RESEARCH PAPER

A METHOD OF MULTI-SITE CALIBRATION OF DISTRIBUTED HYDROLOGICAL MODELS BASED ON THE NASH-SUTCLIFFE EFFICIENCY

Boris I. Gartsman[1, 4*], Dimitri P. Solomatine[1, 2, 3], Tatiana S. Gubareva[1]

[1] Water Problem Institute, Russian Academy of Sciences, Gubkina 3, Moscow, 119333, Russia

[2] Water Resources Section, Delft University of Technology, Mekelweg 5, Delft, 2628CD, Netherlands

[3] Department of Hydroinformatics and Socio-Technical Innovation, IHE Delft Institute for Water Education, Westvest 7, Delft, 2611AX, Netherlands

[4] Vernadsky Crimean Federal University, Prospekt Vernadskogo 4, Simferopol, 295007, Russia

***Corresponding author:** Gartsman@inbox.ru

Received: August 24th 2024 / Accepted: November 22nd 2024 / Published: December 31st 2024

https://doi.org/10.24057/2071-9388-2024-3564

ABSTRACT. Contemporary distributed hydrological models are detailed and mathematically rigorous, but their calibration and testing can be still an issue. Often it is based on the quadratic measure of the calculated and observed hydrographs proximity at one outlet gauge station, typically on the Nash-Sutcliffe model efficiency coefficient (NSE). This approach seems insufficient to calibrate a model with hundreds of spatial elements. This paper presents using a multi-dimensional estimator of modeling quality, being a natural generalization of the traditional NSE but which would aggregate data from several hydrological stations using Principal Component Analysis (PCA). The method was tested on the ECOMAG model developed for a sub-basin (24,400 km^2, with 15 gauges) of the Ussuri River in Russia. The results show that the presented version of the multi-dimensional NSE with PCA in calibration of spatially-distributed hydrological models has a number of advantages compared to other methods: the reduced dimensionality without loss of important information, straightforward data analysis and the automated calibration procedure; objective separation of the deterministic signal from the noise, calibration using the "informational kernel" of data, leading to more accurate parameters' estimates. Additionally, the introduced notion of the "compact" dataset allow to interpret physical-geographical homogeneity of the basins in mathematic manner, which can be valuable for hydrological zoning of the basins, hydrological fields analysis, and structuring the models of large basins. There is no doubt that further development and testing of the proposed methodology is advisable in solving spatial hydrological problems based on distributed models, such as managing a cascade of reservoirs, creating hydrological reanalyses, etc.

KEYWORDS: spatially-distributed hydrological models, multi-site calibration, multi-dimensional Nash-Sutcliffe coefficient, prediction for groups of sub-basins

CITATION: Gartsman B. I., Solomatine D. P., Gubareva T. S. (2024). A Method Of Multi-Site Calibration Of Distributed Hydrological Models Based On The Nash-Sutcliffe Efficiency. Geography, Environment, Sustainability, 4(17), 76-87
https://doi.org/10.24057/2071-9388-2024-3564

ACKNOWLEDGEMENTS: This study was partly carried out under the grant from the Russian Science Foundation, No. 23-27-00236.

Conflict of interests: The authors reported no potential conflict of interest.

INTRODUCTION

Hydrological modeling is currently characterized by the prevalence of spatially distributed models. These models are able to adequately reproduce the runoff generation processes within a river basin, and are more reliable for the assessments going beyond the existing observation ranges. The models are also required to solve many problems of spatial water management that are not limited to local tasks like assessing the parameters of bridge dimensions, water intake characteristics or reservoir volume. Contemporary models may be characterized by the detailed representation of watershed characteristics, rigorous mathematical description of processes, and strong hydrometeorological information support, but the issue of adequate diagnostic and verification of these models is still on the agendas of many researchers.

Calibration and verification techniques are primarily based on the use of objective functions in the form of a few quadratic measures of the calculated and observed hydrographs proximity, the most popular of which are the root squared mean error and the Nash-Sutcliffe model efficiency coefficient (NSE) (Nash and Sutcliffe 1970). Sometimes, the NSE is also combined with more measures, as it is done e.g. in the Kling-Gupta-Efficiency (KGE) (Gupta et al. 2009), which combines bias, variability and correlation components to improve the estimation of the performance error. To analyze the temporal variability of model performance and related parameter sensitivity, their interactions, and other aspects of model identifiability,

one may use the DYNamic Identifiability Analysis (DYNIA) method developed by Wagener et al. (2003), and by Reusser et al. (2011).

A more "physically-based" approach is to consider signature measures which are directly related to catchment functions with the aim to consider the relevance of a certain hydrological component individually (e.g. (Yilmaz et al. 2008; Pokhrel et al. 2012; Mohammed et. al. 2021; Huynh et al. 2023)). For example, signature measures based on flow duration curves (FDC) show a model performance for different discharge levels (Yilmaz et al. 2008; Cheng at al. 2012; Pfannerstill et al. 2014). It is worth noting that a calibration problem can be posed as the single-objective optimization problem, typically using algorithms of randomized search in the space of parameters (e.g., (Solomatine et al. 1999)), but also as a multi-objective optimization problem (Pokhrel et al. 2012; Efstratiadis and Koutsoyiannis 2010) with the use of several performance measures. However, we are leaving consideration of the associated possibilities for further studies, concentrating on single-objective calibration.

NSE has a number of advantages, and is still the most widely used performance measure when a model is estimated by a single variable, e.g. by discharge measured at the outlet gauge station. This approach is seen to be justified for simple models, but it seems to be insufficient to calibrate a model with hundreds of spatial design elements and a developed description of a complex set of interrelated processes in each of them.

Hydrological models spatial organization is an important aspect of their calibration and testing. Realization of this fact has led to the interest to the so-called "multi-site calibration" approaches (Wang et al. 2012; Ashu and Lee, 2023; Serur and Adi 2022; Malik et al. 2022; Xu et al. 2022; Ruiz-Pérez et al. 2017), which include comparison of the model outputs to discharge measurements at multiple locations. As presented in the mentioned papers, multi-site calibration allows for developing more reliable models. However, to the best of our knowledge, these approaches can be mostly seen as realizations of "trial and error" strategy, when the model parameters are adjusted in a attempt to iteratively reduce the model error at all the considered locations. For example, Serur and Adi (2022) describe their approach as follows: "the parameter values were changed repeatedly within the allowed ranges until acceptable agreements between observed and simulated streamflow were obtained for each gauged station". Ahu and Lee (2023) describe this procedure as "trial and error were used to acquire the fitted value until the simulated and observed values were consistent". Advantage of this approach is explicit inclusion of an expert in the process, which has certain advantages. However, there are also deficiencies of this method, and we find it important to introduce more rigor to this procedure. This prompts for developing mathematically strict procedures for multi-site calibration, leading to potentially more accurate models. It is done in the presented paper.

It is important to stress the importance of an accurate account of the spatial aspects in calibration. There is a need for comparison of the distributed models outputs not only to the point-wise measurements, but also to the Earth remote sensing data (ERS), where the data is (evenly) distributed over the measurement and calculated grids, and is not necessarily coinciding with the measurement grid. For example, Ruiz-Pérez et al. (2017) calibrate the Eco-hydrological distributed model TETIS-VEG using Normalized Difference Vegetation Index (NVDI) data obtained from the MODIS satellite. They use the method of empirical orthogonal functions (EOF), which is similar to the Principal Component Analysis (PCA) used in this article. Koch et al. (2015) assess the quality of the models by comparing soil temperature maps (LST, land surface temperature) obtained from the model with those obtained from the MODIS satellite, and also using the EOF at one of the stages. Assessment of the model accuracy based on the calculation of soil moisture by comparing the calculated fields with those ones obtained from satellites data and using the EOF apparatus is also considered in the work (Mascaro et al. 2015).

To the best of our knowledge, the problem of spatial calibration of distributed hydrological models based on direct runoff measurements of several stations using the PCA aggregation and analyzing the catchment specifics on the basis of the calculated principal components has not been solved. We see this as a significant gap requiring the novelty in approach.

Hence the main objective of this paper is to develop mathematically strict procedures for multi-site calibration of distributed models.

The approach we propose seems to contributes several innovative aspects in considering the problem of interest. Firstly, the proposed multi-dimensional estimator of modeling quality is a natural generalization of the traditional NSE to the case of modeling runoff simultaneously across a group of hydrological stations within the basin using a single spatially-distributed model. At the same time, all conceptual advantages of the NSE-based assessment are preserved, but the information content used for the model assessment increases substantially.

Secondly, using the PCA method as applied to a group of the daily discharge series improves the adequacy of the model calibration procedure. This becomes possible due to the fairly effective separation of random noise, which is inevitably present in any observational data, and calibrating the model in accordance to an array of the "cleaned" information, which should lead to more accurate parameter estimates.

Thirdly, this approach opens a possibility of obtaining meaningful interpretation of the PCA method results applied to the group of stations data, which is expressed by the proposed concept of a "compact" basin.

Thus, the framework of study presented includes the following stages: assessing the adequacy of NSE under impact of noise; correct calculation of NSE based on the data from multiple gauges; using the PSA as an "integrator" of data from multiple gauges for noise-free calibration; analysis of the case application of the method on a medium-sized basin. The last section pays attention to the formulation of the concept of a "compact" data system and its relationship with the concept of homogeneity of a river basin.

One of the currently popular points of view, consistently presented by K. Beven (2012), insists on the impossibility of determining the "truth" of multiple models (i.e., calibrated on the same data set, but which parameters still differ slightly). It is suggested to consider a wide range of estimates provided by models solely on the basis of rational decision principles under uncertainty. Without delving into further analysis of this approach, we can note that its prevalence is largely related to the dissonance noted above, when calibration tools appear to be limited to deal with the complexity of the models to be compared. Thus, there is a need to develop model calibration and verification tools matching the complexity of the models employed.

METHODOLOGY

Before presenting the essence of the methodology it is important to consider one aspect of application of NSE, which in absolute majority of practical applications is completely overlooked.

Adequacy of NSE for Model Calibration under Assumption of Noise

Let us consider the NSE score in one of the usual formulations, proposed in (Nash and Sutcliffe 1970) and analyzed in detail by Murphy (1998):

$$R^2_{NS} = 1 - \frac{\sum_{i=1}^{n}\left(Qf_i - Qo_i\right)^2}{\sum_{i=1}^{n}\left(Qo_i - Qa\right)^2} \quad (1)$$

where Qf_i and Qo_i are the simulated and observed discharges for the i th day, Qa is the mean observed discharge for the simulated period of time, and n is the number of days in this period. The second term in Eq. (1) is the ratio of the discrepancy quadratic measure between the simulated discharges and the observed ones to the similar measure of the discrepancy between the observed discharges and its "worst" approximation, obtained on the base of available data only (without any model).

According to Eq. (1) for R^2_{NS}, the mean value of the measured discharges for the period is taken as the "worst" approximation. This assessment is the simplest one, but also the "weakest" and not always an adequate one. The issue of applying various approximations as the "worst" is considered in detail in (Murphy 1998). For example, for the rivers with a predominant snowmelt water supply and a stable seasonal distribution of runoff, it would be reasonable to replace the value of Qa in Eq. (1) by $Qnorm_i$, the average annual flow rate for the day i. Such an approach is employed in meteorology, where all estimates are based on anomalies, i.e. the deviations from climatic values at a given moment of the year, which is mathematically equivalent to using "climate" as the "worst" approximation. Yet another, widely used assessment of hydrological forecasts quality (Moreido et al. 2021) $S/\sigma\Delta$ is of similar nature, where S is a root mean squared error of the forecast, and $\sigma\Delta$ is the standard deviation of the observed values during the forecasting period Δ.

In addition, the problem of perfect modeling accuracy estimation deserves special consideration. Many publications indicate that, R^2_{NS} reaches a value of 1, as provided by Eq. (1), when the simulated and observed discharges are completely equal. This result is thought to be an ideal one, and all efforts are focused on its achievement when calibrating the models, although the actual values are typically less than unity. However, each model experiment has a significant random noise, which should not be reproduced by a deterministic model.

Noise is generated by several sources. Firstly, although the model itself is supposed to be strictly deterministic, the results of its operation are strongly influenced by the quality of the input meteorological information. It is appropriate to present the model output Qf_i as the sum of the "true" simulation result Qm_i which could be obtained by feeding accurate and complete meteorological data of the highest spatio-temporal resolution, and a random error ξf resulting from inaccurate and unrepresentative meteorological data. It is impossible to evaluate directly these two components of the simulated signal due to the fundamental lack of "ideal" meteorological data.

Similarly, the measured hydrograph can be represented as the sum of the real discharges Qr_i and the random observation error ξ_o. It is important to bear in mind, that the Qr_i value still contains some internal uncertainty, which, in fact, can also be considered as random. This uncertainty results from the fact that in the real basin there are processes, often random or chaotic, which are beyond the scope of our knowledge, or, at least, beyond the base concepts of the model in use. Model designers often interpret them as "subgrid" processes.

With the above considerations in mind, the formula for R^2_{NS} can be complemented with the independent random variables ξf and ξo with zero mean values, and can be rewritten as follows:

$$R^2_{NS} = 1 - \frac{\sum_{i=1}^{n}\left(\left(Qm_i + \xi_f\right) - \left(Qr_i + \xi_o\right)\right)^2}{\sum_{i=1}^{n}\left(\left(Qr_i + \xi_o\right) - Qa\right)^2} \quad (2)$$

Opening the brackets, we perform the term-wise summation in the numerator and the denominator, assuming n is sufficiently large. Taking into account independence and zero means of ξf and ξo, we can carry out the following derivations:

$$R^2_{NS} = 1 - \frac{\sum_{i=1}^{n}\left(\left(Qm_i + \xi_f\right) - 2\left(Qm_i + \xi_f\right)\left(Qr_i + \xi_o\right) + \left(Qr_i + \xi_o\right)^2\right)}{\sum_{i=1}^{n}\left(\left(Qr_i + \xi_o\right)^2 - 2Qa\left(Qr_i + \xi_o\right) + Qa^2\right)} =$$

$$= 1 - \frac{\sum_{i=1}^{n}\left(Qm_i^2 - 2Qm_i \cdot Qr_i + Qr_{i^2} + \xi_f^2 + \xi_o^2\right)}{\sum_{i=1}^{n}\left(Qr_i^2 + \xi_o^2 - 2Qa \cdot Qr_i + Qa^2\right)} = \quad (3)$$

$$= 1 - \frac{\sum_{i=1}^{n}\left(\left(Qm_i - Qr_i\right)^2 + \xi_f^2 + \xi_o^2\right)}{\sum_{i=1}^{n}\left(\left(Qr_1 - Qa_i\right)^2 + \xi_o^2\right)}$$

With the ideally accurate simulation ($Qm_i = Qr_i$), by dividing the numerator and the denominator by n, we obtain the following estimate of the extreme value of R^2_{NS}:

$$lim_{Qm \to Qr} R^2_{NS} = 1 - \frac{D_{\xi f} + D_{\xi o}}{D_{Qr} + D_{\xi o}} = \frac{D_{Qr} - D_{\xi f}}{D_{Qr} + D_{\xi o}} \quad (4)$$

where Dx is the variable x variance. This expression could also be derived directly based on the general principles of the random variables algebra.

This result is valid only for the assumptions presented earlier. However if, for example, the considered errors are systematic with a non-zero mean value, or variables are not independent, then the final expression will be much more complicated. We assume here that during the calibration process, the systematic part of the analyzed errors is minimized together with the systematic errors inherent in the model itself, which is developed due to the conceptual schematization of real processes.

In real modelling practice one can inevitably expect certain (often significant) random "noise" in the data and simulations. One can see, that in this case the maximum value of NSE for an "ideal" model is below 1. For example, if the variances of errors in input ξf and output ξo are 4% each, then $lim_{Qm \to Qr} R^2_{NS}$ will be 0.923. An attempt of a modeler to exceed (and even achieve) this value cannot be justified, and if this happens, it points at an excessive flexibility of the model. If the number n is not high enough, then the final formula of $lim_{Qm \to Qr} R^2_{NS}$ will also become more complicated, and the estimate will increase slightly,

reflecting the well-known effect of obtaining a fictitiously "better" result of estimates on shorter data samples. We find it important that these considerations must be taken into account for the correct application of the R^2_{NS} measure in rigorous testing of complex models that claim a high degree of proximity to natural processes. Surprisingly, in the majority of papers on (hydrological) modelling the considerations presented above, and their impact on the use of NSE, are not seen as an important part of the calibration and verification process.

Spatial/Dimensional Aspect of Calibration

We will be now considering a more complicated aspect of calibration and testing of the distributed models of extensive basins. Calibration of such models typically uses the data of several gauges damming sub-basins of various sizes and physiographic conditions. The use of all available data for calibration is advisable for the following reasons. Firstly, for the maximum information utilization. Secondly, simulation of a hydrologic event spatial structure is important for solving various tasks, such as operational event planning during floods or reservoir cascade management, when a model should guarantee acceptable accuracy both in the main stream and other stations of the basin. It is also worth mentioning that it would be methodologically attractive to have a unified procedure for such calibration.

In many applications R^2_{NS} estimates are used for distributed models calibration with the data at the outlet gauge station. Along with this, R^2_{NS} is usually evaluated for other gauging stations by experts, so the choice of the most preferred combination of estimates is quite arbitrary. Averaging (simple or weighted) of R^2_{NS} estimates over several gauges can be also employed. It is worth noting that that several stations data in one basin belonging to nested sub-basins are highly correlated and therefore use of simple averaging would be incorrect.

The following question can be posed: is it possible to solve a problem of a multivariate calibration by using the approach similar to the one utilized for the Nash-Sutcliffe model efficiency coefficient proposition? The following consideration answers this question positively. For that, it is possible to use Eq. (1) for multivariate calibration by changing the value of a single gauge discharge to a vector value, that is an ordered set of discharges from multiple gauges:

$$\square^{*}R^2_{NS} = 1 - \frac{\sum_{i=1}^{n}\left(\overline{Qf}_i - \overline{Qo}_i\right)^2}{\sum_{i=1}^{n}\left(\overline{Qo}_i - \overline{Qa}\right)^2} \qquad (5)$$

The asterisk on indicates that this is a multi-dimensional estimate combining all series of the data system considered. The calculation algorithm and the properties of the estimator remain almost unchanged – as a result, one obtains a scalar quantity that reaches 1 in case of perfect modeling without noise, or 0 with the absence of a relationship between the simulated values and the observed ones, and becomes less than zero when their relationship is inverse. In particular, considering Eq. (5) for the simplest case of a two-dimensional vector (i.e., two gauges), in accordance with the rules of vector operations, and after recombination, we obtain:

$$\begin{aligned}
\square^{*}R^2_{NS} &= 1 - \frac{\sum_{i=1}^{n}\left(\left(Qf1_i, Qf2_i\right) - \left(Qo1_i, Qo2_i\right)\right)^2}{\sum_{i=1}^{n}\left(\left(Qo1_i, Qo2_i\right) - \left(Qa1, Qa2\right)\right)^2} = \\
&= 1 - \frac{\sum_{i=1}^{n}\left(Qf1_i - Qo1_i, Qf2_i - Qo2_i\right)^2}{\sum_{i=1}^{n}\left(Qo1_i - Qa1, Qo2_i - Qa2\right)^2} = \\
&= 1 - \frac{\sum_{i=1}^{n}\left(\left(Qf1_i - Qo1_i\right)^2 + \left(Qf2_i - Qo2_i\right)^2\right)}{\sum_{i=1}^{n}\left(\left(Qo1_i - Qa1\right)^2 + \left(Qo2_i - Qa2\right)^2\right)} = \\
&= 1 - \frac{nD_{Qo1}\left(1 - 1 + \frac{\sum_{i=1}^{n}\left(Qf1_i - Qo1_1\right)^2}{nD_{Qo1}}\right) + nD_{Qo2}\left(1 - 1 + \frac{\sum_{i=1}^{n}\left(Qf2_1 - Qo2_1\right)^2}{nD_{Qo2}}\right)}{n\left(D_{Qo1} + D_{Qo2}\right)} = \\
&= 1 - \frac{D_{Qo1}\left(1 - \square^{1}R^2_{NS}\right) + D_{Qo2}\left(1 - \square^{2}R^2_{NS}\right)}{D_{Qo1} + D_{Qo2}} = \frac{D_{Qo1}\square^{1}R^2_{NS} + D_{Qo2}\square^{2}R^2_{NS}}{D_{Qo1} + D_{Qo2}}
\end{aligned} \qquad (6)$$

where $Qf1$, $Qf2$, $Qo1$, $Qo2$, $Qa1$ and $Qa2$ are the modeled, measured and mean discharges in the 1st and 2nd gauges respectively, and и are the regular one-dimensional Nash-Sutcliffe efficiency coefficients for every gauge. The result is that the multi-dimensional is equal to a weighted average of the , and the weights are the corresponding discharge variances. The considered case of two dimensions can be easily extended to deal with higher dimensions (i.e., more than two gauges).

PCA as an "integrator" of data from multiple gauges for noise-free calibration

To analyse and possibly reduce dimension of multi-dimensional sets of highly correlated data the Principal Components Analysis (PCA) method is often used. The principle of the method is the linear transformation of the original coordinate system, in which the initial matrix of correlated data is given, into a new one, which is typically called *U*-space. The first *U*-space axis is oriented along the highest scatter of the initial data, the second one – along the maximum remaining scatter and is orthogonal to the first axis, and further in a sequence. As a result, the initial set of variables is projected to a new set of variables in the *U*-space that is named principal components (PCs). The number of PCs is equal to the initial number of variables, and PCs are non-correlated and orthogonal to each other. Besides, due to the *U*-space conversion method, PCs variance λ is gradually decreasing from the first one to the last one, and thus amounting in the sum the total variance of the initial dataset.

Those properties often make it possible to limit the analysis to only few first PCs representing an expertly defined proportion of the initial dataset variance, of 0.9 or 0.95, for example. Those PCs are considered to be significant, representing the substantial content of the dataset, whereas the rest of the PCs are random noise. Therefore, when applied properly, the PCA makes it possible to decrease the dimensions of the dataset under study and to separate the content-rich data from the noise. Besides, the PCA enables to identify the latent data structures, making it effective in solving classification and zoning tasks.

Apart from using multi-dimensional NSE, using PCA is forms yet another important part of the methodology proposed. PCA serves as an "integrator" of the observed data from multiple (N) gauges and the simulated data for these gauge locations. The information integrated in the major *K* PCs, is considered to represent the deterministic component of the system, and minor (*N* – *K*) components are considered to be representing the noise. The dataset based on the major *K* PCs can be transformed back into multiple *N*-gauges series, and this transformed data is used for the "noise-free" calibration.

A detailed presentation of the PCA and it practical use is provided in a number of publications (e.g., (Pomerantsev 2014; Harris 2001)). The algorithm is implemented in many software packages; we used Minitab Release 14. The PCA transformation is possible both by using the correlation matrix with preliminary standardization of the initial data, and by using the covariance matrix involving initial data processing. The first option scales all variables and equalizes their "weights". Considering that in our investigation the large rivers are more "weighty" compared to small ones, the second option with covariance matrix is adopted.

This section presents the main components of the methodology for calibration of distributed models proposed in this paper, which is based on using a limited number of PCA components, thus concentrating on the main properties of the basin, and reducing noise.

RESULTS: APPLICATION OF THE METHODOLOGY TO A MEDIUM SIZE BASIN

The Ussuri River basin model is considered. It was developed by using the ECOMAG (Ecological Model for Applied Geophysics), developed in the 1990s by Yu.G. Motovilov in the Water Problems Institute of the Russian Academy of Sciences. ECOMAG is a spatially-distributed hydrological model, based on HRU (Hydrological Response Unit) concept. Its complete details are presented in a number of publications (Motovilov et al. 1999; Danilov-Danilyan et al. 2014; Bugaets et al. 2023). The spatial structure of the ECOMAG model splits watershed into sub-basins based on topography, river network structure, soil and vegetation type, land use, and variability of climate characteristics. The main ECOMAG model equations were adopted from the full spatially distributed model by spatial aggregation at subbasin scale, neglecting secondary terms. Daily resolution time series of precipitation, air temperature and air relative humidity are used as inputs. Computation of river basin hydrological response described by the two main phases: calculation of the effective precipitation for each sub-basin, and then routing it through the river network. Runoff from sub-basin is calculated as sum of the three components: Horton overland (surface) flow, soil flow and groundwater outflow. During warm periods precipitation is partially infiltrated and moves along the hillslopes as interflow. Excess water produces surface flow and moves downslope towards the drainage network. The rest of the water that has not been drained to rivers as lateral or surface flow can be evaporated or percolated into deep aquifers. Within cold and mid-season periods the model describes snowpack evolution and soil freezing-thawing cycle. Spatial aggregation made it possible to reduce model calibration to a small number of parameters (Table 1), most of which are correction factors for hydrophysical characteristics.

The basin of Ussuri River near Kirovskiy is 24,400 km^2 (Fig. 1). Its major part is middle altitude taiga with, about 1/3 part is sub-mountain plain-like territory, partially reforested and considerably plowed up. This is a typical Far Eastern river characterized by frequent pluvial floods (Moreido et al. 2021). There were 15 gauges operating in the basin during various time periods, with the highly correlated data.

The data of simultaneous observations of nine gauges for the period beginning 1978 through 1990 (Table 2) were selected for analysis, because the number of synchronously operating gauges is maximal and the quality of network observations is the best during these years. The gauges with watershed areas less than 1,000 km^2 were not included into the data set. Besides, in case of several gauges along one river, they were selected by the watershed area on the condition that the discharges at the neighboring gauges would differ by at least of factor of two. Equal distribution of gauges within the river basin is of course preferred. Discharge data series for the period beginning 1978 through 1990 were simulated with the ECOMAG model. Both observed and simulated datasets were used after every model run for calibration and final estimations of modeling quality by different versions of NSE.

It is important to mention the following. In this paper we are presenting a method able to deal with multi-site calibration and verification of distributed models. However, we are presenting only the results of model calibration on a single data set, but not its testing (on an unseen data set). This is a certain limitation of this study (partly due to the limited resources allocated for this work), but still allows to fully demonstrate the feasibility of the presented approach to calibration well. The comprehensive modelling framework to be developed should include all stages required by modelling theory.

Table 1. ECOMAG calibrated parameters

Parameter	Short Name
Coef. of vertical saturated hydraulic conductivity	GFB
Coef. of horizontal saturated hydraulic conductivity	GFA
Soil evaporation coefficient	EK
Baseflow constant, mm day^{-1}	GROUND
Coef. of snowmelt intensity, mm day^{-1} °C	ALF
Critical air temperature snow/rain, °C	TCR
Snowmelt air temperature, °C	TSN
Air temperature gradient, °C 100 m^{-1}	TGR
Precipitation gradient, mm 100 m^{-1}	PGR
Coef. of vertical saturated hydraulic conductivity	GFB
Coef. of horizontal saturated hydraulic conductivity	GFA

Fig. 1. Hydrological observation network of Ussuri Basin. The runoff gauge stations analysed: 1 – Ussuri–Kirovskiy; 2 – Ussuri–Koksharovka; 3 – Arsenievka–Yakovlevka; 4 –Ussuri–Novomikhalovka; 5 – Pavlovka–Uborka; 6 – Arsenievka–Anuchino; 7 –Ussuri–Verkhniaya Breevka; 8 – Izvilinka–Izvilinka; 9 – Krylovka–Krylovka

Table 2. Model quality estimation by both one- and multi-dimensional NSE on the base of 9 gauges data in Ussuri River basin for 1978 – 1990 years

River – gauge station		Ussuri – Kirovskiy	Ussuri – Koksharovka	Arsenievka – Yakovlevka	Ussuri – Novomikhalovka	Pavlovka – Uborka	Arsenievka –Anuchino	Ussuri – Verkhniaya Breevka	Izvilinka – Izvilinka	Krylovka– Krylovka
Watershed area, km^2		24,400	9,340	5,180	5,170	3,350	2,480	1,800	1,160	1,070
Discharges in initial coordinate system										
Observed discharge series, m^3/sec	Mean	186.0	72.6	36.5	38.5	27.3	20.3	15.6	11.6	8.0
	Standard deviation	268.8	127.4	71.1	62.3	43.4	41.5	27.0	18.3	16.2
	Total variance percentage	70.6%	15.9%	4.9%	3.8%	1.8%	1.7%	0.7%	0.3%	0.3%
Simulated discharge series, m^3/sec	Mean	212.9	90.8	42.5	53.3	31.6	21.8	19.5	14.1	7.6
	Standard deviation	253.1	122.0	57.4	73.9	48.7	34.4	32.3	23.1	13.5
	Total variance percentage	68.9%	16.0%	3.5%	5.9%	2.5%	1.3%	1.1%	0.6%	0.2%
		0.783	0.722	0.625	0.640	0.579	0.650	0.494	0.409	0.229

Multi-dimensional on 9 series is 0.750										
PCs in *U*-space										
		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PCs of observanions	Mean	210.4	0.59	−5.72	2.81	1.88	0.77	0.08	−0.17	0.29
	Standard deviation	300.9	93.4	38.6	31.6	16.5	12.0	10.3	8.2	3.4
	Total variance percentage	88.5%	8.5%	1.5%	1.0%	0.3%	0.1%	0.1%	0.1%	0.0%
PCs of simulations	Mean	229.2	9.12	−7.67	2.85	6.34	−2.37	3.03	−1.26	−0.389
	Standard deviation	293.0	73.4	21.5	22.7	14.0	17.8	12.4	9.17	4.08
	Total variance percentage	92.3%	5.8%	0.5%	0.6%	0.2%	0.3%	0.2%	0.1%	0.0%
		0.819	0.358	−0.020	0.170	−0.718	−1.7151	−1.301	−0.9678	−1.054
Multi-dimensional on 9 PCs is 0.767, on the first two PCs – 0.778										

The observed data matrix *Qo (I×J)* presents the sets of daily measured discharges during the calibrating period at 9 gauges, where *I* is the series length (number of days), *J* is the number of variables (gauges). Based on the result of the model simulation performed in an ordinary way (i.e. by calibrating by the outlet gauge data with the expert analysis of other gauges), the simulated daily discharge matrix *Qf (I×J)* is composed. The PCA conversion of *Qo (I×J)* into the *U*-space is carried out, obtaining the eigenvectors *P (J×J)* matrix and the PCs matrix of observations *Uo (I×J)*. Subsequently, the *Qf (I×J)* matrix is converted similarly into the same *U*-space through the *P (J×J)* matrix, that resulted in the PCs matrix of simulations *Uf(I×J)*. The individual values (k=1, 2, …, 9) are evaluated by the series of both observed and simulated discharges at every gauge, and hereafter the multi-dimensional evaluation is done by them. The same procedure is applied for the observed and simulated PCs series from 1 through 9, and the multi-dimensional evaluation is carried out as well. Table 2 presents the summary of the evaluation results.

As presented in Table 2, the variances of PCs decrease very quickly – the sum of the variances of PCo1 and PCo2 amounts for 97% of the total variance, exceeding the threshold estimate at 95%, while the other PCs variances are negligible. So, it is possible to consider the first two PCs as the deterministic part of the system, and the remaining seven ones as the random noise. The individual values of , which show the simulation quality of each PC, confirm this hypothesis - for the first PCo1 the estimate is 0.819 (good), for the second PCo2 it is 0.358 (satisfactory), then the estimate decrease, even reaching large negative values.

Note, that observation variables are also graded in the initial matrix *Qo(I×J)* by variance. Sum of variances of two largest gauges is 86.5% of the total variance, that is so close to the sum of variances of PCo1 and PCo2 of observations. However, the values for all gauge series are still high with the slight downward trend at watershed area reduction. Fig. 2 distinctly shows the difference between the gauge hydrographs and the charts of PCs values at the same positions in the matrixes. The hydrographs' values are essentially positive, and the charts have a specific shape similar to each other, whereas the PCs numbered 3 and higher look like random signals with zero mean.

In other words, the content-rich information from the hydrographs of various gauges is evenly distributed, and is duplicated, which is confirmed by high cross-correlation coefficients. The noises are similarly distributed across all gauges. It is obvious that the PCA transformation brings the initial correlated dataset to the system of uncorrelated PCs collecting independent content-rich data into two first ones, whereas the noise is distributed between the rest. It is known that the random variables' total variance being equal to the sum of variances is true for the independent random variables only. Table 2 really shows that this rule is true for PCs, however the total variance of the initial variables is more than three times higher than the sum of their individual variances. This proportion gives a rough indicator of the data "duplication" in the initial data matrix.

It should be emphasized that it is impossible to strictly separate a signal into a content-rich part and the noise by statistical methods (and, note, the 5% threshold level of noise is set arbitrarily). Therefore, it would be more correct to say that the noise content prevails in the removed seven components, and the content-rich data considerably prevails in the first two components compared to the initial dataset.

The Nash-Sutcliffe multi-dimensional coefficient estimation for a traditionally calibrated model does not depend on the coordinate space – it is calculated according to Eq. (6) and is equal to 0.750. The characteristics of the PCs, including their independent status and the distribution between a content-rich signal and noise, make it possible to perform calibration by only PC1 and PC2 when calculating the multi-dimensional values with this two PCs, that is marked by two asterisks. The is equal to 0.778 by the result of a traditional calibration using standard NSE. By further calibration this value can be risen up to 0.804 for a new PCs matrix of simulations *Uf'(I×J)*. Thereby, the final for all gauges reaches 0.780, so is rising by 4%. Once additional calibration is performed for a content-rich part of data, which can be seen as the "information kernel" of the dataset, it is reasonable to expect that the parameters estimates would be more reliable (however, this should be additionally verified).

Supposedly, the PCs from 3 through 9 mainly contain noise, therefore, it is reasonable to discard them. It is done by zeroing all but the first two matrix columns in *Uf'(I×J)*, and the subsequent reverse *U*-space transformation into the initial one. The resulting *Qf''(I×J)* will have the "noise free" modeled values. The resulting values are equal to 0.789, and remain

practically unchanged. However, it is worth mentioning that this procedure considerably effects the distribution of across individual gauges.

With the traditional calibration, there are typically some stations with the small catchment areas, for which the model performance is very low. The reasons for this phenomenon are not very clear. We observed this effect for the gauge station Krylovrka river near Krylovka, where is only 0.229 (Table 2). To improve that, we have transformed the two-component simulated dataset (based on two PCs) back into nine-gauges series (that is the model was calibrated on the two first PCs thus excluding noise represented by the higher-order components). As a result, we obtained = 0.548 which is much better (albeit not ideal). Thereby, and a watershed area are closely related (Fig. 3). It would be reasonable to conclude that under traditional calibration, uncontrolled distribution of noise concentrates at the gauges at small sub-basins whose contribution to the overall flow is also small. This issue is to a large extent resolved when the noise components are discarded.

Fig. 2b presents the two first PCs obtained as the result of the initial data transformation. Their physical meaning is revealed when zoomed at the X axis (Fig. 4a). The first principal component has typical appearance of a discharge hydrograph and, so to say, presents a "generalized hydrograph" of the basin. The second principal component has zero mean and a specific waveform nearby every peak of the "generalized hydrograph". This component indicates obviously that hydrographs are non-synchronous and systematically deviate from the "generalized" one. In other words, the second principal component reflects the space-time structure of the events within the basin.

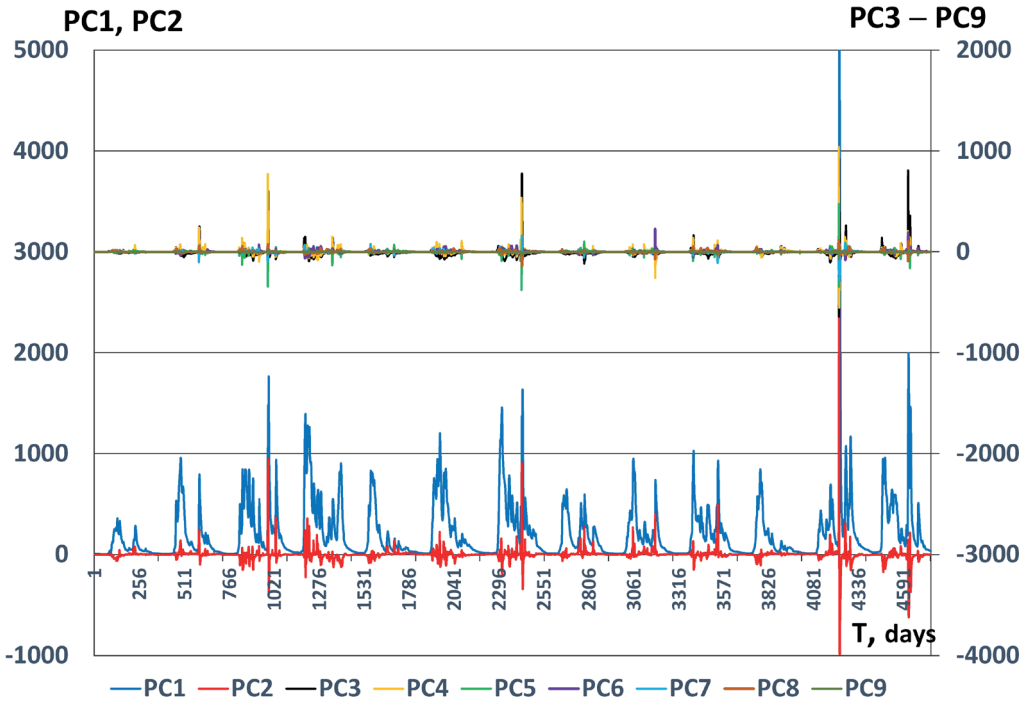

Fig. 2. The series of Ussuri Basin dataset analysed (1978-1990): a) observed daily discharges in initial coordinate system (1 – Ussuri–Kirovskiy; 2 – Ussuri–Koksharovka; 3 – Arsenievka–Yakovlevka; 4 –Ussuri–Novomikhalovka; 5 – Pavlovka–Uborka; 6 – Arsenievka–Anuchino; 7 –Ussuri–Verkhniaya Breevka; 8 – Izvilinka–Izvilinka; 9 – Krylovka–Krylovka); b) PCs in U-space from 1 till 9

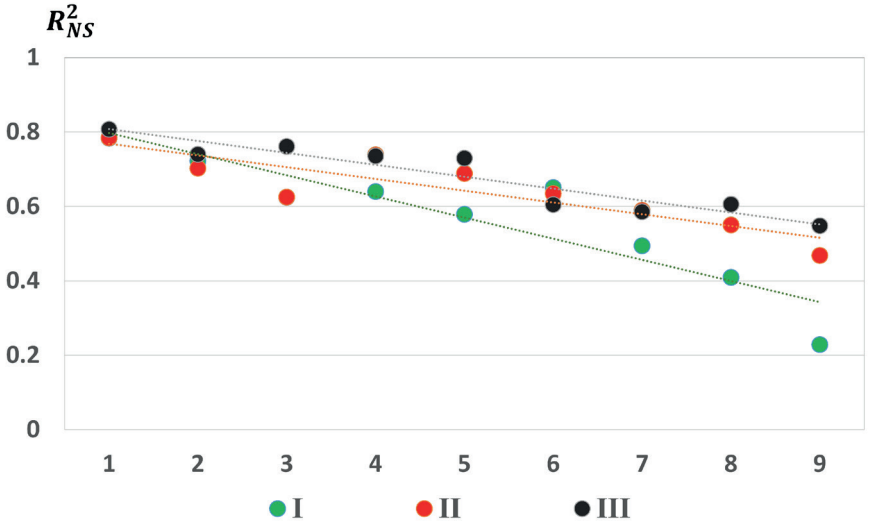

Fig. 3. The one-dimensional Nash-Sutcliffe efficiency coefficients, resulting from different calibration methods, for discharge series: 1 – Ussuri–Kirovskiy; 2 – Ussuri–Koksharovka; 3 – Arsenievka–Yakovlevka; 4 –Ussuri–Novomikhalovka; 5 – Pavlovka–Uborka; 6 – Arsenievka–Anuchino; 7 –Ussuri–Verkhniaya Breevka; 8 – Izvilinka–Izvilinka; 9 – Krylovka–Krylovka. Calibration methods: I – traditional, mainly by discharge measured at outlet gauge station with expert control in others gages; II the same with removing the random noise; III – using the multi-dimensional NSE by all discharge series with removing the random noise

FURTHER EXTENSION: COMPACT DATASETS

It is worth stressing once again some advantages of the multi-dimensional calibration provided in the example above. Firstly, application of a unified measure for a group of gauges reduce the problem of multi-dimensional calibration to the one of one-dimensional problem. Thereby, there is no need in expert evaluation of the relation between model accuracy at various gauges, which facilitates the application of automated procedures. Eq. (6) provides the theoretically justified way of weighting the individual components of NSE by the variances of the corresponding time series, eliminating thus the subjectivity typical of other non-aggregated modeling performance metrices for a group of gauges, designed on the basis of various "practical" considerations.

The example above shows that the PCA procedures allow for considerable reduction of the dimensionality of the datasets for medium-size basins, which are conditionally homogeneous by physiographic conditions and have highly correlated discharge hydrographs at various gauge stations: the useful information is stored in two PCs of varied physical interpretation and the most of the noise can be removed. This opens a possibility to identify the "informational core" of a data set, using which the model calibration would potentially lead to more reliable and stable parameters estimates. It is suggested to name such data sets, as well as their associated basins, the "compact" data sets.

As it may be assumed on the basis of the provided analysis, a "compact" basin is a homogeneous one both by the underlying surface structure and the character of the meteorological forcing, as it is the case for the studied Ussuri River basin. The basin's high homogeneity is expressed by similar response of sub-basins to meteorological forcing, or, in other words, the discharge hydrographs of all sub-basins are similar.

For small basins homogeneity would typically be high. However, we may assume that a certain number of varying sub-basin hydrographs may be averaged without substantial information loss within a basin of few tens of thousands square kilometres (for daily data step), that just is the medium-size. In this case the variance between the hydrographs in different gauges within the basin is determined only by the characteristics of the basin's concentration time and flow in the network. To demonstrate this approach, a simple model demonstrating the properties of a "compact" dataset can be built as follows.

We may use the facts known from hydrological science, such as the linear dependency between discharge and watershed area, and the power dependency between the basin's runoff concentration time and its area, with the exponent close to 0.5 (see, for example, Rodríguez-Iturbe and Rinaldo (1997)). Let us examine the observed hydrograph at the gauge of the Ussuri River near Novomikhaylovka,

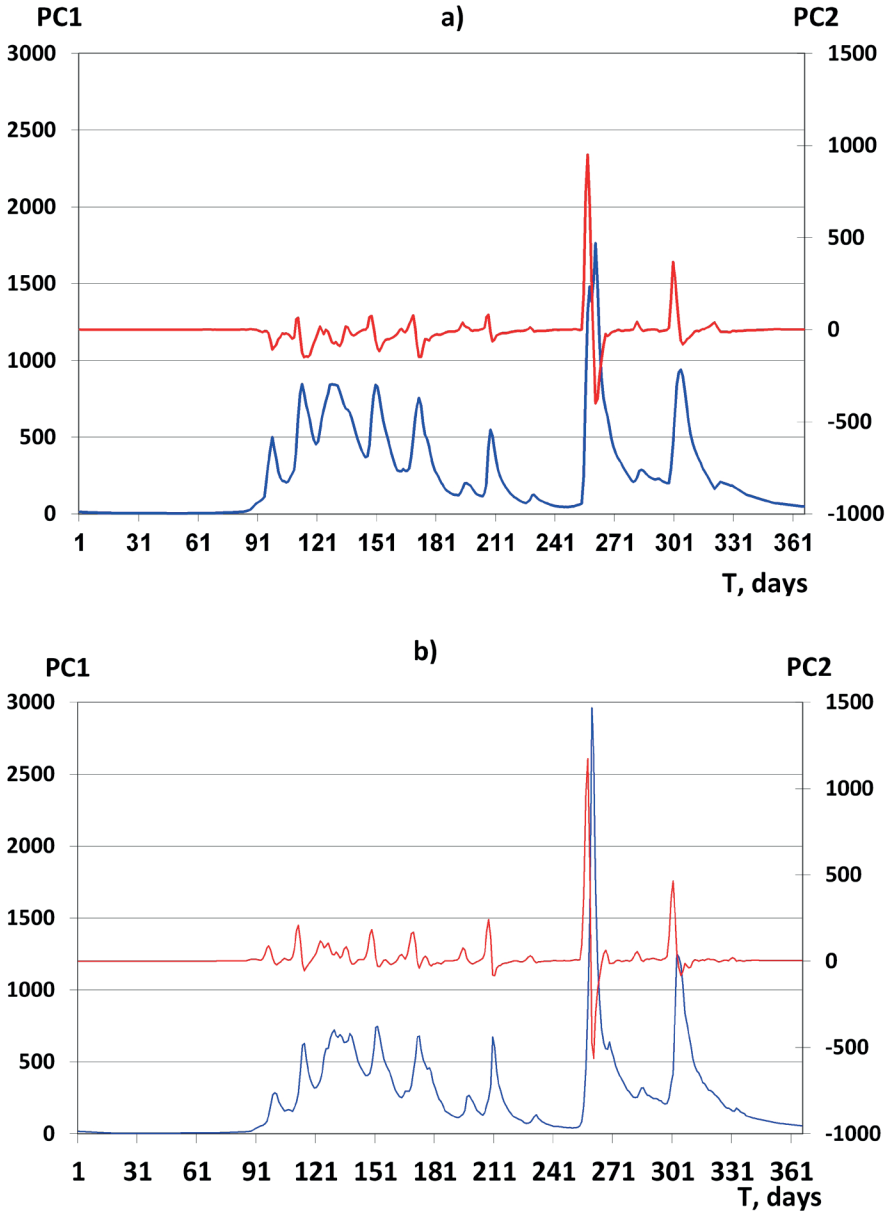

Fig. 4. The example of PC1 (blue line) and PC2 (red line) for the actual observed (a) and the constructed "observed" datasets (b) for 9 gauging stations within Ussuri River Basin (year 1980)

which corresponding sub-basin area is about 5 times higher than that that corresponding to its smallest streams, and 5 times lower than that corresponding to the outlet gauge. We calculate daily runoff specific discharges by the data from the mentioned gauge and transfer them to other gauges according to their watershed areas. As proposed above, the concentration time is proportional to the square root of the area, and we accordingly may shift in time the hydrographs at different gauges relative to one another. Thus, the constructed dataset is obtained from a single observed hydrograph using scaling and shifting transformations only.

For constructing a more realistic dataset, random noise is added to the hydrographs with approximately 20% variance of the main signal. The noise simulates both observational errors and the consequences of other factors, and is modeled by a random number generator with the normal distribution. At each point of time the random noise value is in proportion to discharge and has a considerable inertia described by the coefficient of autocorrelation. The obtained set of hydrographs can be assumed to be the "observed" data. The "simulated" data is the same data, but without the added random noise. This way of constructing the datasets is of course a primitive one, however, it seems sufficient to show the relation between the homogeneity of basin surface and weather conditions, and the statistical "compactness" of observational dataset.

These datasets are processed as described in the previous section regarding the observed and simulated Ussuri River basin data. Table 3 provides the processed results. Tables 2 and 3 mapping confirms that constructed datasets for the period beginning 1978 through 1990 have the properties similar to the actual ones. Thus, "simulated" data have lower variance than the "observed" one. Variance of the constructed series at the gauges is smaller for smaller areas, and the sum of variances for the two largest basins is equal to 88.8% of that for the "observed" data. Individual at gauges vary from 0.872 to 0.316, where the lowest value is related to the smallest basin. estimate is 0.764.

Table 2. Imitation of model quality estimation by both one- and multi-dimensional NSE on the base of constructed data series for 9 gauges in Ussuri River basin for 1978 – 1990 years

River – gauge station		Ussuri – Kirovskiy	Ussuri – Koksharovka	Arsenievka – Yakovlevka	Ussuri – Novomikhalovka	Pavlovka – Uborka	Arsenievka –Anuchino	Ussuri – Verkhniaya Breevka	Izvilinka –Izvilinka	Krylovka– Krylovka
Watershed area, km^2		24,400	9,340	5,180	5,170	3,350	2,480	1,800	1,160	1,070
Discharges in initial coordinate system										
Observed discharge series, m^3/sec	Mean	185	78.4	38.7	41.4	30.9	19.3	12.9	9.2	6.9
	Standard deviation	306	129	72.5	73.6	68.8	34.1	25.6	21.4	10.9
	Total variance percentage	75.4%	13.4%	4.2%	4.4%	3.8%	0.9%	0.5%	0.4%	0.1%
Simulated discharge series, m^3/sec	Mean	192	73.7	40.9	40.8	26.4	19.6	14.2	9.1	8.4
	Standard deviation	312	119	66.2	66.0	42.8	31.7	23.0	14.8	13.7
	Total variance percentage	78.4%	11.5%	3.5%	3.5%	1.5%	0.8%	0.4%	0.2%	0.2%
		0.754	0.818	0.759	0.836	0.777	0.872	0.714	0.769	0.316
Multi-dimensional on 9 series of "observations" and "simulations" is 0.768										
PCs in U-space										
		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PCs of observanions	Mean	–212	–15.4	2.3	–3.4	–3.0	–3.9	–1.0	1.9	1.9
	Standard deviation	335	103	55.2	29.3	24.2	18.3	12.7	6.4	6.4
	Total variance percentage	87.7%	8.3%	2.4%	0.7%	0.5%	0.3%	0.1%	0.0%	0.0%
PCs of simulations	Mean	–217	–8.9	–1.4	–7.5	–5.2	–6.0	–2.2	3.2	3.2
	Standard deviation	340	83.1	29.0	11.4	10.9	13.9	6.9	5.7	5.7
	Total variance percentage	93.3%	5.6%	0.7%	0.1%	0.1%	0.2%	0.0%	0.0%	0.0%
		0.818	0.561	0.231	–0.233	–0.254	0.410	–0.052	–0.102	–0.492
Multi-dimensional on 9 PCs is 0.795, on two first PCs – 0.797										

Being projected into the *U*-space, the constructed data is effectively convoluted, and summarized dispersion of the first two PCs is up to 96.0% for the "observed" hydrographs, and it is 98.9% for the "simulated" ones. Individual for PCs are progressively reducing and exceed the 0.5 threshold for the two first PCs only with the estimate equal to 0.796. The PCs following the first two ones look like random stochastic sequences, whereas the charts of the PCc1 and PCc2 are very close to the ones developed by the actual observed and simulated data (Fig. 4).

It can be seen that the Ussuri River basin, being quite homogeneous in structure and dynamics, allows for demonstrating the main properties and usefulness of the proposed notion of a "compact" dataset. For sure, more studies are needed for the further development of this notion.

CONCLUSIONS AND OUTLOOK

The conducted study demonstrates that the presented version of the multi-dimensional NSE with PCA in calibration of spatially-distributed hydrological models, if compared to other calibration methods, has a number of advantages:

– the reduced dimensionality without loss of important information, straightforward data analysis, and the automated calibration procedure;

– objective separation of the deterministic signal from the noise, calibration using the "information kernel" of data, leading to more accurate parameters' estimates.

Additionally, the introduced notion and the procedure of building the "compact" datasets allows for physical interpretation of "compact" or "noncompact" sub-basins, which is valuable for hydrological zoning of the basins, hydrological fields analysis, and structuring the models of large basins. It should be emphasized once again that the proposed methodology seem of most advantageous along the transition to solving those problems of hydrology that can be solved solely on the basis of spatially distributed models.

The methods presented in this paper should be seen as only one of the possible ways of handling the multi-dimensional time series in the context of space-distributed model diagnosis and calibration. The further research may be aimed at:

(a) testing the hypothesis that the parameters estimated by multi-dimensional calibration are more reliable and adequate that those estimated by other calibration procedures;

(b) exploring variability of the simulation accuracy with time, and the spatially aggregated uncertainty;

(c) testing other techniques for data aggregation and dimensionality reduction;

(d) explicit consideration of data series autocorrelation, which is closely related to the issues of their predictability and the overall simulation accuracy.

The presented methodology (and the framework to be created on its basis) should be developed and tested further: verified on more case studies, include explicit testing of the calibrated models on unseen data sets, and perhaps even extended into multi-objective calibration version.

The presented approach opens yet another potential opportunity. Since the multi-dimensional Nash-Sutcliffe index is a single basin characteristic of the simulation accuracy, it is assumed that it can be evaluated according to data from different groups of stations for different time periods in compliance with the principles of sample evaluation, i.e., while respecting the representativeness of these groups of stations in some sense. This opens up the prospect of expanding the information base for calibration of distributed hydrological models, and this may allow using full multi-year archives of standard runoff observations for all gauges in the same basin, despite the existing gaps in the observations and changes in the observation network configuration.

REFERENCES

Ashu A.B. and Lee S.-I. (2023) Multi-Site Calibration of Hydrological Model and Spatio-Temporal Assessment of Water Balance in a Monsoon Watershed. Water, 15, 360. https://doi.org/10.3390/w15020360.

Beven K.J. (2012) Rainfall-Runoff Modelling: The Primer. 2nd ed. Blackwell: Wiley.

Bugaets A., Gartsman B., Gubareva T., Lupakov S., Kalugin A., Shamov V. and Gonchukov L. (2023) Comparing the Runoff Decompositions of Small Experimental Catchments: End-Member Mixing Analysis (EMMA) vs. Hydrological Modelling. Water, 15(4), 752, https://doi.org/10.3390/w15040752.

Cheng L., Yaeger M., Viglione A., Coopersmith E., Ye S. and Sivapalan M. (2012) Exploring the physical controls of regional patterns of flow duration curves - part 1: Insights from statistical analyses. Hydrol. Earth Syst. Sci., 16, 4435–4446, https://doi.org/10.5194/hess-16-4435-2012.

Danilov-Danilyan V.I., Gelfan A.N., Motovilov Y.G. and Kalugin A.S. (2014) Disastrous flood of 2013 in the Amur Basin: genesis, recurrence assessment, simulation results. Water Resources., 41(2), 115-125, https://link.springer.com/article/10.1134/S0097807814020055.

Efstratiadis A. and Koutsoyiannis D. (2010) One decade of multi-objective calibration approaches in hydrological modelling: A review. Hydrol. Sci. J., 55(1), 58–78, https://doi.org/10.1080/02626660903526292.

Gupta H.V., Kling H., Yilmaz K.K. and Martinez G.F. (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003.

Harris R.J. (2001) A primer of multivariate statistics. 3nd ed. New York: Psychology Press, https://doi.org/10.4324/9781410600455.

Huynh N.N.T, Garambois P.A., Colleoni F. and Javelle P. Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to Mediterranean floods. J. Hydrol., 625, Part A, 129992, https://doi.org/10.1016/j.jhydrol.2023.129992.

Koch J., Jensen K.H. and Stisen S. (2015) Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. Water Resour. Res., 51, 1225–1246, https://doi.org/10.1002/2014WR016607.

Malik M.A., Dar A.Q., Jain M.K. (2022) Modelling streamflow using the SWAT model and multi-site calibration utilizing SUFI-2 of SWAT-CUP model for high altitude catchments, NW Himalaya's. Model. Earth Syst. Environ., 8, 1203–1213, https://doi.org/10.1007/s40808-021-01145-0.

Mascaro G., Vivoni E.R. and Méndez-Barroso L.A. (2015) Hyperresolution hydrologic modeling in a regional watershed and its interpretation using empirical orthogonal functions. Advances in Water Resources., 83, 190–206, http://dx.doi.org/10.1016/j.advwatres.2015.05.023.

Mohammed S.A., Solomatine D.P., Hrachowitz M. and Hamouda M.A. (2021) Impact of dataset size on the signature-based calibration of a hydrological model. Water., 13, 970, https://doi.org/10.3390/w13070970.

Moreido V., Gartsman B., Solomatine D.P. and Suchilina Z. (2021) How well can machine learning models perform without hydrologists? Application of rational feature selection to improve hydrological forecasting. Water., 13, 1696, https://doi.org/10.3390/w13121696.

Motovilov Yu.G., Gottschalk L., Engeland K. and Belokurov A. (1999) ECOMAG – regional model of hydrological cycle. Application to the NOPEX region. Oslo: Univ. of Oslo, https://studylib.net/doc/11453981/ecomag--regional-model-of-hydrological.

Murphy A. (1988) Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. Monthly Weather Review., 116, 2417–2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Nash J.E. and Sutcliffe J.V. (1970) River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10(3), 282–290, http://dx.doi.org/10.1016/0022-1694(70)90255-6.

Pfannerstill M., Guse B. and Fohrer N. (2014) Smart low flow sig-nature metrics for an improved overall performance evaluation of hydrological models. J. Hydrol., 510, 447–458, https://doi.org/10.1016/j.jhydrol.2013.12.044.

Pokhrel P., Yilmaz K.K. and Gupta H.V. (2012) Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. J. Hydrol., 418-419, 49–60, https://doi.org/10.1016/j.jhydrol.2008.12.004.

Pomerantsev A.L. (2014) Chemometrics in Excel. Hoboken, New Jersey: Wiley, https://doi.org/10.1002/9781118873212.

Reusser D.E., Buytaert W. and Zehe E. (2011) Temporal dynamics of model parameter sensitivity for computationally expensive models with Fourier amplitude sensitivity test. Water Resour. Res., 47, W07551, https://doi.org/10.1029/2010WR009947.

Rodríguez-Iturbe I. and Rinaldo A. (1997) Fractal River Basins: Chance and Self-Organization. Cambrige: Cambrige University Press.

Ruiz-Pérez G., Koch J., Manfreda S., Caylor K. and Francés F. (2017) Calibration of a parsimonious distributed ecohydrological daily model in a data-scarce basin by exclusively using the spatio-temporal variation of NDVI. Hydrol. Earth Syst. Sci., 21, 6235–6251, https://doi.org/10.5194/hess-21-6235-2017.

Serur A.B. and Adi K.A. (2022) Multi-site calibration of hydrological model and the response of water balance components to land use land cover change in a rift valley Lake Basin in Ethiopia. Sci. Afr., 15, e01093, https://doi.org/10.1016/j.sciaf.2022.e01093.

Solomatine D.P., Dibike Y. and Kukuric N. (1999) Automatic calibration of groundwater models using global optimization techniques. Hydrol. Sci. J., 44, 879-894, https://doi.org/10.1080/02626669909492287.

Wagener T., McIntyre N., Lees M.J., Wheater H.S. and Gupta, H.V. (2003) Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrol. Process., 17, 455–476, https://doi.org/10.1002/hyp.1135.

Wang S., Zhang Z., Sun G., Strauss P., Guo J., Tang Y. and Yao A. (2012) Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE Model for a large watershed in northern China. Hydrol. Earth Syst. Sci., 16, 4621–4632, https://doi.org/10.5194/hess-16-4621-2012.

Xu Z., Wu Z., Shao Q., He H., and Guo X. (2022) A two-step calibration framework for hydrological parameter regionalization based on streamflow and remote sensing evapotranspiration, Journal of Hydrology, 613, Part A, 128320, https://doi.org/10.1016/j.jhydrol.2022.128320.

Yilmaz K.K., Gupta H.V. and Wagener T. (2008) A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resour. Res., 44, W09417, https://doi.org/10.1029/2007WR006716.