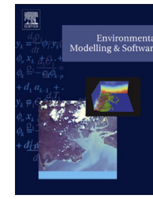
Environmental Modelling & Software 55 (2014) 25–31

Contents lists available at ScienceDirect

Environmental Modelling & Software

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



Short communication

A functional test platform for the Community Land Model

Dali Wang [a,*], Yang Xu [b], Peter Thornton [a], Anthony King [a], Chad Steed [a], Lianhong Gu [a], Joseph Schuchart [c]

[a] Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[b] Department of Geography, University of Tennessee, Knoxville, TN 37966, USA
[c] Joint Institute for Computational Sciences, University of Tennessee, Knoxville, TN 37966, USA

A R T I C L E I N F O

Article history:
Received 8 August 2013
Received in revised form
7 January 2014
Accepted 9 January 2014
Available online 1 February 2014

Keywords:
Community Earth System Model
Community Land Model
Functional test
Photosynthesis

A B S T R A C T

The realistic representation of key biogeophysical and biogeochemical functions is the fundamental of process-based ecosystem models. A functional test platform is designed to create direct linkages between site measurements and the process-based ecosystem model within the Community Earth System Models (CESM). The platform consists of three major parts: 1) interactive user interfaces, 2) functional test models and 3) observational datasets. It provides much needed integration interfaces for both field experimentalists and ecosystem modelers to improve the model's representation of ecosystem processes within the CESM framework without large software overhead.

Published by Elsevier Ltd.

1. Introduction

Over the past several decades, researchers have made significant progress in developing high fidelity earth system models to advance our understanding on earth system, and to improve our capability of better projecting future scenarios (Washington and Parkinson, 2005). The Community Earth System Model (CESM, http://www2.cesm.ucar.edu) is one of the US leading earth system models. CESM is being actively developed to support Department of Energy's climate and environmental research. Within CESM, the Community Land Model (CLM) is the active component to simulate surface energy, water, carbon, and nitrogen fluxes and state variables for both vegetated and non-vegetated land surfaces (Bonan, 1998; Dickinson et al., 2006; Oleson et al., 2010). In order to minimize uncertainty, error and bias in the earth system simulations, it is vital to get the fundamental processes correct and to investigate new theories of ecosystem function and new process representations within the context of earth system behavior. However, the complexity of the current CESM framework (both conceptual design and software implementation) makes function-level testing and exploration very difficult, especially at scales and levels of organization below that of the landscape and whole-ecosystem where many relevant field measurements are made.

The realistic representation of key biogeophysical and biogeochemical functions is the fundamental of process-based ecosystem models. In this paper, we present our approach to create direct linkages between site measurements and the process-based terrestrial ecosystem model (CLM). A functional test platform is designed to eliminate the majority of software complexity to allow scientists to interactively select external forcing, manipulate Plant Functional Type (PFT)-specific ecophysiological parameters and compare the key ecosystem functional representations with measurements and observations. It also preserves the maximum portion of code segment related to those key ecosystem functions. We believe that our experience in the design of the functional testing platform for the CLM can be beneficial to many other research programs which adapt the integrated environmental modeling methodology (Estreguil et al., 2014; Laniak et al., 2013).

2. The software system of the Community Land Model

Within the CESM framework, the CLM is designed to understand how natural and human changes in ecosystems affect climate. The model represents several aspects of the land surface including surface heterogeneity and consists of submodels related to land biogeophysics, the hydrologic cycle, biogeochemistry, human dimensions, and ecosystem dynamics.

* Corresponding author. Tel.: +1 8652418679.
E-mail addresses: wangd@ornl.gov (D. Wang), yxu30@utk.edu (Y. Xu), thorntonpe@ornl.gov (P. Thornton), kingaw@ornl.gov (A. King), steedca@ornl.gov (C. Steed), Lianhong-gu@ornl.gov (L. Gu), joseph.schuchart@zih.tu-dresden.de (J. Schuchart).

1364-8152/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.envsoft.2014.01.015

The software system of the global offline CLM includes two groups: models and scripts. The model group includes physical earth system components, such as the CLM, data atmosphere, stub ocean, stub ice and stub glacier. It contains a driver to configure the parallel computing environment and a coupled simulation system (physical earth system components and flux mapping functions between those components). It also includes several shared software modules and utilities, such as a flux coupler, a parallel Input/Output (IO) library, performance profiling libraries. The schematic diagram of the offline CLM software structure is shown in Fig. 1, which demonstrates that the CLM has to be incorporated with atmosphere and coupler as well as parallel IO, etc.

The whole CLM modeling system consists of more than 1800 source files and over 350,000 lines of source code. Fig. 2 shows the CLM software call tree using Yifan-Hu algorithm for graph layout. Each circle represents an individual subroutine with the area of circle showing the time spent on the subroutine. The directed edges show the procedure of software subroutine calls. The width of each edge indicates the number of subroutine invocations. The upper part is generally the software overhead of CLM, such as the IO, parallel communication, time management, and non-land earth system models setups. The lower part represents the CLM submodels, including several biogeophysical and biogeochemical models. More information on CLM computational characteristics can be found in another paper (Domke and Wang, 2012). Landscape surface is the basic data structure for CLM model development and software design.

Fig. 3 shows the hierarchical data structure for CLM landscape surface. Within this structure, lower level data arrays (e.g. pft-level data arrays) are accessible via references from the higher level data layers. Inside CLM, the land surface is represented by five primary landcover types: glacier, lake, wetland, urban, and vegetated portion. The vegetated portion of a gridcell is further divided into patches of PFTs, each with its own leaf and stem area index and canopy height. This hierarchical data structure makes the CLM initialization very complicated, since the CLM contains over 400 variables across multiple data layers over more than half million landscape surface gridcells.

3. Functional test platform design

In the CLM, PFT is the basic modeling concept. In our project, each biogeophysical or biogeochemical function of PFT is treated as a functional element. The purpose of functional testing is to provide direct comparison between model function and site experimental measurements at each functional element level. As shown in Fig. 3, the CLM landscape surface is represented by a hierarchical data structure, which make the model initialization complicated. In our platform, a single gridcell model is implemented, which consists of all the possible PFTs in the current version of CLM, and keeps the input and output interface of the functional element (e.g., key subroutine or module for ecosystem functions) unchanged, except the method to access global data arrays. Specifically, in our platform, a simplified global data structure is implemented to allocate memory space to host only necessary data for the testing of each individual functional element on a single gridcell landscape.

Fig. 4 shows the key testing data structure, which make the model initialization much simpler. Every data arrays can be accessible directly without hierarchical references. It eliminates the hierarchical data structure from the original CLM model, and the dimension of PFT-level data array is fixed at the maximum number of PFTs (that is 25) within the current version of CLM in order to represent the heterogeneity of each vegetated landscape surface. The functional test platform consists of three major parts: 1) interactive user interfaces, 2) standalone test models for functional elements and 3) query database for observational datasets. Based on those designs, our platform provides intuitive ways to enable direct model verification and model-data validation at individual functional element level.

3.1. Interactive user interfaces

The main function for those interfaces is to provide intuitive ways to setup computational experiments for functional testing, observational database establishment, and model-data comparison. Considering visualization capabilities, cross-platform compatibility, and future web-based presentation, our interfaces are implemented in Java using Java-based open source visualization libraries.

3.2. Standalone test model for functional element

The standalone test model for each key functional element consists of five key parts: 1) a generic functional test driver, which is designed to configure the functional test computing environment, configure a single gridcell model, and initialize physiological parameters and external forcing; 2) a generic functional test data structure, replacing the hierarchical CLM data structure, for a single gridcell model; 3) a initialization and output function for the single gridcell model; 4) a target functional element (e.g., “stomata” function in the following case study); 5) as well as other shared CLM data definitions (e.g. physical and chemical constants) which are used to eliminate unnecessary software structure changes to the target functional element. Users can either keep the original functional element unchanged, or modify the mathematical formula inside the functional element based on their own research. After that users can generate a standalone executable for the target functional element.

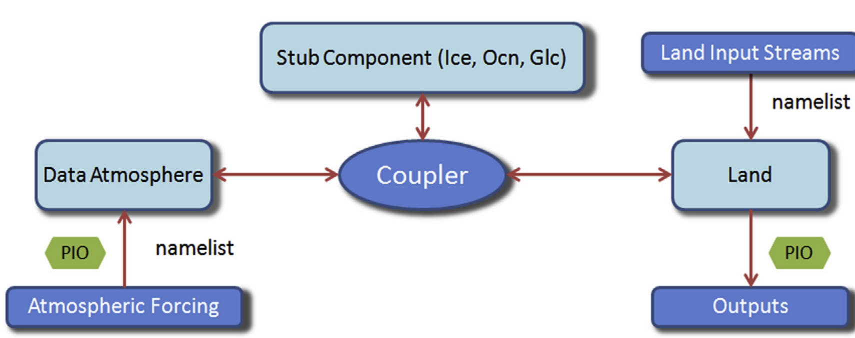


Fig. 1. Software configuration of a global offline CLM simulation. Several earth system model components are listed, including a land model (Land), a data atmospheric model (Data Atmosphere), stub sea ice model (Ice), ocean model (Ocn) and glacier model (Glc). PIO stands for Parallel IO.

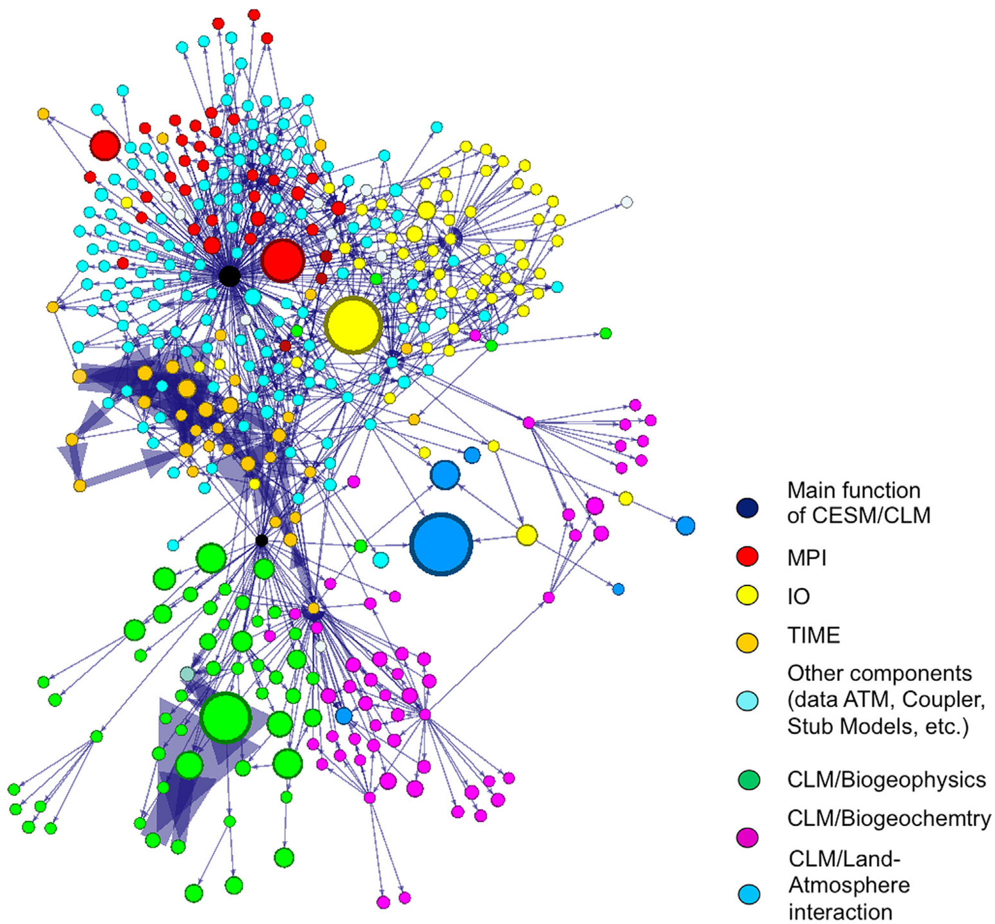


Fig. 2. The software structure (subroutine) of CLM. Each circle represents an individual subroutine with the area of circle showing the time spent on the subroutine with linear representation. The upper part shows the software overhead of CLM, such as the MPI, IO, time management and other earth system model setup. The lower part represents the CLM submodels. The land–atmosphere interactions at a landscape surface are shown in light blue in the middle part. The direct arrows represent the subroutine invocation procedure, and the weight of arrows shows the number of subroutine invocations with a linear representation. Fast subroutines related to Coupler and Fortran string manipulation (total execution time less than 1 s) have been filtered out from the graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. *Query database for observational datasets*

The purpose of the observational database is to provide an interactive search and visualization capability for direct model-data comparison. It consists of a query generation interface, a database, and result demonstration panels. First, users can choose specific datasets to be parsed and inserted into a database. Then Users can then use a dedicated interface to search observation data under given environmental situations.

```
gridcell_data_structure {
      pointers :: arrays of landunit_data_structure (<=5)
      pointers :: gridcell-level data arrays
                  (flux, water and energy, etc.)
}

landunit_data_structure {
      pointers :: arrays of soilcolumn_data_structure (=1)
      pointer :: landunit-level data arrays
}

soilcolumn_data_structure {
      pointers :: arrays of pft_data_structure (<=25)
      pointers :: soilcolumn-level data arrays
}

pft_data_structure {
      pointers :: pft-level data arrays
}
```

Fig. 3. The basic data structure for landscape surface within CLM. It is a single hierarchical data structure for the whole CLM landscape surface. Within this structure, lower level data arrays (e.g. pft-level data arrays) are accessible via references from higher level data layers.

3.4. *Usage scenarios*

Fig. 5 presents the two typical usages of our functional testing. In the first case, the testing interfaces are used to setup computational experiments and to test the target functional element under different environmental settings. The second case is a model-data comparison based on specific environmental settings. In this case, the testing interfaces are used to i) retrieve measured results from observational databases, ii) generate simulated results using the

```
function_test_data_structure {
      pointers :: grid-level data arrays
      pointers :: landunit-level data arrays
      pointers :: soilcolumn-level data arrays
      pointers :: pft-level data arrays
}
```

Fig. 4. The data structure for functional test. It is a globally accessible data structure to host only necessary data for each individual functional testing.

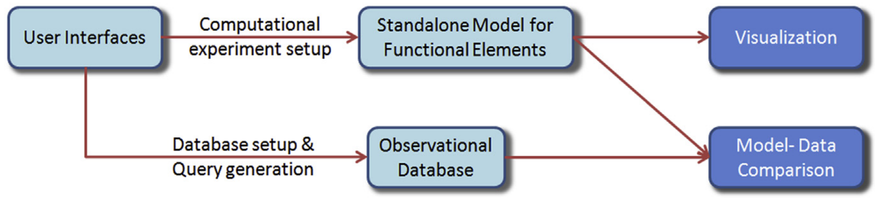


Fig. 5. Typical usages of the functional testing platform: 1) functional testing under different experimental settings, and 2) model-data comparison under similar specific experimental settings.

standalone test model, and then iii) visualize and compare the simulated and measured results together.

4. Case study: photosynthesis

Globally, photosynthesis accounts for the largest flux of CO_2 from the atmosphere into ecosystems and is the driving process for terrestrial ecosystem (http://en.wikipedia.org/wiki/Terrestrial_ecosystem). The importance of accurate predictions of photosynthesis over a range of plant growth conditions led to the development of a C_3 photosynthesis model by Farquhar et al. (1980) (FvCB), which has played a major role in defining the path towards scientific understanding of photosynthetic carbon uptake and the role of photosynthesis on regulating the earth's climate and biogeochemical systems. Within CLM (version 4.0), a dedicated subroutine, stomata, is designed to calculate leaf stomatal resistance and leaf photosynthesis (Thornton and Zimmermann, 2007). Specifically, a steady-state photosynthetic rate model (based on the FvCB model but adapted for C4 Plants following Collatz et al. (1992)) was implemented in a stomata subroutine to relate leaf gas exchange data to underlying limitations to photosynthesis at the leaf-tissue level due to the activity of *Rubisco*, regeneration of *Ribulose Biphosphate (RuBP),* and *Triose Phosphate Utilization (TPU)* limitation. Within CLM, the steady-state leaf photosynthetic carbon assimilation rate is driven by intercepted light, CO_2, temperature, humidity, available soil water as well as leaf processes, such as stomatal conductance.

There are extensive measurements and research reports that estimate photosynthesis for a variety of plant species. The basic components of a photosynthesis measurement system are the gas exchange chamber, infrared gas analyzer, flow meters, gas lines, CO_2 and water vapor filters, power batteries and a console with keyboard, display and memory. The left graph in Fig. 6 shows a typical measurement system (that is LI-6400 from LI-COR Biosciences), in which an air stream that has a known CO_2 concentration is constantly passed through the leaf chamber. The right graph shown in Fig. 6 is a typical curve to illustrate the photosynthetic assimilation rate (*A*) and related intercellular CO_2 concentration (*Ci*).

4.1. User interface

Fig. 7 shows the main interface which allows users to set all the required inputs in order to drive the functional test model for the stomata subroutine and choose the observational datasets. For the stomata subroutine, the input parameters are currently grouped into three sections on the main user interface by considering the parameter type: 1) Input parameters that are explicitly defined by the stomata subroutine interface (top-left panel), 2) Global input variables that are implicitly used by the stomata subroutine (top-right panel), 3) PFT-specific parameters (bottom-left panel).

4.2. Modeled result visualization and comparison

The main user interface allows users to set the range of certain parameter values. Our platform also provides a "Categorical Plot" function to empower users to control two different input parameters simultaneously. For example, by setting "atmospheric CO_2 concentration (*co2*)" as a "Range" variable and "vegetation temperature (*tl*)" as a "Categorical" variable, users could run multiple tests using different values of "*tl*" (e.g., $tl = 294$ K, 299 K, 304 K) by fixing the "*co2*" range. The "Categorical plot" function will demonstrate multiple XY plots as a "Categorical" variable is defined. Fig. 8 shows the modeled photosynthesis – intercellular CO_2 concentration curves (*A*–*Ci*) of deciduous forest tree (PFT = 7) at three different leaf temperatures. These visualizations allow scientists to easily perform the exploratory and sensitivity analysis in a multi-faceted manner.

4.3. Observational datasets and database design

Observational datasets from LeafWeb (http://leafweb.ornl.gov) are used in this project. Leafweb is a service-in-exchange-for-data-sharing project to develop a global database of biochemical,

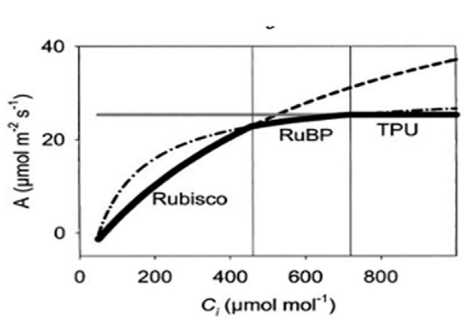


Fig. 6. An example of photosynthesis measurement system and an illustration of *A*–*Ci* curve based on FvCB model. *A*–*Ci* curve is one of key response curves to be measured in the field.

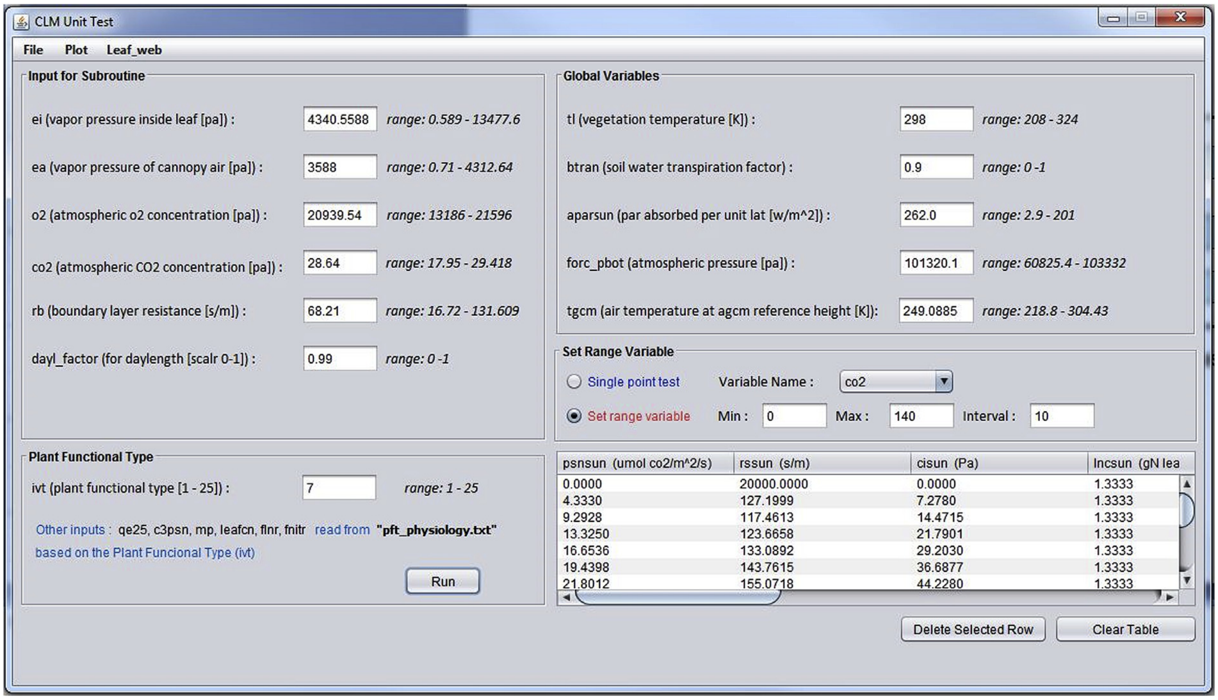


Fig. 7. The main user interface of the "stomata" functional test. This interface allow user to manipulate external forcing (such as temperature), plant physiological parameters (such as PFT type), execute the functional testing model in single or batch mode, and connect to observational database.

physiological, and biophysical properties of single leaves to support studies of plant functions and terrestrial carbon cycle modeling. LeafWeb provides automated numerical analyses of leaf gas exchange measurements. With the approval of the user, the data LeafWeb receives are preserved and captured. This effort is part of ongoing research and data management activities in the area of climate change science at Oak Ridge National Laboratory. As investigators use LeafWeb and contribute their data, the resulting "global leaf database" will grow and be freely available from the Carbon Dioxide Information Analysis Center (CDIAC). It is important to mention that Leafweb also provides its own unique photosynthesis analysis that involves fitting the FvCB model to the data for C3 plants with the newly developed Exhaustive Dual Optimization approach developed by Gu et al. (2010).

4.4. Model-data comparison and analysis

For demonstration purposes, only the datasets from Missouri Ozark Forest AmeriFlux Site (2004–2012) (http://tes-sfa.ornl.gov/node/15) are used in this paper. The majority plant types at this site belong to the temperate deciduous tree PFT (#7) in CLM model (version 4.0). The main user interface allow user to choose a subset or whole set of the observational datasets from Missouri Ameriflux Site. After that, those datasets are parsed and inserted into a

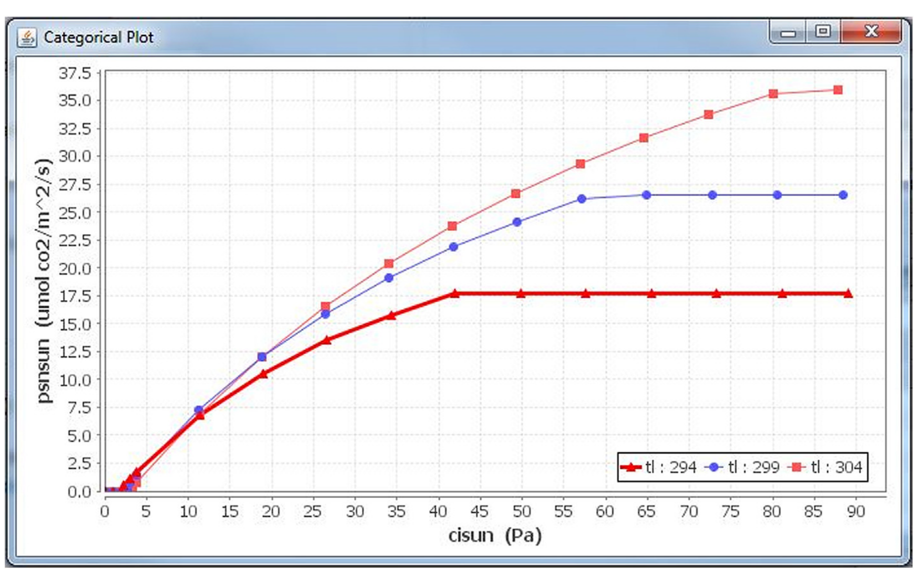


Fig. 8. Categorical plot by setting "vegetation temperature (*tl*)" as a "Categorical" variable. This graph demonstrates *A*–*Ci* curves at different leaf temperatures.

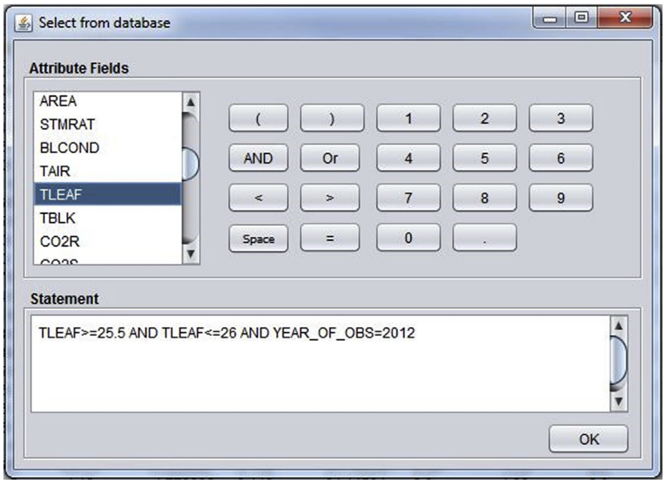


Fig. 9. Database search and query interface of functional test platform. This interface provides intuitive ways for users to generate standard SQL queries.

relational database automatically, and a dedicated interface (shown in Fig. 9) can be used to search the records via standard SQL queries. For example, a query to search all the records measured in the year of 2012, which have measured leaf temperature larger than 25.5 °C and less than 26 °C, returns 78 records from 13 separate measurement series (*A*–*Ci* curves).

Similarly a computational experiment can be conducted using the "stomata" test model in our platform (i.e., leaf temperature = 299 K, PFT = 7, CO_2 ranges from 0 to 140 Pa, soil water transpiration factor = 0.9). The comparison between the model results and observational datasets is shown in Fig. 10.

As show in Fig. 10, the model result is within the range of measurements only using data from year 2012 and given conditions (such as leaf temperature). However, there is a difference in the "CO_2 compensation" point, where the photosynthetic assimilation activities should be near to zero under low *Ci* concentration. In our model, the CO_2 compensation point is around $Ci = 3$ Pa, while the measured values are generally around $Ci = 5$ Pa. Note also that the observations are for net assimilation including leaf dark respiration (resulting in negative assimilation rates at low *Ci* and photosynthetic rates), while stomata returns only the photosynthetic rate. At least some of the variation between observed curves and the difference between observations and model results can likely be attributed to variable environmental conditions, such as water stress, temperature variation, humidity, as well as boundary layer resistance, etc. Refining the query with additional environmental constraints could elucidate these influences, and in future development the query might default to conditions set for the model experiment unless specifically overridden by the user. All the model and query results can be exported into files with comma separated values (CSV), and be further imported into interactive data analysis tools such as the Exploratory Data analysis ENvironment (EDEN) (Steed et al., 2012).

Fig. 11 illustrates an example of using EDEN to explore the relationship between four key variables: photosynthesis, intercellular CO_2 concentration, ambient CO_2 concentration, as well as the leaf temperature. Using mouse gestures, users can interactively query the datasets. In Fig. 11, the query is formed based on range selections for ambience CO_2 concentration and leaf temperature (see the highlighted regions on the last two axes). Similarly, those methods can be applied to model results for further model-data comparison including other environmental variables, such as water stress and humidity etc.

5. Conclusions and future work

This paper presents our approach to create direct linkages between field and laboratory measurements and the process-based/mechanistic ecosystem model (CLM). A functional test platform was designed not only to preserve the maximum portion of code segment related to key function element, but also to allow scientists to interactively select external forcing, manipulate

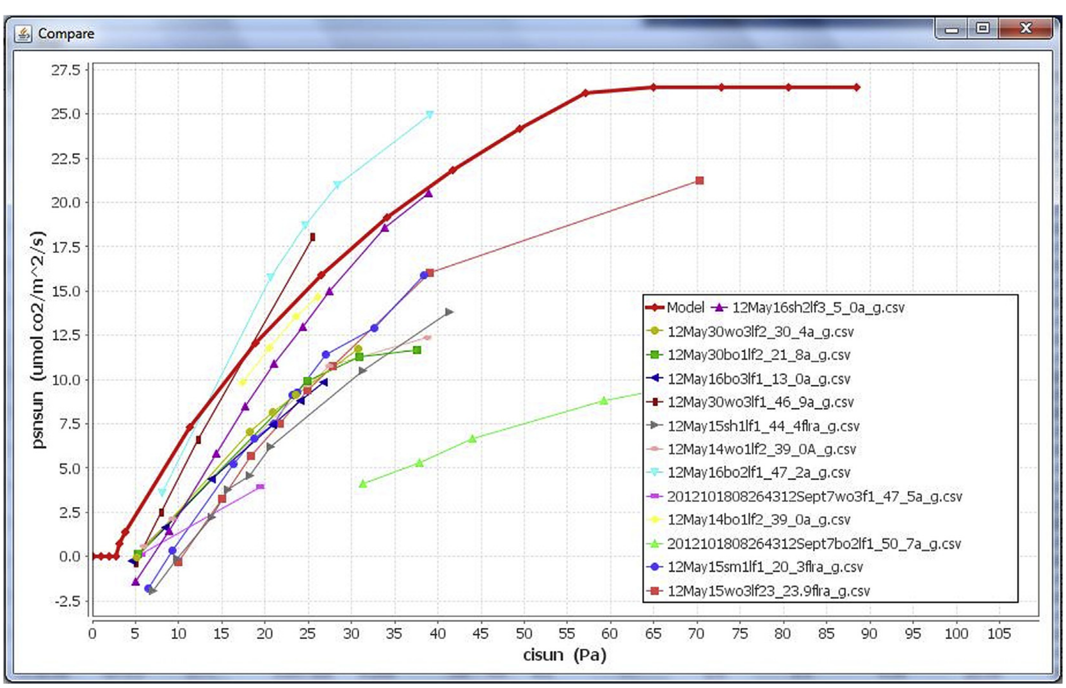


Fig. 10. The comparison between the model results and observational datasets. This graph show one modeled result (red) and 13 measured *A*–*Ci* Curves for temperate deciduous forest under similar environmental conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

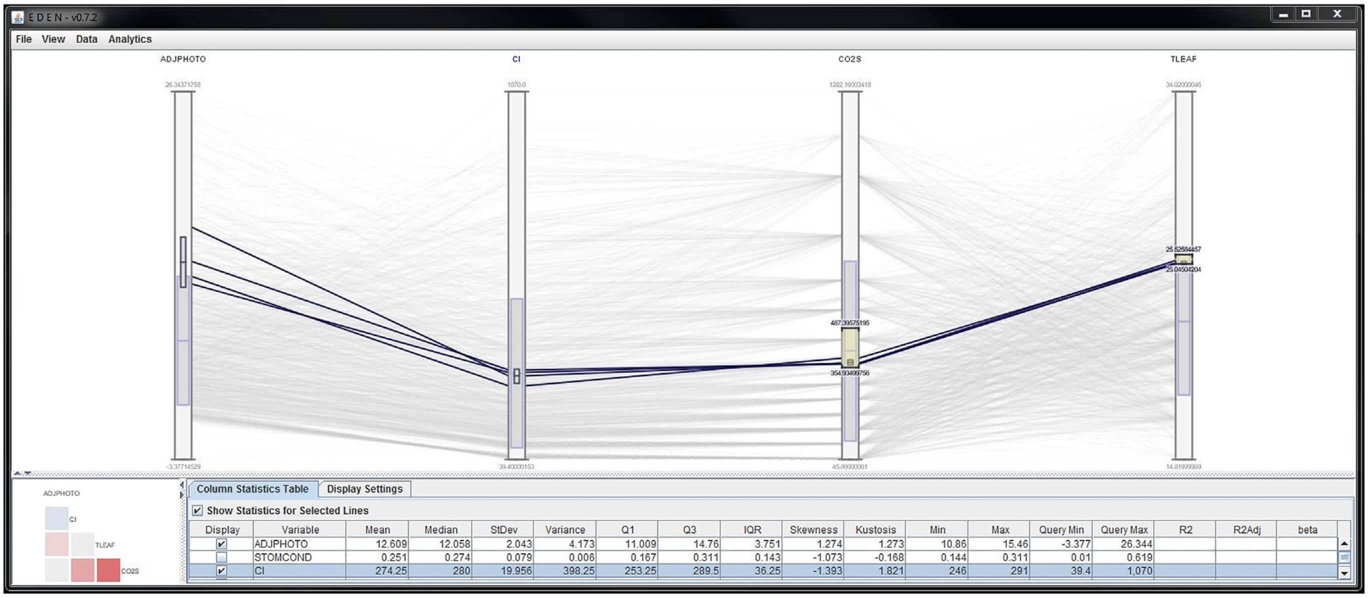


Fig. 11. The illustration of multivariate analysis of the observational datasets.

biogeophysical, biogeochemical, as well as ecophysiological parameters and compare the functional descriptions with measurements and observations. It provides much needed integration interfaces for field experimentalists and ecosystem modelers. The future work will focus on two directions: 1) to further extend our functional test platform for other key functional elements (such as canopy CO_2 and water flux calculation) and observational datasets (such as the datasets from AmeriFlux observation network) for model-data comparison over a variety of vegetation types and under different environmental conditions; 2) to work towards a seamless integration with external multivariate analysis toolkits beyond the parallel coordinates. Also inspired by the strong interests in web-based environmental information system developments (Blower et al., 2013; Demir and Krajewski, 2013) and cloud computing, we are in the process to develop a cloud-based cyberinfrastructure for our comprehensive CLM functional testing and observational database hosting.

The process-based functionality (such as the ecosystem functionality in our case), which incorporates the state-of-the-science understating of nature and human system, is the most significant and vital fundamentals of environmental software system. Therefore, we encourage the adaption of more flexible data structure for environmental software, aiming to facilitate model-data validation and verification on different mechanistic levels (ranging from individual physical process to overall environmental system responses) and at multiple spatial and temporal scales. Our testing platform is available on a variety of desktop computing environments (such as Windows, Mac and Linux). While it will take some time to follow the standard DOE/ORNL procedure to make the functional testing package a part of public accessible, community based software system, readers may contact authors for early distribution.

Acknowledgment

This research was funded by the U.S. Department of Energy (DOE), Office of Science, Biological and Environmental Research (BER). This research used resources of the Oak Ridge Leadership Computing Facility, located in the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the Department of Energy under Contract DE-AC05-00OR22725. Oak Ridge National Laboratory is managed by UT-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

References

Bonan, G.B., 1998. The land surface climatology of the NCAR land surface model coupled to the NCAR community climate model. J. Clim. 11, 1307–1326.
Blower, J.D., Gemmell, A.L., Griffiths, G.H., Haines, K., Santokhee, A., Yang, X., 2013. A web map service implementation for the visualization of multidimensional gridded environmental data. Environ. Model. Softw. 47, 218–224.
Collatz, G.J., Ribos-Carbo, M., Berry, J.A., 1992. Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. Aust. J. Plant Physiol. 19, 519–538.
Demir, I., Krajewski, W.F., 2013. Towards an integrated flood information system: centralized data access, analysis, and visualization. Environ. Model. Softw. 50, 77–84.
Dickinson, R.E., Oleson, K.W., Bonan, G., Hoffman, F., Thornton, P., Vertenstein, M., Yang, Z., Zeng, X., 2006. The Community Land Model and it's climate statistics as a component of the community climate system model. J. Clim. 19, 2302–2324.
Domke, J., Wang, D., 2012. Runtime tracing of the Community Earth System Model: feasibility study and benefits. In: 12th Workshop on Tools for Program Development and Analysis in Computational Science, Omaha, Nebraska, June 2012, Procedia CS 9, pp. 1950–1958.
Estreguil, C., Eigo, D., Caudullo, G., 2014. A proposal for an integrated modelling framework to characterise habitat pattern. Environ. Model. Softw. 52, 176–191.
Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. Planta 149, 78–90.
Gu, L.-H., Pallardy, S.G., Tu, K., Law, B.E., Wullschleger, S., 2010. Reliable estimation of biochemical parameters from C3 leaf photosynthesis–intercellular carbon dioxide response curves. Plant Cell. Environ. 33, 1852–1874.
Laniak, G.F., Olchin, G., Goodall, J., Voinov, V., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated environmental modeling: a vision and roadmap for the future. Environ. Model. Softw. 39, 3–23.
Oleson, K., Lawrence, D., Gordon, B., Flanner, M., Kluzek, E., Peter, J., Levis, S., Swenson, S., Thornton, P., Feddema, J., 2010. Technical Description of Version 4.0 of the Community Land Model (CLM).
Steed, C., Shipman, G., Thornton, P., Ricciuto, D., Erickson, D., Branstetter, M., 2012. Practical application of parallel coordinates for climate model analysis. In: International Conference on Computational Science, Data Mining in Earth Science, June 2012. EDEN is Available at: http://cda.ornl.gov/projects/eden/.
Thornton, P., Zimmermann, N., 2007. An improved canopy integration scheme for a land surface model with prognostic canopy structure. J. Clim. 20 (15), 3902–3923.
Washington, W.M., Parkinson, C.L., 2005. An Introduction to Three-dimensional Climate Modeling, second ed. University Science Books.