Advances in Water Resources 101 (2017) 23–36

Contents lists available at ScienceDirect

Advances in Water Resources

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

Assessment of model behavior and acceptable forcing data uncertainty in the context of land surface soil moisture estimation

Gift Dumedah[a,*], Jeffrey P. Walker[b]

[a] Department of Geography and Rural Development, Kwame Nkrumah University of Science & Technology, Kumasi, Ghana
[b] Department of Civil Engineering, Monash University, Building 60, Melbourne, 3800, Victoria, Australia

ARTICLE INFO

Article history:
Received 27 May 2016
Revised 29 December 2016
Accepted 3 January 2017
Available online 4 January 2017

Keywords:
Soil moisture
Land surface modeling
Uncertainty estimation
Data assimilation
Evolutionary strategy
Hydrologic genomic mapping

ABSTRACT

The sources of uncertainty in land surface models are numerous and varied, from inaccuracies in forcing data to uncertainties in model structure and parameterizations. Majority of these uncertainties are strongly tied to the overall makeup of the model, but the input forcing data set is independent with its accuracy usually defined by the monitoring or the observation system. The impact of input forcing data on model estimation accuracy has been collectively acknowledged to be significant, yet its quantification and the level of uncertainty that is acceptable in the context of the land surface model to obtain a competitive estimation remain mostly unknown. A better understanding is needed about how models respond to input forcing data and what changes in these forcing variables can be accommodated without deteriorating optimal estimation of the model. As a result, this study determines the level of forcing data uncertainty that is acceptable in the Joint UK Land Environment Simulator (JULES) to competitively estimate soil moisture in the Yanco area in south eastern Australia. The study employs hydro genomic mapping to examine the temporal evolution of model decision variables from an archive of values obtained from soil moisture data assimilation. The data assimilation (DA) was undertaken using the advanced Evolutionary Data Assimilation. Our findings show that the input forcing data have significant impact on model output, 35% in root mean square error (RMSE) for 5cm depth of soil moisture and 15% in RMSE for 15cm depth of soil moisture. This specific quantification is crucial to illustrate the significance of input forcing data spread. The acceptable uncertainty determined based on dominant pathway has been validated and shown to be reliable for all forcing variables, so as to provide optimal soil moisture. These findings are crucial for DA in order to account for uncertainties that are meaningful from the model standpoint. Moreover, our results point to a proper treatment of input forcing data in general land surface and hydrological model estimation.

1. Introduction

The estimation of soil moisture through land surface models mainly incorporates a combination of model physics, initial states, parameters, and forcing data. The majority of these components are inherent to the overall makeup of the model, defining how the model handles input forcing data. The model components are fundamental to the evaluation of land surface states in response to meteorological forcing. However, the overall uncertainty in model output is associated with uncertainties of the various model inputs and components, which interact and are strongly linked such that their respective uncertainties are difficult to separate. The input forcing data set is independent from the overall model makeup and has separate uncertainty levels for each variable as defined by their respective observation systems. While the impact of the forcing data on the accuracy of model estimation is universally recognized to be significant (Beven and Binley, 1992; De Lannoy et al., 2006; Durand and Margulis, 2008; He et al., 2011; Mantovan and Todini, 2006; Moradkhani and Hsu, 2005; Raleigh, 2013; Salamon and Feyen, 2009; Seibert, 1997; Steinschneider et al., 2012; Uhlenbrook et al., 1999; Vrugt et al., 2002; Zehe et al., 2005), its quantification remains largely unknown in most modeling procedures. That is, the majority of modeling procedures have limited knowledge about how much of a model's estimation accuracy is attributable to its forcing data uncertainty. Consequently, it is difficult to determine the level of uncertainty in forcing data that is acceptable, in the context of the model, to provide an optimal estimate of soil moisture.

The acceptable level of forcing data uncertainty, while specific to a particular model, will provide for a given model a threshold

* Corresponding author.
E-mail address: dgiftman@hotmail.com (G. Dumedah).

http://dx.doi.org/10.1016/j.advwatres.2017.01.001
0309-1708/

of uncertainty bound beyond which deterioration of the model estimation accuracy will occur. An estimate of the acceptable level of uncertainty in forcing data will separate the inaccuracies in model prediction into two categories: those that are inherent in the model, and those from the forcing data. Note that by acceptable level of uncertainty in forcing data, we are referring to the amount of uncertainty from the forcing data that is admissible/appropriate in the context of the model without deteriorating the model estimation accuracy.

An increased knowledge of forcing data uncertainty will have a crucial impact on water and climate predictions. In particular, an estimate of forcing data uncertainty will provide an important contribution, which is usually missing, to error specification in data assimilation (DA) procedures. Ensemble distributions (i.e., spread) in DA are usually generated through perturbation of input forcing data (Alemohammad et al., 2015; Clark et al., 2006; Wojcik et al., 2014), mostly with limited knowledge of the impact of forcing error on model ensemble spread. Forcing data uncertainty is also critical in climate change studies and forecasting systems (Nagler et al., 2008; Raleigh, 2013; Schär et al., 2004; Steinschneider et al., 2012; Troch et al., 2009), to provide an estimate of the changes (and uncertainties) in model output in response to variability in forcing data. An understanding of the impact of forcing data uncertainty on model prediction will therefore provide the capability to estimate the level of variability that is required in weather systems and forecasts to initiate triggers in water resource systems.

Few studies including Liu and Gupta (2007); Vrugt and Robinson (2007); Wagener et al. (2003); and He et al. (2012) have examined uncertainty in model components, and even fewer (Alemohammad et al., 2015; Maggioni et al., 2011) have actually examined uncertainty in forcing data in relation to the model output. Consequently, this study quantifies the uncertainty threshold in forcing data that can be incorporated into the Joint UK Environment Simulator (JULES) model in the context of soil moisture estimation without a significant deterioration in model estimation accuracy, for the Yanco area in southeast Australia. It also provides a methodology to estimate an acceptable threshold of forcing data uncertainty in the JULES model through three modeling approaches. These modeling approaches are model calibration, data assimilation, and multi-dimensional clustering which is used to assess values in model decision space (i.e., the interval defined by both model parameters and input forcing variables). The calibration and data assimilation procedures employ computational techniques from the state-of-the-art multi-objective evolutionary strategy. Specifically, the calibration is based on the Non-dominated Sorting Genetic Algorithm - II (NSGA-II) developed by Deb et al. (2002), whereas the DA method uses the evolutionary data assimilation (EDA) scheme demonstrated in Dumedah and Walker (2014b); Dumedah et al. (2015); and Dumedah (2015).

The three modeling approaches: calibration, data assimilation, and multi-dimensional clustering used in this study provide unique roles toward the overall goal of quantifying forcing data uncertainty in land surface modeling. Calibration, though subject to a specific time period of observation data, has an important role in determining optimized values in model decision space to generate model outputs which best match observed data. It is noted that the calibration procedure is supplementary and represents an intermediate step to the data assimilation procedure. Data assimilation has been widely credited for its ability to update model predictions through time, and to account for uncertainties in model and observation data. However, the temporal changes in model decision space resulting from data assimilation holds the potential for assessing model behavior under changing hydro-meteorological conditions. The temporal characteristic of DA is crucial in this study in order to assess the temporal evolution of the impact of forcing data uncertainty on the JULES model at different uncertainty levels across time. Consequently, this study uses data assimilation to provide an archive of updated ensemble members in model decision space through several assimilation time periods. The role of the multi-dimensional clustering is to determine commonalities in model decision space for the calibration output and the updated ensemble members.

1.1. Study area, data sets, and the land surface model

The case study demonstration is for soil moisture estimation in the Yanco area in southeast Australia. The study location is at one station (i.e., Y10) out of thirteen OzNet soil moisture monitoring stations in the Yanco area (Smith et al., 2012). The Y10 location has flat topography, along with grassland, scattered trees and loamy textured soil. The study location has extensive soil moisture and meteorological instrumentation, and has provided almost continuous time series of data for validation.

The land cover data set was obtained through the Australian National Dynamic Land Cover Dataset (DLCD) (Lymburner et al., 2011), which was generated from the 16-day Enhanced Vegetation Index composite collected at 250 m spatial resolution from the Moderate Resolution Imaging Spectroradiometer. The soil properties information including texture, bulk density, saturated hydraulic conductivity, and soil layer thicknesses for horizons A and B were obtained from the Digital Atlas of Australian Soils, through the Australian Soil Resource Information System (McKenzie et al., 2000). The meteorological forcing data including incoming short and long wave radiations, air temperature, precipitation, wind speed, pressure, and specific humidity were obtained from the meteorological record at the study location.

The soil moisture estimation model used is JULES, a tiled model of sub-grid heterogeneity for simulating water and energy fluxes between a vertical profile of variable soil layers, land surface, vegetation, and the atmosphere (Best et al., 2011). JULES allows specification of numerous soil layers and variable thickness of soil layers, together with nine land surface types including broadleaf, needleleaf, grass (temperate and tropical), shrub, urban, inland water, bare soil, and ice-covered surfaces. The JULES model requires initialization for variables including the temperatures and the moisture contents of the soil layers; temperature, density, and albedo of the snowpack if present; the temperature and intercepted rain and snow on the vegetation canopy; the temperature and depth of ponded water on the soil surface, and an empirical vegetation growth index.

The JULES model parameters and forcing variables together with their descriptions and intervals are presented in Table 1. The model parameters and forcing variables were allowed to be varied within $\pm$ 10% of their original values through a relative measure. It is noteworthy that the $\pm$ 10% interval is based on the soil texture variability as obtained from McKenzie et al. (2000), and does not represent the actual variability for model parameters and forcing variables. The original values of model parameters and forcing variables were based on the soil, land cover, and meteorological forcing data such that they are physically meaningful for the study location in the context of the JULES model. The soil moisture data set used to drive the calibration and assimilation was the surface 5cm depth of in-situ soil moisture at the Y10 location.

2. Methods

The framework used to assess the acceptable forcing data uncertainty in this study comprises of a number of modeling procedures. Specifically, four procedures were used including: (i) a calibration procedure to estimate model parameters, (ii) a data assimilation procedure using both model parameters and forcing variables, (iii) a data assimilation procedure using perturbed