Geophysical Research Abstracts
Vol. 18, EGU2016-17591, 2016
EGU General Assembly 2016

How well can we forecast future model error and uncertainty by mining past model performance data

Dimitri Solomatine (1,2)

(1) UNESCO-IHE Institute for Water Education, Delft, Netherlands (d.solomatine@unesco-ihe.org), (2) Water Resources Section, Delft University of Technology, Netherlands

Consider a hydrological model Y(t) = M(X(t), P), where X=vector of inputs; P=vector of parameters; Y=model output (typically flow); t=time. In cases when there is enough past data on the model M performance, it is possible to use this data to build a (data-driven) model EC of model M error. This model EC will be able to forecast error E when a new input X is fed into model M; then subtracting E from the model prediction Y a better estimate of Y can be obtained. Model EC is usually called the error corrector (in meteorology - a bias corrector).

However, we may go further in characterizing model deficiencies, and instead of using the error (a real value) we may consider a more sophisticated characterization, namely a probabilistic one. So instead of rather a model EC of the model M error it is also possible to build a model U of model M uncertainty; if uncertainty is described as the model error distribution D this model will calculate its properties - mean, variance, other moments, and quantiles. The general form of this model could be: D = U (RV), where RV=vector of relevant variables having influence on model uncertainty (to be identified e.g. by mutual information analysis); D=vector of variables characterizing the error distribution (typically, two or more quantiles).

There is one aspect which is not always explicitly mentioned in uncertainty analysis work. In our view it is important to distinguish the following main types of model uncertainty:

1. 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. its uncertainty), it is possible to build a statistical or machine learning model of uncertainty trained on this data.
Here 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) the UNEEC method [2,3,7] which takes into account the input variables influencing such uncertainty and uses more advanced machine learning (non-linear) methods (e.g. neural networks or k-NN method)
(c) the recent DUBRAE method (Dynamic Uncertainty Model By Regression on Absolute Error), a autoregressive model of model residuals which first corrects the model residual and then employs an autoregressive statistical model for uncertainty prediction) [5]

2. 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. For real complex non-linear functions (models) implemented in software various versions of the Monte Carlo simulation are used: values of parameters or inputs are sampled from the assumed distributions and the model is run multiple times to generate multiple outputs. 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 [4,6].

3. Structural uncertainty stemming from inadequate model structure.

The paper discusses the possibilities and experiences of building the models able to forecast (rather than analyse) residual and parametric uncertainty of hydrological models.

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.