Bayesian Approach for Uncertainty Analysis of an Urban Storm Water Model and Its Application to a Heavily Urbanized Watershed

Misgana K. Muleta, M.ASCE1; Jonathan McMillan2; Geremew G. Amenu3; and Steven J. Burian, M.ASCE4

Abstract: The significance of uncertainty analysis (UA) to quantify reliability of model simulations is being recognized. Consequently, literature on parameter and predictive uncertainty assessment of water resources models has been rising. Applications dealing with urban drainage systems are, however, very limited. This study applies formal Bayesian approach for uncertainty analysis of a widely used storm water management model and illustrates the methodology using a highly urbanized watershed in the Los Angeles Basin, California. A flexible likelihood function that accommodates heteroscedasticity, non-normality, and temporal correlation of model residuals has been used for the study along with a Markov-chain Monte Carlo-based sampling scheme. The solution of the UA model has been compared with the solution of the conventional calibration methodology widely practiced in water resources modeling. Results indicate that the maximum likelihood solution determined using the UA model produced runoff simulations that are of comparable accuracy with the solution of the traditional calibration method while also accurately characterizing structure of the model residuals. The UA model also successfully determined both parameter uncertainty and total predictive uncertainty for the watershed. Contribution of parameter uncertainty to total predictive uncertainty was found insignificant for the study watershed, underlying the importance of other sources of uncertainty, including data and model structure. Overall, the UA methodology proved promising for sensitivity analysis, calibration, parameter uncertainty, and total predictive uncertainty analysis of urban storm water management models. DOI: 10.1061/(ASCE)HE.1943-5584.0000705. © 2013 American Society of Civil Engineers.

CE Database subject headings: Uncertainty principles; Calibration; Bayesian analysis; Stormwater Management; Parameters; Markov process; Monte Carlo method; Urban areas.

Author keywords: Uncertainty analysis; Calibration; Bayesian approach; Storm water management; Parameter uncertainty; Predictive uncertainty; Markov-chain Monte Carlo.

Introduction

Primarily by increasing imperviousness, urbanization alters natural hydrology of a watershed and negatively impacts ecology, geomorphology, water quality, and socioeconomic functions of the receiving waters (National Research Council (NRC) 2008). Structural and nonstructural methods, generally referred to as storm water control measures (SCMs), are often used to mitigate these impacts. To improve effectiveness of the SCMs, watershed-scale design solutions are advocated as opposed to the conventional approach of selecting and designing SCMs site-by-site (EPA 2007; NRC 2008). Watershed-scale design requires understanding hydrologic and water quality characteristics of individual SCMs as well as the interaction between SCMs of various types, sizes, and relative locations in a watershed, consequently making the design process more challenging.

Computer models could be used for effective design of SCMs at watershed-scale. Models can simulate responses of the watershed and the SCMs considering the factors relevant to the generation and routing of runoff and contaminants. Models, however, must be properly calibrated before their use for planning and management of water resources. The traditional calibration method seeks to identify an optimal set of parameters that forces model simulations closer to the observed counterparts. The basis of this calibration approach is the assumption that the inputs used to build the model, the observations used to evaluate goodness of model simulations, and structure of the model that describes physics of the watershed are all error free. Recent contributions to the water resources literature have seriously questioned the continued usefulness of this classic calibration method (Beven and Freer 2001; Muleta and Nicklow 2005; Kavetski et al. 2006; Vrugt et al. 2008; Montanari et al. 2009). It is acknowledged that hydrologic predictions are plagued with uncertainties arising from errors associated with forcings (inputs), observations, parameters, and model structural inadequacies. Consequently, a prudent evaluation technique is to recognize these uncertainties and to quantify predictive uncertainty and parameter posteriors (Vrugt et al. 2005; Moradkhani and Sorooshian 2008; Gotzinger and Bardossy 2008; Montanari et al. 2009).

During the past two decades, the Generalized Likelihood Uncertainty Estimation (GLUE) technique of Beven and Binley (1992)
and Beven (2006) has found widespread application for uncertainty analysis in water resources. This informal Bayesian approach is simple to implement but has been criticized for being statistically incoherent (Mantovan and Todini 2006; Stedinger et al. 2008; Vrugt et al. 2009b). In response to this, various authors have proposed formal uncertainty analysis (UA) methods that use proper statistics and employ valid likelihood measures (Kuczera and Parent 1998; Thiemann et al. 2001; Vrugt et al. 2003; Schoups and Vrugt 2010). These techniques attempt to provide estimates of the probability density function (PDF) of the model parameters as well as total predictive uncertainty, e.g., through Monte Carlo simulations. For computational efficiency reason, Markov-chain Monte Carlo (MCMC) schemes are preferred to classic Monte Carlo simulations that rely on random sampling.

In addition to efficient and robust sampling schemes, successful UA entails appropriate formulation of the likelihood function. The formal UA applications often reported in the literature make unrealistic assumptions regarding the structure of the residuals (errors) between model simulations and the observed watershed response. Common assumptions include that residuals are: (1) temporally independent (i.e., no correlation between errors of successive time steps); (2) normally distributed; and (3) homoscedastic (i.e., variance of the residuals does not depend on magnitude). Addressing these unrealistic assumptions, a flexible and general formal likelihood (GL) function has been recently described by Schoups and Vrugt (2010).

The objective of this study is to examine effectiveness of a formal Bayesian approach for uncertainty analysis and calibration of the U.S. EPA Storm Water Management Model (SWMM5) (Rossman 2010). The GL function and a recently developed efficient MCMC sampling scheme known as DREAM(za) (Schoups and Vrugt 2010) has been used to identify parameter posteriors and to estimate runoff prediction uncertainty. The methodology is illustrated using the Ballona Creek watershed, a heavily urbanized watershed located in the Los Angeles Basin, California. Effectiveness of the UA method in removing heteroscedasticity and temporal correlation, and in identifying representative PDF for the residuals has been scrutinized. To examine robustness of the UA method for identifying the optimal solutions typically sought by classic calibration approaches, the UA solution [i.e., the maximum likelihood (ML) parameter set and the associated runoff predictions] has been compared with the solution determined by an automated calibration algorithm known as dynamically dimensioned search (DDS) (Tolson and Shoemaker 2007).

Most UA studies in water resources that applied MCMC technique within the Bayesian framework used lumped conceptual models for rainfall-runoff analysis of rural watersheds (Kuczera et al. 2006; Vrugt et al. 2009a; Schoups and Vrugt 2010). Few studies have been reported in spatially distributed modeling (Feyen et al. 2008). Applications to urban watersheds are very limited. Ball (2009) underlined the need for UA-based approaches for evaluation of urban drainage models in the discussion of the conventional calibration effort reported on Ballona Creek watershed by Barco et al. (2008). Freni et al. (2008, 2009a) applied GLUE to an urban drainage model and tested sensitivity of the solutions to likelihood measures (Freni et al. 2009a) and acceptability thresholds (Freni et al. 2008). In a separate study, Freni et al. (2009b) compared performance of Bayesian Monte Carlo method to that of GLUE. Mannina and Viviani (2010) applied GLUE for UA of storm water quality using a conceptual, urban drainage model developed in-house. All these applications of UA to urban drainage models used GLUE, an informal approach whose statistical validity has been questioned (Mantovan and Todini 2006; Stedinger et al. 2008; Vrugt et al. 2008).

Working under the Joint Committee on Urban Drainage established by the International Water Association and the International Association for Hydro-Environment Engineering and Research (IWA/IAHR), the International Working Group on Data and Models has recently published findings of its effort to develop a framework for defining and assessing uncertainties in urban drainage models (Dotto et al. 2012; Deletic et al. 2012). The article by Deletic et al. (2012) underscored the need for consistent use of terminologies and methods for UA of urban drainage models. The authors defined various sources of uncertainties, presented the linkages between the different uncertainty sources, and proposed a framework for UA of urban storm water models. The article by Dotto et al. (2012) compared four different UA methods (three non-Bayesian and one Bayesian) in terms of the posterior PDFs and prediction intervals determined by the methods and their relative computational efficiencies. They showed that the non-Bayesian methods required subjective decisions that affected the UA results, whereas the Bayesian method used erroneous assumption regarding structure of the residuals.

This study is the first, to the best knowledge of the authors, to apply an MCMC scheme that works within formal Bayesian framework for UA of urban watersheds and to apply Bayesian approach to UA of SWMM. Besides demonstrating application of state-of-the-art in UA to urban storm water modeling, the ensuing model will be used for an ongoing study that attempts to develop a simulation-optimization model for watershed-scale design of SCMs for urban watersheds.

Methods and Materials

Uncertainty Analysis and the MCMC Algorithm

The watershed response (e.g., runoff) simulated by a storm water management model, f, such as SWMM5 can be described as

\[ \hat{Y} = f(I, \theta) \]  

(1)

where \( \hat{Y} = n \times 1 \) vector representing the runoff time series \((\hat{y}_1, \ldots, \hat{y}_n)\); \( I \) = matrix of model forcings (e.g., precipitation); and \( \theta \) signifies a d-dimensional vector of model parameters. To test how well \( f \) describes runoff from a watershed, the common practice is to compare the model predictions, \( \hat{Y}_n = \{\hat{y}_1, \ldots, \hat{y}_n\} \) with the corresponding observations, \( Y_n = \{y_1, \ldots, y_n\} \). The difference between the two time series can be represented by a residual vector, \( E_n(\theta) \)

\[ E_n(\theta) = Y - \hat{Y} = \{y_1 - \hat{y}_1, \ldots, y_n - \hat{y}_n\} = \{e_1(\theta), \ldots, e_n(\theta)\} \]  

(2)

The traditional model calibration technique searches for a single optimal combination of parameter values that minimizes this residual time series. With the recognition that model simulations are plagued by many sources of uncertainty, validity of this conventional parameter estimation method has been questioned. A plausible alternative is to account for the various sources of uncertainty and to determine posterior PDF of the parameters, for example, using the Bayesian approach.

From Bayes theorem, the posterior PDF of the parameters, \( p(\theta|I, Y_n) \), can be given as

\[ p(\theta|I, Y_n) \propto L(\theta|Y_n, I)p(\theta) \]  

(3)

where \( L(\theta|Y_n, I) \) denotes the likelihood function that measures how well the model fits the data; and \( p(\theta) = \) prior distribution.
of the model parameters. Different likelihood functions have been proposed, depending on the assumptions made about the statistical properties of the residual vector, \( E_\varepsilon(\theta) \). If the residuals are assumed to be temporally uncorrelated and normally distributed with zero mean and a homoscedastic error standard deviation, \( \sigma_\varepsilon \), the likelihood function takes on the well-known simple least-square (SLS) form (Box and Tiao 1992) as

\[
L(\theta | Y, I) \propto \exp \left( - \frac{1}{2} \sum_{t=1}^{n} e_t(\theta)^2 / \sigma_\varepsilon^2 \right)
\]

(4)

Limitations of the SLS assumptions for hydrologic models have been documented by several authors (Sorooshian and Dracup 1980; Kuczera 1983; Thyer et al. 2009; Schoups and Vrugt 2010), and different proposals have been suggested to relax the assumptions. One of the latest recommendations is the formal likelihood function proposed by Schoups and Vrugt (2010), who described a general error model that embraces temporal correlation, heteroscedasticity, and non-Gaussian nature of the model residuals.

The generalized log-likelihood function of Schoups and Vrugt (2010) can be written as

\[
L(\theta, \varphi | Y, I) = n \log \frac{2\sigma_\varepsilon \omega_{\beta}}{\xi + \xi^{-1}} + \sum_{t=1}^{n} \log \sigma_t - c_\beta \sum_{t=1}^{n} \left[ a_{\xi, t} \right]^{2/(1+\beta)}
\]

(5)

where \( \varphi \) signifies parameters of the error model. Temporal correlation between the residuals is accounted for using a \( p \)th order autoregressive polynomial \( \phi_p(B) \) as

\[
\phi_p(B) e_t = \sigma_t a_t, \quad \text{where} \quad \phi_p(B) = 1 - \sum_{i=1}^{p} \phi_i B_i
\]

(6)

and \( B_i e_t = e_{t-i} \)

where \( \sigma_t \) = standard deviation at time \( t \), and to account for heteroscedasticity, it is assumed to increase linearly with the streamflow \( y_t \) as

\[
\sigma_t = \sigma_0 + \sigma_1 y_t
\]

(7)

where \( \sigma_0 \) and \( \sigma_1 \) are inferred from the data along with model parameters, \( \theta \). Finally, \( a_t \) represents an independent and identically distributed random error with zero mean and a unit standard deviation, whose probability is described by a skew exponential power (SEP) density with skewness (\( \xi \)) and kurtosis (\( \beta \)) parameters

\[
p(a_t | \xi, \beta) = \frac{2\sigma_\varepsilon \omega_{\beta}}{\xi + \xi^{-1}} \exp \left\{ -c_{\beta} \left[ a_{\xi, t} \right]^{2/(1+\beta)} \right\}
\]

(8)

\[
a_{\xi, t} = \xi^{-1} \sign(\mu_t + \sigma_t a_t) (\mu_t + \sigma_t a_t)
\]

(9)

where \( \mu_t, \sigma_\varepsilon, \omega_{\beta}, a_{\xi, t}, \) and \( \omega_{\beta} \) are computed as a function of \( \xi \) and \( \beta \) as described by Schoups and Vrugt (2010).

Besides the likelihood function, a sampling scheme that efficiently identifies posterior PDFs is crucial for effective application of Bayesian-based UAs. Markov-chain Monte Carlo (MCMC) schemes are often used for this application, and improving efficiency of MCMC schemes has been one of the focuses of UA research in the past few years. In this regard, Laloy and Vrugt (2012) developed DREAM(ZS), an MCMC algorithm that capitalized on the strength of the DiffeRential Evolution Adaptive Metropolis (DREAM) (Vrugt et al. 2008). Effectiveness and efficiency of DREAM(ZS) for posterior sampling has been reported in several studies. Schoups and Vrugt (2010) applied the GL function and DREAM(ZS) for rainfall-runoff analysis of two watersheds by using a lumped conceptual model. This study examines DREAM(ZS) and GL for UA of SWMM5 using the Ballona Creek watershed, which is one of the most urbanized watersheds in the world with approximately 83% developed (Bay et al. 2003). Schoups and Vrugt (2010) provides further description of DREAM(ZS).

**Single-Objective Calibration**

Single-objective automated calibration was performed, primarily, to compare solutions of the conventional model calibration technique to those identified by GL and DREAM(ZS). The dynamically dimensioned search (DDS) (Tolson and Shoemaker 2007) was used to identify optimal values of SWMM5 runoff parameters. DDS has been developed to improve efficiency of calibrating computationally demanding models. DDS is a simple, single-objective, heuristic search method that starts by globally searching the feasible region and incrementally localizes the search space as the number of simulations approaches the maximum allowable number of simulations (the only stopping criteria used by the algorithm). Progress from global to local search is achieved by probabilistically reducing the number of model parameters modified from their best value obtained thus far. New potential solutions are created by perturbing the current parameter values of the randomly selected model parameters only. The best solution identified thus far is maintained and is updated only when a solution with superior value of the objective function is found.

DDS requires minimal algorithmic parameter tweaking because the only parameters to set are the maximum number of model evaluations and the scalar neighborhood size perturbation parameter \( r \) that defines the random perturbation size standard deviation as a fraction of the decision variable range. The recommended value of 0.2 (Tolson and Shoemaker 2007) has been used for \( r \) in this study. Efficiency and effectiveness of DDS has been reported by Tolson and Shoemaker (2007) and Muleta (2010), who compared its performance to that of widely used optimization methods including the Shuffled Complex Evolution-University of Arizona (SCE-UA) (Duan et al. 1992) and the Genetic Algorithms (Holland 1975). For this study, DDS has been integrated with SWMM5 to calibrate runoff for the study watershed.

**EPA Storm Water Management Model**

SWMM was first developed in 1971, and it continues to be widely used throughout the world for planning, analysis and design of storm water runoff, combined sewers, sanitary sewers, and other drainage systems (Rossman 2010). SWMM5, the latest version of SWMM, simulates hydrology, hydraulics, and water quality of urbanized and nonurbanized watersheds. The hydrologic processes modeled include precipitation (rainfall or snow fall), evaporation, surface runoff, infiltration, groundwater flow, and snowpacks and snowmelt. Both single event and continuous simulations can be performed, accounting for spatial and temporal variability in the climate, soil, land use, and topography in the watershed. Surface runoff is estimated using the nonlinear reservoir method in which surface runoff occurs only when the depth of the overland flow exceeds the maximum surface storage provided by initial abstractions, including depression storage and interception, in which case the runoff rate is estimated using Manning’s equation, Horton (1937), Green and Ampt (1911), and the Curve Number methods.
(Soil Conservation Service 1964) are available to model infiltration losses.

Runoff quality, including buildup and washoff of pollutants, can be simulated by using various approaches from both developed and nondeveloped land uses. The runoff quantity and quality simulated from a subwatershed and the wastewater loads (if any) assigned to the receiving nodes are added and then transported by using either steady, kinematic wave, or dynamic wave routing through a conveyance system of pipes, channels, storage/treatment devices, pumps, and hydraulic regulators such as weirs, orifices, and other outlet types. Hydraulic conditions of any level of complexity, including those experiencing backwater effect, flow reversal, and pressurized flow, can be accommodated. In addition, the capability to model the commonly used low impact developments (LIDs), including porous pavements, bioretention cells, infiltration trenches, vegetative swales, and rain barrels has been recently added to SWMM5. For this study, source code of SWMM5 has been integrated with the UA and the single-objective calibration methods previously described.

Application Watershed and Data

The Ballona Creek watershed (Fig. 1) is used to illustrate the methods described in this study. Total drainage area of the Ballona Creek watershed is approximately 337 km². For this study, the upper 230 km² of the Ballona Creek watershed (i.e., the portion that drains to the streamflow gauging station used in this study) has been modeled. Approximately 90% of the modeled watershed is developed, and its land-use distribution consists of 60% residential, 10% commercial, 3.5% industrial, and 11% open space (Amenu 2011). The open spaces are in the Santa Monica Mountains, located in the northern part of the watershed. The drainage system is characterized by extensive networks of storm drains that collect storm water from the watershed and convey it to the Ballona Creek, a 14.5 km (9-mi)-long flood protection channel that discharges to the Santa Monica Bay [Los Angeles County Department of Public Works (LACDPW) 2011]. The watershed has been identified as the major source of non-point-source pollution to the Santa Monica Bay (Stenstrom and Strecker 1993; Stein and Tiefenthaler 2005).

The data needed to build SWMM5 have been collected from various sources. A digital elevation model, land-use map, and an imperviousness map were obtained from the USGS seamless data warehouse (USGS 2013), and soil survey geographic (SSURGO) soil map has been obtained from the Natural Resources Conservation Service (NRCS) soil data mart (NRCS 2013). Because of the difficulty to accurately delineate urban subwatersheds from digital elevation models alone (Gironás et al. 2010), the subwatershed delineation obtained from the LACDPW were used for this study. The LACDPW delineation (which was created through a comprehensive hydrologic study based on USGS topo quads, as built drawings, and field surveys) divided the watershed into 134 subwatersheds. For this study, the number of subwatersheds was further reduced to 92 by merging smaller subwatersheds (area less than 0.41 km²) to the adjoining subwatershed. Subcatchment information such as area, slope, and flow length were extracted from the digital elevation model. The soil, land-use, and imperviousness

![Fig. 1. Location map of the Ballona Creek Watershed](image)
maps were superimposed onto the subwatersheds to extract SWMM5 runoff parameters including percent imperviousness, infiltration parameters, and Manning’s roughness coefficient. Rainfall data at three gauges (Fig. 1) and streamflow data for a monitoring station that drains approximately 70% of the Ballona Creek watershed were obtained from the LACDPW for 15 years (i.e., 1996–2010) at 15-min intervals. Both rainfall and streamflow data were collected using automatic gauges that are equipped with real-time data telemetry and electronic data loggers (Amenu 2011). Proximity and altitude criteria were used to define the rain gauge that represents each subcatchment. The climate of the watershed can be characterized as semiarid, with average annual rainfall of approximately 380 mm and temperature of approximately 18°C. Rainfall season for the region spans from October to April. The elevation of the watershed varies from 750 m above mean seal level (AMSL) at the Santa Monica Mountains to 0 m AMSL at its discharge to the Santa Monica Bay.

The watershed consists of an extensive network of storm drains (underground pipes and open channels) designed for flood protection purposes. With the assumption that overland flow from each of the 92 subwatersheds directly flows to a storm drain inlet located at the outlet of the subwatershed, only 72 larger storm drains were considered in this model. In reality, each subwatershed may contain numerous streets, swales, and minor storm drains that can play significant roles in routing of runoff and contaminants within the subcatchment. Neglecting the minor drainage systems and modeling only the major drainage systems, as done in this study, can have an effect on the hydraulics of the drainage system. For example, the travel time could get longer as the faster conduit flows are replaced with the slower overland flows under this assumption. The approach used in this study is, however, commonly used to simplify modeling complexity and to reduce the cost associated with data collection (Gironás et al. 2010). The study by Burian et al. (2000) was used for storm drain information, including shape, size, slope, and length.

Methodology

Simulation Durations and Analysis Methods

For both the DDS and GL-DREAM(\textit{ZS}) methods, rainfall and streamflow data from January 17, 2010, to January 23, 2010, were used for calibration. Verification of the solutions was performed using data from January 23, 2008, to January 28, 2008. The event used for calibration had rainfall depth of approximately 12.4 cm in 7 days, and the verification event produced 12.6 cm of rainfall in 6 days. According to National Oceanic and Atmospheric Administration (NOAA) Atlas 14 (NOAA 2013), for a station in Ballona Creek, mean rainfall depth for a 2-year, 7-day event is approximately 12.1 cm, which is comparable to the calibration/verification storm events used for this study. Both streamflow and rainfall data are available at 15-min intervals. The curve number method was selected for infiltration modeling as the CN values (primary parameter for the curve number method) can be determined more readily, compared to Horton or Green-Ampt parameters, from the land cover and soil maps available for the watershed. Because the GL-DREAM(\textit{ZS}) algorithm used for the UA requires running SWMM5 repetitively (up to hundreds of thousands of runs) to converge, computational time is a significant concern. As such, kinematic wave routing was selected to reduce computational burden of the dynamic wave routing option.

Likewise, continuous (long-term) simulation was not considered because of computational concern. Short-term simulation (i.e., duration of 7 days for calibration and 6 days for verification) were used for both single-objective calibration and uncertainty analysis. Most studies that reported on calibration of urban drainage models used single-event simulations with a typical duration of a day or less (Barco et al. 2008; Fang and Ball 2007). The simulation durations considered in this study for both calibration and verification cases are therefore, significant improvements compared to single-event simulations. Although single-event and the short-duration simulation pursued in this study may suffice for certain applications such as flood control, continuous (e.g., multiple year duration) simulation models are more appropriate for applications that are sensitive to long-term watershed characteristics (e.g., contaminant buildup and washoff processes).

Parameters

A total of 11 SWMM5 runoff parameters were considered for calibration and uncertainty analysis. Values of the parameters vary from subwatershed to subwatershed depending on soil, land use, imperviousness, topography and/or other characteristics of the subwatershed. Initial values of the parameters have been extracted for each subwatershed from the soil, land-use, imperviousness, and topography maps using geographic information system (GIS). During both calibration and uncertainty analysis, these initial (baseline) values were altered by multiplying the parameters by the respective adjustments proposed by the calibration and the UA algorithm. This way, the initial values would be scaled up or down while preserving the heterogeneity determined from watershed characteristics. The parameters were assumed to follow uniform distribution as done in Muleta and Nicklow (2005), and lower and upper percentage adjustment bounds were assigned based on literature and engineering judgment (Rossman 2010; Barco et al. 2008). A list of the parameters and the assumed adjustment ranges are given in Table 1.

Objective Function and Efficiency Criteria

The streamflow measured at the Swatelle station, shown in Fig. 1, was used for calibration and uncertainty analysis. Mean absolute error (MAE), Eq. (10), was used as objective function for the single-objective calibration. MAE was selected as the objective function based on the findings of Muleta (2012), which compared relative effectiveness of the efficiency criteria commonly used in hydrologic modeling to describe goodness of model performances. According to the study, efficiency criteria such as MAE that describe the absolute deviation between observations and model simulations were found to be robust. Goodness of the calibration result was further assessed using additional efficiency criteria including the Nash-Sutcliffe efficiency (NSE) criterion (Nash and Sutcliffe 1970) described in Eq. (11), percent bias (PBIAS) [Eq. (12)], and total volume of runoff

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |Y_i - O_i| \\
\text{NSE} = 1 - \frac{\sum_{i=1}^{N} (Y_i - O_i)^2}{\sum_{i=1}^{N} (O_i - O_{\text{mean}})^2} \\
\text{PBIAS} = 100 \frac{\sum_{i=1}^{N} (O_i - Y_i)}{\sum_{i=1}^{N} O_i}
\]
Table 1. Model Parameters and Ranges Used for Calibration and Uncertainty Analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWMM parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Width</td>
<td>Subcatchment width (m)</td>
<td>−90</td>
<td>150</td>
</tr>
<tr>
<td>Slope</td>
<td>Subcatchment slope (%)</td>
<td>−20</td>
<td>20</td>
</tr>
<tr>
<td>%Imperv</td>
<td>Percentage of impervious area (%)</td>
<td>−15</td>
<td>15</td>
</tr>
<tr>
<td>N-Imper</td>
<td>Manning n for impervious area</td>
<td>−50</td>
<td>100</td>
</tr>
<tr>
<td>N-Perv</td>
<td>Manning n for pervious area</td>
<td>−50</td>
<td>100</td>
</tr>
<tr>
<td>Dstore-Imper</td>
<td>Depression storage for impervious area (mm)</td>
<td>−95</td>
<td>100</td>
</tr>
<tr>
<td>Dstore-Perv</td>
<td>Depression storage for pervious area (mm)</td>
<td>−95</td>
<td>300</td>
</tr>
<tr>
<td>%Zero-Imper</td>
<td>Percent of the impervious area with no depression storage (%)</td>
<td>−100</td>
<td>100</td>
</tr>
<tr>
<td>Curve number</td>
<td>NRCS runoff curve number</td>
<td>−20</td>
<td>20</td>
</tr>
<tr>
<td>Drying Time</td>
<td>Time for a fully saturated soil to completely dry (days)</td>
<td>−95</td>
<td>100</td>
</tr>
<tr>
<td>Conduit n</td>
<td>Manning’s roughness coefficient for conduit</td>
<td>−50</td>
<td>100</td>
</tr>
<tr>
<td>Error parameters</td>
<td>Parameter values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ₀</td>
<td>Heteroscedasticity intercept (m/s)</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>σ₁</td>
<td>Heteroscedasticity slope</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>φ₂</td>
<td>Lag one autocorrelation coefficient</td>
<td>−1</td>
<td>1</td>
</tr>
<tr>
<td>φ₃</td>
<td>Lag two autocorrelation coefficient</td>
<td>−1</td>
<td>1</td>
</tr>
<tr>
<td>β</td>
<td>Kurtosis parameter</td>
<td>−1</td>
<td>1</td>
</tr>
</tbody>
</table>

against which performance of the model simulations is evaluated; and N = number of data points (observations).

Parameter Posteriors and Predictive Uncertainty

Posteriors of the 11 runoff parameters were estimated with GL-DREAM(ZS) and the GL function. Six additional error model parameters were considered for the GL. These include σ₀ and σ₁ in Eq. (7), to explicitly account for heteroscedasticity of model residuals, the kurtosis parameter, β, in Eq. (5) to account for non-normality of the residuals, and φ₁, φ₂, and φ₃ in Eq. (8) to allow for autocorrelation of the residuals. It was initially assumed that the residual distribution is not skewed, setting skewness parameter, ξ, equal to 1, and then this assumption was checked a posteriori. A total of 100,000 SWMM5 simulations were used to sample the posterior distribution of the parameters. Convergence of GL-DREAM(ZS) to a stable posterior PDF was monitored using the R statistic of Gelman and Rubin (1992). Convergence is declared when Rⱼ ≤ 1.2 for all j = 1, . . . , d, where d is the number of model parameters being analyzed (i.e., 17 in this study). The last 5,000 GL-DREAM(ZS) runs that meet the convergence criteria were extracted, and parameter posteriors were determined and reported for each individual parameter.

Once the posterior distribution of the model parameters is known, runoff predictive uncertainty can be estimated by propagating the different samples of the posterior distribution through the SWMM5 model and reporting the respective prediction uncertainty ranges (e.g., 95% confidence interval). This prediction interval, however, represents parameter uncertainty only; it doesn’t consider other sources of error, including model structure, forcing data, and calibration data uncertainty. Total predictive uncertainty was calculated using the methodology described in Schoups and Vrugt (2010) based on the error model parameters determined by GL-DREAM(ZS). The ML parameter values determined by GL-DREAM(ZS) are benchmarked against the calibration results obtained using DDS and are also compared in terms of their ability to fit different parts of the hydrograph.

Results and Discussion

Single-Objective Calibration

Results of the single-objective calibration are summarized in Fig. 2 and Tables 2 and 3. Fig. 2 compares the streamflow simulated using the optimal parameter values identified by DDS with the streamflow observed at Swatelle station for both calibration [Figs. 2(a-d)] and verification [Figs. 2(e-h)] periods. Performance of the calibrated model was also tested using the efficiency criteria given in Table 2. The graphical comparison and the efficiency criteria indicate that the single-objective calibration performed very well for both calibration and verification periods. However, as shown in Fig. 2, a closer look into the characteristics of the residuals (i.e., the difference between the observed streamflow and the streamflow simulated by the calibrated model) depicts that the assumptions (i.e., homoscedasticity, Gaussian distribution, and temporal independence) made by the objective functions almost always used in model calibrations, including the MAE used for this study, are unjustified. The residuals exhibited heteroscedasticity (i.e., they increase with the magnitude of streamflow) as shown in Fig. 2(b), they do not follow Gaussian distribution [Fig. 2(c)], and they are temporally correlated [Fig. 2(d)] for both calibration and verification periods. Similar findings have been reported by other studies including those by Schoups and Vrugt (2010) and regarding characteristics of the residuals generated from solutions of the traditional calibration methods.

In addition to relying on unrealistic assumptions on residuals, the conventional calibration models attempt to identify a single best solution based on the assumption that data, model structure, and model parameters are all error-free. It is now a common knowledge that input and output data are subjected to substantial uncertainty; no model structure is a true representation of the watershed being studied, and optimal parameter sets are not unique for a given watershed (i.e., multiple sets of parameters can be equally good for the watershed). As such, these calibration methods provide no information on reliability of the optimal solution. The uncertainty analysis model described in this study has been developed to address these vital limitations.

Parameter Uncertainty

As previously described, 100,000 SWMM5 simulations were run for the GL-DREAM(ZS) UA model. Fig. 3 shows progress of the R statistic of Gelman and Rubin (1992) that has been used to test convergence of the UA runs. The plot indicates that the 100,000 model runs used for the analysis were sufficient to meet the convergence criteria of R ≤ 1.2 for all the SWMM5 and error model parameters considered for the study. Fig. 4 shows the posterior histograms obtained for the parameters using the last 3,000 simulations. As shown in Fig. 3, the convergence criteria was met after approximately 62,000 model simulations, implying that the last 3,000 simulations used to generate the posterior histograms have met the convergence criteria, and each of these parameter sets represents a reasonable SWMM5 model for the watershed.
Fig. 2. Comparison of observed and simulated streamflow and diagnostic plots of the residuals obtained using DDS for both calibration (a–d) and verification (e–h) periods: (a and e) comparison between observed (dots) and simulated (solid line) streamflow; (b and f) illustration of heteroscedasticity of the residuals; (c and g) comparison of observed PDF of the residuals to normal distribution; (d and h) illustration of autocorrelation of the residuals; dashed lines in (d) and (h) show lower and upper bounds of the 95% confidence interval.

Fig. 4(a) offers important information regarding the relative importance of the SWMM5 parameters considered for the analysis. Except for percentage imperviousness (Imperv), depression storage for impervious subareas (Dstore-Imperv), and percentage of the impervious subarea with no depression storage (% Zero Imperv), the uncertainty bound of the other SWMM5 parameters is very wide. Nonetheless, the streamflow simulated by the last 3,000 parameter sets did not show significant variability as shown in Fig. 5. This suggests that the SWMM5 parameters that exhibited wide uncertainty range do not have substantial effect on rainfall-runoff characteristics of the Ballona Creek watershed. This has practical implication, for example, in terms of prioritizing resources on data collection. Availability of more accurate data that characterize the insensitive parameters may not help in improving accuracy of run-off simulation for the watershed. On the contrary, uncertainty of the model predictions can be further reduced from having more reliable imperviousness data. Unlike the SWMM5 parameters, the error model parameters given in Fig. 4(b) produced narrow uncertainty range indicating their identifiability.

Tables 2 and 3 compare the solution of the GL-DREAM(2S) model with the single-objective calibration results. Performance of the maximum likelihood solution has been summarized using several efficiency criteria in Table 2. The results show that the ML solution performed very well but not as well as the DDS solution. This is not surprising because the objective function used by

| Table 2. Efficiency Criteria Values Obtained Using DDS and GL-DREAM(2S) |
|-----------------------------|----------------|----------------|----------------|
| Method          | Period | MAE (m³/s) | NSE (%) | Bias (%) | Volume (mm) | Volume (mm) |
| DDS             | C      | 6.89       | 0.94    | 11.15    | 59.6        | 66.2        |
|                 | V      | 9.79       | 0.90    | −18.82   | 83.6        | 68.9        |
| GL-DREAM(2S)   | C      | 7.74       | 0.93    | 10.94    | 59.6        | 66.1        |
|                 | V      | 10.57      | 0.88    | −20.40   | 83.6        | 67.6        |

Note: C= calibration; V= verification.
Table 3. Percentage Adjustments Obtained by DDS, the ML, and the 95% Confidence Interval Obtained by GL-DREAM(ZS) and Optimal Parameter Values Determined for One of the Subcatchments in the Ballona Creek Watershed

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DDS</th>
<th>ML</th>
<th>GL-DREAM(ZS) Parameter values*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width</td>
<td>137.1</td>
<td>117.0</td>
<td>-12.6</td>
</tr>
<tr>
<td>Slope</td>
<td>2.7</td>
<td>-7.4</td>
<td>-40.1</td>
</tr>
<tr>
<td>% Imperv</td>
<td>-14.9</td>
<td>-15.0</td>
<td>-15.0</td>
</tr>
<tr>
<td>N-Imperv</td>
<td>32.1</td>
<td>79.3</td>
<td>5.8</td>
</tr>
<tr>
<td>N-Perv</td>
<td>-42.8</td>
<td>38.6</td>
<td>-46.5</td>
</tr>
<tr>
<td>Dstore-Imperv</td>
<td>99.8</td>
<td>98.2</td>
<td>84.9</td>
</tr>
<tr>
<td>Dstore-Perv</td>
<td>234.7</td>
<td>1.8</td>
<td>72.2</td>
</tr>
<tr>
<td>% Zero-Imperv</td>
<td>-99.9</td>
<td>-100.0</td>
<td>-100.0</td>
</tr>
<tr>
<td>Curve Number</td>
<td>-3.2</td>
<td>8.4</td>
<td>-17.5</td>
</tr>
<tr>
<td>Drying Time</td>
<td>-6.9</td>
<td>-93.1</td>
<td>-68.0</td>
</tr>
<tr>
<td>Conduit n</td>
<td>31.0</td>
<td>28.0</td>
<td>34.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>Initial value</th>
<th>Optimal value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width</td>
<td>126.3</td>
<td>460</td>
<td>998</td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>127</td>
<td>1.79</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td>% Imperv</td>
<td>70</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Imperv</td>
<td>99.5</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Perv</td>
<td>99.0</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dstore-Imperv</td>
<td>99.9</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dstore-Perv</td>
<td>293.6</td>
<td>1.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Zero-Imperv</td>
<td>100</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curve Number</td>
<td>69</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drying Time</td>
<td>89.7</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduit n</td>
<td>96.0</td>
<td>0.013</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Initial values refer to the actual initial parameter values assigned to one of the subcatchments, and the optimal values are determined from the ML percent adjustments and the initial parameter values. See Table 1 for units.

This shows that the insensitive parameters have minimal impact on runoff from the watershed and thus on the likelihood function. The ML adjustments recommended for these insensitive parameters are meaningless as any adjustment within the wide range of the posterior PDFs will produce almost identical runoff and likelihood function value.

One interesting observation is the recommendation by both ML and DDS solutions to decrease percent imperviousness of the watershed by approximately 15% so that model simulations closely match observed runoff. Given that only the largest 72 storm drains were considered in the study by ignoring hundreds of smaller storm drains and channels in the watershed, one would intuitively expect a solution that speeds up travel time (e.g., increase in percent imperviousness and steeper slope) to compensate for the ignored storm drains on travel time. The suggested decrease in percent imperviousness is believed to be related to the assumption made regarding connectivity of the impervious and pervious subareas in each subwatershed. Both subarea types were assumed to directly flow to the outlet of each subwatershed, whereas in reality only a fraction of the impervious subareas may be directly connected to the engineered drainage system. This assumption might have led to overestimation of the effective percent imperviousness of the watershed in the model.

Predictive Uncertainty

Understanding the total predictive uncertainty associated with model simulations is very essential for decision makers. Modeling uncertainties are believed to arise from imprecise knowledge of the temporal and spatial variability of input and observed system response, from the assumptions and simplifications made in the simulation model to represent physical processes in the watershed, and owing to the parameter uncertainty described in the previous section. In the past, parameter uncertainty has been assumed to represent all uncertainty sources (Beven and Freer 2001; Muleta and Nicklow 2005). However, as shown in Fig. 5, parameter uncertainty for the study watershed is very minimal indicating that, at least for the Ballona Creek watershed, parameter uncertainty alone cannot represent the total predictive uncertainty associated with simulation models. Similar findings have been reported by previous studies including Kuczera et al. (2006) using different hydrologic models and application watersheds.

Fig. 6(a) shows a 95% confidence interval for the total predictive uncertainty generated for the calibration [Figs. 6(a–d)] and verification [Figs. 6(e–h)] periods. The predictive uncertainty has been determined using the last 3,000 parameter sets identified by the GL-DREAM(ZS) model as previously described. Fig. 6 indicates substantial uncertainty for the watershed especially for low flow simulations. Wider uncertainty bound for low flows might have been obtained because the likelihood function used in the UA (i.e., the GL function) is biased towards peak flows that seem to have been simulated with a higher degree of reliability (i.e., narrower bounds). Overall, the total predictive uncertainty bounds seem reasonably accurate because they bracketed more than 75% of the observed data, albeit lower than the theoretically expected value of 95%, for both calibration and verification periods. This indicates that the MCM-based formal, Bayesian methodology pursued in this study is promising for UA of urban drainage models.

Fig. 6 also shows diagnostic plots of the residuals derived from the ML solution. Fig. 6(b) shows that the residuals are not sensitive to the magnitude of streamflow, indicating that heteroscedasticity has been removed by the GL function. Fig. 6(c) clearly shows the Laplace (i.e., double-exponential) distribution used by the error model is consistent with the PDF of the residuals of the ML.
solution. Temporal dependence of the residuals is shown in Fig. 6(d), which indicates that the residuals still exhibit substantial dependence at lag-one and lag-two autocorrelations. However, the temporal correlation has been significantly reduced compared to the DDS solution shown in Fig. 2(d). Given the short simulation time interval (i.e., 15-min) used for the study, which is typical in urban drainage modeling, the difficulty of removing temporal correlation in its entirety is understandable. Generally, the diagnostic plots demonstrate that the assumptions made by the GL-DREAM\(_{(ZS)}\) model regarding the characteristics of the residuals are consistent with properties of the residuals derived from the ML solution.

**Conclusions**

This paper describes an MCMC-based formal, Bayesian methodology for parameter uncertainty and total predictive uncertainty analysis of a widely used urban storm water management model. The methodology has been illustrated using the Ballona Creek watershed, a heavily urbanized watershed located in the Los Angeles Basin, California. Solution of the UA model has been compared with the optimal solution typically derived by using the traditional calibration methods widely used in water resources modeling. Furthermore, validity of the assumptions commonly made with regard to characteristics of model residuals in the...
Objective functions often used for model calibration have been examined. Flexibility of the likelihood function used in the UA model to accommodate the characteristics of the residuals has been demonstrated. The subsequent paragraphs summarize major conclusions of the study.

The runoff simulated using the optimal solution identified by the single-objective calibration attempt was in good agreement with the observed counterparts when evaluated graphically and by using several goodness-of-fit measures. However, diagnostic analysis of the residuals indicates that the assumptions of homoscedasticity, temporal independence, and Gaussian distribution commonly made in such traditional calibration models are unjustified. On the other hand, the maximum likelihood solutions determined using the UA model produced runoff simulations that are of comparable accuracy with that of the single-objective calibration solutions while accurately characterizing the structure of the model residuals. The assumptions made by the error model used in the UA methodology were found consistent with the characteristics of the residuals generated from the ML solution.

**Fig. 5.** Comparison of observed runoff (dotted) to the 95% confidence interval bounds (lines) determined using GL-DREAM_{ZS}, considering only parameter uncertainty for calibration and verification periods.

**Fig. 6.** Total predictive uncertainty (95% confidence interval) and diagnostic plots of the residuals obtained using GL-DREAM_{ZS} for calibration (a–d) and verification (e–h) periods: (a and e) comparison between observed streamflow (dots) to the 95% confidence bounds (solid line); (b and f) show that the residuals are homoscedastic; (c and g) comparison of observed PDF (dotted) of the residuals to the assumed distribution (solid line) in GL; (d and h) show autocorrelation of the residuals; dashed lines in (d) and (h) show lower and upper bounds of the 95% confidence interval.
In addition to accurately simulating runoff and properly characterizing the residuals, the UA model has successfully determined parameter uncertainty and total predictive uncertainty. The parameter posteriors showed that eight of the 11 SWMM5 parameters considered for the analysis exhibited wide uncertainty bound, whereas the runoff simulated for the watershed considering parameter uncertainty alone showed no appreciable variability. This suggests that runoff is sensitive only to three (i.e., Imperv, Dstore Imperv, % Zero Imperv) of the 11 parameters. Additionally, the ML solution identified by the UA model and the optimal solution determined by DDS showed good agreement only for four (Imperv, Dstore Imperv, % Zero Imperv, Conduit n) of the 11 SWMM5 parameters confirming nonidentifiability of the insensitive parameters. Results also suggest that contribution of parameter uncertainty to total predictive uncertainty is insignificant for the study watershed, underlying the importance of the other sources of predictive uncertainty for Ballona Creek watershed.

The 95% confidence interval determined for total predictive uncertainty using the UA model bracketed the majority of the observed data, demonstrating reasonable accuracy of the UA result. Satisfactory total predictive uncertainty bounds were generated for both calibration and verification periods although the verification period results seem less adequate. Overall, the UA methodology proved promising for sensitivity analysis, calibration, parameter uncertainty, and total predictive uncertainty analysis of urban storm water drainage models at least for the short-simulation durations considered in this study. Applications to additional watersheds in other hydroclimatic regions can help further examine this potential. The subsequent study will investigate application to continuous simulations and on ways to use the predictive uncertainty in decision making such as for optimal SCM design applications.

References


