stanu, stałych, definicji itp. Dane użyte do kalibracji modelu obejmowały stosunkowo długi okres obserwacji meteorologicznych i hydrologicznych (od 1 stycznia 2008 r. do 31 lipca 2014 r.). Kalibrację modelu HBV wykonano automatycznie, z niewielkimi ręcznymi poprawkami następujących parametrów: po-

jemności wodnej gleby, ograniczenia potencjalnej ewaporacji, wykładnika we wzorze obliczającym drenaż gleby, współczynnika recesji oraz współczynników korekcji opadu i topnienia pokrywy śniegu. Ponadto założono, że prawidłowo skalibrowany w momencie rozpoczęcia symulacji model HBV nie będzie wymagał ponownej rekaliibracji w czasie uruchomienia procedury asymilacji, tj. model HBV nie będzie rekaliibrowany z wejściem w postaci obserwacji satelitarnych. Przeprowadzone w ten sposób obliczenia pozwolą na ocenę rzeczywistego wpływu metod korekcji i asymilacji obserwacji satelitarnych na generowanie krótkoterminowych prognoz hydrologicznych.

6.2.2. Interpolacja i korekcja obciążenia (BC) obserwacji meteorologicznych z naziemnej sieci pomiarowo-obszaryjnej i danych satelitarnych

Dane satelitarne z regularnej sieci satelitarnej ($P_{SAT}: P_{03}, P_{04}, P_{0524}, H_{14}: H_{14}^{1,00}, H_{14}^{2,89}$), dane z nieregularnych naziemnych stacji meteorologicznych (P_{OBS} i T_{OBS}) [patrz mapa hydrograficzna zlewni Soły na rys. 2.10 w rozdziale 2.5.1] oraz zmienna stanu (wilgotność gleby) modelu wygenerowana przez model HBV (θ_{HBV}) zostały przestrzennie interpolowane do centroidu zlewni Soły oraz skorygowane za pomocą transformaty dystrybucyjnej (metodą BC) (rys. 6.2). W pracy użyto prostej techniki interpolacji wieloboków Thiessena [Brassel, Reif, 2010]. Decyzję o interpolacji danych wejściowych do modelu podjęto z powodu:

- 1) Trudności interpolacji danych, wynikającej ze specyfiki nieregularnej naziemnej sieci pomiarowej zorganizowanej na obszarze z wyraźnie zmieniającą się wysokością terenu. W takiej sytuacji interpolacja danych z węzłów regularnej siatki do obserwacji naziemnych jest niedokładna (mała liczba naziemnych stacji opadowych położonych głównie w dolinach nie uwzględnia zmiany wartości opadu wraz z wysokością).
- 2) Konieczności skrócenia czasu przetwarzania poprzez redukcję obliczeń do jednego punktu, tj. centroidu (w tym przypadku metoda BC jest obliczana tylko dla centroidu).



Rys. 6.2. Schemat interpolacji i korekcji obciążenia.

Fig. 6.2. The scheme of interpolation and bias correction.

W powyższej fazie obliczeń, tj. interpolacji i korekcji obciążenia obserwacji meteorologicznych z naziemnej sieci pomiarowo-obszaryjnej, zmienne P_{OBS} i wygenerowane przez model HBV ciągi wilgotności gleby, θ_{HBV} , traktowane są jako potencjalne obserwacje *in situ*. Dane o opadach satelitarnych, tj. P_{03}, P_{04} i P_{0524} , porównano z P_{OBS} z naziemnych stacji meteorologicznych, podczas gdy obserwacje satelitarnej wilgotności gleby $H_{14}^{1,00}, H_{14}^{2,89}$ porównano z wartością θ_{HBV} uzyskaną z modelu HBV. Dla każdej zmiennej przyjęto określone granice zmienności (P oznacza opady, θ oznacza wilgotność gleby):

$$P = \begin{cases} P_{OBS} & P_{OBS} \geq 0,1 \text{ (mm)} \\ 0 & P_{OBS} < 0,1 \text{ (mm)} \end{cases}, \quad \theta = \begin{cases} 100 & \theta_{HBV} > 100 \text{ (\%)} \\ \theta_{HBV} & 0 < \theta_{HBV} \leq 100 \text{ (\%)} \\ 0 & \theta \leq 0 \text{ (\%)} \end{cases} \quad (6.1)$$

W ramach analizy statystycznej danych wejściowych do korekcji obciążenia (BC) obserwacje satelitarne zostały przetworzone poprzez dopasowanie teoretycznych rozkładów prawdopodobieństwa do P_{OBS} i θ_{HBV} [szczegóły algorytmu i wyniki przetwarzania w rozdziale 2.5.3]. W tym celu wykorzystano rozkłady prawdopodobieństwa zmiennych losowych [dobowe sumy zaobserwowanego opadu (P_{OBS}) bez uwzględnienia zdarzeń bezopadowych i otrzymanej z symulacji modelu HBV wilgotności gleby (θ_{HBV}) pogrupowanej wg sezonu zimowego i letniego] oraz ich satelitarnych odpowiedników (tj. $P_{MOD}=P_{SAT}$: $P_{03}, P_{04}, P_{05,24}$ i H_{14} : $H_{14}^{1,00}, H_{14}^{2,89}$), które wykazują dodatnią skośność dla opadu i ujemną dla wilgotności gleby oraz nieujemne lewostronne (dolne) ograniczenie rozkładu prawdopodobieństwa [rozdział 2.5.3, tab. 2.14]. Do analizy wybrano trzy trójparametrowe rozkłady, tj. gamma (GA), uogólniony rozkład wykładniczy (GE) i Weibulla (WE) [patrz szczegóły w Dodatku A].

Skorygowane opady satelitarne P_{03}^{BC}, P_{04}^{BC} i $P_{05,24}^{BC}$ zastosowano jako bezpośrednie wejście do symulacji HBV (zmiennie wymuszające), natomiast $H_{14}^{1,00-BC}$ i $H_{14}^{2,89-BC}$ bezpośrednio zastąpiły odpowiednią zmienną stanu HBV w procesie updatingu bez asymilacji satelitarnej wilgotności gleby. Dopasowanie rozkładu prawdopodobieństwa w modelu BC zapewniło optymalne parametry funkcji, która przekształca każdą obserwację do jej skorygowanej postaci. Informacja o optymalnych rozkładach prawdopodobieństwa została wykorzystana do usuwania obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu w ramach procedury zmodyfikowanego filtru EnKF-BC i EnSRF-BC. W powyższej procedurze obserwacje satelitarne traktowane są jako potencjalne obserwacje *in situ*.

6.2.3. Schemat wiązkowego filtru Kalmana (KF) użytego w procedurze asymilacji

Asymilację produktów satelitarnych podzielono na dwa okresy – letni (od 1 sierpnia do 31 października 2014 r.) i zimowy (od 1 listopada 2014 r. do 30 kwietnia 2015 r.). Sezon letni rozpoczął się 1 sierpnia, a nie 1 maja, ze względu na brak godzinnych obserwacji przepływu rzecznej w przekroju wodowskazowym Żywiec na rzece Sole. Podział sezonowy wprowadzono w celu uwzględnienia w procesie asymilacji występowania temperatur powietrza poniżej 0°C i stałej pokrywy śniegu. Umożliwiło to zbadanie wpływu zmian sezonowych na asymilację produktów satelitarnych, zwłaszcza w odniesieniu do obserwacji wilgotności gleby. Procedura asymilacji jest procesem ciągłym, który w tym badaniu wykorzystuje różne równania transformacji, określone oddzielnie dla sezonu letniego i zimowego w ramach metody BC [rozdział 3.8.1.2, algorytm asymilacji danych z metodą korekcji obciążenia, lewa strona rys. 3.5]. Metoda transformaty dystrybucyjnej (BC) wykorzystuje różne typy rozkładów prawdopodobieństwa i parametry, które zostały oszacowane w trakcie kalibracji procedury EnKF-BC.

Schemat metody KF zastosowany w procedurze asymilacji, w tym EnKF-BC, pokazuje w jaki sposób ta metoda, zintegrowana z dynamiczną funkcją modelu HBV, umożliwiła updating częściowych sekwencji prognozowanych zmiennych wilgotności gleby (prawa strona rys. 3.5). Zastosowany algorytm asymilacji satelitarnej wilgotności gleby rozpoczyna się od zainicjowania zmiennych stanu modelu HBV. Wejściem do analizy EnKF jest zbiór zaburzonych prognozowanych zmiennych stanu, tj. wiązka zaburzonych zmiennych z usuniętym obciążeniem, oraz satelitarnych obserwacji wilgotności gleby *in situ*. Analiza EnKF obejmuje: (i) symulację i uruchomienie modelu HBV dla każdego elementu wiązki, (ii) obliczenie zysku Kalmana (K), oraz (iii) updating każdego elementu wiązki z uwzględnieniem K [szczegóły w Załączniku: Dodatek D, wzór 19]. Obliczenia K zawierały Θ^a (mm) [macierz \mathbf{X}^a w Załączniku: Dodatek D], czyli uaktualnione prognozowane zmienne stanu (wilgotności gleby) modelu HBV i wartości tła zmiennej stanu wilgotności gleby θ (mm) [macierz \mathbf{X}^b w Załączniku: Dodatek D] (zmienne stanu wyliczone przez model HBV).

Metoda asymilacji EnKF wymaga określenia liczebności wiązki n , przy jednoczesnym zapewnieniu wystarczającego czasu na przetworzenie rocznego hydrogramu (usuwanie obciążenia metodą BC z zaburzonej wiązki zmiennych stanu jest realizowane za pomocą techniki postprocesowej). Podczas testowania wiązek o liczebności od 10 do 100 określono liczebność wiązki równą $n = 10$. Podstawowym kryterium wyboru liczebności wiązki był krótki czas przetwarzania, gwarantujący równomierny rozkład

zaburzonych hydrogramów w całej wiązce. Asymilacja danych trwa, dopóki są dostępne nowe obserwacje satelitarnej wilgotności gleby.

Do testowania predykcyjnego modelu hydrologicznego HBV z uwzględnieniem updatingu w ramach procedury DA zastosowano metodę hindcastingu [Spis użytych pojęć i stosowane nazewnictwo, pkt. 45]. W ostatnim kroku asymilacji otrzymuje się jest n wektorów zaktualizowanych zmiennych stanu. Następnie symulacyjny model HBV uruchamiany jest n -krotnie, aby uzyskać prognozę hydrologiczną z wejściami w postaci zaobserwowanych (historycznych) danych godzinnych (dane wymuszające), tj. $P_{OBS+1}, \dots, P_{OBS+72}$ i $T_{OBS+1}, \dots, T_{OBS+72}$ i wektorem zmiennych stanu po updatingu. Symulacja modelu HBV dla potrzeb prognozy hydrologicznej nie wykonuje procedury DA. Powyższa metodyka porównuje hydrogramy prognozowane z zaobserwowanymi (historycznymi). Opcjonalnie, w pierwszym kroku prognozy HBV procedura pozwala na m -krotne zaburzenie każdego $n = 10$ elementowego zbioru wiązki prognozowanych zmiennych stanu, uzyskanego w ostatnim kroku updatingu. Teoretycznie szerokość prognozowanej wiązki, tj. prognozy przedziałowej, może osiągnąć liczebność równą $n \times m$. Założono, że prognozy wygenerowane przez model HBV z uwzględnieniem metody DA (updating poprzedzający prognozę) będą przedstawiane w postaci pojedynczego hydrogram uśrednionej wiązki przepływów ($Q_{HBV+1}, \dots, Q_{HBV+72}$) i wiązki hydrogramów.

Przedział ufności symulowanego lub prognozowanego hydrogramu sformułowano za pomocą średniego błędu prognozy *ex post* [Spis użytych pojęć i stosowane nazewnictwo, pkt. 46] przy założonym poziomie istotności (lub wiarygodności prognozy). Problem staje się bardziej skomplikowany, gdy chcemy wyznaczyć prognozę przedziałową w czasie rzeczywistym, ponieważ w chwili stawiania prognozy aktualna wartość prognozowana jest nieznana. Skutkuje to tzw. błędem prognozy *ex ante* [Spis użytych pojęć i stosowane nazewnictwo, pkt. 47]. Wielkość tego błędu opiera się wyłącznie na informacjach dostępnych w czasie prognozy, a mianowicie na informacjach dotyczących oczekiwanych średnich odchyłeń realizacji prognozowanej zmiennej losowej. Rozwinięty w tym badaniu moduł prognozy krótkoterminowej pozwolił na przedstawienie uśrednionego prognozowanego hydrogramu w skali błędów prognozy, tj. >95%, 90-95%, 75-90%, 50-75%, 25-50%, 10-25%, 5-10% i >5%. Definicja prawdopodobieństwa, jako stosunek liczby elementów wiązki, które trafnie przewidują zdarzenie w stosunku do całkowitej liczebności wiązki (rozkład empiryczny), jest użyteczną metodą szacowania stosowaną w celach praktycznych. Powyższe rozwiązanie zakłada, że model dokładnie odzwierciedla rozkład prawdopodobieństwa występowania wezbrania hydrologicznego. Probabilistyczna prognoza generowana w ten sposób powinna być zawsze weryfikowana na podstawie dużych prób losowych. Dzięki temu określono w jakim stopniu prognoza probabilistyczna odnosi się do zaobserwowanych częstości badanego zjawiska. Powyższe rozwiązanie może być stosowane w ramach preprocesora modelu hydrologicznego nawet w sytuacji, w której nie są dostępne jakiegokolwiek dodatkowe obserwacje o prognozowanym zdarzeniu.

6.3. Przykład korekcji obciążenia i asymilacji obserwacji satelitarnych do generowania wiązkowych prognoz hydrologicznych

Wejściami do procedury korekcji obciążenia i asymilacji danych były satelitarne dane opadowe (skumulowany opad na powierzchni gruntu P_{SAT} : $P_{03}, P_{04}, P_{05/24}$) uzyskane z geostacjonarnych satelitów Europejskiej Organizacji Eksploatacji Satelitów Meteorologicznych (EUMETSAT) oraz satelitarne obserwacje wilgotności gleby H_{14} : $H_{14}^{1,00}$ i $H_{14}^{2,89}$ otrzymane w wyniku asymilacji produktów zaawansowanej skaterometrii atmosfery (ASCAT) w Europejskim Centrum Prognoz Średnioterminowych w Zintegrowanym Systemie Prognozowania (ECMWF's IFS). Wszystkie obliczenia i symulacje przedstawione w przykładzie dotyczą części zlewni górskiej rzeki Soły zamkniętej wodowskazem Żywiec [szczegóły w rozdziale 2.5.1].

Dokonano wyboru najlepiej dopasowanej funkcji rozkładu prawdopodobieństwa dla opadu P i wilgotności gleby θ na podstawie testu Kołmogorowa-Smirnowa [szczegóły algorytmu w rozdziale 2.5.3]. Rozkładem prawdopodobieństwa, który najlepiej dopasował się do P_{OBS} był rozkład GE, a do opadów P_{03}, P_{04} i $P_{05/24}$ – rozkład WE. W przypadku wilgotności gleby dopasowanie rozkładów prawdopodobieństwa podzielono na dwa sezony. Zimą najlepiej dopasowanym rozkładem teoretycznym do θ_{HBV} był rozkład WE, do $H_{14}^{1,00}$ – GA i do $H_{14}^{2,89}$ – WE. Latem najlepiej dopasowanym rozkładem do

θ_{HBV} był WE, do $H_{14}^{1,00}$ – GA i do $H_{14}^{2,89}$ – GA. Wybrane rozkłady prawdopodobieństwa wykorzystano w kolejnych etapach przetwarzania, tj. do walidacji, korekcji obciążenia, modelowania i asymilacji za pomocą wiązkowych filtrów Kalmana.

6.3.1. Symulacja przepływu przy użyciu preprocesora modelu HBV

6.3.1.1. Ocena wpływu korekcji obciążenia BC na dokładność modelu i wybór skorygowanych obserwacji satelitarnych do procedury updatingu

Wpływ korekcji obciążenia na dokładność modelu oceniono dla nieskorygowanych obserwacji satelitarnych (np. P_{0524} vs. P_{OBS} i $H_{14}^{1,00}$, $H_{14}^{2,89}$ vs. θ_{HBV}) i ich skorygowanych odpowiedników (np. P_{0524}^{BC} vs. P_{OBS} i $H_{14}^{1,00-BC}$, $H_{14}^{2,89-BC}$ vs. θ_{HBV}). Na podstawie R_0M , $RMSE$ i E_{NS} (tab. 6.1) uznano, że modelowanie P_{OBS} lub θ_{HBV} z uwzględnieniem skorygowanych danych satelitarnych było dokładniejsze niż modelowanie z nieskorygowanymi danymi (tab. 6.1). Na przykład w fazie dopasowania rozkładu miary R_0M , $RMSE$ i E_{NS} dla nieskorygowanego opadu satelitarnego P_{0524} uzyskały wartości odpowiednio 0,908, 0,885 (mm) i $-0,045$, w porównaniu do 0,945, 0,818 i 0,108 dla skorygowanego P_{0524}^{BC} . Podobnie w fazie walidacji w sezonie zimowym R_0M , $RMSE$ i E_{NS} dla nieskorygowanej obserwacji satelitarnej $H_{14}^{1,00}$ uzyskały wartości 0,876, 13,673 (mm) i $-2,112$, w porównaniu do 1,000, 11,487, i $-1,197$ dla skorygowanej wartości $H_{14}^{1,00-BC}$.

Metoda korekcji obciążenia okazała się najmniej skuteczna w przypadku obserwacji satelitarnych H_{14} w sezonie zimowym (tab. 6.1). Zimą metoda poprawiła miarę E_{NS} modelu $H_{14}^{1,00}$ z $-1,728$ do $-1,346$ w fazie dopasowania rozkładu i z $-2,112$ do $-1,197$ w fazie walidacji. Nawet po usunięciu obciążenia symulacja pozostała znacznie gorsza w porównaniu do wilgotności gleby obliczonej bezpośrednio przez model HBV, tj. θ_{HBV} . Jednak latem, mimo ujemnych wartości indeksu E_{NS} uzyskanych w fazach dopasowania rozkładu i walidacji, odpowiednio ($-0,655$ i $-0,419$) dla nieskorygowanej obserwacji $H_{14}^{1,00}$, wartości E_{NS} dla skorygowanej wilgotności gleby $H_{14}^{1,00-BC}$, choć nie wskazują na dużą dokładność ($E_{NS} > +0,5$; wyniosły odpowiednio 0,220 i 0,261), znacząco poprawiły model H_{14} (tab. 6.1). Ta rozbieżność może wynikać z występowania pokrywy śnieżnej i niskich temperatur w glebie na głębokości 0,05 m. Wyniki wskazują na przydatność korekcji obciążenia latem.

Tabela 6.1. Poprawa dokładności modelowania danych H-SAF, tj. opadu i wilgotności gleby P_{03} , P_{04} i $P_{05_{24}}$ oraz $H_{14}^{1,00}$ i $H_{14}^{2,89}$, za pomocą metody korekcji obciążenia. Ocena metody na podstawie RoM (stosunku wartości średnich); RMSE (pierwiastka kwadratowego średniego błędu kwadratowego) i Ens (współczynnika skuteczności Nasha i Sutcliffea). Pogrubione wartości wskazują na poprawę dokładności modelowania danych H-SAF po zastosowaniu korekcji obciążenia.

Table 6.1. Improvement in modelling accuracy with *bias* correction of precipitation and soil moisture H-SAF products P_{03} , P_{04} i $P_{05_{24}}$ and $H_{14}^{1,00}$ and $H_{14}^{2,89}$ based on RoM (ratio of means); RMSE (root mean squared error), and Ens (Nash-Sutcliffe efficiency index). Bold values show the improved accuracy of modelling H-SAF products after bias-correction.

Dane H-SAF		Faza modelowania											
		DOPASOWANIE ROZKŁADÓW						WALIDACJA					
		Bez korekcji obciążenia			Z korekcją obciążenia			Bez korekcji obciążenia			Z korekcją obciążenia		
		RoM	RMSE (mm)	Ens	RoM	RMSE (mm)	Ens	RoM	RMSE (mm)	Ens	RoM	RMSE (mm)	Ens
		od 1 maja 2013 r. to 31 lipca 2014 r.			Opad			od 1 sierpnia 2014 r. to 30 kwietnia 2015 r.					
P_{03}	0,884	0,894	-0,065	0,918	0,815	0,115	0,879	0,743	-0,266	0,905	0,636	0,073	
P_{04}	0,809	0,890	-0,057	0,854	0,793	0,161	0,724	0,678	-0,054	0,828	0,621	0,115	
$P_{05_{24}}$	0,908	0,885	-0,045	0,945	0,818	0,108	0,699	0,617	0,126	1,025	0,512	0,397	
		od 1 maja 2012 r. to 31 lipca 2014 r.			Wilgotność gleby			od 1 sierpnia 2014 r. to 30 kwietnia 2015 r.					
ZIMA	$H_{14}^{1,00}$	0,815	13,572	-1,728	1,007	12,584	-1,346	0,876	13,673	-2,112	1,000	11,487	-1,197
	$H_{14}^{2,89}$	0,876	19,580	-4,679	1,002	12,569	-1,340	0,815	19,793	-5,523	1,001	11,953	-1,397
LATO	$H_{14}^{1,00}$	1,187	14,012	-0,655	0,999	9,615	0,220	1,188	13,219	-0,419	1,000	9,538	0,261
	$H_{14}^{2,89}$	1,201	13,702	-0,583	1,001	7,764	0,492	1,201	13,167	-0,408	1,002	7,944	0,487

Na podstawie rozważanych miar jakości (tj. R_0M , $RMSE$ i E_{NS}) żaden model nie działał dobrze według wszystkich miar, podobnie żadna pojedyncza miara nie okazała się wystarczająca do opisanie, jak dobrze wykonano określony model. Nie było możliwe określenie, które kryterium (miara) było najważniejsze [Houghton-Carr 2009]. Dlatego też na podstawie trzech rozważanych miar uznano, że skuteczność usuwania błędu systematycznego z obserwacji satelitarnych może wpływać na asymilację satelitarnego opadu i wilgotności gleby będących wejściem do modeli hydrologicznych. Uznano, że skorygowany metodą korekcji obciążenia (BC) satelitarny opad P_{0524}^{BC} i skorygowane/nieskorygowane wilgotności gleby H_{14} : $H_{14}^{1,00}$, $H_{14}^{1,00-BC}$, $H_{14}^{2,89}$ i $H_{14}^{2,89-BC}$ będą wejściem do procedury updatingu w preprocesorze modelu HBV (rys. 6.1).

6.3.1.2. Algorytm symulacji przepływu przy użyciu preprocesora modelu HBV

Symulację przepływu rzecznoego w zlewni Soły w przekroju wodowskazowym Żywiec, z uwzględnieniem sezonu letniego (tj. od 1 sierpnia 2014 r. do 31 października 2014 r.) oraz zimowego (tj. od 1 listopada 2014 r. do 30 kwietnia 2015 r.), wykonano za pomocą modelu hydrologicznego HBV. Analizę przeprowadzono w pięciu krokach:

1. Z wymuszeniem w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} bez asymilacji wilgotności gleby (drugie wejście od góry na rys. 6.1).
2. Z wymuszeniem w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} i skorygowaną obserwacją wilgotności gleby ($H_{14}^{1,00-BC}$ lub $H_{14}^{2,89-BC}$) bez asymilacji wilgotności gleby, lecz z updatingiem za pomocą bezpośredniego zastąpienia odpowiedniej zmiennej stanu modelu HBV (trzecie wejście od góry na rys. 6.1).
3. Z wymuszeniem w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} i asymilacją nieskorygowanych satelitarnych obserwacji wilgotności gleby $H_{14}^{1,00}$ lub $H_{14}^{2,89}$ za pomocą filtru EnKF (czwarte wejście od góry na rys. 6.1 – pierwsza opcja).
4. Z wymuszeniem w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} i asymilacją nieskorygowanych satelitarnych obserwacji wilgotności gleby $H_{14}^{1,00}$ lub $H_{14}^{2,89}$ z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV, θ_{HBV} , za pomocą zmodyfikowanych filtrów EnKF-BC i EnSRF-BC (czwarte wejście od góry na rys. 6.1 – druga i trzecia opcja).
5. Z wymuszeniem w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} i asymilacją nieskorygowanych satelitarnych obserwacji wilgotności gleby $H_{14}^{1,00}$ lub $H_{14}^{2,89}$ z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV θ_{HBV} , za pomocą filtru EnSRF/4DVAR (piąte wejście od góry na rys. 6.1).

Uzyskane wyniki symulacji dla określonej konfiguracji preprocesora modelu HBV (tj. komponentów obliczeniowych), zgodnej z rysunkiem 6.1, przedstawiono w tabeli 6.2. Na rysunkach 6.3-6.4 z symulacjami w sezonie letnim i zimowym porównano następujące przepływy:

- 1) Q_{OBS} , obserwowany przepływ.
- 2) Q_{HBV} , symulowany przepływ przez model HBV z obserwowanym opadem P_{OBS} (jako daną wymuszającą model HBV).
- 3) Q_{HBV-BC} , symulowany przepływ przez model HBV ze skorygowanym opadem satelitarnym P_{0524}^{BC} (jako daną wymuszającą model HBV) bez asymilacji wilgotności gleby.

Na rysunkach 6.5-6.6 z symulacjami w sezonie letnim i zimowym porównano:

- 1) θ_{HBV} , obliczoną wilgotność gleby przez model HBV z wymuszeniem w postaci P_{OBS} .
- 2) $H_{14}^{1,00-BC}$ (rys. 6.5) i $H_{14}^{2,89-BC}$ (rys. 6.6), skorygowane satelitarnie wilgotności gleby.
- 3) $\bar{s}r. H_{14}^{1,00}$ (rys. 6.5) i $\bar{s}r. H_{14}^{2,89}$ (rys. 6.6), średnie wartości wiązek zaburzonych wilgotności gleby (bez usuwania obciążenia metodą BC).
- 4) $\bar{s}r. H_{14}^{1,00-BC}$ (rys. 6.5) i $\bar{s}r. H_{14}^{2,89-BC}$ (rys. 6.6), średnie wartości wiązek zaburzonych wilgotności gleby z uwzględnieniem korekcji obciążenia metodą BC.
- 5) Q_{OBS} , obserwowany przepływ.

- 6) Q_{HBV} , symulowany przepływ przez model HBV z obserwowanym opadem P_{OBS} .
- 7) $Q_{\text{HBV-BC-BC}}$, symulowany hydrogram ze skorygowaną wilgotnością: $H_{14}^{1,00-BC}$ (rys. 6.5) i $H_{14}^{2,89-BC}$ (rys. 6.6) bez asymilacji satelitarnej wilgotności gleby.
- 8) $\text{sr.}Q_{\text{HBV-BC-EnKF}}$, średni symulowany hydrogram z asymilacją nieskorygowanych satelitarnych obserwacji wilgotności gleby, $H_{14}^{1,00}$ (rys. 6.5) i $H_{14}^{2,89}$ (rys. 6.6), za pomocą filtru EnKF.
- 9) $\text{sr.}Q_{\text{HBV-BC-EnKF-BC}}$, średni symulowany hydrogram z asymilacją nieskorygowanych satelitarnych obserwacji wilgotności gleby, $H_{14}^{1,00}$ (rys. 6.5) i $H_{14}^{2,89}$ (rys. 6.6), z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu modelu HBV, θ_{HBV} , za pomocą zmodyfikowanego filtru EnKF-BC.

Tabela 6.2. Dokładność estymacji przepływów Soły w przekroju Żywiec przy użyciu modelu HBV, na podstawie obserwowanego opadu, danych H-SAF, tj. opadu i wilgotności gleby z/bez BC (korekcji obciążenia), oraz z filtrowaniem satelitarnej wilgotności gleby za pomocą filtru EnKF-BC (Ensemble Kalman Filter) i EnSRF-BC (Ensemble Square Root Filter).

Table 6.2. Accuracy of the HBV model in estimating stream flow for Żywiec post of Soła River based on observed precipitation, precipitation and soil moisture-related H-SAF products with BC (bias correction) and with filtering the soil moisture HSAF products using EnKF-BC (Ensemble Kalman Filter) and EnSRF-BC (Ensemble Square Root Filter).

Wejście do modelu	BC	Filtry		Wyjście z modelu	SEZON					
		EnKF	EnSRF		LATO 1 VIII 2014 – 31 X 2014			ZIMA 1 XI 2014 – 30 IV 2015		
					EnKF-BC	RoM	RMSE (m^3s^{-1})	ENS	RoM	RMSE (m^3s^{-1})
P_{OBS}	—	—	—	Q_{HBV}	1,274	11,806	-0,045	1,065	18,843	-0,185
P_{0524}	√	—	—	$Q_{\text{HBV-BC}}$	1,271	10,134	0,103	0,801	15,537	0,195
P_{0524}	√	—	—	$Q_{\text{HBV-BC-BC}}$	1,232	9,486	0,326	1,314	14,652	0,284
$H_{14}^{1,00}$	√	—	—	$Q_{\text{HBV-BC-BC}}$	1,232	9,486	0,326	1,314	14,652	0,284
P_{0524}	√	√	—	$Q_{\text{HBV-BC-EnKF}}$	1,192	8,575	0,418	1,359	13,702	0,492
$H_{14}^{1,00}$	—	—	—	$Q_{\text{HBV-BC-EnKF}}$	1,192	8,575	0,418	1,359	13,702	0,492
P_{0524}	√	—	—	$Q_{\text{HBV-BC-EnKF}}$	1,140	6,755	0,658	1,161	11,884	0,529
$H_{14}^{1,00}$	—	√	—	$Q_{\text{HBV-BC-EnKF}}$	1,140	6,755	0,658	1,161	11,884	0,529
P_{0524}	√	—	—	$Q_{\text{HBV-BC-BC}}$	1,158	7,791	0,545	1,139	12,668	0,465
$H_{14}^{2,89}$	√	—	—	$Q_{\text{HBV-BC-BC}}$	1,158	7,791	0,545	1,139	12,668	0,465
P_{0524}	√	√	—	$Q_{\text{HBV-BC-EnKF}}$	1,180	7,764	0,583	1,160	11,882	0,551
$H_{14}^{2,89}$	—	—	—	$Q_{\text{HBV-BC-EnKF}}$	1,180	7,764	0,583	1,160	11,882	0,551
P_{0524}	√	—	—	$Q_{\text{HBV-BC-EnKF}}$	1,143	6,838	0,650	1,139	11,392	0,567
$H_{14}^{2,89}$	—	√	—	$Q_{\text{HBV-BC-EnKF}}$	1,143	6,838	0,650	1,139	11,392	0,567
P_{0524}	√	—	√	$Q_{\text{HBV-BC-EnSR}}$	0,974	4,828	0,825	0,830	12,683	0,461
$H_{14}^{1,00}$	—	—	√	$Q_{\text{HBV-BC-EnSR}}$	0,974	4,828	0,825	0,830	12,683	0,461
P_{0524}	√	—	√	$Q_{\text{HBV-BC-EnSR}}$	0,982	4,746	0,827	0,612	16,760	0,059
$H_{14}^{2,89}$	—	—	√	$Q_{\text{HBV-BC-EnSR}}$	0,982	4,746	0,827	0,612	16,760	0,059

Ostatnim elementem tego rozdziału jest symulacja modelu HBV dla potrzeb prognozy hydrologicznej. Symulacja składa się z dwóch etapów:

- 1) Uaktualniania modelu HBV za pomocą procedury DA,

- 2) Symulacji modelu HBV dla potrzeb prognozy hydrologicznej wykonanej metodą hindcast ze zbiorem zmiennych stanu modelu uzyskanym w ostatnim kroku updatingu.

Wyniki analiz przedstawiono w tabeli 6.3 i na rysunkach 6.7-6.9.

6.3.1.3. Symulacja przepływu z korekcją obciążenia satelitarne opadu bez asymilacji

W pierwszym kroku analizy oceniono wpływ skorygowanego opadu satelitarnego (P_{0524}^{BC}) na dokładność symulacji modelu HBV latem (rys. 6.3) i zimą (rys. 6.4). Porównano obserwowany przepływ (Q_{OBS}) z symulacją HBV (Q_{HBV}) i z symulacją HBV z wymuszeniem satelitarnym P_{0524}^{BC} (Q_{HBV-BC}). W obu sezonach zaobserwowano niewielki wzrost efektywności modelu HBV przy użyciu skorygowanego opadu satelitarnego (tab. 6.2). W porównaniu do Q_{HBV} wymuszenie P_{0524}^{BC} poprawiło symulację modelu HBV (Q_{HBV-BC}) zwiększając E_{NS} latem od $-0,045$ do $0,103$, a zimą od $-0,185$ do $0,195$. W sezonie letnim najwyższą ilościową i jakościową rozbieżność między zaobserwowanym a satelitarnym opadem zarejestrowano w sierpniu, co jest związane z niestabilnością konwekcyjną w atmosferze. Różnice w ilościowym oszacowaniu opadów wpłynęły na rozbieżność między Q_{HBV} i Q_{HBV-BC} . Zauważalną poprawę symulacji modelu HBV z opadami satelitarnymi zaobserwowano w drugiej połowie września i w październiku (rys. 6.3).

Zimą, wraz z pojawieniem się ujemnych temperatur powietrza i pokrywy śnieżnej, ilościowa informacja o opadzie pośrednio wpłynęła na symulację. Głównym determinantem odpływu rzeczne są źródła wody gruntowej, natomiast w okresie zaniku pokrywy śnieżnej i początku topnienia powierzchniowej warstwy gruntu, główną rolę zaczynają odgrywać spływ powierzchniowy i podpowierzchniowy. Dlatego pod koniec sezonu zimowego podczas wezbrań deszczowo-roztopowych zaobserwowano pozytywny wpływ skorygowanego opadu satelitarnego na dokładność symulowanych hydrogramów (rys. 6.4).

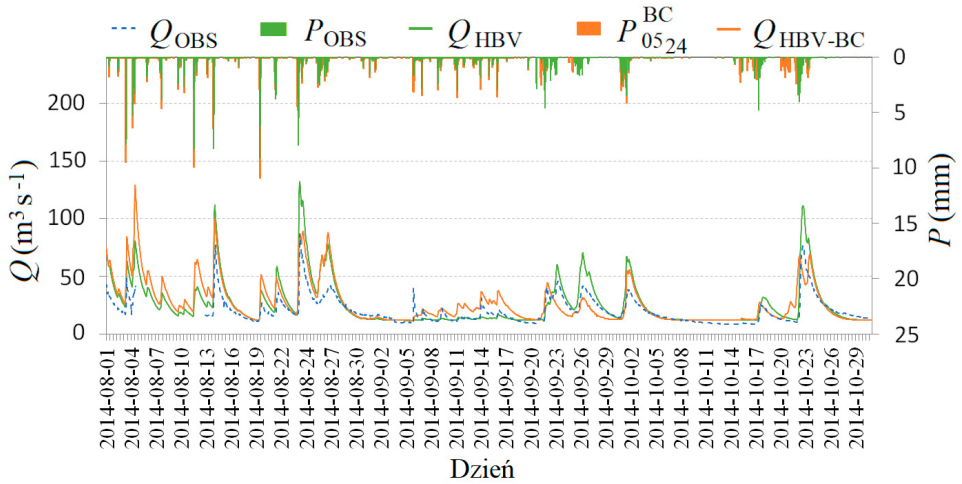
6.3.1.4. Symulacja przepływu z użyciem dwóch metod updatingu, tj.: zastąpienie zmiennej stanu modelu HBV skorygowaną satelitarną wilgotnością gleby i asymilacja satelitarnej wilgotności gleby

Najlepszym satelitarnym opadem do modelowania P_{OBS} był P_{0524}^{BC} (tab. 6.1) i dlatego został użyty jako dana wymuszająca model HBV w kolejnych analizach. W górnej części rysunków 6.5-6.6 porównano przebiegi wilgotności gleby wygenerowanej przez model HBV, θ_{HBV} (zielona przerywana linia), ze skorygowaną satelitarną wilgotnością gleby, $H_{14}^{1,00-BC}$ (latem) lub $H_{14}^{2,89-BC}$ (zimą) (czarna ciągła linia), ze średnią wartością wiązki zaburzonej wilgotności gleby, $\bar{H}_{14}^{1,00}$ lub $\bar{H}_{14}^{2,89}$ (pomarańczowa ciągła linia) i ze średnią wartością wiązki wolnej od obciążenia, $\bar{H}_{14}^{1,00-BC}$ lub $\bar{H}_{14}^{2,89-BC}$ (czerwona ciągła linia). Obliczenia wykonano oddzielnie dla sezonu letniego i zimowego (odpowiednio lewa i prawa część tabeli 6.2).

Na rysunkach 6.5-6.6 pokazano wpływ asymilacji satelitarnych obserwacji wilgotności gleby ($H_{14}^{1,00}$, $H_{14}^{2,89}$) (czwarte wejście od góry na rys. 6.1 – druga opcja) na poprawę dokładności symulacji modelu HBV, z updatingiem za pomocą filtra EnKF-BC, z wyjściem w postaci $\bar{Q}_{HBV-BC-EnKF-BC}$ (czerwona ciągła linia na rys. 6.5 i 6.6). Symulacje HBV z asymilacją za pomocą zmodyfikowanego filtra EnKF-BC porównano do asymilacji za pomocą EnKF (czwarte wejście od góry na rys. 6.1 – pierwsza opcja), $\bar{Q}_{HBV-BC-EnKF}$ (pomarańczowa ciągła linia na rys. 6.5 i 6.6), oraz do symulacji z użyciem tylko skorygowanej satelitarnej wilgotności gleby ($H_{14}^{1,00-BC}$, $H_{14}^{2,89-BC}$) jako wejścia do modelu HBV, Q_{HBV-BC} (czarna przerywana linia na rys. 6.5 i 6.6) (trzecie wejście od góry na rys. 6.1), jak również do symulacji HBV z obserwowanym opadem Q_{HBV} (zielona przerywana linia) (drugie wejście od góry na rys. 6.1) oraz hydrogramem obserwowanym Q_{OBS} (przerywana niebieska linia z podwójnymi kropkami).

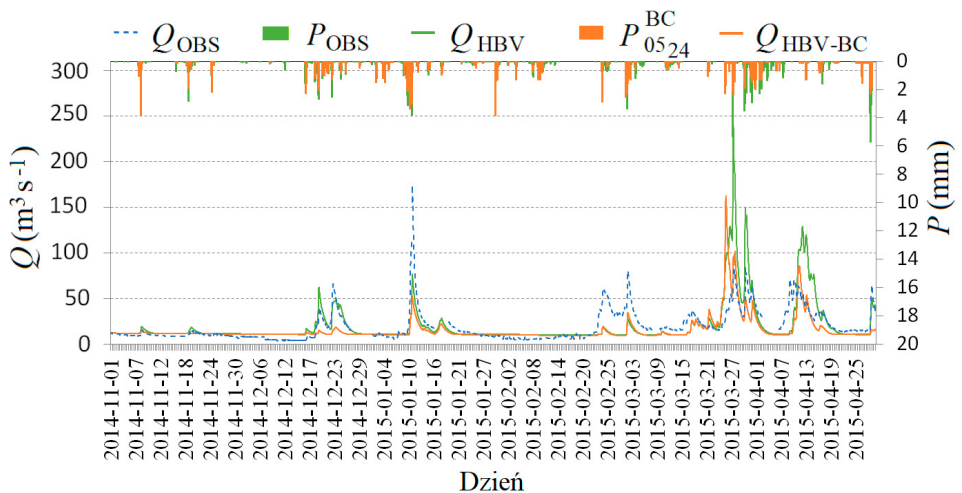
Dynamika wilgotności gleby jest złożona. W HBV modelowany jest tylko wpływ wilgotności gleby na generowanie odpływu w skali całej zlewni. Algorytm opiera się na modyfikacji procedury sortowania kubelkowego, w której zakłada się statystyczny rozkład potencjału magazynowania wody w zlewni [Lindström i in. 1997]. Największy odływ został osiągnięty, gdy wszystkie zbiorniki modelu były pełne i tym samym przyczyniły się do zwiększenia objętości odpływu, np. wezbrania w okresach 22.09-04.10 oraz 21.10-26.10 2014 r. (rys. 6.5). W takich sytuacjach równania transformacji wilgotności

gleby prowadzą do maksymalnego przechowywania wilgoci w glebie. Precyzyjne oszacowanie tej wielkości fizycznej miało duży wpływ na symulację przepływu podczas letnich wezbrań.



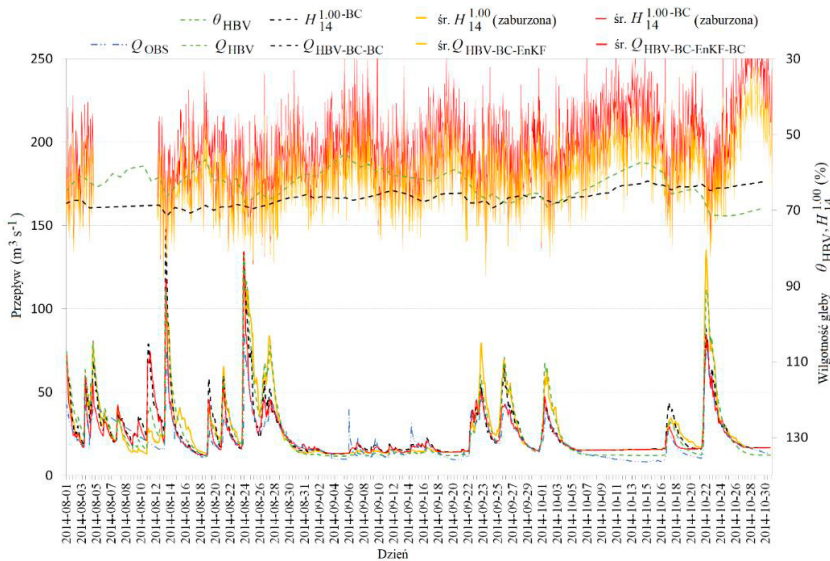
Rys. 6.3. Usuwanie obciążenia z P_{0524} za pomocą transformaty dystrybucyjnej – wezbrania letnie.

Fig. 6.3. Removing bias from P_{0524} by distribution-derived transformation algorithm – summer floods.

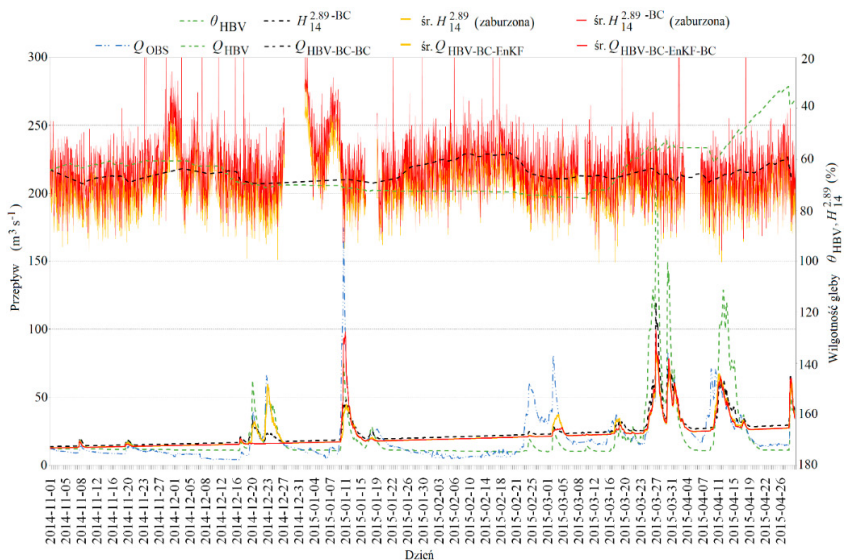


Rys. 6.4. Usuwanie obciążenia z P_{0524} za pomocą transformaty dystrybucyjnej – wezbrania zimowe.

Fig. 6.4. Removing bias from P_{0524} by distribution-derived transformation algorithm -winter floods.



Rys. 6.5. Asymilacja wilgotności gleby $H_{14}^{1,00}$ za pomocą filtra EnKF-BC – wezbrania letnie.
Fig. 6.5. Assimilation the soil moisture $H_{14}^{1,00}$ by EnKF-BC – summer floods.



Rys. 6.6. Asymilacja wilgotności gleby $H_{14}^{2,89}$ za pomocą filtra EnKF-BC – wezbrania zimowe.
Fig. 6.6. Assimilation the soil moisture $H_{14}^{2,89}$ by EnKF-BC – winter floods.

Asymilowana satelitarna wilgotność gleby służy jako obserwacja wpływająca na wartość współczynnika odpływu. Co więcej, błąd oszacowania największych zaobserwowanych przepływów w drugiej części sezonu letniego wyraźnie zmniejszył się dzięki wykorzystaniu preprocesora HBV, powodując ściślejsze dopasowanie symulowanego hydrogramu do obserwacji w ostatniej części symulacji. Sytuacje takie miały szczególne znaczenie w procesie estymacji zbioru początkowych zmiennych stanu, kiedy model HBV jest uruchamiany do prognozowania. Oczywiście przeszacowanie wilgotności gleby

spowodowało przeszacowanie przepływu, co było widoczne w okresie od 11 sierpnia (godz. 12:00) do 26 sierpnia (godz. 02:00) 2014 roku.

Latem pozytywny wpływ korekcji obciążenia opadu i asymilacji obserwacji satelitarnych można było zaobserwować w odniesieniu do dokładności symulowanych hydrogramów. Jednak zimą, tj. od 1 listopada 2014 r. (godz. 06:00) do 30 kwietnia 2015 r. (godz. 23:00) zlewnia Soły była pokryta śniegiem, a temperatury powietrza były ujemne, co pogorszyło jakość $H_{14}^{1,00}$ i $H_{14}^{2,89}$. Spowodowało to wzrost RMSE i zmniejszenie E_{NS} w porównaniu do symulacji HBV z preprocesorem w lecie (tab. 6.2). Zimą, gdy występuje pokrywa śniegu, jakość szacowania wilgotności gleby z ASCAT jest niska. Produkt ECMWF ($H_{14}/SM-DAS-2$) jest walidowany na podstawie pomiarów wilgotności gleby w oparciu o obserwacje *in situ* (m.in. na podstawie pomiarów wilgotności gleby z nizinnej stacji Monitoringu Suszy Rolniczej w Polsce w Łukowie) [szczegóły w rozdziale 2.5.2].

Z rysunku 6.6 wynika, że od listopada do połowy marca wszystkie symulowane hydrogramy niedoszacowały Q_{OBS} , ale przeszacowywały wezbranie kwietniowe. W porównaniu do Q_{HBV} preprocesor HBV z asymilacją wilgotności gleby lepiej modelował wezbranie w marcu i kwietniu 2015 r., co pokazuje duża rozbieżność między niebieską przerywaną linią z podwójnymi kropkami (Q_{OBS}), zieloną przerywaną linią (Q_{HBV}) i czerwoną ciągłą linią (śr. $Q_{HBV-BC-EnKF-BC}$). W trakcie zaniku stałej pokrywy śnieżnej (od końca marca do kwietnia) zaobserwowano pozytywny wpływ opadu satelitarnego, tj. P_{0524}^{BC} (rys. 6.4). W marcu i kwietniu zaobserwowano nieznaczną różnicę między symulowanymi hydrogramami: Q_{HBV-BC} (rys. 6.4), $Q_{HBV-BC-BC}$ a śr. $Q_{HBV-BC-EnKF-BC}$ (rys. 6.6). Wpływ satelitarnej wilgotności gleby i jej asymilacji był widoczny tylko w niewielkim stopniu.

Na rysunku 6.6 można zaobserwować negatywny wpływ procedury DA na odpływ z dolnego liniowego zbiornika modelu HBV. Istnieje wyraźna tendencja wzrostowa symulowanego przepływu śr. $Q_{HBV-BC-EnKF-BC}$ (czerwona ciągła linia) i $Q_{HBV-BC-BC}$ (przerywana czarna linia). W modelu HBV generowany odpływ jest funkcją odpowiedzi, która przekształca nadmiar wody ze zbiornika wilgotności gleby w odpływ. Woda ze zbiornika wilgotności gleby jest dodawana do górnego zbiornika wody gruntowej i następnie transformowana w kierunku zbiornika dolnego. Magazynowana woda w zbiorniku wody gruntowej zlewni jest składnikiem przepływu bazowego. Przy wysokiej wydajności gleby, perkolacja jest niewystarczająca do utrzymania pustego górnego zbiornika. Generowany zrzut wody bezpośrednio przyczynia się do drenażu wody przez powierzchniowe kanały. Dolny zbiornik wody gruntowej reprezentuje proces magazynowania wody, który przyczynia się do wzrostu przepływu bazowego. Konsekwencją powyższych procesów jest utrzymywanie się stałego rosnącego trendu przepływu (rys. 6.6). W celu zmniejszenia przepływu bazowego i tym samym poprawienia początkowych wartości zmiennych stanu modelu HBV w momencie uruchomienia prognozy hydrologicznej, procedura asymilacji powinna zostać rozszerzona o updating zmiennych stanu górnego i dolnego zbiornika wody gruntowej.

6.3.2. Przykład updatingu i prognozy hydrologicznej modelu HBV w zlewni Soły w profilu wodowskazowym Żywiec

Ostatnim elementem badania jest ocena możliwości wykorzystania obserwacji satelitarnych do updatingu prognozy hydrologicznej i porównania prognoz obliczonych z i bez updatingu. Updating uwzględnił korekcję obciążenia danych wymuszających (tj. opadu satelitarnego) oraz asymilację satelitarnych obserwacji wilgotności gleby. Asymilację wykonano za pomocą klasycznego filtra wiązkowego i hybrydowego, z uwzględnieniem postprocesowego usuwania obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu modelu hydrologicznego. Do analizy wybrano wezbranie deszczowo-roztopowe w okresie od 25 marca do 2 kwietnia 2015 roku. Dla wyodrębnionego wezbrania długość updatingu była równa 120 h, tj. od 25 do 30 marca 2015 r., natomiast okno prognozy 72 h, tj. od 31 marca do 2 kwietnia 2015 roku. Dane wejściowe do updatingu zawierały skorygowany opad satelitarny P_{0524}^{BC} , asymilowana była też wilgotność gleby $H_{14}^{1,00}$ i $H_{14}^{2,89}$. Model HBV uruchomiono dla potrzeb prognozy hydrologicznej w ostatnim kroku updatingu. Prognozę hydrologiczną wykonano metodą hindcast, tzn. ze danymi wymuszającymi był zaobserwowany opad P_{OBS} i temperatura powietrza T_{OBS} . W ramach prognozy hydrologicznej nie asymilowano obserwacji satelitarnych.

Prognoza poprzedzona updatingiem osiągnęła lepsze wartości R_0M , RMSE i E_{NS} w porównaniu do prognozy bez updatingu (tab. 6.3).

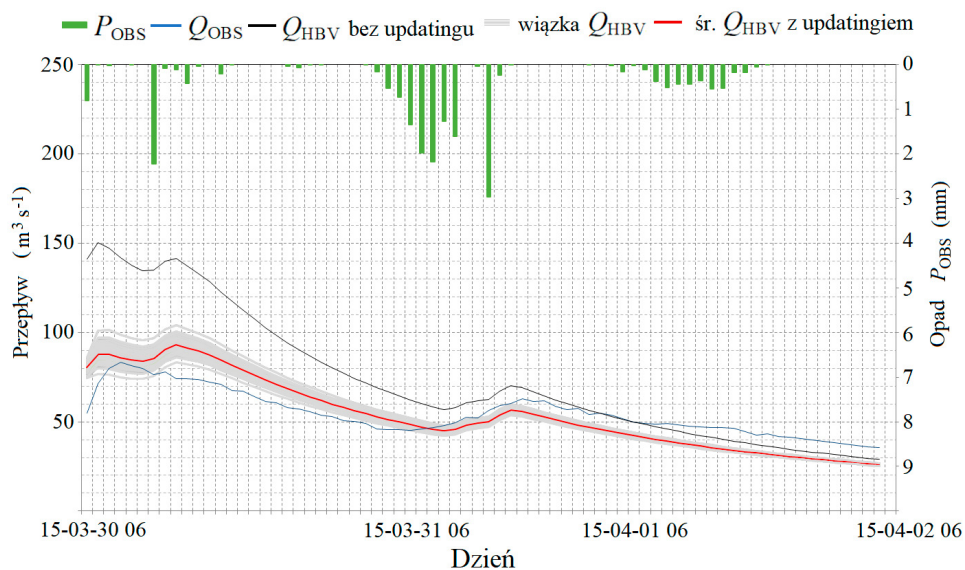
Przykładem bardzo dobrej prognozy hydrologicznej, wynikającej z efektywnego updatingu za pomocą hybrydowego filtru EnSRF/4DVAR, była prognoza wezbrania, które wystąpiło w okresie od 30 marca do 2 kwietnia 2015 roku. Analiza prognozowanego hydrogramu zwraca uwagę na różnicę położenia pozycji startowej obserwowanego hydrogramu (niebieska ciągła linia), hydrogramu bez updatingu (czarna ciągła linia) oraz uśrednionego hydrogramu z updatingiem (czerwona ciągła linia) dla EnKF-BC (rys. 6.7) i EnSRF/4DVAR (rys. 6.8). Konsekwencją jest wyraźne przeszacowanie Q_{HBV} w stosunku do Q_{OBS} i śr. Q_{HBV} . Przeszacowanie Q_{HBV} wynika z faktu, że model HBV uruchomiony bez updatingu nie uwzględnia korekcji opadu oraz asymilacji satelitarnych obserwacji wilgotności gleby. Zmienne wymuszające (opad) oraz zmienna stanu (wilgotność gleby) mają kluczowe znaczenie w procesie symulacji modelu typu opad–odpływ, gdyż mają one bezpośredni udział w wymianie wody i energii. Estymacja rzeczywistego opadu w okresie zimowym i poprzedzającym roztopy, a zwłaszcza opis przestrzennej zmienności opadu w modelowanej zlewni, jest istotnym zagadnieniem w modelu typu opad–odpływ. Obserwacje satelitarne dostarczają dużo większej informacji na powyższy temat w porównaniu do obserwacji prowadzonych w ramach naziemnej sieci pomiarowo-obszaryjnej. W okresie poprzedzającym prognozę, tj. od 25 do 30 marca 2015 r., można zaobserwować pozytywny wpływ skorygowanego opadu $P_{05,24}^{BC}$ oraz asymilowanej satelitarnej wilgotności gleby $H_{14}^{2,89-BC}$ na dokładność symulowanego hydrogramu (odpowiedni rys. 6.4 i 6.6).

Tabela 6.3. Porównanie symulacji modelu HBV dla potrzeb prognozy bez updatingu z P_{OBS} jako zmienną wymuszającą model HBV i z updatingiem z wymuszeniem w postaci $P_{05,24}^{BC}$ i asymilacją satelitarnej wilgotności gleby $H_{14}^{1,00}$ i $H_{14}^{2,89}$ za pomocą filtrów EnKF-BC, EnSRF-BC i EnSRF/4DVAR. Ocena updatingu i prognozy wykonano na podstawie miar R_0M (stosunku wartości średnich), RMSE (pierwiastka kwadratowego średniego błędu kwadratowego) i E_{NS} (współczynnika skuteczności Nasha i Sutcliffe’a).

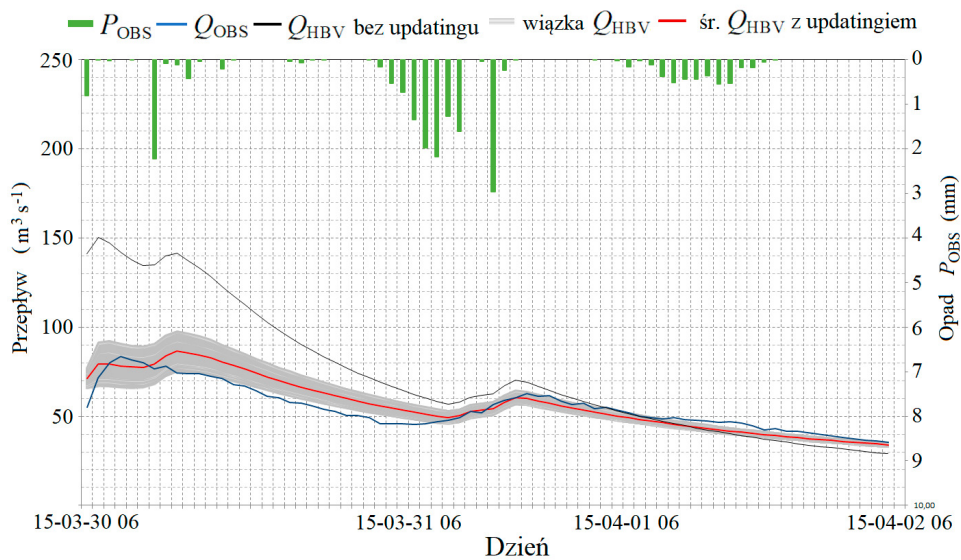
Table 6.3. Comparison of HBV simulation for forecast without updating with P_{OBS} as data forcing and with $P_{05,24}^{BC}$ as forcing data and with filtering the soil moisture $H_{14}^{1,00}$ and $H_{14}^{2,89}$ using EnKF-BC, EnSRF-BC and EnSRF/4DVAR. The assessment was made using measures R_0M (ratio of means), RMSA (root mean squared error) and E_{NS} (Nash-Sutcliffe efficiency index).

Wejście do updatingu	BC	Filtre			Updating			Wejście do prognozy	Prognoza		
		EnKF-BC	EnSRF-BC	En-SRF/VAR	120 h				72 h		
					R_0M	RMSE (m^3s^{-1})	E_{NS}		R_0M	RMSE (m^3s^{-1})	E_{NS}
					25-30 Marzec 2015				30 Marzec-02 Kwiecień 2015		
P_{OBS}	—	—	—	—	2,248	17,163	0,188	P_{OBS}	1,394	11,678	-0,996
$P_{05,24}$	√	—	—	—	1,155	12,959	0,304	P_{OBS}	1,031	9,634	0,353
$H_{14}^{2,89}$	—	√	—	—							
$P_{05,24}$	√	—	—	—	1,106	10,580	0,536	P_{OBS}	0,897	9,345	0,588
$H_{14}^{1,00}$	—	—	√	—							
$P_{05,24}$	√	—	—	—	0,954	10,358	0,556	P_{OBS}	1,072	6,023	0,747
$H_{14}^{1,00}$	—	—	—	√							

W przypadku prognozy z updatingiem metodą hybrydową (rys. 6.8) punkt startowy (czerwona linia) znajduje się najbliżej hydrogramu zaobserwowanego (niebieska linia). Prawie połowa horyzontu prognozy znajduje się wewnątrz lub w dużej bliskości prognozy wiązkowej (szare linie; rys. 6.8). Recesja w drugiej części hydrogramu była porównywalna do hydrogramu zaobserwowanego (niebieska linia) i nieco niższa niż w prognozie z updatingiem z niehybrydowym filtrem (rys. 6.7). Błąd prognozy w ostatnim kroku horyzontu był porównywalny dla obu filtrów. W updatingu i prognozowaniu w okresie zimy nie były dostępne dwie kluczowe ilościowe obserwacje związane z prognozą odpływu, tj. powierzchniowy rozkład pokrywy śnieżnej i rozkład zawartości wody w śniegu. Informacje na temat tych charakterystyk zawarte były tylko w zmiennych stanu modelu, które zostały ponownie przeliczone za pomocą modelu śniegu zastosowanego w HBV. W ramach updatingu nie asymilowano satelitarnych obserwacji pokrywy śniegu.

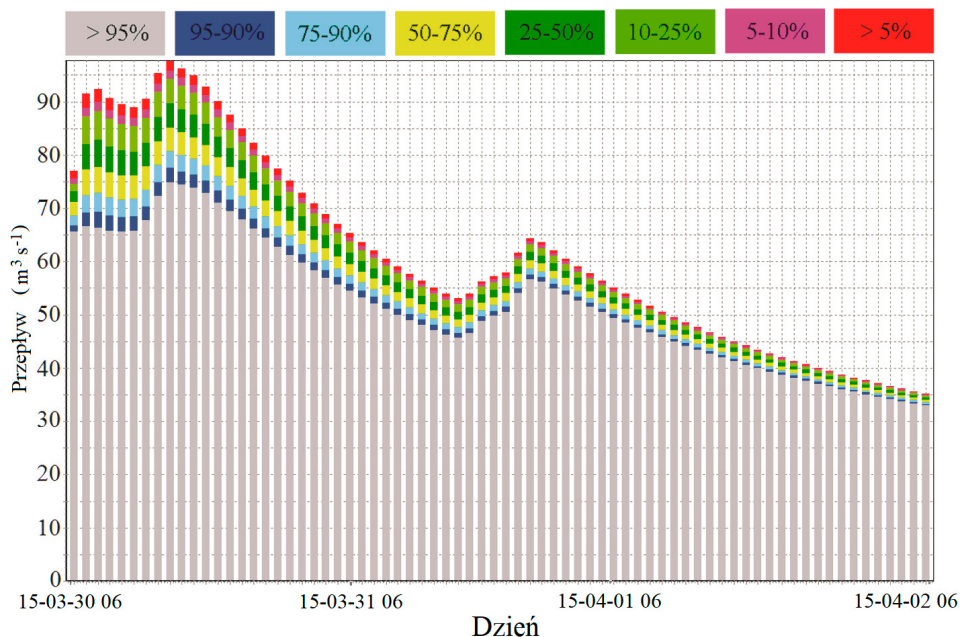


Rys. 6.7. Prognoza modelu HBV z updatingiem za pomocą EnKF-BC.
Fig. 6.7. HBV model forecast with updating by EnKF-BC.



Rys. 6.8. Prognoza modelu HBV z updatingiem za pomocą EnSRF/4DVAR.
Fig. 6.8. HBV model forecast with updating by EnSRF/4DVAR.

Na rysunku 6.9 przedstawiono tą samą prognozę (z updatingiem EnSRF/4DVAR), lecz wykonaną w skali błędów prognozy. Założony jest brak wiedzy o aktualnych (rzeczywistych) wartościach prognozowanego hydrogramu. Błąd prognozy *ex ante* rozłożony jest na hydrogramie w ośmiu klasach prawdopodobieństwa: >95%, 95-90%, 75-90%, 50-75%, 25-50%, 10-25%, 5-10% i >5%.



Rys. 6.9. Prognozowany hydrogram w skali błędów prognozy (niepewność prognozy).

Fig. 6.9. Forecasted hydrograph on the forecast error scale (forecast uncertainty).

6.3.3. Podsumowanie i wnioski

Opisana w niniejszym rozdziale platforma obliczeniowa dotyczyła procedury updatingu modelu hydrologicznego typu opad–odpływ HBV. Danymi wymuszającymi model HBV były, skorygowane metodą transformaty dystrybucyjnej, obserwacje opadu satelitarnego, tj. P_{03}^{BC} , P_{04}^{BC} i P_{0524}^{BC} . Updating hydrologicznego modelu HBV zrealizowano pięcioma metodami:

- 1) Symulacja modelu HBV ze zmienną wymuszającą w postaci skorygowanego opadu satelitarnego P_{0524}^{BC} bez asymilacji wilgotności gleby.
- 2) Bezpośrednie zastąpienie prognozowanych zmiennych stanu modelu HBV skorygowanymi metodą BC satelitarnymi obserwacjami wilgotności gleby $H_{14}^{1,00-BC}$ lub $H_{14}^{2,89-BC}$.
- 3) Asymilowanie nieskorygowanej wilgotności gleby, tj. $H_{14}^{1,00}$, $H_{14}^{2,89}$, za pomocą filtru EnKF.
- 4) Asymilowanie nieskorygowanej wilgotności gleby, tj. $H_{14}^{1,00}$, $H_{14}^{2,89}$, z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV, θ_{HBV} , za pomocą zmodyfikowanych filtrów EnKF-BC i EnSRF-BC.
- 5) Asymilowanie nieskorygowanej wilgotności gleby, tj. $H_{14}^{1,00}$, $H_{14}^{2,89}$ z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV, θ_{HBV} , za pomocą filtru EnSRF/4DVAR.

W metodzie usuwania obciążenia (BC) zastosowano metodę transformaty dystrybucyjnej z nowym algorytmem wykorzystującym trzy trójparametrowe teoretyczne rozkłady prawdopodobieństwa, tj. uo-

gólniony rozkład wykładniczy (GE), gamma (GA) i Weibulla (WE). Do generowania wolnych od obciążenia satelitarnych opadów i obserwacji wilgotności gleby oraz do usuwania obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu modelu HBV w zmodyfikowanych algorytmach wiązkowych filtrów EnKF-BC i EnSRF-BC użyto takich samych teoretycznych rozkładów prawdopodobieństwa (np. GE-GE, GA-GA, WE-WE itd.) lub różnych (np. GE-GA, GE-WE, GA-WE itd.). Oszacowania wilgotności gleby, wynikające z procedury asymilacji danych, były wrażliwe na obserwacje *in situ*, tj. obserwacje satelitarne wilgotności gleby i wariancję błędu modelu. Błąd modelu miał wpływ na kowariancję błędu między obserwacjami i stanem modelu oraz na skuteczność zastosowanej procedury asymilacji.

W ostatniej części rozdziału przedstawiono procedurę asymilacji danych (DA) zastosowanej do updatingu zmiennych stanu modelu HBV w krótkoterminowym prognozowaniu hydrologicznym. Wykonano symulację wezbrania zimowego z uwzględnieniem updatingu za pomocą procedury asymilacji, a prognozowany hydrogram porównano z prognozą bez updatingu. Wymuszeniem modelu HBV w fazie updatingu był skorygowany opad satelitarny P_{0524}^{BC} , a asymilację $H_{14}^{1,00}$, $H_{14}^{2,89}$ wykonano za pomocą filtrów wiązkowych i hybrydowych z uwzględnieniem usuwania obciążenia z zaburzonych wiązek prognozowanych zmiennych stanu. Prognozę hydrologiczną obliczono za pomocą symulacji modelu HBV dla potrzeb prognozowania metodą hindcast, czyli z opadem zaobserwowanym w naziemnej sieci pomiarowo-obserwacyjnej. Zastosowane do asymilacji algorytmy umożliwiły obliczenie prognozy wiązkowej (przedziałowej) i prognozy w skali błędów prognozy.

Zgodnie z uzyskanymi wynikami sformułowano następujące wnioski:

- 1) Najlepszą funkcją transformacji, na podstawie testu zgodności Kołmogorowa-Smirnowa, dla zaobserwowanego godzinowego opadu P_{OBS} był rozkład GE, a dla obserwacji satelitarnych WE.
- 2) Najlepszą funkcją transformacji dla θ_{HBV} był rozkład WE, zarówno zimą jak i latem, podczas gdy dla obserwacji satelitarnej $H_{14}^{1,00}$ rozkład GA, a dla $H_{14}^{2,89}$ odpowiednio WE i GA.
- 3) Usunięcie błędu systematycznego z opadów satelitarnych nieznacznie poprawiło losowe własności obserwacji zarówno podczas dopasowania rozkładów prawdopodobieństwa, jak i walidacji, natomiast usunięcie błędu z satelitarnych obserwacji wilgotności gleby było skuteczne tylko latem podczas dopasowania rozkładów i walidacji. Zimą obserwowano ujemne wartości wskaźnika skuteczności Nasha i Sutcliffea.
- 4) Testowane obserwacje satelitarne P_{0524}^{BC} bez asymilacji wilgotności gleby nie poprawiły symulowanego hydrogramu, szczególnie w lecie (prawdopodobnie ze względu na niepożądany wpływ konwekcyjnego opadu na odpływ). Jakość powierzchniowego rozkładu opadów, szczególnie w zlewni górskiej, była bardzo ważnym czynnikiem wpływającym na jakość symulowanego hydrogramu.
- 5) Po uwzględnieniu asymilacji satelitarnej wilgotności gleby, najlepszy model symulacyjny HBV uwzględnił latem asymilację $H_{14}^{2,89}$ za pomocą EnSRF-BC z wymuszeniem w postaci P_{0524}^{BC} , a zimą asymilację $H_{14}^{2,89}$ za pomocą EnKF-BC.
- 6) Model HBV uruchomiony dla potrzeb prognozy hydrologicznej z updatingiem (filtry EnKF-BC i EnSRF-BC) uzyskał wyższe miary jakości w porównaniu do modelu bez updatingu. Prognozy hydrologiczne wyznaczono na podstawie macierzy zmiennych stanu obliczonej w ostatnim kroku updatingu. Najefektywniejszą metodą updatingu okazało się hybrydowe połączenie filtra EnSRF-BC z metodą wariacyjną 4DVAR.

7. PODSUMOWANIE I WNIOSKI KOŃCOWE

W zaprezentowanej pracy zaprojektowano i przedstawiono dwa złożone modele hydrologiczne składające się z dwóch członów. Pierwszy człon utworzył preprocesor danych wejściowych do standardowego modelu hydrologicznego, natomiast drugi sam model hydrologiczny, przy czym należy zaznaczyć, że preprocesor danych wejściowych był także pewnym modelem probabilistycznym. Preprocesor danych wejściowych zastosowano do modeli regresyjnych i szwedzkiego deterministycznego modelu hydrologicznego typu opad–odpływ (HBV). W ten sposób w pracy utworzono zbiór metod (modeli) probabilistycznych w celu poprawienia dokładności prognoz hydrologicznych opracowywanych w Biurach Prognoz Hydrologicznych. Powyższy efekt uzyskano poprzez uwzględnienie w modelowaniu niepewności. W przeprowadzonych rozważaniach założono, że głównym źródłem błędnej prognozy jest niepewność danych wejściowych do modeli, a co za tym idzie błędy warunków brzegowych i początkowych wartości zmiennych stanu modelu hydrologicznego. W związku z tym, że modele matematyczne są uproszczonymi wersjami rzeczywistości, niedoskonałość matematycznego opisu fizycznych zależności, będących istotą rozpatrywanych w pracy zjawisk przyrody, ma wpływ na niedokładność symulacji modelu. Niebagatelną ujemną rolę w prognozowaniu hydrologicznym odgrywa także zmienność i losowość procesów przyrodniczych.

W pierwszej części pracy przedstawiono ogólne informacje na temat modelowania matematycznego uwzględniającego w obliczeniach niepewność. Zwrócono uwagę na zasady i problemy w modelowaniu procesów przyrodniczych występujących w zlewni, modelowaniu matematycznym oraz na praktykę inżynierską związaną z minimalizowaniem niepewności prognozy hydrologicznej. Ważnym elementem pracy był opis hydrologii testowych zlewni oraz analiza danych. Miało to na celu uzyskanie na ich podstawie użytecznej informacji i wniosków będących podstawą skonstruowania preprocesora danych wejściowych do modeli regresyjnych i deterministycznego modelu typu opad–odpływ HBV.

W ramach preprocesora danych wejściowych do modeli regresyjnych zidentyfikowano zdarzenia hydrologiczne będące przedmiotem modelowania matematycznego oraz wykonano analizę statystyczną zmiennych będących wejściem do modeli. W ramach analizy zbadano jednorodność statystyczną zmiennych prognozowanych i prognozujących, wyznaczono ilościowe charakterystyki prawdopodobieństwa, sprawdzono normalność rozkładu zmiennych prognozowanych i prognozujących oraz przeprowadzono normalizację poprzez ich logarytmowanie. Preprocesor danych wejściowych przetestował ich niezależność, tzn. zbadał czy obserwowana realizacja badanej zmiennej losowej miała w rzeczywistości charakter losowy, czy też jej wybór podlegał tendencji lub cykliczności. W zależności od uzyskanego wyniku zastosowana w pracy procedura uwzględniła testowanie stacjonarności niezależnej lub zależnej zmiennej losowej. W przypadku zależnej zmiennej losowej sprawdzenie stacjonarności poprzedziła identyfikacja trendu, ocena funkcji trendu oraz identyfikacja występowania i wpływu wahań okresowych. Uzyskane wyniki testów statystycznych były podstawą do sklasyfikowania danych jako odpowiednie wejście do statycznych lub dynamicznych modeli regresyjnych.

Preprocesor danych wejściowych do modelu HBV, oprócz opisanych powyżej analiz statystycznych, wybrał najlepiej dopasowane teoretyczne funkcje rozkładu prawdopodobieństwa dla obserwacji opadu pozyskanych z naziemnej sieci pomiarowo-obserwacyjnej i obserwacji satelitarnych oraz dla satelitarnej wilgotności gleby i zmiennej stanu wilgotności gleby modelu HBV. Uzyskane teoretyczne rozkłady prawdopodobieństwa były wejściem do procedury usuwania błędu systematycznego metodą korekcji obciążenia (BC) z satelitarnych obserwacji opadów i wilgotności gleby oraz asymilacji nieskorygowanej satelitarnej wilgotności gleby. Metoda asymilacji wilgotności gleby uwzględniła usuwanie obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV za pomocą zmodyfikowanych wiązkowych filtrów, tj.: EnKF-BC i EnSRF-BC, oraz hybrydowego filtra złożonego z EnSRF i filtra wariacyjnego 4DVAR.

W kolejnych częściach pracy opisano metody i modele użyte do analizy niepewności w prognozowaniu hydrologicznym, przedstawiono warunki ich stosowalności, podano klasyfikację oraz metodykę oceny modeli. Przedstawione algorytmy metod i modeli posłużyły do rozwiązania następujących problemów:

- 1) Modelowania wezbrania deszczowo-roztopowego za pomocą dwóch statystycznych liniowych modeli Bayesowskich (SLMB).

- 2) Modelowania całorocznego hydrogramu i krótkoterminowej prognozy przepływów z wyprzedzeniem 1-, 2- i 3-dobowym przy użyciu dwóch dynamicznych liniowych modeli Bayesowskich (DLMB).
- 3) Poprawy symulacji i krótkoterminowej prognozy przepływu z wyprzedzeniem 1-, 2- i 3-dobowym modelu typu opad–odpływ HBV za pomocą preprocesora danych wejściowych, realizującego korekcję obciążenia meteorologicznych obserwacji satelitarnych i asymilację obserwacji satelitarnej wilgotności gleby.
- 4) Uwzględnienia ilościowej niepewności w prognozowaniu hydrologicznym za pomocą opracowanych złożonych modeli hydrologicznych składających się z preprocesora danych wejściowych i standardowych modeli w postaci modeli regresyjnych i modelu HBV.

Modelowanie niepewności w prognozowaniu średnio- i długoterminowym zrealizowano, wykorzystując wnioskowanie i regułę Bayesowską. Utworzono wielowymiarowe zależności regresyjne z parametrami estymowanymi metodami Bayesowskimi. Prawdopodobieństwo zdarzenia losowego, np. wezbrania hydrologicznego, zostało opisane jako prawdopodobieństwo *a priori*, które na podstawie dodatkowych pomiarów i obserwacji przekształcono w tzw. prawdopodobieństwo wynikowe *a posteriori*, czyli miarę racjonalnego oczekiwania (prognozowania) wystąpienia analizowanego zdarzenia. Opracowano statyczne liniowe modele Bayesowskie (SLMB) w postaci wielokrotnej regresji liniowej (WLR) ze skalarną zmienną prognozowaną i wielowymiarowej wielokrotnej regresji liniowej (WWLR), w której zmienna prognozowana jest wektorem. Parametry występujące w równaniach SLMB oszacowano metodami Bayesowskimi (WLR-B, WWLR-B) i techniką odwrotnego modelowania (WLR-InvB, WWLR-InvB).

Do symulacji i krótkoterminowej prognozy hydrologicznej zaproponowano zmodyfikowane algorytmy wybranych dynamicznych liniowych modeli Bayesowskich, tj. VCR i DWR. Jedną z podstawowych własności tych modeli jest umiejętność odzwierciedlania zmieniającej się dynamiki opisywanego procesu. Realizowane jest to za pomocą liniowego updatingu zmiennych stanu modelu, w sposób zbliżony do dynamicznych zjawisk naturalnych, np. zjawisk będących cyklem hydrologicznym. Z reguły cykl hydrologiczny jest złożonym systemem, zatem dokładne symulacje i prognozy za pomocą modeli statycznych są mało precyzyjne. W niniejszej pracy połączono dynamiczne zależności regresyjne opisujące liniowe procesy z wnioskowaniem Bayesowskim. Opracowano dwa modele dynamiczne, tj. Varying Coefficient Regression (VCR) i Discount Weighted Regression (DWR), uwzględniające dynamiczny updatung zmiennych stanu modelu za pomocą mechanizmu uczenia się w postaci filtru Kalmana. W ramach wykonanej modyfikacji VCR i DWR, w algorytmach obu modeli została uwzględniona możliwość zapisywania zmiennych stanu modelu w każdym kroku przetwarzania, a następnie uruchamiania modelu ze zmiennymi stanu zapisanymi w poprzednim kroku. Zaimplementowane algorytmy dostosowano do przypadku uogólnionego, tj. dla wektora zmiennych prognozowanych. Wykonane modyfikacje rozszerzyły aplikacyjność obu modeli do uruchamiania VCR i DWR nie tylko dla potrzeb symulacji, ale również prognozy hydrologicznej.

Kolejnym ważnym zagadnieniem poruszonym w pracy była poprawa krótkoterminowych prognoz obliczanych za pomocą deterministycznego modelu hydrologicznego typu opad–odpływ (HBV). Zaproponowane rozwiązanie polegało na korygowaniu prognozowanych zmiennych stanu hydrologicznego modelu HBV w ramach preprocesora danych wejściowych. Zmiany wartości zmiennych stanu prowadzą do poprawy zgodności symulacji i obserwacji w momencie stawiania prognozy. W związku z wysokim poziomem niepewności oceny obszarowej pomiarów naziemnych, jako wejście do modelu hydrologicznego wykorzystano obserwacje satelitarne. Rozkład przestrzenny i czasowy danych satelitarnych często znacznie różni się od naziemnych obserwacji, zatem pomiary satelitarne powinny być asymilowane bezpośrednio przez symulacyjny model hydrologiczny albo za pomocą odpowiednio skonstruowanego preprocesora. W pracy skonstruowano matematyczny preprocesor danych wejściowych do modelu HBV, który posłużył do poprawy symulacji hydrologicznej poprzez zwiększenie efektywności updatingu. Skonstruowany preprocesor zawiera procedurę usuwania obciążenia (BC), czyli błędu systematycznego z meteorologicznych produktów satelitarnych opadu i wilgotności gleby oraz procedurę asymilacji wilgotności gleby umożliwiającą updatung prognozowanych zmiennych stanu modelu HBV. Opracowane rozwiązanie wykorzystuje obie procedury do łączenia własności predykcyjnych modelu hydrologicznego z rzeczywistymi danymi wejściowymi.

Warto zwrócić uwagę na innowacyjne elementy zawarte w powyższym rozwiązaniu. W algorytmie transformaty funkcji dystrybuanty użytej do korekcji obciążenia zastosowano trzy trójparametrowe teoretyczne rozkłady prawdopodobieństwa, tj.: uogólniony rozkład wykładniczy (GE), gamma (GA) i Weibulla (WE). Obliczenia wykonano dla takich samych rozkładów zmiennych losowych (np. GE-GE, GA-GA i WE-WE) lub dla różnych rozkładów (np. GE-GA, GE-WE i GA-WE). Posłużyły one do generowania satelitarnego opadu i wilgotności gleby wolnych od błędu systematycznego. Metodę korekcji BC zastosowano również w procesie asymilacji satelitarnych obserwacji wilgotności gleby. Problem asymilacji oparto na technice wiązkowych filtrów (ensemble Kalman filter – EnKF i ensemble square root filter – EnSRF). W filtrze wiązkowym zmienne stanu modelu są syntetycznie zaburzane. W takim przypadku obciążenie symulowanej wiązki zmiennych stanu wynika z nieliniowego charakteru naturalnych procesów hydrologicznych i wiązki zaburzonej szumem Gaussa. Obciążone zaburzenie skorygowano metodą transformaty dystrybuanty w ramach techniki postprocesowej (EnKF-BC i EnSRF-BC). Powyższe rozwiązanie jest nowym podejściem do tych zagadnień. W porównaniu do najczęściej stosowanego na świecie rozwiązania, tj. metody postprocesowej [Ryu i in. 2009], w zaproponowanym algorytmie nie jest uwzględniana średnia wartość obciążenia, ale rozkład prawdopodobieństwa obciążenia. Powyższe rozwiązanie uwzględnia zróżnicowanie obciążenia poszczególnych zaburzonych elementów wiązki w zależności od tego, w którym obszarze dystrybuanty występują.

Zaimplementowane w preprocesorze danych wejściowych do modelu HBV procedury asymilacji satelitarnej wilgotności gleby rozszerzono o deterministyczną 4D wariacyjną metodę asymilacji. 4DVAR połączono hybrydowo z algorytmami wiązkowych filtrów EnKF-BC i EnSRF-BC i tym samym utworzono dwie metody hybrydowe, tj. EnKF/4DVAR i EnSRF/4DVAR. Równania do updatingu prognozowanych zmiennych stanu w filtrach EnKF, EnSRF i w równaniu funkcji celu filtru wariacyjnego 4DVAR zaimplementowano w postaci macierzowej. Postać macierzowa filtrów umożliwiła asymilację nie tylko pojedynczej zmiennej stanu, ale również wektora zmiennych. Mogą to być zarówno zmienne stanu, które mają swoje odpowiedniki w obserwacjach (np. satelitarne obserwacje *pseudo in situ* wilgotności gleby), jak i nieobserwowane (niemierzone) zmienne modelu HBV, np. górne i dolne ograniczenie zawartości zbiornika wody gruntowej. W omówionym rozdziale podjęto próbę rozwiązania jeszcze jednego problemu prognostycznego. Mianowicie do ilościowego opisu niepewności prognozy hydrologicznej zastosowano prognozę przedziałową (wiązkową) i niepewność w skali błędów prognozy. Wynikiem działania filtrów wiązkowych i modeli hybrydowych jest asymilacja nieskorygowanej wilgotności gleby z uwzględnieniem usuwania obciążenia z zaburzonych prognozowanych zmiennych stanu wilgotności gleby modelu HBV. W konsekwencji otrzymywana jest macierz zaktualizowanych zmiennych stanu. Dla każdego elementu macierzy model HBV obliczył prognozowany hydrogram i w ten sposób utworzono prognozę przedziałową.

Opracowane metody i modele dotyczyły symulacji operatorów pojedynczych procesów fizycznych, tj. wybranych charakterystyk ilościowych wezbrania deszczowo-roztopowego i przepływu rzecznej. Zastosowano statystyczne metody badawcze uwzględniające długie ciągi pomiarowe, analizę danych wejściowych i syntezy metody (modelu). Zaproponowane rozwiązania matematyczne należą do kategorii czarnej skrzynki, których merytoryczną poprawność wnioskowania zapewniły prawa przyczynowo-skutkowe. Matematycznie opisano zjawiska i zachodzące między nimi powiązania w pewnym momencie lub w okresie przyjętym za jednostkę czasu oraz w powiązaniu z funkcją czasu, stosując w tym celu odpowiednio modelowanie statyczne i dynamiczne. Funkcje przekształcające we wszystkich zastosowanych rozwiązaniach były liniowe. Wyniki symulacji modeli probabilistycznych przedstawiono w przedziałach uwzględniających niepewność (błąd) oszacowywanych zmiennych losowych. Zastosowane w pracy rozwiązania dotyczyły zagadnień związanych z symulacją hydrologiczną i prognozowaniem krótko- i średnioterminowym.

W ostatniej części pracy przedstawiono zastosowanie złożonych modeli hydrologicznych do symulacji i prognozy hydrologicznej na przykładzie trzech zlewni testowych. Modele SLMB zastosowano do opracowania prognozy średnioterminowej wezbrania deszczowo-roztopowego w małej zlewni nizinnej rzeki Sokoły położonej w północno-wschodnim rejonie Polski. Walidacja modelu SLMB wykazała, że zastosowane szacowanie parametrów regresji metodami Bayesowskimi obniżyło wartość średniego odchylenia standardowego i tym samym zwiększyło dokładność prognozy. Obserwacje zmiennych prognozowanych mieściły się wewnątrz 80% przedziału niepewności modelu. Estymacja modelu SLMB wskazała, jaki wpływ na zmienne prognozowane w postaci skalara i wektora mają wybrane

zmiennie prognozujące. Kolejne kroki analizy Bayesowskiej umożliwiły scharakteryzowanie poszczególnych parametrów modelu przy użyciu wykresów, jak również podstawowych statystyk liczbowych. Zastosowane we wnioskowaniu Bayesowskim założenia aprioryczne, dotyczące różnych prawdopodobieństw *a priori*, wynikają z niejednoznacznej wstępnej wiedzy o analizowanym zjawisku i prowadzą do różnych konkluzji (prawdopodobieństw *a posteriori*). Silna zależność prawdopodobieństwa *a posteriori* od założeń apriorycznych świadczy o słabości wykonanych obserwacji i pomiarów. Jeżeli przyjmujemy, że prawdopodobieństwo *a priori* ma rozkład Gaussa, to rozkład prawdopodobieństwa *a posteriori* będzie należał do tej samej rodziny co rozkład *a priori* [Neal 1997]. W przypadku rozkładu Gaussa, gdy liczebność próby losowej n jest duża, parametry rozkładów *a priori* i finałowego *a posteriori* będą się wyraźnie różnić. Wpływ założeń *a priori* maleje, gdy n jest mała. W takim przypadku średnia z wartości *a posteriori* jest zbliżona do średniej wartości *a priori*. W związku z tym, że próba losowa obserwacji zastosowanych w analizowaniu wezbrania deszczowo-roztopowego w zlewni Sokołdy była małoliczebna, $n = 38$, w opracowaniu przyjęto, że nieinformacyjny rozkład *a priori* jest rozkładem apriorycznym Jeffreys'a opracowanym dla wiarygodności Gaussowskiej i dla przypadku, w którym wektor parametrów modelu θ jest nie znany (tzn. jest estymowany).

Z powyższej analizy wynika, że stosowanie wnioskowania Bayesowskiego do modelowania wezbrań deszczowo-roztopowych powinno być poprzedzone silną analizą statystyczną zbioru zmiennych prognozujących i prognozowanych. Powinna być również brana pod uwagę możliwość modyfikowania specyfikacji modelu, uwzględniająca wpływ innych zmiennych prognozujących, nieliniowe zależności oraz powiększenie liczby obserwacji. Na podstawie przeprowadzonych analiz i uzyskanych wyników oraz przy uwzględnieniu założeń i ograniczeń metody stwierdzono, że ze względów praktycznych stosowanie tych metod jest uzasadnione, zwłaszcza w odniesieniu do systemów hydrologicznych o liniowej odpowiedzi. W przypadku modelowania zlewni z nieliniową odpowiedzią systemu hydrologicznego lepszym rozwiązaniem jest szacowanie parametrów modelu techniką odwrotnego modelowania InvB.

Zmodyfikowane algorytmy modeli VCR i DWR zastosowano do symulacji rocznego hydrogramu i krótkoterminowej prognozy przepływu rzecznego z wyprzedzeniem 1-, 2- i 3-dobowym w przekroju wodowskazowym Wizna na rzece Narew. Oba dynamiczne modele porównano z wielokrotną liniową regresją (WLR) oraz ze sztucznymi sieciami neuronowymi (NN). Wszystkie cztery modele matematyczne miały takie same zbiory danych wejściowych oraz wykonywały obliczenia według tej samej metodyki dotyczącej kalibracji, walidacji i prognozy. Powyższy podział danych na podzbiory do kalibracji i walidacji umożliwił bezpośrednio porównanie wszystkich zastosowanych modeli. Do symulacji rocznego hydrogramu i do obliczenia krótkoterminowej prognozy hydrologicznej algorytmy NN, VCR i DWR wykorzystywały tylko jedno równanie regresji, natomiast WLR w celu osiągnięcia najlepszych wyników wymagała oddzielnych równań do symulacji hydrologicznej i do prognozy. Badanie wykazało, że model DWR dostarczył najdokładniejsze roczne symulacje hydrogramów, a także 1-, 2- i 3-dobowe prognozy, zwłaszcza w okresach wysokiego i niskiego stanu wody.

Modele VCR i DWR, w odróżnieniu od WLR i NN, uwzględniły w swoich algorytmach ewolucję wariancji połączoną z dynamiczną interakcją zmiennych prognozujących w czasie. Modele wzięty pod uwagę dynamiczne zmiany współczynników kierunkowych zmiennych prognozujących i tym samym zapewniły dokładność modelu symulacyjnego uruchamianego dla potrzeb prognozy. VCR i DWR były w stanie wychwycić zmieniającą się w czasie dynamikę modelowanego zjawiska. Umożliwiły uwzględnienie czynnika sezonowego odpowiedzialnego za cykliczność (okresowość) zachowania się wielu danych obserwacyjnych-pomiarowych związanych z system hydrologicznym. Łącząc analizę falkową (wavelet analysis) z modelami VCR i DWR, można utworzyć modele hybrydowe (WAV-DLMB). Otrzymany w ten sposób nowy model matematyczny mógłby lokalizować i wyodrębnić pewne zdarzenia częstotliwościowe z analizowanych szeregów czasowych i tym samym zwiększyć dokładność symulacji hydrologicznej i prognozy krótkoterminowej. Zaproponowane w pracy jednowymiarowe VCR i DWR zostały zaimplementowane w postaci wielowymiarowej, która pozwala na uruchomienie modelu z wektorową zmienną prognozowaną. Następnym etapem badań powinien eksplorować zastosowanie VCR i DWR do symulacji i prognozy hydrogramów dla wektorowej zmiennej prognozowanej oraz do symulacji i prognozy przepływu rzeczego w przekroju podłużnym, jak również w węzle hydrologicznym.

Korekcję obciążenia i asymilację meteorologicznych obserwacji satelitarnych wykonanych w ramach preprocesora danych wejściowych do modelu HBV zastosowano do poprawienia krótkotermino-

wej prognozy przepływu w górskiej zlewni rzeki Soły. Praktycznie dla wszystkich testowanych produktów satelitarnych, tj. P_{0524} , $H_{14}^{1,00}$ i $H_{14}^{2,89}$, metoda korekcji obciążenia w postaci transformaty funkcji dystrybuanty tylko nieznacznie przybliżyła ich poprawione wartości do wartości obserwowanych otrzymanych z naziemnej sieci pomiarowo-obserwacyjnej. Modelowanie opadu P_{OBS} z naziemnej sieci pomiarowo-obserwacyjnej za pomocą skorygowanych danych satelitarnych P_{0524}^{BC} było bardziej dokładne niż modelowanie za pomocą nieskorygowanych danych P_{0524} . Metoda korekcji obciążenia okazała się mało skuteczna w przypadku satelitarnych obserwacji wilgotności gleby $H_{14}^{1,00}$ i $H_{14}^{2,89}$, zwłaszcza w sezonie zimowym. Uzyskana rozbieżność między skutecznością modelowania wilgotności gleby w sezonie letnim a skutecznością modelowania w okresie zimowym wynika z faktu występowania zimą pokrywy śniegu i niskich temperatur w glebie na głębokości 0,05 m.

Skorygowane obserwacje satelitarne opadu P_{0524}^{BC} jako dane wymuszające w modelu HBV, ale bez uwzględnienia asymilacji wilgotności gleby, nie poprawiły symulowanego hydrogramu, szczególnie w lecie. Niepożądany wpływ na symulację odpływu miały opady konwekcyjne. Jakość powierzchniowego rozkładu opadów w zlewni górskiej była bardzo ważnym czynnikiem wpływającym na jakość symulowanego hydrogramu. Procedura asymilacji satelitarnej wilgotności gleby wyraźnie poprawiła symulację HBV. Zdecydowanie słabsze wyniki uzyskano w przypadku prognozy hydrologicznej. Wiązkowe filtry z uwzględnieniem usuwania obciążenia z wiązki zaburzonych zmiennych stanu, tj. EnKF-BC i EnSRF-BC, ogólnie działały słabo. Zdecydowaną poprawę efektywności updatingu zaobserwowano dla modelu hybrydowego EnSRF/4DVAR.

Otrzymane wyniki symulacji HBV i prognozy hydrologicznej wykazały, że zaproponowane procesory mogą być wykorzystane w operacyjnym uruchamianiu modelu HBV. Usuwanie obciążenia z wiązki zaburzonych prognozowanych zmiennych stanu modelu HBV poprawiło zarówno symulację, jak i updating poprzedzający prognozę hydrologiczną. W przyszłości prace badawcze powinny skoncentrować się na możliwości asymilacji szerszego zbioru satelitarnych obserwacji meteorologicznych (np.: pokrywy śniegu, zawartości wody w śniegu oraz nieobserwowanych zmiennych stanu modelu HBV, tj. górnego i dolnego ograniczenia zbiornika wody gruntowej) w celu zwiększenia efektywności updatingu, szczególnie w sezonie zimowym.

W każdym zaproponowanym w niniejszej pracy rozwiązaniu matematycznym niepewność prognozy hydrologicznej określono ilościowo w formie rozkładu prawdopodobieństwa. Rezultaty symulacji i prognozy hydrologicznej obliczonej za pomocą modeli SLMB i DLMB uwzględniły 80% przedział niepewności. Symulacja modelu HBV z asymilacją meteorologicznych obserwacji satelitarnych zawierała niepewność w postaci przedziału utworzonego za pomocą wiązki hydrogramów oraz niepewność w skali błędów prognozy. Opracowano algorytmy metod probabilistycznych spełniające wymagania operacyjnego uruchamiania modeli hydrologicznych. Algorytmy uwzględniły updating parametrów modeli SLMB oraz zmiennych stanu modeli DLMB i HBV. Dodatkowo zapewniono możliwość uruchamiania modeli z uaktualnionymi parametrami oraz zmiennymi stanu, tj. obliczonymi przez model w ostatnim kroku przetwarzania poprzedzającym prognozę hydrologiczną. Utworzono zbiór metod probabilistycznych, który powinien stanowić narzędzie alternatywne lub/i komplementarne do operacyjnie uruchamianych deterministycznych modeli hydrologicznych oraz umożliwić ich zastosowanie z uwzględnieniem szerokiego spektrum zlewni hydrologicznych.

BIBLIOGRAFIA

- Adamowski J., Chan H.F., Prasher S.O., Ozga-Zieliński B., Sliusarieva A., 2012, Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada, *Water Resources*, 48 (1), W01528, DOI: 10.1029/2010WR009945
- Adamowski J., Karapatakis C., 2010, Comparison of multivariate regression and artificial neural networks for peak urban water demand forecasting: the evaluation of different ANN learning algorithms, *Journal of Hydrologic Engineering*, 15 (10), 729-743, DOI: 10.1061/(ASCE)HE.1943-5584.0000245
- Adamowski J., Prokoph A., 2013, Assessing the impacts of the urban heat island effect on streamflow patterns in Ottawa, Canada, *Journal of Hydrology*, 496, 225-237, DOI: 10.1016/j.jhydrol.2013.05.032
- Ahmed K.F., Wang G., Silander J., Wilson A.M., Allen J.M., Horton R., Anyah R., 2013, Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast, *Global and Planetary Change*, 100, 320-332, DOI: 10.1016/j.gloplacha.2012.11.003
- Alvarez-Garreton C., Ryu D., Western A.W., Crow W.T., Robertson D.E., 2014, The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment, *Journal of Hydrology*, 519, 2763-2774, DOI: 10.1016/j.jhydrol.2014.07.041
- Alvarez-Garreton C., Ryu D., Western A.W., Su C.H., Crow W.T., Robertson D.E., Leahy C., 2015, Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes, *Hydrology and Earth System Sciences*, 19, 1659-1676, DOI: 10.5194/hess-19-1659-2015
- Amal S.H., 2006, Goodness-of-fit for the generalized exponential distribution, dostępne online: <http://interstat.stat-journals.net/YEAR/2005/articles/0507001.pdf>
- Anderson R.L., 1941, Distribution of the serial correlation coefficient, *Annual Mathematical Statistics*, 8 (1), 1-13
- Artigue G., Johannet A., Borrell V., Pistre S., 2012, Flash flood forecasting in poorly gauged basins using neural networks: case study of Gardon de Mialet basin (southern France), *Natural Hazards and Earth System Sciences*, 12 (11), 3307-3324, DOI: 10.5194/nhess-12-3307-2012
- Arulampalam M.S., Maskell S., Gordon N., Clapp T., 2002, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing*, 50 (2), 174-188, DOI: 10.1109/78.978374
- Babel M.S., Shinde V.R., 2011, Identifying prominent explanatory variables for water demand prediction using Artificial Neural Networks: A case study of Bangkok, *Water Resources Management*, 25 (6), 1653-1676, DOI: 10.1007/s11269-010-9766-x
- Baguis P., Roulin E., 2017, Soil moisture data assimilation in a hydrological model: a case study in Belgium using large-scale satellite data, *Remote Sensing*, 9 (8), 820, DOI: 10.3390/rs9080820
- Bain L.J., Engelhardt M., 2000, Introduction to probability and mathematical statistics, Cengage Learning, 656 s.
- Barker E., Kelsey J., 2015, Recommendation for random number generation using deterministic random bit generators, National Institute of Standards and Technology, SP800-90A, 128 s., dostępne online: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-90Ar1.pdf>
- Bartnik A., Jokiel P., 1997, Zmiany odpływu na obszarze Polski w latach 1971-1990 w świetle analiza jednorodności szeregów przepływu rzek, *Wiadomości IMGW*, 4, 67-86
- Belayneh A., Adamowski J., Khalil B., Ozga-Zieliński B., 2014, Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models, *Journal of Hydrology*, 508, 418-429, DOI: 10.1016/j.jhydrol.2013.10.052
- Ben C.Y., 1988, Flood hazards for nuclear power plants, *Nuclear Engineering and Design*, 110 (2), 213-219, DOI: 10.1016/0029-5493(88)90025-8
- Ben-Hur A., Horn D., Siegelman H.T., Vapnik V., 2001, Support Vector Clustering, *Journal of Machine Learning Research*, 2, 125-137
- Bennet J.C., Grose M.R., Post D.A., Ling F.L.N., Corney S.P., Bindoff N.L., 2011, Performance of quantile-quantile bias-correction for use in hydroclimatological projections, 19th International Congress on Modelling and Simulation, Perth, Australia, dostępne online: <https://www.mssanz.org.au/modsim2011/F5/bennett.pdf>
- Bergström S., 1976, Development and application of a conceptual runoff model for Scandinavian catchments, Lund Institute of Technology, University of Lund, Sweden, *Bulletin Series A*, 52, 134 s.
- Bergström S., 1992, The HBV model – its structure and applications, Swedish Meteorological and Hydrological Institute, 35 s., dostępne online: <https://www.smhi.se/en/publications/the-hbv-model-its-structure-and-applications-1.83591>
- Bergström S., 1995, The HBV model, [w:] *Computer Models of Watershed Hydrology*, V. Singh (red.), Water Resources Publications, Highlands Ranch, Colorado, USA, 443-476

- Bhadra A., Bandyopadhyay A., Singh R., Raghuwanshi N.S., 2010, Rainfall-runoff modeling: comparison of two approaches with different data requirements, *Water Resources Management*, 24 (1), 37-62, DOI: 10.1007/s11269-009-9436-z
- Box G.E.P., 1953, Non-normality and tests on variances, *Biometrika*, 40 (3-4), 318-335. DOI: 10.1093/biomet/40.3-4.318
- Brassel K.E., Reif D., 2010, A Procedure to generate Thiessen polygons, *Geographical Analysis*, 1, 289-303
- Brzeziński J., 2010, Zastosowanie uogólnionego rozkładu wykładniczego w analizie sezonowych przepływów maksymalnych rocznych, [w:] *Hydrologia w inżynierii i gospodarce wodnej*, B. Więzik (red), Monografie Komitetu Inżynierii Środowiska PAN, 68 (1), 71-81
- Byczkowski A., 1999, *Hydrologia*, t. 1-2, Wydawnictwo SGGW, Warszawa
- Campisi-Pinto S., Adamowski J., Oron G., 2012, Forecasting urban water demand via wavelet-denoising and neural network models. Case study: city of Syracuse, Italy, *Water Resources Management*, 26 (12), 3539-3558, DOI: 10.1007/s11269-012-0089-y
- Cannon A.J., 2008, Probabilistic multisite precipitation downscaling by an expanded Bernoulli-gamma density network, *Journal of Hydrometeorology*, 9, 1284-1300, DOI: 10.1175/2008JHM960.1
- Cannon A.J., 2012, Neural networks for probabilistic environmental prediction: Conditional Density Estimation Network Creation and Evaluation (CaDENCE) in R, *Computers & Geosciences*, 41, 126-135, DOI: 10.1016/j.cageo.2011.08.023
- Cappé O., Moulines E., Rydén T., 2005, *Inference in Hidden Markov Models*, Springer, New York, 652 s.
- Carpenter S., 2003, Regime shifts in lake ecosystems: pattern and variation, *Excellence in Ecology Series*, 15, International Ecology Institute, Luhe, Germany
- Chai T., Draxler R.R., 2014, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, 7 (3), 1247-1250, DOI: 10.5194/gmd-7-1247-2014
- Charemza W., Deadman D., 1997, *Nowa ekonometria, PWE*, Warszawa
- Chib S., Greenberg E., 1995, Understanding the Metropolis-Hastings algorithm, *The America Statistician*, 49 (4), 327-335, DOI: 10.2307/2684568
- Ciupak M., 2004, Rozwiązania modelowe na potrzeby prognozy wezbrania roztopowego i deszczowo-roztopowego. Część I. Analiza zbiorów zmiennych prognozujących i prognozowanych, *Wiadomości IMGW*, 27 (3-4)
- Ciupak M., 2007, Zastosowanie regresji Bayesa w celu polepszenia przedziałowej prognozy wezbrania roztopowego i deszczowo-roztopowego w profilu wodowskazowym Wizna na rzece Narew, *Przegląd Geofizyczny*, 1, 27-43
- Ciupak M., 2009, Zastosowanie regresji Bayesa do oceny niepewności prognozy hydrologicznej, *Komitet Gospodarki Wodnej PAN*, 30, 141 s.
- Ciupak M., Ceran M., Dominikowski M., Ozga-Zielińska M., 2009, Algorytm sprawdzalności prognoz hydrologicznych. Cz. I. Ocena wektorowych prognoz w punkcie i obszarze, *Wiadomości Meteorologii i Hydrologii Gospodarki Wodnej*, 3 (2-3), 49-75
- Ciupak M., 2010a, Statyczne i dynamiczne Bayesowskie modele w prognozowaniu hydrologicznym, [w:] *Hydrologia w inżynierii i gospodarce wodnej*, B. Więzik (red.), Monografie Komitetu Inżynierii Środowiska PAN, 68, 131-141
- Ciupak M., 2010b, Wybrane dynamiczne modele Bayesowskie w hydrologii, *Przegląd Geofizyczny*, 55 (1-2), 21-35
- Ciupak M., Ozga-Zieliński B., Adamowski J., Deo R.C., Kochanek K., 2019, Correcting satellite precipitation data and assimilating satellite-derived soil moisture data to generate ensemble hydrological forecasts within the HBV rainfall-runoff model, *Water*, 11 (10), 2138, DOI: 10.3390/W11102138
- Ciupak M., Ozga-Zieliński B., Adamowski J., Quilty J., Khalil B., 2015, The application of Dynamic Linear Bayesian Models in hydrological forecasting: Varying Coefficient Regression and Discount Weighted Regression, *Journal of Hydrology*, 530, 762-784, DOI: 10.1016/j.jhydrol.2015.10.023
- Clark M.P., Rupp D.E., Woods R.A., Zheng X., Ibbitt R.P., Slater A.G., Schmidt J., Uddstrom M.J., 2008, Hydrologic data assimilation with the ensemble Kalman filter: Use of stream-flow observations to update states in a distributed hydrological model, *Advances in Water Resources*, 31 (10), 1309-1324, DOI: 10.1016/j.advwatres.2008.06.005
- Cohn T.A., England J.F., Berenbrock C.E., Mason R.R., Stedinger J.R., Lamontagne J.R., 2013, A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series, *Water Resources Research*, 49 (8), 5047-5058, DOI: 10.1002/wrcr.20392
- Colin Cameron A., Windmeijer F.G., 1997, An R-squared measure of goodness of fit for some common nonlinear regression models, *Journal of Econometrics*, 77 (2), 329-342, DOI: 10.1016/S0304-4076(96)01818-0
- Darlington R.B., 1990, *Regression and linear models*, McGraw-Hill College, New York, 480 s.
- Dawid A. P., 1981, Some matrix-variate distribution theory: Notational considerations and Bayesian application, *Biometrika*, 68 (1), 265-274, DOI: 10.2307/2335827

- Demirel C.H., Mai J., Mendiguren G., Koch J., Samaniego L., Stisen S., 2017, Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model, *Hydrology and Earth System Sciences*, 22, 1299-1315, DOI: 10.5194/hess-22-1299-2018
- Dickey D.A., Fuller W.A., 1979, Distribution of the estimators for autoregressive time series with unit root, *Journal of the American Statistical Association*, 74, 427-431, DOI: 10.1080/01621459.1979.10482531
- Dowd M., Meyer R., 2003, A Bayesian approach to the ecosystem inverse problem, *Ecological Modelling*, 168 (1-2), 39-55, DOI: 10.1016/S0304-3800(03)00186-8
- Drécourt J-P., 2003, Kalman filtering in hydrological modelling, DAIHM Technical Report 2003-1, dostępne online: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.5877&rep=rep1&type=pdf>
- Durbin J., Watson G.S., 1971, Testing for serial correlation in least squares regression. III, *Biometrika*, 58 (1), 1-19, DOI: 10.2307/2334313
- Efron B., Tibshirani R., 1994, *An Introduction to the Bootstrap*, Chapman & Hally CRC Monographs on Statistics & Applied Probability, 456 s.
- Eisner S., Voss F., Kynast E., 2012, Statistical bias correction of global climate projections – consequences for large scale modeling of flood flows, *Advances in Geosciences*, 31, 75-82, DOI: 10.5194/adgeo-31-75-2012
- EUMESTAT, 2015a, Product User Manual (PUM) for product H03-PR-OBS-3A. Precipitation rate at ground by GEO/IR supported by LEO/MW, 2015, EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management, Doc. No: SAF/HSAF/PUM-03A, Revision 1.2, European Organisation for the Exploitation of Meteorological Satellites, dostępne online: http://hsaf.meteoam.it/documents/PUM/SAF_HSAF_PUM-03A_1_2.pdf
- EUMESTAT, 2015b, Product User Manual (PUM) for product H04-PR-OBS-4A. Precipitation rate at ground by LEO/MW supported by GEO/IR, 2015, EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management, Doc No: SAF/HSAF/PUM-04A, Revision 1.3, European Organisation for the Exploitation of Meteorological Satellites, dostępne online: http://hsaf.meteoam.it/documents/PUM/SAF_HSAF_PUM-04A_1_3.pdf
- EUMESTAT, 2015c, Product User Manual (PUM) for product H05-PR-OBS-5A. Accumulated precipitation at ground by blended MW and IR. EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management, Doc. No: SAF/HSAF/PUM-05A, Revision 1.3, European Organisation for the Exploitation of Meteorological Satellites, dostępne online: http://hsaf.meteoam.it/documents/PUM/SAF_HSAF_PUM-05A_1_3.pdf
- EUMESTAT, 2015d, Product User Manual (PUM) for product H14-SM-DAS-2. Soil Moisture Profile Index in the roots region by scatterometer data assimilation. EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management. Doc. No: SAF/HSAF/PUM-14, Revision 1.1, European Organisation for the Exploitation of Meteorological Satellites, dostępne online: http://hsaf.meteoam.it/documents/PUM/SAF_HSAF_PUM-14_1_1.pdf
- Evensen G., 1992, Using the extended Kalman filter with a multi-layer quasi-geostrophic ocean model, *Journal of Geophysical Research*, 97 (C11), 17905-17924, DOI: 10.1029/92JC01972
- Evensen G., 2003, The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343-367, DOI: 10.1007/s10236-003-0036-9
- Friedrich J.O., Adhikari N.K.J., Beyene J., 2008, The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study, *BMC Medical Research Methodology*, 8, 32, DOI: 10.1186/1471-2288-8-32
- Garen D.C., 1992, Improved techniques in regression-based streamflow volume forecasting, *Journal of Water Resources Planning and Management*, 118 (6), DOI: 10.1061/(ASCE)0733-9496(1992)118:6(654)
- Gelfan A.N., Motovilov Y.G., 2009, Long-Term hydrological forecasting in cold regions: Retrospect, current status and prospect, *Geography Compass*, 3 (5), 1841-1864, DOI: 10.1111/j.1749-8198.2009.00256.x
- Gelman A., Carlin J.B., Stern H.S., Rubin D.B., 1995, *Bayesian Data Analysis*, Chapman and Hall, London, 655 s.
- Genest C., Rémillard B., Beaudoin D., 2009, Godness-of-fit tests for copulas: a review and power study, *Insurance: Mathematics and Economics*, 44 (2), 199-213, DOI: 10.1016/j.insmatheco.2007.10.005
- Gentle J.E., 2003, *Random Number Generator and Monte Carlo Methods*, Springer-Verlag, New York, 381 s.
- Gibbs M., 1997, *Bayesian Gaussian Processes for Regression and Classification*, PhD thesis, University of Cambridge, dostępne online: <http://www.inference.phy.cam.ac.uk/mng10/GP/thesis.ps.gz>
- Górka J., 2007, Kurtoza w procesach generowanych przez model RCA GARCH, Prace i materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, 5, 577-589
- Grubbs F.E., Beck G., 1972, Extension of sample size and percentage points for significance test of outlying observations, *Technometrics*, 14 (4), 847-854, DOI: 10.2307/1267134
- Gudmundsson J.B., Bremnes J.E., Haugen J.E., Engen-Skaugen T., 2012, Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, *Hydrology and Earth System Sciences*, 16 (9), 3383-3390, DOI: 10.5194/Hess-16-3383-2012

- Guidelines for Flood Frequency Analysis. Long Measurement Series of River Discharge, 2005, Institute of Meteorology and Water Management, Warsaw, Poland, WMO HOMS Component I81.3.01, 44 s.
- Gupta A.K., Nagar D.K., 1999, Matrix Variate Distributions, Chapman and Hall/CRC Monographs and Surveys in Pure and Applied Mathematics no. 104, Chapman & Hall/CRC, New York, NY/Boca Raton, FL, 104 s.
- Gupta R.D., Kundu D., 1999, Generalized exponential distributions, Australia and New Zealand Journal of Statistics, 41 (2), 173-188
- Gupta R.D., Kundu D., 2001, Exponentiated exponential family; an alternative to gamma and Weibull, Biometrical Journal, 43, 117-130
- Halbe J., Adamowski J., Bennett E.M., Pahl-Wostl C., Farahbakhsh K., 2014, Functional organization analysis for the design of sustainable engineering systems, Ecological Engineering, 73, 80-91, DOI: 10.1016/j.ecoleng.2014.08.011
- Halbe J., Pahl-Wostl C., Sendzimir J., Adamowski J., 2013, Towards adaptive and integrated management paradigms to meet the challenges of water governance, Water Science and Technology, 67 (11), 2651-2660, DOI: 10.2166/wst.2013.146
- Hamed K.H., 2008, Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis, Journal of Hydrology, 349 (3-4), 350-363, DOI: 10.1016/j.jhydrol.2007.11.009
- Hamil T.M., Snyder C., 2000, A hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme, Monthly Weather Review, 128 (8), 2905-2919, DOI: 10.1175/1520-0493(2000)128<2905:AHEKVF>2.0.CO;2
- Haremba W., Deadman D., 1997, Nowa ekonometria, PWE, Warszawa
- Harvey A.C., 1986, Analysis and generalisation of multivariate exponential smoothing model, Management Science, 32 (3), 374-380, DOI: 10.1287/mnsc.32.3.374
- Hashino T., Bradley A.A., Schwartz S.S., 2007, Evaluation of bias-correction methods for ensemble streamflow volume forecasts, Hydrology and Earth System Sciences Discussions, 11 (2), 939-950, DOI: 10.5194/hess-11-939-2007
- Hawkins D., 1980, Identification of Outliers, Chapman and Hall, London-New York
- Helsel D.R., Hirsch R.M., 2002, Statistical methods on water resources, [w:] Techniques of Water-Resources Investigations of the United States Geological Survey, Book 4, Hydrologic Analysis and Interpretation, Chapter A3
- Hempel S., Frieler K., Warszawski L., Schewe J., Piontek F., 2013, A trend-preserving bias correction – the ISI-MIP approach, Earth System Dynamics, 4 (2), 219-236, DOI: 10.5194/esd-4-219-2013
- Hodges J.L., Lehmann E.L., 1963, Estimates of location based on rank tests, The Annals of Mathematical Statistics, 34 (2), 598-611, DOI: 10.1214/aoms/1177704172
- Holder R.L., 1985, Multiple Regression in Hydrology, Institute of Hydrology, Wallingford, Oxfordshire, 147 s.
- Houghton-Carr H.A., 2009, Assessment criteria for simple conceptual daily rainfall-runoff models, Hydrological Sciences Journal, 44 (2), 237-261, DOI: 10.1080/02626669909492220
- Howson C., Urbach P., 2005, The Bayesian Approach, Open Court Publishing Company, LaSalle, IL, 459 s.
- Hsu K.-L., Gupta H.V., Gao X., Sorooshian S., Imam B., 2002, Self-organizing linear output map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis, Water Resources Research, 38 (12), DOI: 10.1029/2001WR000795
- Jażwiński A.H., 1970, Stochastic Processes and Filtering Theory, Academic Press, New York, 376 s.
- Johansson B., 2013, IHMS Integrated Hydrological Modelling System, Manual Version 6.3, SMHI
- Johnell A., Johansson B., Pers C., Person M., 2011, Beskrivning av HBV – modellens Filer och Fel, SMHI-HBV
- Johnson N.L., Kotz S., Balakrishnan N., 1994, Continuous Univariate Distributions, 1. Wiley-Interscience, Wiley Series in Probability and Statistics, 761 s.
- Kaczmarek Z., 1977, Statistical Methods in Hydrology and Meteorology. (J. Bachrach, transl. of Metody statystyczne w hydrologii i meteorologii) Foreign Scientific Publications Dept. of the National Center for Scientific, Technical and Economic Information, Springfield, VA, 319 s.
- Kala R., 2002, Statystyka dla przyrodników, Poznań, Wydawnictwo Akademia Rolnicza w Poznaniu
- Kalbarczyk R., Kalbarczyk E., 2004, czasowo-przestrzenna struktura opadów atmosferycznych w okresie wegetacji różnych grup wczesności ziemniaka w Polsce, Acta Agrophysica, 4 (4), 687-697
- Kalman R.E., 1960, A new approach to linear filtering and prediction problems, Journal of Basic Engineering, 82 (Series D), 34-45
- Kang K.W., Park J.H., Ham K.J., 1993, Evaluation of hydrological forecasting system based on neural networks model, Proceedings of the 25th Congress of IAHR, IAHR Delft, 257-264
- Kiczko A., Romanowicz R.J., Osuch M., Pappenberger F., 2015, Adaptation of the integrated catchment system to on-line assimilation of ECMWF forecasts, [w:] Stochastic Flood Forecasting System, R.J. Romanowicz, M. Osuch (red.), GeoPlanet: Earth and Planetary Sciences, Springer International Publishing, 173-186, DOI: 10.1007/978-3-319-18854-6_1
- Kim T.-W., Valdés J.B., 2003, Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks, Journal of Hydrologic Engineering, 8 (6), 319-328, DOI: 10.1061/(ASCE)1084-0699(2003)8:6(319)

- Kisi Ö., 2004, River flow modeling using artificial neural networks, *Journal of Hydrologic Engineering*, 9 (1), 60-63
- Kisi Ö., 2007, Streamflow forecasting using different artificial neural network algorithms, *Journal of Hydrologic Engineering*, 12 (5), 532-539, DOI: 10.1061/(ASCE)1084-0699(2007)12:5(532)
- Kotz S., Nadarajah S., 2004, *Multivariate t-distributions and their applications*, Cambridge University Press, Cambridge, UK, 284 s.
- Kruskal W.H., Wallis W.A., 1952, Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association*, 47, 583-621, DOI: 10.2307/2280779
- Kruszewski A., Żelaziński J., 1985, Podsystem operacyjnego prognozowania hydrogramu odpływu powodziowego dla dorzecza Górnej Wisły, *Zeszyty Naukowe Politechniki Śląskiej. Seria: Automatyka*, 78, 81-89
- Krzyżko, M., 2009, *Podstawy wielowymiarowego wnioskowania statystycznego*, Wydawnictwo Naukowe, Poznań, 369 s.
- Kundzewicz Z.W., Napiórkowski J.J., 1986, Nonlinear models of dynamic hydrology, *Hydrological Science Journal*, 31 (2), 163-185, DOI: 10.1080/02626668609491038
- Kurnik B., Kajfež-Bogataj L., Ceglar A., 2012, Correcting mean and extremes in monthly precipitation from 8 regional climate models over Europe, *Climate of the Past Discussions*, 8 (2), 953-986, DOI: 10.5194/cpd-8-953-2012
- Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y., 1992, Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root?, *Journal of Econometrics*, 54 (1-3), 159-178, DOI: 10.1016/0304-4076(92)90104-Y
- Laiolo P., Gabellani S., Campo L., Cenci L., Silvestro F., Delogu F., Boni G., Rudari R., Puca S., Pisani A.R., 2015, Assimilation of remote sensing observations into a continuous distributed hydrological model: Impacts on the hydrologic cycle, 2015 IEEE International Geoscience and Remote Sensing Symposium, 1308-1311, DOI: 10.1109/IGARSS.2015.7326015
- Lambor J., 1962, *Metody prognoz hydrologicznych*, WKiŁ, Warszawa, 34 s.
- Lee H., Seo D-J., Koren V., 2011, Assimilation of stream-flow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states, *Advances in Water Resources*, 34 (12), 1597-1615, DOI: 10.1016/j.advwatres.2011.08.012
- Lee H., Zhang, Y., Seo D-J., Xie P., 2015, Utilizing satellite precipitation estimates for stream-flow forecasting via adjustment of mean field bias in precipitation data and assimilation of stream-flow observations, *Journal of Hydrology*, 529, 779-794, DOI: 10.1016/j.jhydrol.2015.08.057
- Lefebvre M., 2003, Short-term hydrological forecasts using linear regression, *Journal of Water Science*, 16 (2), 255-277, DOI: 10.7202/705507ar
- Lehmann E.L., 1959, *Testing Statistical Hypotheses*, John Wiley and Sons, Inc, New York, 767 s.
- Lehner B., Doll P., Alcamo J., Henrichs T., Kaspar F., 2006, Estimating the impact of global change on flood and drought risks in Europe: a continental, integrated analysis, *Climatic Change*, 75, 273-299, DOI: 10.1007/s10584-006-6338-4
- Lin Z., He L., Wu T., Xu C., 2018, The parameter estimation of the multivariate matrix regression models, *Statistics, Optimization and Information Computing*, 6 (2), 286-291, DOI: 10.19139/soic.v6i2.361
- Lindström G., Johansson B., Persson M., Gardelin M., Bergström S., 1997, Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, 201 (1-4), 272-288, DOI: 10.1016/S0022-1694(97)00041-3
- Lipski C., Kostuch R., Ryzek M., 2005, Charakterystyka hydrologiczna górnej części zlewni Soły na tle warunków fizjograficznych, klimatu i użytkowania, *Infrastruktura i ekologia terenów wiejskich*, PAN, Komisja Technicznej Infrastruktury Wsi, 2, 75-82
- Mahabir C., Hicks F.E., Robichaud C., Fayed A.R., 2006, Forecasting breakup water levels at Fort McMurray, Alberta, using multiple linear regression, *Canadian Journal of Civil Engineering*, 33 (9), 1227-1238, DOI: 10.1139/106-067
- Maraun D., 2013, Bias correction, quantile mapping, and downscaling: revisiting the inflation issue, *Journal of Climate*, 26 (6), 2137-2143, DOI: 10.1175/JCLI-D-12-00821.1
- Mardia K.V., Kent J.T., Bibby J.M., 1979, *Multivariate Analysis*, Academic Press, New York, 518 s.
- Marsaglia G., Tsang W.W., Wang J., 2003, Evaluating Kolmogorov's distribution, *Journal of Statistical Software*, 8 (18), DOI: 10.18637/jss.v008.i18
- Mashayekhi M., Mahjoub M., 1991, Snowmelt runoff forecasting: case study of Karadj reservoir, Iran, *Snow, Hydrology and Forest in High Alpine Areas*, 213-220
- McMillan H., Jackson B., Clark M., Kavetski D., Woods R., 2011, Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *Journal of Hydrology*, 400 (1-2), 83-94, DOI: 101016/j.jhydrol.2011.01.026

- McMillan H.K., Hreinsson E.Ö., Clark M.P., Singh S.K., Zammit C., Uddstrom M.J., 2013, Operational hydrological data assimilation with the recursive ensemble Kalman filter, *Hydrology and Earth System Sciences*, 17 (1), 21-38, DOI: 10.5194/hess-17-21-2013
- Mishra A.K., Desai V.R., 2006, Drought forecasting using feed-forward recursive neural networks, *Ecological Modelling*, 198 (1-2), 127-138, DOI: 10.1016/j.ecolmodel.2006.04.017
- Müller M.F., Thompson S.E., 2013, Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal, *Advances in Water Resources*, 60, 121-134, DOI: 10.1016/j.advwatres.2013.08.004
- Nash J.E., Sutcliffe J.V., 1970, River flow forecasting through conceptual models, Part I: A discussion of principles, *Journal of Hydrology*, 10 (3), 282-290, DOI: 10.1016/0022-1694(70)90255-6
- Neal R.M., 1997, Monte Carlo implementation of Gaussian process models for Bayesian regression and classification, Technical Report 9702, Department of Statistics, University of Toronto, 24 s., dostępne online: <http://www.cs.toronto.edu/~radford/ftp/mc-gp.pdf>
- Neyman J., Scott E., 1971, Outlier proneness of phenomena and of related distributions, *Optimizing Methods in Statistics*, Academic Press New York and London
- O'Hagan A., Forester J.J., 2004, *Kendall's Advanced Theory of Statistic 2B: Bayesian Inference*, Oxford University Press Inc, New York, 332 s.
- Osiewalski J., 1991, Bayesowska estymacja i predykcja dla jednorodniowych modeli ekonometrycznych, *Zeszyty Naukowe. Akademia Ekonomiczna w Krakowie, Seria Specjalna: Monografie*, 100, 193 s.
- Osuch M., Lawrence D., Meresa H.K., Napiórkowski J.J., Romanowicz R.J., 2016, Projected changes in flood indices in selected catchments in Poland in the 21st century, *Stochastic Environmental Research and Risk Assessment*, 31, 2435-2457, DOI: 10.1007/s00477-016-1296-5
- Ozga-Zielińska M., Nawalany M., 1979, *Zagadnienia identyfikacji i weryfikacji integralnego modelu zlewni*, Warszawa, Biblioteka Wiadomości IMUZ, 61
- Ozga-Zielińska M., Brzeziński J., 1997, *Hydrologia stosowana*, PWN, Warszawa, 326 s.
- Ozga-Zielińska M., Brzeziński J., Ozga-Zieliński B., 1997, Zasady obliczania największych przepływów rocznych o określonym prawdopodobieństwie przewyższenia. Długie ciągi pomiarowe, *Materiały Badawcze IMGW. Seria Hydrologia i Oceanologia*, 27
- Ozga-Zieliński B., 1999, Metody analizy niejednorodności ciągów pomiarowych zjawisk hydrologicznych, *Wiadomości Instytutu Meteorologii i Gospodarki Wodnej*, 2, 13-32
- Pearson F.R.S.K., 1931, On the inheritance of mental disease, *Annals of Human Genetics*, 4 (3-4), 362-380, DOI: 10.1111/j.1469-1809.1931.tb02085.x
- Phillips P.C.B., 1982, True characteristic function of F distribution, *Biometrika*, 69 (1), 261-264, DOI: 10.2307/2335882
- Piani C., Haerter J.O., Coppola E., 2009, Statistical bias correction for daily precipitation in regional climate models over Europe, *Theoretical and Applied Climatology*, 99, 187-192, DOI: 10.1007/s00704-009-0134-9
- Piani C., Weedon G.P., Best M., Gomes S.M., Viterbo P., Hagemann S., Haerter J.O., 2010, Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, *Journal of Hydrology*, 395 (3-4), 199-215, DOI: 10.1016/j.jhydrol.2010.10.024
- Piotrowski A., Napiórkowski J.J., Rowiński P.M., 2006, Flash-flood forecasting by means of neural networks and nearest neighbour approach – a comparative study, *Nonlinear Processes in Geophysics*, 13 (4), 443-448, DOI: 10.5194/npg-13-443-2006
- Pramanik N., Panda R.K., Singh A., 2011, Daily river flow forecasting using wavelet ANN hybrid models, *Journal of Hydroinformatics*, 13 (1), 49-63, DOI: 10.2166/hydro.2010.040
- Pruchnicki J., 1976, Zastosowanie testów nieparametrycznych do wykrywania niejednorodności ciągów klimatologicznych, *Przegląd Geofizyczny*, 21 (4), 263-270
- Rao C.R., Toutenburg H., 1999, *Linear Models: Least Squares and Alternatives*, Springer, New York, 427 s.
- Rasmussen J., Madsen H., Jensen K.H., Refsgaard J.C., 2015, Data assimilation in integrated hydrological modelling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance, *Hydrology and Earth System Sciences*, 19 (7), 2999-3013, DOI: 10.5194/hess-19-2999-2015
- Refsgaard J.C., 1997, Validation and intercomparison of different up-dating procedures for real-time forecasting, *Hydrology Research*, 28 (2), 65-84, DOI: 10.2166/nh.1997.0005
- Reichle R.H., Koster R.D., 2004, Bias reduction in short records of satellite soil moisture, *Geophysical Research Letter*, 31 (19), L19501, DOI: 10.1029/2004GL020938
- Reichle R.H., Walker J.P., Koster R.D., Houser P.R., 2002, Extended versus ensemble Kalman filtering for land data assimilation, *Journal of Hydrometeorology*, 3 (6), 728-740, DOI: 10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2
- Reis D.S., Stedinger J.R., Martins E.S., 2005, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation, *Water Resources Research*, 41 (10), DOI: 10.1029/2004WR003445

- Reynolds S.G., 1970, The gravimetric method of soil moisture determination. Part I. A study of equipment and methodological problem, *Journal of Hydrology*, 11 (3), 258-273, DOI: 10.1016/0022-1694(70)90066-1
- Rohatgi V.K., 1976, *An Introduction to Probability Theory and Mathematical Statistics*, Wiley Eastern Limited, 684 s.
- Rojas R., Feyen L., Dosio, A., Bavera D., 2011, Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulation, *Hydrology and Earth System Sciences*, 15 (8), 2599-2620, DOI: 10.5194/hess-15-2599-2011
- Rosen O., Thompson W.K., 2009, A Bayesian regression model for multivariate functional data, *Computational Statistics and Data Analysis*, 53 (11), 3773-3786, DOI: 10.1016/j.csda.2009.03.026
- Royle J.A., Chandler R.B., Sollmann R., Gardner B., 2014, *Spatial capture-recapture*, [w:] *GLMs and Bayesian Analysis*, B. Gardner (red.), Academic Press, 47-85
- Royston P., 1992, Approximating the Shapiro-Wilk W-test for normality, *Statistics and Computing*, 2, 117-119, DOI: 10.1007/BF01891203
- Ryu D., Crow W.T., Zhan X., Jackson T.J., 2009, Correcting unintended perturbation biases in hydrologic data assimilation, *Journal of Hydrometeorology*, 10 (3), 734-750. DOI: 10.1175/2008JHM1038.1
- Salvador M.P., Gargallo P., 2004, Automatic monitoring and intervention in multivariate dynamic linear model, *Computational Statistics and Data Analysis*, 47 (3), 401-431, DOI: 10.1016/j.csda.2003.11.024
- Salvi K., Kannan S., Ghosh S., 2011, Statistical downscaling and bias correction for projections of indian rainfall and temperature in climate change studies, 2011 International Conference on Environmental and Computer Science, IPCBEE, 19, 7-11, dostępne online: <http://www.ipcbee.com/vol19/2-ICECS2011R00006.pdf>
- Samuel J., Coulibaly P., Dumedah G., Moradkhani H., 2014, Assessing model state and forecasts variation in hydrologic assimilation, *Journal of Hydrology*, 513, 127-141, DOI: 10.1016/j.jhydrol.2014.03.048
- Saputro D.R.S., Amalia F., Widyarningsih P., Affan R.C., 2018, Parameter estimation of multivariate multiple regression model using Bayesian with non-informative Jeffreys' prior distribution, *Journal of Physics: Conference Series*, 1022, DOI: 10.1088/1742-6596/1022/1/012002
- Scharnagl B., Vrugt G.A., Vereecken H., Herbst M., 2011, Bayesian inverse modelling of in situ soil water dynamics: using prior information about the soil hydraulic properties, *Hydrology and Earth System Sciences Discussions*, 8, 2019-2063, DOI: 10.5194/hessd-8-2019-2011
- Seibert J., Vis M.J.P., 2012, Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, *Hydrology and Earth System Sciences*, 16 (9), 3315-3325. DOI: 10.5194/hess-16-3315-2012
- Sequi P.Q., Ribes A., Martin E., Habets F., Boé J., 2010, Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins, *Journal of Hydrology*, 383 (1-2), 111-124, DOI: 10.1016/j.jhydrol.2009.09.050
- Sharma D., Gupta A.D., Babel M.S., 2007, Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand, *Hydrology and Earth System Sciences*, 11 (4), 1373-1390, DOI: 10.5194/hess-11-1373-2007
- Sidor K., 2018, Porównanie wysokości opadu SYNOP z produktem satelitarnym IMERG dla 10 wybranych dni w roku 2017, praca magisterska, Politechnika Krakowska, Wydział Inżynierii Środowiska, Instytut Inżynierii i Gospodarki Wodnej, 63 s.
- Sikorska A.E., Scheidegger A., Banasik K., Rieckermann J., 2012, Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, *Hydrology and Earth System Sciences*, 16 (4), 1221-1236, DOI: 10.5194/hess-16-1221-2012
- Sinay M.S., Hsu J.S.J., 2014, Bayesian inference of a multivariate regression model, *Journal of Probability and Statistics*, 673657, DOI: 10.1155/2014/673657
- Solhjell I.K., 2009, *Bayesian Forecasting and Dynamic Models Applied to Strain Data from the Göta River Bridge*, University of Oslo, Statistics Division, 87 s.
- Sollis R., 2012, *Empirical finance for finance and banking*, John Wiley & Sons Ltd., New York, 345 s.
- Spearman C., 1904, The proof and measurement of association between two things, *The American Journal of Psychology*, 15, 72-10. DOI: 10.2307/1412159
- Steeb W.H., 2006, *Problems and Solutions in Introductory and Advanced Matrix Calculus*, World Scientific Publishing, Hackensack, NJ, 239 s.
- Strupczewski W., 1967, Przekształcanie rozkładów zmiennych hydrologicznych i meteorologicznych do rozkładu normalnego, *Wiadomości IMGW*, 2
- Sun L., Nistor J., Seidou O., 2015, Stream-flow assimilation in SWAT model using Extended Kalman Filter, *Journal of Hydrology*, 531 (3), 671-684, DOI: 10.1016/j.jhydrol.2015.10.060
- Sudipto B., 2000, MCMC algorithms for fitting Bayesian models, University of Minnesota, dostępne online: <http://www.biostat.umn.edu/~ph7440/pubh7440/Lecture8Ppt.pdf>
- Tawfik M., 2003, Linearity versus non-linearity in forecasting Nile River flows, *Advances in Engineering Software*, 34 (8), 515-524, DOI: 10.1016/S0965-9978(03)00039-5

- Telford P.J., Braesicke P., Morgenstern O., Pyle J.A., 2008, Technical Note: Description and assessment of a nudged version of the new dynamics Unified Model, *Atmospheric Chemistry and Physics*, 8 (6), 1701-1712, DOI: 10.5194/acp-8-1701-2008
- Thacker W.C., Long R.B., 1988, Fitting dynamics to data, *Journal of Geophysical Research. Oceans*, 93 (C2), 1227-1240, DOI: 10.1029/JC093iC02p01227
- Themeßl M. J., Gobiet A., Leuprecht A., 2011, Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *International Journal of Climatology*, 31 (10), 1530-1544, DOI: 10.1002/joc.2168
- Themeßl M.J., Gobiet A., Heinrich G., 2012, Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, *Climatic Change*, 112, 449-468, DOI: 10.1007/s10584-011-0224-4
- Thibault A., Anctil F., 2015, On the difficulty to optimally implement the Ensemble Kalman filter; An experiment based on many hydrological models and catchments, *Journal of Hydrology*, 529, 1147-1160, DOI: 10.1016/j.jhydrol.2015.09.036
- Triantafyllopoulos K., 2002, On Observational Variance Learning for Multivariate Bayesian Time Series and Related Models, PhD thesis, Department of Statistics, University of Warwick, 279 s., dostępne online: <http://webcat.warwick.ac.uk/record=b1378015~S1>
- Triantafyllopoulos K., Pikoulas J., 2002, Multivariate Bayesian regression applied to the problem of network security, *Journal of Forecasting*, 21 (8), 579-594, DOI: 10.1002/for.844
- Triantafyllopoulos K., 2006, Multivariate discount weighted regression and local level models, *Computational Statistics and Data Analysis*, 50 (12), 3701-3720, DOI: 10.1016/j.csda.2005.07.003
- Trzęsiok M., 2007, Identyfikacja obserwacji oddalonych z wykorzystaniem metody wektorów nośnych, [w:] *Taksonomia*, 14, Klasyfikacja i analiza danych – teoria i zastosowania, K. Jajuga, M. Walesiak (red.), Prace Naukowe, 1169, Wydawnictwo Akademii Ekonomicznej, Wrocław, 350-357
- Uhlenbrook S., 2006, Catchment hydrology – a science in which all processes are preferential, *Hydrological Processes*, 20 (16), 3581-3585, DOI: 10.1002/hyp.6564
- Verzano K., 2009, Climate change impacts on flood related hydrological processes: further development and application of a global scale hydrological model, *Reports on Earth System Science*, DOI: 10.17617/2.993926
- Wald A., Wolfowitz J., 1940, On a test whether two samples are from the same population, *The Annals of Mathematical Statistics*, 11 (2), 147-162, DOI: 10.1214/aoms/1177731909
- Walsh B., 2004, Markov Chain Monte Carlo and Gibbs sampling, *Lecture Notes for EEB*, dostępne online: <http://nitro.biosci.arizona.edu/courses/EEB581-2006/handouts/Gibbs.pdf>
- Welch G., Bishop G., 2006, An Introduction to the Kalman Filter, Technical Report, University of North Carolina at Chapel Hill, Nc, USA
- West M., Harrison P.J., 1999, *Bayesian forecasting and dynamic models*, Springer-Verlag, New York, 680 s.
- Węglarczyk S., 1998, Wybrane problemy hydrologii stochastycznej, Monografia, 235, Seria Inżynieria Sanitarna i Wodna, Wydawnictwo PK, Kraków, 214 s.
- Węglarczyk S., 2010, *Statystyka w inżynierii środowiska*, Wydawnictwo PK, Kraków, 375 s.
- Whitaker J.S., Hamill T.M., 2002, Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, 130 (7), 1913-1924, DOI: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2
- Wilby R.L., Hassan H., Hanaki K., 1998, Statistical downscaling of hydrometeorological variables using general circulation model output, *Journal of Hydrology*, 205 (1-2), 1-19, DOI: 10.1016/S0022-1694(97)00130-3
- Wood A.W., Leung L.R., Sridhar V., Lettenmaier D.P., 2004, Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, 62, 189-216, DOI: 10.1023/B:CLIM.0000013685.99609.9e
- Wöhling T., Lennartz F., Zappa M., 2006, Technical Note: Updating procedure for flood forecasting with conceptual HBV-type models, *Hydrology and Earth System Sciences*, 10 (6), 783-788, DOI: 10.5194/hess-10-783-2006
- Xiong J., Huang X.L., Cao Z.Y., 2011, Assimilating observation data into hydrological model with ensemble Kalman filter, *Advances Materials Research*, 255-260, 3632-3636, DOI: 10.4028/www.scientific.net/AMR.255-260.3632
- Xiong L., O'Connor K.M., 2002, Comparison of four updating models for real-time river flow forecasting, *Hydrological Sciences Journal*, 47 (4), 621-639, DOI: 10.1080/02626660209492964
- Xie X., Zhang D., 2010, Data assimilation for distributed hydrological catchment modelling via ensemble Kalman filter, *Advances in Water Resources*, 33 (6), 678-690, DOI: 10.1016/j.advwatres.2010.03.012
- Yang X., Michel C., 2001, Flood forecasting with a watershed model: A new method of parameter updating, *Hydrological Sciences Journal*, 45 (4), 537-547, DOI: 10.1080/02626660009492355
- Yevjevich V., Jeng R.I., 1969, Properties of non-homogeneous hydrologic series, *Hydrology Papers Colorado State University*, 32, 33 s.

- Young A., 1977, A Bayesian approach to prediction using polynomials, *Biometrika* 64 (2), 309-317, DOI: 10.1093/biomet/64.2.309
- Zhang F., Zhang M., Hansen I, A., 2009, Coupling ensemble Kalman filter with four-dimensional variational data assimilation, *Advances in Atmospheric Sciences*, 26 (1), 1-8, DOI: 10.1007/s00376-009-0001-8
- Zhang Y., Wu H., Cheng L., 2012, Some new deformation about variance and covariance, *Proceedings of 4th International Conference on Modelling, Identification and Control*, Wuhan, Chine, June, 24-26, 1042-1047
- Zhu S., Kai Y., Yihong G., 2007, Predictive matrix variate t models, [w:] *NIPS'07: Advances in Neural Information Processing Systems*, I.C. Platt, D. Koller, Y. Singer, S. Roweis (red.), MIT Press Cambridge, 20, 1721-1728

UNCERTAINTY IN HYDROLOGICAL FORECASTING

Summary

The aim of the work is to present a set of existing and author's own solutions of the probabilistic methods and models taking into account uncertainty modeling and application in hydrological practice to improve the hydrological simulation and forecast. The monograph considers two types of complex hydrological models consisting of a preprocessor of input data to regression models and a Swedish rainfall-runoff model (Hydrologiska Byråns Vattenbalansavdelning) HBV, while the input preprocessor is also a probabilistic model. The preprocessor of input data for regression models identifies hydrological events that are the subject of mathematical modeling, determines quantitative characteristics of the predictands and predictors, and performs statistical analysis of the input variables using statistical methods. On the other hand, the data preprocessor for the HBV model processes satellite precipitation observations, which are the forcing variables the hydrological model, and assimilates satellite observations of soil moisture.

Modeling of snowmelt floods was performed using static linear Bayesian models (SLMB) with parameters estimated by Bayesian method and inverse modeling technique using multiple and multivariate multiple linear regression (MLR-B, MMLR-B, MLR-InvB, MMLR-InvB). The modeling of the simulation of annual hydrographs and short-term hydrological forecasts were calculated using dynamic linear Bayesian models (DLMB), i.e. Varying Coefficient Regression (VCR) and Discount Weighted Regression (DWR). The obtained simulation results of DLMB were compared with multiple linear regression (MLR) and artificial neural networks (NN). To remove the systematic bias error from satellite meteorological observations (precipitation and soil moisture), the distribution-derived transformation algorithm (BC) was used, using three three-parameter probability distributions, i.e. generalized exponential (GE), gamma (GA) and Weibull (WE). The statistical transformation was performed for the same probability distributions (e.g. GE-GE, GA-GA, WE-WE) or for different (e.g. GE-GA, GE-WE, GA-WE, etc.). A novelty in the construction of the input data preprocessor to the HBV model was the use of the BC algorithm in the procedure of assimilation of satellite observations of soil moisture to generate a bias-free ensemble of predicted variables of the HBV model state in Kalman filters (EnKF-BC and EnSRF-BC) and in combination with a 4D variational filter (4DVAR) to create hybrid models (EnKF/4DVAR and EnSRF/4DVAR). The uncertainty of the hydrological forecast in all the models used was quantified in the form of a probability distribution using an interval forecast or a forecast on the error scale.

The proposed methods are not intended to replace of the used deterministic hydrological models, but to create a set of mathematical alternatives or complementary approaches to existing solutions built on the basis of physics. The probability and statistics used look for the regularity, the laws of nature and they extract the maximum of information from the observed phenomena in the past for conclude what should be expected in the future. In order to emphasize the effectiveness and universality of the proposed probabilistic methods in hydrological forecasting, examples of applications include both the large and small lowland catchment area and the mountain catchment area. An important element of this work is the application of all proposed methods in operating mode, i.e., in forecasting mode.

Keywords

Theory and Bayesian inference, mathematical model, uncertainty, bias correction, assimilation of satellite meteorological observations, Kalman filters, variational filter, hybrid models.

ZAŁĄCZNIK

Dodatek A. Brzegowe rozkłady prawdopodobieństwa zastosowane w metodzie usuwania obciążenia

Gupta i Kundu [1999] zauważyli, że uogólniony rozkład wykładniczy (GE) wykazuje podobieństwo w odniesieniu do funkcji gęstości $f(x)$ i funkcji kwantylowej x_p do rozkładów gamma (GA) i Weibulla (WE). Powyższe rozkłady mogą być opisane trzema parametrami: α (gdzie $\alpha > 0$), λ (gdzie $\lambda > 0$) i ε (gdzie $\varepsilon \leq x \leq +\infty$), reprezentującymi odpowiednio współczynnik skali, kształtu i dolne (lewostronne) ograniczenia rozkładu prawdopodobieństwa. Funkcja gęstości i funkcja kwantylowa trójparametrowego rozkładu GA jest zapisana jako:

$$f(x) = \frac{(x-\varepsilon)^{\lambda-1}}{\alpha^\lambda \Gamma(\lambda)} \exp\left(-\frac{x-\varepsilon}{\alpha}\right) \quad (1)$$

$$x_p = \varepsilon + \alpha t_p(\lambda) \quad (2)$$

gdzie: $\Gamma(\lambda) = \int_0^\infty t^{\lambda-1} \exp(-t) dt$ – funkcja gamma Eulera; x – zaobserwowana realizacja zmiennej losowej X ; $t_p(\lambda)$ – kwantyl standaryzowany rozkładu gamma o prawdopodobieństwie przewyższenia p .

Uogólniony rozkład wykładniczy może być zapisany jako:

$$f(x) = \frac{\lambda}{\alpha} \exp\left(-\frac{x-\varepsilon}{\alpha}\right) \left[1 - \exp\left(-\frac{x-\varepsilon}{\alpha}\right)\right]^{\lambda-1} \quad (3)$$

$$x_p = \varepsilon - \alpha \ln \left[1 - (1-p)^{\frac{1}{\lambda}}\right] \quad (4)$$

Rozkład Weibulla określony jest jako:

$$f(x) = \frac{\lambda}{\alpha} \left(\frac{x-\varepsilon}{\alpha}\right)^{\lambda-1} \exp\left[-\left(\frac{x-\varepsilon}{\alpha}\right)^\lambda\right] \quad (5)$$

$$x_p = \alpha \left[-\ln(1 - (1-p))\right]^{\frac{1}{\lambda}} + \varepsilon \quad (6)$$

W sytuacji, w której parametr kształtu λ jest równy 1.0, rozkłady GA i WE upraszczają się do rozkładu wykładniczego, wskazując tym samym, że są różnymi formami uogólnionego rozkładu wykładniczego [Gupta, Kundu 2001]. Uogólnienie polega na innym sposobie wprowadzenia do rozkładu GA i WE parametru kształtu λ [Brzeziński 2010].

Dodatek B. Metoda symulacji Monte Carlo (MCMC) do uzyskiwania rozkładów *a posteriori*

Do uzyskiwania rozkładów *a posteriori* stosowane jest wnioskowanie oparte na metodzie symulacji Monte Carlo, tzw. MCMC (Markov Chain Monte Carlo – metoda Monte Carlo przy użyciu łańcuchów Markowa). Szczegółowymi algorytmami realizującymi takie próbkowanie jest metoda Metropolis–Hastingsa [Chib, Greenberg 1995] lub generator Gibbsa [Walsh 2004]. Metoda MCMC umożliwia wygenerowanie listy wartości parametrów θ_i próbkujących rozkład *a posteriori*. Wartości oczekiwane parametrów uzyskuje się w wyniku uśrednienia po próbkach wylosowanych z rozkładu *a posteriori*. W modelu SLMB zastosowano generator Gibbsa. Generator Gibbsa bierze pod uwagę tylko jednowymiarowe rozkłady warunkowe. Najczęściej jest to rozkład normalny lub odwrotne rozkłady $(\chi^2)^{-1}$ i γ^{-1} . W algorytmie Gibbsa realizowana jest pojedyncza symulacja sekwencji d zmiennych jednowymiarowego warunkowego rozkładu, gdzie d oznacza liczbę parametrów (np. liczbę zmiennych prognozujących dla wielokrotnej regresji liniowej). Rzadziej stosuje się generowanie pojedynczego d wymiarowego wektora zmiennych losowych.

Rozkład *a posteriori* wektora parametrów θ można wyrazić jako $p(\theta|\mathbf{Y}) \propto p(\mathbf{Y}|\theta)p(\theta)$, gdzie: $p(\theta)$ jest rozkładem *a priori* i $p(\mathbf{Y}|\theta)$ oznacza Bayesowską funkcję wiarygodności. Niech $\theta = (\theta_1, \dots, \theta_d)$ będzie zbiorem wszystkich parametrów modelu regresji. Algorytm Gibbsa przy pomocy generatora liczb losowych tworzy zbiór x_g jednorodnie próbkujący przedział $[0,1]$. Następnie traktując te liczby jako wartości dystrybuanty $F(\theta)$ próbkowanego rozkładu przekształca je (odwraca) w próbkę θ_g , $F(\theta) = x_g$. Generator Gibbsa rozpoczyna tworzenie łańcucha Markowa ze zbiorem wartości początkowych $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_d^0)$. Powyższa operacja wykonywana jest g -krotnie dla $g = 0, \dots, M$. Wygenerowanie jednej realizacji θ^g w g -tym cyklu losowania Gibbsa polega na losowaniu z warunkowych rozkładów *a posteriori*. W rezultacie wylosowana jest pełna próba z wielowymiarowego warunkowego rozkładu oraz próbka Gibbsa w postaci wektora $\theta^1, \dots, \theta^M$. Przykładowy algorytm przybiera następującą postać:

```
for(int g = 1; g < M + 1; g++) {
 $\theta_1^g \sim p(\theta_1|\theta_2^{g-1}, \theta_3^{g-1}, \theta_4^{g-1}, \dots, \theta_d^{g-1}; \mathbf{Y})$ 
 $\theta_2^g \sim p(\theta_2|\theta_1^g, \theta_3^{g-1}, \theta_4^{g-1}, \dots, \theta_d^{g-1}; \mathbf{Y})$ 
 $\theta_3^g \sim p(\theta_3|\theta_1^g, \theta_2^g, \theta_4^{g-1}, \dots, \theta_d^{g-1}; \mathbf{Y})$ 
 $\theta_4^g \sim p(\theta_4|\theta_1^g, \theta_2^g, \theta_3^g, \dots, \theta_d^{g-1}; \mathbf{Y})$ 
...
 $\theta_k^g \sim p(\theta_k|\theta_1^g, \theta_2^g, \theta_3^g, \dots, \theta_{k-1}^g, \theta_{k+1}^g, \dots, \theta_d^{g-1}; \mathbf{Y})$ 
...
 $\theta_d^g \sim p(\theta_d|\theta_1^g, \theta_2^g, \theta_3^g, \dots, \theta_k^g, \theta_{k+1}^g, \dots, \theta_{d-1}^g; \mathbf{Y})$ 
}
```

Cykle powtarzane są aż do osiągnięcia zbieżności. Otrzymywanych jest M wektorów z następującym zestawem parametrów:

$$\begin{aligned} \theta^0 &= (\theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0, \dots, \theta_d^0) \\ &\dots \\ \theta^1 &= (\theta_1^1, \theta_2^1, \theta_3^1, \theta_4^1, \dots, \theta_d^1) \\ \theta^2 &= (\theta_1^2, \theta_2^2, \theta_3^2, \theta_4^2, \dots, \theta_d^2) \\ &\dots \\ \theta^M &= (\theta_1^M, \theta_2^M, \theta_3^M, \theta_4^M, \dots, \theta_d^M) \end{aligned}$$

Tak skonstruowany łańcuch powinien być zbieżny, stacjonarny i równoważny rozkładowi, który jest dokładnie rozkładem *a posteriori* $p(\theta|\mathbf{Y})$ [Sudipto 2000]. Początkowe elementy łańcucha zostały pominięte, gdyż uważa się, że ich rozkład prawdopodobieństwa jest obciążony wpływem przyjęcia dowolnego punktu startowego θ^0 [Barker, Kelsey 2015].

Dodatek C. Wielowymiarowe i macierzowe rozkłady prawdopodobieństwa występujące w statycznych i dynamicznych modelach Bayesowskich

Wielowymiarowy rozkład normalny

Dla zmiennej losowej $\mathbf{X} = [x_1, \dots, x_n]^T$ gęstość n -wymiarowego rozkładu normalnego ma postać:

$$f_{\mu, \Sigma}(\mathbf{X}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{X} - \mathbf{m})^T \Sigma^{-1} (\mathbf{X} - \mathbf{m})\right) \quad (7)$$

gdzie: $\mathbf{X} = [x_1, \dots, x_n]^T$ jest wektorem losowym o wektorze wartości oczekiwanych $\mathbf{m} = [m_1, \dots, m_n]^T$ i macierzy kowariancji Σ . Rozkład oznacza się zapisem $\mathbf{X} \sim N(\mathbf{m}, \Sigma)$.

Wielowymiarowy rozkład t-Studenta

Dla zmiennej losowej $\mathbf{X} = [x_1, \dots, x_n]^T$ gęstość n wymiarowego rozkładu t -Studenta o ν stopniach swobody ma postać:

$$f_{\mu, \Sigma, \nu}(\mathbf{X}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)^{n/2} \pi^{n/2} |\Sigma|^{1/2}} \left[1 + \frac{1}{\nu} (\mathbf{X} - \mathbf{m})^T \Sigma^{-1} (\mathbf{X} - \mathbf{m}) \right]^{-(\nu+n)/2} \quad (8)$$

gdzie $\mathbf{X} = [x_1, \dots, x_n]^T$ jest wektorem losowym o wektorze wartości oczekiwanych $\mathbf{m} = [m_1, \dots, m_n]^T$ i $n \times n$ wymiarowej macierzy Σ , a $\Gamma(\cdot)$ oznaczają funkcję gamma Eulera rzędu odpowiednio $(\nu+n)/2$ i $\nu/2$. Macierz kowariancji wyrażona jest jako $\nu/(\nu-2)\Sigma$ dla $\nu > 2$. Rozkład oznacza się zapisem $\mathbf{X} \sim T_\nu(\mathbf{m}, \Sigma)$.

Algorytm macierzowego rozkładu normalnego/odwrotnego rozkładu Wisharta

Macierzowy rozkład normalny/odwrotny rozkład Wisharta $(\boldsymbol{\theta}, \Sigma) \sim \text{NIW}_r(\mathbf{m}, \mathbf{C} \otimes \Sigma)$ jest wielowymiarowym czteroparametrowym ciągłym rozkładem prawdopodobieństwa, gdzie: \mathbf{m} jest macierzą położenia; \mathbf{C} jest macierzą kowariancji; Σ jest symetryczną, dodatnio określoną macierzą skali i r liczbą stopni swobody. Rozkład tworzy sprzężenie *a priori* macierzowego rozkładu normalnego o nieznannej średniej \mathbf{m} , tj. $(\boldsymbol{\theta}, \Sigma) \sim N[\mathbf{m}, \mathbf{C} \otimes \Sigma]$, z nieznaną macierzą wariancji Σ o odwrotnym rozkładzie Wisharta, tj. $\Sigma \sim \text{IW}_r(\mathbf{S})$.

W pierwszym kroku algorytmu szacuje się macierz wariancji Σ za pomocą odwrotnego rozkładu Wisharta. Zaktualizowana macierz Σ do postaci macierzy \mathbf{S} jest wejściem do drugiego etapu procedury numerycznej, która oblicza iloczyn Kroneckera $(\mathbf{S} \otimes \mathbf{C})$ [Mardia i in. 1979; Steeb 2006]. Ostatni etap polega na uruchomieniu generatora zmiennych losowych, opisanych wielowymiarowym rozkładem normalnym ze średnią w postaci macierzy \mathbf{m} zawierającą d współczynników kierunkowych dla l zmiennych prognozowanych oraz wariancją w postaci iloczynu Kroneckera. Macierz \mathbf{m} jest wyznaczana na podstawie informacji historycznej otrzymanej z eksperymentu, a macierz \mathbf{C} może występować w postaci dużego zbioru macierzy diagonalnej odzwierciedlającej niską precyzję (lub wysoką niepewność) w określaniu wartości macierzy $\boldsymbol{\theta}$.

ETAP I. Odwrotny rozkład Wisharta/Inverse Wishart distribution

Dodatnio określona macierz Σ o wymiarach $l \times l$ podlega odwrotnemu rozkładowi Wisharta wtedy i tylko wtedy, jeżeli gęstość rozkładu Σ wyrazimy wzorem:

$$p(\Sigma) \propto |\Sigma|^{-(q+r/2)} \exp[-0,5 \times \text{tr}(r \mathbf{S} \Sigma^{-1})] \quad (9)$$

gdzie $r > 0$ – znana liczba stopni swobody, \mathbf{S} – znana, dodatnio określona, macierz o wymiarach $l \times l$.

Powyższy rozkład dla macierzy Σ oznacza się w postaci:

$$\Sigma \sim \text{IW}_r(\mathbf{S}) \quad (10)$$

gdzie r oznacza liczbę stopni swobody.

ETAP II. \otimes – Iloczyn Kroneckera

Iloczynem Kroneckera macierzy \mathbf{S} o wymiarach $l \times l$ i macierzy \mathbf{C} o wymiarach $d \times d$ nazywamy macierz blokową o wymiarach $ld \times ld$ w postaci:

$$\mathbf{S} \otimes \mathbf{C} = \begin{bmatrix} \sigma_{11} \mathbf{C} & \sigma_{12} \mathbf{C} & \dots \\ \sigma_{21} \mathbf{C} & \sigma_{22} \mathbf{C} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} \sigma_{11} c_{11} & \sigma_{11} c_{12} & \dots & \sigma_{12} c_{11} & \sigma_{12} c_{12} & \dots \\ \sigma_{11} c_{21} & \sigma_{11} c_{22} & \ddots & \sigma_{12} c_{21} & \sigma_{12} c_{22} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ \sigma_{21} c_{11} & \sigma_{21} c_{12} & \dots & \sigma_{21} c_{21} & \sigma_{21} c_{22} & \dots \\ \sigma_{21} c_{21} & \sigma_{21} c_{22} & \dots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix} \quad (11)$$

Macierze \mathbf{S} i \mathbf{C} mogą być dowolnych rozmiarów. Jeżeli $\mathbf{S} = [\sigma_{ij}]$ dla $i = 1, \dots, l, j = 1, \dots, l$ oraz $\mathbf{C} = [c_{ij}]$ dla $i = 1, \dots, d, j = 1, \dots, d$, to wtedy współczynniki macierzy będącej iloczynem Kroneckera dane są wzorem:

$$(\mathbf{S} \otimes \mathbf{C})_{ij} = \sigma_{(i \operatorname{div} l) + 1, (j \operatorname{div} d) + 1} \times C_{(i \operatorname{div} d) + 1, (j \operatorname{div} d) + 1} \quad (12)$$

ETAP III. Uruchomienie generatora zmiennych losowych o wielowymiarowym rozkładzie normalnym Generator zmiennych losowych o wielowymiarowym rozkładzie normalnym uruchamiany jest ze średnią w postaci macierzy \mathbf{m} oraz wariancją w postaci iloczynu Kroneckera:

$$(\boldsymbol{\theta}|\boldsymbol{\Sigma}) \sim N(\mathbf{m}, (\mathbf{S} \otimes \mathbf{C})) \quad (13)$$

Do obliczania rozkładu brzegowego $\mathbf{Y}_t|\mathbf{D}_{t-1}$ stosowany jest wielokrotny rozkład t -Studenta [Kotz, Nadarajah 2004], natomiast do odświeżania rozkładu *a posteriori* (proces prognozowania) zastosowano macierzowy rozkład t -Studenta [Zhu i in. 2007].

Algorytm macierzowego rozkładu normalnego

Niech $\boldsymbol{\theta}$ jest uwarunkowana macierzą $\boldsymbol{\Sigma}$ i podlega macierzowemu rozkładowi normalnemu [Dawid 1981; Gupta, Nagar 1999], więc:

$$(\boldsymbol{\theta}|\boldsymbol{\Sigma}) \sim N(\mathbf{m}, \mathbf{C} \otimes \boldsymbol{\Sigma}) \quad (14)$$

Do wyznaczenia powyższego rozkładu stosowane jest złożenie dwóch funkcji – funkcji obliczającej iloczyn Kroneckera oraz generatora zmiennych losowych o wielowymiarowym rozkładzie normalnym. W pierwszym kroku przetwarzania numerycznego obliczany jest iloczyn Kroneckera ($\mathbf{C} \otimes \boldsymbol{\Sigma}$), który stanowi wejście do generatora zmiennych losowych o wielowymiarowym rozkładzie normalnym uruchamianym ze średnią w postaci macierzy \mathbf{m} oraz wariancją w postaci iloczynu Kroneckera.

Macierzowy rozkład t -Studenta – Matrix T distribution

Obliczenie rozkładu macierzowego t -Studenta polega na złożeniu dwóch funkcji. W pierwszym kroku przetwarzania numerycznego obliczany jest iloczyn Kroneckera ($\boldsymbol{\theta} \otimes \mathbf{S}$), który jest wejściem do generatora zmiennych losowych o wielowymiarowym rozkładzie t -Studenta uruchamianego ze średnią w postaci macierzy \mathbf{m} oraz wariancją w postaci iloczynu. Dodatkowo należy określić parametr w postaci liczby stopni swobody. Macierzowy rozkład t -Studenta zapisywany jest w postaci:

$$\boldsymbol{\theta} \sim T_\nu(\mathbf{m}, \boldsymbol{\theta} \otimes \mathbf{S}) \quad (15)$$

gdzie ν oznacza zdefiniowaną liczbę stopni swobody.

Dodatek D. Równanie do updatingu zmiennych stanu w filtrze – ensemble Kalman filter (EnKF)

Proces updatingu zmiennych stanu modelu przebiega następująco [Clark i in. 2008]:

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{K}[\mathbf{Z} - (\mathbf{H}\mathbf{X}^b)] \quad (16)$$

gdzie: m – liczba zmiennych stanu modelu; l – wymiar wektora obserwacji *in situ*; \mathbf{H} – $l \times m$ wymiarowa macierz operatora, który przekształca stan modelu do przestrzeni obserwacji; \mathbf{K} – $m \times l$ wymiarowa macierz zysku Kalmana; \mathbf{X}^a – $m \times n$ wymiarowa macierz uaktualnionych zmiennych stanu modelu; \mathbf{X}^b – $m \times n$ wymiarowa macierz tła zmiennych stanu modelu (zmiennych obliczonych przez model), która jest zdefiniowana jako:

$$\mathbf{X}^b = (\mathbf{X}_1^b, \dots, \mathbf{X}_n^b) \quad (17)$$

gdzie $\mathbf{X}_1^b, \dots, \mathbf{X}_n^b$ są wektorami wszystkich zmiennych stanu dla każdego n elementu wiązki przed aktualizacją, a \mathbf{Z} jest $l \times n$ wymiarową macierzą obserwacji *in situ* zdefiniowaną jako:

$$\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n) \quad (18)$$

gdzie: $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ są wektorami obserwacji *in situ* dla każdego n elementu wiązki.

Zysk Kalmana \mathbf{K} zdefiniowany jest jako:

$$\mathbf{K} = \frac{\mathbf{P}^b \mathbf{H}^T}{(\mathbf{H} \mathbf{P}^b \mathbf{H}^T) + \mathbf{R}} \quad (19)$$

gdzie \mathbf{P}^b to $m \times m$ wymiarowa macierz kowariancji błędu modelu, która jest obliczana *a priori* bezpośrednio z anomalii (odchylenia) wiązki jako:

$$\mathbf{P} = \frac{1}{n-1} [\mathbf{X}^{b'} \cdot (\mathbf{X}^{b'})^T] \quad (20)$$

gdzie $\mathbf{X}^{b'} = \mathbf{X}^b - \bar{\mathbf{X}}^b$; $\bar{\mathbf{X}}^b = \frac{1}{n} \sum_i^n \mathbf{X}_i^b$, a \mathbf{R} jest $l \times l$ wymiarową macierzą kowariancji błędu obserwacji *in situ*, która jest obliczana jako:

$$\mathbf{R} = \frac{1}{n-1} [\mathbf{Z}' \cdot (\mathbf{Z}')^T] \quad (21)$$

gdzie $\mathbf{Z}' = \mathbf{Z} - \bar{\mathbf{Z}}$; $\bar{\mathbf{Z}} = \frac{1}{n} \sum_i^n \mathbf{Z}_i$.

Macierz \mathbf{H} jest wykorzystywana do przyporządkowania stanów do stanów obserwowanych i do obliczania ekwiwalentu obserwacji na podstawie zmiennych stanu modelu. Jeżeli są to relacje liniowe, macierz \mathbf{H} obliczana jest bezpośrednio, podczas gdy w przypadku relacji nieliniowych należy użyć oddzielnego modelu. Jeżeli istnieje ścisły związek między obserwacjami a ich odpowiednikami w modelu, macierz \mathbf{H} wypełniana jest wartościami 1 dla każdego elementu macierzy, dla którego zaimplementowany jest model predykcji obserwacji, lub 0, jeżeli nie istnieje ekwiwalent obserwacji.

Dodatek E. Równanie do updatngu zmiennych stanu w filtrze - ensemble square root filter (EnSRF)

Równania wynikowe EnSRF są zapisane jako:

$$\bar{\mathbf{X}}^a = \bar{\mathbf{X}}^b + \mathbf{K}(\bar{\mathbf{Z}} - \mathbf{H}\bar{\mathbf{X}}^b) \quad (22)$$

$$\mathbf{X}^{a'} = \mathbf{X}^{b'} + \tilde{\mathbf{K}}(\mathbf{Z}' - \mathbf{H}\mathbf{X}^{b'}) \quad (23)$$

gdzie: $\tilde{\mathbf{K}}$ - $m \times l$ wymiarowa macierz zysku Kalmana użyta do aktualizacji odchylenia (anomalii) od średniej wartości wiązki zmiennych stanu; $\bar{\mathbf{X}}^a$ - $m \times 1$ wymiarowa macierz średnich wartości wiązki uaktualnionych zmiennych stanu modelu; $\bar{\mathbf{X}}^b$ - $m \times 1$ wymiarowa macierz średnich wartości wiązki tła zmiennych stanu modelu; $\mathbf{X}^{a'}$ - $m \times l$ wymiarowa macierz odchyleń macierzy tła zmiennych stanu po aktualizacji od macierzy średnich wartości wiązki uaktualnionych zmiennych stanu modelu, tj. $\mathbf{X}_i^{a'} = \mathbf{X}_i^a - \bar{\mathbf{X}}_i^a$; $\mathbf{X}^{b'}$ - $l \times n$ wymiarowa macierz odchyleń wszystkich elementów wiązki tła zmiennych stanu modelu przed aktualizacją od macierzy średnich wartości wiązki tła zmiennych stanu modelu, tj. $\mathbf{X}_i^{b'} = \mathbf{X}_i^b - \bar{\mathbf{X}}_i^b$; \mathbf{Z}' - $l \times n$ wymiarowa macierz odchyleń każdego elementu wiązki obserwacji *in situ* od średniej wartości wiązki obserwacji, tj. $\mathbf{Z}'_i = \mathbf{Z}_i - \bar{\mathbf{Z}}_i$, gdzie zmodyfikowany zysk Kalmana zdefiniowany jest jako:

$$\tilde{\mathbf{K}} = \left[1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}}} \right]^{-1} \mathbf{K} \quad (24)$$

a czynnik \mathbf{K} (równanie 24) jest skalarem o wartości z przedziału $[0,1]$. Wartość $\tilde{\mathbf{K}}$ jest używana do aktualizacji odchylenia od średniej wartości wiązki i jej wartość jest redukowana relatywnie do wartości \mathbf{K} . W metodzie EnSRF średnia wartość wiązki zmiennych stanu modelu i odchylenie od średniej wartości są aktualizowane niezależnie na podstawie równania (19) i równań (22-24) oraz dla $\mathbf{Z}' = 0$. Algorytm EnSRF wymaga obliczenia macierzy \mathbf{K} , lecz nie jest obciążony obliczeniowo bardziej od metody EnKF wymagającej procedury zaburzania obserwacji. Finalne równanie do aktualizacji zmiennych stanu modelu jest zapisane następująco:

$$\mathbf{X}^a = \bar{\mathbf{X}}^a + \mathbf{X}^{a'} \quad (25)$$

Dodatek F. Metoda wyboru zbioru zmiennych prognozujących, dla którego średnia wartość średniego błędu prognozy przyjmuje minimalną wartość

Wybór zmiennych prognozujących polega na wyszukaniu takiego zestawu zmiennych $\mathbf{X} = [x_1, x_2, \dots, x_d]$, gdzie d oznacza liczbę zmiennych prognozujących, dla którego średnia wartość średniego błędu prognozy przyjmuje minimalną wartość [Kaczmarek 1977; Sollis 2012].

Algorytm postępowania jest podobny do metody wyboru *a priori*. Podstawą porządkowania jest analiza zwykłych współczynników korelacji w pierwszym kroku, w następnych analizuje się cząstkowe współczynniki korelacji. Każdy krok kończy się wyszukaniem zmiennej prognozującej x_j , dla której wartość bezwzględna tego współczynnika jest największa, gdzie $j = 1, \dots, d$. Dla każdego uporządkowanego ciągu określana jest wartość współczynnika korelacji wielokrotnej R_0 (26) i średnia wartość średniego błędu prognozy U_0 (27). Zbiór zmiennych prognozujących, dla którego średnia wartość średniego błędu prognozy przyjmuje minimalną wartość, uważa się za optymalny.

$$R_0 = \sqrt{1 - \frac{|P|}{P_{00}}} \quad (26)$$

$$U_0 = \sigma_w \sqrt{\frac{n+d+1}{n-d-1}} \quad (27)$$

gdzie: $|P|$ – wyznacznik macierzy korelacyjnej utworzonej dla zmiennej prognozowanej; P_{00} – dopełnienie algebraiczne tej macierzy dla 0-go wiersza i 0-wej kolumny; σ – błąd standardowy zmiennej prognozowanej; $\sigma_w = \sigma \sqrt{1 - R_0^2}$ – warunkowe odchylenie standardowe zmiennej prognozowanej.