Water Science and Engineering 2021, 14(2): 167–170

HOSTED BY ELSEVIER

Available online at www.sciencedirect.com

Water Science and Engineering

journal homepage: wse.hhu.edu.cn

Issues influencing accuracy of hydrological modeling in a changing environment

Chong-yu Xu

Department of Geosciences, University of Oslo, Oslo N-0316, Norway

Received 3 June 2021; accepted 15 June 2021
Available online 18 June 2021

Hydrological models have undergone a long period of development and application. Many hydrological models of different temporal and spatial scales have been developed and applied in various fields of hydrological research and hydrological engineering, including hydrological forecasting, water resources estimation, river basin management, reservoir design and operation, and runoff simulation in ungauged regions. They have also been used in research on the impact of land use change and climate change. Recently, hydrological models have taken on the most important problem-solving tasks in hydrology.

In general, several levels of evaluation are necessary before any model should be used to estimate the output from a catchment or a region, or even global hydrological flux, including model selection, model calibration (estimation of the parameter values), model validation/verification (testing of the fitted model to verify its accuracy), and estimation of its range of applicability. Modeling accuracy in all applications has been the core of concern. Various issues influencing the accuracy of modeling, often depending on the tasks of modeling, have been identified and examined over the past decades. The purpose of this short communication is to summarize what we have learned from past studies as well as conclusions that can guide our future endeavours both in research and practical application. New research should be built systematically on previous research and fill gaps, rather than continuing to reaffirm the consensus that has been formed by the prior work of others in the field. The issues discussed below, which we believe to influence the accuracy of modeling results, include the model calibration and validation strategy, the necessity and sufficiency of modeling data, definitions of climatological and discharge days, input data errors, transferability in nonstationary conditions, and regionalization. Other issues will be discussed in future communications. Our conclusions and suggestions are based on ample literature as well as the author's experience in modeling studies. Studies relevant to the readers' particular interest (which may be too many to be listed in this short letter of communication) can be provided upon request. The author takes responsibility for any insufficiency or inaccuracy in the discussion that follows.

1. Model calibration and validation strategy

Of all the issues that influence the accuracy of hydrological modeling, model calibration and validation have received the most attention from the hydrology community. Almost all hydrological models require calibration—estimation of model parameter values so that the model results agree with observed data according to user-specified criteria. Different model calibration and validation strategies have been proposed, including hierarchical schemes (split-sample, differential split-sample, proxy-basin, and proxy-basin differential split-sample), odd/even year calibration/validation, seasonal calibration, dynamic calibration, and regional calibration. While some methods are still at the development and testing stage, some consensus conclusions could/should be used as a starting point for new research and a basis for guiding practical application. For example, we have learned that different hydrological models, even if they produce equally good results in model calibration, produce largely different results in simulating changed conditions, i.e., nonstationary precipitation–runoff relationships. We have also learned that, for the same model, the simulation results depend on the data period of model calibration. Better simulation results are produced as the calibration data/period (in terms of either the climate characteristics or the time period) draws closer to the validation data/period. This means that models built on different concepts and structures have different abilities in simulating changed conditions. It is therefore an important and basic requirement to perform a differential split-sample test and/or comparison of different models, if the models are to be used in climate change study, in order to at least estimate the

This work was supported by the Research Council of Norway (FRINATEK Project 274310).

E-mail address: c.y.xu@geo.uio.no.

Peer review under responsibility of Hohai University.

https://doi.org/10.1016/j.wse.2021.06.005

error that could be expected for the model in use. Although these models have been used in the past to simulate the hydrological impact of climate change without performance of such tests, the results only represent our past level of knowledge and understanding in the relevant areas. There are several reasons to perform the recommended testing, such as learning which model performs better in the changed conditions and why, estimating the expected error when the chosen model is used to simulate changed conditions, and calibrating the model for the data period with climate conditions closest to the future scenario. Moreover, it is worth pointing out that hydrological data are not only influenced by climate change and variability, but also by land use change, as well as change in the number of available stations in different data periods, which results in spatially and temporally varying data quality. We therefore advise using odd/even years to calibrate/validate the model, which is a better practice than using the split-sample method to calibrate and validate the model. Users might have had an experience in which the validation results are significantly better than the calibration results when using the split-sample test. This is a typical problem related to data inconsistency or inhomogeneity rather than the model's ability. The odd/even year method, meanwhile, can get rid of the noise and complement the true ability of the model.

2. Necessity and sufficiency of modeling data

There has always been a question of how much data are necessary and how much data are sufficient to calibrate and validate a model. The answer to this question is of course dependent on several factors, including the purpose of the model (continuous or flood event-based), the nature and complexity of the model, the climate conditions (humid or arid, etc.), and physical characteristics of the region (urban or rural, etc.). Many studies have examined this important issue, and due to the complexity and heterogeneity of both model types and study catchments, the investigation will continue. However, some important consensus conclusions can be drawn from past studies and advice can be provided for practical use. It is only theoretically true that more data used in model calibration produces better modeling results. This assumes that the greater amount of data contains more information and thus allows for more accurate determination of model parameter values and more robust results. However, in the changing world we are facing, it is impossible to maintain homogeneity and consistency of data due to (1) change and variability of the climate; (2) change in measurement methods, instruments, and therefore accuracy; (3) change in land use; and (4) other reasons. It is therefore necessary to find a balance and compromise in meeting the needs and serving the interests. The following conclusions and advice can be drawn from past studies. For continuous modeling, longer data series are needed for arid climates than for humid climates, as well as for catchments undergoing more change in climate and physical characteristics than those that are not. On average, eight to ten years of data are found generally to be necessary and sufficient for model calibration with daily or monthly simulation. It is also suggested that calibration should be done for a data period that is as close to the forecasting period as possible, since an increase in the length of the interval between calibration and simulation periods diminishes the performance of the model, and the degree of diminishment is greater for catchments with nonstationary rainfall—runoff relationships. For event-based modeling, informativeness of data is much more important than the amount. Two to four selected events for calibration may considerably improve flood predictions with regard to accuracy and uncertainty reduction, whereas adding more events beyond this results in small performance gains.

3. Definitions of climatological and discharge days

The performance of hydrological models is affected not only by uncertainty related to observed climatological and discharge data, but also by the inconsistent definitions of the climatological day and discharge day. Although the former has been widely investigated, the effects of the inconsistency of climatological day and discharge day definition on hydrological model predictions have received little attention. The issue arises from the fact that rainfall—runoff models are usually calibrated at a daily resolution, and daily discharge is commonly computed from midnight to midnight (or 08:00 to 08:00) based on instantaneous discharge values, processed from gauge recorded water levels, while daily rainfall observation data and forecasted rainfall data are usually summed up from hourly data starting at different time of the day. Recent studies have shown that this inconsistency between the starting time for the meteorological data and discharge data has implications for hourly or daily flood forecasting results if the basin response time is shorter than a day. Studies have also shown that model performance is more dependent on the definition of the climatological day than on the definition of the discharge day. This mismatch is expected to be more dependent on catchment size and the intraday distribution of rainfall intensity than on the model. Additional research is suggested, and consideration of the impact of these inconsistent definitions on flood forecasting modeling is highly warranted.

4. Input data errors

Input data errors (e.g., with regard to precipitation and potential evaporation) are among the four important uncertainty sources in hydrological modeling, with other three being uncertainty in data used for calibration (data used for comparison with simulated output, e.g., stream flow observations), uncertainty in model parameters (non-optimal parameter values or non-significant parameters), and uncertainty due to an imperfect model structure. The first two error sources depend on the quality of data, whereas the last two are more model-specific. With regard to various discrepancies in model outputs, input errors are perhaps most important, especially in large-scale modeling using gridded global datasets (interpolated based on observation, remote sensing, or based on reanalysis), in climate change impact modeling based on hypothetical scenarios or projected scenarios, and in flood forecasting based on forecasted rainfall. Therefore, the effect of various input data (precipitation and potential evaporation) errors on model prediction results has been one of the hot research topics of recent decades.

For continuous modeling, attention has been paid to the effects of systematic error and random error of input data on the calibrated parameter values and on the modeling performance. Some important consensus conclusions are that the response of the model parameters and model performance to the input data error depends on, among things, the type of the error, the magnitude of the error, physical characteristics of the catchment, the climate of the study catchment, the season, and, at least partly, the model structure. More specific conclusions are as follows: (1) The relative effect of input data error on the performance of hydrological models depends on the physical characteristics of the catchments, and arid areas with low runoff coefficients are more affected by input data error than catchments with high runoff coefficients. (2) The relative effect depends on the season, with dry seasons being more affected than wet seasons. (3) The sensitivity of a rainfall-runoff model to input data error depends partly on the model structure itself. (4) Precipitation errors are much more important than potential evaporation (or pan evaporation) errors. (5) Systematic error in rainfall input affects most of the model parameter values as well as the model performance, with the model results being significantly affected when the systematic rainfall input error is greater than 10%. (6) The change in parameter values affected by random errors in precipitation input is also random, and when the random error in precipitation input is greater than 20%, the changes in parameter values are significant. (7) Both the model parameters and the model performance are insensitive to random error in potential evaporation input, and systematic error in potential evaporation input has only a moderate effect on model results, which is due to the fact that potential evaporation is not a water balance component but is used as a driving force, and serves as an upper limit for calculation of actual evapotranspiration. However, in modeling the hydrological impact of climate change, potential evaporation calculated using temperature-based methods should not be used as a model input since it more often than not shows an incorrect trend when compared with the true evaporation trend, which results from a combined effect of all the meteorological parameters.

Detailed knowledge of the sensitivity of the chosen model to the input data error is of particular importance, especially when the model is used to simulate the hydrological impact of climate change, since changed climate in both trend and variability is similar to systematic and random error in a sensitivity study, and the results are model-dependent.

For an event-based flood forecasting model, focus has been put on the effects of volume error, timing error, and their combination in input data on the forecasting errors in flood peak timing, flood peak discharge, and flood volume. What we have learned is that (1) flood predictions are highly sensitive to volume errors, and are generally more sensitive to positive than to negative rainfall-volume errors; (2) in terms of predicted time-to-peak errors, model performance is only sensitive to rainfall-duration errors, and is more sensitive to negative than to positive duration errors; and (3) the interplay between the two types of errors seems to compensate for the effect of each type to some extent, which means that relatively good flood prediction results do not mean that the rainfall data are free from volume error and duration error.

5. Regionalization

Regionalization methods can provide runoff predictions in ungauged basins, which cover around 50% of the global land area. Due to the importance of this issue, the International Association of Hydrological Sciences launched a “Decade on Predictions in Ungauged Basins (PUB): 2003−2012”. Numerous studies have been carried out during and after the PUB decade, and different method classifications are available. The regionalization methods categorized here are (1) spatial proximity methods (based on the concept that geographically close catchments show similar hydrological behavior); (2) physical similarity methods (based on the concept that catchments with similar physical characteristics have the same hydrological response); (3) the regression method linking model parameters (dependent variables) to physical and climatic catchment characteristics (independent variables) through multiple regression functions; and (4) regional calibration, in which the model parameters and regression equations (relating the catchment attributes to the model parameters) are optimized simultaneously.

Studies have generally recognized that success is dependent on availability and quality of data, including station density; the hydrological model; basin physical characteristics, including climate; and the regionalization method. Study on this topic will continue due to the complexity and diversity of models, data, and study regions. On the other hand, significant progress has been achieved and some important consensus conclusions can be drawn to provide a basis for further study and guidance for practical use: (1) The traditional multiple regression method has a greater chance to stand out in the competition for hydrological models with stronger physical relevance and fewer parameters that are clearly identifiable. Conceptual models with a larger number of parameters usually perform worst with the regression method because they suffer more from parameter interaction/interdependence and therefore equifinality, which results in only a small fraction of the total number of parameters being successfully regressed with basin physical and climatic factors. (2) Spatial proximity and physical similarity methods usually score best for many conceptual models over other methods, and the differences between these two methods are usually small. (3) Of the two transfer options in the spatial proximity and physical similarity methods, the output average option (in which individual parameter sets from donor catchments are used in ungauged catchments and averaged runoff is calculated from the simulations) outperforms the parameter average option (in which averaged model parameter values are calculated from individual parameter sets of donor catchments and then used in ungauged catchments) in all cases, and the difference in performance between these options increases with the number of model parameters. This is a confirmed conclusion in regionalization studies, which is independent of the hydrological model and study region. (4) The number of donor catchments selected for the spatial proximity and physical similarity methods depends on the model, but three to five donor catchments seem to be right in most cases. (5) The regionalization performance depends on the threshold values of model efficiency in the calibration stage used as the basis to select donor catchments, and the intermediate value (equivalent to a Nash-Sutcliffe efficiency coefficient or Kling Gupta efficiency value of 0.5−0.7) for the threshold yields the best regionalization results. Using too high of a threshold value results in too few donor catchments, leading to a remarkable drop in regionalization results, while using too low of a threshold value, or even using all of the catchments, does not improve the results. (6) A regional calibration method considerably improves the regional relationship between model parameters and physical characteristics of the catchments. However, improvement in the regionalization results over the traditional regression method is less significant. This is partly due to the high interdependency of model parameters. (7) The success of regionalization methods depends to some extent on the climate and geographic regions, with better results reported for temperate and humid regions than for arid regions. However, selection of the climatic region has no significant effect on the ranking of regionalization methods.

Declaration of competing interest

The author declares no conflicts of interest.