Lumped hydrological models is an Occam' razor for runoff modeling in large Russian Arctic basins

Georgy Ayzel, Institute of Earth and Environmental Science, University of Potsdam, 14476 Potsdam, Germany; Institute of Water Problems of Russian Academy of Sciences, 119333 Moscow, Russia

Abstract This study is aimed to investigate the possibility of three lumped hydrological models to predict daily runoff of large-scale Arctic basins for the modern period (1979-2014) in the case of substantial data scarcity. All models were driven only by meteorological forcing reanalysis dataset without any additional information about landscape, soil or vegetation cover properties of studied basins. We found limitations of model parameters calibration in ungauged basins using global optimization algorithm and confirmed the hypothesis about its equifinality without robustness violation. Model parameters regionalization based on the whole parameters set transfer across studied watersheds was performed and showed good efficiency for prediction in ungauged basins. We run a blind test of the proposed methodology for ensemble runoff predictions on five subbasins which had only periodical monthly observations, and it showed promising results for current freshwater resources assessment for a broad domain of the Russian Arctic. Further work will focus on gridded daily runoff dataset development based on presented findings. The whole research workflow (from data to figures) was performed reproducibly and freely available on Github (github.com/hydrogo/HSJ_article_1).

1 INTRODUCTION

Arctic basins are amongst the most vulnerable systems to climate change impact (Blaen et al., 2014; Flato et al., 2013; Lammers et al., 2001), but at the same time, they are among the least studied geographical objects because of observational data scarcity and overall inaccessibility for long-term direct geoscientific measurements (Shiklomanov et al., 2002; Vorosmarty et al., 2001). Arctic ecosystems and local communities are strongly affected by freshwater budget from rivers (Harms et al., 2000; Peterson et al., 2002; Holmes et al., 2000; Kane, 1997). In recent decade there were many studies devoted to an assessment of freshwater fluxes to the Arctic ocean. Despite the firm scientific basis all of them provide significantly different results (Couet and Mauer, 2009; Wilkinson et al., 2014; Slater et al., 2007), which indicates both high natural variability (uncertainty) of runoff formation processes in the Arctic basins and the lack of our understanding (and a possibility to describe) of their dynamics (Berezovskaya et al., 2004; Lammers et al., 2001; Gelfan et al., 2015). A movement towards annual to daily runoff estimates, from well-observed to ungauged basins, and from global to local research scale increases the uncertainty of hydrological cycle processes and therefore decreases runoff predictability (Bloschl and Sivapalan, 1995; Gelfan et al., 2015).

One of the main problem facing hydrological community in the past decades has been runoff predictions in ungauged basins. Research decade initialized by International Association of Hydrological Sciences (IAHS) in 2003 and devoted to Predictions in Ungauged Basins (PUB) put a wide range of scientific questions are expected to be resolved within ten years (Sivapalan et al., 2003). Despite great research achievements, there is no community consensus about a universal framework to deal with the problem of continuous streamflow simulations in ungauged basins (Hrachowitz et al., 2013; Parajka et al., 2013). The main PUB science focus - a reduction of predictive uncertainty - is still alive across the globe: there are neither only right land cover and meteorological forcing datasets (Nasonova et al., 2011; Essou et al., 2016; Vu et al., 2016), nor true physically-based model working with the same efficiency in different geographical conditions (Arsenault and Brissette, 2016; Goswami et al., 2006; Duan et al., 2006), nor the best regionalization technique for model parameters transfer (He et al., 2011; Razavi and Coulibaly, 2013). Despite these difficulties in hydrological theory development, the most state-of-art technique for research in PUB still is in the coherent framework of the bundle "data - model - regionalization technique - prediction."

1.1 Hydrological models

There are many types of different hydrological models implemented for a variety of scientific and practical purposes, including (i) water resources assessment, (ii) flood forecasting, (iii) runoff calculations, (iv) climate impact and uncertainties assessment, etc. (Beck et al., 2016a). Nevertheless, modern hydrological model development, testing, and further implementation face the same limitations as five decades ago (Smith et al., 2013; Paniconi and Putti, 2015; Hrachowitz et al., 2013). A lot of physically-based (Stromqvist et al., 2012; Gusev and Nasonova, 2006; Maoyi and Liang, 2006; Semenova et al., 2015), conceptual (Winsemius et al., 2009; Arsenault and Brissette, 2014; Razavi and Coulibaly, 2016; Oudin et al., 2008; Merz and Bloschl, 2004), and data-driven models (Yang et al., 2008; Besaw et al., 2010) use for continuous streamflow predictions in ungauged basins, but only some of them take into consideration issues devoted to model parameters equifinality and model robustness.

1.1.3 Equifinality

The equifinality thesis is related to the situation that there are many acceptable model representations we derived from calibration procedure and our inability to easily reject them on a theoretical or observational basis (Beven, 2006). In other words, the equifinality is the case where different model conditions (are determined by model parameters) lead to similar effect (model efficiency) (Ebel and Loague, 2006). To increase the realism of hydrological model predictions in case of model parameters equifinality we need to incorporate in our research more knowledge about hydrological cycle physics of particular basin or more common physically-based constraints (Loague and VanderKwaak, 2004). For ungauged basins in data-scarce regions, the issue about concrete methodology implementation for appropriate model parameters restraint remains controversial (Yadav et al., 2007; Arsenault and Brissette, 2016; Savenije, 2001; van Emmerik et al., 2015).

1.1.4 Robustness

The robustness is another important issue to deal with the assessment of hydrological model ability to predict streamflow in changing conditions (Maier et al., 2016; Thirel et al., 2015b). Many studies report the lack of robustness both of physically-based (Bloschl and Montanari, 2010) and conceptual models (Coron et al., 2014). There are several strategies to deal with model robustness evaluation based on common principles provided by Klemes (1986). Coron et al. (2012, 2014) proposed the generalized split-sample testing procedure which incorporates efficiency metrics were derived from many sub-periods of calibration and validation, and provide graphical and numerical analysis criteria. Thirel et al. (2015a) proposed an evaluation protocol which extends Klemes' differential split-sample testing technique (Klemes, 1986) and provides a wide range of primary and secondary efficiency metrics and graphical analysis instruments. Despite this there is no scientific consensus about considering complex robustness testing framework into research workflow of hydrological modeling - the vast majority of studies use the simplest approach of robustness evaluation based on efficiency criteria estimation on calibration and validation periods and vice-versa (Gusev et al., 2015; Gelfan et al., 2015).

1.2 Data for modelling

There are no clear guidelines for choosing the best landcover, land-use or meteorological forcing dataset for PUB (Hrachowitz et al., 2013). Every modelers' group choose required datasets according to their own experience, model structure, spatial coverage and scale, specific issues of data availability, model intercomparison project conditions, national obligations, etc., and frequently avoid detailed description about the reason why they choose this particular data source, rather than another. Recent studies provide common results of an intercomparison of various meteorological data sources for hydrological models: (i) global reanalyses have good potential to be used as hydrological models forcing, especially in data-scarce regions (Essou et al., 2016a), (ii) despite datasets differences hydrological models can perform equally well after a specific calibration to each dataset (Essou et al., 2016b).

1.3 Regionalization

Regionalization is the main technique for PUB. The term "regionalization" was used in hydrological literature in quite different meanings (He et al., 2011), but today it is typically used for a group of methods aimed at necessary information transfer from gauged to ungauged basins to perform runoff calculations. There are three more utilized regionalization techniques: (i) regression-based, (ii) spatial proximity, and (iii) physical similarity (He at al., 2011; Razavi and Coulibaly, 2013). In their comprehensive review article Razavi and Coulibaly, reported that spatial proximity and physical similarity regionalization approaches had shown satisfactory results in arid to warm temperate climate (e.g., Australia) and cold, snowy regions (e.g., Canada), in contrast, regression-based approaches have been preferred in warm temperate regions (e.g., Europe). Recent PUB-inspired studies include plenty of researched basins across the globe. Gupta (Gupta et al., 2014) systematized 94 papers in hydrology that used more than 30 catchments (median value is 140), and most of them also referred to PUB initiative. We suppose that the use of a large-sample datasets always outperform hydrological model abilities for continuous streamflow predictions because of dense spread of research basins - there is quite easy to find calibrated and robust model parameters set for ungauged basin which is located at a few hundred kilometers distance from donor basin. Performance of regionalization techniques in rarely gauged regions still remains insufficiently investigated.

1.4 Blind test problem

We suppose that the problem of runoff predictions in ungauged basins is rather far-fetched. Most of PUB-related studies verify proposed hypotheses by the methods of leave-one-out jack-knife cross-validation (Arsenault and Brissette, 2014; Oudin et al., 2008) or reserving some part of dataset for independent evaluation (Young, 2006). Thereby every study actually uses information from gauged basins and transform it in artificial manner to "pseudo-gauged" concept. This problem of inability of proper tested hypotheses verification was discussed in a few studies (Efstratiadis et al., 2014; van Emmerik et al., 2015; Rojas-Serna et al., 2016). Goswami (Goswami et al., 2006) proposed for truly ungauged basin to calculate runoff without any verification, Seibert and Vis (Seibert and Vis, 2016) investigated the ability to use stream level observations for hydrological model calibration. In our opinion frameworks proposed by Seibert and McDonnell (Seibert and McDonnell, 2014) that includes both clever fieldwork and modeller experience, and the method proposed by Rojas-Serna (Rojas-Serna et al., 2016) that combines weighted regional data and local information, are the best for PUB-inspired cases.

1.5 Motivation

Despite the efforts exerted to the PUB research, hydrologic community still faces the need to make stronger attempts to provide a valuable assessment of contemporary water resources based on state-of-art techniques (Hrachowitz et al., 2013). In the presented study, we tried to pass standard research workflow (as mentioned above: "data - model - regionalization technique - prediction") for runoff predictions in ungauged basins for the large domain of Russian Arctic rivers Nadym, Pur, Taz, and their nested basins.

According to the Occam's razor principle *“plurality should not be posited without necessity”* in our study we decided to use well-known lumped hydrological models instead of more complex distributed physically-based or land-surfaces models. If a simpler model can describe daily runoff variations well it is the reason we consider it realizes proper hydrological processes schematization and carries a low rate of structural model uncertainty (Butts et al., 2004). Data availability drives our decision to use only WFDEI (WATCH Forcing Data ERA-Interim) meteorological forcing dataset in our study, and this reflects our choice of unconstrained model calibration towards implementation of automatic parameters searching algorithm which requires only information about model parameters bounds. Because of the small dataset using, there is no intercomparison of different regionalization techniques performances - we focused only on simple model parameters transfer across the researched basins (known as "proxy-basin test") and three-fold split-sample test, both proposed by Klemes (Klemes, 1986). We used five nested ungauged basins for blind test case study where only monthly observations were available for runoff predictions verification.

There are three central questions in these study that we would like to discuss: (i) Does high variability of optimal parameters make hydrological model unstable? (or does equifinality lead to instability?) (ii) Does simple transfer of model parameters set from donor to recipient basin work for large Arctic basins? Does it lead to predictions instability (uncertainty)? (iii) Does parameter downscaling from the main to nested watersheds "as is" provide appropriate results for ensemble runoff prediction in ungauged basins?

2 STUDY AREA

The study area consists of three large river basins of the Russian Arctic: the Nadym, the Pur, and the Taz (NPT domain). Five nested river basins used for blind test of runoff predictions in ungauged basins. Figure 1 shows basins locations, streamflow gauges locations, and covering study area.

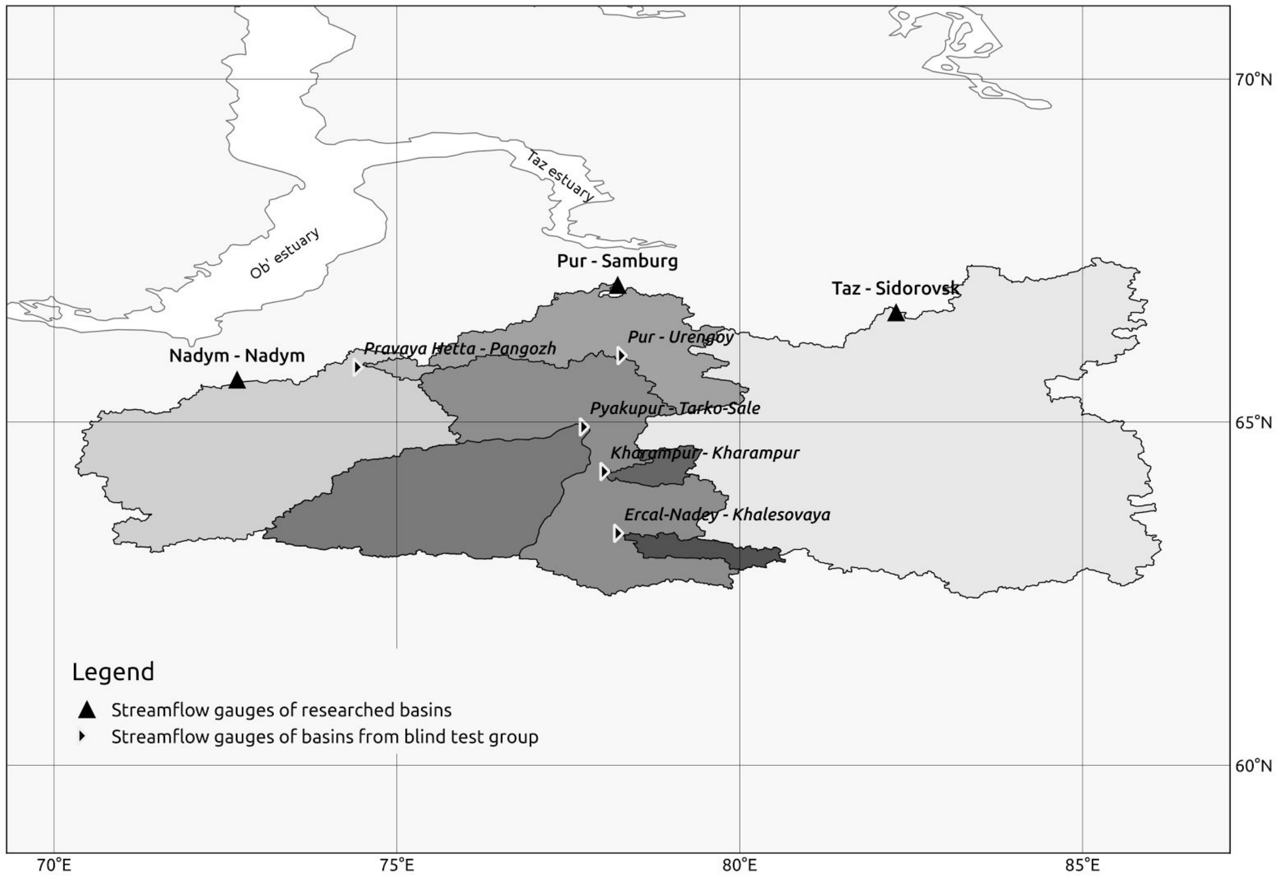

Fig. 1 River basins locations used in this study

NPT domain was and remains the object of high scientific research interest. Recent studies investigated different sides of hydrological cycle features of rivers in NPT domain. Peteet et al. (1998) researched peatland dynamics and climate history for NPT based on a stratigraphic analysis. Zakharova et al. (2011) examined snow cover formation features and estimated proportions of various freshwater sources and its transformation in main rivers of NPT. Karlsson et al. (2014) investigated the behavior of lake size-distribution in permafrost landscape dynamics. Gusev et al. (2015) proposed a robust tool for NPT water balance assessment based on land surface model SWAP. It is worth mentioning that every study noted NPT river basins as a region which has similar natural conditions. Main points of basins description (including annual runoff provided by ArcticRIMS and r-ArcticNET projects (ArcticRIMS, 2016; r-ArcticNET, 2016)) are presented in Table 1.

Table 1. Main characteristics of basins used in this study

#	River	Gauge station	Basin area, km^2	Annual discharge, km^3	Period of runoff observations	Study stage
1	Nadym	Nadym	48000	14.45	1955 - 1991	main
2	Pur	Samburg	95100	28.25	1939 - 1991	main
3	Taz	Sidorovsk	100000	32.99	1962 - 1996	main
4	Pravaya Hetta	Pangozh	1200	0.38	1979 - 1993	blind test
5	Pur	Urengoy	80400	23.99	1961 - 1999	blind test
6	Pyakupur	Tarko-Sale	31400	9.72	1954 - 1999	blind test
7	Kharampur	Kharampur	4330	1.27	1980 - 1985	blind test
8	Ercal-Nadey	Khalesovaya	6600	2.02	1959 - 1995	blind test

3 DATA

3.1. Forcing

Meteorological observation network in Russia is weak, especially in remote Arctic territories. National hydrometeorological data provider (Roshydromet) does not provide continuously updated observational dataset and does not always monitors the quality of the data provided. For the modern studies related to contemporary water resources assessment on vast territories, it is essential to use well-served meteorological datasets, such as global gridded data provided by international scientific collaborations. For this reason, all models were driven by precipitation and temperature data from WFDEI meteorological forcing dataset (1979-2014, 0.5° spatial resolution, Weedon et al., 2014). In the WFDEI framework precipitation data were further enhanced using the monthly Climate Research Unit (CRU) dataset (Harris et al., 2013). Potential evapotranspiration is another required forcing variable for all models, and it was derived based on temperature-based equation proposed by Oudin (Oudin et al., 2005).

3.2. Observed runoff

Daily and monthly observed data were used in this study for a variety of tasks, such as (i) hydrological models parameters calibration, (ii) models efficiency evaluation on different periods, (iii) proxy-basin regionalization approach evaluation, (iv) estimation of ensemble runoff predictions efficiency in ungauged basins. All data were provided by Global Runoff Data Centre (GRDC, 2016).

3.3. Basin schematization and geospatial data

For meteorological forcings averaging procedure and for further runoff modeling results comparison, researched basins were schematized by spatial grid cells of 0.5°×0.5° resolution. In (Boone et al., 2004) it was shown that spatial representation of basin does not significantly affect the result of forcing averaging. Geospatial data used in this study were provided by GRDC (GRDC, 2011).

4 METHODS

Standard PUB workflow determines our research toolkit as a bundle of three lumped conceptual rainfall-runoff models, automatic model parameters calibration procedure, and appropriate regionalization approach. Statistical metrics of model efficiency evaluation are also defined and described.

4.1 The HBV model

The HBV (Hydrologiska Byråns Vattenbalansavdelning) model was used in this study according to its broad implementation for different hydrological applications, flexibility, proven effectiveness for runoff predictions in different geographical conditions, and numerous successful applying for PUB-inspired studies. The HBV model is a typical bucket-type river basin, it has daily time step and simulates river runoff using three meteorological variables as input forcings: temperature, precipitation, and potential evaporation. There are four basic routines are used to represent conceptual water balance processes at basin scale: (i) snow routine, (ii) soil routine, (iii) groundwater routine (iv) routing routine. We slightly modified the HBV model version proposed in (Beck et al., 2016) adding extra parameters related to forcings correction, and also we transformed routing scheme replacing triangular weighting function to Butterworth function. List of all HBV model parameters and its calibration ranges (were based on typical values from previous studies) is shown in Table 2. For the detailed model description, please refer to (Lindstrom, 1997; Seibert, 1999; Seibert and Vis, 2012).

4.2 The GR4J-Cema-Neige model

The GR4J (modele du Genie Rural a 4 parametres Journalier) model is a daily lumped four-parameter rainfall-runoff model (Perrin et al., 2003). It was used in this study according to similar reasons as the HBV: it was widely and successfully implemented in different geographical conditions and for extensively used for case studies related to the PUB initiative. Model routines represent river basin behavior as an interaction between three water storages (for an interception, production, and routing), including interbasin and groundwater exchange, and also incorporate routine scheme based on unit hydrograph methodology. There is no snow accounting module in the original version of the GR4J model. For appropriate using this model in the Arctic basins, we coupled two-parameter snow model Cema-Neige proposed by Valery (Valery, 2010) with the GR4J model in a similar way as in (Coron et al., 2014). List of parameters, its description and calibration ranges for the GR4J-Cema-Neige model is shown in Table 2. For the detailed model description, please refer to (Perrin et al., 2003; Valery, 2010).

4.3 The SIMHYD-Cema-Neige model

The third lumped conceptual hydrological model in our study is SIMHYD (Chiew et al., 2002, 2009). Its conceptual representation is quite similar to the GR4J model despite different approaches to empirically-based water balance processes description and routing schemes distinction. The SIMHYD model also has been used widely for various applications, including regionalization and PUB-related studies (Reichl et al., 2005, 2009; Zhang and Chiew, 2009). As well as for GR4J model, we also incorporated Cema-Neige snow model in SIMHYD simulation procedure. The SIMHYD-Cema-Neige model has in total twelve parameters: eight for main simulation routine; two for routing scheme, and two for snow routine - list of all parameters description and its calibration ranges is shown in Table 2.

Table 2. Hydrological models parameters description and calibration ranges

Parameter	Description	Minimum	Maximum

HBV			
TT (°C)	Threshold temperature when precipitation is simulated as snowfall	-1.5	2.5
SFCF	Snowfall gauge undercatch correction factor	0.4	1.0
CWH	Water holding capacity of snow	0	0.2
CFMAX (mm $°C^{-1}$ d^{-1})	Melt rate of snowpack	1	10
CFR	Refreezing coefficient	0	0.1
FC (mm)	Maximum water storage in the unsaturated zone store	50	500
LP	Threshold of soil moisture value above which actual evaporation reaches potential evaporation	0.3	1.0
BETA	Shape coefficient of recharge function	1	6
UZL (mm)	Threshold parameter for extra outflow from upper zone	0	500
PERC (mm d^{-1})	Maximum percolation to lower zone	0	3
PCORR	Precipitation correction factor	0.5	2
CET	Evaporation correction factor	0	0.3
K0 (d^{-1})	Additional recession coefficient of upper groundwater storage	0.01	0.4
K1 (d^{-1})	Recession coefficient of upper groundwater store	0.01	0.4
K2 (d^{-1})	Recession coefficient of lower groundwater store	0.001	0.15
MAXBAS (d)	Routing scheme parameter	1	7

GR4J-Cema-Neige			
X1 (mm)	Capacity of the production store	0	1500
X2 (mm)	Groundwater exchange coefficient	-10	5
X3 (mm)	Capacity of the nonlinear routing store	1	500
X4 (d)	Unit hydrograph time base	0.5	4
X5	Snowpack thermal state coefficient	0	1
X6 (mm °C^{-1} d^{-1})	Melt rate of snowpack	1	10
SIMHYD-Cema-Neige			
INSC (mm)	Interception store capacity	0	50
COEFF (mm)	Maximum infiltration loss	0	400
SQ	Infiltration loss component	0	10
SMSC (mm)	Soil moisture store capacity	1	1000
SUB	Constant of proportionality in interflow equation	0	1
CRAK	Constant of proportionality in groundwater recharge equation	0	1
K	Baseflow linear recession parameter	0	1
etmul	Potential evaporation correction factor	0.1	3
DELAY (d)	Runoff delay	0.1	5
X_m	Muskinghum transformation parameter	0.01	0.5
X5	Snowpack thermal state coefficient	0	1

X6 (mm $°C^{-1}$ d^{-1})	Melt rate of snowpack	1	10

4.4 Numerical performance metrics for model simulations

The simulated runoff of the models was evaluated with two widely used criteria: the Nash-Sutcliffe (Nash and Sutcliffe, 1970) efficiency (NS, Eq. 1) and the systematic error of runoff estimation (Bias, Eq. 2).

$$NS = \frac{\sum_{\Omega}(X_{sim}-X_{obs})^2}{\sum_{\Omega}(X_{obs}-\overline{X}_{obs})^2}, (1)$$

where X_{sim} and X_{obs} are the simulated and observed values of a variable X (here, runoff), while Ω is a discrete sample set of variable X.

$$Bias = |\frac{\sum_{\Omega}(X_{sim}-X_{obs})}{\sum_{\Omega}X_{obs}}| \times 100\%, (2)$$

Despite recent studies that pointed NS as a flawed metric for model performance evaluation (Jain and Sudheer, 2008), NS remains the most commonly used metric in the field of hydrological modeling. We decided to use Bias as an additional metric that is aimed mostly at base-flow events evaluation. Runoff simulation Bias can be high (in absolute values) even at high NS, that can tell us about some structural discrepancies in model routines of hydrological cycle processes representation (Nasonova et al., 2011; Zhang and Chiew, 2009).

4.5 Calibration procedure

Model parameters (listed in Table 2 for every model) were calibrated by maximizing NS criteria. Differential evolution (DE) algorithm that finds the global minimum of a multivariate function proposed by (Storn and Price, 1997) was implemented. DE algorithm is rarely used in hydrological model calibration, but for known studies, it showed good results (Zhang et al., 2009). DE optimization routine does not use gradient methods, and therefore it can be computationally expensive. For detailed algorithm description, please refer to (Storn and Price, 1997; Price et al., 2006).

For every model, we used three calibration periods of observational runoff data: full period (further referred on figures as “full” or “f”), and two approximately equal half-periods (further referred as “first half” or “h1” and “last half” or “h2”) which are different for every river (Table 3). The use of multiple calibration periods ensures required information for model robustness evaluation and a possibility to use various optimal parameter sets for runoff predictions in ensemble manner.

Table 3. Periods of hydrological models calibration

River	Full period (“f”)	First-half period (“h1”)	Last-half period (“h2”)
Nadym	1979 - 1991	1979 - 1984	1985 - 1991
Pur	1979 - 1991	1979 - 1984	1985 - 1991
Taz	1979 - 1996	1979 - 1987	1988 - 1996

4.6 Interperiod model robustness cross-evaluation

After the calibration procedure, we derived optimal sets of parameters for every model, and for every river basin. The further idea is to verify the efficiency of using every parameter set on every period and river basin (Klemes, 1986; Merz and Bloschl, 2004; Beck et al., 2016b).

In our study, we want to propose to expand the meaning of the "equifinality" and "robustness" terms. Beven and Freer (Beven and Freer, 2001) state that "it may be endemic to mechanistic modeling of complex environmental systems that there are many different model structures and many different parameter sets within a chosen model structure that may be behavioural or acceptable in reproducing the observed behaviour of that system." According to this point of view, every the unique model parameters set represents the unique model structure, which works well on a calibration period ("equifinality"), but it is also necessary (for using equifinality as a working paradigm) to prove model robustness with these different parameters sets. In presented study we propose to use soft threshold criterion - NS > 0.5 - for model robustness verification: if the model with selected optimal parameters set predicts runoff on test periods with efficiency greater than the threshold value - we can talk about robustness confirmation. Otherwise, we can talk about model robustness failure. If there are more than one such optimal parameters set satisfies model robustness condition, we can talk about the situation of "proof of model parameters equifinality without model robustness violation" - and this statement may be used as a strict condition for selection ensemble members for ensemble runoff predictions. A decision about selection of NS threshold value is based on the NS values systemized in (Moriasi et al., 2007) and marked as lower bound for satisfactory modeling results.

4.7 Regionalization approach

Model parameters transfer from gauged to a pseudo-ungauged basin in our study was based on a simple proxy-basin approach which does not use any additional information neither about physiographic characteristics (which is commonly used for physical similarity regionalization approaches), nor basins location information (which is commonly used for spatial proximity regionalization approaches). According to this approach, the entire set of optimal model parameters which were calibrated on the gauged basin is used without any modifications to runoff modeling for ungauged basin (Klemes, 1986; Oudin et al., 2008; Arsenault and Brissette, 2016). According to recent studies (Oudin et al., 2008; Reichl et al., 2009; Singh et al., 2014), the entire optimal model parameters set transfer (based on both physical similarity or spatial proximity indices) is the most successful regionalization approach for PUB.

4.8 Blind test methodology

The main assumption of lumped hydrological model implementation for real-world cases is the uniformity of landscape conditions within researched basins. It is a strong assumption, and it is always in strict conformity with reality, especially for large river basins. We decided to check this hypothesis with the reserved set of nested pseudo-ungauged basins. This blind test routine consists of using optimal parameters sets from the main river basin for runoff predictions in ungauged nested basins ("dumb downscaling"). Since we have three different models and three independent parameters sets for them, we can provide ensemble runoff predictions for every nested ungauged basin. In this study, we do not propose any technique for model outputs averaging and do not search the best solution among calculated runoff realizations. Unfortunately, in this study, we can evaluate the efficiency of proposed ensemble runoff predictions approach only for a few available episodic monthly runoff observations.

4.9 Toolkit for research reproducibility

We think that the main shortcoming issue of the modern hydrological modeling-oriented studies is the lack of research reproducibility. Please, refer to our Github repository (github.com/hydrogo/HSJ_article_1) for more information about the way you can reproduce (replicate) results obtained in the presented study.

5 RESULTS AND DISCUSSION

5.1 Parameters variability, equifinality, robustness

5.1.1 HBV model variability

Sixteen HBV model parameters were obtained for three calibration periods and three main researched rivers (Figure 2). There are four parameters which vary in a more narrow range than calibration range for all three basins: field capacity (FC), recession coefficient of upper groundwater store (K1), routing scheme parameter (MAXBAS), and melt rate of snowpack (CFMAX). A low range of these parameters variation tells us about low uncertainties provided by processes representation which have been parametrized by them and processes proximity to overall NPT domain. The number of parameters which vary in a broad range is higher than relatively stable ones. There are five parameters: shape coefficient of recharge function (BETA), additional recession coefficient of upper groundwater storage (K0), maximum percolation to lower zone (PERC), threshold parameter for extra outflow from the upper zone (UZL), a refreezing coefficient (CFR). Remaining parameters vary in mediate ranges and provide additional uncertainty in runoff predictions.

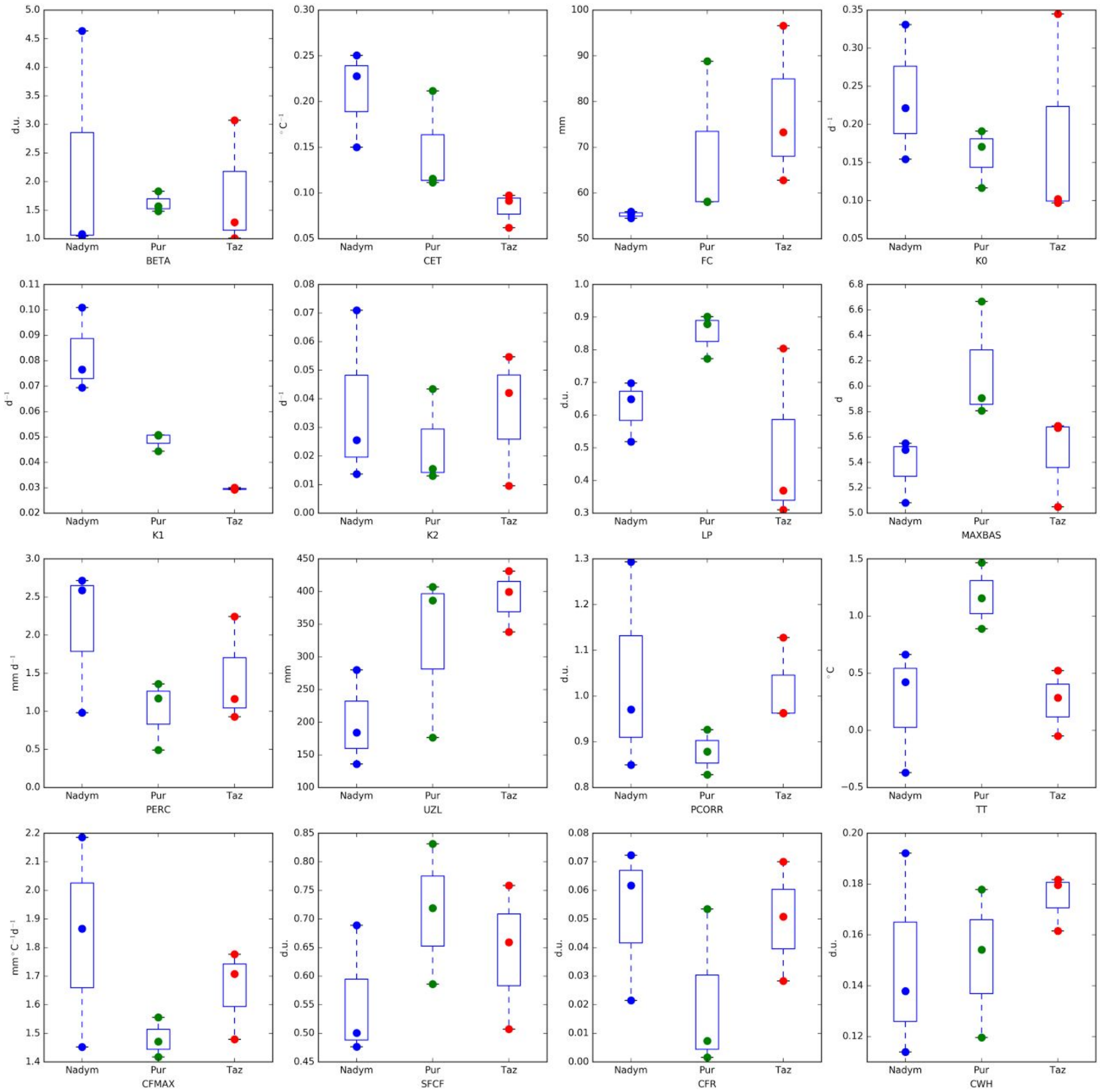

Fig. 2 HBV model parameters variability on different calibration periods

There is no clear picture about similarities between basins: (i) there is no parameter with the same variations range for all basins and (ii) there is no clear pattern of pair-basin similarities - it's impossible to say which basins are more similar to each other than the other one. Almost all HBV model parameters are located on the Figure 2 canvas in cascade manner - this feature can be related with climate continentality rise from west to east for NPT domain, and we can assume that climate forcing characteristics may affect model parameters (and related hydrological cycle processes) both in spatial and temporal manner.

We suppose that high parameters variability is linked with structural limitations provided by lumped HBV model for large basins. Results show that there are not only one best (optimal) model parameters set, but the individual for every calibration period - this point indicates the possibility of many optimal conceptual basin representations which have a similar response to meteorological forcing.

We also suppose that another possible reason of high parameters variability links with parameter identifiability issues provided by uncertainties in initial guessing realized in the optimization procedure. Because of our DE calibration procedure incorporates quasi-random latin hypercube initial guessing we may find (within certain limits, of course) many optimal parameter sets reflecting various conceptual structures of our basins - and, as a result, we may achieve model parameters equifinality. For many procedures with stable initial guessing (e.g., Newtonian optimizers) it is impossible to find many optimal solutions, and this leads to pseudo-equifinality - model parameters vary insignificantly within its uncertainty (error) bounds. We think that model parameters equifinality phenomenon links strongly with the over-parametrization case: a large number of parameters increase the complexity of model parameters surface and may cause computational problems with several objective function minimums availability.

In our study, we follow hypothesis of HBV model parameters stationarity and we based on the assumption that calibration provides the optimal lumped structure of researched basin that can be defined by optimal parameter set and will not change in time. This assumption is in substantial conflict with a hypothesis of HBV parameters non-stationarity described and proved in (Merz et al., 2011). Osuh et al. (2015) provided contradictory results of investigation how model parameters relate to climatic indices, and we think that non-stationarity successful hypothesis testing is in more relation with study area and researched basins characteristics than with possible changing of climate conditions.

Using additional information about basin physiographic and landscape characteristics can provide useful decision rules for model parameters constraining, but for ungauged basins, it can be difficult to obtain any relevant information. We think that using global databases of vegetation and soil properties is a perspective way for hydrological model parameters constraining and therefore for the more physically-based calibration procedure. The problem of model parameters sets stability will be discussed later in Sect. 5.1.4.

5.1.2 GR4J-Cema-Neige parameters variability

In contrast to HBV model, the GR4J-Cema-Neige model has only six parameters. Five of them vary in a wide range, comparable to the calibration range and only one parameter - unit hydrograph time base (X4) - varies in quite narrow range (Figure 3). These results can be explained by the same reasons we supposed for the HBV model parameters variability: basin approximation by the lumped structure leads to high uncertainty (variability) of optimal parameters combination, primarily if we do not use any physically-based constraints for the calibration procedure. For the routing parameter (X4) basin area plays the leading role and regulates its variations in a narrow range.

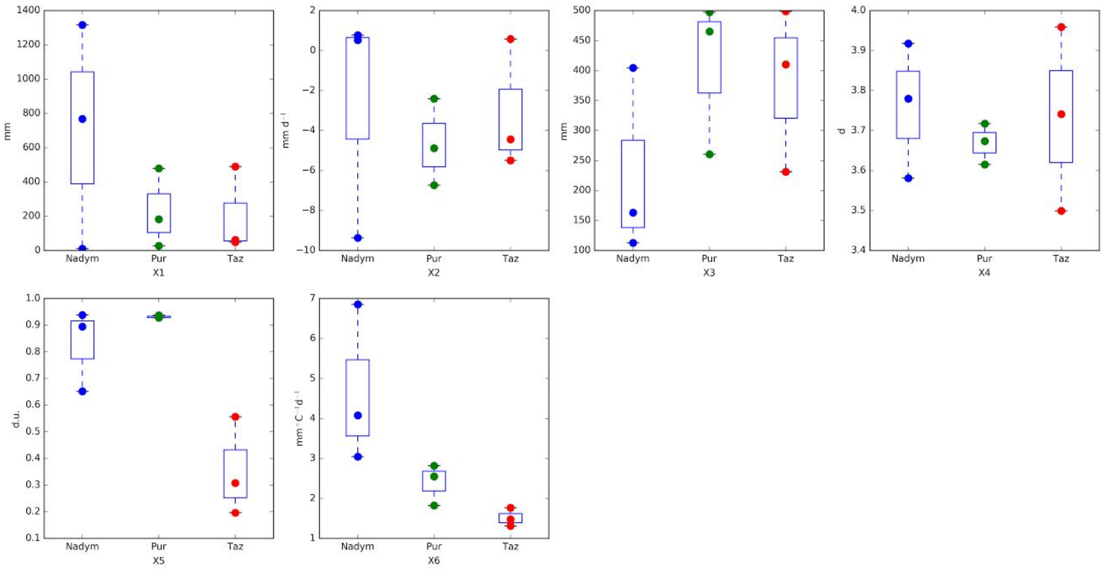

Fig. 3 GR4J-Cema-Neige model parameters variability on different calibration periods

Figure 3 shows the similarity between the Pur and the Taz basins - for most of parameters allocation we can mark the pair of these basins. According to this notice, we can suppose that under-parameterized model can approximate similar runoff formation complexes into the identical model structures. These structures may be very different from the real basins, and it is necessary to evaluate its robustness (will be discussed later in Sect. 5.1.4). For the some HBV model parameters, we noted its cascade look (on Figure 2 canvas). For the GR4J-Cema-Neige model, we can note parabolic look (on Figure 3 canvas) for the most model parameters allocation for different rivers, which strongly corresponds with a hypothesis about similarities between the Pur and Taz rivers and their high contrast from the Nadym river.

The HBV and the GR4J-Cema-Neige models have three pairs of quite similar (but not identical) parameters (i) melt rate of snowpack (CFMAX and X6), (ii) maximum water storage in the unsaturated zone store (FC) and capacity of the production store (X1), (iii) routing parameters (MAXBAS and X4). Intercomparison of these parameters pairs shows a few discrepancies in hydrological processes description provided by models we use. For example, for HBV model, melt rate of snowpack varies (CFMAX) in a quite narrow range and has concave parabolic distribution across basins, but the same parameter of GR4J model (X6) varies in three times wider range and has falling cascade distribution. Thus we see that the different snow routines which have "conceptually true" processes schematization are in sharp conflict with a parameter range and its spatial distribution. The parameters that determine the features of soil moisture routine for the HBV and GR4J-Cema-Neige - FC and X1 - also have mirrored (rising and falling, respectively) cascade distribution across basins. Routing scheme parameters for both models - MAXBAS and X4 - vary in quite narrow ranges, but have mirrored parabolic (convex and concave, respectively) distribution across basins.

We can not find any information about temporal (induced by climate change or uncertainties in hydrological processes representation) model parameters variability in most of the studies related to GR4J model implementation. All of these studies focused on an evaluation of different efficiency metrics in connection with varying factors of impact and did not provide extensive information about every model parameter uncertainty.

We suppose that the reasons of observed high model parameters variability are the same and for GR4J-Cema-Neige model: (i) rough approximation of basin structure by the lumped model, (ii) uncertainties and high complexity of global optimum parameters set search. The effect of model parameters equifinality will be discussed later in Sect. 5.1.4.

5.1.3 SIMHYD-Cema-Neige parameters variability

The SIMHYD-Cema-Neige model takes an intermediate position (between HBV and GR4J-Cema-Neige) according to the number of parameters - there are 12 parameters. Most of them vary in broad ranges, comparable to calibration ones (Figure 4). Only for baseflow linear recession parameter (K), we have an entirely narrow range of variability, that can be related to conservative nature of groundwater recharge in NPT domain.

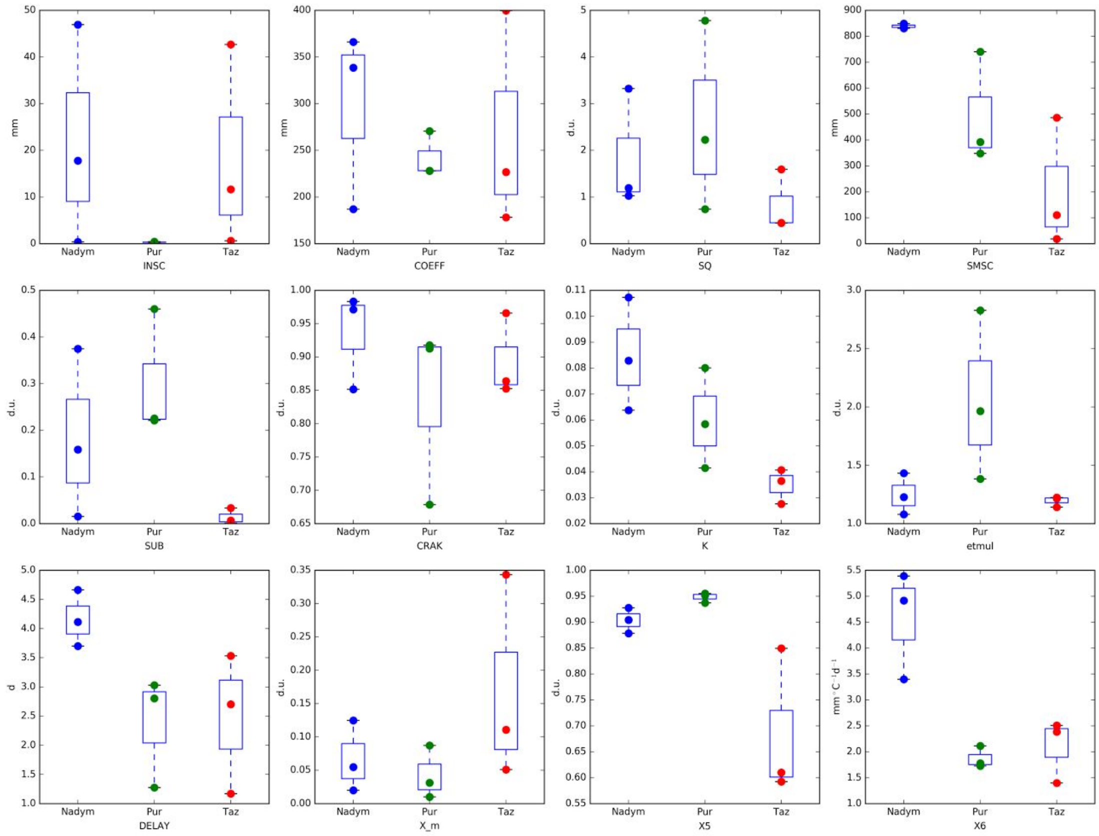

Fig. 4 SIMHYD-Cema-Neige model parameters variability on different calibration periods

There are no apparent similarities between researched basins - every parameter variation has its pattern for every basin. We also need to note that parameters distributions on (Figure 3) canvas tend to be more blurred - cascade and parabolic looks of distributions are not strict because of parameters ranges overlay.

The parameters distribution pattern for GR4J and SIMHYD-based models are similar, while HBV model parameters patterns have a different (mirrored) look. The parameters X5 and X6, which refer to Cema-Neige snow routine, and that are the same for the both GR4J-Cema-Neige and SIMHYD-Cema-Neige models, are in full agreement for both models. We suppose that this point tells us about the relative stability of Cema-Neige snow model and its low impact on structural uncertainty provided by the models coupled with it.

There are a couple of studies which examined parameters variability of the SIMHYD hydrological model. Reichl (Reichl et al., 2005) conducted an experiment based on Monte-Carlo simulations run on 44 Australian catchments. Results showed that equifinality of SIMHYD model parameters amounts to considerable parameter uncertainty and this is the primary reason for regression-based regionalization approaches failure. Li (Li et al., 2015) linked high temporal variability of SIMHYD parameters with forcing changes between different calibration periods and with the inadequate model presentation. In this study also noted that landscape basin characteristics also define parameters stability - for humid basin parameters variability is lower than for arid ones.

We suppose that the reasons of observed high model parameters variability are the same and for SIMHYD-Cema-Neige model (as for the HBV and the GR4J-Cema-Neige models): (i) rough approximation of basin structure by the lumped model, (ii) uncertainties and high complexity of global optimum parameters set search. We believe that reducing of model parameters variability can only be achieved by (i) the use of new information for model parameters constraining on a physical basis, and (ii) the development of more physically-based, process-oriented models with a low number of calibrated parameters.

In the next section, we will show how model parameters variability affected model robustness (in the case of runoff predictions for different time periods).

5.1.4 Inter-period cross-evaluation

The procedure of model robustness evaluation was described in Sect. 4.6. For results representation, we decided to use heat maps of efficiency criteria for different evaluation/calibration periods as provided in (Thirel et al., 2015a).

Obtained results have a clear spatial pattern - when moving east modeling efficiency (NS) increases and Bias decreases - the best daily runoff predictions were obtained for the Taz River, the worst - for the Nadym river. We suppose that this point relates to climate continentality growth in the east direction that determines the low inter-seasonal variability of hydrological processes intensity and, as a result, forms stronger predictability of runoff simulations.

For the Nadym basin, all three lumped hydrological models have a similar temporal pattern of their runoff prediction criteria (Figure 5). For model parameters set calibrated on the full ("full set") period of observations, the best efficiency reaches on the first half of evaluation period and slightly decreases for the last half. It indicates some changes in climate, basin behavior or streamflow measurements quality between selected half-periods. For the parameters sets calibrated on the first ("first set") and the last ("last set") periods slight performance decreasing on independent half-periods is a common situation. All the models with the "last set" and "first set," except GR4J-Cema-Neige, show satisfactory results on the full period. Only one parameter set of all three models - "last set" for GR4J-Cema-Neige - can be marked as non-robust, because it performs well only during the calibration period and fails on the validation ones. Visual patterns of NS and Bias almost identical, the only exception is Bias values pattern of "last set" for the SIMHYD-Cema-Neige model, in this case, Bias on the evaluation periods is smaller than on calibration period. The HBV model performs better calculations than both GR4J-Cema-Neige and SIMHYD-Cema-Neige on every corresponding period with every optimal parameter set. GR4J and SIMHYD -based models, show comparable results.

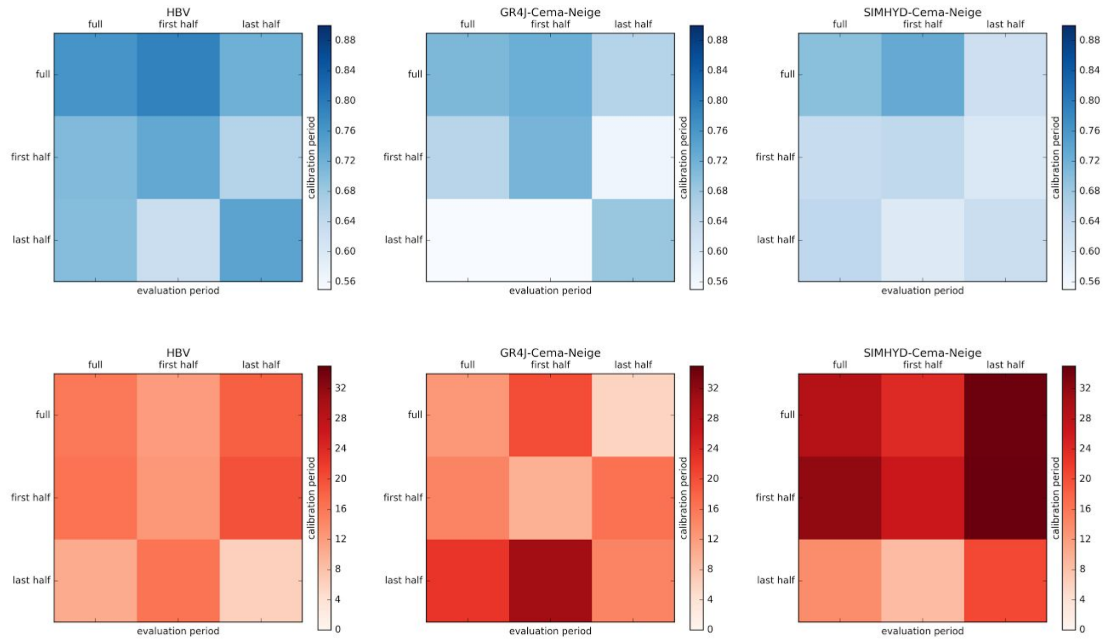

Fig. 5 NS (blue palette) and Bias (red palette) criteria for model evaluation for the Nadym river. In the rows - criteria values for different calibration periods which marked as columns ticks

For the Pur basin, temporal patterns of runoff prediction efficiency are individual for every model (Figure 6). For HBV model for every optimal parameter set the best efficiency reaches for the first half of evaluation period, i.e., not only "full set" works better on the first half (a result that we derived for the Nadym basin for all the models). The HBV model for the Taz basin also tends to be fairly robust one - loss of NS value across validation periods is small, and the model Bias is moderate. For the GR4J-Cema-Neige we can note a standard pattern of decreasing in efficiency for half-periods parameters sets during its validation on independent half-periods. For evaluation on the full period, all sets show satisfactory results. The NS pattern of GR4J-Cema-Neige model is not similar to the Bias pattern - for this model we may notice a weak positive correlation between NS and Bias. For the SIMHYD-Cema-Neige we can note two clear patterns: (i) for every optimal parameter set the best efficiency reaches on the first half of evaluation period (similar result with the HBV model), (ii) "last set" parameters provides a better result than "full set" parameters. The SIMHYD-Cema-Neige Bias values are extremely high for all subperiods and all optimal parameters sets. It is caused by unsatisfactory model simulations for the long period of low flows. The lowest Bias could be derived with "first half" set for every evaluation period. For the Pur basin the HBV model also shows the best results, the SIMHYD-Cema-Neige model works worst (especially because highest Bias), the GR4J-Cema-Neige model takes an intermediate position.

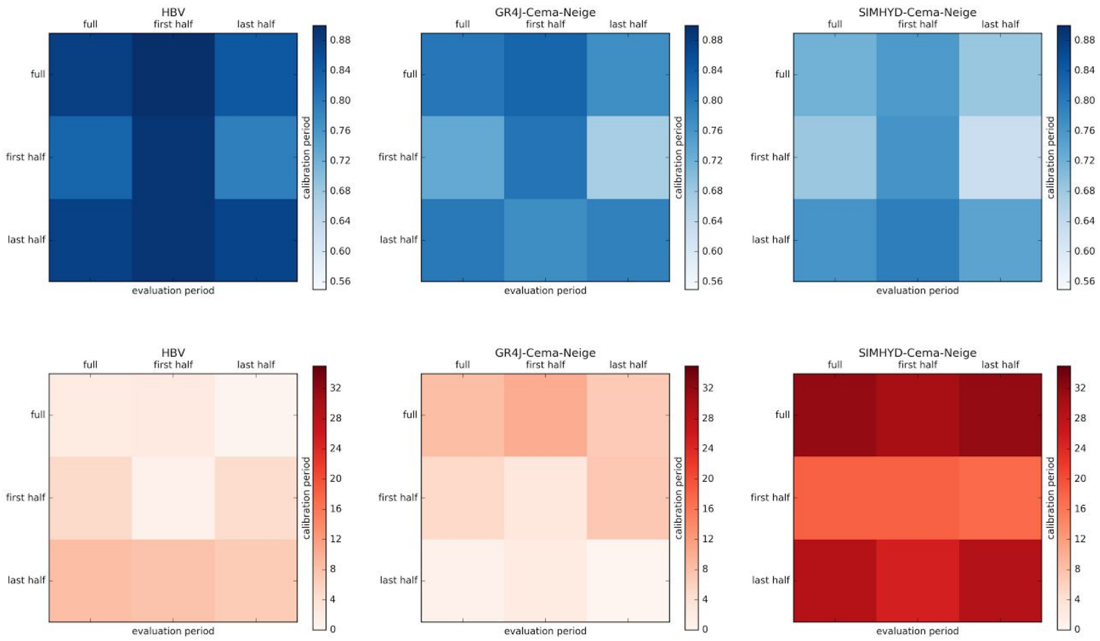

Fig. 6 NS (blue palette) and Bias (red palette) criteria for model evaluation for the Pur river. Designations are the same as in Figure 5

For the Taz basin, efficiency criteria patterns for all the models differ considerably (Figure 7). For all the models average NS efficiency is higher, and Bias is lower for the Taz basin than for the Nadym and the Taz basins. The HBV model also shows highly robust results of inter-period cross-evaluation - only the "first set" parameters show apparent efficiency decreasing on validation periods, the HBV Bias pattern also represents this trait. The GR4J-Cema-Neige model has an interesting cross-shape pattern for both NS and Bias criteria, which shows us mediocre model robustness for using with the "first set" parameters. The SIMHYD-Cema-Neige model has NS pattern, which is similar to the HBV model, but the differences in Bias patterns are apparent: we note slight increasing in Bias values on first-half validation period for all of optimal parameters sets. The SIMHYD-Cema-Neige model only for the Taz river performs better than the GR4J-Cema-Neige model, but the HBV model remains the best among them.

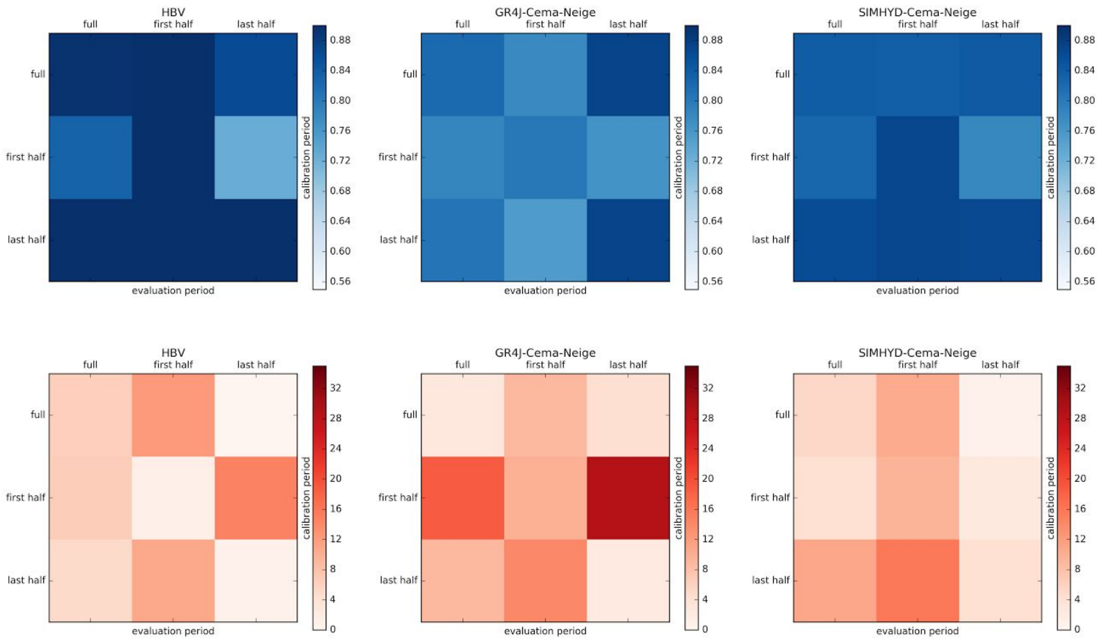

Fig. 7 NS (blue palette) and Bias (red palette) criteria for model evaluation for the Taz river. Designations are the same as in Figures 5, 6

The main result of the proposed inter-period cross-evaluation procedure is the evidence of parameters equifinality of various lumped conceptual hydrological models in a broad sense with the preservation of model robustness. This result shows that there are multiple structures of our researched basins which can represent the complexity of runoff formation processes. We can not say which structure is the best for our purposes until we receive additional information about landscape or climatic characteristics and, on this basis, develop constraining conditions for our model parameters. We suppose that it is normal to get high model parameters uncertainty in our study - we use one of the most straightforward techniques for continuous streamflow prediction for the three large Arctic basins, and we can be ready for difficulties if we want to assess contemporary water resources conditions for our domain. Without any additional information and strict parameters constraints we can use multiple optimal parameters sets for runoff predictions in ungauged basins in ensemble manner - if we can not decide what is the best parameters set, why don't we use all of them? We suppose that it is a good strategy for decision-making in ungauged basins - provide not only single realization, but a whole band of robust realizations (Arsenault and Brissette, 2014; Beck et al., 2016b).

Another significant result is the overall high efficiency of simulated daily runoff predictions with the use of optimal parameters. It was a difficult question at the beginning of our study - even if we find optimal parameters (calibration procedure was successful), can they then be used for predictions? And despite a high rate of individual model parameters variability (Sect. 5.1.1-5.1.3), the whole sets of parameters are stable and allow the model to satisfy the soft robustness criteria. These conceptual models have not been previously used for continuous runoff simulations for this large domain in the Arctic, and verification of its possibility to simulate runoff only with meteorological reanalysis data, which was obtained in our study, itself is of great importance for further water resources assessment for the NPT domain.

Obtained results show that the HBV model is the best choice for continuous runoff predictions in a whole NPT domain. The GR4J and the SIMHYD models had not been initially designed for simulation in snow-dominated regions, unlike the HBV. The Cema-Neige snow model coupling with the GR4J and the SIMHYD seriously improves simulation quality but overparameterized HBV routine is slightly more in line with cold region hydrological cycle processes conceptualization. Li (Li et al., 2015) obtained results that are consistent with our findings for the comparison of HBV, SIMHYD, and XAJ models on two basins in China and Australia.

It should be noted (for all models) that high NS values do not lead to low Bias values and there is another confirmation of necessity to use multiple evaluation criteria in hydrological modeling studies (Moriasi et al., 2007). We also could not find the strong reason why we received the worst results for Nadym river: it may relate to specific geographical features, differences in hydrological regime, poor quality of streamflow data, etc. According to previous studies (Zakharova et al., 2011; Gusev et al., 2015) and our findings, we suppose that inconsistency in measured at gauge station and real streamflow data is the main reason of mediocre results.

At the analysis stage of our research, it became apparent that splitting observational period into a single pair of calibration and validation period (according to Klemes, 1986) is not the best idea. We showed in this section that there are apparent changes in basin configuration between the first and the last periods of model verification, but we cannot correctly decide what kind of natural conditions may influence this difference: changes in climate features or, for example, changes in the riverbed near streamflow gauge station. We suppose that hypothesis of model parameters non-stationarity (Merz et al., 2011) can help us to conceptualize our research basins as evolving systems, which is strictly defined by its parameters that depend on the climate, landscape, and anthropogenic factors. If we accept this hypothesis of model parameters non-stationarity, we need to develop an entirely different framework for model parameters calibration and further robustness estimation which will be able to aggregate strict model parameters constraints (for varying in physically-based boundaries defined by a wide range of physiographic parameters, smoothness of temporal changes, etc.).

5.2 Proxy-basin regionalization for runoff predictions in (pseudo)ungauged basins

Proxy-basin test specification was described in Sect. 4.7. For results representation, we used the same conception of heat maps (as provided in (Thirel et al., 2015a)) for efficiency criteria grouped for every model for different evaluation/calibration periods. Figures 8 shows results for all model parameters regionalization across the basins in NPT domain.

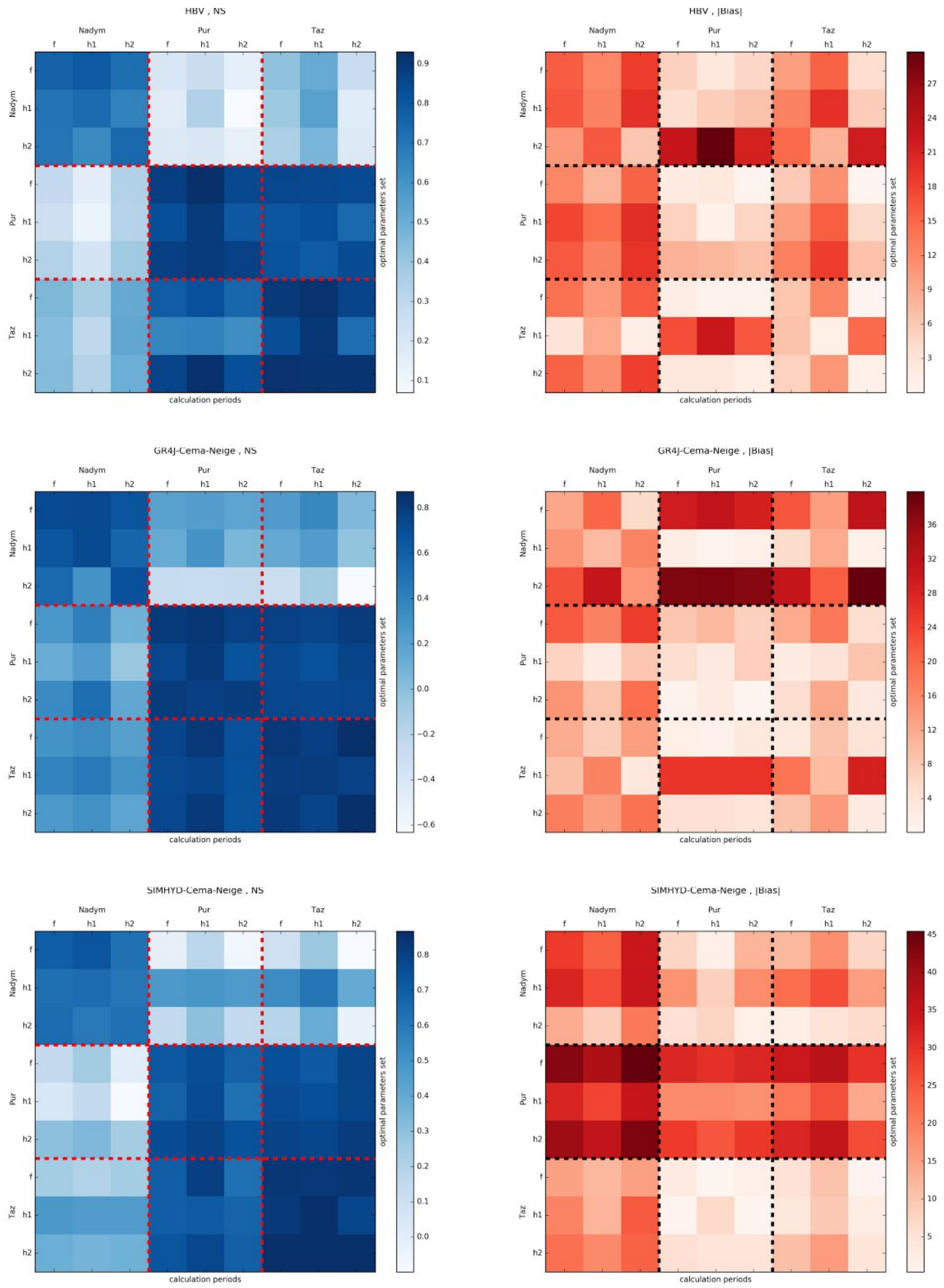

Fig 8. Proxy-basin regionalization results for HBV, GR4J-Cema-Neige, SIMHYD-Cema-Neige models

For all models, we can note similar visual pattern for both NS and Bias - high parameters transferability efficiency for the basin pair of the Pur and the Taz rivers. Optimal model parameters sets which were calibrated on different time periods for the Pur river work well and for each period for the Taz river and vice versa. It also should be noted that the Nadym has non-transferable (across NPT domain) parameters - they work well only for the Nadym basin itself, but runoff simulations with them for the Pur and the Taz rivers show unsatisfactory results. This pattern repeats with the insignificant variations for every model used in this study.

This result proves an effect of parameters equifinality for regionalization efficiency. We suppose that leading roles in proxy-basin regionalization success play: (i) similarity between basins, (ii) scale effects of this similarity, and (iii) internal optimal parameters stability and robustness. Results show that if basins pair ("gauged" - "ungauged") has not only similar conditions but also has comparable drainage areas and stable optimal parameters sets, then it is possible to predict runoff in (pseudo)ungauged basin well.

Obtained insights of model parameters transfer capabilities across NPT domain should be independently tested on very different nested ungauged basins set for proper validation of these assumptions in-depth (inter-basin hypothesis validation).

5.3 Ensemble runoff predictions for truly ungauged basins (blind test results)

Features of cross-NPT model parameters proxy-basin regionalization are also challenges for predictions in ungauged basins. We need to know how stable and robust structure of calibrated parameters sets could be transferred to small nested basins - will we lose prediction efficiency because of scale effects which could determine discrepancies in basin structure representation?

In the presented study, we follow the most accessible approach to predictions in ungauged basins called "dumb downscaling" (please, refer to Section 4.8) in ensemble manner, and also we do not provide any likelihood estimations for our ensemble realizations or some complex averaging strategy. Results of continuous runoff streamflow simulations (monthly averaged hydrographs) for five independent ungauged basins for a long period (1980 - 2014) are presented in Figure 9 with short periods of observed runoff values.

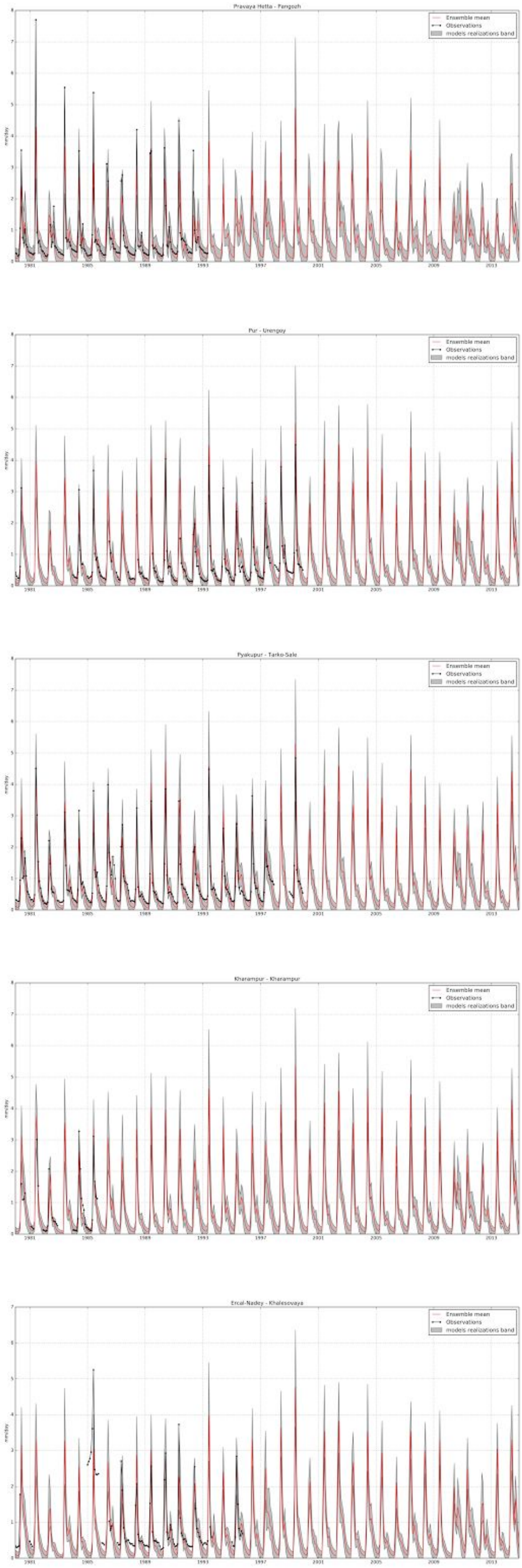

Fig. 9 Ensemble monthly runoff predictions for ungauged basins

For every runoff prediction realization, we calculated NS and Bias evaluation criteria and found the best realization regarding highest NS and lowest Bias values (Table 4).

Table 4. Evaluation criteria statistics for ungauged basins

Ungauged basin	NS			Bias			Highest NS realization	Lowest Bias realization
	Mean	Min	Max	Mean	Min	Max		
Pravaya Hetta	0.72	0.65	0.79	20.5	2.0	41	GR4J, “h1”	GR4J, “f”
Pur	0.80	0.64	0.88	13.2	2.9	23.5	HBV, “h2”	SIMHYD, “h1”
Pyakupur	0.78	0.61	0.87	12.7	1.5	34.2	HBV, “h2”	HBV, “f”
Kharampur	0.74	0.64	0.78	12.0	0.1	31.3	Ensemble mean	HBV, “h1”
Ercal-Nadey	0.23	0.10	0.31	36.7	23.7	52.0	HBV, “h2”	GR4J, “h1”

We suppose that for our weak assumptions proposed in experiment design stage - "dumb downscaling" and ensemble approach - adequate efficiency measure for runoff predictions in ungauged basins is simple entering of prediction values in ensemble realizations band. Visual hydrographs evaluation shows appropriate fit between the predicted band and observed values. Results presented in Table 4, in addition to visual assessment, also confirm the satisfactory efficiency of proposed prediction methodology for all ungauged basins except the Ercal-Nadey river - lower bound of NS criterion for these rivers is higher than 0.61. A common negative feature of obtained results is low capabilities of models used in this study to calculate baseflow for the autumn-winter period - every model underestimates base flow for this period, and it leads to high Bias criterion values. It is impossible to understand specific inconsistencies between predicted and observed runoff for the Ercal-Nadey river without additional information from streamflow gauging network.

6 CONCLUSIONS

Three different lumped conceptual hydrological models were used for runoff predictions in NPT domain for three main and nested basins. The principal findings of the presented work are: (i) global optimization procedure provides very different, but robust model parameter sets which could be used in ensemble runoff predictions, (ii) simple model parameters transfer from donor to ungauged basins works well for runoff predictions in ungauged basins only if the basins are quite similar, (iii) additional information about physical characteristics of researched basins may provide significant improvement of model realism through a reasonable limitation of model parameters bounds and the way for implementing standard similarity metrics for more reasonable model parameters transfer procedure, (iv) hypothesis about model parameters similarity between main and nested basins shows good results for ensemble monthly runoff predictions in ungauged basins in the case of blind test methodology.

We found significant differences between the Nadym basin in one hand and the Pur and the Taz basins in another hand which reflects in hydrological model efficiency and success of proxy-basin regionalization procedure. Obtained results also show an insignificant role of structural differences between models, and prove an effect of parameters equifinality for regionalization efficiency.

Finally, we would like to cite Lem's discourse about mathematical models:

"*Imagine a mad tailor who makes all possible clothes. He knows nothing about people, birds or plants. The world does not interest him; he does not study it. He makes clothes. He does not know for whom. He does not think about it… Mathematics works just like that. It builds structures, but it is not known of what. Perfect, accurate models, but of what - the mathematician does not know. It does not interest him. He does what he does because has turned out possible*" (Lem, 2013).

We think that presented work is vulnerable to such criticism related to models unrealism and lack of physically-based procedures or analyses, but we also sure that obtained results have to help us to find the right way for further improving this field. Future work will focus on creation, validation, and communication of gridded daily runoff dataset for NPT domain.

7 ACKNOWLEDGEMENTS

8 FUNDING

9 REFERENCES

1. Blaen, Phillip J, David M Hannah, Lee E Brown, and Alexander M Milner. 2014. Water source dynamics of high Arctic river basins. *Hydrological Processes* 28, no. 10: 3521-3538.
2. Harms, IH, MJ Karcher, and D Dethleff. 2000. Modelling Siberian river runoff—implications for contaminant transport in the Arctic Ocean. *Journal of Marine Systems* 27, no. 1: 95-115.
3. Peterson, Bruce J, Robert M Holmes, James W McClelland, et al. 2002. Increasing river discharge to the Arctic Ocean. *science* 298, no. 5601: 2171-2173.
4. Flato, Gregory, Jochem Marotzke, Babatunde Abiodun, et al. 2013. Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2013* 5: 741-866.
5. Holmes, RM, BJ Peterson, VV Gordeev, et al. 2000. Flux of nutrients from Russian rivers to the Arctic Ocean: Can we establish a baseline against which to judge future changes?. *Water Resources Research* 36, no. 8: 2309-2320.
6. Butts, Michael B, Jeffrey T Payne, Michael Kristensen, and Henrik Madsen. 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* 298, no. 1: 242-266.
7. Global Runoff Data Centre (2016): River Discharge Time Series / Global Runoff Data Centre. Koblenz, Federal Institute of Hydrology (BfG).
8. Global Runoff Data Centre (2011): Watershed Boundaries of GRDC Stations / Global Runoff Data Centre. Koblenz, Federal Institute of Hydrology (BfG).
9. Zakharova, EA, AV Kouraev, S Biancamaria, et al. 2011. Snow cover and spring flood flow in the northern part of Western Siberia (the Poluy, Nadym, Pur, and Taz rivers). *Journal of Hydrometeorology* 12, no. 6: 1498-1511.
10. Gusev, EM, ON Nasonova, L Ya Dzhogan, and GV Ayzel. 2015. Simulating the formation of river runoff and snow cover in the northern West Siberia. *Water Resources* 42, no. 4: 460-467.

11. Karlsson, Johanna Mård, Steve W Lyon, and Georgia Destouni. 2014. Temporal behavior of lake size-distribution in a thawing permafrost landscape in northwestern Siberia. *Remote sensing* 6, no. 1: 621-636.

12. Peteet, Dorothy, Andrei Andreev, William Bardeen, and Francesca Mistretta. 1998. Long-term Arctic peatland dynamics, vegetation and climate history of the Pur-Taz region, western Siberia. *Boreas* 27, no. 2: 115-126.

13. Shiklomanov, AI, RB Lammers, and Ch J Vörösmarty. 2002. Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Transactions American Geophysical Union* 83, no. 2: 13-17.

14. Lammers, Richard B, Alexander I Shiklomanov, Charles J Vörösmarty, Balázs M Fekete, and Bruce J Peterson. 2001. Assessment of contemporary Arctic river runoff based on observational discharge records. *Journal of Geophysical Research: atmospheres* 106, no. D4: 3321-3334.

15. Berezovskaya, Svetlana, Daqing Yang, and Douglas L Kane. 2004. Compatibility analysis of precipitation and runoff trends over the large Siberian watersheds. *Geophysical Research Letters* 31, no. 21.

16. Vorosmarty, CJ, LD Hinzman, BJ Peterson, et al. 2001. The Hydrologic Cycle and Its Role in the Arctic and Global Environmental Change: A Rationale and Strategy for Synthesis Study, 84 pp. *Arct. Res. Consortium of the US, Fairbanks, Alaska*.

17. Kane, Douglas L. 1997. The impact of hydrologic perturbations on arctic ecosystems induced by climate change. *Global change and arctic terrestrial ecosystems*: 63-81.

18. Couet, Thomas de and Thomas Maurer (2009):Surface Freshwater Fluxes into the World Oceans / Thomas de Couet; Thomas Maurer. Global Runoff Data Centre. Koblenz: Federal Institute of Hydrology (BfG), 2009.

19. Wilkinson, K., Zabern, M. von and J. Scherzer: Global Freshwater Fluxes into the World Oceans : Technical Report prepared for the GRDC. - Koblenz, Federal Institute of Hydrology (BfG), 2014. - (GRDC Report; 44) . – DOI: 10.5675/GRDC_Report_44

20. Gelfan, A et al. "Large-basin hydrological response to climate model outputs: uncertainty caused by internal atmospheric variability." *Hydrology and Earth System Sciences Discussions* 19.6 (2015): 2737-2754.

21. Blöschl, Günter, and M Sivapalan. "Scale issues in hydrological modelling: a review." *Hydrological processes* 9.3-4 (1995): 251-290.

22. Sivapalan, Murugesu, K Takeuchi, SW Franks, et al. 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological sciences journal* 48, no. 6: 857-880.

23. Hrachowitz, M, HHG Savenije, G Blöschl, et al. 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological sciences journal* 58, no. 6: 1198-1255.

24. Parajka, J, A Viglione, M Rogger, et al. 2013. Comparative assessment of predictions in ungauged basins–Part 1: Runoff-hydrograph studies. *Hydrology and Earth System Sciences* 17, no. 5: 1783-1795.

25. Essou, Gilles RC, Florent Sabarly, Philippe Lucas-Picher, François Brissette, and Annie Poulin. 2016. Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling?. *Journal of Hydrometeorology*, no. 2016a.

26. Nasonova, ON, Ye M Gusev, and Ye E Kovalev. 2011. Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components. *Hydrological Processes* 25, no. 7: 1074-1090.

27. Arsenault, Richard, and François Brissette. 2016. Multi-model averaging for continuous streamflow prediction in ungauged basins. *Hydrological Sciences Journal*, no. just-accepted.

28. Goswami, M, KM O'connor, and KP Bhattarai. 2007. Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. *Journal of Hydrology* 333, no. 2: 517-531.

29. Duan, Q, J Schaake, V Andreassian, et al. 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology* 320, no. 1: 3-17.

30. Razavi, Tara, and Paulin Coulibaly. 2013. Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of Hydrologic Engineering* 18, no. 8: 958-975.

31. Beck, Hylke E, Albert IJM van Dijk, Ad de Roo, et al. 2016a. Global evaluation of runoff from ten state-of-the-art hydrological models.

32. Beck, Hylke E, Albert IJM van Dijk, Ad de Roo, et al. 2016b. Global-scale regionalization of hydrologic model parameters. *Water Resources Research*.

33. Paniconi, Claudio, and Mario Putti. 2015. Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research* 51, no. 9: 7090-7129.

34. Gusev, YEUGENIY M, and OLGA N Nasonova. 2006. Simulating runoff from MOPEX experimental river basins using the SWAP land surface model and different parameter estimation techniques. *IAHS PUBLICATION* 307: 188.

35. Strömqvist, Johan, Berit Arheimer, Joel Dahné, Chantal Donnelly, and Göran Lindström. 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. *Hydrological Sciences Journal* 57, no. 2: 229-247.

36. Huang, Maoyi, and Xu Liang. 2006. On the assessment of the impact of reducing parameters and identification of parameter uncertainties for a hydrologic model with applications to ungauged basins. *Journal of Hydrology* 320, no. 1: 37-61.

37. Semenova, O, L Lebedeva, N Volkova, et al. 2015. Detecting immediate wildfire impact on runoff in a poorly-gauged mountainous permafrost basin. *Hydrological Sciences Journal* 60, no. 7-8: 1225-1241.

38. Winsemius, HC, Bettina Schaefli, A Montanari, and HHG Savenije. 2009. On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. *Water Resources Research* 45, no. 12.

39. Arsenault, Richard, and François P Brissette. 2014. Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research* 50, no. 7: 6135-6153.

40. Razavi, Tara, and Paulin Coulibaly. 2016. Improving Streamflow Estimation in Ungauged Basins using Multi-Modelling Approach. *Hydrological Sciences Journal*, no. just-accepted.

41. Oudin, Ludovic, Vazken Andréassian, Charles Perrin, Claude Michel, and Nicolas Le Moine. 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research* 44, no. 3.

42. Merz, Ralf, and Günter Blöschl. 2004. Regionalisation of catchment model parameters. *Journal of Hydrology* 287, no. 1: 95-123.

43. Yang, Ming-Der, Boris PT Chen, and Chang-Shian Chen. 2008. Using artificial neural network for outflow estimation in an ungauged area. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, June 1.

44. Besaw, Lance E, Donna M Rizzo, Paul R Bierman, and William R Hackett. 2010. Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology* 386, no. 1: 27-37.

45. Beven, Keith. 2006. A manifesto for the equifinality thesis. *Journal of hydrology* 320, no. 1: 18-36.

46. Ebel, Brian A, and Keith Loague. 2006. Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. *Hydrological Processes* 20, no. 13: 2887-2900.

47. Yadav, Maitreya, Thorsten Wagener, and Hoshin Gupta. 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources* 30, no. 8: 1756-1774.

48. Savenije, Hubert HG. 2001. Equifinality, a blessing in disguise?. *Hydrological processes* 15, no. 14: 2835-2838.

49. van Emmerik, Tim, Gert Mulder, Dirk Eilander, Marijn Piet, and Hubert Savenije. 2015. Predicting the ungauged basin: Model validation and realism assessment. *Frontiers in Earth Science* 3: 62.

50. Maier, HR, JHA Guillaume, H van Delden, et al. 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together?. *Environmental Modelling & Software* 81: 154-164.

51. Blöschl, Günter, and Alberto Montanari. 2010. Climate change impacts—throwing the dice?. *Hydrological processes* 24, no. 3: 374-381.

52. Coron, L, V Andréassian, C Perrin, M Bourqui, and F Hendrickx. 2014. On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. *Hydrology and Earth System Sciences* 18, no. 2: 727-746.

53. Klemeš, Vit. 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* 31, no. 1: 13-24.

54. Thirel, G, V Andréassian, C Perrin, et al. 2015a. Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal* 60, no. 7-8: 1184-1199.\

55. Essou, Gilles RC, Richard Arsenault, and François P Brissette. 2016b. Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of Hydrology* 537: 334-345.

56. He, Y, A Bárdossy, and E Zehe. 2011. A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences* 15, no. 11: 3539-3553.

57. Gupta, Hoshin Vijai, C Perrin, G Bloschl, et al. 2014. Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences* 18, no. 2: p. 463-p. 477.

58. Young, Andrew R. 2006. Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. *Journal of Hydrology* 320, no. 1: 155-172.

59. Efstratiadis, A, AD Koussis, D Koutsoyiannis, and N Mamassis. 2014. Flood design recipes vs. reality: can predictions for ungauged basins be trusted?. *Natural Hazards and Earth System Sciences* 14, no. 6: 1417-1428.

60. Seibert, Jan, and JJ McDonnell. 2014. Gauging the ungauged basin: relative value of soft and hard data. *Journal of hydrologic engineering* 20, no. 1: A4014004.

61. Seibert, Jan, and Marc Vis. 2016. How informative are stream level observations in different geographic regions?. *Hydrological Processes*.

62. ArcticRIMS, cited 2016: ArcticRIMS—A regional, integrated hy- drological monitoring system for the pan-Arctic land mass. [Available online at http://rims.unh.edu/index.shtml.]

63. r-ArcticNET, cited 2016: r-ArcticNET—A regional, electronic, hydrographic data network for the Arctic region. [Available online at http://www.r-arcticnet.sr.unh.edu/v4.0/index.html]

64. Weedon, Graham P, Gianpaolo Balsamo, Nicolas Bellouin, et al. 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research* 50, no. 9: 7505-7514.

65. Harris, IPDJ, PD Jones, TJ Osborn, and DH Lister. 2014. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology* 34, no. 3: 623-642.

66. Oudin, Ludovic, Frédéric Hervieu, Claude Michel, et al. 2005. Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology* 303, no. 1: 290-306.

67. Boone, A, F Habets, J Noilhan, et al. 2004. The Rhone-Aggregation land surface scheme intercomparison project: An overview. *Journal of Climate* 17, no. 1: 187-208.

68. Rojas-Serna, Claudia, Laure Lebecherel, Charles Perrin, Vazken Andréassian, and Ludovic Oudin. 2016. How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments. *Water Resources Research*.

69. Lindström, Göran, Barbro Johansson, Magnus Persson, Marie Gardelin, and Sten Bergström. 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology* 201, no. 1: 272-288.

70. Bergstrom, S. (1992), The HBV model—its structure and applications, SMHI Rep. RH 4, Swed. Meteorol. and Hydrol. Inst., Norrkoping, Sweden.

71. Seibert, Jan. 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and forest meteorology* 98: 279-293.

72. Seibert, Jan, and MJP Vis. 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences* 16, no. 9: 3315-3325.

73. Smith, Michael, Victor Koren, Ziya Zhang, et al. 2013. The distributed model intercomparison project–Phase 2: Experiment design and summary results of the western basin experiments. *Journal of Hydrology* 507: 300-329.

74. Coron, Laurent, Vazken Andreassian, Charles Perrin, et al. 2012. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. *Water Resources Research* 48, no. 5.

75. Valéry, Audrey. 2010. Modélisation précipitations–débit sous influence nivale. Élaboration d'un module neige et évaluation sur 380 bassins versants. *Agro Paris Tech., Paris, France*.

76. Chiew, F. H. S.,M. C. Peel, and A.W.Western (2002), Application and testing of the simple rainfall-runoff model SIMHYD, in Mathematical Models of Small Watershed Hydrology and Applications,edited by V. P. Singh and D.K. Frevert, pp. 335–367,Water Resour. Publ., Littleton, Colo.

77. Chiew, FHS, J Teng, J Vaze, et al. 2009. Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method. *Water Resources Research* 45, no. 10.

78. Reichl, JPC, AW Western, NR McIntyre, and FHS Chiew. 2009. Optimization of a similarity measure for estimating ungauged streamflow. *Water Resources Research* 45, no. 10.

79. Zhang, Yongqiang, and Francis HS Chiew. 2009. Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research* 45, no. 7.

80. Nash, J Eamonn, and Jonh V Sutcliffe. "River flow forecasting through conceptual models part I—A discussion of principles." *Journal of hydrology* 10.3 (1970): 282-290.

81. Jain, Sharad K, and KP Sudheer. "Fitting of hydrologic models: a close look at the Nash–Sutcliffe index." *Journal of hydrologic engineering* 13.10 (2008): 981-986.

82. Nasonova, Olga N, Yeugeniy M Gusev, and Yeugeniy E Kovalev. "Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: A case study using MOPEX materials." *Journal of Hydrometeorology* 10.5 (2009): 1128-1150.

83. Zhang, Xuesong, Raghavan Srinivasan, Kaiguang Zhao, and Mike Van Liew. 2009. Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrological Processes* 23, no. 3: 430-441.

84. Storn, Rainer, and Kenneth Price. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization* 11, no. 4: 341-359.

85. Price, Kenneth, Rainer M Storn, and Jouni A Lampinen. 2006. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, March 4.

86. Singh, R, SA Archfield, and T Wagener. 2014. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments–A comparative hydrology approach. *Journal of Hydrology* 517: 985-996.

87. Slater, AG, TJ Bohn, JL McCreight, MC Serreze, and DP Lettenmaier. 2007. A multimodel simulation of pan-Arctic hydrology. *Journal of Geophysical Research: Biogeosciences* 112, no. G4.

88. Thirel, Guillaume, Vazken Andréassian, and Charles Perrin. 2015b. On the need to test hydrological models under changing conditions. *Hydrological Sciences Journal* 60, no. 7-8: 1165-1173.

89. Osuch, Marzena, Renata J Romanowicz, and Martijn J Booij. 2014. The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. *Hydrological sciences journal*.

90. Merz, Ralf, Juraj Parajka, and Günter Blöschl. 2011. Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research* 47, no. 2.

91. Perrin, Charles, Claude Michel, and Vazken Andréassian. 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* 279, no. 1: 275-289.

92. Li, Hong, Stein Beldring, and Chong-Yu Xu. 2015. Stability of model performance and parameter values on two catchments facing changes in climatic conditions. *Hydrological Sciences Journal* 60, no. 7-8: 1317-1330.

93. Reichl, JPC, Francis HS Chiew, and AW Western. 2006. Model averaging, equifinality and uncertainty estimation in the modelling of ungauged catchments. *iEMSs Third Biennial Meeting, Summit on Environmental Modelling and Software, International Environmental Modelling and Software Society, Burling*.

94. Beven, Keith, and Jim Freer. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of hydrology* 249, no. 1: 11-29.

95. Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900.

96. Lem, S. (2013). Summa technologiae. Suhrkamp Verlag.