ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA VODNÍHO HOSPODÁŘSTVÍ
A ENVIRONMENTÁLNÍHO MODELOVÁNÍ

Zlepšení predikce sucha prostřednictvím asimilace dat

DIPLOMOVÁ PRÁCE

Vedoucí práce: **doc. Ing. Martin Hanel, Ph.D.**

Diplomant: **Bc. Irina Georgievová**

2020

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Irina Georgievová

Krajinné inženýrství
Environmentální modelování

Název práce

Zlepšení predikce sucha prostřednictvím asimilace dat

Název anglicky

Improving drought forecasting using data assimilation

Cíle práce

Cílem práce je implementace metodiky asimilace pozorovaných dat pro zlepšení simulace hydrologické bilance a predikce sucha v rámci systému HAMR.

Metodika

Ze systému HAMR bude vybráno několik terminálních povodí s pozorovanými daty průtoku. Budou otestovány metody pro zohlednění aktuálního vývoje odtoků/průtoků v simulaci modelu Bilan. Konkrétně budou uvažovány metody založené na ensemblovém Kalman filtru. Validace bude provedena pomocí asimilace pozorovaných dat, která jsou k dispozici v rámci systému HAMR.

Doporučený rozsah práce

40 stran

Klíčová slova

predikce sucha, asimilace dat

Doporučené zdroje informací

Asch, M., Bocquet, M., & Nodet, M. (2016). Data assimilation: methods, algorithms, and applications (Vol. 11). SIAM.

Geir Evensen: Data Assimilation: The Ensemble Kalman Filter. Springer.

Lahoz, W. A., & Schneider, P. (2014). Data assimilation: making sense of Earth Observation. Frontiers in Environmental Science, 2, 16.

Park, S. K., & Xu, L. (Eds.). (2018). Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. III). Springer International Publishing.

Rakovec, O. (2014). Improving operational flood forecasting using data assimilation. Wageningen University.

Shin, S., Kang, J. H., Chun, H. W., Lee, S., Sung, K., Cho, K., ... & Kang, J. S. (2018). Real Data Assimilation Using the Local Ensemble Transform Kalman Filter (LETKF) System for a Global Non-hydrostatic NWP model on the Cubed-sphere. Asia-Pacific Journal of Atmospheric Sciences, 54(1), 351-360.

Předběžný termín obhajoby

2019/20 LS – FŽP

Vedoucí práce

doc. Ing. Martin Hanel, Ph.D.

Garantující pracoviště

Katedra vodního hospodářství a environmentálního modelování

Elektronicky schváleno dne 25. 3. 2020

doc. Ing. Martin Hanel, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 25. 3. 2020

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 27. 03. 2020

Čestné prohlášení:

Prohlašuji, že jsem diplomovou práci na téma *Zlepšení predikce sucha prostřednictvím asimilace dat* vypracovala samostatně a citovala jsem všechny informační zdroje, které jsem v práci použila a které jsem rovněž uvedla na konci práce v seznamu použitých informačních zdrojů.

Jsem si vědoma, že na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů, především ustanovení §35 odst. 3 tohoto zákona, tj. o užití tohoto díla.

Jsem si vědoma, že odevzdáním diplomové práce souhlasím s jejím zveřejněním podle zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů, ve znění pozdějších předpisů, a to i bez ohledu na výsledek její obhajoby.

Svým podpisem rovněž prohlašuji, že elektronická verze práce

V Praze dne ....

Irina Georgievová

Poděkování:

Ráda bych poděkovala doc. Ing. Martinu Hanelovi, Ph.D. za cenné rady, věcné připomínky a vstřícnost při konzultacích a při vypracovávání diplomové práce.

Abstrakt

Metody datové asimilace vzhledem k jejich přesvědčivým výsledkům našly v posledních desetiletích široké uplatnění v hydrologii. Pojem datová asimilace lze shrnout jako metodu začlenění pozorování systému v čase do modelu za účelem získání lepších odhadů stavů systému. Tato práce se konkrétně zabývá implementací sekvenčních metod datové asimilace do celistvého hydrologického modelu BILAN, využívaného v rámci systému HAMR, který slouží pro monitorování a předpověď sucha v České republice. Tento systém byl vyvinut na základě novely vodního zákona schváleného v roce 2019, který deklaruje vznik komise operativního řízení v období nedostatku vody, obdobně jak tomu je v případě povodňových situací. Systém by tak měl sloužit jako všestranný zdroj informací pro přijetí konkrétních opatření zmíněnou komisí. Byly navrženy dva možné koncepty implementace datové asimilace do modelu: (a) kombinace predikcí více modelů metodou Kalmanova Filtru s cílem zlepšení výsledků predikce na týden dopředu, (b) zlepšení výsledků simulace modelu BILAN úpravou parametrů za pomocí metody Ansámblového Kalmanova Filtru s aktualizací po částech. Koncept (a) kombinuje predikce modelu BILAN s AR(1) procesem a byl aplikován na predikcích teploty vzduchu, srážek a potenciální evapotranspirace. Vyhodnocení proběhlo na sto útvarech povrchových vod České republiky za pomocí kritéria MASE. Navrženou metodikou se podařilo v průměru snížit chybu oproti predikcím BILANu u teplot o 10.3 %, v případě srážek o 7.15 % a pro potenciální evapotranspiraci o 9.26 %. Koncept (b) byl vyhodnocen na šesti českých povrchových útvarech metodou křížové validace, přičemž z jednoho roku pro každé povodí bylo vytvořeno 51 testovacích úseků pro následnou validaci. Při vyhodnocení tohoto konceptu došlo ke snížení RMSE simulace o 48.7 % - 88.4 % (v závislosti na zkoumaném povodí) oproti simulační metodice s neměnnými parametry. Oběma navrženým konceptům se tedy podařilo zlepšit přesnost simulací a predikcí oproti běžnému postupu v rámci systému HAMR. Jejich budoucí implementace by mohla přispět k lepšímu hospodaření a distribuci vody jak v extrémních hydrologických situacích, jakou je sucho, tak i při jeho nezbytném monitorování.

Klíčová slova: sucho, BILAN, asimilace dat, Kalmanův Filtr (KF), ansámbl, aktualizace po částech

Abstract

Data assimilation methods have been extensively used in recent decades in the field of hydrology due to their convincing results. The definition of data assimilation can be summarized as a method of incorporating observations of the system over time into the model in order to obtain improved estimates of the system states. This thesis specifically focuses on the implementation of sequential methods of data assimilation into the conceptual hydrological model BILAN, used in the system for monitoring and forecasting of drought in the Czech Republic named HAMR. Said system was developed based on amendments to the Water Act, which were approved in 2019 and entails the assembly of the operative management committee in case of water deficiency as well as floods and therefore is intended to serve as a comprehensive information source for their decision-making. Two possible concepts of data assimilation implementation were proposed: (a) combination of predictions from several models using Kalman Filter in order to improve performance of the one week ahead predictions, (b) improvement of BILAN simulation results by modifying model parameters using partitioned update scheme based on the Ensemble Kalman Filter (PU_EnKF). Concept (a) combines BILAN predictions with predictions of the AR(1) model and has been tested on air temperature, precipitation and potential evapotranspiration. Results were evaluated on one hundred catchments in the Czech Republic using MASE as a criterion. Proposed methodology succeeded in reducing the error of temperature predictions on average by 10.3%, in the case of precipitation by 7.15% and the potential evapotranspiration by 9.26% compared to BILAN predictions. Concept (b) was evaluated on six catchments by cross-validation method. For validation purposes, 51 subsequent test datasets from one year were created for each of the catchments. The evaluation of this concept showed a reduction in RMSE by 48.7% - 88.4% (depending on the catchment) compared to the simulation methodology with constant parameters. Thus, both proposed concepts managed to improve the accuracy of simulations and predictions compared to the usual procedures of the HAMR system. Therefore, their future implementation could contribute to better water management and distribution in extreme hydrological situations such as drought as well as in its essential monitoring routine.

Keywords: drought, BILAN, data assimilation (DA), Kalman Filter (KF), Partitioned Update Ensemble Kalman Filter (PU_EnKF)

Obsah

Úvod **1**

Cíle práce **2**

1 Teorie **3**
1.1 Sucho . . . 3
1.2 Simulace hydrologické bilance a predikce sucha v systému HAMR 4
1.2.1 BILAN . . . 5
1.2.2 Indexy sucha a jejich predikce . . . 8
1.3 Asimilace dat . . . 9
1.3.1 Základní pojmy a označení . . . 9
1.3.2 Kalman Filter . . . 11
1.3.3 Ensemble Kalman Filter . . . 13
1.3.4 Partitioned Update Ensemble Kalman Filter . . . 16

2 Metodika **19**
2.1 Data . . . 19
2.2 Zlepšení predikce pomocí KF . . . 20
2.2.1 Koncept . . . 20
2.2.2 Výsledky . . . 22
2.2.2.1 Ověření konceptu na jednom povodí . . . 22
2.2.2.2 Aplikace na více povrchových útvarech . . . 22
2.3 Zlepšení simulace pomocí PU_EnKF . . . 26

Diskuze **31**

Závěr **32**

Literatura **34**

Přílohy **39**
A Obrázky . . . 40
B Skripty . . . 42

Seznam obrázků

1.1 Dělení sucha a jeho dopadů 3
1.2 Grafické logo systému HAMR 4
1.3 Schéma modelu BILAN v měsíčním kroku 8
1.4 Schéma modelu BILAN v denním kroku 8
1.5 Schéma sekvenční asimilace pro KF, osa x znázorňuje čas a osa y značí hodnoty vektorů analytických a pozorovacích stavů (převzato z Asch et al. (2016)) 12
1.6 Schéma PU_EnKF (převzato z Xie – Zhang (2013)) 17

2.1 Navržená struktura vstupních dat pro předpověď modelem BILAN na týden dopředu, `X` se značí číslo zájmového týdnu 20
2.2 Predikce pomocí AR(1), BILANu a BILANu s použitím KF oproti pozorováním vybraných proměnných hydrologické bilance 23
2.3 Krabicový graf výsledků MASE kritéria na 100 povodích pro proměnné teploty (T), srážek (P) a potenciální evapotranspirace (PET) pro modely AR(1), BILAN a BILAN s aplikaci KF 24
2.4 Výsledek MASE kritéria na 100 povodích pro proměnné teploty (T), srážek (P) a potenciální evapotranspirace (PET) pro modely AR(1), BILAN a BILAN s aplikaci KF 25
2.5 Simulace řad odtoku modelem BILAN za pomocí průměrů z parametrů všech členů ansámblů pro vybraná povodí 27
2.6 Schéma křížové validace metody PU_EnKF 28
2.7 Porovnání kritéria RMSE na testovacích sadách pro vybraná povodí: simulace modelem BILAN s průměry parametrů ansámblů (`BILAN_ENS`), modelem BILAN s parametry vypočtené v rámci systému HAMR (`BILAN_HAMR`) a simulace s aplikací metody datové asimilace (`PU_EnKF`) 28
2.8 Porovnání kritéria RMSE na testovacích sadách pro vybraná povodí po zavedení korekce: simulace modelem BILAN s průměry parametrů ansámblů (`BILAN_ENS`), modelem BILAN s parametry vypočtené v rámci systému HAMR (`BILAN_HAMR`) a simulace s aplikací metody datové asimilace (`PU_EnKF`) 29
2.9 Porovnání vlivu zavedení omezení parametrů do výpočtu PU_EnKF za pomocí kritéria RMSE na testovacích sadech pro vybraná povodí 30

A.1 Průběh simulace modelu BILAN (průměr parametrů ansámblů) a průběh simulace metodou s aplikací PU_EnKF na týdenních testovacích sadech oproti průběhu měřených hodnot odtoků pro vybrané povrchové útvary ČR 40
A.2 Průběh simulace modelu BILAN (průměr parametrů ansámblů) a průběh simulace metodou s aplikací PU_EnKF zavedení korekce na týdenních testovacích sadech oproti průběhu měřených hodnot odtoků pro vybrané povrchové útvary ČR 41

Seznam tabulek

1.1 Parametry modelu BILAN 6
1.2 Vysvětlivky ke schématům modelu BILAN 8

2.1 Technická charakteristika použitých R balíčku 19
2.2 Hodnoty kritérií MAE, MSE a RMSE pro vybrané prvky hydrologické bilance povodí 'OHL_0030' 22
2.3 Procentuální úspěšnost metod podle kritéria MASE na vzorku sta povodí pro vybrané prvky hydrologické bilance 24
2.4 Souhrnné statistiky kritéria NSE, vypočítaných přes jednotlivé členy ansámblů v rámci zájmového povodí 26

Seznam algoritmů

1 Cyklus KF 12
2 Stochastický EnKF 15

Seznam skriptů

B.1 Modelování průběhu časových řad prvků hydrologické bilance modelem BILAN pro období 1981-2019 42
B.2 Tvorba predikcí modelem BILAN pro rok 2019 43
B.3 Odhad kovariančních matic pro metodu KF 45
B.4 Funkce pro datovou asimilaci metodou KF 46
B.5 Funkce pro výpočet kritéria MASE 48
B.6 Aplikace metody KF, výpočet pro sto povrchových útvarů s vyhodnocováním kritéria MASE 48
B.7 Kalibrace parametrů pro jednotlivé členy ansámblů pro každé ze zvolených povodí 50
B.8 Funkce pro datovou asimilaci metodou PU_EnKF 54
B.9 Výpočet metodou PU_EnKF, včetně křížové validace 56

Úvod

Extrémní hydrologické situace, zapřičiněné klimatickými změnami, reprezentované rozsáhlými nebo bleskovými povodněmi, případně déle trvajícími periodami sucha, jsou vzhledem k jejich vážným následkům aktuálním tématem (WMO, 2020b). Lepší pochopení průběhu sucha a metod jeho vyhodnocení je klíčovým krokem k jeho možnému zmírnění. Za tímto účelem se k popisu dynamiky environmentálních systémů používají matematické a numerické modely. Tyto modely lze použít jak pro simulaci současného chování systému, tak i pro predikci chovaní budoucího. Vzhledem k velkému množství zdrojů nejistot v tomto směru však existuje mnoho překážek. Tyto nejistoty jsou spojené se zjednodušením komplexností skutečného světa, spočívajících například v počátečních a okrajových podmínkách, struktuře modelu a jeho parametrech. Toto vede k významným rozdílům mezi předpokládanými a skutečnými stavy systému. Začlenění pozorování systému v čase do modelových rovnic může vést k získání lepších odhadů zmíněných stavů. Cílem datové asimilace je tedy použití měřených pozorování v kombinaci s modelem dynamického systému pro dosažení přesných odhadů jak současných, tak i budoucích stavů systému, současně s odhady nejistoty v těchto odhadnutých stavech (Lahoz et al., 2010).

Metody datové asimilace (DA) vzhledem k jejich přesvědčivým výsledkům našly v posledních desetiletích široké uplatnění v hydrologii (Liu et al., 2012). Především bylo vynaloženo značné úsilí pro zlepšení predikce odtoků, a to převážně metodami, kde predikce odtoku je aktualizována za pomocí chyb spočítaných nezávislým procesem či modelem (například Abebe – Price (2004) a Wu et al. (2011)). Často je odtok uvažován i jako příchozí pozorování a je využit k asimilaci stavových veličin a parametrů, jak popisuje například Clark et al. (2008) a Seo et al. (2009). Existují dále i přístupy, kde odtok vystupuje v roli diagnostické proměnné a tedy nedochází k jeho přímé aktualizaci. V rámci datové asimilace mohou být zkoumány i další prvky hydrologické bilance, a to jak pozorování in situ (Lannoy et al., 2007), tak i Dálkového průzkumu Země (Eicker et al., 2014). Mezi takové prvky hydrologické bilance patří například půdní vlhkost, sněhová pokrývka, teplota zemského povrchu a další (Lahoz et al., 2010).

Kladné výsledky metod datové asimilace jsou tak hlavní motivací pro výzkum její implementace do hydrologického modelu. Tato práce se zabývá zlepšením výsledků simulace a predikce prvků hydrologické bilance, což těsně souvisí se zdokonalením v monitorování a predikci sucha, jelikož dostupné nástroje vyhodnocení jsou založené na těchto prvcích.

Cíle práce

Cílem práce je implementovat metody datové asimilace do hydrologického modelu BILAN (viz kapitola 1.2.1), který se využívá v rámci systému pro monitorování a předpověď sucha HAMR (kapitola 1.2). Cílem je tedy vylepšení simulace prvků hydrologické bilance, což povede ke zpřesnění ve výsledcích indexů sucha, jenž se na základě těchto prvků vypočítávají. To bude mít i přímý pozitivní dopad na výsledky jejich predikcí. Datová asimilace však má uplatnění i při práci výhradně s predikcemi, a to tím, že lze použít ke kombinaci predikcí z více modelů. Z tohoto důvodu praktická část práce bude rozdělena na dva segmenty. První se bude zabývat konceptem zlepšení výsledků predikcí na týden dopředu, a to za pomoci základní metody DA - Kalman Filter (podrobně o metodě kapitola 1.3.2). Druhý segment se bude soustředit na zlepšení výsledků simulace samotného hydrologického modelu BILAN. Z důvodů komplexnosti úlohy se bude navržený koncept zakládat na pokročilejší metodě DA, založené na ansámblovém přístupů. Konkrétně se jedná o úpravu parametrů modelu za pomocí rozšíření Ensemble Kalman Filter (kapitola 1.3.3) - Partitioned Update Ensemble Kalman Filter (kapitola 1.3.4).

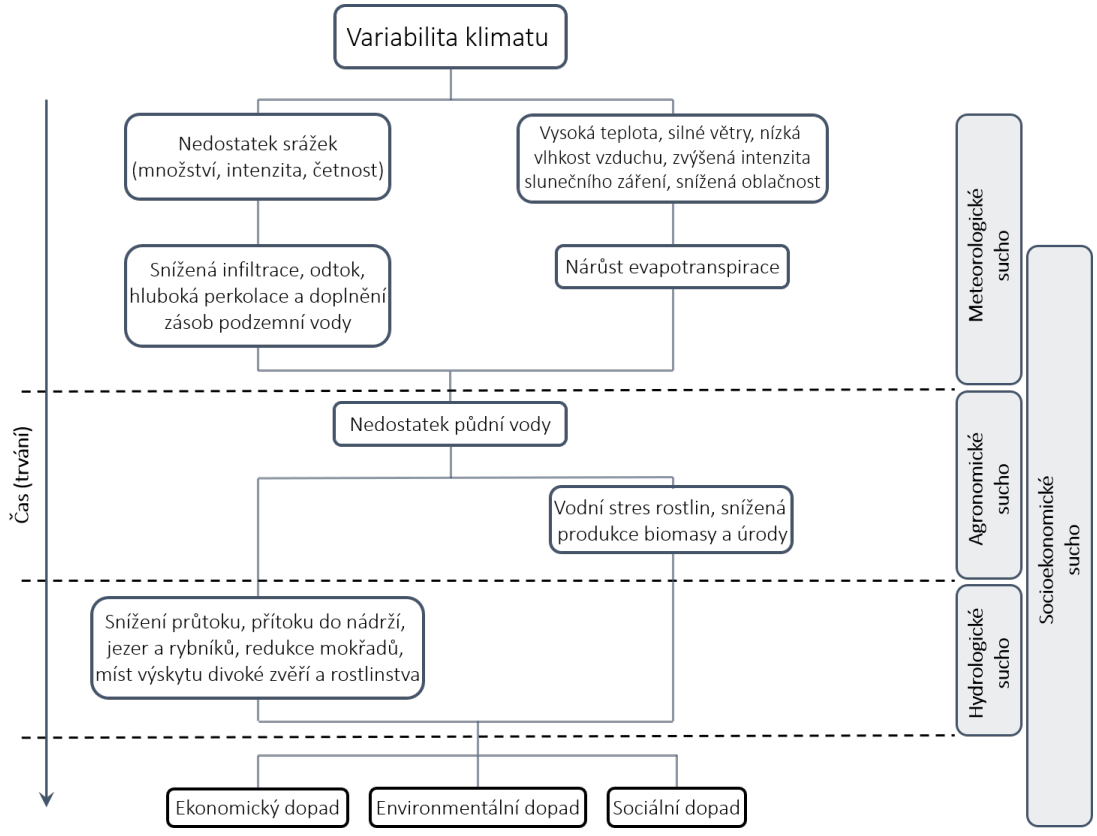
Teoretická část práce (sekce 1) se tak bude soustředit na popis používaného hydrologického modelu BILAN se zasazením do širšího rámce systému pro monitorování a předpověď sucha (kapitola 1.2). Dále bude popsána základní teorie a terminologie datové asimilace včetně vybraných metod (kapitola 1.3). Na popisu těchto metod a postupů se bude zakládat navržení konceptů implementace zmíněných metod do daného hydrologického modelu v navazující praktické části. V sekci 2 tak budou podrobně popsány postupy jednotlivých konceptů, včetně zdůvodnění a následných výsledků.

Očekávány přínos této práce tak spočívá v navržení funkčních konceptů začlenění metod datové asimilace do hydrologického modelu BILAN, čímž se položí základy pro další výzkum možné implementace do stálého provozu v rámci systému HAMR. Vzhledem k předpokládanému využití HAMRu jako zdroje informace pro komisi operativního řízení, by se dal očekávat následný dlouhodobý přínos práce ve formě zlepšení informovanosti v případě extrémních hydrologických situací, jimiž je sucho.

1. Teorie

1.1 Sucho

Sucho je často využívaným pojmem v hydrologii, klimatologii a meteorologii, avšak tento pojem nemá jednotnou definici. Obecně se tím rozumí klimaticky způsobené anomálie srážek a teplot, vedoucí k deficitu vody v atmosféře, s následnou propagací tohoto deficitu do půdy a rostlin (Iglesias et al., 2019). Sucho tak lze charakterizovat z různých pohledů, vzhledem k rozmanitosti faktorů jej způsobujících a následků, které za sebou zanechává. Komplexní popis problematiky lze nalézt například u Tallaksen – van Lanen (2004), Van Loon (2015) a Iglesias et al. (2019).



Obrázek 1.1: Dělení sucha a jeho dopadů

Sucho se propaguje do odlišných částí hydrologického cyklu, proto je důležité odlišovat sucho z tohoto hlediska a uvažovat s tím spojené kritické dopady socioekonomického a přírodního charakteru (Van Lanen et al., 2016). Nedostatek srážek a teplotní anomálie na větším území po delší dobu způsobují *meteorologické sucho*, snížení infiltrační schopnosti půdy v kombinaci se zvýšenou evaporací vede k *půdnímu suchu*, kterému se také říká *agronomické sucho*, a to pokud půda

přichází o svoji plodonosnou schopnost. Následný pokles v doplnění zásob podzemních vod a odtocích vede k *hydrologickému suchu*. Vztahy mezi různými typy sucha jsou znázorněny na schématu 1.1 (Tallaksen – van Lanen, 2004).

V publikacích Brázdil – Kirchner (2007) a Brázdil et al. (2015) byla na téma sucha provedena řada komplexních analýz a sucho se popisuje z pohledu minulosti, současnosti a budoucnosti. Podle autorů v České republice jsou sucha druhým nejvýznamnějším hydrometeorologickým extrémem po povodních a mají za následek výrazné ekonomické a sociální dopady.

1.2 Simulace hydrologické bilance a predikce sucha v systému HAMR

Jedním z vládních kroků ke zvládaní sucha bylo podle tiskové zprávy MŽP ČR (2019) schválení návrhu novely vodního zákonu (č. 254/2001 Sb.). Nová část zákona s názvem „Zvládání sucha a stavu nedostatku vody“ by měla umožnit úřadům větší regulaci odběrů vody a v případě sucha by měl vzniknout systém operativního řízení, obdobně jako tomu je v případě povodňových situací. Platné znění návrhu právního předpisu je k dispozici na Portál ODok (2019). Na základě navržených plánů pro zvládání sucha by měla v období nedostatku vody svolaná komise zvážit přijetí konkretních opatření. Za tímto účelem byl vyvinut systém pro monitorování a předpověď sucha HAMR (Hydrologie, Agronomie, Meteorologie, Retence). Na vzniku systému se podíleli VÚV TGM, v.v.i., ÚVGZ AV ČR - CzechGlobe, ČHMÚ a i KVHEM FŽP ČZU. Grafické logo systému HAMR je uvedeno na obrázku 1.2.

Obrázek 1.2: Grafické logo systému HAMR

Online systém pro zvládání sucha HAMR je založen na propojení modelů *půdního*, *hydrologického* a *vodohospodářské* bilance, které dohromady se vstupními klimatickými daty reprezentují tři typy sucha: meteorologické, zemědělské a hydrologické. Systém poskytuje informace o současném stavu vodní bilance v jednotlivých povrchových útvarech ČR a o stavu podzemních vod. Monitorují se i stavy vodních zdrojů a jejich předpokládaný vývoj. Dále se v rámci systému jednou týdně vyhodnocují indikátory sucha a zároveň se vypracovává jejich předpovídaný stav na dva týdny dopředu, viz kapitola 1.2.2.

Jednotlivé modely jsou propojené za pomoci informací z *půdního* modelu (modelu vláhové bilance půdy) SoilClim (Hlavinka et al., 2011), ze kterého byly získány parametry pro *hydrologický* model BILAN (kapitola 1.2.1), a to pro jednotlivé útvary povrchových vod České republiky. Simulace odtokových poměrů

útvaru vychází z odvozených kvantilů m-denních průtoků (Q_{30d} - Q_{364d}) a dlouhodobého průměrného ročního průtoku (Q_a) Českým hydrometeorologickým ústavem (ČHMÚ). Jelikož tyto parametry postrádají autokorelační strukturu, byly ke kalibraci modelu použity časové řady evapotranspirace a údaje o celkové půdní zásobě vody z modelu SoilClim. Jako kalibrační kritérium při výpočtu slouží Kling-Gupta Efficiency (KGE) a optimalizace probíhá za pomoci globálního optimalizačního algoritmu Shuffled Complex Defferential Evolution. Validace simulovaných průtoků probíhá na časových řadách 156 měrných stanic ČHMÚ. Systém HAMR je dále rozšířen o model *vodohospodářské* soustavy jednotlivých povodí WATERES, který z údajů o nakládání s vodami vyhodnocuje povolené a skutečné hodnoty užívání vody za den či měsíc. Více o vstupních datech do systému, jeho struktuře a použití, viz Vizina et al. (2018).

1.2.1 BILAN

BILAN je hydrologickým modelem vyvíjeným Výzkumným ústavem vodohospodářským T. G. Masaryka. Tento fyzikálně založený celistvý model neuvažuje prostorovou variabilitu parametrů a k popisu hydrologické bilance používá nízký počet parametrů. Výpočet hydrologické bilance povodí či území lze provést v denním či měsíčním časovém kroku. Výpočet však může být úpraven i na simulace v týdenním kroku. Základní bilanční vztahy jsou vyjádřeny na povrchu povodí, v zóně saturace a zóně aerace, do níž je zahrnut i vegetační kryt povodí. (van Lanen et al., 1997, VÚV TGM, v.v.i., 2015) Výchozí vstupní veličiny pro výpočet hydrologické bilance jsou denní či měsíční data:

- srážkové úhrny P [mm]
- průměrné teploty vzduchu T [°C]
- relativní vlhkost vzduchu H [%] [1]
- potenciální evapotranspirace PET [mm] [1,2]
- pozorované odtokové výšky [mm] [1,3]
- užívání vody [mm/měsíc] [1,4]

V případě nezadané vstupní řady potenciální evapotranspirace bude tato řada dopočítána. Odhad může být proveden dvěma způsoby: na základě sytostního doplňku, odvozeného pro danou teplotu a relativní vlhkost vzduchu, a pomocí vztahu odvozeného Oudin et al. (2010) ze slunečního záření a teploty vzduchu (VÚV TGM, v.v.i., 2015). První metoda využívá funkcí ve formě tabulek, které

[1] volitelně

[2] pro Oudinovu metodu stanovení PET nebo pro PET ze vstupních dat

[3] řada bude vypsána ve výstupech, lze ji použít ve vizualizacích

[4] řady odběrů z podzemních vod, neevidovaných odběrů z povrchových vod a vypouštění

Tabulka 1.1: Parametry modelu BILAN

Název	Denní typ modelu	Měsíční typ modelu
Spa	kapacita zásoby půdní vlhkosti [mm]	
Dgm	koeficient mezi teplotou a táním sněhu	
Dgw	—	koeficient pro výpočet množství kapalné vody dostupné na povrchu za zimních podmínek
Alf	parametr určující odtok ze zásoby pro přímý odtok (přímý odtok)	parametr vztahu mezi srážkou a odtokem (přímý odtok)
Soc	parametr rozdělující perkolaci na přímý odtok a na dotaci podzemní vody pro letní podmínky	parametr rozdělující perkolaci na hypodermický odtok a na dotaci podzemní vody pro letní podmínky
Mec	parametr rozdělující perkolaci na přímý odtok a na dotaci podzemní vody pro podmínky tání sněhu	parametr rozdělující perkolaci na hypodermický odtok a na dotaci podzemní vody pro tání sněhu
Wic	—	parametr rozdělující perkolaci na hypodermický odtok a na dotaci podzemní vody pro zimní podmínky
Grd	parametr určující odtok ze zásoby podzemní vody (základní odtok)	

byly odvozeny pro různé vegetační zóny v jednotlivých měsících. Druhá metoda vyžaduje uvedení zeměpisné šířky povodí a využívá tuto informaci k výpočtu hodnoty extraterestriálního slunečního záření pro každý časový krok.

Model simuluje časové řady potenciální evapotranspirace, územního výparu, zásoby vody ve sněhu, infiltrace do půdy a dotace podzemní vody z půdy do podzemní vody. Celkový odtok pro denní krok se skládá z přímého odtoku a základního odtoku. Odtok pro měsíční krok je modelován jako součet přímého, hypodermického a základního odtoku.

Jak již bylo zmíněno, základní struktura modelu je schematizována prostřednictvím tři akumulačních prostorů: povrch, půdní zóna včetně vegetačního krytu (zóna aerace) a zóna podzemní vody (zóna saturace). K výpočtu se používá obecný vztah pro vyjádření bilance akumulačního prostoru: $S_t = S_{t-1} + I_t - O_t$, kde S je zásoba vody, I je vstupní objem do akumulačního prostoru a O je objem výstupní pro daný čas t. Velikost toku mezi nádržemi je řízena algoritmy modelu spojených s šestí volnými parametry pro denní krok a osmi pro měsíční, které jsou považovány za časově invariantní (viz obrázek 1.3 a 1.4 (Máca, 2015), vysvětlivky k nim jsou uvedeny v tabulce 1.2 a seznam parametrů modelu je uveden v tabulce 1.1 (VÚV TGM, v.v.i., 2015)). BILAN rozlišuje tři teplotní režimy: zima, tání sněhu a léto.

Kalibrace parametrů modelu se provádí optimalizačním algoritmem za-

loženým na hledání nejlepší shody mezi simulovanou a pozorovanou řadou odtoku. K dispozici je lokální gradientní algoritmus (binární vyhledávání) a globální optimalizační algoritmus SCDE (Shuffled Complex Differential Evolution).

Gradientní způsob kalibrace parametrů je dvoufázový, přičemž každou fázi lze nastavit individuálně (Vizina, 2014). Tento způsob optimalizace je vysvětlen nedostatečnou shodou *MSE* (Mean Squared Error, 1.1), případně *MAE* (Mean Absolute Error, 1.2) v oblasti malých průtoků a *MAPE* (Mean Absolute Percentage Error, 1.3) v oblasti průměrných průtoků. Kombinací těchto optimalizačních kritérií se dosahuje lepších výsledků při kalibraci modelu a to jak pro nízké, tak i pro průměrné hodnoty průtoku. V první fázi se odhadnou pomocí *MSE* nebo *MAE* parametry reprezentující průměrný odtok (Dgw, Dgm, Spa, Alf) a zbylé parametry ovlivňující distribuci odtoku na složky jsou následovně kalibrovány s využitím *MAPE*. Ke kalibraci lze také použít optimalizační kritéria *NSE* (Nash-Sutcliffe Efficiency, 1.4) a *lnNSE* (Logarithmic Nash-Sutcliffe Efficiency, 1.5).

Kriteriální funkce, kde R vystupuje jako pozorovaný odtok, RM jako odtok modelovaný a n jako počet pozorovaní, lze zapsat následujícím způsobem:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(RM(i) - R(i))^2 \tag{1.1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|RM(i) - R(i)| \tag{1.2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|RM(i) - R(i)|}{R(i)} \tag{1.3}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(RM(i) - R(i))^2}{\sum_{i=1}^{n}(R(i) - \overline{R})^2}, \quad \overline{R} = \frac{1}{n}\sum_{i=1}^{n}R(i) \tag{1.4}$$

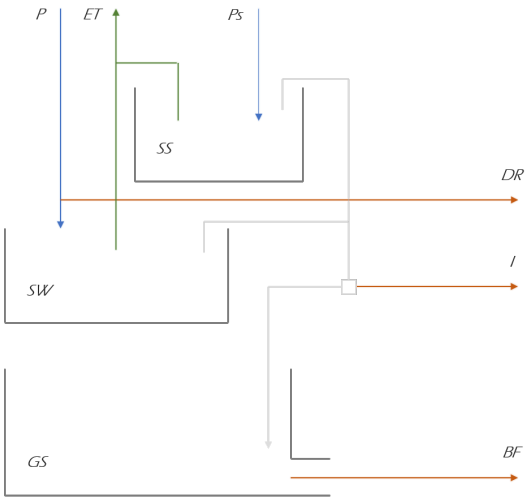
$$lnNSE = 1 - \frac{\sum_{i=1}^{n}(\ln RM(i) - \ln R(i))^2}{\sum_{i=1}^{n}(\ln R(i) - \overline{\ln R})^2}, \quad \overline{\ln R} = \frac{1}{n}\sum_{i=1}^{n}\ln R(i) \tag{1.5}$$

Optimalizační algoritmus SCE (Shuffled Complex Evolution) nebo algoritmus promíchávání popsal Duan et al. (1993) a následně rozšířil Storn – Price (1997) o metodu diferenciální evoluce (DE). Metody byly testovány v rámci dalších publikací, například Vrugt et al. (2009), Mariani et al. (2011) a Máca et al. (2013). Dle Duan et al. (1993) metoda SCE je založená na čtyřech konceptech,

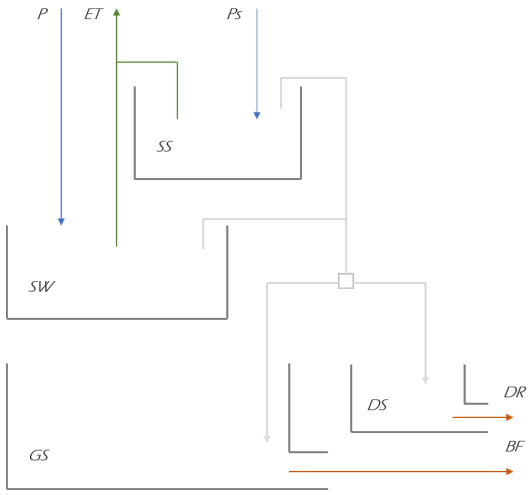
které se osvědčily pro globální optimalizaci. Konkrétně se jedná o (a) kombinaci pravděpodobnostních a deterministických přístupů, (b) shlukování, (c) systematickou evoluci sady parametrů směřujících ke globálnímu zlepšení a (d) konkurenční evoluci. Algoritmus SCDE pracuje s populací sad parametrů a hodnotami kriteriálních funkcí. Promícháváním se rozdělí populace podle hodnot kriteriálních funkcí (tak, aby hodnoty byly rovnoměrně zastoupeny) do komplexů a následně se aplikuje optimalizace pomocí diferenciální evoluce a z vyvinutých komplexů je vytvořena nová populace. Tento postup se opakuje dokud se nedosáhne maximálního počtu iterací.

Tabulka 1.2: Vysvětlivky ke schématům modelu BILAN

P	srážky	I	hypodermický odtok	SW	zásoba vody v nenasycené zóně
ET	územní výpar	BF	základní odtok	GS	zásoba podzemní vody
DR	přímý odtok	SS	zásoba vody ve sněhu	DS	zásoba pro přímy odtok



Obrázek 1.3: Schéma modelu BILAN v měsíčním kroku



Obrázek 1.4: Schéma modelu BILAN v denním kroku

1.2.2 Indexy sucha a jejich predikce

Sucha jsou zjevná pouze po dlouhém období bez deště, je těžké stanovit jejich počátek, rozsah a konec. Obecně jsou sucha interpretována a kategorizována na na *meteorologické*, *agronomické* a *hydrologické*, přičemž každá s těchto skupin se liší dle intenzity, trvání a rozsahu. Jakékoliv sucho má však značný dopad na rozličné oblasti lidské činnosti. Z tohoto důvodu bylo vynaloženo velké úsilí do vývoje technik pro analýzu a monitoring sucha a v současnosti se jako nejvhodnější a nejefektivnější nástroj prokázaly tzv. indexy sucha. Subjektivita v definici

sucha avšak značně znesnadňuje stanovení jedinečného a univerzálního indexu sucha. Pro kvantifikaci, monitoring a analýzu sucha byla proto vyvinuta celá řada indexů a stálé se pracuje na vývoji nových, případně na vylepšení již existujících indexů (Vicente-Serrano et al., 2010).

V rámci systému HAMR je jednou týdně vyhodnocena řada indexů a to jak pro pravidelný monitoring, tak i pro predikci sucha na dva týdny dopředu. Pro hodnocení sucha se používají například standardizované indexy (tzv. indexy typu S) charakterizované typickými veličinami pro zvolenou kategorii sucha. Pro meteorologické sucho se používá *Standardized Precipitation Evapotranspiration Index* (SPEI), při výpočtu kterého se vychází z hodnot srážkových úhrnů a potenciální evapotranspirace za posledních šest týdnů. Hydrologické sucho povrchové se v rámci systému hodnotí dle aktuální hodnoty *Standardized Runoff Index* (SRI), založené na průtocích, zatímco podzemní hydrologické sucho se hodnotí na základě modelovaných hodnot základního odtoku a to pomocí *Standardized Groundwater Index* (SGI) v daném týdnu. Všechny tyto indexy standardizovaným způsobem reprezentují anomálie od normálu. Jedná se o transformaci časových řad na normální rozdělení, více o výpočtu například Vlnas et al. (2014) . Výhodou je jejich relativně snadný výpočet na základě několika proměnných a právě jejich standardizovaný charakter. Nevýhodou je to, že závažnost sucha je vyjádřena pouze relativně, proto se běžně používají spíše jako doplňující informace k dalším ukazatelům (Iglesias et al., 2019). Sucho v rámci systému HAMR je dále hodnoceno na základě například nedostatkových objemů, srážkového deficitu, deficitu vody v půdě a retenční kapacity.

Předpověď indikátorů na dva týdny dopředu je založená na předpovědích pěti modelů, používaných meteorologickými službami řady států, mezi nimiž je i Česká republika. Jedná se o IFS (*Integrated Forecast System*, vyvíjený a spravovaný European Centre for Medium-Range Weather Forecasts), GFS (*Global Forecasting System*, model meteorologické služby USA National Office for Ocean and Atmosphere neboli NOAA), GEM (*Global Earth Model*, Canadian Meteorological Centre neboli CMC), UM (*Unified Model*, z United Kingdom Meteorological Office či UKMO) a ARPEGE (*Action de Recherche Petite Echelle Grande Echelle*, model francouzské meteorologické služby Centre National de Recherches Météorologiques/Meteo France) (Trnka et al., 2017).

1.3 Asimilace dat

1.3.1 Základní pojmy a označení

Datovou asimilaci se označuje přístup či metoda pro kombinování pozorování s výstupem modelu za účelem zlepšení zmíněného modelu (Asch et al., 2016). Datová asimilace mimo jiné umožňuje optimalizaci počátečních podmínek, aktualizaci parametrů a dokonce i vylepšení struktury dynamického modelu. DA, dříve používaná především ve vědách atmosferických a oceánografických, je stále více používána hydrology vzhledem k přesvědčivým výsledkům při korekci hydrologických modelů prováděných v reálném čase (Sun et al., 2016).

Existují dva klasické přístupy datové asimilace: variační a sekvenční (nebo statistický). V obou případech se jedná o hledání optimálního řešení, ale zatímco u statistického přístupu se jedná o nalezení nejnižšího rozptylu, přístup variační minimalizuje vhodnou kriteriální (účelovou) funkci. Vývoj nových schémat a metod je aktuálním tématem napříč matematickými disciplínami. Neustále tedy přibývají novější metody a rozšíření. Mezi ně patří kombinace dvou základních přístupů, které se označují za metody smíšené. Přehledně rozdělení metod DA popisuje například Asch et al. (2016), stručně by se dalo shrnout takto:

Sekvenční:	Variační:	Smíšené:
◦ Kalman Filter (KF)	◦ 3D-Var	◦ 4DEnVar
◦ Ensemble KF (EnKF)	◦ 4D-Var	◦ IEnKS
◦ Extended KF (EKF)	◦ Nudging	◦ EDA

Tato práce se bude zabývat pouze datovou asimilaci sekvenčního typu, o aktuálním vývoje těchto metod v hydrologii píše například Sun et al. (2016).

Standardní zápis problému datové asimilace formuloval Lorenc – Ghil (1997) a tento zápis je často využíván v literatuře o datové asimilaci, například Asch et al. (2016), Lahoz et al. (2010). Pokud uvažujeme diskrétní model pro vývoj fyzického (atmosférického, oceánského, mechanického atd.) systému definovaného z času t_k do času t_{k+1}, pak ho lze popsat dynamickou stavovou rovnicí

$$\boldsymbol{x}^f(t_{k+1}) = \boldsymbol{M}_{k+1}[\boldsymbol{x}^f(t_k)], \tag{1.6}$$

kde $\boldsymbol{x}$ je vektor *skutečných* stavů a $\boldsymbol{M}$ je odpovídajícím *dynamickým operátorem*, který obecně může být časově závislý. Se stavem $\boldsymbol{x}$ souvisí chybová kovariační matice $\boldsymbol{P}$, neboť skutečný stav se bude od simulovaného lišit jak náhodnou tak i systematickou chybou. Pozorování neboli měření v čase t_k jsou definovány

pomocí vztahu

$$\boldsymbol{y}^o = \boldsymbol{H}_k[\boldsymbol{x}^t(t_k)] + \boldsymbol{\varepsilon}_k^o, \tag{1.7}$$

kde $\boldsymbol{H}$ je časově závislý *operátor pozorování* a $\boldsymbol{\varepsilon}_k^o$ je bílý šum s nulovou střední hodnotou a s kovarianční maticí $\boldsymbol{R}$, reprezentující chyby měření. Horní index odpovídá původu vektorů, případně matic v procesu datové asimilace:

- **a** - *analysis* neboli výsledek aproximace skutečného stavu pomocí odhadu a měření (*a posteriori*),
- **b** - *background* neboli počáteční (*a priori*) odhad,
- **f** - *forecast* neboli předpověď,
- **o** - *observation* neboli pozorování, případně měření,
- **t** - *true* neboli skutečný, většinou neznámý stav.

1.3.2 Kalman Filter

Jednou z klasických sekvenčních metod, která je již více jak dvě desetiletí využívána v hydrologické datové asimilaci (například Evensen (1994) nebo McLaughlin (1995)) je metoda Kalmanova Filtru (KF). Metoda je pojmenována dle Rudolpha E. Kalmana, který publikoval v roce 1960 článek „A New Approach to Linear Filtering and Prediction Problems“, popisující rekurzivní řešení časově závislého problému lineárního filtrování založeného na diskrétních datech (Kalman, 1960). Předpokládá se dynamický stochastický systém, který se vyvíjí v diskrétním čase k. Tento systém lze zapsat následovně:

$$\boldsymbol{x}_{k+1} = \boldsymbol{M}_{k+1}[\boldsymbol{x}_k] + \boldsymbol{w}_k, \tag{1.8}$$

kde $\boldsymbol{w}_k$ přestavuje chybu z nezávislého normálního rozdělení s nulovou střední hodnotou $\boldsymbol{w}_k \sim N(0, \boldsymbol{Q}_k)$. Pro měření/pozorování platí

$$\boldsymbol{y}_k = \boldsymbol{H}_k[\boldsymbol{x}_k] + \boldsymbol{v}_k, \tag{1.9}$$

kde $\boldsymbol{v}_k$ je opět chyba z nezávislého normálního rozdělení s nulovou střední hodnotou $\boldsymbol{v}_k \sim N(0, \boldsymbol{R}_k)$. $\boldsymbol{Q}$ a $\boldsymbol{R}$ jsou kovarianční matice u kterých se předpokládá, že jsou známé a které můžou být časově závislé. Sekvenční schéma datové asimilace se skládá ze dvou hlavních kroků: tvorba **predikce** a následná **analýza**. V čase t_k je k dispozici výsledek předchozí predikce $\boldsymbol{x}_k^f$ a vektor dostupných pozorování $\boldsymbol{y}_k$. Na základě těchto dvou vektorů se provedením analýzy získává a posteriori

stav $\boldsymbol{x}_k^a$, který je lineární kombinací a priori stavu ($\boldsymbol{x}_k^f$) a váženým rozdílem mezi skutečným pozorováním ($\boldsymbol{y}_k$) a pozorováním predikčním $\boldsymbol{H}_k[\boldsymbol{x}_k^f]$ Změna je tedy založena na rozdílu mezi měřením a odhadovanou hodnotou v čase k. KF pak má tvar:

$$\boldsymbol{x}_k^a = \boldsymbol{x}_k^f + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k\boldsymbol{x}_k^f), \tag{1.10}$$

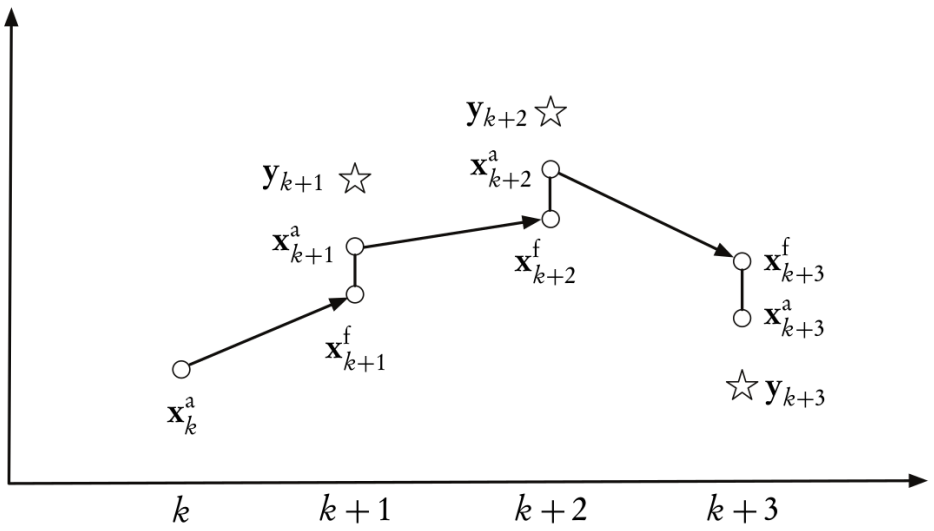
kde $\boldsymbol{K}$ je takzvaná Kalmanova váha, zvolená tak, aby minimalizovala kovarianční chybu analytického stavu. Kalmanovu váhu minimalizující tuto chybu lze získat následovně

$$\boldsymbol{K}_k = \boldsymbol{P}_k^f\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}_k^f\boldsymbol{H}^T + \boldsymbol{R})^{-1}, \tag{1.11}$$

kde $\boldsymbol{P}_k^f$ je kovarianční matice predikčních chyb. Takto získaný odhad je BLUE neboli *best linear unbiased estimator*. Podrobný popis s důkazy uvádí například Asch et al. (2016). Postup KF lze sepsat do podoby algoritmu (viz 1) nebo shrnout do schématu (obrázek 1.5).

Algoritmus 1: Cyklus KF

1 Inicializace odhadů $\boldsymbol{x}_k^a$ a $\boldsymbol{P}_k^a$
2 **for** $k = 0, \ldots, K$ **do**
3 Predikční fáze analýzy:
predikce stavu $\boldsymbol{x}_{k+1}^f = \boldsymbol{M}\boldsymbol{x}_k^a$
predikce kovarianční chyby $\boldsymbol{P}_{k+1}^f = \boldsymbol{M}\boldsymbol{P}_k^a\boldsymbol{M}^T + \boldsymbol{Q}$
4 Aktualizace pozorování, „korekční“ fáze:
výpočet Kalmanovy váhy
$\boldsymbol{K}_{k+1} = \boldsymbol{P}_{k+1}^f\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}_{k+1}^f\boldsymbol{H}^T + \boldsymbol{R})^{-1}$
aktualizace odhadu pozorováním
$\boldsymbol{x}_{k+1}^a = \boldsymbol{x}_{k+1}^f + \boldsymbol{K}_{k+1}(\boldsymbol{y}_{k+1} - \boldsymbol{H}\boldsymbol{x}_{k+1}^f)$
aktualizace kovarianční chyby $\boldsymbol{P}_{k+1}^a = (\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{H})\boldsymbol{P}_{k+1}^f$
5 **end**



Obrázek 1.5: Schéma sekvenční asimilace pro KF, osa x znázorňuje čas a osa y značí hodnoty vektorů analytických a pozorovacích stavů (převzato z Asch et al. (2016))

1.3.3 Ensemble Kalman Filter

Ensemble Kalman Filter (EnKF) je další sekvenční metodou datové asimilace. Tato metoda byla prvně navržena Evensen (1994) jako stochastická nebo Monte Carlo alternativa Kalmanova Filtru. EnKF je založen na konceptu dílčích součástí, souboru stavových vektorů, které jsou nazývány členy ansámblu. Ve snaze se vyhnout propagaci velkých kovariančních matic, je nejistota systému reprezentována variabilitou těchto členů. Existují dva přístupy EnKF: deterministický a stochastický. V rámci analýzy ve stochastickém přístupu jsou generovány náhodné perturbace, zatímco v deterministickém přístupu nikoliv. Stochastický EnKF se tak snaží napodobit postup analýzy KF z kapitoly 1.3.2, kovarianční matice jsou však nahrazeny již zmíněnými ansámbly. Cílem je provést u každého členu ansámblu analýzu následujícím způsobem (Asch et al., 2016):

$$\boldsymbol{x}_i^a = \boldsymbol{x}_i^f + \boldsymbol{K}\left[\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}_i^f)\right], \tag{1.12}$$

kde $i = 1, \ldots, m$ je index členu ansámblu a $\boldsymbol{x_i^f}$ je vektorem predikcí pro člen ansámblu i, který reprezentuje počáteční odhad nebo stav před analýzou. Pro napodobení KF, K je označována Kalmanovou váhou, kterou je potřeba odhadnout ze statistik ansámblu:

$$\boldsymbol{K} = \boldsymbol{P}^f \boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{P}^f \boldsymbol{H}^T + \boldsymbol{R}\right)^{-1} \tag{1.13}$$

Lze stanovit kovarianční matici chyby jako

$$\boldsymbol{P}^f = \frac{1}{m-1}\sum_{i=1}^{m}\left(\boldsymbol{x}_i^f - \bar{\boldsymbol{x}}^f\right)\left(\boldsymbol{x}_i^f - \bar{\boldsymbol{x}}^f\right)^T, \text{ kde } \bar{\boldsymbol{x}}^f = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^f.$$

Tento odhad chybové kovarianční matice lze pak zapsat

$$\boldsymbol{P}^f = \boldsymbol{X}_f\boldsymbol{X}_f^T,$$

kde $\boldsymbol{X}_f$ je matice o velikosti $n \times m$, sloupce, které jsou představeny normalizovanými perturbacemi pro $i = 1, \ldots, m$,

$$[\boldsymbol{X}_f]_i = \frac{\boldsymbol{x}_i^f - \bar{\boldsymbol{x}}^{\boldsymbol{f}}}{\sqrt{m-1}}.$$

Ze vztahu 1.12 lze stanovit „analytické“ stavy ansámblu $\{\boldsymbol{x}_i^a\}_{i=1,\ldots,m}$, ze kterých lze následně odhadnout statistiky odhadů. Lze tedy spočítat jak odhady, tak i ansámbl perturbací odhadů pomocí vztahů:

$$\bar{\boldsymbol{x}}^a = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^a, \quad [\boldsymbol{X}_a]_i = \frac{\boldsymbol{x}_i^a - \bar{\boldsymbol{x}}^a}{\sqrt{m-1}}.$$

Normalizované *anomálie*, $\boldsymbol{X}_i^a \equiv [\boldsymbol{X}_a]_i$, tj. normalizované odchylky členů ansámblu od průměru, jsou získány z rovnice 1.12 minus aktualizovaná střední hodnota,

$$\boldsymbol{X}_i^a = \boldsymbol{X}_i^f + \boldsymbol{K}\left(\boldsymbol{0} - \boldsymbol{H}\boldsymbol{X}_i^f\right) = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H})\boldsymbol{X}_i^f, \tag{1.14}$$

kde $\boldsymbol{X}_i^f \equiv [\boldsymbol{X}_f]_i$, což vede k chybové kovarianční matici analýzy

$$\boldsymbol{P}^a = \boldsymbol{X}_a\boldsymbol{X}_a^T = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H})\boldsymbol{X}_f\boldsymbol{X}_f^T(\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H})^T = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H})\boldsymbol{P}^f(\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H})^T. \tag{1.15}$$

Toto avšak nesplňuje BLUE charakteristiku odhadu, jak tomu bylo u KF, dochází k podhodnocení odhadu chybových kovariancí, což vede k divergence řešení, když je schéma cyklováno. Toto se řeší perturbací vektoru pozorovaných hodnot každého členu ansámblu: $\boldsymbol{y} \to \boldsymbol{y}_i \equiv \boldsymbol{y} + \boldsymbol{u}_i$, kde $\boldsymbol{u}_i$ pochází z Gaussova rozdělení a tedy $\boldsymbol{u}_i \sim N(0, \boldsymbol{R})$. Střední hodnota vzorku $\boldsymbol{u}_i$ je pak definovaná jako $\bar{\boldsymbol{u}}$ a změny v perturbacích

$$[\boldsymbol{Y}_f]_i = \frac{\boldsymbol{H}\boldsymbol{x}_i^f - \boldsymbol{u}_i - \boldsymbol{H}\bar{\boldsymbol{x}}^f + \bar{\boldsymbol{u}}}{\sqrt{m-1}}, \tag{1.16}$$

z čehož je patrné, že předpovězené pozorování $\boldsymbol{H}\boldsymbol{x}_i^f$ jsou vskutku perturbovány. Vektor anomálií z analýzy je modifikován odpovídajícím způsobem:

$$\boldsymbol{X}_i^a = \boldsymbol{X}_i^f - \boldsymbol{K}\boldsymbol{Y}_i^f = (\boldsymbol{I}_n - \boldsymbol{K}\boldsymbol{H})\boldsymbol{X}_i^f + \frac{\boldsymbol{K}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})}{\sqrt{m-1}}. \tag{1.17}$$

Kalmanova váha pak může být definována pouze z hlediska matic anomálií,

$$\boldsymbol{K} = \boldsymbol{X}_f \boldsymbol{Y}_f^T \left(\boldsymbol{X}_f \boldsymbol{Y}_f^T\right)^{-1}. \tag{1.18}$$

Z tohoto zápisu je patrné, že aktualizované perturbace jsou lineární kombinací perturbací předpovězených. Postup EnKF lze tedy sepsat do několika kroků (viz algoritmus 2).

Algoritmus 2: Stochastický EnKF

Vstup: Pro $k = 0, \ldots, K$: chybová kovarianční matice pozorování $\boldsymbol{R}_k$, operátor pozorování $\boldsymbol{H}_k$, dynamický operátor $\boldsymbol{M}_k$

1 Inicializace ansámblu $\{\boldsymbol{x}_{i,0}^f\}_{i=0,\ldots,m}$

2 **for** $k = 0, \ldots, K$ **do**

3 Generace statisticky konzistentní sady pozorování:
pro $i = 1, \ldots, m$: $\boldsymbol{y}_{i,k} = \boldsymbol{y}_k + \boldsymbol{u}_i, \qquad \boldsymbol{u}_i \sim N(0, \boldsymbol{R}_k)$

4 Výpočet středních hodnot ansámblu:
$\bar{\boldsymbol{x}}_k^f = \frac{1}{m}\sum_{i=1}^m \boldsymbol{x}_{i,k}^f, \quad \bar{\boldsymbol{u}} = \frac{1}{m}\sum_{i=1}^m \boldsymbol{u}_i, \quad \bar{\boldsymbol{y}}_k^f = \frac{1}{m}\sum_{i=1}^m \boldsymbol{H}_k(\boldsymbol{x}_{i,k}^f)$
normalizace anomálií
$[\boldsymbol{X}_f]_{i,k} = \dfrac{\boldsymbol{x}_{i,k}^f - \bar{\boldsymbol{x}}_k^f}{\sqrt{m-1}}, \qquad [\boldsymbol{Y}_f]_{i,k} = \dfrac{\boldsymbol{H}_k(\boldsymbol{x}_{i,k}^f) - \boldsymbol{u}_i - \bar{\boldsymbol{y}}_k^f + \bar{\boldsymbol{u}}}{\sqrt{m-1}}.$

5 Výpočet Kalmanovy váhy: $\boldsymbol{K} = \boldsymbol{X}_k^f \left(\boldsymbol{Y}_k^f\right)^T \left\{\boldsymbol{Y}_k^f \left(\boldsymbol{Y}_k^f\right)^T\right\}^{-1}$

6 Aktualizace ansámblu:
pro $i = 1, \ldots, m$: $\boldsymbol{x}_{i,k}^a = \boldsymbol{x}_{i,k}^f + \boldsymbol{K}_k\left(\boldsymbol{y}_{i,k} - \boldsymbol{H}_k(\boldsymbol{x}_{i,k}^f)\right)$

7 Výpočet předpovědí pro ansámbl:
pro $i = 1, \ldots, m$: $\boldsymbol{x}_{i,k+1}^f = \boldsymbol{M}_{k+1}(\boldsymbol{x}_{i,k}^a).$

8 **end**

1.3.4 Partitioned Update Ensemble Kalman Filter

Partitioned update schéma, neboli aktualizace po částech, je rozšířením Ensemble Kalman Filter (zkráceně PU_EnKF), které bylo představené Xie – Zhang (2013) a používá se k aktualizaci jak hydrologických proměnných, tak i parametrů modelu. Během asimilace nových pozorování je tak vektor parametrů aktualizován najednou se stavovými proměnnými a to jakmile je nové pozorování k dispozici. Při takovéto aktualizaci ansámblu parametrů je soubor náchylný ke tzv „shrinkage“ (tj zmenšení rozptylu parametrů). PU_EnKF proto uvádí řešení tohoto problému pomocí kernel vyhlazení vzorků parametrů, metodologii představené West (1993) a později rozšířené Liu (2000). Xie et al. shrnují řešení následujícím způsobem:

$$\theta_t^{i-} = \alpha\theta_{t-1}^{i+} + (1-\alpha)\overline{\theta}_{t-1}^{+} + \tau_t^i, \qquad \tau_t^i \sim N(0, T_t) \tag{1.19}$$

$$\overline{\theta}_{t-1}^{+} = \frac{1}{n}\sum_{i=1}^{n}\theta_{t-1}^{i+}$$
$$T_t = h^2 \mathrm{var}(\theta_{t-1}^{+})$$

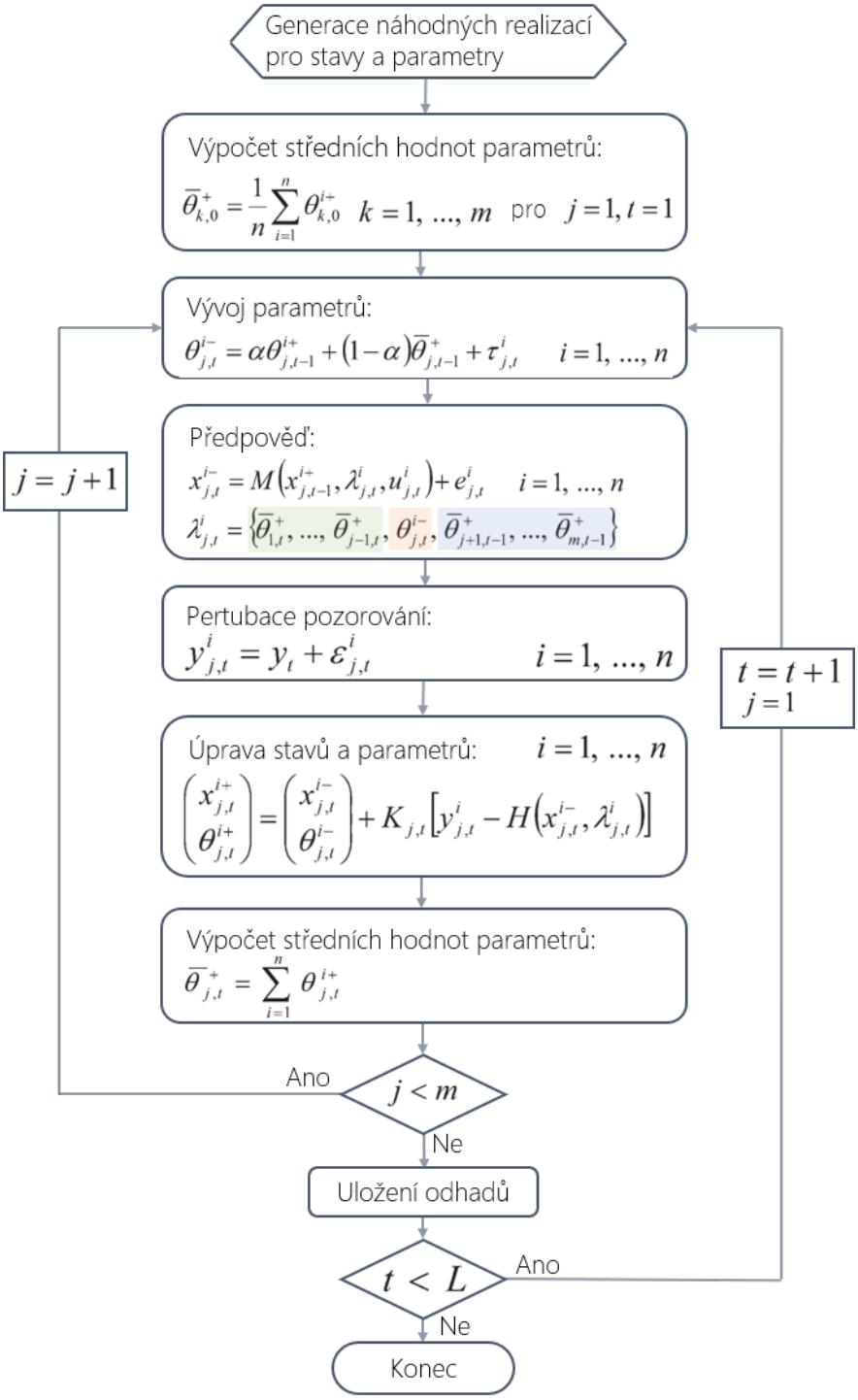
kde α reprezentuje „shrinkage factor“ rozsahu $(0, 1]$ (typicky nabývá hodnot v rozmezí $[0.95, 0.99]$), h je vyhlazovací faktor, který se většinou odvodí ze vztahu $\alpha^2 + h^2 = 1$ a T_t je kovariance odvozená z v vyhlazovacího faktoru a variance ansámblu parametrů $\mathrm{var}(\theta_t^+)$. Proměnná h se stanovuje dle autorů metody experimentálně a poněkud subjektivně, je závislá na velikosti rozptylu ansámblu parametrů $\mathrm{var}(\theta_{t-1}^+)$ a musí být zvolená tak, aby zamezovala či naopak přispívala k rozšíření ansámblu.

Metoda aktualizace po částech rozděluje parametry na několik typů podle jejich citlivosti a každý z těchto typu je aktualizován samostatně v rámci m cyklu (kde m se značí počet parametrů). Shrnutí postupu metody je zobrazeno na schématu 1.6, v němž n je počet realizací ansámblu, parametry jsou značeny θ, L je délka časového úseku, x se značí vektor předpovězených stavů a y se značí vektor pozorovaných stavů. Vektor parametrů λ se skládá ze tří částí:

(1) průměry nově aktualizovaných parametrů z předchozích cyklů v současném časovém kroku t: $(\overline{\theta}_{1,t}^{+}, \ldots, \overline{\theta}_{j-1,t}^{+})$

(2) cílový typ parametru, který má být aktualizován v současném cyklu: $(\theta_{j,t}^{i-})$

(3) průměr parametrů odhadnutých v předchozím časovém kroku $t-1$: $(\overline{\theta}_{j+1,t-1}^{+}, \ldots, \overline{\theta}_{m,t-1}^{+})$.



Obrázek 1.6: Schéma PU_EnKF (převzato z Xie – Zhang (2013))

Perturbace pozorovaní y pak probíhá obdobně jak tomu bylo v kapitole 1.3.3, tedy $y^i_{j,t} = y_t + \varepsilon^i_{j,t}$, kde $\varepsilon^i_{j,t} \sim N(0, \boldsymbol{R}_t)$, kde $\boldsymbol{R}_t$ je chybová kovariančí matice pozorování. Operátor pozorování $\boldsymbol{H}$ je zjednodušen a odpovídá $[1, 0]$.

Kalmanova váha $\boldsymbol{K}$ pak může být vyjádřena následujícím způsobem:

$$K_{j,t} = \begin{bmatrix} \mathrm{cov}\big(X^-_{j,t}, X^-_{j,t}\big) \\ \\ \mathrm{cov}\big(X^-_{j,t}, \theta^-_{j,t}\big) \end{bmatrix} \frac{1}{\mathrm{cov}\big(X^-_{j,t}, X^-_{j,t}\big) + R_t} \tag{1.20}$$

Aktualizaci stavu a parametru lze dále zapsat:

$$\begin{pmatrix} x^{i+}_{j,t} \\ \\ \theta^{i+}_{j,t} \end{pmatrix} = \begin{pmatrix} x^{i-}_{j,t} \\ \\ \theta^{i-}_{j,t} \end{pmatrix} + K_{j,t}\Big[y^i_{j,t} - H\big(x^{i-}_{j,t}, \lambda^i_{j,t}\big)\Big], \qquad \text{pro } i = 1, \ldots, n \tag{1.21}$$

Předpověď stavu je tak podmíněná cílovým parametrem, jehož odhad se vzhledem k iterativnímu charakteru aktualizace posouvá k optimální hodnotě. Autoři metody resumují, že metoda umožňuje redukci počtu stupňů volnosti a snížení rušení různými typy parametrů a tímto se očekává dosažení přesnější reprezentace kovariancí mezi parametry a dynamickými stavy. Optimální odhad každého stavu pak lze aproximovat průměrem všech m členů ansámblu.

2. Metodika

2.1 Data

Data vstupující do výpočtu hydrologické bilance modelem BILAN byla získána ze zdrojového FTP serveru, spravovaného CzechGlobe – Ústavem výzkumu globální změny AV ČR, v. v. i., kde se nachází databáze výstupů z fyzikálně založeného matematického modelu SoilClim. Jedná se o výstupy úhrnů srážek, teplot vzduchu a evapotranspirace pro jednotlivé útvary povrchových vod České republiky. Obdržené časové řady jsou v časovém rozmezí od roku 1961 až do konce roku 2019.

Dále byly od Českého hydrometeorologického ústavu obdrženy záznamy o měřeném odtoku pro 74 vodoměrných stanic ČR a to pro období od 01-01-1981 až 31-10-2010. Data jsou následně přiřazená k 70 povrchovým útvarům. Často se ovšem rozvodnice povodí nekryje s rozvodnicí ke stanici. Nejedná se však o velké rozdíly a tedy odtok na plochu povodí se nemění nijak významně.

Veškeré skripty použité pro výpočet jsou napsány v programovacím jazyce R, konkretně verze 3.6.1 (2019-07-05) a to převážně za použití balíčků uvedených v tabulce 2.1. Veškeré skripty použité pro výpočty jsou uvedeny v příloze B.

Tabulka 2.1: Technická charakteristika použitých R balíčku

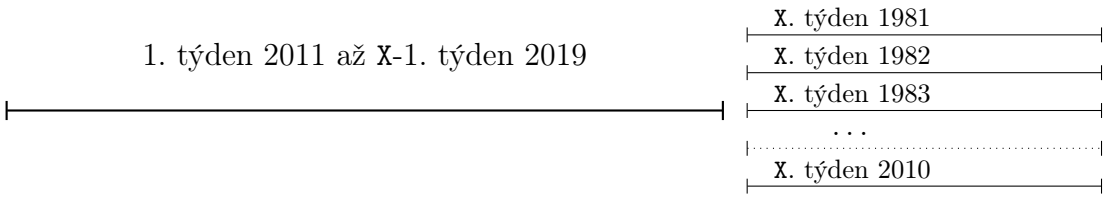
Název	Verze	Popis
`bilan`	2018-11-29	Bilan water balance model
`hydroGOF`	0.3-10	Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series
`data.table`	1.12.2	Extension of `data.frame`
`dplyr`	0.8.3	A Grammar of Data Manipulation
`lubridate`	1.7.4	Make Dealing with Dates a Little Easier
`reshape2`	1.4.4	Flexibly Reshape Data: A Reboot of the Reshape Package

2.2 Zlepšení predikce pomocí KF

2.2.1 Koncept

Pro tvorbu jak simulací tak i předpovědí byla použita metodika vyvinutá hydrologickým oddělením Výzkumného Ústavu Vodohospodářského T. G. Masaryka a KVHEM ČZU. Použité parametry modelu v této části práce byly naučeny jako součást projektu HAMR (viz kapitola 1.2) na základě informací získaných z modelu SoilClim a to pro každý z 1121 útvarů povrchových vod (povodí) České republiky. Tyto parametry byly následnými vstupy do hydrologického modelu BILAN v denním kroku a byly tedy použity k modelování průběhu prvků hydrologické bilance od roku 1981 do současnosti včetně předpovědního období (rok 2019). Konkrétní postup simulace za použití předpočítaných parametrů je uveden ve skriptu B.1.

Postup pro předpověď na týden dopředu je založen na simulaci dat, složených z (a) dat od roku 2011 až k týdnu, předcházejícímu týden předpovídaný a (b) týdnů odpovídajících předpovídanému z let 1981 až 2010 (viz schéma 2.1). Výsledných třicet simulací se omezí na zájmový týden a následně se průměrují. Postup byl dále opakován pro všechny týdny v roce.



Obrázek 2.1: Navržená struktura vstupních dat pro předpověď modelem BILAN na týden dopředu, X se značí číslo zájmového týdnu

Koncept navržené metody se zabývá kombinací úprav predikčního modelu na základě výsledků alternativního modelu. Jako alternativní model byl zvolen jednoduchý autoregresní proces prvního stupně $AR(1)$, který lze zapsat:

$$x_k = \alpha x_{k-1} + w_k, \tag{2.1}$$

kde α je autoregresním koeficientem a w_k je bílý šum a tedy $w_k \sim N(0, \sigma^2)$. Příklad obdobné úlohy popisuje například Wu et al. (2017). Zmíněný proces byl následně použit k tvorbě jednotýdenní předpovědi a tato předpověď je následně asimilována s předpovědí získanou z modelu BILAN pomocí Kalmanova filtru. Tento postup je aplikován pro vybrané proměnné hydrologické bilance, tedy atmosferické srážky (P), teploty vzduchu (T) a potenciální evapotranspiraci (PET). Lepší předpověď zmíněných proměnných by vedla k následnému zlepšení předpo-

vědí modelovaného odtoku, jelikož tyto proměnné působí jako základní vstupy do modelu BILAN. Proměnné P a PET jsou zároveň klíčovými při výpočtu indexu sucha SPEI, zatímco k výpočtu SRI je potřeba již zmíněný modelovaný odtok.

Pro aplikaci metody KF (viz kapitola 1.3.2) je potřeba znalost hodnot kovariančních matic $\boldsymbol{R}_k$ a $\boldsymbol{Q}_k$ a znalost hodnoty dynamického operátoru $\boldsymbol{M}_k$. Tyto veličiny byly stanoveny odhadem z předchozího období, tedy na roce 2018. Za tímto účelem byly odhadnuty týdenní predikce pro rok 2018 a to již výše popsaným způsobem (uvedeným ve skriptu B.2), s příslušnou úpravou časového horizontu. Na těchto datech se stanoví autoregresní koeficienty, které použijeme jako dynamický operátor $\boldsymbol{M}_k$. Dále $\boldsymbol{R}_k$ neboli chybová kovarianční matice je představena rozptylem chyby predikce modelu BILAN a kovarianční matice pozorování $\boldsymbol{Q}_k$ je představena rozptylem chyby AR modelu (viz skript B.3).

Na základě popsané teorii a algoritmu z kapitoly 1.3.2 byla napsána funkce postupu základního KF, která je následně využitá ve funkci při tvorbě predikčního kroku. Dohromady jsou součásti funkce, jejíž požadovanými vstupy jsou zájmová proměnná, měřená a predikovaná data, taktéž jako soubor kovariančních matic a hodnoty dynamického operátoru. Součásti tohoto výpočtu je dále získání predikce na týden dopředu pomocí $AR(1)$ procesu a její následná asimilace s predikcí dle BILANu. Funkce jsou následně zahrnuty do jednotného výpočtu přes všechny zájmové proměnné období roku 2019, jenž je uveden ve skriptu B.4.

2.2.2 Výsledky

2.2.2.1 Ověření konceptu na jednom povodí

Ověření navrženého konceptu bylo prvně provedeno na jednom povodí a to konkretně na povodí s identifikátorem OHL_0030. Vyhodnocení chyb bylo prováděno na základě několika kritérií (*MAE* (1.2) , *MSE* (1.1) a *RMSE*), přičemž se vyhodnocoval výsledek předpovědí autoregresního procesu, modelu BILAN a kombinace modelu BILAN a Kalmanova Filtru pro každou ze zvolených proměnných hydrologické bilance. Kritérium *RMSE* (Root Mean Square Error) se stanoví podle vzorce:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \tag{2.2}$$

kde $\hat{y}_i$ jsou hodnoty predikované a y_i jsou hodnoty pozorované. Za pozorované hodnoty v této části práci se považují výstupy simulace modelu BILAN pomocí parametrů získaných podle metodiky uvedené v kapitole 1.2 a skriptu B.1. Vizualizace výsledků zmíněných predikčních metod oproti „pozorovaným“ hodnotám je zobrazena na obrázku 2.2 a výsledné hodnoty všech tři kritérií jsou uvedeny v tabulce 2.2.

Tabulka 2.2: Hodnoty kritérií MAE, MSE a RMSE pro vybrané prvky hydrologické bilance povodí ‘OHL_0030’

	T		
	AR(1)	BILAN	KF
MAE	2.396	2.474	2.233
MSE	8.346	8.566	7.180
RMSE	2.889	2.927	2.680

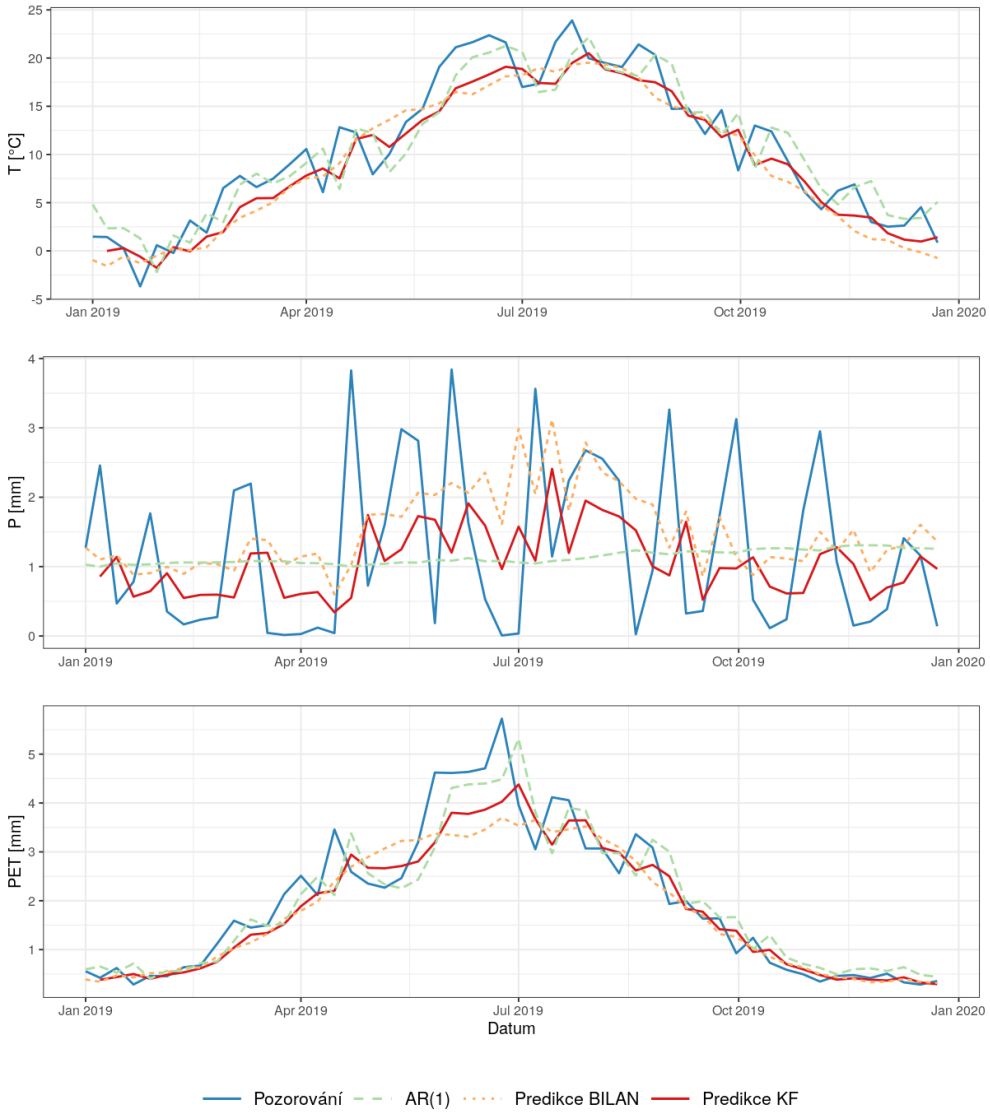
	P		
	AR(1)	BILAN	KF
MAE	1.024	0.995	0.966
MSE	1.439	1.454	1.436
RMSE	1.200	1.206	1.198

	PET		
	AR(1)	BILAN	KF
MAE	0.402	0.403	0.379
MSE	0.314	0.343	0.281
RMSE	0.561	0.585	0.530

Z výsledků v tabulce je patrné, že navržený přístup dosahuje nejnižší chyby oproti zbylým metodám, a to přes všechna kritéria a pro všechny proměnné. Tedy na základě uvedených výsledků lze označit navržený koncept metody za funkční a přejít k testování na více povrchových útvarech.

2.2.2.2 Aplikace na více povrchových útvarech

Pro ověření tvrzení o funkčnosti navrženého konceptu bylo potřeba ho ověřit na více povrchových útvarech, z tohoto důvodu bylo náhodně zvoleno sto takových útvarů. Jako míra přesnosti předpovědí se tentokrát zvolilo kritérium *MASE* (Mean Absolute Scaled Error), navržený Hyndman – Koehler (2006) a to



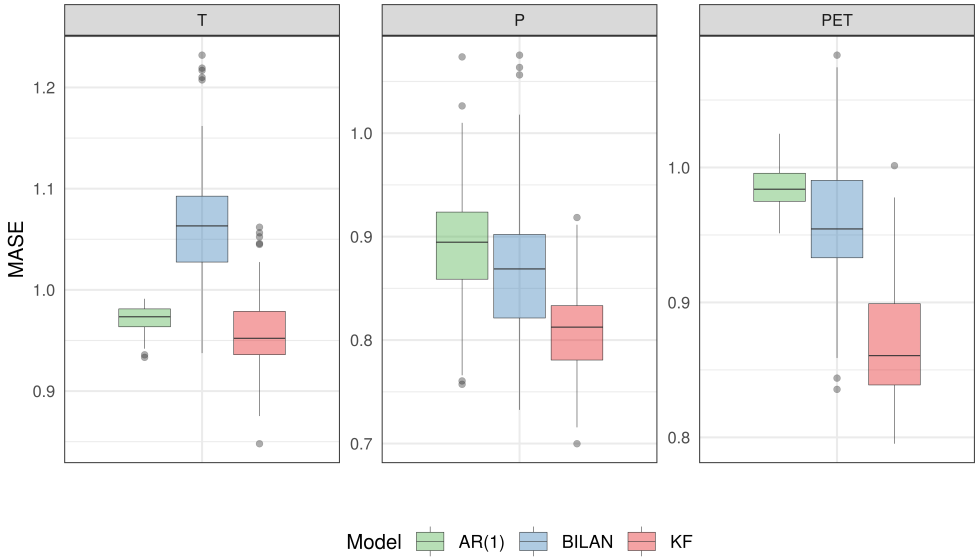
Obrázek 2.2: Predikce pomocí AR(1), BILANu a BILANu s použitím KF oproti pozorováním vybraných proměnných hydrologické bilance

hlavně z důvodu měřítkové invariance a tedy porovnatelnosti výsledků napříč povodími. Výpočet probíhá podle vzorce:

$$MASE = \frac{\frac{1}{J}\sum_{j=1}^{J}|\hat{y}_j - y_j|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}, \tag{2.3}$$

kde J je délka predikčního horizontu a n je délka časové řady. Zápis ve formě funkce je uveden ve skriptu B.5. Grafická vizualizace výsledků pro všech 100 povrchových útvarů je uvedena na obrázku 2.4. Vyhodnocení formou krabicového grafu je na obrázku 2.3. Dle tohoto grafu je patrné, že kombinace predikcí pomocí KF v průměru snížila chyby. Dále procentuální úspěšnost dosažení nejnižší chyby

přes všechna zvolená povodí pro každou z metod je shrnuta v tabulce 2.3, a zde opět je úspěšnost kombinace potvrzena. Kombinaci pomocí KF se podařilo získat nejnižší predikční chybu v 68% povodí za teplotu, 88% za srážky a ve 100% případů u potenciální evapotranspirace.



Obrázek 2.3: Krabicový graf výsledků MASE kritéria na 100 povodích pro proměnné teploty (T), srážek (P) a potenciální evapotranspirace (PET) pro modely AR(1), BILAN a BILAN s aplikaci KF

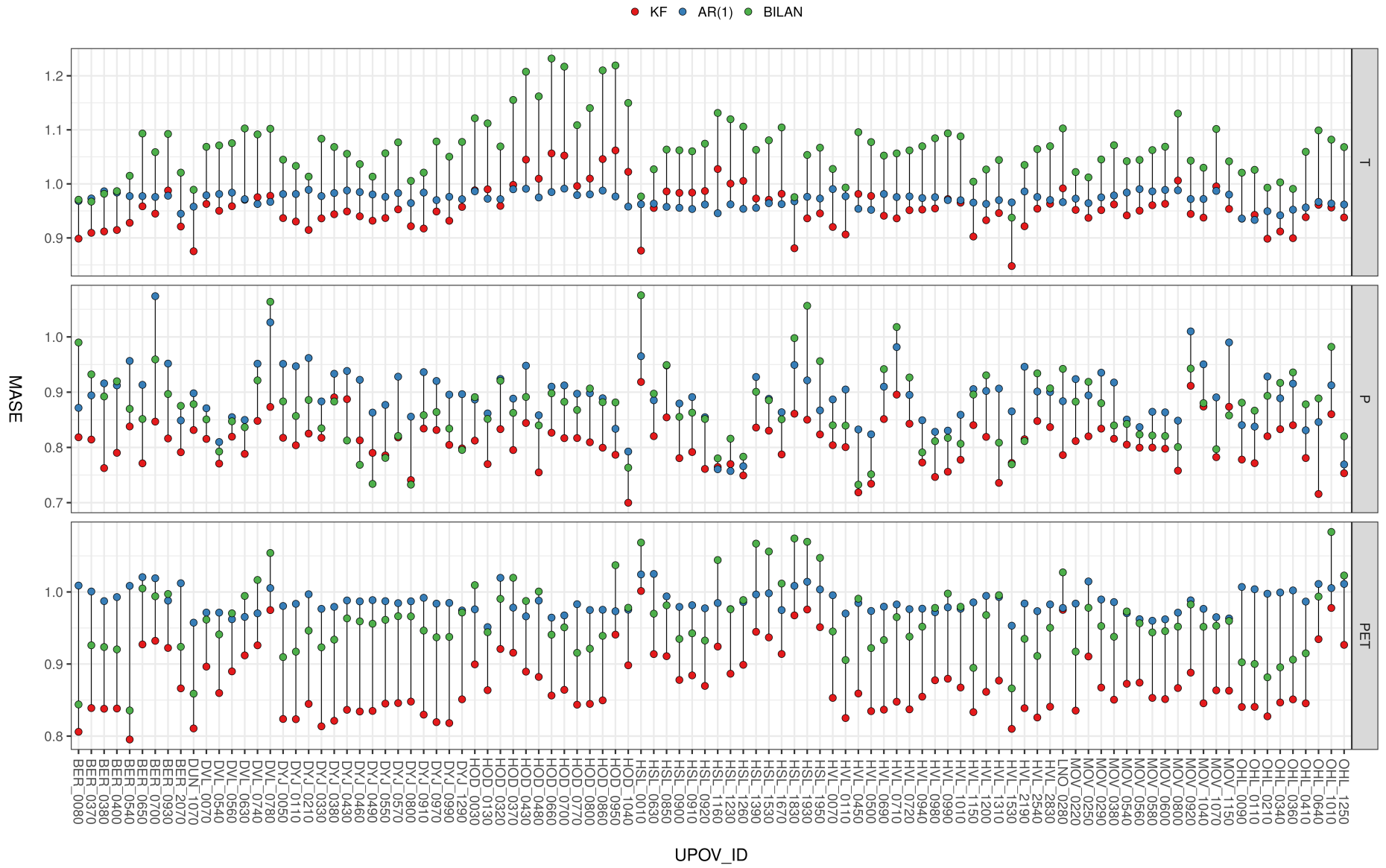
Tabulka 2.3: Procentuální úspěšnost metod podle kritéria MASE na vzorku sta povodí pro vybrané prvky hydrologické bilance

T		
KF	AR(1)	BILAN
68%	32%	0%

P		
KF	AR(1)	BILAN
88%	2%	10%

PET		
KF	AR(1)	BILAN
100%	0%	0%

Z výsledků je tedy zřejmé, že navržená metoda s aplikaci KF z převážné části vedla k lepším výsledkům, než další představené metody. Nižší predikční schopnost metody KF v případě teploty vzduchu lze vysvětlit faktem, že podle tiskových zpráv NASA (Brown, 2020) a WMO (WMO, 2020a) byl rok 2019 druhým nejteplejším zaznamenaným rokem. Parametry $\boldsymbol{R}_k$, $\boldsymbol{Q}_k$ a $\boldsymbol{M}_k$ odhadnuté na roce 2018 tedy tuto skutečnost nemohly postihnout. Řešením by mohla být změna období či samotné metody odhadu těchto parametrů. V případě srážek nízká sériová závislost mezi jednotlivými pozorováními ovlivňuje predikční schopnost AR procesu a tedy se snižuje i vliv tohoto modelu v rámci celkové metodiky.



Obrázek 2.4: Výsledek MASE kritéria na 100 povodích pro proměnné teploty (T), srážek (P) a potenciální evapotranspirace (PET) pro modely AR(1), BILAN a BILAN s aplikaci KF

2.3 Zlepšení simulace pomocí PU_EnKF

Navržený koncept pro aplikaci partitioned update schématu je založen na ansámblu o šedesáti členech. Pro kalibraci modelem BILAN byla použita data tvořená dvojicemi let od roku 1981 do roku 2000, konkretně deset po sobě jdoucích dvojic a padesát dvojic náhodně kombinovaných. Kalibrace parametrů modelu se provedla metodou diferenciální evoluce (DE). Takto koncipované ansámbly byly vytvořeny pro šest náhodně zvolených povodí. Výsledky kalibrace vyhodnocené na základě kritéria *NSE* (Nash-Sutcliffe efficiency, (1.4)) pro všechna vybraná povodí jsou uvedeny v tabulce 2.4. Skript pro výpočet ansámblů všech povodí je uveden v příloze B.7. Na obrázku 2.5 je pak znázorněn modelovaný odtok za pomocí průměrů z parametrů všech členů ansámblů oproti měřenému odtoku na povodích.

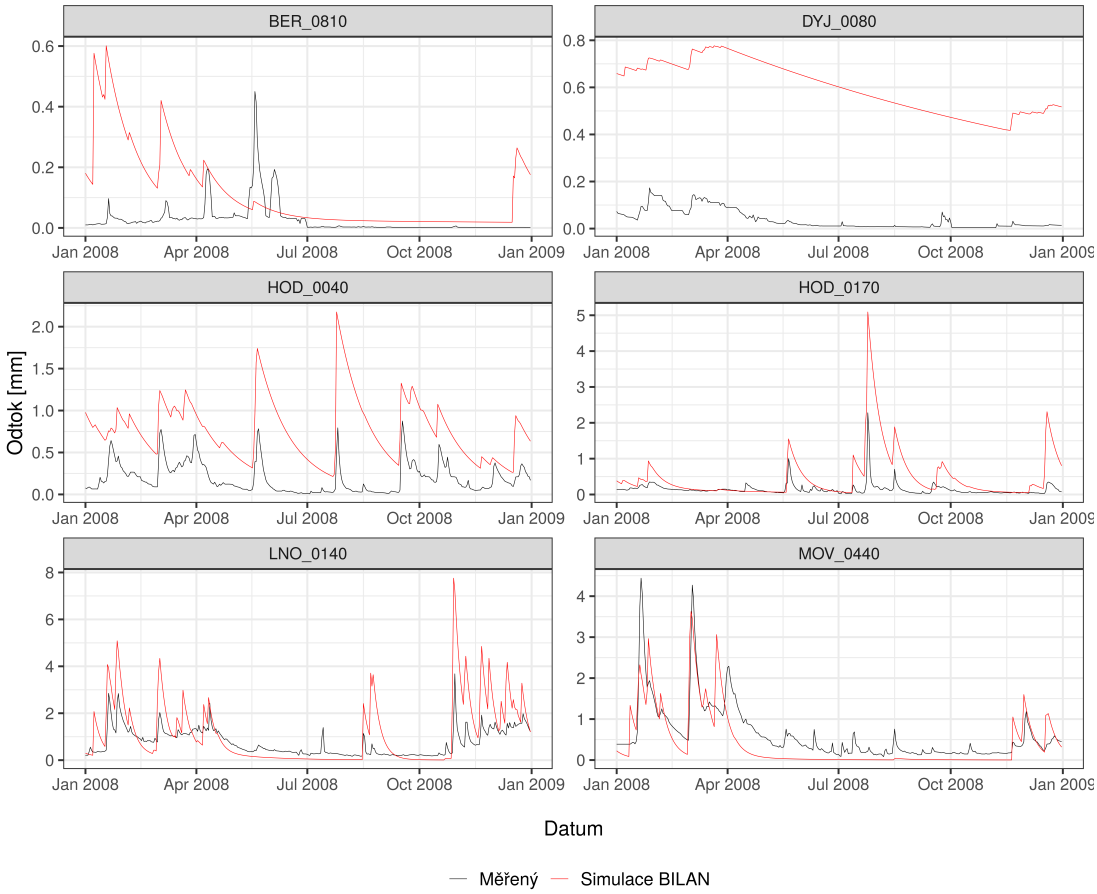
Tabulka 2.4: Souhrnné statistiky kritéria NSE, vypočítaných přes jednotlivé členy ansámblů v rámci zájmového povodí

UPOV_ID	Minimum	Průměr	Medián	Maximum
BER_0810	0.0005	0.0118	0.0099	0.0499
DYJ_0080	0.0004	0.0036	0.0029	0.0116
HOD_0040	0.0188	0.1146	0.1009	0.3403
HOD_0170	0.0037	0.1004	0.0633	0.5468
LNO_0140	0.0776	0.2176	0.1967	0.4158
MOV_0440	0.0956	0.5179	0.3818	1.3928

Pro simulaci řad odtoku modelem BILAN byla napsána funkce, vstupy do které jsou pouze datový soubor, obsahující časové řady srážkových úhrnů, teplot vzduchu a potenciální evapotranspirace a parametry. Tento krok je dále uplatněn v rámci predikční fáze analýzy, a tedy začleněn do funkce, která byla sestavena dle schématu 1.6. Obě funkce jsou uvedeny ve skriptu B.8. Jako vstupy do funkce slouží parametry ansámblů, simulace tvořené těmito parametry, měřená pozorování odtoku na povodí, časové rozmezí dat, chybová kovarianční matice $\boldsymbol{R}_t$ a takzvaný „shrinkage factor“ α. Odchýlením od postupu uvedeném v kapitole 1.3.4, je ponechávání parametru beze změny, pokud variance a priori stavu se blíží nule, jelikož změna parametru za takových podmínek je nestabilní a tedy nepřispívá ke změně odhadu.

Chybová kovarianční matice $\boldsymbol{R}_t$ a „shrinkage factor“ α se stanovily experimentálně, a to tak, že se vyzkoušela řada možných kombinaci mimo zájmové období, přičemž se ukázalo, že změnu hodnoty α (v rozmezí 0.95 až 0.99 s rozestupem 0.01) by se dalo označit za bezvýznamnou. Optimální hodnota $\boldsymbol{R}_t$ z testovaného rozmezí odpovídá 0.02, s 90% úspěšnosti jak na testovací, tak i na trénovací

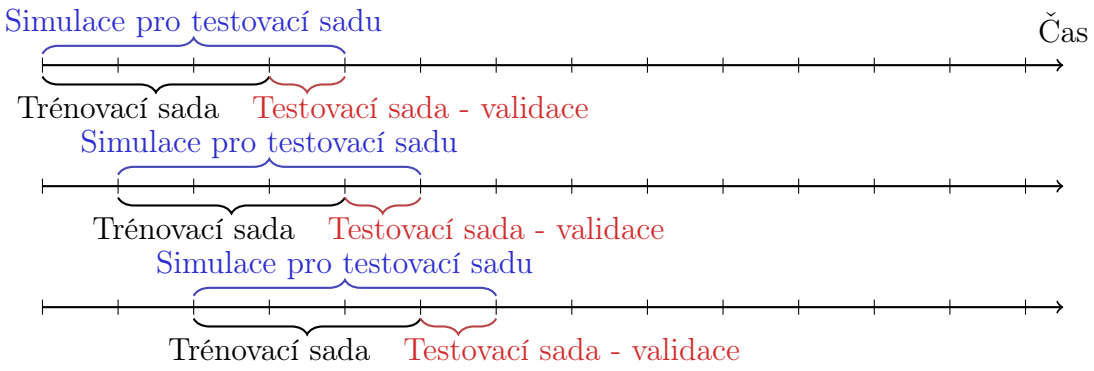
sadě v rámci zvoleného období.



Obrázek 2.5: Simulace řad odtoku modelem BILAN za pomocí průměrů z parametrů všech členů ansámblů pro vybraná povodí

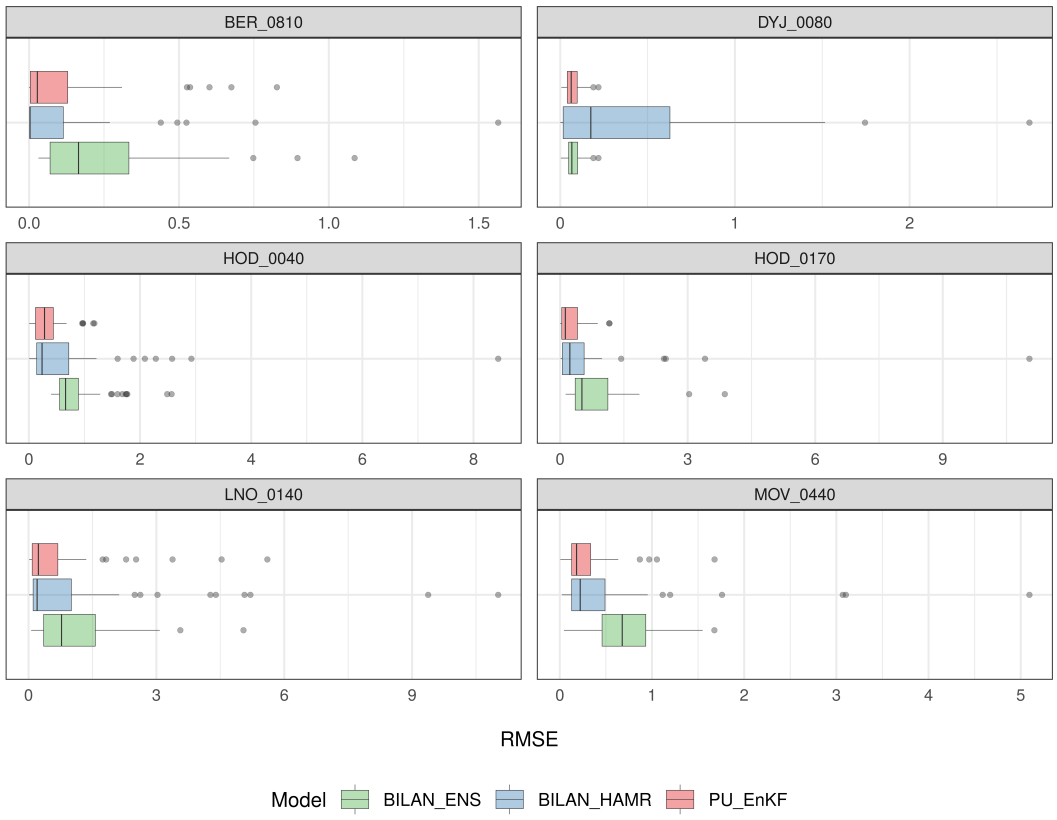
Validace navrženého konceptu proběhla za pomocí metody křížové validace. Rok 2008 byl rozdělen na 51 úseků, a to klouzavým oknem o délce tři týdny s jednotýdenním posunem. Tyto úseky jsou dále označovány za trénovací. Parametry odhadnuté na trénovací sadě následně vstupují do výpočtu pro testovací sadu. Simulace pro testovací sadu začíná na začátku trénovacího období a skládá se tak z třech týdnů trénovací sady a jednoho týdne na ně navazujícího (viz schéma 2.6). Vyhodnocení chyb je prováděno pouze na tomto navazujícím týdnu, a to za použití kritéria *RMSE* (2.2). Skript popsaného postupu je uveden v příloze B.9.

Shrnutí výsledků chyb vypočtených formou *RMSE* do formy krabicového grafů je uvedeno na obrázku 2.7. Mezi vyhodnocení chyb byla zahrnutá i simulace modelem BILAN s parametry vypočtenými pro jednotlivá povodí v rámci systému HAMR. Z grafu je patrné, že navržený koncept metody PU_EnKF při porovnání s výsledky průměrů parametrů ansámblů, přispívá ke značnému snížení mediánu chyby a jejího rozptylu v každém ze zkoumaných povrchových



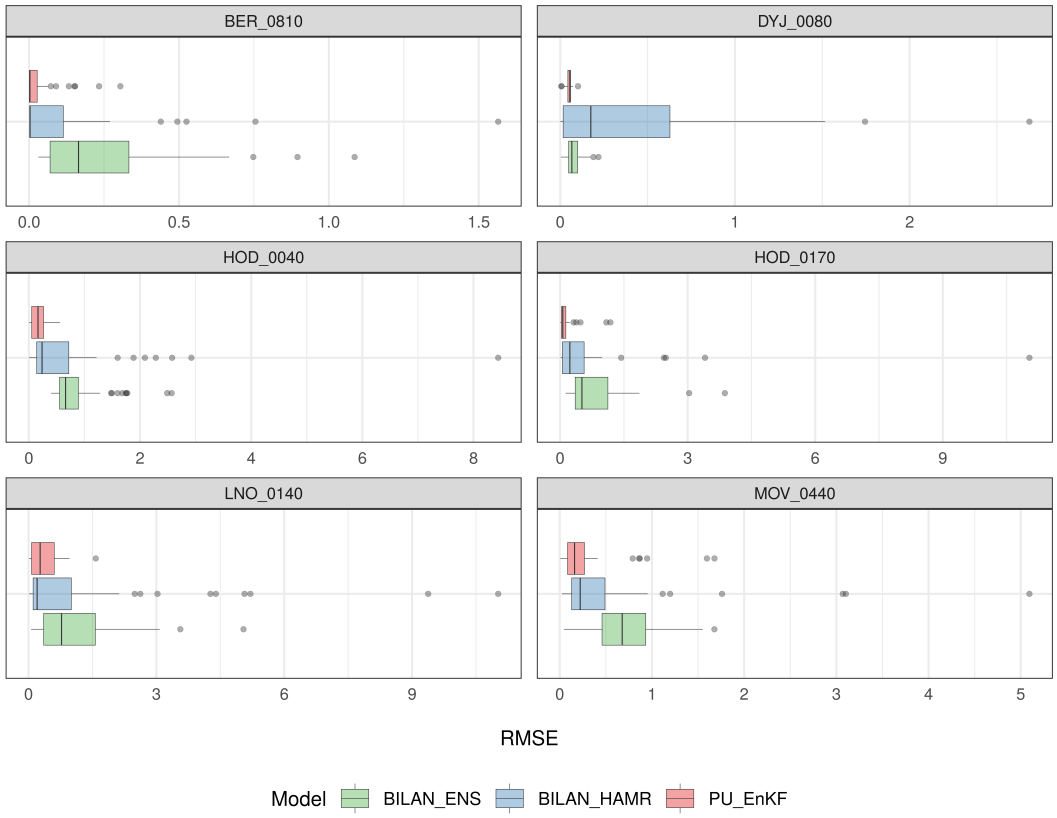
Obrázek 2.6: Schéma křížové validace metody PU_EnKF

útvarů. Při srovnání dosažených výsledků metody PU_EnKF s výsledky ze systému HAMR, mediány chyb jsou dosti srovnatelné, místy vyšší, než v případě parametrů z HAMRu, stále však došlo ke značnému snížení rozptylu chyby.



Obrázek 2.7: Porovnání kritéria RMSE na testovacích sadách pro vybraná povodí: simulace modelem BILAN s průměry parametrů ansámblů (`BILAN_ENS`), modelem BILAN s parametry vypočtené v rámci systému HAMR (`BILAN_HAMR`) a simulace s aplikací metody datové asimilace (`PU_EnKF`)

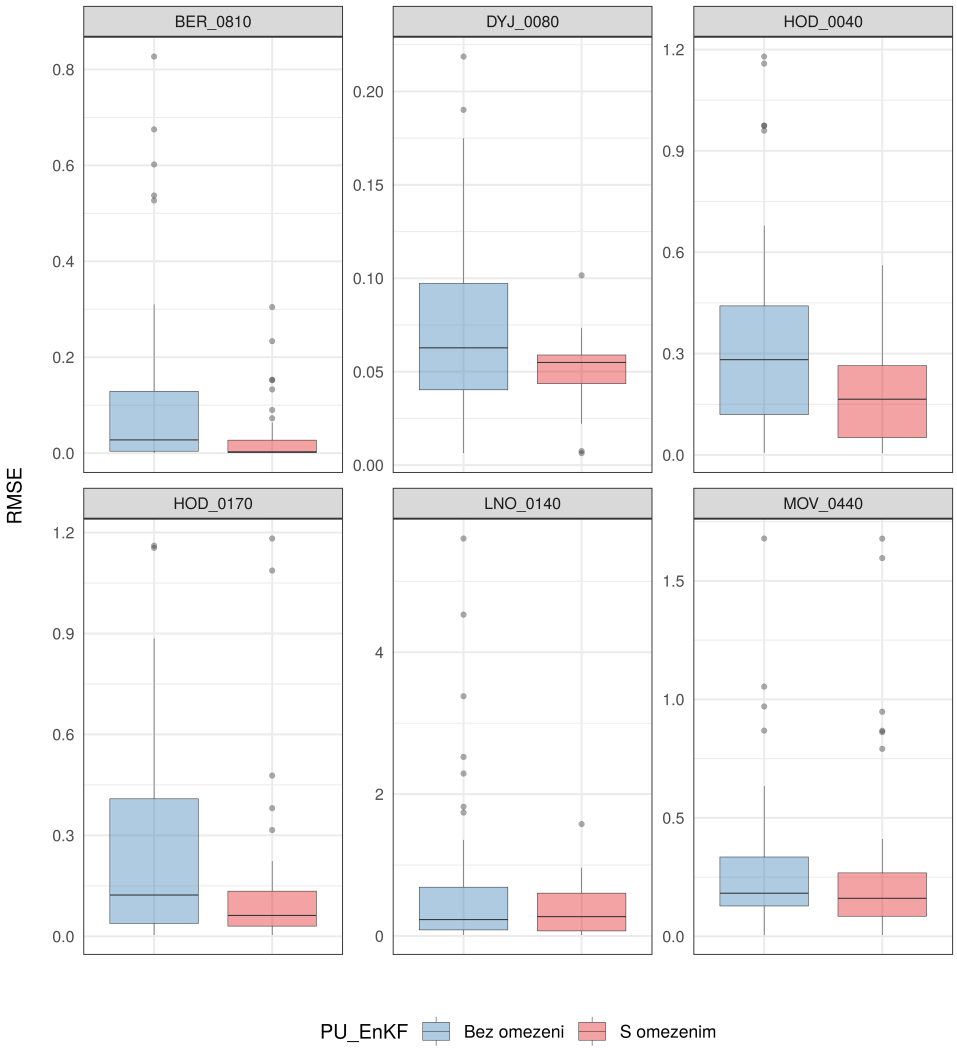
Vizualizace výsledků simulací oproti měřenému odtoku pro všech 51 testovacích období pro každé ze zvolených povodí jsou uvedeny v příloze na obrázku A.1. Zobrazené jsou pouze týdny, na kterých probíhala validace. U povrchového útvaru s identifikátorem `LNO_0140` došlo v úsecích 42 a 44 k nadhodnocení odtoků. Při bližším zkoumání chování parametrů v tomto období se ukázalo, že hodnoty jednoho z parametrů se dostaly do záporných hodnot. Z tohoto důvodu bylo rozhodnuto o zavedení korekce do výpočtu, obdobně postup aplikuje například Clark et al. (2008). Tedy pro dodržení fyzikální reality byl stanoven rozsah parametrů na hodnoty 0 až 200. Přidání dalších omezení pro horní hranici parametrů se však na zkoumaných povodích prokázalo jako nevýznamné. Vliv na výsledek výpočtu tak mělo především omezení dolní hranice parametrů. Skript uvedený v příloze B.8 tyto změny již zahrnuje.



Obrázek 2.8: Porovnání kritéria RMSE na testovacích sadách pro vybraná povodí po zavedení korekce: simulace modelem BILAN s průměry parametrů ansámblů (`BILAN_ENS`), modelem BILAN s parametry vypočtené v rámci systému HAMR (`BILAN_HAMR`) a simulace s aplikací metody datové asimilace (`PU_EnKF`)

Výsledky chyb ve formě krabicového grafu po zavedení omezení parametrů jsou zobrazeny na obrázku 2.8, porovnání chyb samotné metody PU_EnKF s omezením a bez je pak k dispozici na obrázku 2.9. Z grafu je patrné, že došlo k plošnému snížení variability chyb metody PU_EnKF. To samé lze říct i o medi-

ánu chyb, s výjimkou povodí LNO_0140, kde výsledek obou metod je srovnatelný. Tyto výsledky tak indikují pozitivní dopad zavedení omezení hodnot parametrů. Vizualizace simulací pro všech 51 testovacích úseků jsou opět uvedeny v příloze A.2. Vychýlení v odtoku, jak tomu bylo v obdobích 42 a 44 předchozího výpočtu, již na žádném z povodí zaznamenáno nebylo. Navržený koncept metody datové asimilace PU_EnKF tedy lze označit za vhodný k rozšíření a možné aplikaci v rámci většího systému.



Obrázek 2.9: Porovnání vlivu zavedení omezení parametrů do výpočtu PU_EnKF za pomocí kritéria RMSE na testovacích sadech pro vybraná povodí

Diskuze

Metody datové asimilace a jejich přesvědčivé výsledky byly hlavní motivací za výzkumem jejích implementace do hydrologického modelu BILAN. Jejich použiti následně přineslo převážně pozitivní výsledky v rámci zkoumaných metod této práce.

Koncept zlepšení predikcí pomocí kombinace dvou modelů za použití Kalman Filter se ukázal na vzorku sta povodí jako celkově úspěšný. Stále však existuje potenciál k jeho vylepšení. Konkrétně by se jednalo o nalezení reprezentativního období pro odhad parametrů, schopného postihnout jak extrémní situace, tak i běžný stav dat. Dále by bylo přínosné vyzkoumat možnosti jiných metod při odhadů parametrů a vliv jiných modelů či procesů při získání následně asimilovaných predikcí, než jednoduchého $AR(1)$ procesu.

Při implementaci metody partitioned update schématu PU_EnKF, hodnoty jednoho z parametrů se dostali pod dolní mez stanovenou modelem BILAN. Toto chovaní bylo ošetřeno omezením parametrů pouze na povolené rozmezí, což vedlo ke zlepšení ve výsledcích. Důležité je v tomto případě zmínit, že i přestože navržený koncept vychází z metody Xie – Zhang (2013), používá se v této práci pouze pro úpravu parametrů modelu a tedy neuplatňuje autory původně navržené upravení samotných stavů. Toto je vlastním důvodem k zavedení omezení parametrů, které jsou následně využity modelem BILAN pro tvorbu odhadu těchto stavů.

Předmětem dalšího výzkumu v rámci konceptu založeného na PU_EnKF by mohla být volba vhodné délky asimilovaného období. Tedy období, při kterém by nedošlo k přeučení parametrů a které by bylo zároveň dostatečně rozsáhlé, aby z něj model obdržel co nejvíce informací.

Závěr

Cílem práce bylo implementovat metody datové asimilace do hydrologického modelu BILAN. Vyzkoušeny byly metody pro zlepšení jak predikcí, tak i simulací modelu. Prvně byl navržen koncept kombinující predikce modelu BILAN s výsledky alternativního modelu, konkrétně jednoduchým $AR(1)$ procesem, pomocí sekvenční metody datové asimilace - Kalman Filter. Koncept byl aplikován na predikcích tři prvků hydrologické bilance: teplotách vzduchu, atmosferických srážkách a potenciální evapotranspiraci. Vyhodnocení výsledků proběhlo na sto náhodně zvolených povrchových útvarech České republiky pomocí kritéria *MASE*. Při vyhodnocení výsledků se ukázalo, že v případě predikce pomocí metody s aplikací KF došlo ke snížení chyby o 10.3% u teplot, o 7.15% u srážek a o 9.26% u potenciální evapotranspirace oproti běžné predikční metodice modelu BILAN.

Dále byl navržen koncept pro zlepšení simulací odtoku modelu BILAN, a to za pomoci úprav parametrů modelu rozšířenou metodou sekvenční datové asimilace založené na ansámblovém přístupu - Partitioned update Ensemble Kalman Filter. Pro šest náhodně zvolených povodí byly vytvořeny šedesáti-členné ansámbly parametrů, kalibrovaných modelem BILAN. Pro asimilaci byla použita data odtoku z měrných stanic poskytnutá ČHMÚ. Pro vyhodnocení metody byla použita křížová validace s trénovací sadou o délce tři týdny, aby se zamezilo přeučení parametrů, a testovací sadou o délce jednoho týdne. Posunutím trénovací sady formou klouzavého okna s jednotýdenním posunem proběhla validace po délce celého roku. K následnému vyhodnocení se tedy vytvořilo celkem padesát jedna výsledků z testovacích sad za každé povodí. Vyhodnocení proběhlo za pomocí kritéria *RMSE* a výsledky ukazují, že za použití metody PU_EnKF se podařilo snížit chyby minimálně o 48.7%, a to v případě povodí `MOV_0440`, a maximálně až o 88.4%, v případě povodí `DYJ_0080`, oproti běžné simulační metodice modelu HAMR s neměnnými parametry. Dalo by se to vysvětlit zvyšující se flexibilitou modelu při uvolňování parametrů, navazující na zvyšující se schopnost modelu přizpůsobit se k nastávajícím klimatickým změnám a postihnout i menší, například sezonální výkyvy.

Výsledky by však měly být popsány i z hlediska jejich limitací. Byla testována pouze mála část povodí České republiky a tedy při dalším výzkumu by se mohla odhalit nezbytnost dalších korekcí konceptů. Dále v případě metody PU_EnKF existuje prostor ke zlepšení výsledků kalibrace parametrů při tvorbě členů ansámblů, což by následně mělo vést k celkovému zlepšení výsledků. Vzhledem k momentálnímu koncipování metod a jejich funkcí, které je striktně zaměřené na jejich testování a následnou validační hodnotu výsledků, bude

potřeba vzít v potaz nezbytné změny v jejich struktuře před možným plošným zavedením do výpočtu v rámci systému HAMR. Mezi nimi je například i navržení postupu výpočtu v případě složených povodí.

Zlepšení výsledků predikcí prvků hydrologické bilance těsně souvisí s následným zlepšením předpovědí jak modelovaného odtoku modelem BILAN, tak i některých indexů sucha (například SPEI, SRI), vzhledem k tomu, že jejich metodika výpočtu je postavená na těchto prvcích. Přesnější výsledky jak simulací, tak i predikcí hydrologického modelu BILAN vede ke zdokonalení v monitorování a predikci sucha v rámci systému HAMR. Toto následně vede k větší informovanosti při přijetí opatření proti suchu či regulaci odběrů vody v případě jeho nastání. V této práci byly navrženy koncepty implementující metody datové asimilace do modelu BILAN. Jejich funkčnost pokládá základy pro implementaci těchto konceptů do stálého provozu a otevírá cesty pro další výzkum ve vymezené problematice.

Literatura

ABEBE, A. J.. – PRICE, R. K.. Information Theory and Neural Networks for Managing Uncertainty in Flood Routing. 2004. doi: 10.1061/(ASCE)0887-3801(2004)18:4(373).

ASCH, Mark. – BOCQUET, Marc. – NODET, Maëlle. *Data Assimilation: Methods, Algorithms, and Applications*. Philadelphia, PA : Society for Industrial and Applied Mathematics, Prosinec 2016. doi: 10.1137/1.9781611974546. ISBN 978-1-61197-453-9 978-1-61197-454-6.

BRÁZDIL, Rudolf. – KIRCHNER, Karel. *Vybrané přírodní extrémy a jejich dopady na Moravě a ve Slezsku*. Brno; Praha; Ostrava : Masarykova univerzita ; Český hydrometeorologický ústav ; Ústav geoniky Akademie věd České republiky, 2007. ISBN 978-80-210-4173-8.

BRÁZDIL, R. et al. Sucho v Českých Zemích: Minulost, Současnost a Budoucnost. *Historie počasí a podnebí v českých zemích*. 2015, , XI., s. 402.

BROWN, Katherine. NASA, NOAA Analyses Reveal 2019 Second Warmest Year on Record. Dostupné z: http://www.nasa.gov/press-release/nasa-noaa-analyses-reveal-2019-second-warmest-year-on-record, Leden 2020.

CLARK, M. P. et al. Hydrological Data Assimilation with the Ensemble Kalman Filter: Use of Streamflow Observations to Update States in a Distributed Hydrological Model. *Advances in Water Resources*. Říjen 2008, 31, 10, s. 1309–1324. ISSN 0309-1708. doi: 10.1016/j.advwatres.2008.06.005.

DUAN, Q. Y.. – GUPTA, V. K.. – SOROOSHIAN, S.. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and Applications*. Březen 1993, 76, 3, s. 501–521. ISSN 0022-3239, 1573-2878. doi: 10.1007/BF00939380.

EICKER, A. et al. Calibration/Data Assimilation Approach for Integrating GRACE Data into the WaterGAP Global Hydrology Model (WGHM) Using an Ensemble Kalman Filter: First Results. *Surveys in Geophysics*. Listopad 2014, 35, 6, s. 1285–1309. ISSN 1573-0956. doi: 10.1007/s10712-014-9309-8.

EVENSEN, Geir. Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*. 1994, 99, C5, s. 10143. ISSN 0148-0227. doi: 10.1029/94JC00572.

HLAVINKA, P. et al. Development and Evaluation of the SoilClim Model for Water Balance and Soil Climate Estimates. *Agricultural Water Management*. Květen 2011, 98, 8, s. 1249–1261. ISSN 03783774. doi: 10.1016/j.agwat.2011.03.011.

HYNDMAN, Rob J.. – KOEHLER, Anne B.. Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*. Říjen 2006, 22, 4, s. 679–688. ISSN 01692070. doi: 10.1016/j.ijforecast.2006.03.001.

IGLESIAS, Ana. – ASSIMACOPOULOS, Dionysis. – LANEN, Henny A.J. Van. Hydrometeorological Extreme Events - Drought: Science and Policy, 2019.

KALMAN, Rudolph Emil. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering*. 1960, 82, Series D, s. 35–45.

LAHOZ, William Albert. – KHATTATOV, Boris. – MÉNARD, Richard. (Ed.). *Data Assimilation: Making Sense of Observations*. Heidleberg ; New York : Springer, 2010. ISBN 978-3-540-74702-4 978-3-540-74703-1.

LANNOY, G. J. M. D. et al. Correcting for Forecast Bias in Soil Moisture Assimilation with the Ensemble Kalman Filter. *Water Resources Research*. 2007, 43, 9. ISSN 1944-7973. doi: 10.1029/2006WR005449.

LIU, Fang. *Bayesian Time Series: Analysis Methods Using Simulation-Based Computation*. PhD thesis, 2000.

LIU, Y. et al. Advancing Data Assimilation in Operational Hydrologic Forecasting: Progresses, Challenges, and Emerging Opportunities. *Hydrology and Earth System Sciences*. Říjen 2012, 16.

LORENC, Andrew. – GHIL, Michael. Unified Notation for Data Assimilation: Operational, Sequential and Variational. 1997, s. 21.

MÁCA, Petr. *Hydrologické Modely a Automatická Optimalizace Parametru: Neuronové Síte, Konceptuální Modely Globální Optimalizacní Algoritmy*. Habilitační Práce, ČZU Praha, 2015.

MÁCA, Petr. – VIZINA, Adam. – HORÁČEK, Stanislav. OPTIMALIZACE PARAMETRŮ MODELU BILAN METODOU SCDE. 2013, s. 12.

MARIANI, V. C. et al. A Hybrid Shuffled Complex Evolution Approach Based on Differential Evolution for Unconstrained Optimization. *Applied Mathematics and Computation*. Únor 2011, 217, 12, s. 5822–5829. ISSN 00963003. doi: 10.1016/j.amc.2010.12.064.

MCLAUGHLIN, Dennis. Recent Developments in Hydrologic Data Assimilation. *Reviews of Geophysics*. 1995, 33, S2, s. 977–984. ISSN 1944-9208. doi: 10.1029/95RG00740.

MŽP ČR. Udělali jsme další krok ke zvládání sucha. Vláda schválila návrh novely vodního zákona - Ministerstvo životního prostředí. Dostupné z: https://www.mzp.cz/cz/news_20190708_sucho-vodni-zakon, Červenec 2019.

OUDIN, L. et al. Estimating Potential Evapotranspiration without Continuous Daily Data: Possible Errors and Impact on Water Balance Simulations. *Hydrological Sciences Journal*. Březen 2010, 55, 2, s. 209–222. ISSN 0262-6667, 2150-3435. doi: 10.1080/02626660903546118.

Portál ODok. Materiál - Portál Aplikace ODok. Dostupné z: https://apps.odok.cz/veklep-detail?p_p_id=material_WAR_odokkpl&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=3&_material_WAR_odokkpl_pid=KORNB5ZDUQCE&tab=detail, 2019.

SEO, D.-J. et al. Automatic State Updating for Operational Streamflow Forecasting via Variational Data Assimilation. *Journal of Hydrology*. Duben 2009, 367, 3-4, s. 255–275. ISSN 00221694. doi: 10.1016/j.jhydrol.2009.01.019.

STORN, Rainer. – PRICE, Kenneth. Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. 1997, s. 15.

SUN, L. et al. Review of the Kalman-Type Hydrological Data Assimilation. *Hydrological Sciences Journal*. Říjen 2016, 61, 13. doi: 10.1080/02626667.2015.1127376.

TALLAKSEN, Lena M.. – LANEN, Henny A. J.. *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*. Elsevier, 2004. ISBN 978-0-444-51767-8.

TRNKA, M. et al. Metodika pro praxi. 2017, s. 66.

van Lanen, H. A. J. et al. Hydrological Drought Analysis in the Hupsel Basin Using Different Physically-Based Models. *FRIEND'97: Regional Hydrology: Concepts and Models for Sustainable Water Resource Management*. 1997, 246, s. 189.

VAN LANEN, H. A. et al. Hydrology Needed to Manage Droughts: The 2015 European Case: INVITED COMMENTARY. *Hydrological Processes*. Srpen 2016, 30, 17, s. 3097–3104. ISSN 08856087. doi: 10.1002/hyp.10838.

VAN LOON, Anne F.. Hydrological Drought Explained. *Wiley Interdisciplinary Reviews: Water*. Červenec 2015, 2, 4, s. 359–392. ISSN 20491948. doi: 10.1002/wat2.1085.

Vicente-Serrano, Sergio M.. – BEGUERÍA, Santiago. – López-Moreno, Juan I.. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*. Duben 2010, 23, 7, s. 1696–1718. ISSN 0894-8755, 1520-0442. doi: 10.1175/2009JCLI2909.1.

VIZINA, Adam. *ANALÝZA PROPAGACE SUCHA V ČESKÉ REPUBLICE*. Disertační práce, ČZU Praha, 2014.

VIZINA, A. et al. HAMR: Online Systém pro Zvládání Sucha – Operativní Řízení Během Suché Epizody. *Vodohospodářské technicko-ekonomické informace*. 2018, 60, 5, s. 22–28. ISSN 0322–8916.

VLNAS, R. et al. *Metodika pro Stanovení Mezních Hodnot Indikátorů Hydrologického Sucha*. Praha : Výzkumný ústav vodohospodářský TG Masaryka, veřejná výzkumná instituce, 2014.

VRUGT, J.A.. – ROBINSON, B.A.. – HYMAN, J.M.. Self-Adaptive Multimethod Search for Global Optimization in Real-Parameter Spaces. *IEEE Transactions on Evolutionary Computation*. Duben 2009, 13, 2, s. 243–259. ISSN 1941-0026, 1089-778X. doi: 10.1109/TEVC.2008.924428.

VÚV TGM, v.v.i. Příručka: model hydrologické bilance Bilan. Technical report, Srpen 2015.

WEST, Mike. Mixture Models, Monte Carlo, Bayesian Updating and Dynamic Models. 1993, s. 11.

WMO. WMO Confirms 2019 as Second Hottest Year on Record. Dostupné z: https://public.wmo.int/en/media/press-release/wmo-confirms-2019-second-hottest-year-record, Leden 2020a.

WMO. WMO Statement on the State of the Global Climate in 2019. 2020b, s. 44. ISSN 978-92-62-11248-5.

WU, J. et al. Kalman Filter Based Time Series Prediction of Cake Factory Daily Sale. In *2017 10th International Congress on Image and Signal Processing,*

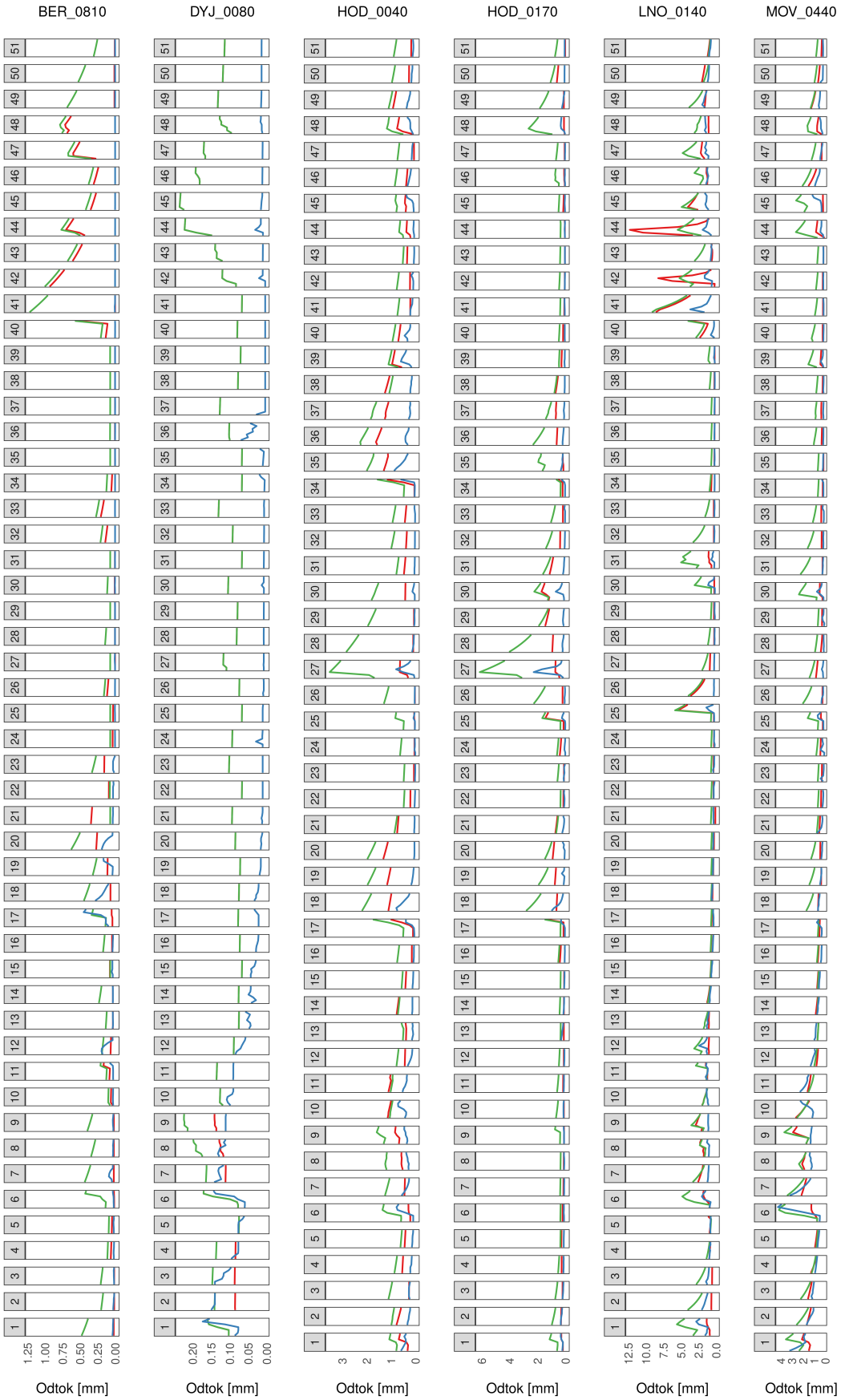
BioMedical Engineering and Informatics (CISP-BMEI), s. 1–7, Shanghai, Říjen 2017. IEEE. doi: 10.1109/CISP-BMEI.2017.8302108. ISBN 978-1-5386-1937-7.

WU, S.-J. et al. Real-Time Correction of Water Stage Forecast during Rainstorm Events Using Combination of Forecast Errors. *Stochastic Environmental Research and Risk Assessment*. 2011. doi: 10.1007/s00477-011-0514-4.

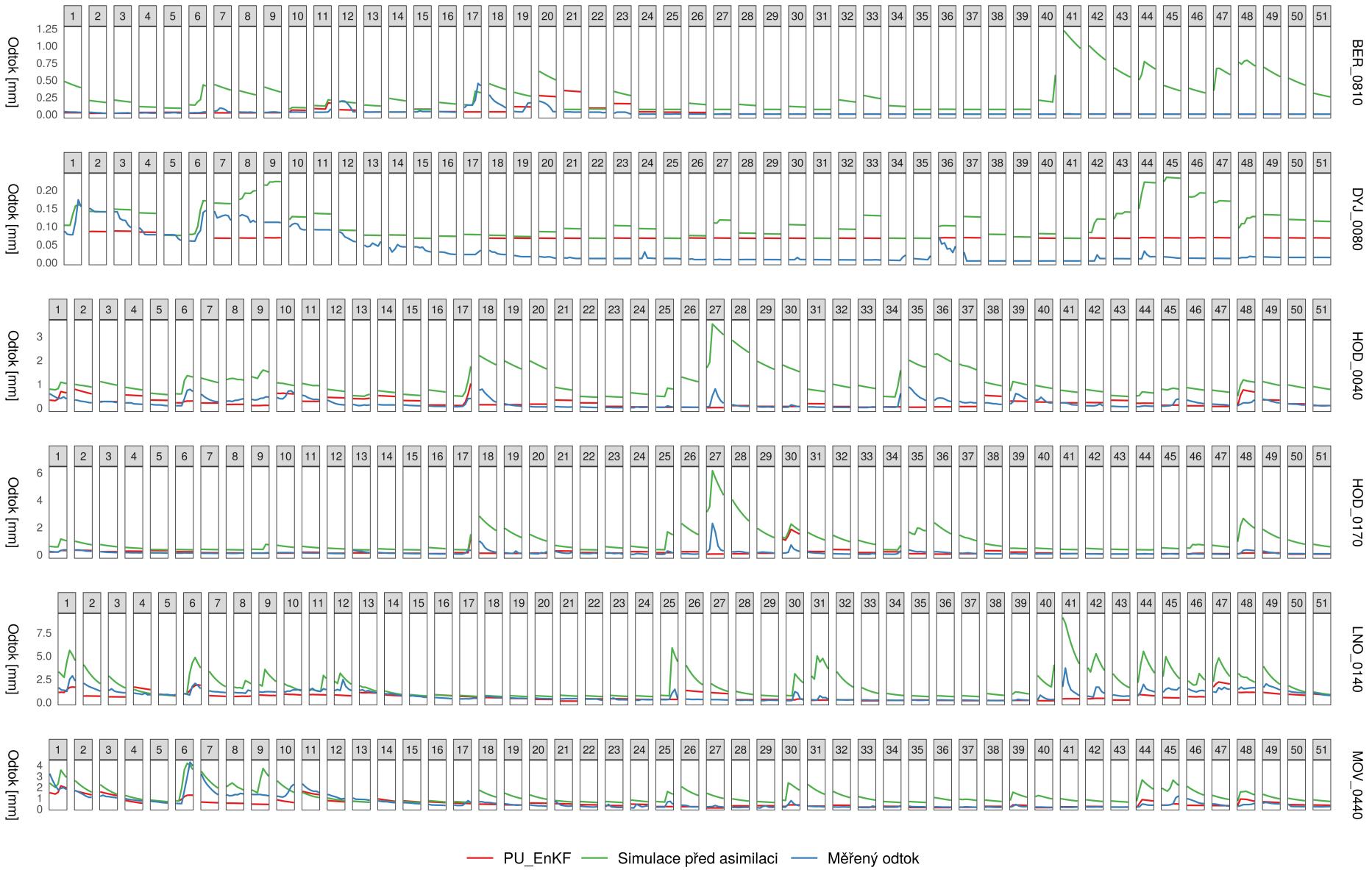
XIE, Xianhong. – ZHANG, Dongxiao. A Partitioned Update Scheme for State-Parameter Estimation of Distributed Hydrologic Models Based on the Ensemble Kalman Filter: A Novel Scheme for State-Parameter Estimation. *Water Resources Research*. Listopad 2013, 49, 11, s. 7350–7365. ISSN 00431397. doi: 10.1002/2012WR012853.

Přílohy

A Obrázky



Obrázek A.1: Průběh simulace modelu BILAN (průměr parametrů ansámblu) a průběh simulace metodou s aplikací PU_EnKF na týdenních testovacích sadech oproti průběhu měřených hodnot odtoků pro vybrané povrchové útvary ČR



Obrázek A.2: Průběh simulace modelu BILAN (průměr parametrů ansámblů) a průběh simulace metodou s aplikací PU_EnKF zavedení korekce na týdenních testovacích sadech oproti průběhu měřených hodnot odtoků pro vybrané povrchové útvary ČR

B Skripty

Pozn.: `filepath` je vlastní cestou k adresáři.

Skript B.1: Modelování průběhu časových řad prvků hydrologické bilance modelem BILAN pro období 1981-2019

```
library(data.table); library(bilan)
pars <- readRDS(file.path(filepath, 'data/pars.rds'))
for (i in seq_along(pars)) {
  UPOV.pars <- pars[[i]]
  UPOV.dta <-
    as.data.table(readRDS(paste0(
      filepath, 'data/rds/', names(pars[i]), '.rds'
    )))
  UPOV.dta <- UPOV.dta[, c('UPOV_ID', 'DTM', 'Rain',
                           'Tavg', 'ETr')]
  UPOV.dta <- UPOV.dta[DTM >= '1981-01-01' ,]
  UPOV.dta <-
    UPOV.dta[, DTM := seq(as.Date('1981-01-01',
                                  format = '%Y-%m-%d'),
                          by = 'month', length.out = .N)]
  b <- bil.new(type = 'd')
  p <- UPOV.pars$pars$current
  names(p) <- as.character(UPOV.pars$pars$name)
  bil.set.params.curr(model = b, params = p)
  bil.set.values(
    model = b,
    input_vars = data.frame(P = UPOV.dta[, Rain],
                            T = UPOV.dta[, Tavg],
                            PET = UPOV.dta[, ETr]),
    init_date = '1981-01-01',append = FALSE)
  bil.run(b)
  out <- as.data.table(bil.get.values(b)$vars)
  out <- out[, DTM := seq(as.Date('1981-01-01',
                                  format = '%Y-%m-%d'),
                          by = 'day',
                          length.out = .N)]
  saveRDS(out, paste0(filepath, 'data/BILout_day/',
                      names(pars[i]), '.rds'))
  print(paste('BILAN', i))
}
```

Pozn.: skript B.2 vyžaduje spouštění v terminálu pomocí `Rscript` pro úspornější nakládání s operační pamětí.

Skript B.2: Tvorba predikcí modelem BILAN pro rok 2019

```
library(data.table); library(bilan); library(lubridate)

pars <- readRDS(paste0(filepath, 'pars.rds'))
SEL <-
  sort(as.character(read.table(
    paste0(filepath, 'selectedUPOVs.txt'),
    sep = ',',
    stringsAsFactors = F
  )))

done <- gsub('.rds', '',
             dir(paste0(filepath, 'BILprediction2019')))
iter <- which(names(pars) %in% SEL[!(SEL %in% done)])

if (length(iter) != 0) {
  i = iter[1]

  UPOV.pars <- pars[[i]]
  UPOV.dta <-
    as.data.table(readRDS(paste0(filepath, 'rds/',
                                 names(pars[i]), '.rds')))
  UPOV.dta <- UPOV.dta[DTM >= as.Date('1961-01-01') , ]
  UPOV.dta[, week := week(DTM)]
  UPOV.dta[, rok := year(DTM)]

  roky <- 1981:2010
  out_fin <- NULL
  out2 <- NULL
  for (TYD in 1:52) {
    for (ROK in roky) {
      UPOV.dta.p <- UPOV.dta[DTM > as.Date('2010-12-31')]
      UPOV.dta.p <-
        UPOV.dta.p[DTM <= as.Date('2018-12-31') + weeks(TYD
           - 1)]
      UPOV.dta.p <-
        rbind(UPOV.dta.p, UPOV.dta[rok == ROK & week %in%
           TYD])

```

```
      b <- bil.new(type = 'd')

      p <- UPOV.pars$pars$current
      names(p) <- as.character(UPOV.pars$pars$name)

      bil.set.params.curr(model = b,
                          params = p)

      init.year22 <- 2011
      bil.set.values(
        model = b,
        input_vars = data.frame(
          P = UPOV.dta.p[, Rain],
          T = UPOV.dta.p[, Tavg],
          PET = UPOV.dta.p[, ETr]
        ),
        init_date = paste0(init.year22, '-01-01'),
        append = FALSE
      )

      bil.run(b)
      out <- as.data.table(bil.get.values(b)$vars)
      out[, id_rok := paste0(ROK)]
      out <- out[DTM > (as.Date('2018-12-31') + weeks(TYD -
         1))]
      out2 <- rbind(out2, out)
      gc()
    }
    if (TYD %% 5 == 0) {
      cat(TYD, 'T', collapse = ' ', sep = '')
    }
  }
  out_fin <-
    rbind(out_fin, out2[, lapply(.SD, mean), .SDcols =
       colnames(out)[-c(1, 20)], by = DTM])
  saveRDS(out_fin,
          paste0(filepath, 'BILprediction2019/',
                 names(pars[i]), '.rds'))
  cat(names(pars[i]), collapse = '\n')
}
q()
```

Skript B.3: Odhad kovariančních matic pro metodu KF

```
require(dplyr); require(lubridate)
SEL <- sort(as.character(read.table(paste0(filepath,
    'selectedUPOVs.txt'), sep = ',', stringsAsFactors = F)))
kf_in_est <- list()
for (upov in SEL) {
  dta2018 <- readRDS(paste0(filepath,
                      'BILout_day/', upov, '.rds'))
  BIL.predic <- readRDS(paste0(filepath,
                         'BILprediction2018/', upov, '.rds'))
  VAR <- c('T', 'P', 'PET')
  BIL.predic <- BIL.predic[BIL.predic$DTM < '2019-01-01',
                           c("DTM", VAR)]
  AR_est <- R_est <- Q_est <- NULL
  for (i in VAR) {
    dta <- dta2018 %>% select(value = i, DTM) %>%
      filter(year(DTM) == 2018) %>% group_by(week(DTM)) %>%
      summarise(DTM = first(DTM), value = mean(value)) %>%
      ungroup() %>% select(DTM, value)
    x <- dta$value
    predikce2018 <- BIL.predic %>%
      select(DTM, value = i) %>% mutate(week = week(DTM),
         year = year(DTM)) %>%
      group_by(year, week) %>% summarise(DTM = first(DTM),
         value = mean(value)) %>% ungroup()
    y <- predikce2018$value
    if (length(ar(x, order.max = 1)$ar) != 0) {
      AR_est[i] <- ar(x, order.max = 1)$ar
      Q_est[i] <- mean((ar(x, order.max = 1)$resid) ^ 2 ,
                       na.rm = T)
    } else {
      estPerc <- lm(data = dta, value ~ lag(value) - 1)
      AR_est[i] <- as.numeric(coef(estPerc))
      Q_est[i] <- mean(residuals(estPerc) ^ 2 , na.rm = T)
    }
    R_est[i] <- sum((y - x) ^ 2) / length(x)
  }
  kf_in_est[[upov]] <- rbind(AR_est, R_est, Q_est)
}
saveRDS(kf_in_est, file = paste0(filepath, 'KF_params.rds'))
```

Skript B.4: Funkce pro datovou asimilaci metodou KF

```
kalmanf <- function(s) {
  if (is.null(s$y))
    warning("No observation")
  return
  if (is.null(s$x))
    s$x <- rep(NA, length(s$y))
  if (is.null(s$P))
    s$P <- NA
  if (is.null(s$A))
    s$A <- diag(length(s$x))
  if (is.null(s$Q))
    s$Q <- matrix(0, length(s$y), length(s$y))
  if (is.null(s$R))
    warning("Observation covariance matrix is missing")
  return
  if (is.null(s$H))
    s$H <- diag(length(s$x))
  if (all(is.na(s$x))) {
    s$x <- solve(s$H) %*% s$y
    s$P <- solve(s$H) %*% s$R %*% solve(s$H) } else {
    s$x <- s$A %*% s$x
    s$P <- s$A %*% s$P %*% t(s$A) + s$Q
    s$K <- s$P %*% t(s$H) %*% solve(s$H %*% s$P %*% t(s$H) +
        s$R)
    s$x <- s$x + s$K %*% (s$y - s$H %*% s$x)
    s$P <- s$P - s$K %*% s$H %*% s$P
  }
  return(s)
}
```

```
kf_pred <- function(s, y, x) {
  state_1 <- list(s); x_kf <- NULL
  for (i in seq(2, 52)) {
    state_1[[i - 1]]$y <- y[i]
    state_1[[i - 1]]$x <- x[i - 1]
    state_1[[i]] <- kalmanf(state_1[[i - 1]])
    x_kf <- c(x_kf, state_1[[i]]$x)
  }
  return(x_kf)
}
```

```r
KFsolver <- function(var, obs, predic, s) {
  dta <-
    obs %>% select(value = var, DTM) %>% filter(year(DTM) >=
        2018)
  dta <-
    dta %>% mutate(week = week(DTM), year = year(DTM)) %>%
       group_by(year, week) %>%
    summarise(DTM = first(DTM), value = mean(value)) %>%
       ungroup()
  x <- dta$value
  ar1_fl <- NULL
  for (i in 53:104) {
    ar1_h <- ar(x[(i - 52):i], order.max = 1)
    ar1_fl <- c(ar1_fl, predict(ar1_h, n.ahead = 1)$pred)
  }
  x_2019 <- dta[dta$year == 2019, ]$value
  y_prB <-
    predic %>% select(DTM, value = var) %>% mutate(week =
       week(DTM), year = year(DTM)) %>% filter(week != 53)
       %>%
    group_by(year, week) %>% summarise(DTM = first(DTM),
       value = mean(value)) %>% ungroup()
  y <- y_prB$value

  x_kf <- kf_pred(s, y, x_2019)
  sim <-
    data.frame(
      DTM = y_prB$DTM,
      prKF = c(NA, x_kf),
      prBIL = y_prB$value,
      AR1 = ar1_fl,
      stringsAsFactors = F
    )
  real <-
    list(full = data.frame(DTM = dta$DTM,value = dta$value),
         half = data.frame(dta[dta$year == 2019,
                               c("DTM", "value")]))
  return(list(real = real, sim = sim))
}
```

Skript B.5: Funkce pro výpočet kritéria MASE

```
MASE <- function(sim, obs){
  e <- mean(abs(obs - sim), na.rm = T)
  y <- matrix(c(obs[2:length(obs)], obs[1:length(obs)-1]),
     byrow=F, ncol=2)
  X <- e/mean(abs(y[,1] - y[,2]))
  return(X)
}

MASEsolver <- function(simDT, obs, COLs){
  X <- data.frame(A = MASE(simDT[,1], obs),
                  B = MASE(simDT[,2], obs),
                  C = MASE(simDT[,3], obs))

  colnames(X) <- COLs
  return(X)
}
```

Skript B.6: Aplikace metody KF, výpočet pro sto povrchových útvarů s vyhodnocováním kritéria MASE

```
library(dplyr); library(lubridate); require(hydroGOF)
source(paste0(filepath, 'KF_fun.R'))
source(paste0(filepath, 'err_est_fun.R'))
SEL <-
  sort(as.character(read.table(
    paste0(filepath, 'data/selectedUPOVs.txt'),
    sep = ',',
    stringsAsFactors = F
  )))

mase_est <- NULL
VAR <- c('T', 'P', 'PET')

for (upov in SEL) {
  UPOV.dta <- readRDS(paste0(
    filepath, 'data/', 'BILout_day/', upov, '.rds'))
  BIL.predic <- readRDS(paste0(
    filepath, 'data/', 'BILprediction2019/', upov, '.rds'))
  KF_in_est <- readRDS(paste0(
    filepath, 'data/', 'KF_params.rds'))

```

```
  for (i in VAR) {
    s <- list(
      Q = matrix(KF_in_est[[upov]][3, i], 1, 1),
      A = matrix(KF_in_est[[upov]][1, i], 1, 1),
      P = matrix(1, 1, 1),
      R = matrix(KF_in_est[[upov]][2, i], 1, 1)
    )

    OUT <-
      KFsolver(
        var = i,
        obs = UPOV.dta,
        predic = BIL.predic,
        s = s
      )

    simDT <-
      cbind(
        KFmodel = OUT$sim$prKF,
        ARmodel = OUT$sim$AR1,
        BILmodel = OUT$sim$prBIL
      )
    OBS <- OUT$real$half$value

    mase_est <-
      rbind(mase_est, cbind(
        UPOV_ID = upov,
        MASEsolver(simDT[2:nrow(simDT), ],
                   OBS[2:length(OBS)],
                   c("KF", "AR(1)", "BILAN")),
        VAR = i
      ))
  }
}
```

Skript B.7: Kalibrace parametrů pro jednotlivé členy ansámblů pro každé ze zvolených povodí

```
library(data.table); library(bilan); library(lubridate)
set.seed(100)

randy <-
  matrix(sample(seq(1981, 2000), 104, replace = T),
         byrow = T,
         ncol = 2)
randy <- randy[!duplicated(randy), ]
randy <- randy[1:50, ]

mpa <-
  as.data.table(readRDS(paste0(
    filepath, 'data/merena_pro_asimilaci.rds'
  )))

for (upov in c("MOV_0440", "LNO_0140", "HOD_0170",
               "HOD_0040", "BER_0810", "DYJ_0080")) {
  ns <- NULL
  mpa <- mpa[UPOV_ID == upov, c('DTM', 'Q')]

  DT <-
    as.data.table(readRDS(paste0(filepath, 'data/rds/', upov
       , '.rds')))

  init.year <- seq(1981, 2000, 2)
  i = 1
  for (rok in init.year) {
    UPOV.dta <-
      DT[DTM >= paste0(rok, '-01-01') &
           DTM < (as.Date(paste0(rok, '-01-01')) + years(2))
              ,]
    real <-
      mpa[DTM >= paste0(rok, '-01-01') &
            DTM < (as.Date(paste0(rok, '-01-01')) + years(2)
               ),]

    b = bil.new(type = 'd')

    bil.set.values(
```

```
      model = b,
      input_vars = data.frame(
        P = UPOV.dta[, Rain],
        T = UPOV.dta[, Tavg],
        PET = UPOV.dta[, ETr],
        R = real[, Q]
      ),
      init_date = paste0(rok, '-01-01'),
      append = FALSE
    )

    bil.set.optim(b, method = 'DE')

    bil.optimize(b)
    out <-
      list(data = bil.get.values(b)$vars, params = bil.get.
         params(b))

    if (!dir.exists(paste0(filepath, 'data/Ensemble/', upov)
       )) {
      dir.create(paste0(filepath, 'data/Ensemble/', upov))
      dir.create(paste0(filepath, 'data/Ensemble/', upov, '/
         params'))
    }

    ns <- c(ns, bil.get.values(b)$crit)

    saveRDS(out,
            paste0(
              filepath,
              'data/Ensemble/',
              upov,
              '/params/',
              'set_',
              i,
              '.rds'
            ))
    print(paste0('set_', i))

    i = i + 1
  }
```

```
75
76  j = 11
77
78  for (i in 1:nrow(randy)) {
79    UPOV.dta <- DT[year(DTM) %in% c(randy[i, 1], randy[i,
        2]),]
80    real <- mpa[year(DTM) %in% c(randy[i, 1], randy[i, 2]),]
81
82    b = bil.new(type = 'd')
83
84    bil.set.values(
85      model = b,
86      input_vars = data.frame(
87        P = UPOV.dta[, Rain],
88        T = UPOV.dta[, Tavg],
89        PET = UPOV.dta[, ETr],
90        R = real[, Q]
91      ),
92      init_date = paste0(min(randy[i, ]), '-01-01'),
93      append = FALSE
94    )
95
96    bil.set.optim(b, method = 'DE')
97
98    bil.optimize(b)
99    out <-
100     list(data = bil.get.values(b)$vars, params = bil.get.
          params(b))
101
102   ns <- c(ns, bil.get.values(b)$crit)
103
104   saveRDS(out,
105           paste0(
106             filepath,
107             'data/Ensemble/',
108             upov,
109             '/params/',
110             'set_',
111             j,
112             '.rds'
113           ))
```

```
114     print(paste0('set_', j))
115
116     j = j + 1
117   }
118
119   write.table(
120     ns,
121     paste0(filepath, 'data/Ensemble/', upov, '/', 'NS.txt'),
122     dec = '.',
123     col.names = F,
124     row.names = F
125   )
126
127   par_est <- list()
128   for (i in dir(paste0(filepath, 'data/Ensemble/', upov, '/
          params/'))) {
129     par_est[[gsub('set_|.rds', '', i)]] = readRDS(paste0(
            filepath, 'data/Ensemble/', upov, '/params/', i))[['
            params']]
130   }
131
132   saveRDS(par_est,
133           paste0(filepath, 'data/Ensemble/', upov, '/
                  estimated_pars.rds'))
134
135   PARS <- as.data.frame(par_est)
136   PARS <- PARS[, grep('current', colnames(PARS))]
137   colnames(PARS) <- gsub('X|.current', '', colnames(PARS))
138   rownames(PARS) <- c('Spa', 'Alf', 'Dgm', 'Soc', 'Mec', '
          Grd')
139   PARS <- PARS[, c(as.character(1:length(par_est)))]
140   saveRDS(PARS,
141           paste0(
142             filepath,
143             'data/Ensemble/',
144             upov,
145             '/estimated_pars_unlisted.rds'
146           ))
147 }
```

Skript B.8: Funkce pro datovou asimilaci metodou PU_EnKF

```
bilEval <- function(UPOV.dta, lambda) {
  UPOV.dta <- as.data.frame(UPOV.dta)
  b <- bil.new(type = 'd')
  bil.set.params.curr(model = b,
                      params = lambda)
  bil.set.values(
    model = b,
    input_vars = data.frame(P = UPOV.dta[, 'Rain'],
                            T = UPOV.dta[, 'Tavg'],
                            PET = UPOV.dta[, 'ETr']),
    init_date = '2000-01-01',
    append = FALSE
  )
  bil.run(b)
  out <- bil.get.values(b)$vars$RM
  return(out)
}
```

```
puenkf <- function(PARS, ENS, OBS, dtm_range, R, alpha) {
  H <- matrix(c(1, 0), nrow = 1)
  Tt <- (1 - 0.95) * apply(PARS, 1, var)
  Theta_avg <- rowMeans(PARS)
  lambda <- Theta_avg
  names(lambda) <- rownames(PARS)
  param_ev <- PARS
  lambda_evolve <- list()
  iter <- 0
  tm_range <- which(OBS$DTM %in% dtm_range)
  for (tm in tm_range) {
    iter <- iter + 1
    for (j in 1:nrow(PARS)) {
      x_f <- rep(0, n_ens)
      lambda_ens <- lambda
      for (i in 1:n_ens) {
        lambda_ens[j] <- param_ev[j, i]
        x_f[i] <-
          bilEval(UPOV.dta[tm, , drop = FALSE], lambda_ens)
      }
      if (any(is.na(x_f)))
        break
```

```
23        if (round(var(x_f), 3) != 0) {
24          prtrb <- rnorm(n_ens, mean = 0,
25                         sd = 0.001) + OBS[tm, 'Q']
26          K <- matrix(c(var(as.numeric(x_f)),
27                        cov(as.numeric(x_f),
28                            as.numeric(param_ev[j,])
29                        )), ncol = 1) %*%
30            ((1 / var(as.numeric(x_f))) + R)
31          state <- list()
32          for (i in 1:(n_ens)) {
33            state_fin <-
34              matrix(c(x_f[i],
35                       param_ev[j, i])) + K %*%
36              (prtrb[i] - (H %*% matrix(c(x_f[i], lambda[j])))
                 )
37            state_fin[2] <-
38              ifelse(state_fin[2] < 0, 0, state_fin[2])
39            state_fin[2] <-
40              ifelse(state_fin[2] > 200, 200, state_fin[2])
41            state[[i]] <- state_fin
42          }
43          lambda[j] <- mean(unlist(lapply(state, function(x)
44            x[2])))
45          if (lambda[j] < 0)
46            break
47          if (any(is.na(lambda))) {
48            stop(paste0(tm, sep = '\n'))
49          }
50          param_ev[j,] <-
51            alpha * unlist(lapply(state, function(x)
52              x[2])) + (1 - alpha) * lambda[j] +
53            rnorm(n_ens, 0, Tt[j])
54        }
55        lambda_evolve[[iter]] <- lambda
56      }
57    }
58    return(list(lambda_evolve = lambda_evolve,
59                param_ev = param_ev))
60  }
```

Skript B.9: Výpočet metodou PU_EnKF, včetně křížové validace

```
library(bilan); library(lubridate)
source(paste0(filepath, 'PU_EnKF_fun.R'))

OBS_nf <- readRDS(paste0(filepath, 'data/merena_pro_
   asimilaci.rds'))

for (upov in c("MOV_0440",
               "LNO_0140",
               "HOD_0170",
               "HOD_0040",
               "BER_0810",
               "DYJ_0080")) {
  OBS <-
    OBS_nf[OBS_nf$UPOV_ID == upov &
             OBS_nf$DTM >= '2000-01-01', c('DTM', 'Q')]
  ENS <- readRDS(paste0(filepath, 'data/Ensemble/',
                        upov, '/ensemble_sim.rds'))

  n_ens <- ncol(ENS)

  PARS <-
    readRDS(paste0(
      filepath,
      'data/Ensemble/',
      upov,
      '/estimated_pars_unlisted.rds'
    ))

  UPOV.dta <- readRDS(paste0(filepath, 'data/rds/',
                             upov, '.rds'))
  UPOV.dta <- UPOV.dta[, c('UPOV_ID', 'DTM',
                           'Rain', 'Tavg', 'ETr')]
  UPOV.dta <-
    UPOV.dta[UPOV.dta$DTM >= '2000-01-01' &
               UPOV.dta$DTM <= '2010-10-31', ]

  first_week <- seq(as.Date('2008-01-01'),
                    as.Date('2008-01-01') + weeks(3), 1)
  dtm_set <- list()
  dtm_set[[1]] <- first_week
```

```
for (tyd in 1:50) {
  dtm_set[[tyd + 1]] <- first_week + weeks(tyd)
}

EnsParsMean <- rowMeans(PARS)
results_pars <- list()
i = 1

dir.create(paste0(filepath, 'data/Ensemble/',
                  upov , '/res'))

for (dtm_range in dtm_set) {
  next_period <-
    seq(dtm_range[length(dtm_range)] + days(1),
        dtm_range[length(dtm_range)] + days(7), 1)
  selected <- which(c(dtm_range, next_period) %in%
                      next_period)

  bilan_train <-
    bilEval(UPOV.dta[UPOV.dta$DTM %in% dtm_range, ],
       lambda = EnsParsMean)
  skut_train <- OBS[OBS$DTM %in% dtm_range, c('DTM', 'Q')]

  bilan_test <-
    bilEval(UPOV.dta[UPOV.dta$DTM %in% c(dtm_range,
                                         next_period), ],
            lambda = EnsParsMean)
  skut_test <-
    OBS[OBS$DTM %in% c(dtm_range, next_period),
        c('DTM', 'Q')]

  PU_ParsMean <-
    rowMeans(puenkf(PARS, ENS, OBS, dtm_range,
                    alpha = 0.95, R = 0.02)$param_ev)

  pu_enkf_train <-
    bilEval(UPOV.dta[UPOV.dta$DTM %in% dtm_range, ],
       lambda = PU_ParsMean)
  pu_enkf_test <-
    bilEval(UPOV.dta[UPOV.dta$DTM %in% c(dtm_range,
                                         next_period), ],
```

```
                lambda = PU_ParsMean)

    results_pars[[i]] <-
      list(pars = PU_ParsMean,
           data = reshape2::melt(
             data.frame(
               DTM = UPOV.dta[
                 UPOV.dta$DTM %in% next_period,
                 ]$DTM,
               R_n = pu_enkf_test[selected],
               R_o = bilan_test[selected],
               R_sk = skut_test[selected, 'Q']
             ),
             id.vars = 'DTM'
           ))
    i = i + 1
    rm(pu_enkf_train)
    gc()
  }
  saveRDS(
    results_pars,
    paste0(
      filepath,
      'data/Ensemble/',
      upov,
      '/res/res_pars_list_meze.rds'
    )
  )
}
```