Published version can be accessed at https://doi.org/10.1016/j.envsoft.2020.104807

# GDNDC: An integrated system to model water-nitrogen-crop processes for agricultural management at regional scales

Xiao Huang[a], Shaoqiang Ni[b], Chao Wu[c], Conrad Zorn[d,e],

Wenyuan Zhang[f], Chaoqing Yu[b,g,*]

[a] *Norwegian Institute of Bioeconomy Research, Saerheim, Norway*

[b] *Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China*

[c] *Department of Ecology and Evolutionary Biology, Yale University, New Haven, USA*

[d] *Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand*

[e] *Environmental Change Institute, University of Oxford, Oxford, UK*

[f] *Department of Zoology, University of Oxford, Oxford, UK*

[g] *AI for Earth Laboratory, Cross-strait Tsinghua Research Institute, Xiamen, China*

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

**Abstract:** Agroecosystem modelling has increasingly focused on the integration of soil biogeochemical processes and crop growth. However, few models are available that offer high computing efficiencies for region-scale simulations, integrated decision support tools, and a structure that allows for easy extension. This paper introduces a new modeling tool to fill this gap: the GDNDC (Gridded DNDC) system for gridded agro-biogeochemical simulations. Based on the established DeNitrification and DeComposition (DNDC) model version-95, its main advancements include (i) implementation of parallel computation to significantly reduce computation time across multiple scales; (ii) a built-in parameter optimization algorithm to improve the predictive accuracy, and (iii) several decision support tools. We demonstrate each of these for county-level maize growth simulations in Liaoning Province (China) and reveal the potential of this new modeling tool to guide both long-term policy decisions regarding optimal fertilizer application and near-term crop yield forecasting for reactive decisions required in times of drought.

**Keywords:** GDNDC; parallel computation; parameter optimization; optimal fertilization; decision support;

## 1. Introduction

In past decades, the expansion of irrigation area and fertilizer use for agriculture has significantly improved global food production especially under drought and nutrient depleted conditions (Schultz et al., 2005; Stewart et al., 2005; Yu et al., 2018). More food has to be produced sustainably to meet the demand of growing population by the middle of this century (Godfray et al., 2010). However, surplus nutrients from cropland, including nitrogen (N) and phosphorous (P), have led to severe environmental problems in both the hydrosphere and atmosphere (Cordell et al., 2009; Yu et al., 2019). For example, the loadings of N and P from cropland into surrounding water systems (rivers, lakes and coastal ocean) can result in eutrophication (Paerl et al., 2011). In addition, greenhouse gas (GHG) emissions from agriculture, such as nitrous oxide ($N_2O$) and methane ($CH_4$) gas emissions from rice cultivation, can contribute to global climate change (Cai et al., 1997). On top of excessive inputs into the surrounding environment, agriculture can also detrimentally remove resources from the surrounding environment. Excessive extraction of water for agricultural irrigation has been observed to contribute to groundwater depletion in some regions (e.g. the North China Plain and Northern India) (Famiglietti, 2014). It is therefore of great importance to improve our fertilization practices and irrigation management to minimize environmental impacts while maintaining food production for the population growth (Tilman, 1999).

Field experiments provide important information about the relationship between crop growth and environmental factors (e.g. climate and soil properties). Experiments which investigate various management interventions (e.g. fertilization, irrigation and tillage) at different phenological stages can test the response of crop development and evaluate the effectiveness of different options (Geerts et al., 2008; Gao et al., 2012). Such controlling experiments have become popular tools for determining the optimal management of both fertilization and irrigation in the long term to minimize the environmental impacts for many important crop species, including rice, maize, wheat, soybean, etc. Further, increasingly advanced approaches, including global positioning system (GPS), wireless sensor networks and unmanned aerial vehicles (UAV), have been utilized to provide accurate monitoring of field locations, crop growth conditions and soil properties (Zhang et al., 2002; Wang et al., 2006; Gómez-Candón et al., 2014). Such approaches facilitate the collection of large amounts of data at a high spatial-temporal resolution. Thus the integration of both these advanced technological approaches and field experiments can lead to the development of improved real-time management strategies.

Many of these experimental and technological approaches are most beneficial at the local scale, with high costs associated with labor and equipment, as well as the need for specialized skills, which have prevented the wide use of such approaches over regional scales (Zhang and Kovacs, 2012). Simply upscaling local data to a regional level is not often possible (or advised) due to the significant heterogeneity in soil and crop conditions, thus leading to a high amount of uncertainty in the resulting data. In addition, without long-term or good quality historical data, these approaches are limited in their predictive performances, especially during the meteorological extremes (e.g. extreme drought). A solution to these issues can be found by using process-based crop models, which are developed through a combination of mathematical equations describing the interaction between crop growth, soil nutrient dynamics and agricultural management (Rauff and Bello, 2015). For example, global gridded crop models (GGCMs) can be used to project the yield potential under climate change at regional or global scales (Rosenzweig et al., 2014). Other models, e.g. AquaCrop, WOFOST, DeNitrification and DeComposition (DNDC), are widely applied for deficit irrigation, optimal fertilization schemes and estimation of GHG emissions (Miao et al., 2006; García-Vila et al., 2009; Uzoma et al., 2015). With field experiments or monitoring providing observed facts for model calibration, models can be used to upscale the results and offer timely information about regional conditions. Driven by reliable input database (e.g. climate forecast or reanalysis), crop models can also be used to predict the potential crop growth under different scenarios and calculate the long-term climate risk for better agricultural management (Huang et al., 2018).

Though originally developed and validated at field scale, process-based crop/biogeochemical models are becoming more popular in regional-scale simulations (Holzworth et al., 2015). Yu et al., (2019) used the DNDC model to quantify the provincial-level N discharge from cropland in China and evaluated the contribution of optimal fertilization to water quality. At the global scale, Liu et al., (2016) analyzed the response of wheat yield to rising temperature at a 0.5° spatial resolution based on the simulations of seven crop models. Elliott et al., (2014) projected the global water limitation to maize, soybean, wheat and rice productivity under climate change by combining 16 global hydrological and crop models and then assessed the adaptation potential by irrigation improvement. Overall, regional simulations using process-based models have been proven as a powerful approach in predicting the effects of climate change on crop productivity and the response of agroecosystems to different management practices (Deryng et al., 2011; Zhao et al., 2013; Drewniak et al., 2015; Müller et al., 2015; Bowles et al., 2018). As such, these models have the potential to play an important role in policy making regarding food security, climate change mitigation and environmental protection. The utilization of these models continues to expand, due in part to the many agricultural modelling systems (or software) providing user-friendly tools for various applications (Gerber et al., 2008; Yu et al., 2014; Capalbo et al., 2017; Han et al., 2017; Rurinda et al., 2020). However, there are still several key challenges:

(i) Computing efficiency prohibits the use of models in regional simulation with very high resolution (i.e. global-scale simulations with 0.1° grid cells) over decadal time periods. The traditional approach, where the computation proceeds grid cell by grid cell is time intensive. Some crop models (e.g. PaSim, APSim) adopt high performance computing (HPC) technology to accelerate the model run time by using parallel computing, where independent grid cells are processed at once across multiple CPUs (Vital et al., 2013; Zhao et al., 2013). For integrated modelling systems, Buahin et al., (2019) cloned each component in a water temperature model and designed a parallel execution framework to achieve high computing efficiency. However, most crop models (e.g. WOFOST, AquaCrop) and more complex biogeochemical models (e.g. DNDC, DayCent) do not have open-access parallel versions compatible with different operating environments.

(ii) There are few agricultural modelling systems available for users with all necessary components to perform a complete end-to-end simulation, from model calibration to scenario prediction and finally optimal management assessment. Most studies only focus on one aspect, such parameter optimization (Iizumi et al., 2009; Abbaspour, 2013), drought prediction (Yu et al., 2014), improved practices for ecosystem service (Chen et al., 2016) and water quality (Kaini et al., 2012). However, it is a difficult and time consuming process for users to perform these tasks independently with different software packages or source codes – something that could be changed by using a coupled system.

(iii) The structure of most modelling systems does not easily allow for further extension. Even when using the same original model code base, researchers will develop the model in different directions relevant to their own research interests. For example, based on the DSSAT model, Han et al., (2017) developed the CAMDT software to provide the seasonal forecast of crop growth and adaptation of managements, while Nguyen et al., (2017) applied the ant colony algorithm to optimize the irrigation and fertilization schedules. Although each application makes a novel contribution, combining both approaches could lead to even greater insights; however, such integration would be near impossible due to the disparate approaches, methods, and software used in each study. Even with very powerful processing systems, such integration would remain insurmountable. Therefore, a flexible structure is critical for the sustainable development of agricultural modelling system.

This paper seeks to address these challenges by developing an integrated modelling system, entitled Gridded- DeNitrification and DeCompostion (GDNDC). This is based on the established DNDC model, which is a nitrogen-based biogeochemical model for agroecological processes (Li et al., 1992). It models crop growth, soil water dynamics, soil carbon and nitrogen cycles under different management practices, with widespread use across GHG emission estimation (Li et al., 2001), yield prediction (Yu et al., 2014; Huang et al., 2018) and N leaching (Qiu et al., 2011; Yu et al., 2019) at regional scales. Using the DNDC model as the emulator for agro-biogeochemical processes, we aim to: (i) present a new structure for agricultural modelling systems by introducing a central coupler to integrate existing and potential future modules; (ii) enable parallel simulations with MPI (Message Passing Interface) protocol to increase computing efficiency for simulating tasks with high computational expenses; (iii) couple a number of additional modules to the model including a parameter optimization module using SCE-UA algorithm (Duan et al., 1992), a tool for scenario-based drought prediction and risk analysis of yield, and finally an optimal fertilization estimator for decision support.

In Section 2 of this paper, we introduce the newly developed structure of GDNDC, which now mainly depends on the dispatch of the coupler. In Section 3, we describe the detailed methods used in different modules including parallel running, parameter optimization, optimal fertilization estimate, drought scenarios settings and risk calculation. In Section 4, case studies for regional scale applications are presented to illustrate the whole workflow for using GDNDC. Finally, we discuss potential improvements and summarize the characteristics of our system in Section 5 and 6.

## 2. Framework of the GDNDC system

### 2.1 Overview of the GDNDC system

The current version of GDNDC system is developed using C++. With only standard libraries (normally compilable for most common compilers) invoked across the whole program, the system is compatible with different operating systems (Windows and Linux) and hardware environment (PC and cluster). Similar to DNDC 95, users of GDNDC are able to perform both field-scale simulations and regional-scale simulations. In regional-scale simulations, users can split their study regions (e.g. state, nation, globe) into a larger number of grid cells at a defined spatial resolution from 0.01° to 0.5°, according to the corresponding resolution of input data (e.g. soil map, climate data). The temporal scale is also defined by users from one month to over 100 years. Compared with DNDC 95, the parallel computing mode has been developed for regional-scale simulation to accelerate the computing efficiency. In addition to this development, we have coupled several additional modules in this system, in which users can use for predicting crop yield and the risk under drought events, as well as proposing improved N fertilization schemes to protect water quality. The structure of GDNDC enables convenient extension for other applications (see section **2.3** and **2.4**).

### 2.2 Modules in the GDNDC system

The GDNDC system consists of five modules (see Fig. 1b):

(1) *Coupler* module: The *Coupler* works as the trunk of GDNDC system to couple other modules together. Initially it recognizes the input settings from modelling tasks with different goals, and begins to initializes the corresponding modules. Throughout the simulation process, the *Coupler* collects the outputs and delivers relevant information between working modules. Further detail is explained in section **2.3**.

(2) *DNDC* module: This module is responsible for the calculation of all biogeochemical

processes from the DNDC model. This only includes the original process-based parts of the DNDC95 version with the rest such as the input/output (I/O) integrated into the *I/O* module. It therefore makes it a pure emulator in this system.

(3) *I/O* module: The *I/O module* reads the settings of a modelling task and input database and writes the outputs to be exported. The detailed description of the *I/O* files is presented in Table 1.

(4) *Parameter optimization* module: This module uses an optimization algorithm to determine the optimal parameters to reduce the discrepancy between model outputs and corresponding observation data. Users can improve the predictive capacity for targeted outputs given the spatial heterogeneity at regional scales. We explain the mathematical background of this module in section **3.2**.

(5) *Decision support* modules: It includes the *Optimal fertilization*, *Scenario prediction* and *Risk analysis* modules. They are developed to realize the estimation of optimal fertilization schemes, scenario-based prediction and yield loss analysis, respectively. The methods used in GDNDC to realize these functions are shown in sections **3.3-3.5**.

## 2.3 Module coupling

While different modules can be directly coupled into the DNDC source code to extend the corresponding functions (see Fig. 1a), following such an approach has a number of disadvantages. Firstly, as the source code is bounded together, the program becomes increasingly complicated. As such, further modification can become challenging if previous alterations not be documented properly, and developers fail to remember how modules are coupled together. Secondly, to extend the code, a developer requires a deep understanding of almost every process in the system in order to make their required changes without compromising the wider code base – an inherently complicated and time consuming task. Finally, for models like DNDC with many users across the world, incorporating all of the valuable contributions into one single codebase is not a trivial task.

On the other hand, for Earth system models (e.g. Community Earth System Model, CESM) with several complex components (e.g. land surface model, atmosphere model and ocean model), a *coupler* is used as the trunk of the system to communicate with all the other components. The outputs of a certain component are firstly delivered to the *coupler*, which then will send the required information in suitable format to initialize and activate another component. Such a structure keeps all process-based components independent from each other (and able to run in parallel or further developed in isolation) while the *coupler* is primarily used for information exchange between them. Following this, we added a simple *coupler* as the kernel to coordinate the processes among different modules in GDNDC. The general structure of GDNDC is presented in Fig. 1b.

In the GDNDC system, the *coupler* consists of four main components: *Mode control*, *Data stream*, *Task manager*, and *Timer* (see descriptions in Table 2). In the general workflow of this system, the *I/O* module is first called by the *coupler* to read the setting file (see Table 1). All the information is packed as a structure and delivered into coupler. Then in the *coupler*, *mode control* recognizes which computing mode (serial or parallel) is used, the *Timer* calculates the time nodes to read/write data, while *Task manager* initializes *DNDC* and *other* modules. Following these steps, the modelling process starts. For every individual day within the simulated time period, the *Timer* checks if the system needs to update the input data (e.g. parameter, climate and management practices) from the input database. If so, *I/O* will be called again to read the corresponding data (Table 1, [1.2]) and it transmits the data into *Data stream*. Then the data will be handled by *Data stream* and delivered to *DNDC* to activate and enable the modelling process. After completing the calculation for one day, model outputs (e.g. aboveground biomass, soil moisture, leaching, $N_2O$ emission, amongst others) are collected in *Data stream* for inputs into other targeted modules:

(1) For parameter optimization, model outputs are transported from *Data stream* to *Parameter optimization* module and then compared with observation data. Then new parameter sets can be updated and passed to *Data stream* and then to *DNDC* module for the next iteration of the simulation.

(2) For estimating the optimal fertilization strategy, the *Optimal fertilization* module generates different levels of fertilizer application and different kinds of fertilization methods. These combinations are transported into *Data stream* and used to replace the fertilization scheme. *Data stream* delivers the new management information to *DNDC* module to test the performance of new fertilization schemes.

(3) For scenario-based prediction, *Timer* provides the time information to *Scenario prediction* module, in which the future climatic scenarios are generated and then used to update the climatic information in *Data stream.* Afterwards the climate scenarios are transported to *DNDC* which enables the yield modelling.

(4) For yield loss estimation, *Risk analysis* module receives the simulated yield values in different irrigation and fertilization levels and then calculates the corresponding return period of yield loss at different spatial scales.

### 2.4 Advantages over the DNDC

The structure of the GDNC, based on the coordinating *coupler* shows a number of advantages over DNDC 95 for maintenance and expansibility purposes. These include:

(i) In the regional simulation mode in DNDC 95, the model reads all input data at the start of the simulation and proceeds to perform all numerical calculation from start to finish. For long-term simulation, the management settings (e.g. fertilizer level) in each year are kept constant, which does not reflect reality. If users want to update their simulation with new data available, they instead have to start the simulation from the beginning year every time. Whereas in the GDNDC system, the *I/O* process is an independent module controlled by the *coupler*, which in turn enables the dynamic update of new management information for each year of simulation whilst reloading key state variables (e.g. soil moisture, N/C pools) from the previous timestep.

(ii) All the other modules only exchange information with the *coupler*, keeping the program clear and understandable for efficient maintenance. Developers can focus on the single module of interest and do not need to consider others, thus enabling the parallel development of GDNDC from users across different specialties.

(iii) The opportunity for developing custom modules and enhancing existing modules in GDNDC will strengthen its power as an agricultural modelling system. For example, in the *I/O* module, developers can couple numerical climate models (e.g. Weather Research and Forecasting model, WRF) to provide short-term climate predictions for the *DNDC* module. Similarly, different algorithms can be supplemented into the *Parameter optimization* module. Modifying the data exchange interface in *coupler* would allow lots of other models (e.g. agent-based, water quality and economic models) to be integrated as additional modules to extend the application of GDNDC.

# 3. Methodology

## 3.1 Parallel computing

Across the components of the GDNDC system, the *DNDC* model has the greatest computational expenses as it runs at an hourly resolution and includes lots of numerical calculation for soil dynamics. Therefore, by enabling the *DNDC* model to run in parallel will greatly reduce the simulation run time. We develop two options for users: the serial mode and parallel mode. In the serial mode, a multiple of grid cells (e.g. regular 0.05° grids or irregular administrative grids) are allocated with one single process. The computation of certain grid only starts after the completion of the previous one (see Fig. 2a). This mode is recommended for field-scale simulations and debugging. Whereas in parallel mode, a number of processes (user defined within cluster's capacity) can be initialized simultaneously using MPI protocol. All the grid cells are matched to these processes uniformly, and each process can independently perform its calculations in parallel (see Fig. 2b). Users can expect significant improvements in the efficiency of regional-scale simulations.

## 3.2 Parameter optimization

In GDNDC, we couple the global optimization algorithm SCE-UA to automatically calibrate the model performance and obtain the optimal parameter sets. SCE-UA is a global optimization method to solve nonlinear problems in high-dimension space by combining deterministic and probabilistic approaches. In this algorithm, multiple "complexes" are initialized with their points randomly sampled from the search space. The downhill simplex algorithm (Nelder and Mead, 1965) is applied for evolving each complex independently in the direction of global improvement. Meanwhile, these complexes are periodically shuffled and all the points are reassigned to avoid the search getting trapped in local optima (for detailed mathematical processes see Duan et al., 1992; Duan et al., 1994). It enables the search progress to converge towards the global optimum with high efficiency. SCE-UA was initially developed for the hydrological models (Sorooshian et al., 1993; Duan et al., 1994; Yang et al., 2008), and later became popular for crop models and biogeochemical models (Ueyama et al., 2016; Jin et al., 2018; Cui and Wang, 2019). For consistency with the wider GDNDC system, the Fortran version of the SCE-UA source code was translated into C++ before being adopted as a module.

In Table 3, eight crop-related parameters which are sensitive in the modelling of water and nitrogen dynamics are listed. These parameters include: (i) *MaxY* for the theoretical rate of daily N uptake and model's response to N supply; (ii) *TDD* for the phenological process; (iii) *WD* for the theoretical rate of daily water uptake and model's response to drought; (iv) *G_CN, L_CN, G_Fra and L_Fra* for the biomass accumulation and allocation in different organs; and (v) *VarY* for the influence of technology improvement (e.g. breeding). The relevant input file (see Table 1, [1.4]) is designed for users to select any combination of these eight parameters for optimization, while other parameters adopt default values from the regional database. The algorithm minimizes the RMSE (root-mean-squared-error) as the objective function:

$$\text{obj RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \rightarrow \min \quad (1)$$

where $\hat{y}_i$ and $y_i$ are the predicted and observed variables (e.g. yield, soil moisture) at the $i^{th}$ time step, respectively. By running the optimization module, the DNDC model will be called iteratively with a set of parameters from SCE-UA. After each iteration, model outputs are fed back to the *coupler* and then used for deriving a new set of parameters to minimize the objective function in Eqn (1). The optimization process stops when it reaches user-defined convergence standard or maximum iteration.

### 3.3 Optimal fertilization

The *Optimal fertilization* module determines the minimum fertilizer application required to maintain targeted yield levels while minimizing the environmental costs, including $N_2O$ emission and N leaching. Compared with the *n*-dimension search for optimal parameters in section **3.2**, the *1*-dimension search for optimal fertilizer amount is much less demanding. We adopt the method of bisection with the workflow given in Fig. 3. In the first step, the system simulates the yield level using the current fertilization level (see Table 1, [1.5]) and sets it as the target. The range of optimal fertilizer amount is set between 0 and current level. By using the method of bisection, the module compares the targeted yield with the simulated yield using the mid-range of the fertilizer. By this approach the fertilizer range is narrowed down until an optimal fertilizer amount is obtained. The default maximum number of iterations is set to 15 as this guarantees a final precision of ~0.1 kgN/ha.

### 3.4 Scenario-based yield prediction

Given the uncertainties involved in a regional climate projection, the GDNDC system adopts climatic scenarios from a historical database to drive the prediction of crop growth particularly under drought conditions. Following Yu et al., (2014) and Huang et al., (2018), we assumed the climatic forcing from a given time up until harvest follows one of three scenarios:

(1) **Ideal scenario**: The water deficit for crop growth ceases immediately after the current day. The water demand is thus fully met until the harvest. With this setting, the potential yield loss can be derived;

(2) **Drought continuing scenario**: A period without rain (e.g. 3 days, 10 days) following the current timestep of interest can be specified in Table 1 ([1.6]). After this period, the climate returns to the ideal condition. So the potential yield loss for the following drought can be estimated;

(3) **Historic scenario**: The climatic data in typical year in history (including historical wet, medium and dry year) are used to drive the simulation of yield. The yield losses under representative climate conditions can provide useful information to compare the severity of a current drought to others in recorded history

### 3.5 Risk analysis

Based on the dynamic update of yield predictions in section 3.4, the corresponding return period of yield loss can be estimated to demonstrate the impacts of droughts. The return period, often used to quantify the severity of natural disasters, including floods (Hirabayashi et al., 2013), droughts (Kwon and Lall, 2016) and wind storms (Della-Marta et al., 2009) is calculated as the inverse of the frequency of a certain event. It therefore represents the average recurrence interval of that particular event. For example, a 50-year drought implies that a drought event with equal severity has a 2% probability to occur in any year, or simply put, it could be expected to occur every 50 years on average. The GDNDC system follows three steps to quantify the agricultural drought return period.

Firstly, with the optimal parameters in section **3.2**, the model runs a long-term yield simulation over the past 50 years using historical climate data and current management practices (e.g. irrigation and fertilization). The yield outputs over a 50-year timespan for each grid cell constitutes the baseline yield database.

Secondly, in this system, the GEV (Generalized Extreme Value) distribution (see Eqn 2) is adopted as the default probability distribution curve for yield (Yu et al., 2014). For each grid cell, the baseline yield records are used to estimate the optimal parameters k, μ, and σ such that:

$$F(x) = \begin{cases} \exp\left(-\left(1 + k\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/k}\right) & k \neq 0 \\ \exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right) & k = 0 \end{cases} \tag{2}$$

where $F(x)$ is the cumulative probability function; $k$, $\mu$ and $\sigma$ are the shape, location and scale parameters of GEV distribution, respectively; and $x$ is the simulated yield or yield loss in this case.

Finally, after determining distribution parameters for each grid cell, we can calculate the value of $F(x_i)$ with the predicted yield $x_i$ driven by the $i^{th}$ climate scenario (e.g. drought continues 10 day without rain, as described in section **3.4**). The return period is then computed as $T(x_i) = 1/F(x_i)$.

# 4. Regional Scale Demonstration

## 4.1 Gridded modelling in parallel mode

In this section, we demonstrate the regional simulation performed for the Liaoning Province, China to illustrate the computing efficiency of the new parallel mode developed in GDNDC. 30 counties in this region are randomly selected to model the annual maize yield during 1996-2008, with each county as an independent grid. The whole numerical experiment is based on the Intel i7-8700 (3.20GHz) CPU cluster. To compare parallel and serial mode run times, we run the model eight times for the serial mode and for each of the parallel modes with 2, 3, 4, 5, and 6 MPI processes.

The numerical experiment in Fig. 4 explicitly demonstrates the significant improvement in the computation efficiency with the increase of MPI processes. The variations between each of the eight repeats are negligible. Therefore, the running time is expected to be greatly shortened with enough computing resources, especially for large-scale or global-scale simulation with thousands of grid cells. The enhanced computing capacity further ensures the effective performance of some other functions including parameter optimization and uncertainty analysis, which requires much more computation. The theoretical running time, computed as the average running time for one process (i.e. serial mode) divided by the number of processes, is also presented in Fig 4. We find in Fig. 4 that the run time in reality (practical running time) is slightly longer than the theoretical running time. We attribute the extra time to the computational requirements for communication between different processes. This could be increased further in a large cluster if the allocated nodes are physically far from each other. However, it is not significant considering the overall time.

### 4.2 Parameter optimization module for maize yield prediction

To demonstrate the improvement in predictive accuracy by incorporating the SCE-UA algorithm into the GDNDC, we carry out two model runs over the all 42 counties with maize plantation in Liaoning Province for the time period 1998-2008. The first simulation adopts the default values from the regional database for each crop-related input parameter. The second simulation instead uses the SCE-UA algorithm to optimize all eight parameters (as given in Table 3) over a maximum of 1000 iterations.

Results are presented at both the county level and aggregated together to form a provisional level estimation in Fig. 5. Bias correction methods have not been applied to the simulated results as a post-process, although doing so would be expected to improve the accuracy of the yield produced by the model (especially when using default parameters). We present the original outputs here as our system is also designed for water- or N-related simulations and any post-processing to yield outputs will cause a mass imbalance of the system when continuing model simulations for other applications.

By comparing the county-level simulated yields with observed statistical records in Fig. 5a and Fig. 5c, we can find the parameter optimization approach effectively enhanced the $R^2$ from 0.505 to 0.706 while reducing the RMSE from 1836 kg/ha to 1347 kg/ha. The number of outliers (distant from the 1:1 line) also decreases by using the optimal parameters. Similarly, for the province-level aggregation (Fig. 5b and 5d), the yield simulations using parameter optimization also correspond better to the observations – particularly in the recorded drought years 2000 and 2006.

### 4.3 Return period of yield loss in droughts

To demonstrate the *Risk analysis* module, GDNDC is used to simulate annual maize yields over 42 counties in Liaoning province across a 50-year period from 1961 to 2010. The optimal parameters obtained in section **4.2** are used to drive the model while the ideal maximum grain biomass is set to the 2008 level. Both the county-level outputs and province-level aggregation are used to derive the parameters of the GEV distribution (section **3.5)**. The province-level return period of maize yield in this region is shown in Fig. 6.

The most significant drought across the simulation time period was observed in 2000 with a recurrence interval of nearly 60 years. This is consistent with reality given the extreme summer drought that occurred across Liaoning that year. The droughts of the 1960's are estimated with around 15-year return periods – consistent with the conclusions of (Yu et al., 2018) who acknowledged that besides the natural drought conditions, socioeconomic factors also played an important role in the food deficit during that period.

Taking the 2000 drought, we demonstrate the workflow of the scenario-based dynamic yield prediction. Assuming the drought period started July 1st, (approximately the beginning of the productive stage for maize growth), we adopt observed climate data up until this date. From July 1st onwards, different climatic scenarios are generated (according to the scenarios listed in section **3.4**) such that simulation can proceed until harvest. In Fig. 7, the drought-induced yield losses and corresponding return periods under different scenarios are shown. We calculate the yield loss as followed:

$$Y\_loss_i = \frac{Y_{ideal} - Y_i}{Y_{ideal}} \times 100\% \quad (3)$$

where $Y\_loss_i$ is the relative yield loss under $i^{th}$ scenario (including the drought-continuing scenarios and typical-year scenarios); $Y_{ideal}$ is the simulated yield under the ideal scenario without any water deficit since the current day; and $Y_i$ is the simulated yield under $i^{th}$ scenario.

We find the drought-induced yield loss, as well as the corresponding return period, increases with the assumed length of drought. The next 10-20 days is the critical period for hazard mitigation, during which drought conditions are likely to cause further losses (from <15% at current stage to >30% 20 days later) which makes the magnitude of yield loss equal to the driest level in history. After 20 days, no further yield losses are observed since irreversible damage has been generated in the first 20 days. Special attention should be paid to the western and northern areas of this province given the areas seem to be more sensitive to drought conditions and therefore potentially more yield loss. Such dynamic maps for yield prediction are able to provide useful information and forecasts for decision makers.

## 4.4 Improved nitrogen use efficiency by optimal fertilization

Here the annual optimal fertilizer amounts from 2000-2008 are derived by GDNDC for the maize plantation of the 42 counties in Liaoning. We set the fertilization level of this region in 2008 (~227 kgN/ha synthetic fertilizer and ~20 kgN/ha manure) as the baseline for maize production and then calculate the minimal fertilizer amount which can still maintain the production while increase the nitrogen use efficiency (NUE). The calculation of NUE is defined as followed:

$$\mathrm{NUE} = \frac{N_{yield}}{N_{fer} + N_{dep} + N_{man} + N_{fix}} \quad (4)$$

where $N_{yield}$, $N_{fer}$, $N_{dep}$, $N_{man}$ and $N_{fix}$ refer to the nitrogen in yield, fertilizer, deposition, *manure*, and biological fixation, respectively. In Fig. 8, we show the long-term annual average of (i) the fertilizer reduction rate by optimal fertilization compared with baseline level, and (ii) the NUE at both the baseline and optimal levels. It reveals the over-fertilization still exists in Liaoning and a 14% reduction of N fertilizer application can be achieved without lowering the production level. The west and north counties in Liaoning have a relatively lower rate of fertilizer reduction because more N is required to maintain the higher maize yield compared with the counties in the east. Besides, the NUEs at county level are also improved significantly by optimal fertilization (Table 4). The averaged NUE in Liaoning increases from 0.19 to 0.42 by optimizing fertilizer application. Therefore, it is expected to effectively save monetary and energy costs associated with fertilizer application whilst improving the regional environment by reducing the surplus N load to groundwater and surface water. Although the NUE values vary annually due to meteorological factors (e.g. heavy precipitation and runoff), GDNDC has the advantage of being able to compute the optimal fertilizer amount year-by-year based on the climatic and management conditions.

## 5. Discussion

DNDC model has been widely used for the regional-scale simulation for agro-biogeochemical dynamics in the past decade. While improvements have been made to the scientific processes of the model, its serial computing mode limits its application for modelling tasks with high computational demand. At the same time, the general structure has been maintained in its original form – originally intended for field-scale applications. It combined I/O processes, biogeochemical processes, and some other functions for decision support, which makes the whole program difficult to understand. Researchers who are not familiar with the detailed processes in this model must invest significant time familiarizing themselves with it before embedding their contributions into the source code. Subsequently, many unique versions with the same underlying model have been developed as it is not possible for the current structure to integrate all modifications by different individuals. It leads to issues with version control and is not sustainable for DNDC's development.

The *coupler* developed in GDNDC is to substitute the previous structure and coordinate the cooperation between different modules. As the process-based module (*DNDC*) and application modules (e.g. *Optimal fertilization*) are all independent from each other, both the developing efficiency and maintenance of different versions could be significantly improved. Apart from its basic use for biogeochemical modelling, a more integrated system can be achieved in the future for hazard prediction and resource management by coupling other modules (e.g. regional climate model and agent-based model) in a similar way.

The compatibility for both the serial mode and parallel mode is achieved in GDNDC. Unlike the previous work by Huang et al., (2018), which parallelized the DNDC in a unique supercomputer platform, the MPI method used in GDNDC is more compatible in universal computing environments, including PC and large HPC clusters. Now users of this model are able to choose between serial mode for debugging or small-scale simulation, or using parallel modes to accelerate the computation for regonal-scale modelling. Furthermore, GPU-based accelerating approaches have the potential to further speed up the calculation of these processes across multiple soil layers, however, this has not been coupled to GDNDC given the heavy reliance on specific hardware and therefore compatibility/usability.

The modules *Parameter optimization* and *Scenario prediction* are integrated in GDNDC to improve the modelling accuracy and quantify future potential yield loss, respectively. As crop N uptake is one of the most important components for both the crop growth dynamic and soil N balance, the optimization in the current version of GDNDC only focuses on these parameters which are sensitive to crop growth. Further development could be made by adding other parameters if more accurate simulations are required for GHG emission, N leaching, or soil organic carbon. Compared with the single-objective optimization, multi-objective optimization could not only improve the predictive accuracy of multiple metrics of model simulations, but also contribute to more complex management goals when users have to consider yield productivity, soil quality, and environmental effects simultaneously. Relevant algorithms like NSGA-II (Deb et al., 2002) and MOEA/D (Zhang and Li, 2007) are targeted additions to the system. Further development is also focused on a data assimilation module. As the predictive bias can still accumulate in the long-term running (even when adopting optimal parameters), this module will utilize real-time satellite data (e.g. Modis LAI) to correct the model state variables. Additionally, considering the uncertainty of the climatic scenarios derived from historical datasets, the online data extraction for climate observations and forecast will also be supplemented into the following version.

A method of bisection is used in the algorithm to derive the minimal N fertilizer amount while maintaining the production level. With this approach, an optimal nitrogen use can be obtained with the overall environmental cost considered. However, users may consider the term “optimal fertilization” to have a broader scope than the minimal fertilizer use defined in GDNDC. As a result, the module will be enhanced over time to incorporate additional targets based on the practical demand in the future. For the risk analysis module, the return period metric provides a readily useable and understandable metric for local governments seeking to mitigate the impacts of drought. Others, e.g. Huang et al., (2018) and Gaupp et al., (2017) have used a Copula function to derive the joint probability of yield losses among multiple region. Thus far, it has not been included in GDNDC because of the dependence on both the distribution curve and Copula function, and therefore the information is not always easily translated for dissemination to the public and policy makers.

GDNDC system integrates different modules together to provide useful information for decision support. Compared with other agricultural modelling system concentrating on a specific application, GDNDC system connects the whole workflow from parameter optimization to drought prediction, optimal management strategy and risk analysis. It provides convenience to users with different backgrounds as they do not need to switch between software or applications to achieve their desired results. Meanwhile, the new structure of GDNDC presented in this research creates a user-friendly environment for joint collaboration among the community of DNDC users. It does not require expertise across the whole system before developers can start to develop their own modules. Unlike some agricultural modelling systems which may be maintained by a professional team or stop seeing further support/development after completion of project, we believe GDNDC is suitably structured to allow widespread international collaboration and development and advance the science of agricultural systems modelling.

## 6. Conclusion

In this research, we presented the new GDNDC system based on crop-DNDC95 for regional simulation on agro-biogeochemical processes. The original structure of this model is substituted with the new framework and a coupler as its kernel to coordinate the interaction between different modules. We believe that the GDNDC system can significantly improve the efficiency of development for both the scientific and practical purposes among different developers and contribute to the version control of this model. Users can run simulations in both serial and parallel modes which are embedded into GDNDC, of which the significant benefits of parallelization have been demonstrated. In addition, several modular functions including parameter optimization, scenario prediction, optimal fertilization and risk analysis, which are all frequently applied by third-party software in research or practical application, are now integrated into GDNDC by default. With application to Liaoning Province, we demonstrate the effectiveness of GDNDC in providing useful information about crop yield prediction, drought hazard assessment, and fertilization guidance. While further improvements for GDNDC are in progress to integrate further state-of-the-art techniques and data products, we have demonstrated that the new GDNDC in its current form still enhances the accessibility and convenience for users from different sectors. Overall, the GDNDC is in a position to now provide timely and trustworthy simulation outputs and forecasts that stakeholders, including researchers, farmers, policy makers and insurance companies, need for both long term decision making to reduce the agricultural sectors effects on the environment and advise reactive decisions in times of severe drought to minimize yield loss.

Table 1

The information about input/output files

| **Input files (.txt)** | |
|---|---|
| [1.1] Setting file | (1) Goal of modelling task (e.g. long-term modelling, parameter optimization, etc);<br>(2) Simulating period;<br>(3) Running mode (serial or parallel);<br>(4) Path of input database;<br>(5) Time interval to read input;<br>(6) Path of output file;<br>(7) Time interval to write output; |
| [1.2] Input database<br>(for regional simulation, the same property of all grids are merged into one file) | (1) Soil property file;<br>(2) Crop parameter file (default);<br>(3) Planting structure file;<br>(4) Fertilizer amount file;<br>(5) Fertilization method file;<br>(6) Manure amount file;<br>(7) Irrigation ratio file;<br>(8) Planting/harvest date file;<br>(9) Tillage information file;<br>(10) Climatic data files; |
| [1.3] Output selection file | The names of over 120 variables are listed in this file, regarding to soil water, carbon, nitrogen cycles and crop growth. Users can select among them and decide what to write out. |
| [1.4] Parameter optimization file (if used) | (1) Selected model parameters;<br>(2) The prior interval of parameter value;<br>(3) Parameters for SCE-UA;<br>(4) Observations; |
| [1.5] Optimal fertilizer file<br>(if used) | (1) The current level of fertilizer amount;<br>(2) Maximum iteration number; |
| [1.6] Scenario prediction file<br>(if used) | (1) Typical year (dry, wet, mid);<br>(2) User-defined drought continuing days; |
| **Output file (.dat)** | |
| [2.1] Restart file | The state variable on the end day of simulating period. It is used to restart the simulation. |
| [2.2] Output file | It contains information of the selected outputs in 1.3 |

Table 2

The description of the main components of coupler

| **Component** | **Role** |
|---|---|
| *Mode control* | To switch between serial mode and parallel mode and allocate computing processes for numerical calculation; |
| *Data stream* | For the data distribution among different modules; |
| *Task manager* | To dispatch different task according to user’s requests; |
| *Timer* | To control the progress of system running at different time nodes; |

Table 3

The key parameters in GDNDC available for optimization

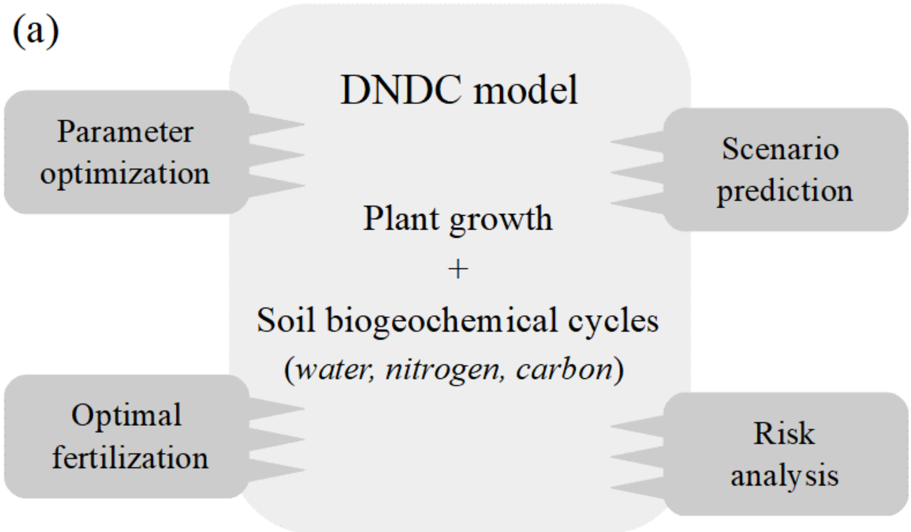
| **Parameter** | **Meaning** | **Unit** | **Range*** |
|---|---|---|---|
| MaxY | The maximum biomass of grain at harvest | KgC/ha | (0.5, 1.5) |
| TDD | Thermal degree days required to reach maturity | °C/day | (0.8, 1.2) |
| WD | Water demand for crop growth | Kg | (0.7, 1.3) |
| G_CN | C:N ratio of grain | KgC/KgN | (0.8, 1.2) |
| L_CN | C:N ratio of leaf | KgC/KgN | (0.8, 1.2) |
| G_Fra | The allocation coefficient of biomass for grain | - | (0.8, 1.2) |
| L_Fra | The allocation coefficient of biomass for leaf | - | (0.8, 1.2) |
| VarY | The annual variation in maximum yield considering cultivar improvement | % | (0.0, 5.0) |

* It means the multiplier to the default value in DNDC's regional database of crop properties.

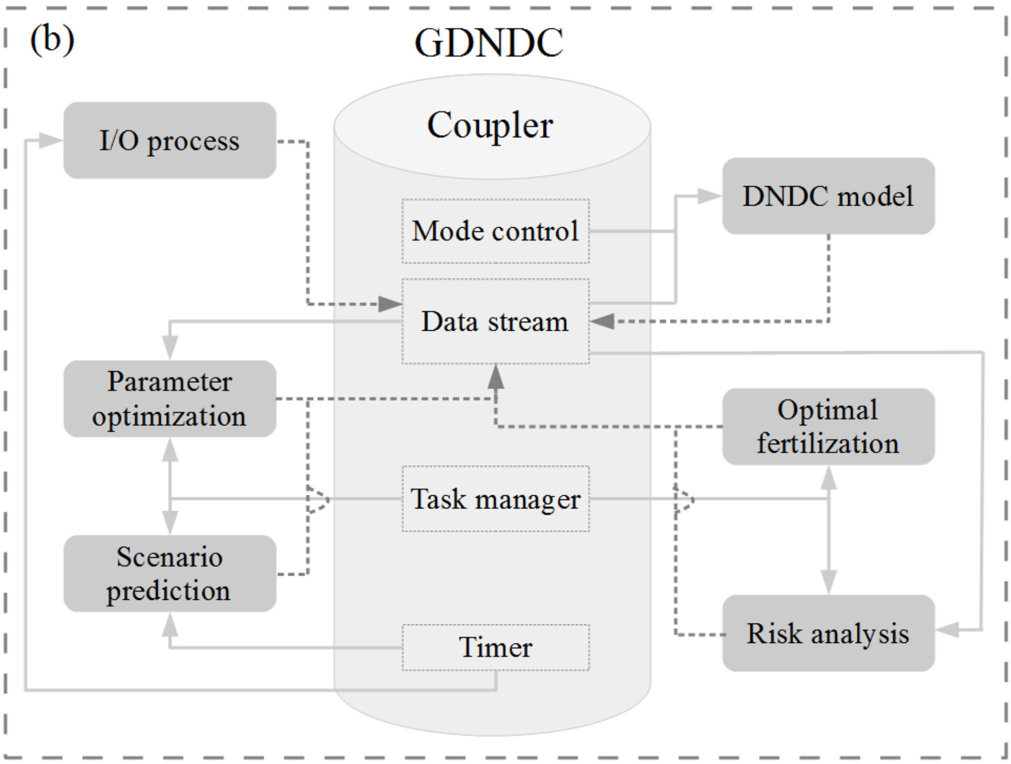
Table 4

The province-level annual NUE in both baseline and optimal levels

| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | Average |
|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | 0.14 | 0.18 | 0.18 | 0.19 | 0.20 | 0.23 | 0.19 | 0.20 | 0.22 | 0.19 |
| Optimal | 0.33 | 0.40 | 0.40 | 0.42 | 0.41 | 0.49 | 0.41 | 0.43 | 0.48 | 0.42 |




652




653
654 Figure 1. (a) Traditional way of coupling process-based models with other functions;
655 (b) The new framework of GDNDC system based on coupler coordination.
656

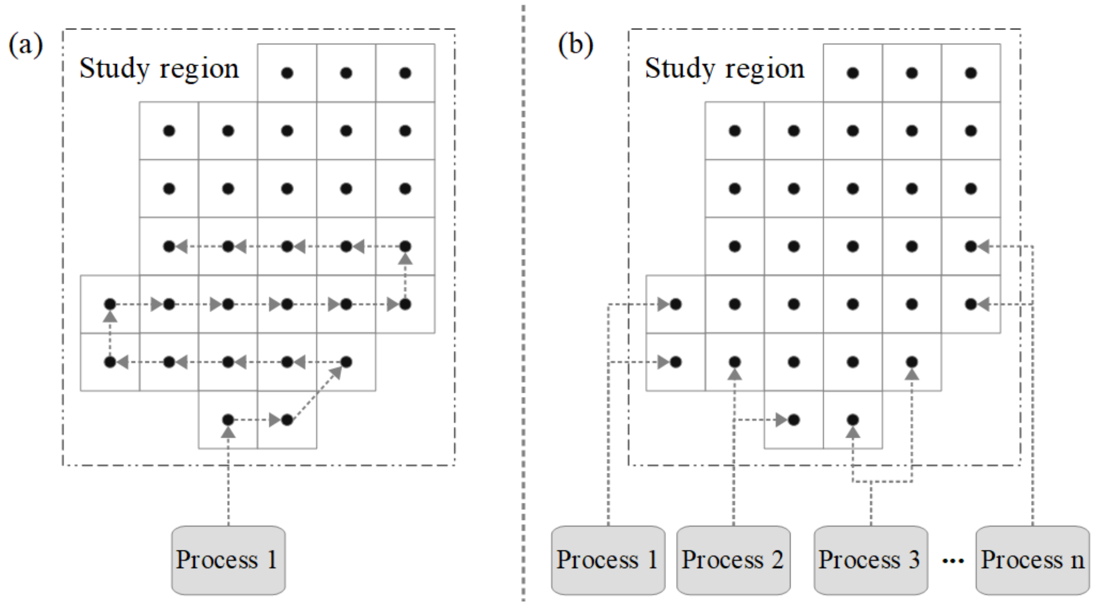



Figure 2. The description of two computing modes for *DNDC* module: (a) serial mode with one process from start to finish; and (b) parallel mode with multiple processes operating simultaneously to significantly reduce the model simulation time.

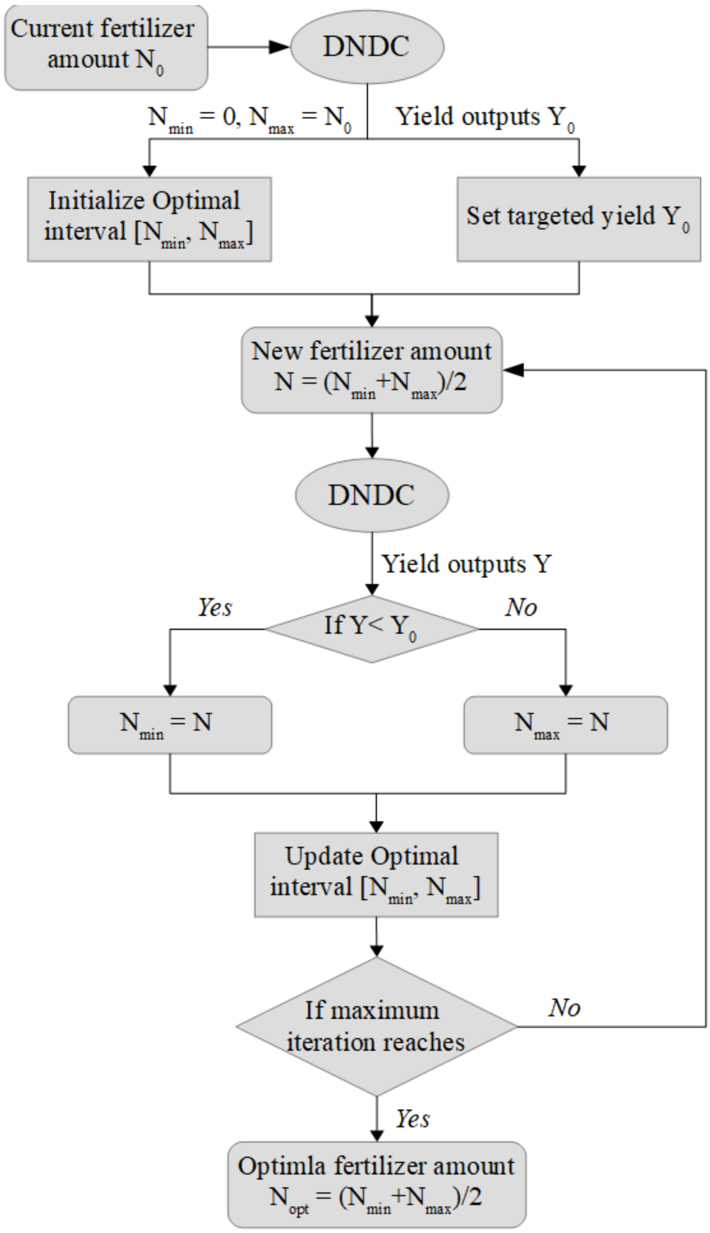



Figure 3. The workflow to determine the optimal fertilizer amount in GDNDC

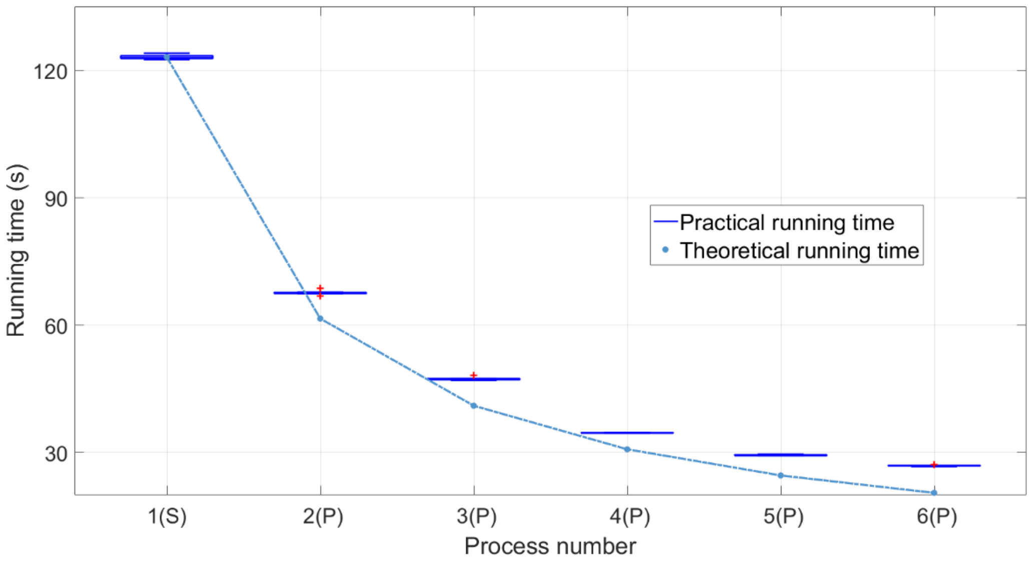



Figure 4. Boxplot of the running time using different numbers of process. S: serial mode; P: parallel mode. Theoretical runtimes for parallel processes are calculated as the practical (observed) runtime from one process (serial) divided by the number of processes in total.

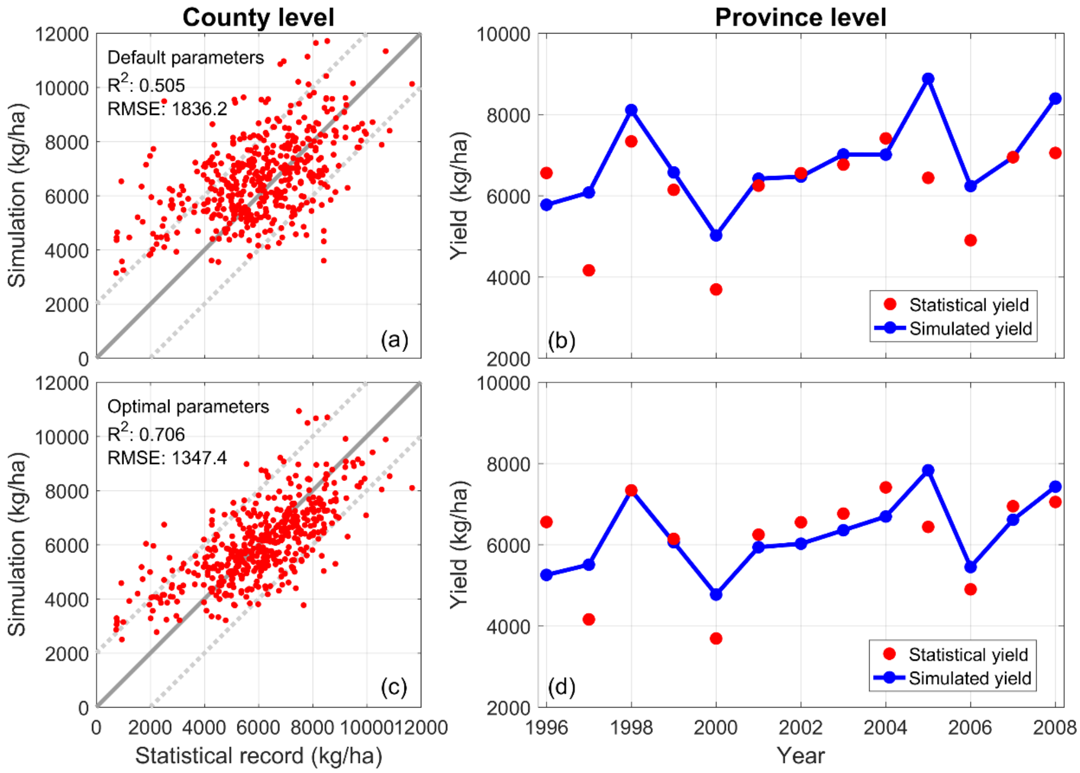



Figure 5. The performance of yield simulation using (a) default parameters at the county level, (b) default parameters with yield aggregated to the provincial level; (c) optimal parameters at the county level, and (d) optimal parameters aggregated to the provincial level.

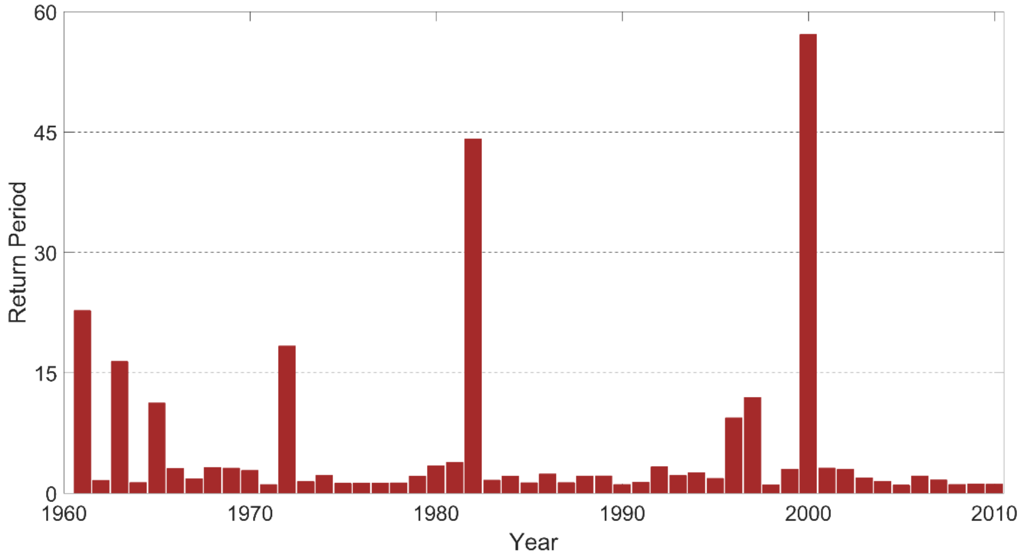



Figure 6. Estimated return periods of the province-level maize yield in Liaoning, China for 1960-2010.

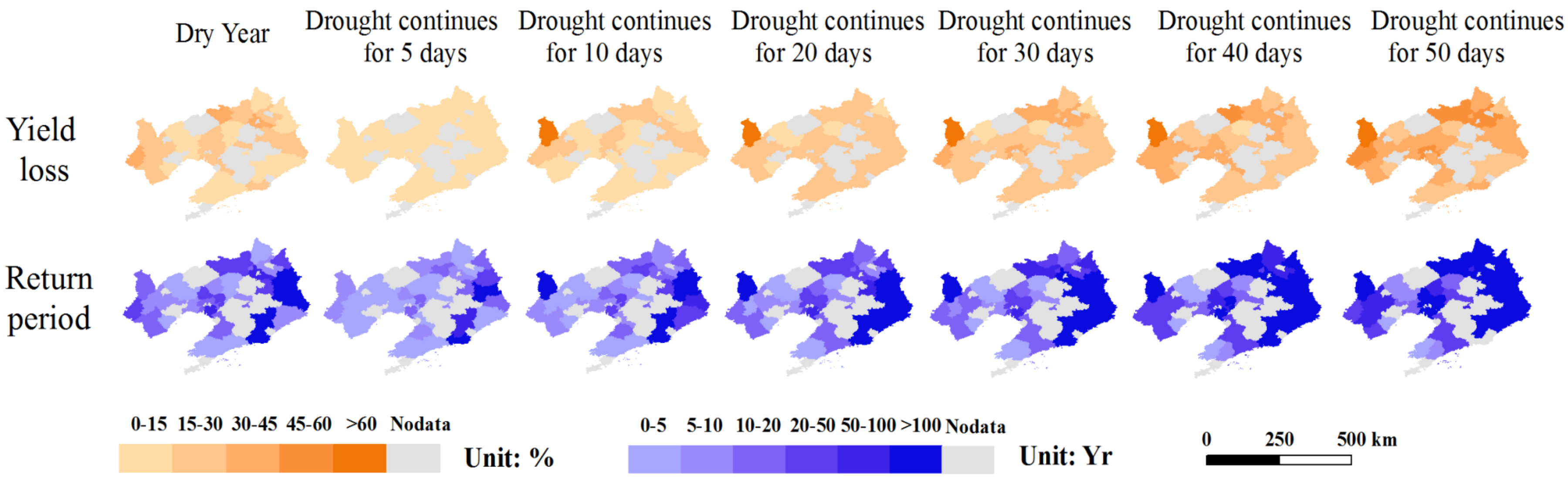



Figure 7. County-level predictions of both the yield loss and return period under different climate scenarios on July 1st, 2000

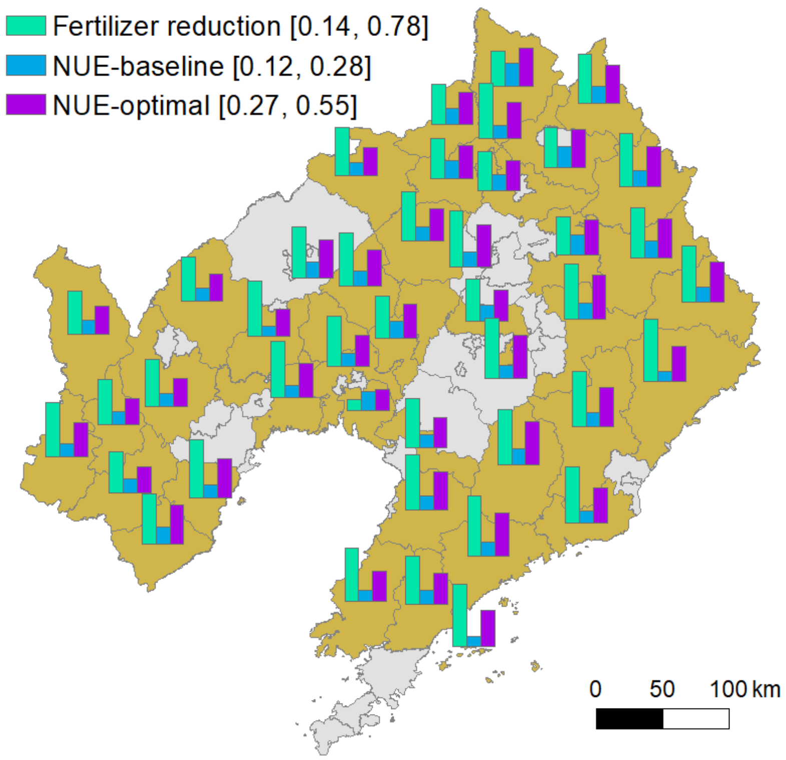



Figure 8. The county-level annual average during 2000-2008 of fertilizer reduction by optimal fertilization and the nitrogen use efficiencies (NUEs) at both the baseline and optimal level

# References:


[1] Abbaspour, K.C., 2013. SWAT-CUP 2012. SWAT Calibration and uncertainty program—a user manual.

[2] Bowles, T.M., Atallah, S.S., Campbell, E.E., Gaudin, A.C.M., Wieder, W.R., Grandy, A.S., 2018. Addressing agricultural nitrogen losses in a changing climate. Nature Sustainability 1, 399-408.

[3] Buahin, C.A., Horsburgh, J.S., Neilson, B.T., 2019. Parallel multi-objective calibration of a component-based river temperature model. ENVIRON MODELL SOFTW 116, 57-71.

[4] Cai, Z., Xing, G., Yan, X., Xu, H., Tsuruta, H., Yagi, K., Minami, K., 1997. Methane and nitrous oxide emissions from rice paddy fields as affected by nitrogen fertilisers and water management. PLANT SOIL 196, 7-14.

[5] Capalbo, S.M., Antle, J.M., Seavert, C., 2017. Next generation data systems and knowledge products to support agricultural producers and science-based policy decision making. AGR SYST 155, 191-199.

[6] Chen, H., Yu, C., Li, C., Xin, Q., Huang, X., Zhang, J., Yue, Y., Huang, G., Li, X., Wang, W., 2016. Modeling the impacts of water and fertilizer management on the ecosystem service of rice rotated cropping systems in China. Agriculture, Ecosystems & Environment 219, 49-57.

[7] Cordell, D., Drangert, J., White, S., 2009. The story of phosphorus: Global food security and food for thought. Global Environmental Change 19, 292-305.

[8] Cui, G., Wang, J., 2019. Improving the DNDC biogeochemistry model to simulate soil temperature and emissions of nitrous oxide and carbon dioxide in cold regions. SCI TOTAL ENVIRON 687, 61-70.

[9] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE T EVOLUT COMPUT 6, 182-197.

[10] Della-Marta, P.M., Mathis, H., Frei, C., Liniger, M.A., Kleinn, J., Appenzeller, C., 2009. The return period of wind storms over Europe. INT J CLIMATOL 29, 437-459.

[11] Deryng, D., Sacks, W.J., Barford, C.C., Ramankutty, N., 2011. Simulating the effects of climate and agricultural management practices on global crop yield. GLOBAL BIOGEOCHEM CY 25.

[12] Drewniak, B.A., Mishra, U., Song, J., Prell, J., Kotamarthi, V.R., 2015. Modeling the impact of agricultural land use and management on US carbon budgets. BIOGEOSCIENCES 12, 2119-2129.

[13] Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. WATER RESOUR RES 28, 1015-1031.

[14] Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J HYDROL 158, 265-284.

[15] Elliott, J., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., Glotter, M., Flörke, M., Wada, Y., Best, N., Eisner, S., Fekete, B.M., Folberth, C., Foster, I., Gosling, S.N., Haddeland, I., Khabarov, N., Ludwig, F., Masaki, Y., Olin, S., Rosenzweig, C., Ruane, A.C.,

Satoh, Y., Schmid, E., Stacke, T., Tang, Q., Wisser, D., 2014. Constraints and potentials of future irrigation water availability on agricultural production under climate change. Proceedings of the National Academy of Sciences 111, 3239.

[16] Famiglietti, J.S., 2014. The global groundwater crisis. NAT CLIM CHANGE 4, 945.

[17] Gao, Q., Li, C., Feng, G., Wang, J., Cui, Z., Chen, X., Zhang, F., 2012. Understanding Yield Response to Nitrogen to Achieve High Yield and High Nitrogen Use Efficiency in Rainfed Corn. AGRON J 104, 165-168.

[18] García-Vila, M., Fereres, E., Mateos, L., Orgaz, F., Steduto, P., 2009. Deficit Irrigation Optimization of Cotton with AquaCrop. AGRON J 101, 477-487.

[19] Gaupp, F., Pflug, G., Hochrainer-Stigler, S., Hall, J., Dadson, S., 2017. Dependency of Crop Production between Global Breadbaskets: A Copula Approach for the Assessment of Global and Regional Risk Pools. RISK ANAL 37, 2212-2228.

[20] Geerts, S., Raes, D., Garcia, M., Vacher, J., Mamani, R., Mendoza, J., Huanca, R., Morales, B., Miranda, R., Cusicanqui, J., Taboada, C., 2008. Introducing deficit irrigation to stabilize yields of quinoa (Chenopodium quinoa Willd.). EUR J AGRON 28, 427-436.

[21] Gerber, P.J., Carsjens, G.J., Pak-uthai, T., Robinson, T.P., 2008. Decision support for spatially targeted livestock policies: Diverse examples from Uganda and Thailand. AGR SYST 96, 37-51.

[22] Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. SCIENCE 327, 812.

[23] Gómez-Candón, D., De Castro, A.I., López-Granados, F., 2014. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. PRECIS AGRIC 15, 44-56.

[24] Han, E., Ines, A.V.M., Baethgen, W.E., 2017. Climate-Agriculture-Modeling and Decision Tool (CAMDT): A software framework for climate risk management in agriculture. ENVIRON MODELL SOFTW 95, 102-114.

[25] Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. NAT CLIM CHANGE 3, 816-821.

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

[27] Huang, X., Yu, C., Fang, J., Huang, G., Ni, S., Hall, J., Zorn, C., Huang, X., Zhang, W., 2018. A dynamic agricultural prediction system for large-scale drought assessment on the Sunway TaihuLight supercomputer. COMPUT ELECTRON AGR 154, 400-410.

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

[29] Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data

assimilation of remote sensing and crop models. EUR J AGRON 92, 141-152.

[30] Kaini, P., Artita, K., Nicklow, J.W., 2012. Optimizing Structural Best Management Practices Using SWAT and Genetic Algorithm to Improve Water Quality Goals. WATER RESOUR MANAG 26, 1827-1845.

[31] Kwon, H., Lall, U., 2016. A copula-based nonstationary frequency analysis for the 2012–2015 drought in California. WATER RESOUR RES 52, 5662-5675.

[32] Li, C., Frolking, S., Frolking, T.A., 1992. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. Journal of Geophysical Research: Atmospheres 97, 9759-9776.

[33] Li, C., Zhuang, Y., Cao, M., Crill, P., Dai, Z., Frolking, S., Moore, B., Salas, W., Song, W., Wang, X., 2001. Comparing a process-based agro-ecosystem model to the IPCC methodology for developing a national inventory of N2O emissions from arable lands in China. NUTR CYCL AGROECOSYS 60, 159-175.

[34] Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., Lobell, D.B., Martre, P., Ruane, A.C., Wallach, D., Jones, J.W., Rosenzweig, C., Aggarwal, P.K., Alderman, P.D., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A., Deryng, D., Sanctis, G.D., Doltra, J., Fereres, E., Folberth, C., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Kimball, B.A., Koehler, A., Kumar, S.N., Nendel, C., O Leary, G.J., Olesen, J.E., Ottman, M.J., Palosuo, T., Prasad, P.V.V., Priesack, E., Pugh, T.A.M., Reynolds, M., Rezaei, E.E., Rötter, R.P., Schmid, E., Semenov, M.A., Shcherbak, I., Stehfest, E., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wall, G.W., Wang, E., White, J.W., Wolf, J., Zhao, Z., Zhu, Y., 2016. Similar estimates of temperature impacts on global wheat yield by three independent methods. NAT CLIM CHANGE 6, 1130-1136.

[35] Miao, Y., Mulla, D.J., Batchelor, W.D., Paz, J.O., Robert, P.C., Wiebers, M., 2006. Evaluating management zone optimal nitrogen rates with a crop growth model. AGRON J 98, 545-553.

[36] Müller, C., Elliott, J., Chryssanthacopoulos, J., Deryng, D., Folberth, C., Pugh, T.A.M., Schmid, E., 2015. Implications of climate mitigation for future agricultural production. ENVIRON RES LETT 10, 125004.

[37] Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. The Computer Journal 7, 308-313.

[38] Nguyen, D.C.H., Ascough, J.C., Maier, H.R., Dandy, G.C., Andales, A.A., 2017. Optimization of irrigation scheduling using ant colony algorithms and an advanced cropping system model. ENVIRON MODELL SOFTW 97, 32-45.

[39] Paerl, H.W., Xu, H., McCarthy, M.J., Zhu, G., Qin, B., Li, Y., Gardner, W.S., 2011. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. WATER RES 45, 1973-1983.

[40] Qiu, J., Li, H., Wang, L., Tang, H., Li, C., Van Ranst, E., 2011. GIS-model based estimation of nitrogen leaching from croplands of China. NUTR CYCL AGROECOSYS 90, 243-252.

[41] Rauff, K.O., Bello, R., 2015. A review of crop growth simulation models as tools for agricultural meteorology. Agricultural Sciences 6, 1098.

[42] Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proceedings of the National Academy of Sciences 111, 3268.

[43] Rurinda, J., Zingore, S., Jibrin, J.M., Balemi, T., Masuki, K., Andersson, J.A., Pampolino, M.F., Mohammed, I., Mutegi, J., Kamara, A.Y., Vanlauwe, B., Craufurd, P.Q., 2020. Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa. AGR SYST 180, 102790.

[44] Schultz, B., Thatte, C.D., Labhsetwar, V.K., 2005. Irrigation and drainage. Main contributors to global food production. IRRIG DRAIN 54, 263-278.

[45] Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. WATER RESOUR RES 29, 1185-1194.

[46] Stewart, W.M., Dibb, D.W., Johnston, A.E., Smyth, T.J., 2005. The Contribution of Commercial Fertilizer Nutrients to Food Production. AGRON J 97, 1-6.

[47] Tilman, D., 1999. Global environmental impacts of agricultural expansion: The need for sustainable and efficient practices. Proceedings of the National Academy of Sciences 96, 5995.

[48] Ueyama, M., Tahara, N., Iwata, H., Euskirchen, E.S., Ikawa, H., Kobayashi, H., Nagano, H., Nakai, T., Harazono, Y., 2016. Optimization of a biochemical model with eddy covariance measurements in black spruce forests of Alaska for estimating CO2 fertilization effects. AGR FOREST METEOROL 222, 98-111.

[49] Uzoma, K.C., Smith, W., Grant, B., Desjardins, R.L., Gao, X., Hanis, K., Tenuta, M., Goglio, P., Li, C., 2015. Assessing the effects of agricultural management on nitrous oxide emissions using flux measurements and the DNDC model. Agriculture, Ecosystems & Environment 206, 71-83.

[50] Vital, J., Gaurut, M., Lardy, R., Viovy, N., Soussana, J., Bellocchi, G., Martin, R., 2013. High-performance computing for climate change impact studies with the Pasture Simulation model. COMPUT ELECTRON AGR 98, 131-135.

[51] Wang, N., Zhang, N., Wang, M., 2006. Wireless sensors in agriculture and food industry—Recent development and future perspective. COMPUT ELECTRON AGR 50, 1-14.

[52] Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. J HYDROL 358, 1-23.

[53] Yu, C., Huang, X., Chen, H., Godfray, H.C.J., Wright, J.S., Hall, J.W., Gong, P., Ni, S., Qiao, S., Huang, G., Xiao, Y., Zhang, J., Feng, Z., Ju, X., Ciais, P., Stenseth, N.C., Hessen, D.O., Sun, Z., Yu, L., Cai, W., Fu, H., Huang, X., Zhang, C., Liu, H., Taylor, J., 2019. Managing nitrogen to restore water quality in China. NATURE 567, 516-520.

[54] Yu, C., Huang, X., Chen, H., Huang, G., Ni, S., Wright, J.S., Hall, J., Ciais, P., Zhang, J., Xiao, Y., Sun, Z., Wang, X., Yu, L., 2018. Assessing the Impacts of Extreme Agricultural Droughts in China Under Climate and Socioeconomic Changes. Earth's Future 6, 689-703.

[55] Yu, C., Li, C., Xin, Q., Chen, H., Zhang, J., Zhang, F., Li, X., Clinton, N., Huang, X., Yue, Y., Gong, P., 2014. Dynamic assessment of the impact of drought on agricultural yield and scale-dependent return periods over large geographic regions. ENVIRON MODELL SOFTW 62, 454-464.

[56] Zhang, C., Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. PRECIS AGRIC 13, 693-712.

[57] Zhang, N., Wang, M., Wang, N., 2002. Precision agriculture—a worldwide overview. COMPUT ELECTRON AGR 36, 113-132.

[58] Zhang, Q., Li, H., 2007. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. IEEE T EVOLUT COMPUT 11, 712-731.

[59] Zhao, G., Bryan, B.A., King, D., Luo, Z., Wang, E., Bende-Michl, U., Song, X., Yu, Q., 2013. Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. ENVIRON MODELL SOFTW 41, 231-238.