Comments by Shlomo P. Neuman on

A robust objective function for calibration of groundwater models in light of deficiencies of model structure and observations

by Raphael Schneider, Hans Jørgen Henriksen, and Simon Stisen

The authors propose using Continuous Ranked Probability Score (CRSP) as a criterion for calibrating groundwater models against hydraulic head data. They reason that large residual calibration errors, which often result in part from structural model errors, would dominate CRSP calibration results to a lesser degree than they do when one uses standard criteria such as mean square error (MSE) or mean absolute error (MAE); CSRP would assign greater weight to the majority of smaller residuals than to a few larger residuals at the edges of their cumulative distribution. Whereas CSRP is designed to work with ensembles of predictions, the authors suggest applying it to a single realization of calibration errors across a model space-time horizon. To test their idea, the authors apply CSRP to two regional scale coupled surface-groundwater models to conclude that their proposed criterion results in lesser calibration bias than do MSE or MAE.

I find the idea of using CRSP as a calibration criterion interesting but consider the authors' attempt to demonstrate its utility unconvincing. My reasons are as follows:

1. Applying the probabilistic CRSP criterion to a single realization of calibration error requires an assumption of ergodicity. There is no discussion of this potential restriction in the manuscript.
2. Groundwater flow models differ fundamentally from most surface water models in that parameters entering the former (hydraulic conductivity or transmissivity, specific storage or drainable porosity) tend to have reasonably well-defined physical meanings and can often be estimated, independently of the calibrated model, through methods such as pumping tests and geostatistical interpolation. This makes it possible, and often necessary, to regularize the model calibration process with the aid of parameter plausibility criteria based either on such independent prior parameter estimates or on functional criteria such as smoothness. One purpose of such regularization criteria is to ensure that large calibration errors do not dominate the parameter estimation process. Would CRSP be still necessary, and/or useful, in this context? The manuscript does not address this question.
3. The two case studies fail to provide information about the reliability of parameters estimated using either CSRP, MSE or MAE. To validly compare these three criteria, one would need to test them on synthetic systems having known structures, parameters and forcing terms that are corrupted by known random and/or systematic errors of realistic kinds and magnitudes. One would further need to explore CSRP in the context of regularization criteria such as those commonly used in groundwater model calibration. Only then would it make sense to demonstrate the utility of CSRP on partially defined field problems such as those in the two case studies described.