Geophysical Research Abstracts
Vol. 18, EGU2016-15870, 2016
EGU General Assembly 2016

Uncertainty "escalation" and use of machine learning to forecast residual and data model uncertainties

Dimitri Solomatine (1,2)

(1) UNESCO-IHE Institute for Water Education, IWSG, Delft, Netherlands, (2) Water Resources Section, Delft University of Technology, Netherlands

When speaking about model uncertainty many authors implicitly assume the data uncertainty (mainly in parameters or inputs) which is probabilistically described by distributions. Often however it is look also into the residual uncertainty as well. It is hence reasonable to classify the main approaches to uncertainty analysis with respect to the two main types of model uncertainty that can be distinguished:

A.
The residual uncertainty of models. In this case the model parameters and/or model inputs are considered to be fixed (deterministic), i.e. the model is considered to be optimal (calibrated) and deterministic. Model error is considered as the manifestation of uncertainty. If there is enough past data about the model errors (i.e. it uncertainty), it is possible to build a statistical or machine learning model of uncertainty trained on this data.

The following methods can be mentioned:
(a) quantile regression (QR) method by Koenker and Basset in which linear regression is used to build predictive models for distribution quantiles [1]
(b) a more recent approach that takes into account the input variables influencing such uncertainty and uses more advanced machine learning (non-linear) methods (neural networks, model trees etc.) – the UNEEC method [2,3,7]
(c) and even more recent DUBRAUE method (Dynamic Uncertainty Model By Regression on Absolute Error), a autoregressive model of model residuals (it corrects the model residual first and then carries out the uncertainty prediction by a autoregressive statistical model) [5]

B.
The data uncertainty (parametric and/or input) – in this case we study the propagation of uncertainty (presented typically probabilistically) from parameters or inputs to the model outputs. In case of simple functions representing models analytical approaches can be used, or approximation methods (e.g., first-order second moment method). However, for real complex non-linear models implemented in software there is no other choice except using some variant of the Monte Carlo simulation when values of parameters or inputs are sampled from the assumed distributions and the model is run multiple times to generate multiple outputs. This is the most widely used approach.

The data generated by Monte Carlo analysis can be used to build a machine learning model which will be able to make predictions of model uncertainty for the future his method is named MLUE (Machine Learning for Uncertainty Estimation) and is covered in [4,5]

With this in mind, one may consider the following framework based on the stepwise "building up" (or "escalation") of the model uncertainty:
• first consider the residual uncertainty of an optimal model M (X, p*)
• then add and consider the model uncertainty due the parameters uncertainty (p)
• then add and consider the model uncertainty due the data (mainly, input) uncertainty (X)
• then add and consider the structural uncertainty of the model M (X, p).

The paper presents the details of this framework and examples if its application in hydrological forecasting.

This study is partly supported by the FP7 European Project WeSenseIt Citizen Water Observatory

(www.http://wesenseit.eu/).

References

[1] Koenker, R., and G. Bassett (1978). Regression quantiles. Econometrica, 46(1), 33– 50, doi:10.2307/1913643.
[2] D.L. Shrestha, D.P. Solomatine (2006). Machine learning approaches for estimation of prediction interval for the model output. Neural Networks J., 19(2), 225-235.
[3] D.P. Solomatine, D.L. Shrestha (2009). A novel method to estimate model uncertainty using machine learning techniques. Water Resources Res. 45, W00B11.
[4] D. L. Shrestha, N. Kayastha, and D. P. Solomatine. A novel approach to parameter uncertainty analysis of hydrological models using neural networks. Hydrol. Earth Syst. Sci., 13, 1235–1248, 2009.
[5] F. Pianosi and L. Raso (2012). Dynamic modeling of predictive uncertainty by regression on absolute errors. WRR, 48, W03516.
[6] Shrestha, D.L., Kayastha, N., Solomatine, D., Price, R. (2014). Encapsulation of parametric uncertainty statistics by various predictive machine learning models: MLUE method. J Hydroinformatics, 16 (1), 95-113.
[7] N. Dogulu, P. López López, D. P. Solomatine, A. H. Weerts, and D. L. Shrestha (2014). Estimation of predictive hydrologic uncertainty using quantile regression and UNEEC methods and their comparison on contrasting catchments, Hydrol. Earth Syst. Sci. Disc, 11, 10179-10233.