

# **Structural Uncertainty in Hydrological Models**

**Abhinav Gupta**

## **Abstract**

All hydrologic models incur various uncertainties that can be broadly classified into three categories: measurement, structural, and parametric uncertainties. Measurement uncertainty exists due to error in measurements of properties and variables (e.g. streamflows that are typically an output and rainfall that serves as an input to hydrological models). Structural uncertainty exists due errors in mathematical representation of real-world hydrological processes. Parametric uncertainty exists due to structural and measurement uncertainty and limited amount of data availability for calibration.

Several studies have addressed the problem of measurement and parametric uncertainties but studies on structural uncertainty are lacking. Specifically, there does not exist any model that can be used to quantify structural uncertainties at an ungauged location. This was the first objective of the study: to develop a model of structural uncertainty that can be used to quantify total uncertainty (including structural uncertainty) in streamflow estimates at ungauged locations in a watershed. The proposed model is based on the idea that since the effect of structural uncertainty is to introduce a bias into the parameter estimation, one way to accommodate structural uncertainty is to compensate for this bias. The developed model was applied to two watersheds: Upper Wabash Busseron Watershed (UWBW) and Lower Des Plaines Watershed (LDPW). For UWBW, mean daily streamflow data were used while for LDPW mean hourly streamflow data were used. The proposed model worked well for mean daily data but failed to capture the total uncertainties for hourly data likely due to higher measurement uncertainties in hourly streamflow data than what was assumed in the study.

Once a hydrological and error model is specified, the next step is to estimate model- and error-parameters. Parameter estimation in hydrological modeling may be carried out using either formal Bayesian methodology or informal Bayesian methodology. In formal Bayesian methodology, a likelihood function, motivated from probability theory, is specified over a space of models (or residuals) and, a prior probability distribution is assigned over the space of models. There has been significant debate on whether the likelihood functions used in Bayesian theory are justified in

hydrological modeling. However, relatively little attention has been given to justification of prior probabilities. In most hydrological modeling studies, a uniform prior over hydrological model parameters is used to reflect a complete lack of knowledge of a modeler about model parameters before calibration. Such a prior is also known as a non-informative prior. The second objective of this study was to place the assumption of a uniform prior as being non-informative to scrutiny using the principle of maximum information gain. This principle was used to derive non-informative priors for several hydrological models, and it was found that the obtained prior was significantly different from a uniform prior. Further, the posterior distributions obtained by using this prior were significantly different from those obtained by using uniform priors.

The information about uncertainty in a modeling exercise is typically obtained from residual time-series (the difference between observed and simulated streamflows) which is an aggregate of structural and measurement uncertainties for a fixed model parameter set. Using this residual time-series, an estimate of total uncertainty may be obtained but it is impossible to separate structural and measurement uncertainties. The separation of these two uncertainties is, however, required to facilitate the rejection of deficient model structures, and to identify whether the model structure or the measurements need to be improved to reduce the total uncertainty. The only way to achieve this goal is to obtain an estimate of measurement uncertainty before model calibration. An estimate of measurement uncertainties in streamflow can be obtained by using rating-curve analysis but it is difficult to obtain an estimate of measurement uncertainty in rainfall. In this study, the classic idea of repeated sampling is used to get an estimate of measurement uncertainty in rainfall and streamflows. In the repeated sampling scheme, an experiment is performed several times under identical conditions to get an estimate of measurement uncertainty. This kind of repeated sampling, however, is not strictly possible for environmental observations, therefore, repeated sampling was used in an approximate manner using a machine learning algorithm called random forest (RF). The main idea is to identify rainfall-runoff events across several different watersheds which are similar to each-other such that they can be thought of as different realizations of the same experiment performed under identical conditions. The uncertainty bounds obtained by RF were compared against the uncertainty band obtained by rating-curve analysis and runoff-coefficient method. Overall, the results of this study are encouraging in using RF as a pseudo repeated sampler.

In the fourth objective, importance of uncertainty in estimated streamflows at ungauged locations and uncertainty in measured streamflows at gauged locations is illustrated in water quality modeling. The results of this study showed that it is not enough to obtain an uncertainty bound that envelops the true streamflows, but that the individual realizations obtained by the model of uncertainty should be able to emulate the shape of the true streamflow time-series for water quality modeling.