Environmental Modelling & Software 74 (2015) 48–65

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Setting up a hydrological model of Alberta: Data discrimination analyses prior to calibration

Monireh Faramarzi [a,*], Raghavan Srinivasan [b], Majid Iravani [a,c], Kevin D. Bladon [a,d], Karim C. Abbaspour [e], Alexander J.B. Zehnder [f,g], Greg G. Goss [a,h]

[a] Water Initiative, University of Alberta, Edmonton, Alberta, Canada
[b] Spatial Science Laboratory, Department of Ecosystem Science and Management, Texas A&M University, College Station, TX, USA
[c] Alberta Biodiversity Monitoring Institute, University of Alberta, Edmonton, Alberta, Canada
[d] Department of Forest Engineering, Resources, and Management, Oregon State University, Corvallis, OR, 97331, USA
[e] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland
[f] Alberta Innovates-Energy and Environment Solutions, Edmonton, Alberta, Canada
[g] Nanyang Technological University (NTU), Sustainable Earth Office, Singapore, 637459, Singapore
[h] Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

ARTICLE INFO

Article history:
Received 10 April 2015
Received in revised form
1 September 2015
Accepted 3 September 2015
Available online xxx

Keywords:
SWAT
Watershed modeling
Climate data
Hydrological processes
Geo-spatial data

ABSTRACT

Failure to setup a large-scale hydrological model correctly may not allow proper calibration and uncertainty analyses, leading to inaccurate model prediction. To build a model with accurate accounting of hydrological processes, a data discrimination procedure was applied in this study. The framework uses a hydrological model of Alberta built with the Soil and Water Assessment Tool (SWAT) program. The model was used to quantify the causes and extents of biases in predictions due to different types of input data. Data types represented different sources of errors, including input data (e.g., climate), conceptual model (e.g., potholes, glaciers), and control structure (e.g., reservoirs, dams). The results showed that accounting for these measures leads to a better physical accounting of hydrological processes, significantly improving the overall model performance. The procedure used in this study helps to avoid unnecessary and arbitrary adjustment of parameters to compensate for the errors in the model structure.

Software availability

SWAT program is available for use at the following URL http://swat.tamu.edu/

1. Introduction

Physically-based, distributed hydrological models have been widely used for water resources management and planning. They have been extensively applied to study the impact of climate change and landuse change on water quality and quantity, water related activities, and adaptation measures among others (Li et al., 2009; Faramarzi et al., 2010a, 2010b, Van Griensven et al., 2012; Faramarzi et al., 2013; Eum et al., 2014; Xue et al., 2014). The reliability of such applications depends on the accuracy of hydrological models in representing the physical processes (Beven, 2000; Muleta and Nicklow, 2005), correct input data, and proper model calibration. As such, a key challenge is initially to set up an accurate hydrological model, which correctly represents the site's actual physical processes (Gupta and Sorooshian, 1998; Perrin et al., 2001; Blasone et al., 2008; Moradkhani et al., 2012; Houska et al., 2014; Guse et al., 2014; Gabriel et al., 2014).

Calibration of distributed models is often difficult and subjective when there is a considerable simplification in model setup. It is standard practice in watershed modeling studies that the physical parameters are adjusted to achieve the optimal fit to the measured data. However, simplification of the models, especially in large scale watersheds (where a considerable heterogeneity exist in climate, vegetation, soil, physiography, and management activities), might result in a wrong parameter estimation (Schuol et al., 2008b; Faramarzi et al., 2009). In large scale models where a vast number of adjustable physical-parameters are allowed to vary within a broad range of values, a seemingly good simulation can be obtained with erroneous parameter values (Abbaspour et al., 2007). In other words, wrong model structure and inappropriate input data can be compensated by unrealistic model parameters. Such models could

* Corresponding author. University of Alberta Water Initiative, G-222 Biological Sciences Building, 11455 Saskatchewan Dr., Edmonton, Alberta, T6G 2E9, Canada.
E-mail address: faramarz@ualberta.ca (M. Faramarzi).

http://dx.doi.org/10.1016/j.envsoft.2015.09.006
1364-8152/

produce misleading results in scenario analyses, even though typical performance criteria are satisfied during calibration. One way to detect these over calibration problems is by validation of the model for a reasonable time period where major hydrological events (e.g., wet years, dry years) are presented.

A correct model setup, accurately representing the actual hydrological processes, can limit uncertainty in parameter estimation. In literature, to limit uncertainties in parameter estimation, various measures through automated calibration techniques have been examined. These include multi-variable calibration procedure (Gupta and Sorooshian, 1998; Xie et al., 2012; Qiao et al., 2013; Samuel et al., 2014), use of multiple calibration sites rather than only catchment integrated behavior (Abbaspour et al., 1999, 2007; Cao et al., 2006; Schuol et al., 2008a, 2008b), a multi-objective formulation by including different variables in the objective function (Gupta and Sorooshian, 1998; Madsen, 2003; White and Chaubey, 2005), and use of various techniques to increase the computational efficiency of the large scale hydrological models (Wu et al., 2013; Ercan et al., 2014). Although the schemes are beneficial in limiting uncertainties in the predictions, a more reliable result can be achieved through building an accurate model. Building a correct model, especially in large scale and complex watersheds, is an important practice to represent correct processes inside a watershed. A correct model is one that adheres to the principle of "correct neglect", where only unimportant processes are neglected in the model and all important processes should be included. Therefore, it is inevitable that large scale models should go through careful data discrimination scheme to ensure most of the important processes are represented prior to calibration. These include: (i) gathering and compiling appropriate input data (e.g., climate data in mountainous regions); (ii) including management control structures that can disrupt natural processes (e.g., dams that regulate downstream water flow); and (iii) incorporating local knowledge about the natural complexity and anthropogenic changes into watershed models. These are all key factors that can reduce the uncertainty in model predictions and avoid unnecessary and arbitrary adjustment of the parameters.

Overall, the majority of researchers have focused on elaboration of the importance of robust calibration schemes in parameter estimation (e.g., Joseph and Guillaume, 2013) and prediction uncertainty, while much fewer studies have addressed proper model setup and choice of appropriate input datasets. Later group are those that focused on modifying the existing climate datasets to better represent the effect of altitude on precipitation (Masih et al., 2011; Galvan et al., 2014) and those that examined the effect of input data quality and quantity on parameter estimation and model calibration (Getirana et al., 2011; Strauch et al., 2012; Yalew et al., 2013; Gabriel et al., 2014; Rouholahnejad et al., 2014; Yen et al., 2014; Abbaspour et al., 2015; Leta et al., 2015).

With an area of about 660,000 km^2, Alberta encompasses 17 river basins that principally originate from the east slopes of the Canadian Rocky Mountains and the majority drain east to Hudson Bay through the provinces of Saskatchewan and Manitoba and north to the Arctic Ocean. The heterogeneous hydro-climatic conditions and the diverse land management practices in combination with the scarcity of data, especially in the northern remote areas and western mountainous region, make hydrological modeling challenging in this region. To the best of our knowledge a high resolution and province-wide hydrological model has not been developed for Alberta. Most of the previous studies in Alberta have been conducted at a catchment (e.g., Kienzle et al., 2012; Marshall, 2014) or river basin (e.g., Islam and Gan, 2014; Eum et al., 2014) scale.

The model of choice for this project was "Soil and Water Assessment Tool" (SWAT) (Arnold et al., 1998). SWAT has been developed to quantify the impact of land management practices and climate on water, sediment, and agricultural chemical yields in large complex watersheds with varying soils, landuses, and management conditions over long periods of time. The program, therefore, lends itself easily to climate and landuse change analyses. SWAT is a valuable watershed-scale management tool and we chose this program for our purposes because: i) it integrates many components such as hydrology, climate, nutrient, soil, sediment, crop, pesticide, and agricultural management, ii) it has been successfully applied worldwide in many different climate and landuse situations (Arnold et al., 1999; Gosain et al., 2006; Schuol et al., 2008a,b; Rouholahnejad et al., 2014; Abbaspour et al., 2009, 2015), iii) the program is actively maintained and continuously updated with new and up-to-date knowledge of watershed processes, and iv) many side programs are written for SWAT from calibration and uncertainty analysis to graphic packages for visualization and animation of the results. Hence, over a 50-year period, a global consensus is built around the accuracy and usefulness of the program as there exist over 3000 scientific publications where SWAT has been used to address numerous watershed issues (Gassman et al., 2007, 2010).

We used the SWAT hydrological model of Alberta as an example to demonstrate that proper model setup could produce more accurate model outputs and represent most of the natural and anthropogenic processes. However, one hypothesis would be how a model with a better performance would guarantee that it will be actually the best option after calibration. We address in this paper the fact that building a correct model is a key step prior to calibration to avoid compensation through subjective and challenging parameter estimation and this will provide the best performance model.

Objectives of this paper are: (i) to build various SWAT projects to test the effects of including alternative climate and geo-spatial datasets available from global and regional sources; (ii) to evaluate the performance of the model predictions using combination of multiple datasets from different sources, (iii) to define the procedures by which raw datasets are evaluated for inclusion or exclusion in the model; and (iv) to calibrate and validate all of the model scenarios for the Athabasca River basin as an example hydrological region, thereby allow us to test how an accurate model will perform best after calibration. It is important to point out that the above SWAT models are tested against each other prior to calibration, as over calibration and over fitting of model parameters would mask the input data and model structure effects and will not allow a proper discrimination of initial model setups (Dile and Srinivasan, 2014; Abbaspour et al., 2015).

2. Materials and methods

2.1. Study area

Alberta, with an area of about 660,000 km^2, is located between 49–60 °N and 110–120 °W where altitude varies from 3747 m (Mount Columbia) to 152 m (Slave River-Wood Buffalo National Park) (Fig. 1a). Geographically, the province spans >1200 km from north to south and large-scale climate anomalies, originating from Pacific Ocean, have a considerable influence on climate diversity (Lapp et al., 2013). Air temperatures can drop to as low as −54 °C during the winter (northern Alberta), and rise to as high as 40 °C during the summer (southern Alberta). Average annual precipitation ranges from 300 mm in the southeast to 600 mm in the foothills of the Rocky Mountains (AENV-GA, 2008; Mwale et al., 2009).

The province has 17 river basins (Fig. 1a; AENV-GA, 2008) with the northern rivers of the province having comparatively larger areas and therefore higher discharge rates than the southern rivers that flow through regions that receive much lower annual precipitation. For instance, the average flow of Peace River in the north is