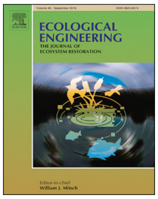
Ecological Engineering 122 (2018) 16–26

Contents lists available at ScienceDirect

Ecological Engineering

journal homepage: www.elsevier.com/locate/ecoleng



Input uncertainty on watershed modeling: Evaluation of precipitation and air temperature data by latent variables using SWAT

Haw Yen[a,*], Ruoyu Wang[b], Qingyu Feng[c], Chih-Chieh Young[d,e], Shien-Tsung Chen[f], Wen-Hsiao Tseng[g], June E. Wolfe III[a], Michael J. White[h], Jeffrey G. Arnold[h]

[a] *Blackland Research and Extension Center, Texas A&M University, 720 East Blackland Road, Temple, Texas 76502, USA*
[b] *Department of Land, Air and Water Resources, University of California, Davis, Davis 95616, USA*
[c] *Department of Agricultural and Biological Engineering, Purdue University, West Lafayette IN 47907, USA*
[d] *Department of Marine Environmental Informatics, National Taiwan Ocean University, Keelung City 202, Taiwan*
[e] *Center of Excellence for Ocean Engineering, National Taiwan Ocean University, Keelung City 202, Taiwan*
[f] *Department of Water Resources Engineering and Conservation, Feng Chia University, Taichung City 40724, Taiwan*
[g] *Water Resources Planning Institute, Water Resources Agency, Ministry of Economic Affairs, Taiwan*
[h] *Grassland, Soil & Water Research Laboratory, USDA-ARS, 808 East Blackland Road, Temple, Texas 76502, USA*

ARTICLE INFO

Keywords:
Input uncertainty
Latent variables
Model calibration
Precipitation
Air temperature
SWAT
IPEAT

ABSTRACT

Latent variables (i.e., normally distributed random noise) provide valuable information regarding model input uncertainty. Watershed processes have been explored with sophisticated simulation models in the past few decades and researchers have found that incorporating the uncertainty attributed to forcing inputs, model parameters, and measured data, can help improve simulation results, however, not in all cases. Latent variable use requires careful consideration to determine if results are better or worse. In this study, latent variables were implemented to both precipitation and air temperature data to investigate the influence on model predictions and associated predictive uncertainty by using the Soil and Water Assessment Tool (SWAT). Results indicated that model predictions in terms of statistics, behavior solutions, and predictive uncertainty were substantially affected by applying latent variables on precipitation data but it does not guarantee improved performance. On the other hand, model responses did not denote similar performance by conducting the same approach to air temperature data. Ultimately, incorporating latent variables *a priori* proportionally may or may not improve model predictive uncertainty. Researchers should carefully consider latent variable potential benefits on model predictions before committing to further work or making important model-supported decisions.

1. Introduction

Large-scale watershed simulation models are being used to provide scientifically credible solutions for various challenging environmental issues such as the Toledo water crisis (Dungjen and Patch, 2014; Feng et al., 2018). Over the past few decades, intensive efforts have been made to develop/improve hydrological models (Rossman 2005; Arnold et al., 2012; Williams et al., 2012). Successfully setting up, executing, and interpreting the results of these complex watershed models requires significant effort and time. Preparing model inputs into specified format and files is labor intensive (Neitsch et al., 2011) and accurate model calibration is needed to identify parameter values (Yen et al., 2015; Guo et al., 2018a,b). When taking a conventional deterministic watershed modeling approach, the importance of uncertainty sources such as: forcing inputs, measurement data, model parameters and structure, is paramount (Harmel et al., 2010). Particularly, the interpretation of modeling results may lead to poor decisions without taking the uncertainty sources into account (Yen et al., 2014a; Wang et al., 2014).

In general, precipitation and air temperature are among the most important inputs for hydrological models (Neupane and Kumar, 2015). Precipitation is the primary driving force of surface and subsurface flow, sediment movement, and nutrient cycling processes (Balin et al., 2010; Wang, 2010). Variations in temperature heavily influence evapotranspiration and humidity, drives wind and rainfall patterns, and affects ice formation and melting processes. As a result, precipitation and temperature both heavily influence streamflow, which is often used to judge the successfulness of a hydrological model in accurately representing the system of interests (Ficklin et al., 2014). On the other hand, quality of precipitation and air temperature data may be compromised spatially or temporally by other factors including: wind,

---

* Corresponding author.
E-mail address: hyen@brc.tamus.edu (H. Yen).

https://doi.org/10.1016/j.ecoleng.2018.07.014
Received 23 February 2018; Received in revised form 25 June 2018; Accepted 17 July 2018

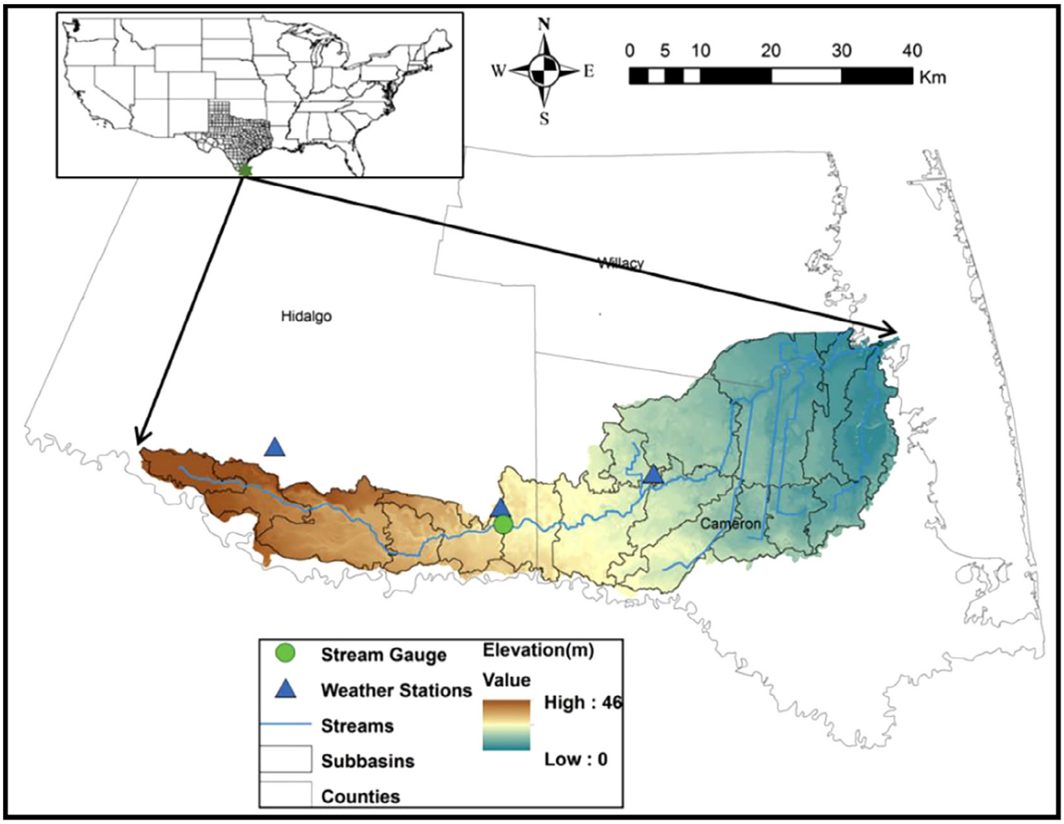


Fig. 1. Stream gauge and weather station locations of the Arroyo Colorado Watershed, Texas, USA (Seo et al., 2014).

mechanical matters, and other reasons (McMillan et al., 2011). In general, temperature data is less variable since its measurement is less affected by mechanical issues. A common approach of performing model simulation is to use input data with mixed confidence levels (e.g., precipitation data may be assumed to be more uncertain than temperature data). However, studies in exploring the significance between different data sources, in terms of impact on model predictions can be rarely found.

Input uncertainty on climate data is typically studied in scenario-based approaches such as climate change scenarios in which climate data has been clustered in varying categories with deterministic format (Awan and Ismaeel, 2014). For instance, the Representative Concentration Pathways (RCP) emission scenarios are conducted for investigation of projected changing climate in predefined levels such as RCP4.5 or RCP8.5 (Knutti and Sedlacek, 2013; Wang et al., 2015). However, input uncertainty is infrequently examined in explicit scheme with stochastic components while conducting watershed modeling. The concept of latent variables was first proposed by Kavetski et al. (2002) to incorporate precipitation data and the associated input uncertainty into the Bayesian Total Error Analysis (BATEA) framework. The values of input data (precipitation) is reformed by white noise (normally distributed random number) with predefined mean (θ) and variance (σ^2) as latent variables. However, BATEA is difficult to apply in practice since latent variables are required in each time step so that the total number of latent variables increases substantially with the length of temporal data. The use of latent variables was further modified by Ajami et al. (2007) in the Integrated Bayesian Uncertainty Estimator (IBUNE) framework where the same set of latent variable ($\theta \in [0.9, 1.1]$; $\sigma^2 \in [10^{-5}, 10^{-3}]$) will be assigned in all temporal series. In addition, the Integrated Parameter Estimation and Uncertainty Analysis Tool (Yen et al., 2014a), which is a recently developed framework for calibration and uncertainty analysis, is adopting the same calculation pattern and linked with large-scale watershed simulation model (e.g., Soil and Water Assessment Tool, SWAT) for broader explorations (e.g., measurement uncertainty (Yen et al., 2014a)).

Currently, applications of latent variable are limited to examine input uncertainty attributed to precipitation data with no assessment of other model input such as air temperature (Yen et al., 2014b). Therefore, the knowledge of interactions among different input data upon model predictions and the associated predictive uncertainty is still not thoroughly investigated. The primary goal of this study is to explore the potential influence of watershed simulation model performance by incorporating latent variables (stochastic approach) on precipitation and air temperature inputs. Limitations of deterministic approaches (or, scenario-based) can be resolved since the values of latent variables will change during the model calibration process. Specifically, two objectives are defined: (i) to examine model performance affected by latent variable on precipitation and air temperature data in terms of error statistics; and (ii) to quantify and compare predictive uncertainty in two sources of input data. The proposed research is one-of-a-kind since air temperature and the corresponding impact on model predictions was never rigorously evaluated (in terms of explicitly considering input uncertainty with watershed simulation process). In this study, the SWAT model is used to conduct watershed simulation on the Arroyo Colorado Watershed (ACW) in Southern Texas, USA. Eight scenarios (four for precipitation and another four for air temperature scenarios) were employed for evaluation and corresponding comparisons by implementing statistical measures, behavior definition, and uncertainty analysis.

2. Materials and methods

2.1. Watershed simulation model

To evaluate the environmental impact caused by various sources of pollutants (e.g., point and nonpoint sources), the SWAT model was