Pathologies of hydrological models used in changing climatic conditions: a review

L. CORON[1,2], V. ANDRÉASSIAN[2], M. BOURQUI[1], C. PERRIN[2] & F. HENDRICKX[1]

1 *EDR R&D, Chatou, France*
laurent.coron@edf.fr

2 *Cemagref, Antony, France*

Abstract Characterizing the impact of climate change on hydrology is not as simple as feeding a previously calibrated hydrological model with future climate scenarios. Nevertheless, hydrological modelling is often considered as a small contributor to the overall uncertainty in climate change impact studies. Running a model under conditions that can be significantly different from those used for calibration raises questions relative to the actual extrapolation capacity of the model. As hydrological models (as complex as they may be) are always a simplification of reality, they can never fully integrate all aspects of the rainfall–runoff relationship. Consequently, we prefer to consider them as patients that can certainly be in good health in average conditions, but may also be affected with pathologies when exposed to unusual conditions (namely conditions they have not been properly trained or structured for). Focusing on the robustness issues linked with non-stationary climatic conditions, this paper reviews some of the typical pathologies rainfall–runoff models can suffer from when asked to predict discharges under climate conditions different from the calibration ones.

Key words hydrological model; climate change; model calibration; parameter stability

INTRODUCTION

Characterizing the impact of climate change on hydrology is an increasing field of study throughout the world. However, this task is sensitive to the numerous uncertainties induced by the complexity of the modelling chain (emissions scenarios → global circulation models → downscaling techniques → rainfall–runoff models). Several authors agree that the major sources of uncertainty are the first steps of the modelling chain (mainly emissions scenarios and climate models) and that hydrological modelling is a lower contributor to total uncertainty (Wilby, 2005; Wilby & Harris, 2006; Kay *et al*., 2009; Prudhomme & Davies, 2009). Whether or not progress is made by climate modellers to improve the estimation of future climatic forcings, it is our responsibility as hydrologists to provide trustworthy simulations when a rainfall–runoff (RR) model is run under conditions that may be significantly different to those used for calibration (e.g. future *vs* current conditions). Indeed, many unknowns still remain concerning the actual climate extrapolation capacity of hydrological models. In this paper, we review the main robustness issues of RR models which are of importance when models are used in a climate change context.

The ideal case: the healthy model

We will not discuss in detail here how an ideal model should look (number of modules, stores, spatial representation of inputs, etc.). Let us just mention that such a model should be able to integrate correctly the various aspects of the rainfall–runoff relationship, and that it should then be able to simulate correct discharges when provided with correct inputs. Within the context of this article, one should insist on this latter aspect by adding that the model performances should remain good at the application stage, however climatically-different the simulation period may be.

The hard reality: models are full of simplifications that make them dependent on the calibration period

Due to the lack of knowledge on the true functioning of the hydrological system, the lack of data for model construction and the complexity of measuring each process within the hydrological sphere, a RR model incorporates large simplifications (Murphy *et al*., 2006). All hydrological models remain at some point conceptual and empirical. Thus, their parameters require manual or automatic calibration. One would ideally like to measure these parameters in the field, but due to the model conceptualisation, direct use of field parameter estimations is rarely possible in practice and calibration remains necessary to reduce bias (Merz *et al.*, 2011). Moreover, whichever approach is used, the estimated parameters may not be the "true" representative parameters of the processes in the catchment, leading to uncertainties in the prediction of the model variable (Abebe *et al.*, 2010). As discussed below, model parameters may indeed be affected to various extents by various elements during calibration, which may divert them from their correct values, i.e. those values that would be optimal over the long term.

A REVIEW OF COMMON PATHOLOGIES

Dependency of model parameters on the inputs quality and availability

Quite logically, the quality of a RR model is highly dependent on the inputs it was fed with during calibration. Therefore, incorrect estimation of the input or a too short calibration period may affect the model parameter determination and thus bias simulations.

Several studies have assessed the impact of input quality on the optimal parameter set obtained from calibration. Among them, one can mention the work of Andréassian *et al.* (2001, 2004) and Oudin *et al.* (2006). They studied the impact that random and systematic errors in rainfall and potential evapotranspiration (PE) estimates can have on model performance and parameter values for simple RR models. They showed how the models can sometimes use their free parameters to smooth input errors, without decreasing the quality of simulation. Such conclusions are in agreement with the findings of Kokkonen & Jakeman (2001) who worked on the impact of PE estimates on model performance.

Apart from errors in input rainfall depths, several works were made on the effect of incorrect estimation of spatial variability (e.g. poor coverage of the basin with a gauging network), as reviewed by Brath *et al.* (2004). These authors also carried out some tests on the influence of reducing the rainfall gauging network for a distributed model. Interestingly, they found that model performance did not noticeably decrease when fed with a spatially uniform rainfall, as long as the network size remained sufficient to measure the correct amount of overall rainfall. These results contrast with the previous findings of Chaubey *et al.* (1999) who observed a decrease in performance of the Agricultural Non-Point Source pollution model (AGNPS) and a higher variability in the estimated parameters when the spatial variability of rainfall was deteriorated.

The dependency of model parameters on input quality may also exist for inputs other than climate forcings, as illustrated by Apaydin *et al.* (2006). These authors studied the transferability of the parameters of the SLURP distributed model between two temporally spaced periods. They observed a greater performance loss when transferring model parameters from the oldest to the most recent period than the contrary. They attributed this finding to the wider range of precipitation, but also to the better distribution of land cover data in the recent period. This yielded a more accurate estimation of the model parameters.

Finally, the length of the calibration period may significantly affect model calibration if it is too short. A large number of studies were carried out on this topic and were reviewed by Perrin *et al.* (2007). The general rule about calibration length is to have climatic and flow conditions sufficiently diverse to give a representative picture of their natural variability, thus allowing for an exhaustive activation of the hydrological processes at work in the basin. Beyond this statement, very few recommendations exist on the minimum length required for calibration. Indeed, depending on the studies and models used, this length may vary from two to ten years (Yapo *et al.*, 1996; Anctil *et al.*, 2004; Brath *et al.*, 2004; Perrin *et al.*, 2007).

Dependency of model parameters on the climate of the calibration period

Once the free parameters have been calibrated, the model structure should ideally provide a good simulation of the rainfall–runoff relationship. Therefore, it should be able to simulate discharges without significant performance loss, when fed with other climate inputs. In reality, model parameters may sometimes be dependent on the climate they have experienced during the calibration period. However, one must say there is currently no consensus in the literature on this dependency or on the impact it may have on the model simulations.

Merz *et al*. (2011) calibrated the semi-distributed HBV model on a series of 5-year periods between 1976 and 2006 for 273 Austrian catchments. They showed how the parameters representing snow and soil moisture processes could vary depending on the calibration period. They established a link between these variations and climatic characteristics such as air temperature and potential evapotranspiration. Similarly, Wagener *et al*. (2003) found in their test using the DYNIA approach (see next section) that the model parameter that controls rapid groundwater recharge or rapid surface runoff was dependent on the calibration climate. Furthermore, they illustrated how this parameter had two optima both with relatively high identifiability: small values were required during summer periods whereas high values were needed during storm events.

With regard to these findings, questions can be raised on the validity of using a model to simulate discharges under climatic conditions that are different from the calibration ones. Several authors investigated the topic carrying out various tests derived from the differential split-sample test proposed by Klemeš (1986). One can mention the work of Wilby (2005) on uncertainties in climate change impact studies. He found that the projections' uncertainty due to the choice of the calibration period could be of the same order as the uncertainty due to greenhouse gas emission scenarios. He concluded that the transferability of model parameters was dependent on the representativeness of the calibration period. Also working on this issue of climatic parameter transferability, Vaze *et al*. (2010) calibrated four commonly-used models (SIMHYD, Sacramento, MARG, IHACRES) on 61 catchments in southeast Australia over the wettest and driest 10, 20, 30 and 40 years and simulated the other periods. They illustrated how the performance decrease and bias increase between calibration and validation could be related to the difference in annual rainfall. They concluded that models calibrated on average or wet conditions had more difficulty in simulating dry periods than the other way round. Finally, one can mention the study by Choi & Beven (2007), who used the Generalized Likelihood Uncertainty Estimator (GLUE) approach to evaluate TOPMODEL over one basin in South Korea. After sampling the time series into several clusters according to their hydrological similarities, they found that parameters that were optimal for some clusters were not convenient for use on others (particularly those obtained on the dry clusters, which contrasts with the findings of Vaze *et al*. (2010)).

However, some authors have drawn conclusions regarding the ability of hydrological models calibrated in specific climatic conditions to be used in others. This is the case for Chiew *et al*. (2009) who calibrated the model SIMHYD over a long period and used it to simulate discharges on particularly wet and dry periods. The authors concluded on the suitability of RR models for climate change impact studies when they are calibrated against a sufficiently long period. However, clear decreases of model performance were observed in some cases between calibration and simulation, and the authors considered the subject as worthy of further investigation. One can also mention the case of Niel *et al*. (2003), who studied 17 African catchments and noted time stable parameters for about two thirds of the basins, and found no obvious link between parameter values and climate for the last third, although rainfall and runoff significantly decreased over the years.

Although no general consensus has been reached yet, and despite the differences in model sensitivity, it seems that parameters cannot be directly transferred from one period to another without paying attention to the climatic differences. Obviously this applies to all climate change impact studies.

Low identifiability of parameter values

The identifiability level of a parameter expresses how well the parameter is defined within a model structure (Abebe *et al*., 2010). This level is high if changes in the parameter value have significant effects on the simulated discharges and thus on the calibration criteria, allowing an accurate estimation of the parameter correct value. Conversely, this level can be low if it has only little effect on the discharges. Potentially, a random value could then be considered as optimal. Methods to analyse parameters identifiability are usually based on a Monte Carlo approach. Wilby (2005) carried out such an analysis on the CATCHMOD model by testing random parameter sets and looking at the evolution of the simulation's quality relative to individual parameter values. Another method proposed to investigate this question is the Dynamic Identifiability Analysis (DYNIA) of Wagener *et al*. (2003) that aims to identify the most informative parts of the hydrograph for the calibration of each model parameter. Applying this method on a five-parameter RR model, the authors showed that some parts of the hydrograph (recession limbs, rainy periods, etc.) contain more information for the identification of some parameters.

Low identifiability may happen in the case of structural problems in the model, when a parameter does not have a precise role in the model functioning (in spite of the modeller's intention), or when several parameters interact and compensate each other (Wilby, 2005; Abebe *et al*., 2010). However, low identifiability issues are generally not independent from the two "pathologies" mentioned above. Indeed, insufficient input quality or climate variability (linked with calibration length) will induce problems on the determination of the model parameters.

The low identifiability of a parameter may or may not have an impact on model performance: it will have no impact if the process represented by the parameter is insignificant in the case study (e.g. the parameters of a snow module on a basin without snow influence). But if the process in question is significant in some periods and not in others, then it may have a serious impact, because the choice of the calibration period will influence the correctness of the parameter estimate. Wilby (2005) found that many parameters of the CATCHMOD model had a low identifiability on the Thames basin. However, some of them showed better identifiability when calibrated on wet periods. Abebe *et al*. (2010) applied DYNIA on the HBV model. They also found that identifiability was higher on wet periods than on dry ones. Using the same methodology on the WaSiM-ETH model, Wriedt & Rode (2006) illustrated how the snowmelt runoff parameter was only identifiable during winter runoff and how low-flow conditions were not suitable to calibrate parameters controlling fast runoff processes.

One could argue that the above results are trivial and that nothing will be problematic as long as the calibration period is rich enough in terms of climate variability. However, this is not completely true: hydrologists may face cases where a process that was insignificant during the calibration period (in spite of its length), becomes relevant in the period of interest for simulation. This is particularly true in climate change impact studies. For example, if groundwater exchanges between catchments have a visible impact on discharges only during extended periods of drought, they will not be properly accounted for in a model that was calibrated on a period where droughts were not severe enough. Similar examples could be made with glacier melt processes in mountainous areas, or regarding the role of evapotranspiration in the water balance.

DISCUSSION

The various pathologies listed in this paper cause real problems to hydrologists but should not be seen as fatal. Some preventive and/or curative measures may exist, or they could be found, provided that the problem source is clearly identified first. Following Andréassian *et al*. (2001), we propose to classify model parameters in four categories (considering that calibrations are always made with sufficient input data, i.e. several years):

- **Category A** The parameter remains stable over time (it is not affected by changes in the climate of the calibration period or variance in the quality of input estimates). An example of such a parameter can be found in the simplest models, where one unique parameter controls the time shift between rainfall and flow (see Oudin *et al*., 2006).

- **Category B** The parameter is affected by wrong estimates of the inputs from the calibration period. However, it converges towards a single value when the input quality is improved. Examples of such parameters are the evapotranspiration module parameters, the rainfall multiplication factor or the exchange with groundwater if it is the only way to adjust the water balance (see Andréassian *et al.*, 2001).
- **Category C** The parameter is affected by the climatic properties of the calibration period. This may happen either because the level of identifiability varies depending on the climate, or because the model parameters depend on the climate of the calibration period due to model conceptualization deficiency (Wriedt & Rode, 2006; Abebe *et al*., 2010; Merz *et al*., 2011).
- **Category D** The parameter has an erratic behaviour with no apparent link to the inputs. This reveals the poor adaptation of the model for the case study. It may happen if the process represented by the parameter is not significant, or if the model shows conceptual problems inducing equifinality through parameters compensation (Mo *et al*., 2006; Abebe *et al*., 2010). Parameters of this category are clearly undesirable in a hydrological model.

Parameters from category A are the ideal ones for modellers since they are easy to calibrate and can be globally trusted whatever the future conditions of model use.

Parameters from category B may be the source of problems if they are transferred between periods having different inputs quality/availability levels (e.g. the measurements network has evolved and provide more accurate inputs). However, the problem can be addressed by working on the input quality and choosing cautiously the calibration period to ensure that the model parameters do not compensate for incorrect input estimates. This may lead to the use of only the most recent years, as suggested by Apaydin *et al*. (2006).

Parameters from categories C and D are the most problematic, particularly in the case of climate change impact studies where the model use conditions are known to be different from the calibration ones. However, some sub-distinctions can be made:

1. The parameter falls in category C or D because the process it represents is not significant for the available time series, although the model structure is parsimonious enough to ensure the need for the parameter. If possible, one should first try to increase the length of the calibration period (and thus the climate variability), which may add periods where the process is significant. If all available data are already used, then the problem can be solved either by setting the parameter to an *a priori* coherent value or simply by inactivating the related part of the model (if the process in question is believed to remain insignificant).
2. The parameter falls in category C because it has a clear dependence on the climatic conditions of the calibration period, although its identifiability is high. One solution is to extract from the historical data the sub-periods that are climatically closest to the test period, and calibrate the model on those only. However, this remains problematic if the test conditions are different from all the available data, as is often the case in climate change impact studies. In this case, a second solution could be to use the established relation between parameter values and climate to extrapolate the parameter values corresponding to the future climate. However, this option is qualified by Merz *et al*. (2011) as *inelegant* since it does not respect the usual philosophy of modelling which is to have time invariant parameters that are able to work correctly under time variant conditions (such as the rainfall or the air temperature).
3. For all the other situations, the impossibility of determining the parameter's true value is closely linked with the inappropriate conceptualization of the model. Therefore, no simple solution exists because modellers may be reluctant to change the structure of a model they have been using for years or decades.

CONCLUSIONS

All hydrological models remain simplifications of the real world and incorporate parameters that need to be calibrated to some extent. However, these parameters may sometimes be dependent on the calibration conditions and, therefore, not be appropriate for use under other conditions. Making a parallel with the medical domain, we reviewed here some of the "pathologies" a model may suffer from due to the parameters' dependency on the calibration conditions (e.g. input quality, average climate, diversity of hydrological processes, etc.). Although these findings are highly dependent on the model and given case study, they raise questions regarding the validity of transferring model parameters from one period to another that is temporarily spaced and/or climatically different. This is of particular concern for climate change studies where the model application conditions are known to be different from the calibration ones.

REFERENCES

Abebe, N. A., Ogden, F. L. & Pradhan, N. R. (2010) Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *J. Hydrol.* **389**(3-4), 301–310.

Anctil, F., Perrin, C. & Andréassian, V. (2004) Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall–runoff forecasting models. *Environ. Model. Softw.* **19**(4), 357–368.

Andréassian, V., Perrin, C. & Michel, C. (2004) Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *J. Hydrol.* **286**(1-4), 19–35.

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. & Lavabre, J. (2001) Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* **250**(1-4), 206–223.

Apaydin, H., Anli, A. S. & Ozturk, A. (2006) The temporal transferability of calibrated parameters of a hydrological model. *Ecological Modelling* **195**(3-4), 307–317.

Brath, A., Montanari, A. & Toth, E. (2004) Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *J. Hydrol.* **291**(3-4), 232–253.

Chaubey, I., Haan, C. T., Grunwald, S. & Salisbury, J. M. (1999) Uncertainty in the model parameters due to spatial variability of rainfall. *J. Hydrol.* **220**(1-2), 48–61.

Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C. & Viney, N. R. (2009) Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method. *Water Resour. Res.* **45**, W10414.

Choi, H. T. & Beven, K. (2007) Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *J. Hydrol.* **332**(3-4), 316–336.

Kay, A., Davies, H., Bell, V. & Jones, R. (2009) Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic Change* **92**(1-2), 41–63.

Klemeš, V. (1986) Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **31**(1), 13–24.

Kokkonen, T. S. & Jakeman, A. J. (2001) A comparison of metric and conceptual approaches in rainfall–runoff modeling and its implications. *Water Resour. Res.* **37**(9), 2345–2352.

Merz, R., Parajka, J. & Blöschl, G. (2011) Time stability of catchment model parameters – implications for climate impact analyses. *Water Resour. Res.* **47**, W02531.

Mo, X., Pappenberger, F., Beven, K. J., Liu, S., De Roo, A. & Lin, Z. (2006) Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model. *Hydrol. Sci. J.* **51**(1), 45–65.

Murphy, C., Fealy, R., Charlton, R. & Sweeney, J. (2006) The reliability of an 'off-the-shelf' conceptual rainfall runoff model for use in climate impact assessment: uncertainty quantification using Latin hypercube sampling. *Area* **38**(1), 65–78.

Niel, H., Paturel, J. E. & Servat, E. (2003) Study of parameter stability of a lumped hydrologic model in a context of climatic variability. *J. Hydrol.* **278**(1-4), 213–230.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. & Michel, C. (2006) Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *J. Hydrol.* **320**(1-2), 62–83.

Perrin, C., Oudin, L., Andréassian, V., Rojas-Serna, C., Michel, C. & Mathevet, T. (2007) Impact of limited streamflow data on the efficiency and the parameters of rainfall–runoff models. *Hydrol. Sci. J.* **52**(1), 131.

Prudhomme, C. & Davies, H. (2009) Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate. *Climatic Change* **93**(1), 177–195.

Vaze, J., Post, D. A., Chiew, F., Perraud, J. M., Viney, N. & Teng, J. (2010) Climate nonstationarity – validity of calibrated rainfall–runoff models for use in climatic changes studies. *J. Hydrol.* **394**(3-4), 447–457.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. & Gupta, H. V. (2003) Towards reduced uncertainty in conceptual rainfall–runoff modelling: dynamic identifiability analysis. *Hydrol. Processes* **17**(2), 455–476.

Wilby, R. L. (2005) Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrol. Processes* **19**(16), 3201–3219.

Wilby, R. L. & Harris, I. (2006) A framework for assessing uncertainties in climate change impacts: low-flow scenarios for the River Thames, UK. *Water Resour. Res.* **42**, W02419.

Wriedt, G. & Rode, M. (2006) Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH. *Adv. Geosci.* **9**, 145–150.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1996) Automatic calibration of conceptual rainfall–runoff models: sensitivity to calibration data. *J. Hydrol.* **181**(1-4), 23–48.