Development of mpi_EPIC Model for Global Agroecosystem Modeling

S. Kang ^{a,*}, D. Wang ^a, J.A. Nichols ^b, J. Schuchart ^c, K.L. Kline ^a, Yaxing Wei ^a, D.M. Ricciuto ^a, S.D. Wulschleger ^a, W.M. Post ^a, R.C. Izaurralde ^d

^a Climate Change Science Institute, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

^b Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

^c Center for Information Services and High-Performance Computing (ZIH), the Dresden University of Technology, 01062 Dresden, Germany

^d Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, 20740, USA

*Corresponding author: Telephone: 1-865-574-5948; email: kangs@ornl.gov

Abstract

Agroecosystem models that can incorporate management practices and quantify environmental effects are necessary to assess sustainability-associated food and bioenergy production across spatial scales. However, most developed agroecosystem models are at a plot scale. Tremendous computational need on simulations and datasets is necessary when large scales of high-resolution spatial simulations are conducted. We used the message passing interface (MPI) parallel technique and developed a master-slave scheme for an agroecosystem model, EPIC on global food and bioenergy studies. Simulation performance was further enhanced by applying the Vampir framework. On a Linux-based supercomputer, Cray XT7 Titan, we used 2048 cores and successfully shortened the running time from days to 30 minutes for a global 30 years of

modeling of a bioenergy crop at the resolution of half-degree (62,482 grids) with the message passing interface based EPIC (mpi_EPIC). The results illustrate that mpi_EPIC using parallel design can balance simulation workloads and facilitate large-scale, high-resolution analyses of agricultural production systems, management alternatives and environmental effects.

Keywords: bioenergy, food, high performance computing (HPC), message passing interface (MPI), parallel design, sustainability

1. Introduction

The intertwined global issues of food security, bioenergy and climate change call for large-scale high-resolution analysis of agricultural landscapes (Lynd et al., 2011; Lobell et al., 2011; Savage, 2013). Agroecosystem modeling provides opportunities for decision makers to evaluate trade-offs among various scenarios and management options. Spatial variability in climate (e.g., precipitation, temperature), soils, and crop management practices (e.g., fertilizer, irrigation, tillage, chemical control and harvest) affect crop productivity and environmental effects. Ideally, decision makers would prefer to guide management practices and production systems toward more sustainable options. But enormous data sets must be integrated and processed if modeling is to explore the implications of alternative land management across spatially heterogeneous scales. This is especially true when land managers or decision-makers want to evaluate trade-offs between management practices, environmental impacts and aspects of the sustainability of production systems in different locations.

The incorporation of spatially explicit data for climate, biophysical and crop management variables requires enormous computational capacity, particularly when the scope of analysis

1 expands from plot-scale to larger scales, such as county, state or global. Traditional
2 agroecosystem models, particularly biophysical process-based models, are designed for single-
3 site agricultural research tests, production potential assessment and management practice
4 optimization. They are valuable tools for studying novel cropping systems, new management
5 practices at specific sites or local analyses. However, applying these process-based models at
6 national and global scales is challenging due to the requirements for input data, calibration,
7 validation and simulation setups appropriate for each of thousands to millions of spatial points
8 (Nichols et al., 2011; Wang et al., 2012).

9 Super-computing advances have opened the door to detailed, global-scale modeling
10 simulations that were previously considered impractical if not impossible. Several preliminary
11 studies have demonstrated the potential of high-performance computing (HPC) application to
12 large-scale data management, modeling and analysis of agroecosystem functions. For example,
13 Hawick et al. (1997) proposed to use an HPC-based geographic information system for storage,
14 access and processing of weather, soil, and land management data. Kuo et al. (2004) designed an
15 HPC-visualization tool for use in displaying a three-dimensional environment for farming system
16 analyses. Wang et al. (2005) describe component-based software architecture for spatially
17 explicit ecosystem models that could be adapted to address the large scale high-resolution
18 ecosystem simulation challenge. High-resolution modeling with a parallel job launcher was
19 applied to study climate change impacts on pasturelands on a 200-processors cluster machine
20 (Vital et al., 2013). HPC can facilitate detailed management optimization at local scales, and can
21 help aggregate regional, national and global distributions of production and environmental
22 impacts as a tool for decision makings at larger spatial scales.

Demonstrations of regional and national high-resolution agroecosystem modeling under an HPC environment are limited (Nichols et al., 2011; Zhao et al., 2012). Nichols et al. (2011) constructed a modeling system to conduct regional high-resolution (30 m) assessment of production and environmental effects for cropping systems in the Midwest US with the high-performance computing Environmental Policy and Integrated Climate (HPC-EPIC) model. The study assigned jobs or simulation packages concurrently to multiple processors by “embarrassingly parallel” design – also referred to as “pleasingly parallel” for work that can be distributed because it requires little or no communication of results among tasks. This approach used Linux-based computing cluster and achieved a 40-fold reduction in the time required to run 140,000 EPIC simulations. A hybrid HPC approach was adopted to simulate national agricultural systems of Australia on a heterogenous distributed computing grid (Zhao et al., 2012). Similar to the embarrassingly parallel design in HPC-EPIC, jobs or simulation packages were sent to the individual computers on the grid. The hybrid HPC approach accelerates most processing by 1000-fold and complete jobs within a few days rather than months. However, both approaches are challenged with simulation load balancing which can delay overall execution time due to just a few extremely slow or failed executions. Slow or failed executions on a few nodes in clusters or an individual computer on a hybrid computing grid often occur (Nichols et al., 2011; Zhao et al., 2013). A few failed executions can cause significant processing delays when millions of simulations are conducted for large scale of high-resolution ecosystem studies involving massive data management (I/O processing) and bandwidth/memory.

To meet the heavy computational load of large-scale high-resolution agroecosystem simulations, we developed a design to address the problem of simulation load balancing using the message passing interface (MPI) on clusters or supercomputers. We directly coded MPI

1 routines into a widely applied agroecosystem model, EPIC, using a master/slave scheme to
2 dynamically manage model input, outputs and simulations. An application of global half-degree
3 biomass productivity analysis was used to test and evaluate this new mpi_EPIC approach to
4 agroecosystem modeling.

5 **2. mpi_EPIC development and computational improvement**

6 *2.1. Description of the EPIC model*

7 The EPIC model is a process-based agroecosystem model representing soil-crop-atmosphere
8 interrelationship with various management scenarios (Williams et al., 1984). Major processes in
9 the model include plant growth, development and production, nutrient cycling and nonpoint
10 sources pollution, emissions of greenhouse gases, and plant management practices. With
11 continuing testing and improvement, the EPIC model has become a widely applied tool designed
12 to meet needs for production estimation, environmental effect assessment, and climate change
13 studies (Easterling et al., 2001; Mausbach et al., 2004; Gassman et al., 2005; Liu et al., 2008;
14 Balkovič et al., 2013; Rosenzweig et al., 2014). Currently, EPIC is able to simulate over 100
15 crops and has been calibrated and applied in over 30 countries and regions (Gassman et al.,
16 2005).

17 *2.2. mpi_EPIC design*

18 Each EPIC simulation represents a single location or point, described by different input
19 variables such as site characteristics (e.g. location, slope, and soil), crop rotation, management
20 practices, and weather. Usually, the individual simulations are of short duration, typically 20-30
21 seconds for a 30 year daily time step simulation. However, the computational time can vary
22 widely depending on the complexity of involved biogeochemical processes. For example, the

numerical processes simulating soil water and nutrient transport, and gas diffusion in soils can be time-consuming due to multiple layers of soils and management practices. Since some points are more computationally demanding than others, the total time required to complete a full set of simulations can be reduced by distributing the work load to processors that have completed one job, and are available for another. This “load balancing” can be done under parallelization schemes, for example, by managing jobs among available cores on a supercomputer.

To accelerate high-resolution modeling at large scales, e.g., national and global ecosystem studies, we redesigned the EPIC module structure and calling sequence to be capable of dynamically balancing simulations across multiple cores of clusters or supercomputers. In this study, we adopt MPI to distribute site-specific simulation jobs among cores. A master/slave paradigm is applied to balancing job assignments and input control (Fig. 1). The master process generates and schedules thousands of jobs and sends them to each slave process for execution. The slave processes dynamically communicate with the master process for job completion and new job assignments. If a simulation fails, the simulation ID is sent to master process for check by users, and the next simulation will be continued. The failed simulations can be further examined through output log information.

2.3. Titan and Lustre file system advantages associated with high-resolution simulations

The computational platform used in this study is the Cray XT7 Titan supercomputer at the National Center for Computational Sciences (NCCS) at Oak Ridge National Laboratory (ORNL). Each Titan node consists of a 16-core AMD 6274 Opteron processor in conjunction with Nvidia Tesla K20X GPUs to boost its computational capabilities and improve energy efficiency. The 18,688 of these nodes perform at a theoretical peak of 27 PetaFLOPS (27

quadrillion floating point operations per second). A center-wide Lustre file-system provides 5 PB of disk space for all NCCS computing resources. The NCCS has implemented customized queue policies to encourage large jobs to run in a timely fashion on the systems at ORNL.

In order to obtain detailed information about how data are processed in our global simulation using mpi_EPIC in a parallelization design, the Vampir framework (Knüpfer et al., 2006) is used to gather information on simulation characteristics for diagnosis. The Vampir framework consists of two components, VampirTrace and Vampir. VampirTrace is used to instrument the source code of the program and to manage the recording of events while the instrumented program is running. Vampir is used to visualize the data/events gathered during the program execution. Within the Vampir framework, the instrumentation could be done via compiler instrumentation or using the Tuning Analysis Utility (TAU) instrumentor (Shende et al., 2006). A detailed explanation on how to configure the Vampir framework and TAU for environmental simulation can be found in Domke and Wang (2012).

2.4. mpi_EPIC parallel coding, testing and computational improvement

Our tests indicate that introducing a MPI parallel scheme to the existing sequential Fortran-based code of EPIC represents an important improvement in computational efficiency. This was achieved by splitting the standard EPIC model into two sub-modules. A new main routine (master) handles inputs and outputs (e.g., site, weather, parameters) and coordinates task assignments. The master also includes the Fortran-based MPI parallel scheme, which employs a master/slave algorithm. The slave module receives information from the master and reports results back. The biophysical processes simulations run by each slave are unchanged from the original EPIC.

After the coding was completed, a series of tests were conducted to examine the effectiveness of the parallel design, outputs, and load balancing. We used the parallel debug tool, Distributed Debugging Tool (DDT) to diagnose the performance and debug the errors on Titan. The outputs including format and values from mpi_EPIC were carefully compared with the outputs from the EPIC model (without the MPI modification) when processing the same simulation input data. We found that mpi_EPIC was able to duplicate the format and values of EPIC without error.

We examined load balancing (Fig. 2a) through a series of standard test jobs on mpi_EPIC. The tests illustrated that it took significantly different amounts of time to execute the same number but different simulations. For example, a job on processor 175 used over 120 seconds to complete, but it took under 40 seconds for another job with the same number of simulations on processor 170 (Fig. 2a). After we examined the execution time of those model components that were most time-consuming, we found that some simulations took longer because of numerical procedures used to solve subsurface biogeochemical processes. The simulations illustrated in Fig. 2a applied a standard job size of 20. Through a series of tests considering two to one hundred simulations per job, we found that a job size with five was the optimum for the mpi_EPIC simulations on Titan. We therefore adjusted the job size (number of simulations sent to a slave) to five when the master/slave scheme was applied and found similar execution times across cores (Fig. 2b). The mpi_EPIC using parallel design and optimized job size required 28 minutes to complete the same global simulation that required over two weeks using a sequential execution on a single core (Table 1). We also documented the time required to complete the same simulation test package with HPC-EPIC (embarrassingly parallel design) (Nichols et al., 2011). Table 1 illustrates that mpi_EPIC also reflects an improvement relative to HPC-EPIC,

despite the different types of cores and I/O systems used in Titan and Oak Ridge Institute Cluster (OIC). Of the approaches tested, mpi_EPIC offers the greatest capacity to efficiently complete a large number of agroecosystem simulations when using computing clusters or supercomputers.

Next, we conducted a scalability analysis of mpi_EPIC to determine to what extent we could reduce the execution time by adding computing power (additional cores) to complete simulations. The total computational time for identical global simulation setups with computing cores ranging from 16 to 2,048 was measured (Fig. 3). We observed nearly linear scalability in terms of proportional time reduction with each additional core when 128 or fewer cores were applied in the tests. This scalability fell slightly when more computing cores beyond 128 cores were used (Fig. 3). As more cores or computing resources are available, the speed of global simulation can be considerably increased. Although these tests were conducted on Titan, a super computer at ORNL, the model can be used similarly in the other computing clusters for applications.

3. mpi_EPIC application to large-scale agroecosystem modeling

3.1. Modeling workflow

In order to test mpi_EPIC, we conducted a study of global biomass productivity on Titan. The workflow was applied for a global half-degree assessment of production and environmental effects (Fig. 4). This design integrates global natural resources data sets, local management scenarios, model calibration and validation, simulation, and simulation output processing, and result analysis. The database of natural resources and management provided model inputs of weather, soil property, slope, land use and various management practices commonly used by

biophysical crop models. This workflow can also be applied to other large-scale, high-resolution crop modeling assessment to consider effects on agroecosystems.

3.2. mpi_EPIC Model input processing and calibration

Model inputs used for mpi_EPIC include three major categories of data (weather datasets, soil properties, and landscape attributes). We generated 62,482 simulation units using the half-degree global land mask from the Climate Research Unit – National Centers for Environmental Protection (CRU-NCEP) (Viovy, 2008) and 30-year daily weather input data (1980-2010) for each simulation unit. Daily CRU-NECP data include four sets of 6-hour weather data, total shortwave solar radiation (W m^{-2}), air pressure (Pa), temperature (K), U-wind (m s^{-1}), V-wind (m s^{-1}), and precipitation (mm). The daily radiation ($\text{MJ m}^{-2} \text{ d}^{-1}$), maximum temperature and minimum temperature ($^{\circ}\text{C}$), precipitation (mm), average relative humidity (%) and average wind speed (m s^{-1}) were calculated from CRU-NECP data. Next, we processed soil property data for each simulation unit using the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2008). Contents of sand, silt, clay, gravel content, bulk density, and soil organic carbon as well as cation exchange capacity, texture, and pH of two soil layers (0-50 cm, 50-100 cm) from the dominant soils were assigned to each simulation unit from the polygon-based HWSD. Slope data are also needed in the EPIC model and were calculated from 30 arc-second global elevation data, GTOPO30 (USGS, 2008) using ArcGIS10 (ESRI, 2011). The first simulation year, 1980 was treated as a spin-up or initialization to balance the most concerned soil variables such as soil water content and nitrogen that are sensitive to physiological processes in EPIC. Therefore, this year is excluded from simulation analysis.

In this study, we used mpi_EPIC to assess global biomass productivity and environmental effect of a bioenergy crop, switchgrass (*Panicum virgatum* L). This bioenergy crop has been recommended as one of the major bioenergy crop for biomass production in the US (Parrish and Fike, 2005). Agronomic characteristics and management for switchgrass cultivars in different regions were obtained from the field trial database developed at https://bioenergykdf.net/sites/default/files/Introduction_global_switchgrass_trial_dataset.pdf. The database consists of over 2,000 global observations from different continents, including the detailed information of site, cultivar, management practices, and biomass production. We designed 80 management files representing the various switchgrass cultivars and management practices (e.g., planting dates, fertilizer rates and harvesting dates) based on the studies from Kiniry et al. (2006), Wullschleger et al. (2010) and Nichols et al. (2011).

mpi_EPIC was carefully calibrated with the global switchgrass agronomic databases. Over 50 crop parameters would be considered for modification to different switchgrass cultivars in different ecological zones. Among them, the crop physiological parameters (radiation use efficiency, leaf area indices, harvest index, base and optimal temperature, nutrient parameters, root ratio) are critical to calibration. In this study, we used two common indicators of model performance evaluation using the REG procedure in SAS (SAS Institute, 2012). Our calibration results indicated that mpi_EPIC was able to simulate switchgrass productivity well with $r^2 = 0.67$ and root mean square errors (RMSE) less than 3.69 Mg ha^{-1} (Fig. 5). Most outliers of calibration for switchgrass cultivars were within one standard deviation of the observed values. Overall, the calibrations for the major ecological zones were acceptable although further improvement in the continental calibration are necessary for some zones and can be made when more experimental data are available.

3.3. *Simulation analysis of switchgrass biomass production*

We illustrate the simulated biomass production potential of switchgrass in Fig. 6. The average switchgrass productivity ranges from near zero in boreal and desert areas to a maximum exceeding 30 Mg ha⁻¹ in moist tropics. However, these extreme low and high values are uncertain because they are simulated for areas where reliable field-trial data were not available to support crop parameter calibration.

The mpi_EPIC model simulates production potential based on biophysical factors across the globe, regardless of current land cover, markets and zoning. High simulated productivity does not necessarily identify the best opportunities for cultivating switchgrass. This case study simulates high potential productivity in areas with favorable climate and soils even though other local conditions and current land cover may limit or prohibit cultivation. Areas with high productive potential also include lands that have been under cultivation for food crop production. This is not surprising because agricultural land is expected to represent relatively fertile areas in the ecological zones. Furthermore, a series of environmental variables such as soil organic carbon, nonpoint source pollutions are not included in this discussion because they aren't calibrated and validated with less available experiment data of switchgrass as a new bioenergy crop. Therefore, the enhanced global field experiment database for calibration and validation will be critical to further improve the global applications to food and bioenergy ecosystem studies.

4. Conclusion

To advance capabilities for analyses of global productivity and sustainability of food, bioenergy and climate change, we used the MPI parallelization technique and applied it to the mpi_EPIC model. The master/slave scheme in the model was capable of managing job allocation

1 and execution on an ORNL supercomputer. A case study to test capacity to complete high-
2 resolution agroecosystem simulations using the mpi_EPIC model was conducted for a bioenergy
3 crop, switchgrass. A global 0.5-degree resolution crop production assessment involved 62,482
4 simulations and was completed in less than 30 minutes. The simulation processed approximately
5 11 gigabytes of information in the model database. The results show that the mpi_EPIC model is
6 capable of serving a valuable role in exploring production and management scenarios and
7 corresponding environmental effects for food and bioenergy production under changing climate
8 conditions. Indeed, the mpi_EPIC modelling platform demonstrated that it can effectively
9 manage the 2.2 billion 1-km scale agroecosystem simulations projected by Wang and Kang
10 (2013) as required to assess global effects of different management options and production
11 scenarios. This study provides an insight of how traditional sequential ecosystem models can
12 effectively be integrated into available HPC systems for various ecosystem assessments

13 This study targets large scale of high-resolution agroecosystem modeling for regional and
14 global agricultural production and sustainability analysis, but the demand on intensive
15 computational resources, large dataset management and knowledge of HPC application on
16 clusters or supercomputers would be a challenge of its wide applications. Some HPC-based code
17 in mpi_EPIC associated with different module versions and libraries on machines or platforms
18 may have to be adjusted and validated. Limited training on parallel programming and experience
19 on clusters will facilitate solving the problem of applications. Additionally, detailed model input
20 data and experiment data for calibration and validation are the other challenges of regional and
21 global applications.

1 **Acknowledgement**

2 This research was thanks to computational resources made available through the Oak Ridge
3 Leadership Computing Facility, located in the National Center for Computational Sciences at
4 Oak Ridge National Laboratory. Contributions were also supported by the U.S. Department of
5 Energy (DOE) Bioenergy Technologies Office. Oak Ridge National Laboratory is managed by
6 UT-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

7 **References**

- 8 Balkovič, J., van der Velde, M. Schmid, E., Skalsky, R., Khabarov, N., Obersteiner, M.,
9 Sturmer, B., Xiong, W., 2013. Pan-European crop modelling with EPIC: Implementation,
10 up-scaling and regional crop yield validation, *Agricultural Systems* 120, 61-75.
- 11 Domke, J., Wang, D., 2012. Runtime Tracing of the Community Earth System Model:
12 Feasibility Study and Benefits, 12th Workshop on Tools for Program Development and
13 Analysis in Computational Science, Omaha, Nebraska, *procedia CS* 9: pp1950-1958, 2012
- 14 Easterling, W.E., Mearns, L.O., Hays, C.J., Marx, D., 2001. Comparison of agricultural impacts
15 of climate change calculated from high and low resolution climate change scenarios,
16 *Climate Change* 51, 173-197.
- 17 ESRI, 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research
18 Institute.
- 19 FAO/IIASA/ISRIC/ISSCAS/JRC, 2008. Harmonized World Soil Database (version 1.2). FAO,
20 Rome, IT and IIASA, Laxenburg, AT.
- 21 Forest Resources Assessment Program (FRA) of UN-FAO, 2000. Global ecological zoning for
22 the global forest resources assessment (FRA 2000). Rome, IT.

1 Gassman, P.W., Williams, J.R. Benson, V.W., Izaurrealde, R.C. 2005. Historical development and
2 applications of the EPIC and APEX models. CARD Working Paper 05-WP. Center for
3 Agricultural and Rural Development Iowa, US.

4 Hawick, K.A., James, H.A., Maciunas, K.J., Vaughan F.A., Wendelborn, A.L., Buchhorn, M.,
5 Rezny, M., Taylor, S.R., Wilson, M.D., 1997. Geographic information systems
6 applications on an ATM-based distributed high high-performance computing system,
7 High-performance Computing and Networking 1225, 1035-1037.

8 Kiniry, J.R., Sanderson, M.A., Williams, J.R., 2006. Simulating Alamo switchgrass with the
9 ALMANAC model, Agronomy Journal 88, 602-606.

10 Knüpfer, A., Brendel, R., Brunst, H., Mix, H., Nagel, W. E., 2006. Introducing the Open Trace
11 Format (OTF), in: V. N. Alexandrov, G. D. van Albada, P. M. A. Soot, J. Dongarra
12 (Eds.), International Conference on Computational Science (2), Vol. 3992 of Lecture
13 Notes in Computer Science, Springer, pp. 526–533.

14 Kuo, C., Shiau, Y., Huang, C., Shen, C., Tsai, C., 2004. Application of virtual reality in
15 ecological farmland navigating system. Proceedings of the Seventh International
16 Conference on High Computing and Grid in Asia Pacific Region, 0-7695-2138-X/04.

17 Liu, J., Fritz, S., van Wesenbeeck, C.F.A., Fuchs, M., You, L., Observeiner, M., Yang, H., 2008.
18 A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan
19 Africa in the context of global change, Global and Planetary Change 64, 222-235.

20 Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate trends and global crop production
21 since 1980, Science 333, 616-620.

22 Lynd, L.R., Aziz, R.A., de Brito Cruz, C.H., Chimpango, A.F.A., Cortez, L.A.B., Faaij, A.,
23 Greene, N. , Keller, M., Osseweijer, P., Richard, T.L., Sheehan, J., Chugh, A., van der

1 Wielen, L., Woods, J., van Zyl, W.H., 2011. A global conversation about energy from
2 biomass: the continental conventions of the global sustainable bioenergy project, *Interface*
3 *Focus* 1, 271-279.

4 Mausbach, J.M., Dedrick, A.R., 2004. The length we go: measuring environmental benefits of
5 conservation practices in the CEAP, *Journal of Soil and Water Conservation* 59, 96.

6 Nichols, J., Kang, S., Post, M., Wang, D., Bandaru, V., Manowitz, D., Zhang, X., Izaurrealde, R.,
7 2011. HPC-EPIC for high resolution simulations of environmental and sustainability
8 assessment, *Computers and Electronics in Agriculture* 79, 112-115

9 Parrish, D.J., Fike, J.H., 2005. the biology and agronomy of switchgrass for biofuels. *Critical*
10 *Reviews in Plant Sciences* 24, 423-459.

11 Rosensweig, C., Elliott, J., Deryng, D., Ruane, A.C., Muller, C., Arneth, A., Boote, K.J.,
12 Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Pionteck, F., Pugh, T., Schmid, E.,
13 Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in
14 the 21st century in a global gridded crop model intercomparison. *Proceeding of the*
15 *National Academy of Science of the United States of America* 111, 3268-3273.

16 SAS Institute, 2012. The SAS system for Windows. Release 9.2. SAS Inst., Cary, NC

17 Savage, N., 2013. Modeling: predictive yield, *Nature* 501, S10-S11.

18 Shende, S.S., Malony, A.D., 2006. The Tau Parallel Performance System. *The International*
19 *Journal of High Performance Computing Applications* 20, 287–331.

20 United States Geological Survey (USGS), 2008. GTOPO30, Earth Resources Observation and
21 Science (EROS) Center, USGS.

22 Viovy, N., 2008. CRU-NCEP reanalysis data, <http://www.cru.uea.ac.uk/cru/data/ncep/>.

- 1 Vital, J., Gaurut, M., Lardy, R., Viovy, N., Soussana, J., Bellocchi, G., Martin, R., 2013. High-
2 performance computing for climate change impact studies with the pasture simulation
3 model, *Computers and Electronics in Agriculture* 98, 131-135.
- 4 Wang, D., Carr, E., Gross, L.J., Berry, M.W., 2005. Toward ecosystem modeling on computing
5 grids, *Computing in Science & Engineering* 7, 44-52.
- 6 Wang, D., Kang, S., 2013. Modern accelerator technologies for spatially-explicit integrated
7 environmental modeling, In: X. Shi et al. (eds.), *Modern Accelerator Technologies for*
8 *Geographic information Science*, Springer Science Business Media, New York, pp 237-
9 251.
- 10 Wang, D., Kang, S., Nichols, J., Post, W., Liu, S., Zhao, Z. 2012. A computational framework
11 for spatially-explicit agroecosystem modeling: application to regional simulation, *Journal*
12 *of Computational Science* 4, 386-392.
- 13 Williams, J.R., Jones, C.A., Dyke, P.T., 1984. A modeling approach to determining the
14 relationship between erosion and soil productivity, *Transaction of the ASAE* 27, 129-144.
- 15 Wullschleger, S.D., Davis, E.B., Borsuk, M.E., Gunderson, C.A., Lynd, L.R., 2010. Biomass
16 production in switchgrass across the United States: database description and determinants
17 of yield, *Agronomy Journal* 102, 1158–1168.
- 18 Zhao, G., Bryan, B.A., King, D., Song, X., Yu, Q., 2012. Parallelization and optimization of
19 spatial analysis for large scale environmental model data assembly. *Computers and*
20 *Electronics in Agriculture* 89, 94–99.

1 **Figure Captions**

2 Fig. 1. Message passing interface (MPI) parallel design of the mpi_EPIC model for large scale
3 high-resolution agroecosystem modeling. Through a mast/slave scheme, jobs are dynamically
4 assigned to slave cores by the master core on computing clusters or supercomputers.

5 Fig. 2. Diagnosis and improvement of parallel design of the mpi_EPIC model using Vampir tools
6 on the Titan, a supercomputer at Oak Ridge National Laboratory: (a) execution time (in seconds)
7 for each job in different cores (processors) before the improved master-slave design, and (b)
8 execution time (in seconds) in different cores (processors) after the improved master-slave
9 design. Note that a different time scale is used in (b). The green represents physical-process
10 simulation, the yellow for I/O, and the red for master core communication processing.

11 Fig. 3. Parallel computing scalability analysis of the mpi_EPIC model on Titan. Speedup refers
12 to how much a parallel design is faster than a corresponding sequential algorithm when
13 processing the same workload.

14 Fig. 4. Workflow of data management, calibration-validation, and simulation and result analysis
15 of the mpi_EPIC model.

16 Fig. 5. Calibration and validation of the mpi_EPIC model for large-scale high-resolution
17 agroecosystem analysis. The solid line is 1:1 line, and the dotted lines show one standard
18 deviation from the 1:1 line.

19 Fig. 6. Potential global biomass productivity of switchgrass simulated by the mpi_EPIC model.

20

1

2

1

2

1

2

1

2

1

2

1

2