EXECUTIVE SUMMARY

The ultimate goal of our project is to improve the predictability of regional and global crop yields by integrating advanced remote-sensing and process-based modeling. We have developed the CLM–APSIM model framework, which combines the strengths of Earth system models and agronomy crop models. We conducted parameter sensitivity analysis and spatially explicit parameter optimization using satellite-based constraints and ran a set of historical simulation and future projection experiments. For remote sensing, we developed a multisource satellite data fusion algorithm (STAIR, or SaTellite DAta IntegRation) that can generate daily cloud-free high-resolution surface reflectance images. We have also developed a machine-learning framework that ingests a time series of high-resolution satellite data for field-level in-season crop-type identification. The combined advancements in crop modeling and remote sensing will allow us to develop a seasonal forecasting system for crop productivity in the U.S. Corn Belt.

RESEARCH CHALLENGE

Global food security is under continuing pressure from increased population and climate change [5–8]. Crop productivity forecasting offers a promising approach for addressing food-related early warning and decision-making. Crop models, either statistical or process-based, are the most essential component in any crop productivity forecasting system. Compared to statistical models, process-based crop models are better tools for studying the impacts of historical and future climate on regional and global food production, for assessing the effectiveness of possible adaptations and their potential feedback to climate, and for attributing different pathways through which climate can impact crop yields. Nevertheless, current process-based crop models still have large uncertainties when considering the demands from a forecasting perspective. Moreover, an in-season field-level crop-type classification product is also missing and remains a challenge for seasonal forecasting of crop productivity.

METHODS & CODES

We are targeting better prediction performance for crop yield in the U.S. Corn Belt through combining advanced process-based modeling and remote-sensing observations. We first improved the crop growth representation in one of the leading Earth System Models, the Community Earth System Model (CESM) [2]. We combined the strengths of the Community Land Model (CLM) [4], the land component of CESM, and the state-of-the-art agronomy crop model APSIM in the CLM–APSIM crop model. We coupled the SALib Python package [1] with the CESM by using its multi-instance configuration for parameter sensitivity analysis, and coupled the PyDREAM Python package [3] with the CESM for Bayesian parameter calibration of the CLM–APSIM model. We are also developing a remote-sensing monitoring system for U.S. croplands, which includes two major components: the satellite data preprocessing and fusion component, and the machine learning-based classifier component. The data preprocessing component is responsible for harmonizing the multisource satellite images, including data encoding, geo-projection, geo-registration, and quality control of the images. The multisource satellite data fusion component uses our newly developed STAIR algorithm to generate cloud/gap-free high-resolution (daily and <=30m) surface reflectance images. For the machine-learning-based classifier component, we have implemented a machine-learning framework that can ingest time series high-resolution satellite images to provide in-season crop-type information. All of these workflows were developed on Blue Waters.

RESULTS & IMPACT

The CLM–APSIM combines the strengths of both the CLM version 4.5 and the APSIM models (Peng, et al., 2018, Fig. 1). An evaluation of results at the AmeriFlux sites located in the U.S. Corn Belt show that the CLM–APSIM model performs better than the original CLM4.5 in simulating phenology, surface fluxes, and especially biomass partition and maize yield. The CLM–APSIM model corrects a serious deficiency in CLM4.5 related to underestimation of belowground biomass (i.e., overestimation of aboveground biomass) and overestimation of the Harvest Index, which leads to a reasonable yield estimation with the wrong mechanisms. We are conducting parameter sensitivity analysis and spatially explicit parameter optimization using satellite-based constraints for the CLM–APSIM model. We are also running a set of historical simulation and future projection experiments aimed at disentangling the contribution of different mechanisms to high-temperature impact on crop yield.

This method is computationally efficient and ready to be scaled up to continental scales. It is also sufficiently generic to easily include various optical satellite data for fusion. We envision this novel algorithm can provide the effective means to leverage historical optical satellite data to build long-term daily, 30-m surface reflectance records (e.g., from 2000 to the present) at continental scales for various applications. In addition, it will produce operational near-real-time daily and high-resolution data for future earth observation applications. The machine learning-based crop-type and area-mapping system has also been tested for Champaign County, Illinois (Cai, et al., 2018). Specifically, we used the U.S. Department of Agriculture’s (USDA) Common Land Units (CLUs) to aggregate spectral information for each field based on a time-series Landsat image data stack. This largely overcomes the cloud contamination issue while exploiting a machine-learning model based on Deep Neural Network (DNN) and high-performance computing for intelligent and scalable computation of classification processes. We used a total of 1,322 Landsat multitemporal scenes including all the spectral bands from 2000 to 2015. Our results show the inclusion of temporal phenology information and evenly distributed spatial training samples in the study domain improve classification performance.Benchmarked with USDA’s Crop Data Layer, our algorithm shows a relatively high overall accuracy of 96% for classifying corn and soybeans across all CLU fields in Champaign County from 2000 to 2015. Furthermore, our approach achieved 95% overall accuracy for classifying corn and soybeans by late July of each year.

WHY BLUE WATERS

Blue Waters is essential for our research since other resources, such as those available from XSEDE, are not suitable for our project considering the petabyte-level storage demand, data availability, and intensive I/O and computational demands.

PUBLICATIONS & DATA SETS