



Development of a generic auto-calibration package for regional ecological modeling and application in the Central Plains of the United States



Yiping Wu ^{a,*}, Shuguang Liu ^{b,c,*}, Zhengpeng Li ^d, Devendra Dahal ^e, Claudia J. Young ^f, Gail L. Schmidt ^e, Jinxun Liu ^e, Brian Davis ^e, Terry L. Sohl ^b, Jeremy M. Werner ^b, Jennifer Oeding ^e

^a ASRC Research and Technology Solutions, U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, USA

^b U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, USA

^c Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

^d Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD 20740, USA

^e Stinger Ghaffarian Technologies, USGS EROS Center, Sioux Falls, SD 57198, USA

^f ERT Inc., USGS EROS Center, Sioux Falls, SD 57198, USA

ARTICLE INFO

Article history:

Received 19 September 2013

Received in revised form 15 November 2013

Accepted 20 November 2013

Available online 28 November 2013

Keywords:

Biogeochemical modeling

FME

Parameter optimization

R

Remote sensing

SCE

ABSTRACT

Process-oriented ecological models are frequently used for predicting potential impacts of global changes such as climate and land-cover changes, which can be useful for policy making. It is critical but challenging to automatically derive optimal parameter values at different scales, especially at regional scale, and validate the model performance. In this study, we developed an automatic calibration (auto-calibration) function for a well-established biogeochemical model—the General Ensemble Biogeochemical Modeling System (GEMS)—Erosion Deposition Carbon Model (EDCM)—using data assimilation technique: the Shuffled Complex Evolution algorithm and a model-inversion R package—Flexible Modeling Environment (FME). The new functionality can support multi-parameter and multi-objective auto-calibration of EDCM at the both pixel and regional levels. We also developed a post-processing procedure for GEMS to provide options to save the pixel-based or aggregated county-land cover specific parameter values for subsequent simulations. In our case study, we successfully applied the updated model (EDCM-Auto) for a single crop pixel with a corn–wheat rotation and a large ecological region (Level II)—Central USA Plains. The evaluation results indicate that EDCM-Auto is applicable at multiple scales and is capable to handle land cover changes (e.g., crop rotations). The model also performs well in capturing the spatial pattern of grain yield production for crops and net primary production (NPP) for other ecosystems across the region, which is a good example for implementing calibration and validation of ecological models with readily available survey data (grain yield) and remote sensing data (NPP) at regional and national levels. The developed platform for auto-calibration can be readily expanded to incorporate other model inversion algorithms and potential R packages, and also be applied to other ecological models.

Published by Elsevier B.V.

1. Introduction

Mathematical models, especially process-oriented models, are widely used to support decision making in environmental disciplines because they can characterize and predict landscape processes and consequences (Liu et al., 2008b). The majority of ecological models are built for research purposes, but increasingly also for forecasting and management purposes (Marta-Almeida et al., 2012; Rykiel, 1996). These numerical models, however, usually contain parameters which may be hard to measure (Janssen and Heuberger, 1995; Luo et al., 2001;

Van Oijen et al., 2005; Wu and Liu, 2012a; Yuan et al., 2012) or cannot be determined using field measurement because of scaling effects (discrepancies between measurement and modeling scales) (Beven, 2001; Juston et al., 2010). Hence, model calibration is required to estimate and adjust model parameters and constants to improve the agreement between simulations and the corresponding observations (Janssen and Heuberger, 1995; Larssen et al., 2007; Pereira et al., 2008; Rykiel, 1996; Zhang et al., 2009a). This procedure is fundamental and critical for ecological model applications (Janssen and Heuberger, 1995; Larssen et al., 2007; Mazzotti and Vinci, 2010; Rykiel, 1996).

Conventional manual calibration can incorporate the knowledge and experience of modelers through analysis of results and common sense reasoning when searching new values for each parameter (Pereira et al., 2008). However, it is subjective, time-consuming, and

* Corresponding authors.

E-mail addresses: ywu@usgs.gov, rocky.ypwu@gmail.com (Y. Wu), sliu@usgs.gov (S. Liu).

challenging when calibrating more than one parameter and satisfying multiple constraints (i.e., multi-objective calibration) for biogeochemical models (Liu et al., 2008b). Further, it could be impractical to obtain the globally optimal set of parameter values for non-linear models using a manual approach which is labor intensive and associated with subjective experience. Additionally, the size of a study area can dramatically increase the difficulty to derive spatially-explicit parameter values (i.e., the spatial variability of a parameter) (Liu et al., 2008b).

With the development of computer technology, auto-calibration was proposed and attracted increasing attention over the past decades (Gupta et al., 1999; Vrugt et al., 2003; Zhang et al., 2009b). Simple techniques such as controlled random search or linear regression methods can be useful for some simple data-oriented models. However, because these techniques are based on systematic and exhaustive generation of parameter arrays and thus require a huge number of model iterations, they are inapplicable of deriving the globally optimal values for non-linear, process-oriented models which are usually computationally-intensive. From literature, there are a number of global optimization techniques available, such as Shuffled Complex Evolution (SCE) (Duan et al., 1992), Artificial Neural Network (ANN), Genetic Algorithm (GA) (Goldberg, 1989; Holland, 1975), Multi-objective Evolutionary Algorithm (MOEA), Kalman Filter, and the Flexible Modeling Environment (FME) (Soetaert and Petzoldt, 2010), which was developed and implemented using the General Public License (GPL) R software (R Development Core Team, 2009). For example, Wu and Liu (2012a) developed a universal modeling framework to incorporate the R-based FME into a Fortran-based model such as the Soil and Water Assessment Tool (R-SWAT-FME) to conduct parameter optimization, sensitivity, and uncertainty analysis. Although this framework was tested for SWAT, it can be readily applied to other environmental models. Ricciuto et al. (2008) performed a Bayesian calibration of a simple carbon cycle model at global-scale using the Monte Carlo Markov Chain (MCMC) technique. For the well-established biogeochemical model CENTURY, Liu et al. (2008b) adopted a nonlinear inversion technique—PEST (PEST, 2003)—to match the simulated net primary production (NPP) with Moderate-Resolution Imaging Spectroradiometer (MODIS) NPP. Chen et al. (2008) used another technique—smoothed ensemble Kalman Filter (SEnKF)—to achieve the model-data assimilation for the same model.

In this study, the primary objective is to develop automatic calibration (auto-calibration) for the General Ensemble Biogeochemical Modeling Systems-Erosion Deposition Carbon Model (GEMS-EDCM) (Liu et al., 2003b, 2004), at multiple (especially the regional) scales, using the widely-used SCE algorithm and the R package FME. The updated EDCM is named EDCM-Auto, which can be implemented to derive the optimal parameter values for each individual pixel. GEMS, which originally manages the spatial data and prepares input for EDCM, is modified to provide observation data for the auto-calibration procedure and include options to either save these derived pixel-based parameters by EDCM-Auto or further derive the county-land cover specific parameter values through aggregation (i.e., parameter post-process) before model application. The developed EDCM-Auto can accommodate multi-parameter and multi-objective (i.e., satisfying multiple model output variables such as grain yield, NPP, and biomass) calibration and deal with land-cover changes. In a case study, however, we adopted the most sensitive parameter (i.e., maximum potential primary productivity) to demonstrate how the EDCM-Auto works for model auto-calibration due to the availability of observation data (grain yield for croplands and MODIS NPP for other land covers) at the regional scale.

2. Materials and methods

2.1. GEMS-EDCM description

2.1.1. GEMS description

The GEMS was developed by the U. S. Geological Survey (USGS) to provide spatially explicit biogeochemical simulations over large areas

via integrating well-established ecosystem biogeochemical models and various spatial databases (Liu et al., 2004). GEMS is a new type of land-use land-cover change (LULCC)-oriented, regional-level biogeochemical simulation system assimilating spatially dynamic databases such as climate, land cover, management, and disturbances, etc. GEMS has two major components: Input/output database management and multiple encapsulated ecosystem models. As an interface and platform, GEMS framework assists users with getting standardized data into and out of the biogeochemical models that are staged on the GEMS platform. Fig. 1 shows the schematic diagram of the GEMS model.

The current GEMS has encapsulated multiple site-scale biogeochemical models such as the CENTURY model (Parton et al., 1987, 1994), the EDCM model (Liu et al., 2003b), and the spreadsheet or paint-by-number model (PBN) (Liu et al., 2012a). Additional models can be added into the GEMS framework and share common input data layers. GEMS can drive these models simultaneously to perform ecosystem dynamics simulation over time and space, and it has been successfully used to simulate carbon dynamics (e.g., CO₂ and CH₄ fluxes and changes of carbon pools) in vegetation and soil at various spatial scales and for different regions (Dieye et al., 2012; Liu et al., 2008a, 2011; Tan et al., 2005, 2009, 2010; Zhao et al., 2010a,b).

Technically, the GEMS framework has relatively few pre-requisites that include HDF5 and netCDF4 libraries as well as the GNU XML parser libraries. Individual models are linked into the framework at compile time as standard C or C++ function calls. By using the control file which defines the input and output variables and other settings (e.g. the sampling rate, the size of the moving window, and the simulation period), a user can invoke any of the models with a single executable. The major function of GEMS can be described briefly as follows:

- 1) Reads various spatiotemporal database (climate, land cover, etc.) with standardized NetCDF4 format into data arrays and an attribute structure and pass them into a specific model library (e.g., EDCM library) which is responsible for writing out the required input files for a model (EDCM),
- 2) Drives a model (e.g., EDCM, CENTURY, and PBN) run with or without data assimilation followed by opening a series of output netCDF4 files and provides a function to write data arrays to those output files,
- 3) Utilizes a “moving window” to break up requested model tins into manageable sizes given hardware resources, yet in the end provides output files that are continuous for the entire requested geographical area.
- 4) Implements sampling approach to speed model simulations in addition to conventional approach that supports wall-to-wall or per-pixel only simulations. With the sampling approach, users can choose different sampling densities to run GEMS to meet their needs.

2.1.2. EDCM description

EDCM (Liu et al., 2003b) is a modified version of the CENTURY model (version IV) (Parton et al., 1987, 1994). Although EDCM retains the basic input and output file structures of CENTURY, many changes have been made in input parameters and simulation algorithms of the underlying biogeochemical processes. First, EDCM uses up to ten soil layers to simulate the soil organic carbon (SOC) dynamics in the whole soil profile, instead of one single top-layer (20-cm) structure of CENTURY. In each soil layer, EDCM characterizes the quantity and quality of SOC, following the practice used by CENTURY for the top soil layer. This treatment retains the proven applicability of CENTURY model for the topsoil layer and provides consistency between the top layer and the deeper layers in EDCM simulations. Second, EDCM can dynamically keep track of the evolution of the soil profile and carbon storage as influenced by soil erosion and deposition (Liu et al., 2003b), though CENTURY does not have the capability of simulating SOC dynamics in depositional environments. Third, instead of using a simple bucket hydrological sub-model of CENTURY, EDCM has improved the routine for simulating vertical water fluxes and soil moisture content to enhance the prediction of

Download English Version:

<https://daneshyari.com/en/article/4374949>

Download Persian Version:

<https://daneshyari.com/article/4374949>

[Daneshyari.com](https://daneshyari.com)