Journal of Hydrology 527 (2015) 458–470

Contents lists available at ScienceDirect

Journal of Hydrology

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

Assessment of model predictions and parameter transferability by alternative land use data on watershed modeling

Haw Yen [a,b,*], Amirreza Sharifi [c], Latif Kalin [d,e], Golbahar Mirhosseini [f], Jeffrey G. Arnold [a]

[a] Grassland, Soil & Water Research Laboratory, USDA-ARS, 808 East Blackland Road, Temple, TX 76502, USA
[b] Blackland Research and Extension Center, Texas A&M University, 720 East Blackland Road, Temple, TX 76502, USA
[c] Department of Environmental Science and Technology, University of Maryland, College Park, MD 20742, USA
[d] School of Forestry and Wildlife Science, Auburn University, 602 Duncan Dr., Auburn, AL 36849, USA
[e] Civil Engineering Department, Canik Basari University, Gürgenyatak köyü 55080 Canik/SAMSUN, Turkey
[f] Biosystems Engineering Department, Auburn University, Auburn, AL 36849, USA

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form 27 April 2015
Accepted 29 April 2015
Available online 8 May 2015
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Alternative land use data
Input uncertainty
SWAT
IPEAT

SUMMARY

In recent years, complex large-scale watershed models have been developed to perform simulations of hydrologic and nutrient processes. The potential impact caused by human activities such as agricultural implementations against the environment can be evaluated under future scenarios. Meanwhile, large amount of input data are required to enhance the performance of simulated results. For some natural or urban regions, it is possible to have multiple sources of geophysical data available but the associated effects of using alternating data sources on modeling results is not yet evaluated. In this study, three sources of land use data (Mid-Atlantic Regional Earth Science Applications Center (RESAC 2000), National Land Use Cover Dataset (NLCD 2001), and State Land Use/Cover Maps (STATE) were implemented on the Greensboro watershed, Maryland, USA. The Alternative Dataset Scheme (ADS) and the Parameter Transferability Scheme (PTS) were applied to investigate model predictive uncertainty and the potential impact of cross transferring optimal calibration parameters between models. It was demonstrated that model predictions simulated by SWAT model had better performance when RESAC land use map was used, followed by STATE, and NLCD land use maps. In addition, calibrated best parameter set from RESAC has presented relatively more transferable compared to NLCD and STATE. The use of varying data source may not only alter model predictions and the associated predictive uncertainty but also have direct impact on the transferability of model parameters. The major findings in this study may help future modelers and decision makers to recognize the importance of alternative data source selection. Therefore, the quality of subsequent research work, engineering applications or policies can be further improved.

1. Introduction

In recent years, complex watershed simulation models have been developed and implemented to solve difficult environmental issues associated with wide variety of human activities (Kalin and Hantush, 2006). Along with great advances in computer technology, hydrological and nutrient processes can be simulated in fairly short time intervals (e.g. sub daily) for decades of long period based on historical data. Among all models, sophisticated, semi-distributed hydrological and water quality models such as Hydrological Simulation Program-Fortran (HSPF, Bicknell et al., 1997), Soil & Water Assessment Tool (SWAT, Arnold et al., 2012) and Agricultural Policy/Environmental Extender Tool (APEX, Williams et al., 2012) are specifically designed to address challenging subjects in the field of agricultural and water resources. By utilizing model predictions under future scenarios, the quality of the decision making processes can be enhanced substantially for decision makers (Winsemius et al., 2009).

One of the major challenges of using a complex watershed simulation model is the potential difficulties in model calibration. Calibration is typically required before implementing complex watershed simulation models; as there are large amount of parameters involved in simulating streamflow, sediment transport, and nutrient processes, many of which cannot be measured directly (Yen et al., 2014b). The calibration processes can be performed by manual or auto-calibration approaches. It has been demonstrated that manual calibration approach may be extremely

* Corresponding author at: Haw Yen, Assistant Research Scientist, Blackland Research and Extension Center, Texas A&M University, 720 East Blackland Road, Temple, TX 76502, USA. Tel.: +1 (254) 774 6004.
E-mail address: hyen@brc.tamus.edu (H. Yen).

http://dx.doi.org/10.1016/j.jhydrol.2015.04.076
0022-1694/© 2015 Elsevier B.V. All rights reserved.

challenging while conducting high-dimensional problems (Yen et al., 2014b), consequently, numerous parameter estimation techniques have been developed to solve this issue efficiently (Yen et al., 2014b). In general, parameter estimation techniques (e.g. IPEAT (Yen et al., 2014d), APEX-CUTE (Wang et al., 2014)) are not only beneficial to complex watershed models, but also provide additional functions such as sensitivity or uncertainty analysis, so users are able to obtain more information regarding watershed responses to future scenarios.

In addition to proper calibration, sophisticated watershed simulation models require detailed information from study area to properly reflect the actual behavior of the watershed. In the last decade, Geographic Information Systems (GIS) have successfully provided distributed watershed models with spatially heterogeneous watershed information (e.g. land use, soil type and elevation records) through user-friendly interfaces such as ArcSWAT and ArcAPEX (Chaubey et al., 2005). Model predictions and the associated predictive uncertainty are influenced by input data such as land use (different land use types such as pasture, forest) (Cotter et al., 2003; Sharifi and Kalin, 2010), digital elevation model resolution (coarser to higher resolution) (Lin et al., 2010) and soil map (e.g. SSURGO, STATSGO data) (Mukundan et al., 2010) in varying spatial resolution (Chaubey et al., 2005). Another source of uncertainty that requires further investigation is the potential impact caused by different sources of available data on model predictive uncertainty. There are several institutions in the US and around the globe (e.g. Federal and State agencies, private entities) that collect soil, land use and land elevation data for their own benefit, to the extent that it is possible to have alternative data sets for the same study area. The potential impact caused by different sources of available data toward model prediction (flow) is still rarely being investigated (Heathman et al., 2009) not to mention the corresponding influences against nutrient processes.

In this study, three independent land use maps were used to build autonomous SWAT models for the same study area. The major goal of this study is to investigate the performance of model predictions caused by altering sources of available land use data for the same watershed. Specifically, the following objectives are defined: (i) to examine statistical model performance (goodness of fit) using different land use data; (ii) to cross compare statistical performance by transferring best parameter sets of one model to others; (iii) to explore results of streamflow, sediment, and nutrient processes by temporal magnitude and percentiles in all scenarios. In this study, three sets of land use data (NLCD 2001, RESAC 2000 and STATE 2002) are employed to construct autonomous SWAT models for Greensboro watershed, Maryland, USA. The Integrated Parameter Estimation and Uncertainty Analysis Tool (IPEAT) is adopted as the optimization tool for model calibration. The purpose of parameter transferring is to handle situation that requires reliable and fast results within very limited time frame. Thus, the validation of parameter transferring may provide great help in the field of water resources engineering and the associate applications.

2. Materials and methods

2.1. SWAT model

SWAT (Arnold et al., 2012) is a large-scale, semi-distributed, continuous-time (sub daily to daily basis) model developed by the United States Department of Agriculture – Agricultural Research Services (USDA–ARS). The mechanism of SWAT operation is applying the Hydrologic Response Units (HRUs) to pre-defined (delineated through ArcSWAT interface) subbasins. HRUs are the elementary simulation units which integrate comprehensive information of land use, soil type, and slope (Gassman et al., 2007). It has been reported that SWAT is capable of providing reliable information to support the decision making processes (Arnold et al., 2012). Further applications of SWAT in the field of water science can be found in literature (Hoque et al., 2012; Yen et al., 2014c; Huang et al., 2015; Panagopoulos et al., 2015) and modifications have also been developed to solve challenges in varying topics (Nikolaidis et al., 2013; Foy et al., 2015; Yen et al., 2014a). Readers are referred to Arnold et al. (2012) for involved processes and theoretical details of SWAT model.

2.2. Study area and data collection

2.2.1. Basin information of study area

The methodology proposed in this study was applied to Greensboro Watershed, a mid-sized basin (294 km^2) with a long track of monitored discharge and water quality data, located on the Eastern shore of Chesapeake Bay. Greensboro watershed lies within the Mid-Atlantic coastal plain physiographic region and is predominantly covered by agricultural fields (54%) and forests (38%). Soils range from hydric and hypoxic poorly drained clays to well-drained, oxic sandy loams (Norton and Fisher, 2000). Greensboro watershed originates from Kent County, Delaware and flows southwest, down to the township of Greensboro, Maryland, where a USGS gauging station (#01491000) has long been monitoring hydrology (since 1948) and water chemistry (since 1964). Over 50% of streamflow is typically contributed by groundwater in the form of baseflow (Lee et al., 2001; Fisher et al., 2006). Greensboro watershed is part of the Choptank river basin, the largest tributary of Chesapeake Bay on Maryland's Eastern shore. Choptank River's water quality has declined over the past decades due to excessive total nitrogen (TN) and total phosphorus (TP) export from intensive agriculture and human waste disposal (Fisher et al., 2010; McCarty et al., 2008), resulting in portions of the Choptank River to be listed as "impaired waters" under Section 303(d) of the Federal Clean Water Act. Subsequently, several best management practices have been implemented in the basin to reduce nitrogen and phosphorus exports from crop fields (Fisher et al., 2010).

2.2.2. Mid-Atlantic Regional Earth Science Applications Center (RESAC 2000)

RESAC 2000 land use map was developed by the mid-Atlantic Regional Earth Science Applications Center (RESAC) to provide improved land cover mapping and ecological modeling capabilities within the 178,000 km^2 Chesapeake Bay watershed (Goetz et al., 2000). The RESAC land use database is based on Multi-temporal Landsat-7 ETM+ imagery with a 30 m resolution. For image classification, a decision tree classifier algorithm (DeFries and Chan, 2000) was implemented along with extensive field measurements and historical air photos to obtain more accurate LULC discrimination. The land cover classification is modified from Anderson Level II (Anderson et al., 1976) with more number of classes (21 classes) compared to NLCD 2001 (16 classes). In addition, the land use categories defined in data layers are manually linked to ArcSWAT classes.

2.2.3. National Land Use Cover Dataset (NLCD 2001)

The NLCD land use map is fairly consistent with the data format in ArcSWAT interface. The NLCD 2001 land use map was developed by MRLC (Multi-Resolution Land Characteristics Consortium), with intentions to deliver a full scale land cover database across all 50 states and Puerto Rico. The land use map was derived from 30 m resolution multi-temporal Landsat 5 and 7 imagery (Homer et al., 2007), where three image dates for each Landsat path-row footprint was collected to capture critical stages of vegetation growth