Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

On the need to test hydrological models under changing conditions

Guillaume Thirel, Vazken Andréassian & Charles Perrin

To cite this article: Guillaume Thirel, Vazken Andréassian & Charles Perrin (2015) On the need to test hydrological models under changing conditions, Hydrological Sciences Journal, 60:7-8, 1165-1173, DOI: 10.1080/02626667.2015.1050027

To link to this article: https://doi.org/10.1080/02626667.2015.1050027

Published online: 14 Aug 2015.

http://dx.doi.org/10.1080/02626667.2015.1050027
Special issue: *Modelling Temporally-variable Catchments*

EDITORIAL

On the need to test hydrological models under changing conditions

Guillaume Thirel, Vazken Andréassian and Charles Perrin

Irstea, Hydrosystems and Bioprocesses Research Unit (HBAN), Antony, France
guillaume.thirel@irstea.fr

Received 5 December 2014; accepted 27 April 2015

Editor D. Koutsoyiannis

Abstract The ability of hydrological models to deal with changing conditions should not be taken for granted: it is an unfortunate but well-known problem of hydrology that the model structure and/or parameters optimized for certain conditions may not be transferable in time. Consequently, it is essential that, for application under changing conditions (e.g. in climate change impact studies), models be thoroughly assessed for their extrapolation capacity using adequate protocols. This editorial provides an overview of the Special Issue of *Hydrological Sciences Journal* compiled after a workshop on this theme held during the General Assembly of the International Association of Hydrological Sciences (IAHS) in Gothenburg (Sweden) in 2013. The Workshop participants had been invited to apply a calibration and evaluation protocol to their own models on a given set of changing basins. The results show that this protocol is an appropriate and instructive way of assessing the suitability of hydrological models to be applied under changing conditions. This special issue also includes papers following alternative testing methodologies, as well as an opinion paper on the definition of non-stationarity.

Key words calibration protocol; hydrology under change; hydrological models; comparison; evaluation

De la nécessité de tester les modèles hydrologiques sous des conditions changeantes

Résumé La capacité des modèles hydrologiques à traiter des conditions changeantes ne devrait pas être considérée comme garantie : c'est un inconvénient bien connu en hydrologie que les modèles (leurs structures et/ou paramètres) optimisés pour certaines conditions peuvent ne pas être transférables dans le temps. Par conséquent, il est essentiel que, pour une application dans des conditions changeantes (par erxemple dans des études d'impact du changement climatique), les modèles soient consciencieusement évalués avec des protocoles adéquats en ce qui concerne leur capacité d'extrapolation. Cet éditorial présente un numéro spécial du Journal des Sciences Hydrologiques organisé après un atelier qui s'est tenu à Göteborg (Suède) lors de l'Assemblée générale de l'Association Internationale des Sciences Hydrologiques (AISH) en 2013. Pour cet atelier, les participants avaient été invités à appliquer un protocole de calage et d'évaluation de leurs propres modèles sur un échantillon de bassins changeants qui leur avait été fourni. Les résultats montrent que ce protocole est une manière appropriée et instructive d'évaluer la pertinence de modèles hydrologiques appliqués dans des conditions changeantes. Ce numéro spécial inclut aussi des articles suivant des protocoles alternatifs ainsi qu'un papier d'opinion sur la définition de la (non-)stationnarité.

Mots clefs protocole de calage ; hydrologie sous conditions changeantes ; modèles hydrologiques ; comparaison ; évaluation

1 INTRODUCTION

1.1 The need for enhanced evaluation of hydrological models under change

Predicting the impact of environmental changes on basins has become a widespread activity for hydrologists, who are, however, increasingly concerned by the fact that the models they use for simulation, forecasting, or projection in impact assessment and as decision-support tools, might not be well suited to deal with change.

The meaning of the word "change" in hydrology is rather imprecise. Here, for river basins, change refers to significant modifications of land cover, or of the statistical characteristics of the climatic conditions of the basin, potentially resulting in modification of catchment behaviour (i.e. in the ability of the catchment to transform precipitation into streamflow). The construction of water management infrastructure, such as storage reservoirs, can also significantly change the streamflow regime and should be included, as well as surface water and groundwater water withdrawal. Obviously, changes occur over various time scales, as discussed by Koutsoyiannis (2013).

We are concerned by change because it may have detrimental consequences on floods (Kundzewicz *et al*. 2013), droughts, and the design, planning and management of water resources systems. Obviously, our models should be able to deal with change consistently. If our models were "perfect" physical representations of complex hydrological systems, changes should not be a problem for them, since known physical laws do not change. However, in essence, a model is an imperfect and simplified representation of a natural system, whatever label we give it (conceptual, black-box, process-oriented) and regardless of whether we have calibrated it "empirically" or assigned its parameter values in a "physical" manner. Thus, models always carry the risk of not reacting to change in an accurate way. This seems particularly obvious when the model does not explicitly account for a given change, though, unfortunately, explicit accounting is not a sufficient guarantee against possible problems without appropriate testing.

1.2 Organization of the workshop

During the 2013 General Assembly of the International Association of Hydrological Sciences (IAHS) in Gothenburg, the Workshop: *Testing simulation and forecasting models in non-stationary conditions* was organized to encourage collaborative work on hydrology under changing conditions. To foster a common reflection on this issue, a dataset of 14 catchments and a modelling protocol were proposed to participants for their model applications, but contributions not based on this common material were also welcomed.

The modellers were invited to apply a calibration and evaluation modelling protocol of their models to the given dataset. Contrasting periods (one complete period and five sub-periods) were selected, on which the models had to be calibrated (Thirel *et al*. 2015). After each calibration, the model had to be run in simulation on each of the other periods. Through the computation of adapted metrics, and plots highlighting the performance and robustness of the models, this protocol intended to provide a framework within which modellers would be able to efficiently test the ability of the modelling options they had chosen to better handle changes.

The modellers were then invited to submit the results of their work to this Special Issue of *Hydrological Sciences Journal* (*HSJ*). Eighteen papers were peer-reviewed and, after revision, accepted for publication, 16 of them presenting modelling results. Among these, 13 papers make use of the proposed dataset and three present modelling results obtained from different basins. More general issues are discussed in an opinion paper (Koutsoyiannis and Montanari 2015) and in the paper presenting the protocols and dataset (Thirel *et al*. 2015).

1.3 Key questions

By proposing a calibration and evaluation protocol on a dataset of catchments for which changes in climate or land use were documented, the aim was to provide a playground for scientists to address the following questions (Thirel *et al*. 2015):

- Can changes in model parameters calibrated over different periods tell us whether a catchment is changing, or are there too many numerical artefacts to answer this question (due to poor parameter identifiability, model overparameterization, etc.)?
- Are our models robust and/or flexible enough to be used under changing conditions?
- What approaches should be tried in the future to better handle hydrological modelling under change?

In the next section, we first present a brief review of studies on hydrology under change, followed by an overview of the Workshop and the papers in this Special Issue of *HSJ*. Lastly, conclusions and perspectives are given.

2 HYDROLOGICAL MODELS UNDER CHANGE: A SHORT OVERVIEW

2.1 What is a good model?

Given the diversity of existing hydrological models, recurrent questions asked by practitioners are: What is the best model for a given objective?

Which model should we rely upon when making decisions?

Coron *et al*. (2011) listed three model parameter "pathologies" that may prevent a model from performing well: dependency on the input data quality and availability (Oudin *et al*. 2006b, Perrin *et al*. 2007), dependency on statistical characteristics of the hydroclimatic data used in calibration (Klemeš 1986, Merz *et al*. 2011), and low identifiability (Abebe *et al*. 2010). These three limitations may be particularly acute when dealing with climate change studies or with historical hydrological reconstructions.

Bearing these elements in mind, it is easier to understand that, in addition to having precision (i.e. closeness of simulations to observations), hydrological models need to be transferable. This transferability (Klemeš 1986) is necessary to ensure that the hydrological models do not overfit the specific conditions of a calibration period (Andréassian *et al*. 2012). The generalization of results is a necessary condition for applying models to climate change applications.

2.2 Model intercomparison as a tool for improvement

Kundzewicz and Gerten (2014) recently pointed out that hydrologic model intercomparisons are an important challenge in the assessment of the impact of climate change. A few such intercomparisons have already been carried out with global hydrological models (GHMs). Haddeland *et al*. (2011) made a comparison between several land surface models and global hydrological models (WaterMIP project). Schewe *et al*. (2014) applied 11 GHMs and five global circulation models (GCMs; ISI-MIP project). Both studies showed that the spread (i.e. uncertainty) of hydrological projections due to GHMs is a bit larger than that due to GCMs.

Comparative studies have also been carried out with catchment-based hydrological models, ranging from the pioneering work led by the World Meteorological Organization (WMO 1975, 1986, 1992) to the more recent comprehensive analyses proposed by Smith *et al*. (2004, 2012), or Nicolle *et al*. (2014). The high dependency of models on the climate of the calibration period was observed by Viney *et al*. (2009) with several hydrological models. Model intercomparisons are definitely an efficient way to identify model flaws, to understand them and to propose improvements based on these observations.

2.3 Preliminary solutions for evaluating or improving models for application in changing conditions

To better handle the modelling issues highlighted above, several options are possible: improving data quality, improving the hydrological models and their parameterization, performing post-treatment on the hydrological outputs, developing multi-model approaches, or combinations of these. Quantifying the dependency of models on the climate conditions of the calibration period, or on the land-use conditions, and subsequently evaluating innovative solutions to reduce these dependencies requires adequate calibration and evaluation protocols.

Although this assertion may seem obvious, only a few studies have described such protocols (WMO 1975, Smith *et al*. 2004). Others, like the study by Seiller *et al*. (2012), assessed the temporal transposability of 20 hydrological models under contrasting climate conditions with the differential split-sample test (DSST; Klemeš 1986). The models were assessed separately and together. The transposability of models over contrasting periods was often rather poor. However, the average simulation of the 20-model ensemble was shown to be more transposable. Some models outperformed this average, but only for one transposition between two periods, or only for a single catchment, meaning that single models were actually less robust. Even badly performing single models were shown to be able to positively contribute to the ensemble's quality. This work showed two things: solutions exist to overcome the limitations of models, and it is necessary to continue improving our models. Similarly, Nicolle *et al*. (2014) used five hydrological models following an evaluation framework for low-flow simulation and forecasting purposes. They showed that no model outperformed the others, and that a multi-model averaging provides more robust results and better performance.

Different options are currently being explored to improve the applicability of hydrological models to changing conditions. First, better identification of the optimum values of parameters should allow better transfer to contrasting periods (Gupta *et al*. 1998, 2009). Multi-objective calibration is a possible solution to transferability issues, because it allows one to take into account the hydrological response regarding different aspects at the same time (Efstratiadis and Koutsoyiannis 2010). Another path investigated in recent years consists in combining hydrological models that are optimized for opposite/complementary objectives. This represents a kind of selective ensemble approach, which we could expect to perform better than classic averaging ensemble methods. Both Kayastha *et al*. (2013) and Oudin *et al*. (2006a) calibrated a single model twice: once on low flows, and then on high flows. Then, according to the given regime, one model or the other, or a combination of both, was chosen for flow simulation. Both studies showed that this method yields better results than those obtained by individual models.

3 OVERVIEW OF THE *HSJ* SPECIAL ISSUE

3.1 The general papers

In their opinion paper, Koutsoyiannis and Montanari (2015) point out the misuse of the expression "non-stationarity", which is very often wrongly understood as a synonym of change. They state that "*stationarity and nonstationarity apply only to models, not to the real world, and are defined within stochastics*", and that non-stationarity should be used only to describe a future that can be predicted in deterministic terms.

The paper by Thirel *et al*. (2015) presents the detailed objectives of the workshop, the calibration and evaluation protocol and the graphical tools proposed to evaluate the model performances. A detailed description of the dataset is also given in the Supplementary material of Thirel *et al*. (2015).

3.2 The 16 papers presenting modelling results

Most participants used lumped models (see Table 1), and only a few are sub-basin-based or grid-based models. The number of free parameters varies greatly between models, ranging from 1 to 95. All models but three were used at a daily time step, which corresponds to the data time step provided to the modellers. The monthly, annual and multi-annual time steps chosen for three of the studies are of interest for water balance questions, and were mainly chosen, unsurprisingly, for basins facing a decrease in rainfall and, consequently, streamflow. Regarding the modelling choices, conceptual models dominate, followed by stochastic models, process-oriented models, artificial neural networks, water balance models and kinematic wave models. These choices reflect the diversity of modelling approaches in hydrology. A summary of the key points of the papers included in this special issue is presented in Table 1.

3.3 Exposing the models to new conditions

While the general aim of the modellers was to expose their model to changing conditions and to assess its transferability to contrasting periods, some modellers fulfilled a second aim: applying their favourite model under conditions they generally do not face, typically in basins from different countries. Indeed, basins from six countries and four continents were proposed to the modellers. For example, Hughes (2015) applied the Pitman model to Australian and Sahelian basins, although it is mostly applied to South African basins (Hughes 2013). Yu and Zhu (2015) applied two models (AWBM and SimHyd), usually employed in Australia, to basins from France and the USA. Tanaka and Tachikawa (2015) applied the DHM-KWMSS model to two basins from France and Australia, although it is generally applied to Japanese basins.

This spatial transferability of the models, while representing here an extreme view of the application of the proxy-basin test of Klemeš (1986), can be seen as a further validation of the modelling choices made by the hydrologists.

3.4 Original calibration protocols

While the calibration and evaluation protocol proposed for the workshop (Thirel *et al*. 2015) was generally used, some modellers used other protocols. Magand *et al*. (2015) modified the periods of calibration and evaluation to run the distributed CLSM model because of computer time constraints. Brigode *et al*. (2015) used a block boot-strapping method in addition to the proposed protocol. This method, by generating more variability among calibration periods, is interesting for obtaining more robust parameters (Ebtehaj *et al*. 2010, Li *et al*. 2010, Brigode *et al*. 2014). Gelfan *et al*. (2015) applied a protocol originally proposed by Kuchment and Gelfan (2009) that makes use of periods of increasing lengths. The objective is to identify a period length for which there is no more improvement when adding data during the calibration process. Coron *et al*. (2015) chose a visualization of their results on sliding windows. This way of viewing results can be connected to the calibration protocol introduced by Coron *et al*. (2012) on sliding

Table 1 Summary of the papers in this Special Issue presenting modelling results. The number of free parameters, spatial resolution and model time steps detailed here are those used in the works reported herein; other configurations may exist for the tested models. (stoch.: stochastic; det.: deterministic; simul.: simulation).

Reference	Basin(s)	Model(s)	Free parameters	Type of model	Spatial resolution	Model time step	Comments
Yu and Zhu (2015)	Rivers Rimbaud and Fernow	AWBM and SimHyd	7–8	Conceptual rainfall–runoff	Lumped	Daily	–
Folton *et al*. (2015)	River Rimbaud	GR4 J	4	Conceptual rainfall–runoff	Lumped	Daily	Uses a streamflow–streamflow model
Semenova *et al*. (2015)	A Russian basin	Hydrograph	36	Process-oriented	Scale-dependent grid	Daily	–
Taver *et al*. (2015)	Rivers Fernow and Durance	Feed-forward and recurrent multilayer perceptron	~70 (Fernow) and 95 (Durance)	Artificial neural network	Lumped	Daily	Uses data assimilation
Gelfan *et al*. (2015)	River Garonne and Obyån Creek	ECOMAG	20	Process-oriented	Sub-basin	Daily	–
Hughes (2015)	Axe Creek, Wimmera, Bani, and South African basin	Pitman	9	Conceptual rainfall–runoff	Lumped	Monthly	–
Osuch *et al*. (2015)	Axe Creek, Rivers Kamp and Wimmera and a Polish basin	HBV	14	Conceptual rainfall–runoff	Lumped	Daily	–
Li *et al*. (2015)	River Wimmera and a Chinese basin	SimHyd, HBV and Xinanjiang	9, 13 and 15	Conceptual rainfall–runoff	Lumped	Daily	–
Baahmed *et al*. (2015)	An Algerian basin	Schreiber	1	Water balance	Lumped	Pluri-annual	Discharge decrease
Zhou *et al*. (2015)	An Australian basin	AWRA-L, XAJ and GR4 J	17, 14 and 4	Conceptual rainfall–runoff	Grid-based or lumped	Daily	Impact of bushfires
Tanaka and Tachikawa (2015)	Rivers Garonne and Flinders	DHM-KWMSS	5–6	Rainfall–runoff based on kinematic wave flow	Grid-based	Daily	–
Kling *et al*. (2015)	11 (all but the American basins)	COSERO	8 (Europe) or 7 (elsewhere)	Conceptual rainfall–runoff	Lumped (semi-distr. snow model)	Daily	–
Coron *et al*. (2015)	Rivers Fernow, Allier, Garonne and Bani	Turc–Budyko	1	Water balance	Lumped	Annual	–
Magand *et al*. (2015)	River Durance	CLSM	4	Process-oriented	Sub-basin	Daily	–
Brigode *et al*. (2015)	River Kamp	MORDOR (+ SCHADEX)	22 (+ ~ 20)	Conceptual rainfall–runoff (within stoch. flood simul. framework)	Lumped	Daily	–
Efstratiadis *et al*. (2015)	Ferson Creek	Hydrogeios + Castalia	16	Nonlinear stoch. (det. conceptual model + stoch. models for inputs and errors)	2 HRUs with time-varying areas	Daily	Tested in calibration and stochastic simulation

windows, which was used by Folton *et al*. (2015). This methodology enables one to visualize long-term evolutions. Osuch *et al*. (2015) similarly chose to use calibration moving windows.

3.5 Giving physical meaning to parameter values

Even with conceptual models, trying to link parameter values with meteorological or physical basin characteristics is tempting, and is often seen as an opportunity to improve the consistency of models. Having basins with significant changes was a good opportunity for modellers to perform such an analysis. Unfortunately, it is generally difficult to understand these relationships because of the interactions between model parameters, or equifinality (Gelfan *et al*. 2015, Hughes 2015, Yu and Zhu 2015, Zhou *et al*. 2015). However, Magand *et al*. (2015) managed to link the snow-related parameters of their model to the actual snow conditions on the specific contrasting period for the mountainous Durance catchment. More specifically, the parameter used in the model snow depletion curve (with larger values indicating a more heterogeneous snow cover during ablation) was found to be higher during the coldest and snowiest sub-period, resulting in a longer snow cover being simulated by the model. Osuch *et al*. (2015) tried to link the parameters of the HBV model with climatic indices. They showed that the FC parameter (i.e. the maximum soil moisture storage) could be significantly correlated with temperature-based indices, whereas the β parameter (the nonlinear runoff parameter) could be correlated with precipitation-related indices, the other HBV parameters showing no strong correlation. Semenova *et al*. (2015) did not automatically calibrate their model parameters, but used instead parameter values from the literature corresponding to post-fire conditions.

Introducing prior knowledge into model parameterization could also lead to great improvement in the identification of parameters (Gelfan *et al*. 2015).

3.6 Comparing different models

Several modellers performed their own model comparisons: Yu and Zhu (2015) compared AWBM and SimHyd; Li *et al*. (2015) compared SimHyd, HBV and Xinanjiang; Kling *et al*. (2015) compared COSERO and MORDOR6; and Zhou *et al*. (2015) compared AWRA-L, XAJ and GR4J. Even though some models exhibited better (or worse) performance than others (see Li *et al*. 2015), the main factor influencing the performance and robustness of a model was the studied basin rather than the selected model.

Moreover, two papers compared different modelling options in order to assess the choices made and to answer the following question: If the model is "improved", does its robustness increase? Taver *et al*. (2015) chose an original approach using artificial neural networks (ANN). They tested different options specifically designed to better handle change (adaptivity and data assimilation) and compared them with classical ANN approaches, confirming their better performance. Efstratiadis *et al*. (2015) tried to make the most of a combination of a deterministic approach, parameterized to take into account the growing urbanization of the basin, and a stochastic approach. They showed that the proposed framework can effectively account for systematic changes in water resources systems and help to better represent the variability of the catchment response.

3.7 The case of semi-arid catchments

Among all the proposed catchments, the semi-arid catchments, namely the Australian basins that faced the "Millennium Drought", or the Sahelian River Bani that faced a severe drought, were probably the most challenging for modellers. These basins represent areas among the most affected by climate change and human activities (He and Hogue 2012, Yang *et al*. 2013, Zeroual *et al*. 2013). However, it is known that the model parameters are usually oversensitive to the choice of calibration period: a model calibrated over a wet period usually overestimates discharge during a drier period (e.g. Hughes 2015, Kling *et al*. 2015, Osuch *et al*. 2015).

Such model failures may come from a modification of the processes involved in discharge generation. As a consequence, a solution could be to adapt the models to these different conditions by including different processes in the model implementation. For example, Kling *et al*. (2015) tried an alternative version of the COSERO model, removing the distinction between surface flow and interflow while adding routing, in order to better take into account the specificities of such basins. For another type of change —urbanization—Efstratiadis *et al*. (2015) directly took the urbanization rate into account in their modelling. The difficulty in accurately dealing with semi-arid catchments could also be related to the absence of water abstraction data, which consequently could not be taken into account, as noted by Hughes (2015).

3.8 Models as tools for understanding hydrological changes

Baahmed *et al*. (2015) applied an annual water balance model to an arid Algerian basin facing decreasing rainfall and discharge since the mid-1970s. They showed that, although a simple water balance model facilitates the understanding of the decrease in discharge, other factors, such as water withdrawals, could induce the decrease in discharge. Zhou *et al*. (2015) applied three conceptual models to three Australian catchments affected by bushfires. This type of change is present in the Workshop dataset with the Rimbaud catchment (see the Supplementary materials of Thirel *et al*. 2015). Zhou *et al*. (2015) used two periods: before the bushfires, and after. They showed that the discharge variations could be explained by the bushfires up to 15 years after their occurrence and by the climate afterwards. Semenova *et al*. (2015) worked on a Russian basin also affected by both wildfire and permafrost. The study revealed an increase in summer flows from the upstream part of the basin that was not observed in the downstream part, as well as an increase in thaw depth and surface flow, and a decrease in total evapotranspiration during summer. Folton *et al*. (2015) used two alternative methods to try to disentangle the causes of apparent changes: with a rainfall–runoff model, they accounted for climatic variability only, while with the paired-catchment model, they accounted for both climatic and catchment behaviour changes. The results are rather surprising and showed that the prolonged drought experienced by the Rimbaud catchment had a more significant hydrological impact than the wildfire (at least on daily discharge values).

4 CONCLUSIONS AND PERSPECTIVES

A calibration and evaluation protocol based on contrasting periods was proposed to modellers on a set of catchments during a Workshop held during the 2013 IAHS General Assembly. A number of papers related to this Workshop (one presentation paper, one opinion paper about the definition of non-stationarity and 16 modelling papers) are presented in this Special Issue of *HSJ*. The need for protocols was highlighted by the wide use of the proposed protocol to compare models or to confront them with new and changing conditions, but also by the fact that additional protocols were proposed by the participants. Understanding hydrological change or giving physical meaning to parameter values seem to be eased by adequate protocols. However, this set of papers shows that efforts are still needed to properly undertake hydrological modelling under changing conditions, especially under semi-arid conditions.

During the Workshop, four working groups were organized. Several issues and ways forward were discussed regarding hydrological modelling under change. First, the issue of model calibration and ways to improve it were mentioned, including:

- the use of data with multiple spatial and temporal scales to perform calibration;
- the use of approaches other than the differential split-sample test to split the observation data series (e.g. the application of bootstrapping techniques to obtain multiple test sub-periods);
- the use of diverse sets of parameters adapted to different hydrological answers;
- the importance of adding further constraints for model calibration; and
- the use of adequate protocols and visual tools to calibrate and validate models.

All the above solutions follow one main goal: increasing the model robustness.

A second major concern involves the improvement of our process understanding. It is clear that a more explicit accounting of groundwater would improve the surface model performances. Similarly to groundwater, evapotranspiration is a second limiting condition strongly affecting flow generation. Constraining the actual evapotranspiration by catchment water balance methods and using additional data sources (fluxnet point data, satellite data, biomass) may bring more robustness. Robustness and accuracy are two expected model qualities. Their improvement should logically positively influence the reliability of models under change and the confidence we have in them. Better exploiting extreme climate conditions could also help to define model limitations. The use of large-sample hydrology is one research area to be explored for advancing these topics (Gupta *et al*. 2014).

Despite the obvious positive impact brought by a better process understanding of basins, and its subsequent better representation, the more accurate representation of two sources of change have to be considered as the major challenge for improving model predictions under changing conditions. First, accounting for human influences in hydrological models is still a real challenge: human influences are not well-documented, access to data—when available—is often a problem, and modelling these influences is not straightforward. Second, natural variations and changes in climate conditions are still difficult to anticipate and, as such, they represent a major source of uncertainty for hydrological predictions or projections.

Lastly, handling hydrological processes within stochastic frameworks and better representing sources of uncertainty are likely to be key approaches to further improve hydrological models in the future.

Acknowledgements We wish to thank all the modellers for participating in this modelling exercise on a voluntary basis (without specific funding), all workshop attendees for the fruitful comments and interesting discussions they generated, the data providers, IAHS for its support in organizing the workshop, the *HSJ* Editors for offering the opportunity to publish this Special Issue, and the IAHS editorial office for its efficiency and support in preparing the Special Issue. The present work is a contribution to the Panta Rhei decade coordinated by IAHS. The authors also thank Andreas Efstratiadis, Ioannis Nalbantis and another anonymous reviewer for their comments, which helped improving the quality of this manuscript.

Disclosure statement No potential conflict of interest was reported by the authors.

REFERENCES

Abebe, N.A., Ogden, F.L., and Pradhan, N.R., 2010. Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: implications for parameter estimation. *Journal of Hydrology*, 389 (3–4), 301–310. doi:10.1016/j.jhydrol.2010.06.007

Andréassian, V., *et al*., 2012. All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes*, 26 (14), 2206–2210. doi:10.1002/hyp.9264

Baahmed, D., Oudin, L., and Errih, M., 2015. Current runoff variations in the Macta catchment (Algeria): is climate the sole factor? *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.975708

Brigode, P., *et al*., 2014. Sensitivity analysis of SCHADEX extreme flood estimations to observed hydrometeorological variability. *Water Resources Research*, 50 (1), 353–370. doi:10.1002/2013WR013687

Brigode, P., *et al*., 2015. Dependence of model-based extreme flood estimation on the calibration period: case study of the Kamp River (Austria). *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2015.1006632

Coron, L., *et al*., 2011. Pathologies of hydrological models used in changing climatic conditions: a review. *In*: S.W. Franks, *et al*., eds. *Hydro-climatology: variability and change*. Wallingford International Association of Hydrological Sciences, IAHS Publ. 344, 39–44.

Coron, L., *et al*., 2012. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. *Water Resources Research*, 48 (5). doi:10.1029/2011wr011721

Coron, L., *et al*., 2015. Graphical tools based on Turc-Budyko plots to detect changes in catchment behaviour. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.964245

Ebtehaj, M., Moradkhani, H., and Gupta, H. V., 2010. Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling. *Water Resources Research*, 46 (7). doi:10.1029/2009WR007981

Efstratiadis, A. and Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal*, 55 (1), 58–78. doi:10.1080/02626660903526292

Efstratiadis, A., Nalbantis, I., and Koutsoyiannis, D., 2015. Hydrological modelling of temporally-varying catchments: facets of change and the value of information. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.982123

Folton, N., Andréassian, V., and Duperray, R., 2015. Hydrological impact of forest-fire from paired-catchment and rainfall–runoff modelling perspectives. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2015.1035274

Gelfan, A., *et al*., 2015. Testing robustness of the physically-based ECOMAG model with respect to changing conditions. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.935780

Gupta, H.V., Sorooshian, S., and Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research*, 34 (4), 751–763. doi:10.1029/97WR03495

Gupta, H.V., *et al*., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377 (1–2), 80–91. doi:10.1016/j.jhydrol.2009.08.003

Gupta, H.V., *et al*., 2014. Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18 (2), 463–477. doi:10.5194/hess-18-463-2014

Haddeland, I., *et al*., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology*, 12 (5), 869–884. doi:10.1175/2011JHM1324.1

He, M. and Hogue, T.S., 2012. Integrating hydrologic modeling and land use projections for evaluation of hydrologic response and regional water supply impacts in semi-arid environments. *Environmental Earth Sciences*, 65 (6), 1671–1685. doi:10.1007/s12665-011-1144-3

Hughes, D.A., 2013. A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall–runoff model. *Journal of Hydrology*, 501, 111–124. doi:10.1016/j.jhydrol.2013.07.043

Hughes, D.A., 2015. Simulating temporal variability in catchment response using a monthly rainfall–runoff model. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.909598

Kayastha, N., *et al*., 2013. Fuzzy committees of specialized rainfall–runoff models: further enhancements and tests. *Hydrology and Earth System Sciences*, 17 (11), 4441–4451. doi:10.5194/hess-17-4441-2013

Klemeš, V., 1986. Operational testing of hydrological simulation-models. *Hydrological Sciences Journal*, 31 (1), 13–24. doi:10.1080/02626668609491024

Kling, H., *et al*., 2015. Performance of the COSERO precipitation–runoff model under non-stationary conditions in basins with different climates. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.959956

Koutsoyiannis, D., 2013. Hydrology and change. *Hydrological Sciences Journal*, 58 (6), 1177–1197. doi:10.1080/02626667.2013.804626

Koutsoyiannis, D. and Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.959959

Kuchment, L. and Gelfan, A., 2009. Assessing parameters of physically-based models for poorly gauged basins. *In*: K.K. Yilmaz, et al., eds. *New approaches to hydrological prediction in data-sparse regions*. Wallingford: International Association of Hydrological Sciences, IAHS Publ. 333, 3–10.

Kundzewicz, Z. and Gerten, D., 2014. Grand challenges related to the assessment of climate change impacts on freshwater resources. *Journal of Hydrologic Engineering*, 20, A4014011. doi:10.1061/(ASCE)HE.1943-5584.0001012

Kundzewicz, Z.W., *et al.*, 2013. Flood risk and climate change: global and regional perspectives. *Hydrological Sciences Journal*, 59 (1), 1–28. doi:10.1080/02626667.2013.857411

Li, H., Beldring, S., and Xu, C.-Y., 2015. Stability of model performance and parameter values on two catchments facing changes in climatic conditions. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.978333

Li, Z., *et al.*, 2010. Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: a case study of SWAT model applied to Yingluoxia watershed in northwest China. *Journal of Hydrology*, 385 (1–4), 76–83. doi:10.1016/j.jhydrol.2010.01.025

Magand, C., *et al.*, 2015. Parameter transferability under changing climate: case study with a land surface model in the Durance watershed, France. *Hydrological Science Journal*, 60 (7–8). doi:10.1080/02626667.2014.993643

Merz, R., Parajka, J., and Blöschl, G., 2011. Time stability of catchment model parameters: implications for climate impact analyses. *Water Resources Research*, 47 (2). doi:10.1029/2010WR009505

Nicolle, P., *et al.*, 2014. Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. *Hydrology and Earth System Sciences*, 18 (8), 2829–2857. doi:10.5194/hess-18-2829-2014

Osuch, M., Romanowicz, R.J., and Booij, M.J., 2015. The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.967694

Oudin, L., *et al.*, 2006a. Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations. *Water Resources Research*, 42 (7). doi:10.1029/2005wr004636

Oudin, L., *et al.*, 2006b. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology*, 320 (1–2), 62–83. doi:10.1016/j.jhydrol.2005.07.016

Perrin, C., *et al.*, 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall–runoff models. *Hydrological Sciences Journal*, 52 (1), 131–151. doi:10.1623/hysj.52.1.131

Schewe, J., *et al.*, 2014. Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, 111 (9), 3245–3250. doi:10.1073/pnas.1222460110

Seiller, G., Anctil, F., and Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrology and Earth System Sciences*, 16 (4), 1171–1189. doi:10.5194/hess-16-1171-2012

Semenova, O., *et al.*, 2015. Detecting immediate wildfire impact on runoff in a poorly-gauged mountainous permafrost basin. *Hydrological Sciences Journal*, 60, 7–8. doi:10.1080/02626667.2014.959960

Smith, M.B., *et al.*, 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. *Journal of Hydrology*, 298 (1–4), 4–26. doi:10.1016/j.jhydrol.2004.03.040

Smith, M.B., *et al.*, 2012. The distributed model intercomparison project—Phase 2: motivation and design of the Oklahoma experiments. *Journal of Hydrology*, 418–419, 3–16. doi:10.1016/j.jhydrol.2011.08.055

Tanaka, T. and Tachikawa, Y., 2015. Testing the applicability of a kinematic wave-based distributed hydrological model in two climatically contrasting catchments. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.967693

Taver, V., *et al.*, 2015. Feed-forward *vs* recurrent neural network models for non-stationarity modelling using data assimilation and adaptivity. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.967696

Thirel, G., *et al.*, 2015. Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.967248

Viney, N.R., *et al.*, 2009. Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions. *Advances in Water Resources*, 32 (2), 147–158. doi:10.1016/j.advwatres.2008.05.006

WMO, 1975. *Intercomparison of conceptual models used in operational hydrological forecasting*. Geneva: World Meteorological Organization, Operational Hydrology Report No. 7, WMO No. 429.

WMO (World Meteorological Organization), 1986. *Intercomparison of models of snowmelt runoff*. Geneva: World Meteorological Organization, Operational Hydrology Report No. 23, WMO No. 646.

WMO, 1992. *Simulated real-time intercomparison of hydrological models*. Geneva: World Meteorological Organization, Operational Hydrology Report No. 38, WMO No. 779.

Yang, X., *et al.*, 2013. Hydrological response to land-use changes in a semi-arid region in China. *In*: E. Boegh, *et al.*, eds. *Climate and land surface changes in hydrology.* Wallingford: International Association of Hydrologial Sciences, IAHS Publ. 359, 397–402.

Yu, B. and Zhu, Z., 2015. A comparative assessment of AWBM and SimHyd for forested watersheds. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.961924

Zeroual, A., Meddi, M., and Bensaad, S., 2013. The impact of climate change on river flow in arid and semi-arid rivers in Algeria. *In*: E. Boegh, et al., eds. *Climate and land surface changes in hydrology.* Wallingford: International Association of Hydrologial Sciences, IAHS Publ. 359, 105–110.

Zhou, Y., *et al.*, 2015. Impact of bushfire and climate variability on streamflow from forested catchments in southeast Australia. *Hydrological Sciences Journal*, 60 (7–8). doi:10.1080/02626667.2014.961923