Uncertainty Analysis in Surface Water Hydrology

Abhirama Gorti 20AG38002

Abstract

Uncertainty analysis in surface water hydrology is a set of methods that can be used to quantify and manage Uncertainty in hydrological models and predictions. Uncertainty can arise from various sources, including input data uncertainty, model structure uncertainty, parameter uncertainty, and natural variability. Uncertainty in surface water hydrology can significantly affect water management decisions, such as flood forecasting and water resources planning.

This term paper will provide a comprehensive overview of uncertainty analysis in surface water hydrology. It will cover the different types of Uncertainty, the sources of luck, and the methods that can be used to quantify and manage Uncertainty. The paper will also discuss the implications of uncertainty analysis for water management decision-making.

1 Introduction

Uncertainty is inherent in all natural phenomena, and hydrology is no exception. Various factors influence hydrological processes, including climate, land use, and human activities. These factors are complex and interconnected, and their interactions are often poorly understood. This complexity and interconnectedness lead to a high uncertainty in hydrological processes and predictions.

Kirchner et al. (2021), explains, in brief, the different procedures in which uncertainties can be estimated. It is by asking ourselves three questions:

1. **Nature** of the uncertainty
2. **Type** of the uncertainty
3. **Location** of uncertainty

Based on the nature of Uncertainty, It can be classified into two types: epistemic and aleatory. Epistemic Uncertainty, according to Gupta and Govindaraju (2023), refers to the delay due to a lack of knowledge or understanding about the system. It can be reduced by gathering more information. The possible sources can be random/systemic errors in measurement or imprecision derived from qualitative modelling. The other type is Aleatory Uncertainty, which refers to the fundamental part of total Uncertainty. It occurs due to the variability over space and time, Inherent randomness in a system, also referred to as entropy of the system or Chaos in the system. It can be approximated using statistical distributions and Bayesian Statistics.

Although Grey S. Nearing and Weijs (2016) says that all the uncertainties are strictly epistemic because there are specific physical reasons behind having those random errors since considering all those would become highly complex, we have resorted to assuming that the uncertainties are of two types.

While it comes to the types of Uncertainty, it is pretty natural that all the different kinds of Uncertainties that stem in the hydrological context and are important essentially would be Epistemic. Under epistemic, based on the error type, the unc certainty can be classified as follows:

- **Input data uncertainty:** This is the Uncertainty associated with the accuracy and completeness of the data used to drive hydrological models. Input data uncertainty can arise from various sources, such as the measurement errors of sensors, the spatial and temporal resolution of data, and the methods used to interpolate data.
- **Model structure uncertainty:** This is the Uncertainty associated with the mathematical equations and assumptions used to represent hydrological processes in hydrological models. Model structure uncertainty arises because it is impossible to perfectly represent the complex interactions of all factors that influence hydrological processes.
- **Parameter uncertainty:** This is the Uncertainty associated with the parameters' values used in hydrological models. Parameter uncertainty arises because it is difficult to measure the parameters accurately and because they may vary in space and time.

In the early days of hydrological research, the focus was on developing models that could accurately predict hydrological processes. However, it soon became clear that these models were highly uncertain. As a result, researchers began to focus on developing methods to quantify and manage Uncertainty in hydrological models and predictions.

Researchers have developed many methods for quantifying and managing Uncertainty, including Monte Carlo simulation, Bayesian statistics, and ensemble modelling, as discussed in Moges et al. (2021). These methods have been applied to various hydrological problems, including flood forecasting, water resources planning, and climate change impact assessment.

In the next part of this term paper, we will discuss how hydrological models have evolved and how Uncertainty is produced at the first step. We will discuss the methods that can be used to quantify and manage these uncertainties. We will also discuss the implications of uncertain analysis for water management decision-making.

2 Hydrological Models

A Hydrological model is a simplified representation of the complex physical characteristics of a catchment, which aids in the decision-making process. They are mathematical models representing the physical process using mathematical equations expressing the relations between input, parameters and output.

However, since it is a simplified version with parameters involved, it brings Uncertainty into discussion. As explained earlier, there are three uncertainties, i.e. through input data, model structure and parameters.

McMillan et al. (2018) mentions multiple sources of uncertainties such as interpolation uncertainty, data management uncertainty, etc., and their range lies around 10-40%. Zhou et al. (2021) focuses on two main types of errors: Input Error and Structural. The study investigates the impact of these uncertainties on flows with different magnitudes using three models (HyMod, XAJ and HBV). The variance decomposition method was used to quantify the Uncertainty of runoff emulation between input levels and its interaction with the hydrological model. The results show that different rain gauge station input levels and hydrological models dynamically affected the hydrological simulation due to an uneven spatio-temporal distribution of precipitation. Moreover, increasing rainfall station input under a certain threshold could significantly improve the accuracy of hydrological simulation.

On a broad scale, it is essential to understand that all types of uncertainties are interlinked. Starting with input data, irregular calibration data to train a model added with the model's structural Uncertainty lead to parameter uncertainty, which is the most important of all of the above.

The elephant in the room is the model type and the calibration parameters that complete the uncertainty cycle. Let us look at a few papers that discuss it.

2.1 Impact of type of model

One of the most impacting papers by Keith Beven, Binley et al. (1991) discusses the Uncertainty in hydrological model predictions in the case of a physical model or, in general, non-linear hydrological models. His works cover a major portion of psychological aspects, such as the Equifinality Theory, which states that multiple acceptable models can provide a good fit to observational data. This concept challenges the traditional scientific approach of working towards a single correct description of reality.

He explains that the traditional calibration methods for hydrological models are incomplete and that the potential for multiple acceptable models should be considered more seriously. He proposes some techniques for an extended Generalised Likelihood Uncertainty Estimation (GLUE) methodology to make it more rigorous and outlines some of the research issues still to be resolved.

This concept also extends to the modelling of GCMs, which is not very important in our study.

Brirhet and Benaabidate (2016) compares the adaptability between a distributed model and a lumped model in the Moroccan context to generalise the selected model to the entire watershed. Although the results from the validation phase during flooding periods were satisfactory for both models, it concludes that depending on the context, the globally distributed model always has more accurate predictions, which is true because distributed ones divide the catchment into sub-basins and make predictions at better scale than lumped ones which work by taking an ensemble of all the catchment characteristics.

Since the context of model structure has come, it brings an inherent quality of complexity.Orth et al. (2015) validates and compares three hydrological models of different complexities: a SWBM, the semi-distributed model HBV and a spatially distributed model PREVAH. The set of governing equations was the same for all. The results show that more complex models outperformed runoff predictions, but for extreme events and soil moisture predictions, the case was reversed, and the conceptual lumped SWBM was ranked 1. The authors attribute the reasons to differences in forcing variables and Uncertainty of calibrated parameters. Thus, added complexity does not always result in better predictions and depends on other unpredictable hydrological conditions.

The Uncertainty behind mapping these parameters can be well understood in the next section, where the methods for parameter estimations and uncertainty analysis are discussed as in the literature.

3 Uncertainty Analysis Methods

Bringing from the previous sections in the paper, we have covered various uncertainty analysis methods, which researchers have used to bridge the gap for better predictions on hydrological models.

Starting with Moges et al. (2021), the authors have done a great job covering all the methods very lucidly. The commonly used UA methods can be categorised into six broad classes.

1. **Monte Carlo Analysis and GLUE**
2. **Bayesian Statistics**
3. **Multi-objective analysis**
4. **Least-squares-based inverse modeling**
5. **Response-surface-based techniques**
6. **Multi-modeling analysis**

While there are other novel methods, such as the use of neural networks, Markov chains, and fuzzy set theory, the above techniques are covered in great detail, and let us discuss them.

3.1 Monte Carlo Analysis and GLUE

Monte Carlo analysis uses a boosting model approach and analyses complex systems. Different output values are generated using a range of input values in our context. By judging the range of possible outcomes, one can conclude the amount of data required and the number of simulation runs needed to get to a good prediction range.

William Scharffenberg, one of the pioneers behind the HEC-HMC model, has written a summary in one of his papers. The model faced knowledge uncertainty, a counterpart of epistemic Uncertainty, which is our inability to capture the whole hydrological process despite having many parameters such as precipitation, snow melt, transpiration, etc. They have used a Monte Carlo-based uncertainty analysis tool embedded in the model to address this. More accurate predictions could be made by statistically sampling model parameters and considering their natural variability and knowledge uncertainty.

Without any doubt, the best use of Monte-Carlo simulations has been made in GLUE (Generalized Likelihood Uncertainty Estimation) methodology by Beven and Binley (1992) which maps the Uncertainty in the modelling process onto the parameter space and operates within the context of Monte Carlo analysis coupled with likelihood functions.

Building upon this, better methods have been proposed. Some of them include Khu and Werner (2003), which offers to use a Genetic Algorithm. This search heuristic reflects natural selection from Charles Darwin's evolutionary theory and Neural networks as a surrogate model to reduce the number of simulation runs and increase the sampling efficiency of GLUE. However, GLUE has been criticised for its subjectivity in choosing a behavioural threshold and its lack of a formal statistical foundation, i.e., it might be sensitive to the selection of likelihood function and prior distribution and hence cannot always be the best solution.

3.2 Bayesian Statistics

Bayesian statistics is a statistical approach that deals with Uncertainty by incorporating prior knowledge and updating it with new data. This constant updating helps constrain the range of plausible parameter values and reduce Uncertainty. This can be combined with multiple models to produce an ensemble of "Best" predictors for different parameter sets.

The predictions are combined using a weighted approach as discussed in Adrian E. Raftery and Hoeting (1997), where the weights are derived from the model's previous performances on the parameter set. This helps in dealing with the structural uncertainty component of the total Uncertainty very well rather than the initial attempts where all the models were assigned equal weights. This is the basis for the DREAM model system, which takes data on Runoff, Evapotranspiration, and Antecedent soil Moisture daily.

3.3 Multi-objective Analysis

Multi-objective analysis in the context of hydrological models is a method used to address the issue of Uncertainty. It involves optimising more than one performance measure simultaneously, resulting in multiple sets of parameters that provide optimal solutions.

This method was used and explained by Smith et al. (2019). The analysis included GR4J, a daily-lumped rainfall-runoff model used for stream flow reconstructions with four parameters. It was applied across 303 catchments and six evaluation metrics to cover a multi-objective analysis and did pretty well on most of the catchments. The authors also reconstructed previous extreme events, such as floods and droughts across England and Scotland.

3.4 Least-squares-based inverse modeling

Least-squares-based inverse modelling solves inverse problems by minimising the sum of the squared residuals between the model predictions and the observed data. Inverse problems are problems in which one tries to infer the cause of a system from its observed effects. This can be used to estimate the parameters of a hydrological model, such as the infiltration rate and the roughness coefficient, from observed data, such as precipitation and runoff.

The least-squares-based inverse modelling algorithm works as follows:
Algorithm

- An initial set of model parameters is selected.
- The model is run with the initial set of parameters to produce a set of model predictions.
- The residuals between the model predictions and the observed data are calculated.
- The model parameters are adjusted to minimise the sum of the squared residuals.
- Steps 2-4 are repeated until the model predictions converge to values consistent with the observed data.

The method is relatively simple and provides a straightforward criterion of squared error. The most famous use case is the PEST (**P**arameter **EST**imation) model. It is an ensemble of the Monte-Carlo methodology of running simulations using random sampling techniques to arrive at a prediction.

However, the major downside of this model and method is that it carries a lot of bias. As discussed in **??**, the equifinality theory suggests a range of valid target values from many parameter sets in the space. By using the model, we are constraining ourselves to less data.

4 New Challenges Ahead

Reducing uncertainties in model predictions is a significant challenge in various fields, including climate modelling, demand prediction, and hydrological modelling. Some of the key challenges include:

- **Model Imperfections**: These lead to drift and errors in near-term initialised climate prediction systems and uncertainties in long-term future projections. Techniques such as bias correction and drift removal have been developed to reduce the impact of model imperfection.
- **Data Aggregation**: The failure of demand prediction models is often rooted in the fact that they do not consider how data is generated but explore apparent relationships in aggregated data.
- **Complex Dynamics**: Managing prediction uncertainty presents significant challenges in environments characterised by complex dynamics.

Recent papers offer solutions in various domains, such as using Fuzzy logic and Soft computing models Chandwani et al. (2015), Sequential models Zecchin et al. (2023), and groundbreaking neural network architectures such as transformer network Amanambu et al. (2022). It has outperformed traditional models at many criteria for considering data from a longer timescale with recurrence.

We hoped to have explained lucidly the uncertainties associated with model predictions, the different types of models responsible for projections and the solutions offered in decreasing order of complexity to address this gap of Uncertainty. Uncertainty extends to Information theory, routing to another realm of hydrology: the study of extreme events. We have left it for good reasons as it might lead to other mathematical concepts, which may take another paper to discuss if discussed.

References

Adrian E. Raftery, D. M. and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191.

Amanambu, A. C., Mossa, J., and Chen, Y.-H. (2022). Hydrological drought forecasting using a deep transformer model. *Water*, 14(22).

Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298.

Binley, A., Beven, K., Calver, A., and Watts, L. (1991). Changing responses in hydrology: Assessing the uncertainty in physically based model predictions. *Water Resources Research - WATER RESOUR RES*, 27:1253–1261.

Brirhet, H. and Benaabidate, L. (2016). Comparison of two hydrological models (lumped and distributed) over a pilot area of the issen watershed in the souss basin, morocco. *European Scientific Journal, ESJ*, 12(18):347.

Chandwani, V., Vyas, S., and Sharma, G. (2015). Soft computing approach for rainfall-runoff modelling: A review. *AQUATIC PROCEDIA*, 4:1054–1061.

Grey S. Nearing, Yudong Tian, H. V. G. M. P. C. K. W. H. and Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, 61(9):1666–1678.

Gupta, A. and Govindaraju, R. S. (2023). Uncertainty quantification in watershed hydrology: Which method to use? *Journal of Hydrology*, 616:128749.

Khu, S.-T. and Werner, M. G. F. (2003). Reduction of monte-carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrology and Earth System Sciences*, 7(5):680–692.

Kirchner, M., Mitter, H., Schneider, U. A., Sommer, M., Falkner, K., and Schmid, E. (2021). Uncertainty concepts for integrated modeling - review and application for identifying uncertainties and uncertainty propagation pathways. *Environmental Modelling Software*, 135:104905.

McMillan, H. K., Westerberg, I. K., and Krueger, T. (2018). Hydrological data uncertainty and its implications. *WIREs Water*, 5(6):e1319.

Moges, E., Demissie, Y., Larsen, L., and Yassin, F. (2021). Review: Sources of hydrological model uncertainties and advances in their analysis. *Water*, 13(1).

Orth, R., Staudinger, M., Seneviratne, S., Seibert, J., and Zappa, M. (2015). Does model performance improve with complexity? a case study with three hydrological models.

Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J. (2019). A multi-objective ensemble approach to hydrological modelling in the uk: an application to historic drought reconstruction. *Hydrology and Earth System Sciences*, 23(8):3247–3268.

Zecchin, M., Park, S., and Simeone, O. (2023). Forking uncertainties: Reliable prediction and model predictive control with sequence models via conformal risk control.

Zhou, S., Wang, Y., Li, Z., Chang, J., and Guo, A. (2021). Quantifying the uncertainty interaction between the model input and structure on hydrological processes. *Water Resources Management*, 35(12):3915–3935.