Manuscript Details

Manuscript number	COMPAG_2018_396_R1
Title	A dynamic agricultural prediction system for large-scale drought assessment on the Sunway TaihuLight supercomputer
Article type	Research Paper

Abstract

Crop models are widely used to evaluate the response of crop growth to drought. However, over large geographic regions, the most advanced models are often restricted by available computing resource. This limits capacity to undertake uncertainty analysis and prohibits the use of models in real-time ensemble forecasting systems. This study addresses these concerns by presenting an integrated system for the dynamic prediction and assessment of agricultural yield using the top-ranked Sunway TaihuLight supercomputer platform. This system enables parallelization and acceleration for the existing AquaCrop, DNDC (DeNitrification and DeComposition) and SWAP (Soil Water Atmosphere Plant) models, thus facilitating multi-model ensemble and parameter optimization and subsequent drought risk analysis in multiple regions and at multiple scales. The high performance capability also opens up the possibility of real-time simulation during droughts, providing the basis for more effective drought management. Initial testing with varying core group numbers shows that computation time can be reduced by between 2.6-3.6 times. Based on the powerful computing capacity, a county-level model parameter optimization (2043 counties for 1996-2007) by Bayesian inference and multi-model ensemble using BMA (Bayesian Model Average) method were performed, demonstrating the enhancements in predictive accuracy that can be achieved. An application of this system is presented predicting the impacts of the drought of May-July 2017 on maize yield in North and Northeast China. The spatial variability in yield losses is presented demonstrating new capability to provide high resolution information with associated uncertainty estimates.

Keywords	drought; dynamic prediction; supercomputer; accuracy; risk analysis.
Corresponding Author	chaoqing yu
Corresponding Author's Institution	Tsinghua University
Order of Authors	Xiao Huang, chaoqing yu, Jiarui Fang, guorui huang, Shaoqiang Ni, Jim Hall, Conrad Zorn, Xiaomeng Huang, Wenyuan Zhang
Suggested reviewers	Haohuan Fu, Qingyun Duan, Enli Wang, Jasper Vrugt

Submission Files Included in this PDF

File Name [File Type]

Cover_letter.docx [Cover Letter]

Response_Huang.docx [Response to Reviewers]

Highlights.docx [Highlights]

Manuscript_Xiao_Huang_revision.docx [Manuscript File]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
Data will be made available on request

Dear editor,

On behalf of all the authors, I hereby submit our original paper "A dynamic agricultural prediction system for large-scale drought assessment on the Sunway TaihuLight supercomputer" as a research paper to "Computers and Electronics in Agriculture".

Drought hazards with increasing frequency and severity in global climate change are likely to cause significant losses to agricultural production. Using process-based crop models for real-time crop growth simulation and prediction will be helpful for drought impact assessment and mitigative actions. The manuscript presents an integrated system for the dynamic prediction and assessment of agricultural yield using the top-ranked Sunway TaihuLight supercomputer platform. The main characteristics of this system includes: (i) further acceleration of crop models and high-performance computing for large-scale crop modeling; (ii) combination of Bayesian inference and Bayesian model average to improve predictive accuracy; (iii) real-time simulation and prediction based on observational and scenario forces; (iv) risk analysis of yield losses in multiple scales and its spatial dependency. We showed several real applications to demonstrate the effectiveness of our system in such powerful platform.

This article has not been published elsewhere in whole or in part. All authors have read and approved the content, and agree to submit for consideration for publication in the journal. No ethical/legal conflicts are involved in the article.

Thank you for your time and consideration of this manuscript. We are looking forward to your reply.

Sincerely yours,
Chaoqing Yu

-Reviewer 1

[1] Kindly give more details on soil parameters considered in the paper for crop yield prediction? Have you considered fertilizer parameters?

Response:

(i) The soil parameters (including clay content, hydraulic conductivity, maximum soil depth, etc) are from the database "The Soil Database of China for Land Surface Modeling" (Shangguan et al., 2014). We calculate the value of each county by the average of all grids in the corresponding county.

(ii) Only DNDC model needs fertilizer input. It is derived from the statistical record in yearbook and we use the averaged fertilizer amount as the input.

(iii) We only optimize the parameters related to crop growth. These parameters for optimization are in accordance with Yu's publication in 2018 (Yu et al., 2018).

(iv) The detailed information of model setting is added in section 4.2, **Line 384-392**.

[2] The authors have claimed greater accuracy in both calibration and validation periods with a higher coefficient-of-determination (R^2) value and lower root-mean-square error (RMSE) value, but results are not presented for R^2 & RMSE

Response:

The results of RMSE and R^2 are presented in Fig. 6. For example, for wheat yield simulation, the RMSE of BMA ensemble is at least 0.1 ton/ha lower than any individual model simulation while R^2 at least 0.03 higher. We add a more explicit statement in **Line 406-409**.

[3] Authors claim the Speed up of 2.6-3.6, but lacks interpretation in results section

Response:

(i) The values of 2.6 and 3.6 mentioned here are the speedup ratios of the total run time (MPE-based module + CPE-based module). The performance of acceleration for CPE-based module is over 8 times. We add **Table 2** to explicitly show it.

(ii) The reason that the overall speed up ratio is lower than that of CPE-based module is that MPE-based module cannot get accelerated and this part accounts for over 10% of computing task. Besides, the communication between CPEs and MPE generates extra costs (e.g. copying data from main memory to LDM). It is added in **Line 372-376**.

(3) The reason why the speedup of DNDC is 3.6 while SWAP is 2.8 has been explained in **Line 376-379**.

36

Response:

(i) The related work about crop model uncertainty has been listed in section 1, **Line 67-73**, which includes Bayesian inference, data assimilation, multi-model ensemble and climatic input ensemble.

(ii) The related work about HPC for crop modelling is listed in section1, **Line 93-95**. But we have to admit that it is not clear. We add a statement in **Line 95-97**.

(iii) The related work about probability analysis of yield loss is presented in section 1, **Line 107-115**, which includes marginal distribution and joint distribution.

Response:

This error is fixed. Besides, in section 2, we add the full names of some abbreviations (e.g. DDR3, PCI-E)

-Reviewer 2

Response:

To perform HPC for large-scale modelling, we should consider not only the run time of the simulation, but also the energy and monetary costs of such task. So the general purpose of the MPE-CPE design is to minimize the power consumption while maintain the computing capacity. Compare with the MPE with 8G main memory, each CPE is equipped with only 64KB local memory but owns similar computing capacity of MPE. As a result, a large amount of electricity to maintain data in memory can be saved. In the TOP 500 list (https://www.top500.org/lists/), we can find the power efficiency of the Sunway TaihuLight supercomputer is better than many other computing platforms based on Intel structure, which demonstrates the strength of such MPE-CPE hybrid structure. The brief explanation is added in **Line 160-161**.

Response:

(i) The main driving elements of crop modelling in AquaCrop, DNDC and SWAP models are water, nitrogen and radiation, respectively, which are also the main factors limiting crop growth in China (Huang et al., 2017). As a result, we believe the multi-model ensemble prediction based on these three models can be applicable in the complex environment of China. The reason of using these three models in this research is added in **Line 198-200**.

(ii) We have to admit the crop growth module in SWAP model is modified from WOFOST model. However, I'm afraid that, in the source code of WOFOST model, the simulation for crop rotation system (e.g. the maize-wheat rotation in North China Plain) is still lacking. However, in SWAP model, such function for crop rotation system is realized (Kroes et al., 2009). That's why we decide to use SWAP model instead of WOFOST model.

Response:

A figure to illustrate the detailed steps of model acceleration for DNDC and SWAP models is added as Fig. 4. The statement of this figure is added in **Line 254**.

Response:

In SWAP model, the modelling of soil water dynamic and soil temperature are both based on *continuous differential equations*. The original matrix for numerical calculation can be decomposed into several small matrix blocks, and therefore CPE acceleration can be used in such matrix block level. However, in DNDC model, the simulations of these processes are based on *empirical equations*. The calculation in lower soil layer depends on the completion of upper layer, which is practically difficult for further acceleration.

Response:

In Bayesian inference, it is assumed:

$$Y = E[Y'|\theta,F] + \varepsilon_y$$

Where Y is the observational value; Y' is the model prediction; θ is the model parameter; F is the input driving force and ε_y is the Gaussian noise with $\varepsilon_y \sim N(0, \sigma_y^2)$. For parameter optimization, we have to maximize the value of Eqn 1. By Jeffreys' prior, we can simplify Eqn 1 to Eqn 2 without assuming the specific value of σ_y (Kavetski et al., 2006; Box and Tiao, 2011). However, if we don't use Jeffreys' prior, the value of σ_y is necessary. I think the discussion of σ_y is far beyond the scope of this study and Jeffreys' prior is preferred in large-scale simulation. The brief explanation is added in **Line 270-272**.

In Section 4.1, Line 358, the author provides a figure to illustrate the acceleration performance. I think it is the overall running time combining the MPE-based module and the CPE-based module shown in Table 1. If so, I'm wondering what the speed-up ratio for CPE module is, as the CPE acceleration part is the highlight of this paper after all.

Response:

Yes, the acceleration performance in section 4.1 refers to the total run time. The speedup ratio of CPE-based module is over 8 times. Such information related to the CPE-based module is added as Table 2. The statement for this table is added in **Line 358-360**.

[7] In Section 4.4, Line 427, how does the author determine the parameter values of GEV distribution, including the location, scale and shape parameters? The evaluation of return period is sensitive to these parameters. So more detail should be explained.

Response:

In this research, we firstly perform the long-term yield simulation from 1960 to 2010 using optimal model parameter values obtained in section 4.2. Then parameters of GEV distribution are estimated for each county by maximum likelihood method to fit the 50-year ensemble simulations. Eventually, in dynamic assessment, we can use these parameters to calculate the probability corresponding to the real-time yield loss. The process is briefly explained in **Line 456-459**.

[8] In Section 4.4, Fig. 9(a), why is the return period of yield loss of Inner Mongolia so high? Besides, in Line 464, the results calculated from joint distribution and marginal distribution are compared to illustrate the effectiveness of Copula method. But why is the joint distribution better when there is obvious discrepancy within the two results?

Response:

(i) The high return period of yield loss in Inner Mongolia could be due to the irrigation ratio in our database is not very accurate in this region. In the past years, the irrigation ratio has increased in several counties of Inner Mongolia, while it is difficult for us to update our database to the real-time level. This issue is discussed in section 5 for further improvement. A brief explanation of this high return period is added in **Line 466-468**.

(ii) For Gumbel Copulas, the independent situation (the product of marginal distribution) is the special case when the Kendall coefficient τ equals 0. So if the long-term data reveals independence between two regions, the result of copula function will be equal to the product of marginal distribution. However, in this

5

165 research, the Kendall coefficient of 50-year yield between North China and Northeast
166 China approaches 0.7, which implies significant correlation between these two
167 regions. As a result, we believe the return period estimated by Copula is more
168 objective.

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185 # References:

186 [1] Box, G.E., Tiao, G.C., 2011. Bayesian inference in statistical analysis. John Wiley & Sons.

187 [2] Huang, X., Huang, G., Yu, C., Ni, S., Yu, L., 2017. A multiple crop model ensemble for
188   improving broad-scale yield prediction using Bayesian model averaging. FIELD CROP RES
189   211, 114-124.

190 [3] Kavetski, D., Kuczera, G., Franks, S.W., 2006. Bayesian analysis of input uncertainty in
191   hydrological modeling: 1. Theory. WATER RESOUR RES 42.

192 [4] Kroes, J.G., Van Dam, J.C., Groenendijk, P., Hendriks, R., Jacobs, C., 2009. SWAP version
193   3.2. Theory description and user manual. Alterra.

194 [5] Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth
195   system modeling. J ADV MODEL EARTH SY 6, 249-263.

196 [6] Yu, C., Huang, X., Chen, H., Huang, G., Ni, S., Wright, J.S., Hall, J., Ciais, P., Zhang, J.,

197   Xiao, Y., Sun, Z., Wang, X., Yu, L., 2018. Assessing the Impacts of Extreme Agricultural

198   Droughts in China Under Climate and Socioeconomic Changes. Earth's Future.

1) Further acceleration of crop models and high-performance computing for large-scale crop modeling;

2) Combination of Bayesian inference and Bayesian model average to improve predictive accuracy;

3) Real-time simulation and prediction based on observational and scenario forces;

4) Risk analysis of yield losses in multiple scales and its spatial dependency.

1 # A dynamic agricultural prediction system for large-scale

2 # drought assessment on the Sunway TaihuLight supercomputer

3 Xiao Huang[a, b], Chaoqing Yu[a,*], Jiarui Fang[c, d], Guorui Huang[a, e], Shaoqiang Ni[a],

4 Jim Hall[b], Conrad Zorn[b], Xiaomeng Huang[a, d], Wenyuan Zhang[a]

5

6 [a] *Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System*
7 *Science, Tsinghua University, Beijing, China*
8 [b] *Environmental Change Institute, Oxford University, Oxford, UK*
9 [c] *Department of Computer Science and Technology, Tsinghua University, Beijing, China*
10 [d] *National Supercomputing Center in Wuxi, Wuxi, China*
11 [e] *Twente Water Centre, University of Twente, Twente, Netherlands*

12

13 [*]*Corresponding to* chaoqingyu@gmail.com

14

15

16

Abstract: Crop models are widely used to evaluate the response of crop growth to drought. However, over large geographic regions, the most advanced models are often restricted by available computing resource. This limits capacity to undertake uncertainty analysis and prohibits the use of models in real-time ensemble forecasting systems. This study addresses these concerns by presenting an integrated system for the dynamic prediction and assessment of agricultural yield using the top-ranked Sunway TaihuLight supercomputer platform. This system enables parallelization and acceleration for the existing AquaCrop, DNDC (DeNitrification and DeComposition) and SWAP (Soil Water Atmosphere Plant) models, thus facilitating multi-model ensemble and parameter optimization and subsequent drought risk analysis in multiple regions and at multiple scales. The high computing capability also opens up the possibility of real-time simulation during droughts, providing the basis for more effective drought management. Initial testing with varying core group numbers shows that computation time can be reduced by between 2.6-3.6 times. Based on the powerful computing capacity, a county-level model parameter optimization (2043 counties for 1996-2007) by Bayesian inference and multi-model ensemble using BMA (Bayesian Model Average) method were performed, demonstrating the enhancements in predictive accuracy that can be achieved. An application of this system is presented predicting the impacts of the drought of May-July 2017 on maize yield in North and Northeast China. The spatial variability in yield losses is presented demonstrating new capability to provide high resolution information with associated uncertainty estimates.

Keywords: drought; dynamic prediction; supercomputer; accuracy; risk analysis.

1. Introduction

The growing pressure on food security is expected to continue in the coming decades due to global climate change and population growth (Adams et al., 1998; Bloom, 2011; Field, 2012). The risk of agricultural production, posed by the increasing frequency and severity of extreme events (Howden et al., 2007; Lobell et al., 2008), calls for more accurate and effective predictions of crop growth dynamic on large scales (Rosenzweig and Tubiello, 2007; Yu et al., 2018). Process-based crop models are increasingly popular tools for climate change impact and adaptation studies (Rosenzweig et al., 2013; Rosenzweig et al., 2014). A real-time decision support system based on crop model simulations and predictions could provide useful information for users with different goals (e.g. farmers, decision makers), and hence improve our ability for disaster mitigation (Van Ittersum et al., 2008; Jakku and Thorburn, 2010; Yu et al., 2014). However, such applications still face a number of key challenges (Holzworth et al., 2015). We identify at least three of these as: (i) reducing predictive uncertainty, (ii) the computing requirement to enable simulations at sufficiently high spatial and temporal resolutions, and (iii) the evaluation of yield risk.

The first challenge is how to reduce the uncertainty of model simulations and provide more accurate predictions. As most crop models are primarily developed based on site experiments, their scaling and application to larger areas, such as county, state or nation, are likely to introduce predictive uncertainties (Ewert et al., 2015), arising from assumptions regarding model structures, model parameters, and calibration/validation data, amongst others. Several methods have been explored to address this issue, including Bayesian inference (Iizumi et al., 2009; Dumont et al., 2014) and parameter optimization (Guo et al., 2006) for parameter uncertainty, multi-model ensemble predictions for model structure uncertainty (Martre et al., 2015; Huang et al., 2017), remote sensing data assimilation to correct the state variables (De Wit and Van Diepen, 2007) and the ensemble of climatic force for input uncertainty (Baigorria et al., 2008; Tao et al., 2009). However, most of these studies only address an individual source of uncertainty with fewer integrating multiple sources of uncertainty systematically. Therefore crop predictions in large region still remain highly uncertain.

The second challenge relates to computing efficiency and the need to design an advanced tool for large-scale crop simulation with high performance computing (HPC) technology. There are a number of major obstacles to achieve this: (i) The computational requirement for crop modeling is extremely large, both in the real-time simulation and scenario prediction at high spatial and temporal resolutions and for robust uncertainty analyses such as Markov Chain Monte Carlo (MCMC) methods. (ii) Unlike global climate models (Neale et al., 2010) or land models (Oleson et al., 2010) which consider the possibility of parallel optimization across a computer cluster at the beginning of their design, most traditional crop models are developed targeting to

87   deploy on one single node. The much needed redesigning of workflows in these
88   traditional models will require significant effort to exploit parallelism with large-size
89   computing resources (Holzworth et al., 2015). (iii) The computing kernels inside crop
90   models work in serial mode on general proposed processors. Complex dependencies
91   exist between different kernels, which make it more difficult to get accelerated in
92   parallel by multiple-threading method on powerful accelerators like Graphic
93   Processor Units (GPUs) and Intel Xeon Phis. Several studies have made attempt to
94   use HPC technologies in large clusters for regional simulation (Vital et al., 2013;
95   Zhao et al., 2013; Elliott et al., 2014). These researches mainly focus on the parallel
96   computation in grid level or between different models, as well as the high-speed
97   transformation of model input/output files into standard format. However, we find no
98   evidence in the literature attempting to further accelerate the scientific algorithms of
99   crop models, which account for the majority of simulation time in crop modeling
100   processes.
101
102   The final challenge identified relates to the post-processing of simulation results into
103   useful information for users – in particular the spatial distribution of production losses
104   during periods of drought and the severity of yield losses for given return periods. The
105   spatial distribution of production losses provided by the crop model simulation across
106   large scale is essential for the monitoring of drought evolution (Báez-González et al.,
107   2002; Launay and Guerif, 2005). To reveal the severity of yield losses in the local
108   history, return periods seem to be a more useful index because it clearly demonstrates
109   the comparable features in longer climatic background and is widely accepted by the
110   public (Fernández and Salas, 1999; Bonaccorso et al., 2003). Besides the marginal
111   distribution of loss in each region (Yu et al., 2014; Skakun et al., 2016), the joint
112   distribution in multiple areas (Bárdossy and Pegram, 2009; AghaKouchak et al., 2010;
113   Gaupp et al., 2016) can offer more critical information of the spatial correlation of
114   agricultural production, and it will lead the decision makers to a comprehensive
115   assessment of the overall agricultural risk.
116
117   In this paper, we seek to address these three challenges through the development of a
118   novel integrated decision support system. The crop models AquaCrop (Steduto et al.,
119   2009), SWAP (Kroes et al., 2000) and the biogeochemical model DNDC (Li et al.,
120   1992), which are all validated from field scale to region scale for yield prediction (De
121   Wit and Van Diepen, 2007; Heng et al., 2009; Yu et al., 2014; Huang et al., 2015;
122   Huang et al., 2017), are mapped into a supercomputer platform, entitled Sunway
123   TaihuLight (Fu et al., 2016; Zhao et al., 2018). The crop simulation in each model is
124   modified to fit the hardware architecture of this supercomputer for model acceleration.
125   Bayesian inference and BMA method (Huang et al., 2017) are used in model
126   prediction to reduce the parameter uncertainty and model structure uncertainty
127   respectively. We apply a scenario analysis method (Yu et al., 2014) to assess the
128   evolution of crop growth under uncertain drought development. Finally, the system
129   estimates both the marginal distribution of yield loss in different scales and the joint
130   distribution of different regions using a Copula method.

Section 2 briefly introduces the architecture of the Sunway TaihuLight supercomputer. Section 3 describes the integrated framework for agricultural risk analysis, including the methodology and technological process. In section 4, the case studies about model acceleration, parameter optimization, dynamic drought impact assessment and probability analysis of using this system for real practice are illustrated. Finally the discussion and summary is presented in section 5.

2. The platform: Sunway TaihuLight supercomputer

The Sunway TaihuLight supercomputer based at the National Supercomputing Center in Wuxi, China (Fu et al., 2016) has held the 47th to 50th (most recent) TOP500 rankings of global supercomputers (https://www.top500.org/lists/). It achieves a Linpack performance (Dongarra et al., 2003) of 93 PetaFlops from a theoretical peak of 125 PetaFlops, with computing power originating from 40,000 SW26010 processors arranged in a unique heterogeneous many-core architecture and memory hierarchy (Fig. 1). This processor contains 4 Core Groups (CGs) connected via a network on chip (NoC). Each CG consists of 1 management processing element (MPE) and 64 computing processing elements (CPEs) organized into an 8×8 CPE cluster (Fang et al., 2017). Both MPEs and CPEs are complete 64-bit Reduced Instruction Set Computing (RISC) cores working at a frequency of 1.45 GHz, but they adopt different memory hierarchies. MPE has a 32 KB L1 data cache, a 32 KB L1 instruction cache and a 256 KB L2 cache. CPE has a 16 KB L1 instruction cache and a 64KB local directive memory (LDM) as the user-controlled fast buffer. Users need to explicitly control data placement in the LDM. CPE clusters account for most of computing power and have access to the main memory of CGs indirectly through LDM, while MPE can directly access main memory and system interface for communication. In addition, a 128-bit memory controller is equipped with each CG to access the 8 GB double-Data-Rate Three Synchronous Dynamic Random Access Memory (DDR3) local memory with a theoretical peak bandwidth of 34 GB/s. Such design of processor enables relatively good power efficiency while maintaining high computing capacity.

The 40,000 SW26010 processors are connected using Peripheral Component Interconnect Express (PCI-E 3.0) connections in a customized Sunway Network. Peer-to-peer bidirectional communication between two processors via Message Passing Interface (MPI) is at 12 GB/s and a latency of about 1 us. Further technical detail is provided by (Fu et al., 2016). The supercomputer is used for a wide range of earth system modelling, including global atmosphere dynamics (Fu et al., 2017), earthquake simulation (https://awards.acm.org/bell), sea ice modelling (Li et al., 2017) and ocean cycle (Qiao et al., 2016). We choose this supercomputer for our large-scale agricultural prediction because of its unique architecture of hardware for earth system modelling and the abundant computing resources available for real-time simulation, as well as the potential expandability of our system coupling with other earth system modules (e.g. atmosphere model) on this platform.

3. Methodology

3.1 The main structure of the system

We develop an agricultural drought monitoring and yield prediction system using HPC technology on the Sunway TaihuLight supercomputer for the dynamic prediction of drought-induced agricultural losses. The framework of this system comprises three components: model optimization, dynamic prediction and result visualization (Fig. 2). The model optimization component provides the parameterization schemes for each model and the model weights for the ensemble. The dynamic prediction component collects climatic observations at daily time steps and monitors the development of precipitation deficits in areas with drought potential. Drought scenarios are used to evaluate the impacts of future drought development on yield formation. The climate data including the observations and future scenarios to the harvest time will be generated to drive the multi-model system for ensemble predictions. Finally, the model outputs are converted to the map of spatial distribution and risk assessment for drought mitigation. The following sections will explain each of these components in more detail.

3.2 Mapping multiple crop models onto the Sunway TaihuLight supercomputer

In this study, we use AquaCrop, SWAP and DNDC models for the prediction of crop growth. With water, radiation and nitrogen as the main driving element for each model, respectively (Huang et al., 2017), the multi-model ensemble prediction can be applicable in most cropland in China. For large-scale simulation, target regions are divided into a number of equal area rectangular grid cells or irregular grids based on the administrative boundary, assumed to be homogeneous in its environmental condition (e.g. soil, climate, farming practice) and independent from neighboring cells. Across the models, the crop growth modules and soil dynamic modules are usually the two main components with feedbacks between each other. In the growth modules, the above-ground canopy of crop is depicted as a whole unit to model the processes of radiation interception, photosynthesis and biomass partition into different organs, which have a straightforward influence on soil processes (e.g. soil evaporation and nutrient uptake). In the soil modules, a soil column is divided into multiple layers to simulate the vertical dynamic of water and nutrition, and it in return reveals the water stress and root length conditions for crop growth. Simulation run times differ across each model due to additional complexities. In DNDC, the nitrification and denitrification processes account for over 70% of the running time, as such simulation is performed in hour step for each soil layers (usually greater than 20 layers) involving multiform chemical processes. For SWAP, the Richards equation for the soil water flow is the dominant part of calculation because of the iteration process for the convergence of matrix calculation (Kroes et al., 2009). These computing modules can be speed up by proper parallelization schemes.

219

220 As shown in Fig. 3, we design our paralleled crop models based on the hardware
221 characteristics of the Sunway TaihuLight system in 3 steps:

222

223 (1) Similar to land models (e.g. Community Land Model, (Lawrence et al., 2011)), we
224 adopt a multiple-processing strategy to distribute computing tasks of a number of
225 grids on multiple CGs. When a process is launched on each CG to perform simulation
226 for each independent grid, the driving data (e.g. climate, management, parameter
227 values) is shared among different grids by communication with MPI. In this way,
228 computation for different grids inside the targeting region can be conducted
229 simultaneously.

230

231 (2) To exploit parallelism capacity inside CGs, we classify the scientific algorithms of
232 crop modeling into two categories: (i) MPE-based modules that feature light
233 computation and heavy communication and (ii) CPE-based modules that feature
234 heavy computation and light communication. CPE-based modules include the
235 independent calculation of state variable in each soil layer and the matrix calculation
236 that can be divided into small blocks. The detailed classifications of each model are
237 presented in Table 1.

238

239 (3) We apply a multiple-threading strategy to accelerate the CPE-based modules on
240 the CPE cluster. A light-weight thread can be launched on each CPE to process data
241 stored in its LDM in parallel of other CPEs of this cluster. Once the computation of
242 CPE-based module finished, we explicitly copy all the relevant variables from the
243 main memory to LDM. For instance, the values of variables (including water content,
244 temperature, etc.) in the ith soil layer at current step will get passed to the LDM of the
245 ith CPE for the calculation of nitrification process in DNDC, as well as the small
246 blocks of the water head matrix in water flow simulation in SWAP. Then different
247 CPEs use these data in LDM to calculate the substantial processes in parallel, and
248 return the final results to the main memory for the next round of computation. The
249 multi-layer acceleration is achieved by the AThread Library for C programming while
250 matrix calculation by the BLAS (Basic Linear Algebra Subprograms) package in the
251 Sunway TaihuLight supercomputer (Jiang et al., 2017).

252

253 Through the three steps above, we realize the parallelization and acceleration of our
254 models. The detailed acceleration steps using this MPE+CPEs mode are shown in Fig.
255 4. Our system can be scaled to 20,000 processes with MPI for global-level simulation.
256 It should be noted that from Table 1, only MPE is used for AquaCrop model due to
257 the proportion of CPE-based modules is relatively low. The use of CPEs may lead to
258 extra run time (time for data communication between main memory and LDM)
259 according to our tests.

260

261 **3.3 Model uncertainty**

262

263 Here, the optimization of model parameters for each individual model is based on the
264 maximization of the posterior probability density function (PDF) of model parameters.
265 According to Bayes' theorem, the posterior PDF can be expressed as:
266

267 $$p(\boldsymbol{\theta}\,|\,\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}\,|\,\boldsymbol{\theta}) * p(\boldsymbol{\theta})}{p(\boldsymbol{Y})} \qquad (1)$$

268

269 Where $\boldsymbol{\theta}$ is the vector of model parameters, $\boldsymbol{Y}$ is the series of observations and $p(\boldsymbol{\theta}$
270 $)$ is the parameter prior. To avoid further assumption of residual error in large-scale
271 simulation, the Jeffreys' uninformative prior (Box and Tiao, 2011) is applied to Eqn 1
272 and it leads to:
273

274 $$p(\boldsymbol{\theta}\,|\,\boldsymbol{Y}) \propto \left(\sum_{i=1}^{N} (Y_i - \tilde{Y}_i(\boldsymbol{\theta}))^2 \right)^{-\frac{N}{2}} \qquad (2)$$

275

276 Where N is the length of observations, and $\tilde{Y}$ is the model simulation. For Eqn 2,
277 we use the accelerated Markov Chain Monte Carlo (MCMC) algorithm 'Dream'
278 (Vrugt et al., 2009) to obtain the posterior distribution of parameters and the optimal
279 parameter sets for each crop model.
280

281 To address the uncertainty of model structure, we use BMA method (Huang et al.,
282 2017) to generate the multi-model ensemble from the individual model predictions.
283 For each ensemble member, it is assumed there is a conditional PDF $p(Y|\tilde{Y}^j)$ of the
284 yield $\boldsymbol{Y}$ with the jth model prediction $\tilde{Y}^j$ considered. Following (Raftery et al., 2005)
285 and the law of total probability, the BMA conditional PDF of the ensemble on all the
286 individual models can be written as:
287

288 $$p(\boldsymbol{Y}\,|\,\tilde{Y}^1,...,\tilde{Y}^M) = \sum_{j=1}^{M} w_j p(\boldsymbol{Y}\,|\,\tilde{Y}^j) \qquad (3)$$

289

290 Where M is the total number of our crop models, and w_j is the posterior probability

291 (weight) for the jth model with $\sum w_j = 1$. Further assuming the conditional PDF $p(Y|$

292 $\tilde{Y}^j)$ follows a Gaussian distribution provides:
293

294 $$\boldsymbol{Y}|\tilde{Y}^j \sim \boldsymbol{N}\big(a_j \tilde{Y}^j + b_j, \sigma_j^2\big) \qquad (4)$$

295

296   Where a_j and b_j are the coefficients of linear bias correction of $\tilde{Y}^j$. Then we can

297   obtain the value of w_j and σ_j by maximizing Eqn 3. The ensemble member $\tilde{Y}^j$

298   used here is the optimal prediction with the maximum posterior PDF in Eqn 2.

299

300   **3.4 Scenario based prediction of crop yields**

301

302   The climate forecast is one of the most important factors when crop models are used
303   for agricultural predictions. However, most seasonal weather forecasts still remain
304   uncertain for crop growth prediction while short-term forecast are insufficient to meet
305   this demands for its limited duration (Hansen et al., 2006). Addressing the uncertainty
306   of drought development, we use scenario analysis approach (Yu et al., 2014) to
307   estimate the potential interval of drought impacts on crop yields, so as to provide the
308   best and the worst outcomes to decision makers. For future drought development,
309   future climates are assumed to follow one of three scenarios:

310

311   (i) The drought ceases after the current day, and the water demand of crop growth is
312   fully satisfied until harvest. The maximum yield to date (or the unrecoverable yield
313   losses) can be derived in this scenario.

314

315   (ii) The drought continues in a user-specific period of time (e.g. 5, 10 or 20 days), but
316   will return to the ideal condition afterwards. The extra yield loss in this given period
317   of time can be derived from the modeling results.

318

319   (iii) The climate data after the current day will be replaced with representative climate
320   in history (e.g. the historical wet-, medium- or dry-year data) so that users can
321   compare the current drought impacts with that in historical series.

322

323   The crop models receive the latest real-time climatic observations at each time step
324   and update the outputs for the scenario analysis. Though extreme events are likely to
325   occur beyond the scope of historical record, the long historical series are still expected
326   to cover most possible conditions. This scenario-based assessment provides
327   comprehensive information about the potential trajectory of crop growth, based on
328   which mitigation management can be made combined with the expert's judgment.

329

330   **3.5 Risk analysis**

331

332   We use a copula model of the dependence structure between spatial locations to
333   compute the joint distribution of the relative production loss in multiple regions
334   (Vedenov, 2008; Gaupp et al., 2016). Sklar's theorem (Sklar, 1959) states the joint

335   distribution $F(x_1,...,x_n)$ of n-dimension random variables $\{X_1,...,X_n\}$ and the

marginal distribution $F_i(x_i)$ for each individual random variable X_i can be connected by the copula function as (Nelsen, 1999):

$$F(x_1,...,x_n) = C[F_1(x_1),...,F_n(x_n)] \qquad (5)$$

Where $C(\cdot)$ is the copula function with $C = [0,1]^n \rightarrow [0,1]$. When $F(\cdot)$ and $C(\cdot)$ are both differentiable, the joint PDF can be written as:

$$f(x_1,...,x_n) = c[F_1(x_1),...,F_n(x_n)] \cdot f_1(x_1)... \cdot f_n(x_n) \qquad (6)$$

Where $c(\cdot)$ is the PDF of the copula function as:

$$c(\mu_1,...,\mu_n) = \frac{\partial^n}{\partial\mu_1...\partial\mu_n} C(\mu_1,...,\mu_n) \qquad (7)$$

The advantage of using copula method is that users can only consider the marginal distribution of each random variable with fixed copula type. In our system, we employ a Gumbel copula method and pairwise ordered coupling as the copula structure (Gaupp et al., 2016). Users are able to choose corresponding predictions of yield loss in the regions of interest (e.g. multiple counties or provinces) for the analysis of joint distribution.

4. System testing and applications

4.1 Computational performance

To demonstrate the performance of model acceleration with SW26010 on this platform, we perform a 10-year simulation of maize yield in Changtu county with MPE-only and MPE-CPEs hybrid modes. The speedup ratio of CPE-based module is over 8, as shown in Table 2.

Then we randomly selected 512 counties in China for county-level modeling. For each model, the simulation period is 1996-2007, with corresponding run times for MPE-only and MPE-CPEs hybrid modes shown in Fig. 5, for CGs increasing from 8 to 512. In this experiment, we ran 10 iterations for each model and summated the run times to reduce the random error of run time measurement.

In Fig. 5, the run time for each trial decreases with the increase of CGs numbers. However, the actual accelerating efficiency is lower than the ideal speedup with the increasing CGs numbers, because the costs of both communication between different processes and I/O (input/output) processes will expanded rapidly when more CGs are used simultaneously. In general, by using CPEs, the total run time of DNDC model is

accelerated by about 3.6 times while SWAP by 2.6 times. The overall performance is lower than the CPE-based part because of the investable part of MPE-based module, as well as the extra cost for the communication between MPE and CPEs (e.g. copying data from main memory to LDM). The difference in model algorithms (e.g. soil water simulation), the proportions of MPE-based module to the whole computing part and the detailed model setting (e.g. number of soil layers, convergence condition) is believed to contribute to the different efficiencies of speedup for these models. The overall reduction in computation time demonstrates the potential for near real-time predictions across large scales using the Sunway TaihuLight supercomputer resource.

4.2 Parameter optimization and multi-model ensemble for China

We simulate the annual yields of wheat, rice and maize for all 2403 counties in China. The soil properties (including clay content, hydraulic conductivity, maximum soil depth, etc) are from the database "The Soil Database of China for Land Surface Modeling" (Shangguan et al., 2014). The climatic observation from weather station (http://www.cma.gov.cn/2011qxfw/2011qsjgx/), county-level yield records and averaged fertilizer amount from 1998-2010 statistical yearbooks (http://navi.cnki.net/KNavi/Yearbook.html) are used for parameter optimization and multi-model ensemble based on the methods in section 3. Only parameters related to crop growth are selected for optimization and they are set in accordance with Yu's literature (Yu et al., 2018). The period 1998-2007 is used for calibration and 2008-2010 for validation.

In this case, 6000 CGs are applied collectively to generate 4000 samples for each crop in each county, with MCMC simulations for the three crop models taking approximately 18 hours. We extract the predictions with maximum posterior PDF in each county and display the simulation performance in Fig. 6. And the parameter sets obtained here can be used for the real-time prediction afterwards.

Each individual model is shown to perform well at the national scale largely matching observations. However, for all the ensemble members, there are still significantly overestimated or underestimated results, especially in validation period. This implies great uncertainty of model structure for agricultural simulation and the limited skill of a single model for large-scale prediction. In comparison to individual models, the multi-model ensemble prediction in Fig. 6 shows greater accuracy in both calibration and validation periods with a higher coefficient-of-determination (R^2) value and lower root-mean-square error (RMSE) value. For example, the ensemble simulation of wheat is obviously better than any individual model. The over- and under-estimations of yield prediction decrease and the residual errors appears to have a normal distribution with smaller variance. These observations indicate that the Bayesian inference and BMA methods in this system effectively reduce the predictive uncertainty of crop growth on the basis of current crop models.

4.3 Dynamic drought impact assessment in North and Northeast China

Between April and July 2017, a severe drought occurred in North and Northeast China, including Jilin, Liaoning, Heilongjiang, Inner Mongolia, Shandong and Hebei provinces (Fig. 7). Combined, these regions produce over 60% of China's annual maize production, while any drought during this period is likely to cause significant decrease to the maize production later in the year. Collating daily climate data from June 1st, we monitored the current production losses and predicted the potential evolution of agricultural drought based on our three drought scenarios (Section 3.4).

In Fig. 8, the average yield for each scenario is predicted for July 5th and July 30th. On July 5th (Fig. 8(a)), given the study region has been in meteorological drought for over 2 months, the predictions in most scenarios are below the historic medium-level records. Under the scenario where drought lasts less than 10 days, the final yield loss varies within 10% of the historical average. Theoretically, if future precipitation is sufficient, the final yield could even be slightly higher than the historic average. However, as maize growth just proceeds into the reproductive stage at this time of year, the uncertainty among these scenario predictions remains large due to the sensitivity of crop growth to subsequent weather conditions. On July 30th, the phenology of maize growth is closer to maturity than 25 days ago (July 5th), and the rate of photosynthetic assimilation typically decreases since this point. As a result, the range of yield of different scenarios in Fig. 8(b) converges to smaller range compared with Fig. 8(a). In the whole month in July, the drought condition still existed in many counties of this region, and it is inevitable that the average yield in 2017 will drop below the historical mean value, even with sufficient precipitation thereafter.

The distributions of up-to-date yield losses from June 15st to July 15th in about 10-day step are displayed in Fig. 9. The model simulations indicate the yield at any time before early June could still be recovered to the average level if the drought condition ends then. However, the sustained drought condition in late June and July was critical to the final grain formation, and the water stress in this period led to the irreversible damage to maize growth. This knowledge can be used by the local department of agriculture to advise mitigation actions to minimize the impacts of the drought. Fig. 9 further shows the agricultural loss in generally more severe in Northeast China compared to North China. We expect this is due to different meteorological conditions and irrigation levels.

4.4 The probability of yield losses

The marginal distributions of drought-induced yield loss can be estimated through the generalized extreme value (GEV) distribution and the joint distribution of regional yield losses by Gumbel copulas. The long-term yield simulations from 1960 to 2010 in each county are performed using the optimal model parameters in section 4.2, based on which the related parameter values of probability distributions are estimated

by maximum likelihood method. We calculate the return periods (the inverse of cumulative probability) for county and province levels in these regions in Fig. 10(a) and the joint cumulative probability for North and Northeast China in Fig. 10(b) based on the scenario-analysis results on July 30th.

Fig. 10(a) shows that the drought caused the most severe damage to agricultural production in Inner Mongolia, with the return periods over 50 and 100 years in many counties as well as the province level. The magnitude of yield loss seems a bit overestimated as the irrigation level in our database is actually lower than the current condition in this region. But generally this result is consistent with the meteorological drought in Fig. 7, because precipitation deficit in 2017 in this province shows a higher spatial and temporal constancy than other provinces. It is important to note that some provinces or counties with relatively low irrigation levels (e.g. Jilin and Heilongjiang provinces) have lower return period values than those with higher irrigation levels (e.g. Hebei and Shandong provinces). It is mainly because the lack of irrigation is more likely to cause the variability in historical yield simulation more frequent and severe, which makes the yield loss in 2017 less sensitive. The difference between yield loss and return period consequently implies the advantage of risk analysis for drought assessment in longer time series than then absolute value.

Fig. 10(b) clearly demonstrates the trajectory of joint probability in different scenarios. For decision makers concerned with food security, instead of treating those regions independently, we suggest using information on the joint condition of different production areas as the increase of production in some regions can compensate for the loss of others. In this case, for example, the marginal cumulative probabilities of current yield loss on July 30th are 0.1585 and 0.1393 for North and Northeast China respectively, while the joint cumulative probability is about 0.0311 (approx. 32-year return period). Compared with the joint distribution, the product (independent assumption) of the marginal probability 0.0221 (approx. 45-year return period) seems to overestimate the yield loss significantly. As the overall meteorological conditions for crop growth in North and Northeast China are similar due to geographic proximity, the spatial dependency of yield loss cannot be ignored for risk analysis, and our system provides an objective evaluation by Copula method.

5. Concluding remarks

In this study, we have developed an integrated agricultural risk system for daily agriculture dynamics on the Sunway TaihuLight supercomputer, including the parallelization and acceleration for existing crop models, model optimization and multi-model ensemble for yield prediction and multi-scale analysis of yield loss risk.

While traditional crop models are relatively computationally inefficient, through the detailed classification of model algorithms, we have observed an approximate 3x decrease in run times using the SW26010 processor. With over 40,000 processors

509 available in our platform, we can conduct parameter uncertainty analyses and multi-
510 model ensemble based on the MCMC method to improve the accuracy and precision
511 of our model predictions. We have demonstrated the feasibility of using our approach
512 for monitoring the effect of drought evolution on agricultural production across large
513 scales through application to the May-July 2017 drought in North and Northeast
514 China. The combination of yield losses and corresponding return periods at different
515 scales is able to provide objective and useful evaluation of loss magnitude for the
516 public and the local decision makers (e.g. local Departments of Agriculture) while the
517 probability of joint distribution is helpful for higher-level decision makers such as the
518 Ministry of Agriculture to coordinate mitigation or response actions in different
519 regions.
520
521 To reduce the cost of I/O processes, which account for a considerable share of the
522 total run time (especially when the number of processes increases), future
523 improvements can be made by optimizing the format of I/O files (e.g. mergence of
524 files in small size, binary format) of our crop models and the introduction of new
525 techniques for the parallel I/O (e.g. PnetCDF) into our system. The quality and
526 temporal frequency of input regional data during drought conditions is critical for the
527 accuracy of the dynamic prediction. Such real-world information is not limited to
528 replanting schedules, and changes to crop species, irrigation levels, and planting times.
529 Potential techniques for collecting such information need to be explored, such as
530 using online big data collection techniques to continually update the database. Data
531 assimilation using satellite observations should also be adopted to improve predictive
532 accuracy.
533
534

535 Table 1

536 The classification of MPE-module and CPE-module for each model

537

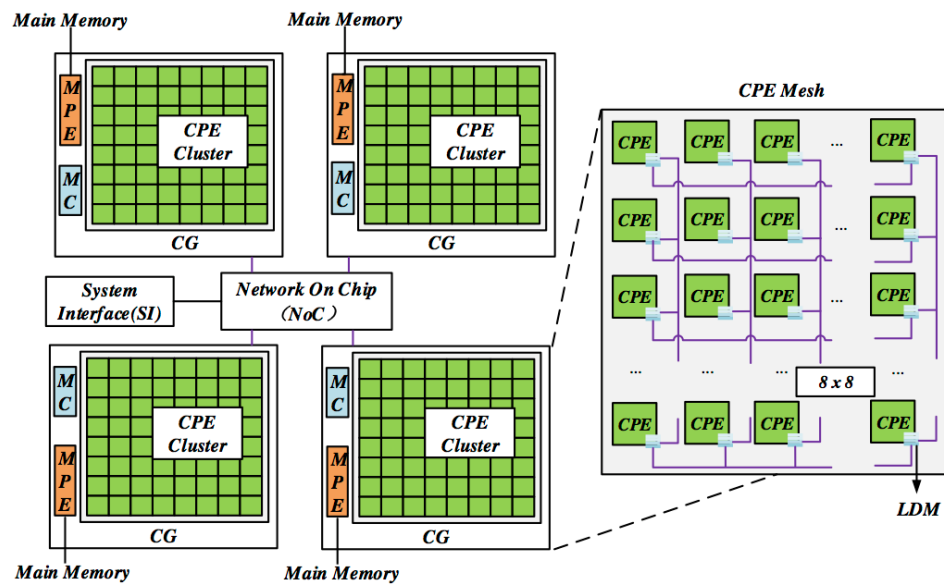| Model	MPE-based module	CPE-base module
SWAP	crop growth ET and water uptake	soil water flow soil temperature conductance
DNDC	crop growth soil water flow soil temperature conductance ET and water uptake	nitrification and denitrification decomposition
AquaCrop	All	None

538

539

540   Table 2

541   The proportions of MPE-based module and CPE-base module to the total run time and

542   the speedup ratio for CPE-module

Model	MPE-based module	CPE-base module in MPE mode	CPE-base module in MPE+CPEs mode	Speedup ratio
SWAP	20%	80%	8%	8
DNDC	15%	85%	10%	8.5

543

544

545

546
547 Figure 1. The architecture of the SW26010 processor, modified from (Jiang et al.,
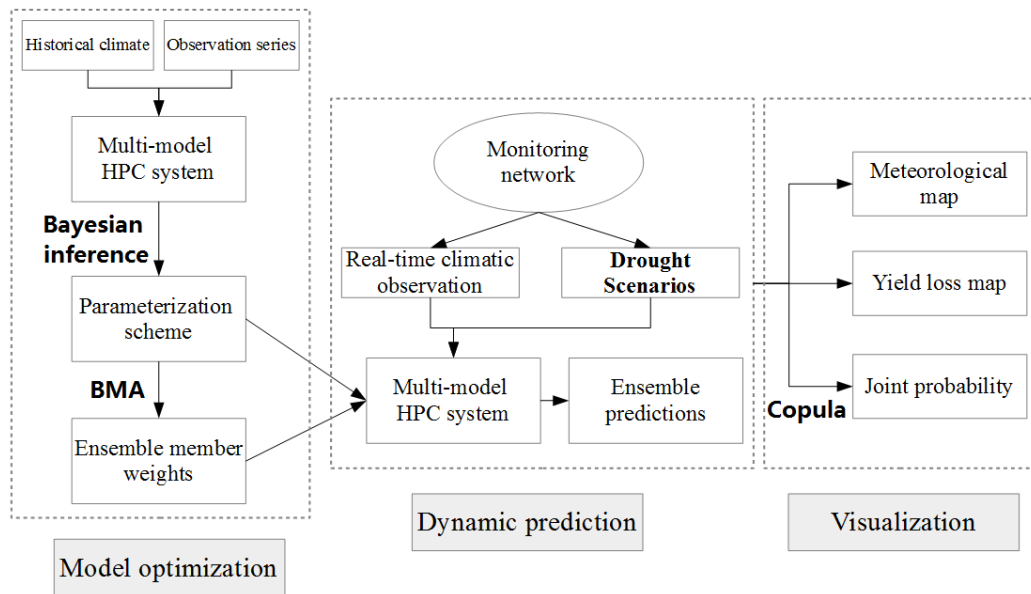548 2017)
549
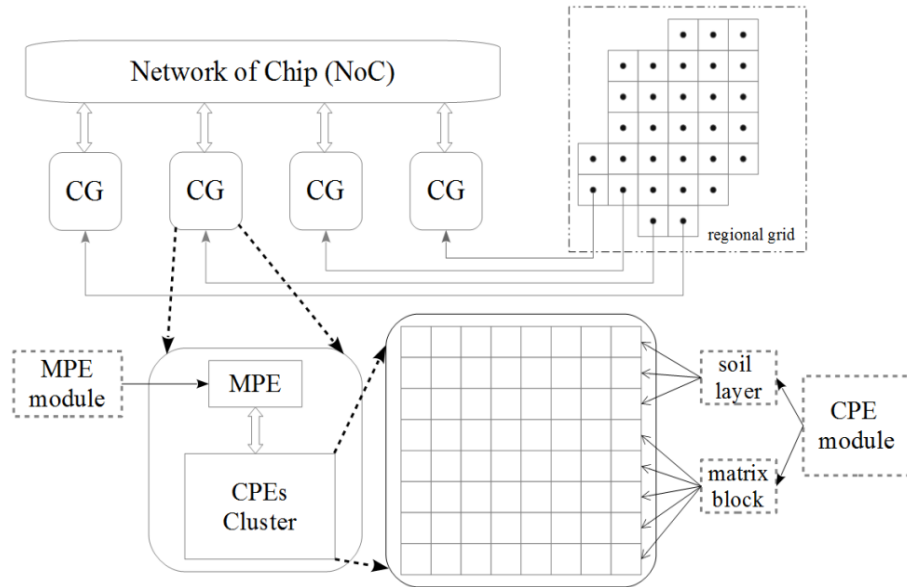
Figure 2. The main structure of the system

550

551

552

553

Network of Chip (NoC)

CG CG CG CG

regional grid

MPE module

MPE

CPEs Cluster

soil layer

matrix block

CPE module

554

555   Figure 3. Parallel strategy for CPE-based module and MPE-based module
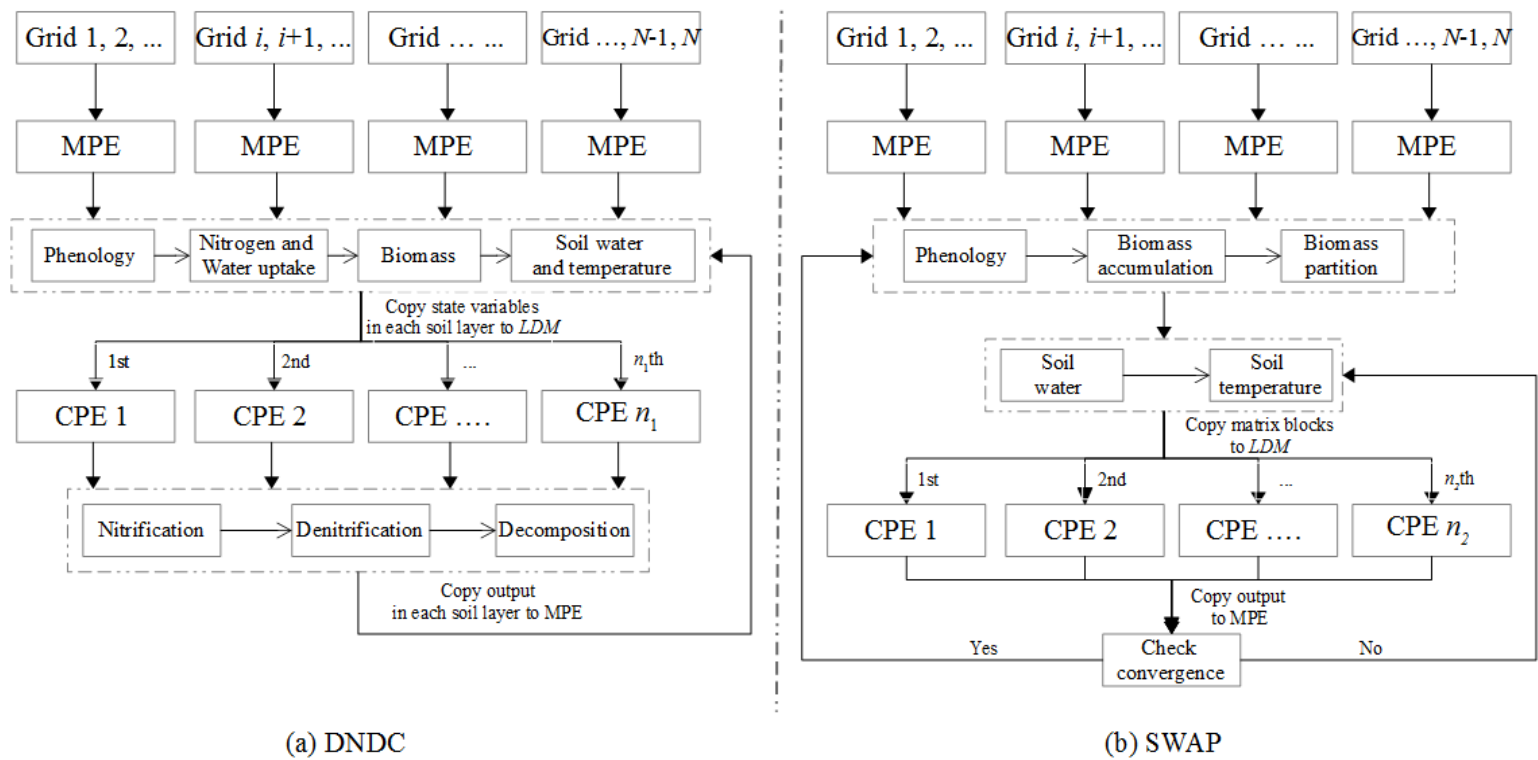
(a) DNDC

(b) SWAP

556

557 Figure 4. The detailed acceleration steps using MPE+CPEs mode for (a) DNDC model and (b) SWAP model
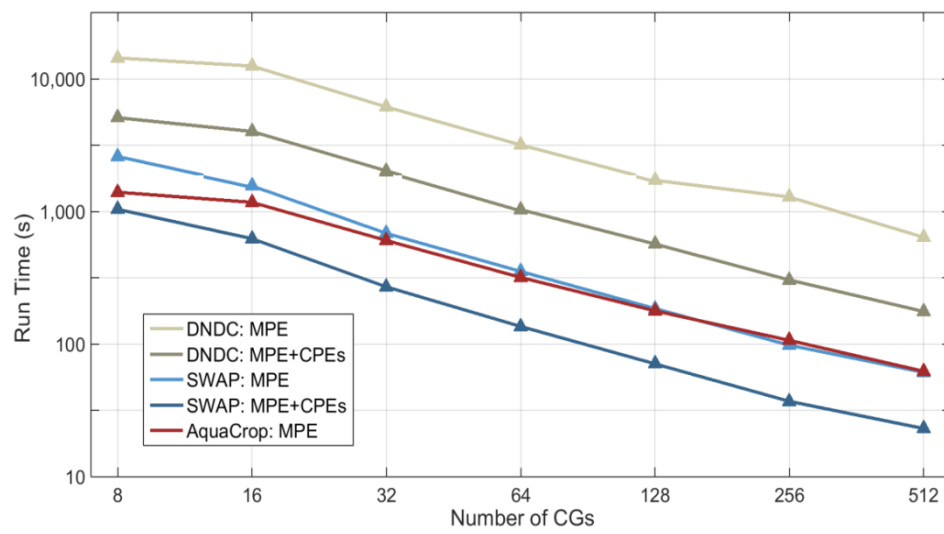
558

559

560

Figure 5. The run time (10 iterations) of three models for 512 counties
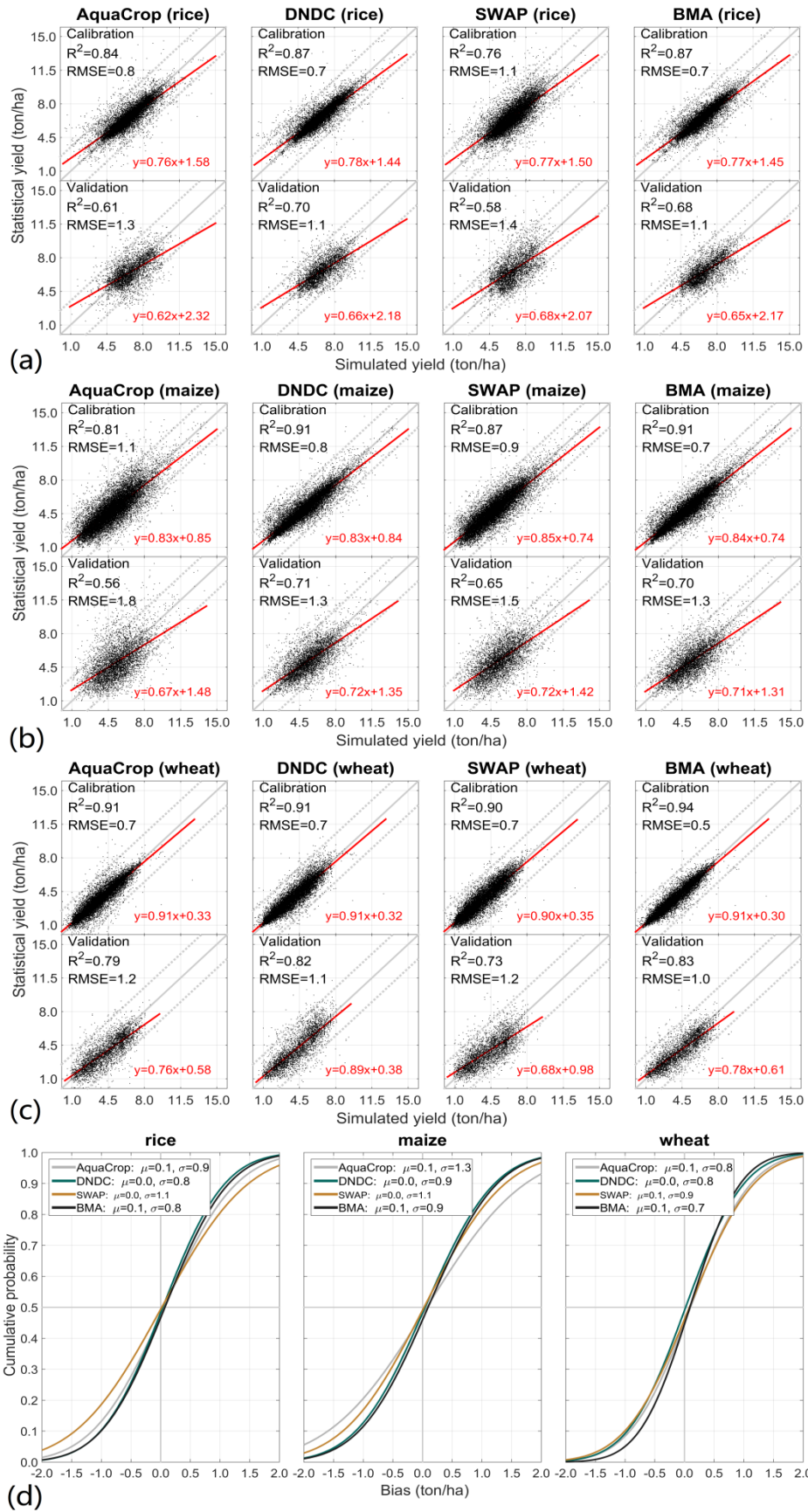in MPE-only and MPE+CPEs modes

561
562
563

564
565 Figure 6. (a)(b)(c) statistical yield observation(X axis, t/ha) VS model yield
566 simulation(Y axis, t/ha); (d) The cumulative distribution of residual error;
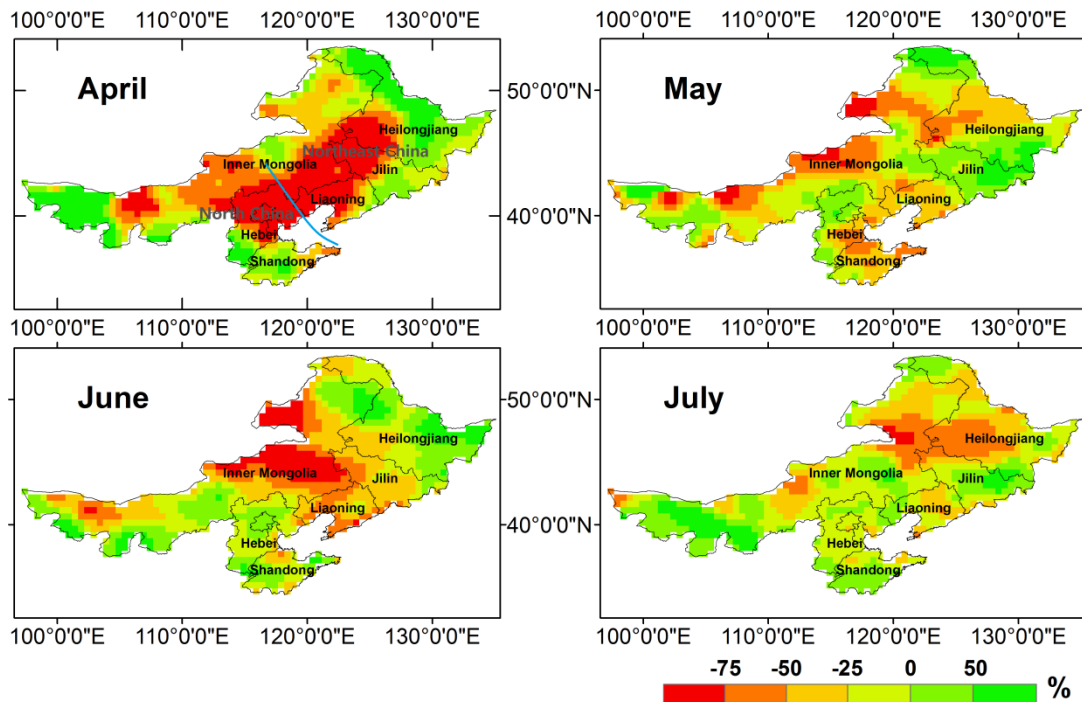
567

568 Figure 7. The monthly precipitation anomaly in North and Northeast China from April
569 to July in 2017 (departure from the average values of 2000-2016). Data source: CPC
570   0.5°×0.5° Global Daily Gauge-Based Analysis of Precipitation,
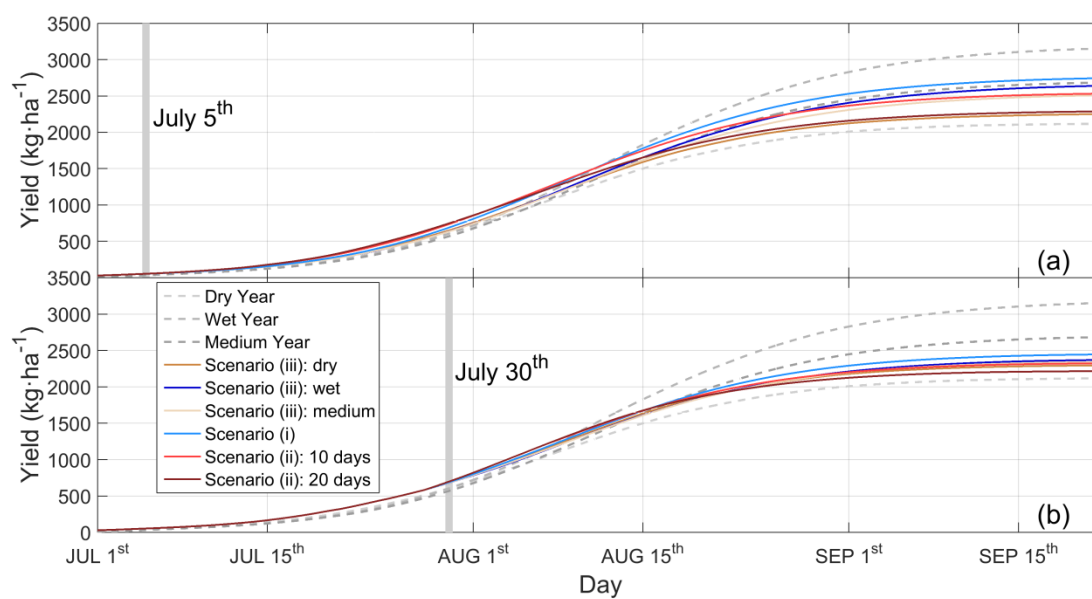571   https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globalprecip.html

572

Figure 8. The average yield predictions using different scenarios in this whole region: (a) real-time observation updated to July 5th; (b) real-time observation updated to July 30th
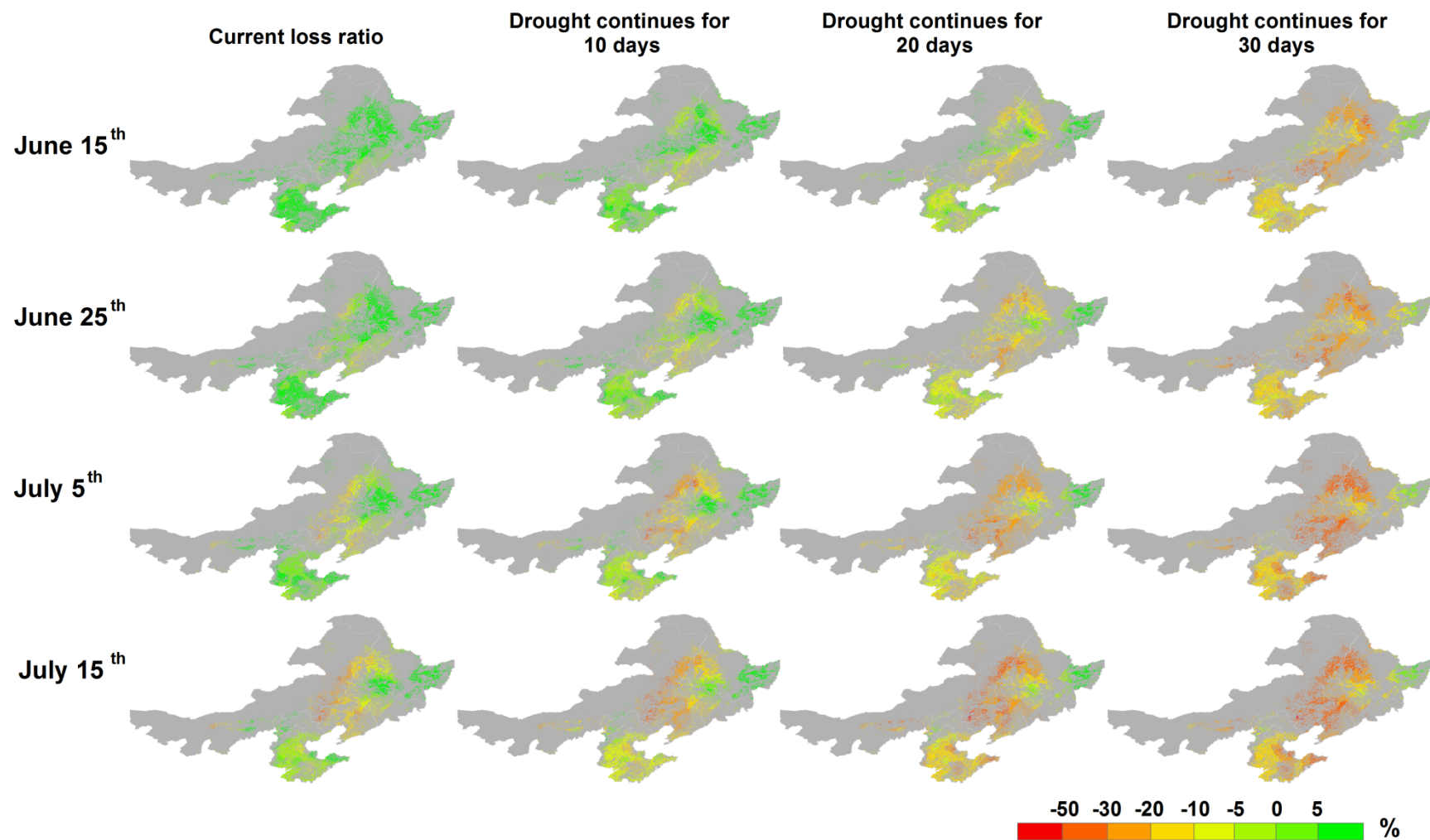
Figure 9. Distributions of the expected maize yield losses (the modeled yield departure from the average county-level values of 2000-2016

579

580

581

(a)

(b)

Figure 10. (a) The return period of yield loss on July 30th in county and province levels; (b) The cumulative probability (contour line) of joint distribution of yield losses in both North and Northeast China

582

583

584

References:

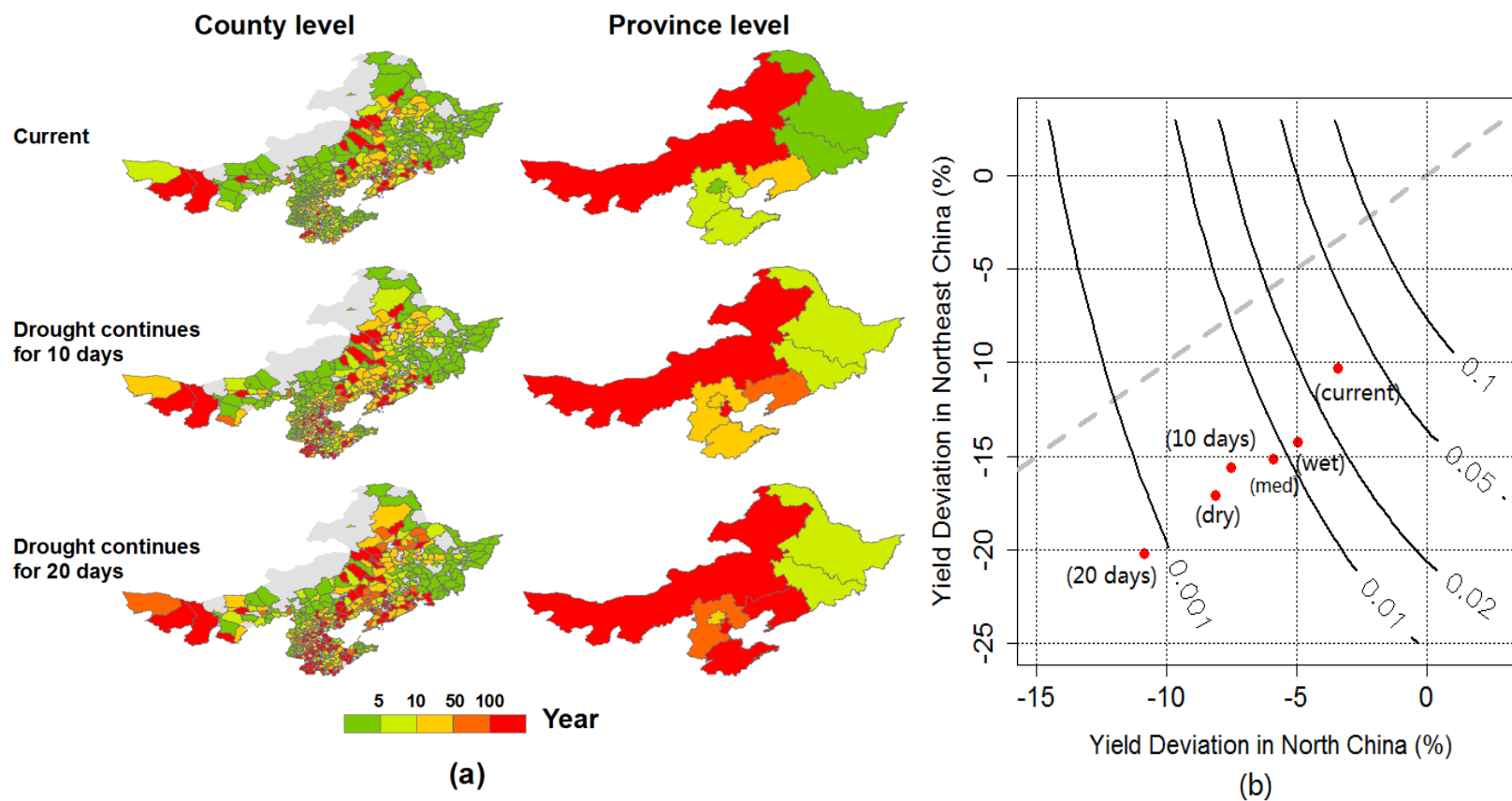[1] Adams, R.M., Hurd, B.H., Lenhart, S., Leary, N., 1998. Effects of global climate change on agriculture: an interpretative review. CLIM RES 11, 19-30.

[2] AghaKouchak, A., Bárdossy, A., Habib, E., 2010. Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. ADV WATER RESOUR 33, 624-634.

[3] Báez-González, A.D., Chen, P., Tiscareño-López, M., Srinivasan, R., 2002. Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. CROP SCI 42, 1943-1949.

[4] Baigorria, G.A., Jones, J.W., O Brien, J.J., 2008. Potential predictability of crop yield using an ensemble climate forecast by a regional circulation model. AGR FOREST METEOROL 148, 1353-1361.

[5] Bárdossy, A., Pegram, G., 2009. Copula based multisite model for daily precipitation simulation. HYDROL EARTH SYST SC 13, 2299.

[6] Bloom, D.E., 2011. 7 billion and counting. SCIENCE 333, 562-569.

[7] Bonaccorso, B., Cancelliere, A., Rossi, G., 2003. An analytical formulation of return period of drought severity. STOCH ENV RES RISK A 17, 157-174.

[8] Box, G.E., Tiao, G.C., 2011. Bayesian inference in statistical analysis. John Wiley & Sons.

[9] De Wit, A.D., Van Diepen, C.A., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. AGR FOREST METEOROL 146, 38-56.

[10] Dongarra, J.J., Luszczek, P., Petitet, A., 2003. The LINPACK benchmark: past, present and future. Concurrency and Computation: practice and experience 15, 803-820.

[11] Dumont, B., Leemans, V., Mansouri, M., Bodson, B., Destain, J., Destain, M., 2014. Parameter identification of the STICS crop model, using an accelerated formal MCMC approach. ENVIRON MODELL SOFTW 52, 121-135.

[12] Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., Foster, I., 2014. The parallel system for integrating impact models and sectors (pSIMS). ENVIRON MODELL SOFTW 62, 509-516.

[13] Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., van Ittersum, M.K., Janssen, S., Rivington, M., 2015. Crop modelling for integrated assessment of risk to food production from climate change. ENVIRON MODELL SOFTW 72, 287-303.

[14] Fang, J., Fu, H., Zhao, W., Chen, B., Zheng, W., Yang, G., 2017. swDNN: A Library for Accelerating Deep Learning Applications on Sunway TaihuLight., 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 615-624.

[15] Fernández, B., Salas, J.D., 1999. Return period and risk of hydrologic events. I: mathematical formulation. J HYDROL ENG 4, 297-307.

[16] Field, C.B., 2012. Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. Cambridge University Press.

[17] Fu, H., Liao, J., Ding, N., Duan, X., Gan, L., Liang, Y., Wang, X., Yang, J., Zheng, Y., Liu, W., 2017. Redesigning CAM-SE for peta-scale climate modeling performance and ultra-high resolution on Sunway TaihuLight., Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. ACM, p. 1.

[18] Fu, H., Liao, J., Yang, J., Wang, L., Song, Z., Huang, X., Yang, C., Xue, W., Liu, F., Qiao, F.,

628   2016. The Sunway TaihuLight supercomputer: system and applications. Science China
629   Information Sciences 59, 72001.

630   [19] Gaupp, F., Pflug, G., Hochrainer Stigler, S., Hall, J., Dadson, S., 2016. Dependency of crop
631   production between global breadbaskets: a copula approach for the assessment of global and
632   regional risk pools. RISK ANAL.

633   [20] Guo, Y., Ma, Y., Zhan, Z., Li, B., Dingkuhn, M., Luquet, D., De Reffye, P., 2006. Parameter
634   optimization and field validation of the functional–structural model GREENLAB for maize. ANN
635   BOT-LONDON 97, 217-230.

636   [21] Hansen, J.W., Challinor, A., Ines, A., Wheeler, T., Moron, V., 2006. Translating climate forecasts
637   into agricultural terms: advances and challenges. CLIM RES 33, 27-41.

638   [22] Heng, L.K., Hsiao, T., Evett, S., Howell, T., Steduto, P., 2009. Validating the FAO AquaCrop
639   model for irrigated and water deficient field maize. AGRON J 101, 488-498.

640   [23] Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G.,
641   White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: current
642   status and future prospects. ENVIRON MODELL SOFTW 72, 276-286.

643   [24] Howden, S.M., Soussana, J., Tubiello, F.N., Chhetri, N., Dunlop, M., Meinke, H., 2007. Adapting
644   agriculture to climate change. Proceedings of the national academy of sciences 104, 19691-19696.

645   [25] Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., Wu, W., 2015. Jointly assimilating
646   MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. IEEE J-
647   STARS 8, 4060-4071.

648   [26] Huang, X., Huang, G., Yu, C., Ni, S., Yu, L., 2017. A multiple crop model ensemble for
649   improving broad-scale yield prediction using Bayesian model averaging. FIELD CROP RES 211,
650   114-124.

651   [27] Iizumi, T., Yokozawa, M., Nishimori, M., 2009. Parameter estimation and uncertainty analysis of
652   a large-scale crop model for paddy rice: Application of a Bayesian approach. AGR FOREST
653   METEOROL 149, 333-348.

654   [28] Jakku, E., Thorburn, P.J., 2010. A conceptual framework for guiding the participatory
655   development of agricultural decision support systems. AGR SYST 103, 675-682.

656   [29] Jiang, L., Yang, C., Ao, Y., Yin, W., Ma, W., Sun, Q., Liu, F., Lin, R., Zhang, P., 2017. Towards
657   Highly Efficient DGEMM on the Emerging SW26010 Many-Core Processor., Parallel Processing
658   (ICPP), 2017 46th International Conference on. IEEE, pp. 422-431.

659   [30] Kroes, J.G., Van Dam, J.C., Groenendijk, P., Hendriks, R., Jacobs, C., 2009. SWAP version 3.2.
660   Theory description and user manual. Alterra.

661   [31] Kroes, J.G., Wesseling, J.G., Van Dam, J.C., 2000. Integrated modelling of the soil–water–
662   atmosphere–plant system using the model SWAP 2· 0 an overview of theory and an application.
663   HYDROL PROCESS 14, 1993-2002.

664   [32] Launay, M., Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve
665   predictive performance for spatial applications. Agriculture, ecosystems & environment 111, 321-
666   339.

667   [33] Lawrence, D.M., Oleson, K.W., Flanner, M.G., Thornton, P.E., Swenson, S.C., Lawrence, P.J.,
668   Zeng, X., Yang, Z.L., Levis, S., Sakaguchi, K., 2011. Parameterization improvements and
669   functional and structural advances in version 4 of the Community Land Model. J ADV MODEL
670   EARTH SY 3.

671   [34] Li, B., Li, B., Qian, D., 2017. PFSI. sw: A programming framework for sea ice model algorithms

672   based on Sunway many-core processor., Application-specific Systems, Architectures and

673   Processors (ASAP), 2017 IEEE 28th International Conference on. IEEE, pp. 119-126.

674 [35] Li, C., Frolking, S., Frolking, T.A., 1992. A model of nitrous oxide evolution from soil driven by

675   rainfall events: 1. Model structure and sensitivity. Journal of Geophysical Research: Atmospheres

676   97, 9759-9776.

677 [36] Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008.

678   Prioritizing climate change adaptation needs for food security in 2030. SCIENCE 319, 607-610.

679 [37] Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C.,

680   Thorburn, P.J., Cammarano, D., 2015. Multimodel ensembles of wheat growth: many models are

681   better than one. GLOBAL CHANGE BIOL 21, 911-925.

682 [38] Neale, R.B., Chen, C., Gettelman, A., Lauritzen, P.H., Park, S., Williamson, D.L., Conley, A.J.,

683   Garcia, R., Kinnison, D., Lamarque, J., 2010. Description of the NCAR community atmosphere

684   model (CAM 5.0). NCAR Tech. Note NCAR/TN-486+ STR.

685 [39] Nelsen, R.B., 1999. Introduction., An Introduction to Copulas. Springer, pp. 1-4.

686 [40] Oleson, K.W., Lawrence, D.M., Gordon, B., Flanner, M.G., Kluzek, E., Peter, J., Levis, S.,

687   Swenson, S.C., Thornton, E., Feddema, J., 2010. Technical description of version 4.0 of the

688   Community Land Model (CLM).

689 [41] Qiao, F., Zhao, W., Yin, X., Huang, X., Liu, X., Shu, Q., Wang, G., Song, Z., Li, X., Liu, H.,

690   2016. A highly effective global surface wave numerical simulation with ultra-high resolution.,

691   High Performance Computing, Networking, Storage and Analysis, SC16: International

692   Conference for. IEEE, pp. 46-56.

693 [42] Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model

694   averaging to calibrate forecast ensembles. MON WEATHER REV 133, 1155-1174.

695 [43] Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth,

696   C., Glotter, M., Khabarov, N., 2014. Assessing agricultural risks of climate change in the 21st

697   century in a global gridded crop model intercomparison. Proceedings of the National Academy of

698   Sciences 111, 3268-3273.

699 [44] Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M.,

700   Nelson, G.C., Porter, C., Janssen, S., 2013. The agricultural model intercomparison and

701   improvement project (AgMIP): protocols and pilot studies. AGR FOREST METEOROL 170,

702   166-182.

703 [45] Rosenzweig, C., Tubiello, F.N., 2007. Adaptation and mitigation strategies in agriculture: an

704   analysis of potential synergies. MITIG ADAPT STRAT GL 12, 855-873.

705 [46] Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth system

706   modeling. J ADV MODEL EARTH SY 6, 249-263.

707 [47] Skakun, S., Kussul, N., Shelestov, A., Kussul, O., 2016. The use of satellite data for agriculture

708   drought risk quantification in Ukraine. Geomatics, Natural Hazards and Risk 7, 901-917.

709 [48] Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ.

710   Paris 8, 229-231.

711 [49] Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to

712   simulate yield response to water: I. Concepts and underlying principles. AGRON J 101, 426-437.

713 [50] Tao, F., Yokozawa, M., Zhang, Z., 2009. Modelling the impacts of weather and climate variability

714   on crop productivity over a large area: a new process-based model development, optimization, and

715   uncertainties analysis. AGR FOREST METEOROL 149, 831-850.

716 [51] Van Ittersum, M.K., Ewert, F., Heckelei, T., Wery, J., Olsson, J.A., Andersen, E., Bezlepkina, I.,
717 Brouwer, F., Donatelli, M., Flichman, G., 2008. Integrated assessment of agricultural systems–A
718 component-based framework for the European Union (SEAMLESS). AGR SYST 96, 150-165.
719 [52] Vedenov, D., 2008. Application of copulas to estimation of joint crop yield distributions.,
720 American Agricultural Economics Association Annual Meeting, Orlando, FL, pp. 27-29.
721 [53] Vital, J., Gaurut, M., Lardy, R., Viovy, N., Soussana, J., Bellocchi, G., Martin, R., 2013. High-
722 performance computing for climate change impact studies with the Pasture Simulation model.
723 COMPUT ELECTRON AGR 98, 131-135.
724 [54] Vrugt, J.A., Ter Braak, C., Diks, C., Robinson, B.A., Hyman, J.M., Higdon, D., 2009.
725 Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive
726 randomized subspace sampling. INT J NONLIN SCI NUM 10, 273-290.
727 [55] Yu, C., Huang, X., Chen, H., Huang, G., Ni, S., Wright, J.S., Hall, J., Ciais, P., Zhang, J., Xiao,
728 Y., Sun, Z., Wang, X., Yu, L., 2018. Assessing the Impacts of Extreme Agricultural Droughts in
729 China Under Climate and Socioeconomic Changes. Earth's Future.
730 [56] Yu, C., Li, C., Xin, Q., Chen, H., Zhang, J., Zhang, F., Li, X., Clinton, N., Huang, X., Yue, Y.,
731 2014. Dynamic assessment of the impact of drought on agricultural yield and scale-dependent
732 return periods over large geographic regions. ENVIRON MODELL SOFTW 62, 454-464.
733 [57] Zhao, G., Bryan, B.A., King, D., Luo, Z., Wang, E., Bende-Michl, U., Song, X., Yu, Q., 2013.
734 Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining
735 grid computing and parallel processing. ENVIRON MODELL SOFTW 41, 231-238.
736 [58] Zhao, W., Fu, H., Fang, J., Zheng, W., Gan, L., Yang, G., 2018. Optimizing Convolutional Neural
737 Networks on the Sunway TaihuLight Supercomputer. ACM Trans. Archit. Code Optim. 15, 1-26.
738
739