Chapter 61
Application of Data Fusion for Uncertainty and Sensitivity Analysis of Water Quality in the Shenandoah River

Mbongowo Joseph Mbuh
University of North Dakota, USA

ABSTRACT

This article is aimed at demonstrating the feasibility of combining water quality observations with modeling using data fusion techniques for efficient nutrients monitoring in the Shenandoah River (SR). It explores the hypothesis; "Sensitivity and uncertainty from water quality modeling and field observation can be improved through data fusion for a better prediction of water quality." It models water quality using water quality simulation programs and combines the results with field observation, using a Kalman filter (KF). The results show that the analysis can be improved by using more observations in watersheds where minor variations to the analysis result in large differences in the subsequent forecast. Analyses also show that while data fusion was an invaluable tool to reduce uncertainty, an improvement in the temporal scales would also enhance results and reduce uncertainty. To examine how changes in the field observation affects the final KF analysis, the fusion and lab analysis cross-validation showed some improvement in the results with a very high coefficient of determination.

INTRODUCTION

The behavior of environmental processes is difficult to predict with much certainty from the field observation data because the results of the time series models lack information on the physical knowledge of the process (Weijs, S.V., 2014). High levels of uncertainty in the field data due to space and time variability also make it difficult to use the data to reconstruct the spatial and temporal patterns (Drécourt and Rosbjerg, 2004). In order to obtain consistent spatial and temporal results, deterministic or stochastic physically based models have been used, but these models also come with their shortcomings because

DOI: 10.4018/978-1-5225-8054-6.ch061

they cannot accurately reproduce the available measurement (Liu, Y., & Gupta, H. V., 2007). Since the model for an efficient reconstruction requires information from field data, and vice versa, this makes an analysis, and models complementary through an integration of the uncertain measurement and uncertain models through data fusion (Kistler et al., 2001; Compo et al., 2006). When predicting water quality problems with models, integration of observations is a critical issue for model quality (Errico, 1999; Errico et al., 2000). Also, water quality models used to predict the spatial hydrologic system variations are often weak due to model initialization, state errors and inadequate model physics and/or resolution (Walker and Houser, 2005). Also, satellite data retrievals of water quality are subject to errors and cannot provide complete space-time coverage. As the great statistician George Box noted: "All models are wrong, but some are useful." From results obtained from the water quality simulation program (WASP) (Wool et al., 2006), there are so many assumptions and "parameterizations of our ignorance" that go into the models, we cannot use our results with confidence when making management decisions due to their degree of subjectivity. However, addressing George Box's concern,, data fusion can be used to solve this problem.

Hydrologic modeling with data fusion methods is a quite recent developmetn; as such, there is an absence of existing general guidance on how to choose the best data fusion approach, which considers uncertainty correctly. This has been a limitation to extensive data fusion for hydrologic applications (Liu and Gupta, 2007). Data fusion started as a military project in the 60s, with the aim of controling the trajectory of missiles. Using a model alone would lead to erroneous trajectories because of the incomplete knowledge of atmospheric conditions, and it was impossible to collect data accurate enough to rely solely on them (Drécourt and Rosbjerg, 2004). This method was therefore designed to take the best of both worlds: where there is no observation, a physical model is used and relied upon. Where good data are available, they are used to represent the system, and, above all, the uncertainty of both the data and the model are taken into account (Drécourt and Rosbjerg, 2004). Such an approach is currently used in numerical weather prediction (NWP) and is a technique of merging observation data with prediction model data to more precisely predict the state of a system (Rabier, 2005). Its usage has also been successful in oceanography and hydrology. Monitoring networks for water quality modeling can be improved to reduce modeling uncertainty using data fusion (Yangxiao et al., 2006).

Hydrologic data fusion is relatively new, although deterministic hydrological prediction and parameter estimation have become reasonably mature (Bennett, 1992), with soil moisture being one of the primary areas of hydrologic data fusion application. Surface temperature, terrestrial water storage, snow, and stream flow have only been used, in more recent applications. The available literature on hydrologic data fusion ranges from characterization of soil moisture and/or surface energy balance (Entekhabi et al., 1994; Houser et al., 1998; Entekhabi et al., 1999; Dunne and Entekhabi, 2005), to rainfall-runoff modeling (Restrepo, 1985; Vrugt et al., 2005), to flood foresting (Kitanidis and Bras, 1980; Young, 2002), to estimation of hydraulic conductivity (Katul et al., 1993; Lee et al., 1993), transport problems and groundwater flow (McLaughlin et al., 1993), to estimation of water table elevations (Van Geer et al., 1991; Yangxiao et al., 1991).

In 1960, R.E. Kalman first presented the original Kalman Filter (KF) designed for linear models. The Extended Kalman filter was later developed based on linearization of the nonlinear model using the Jacobian, but it was not suitable for large-scale problems or problems that were too nonlinear.A water quality model, however, provides both spatial and temporal near-surface data at the model resolution, including errors following inadequate model physics, parameters and forcing data (Walker ans Houser, 2005). This method is useful for real-time applications as compared to the Kalman filter, which is limited