Computers & Geosciences 58 (2013) 80–90

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

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

Parameterization for distributed watershed modeling using national data and evolutionary algorithm

Xuan Yu [a], Gopal Bhatt [a], Christopher Duffy [a,*], Yuning Shi [b]

[a] Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA
[b] Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA

ARTICLE INFO

Article history:
Received 27 November 2012
Received in revised form
23 April 2013
Accepted 30 April 2013
Available online 30 May 2013

Keywords:
Calibration
Physics-based hydrological modeling
Little Juniata river
PIHM
SSHCZO
Young Womans creek

ABSTRACT

Distributed hydrologic models supported by national soil survey, geology, topography and vegetation data products can provide valuable information about the watershed hydrologic cycle. However numerical simulation of the multi-state, multi-process system is structurally complex and computationally intensive. This presents a major difficulty in model calibration using traditional techniques. This paper presents an efficient calibration strategy for the physics-based, fully coupled, distributed hydrologic model Penn State Integrated Hydrologic Model (PIHM) with the support of national data products. PIHM uses a semi-discrete Finite Volume Method (FVM) formulation of the system of coupled ordinary differential equations (e.g. canopy interception, transpiration, soil evaporation) and partial differential equations (e.g. groundwater-surface water, overland flow, infiltration, channel flow, etc.). The matrix of key parameters to be estimated in the optimization process was partitioned into two groups according to the sensitivity to difference in time scales. The first group of parameters generally describes hydrologic processes influenced by hydrologic events (event-scale group: EG), which are sensitive to short time runoff generation, while the second group of parameters is largely influenced by seasonal changes in energy (seasonal time scale group: SG). The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is used to optimize the EG parameters in Message Passing Interface (MPI) environment, followed by the estimation of parameters in the SG. The calibration strategy was applied at three watersheds in central PA: a small upland catchment (8.4 ha), a watershed in the Appalachian Plateau (231 km^2) and the Valley and Ridge of central Pennsylvania (843 km^2). A partition calibration enabled a fast and efficient estimation of parameters.

1. Introduction

Physics-based, fully coupled, distributed hydrologic models seek to simulate hydrologic states in space and time. Since the representations of hydrologic processes and parameters involved have physical meanings, ideally the model should not require calibration. This will be possible if all the parameters were available through experimentation, field measurements and national data coverage sets. One example, Soil Survey Geographic Database (SSURGO) with soil textural information was tested as a useful soil physical property for a-priori parameter estimation of distributed hydrological modeling (Anderson et al., 2006). Such national datasets provide a measure of spatial variations and can potentially meet the data requirements of distributed hydrological models. However, due to the high uncertainty of spatially distributed soils and geologic properties estimation (Vereecken et al., 2010), calibration is still an indispensable part of the physics-based hydrologic modeling. Similarly National Land Cover Database (NLCD) and Global Land Cover (UMDGLC) provide distributed vegetation information for the models. Clearly, calibration is of huge potential for improving representations of the soils and vegetation information (Duan et al., 2006). Most physics-based, distributed hydrologic models are computationally intensive, as they seek to simulate complex non-linear interaction between multi-state, multi-process, hydrological system, while incorporating spatially explicit data such as topography, soil, geology, land-use, climate, etc. Traditionally, computational requirements for the numerical solution often resulted in trial and error techniques for parameter estimation with limited success and poor understanding of parameter sensitivity (Ivanov et al., 2004; Qu and Duffy, 2007; Du et al., 2007; Li et al., 2008; Takeuchi et al., 2010; Shih and Yeh, 2011).

In the past decades, computational methods were widely imported in model calibration frameworks, including Monte Carlo Analysis, Genetic Algorithm (GA) and Evolutionary Strategy (ES) (Tolson and Shoemaker, 2007; Nicklow et al., 2010; Reed et al., 2012). Comparative

* Corresponding author. Tel.: +1 814 863 4384; fax: +1 814 863 7304.
E-mail address: cxd11@psu.edu (C. Duffy).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.04.025

studies between different optimization algorithms have improved the understanding of models and parameter properties and benefited watershed modeling applications (Tolson and Shoemaker, 2007). Physics-based distributed model usually requires huge computation resources. Therefore parameter partition is necessary for automated calibration (Lei et al., 2012). However, relatively little of this work has been applied to parameter estimation of physics-based, fully coupled, distributed hydrologic models.

In this study the physics-based, fully coupled, distributed hydrologic model PIHM is analyzed by a sensitivity-based Partition Calibration Strategy (PCS) for efficient model parameter optimization. The model uses a semi-discrete FVM to form the coupled equations for Noah_LSM (Chen and Dudhia, 2001), 2-D overland flow, 1-D unsaturated flow and 2-D subsurface flow to streams. PIHM is an open source distributed hydrologic model (http://www.pihm.psu.edu) and has been applied at multi-scale hydrologic, hydrodynamics modeling (Qu and Duffy, 2007; Li and Duffy, 2011; Bhatt et al., 2010; Bhatt, 2012; Duffy and Bhatt, 2007; Duffy et al., 2010; Duffy et al., 2011; Duffy, 2004). The objective of this paper is to design a calibration framework for a computationally intensive hydrologic model and test the robustness of the method. Section 2 gives a short introduction of PIHM and its national data support. Section 3 describes sensitivity analysis of parameters and two-step calibration. Section 4 shows the application of the method at watersheds ranging from a small-scale catchment to an upland watershed and to a mesoscale watershed.

2. Model formulation and national data set support

2.1. PIHM

PIHM is a physics-based, fully coupled, distributed hydrologic model. It simulates interception, throughfall, infiltration, recharge, evapotranspiration, surface runoff, groundwater flow, and channel routing in a fully coupled scheme. The spatial domain decomposition as quality triangular mesh uses the triangle (Shewchuk, 1997) that is an implementation of Delaunay triangulation algorithm. The resolution of spatial domain decomposition can be varied according to the geomorphological or hydrological characteristics of the watershed. The spatial domain decomposition can be constrained by hydrologic features such as observation point, boundary conditions, etc. (Kumar et al., 2009). Hydrologic equations that include partial differential equations (PDEs) for overland flow, subsurface flow, and channel routing, and ordinary differential equations (ODEs) for interception, infiltration, recharge, and evapotranspiration (ET) are assembled over each control volume. PDEs are discretized to ODEs using finite volume method. This results in a local system of ordinary differential equations corresponding to each model grids. The local systems of ODEs are assembled over the entire model domain to form a global system of ODEs and solved using SUNDIALS (Cohen and Hindmarsh, 1996). For a detailed description of the modeling approach and formulation, the reader should consult Qu and Duffy (2007) and Kumar (2009).

2.2. Hydrological processes

A short introduction about hydrological processes and corresponding model parameters are provided in this section. Fig. 1 illustrates hydrological processes of PIHM and the decomposition and kernel of a real watershed, Shale Hills watershed.

Overland flow: The governing equations for surface flow are the 2-D estimation of St. Venant equations (Qu and Duffy, 2007). The equations are approximated in semidiscrete form:

$$\frac{dh_0}{dt}\Big|_m = p_t - q^+ - e + \sum_{j=1}^{3} q_j^s\Big|_m \tag{1}$$

where h_0 is the shallow water depth above the ground surface, q_j^s is the normalized lateral flow rate from element to its neighbor j. The terms p_t, q^+, and e are throughfall, infiltration, and evaporation, respectively. Subscript m represent the spatial grid, ranging from 1 to the total number of triangles.

Subsurface flow: The model assumes that each subsurface layer can have both unsaturated and saturated storage components (Qu and Duffy, 2007). With the estimation (only vertical flow in unsaturated zone) and integration, the balance equations are formed:

$$\theta_s \frac{dh_u}{dt}\Big|_m = q^+ - q^0\Big|_m \tag{2}$$

$$\theta_s \frac{dh_g}{dt}\Big|_m = q^0 + \sum_{j=1}^{3} q_j^g\Big|_m \tag{3}$$

where θ_s is the moisture content, h_u is the unsaturated storage depth, h_g is the groundwater depth, q^0 is flux between unsaturated–saturated zone (Kumar, 2009), q_j^g is the normalized lateral groundwater flow rate from element i to its neighbor j. Here Van Genuchten (1980) formulation was used in discretized form to improve the computation performance (Qu and Duffy, 2007).

Channel routing: The same semi-discrete finite volume approach is applied to the 1-D estimation of St. Venant equations (Qu and Duffy, 2007):

$$\frac{dh_c}{dt}\Big|_k = p - e + \sum_{j=1}^{2} (q_l^s + q_l^q) + q_{in}^c - q_{out}^c\Big|_k \tag{4}$$

where h_c is depth of water in the channel, p and e are precipitation and evaporation from the channel segment, respectively, and q_l^s and q_l^g are the lateral surface flow and groundwater interaction with the channel, respectively, from each side of the channel. The upstream and downstream flow for each channel segments are $q_{in}{}^c$ and $q_{out}{}^c$, respectively. Subscript k represent the channel segment, ranging from 1 to the total number of channel segments.

Evapotranspiration (*ET*): The total evaporation is the sum of evaporation from canopy interception (e_c), transpiration from vegetation (e_t), and evaporation from soil (e_s). The Penman–Monteith approach is used for the calculation of the potential evaporation

$$e_p = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(\varepsilon_s - \varepsilon_a)}{r_a}}{\Delta + \gamma\left(1 + \frac{r_s}{r_a}\right)} \tag{5}$$

Here e_p refers to potential evapotranspiration, R_n is net radiation at the vegetation surface, G is soil heat flux density, $\varepsilon_s - \varepsilon_a$ represents the air vapor pressure deficit, and ρ_a is the air density, C_p is specific heat of the air. Δ is the slope of the saturation vapor pressure–temperature relationship, γ is the psychometric constant, and r_s, r_a are the surface and aerodynamic resistances. The ET calculation equations are adapted from Noah_LSM (Chen and Dudhia, 2001) for computing the actual evapotranspiration

$$e_c = \sigma_f e_p \left(\frac{W_c}{S}\right)^{0.5} \tag{6}$$

$$e_t = \sigma_f e_p B_c \left[1 - \left(\frac{W_c}{S}\right)^{0.5}\right] \tag{7}$$

$$e_s = (1 - \sigma_f)\beta e_p \tag{8}$$

where σ_f refers to vegetation fraction, W_c is the intercepted