UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

MODELLING TERRESTRIAL CARBON FLUXES AND CROP PRODUCTION WITH REMOTE SENSING AND *IN-SITU* OBSERVATIONS

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By

XIAOCUI WU
Norman, Oklahoma
2020

MODELLING TERRESTRIAL CARBON FLUXES AND CROP PRODUCTION WITH REMOTE SENSING AND *IN-SITU* OBSERVATIONS

A THESIS APPROVED FOR THE

DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Xiangming Xiao, Chair

Dr. Jeffrey Basara

Dr. Heather McCarthy

Dr. Lara Souza

Dr. Jean Steiner

This dissertation is dedicated to my families for their continuous understanding, unconditional love and support, to my advisor Xiangming Xiao, for his patience, encouragement, and support over past years.

Acknowledgements

Throughout completing of this dissertation, I have received a lot of support and assistance. First, I really appreciate my advisor, Dr. Xiangming Xiao, for his patience, encouragement, and guidance during the five years. Without his continuous instruction and support, it is impossible for me to complete my Ph.D. study with so many difficulties.

I would like to express my thanks to EOMF families for their mental and intellectual support. Dr. Yao Zhang taught me everything at the very beginning when I joined the lab. Without his help, I can't learn the R/python language so quickly, and I can't get familiar with vegetation photosynthesis model simulation and solar induced fluorescence dataset in a quick way. Dr. Jie Wang provided me a lot of great suggestions for my life, and also valuable comments on organizing scientific papers. Dr. Yuanwei Qin was helpful in obtaining EOMF resources and provided me useful feedbacks for my Chapter 3 and 4. Dr. Russell Doughty taught me a lot of background information about TROPOMI SIF and how to process the data. Dr. Rajen Bajgain provided me everything about the eddy covariance systems in the field. Dr. Zhenhua Zou and Ling Du shared their experience to deal with all those stuffs related to graduation. During the past five years, I enjoyed all the good time with EOMF families when hanging out for food, discussing life and research.

I would like to thank other members of my advisory committee, Prof. Jeffrey Basara, Pro. Heather McCarty, Prof. Jean Steiner and Prof. Lara Souza for their instruction and help in my study over last five years.

I would like to express my heartfelt thanks to my beloved family for their continuous understanding and unconditional love all through these years. My parents, my elder sister and my younger brother gave me the most precious freedom to pursue my dream. I also owe my heartfelt thanks to my friends for their support and caring. Dr. Mingzhu He shared me a lot of living and working experience with me at my first three years. My roommates brought me a lot of joy when living together.

Bingbing Zhou, my husband, is always my sunshine. His passion in scientific research and persistence in overcoming difficulties has encouraged and comforted me a lot during the past years. I wish he would be happy and successful after the Ph.D. study and in the long run of his life.

Table of Contents

List of Figures

List of Tables

Abstract

Plants fix carbon through photosynthesis, sequestering carbon dioxide from the atmosphere and substantially mitigating the climate warming effect induced by anthropogenic CO_2 emissions. Terrestrial gross primary production (GPP) through photosynthesis is crucial for understanding the land-atmospheric carbon exchange, which is the largest component and one of the most uncertain aspects of the global carbon cycle. Thus, accurate estimation of GPP can help better understand the global carbon budget, and the ecosystem sensitivity to the global climate change. Data driven models that utilize the climate data and remote sensing-based observations can provide reasonable estimates of GPP. The emergence of the solar induced chlorophyll fluorescence (SIF) from both *in-situ* and satellite observations provides another tool to understand and estimate plant photosynthesis. Remote sensing-based observations and models are also widely used in crop monitoring. Timely and accurate crop production estimation are needed to sustain global food security under the background of climate change. My overall objective is to improve the data-driven models to provide better GPP estimates, to combine SIF with other data sources to advance our understanding of the photosynthesis process and ecosystem sensitivity to droughts, and to investigate the potential of a data-driven model, specifically, the vegetation photosynthesis model (VPM), in crop monitoring.

In Chapter 2, we investigated the seasonal dynamics of eddy flux-derived GPP (GPP_{EC}), solar-induced chlorophyll fluorescence (SIF), and four vegetation indices (VIs) and their relationships in a tall grassland site. We also examined drought impact on those structural and physiological proxies of plant photosynthesis. We found SIF explained 49% of the GPP variability at the seasonal scale, and had a stronger consistency with GPP than

the four VIs. Among the four VIs, the soil background corrected VIs, near-infrared reflectance of vegetation (NIRv) and enhanced vegetation index (EVI) showed the best consistency with both GPP and SIF. In addition, SIF is more sensitive to drought than the VIs. This study suggested that the potential of SIF in tracking photosynthesis in grassland and drought impact on photosynthesis.

In Chapter 3, we improved the vegetation photosynthesis model (VPM) by considering the difference of the maximum light use efficiency for C_3 and C_4 croplands. Model validation against GPP_{EC} in multiple sites distributed over CONUS suggests better accuracy of GPP simulated by VPM (GPP_{VPM}) in tracking the cross-site variability and interannual variability ($R^2 = 0.84$ and 0.46, respectively) when compared to MOD17 GPP. We also assessed the spatial and temporal (seasonal) consistency of GPP_{VPM}, MOD17 GPP and other two common-used GPP products with the Global Ozone Monitoring Experiment-2 (GOME-2) SIF. We found good consistency of GPP_{VPM} with SIF across space and time. Anomaly analyses for those GPP products and GOME-2 SIF showed that high GPP during the 2012 spring compensated for low GPP during the summer, resulting in near-neutral changes in annual GPP for the CONUS. This study demonstrates the need to improve light use efficiency models by incorporating C_3 and C_4 plant functional types, and the importance of assessing the impacts of different types and timing of climate extremes on GPP.

In Chapter 4, we evaluated the interannual dynamics of maize and soybean cropping areas and production over the CONUS during 2008-2018. We found an increase of maize and soybean planted areas during this period, mainly driven by markets and international trade. We further investigated the relationship between production and the total GPP derived from VPM simulation and the area statistics from cropland data layer (CDL). We found strong relationship between them. Additionally, the ratio between grain production and GPP_{VPM}, named harvest index derived from GPP (HI_{GPP}), ranged from 0.25 (2012) to 0.36 for maize and from 0.13 to 0.15 for soybean. The seasonal linear regression models between grain production and cumulative GPP_{VPM} (GPP_{VPM_acc}) over time at 8-day resolution shows that GPP_{VPM_acc} accounted for ~90% variance of grain production by the end of July. Our findings suggest that the potential of VPM and GPP_{VPM} data product in monitoring grain production to help farmers, decision makers, stakeholders and the public.

In Chapter 5, we assessed the consistency of winter wheat cropped areas from NASS statistics and planted area derived from CDL data. We found strong spatial-temporal consistency between the CDL and NASS datasets for the planted acreage. However, there is a significant difference (>40%) between the NASS harvested area and planted area from NASS statistics or CDL data in the Southern Great Plains, where winter wheat is used as both a grain crop and forage crop. We then evaluated the relationship between annual GPP_{VPM} and grain production, we found a moderate linear relationship between them for winter wheat, with R^2 value of 0.68 at county scale over the CONUS. But this relationship improved when excluding those counties with large difference between the CDL planted area and NASS harvested area. Our findings suggest that the importance of providing reliable in-season crop type classification in considering the difference between planted and harvested acreage, and the potential of GPP_{VPM} in crop monitoring.

Chapter 1: Introduction

1.1 Research background

Stabilization of atmospheric carbon dioxide (CO_2) concentrations is required to reduce the increasing impacts of climate changes. Terrestrial ecosystems play a major role in the global carbon cycle, offsetting approximately 25-30% of the carbon dioxide (CO_2) emitted by human activities since 1950s (Le Quéré et al. 2009). However, it is projected that there will be an increase of the duration or/and frequency of climate extremes in the future, which could potentially affected the composition, structure and functioning of terrestrial ecosystems and how they will counteract/offset atmospheric CO_2 concentrations (Frank et al. 2015). It is a grand challenge to understand and project the response of terrestrial ecosystems to climate extremes (Reichstein et al. 2013).

To quantify ecosystem carbon fluxes under current and future climates, particularly the role of terrestrial vegetation, scientists have used the eddy-covariance flux towers since the early 1990s (Baldocchi et al. 2001) to provide continuous measurements of carbon and water fluxes between the terrestrial ecosystems and the atmosphere. At present, over 500 flux tower sites are operating on a long-term basis around the globe (https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=9). These sites are located across different climate zones and biomes (Baldocchi et al. 2001; Baldocchi 2003), and provide probably the best estimates of ecosystem-level carbon fluxes. The flux sites measure directly the net ecosystem exchange (NEE), which can be separated into two major components: GPP and ecosystem respiration (ER) (Lasslop et al. 2010; Reichstein et al. 2005). However, the flux towers have a limited footprint of 100-2000 m (Schmid et al. 2000), making it challenging to assess the surface carbon fluxes at regional or global scale.

The emergence of solar-induced chlorophyll (SIF) observations provide another way to assess plant photosynthesis (GPP) at both regional and global scales. SIF is a small amount of light generated during the photosynthetic process. It is well known that photosynthesis involves two main sets of reactions: the light reactions and the carbon fixation reactions. During the light reactions, the light absorbed by the chlorophyll follows one of three pathways: used for photosynthesis, re-emitted at longer wavelength as SIF, and dissipated as heat (Maxwell and Johnson 2000; Porcar-Castell et al. 2014). Thus, SIF can provide information about how the photosystem is using the energy absorbed by chlorophyll. As this part of energy using by the photosystem is indicative of the overall rate of photosynthesis under many conditions, SIF observations has the potential to help to infer plant photosynthesis. Previous studies have shown a quasi-linear relationship between SIF and GPP from both the satellite-based (Guan et al. 2016; Joiner et al. 2013; Sun et al. 2017; Zhang et al. 2016a), and ground-based SIF measurements (Miao et al. 2020; Yang et al. 2015). Multiple applications have used this empirical relationship to infer GPP from SIF directly or indirectly (Guan et al. 2016; Guanter et al. 2014). However, the underlying mechanisms behind the SIF and GPP relationship at different scales and various environmental limitations still requires elucidation. Miao et al. (2018) found the SIF and GPP relationship could change substantially in a day due to the light change. Wu et al. (2020) found a stronger linear relationship between GPP and SIF at a long time scale (monthly) rather than short time scale (hourly).

In addition to flux observations, scientists also have developed various models to estimate surface fluxes. Two approaches have been widely employed to investigate the spatial and temporal variability in GPP using remotely sensed data: (1) process-based models, and (2) data-driven models (Anav et al. 2015). The former is mostly based on the mechanistic description of the photosynthetic biochemical processes and scales the instantaneous leaf-level biochemical model (Farquhar et al. 1980) to the canopy level using multiple scaling approaches. A number of process-based models have been successfully applied to simulate GPP at regional and global scales by utilizing remotely-sensed vegetation parameters, such as leaf area index (LAI), as inputs. However, the application of these process-based models is limited by the model complexity and uncertainty of their parameterization (Li et al. 2018). In contrast, data driven models, one major category is light use efficiency (LUE) models, were developed based on the assumption that plant productivity is linearly related to the amount of the absorbed photosynthetically active radiation (APAR) (Monteith 1972, 1977). Those models have been reported to quantify the spatial and temporal variation of GPP reasonably well in both site- and regional- level studies (Potter et al. 1993; Running et al. 2004; Yuan et al. 2007; Zhang et al. 2017) . The linear relationship works particularly well at a long temporal scale (weekly or monthly) rather than at a short time scale (hourly, daily), due to the uneven light distribution in the canopy and other environmental limitation at short temporal scale (Song et al. 2013; Wu et al. 2015).

Compared with process-based models, LUE models simplified many complicated biochemical processes and just have a few parameters and environmental drivers to represent different environmental conditions. One key issue is that parameterization of LUE models affects the simulation results a lot. Specifically, the maximum light use efficiency (ε_{max}), the most sensitive parameters in the LUE models, requires more careful representation for the LUE models, especially for C_4 croplands (Xin et al. 2013; Yuan et al. 2007).

Accurate quantification of crop production at regional to global scales is important in supporting policy- and decision-making in agriculture (Guan et al. 2016; Lesk et al. 2016). Annual production of one specific crop in a given region is the product of crop's yield and harvested area (Cohn et al. 2016). Thus, to quantify crop production for certain crops, it requires both the area and yield information. The area information can be obtained through remote-sensing based mapping, though detailed high-resolution crop maps are lacking for the entire globe. For yield information, two common approaches with satellite inputs have been widely used: (1) empirical relationships between vegetation biomass and remote sensing spectral vegetation indices to estimate yields (Lobell and Ortiz-Monasterio 2008; Zhao et al. 2015); (2) yield estimated as the product of crop gross primary productivity (GPP) and an empirical crop-specific harvest index (HI) (Guan et al. 2016; He et al. 2018b; Lobell et al. 2002). The first approach is fundamentally simple but the empirical relationship is specific to the limited areas and conditions from which they were developed, and should not be extended to other areas. The second approach depends on the accuracy of GPP products and HI.

1.2 Research objectives

The objectives of this dissertation are to explore the relationship between SIF and GPP using both *in-situ* measurement and remote sensing data, and to assess how drought affect GPP and SIF, and to apply the regional GPP products to estimate crop production.

1.3 Organization of the dissertation

This dissertation consists of one introductory chapter, four main chapters, and one summary chapter. Chapters 3 has been published in a peer-reviewed journal. Chapter 4 is under the second round of review after major revision in Agricultural and Forest Meteorology. Chapter 5 has been circulated to coauthors for internal revision and will soon be submitted. Chapter 2 remains in preparation for journal submission.

***Chapter 2.* Divergence of seasonal and diurnal variability of plant photosynthesis detected by physiological and structural proxies in a tall grass prairie ecosystem.**

Many studies have shown that GPP and SIF have a quasi-linear relationship from both satellite- and ground- based SIF measurements, and APAR is the dominant factor of this relationship. However, those studies are mainly focused on cropland ecosystems, and a few forest sites. Moreover, there are few studies that explore drought impacts with in-situ SIF data. In this chapter, I explored seasonal dynamics and relationships of GPP, SIF and four vegetation indices at seasonal and diurnal scale. I also assessed the drought impact on GPP, SIF and VIs. This study will help us to better understand the GPP-SIF relationship and the drought response in a grassland site.

***Chapter 3.* Divergent impacts of spring warming and summer drought on gross primary production across the Contiguous United States in 2012.**

Wolf et al. (2016) found a compensatory effect of spring warming and summer drought on the net ecosystem production. However, the spring warming and summer drought effect on the GPP is not well understood. In this study, we improved the VPM model by incorporating different LUE for C_3 and C_4 croplands in CONUS, and compared the simulation result with three common GPP products and SIF to show the consistency between these models.

<u>Chapter 4.</u> Spatial-temporal dynamics of maize and soybean planted area, harvested area, gross primary production, and grain production in the Contiguous United States during 2008-2018.

Crop production is not well studied compared with crop yield, one important reason is the lack of detailed crop area data. The cropland data layer (CDL) provided by the USDA NASS provides us this kind of area information. Evaluating corn and soybean in the CONUS as a case study, we estimated their production from CDL area data and VPM GPP using the NASS production as a benchmark.

<u>Chapter 5.</u> Spatial-temporal dynamics of winter wheat croplands in the Contiguous United States: Consistency between agricultural statistical and satellite approaches.

Winter wheat is a major staple crop grown in the US, and the US is the third largest wheat exporter globally. Timely and reliable winter wheat production prediction in the US is important for regional and global food security. In this study, I utilize the VPM GPP data, CDL derived planted area to monitor winter wheat production with NASS crop acreages and grain production as a benchmark.

1.4 List of publications from the Dissertation

Chapter 2

Wu, X., Xiao, X., Yang, Z., et al. Diurnal and seasonal dynamics of vegetation indices, SIF and GPP in a tallgrass prairie ecosystem. *In prep.*

Chapter 3

Wu, X., Xiao, X., Zhang, Y., He, W., Wolf, S., Chen, J., He, M., Gough, C.M., Qin, Y., Zhou, Y. and Doughty, R., 2018. Spatiotemporal consistency of four gross primary production products and solar-induced chlorophyll fluorescence in response to climate extremes across CONUS in 2012. Journal of Geophysical Research: Biogeosciences, 123(10), pp.3140-3161.

Chapter 4

Wu, X., Xiao, X., Yang, Z., et al. Spatio-temporal dynamics of maize and soybean croplands in the Contiguous United States: Consistency between agricultural statistical and satellite approaches. Agricultural and Forest Meteorology, *in review after major revision*.

Chapter 5

Wu, X., Xiao, X., Yang, Z., et al. Spatial-temporal dynamics of winter wheat croplands in the Contiguous United States: Consistency between agricultural statistical and satellite approaches. *To be submitted*

Chapter 2: Diurnal and seasonal dynamics of vegetation indices, SIF and GPP in a tallgrass prairie ecosystem

Abstract

Remote sensing of far-red solar induced chlorophyll fluorescence (SIF) has emerged as an important tool for studying gross primary production (GPP) at field scale. However, the relationship between SIF and GPP has only been investigated on multiple cropland sites and a few forest sites. In addition, it is unknown how SIF responds to drought at field scale. In this study, using data from an integrated SIF measurement system together with eddy flux and meteorological observations from a tallgrass prairie site in Oklahoma, we investigated the seasonal dynamics and relationships of SIF, GPP and four vegetation indices. We found that SIF correlated with GPP derived from eddy covariance (GPP_{EC}) at seasonal scales (R^2 = 0.49). Among the four vegetation indices, the near-infrared reflectance of vegetation (NIR_v) and enhanced vegetation index (EVI) were most significantly correlated with both GPP and SIF than normalized vegetation indices (NDVI) and MERIS terrestrial vegetation index (MTCI). During the drought period, the physiological-based proxies of photosynthesis, including SIF and GPP, are more sensitive than the structural-based vegetation indices. The results of this study highlighted the importance of using SIF to track plant photosynthesis and the physiological and structural limitation during drought.

2.1 Introduction

Plant chlorophyll absorbs sunlight in the 400-700 nm range of the electromagnetic spectrum (photosynthetically active radiation), and emits a small amount of light in the red and far-red (650-850 nm) range known as solar-induced chlorophyll fluorescence (SIF) (Porcar-Castell et al. 2014). This phenomenon provides an opportunity to estimate gross primary productivity (GPP), the total carbon fixed by plants. The recent advances in measuring SIF from both sensors onboard satellites (Guanter et al. 2014; Joiner et al. 2013; Sun et al. 2017; Zhang et al. 2016a) and in-situ sensors (Miao et al. 2020; Yang et al. 2017; Yang et al. 2015) promotes the use of SIF to estimate GPP across different spatial and temporal scales.

Previous studies have explored the empirical relationship between GPP and SIF (Guanter et al. 2014; Joiner et al. 2013; Sun et al. 2017; Zhang et al. 2016a) at regional scale. Some field studies have explored the mechanistic understanding of this relationship at site scale, mostly in cropland sites (Dechant et al. 2020; Miao et al. 2020; Miao et al. 2018; Wu et al. 2020) and forest sites (Yang et al. 2015). Based on the light use efficiency framework, GPP can be thought as a product of the amount of photosynthetically active radiation (PAR), the fraction of PAR absorbed by the canopy (fPAR) and the light use efficiency (LUE) (Monteith 1972, 1977), while SIF can be thought of a product of PAR, fPAR, fluorescence yield ($\mathrm{SIF_{yield}}$) and escaping ratio (Guanter et al. 2014).

$$\mathrm{GPP} = \mathrm{APAR} \times \mathrm{LUE} \qquad (2.1)$$

$$\mathrm{SIF} = \mathrm{PAR} \times \mathrm{fPAR} \times \mathrm{SIF_{yield}} \times \mathrm{f_{esc}} \qquad (2.2)$$

$$\mathrm{APAR} = \mathrm{fPAR} \times \mathrm{PAR} \qquad (2.3)$$

where $\mathrm{SIF_{yield}}$ is the emitted SIF per photon absorbed. $\mathrm{f_{esc}}$ is the proportion of observed top-of-canopy SIF to the total emitted SIF. f_{esc} is usually affected by the canopy structure and the light quality (diffused vs. direct radiation). The relationship between instantaneous GPP and SIF, is thus controlled by the relationship between instantaneous

LUE and $SIF_{yield} \times f_{esc}$. At a longer time scale, the ratio of GPP: SIF is jointly determined by the variations of APAR and the ratio of LUE: $SIF_{yield} \times f_{esc}$. However, as the variations of APAR are usually larger than the variations of LUE: SIF_{yield}, previous studies showed that APAR is the dominant factor for the linear relationship between GPP and SIF (Miao et al. 2018; Yang et al. 2015). However, it is still unknown whether this linear relationship is a universal one at regional scale (Xiao et al. 2019). Considering the difference of canopy structure over biomes, it is necessary to explore such relationship for different biomes.

Previous studies have shown the drought effects on photosynthesis can be attributed to canopy limitation and physiological change (Zhang et al. 2016b). The former is caused by the decrease in leaf area or the chlorophyll pigments; the latter is related to stomatal (stomatal closure to prevent water loss) or non-stomatal (e.g., the reduction in enzyme activity) limitation. Vegetation indices may be able to track the canopy changes, but most remotely sensed indicators are not sensitive to physiological limitation. However, SIF and GPP are sensitive to such physiological limitation. Therefore, a combination of VIs, SIF and GPP can provide a comprehensive understanding about drought effects on photosynthesis.

The Southern Great Plains (SGP), composed of Kansas, Oklahoma, and Texas, experiences frequent droughts, which provides the opportunity to investigate the effect of drought on plant photosynthesis. In this study, we set up a state-of-the-art SIF observation system (FluoSpec), together with an eddy covariance flux tower and climate measurements, to investigate the diurnal and seasonal dynamics of vegetation indices, SIF and GPP and their responses to drought in a tall grass prairie. Specifically, we addressed three questions: (1) what is the seasonal and diurnal dynamics of VIs (NDVI, EVI, MTCI, NIRv), SIF and GPP? (2) what is the relationship between the indicators? and (3) how do Vis, SIF, and GPP respond to drought?

2.2 Materials and methods

2.2.1 Study area

Our experiment site was Kessler Atmospheric and Ecological Field Station (KAEFS)in central Oklahoma, USA, approximately 28 km southwest of the University of Oklahoma (https://www.ou.edu/kaefs) (Fig. 2.1). The ecosystem in KAEFS is a mixed tallgrass prairie, woodland, riparian areas, and pastures. The grass species includes both C_3 and C_4 grasses, and the proportion changes within the growing season, with C_3 grasses dominating during the early and late growing season (cool months), and C_4 grasses dominating the mid-growing season (May to July) (Castillioni et al. 2020). The region has a sub-humid climate, with the annual precipitation around 850 mm and the mean annual air temperature around 16℃.

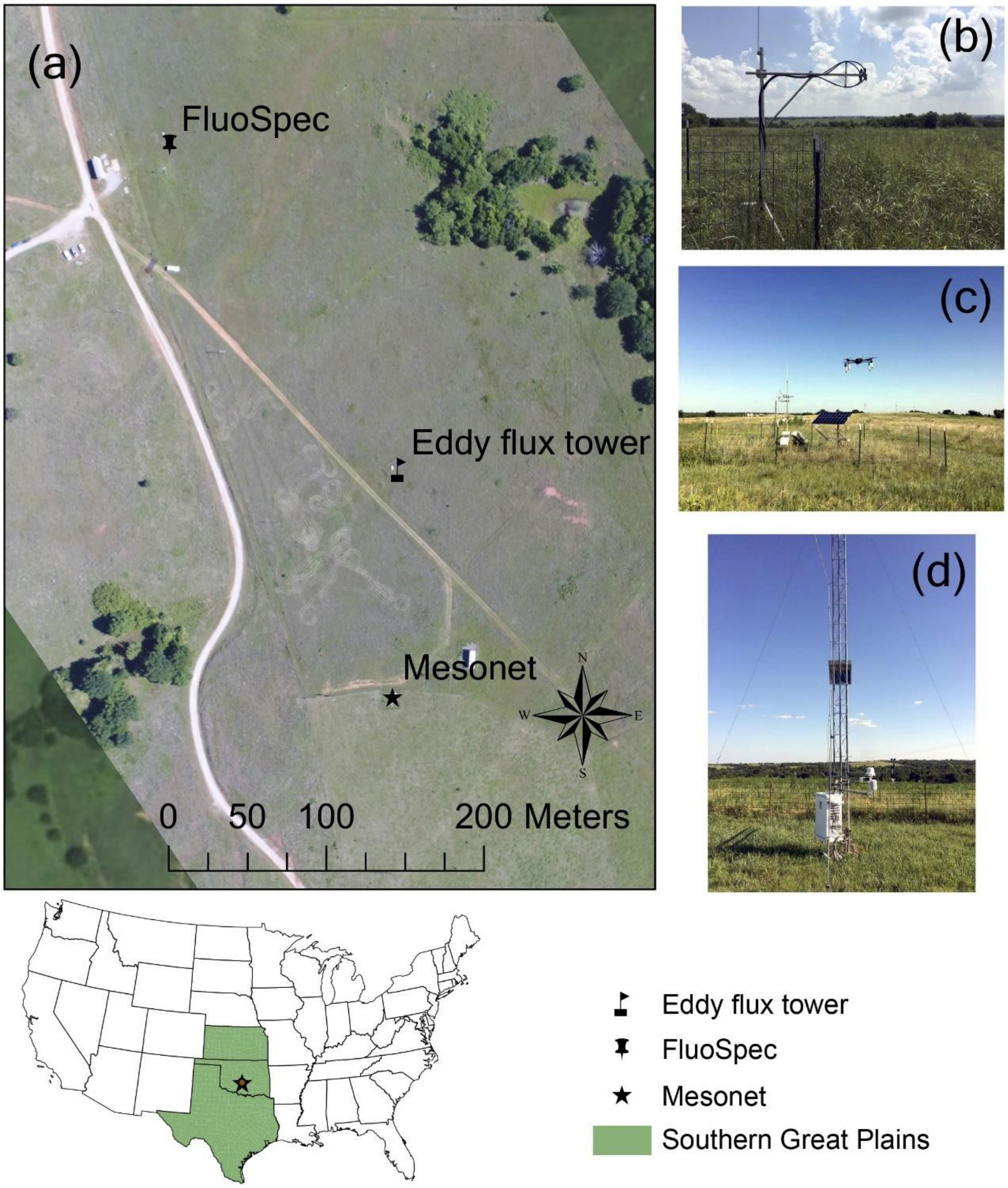

Figure 2.1 The location of the study sites, including the FluoSpec solar-induced chlorophyll fluorescence measurement site, an eddy covariance flux site and an Oklahoma Mesonet site.

2.2.2 FluoSpec system

A continuous canopy-level solar induced chlorophyll fluorescence measurement system --- FluoSpec, was deployed at KAEFS. The key components of the system are two high resolution spectrometers. One spectrometer is QE-Pro, targeted for SIF retrieval, with a spectral resolution of around ~0.04 nm (full width half maximum, FWHM) between 730 nm to 788 nm. The other spectrometer is HR2000+, with a spectral resolution of ~0.5 nm between 187 nm to 1105 nm. We used the spectral measurement from the HR2000+ to calculate several vegetation indices as followings:

$$\mathrm{NDVI} = \frac{\rho_{860}-\rho_{655}}{\rho_{860}+\rho_{655}} \tag{2.4}$$

$$\mathrm{EVI} = 2.5\frac{\rho_{860}-\rho_{655}}{\rho_{860}+6\rho_{655}-7.5\rho_{470}+1} \tag{2.5}$$

$$\mathrm{MTCI} = \frac{\rho_{750}-\rho_{710}}{\rho_{710}-\rho_{680}} \tag{2.6}$$

$$\mathrm{NIRv} = \frac{\rho_{860}-\rho_{655}}{\rho_{860}+\rho_{655}} \times \rho_{860} \tag{2.7}$$

where NDVI is the normalized vegetation index, EVI is the enhanced vegetation index, MTCI is the Merris terrestrial chlorophyll index, and $\mathrm{NIR_v}$ is the near-infrared reflectance of vegetation. NDVI is believed to be related to canopy structure and chemical content, while EVI and MTCI is sensitive to the variation of chlorophyll content (Zhang et al. 2018). $\mathrm{NIR_v}$ introduced the near-infrared band to eliminate the effect of soil background and is found to be closely related to the f_{esc} (Zeng et al. 2019). ρ with a numeric subscript represents the vegetation canopy reflectance at that wavelength.

For each spectrometer, there are two optical fibers and one shutter (FOS-2×2-TTL, OceanOptics, Inc.) to switch the light path, so the spectrometer can measure the downward solar irradiance and upward vegetation radiance in turns. The optical fiber facing up to the sky is integrated with a opaline glass cosine corrector (CC-3, OceanOptics, Inc.) to get the downwelling irradiance of the sun; the other bare optical fiber pointing down to the land surface has a field of view (FOV) of 25° at the height of

~3 m. The land surface within the FOV of the bare fiber, about 1.3m × 1.3m, is dominated by switchgrass (>80%).

Both spectrometers are controlled by a small computer called Raspberry Pi 3 (RP3). Every 15 minutes, the system completed a measurement cycle. In each 15-minutes measurement cycle, two spectrometers operate in turns to get 5 pairs of irradiance and radiance measurements. The system collected data during from March 13 – September 17, 2017, from 6:00 am – 20:00 pm. Careful radiometric calibration was performed prior to and one time during the field campaign using a light source (HL-2000-CAL, OceanOptics, Inc). We also used a white board to calibrate the system every two months during the growing season.

2.2.3 Hyperspectral data processing

The output data from the FluoSpec are digit number (DN) values, rather than absolute radiance or irradiance. To convert DN values to irradiance/radiance, all measurements were applied for dark current and a non-linearity correction (Yang et al. 2015). After getting the absolute irradiance and radiance, a spectral fitting methods (SFM) is used to extract SIF by exploiting the oxygen absorption band at 760 nm (Meroni et al. 2009).

2.2.4 Weather measurements from Mesonet and eddy covariance (EC) flux tower site

The Mesonet station provided a variety of weather variables, such as air temperature, solar radiation, precipitation, vapor pressure deficit, and soil moisture. To identify droughts, a water index called the fractional water index (FWI) was applied. FWI is a unitless soil water indicator ranging from 0 to 1, with 0.00 representing very dry soil to 1.00 for soil at field capacity (Schneider et al. 2003). We calculated FWI at 5cm, 25cm, and 60cm soil depth based on measurements from three Campbell Scientific 229 - L sensors at 5cm, 25cm, and 60cm depth, so it can cover most of the root zone for the grasslands (Illston et al. 2008). We identified two drought events during the 2017 growing season (Fig. 2.2, Table 2.1).

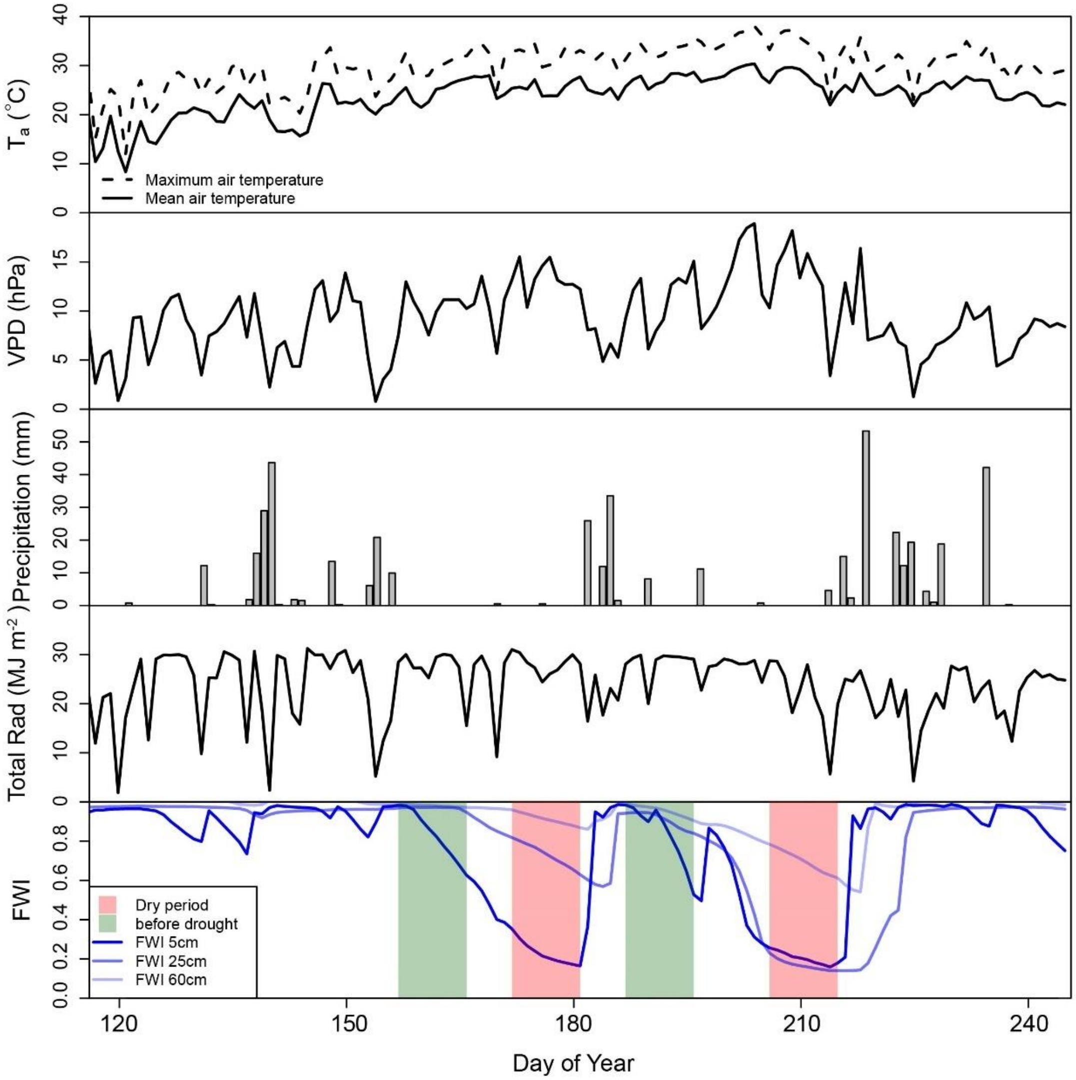

Figure 2.2 The seasonal dynamics of weather variables (air temperature: Ta, Vapor pressure deficit: VPD, precipitation, solar radiation) and fractional water index (FWI) at different depth (5cm, 25cm, 65cm) during the plant growing season.

Table 2.1 Summary of the four periods during the peak growing season

	Pre-drought period 1 (P1)	Dry period 1 (D1)	Pre-drought period 2 (P2)	Dry period 2 (D2)
Start (DOY)	156	170	188	205
End (DOY)	165	179	197	214
Average Tmax (℃)	30.1	32.1	33.6	33.4
Average VPD (hPa)	10.2	13.2	10.8	12.7
Average FWI	0.93	0.64	0.87	0.36

2.2.5. Carbon flux data processing

The EC flux data were processed using a standardized protocol: (1) the raw 10 Hz data were processed with the EddyPro software from Li-Cor (LI-COR, Inc.). (2) the output from the first step underwent an initial data quality check where the upper and lower 1% Net Ecosystem Exchange (NEE), and Latent Energy (LE) observations were rejected to remove noises. (3) after the data quality check, the 30-min flux values were combined with the weather observations and gap-filled and partitioned using the "REddyProc" package (https://www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWebRPackage) in R.

2.3 Results

2.3.1 Seasonal dynamics of vegetation indices, SIF, and GPP and their relationships

Tower-based daily SIF gradually increased in March and peaked to ~0.6 mW m^{-2} sr^{-1} nm^{-1} during the summer, then declined by the end of the growing season (Fig. 2.3). Interestingly, daily GPP_{EC} showed a different pattern than daily SIF. GPP_{EC} showed a similar seasonal cycle with SIF from March to early August, but it increased during August to September, while SIF decreased, which may be caused by the growth of C_3 grasses during the late growing season. The seasonal variability of VIs had a similar pattern with SIF, increasing from March, peaking in the summer and declining from August (~DOY 200).

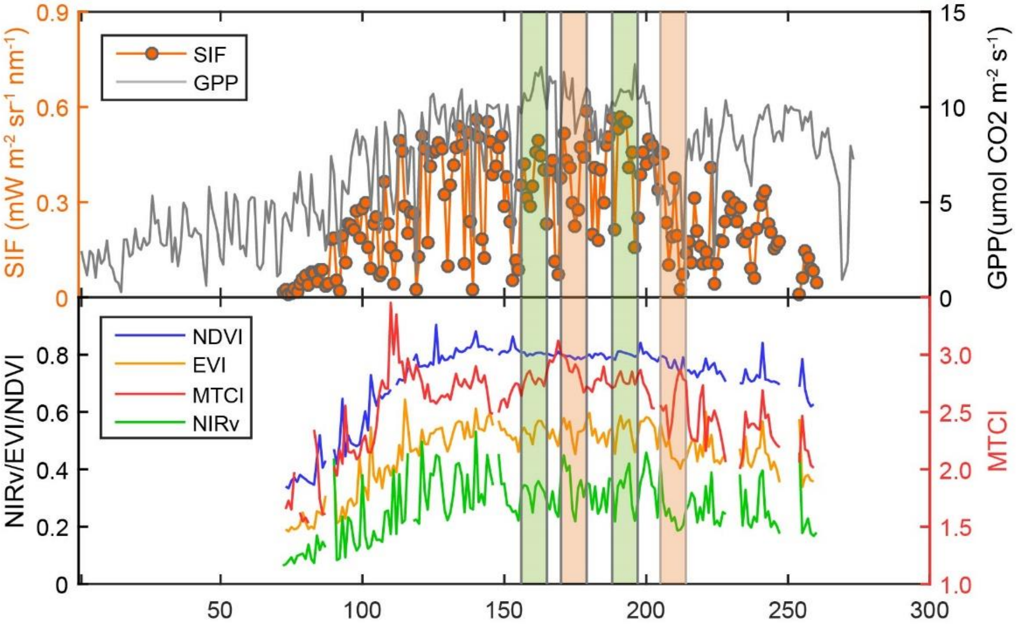

Figure 2.1 Seasonal dynamics of GPP, SIF and four vegetation indices, including NDVI, EVI, NIRv and MTCI.

The seasonal dynamics (DOY 72 to 260) of SIF agrees reasonably well with that of daily GPP_{EC} (R^2 = 0.53) (Fig. 2.4). The seasonal pattern of VIs showed a less tightly relationship with daily GPP_{EC} (Fig. 2.5), compared with that of GPP and SIF. EVI and NIRv showed a more significant relationship than MTCI and NDVI. For the relationship of VIs and SIF, the R^2 ranges from 0.20 to 53, with the tightest relationship for NIRv (0.53) and EVI (0.45), following by NDVI and MTCI.

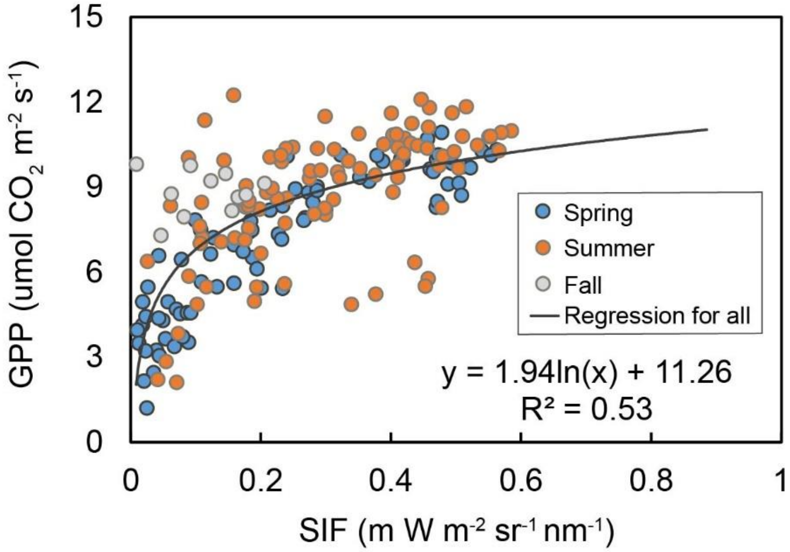

Figure 2.2 Linear regression results between daily SIF and GPP in 2017 (with p-value < 0.05)

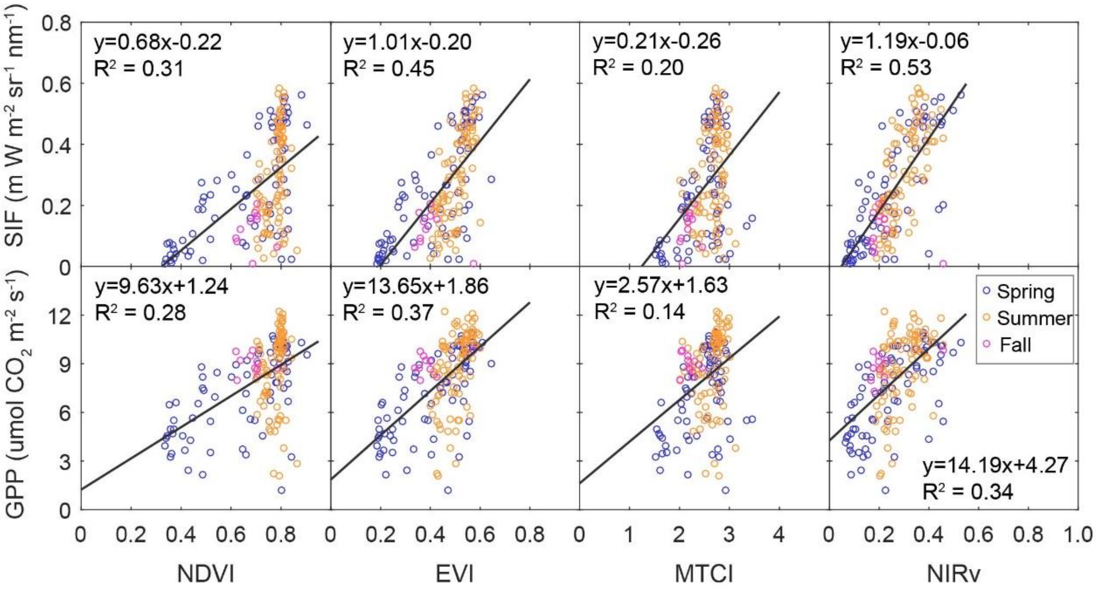

Figure 2.3 Linear regression results between SIF/GPP and the four VIs (NDVI, EVI, NIRv and MTCI), with all p-values<0.05

2.3.2 Diurnal dynamics of vegetation indices, SIF, and GPP and drought response

At the diurnal scale, both SIF and GPP showed a cosine variation, with a peak value at noon time (Fig. 2.6). During the first drought period, both SIF and GPP were minimally affected, with a similar value to the before-drought period. During the second drought period, both GPP and SIF showed significant decreases. the product of SIFyield × fPAR and LUE x fPAR showed a “U” shaped pattern, with higher value at early morning and late afternoon and lower value in the afternoon. Similarly, both SIFyield × fPAR and LUE × fPAR were minimally affected during the first drought, but decreased during the second, more severe, drought period. The decrease was stronger in the midday than in the morning and late afternoon.

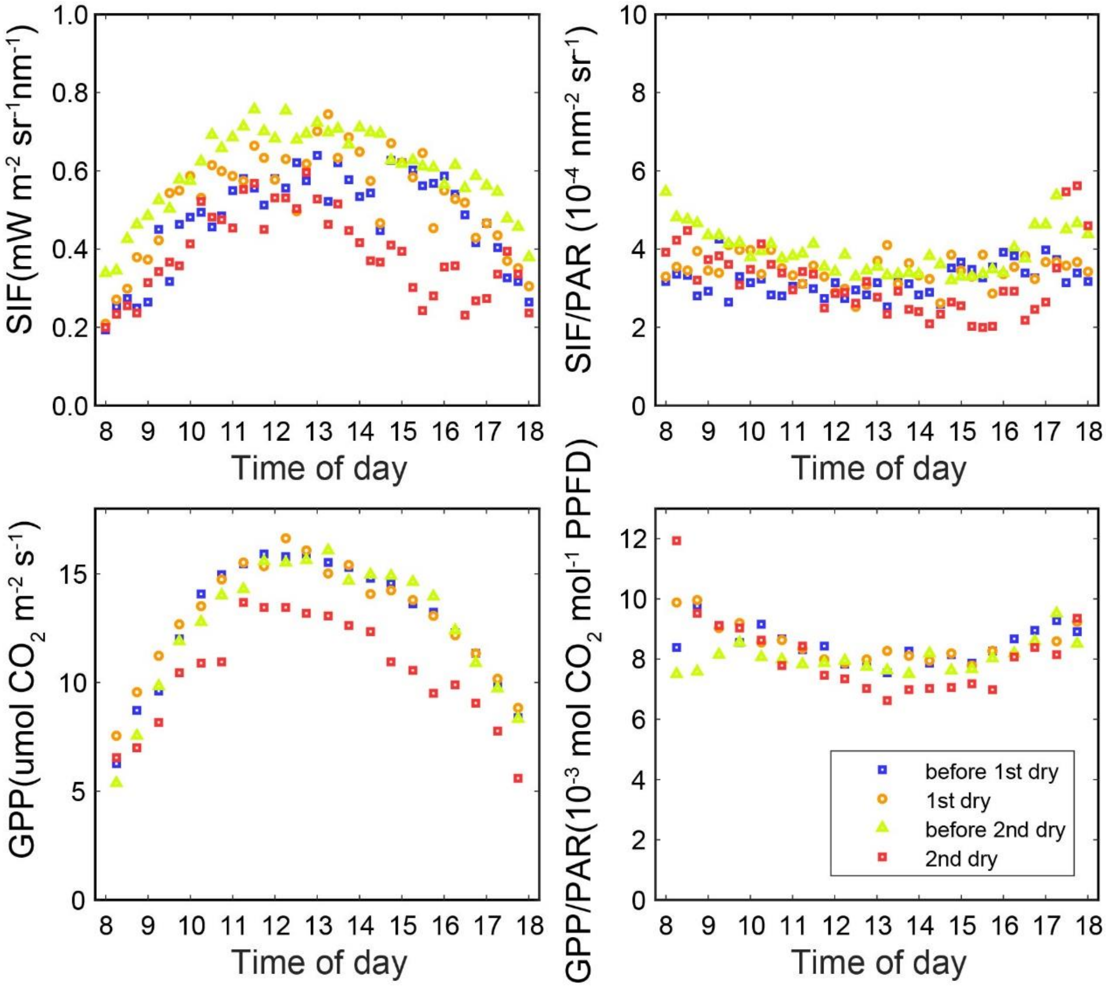

Figure 2.4 Diurnal dynamics (8:00 am-18:00 pm) of SIF, SIF/PAR, GPP, and GPP/PAR during the four identified periods.

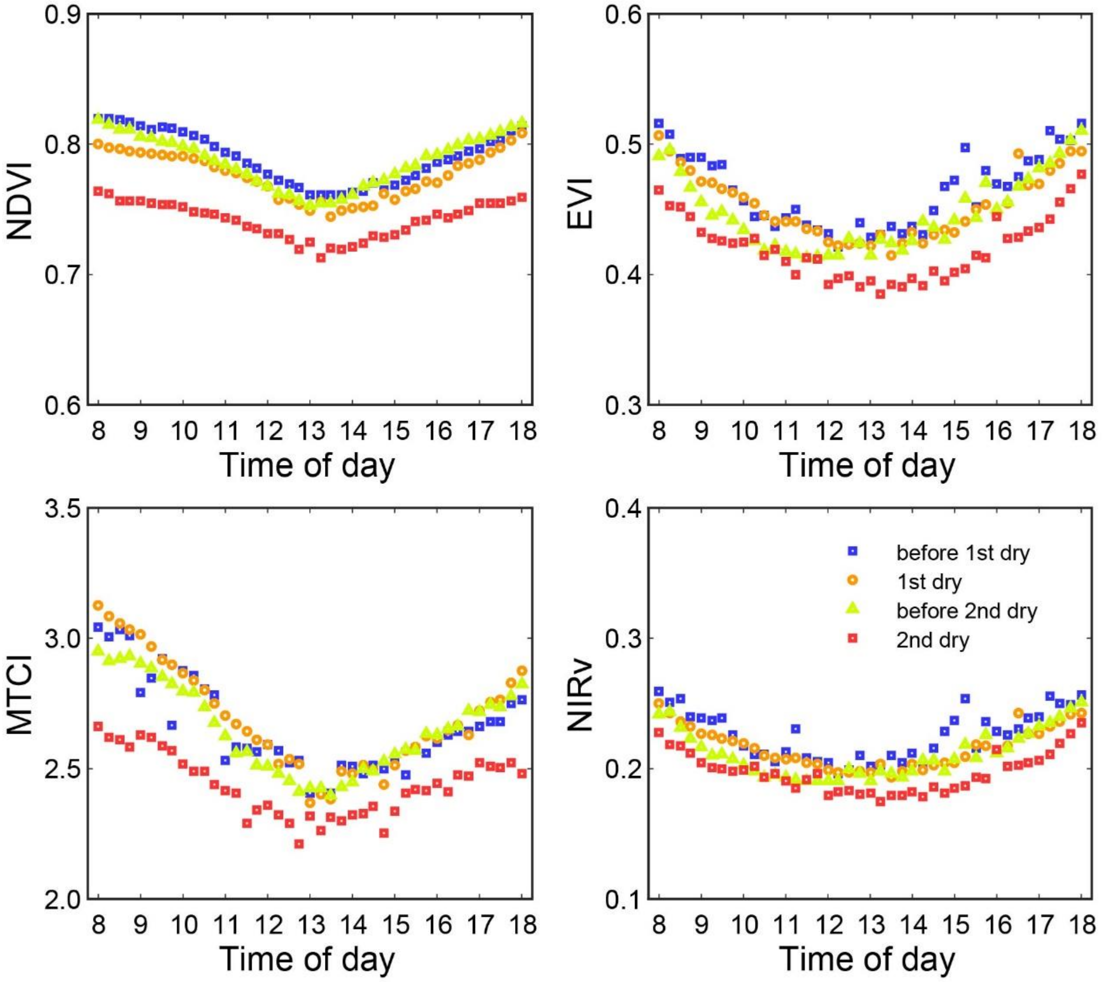

Figure 2.5 Diurnal dynamics (8:00am-18:00 pm) of NDVI, EVI, MTCI and NIRv during the four identified periods.

All the VIs showed a “U” shaped pattern at diurnal scale (Fig. 2.7). Similarly, all the VIs were minimally affected by the first drought, indicated by values close to the pre-drought values. During the second drought period, all the VIs showed a significant decrease. However, compared to the SIF and GPP reduction in the second drought, all the VIs still showed a less reduction (Fig. 2.8).

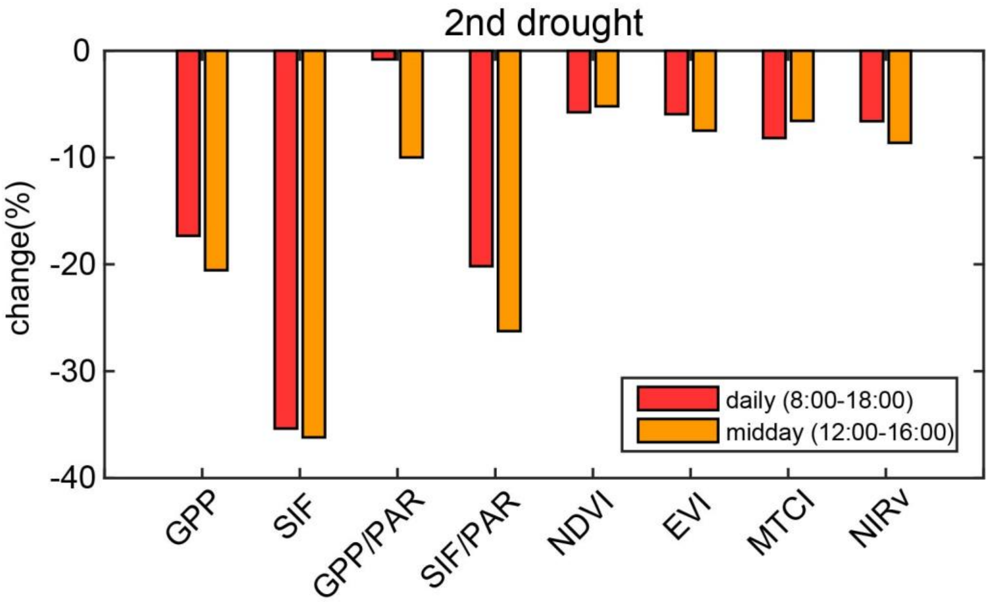

Figure 2.6 Comparison between SIF, GPP and VIs reduction during second drought period to pre-drought period.

2.4 Discussion

2.4.1 The relationships between VIs, SIF and GPP

In this study, we found a moderate nonlinear relationship between GPP and SIF at daily scale. The inconsistency of GPP and SIF could be mainly caused by the different footprints of the two observations, with the GPP site covering a diverse mixed grasses and forbs over hundreds of meters, while the SIF site having less diverse grasses (mainly C_4 grasses) in a much smaller field of view. As the composition of grass species (C_3 and C_4) within the footprint of the EC tower site, can shift during the entire growing season, we found a regrowth period in the late growing season from the GPP data (Fig. 2.3). However, the underlying mechanisms associated with the spatiotemporal variability of SIF and GPP:SIF relationship are still largely unknown. We do know that chlorophyll pigments absorb photons that power photosynthesis, and some of photons are reemitted at a longer wavelength as SIF. Thus, SIF is considered to physiologically link with GPP. However, the relationship between SIF and GPP is complicated because the observed SIF at the top of canopy is only a fraction of total emitted SIF and this fraction is sensitive to the direction of observation and canopy structure, including canopy leaf area index (LAI), clumping index, and leaf orientation (Dechant et al. 2020). GPP, unlike SIF, is related to canopy-level gas exchange processes, thus the observed GPP at the top of canopy is the cumulative GPP of all leaves, which is not affected by the canopy structure and the direction of observation. So the observed pattern of SIF should have some difference with that of GPP theoretically. Numerous studies showed a linear GPP:SIF relationship mainly in croplands (Li et al. 2020; Liu et al. 2020; Miao et al. 2020; Miao et al. 2018) with fewer in forest sites (Yang et al. 2015), with a R^2 ranging from 0.28 to 0.80. Our study showed a moderate relationship between SIF and GPP in the mixed grasses, with a R^2 value of 0.49 for linear regression and 0.53 for nonlinear regression. However, there is still not a clear understanding of the relative contributions of physiological and structural components of SIF relate to photosynthetic light use efficiency over different biomes. Therefore, decoupling the effects of canopy structure and physiological changes on SIF is still needed.

Both NIRv and EVI showed a stronger correlation with SIF (GPP) than NDVI and MTCI (Fig. 2.5). EVI reduces the effects of environmental factors such as atmospheric conditions and soil background through introducing the blue band and addressing the non-linear, differential band transferring process (Huete et al. 1997). Previous studies have documented strong EVI-GPP relationships and thus EVI has been widely used modelling GPP (Wu et al. 2014; Zhang et al. 2017). Compared to NDVI, NIRv minimized the effects of soil background by introducing the NIR band (Zeng et al. 2019). Previous studies reported that NIRv could explain over 90% variability of GPP at monthly, global scales (citation). At site scale, the relationship between GPP and NIRv varied from sub-daily to daily scales, with a R^2 ranging from 0.4-0.7, which is similar for the relationship between GPP and EVI (Wu et al. 2020). In our results, NIRv did not have a clear advantage over EVI in correlation to GPP, but NIRv has a stronger relationship with SIF than EVI. NIRv is closely related to the fraction of escaping photons (f_{esc}) through scattering and absorbing in the canopy and thus provides an ideal bridge for linking observed top-of-canopy SIF and total emitted SIF (Zeng et al. 2019; Dechant et al. 2020).

2.4.2 Response of physiological and structural proxies to drought

Drought impacts plant photosynthesis from both physiological control and canopy control (Zhang et al. 2016b). Physiological control includes a reduction of enzyme activity and stomatal conductance (Reichstein et al. 2002), which is often related to the variation of air temperature, and vapor pressure deficit and soil water deficit. Canopy control is more visible, indicated by the changes of canopy such as the decrease of leaf area index, wilting, leaf curling, and loss of chlorophyll indicated by leaves turning yellow. Previous studies reported that SIF is sensitive to physiological variations during drought at a regional scale (Chen et al. 2019; Sun et al. 2015). The SIF and hyperspectral observations in this tall grass site and the drought events in 2017 provides an unique opportunity to investigate the drought impact on plant photosynthesis at site scale.

Physiological and canopy controls of GPP take place across differing temporal scales. Physiological control can occur rapidly, from minutes to days. In our study, the first drought had limited affect on photosynthesis from either physiological and canopy control. Figure 2.2 indicated that only the shallowest soil layer had dried during the first drought period. During the second drought period, both the surface and deep soil water declined, which may lead to both the physiological and canopy response indicated by GPP, SIF, and VIs. However, the VIs only showed a slightly decrease in response to drought compared with the strong decrease of GPP and SIF. As the canopy change is reflected by the change of pigment or the change of total leaf area index and the VIs indicate small change of the LAI, we suspected that the canopy change is mainly driven by the change of pigment rather than the leaf quantity. It is interesting to see a strong decrease of SIF/PAR (fPAR × SIFyield) in the second drought, indicating that the drought response was mostly physiologically since the SIF/PAR sustained a pretty high level in the early morning but dropped dramatically in the afternoon when the strongest environmental stress occurred.

2.5 Conclusion

SIF has proved to be a good proxy of photosynthesis at both regional and site scales. In this study, by comparing the SIF and vegetation indices derived from the FluoSpec system with GPP_{EC}, we found that SIF can track the seasonal dynamic of GPP tbetter han the structural-based VIs. Among the four VIs, the soil background adjusted VIs, NIRv and EVI, were more consistent with GPP and SIF than NDVI and MTCI. Canopy SIF also showed higher sensitivity to drought stress than those structural based VIs by including both physiological and structural limitation information on photosynthesis. The study indicates the potential of using SIF to track photosynthesis in grassland and monitor drought impact on photosynthesis.

Chapter 3: Spatiotemporal consistency of four gross primary production products and solar-induced chlorophyll fluorescence in response to climate extremes across CONUS in 2012

Abstract

Large spatial-scale effects of climate extremes on gross primary production (GPP), the largest terrestrial carbon flux, are highly uncertain even as these extremes increase in frequency and extent. Here we report the impacts of spring warming and summer drought in 2012 on GPP across the contiguous US (CONUS) using estimates from four GPP models: Vegetation Photosynthesis Model (VPM), MOD17A2H V006, Carnegie-Ames-Stanford Approach (CASA), and Simple Biosphere/Carnegie-Ames-Stanford Approach (SiBCASA). VPM simulations are driven by Moderate Resolution Imaging Spectroradiometer (MODIS), North American Regional Reanalysis (NARR) climate data, and C_3 and C_4 cropland maps from the United States Department of Agriculture (USDA) Cropland Data Layer (CDL) dataset. Across 25 eddy covariance flux tower sites, GPP estimates from VPM (GPP_{VPM}) showed better accuracy in terms of cross-site variability and interannual variability ($R^2 = 0.84$ and 0.46, respectively) when compared to MOD17 GPP. We further assessed the spatial and temporal (seasonal) consistency between GPP products and the Global Ozone Monitoring Experiment-2 (GOME-2) solar-induced chlorophyll fluorescence (SIF) over CONUS during 2008-2014. The results suggested that GPP_{VPM} agrees best with SIF across space and time, capturing seasonal dynamics and interannual variations. Anomaly analyses showed that increased GPP during the spring compensated for the reduced GPP during the summer, resulting in near-neutral changes in annual GPP for the CONUS. This study demonstrates the importance of

assessing the impacts of different types and timing of climate extremes on GPP, and the need to improve light use efficiency models by incorporating C_3 and C_4 plant functional types.

3.1 Introduction

Terrestrial ecosystems play a major role in the global carbon cycle, offsetting approximately 25-30% of the CO_2 emitted by human activities since the 1950s (Le Quéré et al. 2009). Gross primary production (GPP), the amount of CO_2 sequestered by vegetation through photosynthetic assimilation before accounting for respiratory losses, is the largest component of the global terrestrial carbon flux (Beer et al. 2010). Therefore, a small fluctuation in GPP could have significant impact on atmospheric CO_2 concentrations. However, the composition, structure, and functioning of terrestrial ecosystems are expected to be substantially altered by increases in the duration or/and frequency of climate extremes such as droughts, heatwaves, or intense precipitation events (Frank et al. 2015). It is a major challenge to understand and project the response of terrestrial ecosystems to climate extremes (Reichstein et al. 2013). In particular, droughts, together with the frequently co-occurring heatwaves, are among the most widespread natural disasters, and could have large impacts on annual GPP, ecosystem respiration, and net carbon balance (Frank et al. 2015; van der Molen et al. 2011).

The impacts of climate extremes, especially heatwaves and droughts, on GPP have been thoroughly investigated for selected events (Ciais et al. 2005; Parazoo et al. 2015; Wolf et al. 2016; Yuan et al. 2016a). However, how climate extremes affect the carbon cycle is still poorly known at the landscape, regional, and global scales (Pan and Schimel

2016). To investigate the impacts of climate extremes on GPP at ecosystem and landscape scales, three approaches have been separately or jointly applied: eddy covariance (EC) flux tower measurements (von Buttlar et al. 2017; Welp et al. 2007), remote-sensing data (Hilker et al. 2014), and biogeochemical models (Zscheischler et al. 2014). Since the 1990s, the EC flux tower method has provided directly observed evidence for the seasonal changes of terrestrial carbon fluxes, which increases our understanding of the underlying mechanisms of terrestrial ecosystem responses and their feedbacks to climate extremes at the site scale (Reichstein et al. 2007). However, in-situ EC sites are limited by their relatively moderate-size footprints of observation and the number and distribution of FLUXNET sites are limited, making it difficult to assess the impacts of climate extremes on the carbon cycle at regional, continental, and global scales. The GPP data derived from eddy covariance flux towers (GPP_{EC}), though limited in their spatial coverage, are currently the best available data to validate GPP estimates from process-based and data-driven GPP models. In contrast, optical and microwave remote-sensing data provide larger scale insights into the vegetation structure, including leaf area index, and light absorption by canopy (Chen 1996; Disney et al. 2006; Ollinger 2011). Recently, solar-induced chlorophyll fluorescence (SIF) data have been derived from satellite-based observations to estimate GPP, as it is tightly linked with photosynthesis (Frankenberg et al. 2011; Porcar-Castell et al. 2014). However, SIF has a very weak signal and accounts for about 2% of the total light absorbed by vegetation. Satellite retrieved SIF measurements have comparatively large amounts of noise, and the recent SIF data products are often aggregated in temporal and spatial domains resulting in a coarse spatial and temporal resolution (monthly, 0.5°× 0.5° for Global Ozone Monitoring Instrument 2, GOME-2) (Joiner et al. 2013). The coarse spatial resolution of SIF data products limits its application because 0.5^{o} gridcells (~50 km at Equator) are often highly heterogeneous. A final approach uses terrestrial biosphere models to estimate GPP and ecosystem respiration for a variety of ecosystems at multiple scales. However, the reliability of these models is constrained by input datasets, model parameters, and model structures (Schaefer et al. 2012; Schwalm et al. 2010). Hence, a synthesis and comparison of the different approaches can reveal the shortcomings of individual approaches, and help to reach a more reliable assessment of the multiple-scale responses of ecosystems to climate extremes (Pan and Schimel 2016).

In 2012, the Contiguous United States (CONUS) experienced an abnormally warm spring and dry summer (Hoerling et al. 2014; Knutson et al. 2013). Record-breaking temperatures were observed across 34 states during spring and a severe summer drought followed, especially across the Great Plains and the Midwest Corn Belt. The 2012 US drought was reported as one of the worst droughts since 1988 and had a comparable magnitude and spatial extent of those during the 1930s and 1950s (Hoerling et al. 2014; Rippey 2015). Impacts of this spring warming and summer droughts on terrestrial carbon fluxes in CONUS have been investigated, using the data from eddy covariance flux tower sites, GPP from the MOD17 data product, and net ecosystem production (NEP) from CarbonTracker (CTE2014 and CTE2015) (Wolf et al. 2016). They found that the losses of NEP in the summer were offset by an unusually large increase of NEP in spring, resulting in a small gain of annual NEP over CONUS (0.11 Pg C). They also reported that the decrease in GPP during summer was much larger than the increase of spring GPP, resulting in a moderate loss of annual GPP (-0.38 Pg C) over CONUS in 2012. However, there are large uncertainties among the various GPP products (Schaefer et al. 2012); for example, the MOD17 GPP product has large uncertainties in croplands (Turner et al. 2006; Xin et al. 2015). Therefore, there is a need to evaluate various GPP models and their GPP data products, which will help us to better understand and assess GPP responses to spring warming and summer drought in 2012.

In this study, we analyzed GPP data products from four GPP models: (1) the Vegetation Photosynthesis Model (VPM) (Xiao et al. 2004a; Xiao et al. 2004b), which has been well validated at both site (Dong et al. 2015; Doughty et al. 2018; Jin et al. 2013; Wagle et al. 2015) and regional scales (Zhang et al. 2016a; Zhang et al. 2017) in previous studies. In this study, we modified the model for cropland by separating C_3 and C_4 crops with detailed Cropland Data Layer data; (2) MOD17 (Running et al. 2004), which is also used to evaluate the 2012 spring warming and summer drought impact on GPP in Wolf et al. (2016); (3) SiBCASA-GFED4 (van der Velde et al. 2014), and (4) CASA-GFED3 (van der Werf et al. 2006; van der Werf et al. 2010). SiBCASA-GFED4 and CASA-GFED3 models are biosphere models used in CarbonTracker Europe (CTE2014) (van der Laan-Luijkx et al. 2017) and CarbonTracker (CT2014) (Peters et al. 2007), respectively, which provided the prior biosphere carbon fluxes (NEP, GPP - Respiration) in the two carbon tracker systems. We evaluated the GPP estimations from the four datasets with in-situ GPP data from eddy covariance flux tower sites and SIF data from GOME-2. The objectives of this study are threefold: (1) to demonstrate the potential of differentiating C_3 and C_4 croplands for improving GPP estimates (using VPM as an example) and validate the GPP estimates against FLUXNET data; (2) to quantify and understand the spatial-temporal consistency of GOME-2 SIF data and GPP estimates from various models; and (3) to assess the impacts of spring warming and summer drought on GPP at the pixel, biome, and continental scales.

3.2 Materials and methods

3.2.1 Study area

We used the VPM model (Xiao et al. 2004a; Xiao et al. 2004b) to estimate GPP from 2008 to 2014 over CONUS. We followed the original model framework but further differentiated between C_3 and C_4 croplands, since C_3 and C_4 crops have different maximum light use efficiencies (ε_{max}). The National Agricultural Statistics Service (NASS) Cropland Data Layers (CDL) from the United States Department of Agriculture (USDA) were used to calculate the area percentages of C_3 and C_4 croplands within each 500 m pixel over individual years (Boryan et al. 2011). According to the USDA report, the major C_4 crop-types included corn, sorghum, sugarcane, and millet, and other crop-types were considered as C_3 croplands. The GPP of each pixel was estimated by area-weighted averaged GPP (Equation 1), which was derived from area fraction maps of C_3 and C_4 croplands and MCD12Q1 land use datasets:

$$GPP = [(fC_3 \times \varepsilon_{max\text{-}C3} + fC_4 \times \varepsilon_{max\text{-}C4}) \times T_{scalar} \times W_{scalar}] \times APAR_{chl} \qquad (3.1)$$

where fC_3 and fC_4 were the area fraction of C_3 and C_4 crops inside each cropland pixel, respectively. $APAR_{chl}$ is photosynthetic active radiation (PAR) absorbed by chlorophyll in the canopy and is estimated from enhanced vegetation index (EVI) (Huete et al. 1997) as following:

$$APAR\_chl = 1.25 \times (EVI - 0.1) \qquad (3.2)$$

This equation was modified from the previous model version (Xiao et al. 2004a; Xiao et al. 2004b), and has been applied in generating a global GPP product (Zhang et al. 2017). The coefficients 0.1 and 1.25 were used to adjust for sparsely vegetated or barren land and have been evaluated using the solar-induced chlorophyll fluorescence data.

The maximum light used efficiency values for C_3 croplands ($\varepsilon_{max\text{-}C3}$) and C_4 croplands ($\varepsilon_{max\text{-}C4}$) were specified as 0.035 mol CO_2 mol^{-1} PAR (~1.8 g C m^{-2} day^{-1} MJ^{-1} PAR), and 0.053 mol CO_2 mol^{-1} (PAR) (~ 2.7 g C m^{-2} day^{-1} MJ^{-1} (PAR)) (1.5 times larger than C_3 types), respectively (Li et al. 2013). T_{scalar} and W_{scalar} are the temperature and water regulation factor and calculated as:

$$T_{scalar}=\frac{(T-T_{min})(T-T_{max})}{[(T-T_{min})(T-T_{max})]-(T-T_{opt})^2} \tag{3.3}$$

$$W_{scalar}=\frac{1+LSWI}{1+LSWI_{max}} \tag{3.4}$$

where T is the air temperature, derived from the NCEP/NARR climate data. T_{min}, T_{max}, and T_{opt} represent the minimum, maximum, and optimum temperatures for photosynthesis, respectively, which are biome-specific and assigned values as in Zhang et al. (2016a). $LSWI_{max}$ is the maximum land surface water index within the plant growing season, and we applied a temporal smoothing method using nearby four years (two years before and two years after) to eliminate potential bias (Zhang et al. 2017).

3.2.2 Input datasets for VPM simulations in CONUS during 2008–2014

Regional simulations of VPM model require climate, vegetation indices, and land cover data. Here we briefly describe the input datasets used: (1) NCEP/NARR reanalysis meteorological data, (2) MODIS surface reflectance and land cover data, and (3) NASS CDL data.

The NCEP/NARR data was downloaded from (http://www.esrl.noaa.gov/psd). It contains meteorological variables such as air temperature, precipitation, and downward shortwave radiation from 1979 to present at a spatial resolution of 32 km and a temporal resolution of 3 hours. The original 3-hourly NARR data were aggregated into daily data by calculating the maximum, mean, and minimum air temperature in a day (^{o}C), and the cumulative sum of downward shortwave radiation in a day. The resulting daily data were further aggregated to 8-day intervals (following the MODIS 8-day temporal resolution) by calculating the maximum, mean, and minimum temperature (^{o}C), and the cumulative sum of downward shortwave radiation within an 8-day period. We also interpolated these climate variables (32-km spatial resolution) to 500-m using the same algorithm reported in a previous publication (Zhang et al. 2016a). As previous studies have shown, the NARR downward shortwave radiation is systematically overestimated, so we adjusted it by applying a correction factor of 0.8 as proposed in a previous study (Jin et al. 2015).

The latest version of MODIS surface reflectance product, MOD09A1 V006, was used to calculate EVI (Huete et al. 1997) and Land Surface Water Index (LSWI) (Xiao et al. 2004b). A temporal algorithm was applied to EVI to gap-fill the missing data or bad-quality data (Zhang et al. 2016a).

The MODIS land cover product (MCD12Q1 V005) provides annual global maps of land cover at 500-m spatial resolution during 2001-2013 (Friedl et al. 2010). We used the MCD12Q1 data at 2013 to represent year 2014. The IGBP land cover classification scheme in the MCD12Q1 is used in this study (see Fig. 1a). The IGBP land cover map was then used to derive biome-specific model parameter information for VPM simulations.

Annual national CDL datasets with a spatial resolution of 30-m were available for our study period (2008–2014) (https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php). The CDL datasets contain over 100 cropland types, and have very high classification accuracies for most crops (over 90% accuracy for major crop types such as soybean and corn) (Boryan et al. 2011). For the VPM simulations, annual CDL datasets in 2008–2014 were aggregated to generate data layers at 500-m spatial resolution that represent the ratio of C_3 and C_4 vegetation within individual 500-m gridcells for each year (Fig. 1b). The C_4 cropland layer included corn, sorghum, sugarcane, and millet, and all other crops were C_3.

3.2.3 Evaluation of GPP estimates during 2010–2014 from VPM

Eddy covariance data from the FLUXNET2015 dataset were used to assess GPP_{VPM}. We used 25 FLUXNET sites across CONUS according to their data availability during 2008–2014, for which a summary about these sites is shown in Table 3.1 and Fig. 3.1a. The FLUXNET2015 dataset used a standard workflow to process the data from the eddy covariance flux tower sites (http://fluxnet.fluxdata.org/data/). The net ecosystem exchange (NEE) of CO_2 between ecosystems and the atmosphere was gap-filled and then partitioned into GPP and ecosystem respiration (ER) using two methods, the nighttime based and the daytime based approaches (Lasslop et al. 2010; Reichstein et al. 2005). We calculated average daily GPP_{EC} as the average of daily GPP estimated by the two methods. Then, we calculated 8-day average GPP_{EC} by aggregating the average daily GPPs. For each 8-day interval, only the shortwave radiation and net ecosystem exchange (NEE) observations with more than 75% of good quality, gap-filled data were kept.

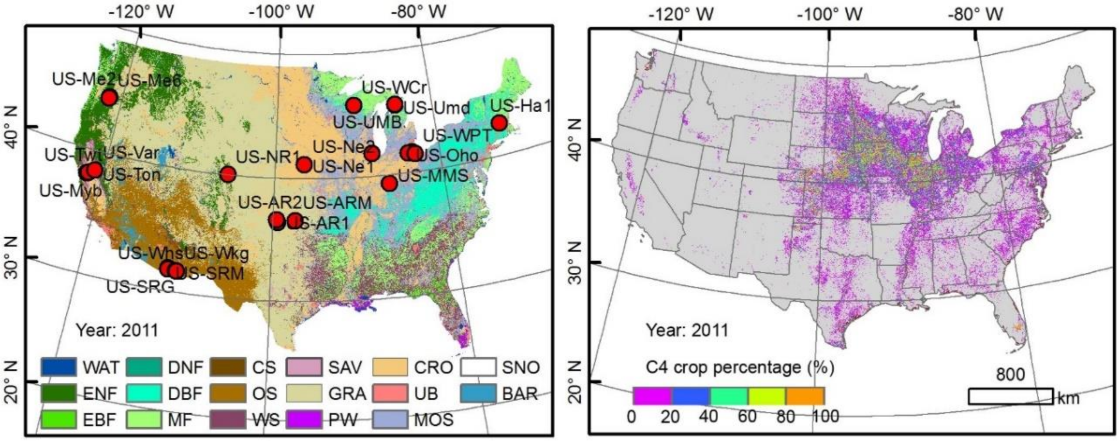

Figure 3.7 (a) Land cover map of CONUS derived from MCD12Q1 in 2011 and (b) the C_4 crop percentage within a 500-m MODIS pixel derived from 30-m cropland data layer. Abbreviations denote the IGBP land-use classes. WAT: Water, ENF: Evergreen Needleleaf Forest, EBF: Evergreen Broadleaf Forest, DNF: Deciduous Needleleaf Forest, DBF: Deciduous Broadleaf Forest, MF: Mixed Forest, CS: Closed Shrublands, OS: Open Shrublands, WS: Woody Shrublands, SAV: Savannas, GRA: Grasslands, PW: permanent wetlands, CRO: Croplands, UB: Urban and Built-up, MOS: Cropland/Natural vegetation mosaic; SNO: Snow and Ice; BAR: Barren or sparsely vegetated. In Fig. 1a, we also labeled the locations of the eddy covariance flux tower sites used in this study.

Table 3.2 Name, location, vegetation type, and available years (within 2010–2014 study period) of 25 eddy covariance flux tower sites in this study. RMSE, R^2, and slope are the root mean square error, coefficient of determination, and regression slope of the regression analysis, respectively, between tower-derived GPP and simulated GPP from VPM and MOD17.

Site ID	Latitude	Longitude	IGBP	Time	slope		R^2		RMSE ($g\ C\ m^{-2}\ d^{-1}$)		References	DOI
					VPM	MOD17	VPM	MOD17	VPM	MOD17		
US-AR1	36.4267	-99.4200	GRA	2009–2012	1.07	0.59	0.55	0.06	1.35	1.35	-	http://dx.doi.org/10.17190/AMF/1246137
US-AR2	36.6358	-99.5975	GRA	2009–2012	1.08	0.62	0.43	-0.12	1.14	1.20	-	http://dx.doi.org/10.17190/AMF/1246138
US-IB2	41.8406	-88.2410	GRA	2009–2011	1.21	0.84	0.87	0.86	2.14	1.52	(Matamala et al. 2008)	http://dx.doi.org/10.17190/AMF/1246066
US-SRG	31.7894	-110.8277	GRA	2009–2014	0.78	0.51	0.69	-0.35	0.75	1.16	(Scott et al. 2015)	http://dx.doi.org/10.17190/AMF/1246154
US-Var	38.4133	-120.9507	GRA	2009–2014	0.66	0.88	0.66	0.42	1.71	1.65	(Ma et al. 2007)	http://dx.doi.org/10.17190/AMF/1245984
US-Wkg	31.7365	-109.9419	GRA	2009–2014	0.76	0.63	0.82	0.39	0.54	0.69	(Scott et al. 2010)	http://dx.doi.org/10.17190/AMF/1246112
US-ARM	36.6058	-97.4888	CRO	2009–2012	0.79	0.64	0.57	0.47	1.46	1.56	(Fischer et al. 2007)	http://dx.doi.org/10.17190/AMF/1246027
US-CRT	41.6285	-83.3471	CRO	2011–2013	0.68	0.45	0.78	0.64	3.00	4.23	(Chu et al. 2014)	http://dx.doi.org/10.17190/AMF/1246156
US-Ne1	41.1651	-96.4766	CRO	2009–2012	0.96	0.31	0.95	0.51	1.80	7.62	(Suyker et al. 2005)	http://dx.doi.org/10.17190/AMF/1246084
US-Ne2	41.1649	-96.4701	CRO	2009–2012	0.89	0.29	0.96	0.41	1.90	8.19	(Suyker et al. 2005)	http://dx.doi.org/10.17190/AMF/1246085
US-Twt	38.1087	-121.6530	CRO	2009–2014	0.87	0.53	0.43	-0.93	3.17	4.24	(Hatala et al. 2012)	http://dx.doi.org/10.17190/AMF/1246140
US-Ha1	42.5378	-72.1715	DBF	2009–2012	0.93	0.66	0.93	0.74	1.32	2.89	(Urbanski et al. 2007)	http://dx.doi.org/10.17190/AMF/1246059
US-MMS	39.3232	-86.4131	DBF	2009–2014	1.07	0.85	0.91	0.71	1.58	2.33	(Schmid et al. 2000)	http://dx.doi.org/10.17190/AMF/1246080
US-Oho	41.5545	-83.8438	DBF	2009–2013	0.87	0.77	0.92	0.83	1.63	2.37	(van Gorsel et al. 2009)	http://dx.doi.org/10.17190/AMF/1246089
US-UMB	45.5598	-84.7138	DBF	2009–2014	1.09	1.20	0.96	0.93	1.15	1.87	(Gough et al. 2008)	http://dx.doi.org/10.17190/AMF/1246107
US-Umd	45.5625	-84.6975	DBF	2009–2014	0.91	0.90	0.88	0.84	1.64	1.70	(Gough et al. 2013)	http://dx.doi.org/10.17190/AMF/1246134
US-WCr	45.8059	-90.0799	DBF	2011–2014	1.25	0.84	0.90	0.80	2.46	2.03	(Cook et al. 2004)	http://dx.doi.org/10.17190/AMF/1246111
US-Me2	44.4523	-121.5574	ENF	2009–2014	0.58	0.72	0.74	0.79	2.81	2.08	(Law et al. 2004)	http://dx.doi.org/10.17190/AMF/1246076
US-Me6	44.3233	-121.6078	ENF	2010–2014	0.56	0.95	0.51	0.56	1.53	1.26	(Law et al. 2000)	http://dx.doi.org/10.17190/AMF/1246128
US-NR1	40.0329	-105.5464	ENF	2009–2014	0.84	0.91	0.86	0.84	1.04	1.03	(Monson et al. 2002)	http://dx.doi.org/10.17190/AMF/1246088
US-Myb	38.0498	-121.7651	WET	2011–2014	1.27	0.92	0.36	-0.31	3.49	2.85	(Sturtevant et al. 2016)	http://dx.doi.org/10.17190/AMF/1246139
US-SRC	31.9083	-110.8395	OSH	2009–2014	0.97	0.99	0.32	-1.31	0.39	0.42	(Cavanaugh et al. 2011)	http://dx.doi.org/10.17190/AMF/1246127
US-Whs	31.7438	-110.0522	OSH	2009–2014	0.70	0.71	0.72	-0.28	0.45	0.56	(Scott et al. 2015)	http://dx.doi.org/10.17190/AMF/1246113
US-SRM	31.8214	-110.8661	WSA	2009–2014	0.90	0.61	0.74	-1.18	0.49	0.81	(Scott et al. 2009)	http://dx.doi.org/10.17190/AMF/1246104
US-Ton	38.4316	-120.9660	WSA	2009–2014	0.68	1.01	0.61	0.73	1.28	0.91	(Baldocchi et al. 2004)	http://dx.doi.org/10.17190/AMF/1245971

We evaluated the seasonal and cross-site performance of GPP_{VPM} across biomes at 8-day and interannual scales. We classified the land cover maps into four major types: forest (FOR), grassland (GRA), cropland (CRO), and others (OTH) based on the MCD12Q1 landcover data. The evergreen needleleaf forest, evergreen broadleaf forest, deciduous broadleaf forest, decidous needleaf forest, and mixed forest were lumped together as forest. Grassland and cropland were the same classification scheme as MCD12Q1, while all the other land cover types, such as savannas, shrublands, wetlands, and sparsed vegetated area, were considered as OTH. To examine the ability of the model to capture the interannual variability of GPP, we compared the anomaly of annual GPP for GPP_{EC} and GPP_{VPM}. Specifically, we compared GPP_{VPM} and GPP_{MOD17} to the anomaly between GPP_{EC} in each site year and average GPP_{EC} over all the site years for each site. The slope, root mean square error (RMSE), and R^2 of the regression models were used to evaluate the difference between modeled and eddy-covariance derived GPP.

SIF is a very small amount of energy emitted by plants and has been demonstrated to be highly correlated with GPP (Guanter et al. 2014; Wagle et al. 2016; Zhang et al. 2016a). In this study, we used the monthly GOME-2 SIF data (V26) during 2008–2014 (Joiner et al. 2013). GOME-2 measurements are in the ultraviolet and visible part of the spectrum (240 - 790 nm) with a high spectral resolution between 0.2 and 0.5 nm and with the footprint size of 80×40 km^2. SIF is retrieved using a principle component analysis method in the 734 to 758 nm spectral window which overlaps the second peak of the SIF emission. The retrievals are quality-filtered and aggregated into 0.5° grids and a monthly interval (Joiner et al. 2013).

We compared GPP_{VPM} with the latest version of MOD17 GPP product (Running et al. 2004), MOD17A2H V006 (GPP_{MOD17}) at both site and regional scales. GPP_{MOD17} is estimated at a spatial resolution of 500-m and a temporal resolution of 8-day, which matches the spatial and temporal resolutions of GPP_{VPM}. MOD17 is also a LUE model and simulates GPP as the product of $APAR_{canopy}$ and light use efficiency (ε_g). ε_g is determined by ε_{max} and scalars that capture environmental limitations such as vapor pressure deficit and air temperature. ε_{max} values are specific for different biome types (e.g., forest, shrub, grass, crop) (Running et al. 2004), but the product does not account for the differences of ε_{max} between C_3 and C_4 croplands, and ε_{max} for croplands is substantially too low (Turner et al. 2006; Xin et al. 2015).

We also compared GPP_{VPM} with GPP simulated by CASA-GFED3 (GPP_{CASA}). CASA estimates Net Primary Productivity (NPP) based on the light use efficiency method (Monteith 1972, 1977) and further estimates GPP with an assumption GPP = 2 * NPP. ε_{max} for predicting NPP in CASA is set uniformly (0.55 g C MJ^{-1} PAR) for different biomes (Potter et al. 2012; Potter et al. 1993; Randerson et al. 1996). The CASA-GFED3 GPP product used a calibrated ε_{max} for the Midwestern region, which was derived from crop yield observations, meteorological data, and remotely sensed FPAR (Lobell et al. 2002), and thus corresponds with much higher GPP values (roughly 45%) over the Midwestern US (Hilton et al. 2015). GPP_{CASA} is used to generate prior biogenic CO_2 fluxes for the CarbonTracker system (Peters et al. 2007) at a spatial resolution of 1°×1.25° every 3 hour. We resampled the data into 1°×1° and aggregated them into monthly values in this study.

The GPP estimates by the SiBCASA-GFED4 model ($GPP_{SiBCASA}$) (van der Velde et al. 2014) were also compared with regional GPP_{VPM}. $GPP_{SiBCASA}$ is used to generate prior biogenic CO_2 fluxes in the Carbon Tracker Europe system (van der Laan-Luijkx et al. 2017). SiBCASA combines the biophysical and GPP components from the Simple Biosphere model (version 2.5) with the heterotrophic respiration (R_H) from CASA model, and calculates the exchange of carbon, energy, and water at a temporal resolution of 10-min and at a spatial resolution of 1°×1° (Schaefer et al. 2008; van der Velde et al. 2014). GPP is calculated for both C_3 and C_4 plants by implementing a modified version of the C_3 enzyme kinetic model (Farquhar et al. 1980) and the C_4 photosynthesis model (Collatz et al. 1992). The C_4 distribution map used in SiBCASA is a static map with the mean C_4 fraction in global 1°×1° grids (Still et al. 2003). The aggregated monthly $GPP_{SiBCASA}$ data is used for the comparison.

The impact of climate extremes on the GPP and SIF over the CONUS was evaluated using the four GPP datasets and GOME-2 SIF data. The seasonal cycle and anomaly of GPP_{VPM}, GPP_{MOD17}, $GPP_{SiBCASA}$, GPP_{CASA} and SIF in the year 2012 were compared to that in the baseline year (the average of the year 2008, 2009, 2010, 2013 and 2014). The uncertainty range of the anomaly was calculated as the standard deviation (SD) of the anomaly between 2012 and selected different baselines. We randomly chose at least three years from the year 2008, 2009, 2010, 2013 and 2014 to calculate the baseline, so there are 16 options ($C_5^3 + C_5^4 + C_5^5$). As $GPP_{SiBCASA}$, GPP_{CASA} and SIF datasets have a spatial resolution of 1.0°×1.0°, both GPP_{VPM} and GPP_{MOD17} datasets (500-m spatial resolution) were aggregated to 1.0°×1.0°. The SIF data (0.5°×0.5°) was also aggregated to 1.0°×1.0°.

We then used the area-weighted method to calculate annual total GPP (Pg C year^{-1}) and average SIF over CONUS.

3.3 Results

3.3.1 Seasonal dynamics and interannual variation of GPP at flux tower sites

GPP_{VPM} agreed reasonably well with the seasonal dynamics and peak values of GPP_{EC} at most sites (Fig. 3.2). The coefficients of determination (R^2) varied from 0.32 (US-SRC) site to 0.96 (US-Ne2 and US-UMB). GPP_{VPM} showed very high accuracy for the cropland sites relative to GPP_{MOD17} (see Fig. 3.2 and Table 3.1). At the US-Ne1 and US-Ne2 maize sites, the regression between GPP_{VPM} and GPP_{EC} show a high R^2 value (> 0.95) and a low RMSE value (< 2.0 g C m^{-2} day^{-1}), while the regression between GPP_{MOD17} and GPP_{EC} show a moderate R^2 value (~0.50) and a large RMSE value (7.0 g C m^{-2} day^{-1}) (Table 3.1).

At the 8-day scale, GPP_{VPM} agrees better with GPP_{EC} than does GPP_{MOD17}, and GPP_{VPM} effectively captures the seasonal dynamics of GPP for all the four biomes (Fig. 3.3a, b). For croplands, GPP_{MOD17} shows significant underestimation with a slope of 0.37 while GPP_{VPM} presents only slight underestimation with a slope of 0.97. The improvement in GPP_{VPM} is most prominent in these C_4 cropland sites, such as US-Ne1 and US-Ne2 (Fig. 3.2 and Table 3.1), with peak value of GPP_{VPM} and GPP_{EC} in the growing season that are larger than 20 g C m^{-2} day^{-1}, while that of GPP_{MOD17} is less than 10 g C m^{-2} day^{-1}. Across all 25 sites, GPP_{VPM} explains about 84% of the seasonal dynamics of GPP_{EC} with RMSE of 1.7 g C m^{2} day^{-1}, while GPP_{MOD17} only explains only about 55% with a RMSE value of 2.6 g C m^{-2} day^{-1}.

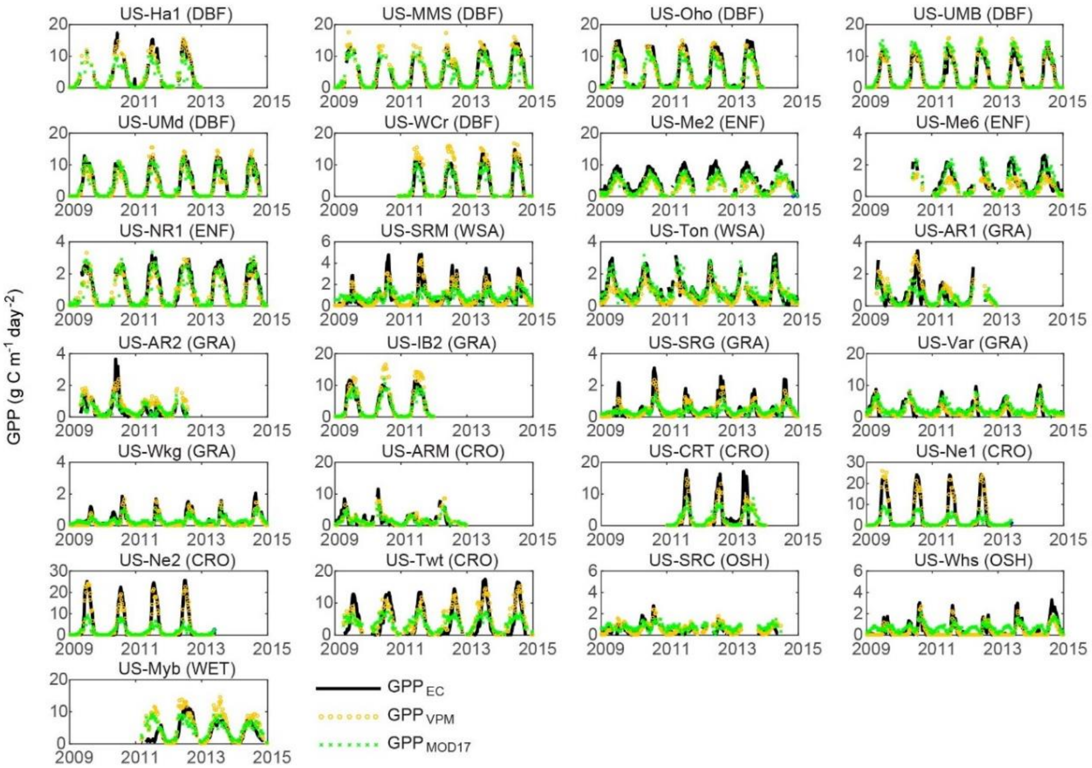

Figure 3.8 Seasonal dynamics and interannual variations of the tower-based GPP (GPP_{EC}), GPP simulated by VPM (GPP_{VPM}), and GPP simulated by MOD17 (GPP_{MOD17}) at 25 flux sites at 8-day intervals (please note the different y-axis scales).

The interannual variation of GPP_{VPM} was best for croplands, followed by forest, grasslands, and other biomes (Fig. 3.3c). In addition, the anomaly of annual GPP_{VPM} in croplands, grasslands, and forest biomes has much higher consistency with GPP_{EC} than does GPP_{MOD17} (Fig. 3.3c, d). In other biomes (5 sites), both GPP_{VPM} and GPP_{MOD17} had relatively low accuracy.

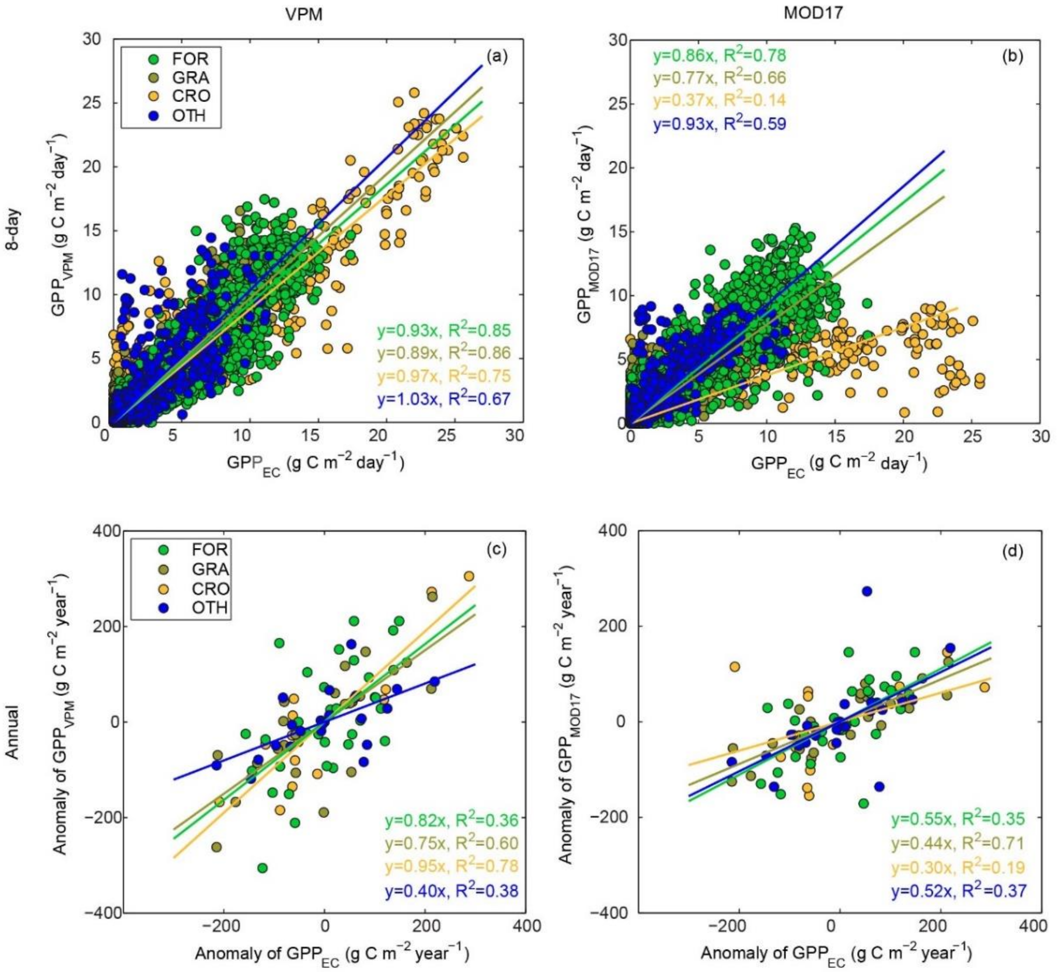

Figure 3.9 Comparison of GPP$_{EC}$, GPP$_{VPM}$, and GPP$_{MOD17}$ across eddy covariance flux tower sites (forest, grassland, cropland, and others) during 2010 to 2014: (a) 8-day GPP$_{EC}$ and GPP$_{VPM}$, (b) 8-day GPP$_{EC}$ and GPP$_{MOD17}$, (c) anomaly of annual GPP$_{EC}$ and GPP$_{VPM}$, and (d) anomaly of annual GPP$_{EC}$ and GPP$_{MOD17}$. FOR: forests, CRO: croplands, GRA: grasslands, OTH: other types. When all the sites were combined, the relationship between GPP$_{VPM}$ and GPP$_{EC}$ was y = 0.92 x (R^2 = 0.84, RMSE = 1.7 g C m^{-2} d^{-1}) at the 8-day time scale, while the relationship between GPP$_{MOD17}$ and GPP$_{EC}$ was y = 0.68 x (R^2 = 0.55, RMSE = 2.6 g C m^{-2} day^{-1} at the 8-day time scale. At the inter-annual scale, the relationship between the annual anomaly of GPP$_{VPM}$ and GPP$_{EC}$ is y = 0.73 x (R^2 = 0.48) while the relationship between the annual anomaly of GPP$_{MOD17}$ and GPP$_{EC}$ was y = 0.45 x (R^2 = 0.37).

3.3.2 Spatial-temporal consistency between model-based GPP and SIF over CONUS in the baseline years and drought year 2012

We compared the spatial distribution of maximum monthly mean GPP (g C m^{-2} day^{-1}) from the four GPP products and annual maximum monthly mean SIF in the baseline year and drought year 2012 at 1°×1° resolution (Fig.3.4a–j). The maximum monthly mean GPP in 2008, 2009, 2010, 2013, and 2014 were used as baseline year. The three GPP products (GPP_{VPM}, GPP_{CASA} and $GPP_{SiBCASA}$) and SIF show the peak photosynthesis in the Mid-western corn-belt region (Fig. 3.4a-j), which was consistent with the results reported by Hilton et al. (2017). GPP_{MOD17} did not have such a spatial pattern for maximum monthly GPP because it did not include higher photosynthetic capacity for C_4 vegetation as did the other three models (VPM, CASA and SiBCASA). Compared to the baseline years, most of gridcells had lower GPP and SIF values during the drought in 2012. The correlation analysis (Fig. 3.5a–d; Fig. 3.5e–h) showed that the maximum monthly GPP_{VPM} and SIF have the strongest linear relationship, followed by SIF/$GPP_{SiBCASA}$, SIF/GPP_{CASA}, and SIF/GPP_{MOD17}.

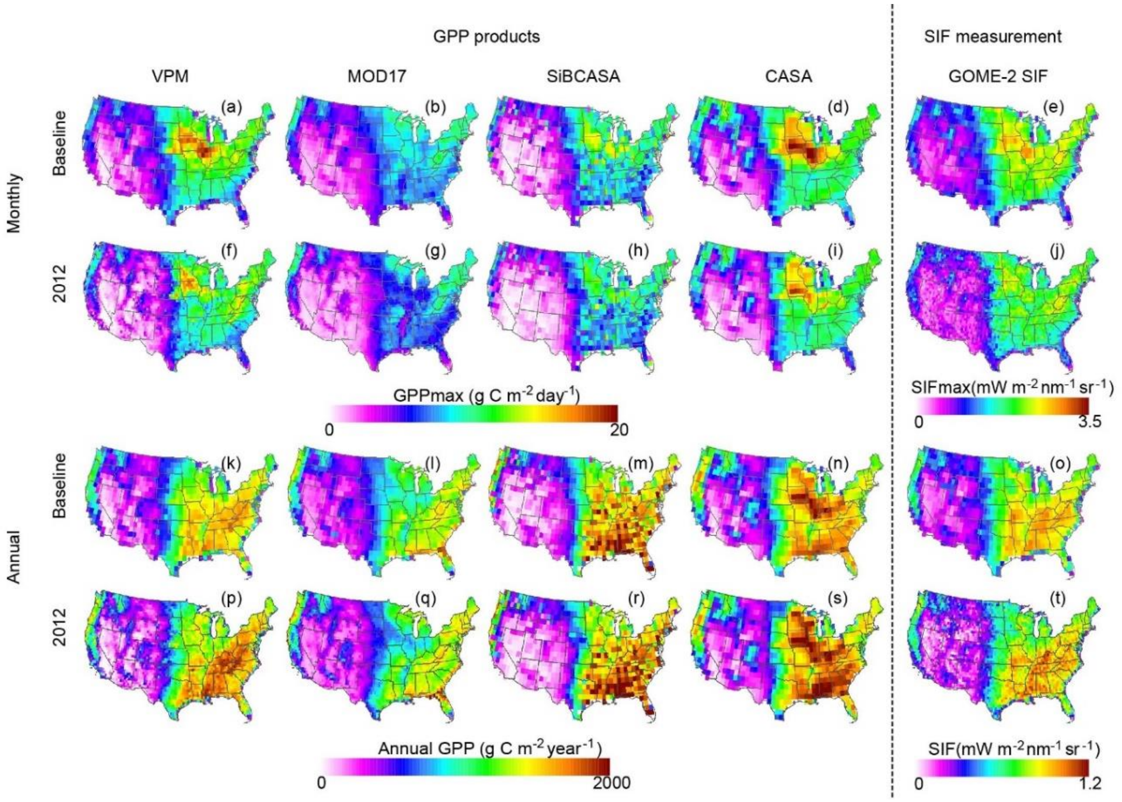

Figure 3.10 Spatial distribution of maximum monthly mean GPP (a–d; f–i) from GPP models (VPM, MOD17, SiB-CASA and CASA) and maximum monthly mean SIF (e, j) from GOME-2 (e, j) in the baseline years (the average of 2010, 2011, 2013, 2014) and drought year 2012, and spatial distributions of annual GPP (k–n; p–s) from GPP models and annual mean SIF from GOME-2 (o, t) in the baseline years and drought year 2012.

For annual total GPP, all four GPP products showed very similar spatial patterns with SIF, with relatively high annual GPP (> 1500 g C m^{-2} $year^{-1}$) in the forested Southeastern US and low annual GPP in the western regions where grasslands and deserts are dominant (Fig. 3.4k–t). In 2012, GPP_{VPM} had a decrease in the Midwestern corn-belt region and Great Plains, and an increase in the eastern temperate forest region, which is consistent with the spatial patterns of SIF. Annual GPP_{MOD17} had an obvious decrease in the Midwestern corn-belt area but a slight increase in the eastern forest area in 2012. Annual $GPP_{SiBCASA}$ had no significant differences between the baseline and drought year

2012. Annual GPP_{CASA} had large increases in both the Midwestern corn-belt region and temperate forest area. The correlation analysis (Fig. 3.5i–l) showed that annual GPP_{VPM} had a stronger linear relationship with SIF ($R^2 = 0.94$) in the baseline years than SIF/$GPP_{SiBCASA}$ ($R^2 = 0.76$), SIF/GPP_{CASA} ($R^2 = 0.75$), and SIF/GPP_{MOD17} ($R^2 = 0.70$). We found similar results for the drought year 2012 (Fig. 3.5m–p), which suggested that the models performed similarly during baseline and drought years.

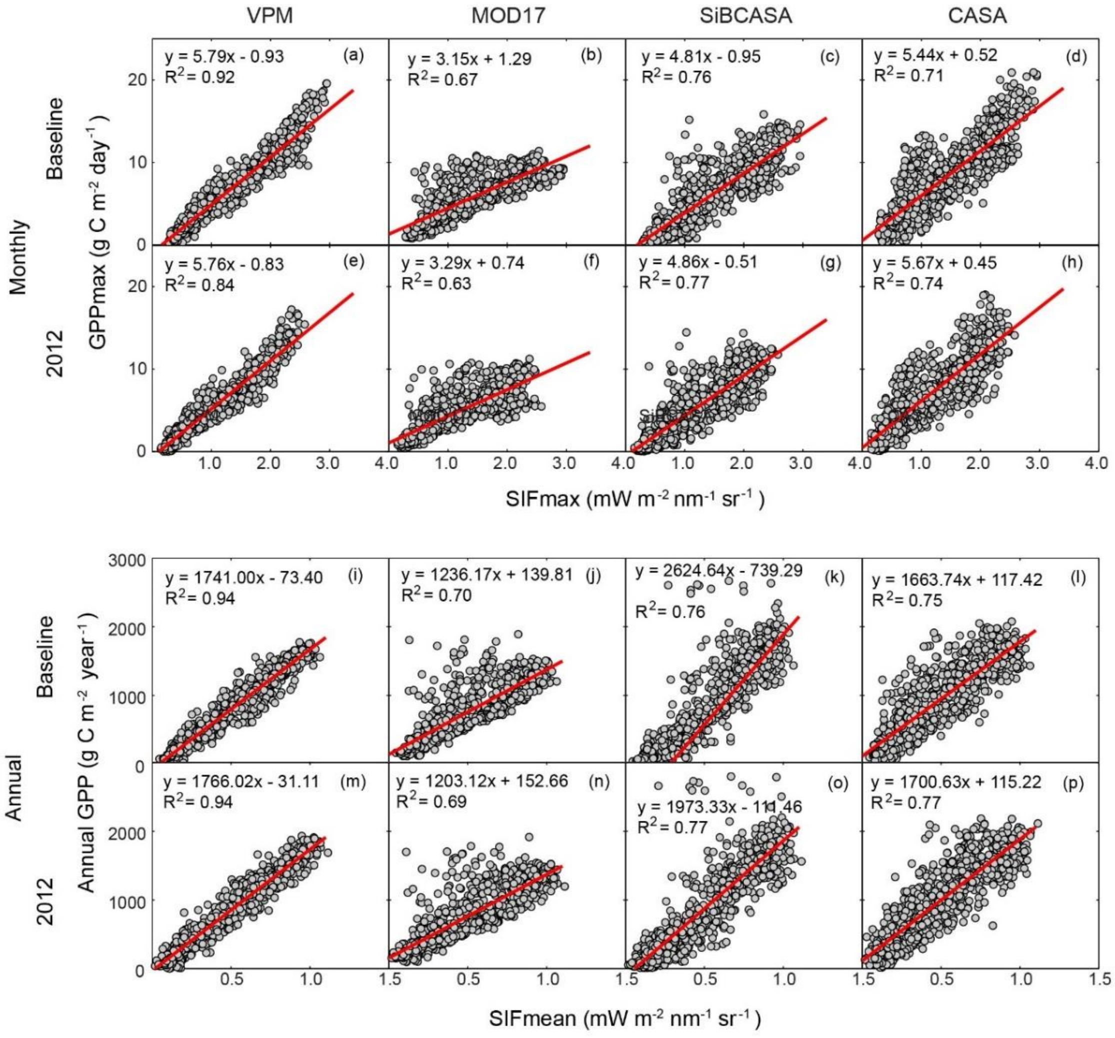

Figure 3.11 Relationships between the maximum monthly mean GPP (a–d; e–h) from GPP models (VPM, MOD17, SiBCASA and CASA) and monthly mean SIF from GOME-2 for each pixel across CONUS during the baseline years (the average

of 2008, 2009, 2010, 2013 and 2014) and drought year 2012, and relationship between total annual GPP (i–l; m–p) from GPP models (VPM, MOD17, SiBCASA, and CASA) and mean annual SIF from GOME-2 in the baseline year (the average of 2008, 2009, 2010, 2013 and 2014) and drought year 2012 (all of the relationships are significant with $p<0.001$).

GPP estimates from all models had a high correlation with SIF (>0.9) in the wetter eastern region but a low correlation in the dry western region, partly due to the very low SIF signal and relatively large signal-to-noise ratio (Fig. 3.6a–h). The percentages of the total number of gridcells with a Pearson correlation coefficient larger than 0.9 in the baseline year was ~65% for SIF/GPP$_{VPM}$, ~55% for SIF/GPP$_{CASA,}$ ~50% for SIF/GPP$_{MOD17,}$ and ~47% for SIF/GPP$_{SiBCASA}$ (Fig. 3.6i–l). The four GPP models had no obvious differences in simulating the seasonal dynamics of GPP between the baseline year and drought year 2012 (Fig. 3.6).

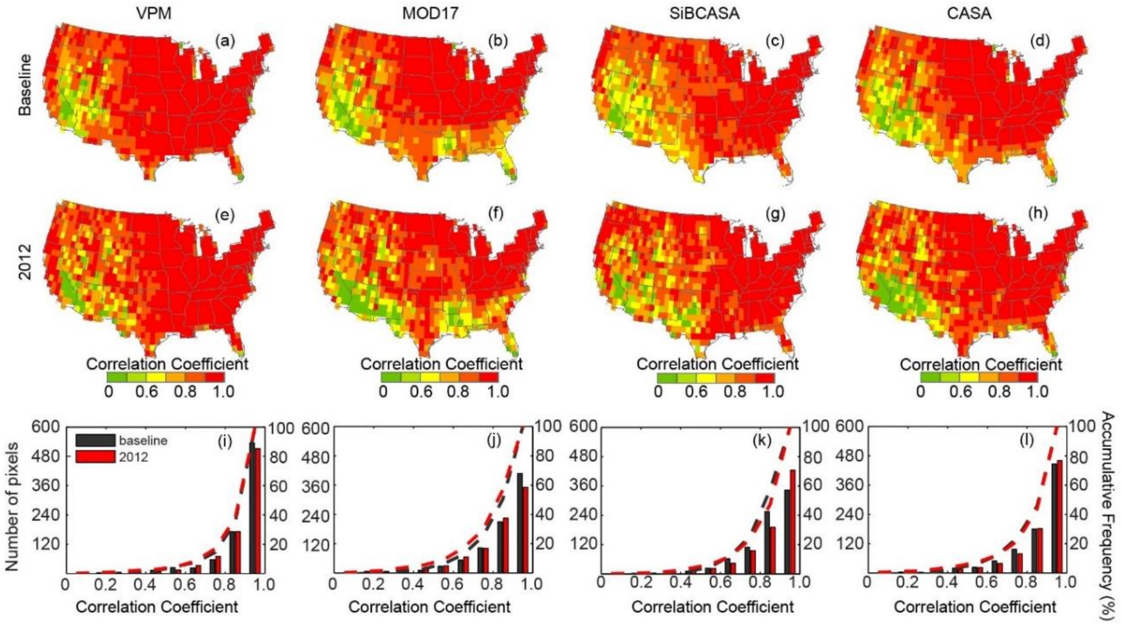

Figure 3.12 Spatial distribution of Pearson correlation coefficient between monthly SIF and GPP products from VPM, MOD17, SiBCASA, and CASA for baseline year (the average of 2008, 2009, 2010, 2013, and 2014) and drought year 2012, and the corresponding frequency distribution (black and red bars) and accumulative

frequency (black and red dashed lines) of the Pearson correlation coefficient for the four models in the baseline years and 2012.

The histograms of the slope values (GPP = a × SIF + b) among these four GPP models differed substantially. The slope values for the SIF/GPP$_{VPM}$ were concentrated between 4 and 7 g C mW^{-1} nm^{-1} sr^{-1} (~53% of all gridcells), while that for SIF/GPP$_{MOD17}$ were between 2 to 5 g C mW^{-1} nm^{-1} sr^{-1} (~60% of all gridcells). The slope values for the SIF/GPP$_{CASA}$ and SIF/GPP$_{SiBCASA}$ were more evenly distributed than that of SIF/GPP$_{VPM}$. Sun et al. (2017) found the GPP-SIF relationship is consistent across different vegetation types when comparing SIF with GPP$_{EC}$, but it is more divergent when comparing SIF with modelled GPP because of the systematic GPP biases. The GPP-SIF slope for the four GPP products in this study is also divergent over CONUS, but the VPM GPP-SIF slope is more convergent than the other three models (Fig. 3.7).

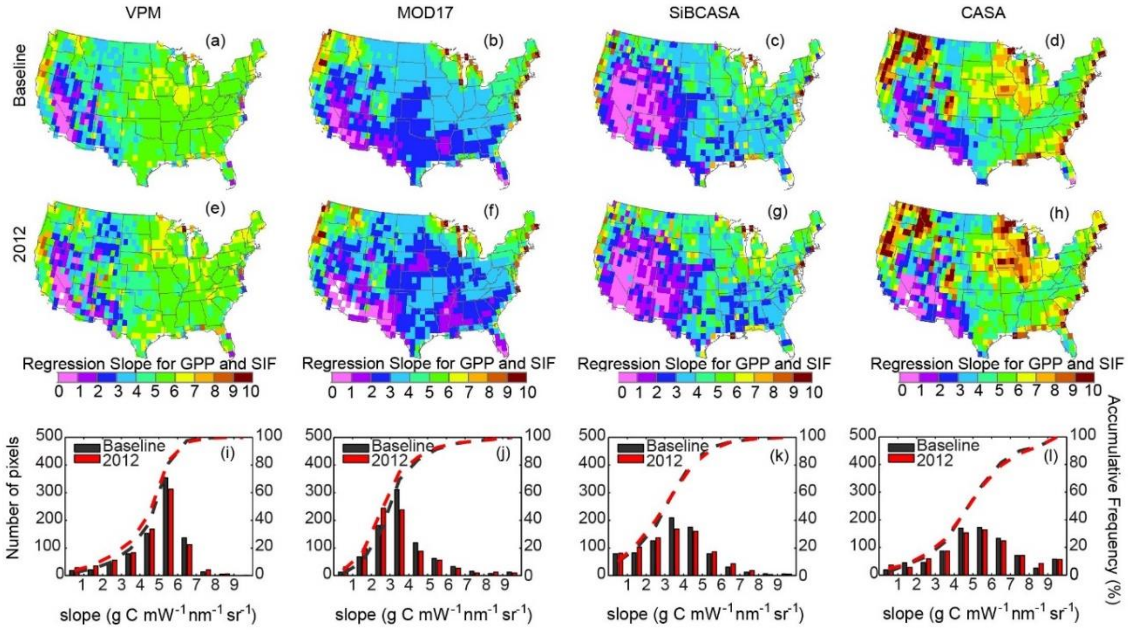

Figure 3.13 Spatial distribution of the regression slope between monthly SIF and GPP products from VPM, MOD17, SiBCASA, and CASA for the baseline year (the average of 2008, 2009, 2010, 2013, and 2014) and drought year 2012, and the

corresponding frequency distribution (black and red bars) and accumulative frequency (black and red dashed lines) of the Pearson correlation coefficient for the four models in the baseline years and 2012.

3.3.3 Spatial-temporal consistency of GPP and SIF anomalies over CONUS in 2012

To evaluate the impacts of spring warming and summer drought on GPP in 2012, we compared the anomalies of GPP from $GPP_{VPM,}$ GPP_{MOD17}, $GPP_{SiBCASA}$, and GPP_{CASA} to the anomalies of SIF in the spring, summer, and the entire year at 1°×1° (latitude and longitude) resolution (Fig. 3.8). The anomalies of GPP and SIF were calculated as the differences between year 2012 and the baseline year. The baseline year was calculated as the average of 2008, 2009, 2010, 2013, and 2014. Geographically, the anomaly of all the four GPP products and SIF showed very interesting spatial patterns at the seasonal and annual scales (Fig. 3.8).

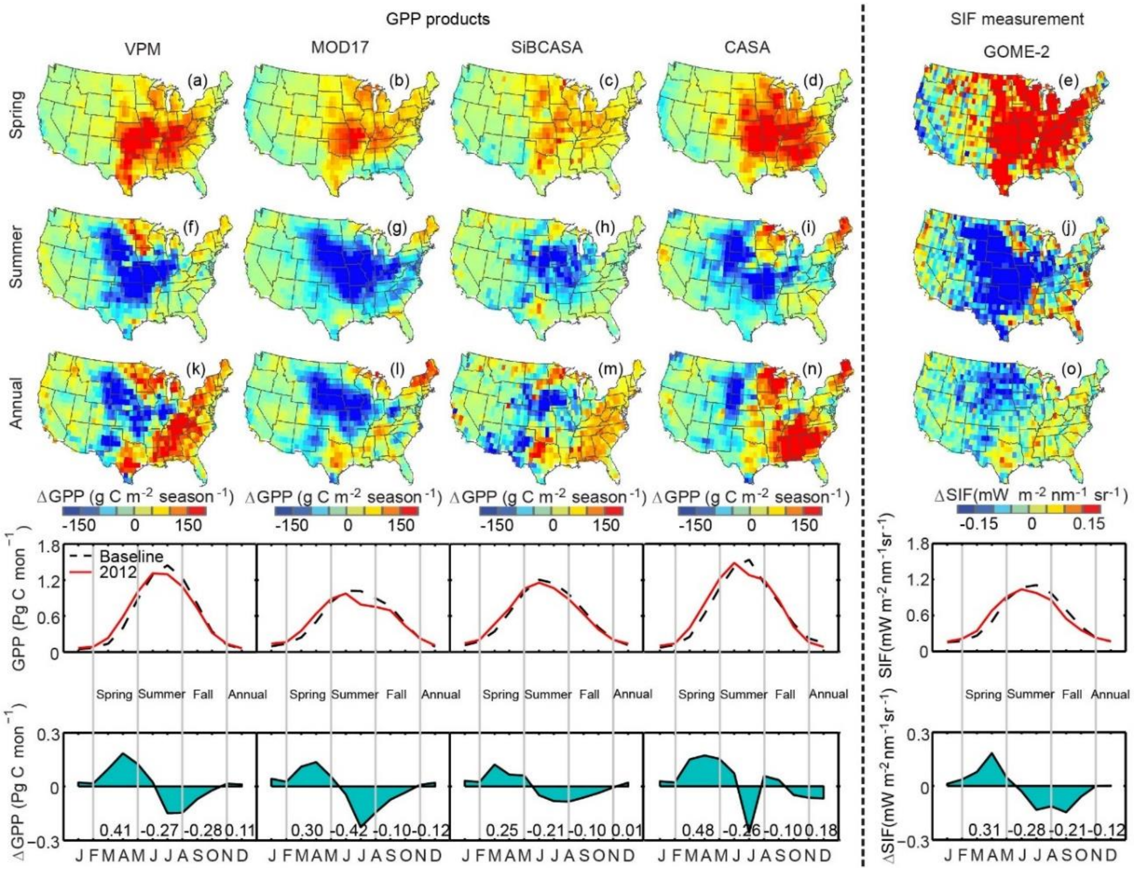

Figure 3.14 Spatial-temporal anomalies of GPP$_{\text{VPM}}$, GPP$_{\text{MOD17}}$, GPP$_{\text{SiBCASA}}$, GPP$_{\text{CASA}}$, and SIF during spring, summer, and annually across CONUS in 2012 relative to the baseline (2008, 2009, 2010, 2013 and 2014). Seasonal cycle and anomaly of total monthly GPP$_{\text{VPM}}$, GPP$_{\text{MOD17}}$, GPP$_{\text{SiBCASA}}$, GPP$_{\text{CASA}}$ and SIF in 2012 relative to the baseline. Numbers shown in the last row of graphs are the anomaly of total GPP in spring (March–May), summer (June–August), fall (September–November) and the whole year (January to December).

In the spring season, the middle and eastern CONUS experienced an increase in GPP anomaly while western CONUS experienced a decrease, which was consistent with the spatial pattern of SIF anomaly (Fig. 3.8a–e). The magnitudes and spatial extent of GPP anomaly vary among the four GPP models. For GPP$_{\text{VPM}}$ and GPP$_{\text{MOD17}}$, the large increases in GPP (larger than 100 g C m^{-2} season^{-1}) occurred mostly in the Southern Great Plains and part of the Midwestern corn-belt region. For GPP$_{\text{CASA,}}$ large increases in GPP occurred mostly in the Midwestern and Southeast regions. The correlation analyses between GPP products and SIF (Fig. 3.9a–d) showed that GPP_{VPM} and SIF had the strongest linear relationship (R^2 = 0.67), followed by SIF/GPP_{MOD17} (R^2 = 0.58), SIF/GPP_{CASA} (R^2 = 0.56), and SIF/$GPP_{SiBCASA}$ (R^2 = 0.48).

Table 3.3 The anomaly of total GPP and the uncertainty range of the anomaly between 2012 and the baseline (the average of 2008, 2009, 2010, 2013 and 2014) in spring (March–May), summer (June–August), fall (September–November) and the whole year. The uncertainty range was calculated as the standard deviation (SD) of the anomaly between 2012 and different baselines. We randomly chose at least three years from the year 2008, 2009, 2010, 2013 and 2014 to calculate the baseline.

Anomaly of GPP (Pg C)	VPM	MODIS	CASA	SiBCASA	Anomaly of SIF (mW m^{-2} nm^{-1} sr^{-1})
Spring	0.41 ± 0.04	0.30 ± 0.03	0.48 ± 0.05	0.25 ± 0.03	0.31 ± 0.05
Summer	-0.27 ± 0.05	-0.42 ± 0.02	-0.26 ± 0.06	-0.21 ± 0.04	-0.28 ± 0.05
Annual	0.11 ± 0.08	-0.12 ± 0.02	0.18 ± 0.10	0.01 ± 0.08	-0.12 ± 0.10

In the summer season, most regions in CONUS experienced decreased GPP and SIF associated with drought (Fig. 3.8f–j). The Great Plains and Midwestern corn-belt regions experienced the largest reductions in GPP (larger than 150 g C m^{-2} $season^{-1}$). The spatial extents of decreased GPP in GPP_{VPM} and GPP_{MOD17} were greater than those in $GPP_{SiBCASA}$ and GPP_{CASA}. GPP_{VPM}, GPP_{CASA}, and $GPP_{SiBCASA}$ displayed strong increases in the southeast regions, which was consistent with the spatial pattern of SIF anomaly. Overall, GPP_{VPM} agreed best with SIF (R^2 = 0.71), followed by SIF/GPP_{CASA} (R^2 = 0.50), SIF/GPP_{MOD17} and R^2 of 0.45), and SIF/$GPP_{SiBCASA}$ (R^2 = 0.19) (Fig. 3.9e–h).

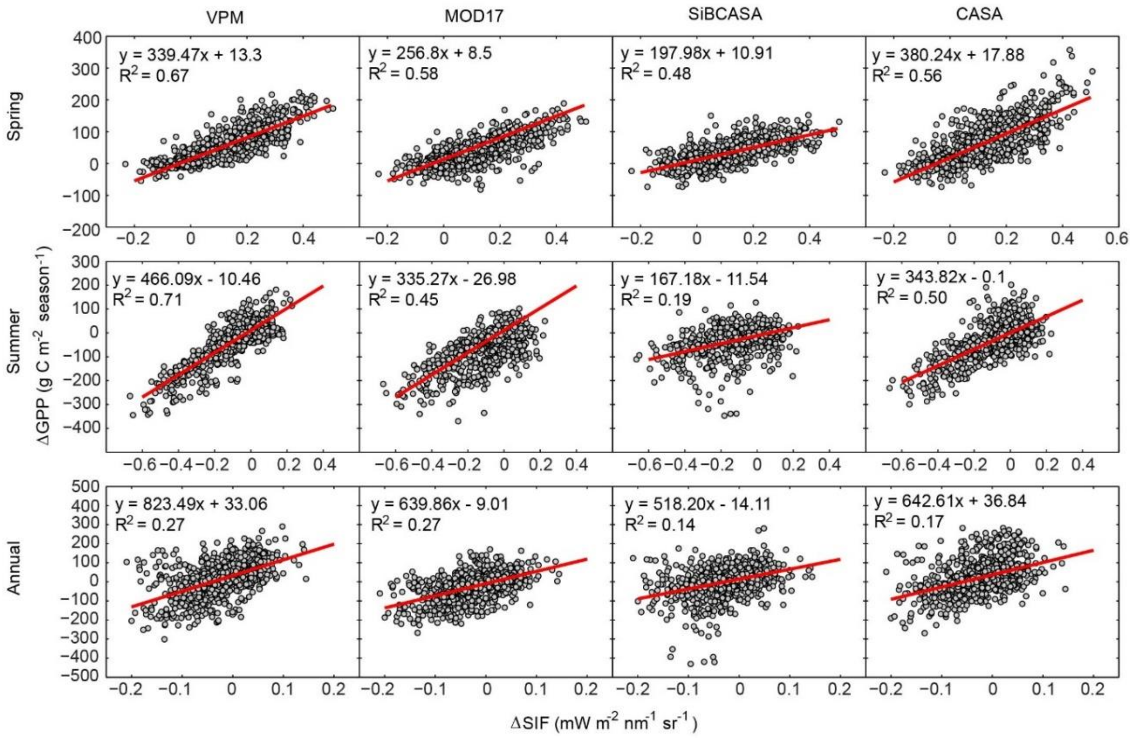

Figure 3.15 Correlation between the anomaly of seasonal/annual GPP from GPP models (VPM, MOD17, SiBCASA, and CASA) and the anomaly of seasonal/annual mean SIF from GOME-2 across CONUS during the baseline years (the average of 2008, 2009, 2010, 2013 and 2014) and drought year 2012 (all of the correlations are significant with p<0.001)

For the entire year, annual GPP_{VPM}, GPP_{MOD17}, and $GPP_{SiBCASA}$ mainly decreased in the western US and corn-belt regions, and annual GPP increased mainly in the eastern and southern forest area, which was consistent with the spatial pattern of SIF (Fig. 3.8k–o). Only GPP_{CASA} reported strong increases in GPP in the corn-belt region. The correlation analysis showed that none of the four GPP products agreed well with the spatial pattern of annual mean SIF anomaly at the annual scale, with R^2 values varying from 0.14 to 0.27 (Fig. 3.9i–l).

When aggregated over the entire CONUS by season, the four GPP products and SIF clearly showed an increase in GPP in the spring and a reduction in the summer, indicating the warm spring and droughty summer had opposite effects on GPP (Fig. 3.8 and Table 3.2). The spring warming led to an increase in GPP by 0.25-0.48 Pg C season^{-1}, where GPP_{CASA} showed the largest increase and $GPP_{SiBCASA}$ showed the least. During the summer, the four GPP products showed a decrease in GPP by 0.21-0.42 Pg C season^{-1}, where GPP_{CASA} decreased the most and $GPP_{SiBCASA}$ decreased the least. The annual total GPP_{VPM} and GPP_{CASA} had an increase of 0.11 and 0.18 Pg C year^{-1}, while the annual total GPPMOD17 and mean SIF had a decrease of 0.12 Pg C year^{-1} and 0.12 mW m^{-2} nm^{-1} sr^{-1}. The annual total $GPP_{SiBCASA}$ remained neutral in 2012.

3.3.4 Impacts of spring warming and summer drought on GPP by biomes in 2012

To quantify the impact of spring warming and summer drought on GPP across biomes, we calculated total GPP from the four models for the four main biomes. In the spring of 2012, all four models showed increased GPP for the four biomes (Fig. 3.10 and Table 3.3), and the non-forest experienced a stronger increase in GPP than forest. In the four models, CASA showed a larger increase in GPP in the spring than other three models in the four biomes, while SiBCASA showed the lowest increase in GPP over most biomes. For the drought summer, all four models showed strong decreases in GPP, and the grassland and cropland experienced the strongest decrease, followed by other biomes. Among the four models, MOD17 showed the largest decrease in GPP in the summer, while SiBCASA showed the least decrease. For the entire year, grassland and cropland experienced a decrease in GPP, while forest and other biomes experienced an increase or no change.

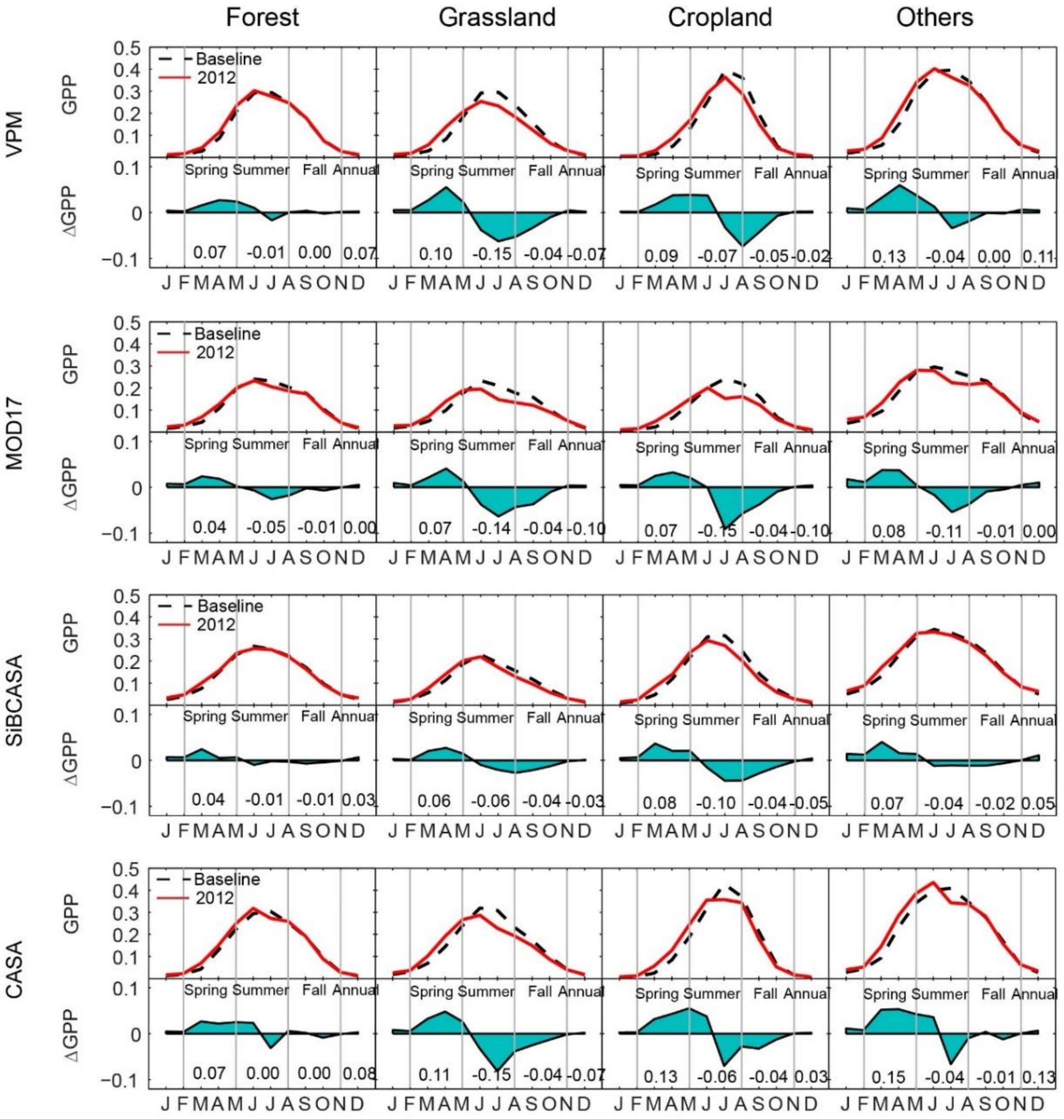

Figure 3.16 Seasonal cycle and anomaly of total monthly GPP_{VPM} and GPP_{MOD17} in (a) forest, (b) grassland, (c) cropland and (d) others. Numbers shown in the bottom panel in each row are the anomalies of total GPP for each biome in spring (March–May), summer (June–August), fall (September–November) and the whole year.

Table 4.3 The anomaly of total GPP estimates from VPM/MOD17 for different biomes between 2012 and the baseline (the average of 2008, 2009, 2010, 2013 and 2014) in spring (March–May), summer (June–August), fall (September–November) and the whole year.

Anomaly of GPP (Pg C)		Spring	Summer	Fall	Annual
VPM	Forest	0.07	-0.01	0.00	0.07
	Grassland	0.10	-0.15	-0.04	-0.07
	Cropland	0.09	-0.07	-0.05	-0.02
	Others	0.13	-0.04	0.00	0.11
MOD17	Forest	0.04	-0.05	-0.01	0.00
	Grassland	0.07	-0.14	-0.04	-0.10
	Cropland	0.08	-0.15	-0.04	-0.10
	Others	0.08	-0.11	-0.01	0.00
SiBCASA	Forest	0.04	-0.01	-0.01	0.03
	Grassland	0.06	-0.06	-0.04	-0.03
	Cropland	0.08	-0.10	-0.04	-0.05
	Others	0.07	-0.04	-0.02	0.05
CASA	Forest	0.07	0.00	-0.01	0.08
	Grassland	0.11	-0.15	-0.04	-0.07
	Cropland	0.13	-0.06	-0.04	0.03
	Others	0.15	-0.04	-0.01	0.13

3.4 Discussion

3.4.1 Improving GPP estimates of C3 and C4 croplands

Accurate estimation of cropland GPP is important for food production and security. The MOD17 GPP data products have been widely used for crop studies (Guan et al. 2016; Xin et al. 2015). However, several studies have reported that the MOD17 data product substantially underestimates GPP in croplands. One reason is that ε_{max} for croplands in the MOD17 model is too low (~1.04 g C MJ^{-1}) (Turner et al. 2006; Wagle et al. 2016; Xin et al. 2015). Site-level studies have indicated that the typical ε_{max} for C_3 crops range from 1.43 to 1.96 g C MJ^{-1} (Chen et al. 2011; Kalfas et al. 2011; Yuan et al. 2015), and ε_{max} for C_4 crops range from 2.25 to 4.06 g C MJ^{-1} (Xin et al. 2015; Yuan et al. 2015). Several model comparison studies have also showed that both process-based GPP models and LUE models have poor performance when estimating GPP in croplands (Schaefer et al. 2012; Verma et al. 2014). Recently, Guanter et al. (2014) used GOME-2 SIF to estimate GPP in croplands using the linear relationship between SIF and tower-based GPP at flux tower sites, and they found these SIF-based GPP estimates in croplands were 50-60% higher than GPP estimates from the ecosystem models over the US Corn Belt. In this study, our ε_{max} values for C_3 croplands (1.80 g C MJ^{-1}) and C_4 croplands (2.7 g C MJ^{-1}) were based on previous site-level studies (Li et al. 2013; Xin et al. 2015). The improved ability of VPM to capture the seasonal dynamics and interannual variability of croplands was partly attributed to more appropriate choices of ε_{max} values.

Another reason for the large error in estimating cropland GPP by the MOD17 and other models can be attributed to the fact that we have very limited knowledge on the spatial distribution of the C_3 and C_4 croplands within individual 500-m MODIS pixels and their temporal dynamics over years (Reeves et al. 2005; Still et al. 2003; Wang et al. 2013). However, in this study we used the fine-resolution, annual CDL cropland maps. The results demonstrated the potential of annual C_3/C_4 cropland maps at high spatial resolution to improve cropland GPP estimates from the individual pixel to country-wide scales. Although there are several existing global C_3/C_4 maps, they are relatively coarse in spatial resolution and produced only for a specific year. An early study developed a static C_3/C_4 fraction map with a spatial resolution of 1°×1° by defining the favorable climate zones for C_3/C_4 and combing the global spatial distribution of crop fractions and national harvest area data for major crop types (Still et al. 2003). Another study developed a global distribution map of croplands and pastures at a 5 min by 5 min (~10 km) spatial resolution in 2000 by combining agricultural inventory data and satellite-derived land cover data (Ramankutty et al. 2008). Recently, several studies made very limited progress in mapping C_3 and C_4 plants (both croplands and grasses) when using remote sensing data and simple algorithms (Foody and Dash 2007; Wang et al. 2013). The development of CDL datasets include the use of satellite-based imagery, supervised image classification methodology, and numerous high-quality ground truth data collected to help determine the multi-spectral rules from time-series imagery that best predicted the land cover category. For grasslands, it was reported that there was a strong linear relationship between the percentage of C_3 grass and the season-long cumulative vegetation index (Foody and Dash 2007). These phenological features and time-series MODIS data were used to classify C_3 and C_4 grasslands in the Great Plains (Wang et al. 2013). Given the importance of C_3 and C_4 plant function types in estimating GPP, it is important for the remote sensing community to increase its effort in mapping C_3 and C_4 croplands and grasslands at site, regional, and global scales.

3.4.2 The timing and location of climate extremes and their impacts on terrestrial ecosystems

Climate extremes such as heatwaves and droughts can reduce vegetative growth, trigger large-scale tree mortality, and turn terrestrial ecosystems from carbon sinks into sources (Ciais et al. 2005; Yuan et al. 2016a). The warm spring and hot and dry summer in 2012 over CONUS offered a unique opportunity to investigate several major questions on the impacts of climate extremes on terrestrial carbon cycle at the regional and continental scales (He et al. 2018c; Sippel et al. 2016; Wolf et al. 2016). Many studies have reported that terrestrial ecosystems in CONUS have served as carbon sinks in recent decades (Hurtt et al. 2002; Pacala et al. 2001), ranging from 0.30 to 0.58 Pg C per year during the

1980s and 1990s, which accounts for 30% of fossil-fuel emissions from the USA. Wolf et al. (2016) analyzed MOD17 GPP data and NEP data from CTE2014 during 2001–2012 and reported that the increase of NEP in the spring compensated for the loss of NEP in the summer, which resulted in a small carbon sink (0.11 Pg C year^{-1} in 2012) for CONUS. This result suggests the importance of assessing the impacts of climate extremes, which depend on timing, duration, and location, on terrestrial carbon budgets at the annual and continental scales (Sippel et al. 2017; von Buttlar et al. 2017).

Wolf et al. (2016) analyzed MOD17 GPP data in 2001-2012 and reported that GPP loss in summer in 2012 over CONUS was twice as large as the increase in GPP in the spring of 2012, resulting in a large annual loss of GPP (-0.38 Pg C). Though we used a different baseline, our analysis of MOD17 GPP data in 2008–2014 also shows that the decrease in GPP in the summer of 2012 was substantially larger than the increase in GPP in the spring of 2012, resulting in large annual loss of GPP (-0.12 Pg C) (Fig. 8). However, the results from GPP_{VPM}, $GPP_{SiBCASA}$, and GPP_{CASA} showed that the GPP increase in the spring is close or slightly larger than GPP loss in the summer of 2012, the annual GPP anomaly ranging from 0.01 ($GPP_{SiBCASA}$), to 0.11 Pg C (GPP_{VPM}), to 0.18 Pg C (GPP_{CASA}), while the GOME-2 SIF anomaly showed a decrease in 2012 (Fig. 3.7). The differences in modeling GPP responses to spring warming and summer drought among these four models are likely to affect our understanding of the responses of ecosystem respiration (ER) to spring warming and summer drought. As NEP is the sum of GPP (carbon gains) and ER (carbon losses), the large decrease in GPP (e.g., -0.38 Pg C year^{-1} in 2012, GPP_{MOD17}) from the previous study (Wolf et al. 2016) implied a slightly larger decrease in ER, which could then result in a small carbon sink (0.11 Pg C year^{-1} in 2012). In addition, since CASA-GFED3 and SiBCASA-GFED4 are the biosphere models used by CarbonTracker (CT2014) and CarbonTracker Europe (CTE2014) to generate prior biosphere carbon fluxes, the spatial-temporal differences in GPP distribution, magnitude, and anomaly from these two models are likely to affect CarbonTracker and CarbonTracker Europe outputs. Previous studies have reported that atmospheric CO_2 inversions are sensitive to the land surface prior fluxes, especially at fine scales and the areas with sparse or no available observations (Peylin et al. 2013; Zhu et al. 2014). Therefore, methods to incorporate more reliable carbon flux estimates from atmospheric CO_2 inversions is critically needed for us to better understand the terrestrial carbon cycle.

3.4.3 Differential responses to climate extremes across biomes

Numerous studies have reported the negative impacts of high temperature and droughts on vegetation productivity (Ciais et al. 2005; Welp et al. 2007; Wolf et al. 2016; Yuan et al. 2016a). Short-term drought or heatwaves lead to stomatal closure, membrane damage, and disruption of photosynthetic enzyme activities, all of which reduce GPP. If plants experience continuous drought, they may respond to drought stress by structural or physiological adjustments such as decreased leaf area index, changes in the root-shoot ratio, or changes in leaf angle (Frank et al. 2015). But different species have adopted different strategies to deal with water stress. These strategies can be broadly classified as dehydration tolerance or dehydration avoidance (Bacelar et al. 2012). Plants with a dehydration tolerance strategy usually grow rapidly when water is available but will senesce and/or become dormant during drought. Plants with a dehydration avoidance strategy tend to grow more slowly and maintain greenness during drought by increasing water extraction from the soils and reducing water loss from transpiration. Our study showed that the impacts of spring warming and summer drought on the change in GPP varied across biomes (Fig. 3.10). This change was not only due to the characteristics (timing, magnitude) of the heatwaves and drought at specific regions (Fig. 3.11), but also species-specific plant drought responses and strategies (von Buttlar et al. 2017; Wolf et al. 2014). Our results show that grasslands experienced the largest reduction in GPP while forests had the largest increase. This difference may be explained by the observation that grasslands are drought sensitive, and more susceptible to heatwaves and droughts as they have less accessibility to soil water (shallow roots) and higher turn-over rates (Frank et al. 2015). Trees usually have deeper roots and better access to soil water, thus forests are considered to be less affected by heatwaves and drought (Frank et al. 2015; van der Molen et al. 2011; Zhang et al. 2016b). Grasslands occur in the most severe drought-affected areas, while most forests are in the northwestern and eastern part of CONUS, which were either not affected by the 2012 drought or were classified as abnormally dry (D0) by the U.S. drought monitor (Fig. 11). Cropland systems are different from natural systems by the frequent human intervention (for example, irrigation or changing planting date). Consequently, the impacts of climate extremes on croplands are expected to be highly modulated by human management (Lobell et al. 2012; van der Velde et al. 2010). However, cropland over the Corn Belt, the most important crop area in the US, is mainly rainfed (Leng et al. 2016), leading to a similar GPP response to drought for cropland and grasslands.

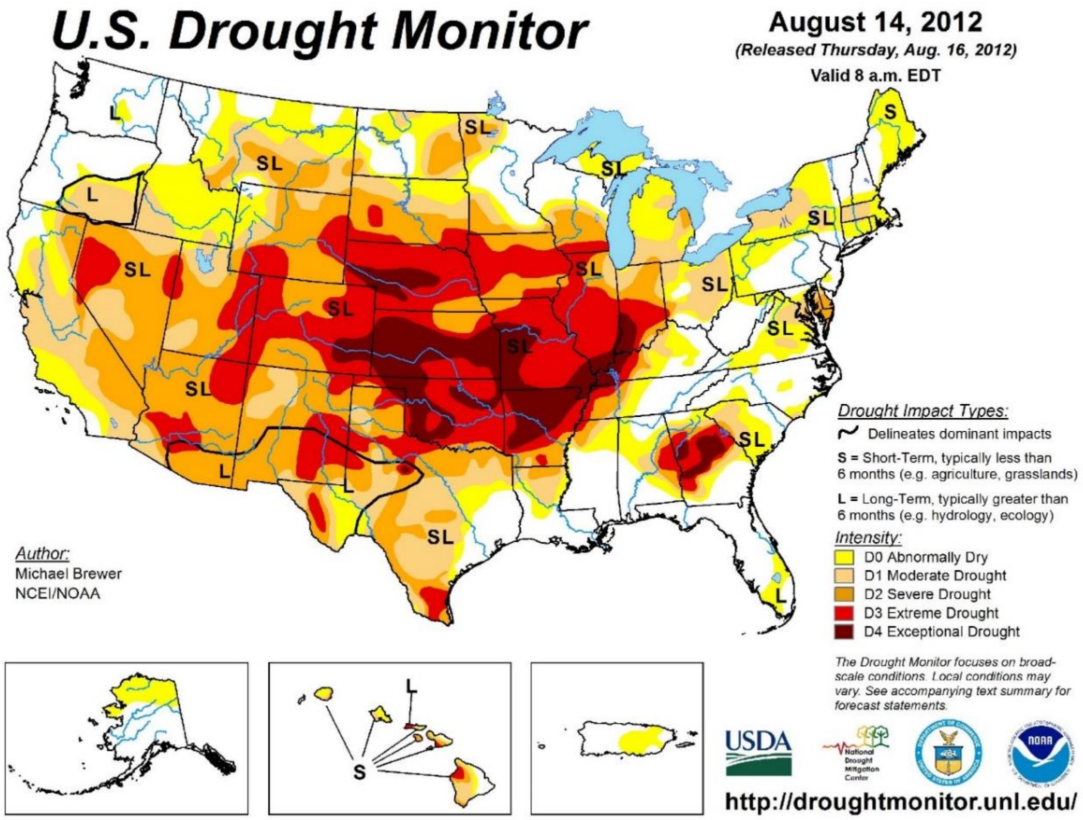

Figure 3.17 Drought-affected areas over CONUS on August 14, 2012

3.4.4 Uncertainties and remaining issues

The uncertainty of ecosystem models remains a challenge for carbon cycling research. Extreme climate events were found to dominate the global interannual variability of GPP (Zscheischler et al. 2014). At present, most ecological models do not accurately represent the responses of major ecosystem processes to climate extremes and do not accurately track the interannual variability of GPP (Reichstein et al. 2013). For example, previous studies indicated that improving GPP estimates for most models requires better representation of water stress effects on photosynthesis (Schaefer et al. 2012; Verma et al. 2014; Yuan et al. 2014). In this study, VPM, MOD17, and CASA are all light use efficiency models, but use different water regulation scalars. VPM uses a water-related vegetation index (LSWI) as the water constraint, MOD17 uses vapor pressure deficit (VPD), and CASA uses the evapotranspiration supply/demand ratio (actual evapotranspiration/potential evapotranspiration). LSWI is found to be a good indicator of soil moisture when taking all the biomes into consideration (Zhang et al. 2015). However, it may not work well for forested areas because of the lower spectral sensitivity to water stress (Sims et al. 2014). VPD represents the impacts of atmospheric dryness on vegetation photosynthesis because stomatal conductance changes with VPD. However, soil moisture also plays an important role in regulating GPP by affecting leaf cell turgor pressure or stomatal conductance, thereby directly affecting photosynthesis (Hashimoto et al. 2013; Leuning et al. 2005). The evapotranspiration ratio requires well simulated hydrologic fluxes in soils where additional information (e.g. soil texture, soil/rooting depth) is required. This information is usually not easy to collect and comes with uncertainties. Therefore, more effort is needed to quantify the model uncertainties and improve model structure.

Since SIF can be directly observed from space, has a very good relationship with GPP (Guanter et al. 2014; Wagle et al. 2016; Zhang et al. 2016a), and is a good indicator of agricultural drought (Sun et al. 2015), we used SIF as a reference to which we compared the impacts of spring warming and summer droughts on vegetation photosynthesis. However, we acknowledge that GOME-2 SIF has some uncertainties, especially in the western part of CONUS (Fig. 3.6) due to the relatively large signal-to-noise ratio (Zhang et al. 2016a). SIF retrievals from recently launched satellites (OCO-2, Sentinel-5 Precursor, and FLEX-Fluorescence Explorer) with higher spatial resolutions and observations tailored for SIF may improve our understanding of the impacts of climate extremes on vegetation.

In this study, we only considered the impacts of climates extremes on terrestrial ecosystems within a year. However, droughts may affect terrestrial ecosystems across months or even years, depending upon plant functional types (Frank et al. 2015; von Buttlar et al. 2017). Extreme events could cause plant functional loss, changes in the community structure of ecosystems, increased wildfires, and pest and pathogen outbreaks, all which may necessitate a long recovery period (van der Molen et al. 2011). Further, species' response to climate extremes vary widely, and some impacts could persist long after extreme events (Rammig et al. 2014). Analysis of the responses of terrestrial ecosystems to climate extremes should be conducted over the next few years.

3.5 Conclusion

The spring warming and summer drought of 2012 across CONUS had substantial impacts on the terrestrial carbon cycle and offered a unique opportunity to investigate the responses of photosynthesis (GPP) and respiration processes at large scales. We presented an improved VPM model that incorporates C_3 and C_4 croplands and can better capture the seasonal dynamics and interannual variation of GPP than the MOD17 product when these models are compared to GPP_{EC} data from eddy covariance flux tower sites. Spatial-temporal comparisons among GOME-2 SIF, GPP_{MOD17}, and GPP_{VPM} products during 2008–2014 showed strong consistency between GOME-2 SIF and GPP_{VPM} data products. Anomaly analyses of (1) annual GPP from three other models (VPM, SiBCASA, and CASA) and (2) GOME-2 SIF data between the baseline years (2008, 2009, 2010, 2013, 2014) and drought year 2012 suggested that increased GPP during the warm spring compensated for decreased GPP during the dry and hot summer, resulting in close to net neutral changes in annual GPP. The results from this study clearly highlight the importance of assessing the impacts of co-occurring climate extremes at seasonal and annual scales over large spatial domains. Our results demonstrate the need to further improve GPP models, which could increase the accuracy and reduce uncertainties in GPP estimates of terrestrial ecosystems.

Chapter 4: Spatial-temporal dynamics of maize and soybean planted area, harvested area, gross primary production, and grain production in the Contiguous United States during 2008-2018

Abstract

The United States of America ranked first in maize export and second in soybean export among the countries in the world. Accurate and timely data and information on maize and soybean production in the Contiguous United States (CONUS) are important for food security at the regional and global scales. There is a need to better understand the spatial-temporal dynamics of maize and soybean production in the CONUS under changing climate, land use, and market. In this study, we evaluate the interannual dynamics of maize and soybean planted area and harvested area in the CONUS during 2008-2018. We find the increases of maize and soybean planted areas in mid-2010s, driven by markets and international trade. The results also show that severe summer drought in 2012 had little impact on soybean planted and harvested areas and maize planted area, but it substantially reduced maize harvested area and grain production. We use the Vegetation Photosynthesis Model (VPM), the Crop Data Layer (CDL), climate, and image data to estimate 8-day gross primary production (GPP) of maize and soybean in the CONUS during 2008-2018 (GPP_{VPM}). Annual GPP_{VPM} (GPP_{VPM_Year}) had strong linear relationships with maize and soybean grain production from the agricultural statistic data at the county scale. The Harvest Index, defined as the ratio between grain production and GPP_{VPM} (HI_{VPM_GPP}), ranged from 0.25 (2012) to 0.36 for maize and from 0.13 to 0.15 for soybean. The linear regression models between grain production and cumulative GPP_{VPM} (GPP_{VPM_acc}) over time at 8-day resolution show that by the end of July,

$\mathrm{GPP}_{\mathrm{VPM\_acc}}$ accounted for ~90% of variance in maize and soybean grain production during 2008-2018 at the county scale, which was approximately two months before farmers started to harvest maize and soybean. Our findings suggest that the VPM and $\mathrm{GPP}_{\mathrm{VPM}}$ data product are useful tools and data for farmers, decision makers, stakeholders and the public.

4.1 Introduction

Crop production and food security is one of fundamental challenges in our society due to the rising global population, dietary change, and increasing biofuel production that uses crops as feedstock (Ray et al. 2013). Maize (*Zea maize, L*) and soybean (*Glycine max*) are two of the major sources of caloric energy for human and are critical for world food supply. The United States of America (USA) is the largest maize and soybean producer in the world (Meade et al. 2016) and ranked first in maize export and second in soybean export in the world. Inter-annual variation of maize and soybean area and grain production in the USA affect the world grain trade market (Gardiner 2016). Therefore, accurate and timely information and knowledge on planted area, harvested area, grain production and grain yield of maize and soybean in the USA is crucial for agriculture, food security, and international trade (Iizumi and Ramankutty 2015; Tilman et al. 2011).

Crop production, the amount of specific crop produced in one calendar year in a region, is the product of crop harvested area and crop yield. For crop production estimation, it is essential to have information of both crop area and crop yield. The National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) provide annual crop statistical reports for crop planted area, harvested area, grain yield, and grain production in a year at various administration levels (e.g., national, state and county). The data collection through sample-based agricultural surveys is not only time consuming and costly but also has long time lags and data gaps (Doraiswamy et al. 2003). Satellite-based remote sensing has been used to monitor cropping area, grain yield, and grain production since the early 1970s (Atzberger 2013; Fritz et al. 2019; Lobell 2013). Notable progress has been made in satellite-based mapping of cropland areas (planted area and/or harvested areas) (Cai et al. 2018; Massey et al. 2017; Wang et al. 2019a; Wardlow and Egbert 2008; Zhong et al. 2014) at various spatial scales. A number of studies reported annual maps of maize and soybean area for a few states or counties in the corn-belt region (Cai et al. 2018; Wang et al. 2019a; Wardlow and Egbert 2008; Zhong et al. 2014). A few studies generated annual maps of croplands over the contiguous United States (CONUS) with a spatial resolution of 250-m (Massey et al. 2017). The overall accuracy of those crop maps ranged from 60 to 96% (Cai et al. 2018; Massey et al. 2017; Wang et al. 2019a; Wardlow and Egbert 2008; Zhong et al. 2014). The USDA/NASS also generated the Cropland Data Layer (CDL) product (Boryan et al. 2011), which includes all major crop types in CONUS. The CDL dataset that covers all the states in CONUS at 30-m spatial resolution started in 2008, and it is widely used in the studies of crop yields in various states (Guan et al. 2016; He et al. 2018a; Marshall et al. 2018). To date, no study has investigated the spatial-temporal dynamics of maize and soybean planted area, harvested area, and grain production in the entire CONUS during 2008-2018, which are likely affected by weather and climate (e.g., drought, flood), market and prices.

Grain yield (ton/ha) and grain production (ton) of maize and soybean crops are related to growing conditions and crop properties including leaf area index, aboveground biomass, gross and net primary production (GPP, NPP), which can be estimated by satellite images and models (Guan et al. 2016; He et al. 2018a; Marshall et al. 2018; Sakamoto et al. 2014; Xin et al. 2013). Several studies evaluated the relationship between vegetation indices and grain yields of maize and soybean at county scale from the NASS crop statistics (Becker-Reshef et al. 2010; Burke and Lobell 2017; Johnson 2016). Some studies used vegetation indices to estimate crop aboveground biomass and then used the biomass-based harvest index (HI_{AGB}), which is calculated as the ratio between grain yield and crop aboveground biomass, to estimate grain yield, and then compared the resultant yield estimates with the yield data from the NASS crop statistics at county scale (Guan et al. 2016; Lobell et al. 2002). GPP can be estimated by using a light use efficiency (LUE) model driven by remote sensing images and climate data, and several GPP data products are available to the public (Running et al. 2004; Wu et al. 2018; Zhang et al. 2017). Some studies used the model-based GPP and harvest index to estimate grain yield, and then compared the resultant yield estimates with the yield data from the NASS crop statistics, for example, croplands in the Midwest during 2009-2012 (Xin et al. 2013), and croplands in the CONUS during 2010-2015 (Marshall et al. 2018) at county scale. These studies at county scale reported reasonably good relationship between the model-based yield estimates and the yield data from NASS crop statistics, with a range of R^2 values from 0.5 to 0.7. Several studies showed that harvest index (HI_{AGB}) varies among crop types and environment (Hay 1995; Lobell et al. 2002), which may affect the yield estimates in those studies. In addition, note that NASS crop statistics is based on the samples in a county, thus the distribution of these samples would affect the yield estimates in a county. One study used model-based GPP and harvest index (HI_{GPP}, calculated as the ratio between GPP and grain production) to estimate grain yields of several crops and compare them with the yield data from the NASS statistics reports for Montana during 2008-2015 (He et al. 2018a). To date, no study has directly evaluated the relationship between GPP and grain production at county and state scales in CONUS during 2008-2018, which are also affected by weather, climate, land use and land management (planted area and harvested area).

In this study we addressed three research questions on maize and soybean croplands in the CONUS. First, we asked what the spatial-temporal dynamics of maize and soybean planted area, harvested area, and grain production in the CONUS during 2008-2018 are? We analyzed the agricultural statistical data of maize and soybean (planted area, harvested area, grain production) from the USDA NASS and satellite-based planted area of maize and soybean from the USDA Cropland Data Layer (CDL) dataset during 2008-2018. The analysis will show the interannual variations (increase, decrease, no change) of maize and soybean planted area, harvested area and grain production during 2008-2018, and assess the effect of severe drought (e.g., summer drought in 2012) on maize and soybean planted area, harvested area, and grain production at county, state and national scales. Second, we asked what are the relationships between GPP and grain production at county scale during 2008-2018? We analyzed the GPP data from the satellite-based Vegetation Photosynthesis Model (VPM) and MOD17 algorithm, and grain production data from the NASS. The analysis quantified the relationships between maize and soybean GPP and grain production from NASS crop statistics at county-, state-, and national-level under varying climate conditions (drought year versus normal years). During 2008-2018, severe drought and heat wave events occurred in various regions of CONUS, for example, the 2012 summer drought, which was reported as one of the worst droughts since 1988. Third, we asked to what degree GPP can be used to estimate grain production in a year? We calculated cumulated GPP_{VPM} over the maize and soybean growing season and analyze the relationship between cumulated GPP_{VPM} and grain production of maize and soybean in CONUS at county scale. The analysis will develop a simple linear regression model, which is based on the relationship between GPP_{VPM} and grain production (harvest index, defined as HI_{GPP}) which can be used to forecast (early prediction) grain production of maize and soybean croplands in CONUS before farmers start to harvest maize and soybean crops.

4.2 Materials and methods

4.2.1 *Study Area*

The study area is the contiguous United States (CONUS). It covers 48 states and 3,233 counties. Climate in the CONUS ranges from subtropical climate in the southern region (e.g., Florida) to temperate climate in the northern region. Crop cultivation is mainly concentrated in the Great Plains (wheat, maize, soybean) and in the regions around the Great Lakes known as Corn-Belt (maize, soybean). Crop cultivation is dominated by a single crop per year, and major crop types include maize (~35%), soybean (~33%), winter wheat (~22%), and sorghum (~3%).

4.2.2 *USDA-NASS statistical data of cropland planted area, harvested area, and grain production during 2008-2018*

The annual county-, state- and national- level statistics data of crop planted area (acre), harvested area (acre), grain production (bush) and grain yield (bu/acre) for maize and soybean during 2008-2018 were downloaded from the USDA-NASS Quick Stats Database (https://quickstats.nass.usda.gov/). For summer crops, the NASS planted and harvested area estimates were mostly based on the June Agricultural Survey (JAS) data. During the first two weeks of June, producers in the designated sample farms are asked by investigators about the acreage and other information by crop, including planted and/or intend-to-plant areas, and the acreage they intend to harvest (USDA 2014). The yield statistics were based on two large panel surveys that are annually conducted throughout the growing season. One is the Agricultural Yield Survey (AYS), which is based on farmers' reported yield information for most crops. Each year, a subsample of farmers who responded to the list portion of the JAS, are contacted monthly by phone during the growing season (August to November) and asked to provide expected crop yield. The other is the Objective Yield Survey (OYS), which provides independent yield estimates by aggregating field biophysical crop measurements into a model (USDA 2014). These biophysical crop measurements, such as plant counts per unit area, grain size, were sampled in the fields across the major crop growing areas. The OYS is very costly and is conducted only in the top crop production states. Ultimately, the results from both the AYS and OYS surveys are analyzed by the NASS Agricultural Statistics Board (ASB) to establish the yield estimates. The NASS crop grain production is estimated from the expected harvested area and yield at the survey reference date and predicted assuming normal conditions for the remainder of the crop growing season. For the unit of cropping area, we converted acres to hectares by applying a scale factor of 0.405. For the unit of grain production, we converted bushel to metric ton by using a scale factor of 0.0254 and 0.0272 ton/bu for maize and soybean when, respectively.

4.2.3 *Cropland area dataset from the USDA-NASS cropland data layer (CDL) at 30-m spatial resolution during 2008-2018*

The annual CDL dataset at 30-m spatial resolution is a remote sensing-based land cover product. The CDL product utilizes both in-situ ground reference data and multiple satellite imagery to identify and map field crops. The major sources of agricultural and non-agricultural ground reference data, which were used as training data in the supervised classification, includes the USDA's Farm Service Agency (FSA) Common Land Unit (CLU) data and the National Land Cover Dataset (NLCD) 2001. The CLU-based data were collected in every growing season when producers reported crop types and crop acreage in their fields to the FSA county offices. The major remote sensing images used by the CDL classifier include AWiFS, Landsat TM and ETM+, Deimos-1 and UK-DMC-2 and MODIS satellite data. Before 2009, the 56-m AWiFS data was the primary source of imagery and the 30-m Landsat data were used as supplementary source because AWiFS has a higher temporal resolution of 5-day than Landsat (16-day), which provides the opportunity for having more cloud-free observations throughout the crop growing season. The resultant CDL data in 2008-2009 had a spatial resolution of 56-m. During the growing season of 2009, some technical issues happened in AWiFS and the increased competition from international customers disrupted the continuing use of AWiFS data. As a result, Landsat became a primary source of images after 2009 (Boryan et al. 2011).

The CDL data in 2010-2018 has a spatial resolution of 30 meter. In 2018, the CDL data from 2008-2009 were aggregated to 30 meters. In this study, we used the 30-m CDL data throughout 2008-2018 to keep our analysis in a consistent spatial resolution.

The CDL data include more than 100 crop types, with classification accuracy higher than 90% for major crops (maize, soybean and winter wheat) (Boryan et al. 2011). It has been widely used in applications related to land use and land cover change, agricultural sustainability, and agricultural production decision-making. To make use of the dataset more effective and efficient, CropScape, an interactive Web-based CDL data portal, was developed to visualize, query, and analyze CDL data through standard geospatial web services in a publicly accessible online environment (Han et al. 2012). In this study, CropScape was used to obtain the acreage statistics for all the crop categories for each county during 2008-2018. We calculated the annual total area for maize and soybean for each county as the sum of both the single cropping and double cropping area, respectively. The state and national total areas for each crop in each year was then calculated by adding the annual total area over all the counties in the state and nation.

4.2.4 *The input datasets for regional simulation of the Vegetation Photosynthesis Model*

The input datasets for simulations of the VPM model include climate (air temperature and radiation), vegetation indices (VIs), and land cover data. The NCEP climate dataset (https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html) was used for simulations of VPM at the global scale (Zhang et al. 2017) and the NCEP/NARR climate dataset was used for simulation of VPM in the CONUS (Wu et al. 2018). The original 3-hourly NARR data with a spatial resolution of 32 km were first aggregated into daily maximum/minimum/mean air temperature, daily daytime mean air temperature, and daily shortwave radiation. The resultant daily data were further aggregated to 8-day intervals to match the MODIS data by calculating the averages for air temperature and the sum for downward shortwave radiation in each 8-day period. The 8-day climate data with a coarse resolution of 32 km were then interpolated to 500-m by applying a weighted distance factor to the nearest four grid cells (Wu et al. 2018; Zhang et al. 2017).

Two vegetation indices, Enhanced Vegetation Index (EVI) and Land Surface Water Index (LSWI), were derived from the MOD09A1 surface reflectance data (Collection 6). EVI and LSWI were processed with a rigorous quality check by identifying those observations affected by cloud, cloud shadow and aerosol as bad-quality observations, based on the quality assurance layer (QA) in the MOD09A1 dataset (Zhang et al. 2016a). The bad-quality observations in the EVI time series data were then gap-filled by applying the “Best Index Extraction Algorithm” (BISE) (Zhang et al. 2017). In this algorithm, a standard seasonal pattern for each pixel was first generated by extracting the median values of all the valid data for each day of year (DOY) across 19 years (2000-2018). Gaps were then filled with a linear interpolation and smoothed using a Savitzky-Golay filter. The International Geosphere-Biosphere Programme (IGBP) land cover classification from MOD12Q1 is used to provide information of biomes. In this study, we also used the CDL dataset to determine the percentage of maize and soybean within each 500-m MODIS pixel (see section 2.4).

4.2.5 *GPP data from the Vegetation Photosynthesis Model (VPM)*

The VPM model is a light use efficiency (LUE) and estimates daily GPP as a product of LUE and the amount of photosynthetically active radiation absorbed by chlorophyll in the canopy ($APAR_{chl}$). The detailed description of the VPM model can be found in previous publications (Xiao et al. 2004a; Xiao et al. 2004b) (Zhang et al., 2017). We used the improved VPM model, which differentiates between C_3 and C_4 crops based on the cropland data layers.

$$GPP = APAR_{chl} * LUE \qquad (4.1)$$

$$APAR_{chl} = FPAR_{chl} * PAR \qquad (4.2)$$

$$LUE = LUE_0 * T_{scalar} * W_{scalar} \qquad (4.3)$$

$$LUE_0 = LUE_{0\_C3} * \%C_3 + LUE_{0\_C4} * \%C_4 \qquad (4.4)$$

where $\%C_3$ and $\%C_4$ are the area fractions of C_3 and C_4 plants within a pixel, LUE_{0_C3} and LUE_{0_C4} are the maximum LUE values for C_3 and C_4 plants, respectively.

A number of publications have reported the evaluation of GPP_{VPM} with GPP_{EC} estimates from the eddy flux tower sites, including maize (Dong et al. 2015; Kalfas et al. 2011), soybean (Jin et al. 2015; Wagle et al. 2015), winter wheat (Doughty et al. 2018; Yan et al. 2009) and paddy rice (Xin et al. 2017). All these publications reported strong agreement between GPP_{VPM} and GPP_{EC}, with a range of R^2 values from 0.70 to 0.98. We also ran VPM simulations with NCEP/NARR climate data, MODIS images and the CDL data to estimate daily GPP in the CONUS during 2008-2014 (Wu et al. 2018). We compared the resultant GPP_{VPM} data with the other three GPP data products: GPP from the MOD17 data product (GPP_{MOD17}), GPP from CASA model (GPP_{CASA}), and GPP from SiBCASA ($GPP_{SiBCASA}$) (Wu et al. 2018), and the results showed that GPP_{VPM} had the stronger relationships with GOME-2 solar-induced chlorophyll fluorescence (SIF) data in both normal years and drought year (2012) than do the other three GPP datasets.

In this study, we ran the VPM simulations with NCEP/NARR climate, MODIS images and the CDL data during 2008-2018 to generate 8-day GPP data (g C m^{-2} day^{-1}) at 500-m spatial resolution. Each pixel has 46 GPP_{VPM} data in a year and we calculated annual sums of GPP_{VPM} for individual pixels. We also aggregated annual sums of GPP_{VPM} by crop types and their planted areas. The average annual GPP (GPP_{VPM_avg}) values of individual crop types (maize, soybean) in a county were calculated by the area fraction of specific crop type from the CDL datasets in each 500-m pixel. The total annual GPP (GPP_{VPM_Year}) (January - Dececmber) for a crop type for each county was calculated by multiplying GPP_{VPM_avg} with the total area of all those pixels with a specific crop located in the county.

4.2.6 *MOD17 GPP dataset*

We used the MOD17 GPP product in this study (Running et al. 2004), MOD17A2H V006 (GPP_{MOD17}). GPP_{MOD17} has the same spatial resolution (500-m) and temporal resolution (8-day) with GPP_{VPM}. In the MOD17 data product, GPP is estimated as the product of $APAR_{canopy}$ and light use efficiency (LUE) (Running et al. 2004), but the land cover data product used in the MOD17 does not differentiate between C_3 and C_4 croplands for the CONUS, and C_3 LUE_0 parameter value was used for cropland simulations (Wu et al. 2018). For simple comparison purpose, we also calculated the average annual MOD17 GPP (GPP_{MOD17_avg}) values of individual crop types (maize, soybean) and the total annual GPP (GPP_{MOD17_Year}) (Jan - Dec) in a county and state, based on the CDL crop type data.

4.2.7 *Statistical analyses*

Simple linear regression models were used to characterize the relationship between grain production and GPP, and between grain production and cropping areas at county, state and national scales, at a minimum 0.05 significance threshold (*p*-value). Model performance was evaluated using the coefficient of determination (R^2), bias and RMSE.

A number of studies have used the vegetation indices over time to predict grain yields of crops in a field or a county (Peng et al. 2018; Zhao and Lobell 2017). In this study, we used the cumulative GPP over time to predict grain production in a county. A simple linear regression model was used to assess the relationship between grain production and cumulative GPP over time at 8-day temporal resolution at county-scale (see Equation 4.5). The model was run at 8-day time step over a year across all the counties in the CONUS during 2008-2018. We calculated the averaged R^2 value among all the counties at each time step, and then plot the R^2 values as a function of time. Based on the time course of R^2 value in a year, we assess the performance of using cumulative GPP_{VPM} to predict grain production over time at county scale. This will address the research questions related to in-season grain production forecasting: (1) at what day in a year the model starts to predict grain production at county scale with reasonable accuracy, and (2) to what degree weather and climate (e.g., drought, flood) affect the model prediction over years.

$$Grain\ Production = a * \sum_{1}^{t}(GPP_t \times k) + b \qquad t = 1, 2, 3, \ldots, 46 \tag{4.5}$$

Where t is the number of time steps in a year, which ranges from 1 to 46, as time series GPP has 46 data points in a year; k is the number of days in each time step, k is equal to 8 days when t ranges from 1 to 45, and k is 5 (non-leap year) or 6 (leap year) when t is 46. In this study, we simply used a calendar year schedule to run the statistical models for all counties.

4.3 Results

4.3.1 *Spatial-temporal changes of maize and soybean planted area, harvested area, and grain production during 2008-2018*

At the CONUS scale, Fig. 4.1a,b shows the interannual variation of maize and soybean planted area (CDL, NASS), harvested area (NASS) and grain production (NASS) in the CONUS during 2008-2018. There are very small differences (0.4 - 6.5%, 0.1 - 5.8%) in maize and soybean planted areas between the CDL and NASS datasets (Fig. 4.1a, b), which supports the use of the CDL dataset as input data for model simulations. The differences between NASS planted area and harvested areas are also small, except 2012 for maize crop (Fig. 4.1a, b). We calculated the mean values of planted area, harvested area and grain production over years (excluding the drought year 2012) and the deviation (anomalies) to the mean values for individual years (Fig. 4.1c, d, e, f, g). The anomalies of maize planted area from both CDL and NASS have similar dynamics during 2008-2018 (Fig. 1c,d). The maize planted area gradually increased between 2008 and 2012 and varied moderately over 2013-2018 (Fig. 4.1c). Maize harvested area had a similar anomaly dynamics with the planted area (Fig. 4.1e). The anomaly of annual maize grain production does not follow the same patterns of maize planted area and harvested area (Fig. 1f), which suggests the effect of climate and crop management on grain production. The anomalies of soybean planted area from both CDL and NASS also have similar dynamics during 2008-2018 (Fig. 4.1c,d). Soybean planted area varied slightly during 2008-2013 but started to have large increase in 2014 (Fig. 4.1d). The step-wise increase in soybean planted area in 2014 and 2017 were largely driven by international market demand. The anomaly of soybean harvested area agreed well with that of soybean planted area with a smaller magnitude of variation. The anomaly of soybean grain production had a similar dynamics with the harvested area during 2008-2018, except 2016, a very warm and wet year (National Climate Report - Annual 2016; https://www.ncdc.noaa.gov/sotc/national/201613).

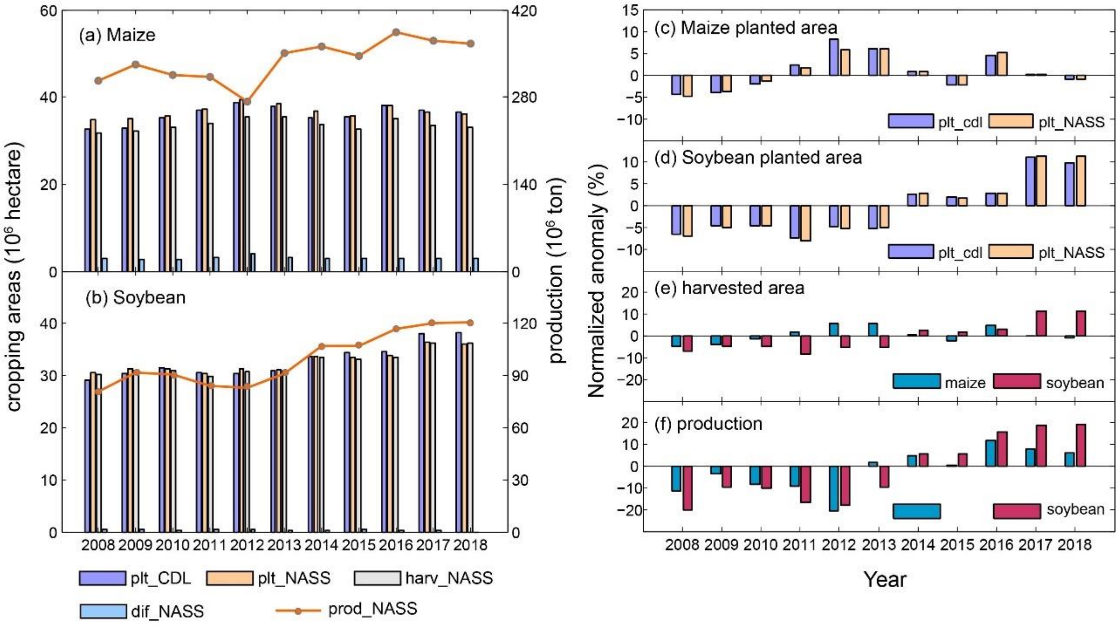

Figure 4.18 Annual national planted area from CDL maps (plt_CDL), planted area from NASS statistics (plt_NASS), harvested area from NASS statistics (harv_NASS), difference between NASS planted and harvested area (dif_NASS), and grain production for a) maize and b) soybean; (c) normalized anomaly of planted area derived from CDL and NASS for maize (d) normalized anomaly of planted area derived from CDL and NASS for soybean (e)normalized anomaly of harvested area from NASS for maize and soybean (f)normalized anomaly of production from NASS for maize and soybean.

At the county scale, Fig. 4.2 shows the spatial distributions of planted area and harvested area of maize and soybean crops in 2010 across all counties of the CONUS. There were strong spatial consistencies in planted areas between the CDL and NASS datasets in CONUS for both maize and soybean (Fig. 4.2), The CDL planted area estimates were highly consistent (only ~1% to 3% discrepancy) with NASS planted and harvested area estimates for both maize and soybean crops during 2008-2018 at the county and state scales (Fig. 4.3). The relationships between CDL planted area and NASS planted areas in individual years were relatively stable at the county and state scales (Table 4.1 and Table 4.2), with moderate differences in 2008 and 2009, which is consistent with previous studies (Lark et al. 2017; Pritsolas and Pearson 2018). These results further support the use of the CDL dataset as input data for model simulations. We further calculated the interannual trend of maize and soybean planted areas from the CDL and NASS datasets during 2008-2018 (Fig. 4.4). A large number of counties in the NASS dataset do not have data for all the 11 years, and they were thus not included in the analysis of interannual trends of planted area. For those counties with 11 years of maize and soybean planted area data, the spatial pattern of the interannual trends (slope values) from the NASS dataset has some similarity with that from CDL dataset (Fig. 4.4). According to the CDL dataset, interannual trends (slope values) of soybean planted area differed to large degree from those of maize planted area (Fig. 4.4). Most counties over the CONUS showed an increase of soybean planted area during 2008-2018, while fewer of them showed such increase of maize planted area. In Illinois, maize planted area decreased while soybean planted area increased during 2008-2018. In the northern Great Plains, maize and soybean planted area increased in recent years.

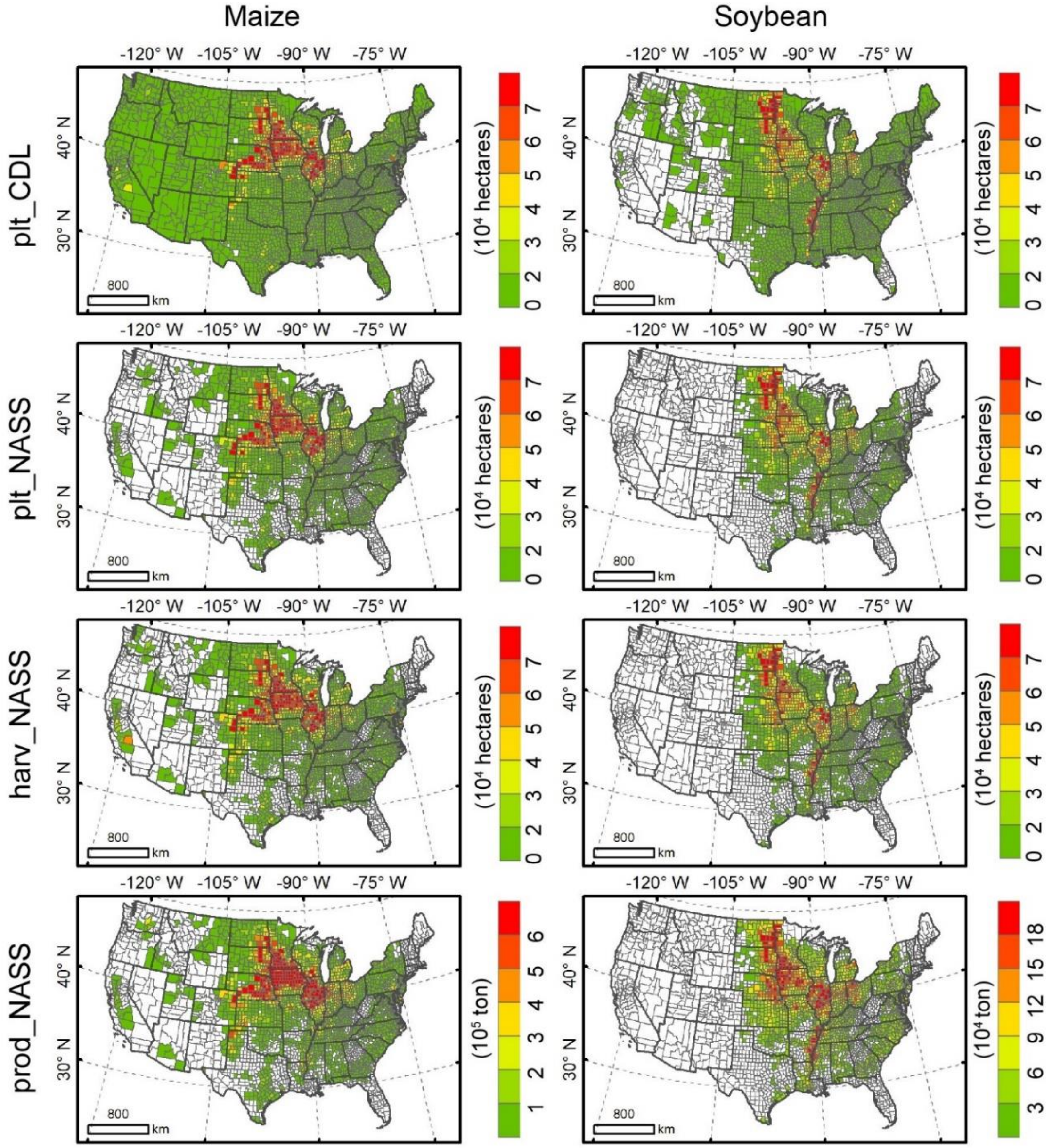

Figure 4.19 Spatial distribution of CDL planted area, NASS planted/harvested area, and NASS production over CONUS in 2010

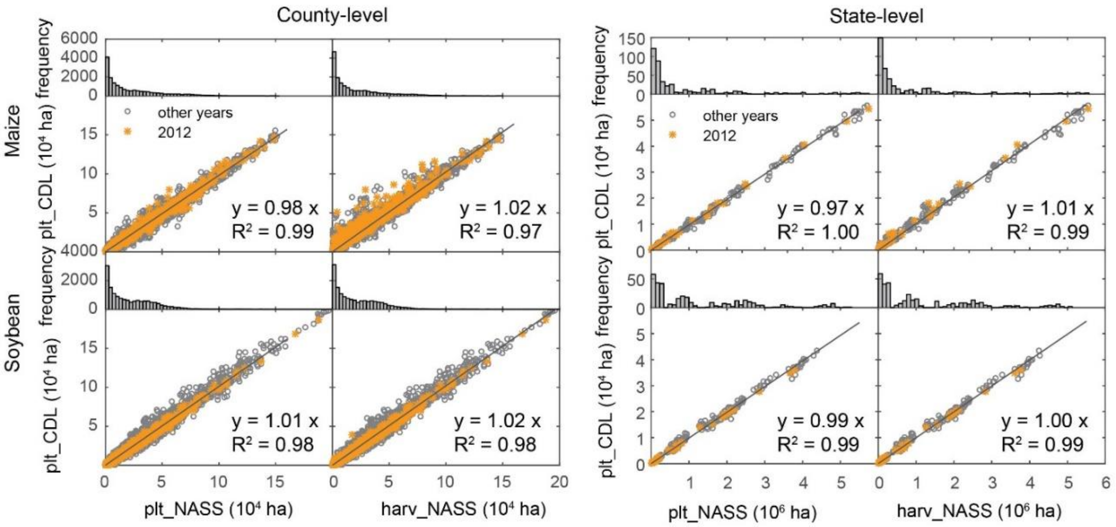

Figure 4.20 Relationship between NASS planted/harvested area and cdl plant area at county- and state- level from 2008-2018

Table 4.5 The regression statistics between the CDL planted area and the NASS planted area and harvested area of maize and soybean in the CONUS during 2008-2018 at the county scale. We used a simple linear regression model $y = a * x$. All the regression models have p-value < 0.001.

Year	Maize							
	plt_CDL vs plt_NASS				plt_CDL vs harv_NASS			
	slope	R^2	bias (ha)	RMSE (10^3 ha)	slope	R^2	bias (ha)	RMSE (10^3 ha)
2008	0.94	0.977	-1362.9	32.72	0.98	0.967	179.7	32.3
2009	0.93	0.984	-1349.7	32.63	0.97	0.975	-26.6	32.24
2010	0.98	0.99	-232.6	32.55	1.02	0.98	1034.6	32.16
2011	1.00	0.992	-62.2	34.59	1.04	0.977	1551.3	34.22
2012	0.98	0.99	-383.8	35.41	1.04	0.966	1660.2	34.64
2013	0.99	0.991	-361	35.67	1.03	0.977	1223.4	35.13
2014	0.97	0.99	-735.3	33.9	1.02	0.98	683.1	33.36
2015	0.99	0.992	-85.3	34.35	1.04	0.977	1426.7	33.88
2016	1.00	0.992	52.5	34.79	1.04	0.978	1548.5	34.37
2017	1.01	0.99	320.1	33.62	1.05	0.973	1813.9	33.19
2018	1.00	0.99	321.8	33.96	1.04	0.971	1819.5	33.55

Year	Soybean							
	plt_CDL vs plt_NASS				plt_CDL vs harv_NASS			
	slope	R^2	bias (ha)	RMSE (10^3 ha)	slope	R^2	bias (ha)	RMSE (10^3 ha)
2008	0.96	0.977	-1126.7	29.68	0.96	0.977	-1126.7	29.68
2009	0.98	0.98	-673.5	30.29	0.98	0.98	-673.5	30.29
2010	1.00	0.986	36.2	29.94	1	0.986	36.2	29.94
2011	1.03	0.986	193.4	29.88	1.03	0.986	193.4	29.88
2012	0.99	0.992	-490.5	29.7	0.99	0.992	-490.5	29.7
2013	1.01	0.988	12.8	30.73	1.01	0.988	12.8	30.73
2014	1.02	0.986	113.5	33.08	1.02	0.986	113.5	33.08
2015	1.02	0.985	666.5	33.04	1.02	0.985	666.5	33.04
2016	1.02	0.987	592.8	32.69	1.02	0.987	592.8	32.69
2017	1.03	0.982	1053.7	35.73	1.03	0.982	1053.7	35.73
2018	1.05	0.984	1573.8	35.77	1.05	0.984	1573.8	35.77

Table 4.6 The regression statistics between the CDL planted area and the NASS planted area and harvested area of maize and soybean in the CONUS during 2008-2018 at the state scale. We used a simple linear regression model y = a * x. All the regression models have p-value < 0.001.

Year	Maize							
	plt_CDL vs plt_NASS				plt_CDL vs harv_NASS			
	slope	R^2	bias (10^3 ha)	RMSE (10^6 ha)	slope	R^2	bias (10^3 ha)	RMSE (10^6 ha)
2008	0.92	0.995	-57.72	1.75	0.97	0.99	14.9	1.73
2009	0.92	0.997	-56.49	1.77	0.96	0.993	10.93	1.75
2010	0.98	0.998	-11.33	1.85	1.01	0.994	54.66	1.83
2011	0.99	0.998	-4.87	1.94	1.04	0.995	74.38	1.92
2012	0.97	0.998	-18.29	1.97	1.04	0.99	79.78	1.92
2013	0.98	0.999	-17.01	1.9	1.03	0.993	60.81	1.87
2014	0.97	0.999	-34.57	1.86	1.01	0.994	38.73	1.84
2015	0.99	0.999	-5.54	1.85	1.03	0.994	65.84	1.83
2016	0.99	0.999	1.69	1.9	1.04	0.994	72.87	1.88
2017	1	0.999	14.09	1.85	1.05	0.992	87.46	1.82
2018	0.99	0.998	14	1.82	1.04	0.992	86.67	1.79

Year	Soybean							
	plt_CDL vs plt_NASS				plt_CDL vs harv_NASS			
	slope	R^2	bias (10^3 ha)	RMSE (10^6 ha)	slope	R^2	bias (10^3 ha)	RMSE (10^6 ha)
2008	0.93	0.992	-51.06	1.46	0.94	0.992	-37.52	1.45
2009	0.95	0.994	-31.93	1.48	0.96	0.992	-17.85	1.48
2010	0.98	0.996	2.98	1.52	0.99	0.995	13.34	1.51
2011	1	0.995	7.22	1.51	1.01	0.995	23.8	1.5
2012	0.96	0.998	-25.36	1.49	0.97	0.998	-11.61	1.49
2013	0.99	0.995	-2.97	1.51	0.99	0.994	4.69	1.51
2014	0.99	0.993	0.85	1.61	1	0.993	9.79	1.61
2015	1	0.993	27.48	1.59	1.01	0.992	39.46	1.58
2016	1	0.995	25.92	1.62	1.01	0.994	35.55	1.61
2017	1.01	0.993	47.97	1.75	1.02	0.993	56.07	1.75
2018	1.03	0.994	67.03	1.75	1.04	0.993	87.57	1.73

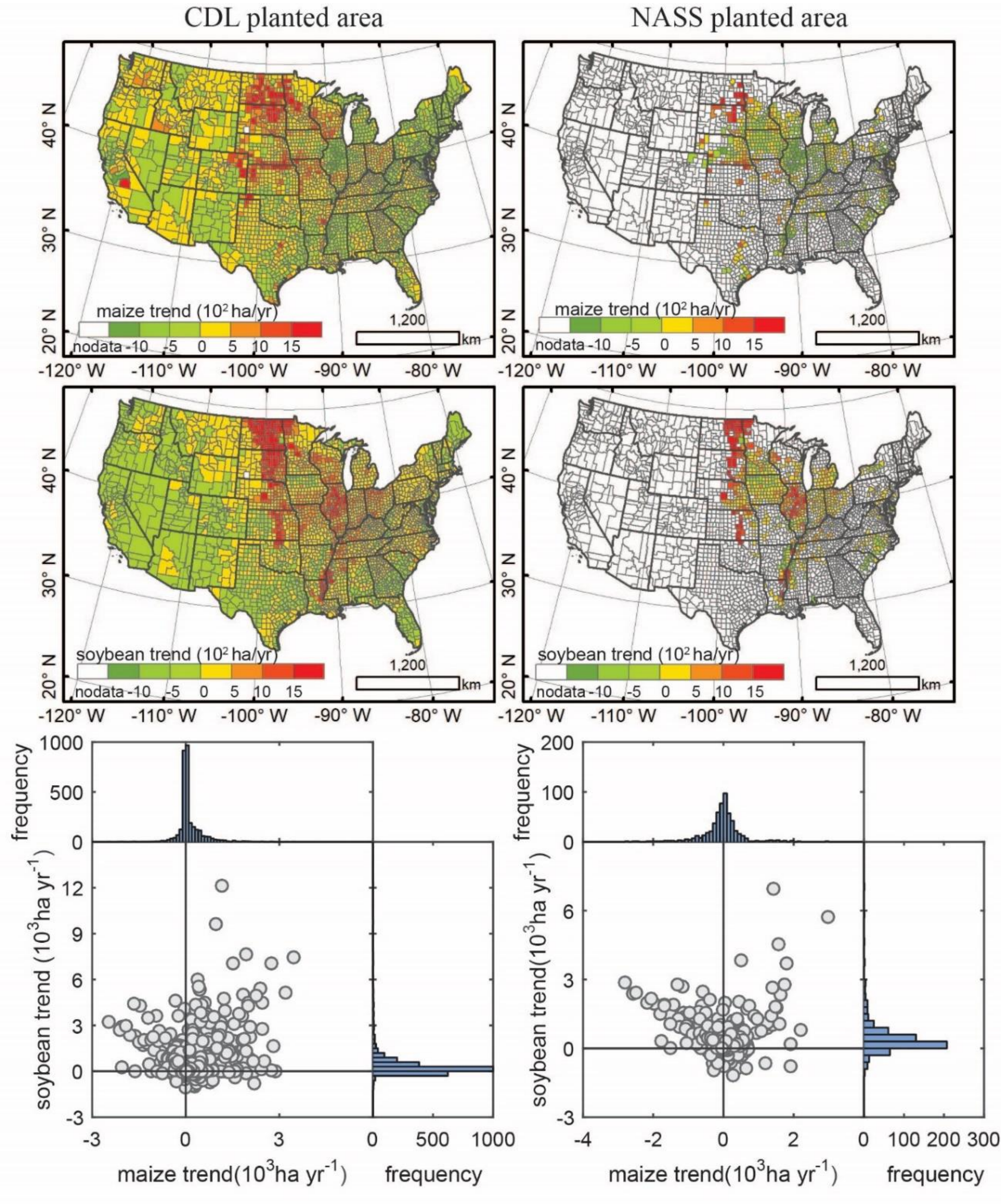

Figure 4.21 Changing trend of planted area for maize and soybean from 2008-2018

Fig. 4.5 shows the relationships among maize and soybean planted area, harvested area and grain production during 2008-2018 at the county and state scales. The slope values in the regression models represents the average yields (ton ha^{-1}) in the CONUS during 2008-2018, which was calculated by either planted area or harvested area and they have very small variations among the methods (Fig. 4.5). The relatively small scattering among the data points (Fig. 4.5) indicate very good management of maize and soybean croplands among the crop producers in the CONUS. For maize, the average yields at the county scale (Table 4.3 & Table 4.4) and the state scale (Table 4.5 & Table 4.6) among individual years had a small variation (< 10%), except for 2012. Because of extreme heat and drought in the summer of 2012, the average yield of maize in 2012 was 8.24 ton ha^{-1} at the county scale (Table 4.3) and 7.91 ton ha^{-1} at the state scale (Table 4.5), which is substantially (more than 20%) lower than multi-year average yields (10.7 ton ha^{-1} or 10.6 ton ha^{-1}). For soybean, the average yields at the county scale (Table 4.4) and the state scale (Table 4.6) also had a small variation among individual years, and the 2012 drought only resulted in ~10% drop in comparison to multi-year average yields. The results indicate that the soybean crop was less impacted than the maize crop in the 2012 severe drought.

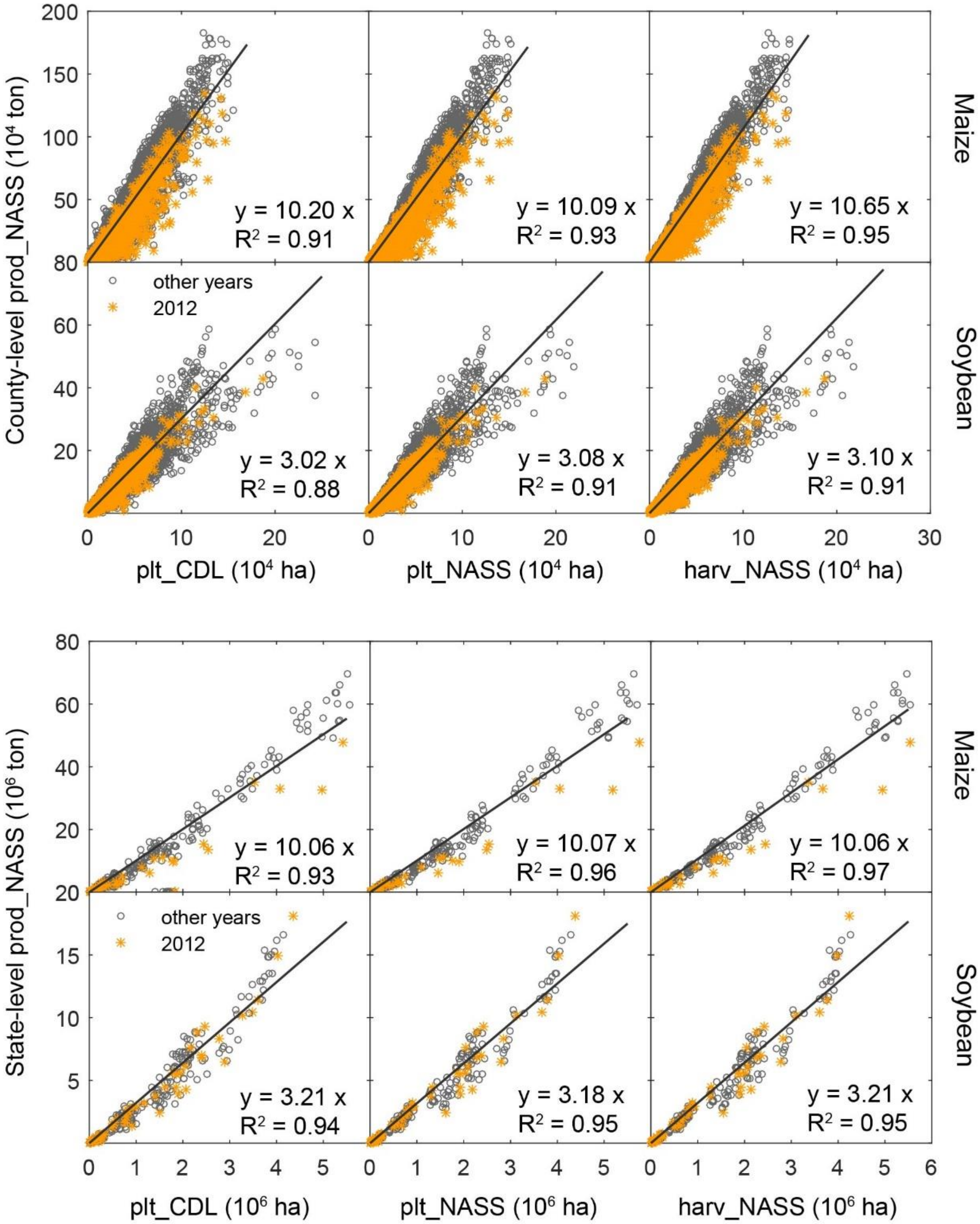

Figure 4.22 Relationship between county- and state- level production and cropping areas from 2008-2018

Table 4.7 The regression statistics between county-level NASS production and cropping areas for maize from 2008 to 2018. All the regression models have p-value < 0.001.

Year	Maize											
	prod_NASS vs plt_CDL				prod_NASS vs plt_NASS				prod_NASS vs harv_NASS			
	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)
2008	10.34	0.931	-14.11	344.52	9.83	0.950	-17.56	344.27	10.36	0.969	-12.77	348.85
2009	11.05	0.945	-12.45	362.17	10.43	0.963	-14.21	362.77	10.92	0.978	-10.10	366.68
2010	9.81	0.949	-14.31	329.46	9.71	0.961	-14.52	330.46	10.14	0.970	-10.42	333.61
2011	9.34	0.930	-16.05	339.13	9.37	0.942	-17.26	339.57	9.87	0.963	-12.17	344.40
2012	7.73	0.864	-12.24	290.68	7.63	0.867	-12.98	290.50	8.24	0.897	-9.82	294.77
2013	9.80	0.925	-7.39	359.88	9.73	0.940	-9.41	360.14	10.28	0.959	-5.56	364.24
2014	10.81	0.940	-7.16	373.26	10.57	0.952	-9.99	372.73	11.16	0.970	-6.67	376.55
2015	10.66	0.937	-14.20	379.85	10.64	0.952	-14.89	381.03	11.21	0.967	-10.32	385.29
2016	11.18	0.939	-18.98	406.71	11.22	0.953	-19.16	408.20	11.78	0.970	-14.43	412.94
2017	11.08	0.929	-15.84	390.59	11.23	0.946	-15.63	392.59	11.82	0.964	-11.15	397.27
2018	11.16	0.920	-16.33	397.86	11.31	0.943	-16.08	400.57	11.90	0.963	-11.52	405.55

Table 4.8 The regression statistics between county-level NASS production and cropping areas for soybean from 2008 to 2018. All the regression models have p-value < 0.001.

Year	Soybean											
	prod_NASS vs plt_CDL				prod_NASS vs plt_NASS				prod_NASS vs harv_NASS			
	slope	R^2	bias (ton)	RMSE (10^3 ton)	slope	R^2	bias (ton)	RMSE (10^3 ton)	slope	R^2	bias (ton)	RMSE (10^3 ton)
2008	2.76	0.894	828.19	82.71	2.68	0.926	-570.07	82.39	2.71	0.928	-491.39	82.50
2009	2.92	0.883	2260.98	89.25	2.89	0.921	807.65	89.09	2.93	0.925	853.93	89.20
2010	2.90	0.895	22.98	89.77	2.94	0.924	-794.78	89.93	2.97	0.926	-754.80	90.00
2011	2.68	0.843	1872.97	84.23	2.80	0.891	-59.08	83.99	2.83	0.898	37.05	84.20
2012	2.72	0.907	216.27	82.62	2.69	0.909	-422.19	82.24	2.72	0.914	-311.26	82.43
2013	2.88	0.894	1730.69	90.79	2.94	0.921	567.62	90.62	2.95	0.922	644.31	90.70
2014	3.10	0.901	1824.62	105.62	3.19	0.931	255.89	105.38	3.21	0.931	347.42	105.46
2015	3.16	0.896	-258.61	108.45	3.25	0.920	-323.39	109.11	3.28	0.923	-155.46	109.30
2016	3.40	0.919	110.10	114.83	3.51	0.948	-457.77	115.32	3.54	0.949	-373.80	115.41
2017	3.13	0.885	1413.64	116.07	3.27	0.910	1161.90	116.66	3.28	0.911	1235.02	116.73
2018	3.28	0.877	63.52	124.04	3.48	0.904	29.68	124.89	3.51	0.906	260.35	125.12

Table 4.9 The regression statistics between state-level NASS production and cropping areas for maize from 2008 to 2018. All the regression models have p-value < 0.001.

Year	Maize											
	prod_NASS vs plt_CDL				prod_NASS vs plt_NASS				prod_NASS vs harv_NASS			
	slope	R^2	bias(10^3 ton)	RMSE(10^6 ton)	slope	R^2	bias (10^3 ton)	RMSE(10^6 ton)	slope	R^2	bias (10^3 ton)	RMSE (10^6 ton)
2008	10.63	0.973	-945.24	18.27	9.85	0.979	-876.9	18.3	10.36	0.989	-570.2	18.5
2009	11.03	0.952	-1049.75	19.24	10.43	0.987	-776.6	19.5	10.89	0.995	-445.6	19.7
2010	9.60	0.953	-896.99	17.97	9.61	0.988	-651.8	18.2	10.01	0.992	-349.0	18.4
2011	9.22	0.944	-1043.77	18.43	9.39	0.976	-879.3	18.7	9.86	0.987	-532.3	18.9
2012	7.36	0.913	-583.77	14.80	7.37	0.950	-399.3	15.0	7.91	0.958	-154.2	15.2
2013	9.64	0.936	-742.34	18.72	9.75	0.979	-591.1	19.0	10.27	0.988	-296.3	19.2
2014	10.68	0.939	-781.61	20.14	10.63	0.976	-681.8	20.4	11.18	0.988	-367.9	20.6
2015	10.47	0.943	-1017.06	19.86	10.64	0.983	-800.2	20.2	11.16	0.989	-472.0	20.4
2016	10.94	0.940	-1222.26	21.49	11.20	0.979	-986.7	21.8	11.71	0.987	-643.1	22.0
2017	10.78	0.934	-1117.69	20.70	11.15	0.974	-856.7	21.0	11.73	0.983	-524.5	21.3
2018	10.96	0.929	-1284.46	20.63	11.25	0.971	-942.2	21.0	11.82	0.983	-602.1	21.3

Table 4.10 The regression statistics between state-level NASS production and cropping areas for soybean from 2008 to 2018. All the regression models have p-value < 0.001.

Year	Soybean											
	prod_NASS vs plt_CDL				prod_NASS vs plt_NASS				prod_NASS vs harv_NASS			
	slope	R^2	bias(10^3 ton)	RMSE(10^6 ton)	slope	R^2	bias(10^3 ton)	RMSE(10^6 ton)	slope	R^2	bias(10^3 ton)	RMSE(10^6 ton)
2008	3.00	0.954	-206.69	4.35	2.80	0.971	-166.90	4.40	2.83	0.972	-160.00	4.40
2009	3.18	0.955	-164.03	4.71	3.04	0.973	-123.97	4.76	3.07	0.975	-112.30	4.77
2010	3.15	0.955	-264.67	4.90	3.10	0.968	-205.81	4.96	3.12	0.969	-201.61	4.96
2011	2.92	0.937	-163.85	4.57	2.94	0.952	-159.03	4.59	2.98	0.956	-147.83	4.60
2012	2.81	0.959	-90.00	4.20	2.71	0.959	-60.07	4.22	2.74	0.963	-54.40	4.22
2013	3.03	0.953	-82.90	4.64	3.01	0.968	-68.06	4.67	3.02	0.969	-63.49	4.67
2014	3.27	0.949	-108.59	5.37	3.26	0.965	-95.65	5.40	3.28	0.966	-92.46	5.40
2015	3.36	0.950	-270.45	5.51	3.37	0.963	-188.06	5.59	3.39	0.964	-176.84	5.60
2016	3.61	0.963	-255.44	6.02	3.63	0.978	-184.06	6.09	3.66	0.978	-177.66	6.09
2017	3.32	0.942	-214.05	6.00	3.37	0.958	-115.37	6.10	3.39	0.958	-111.88	6.10
2018	3.38	0.941	-271.06	6.18	3.49	0.957	-175.53	6.27	3.54	0.955	-163.77	6.28

4.3.2 *The relationship between GPP and NASS grain production during 2008-2018*

Fig 4.6 shows the spatial distributions of GPP_{VPM} at 500-m, county, and state scales in the CONUS in 2010, which were highly consistent with the spatial distributions of NASS grain production, planted area, and harvested areas (Fig. 4.6). For maize croplands, the pixels, counties, and states with the high GPP_{VPM} occurred in the Midwest region. For soybean croplands, the pixels, counties and states with the high GPP_{VPM} occurred in the Midwest region and along the Mississippi delta area.

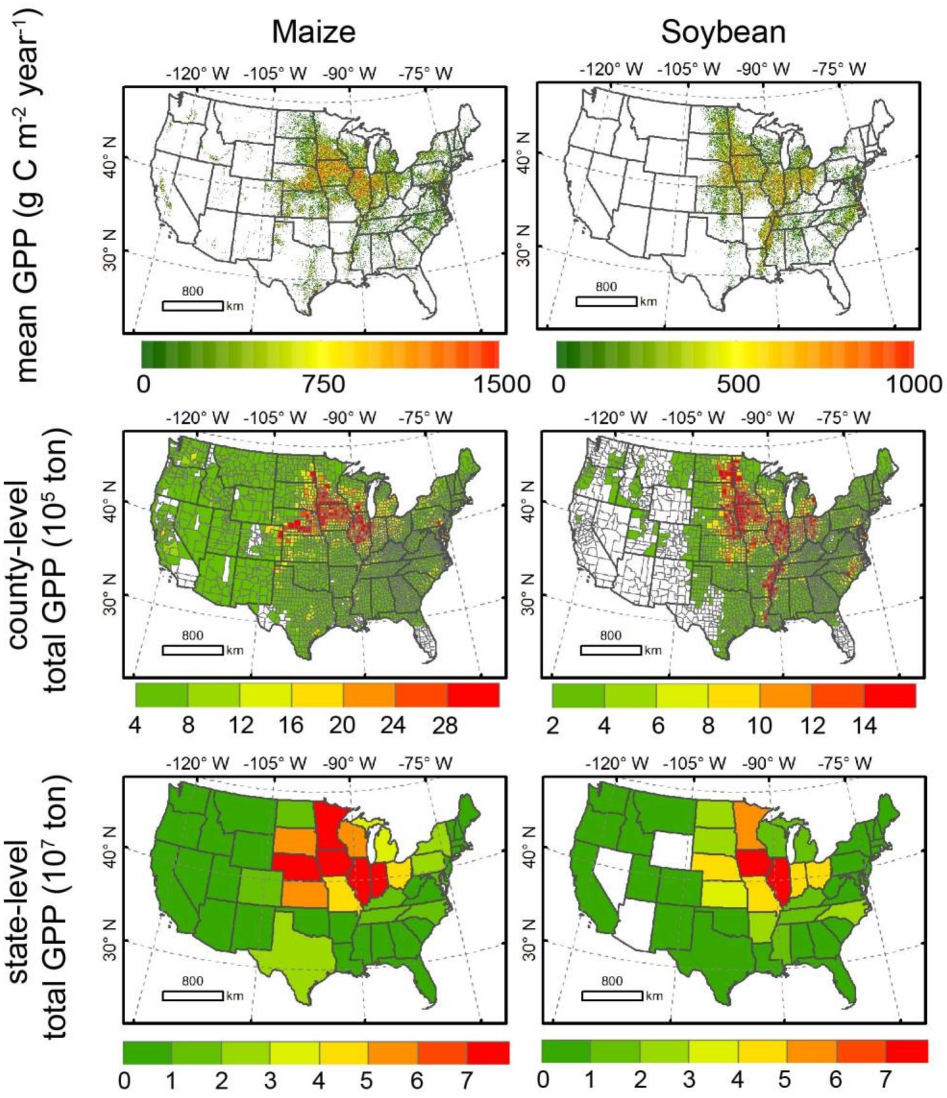

Figure 4.23 Spatial distribution of GPP simulated by VPM, and total GPP calculated by mean GPP and CDL cropping map at 500-m spatial resolution and county-level scale in 2010.

Fig. 4.7 shows the relationship between GPP_{VPM_Year} (and GPP_{MOD17_Year}) and NASS grain production of maize and soybean in the CONUS during 2008-2018 at the county and state scales. The slope values in the regression models represent the average harvest index (HI_{GPP}) of maize and soybean in the CONUS during 2008-2018 at the county and state scales. For maize, GPP_{VPM_Year} during 2008-2018 explained the 93% variation of

NASS grain production at the county scale, with an average HI_{VPM_GPP} of 0.31 (Fig. 4.7). Because of severe drought in 2012, HI_{VPM_GPP} in 2012 (0.25) was substantially (19%) lower than the average HI_{VPM_GPP} (0.31), but R^2 value was still relatively high ($R^2 = 0.89$, p-value <0.001) (Table 4.7). For soybean, GPP_{VPM_Year} during 2008-2018 explained the 91% variation of NASS grain production at the county scale with an average HI_{VPM_GPP} of 0.13 (Fig. 4.7). The HI_{VPM_GPP} in 2012 (0.13) was similar to 2011 but slightly lower than other years (0.13 - 0.15) (Table 4.7). In comparison, GPP_{MOD17_Year} also had strong relationships with NASS grain production at the county and state scales (Fig. 4.7). For soybean crop (C_3 plant), HI_{MOD17_GPP} (0.17-0.18) values are moderately larger than HI_{VPM_GPP} (0.14-0.15). However, for maize crop (C_4 plant), HI_{MOD17_GPP} values (0.60, 0.60) are substantially larger than HI_{VPM_GPP} (~0.31, 0.31) at the county and state scales (Fig. 4.7). This issue with MOD17 dataset can be attributed to the fact that the land cover dataset used in the MOD17 data product does not identify croplands in CONUS as C_4 and C_3 plant function types, separately.

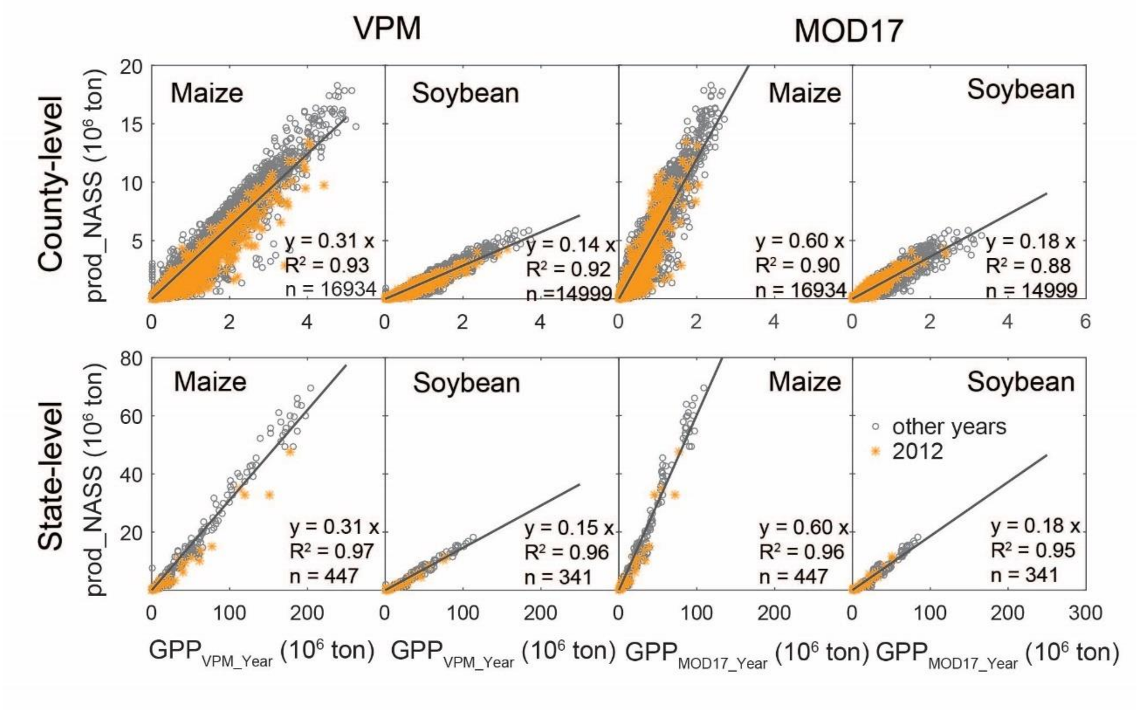

Figure 4.24 Relationship between county-level crop production from NASS statistics and annual total GPP estimates calculated from simulated mean GPP by VPM and MOD17 multiplying by CDL pixel-counting acreage for maize and soybean from 2008-2018, the black solid line is the regression line for all the county-level production and total GPP over 2008-2018. All statistics with $p < 0.001$.

Table 4.11 The regression statistics between county-level NASS production and total GPP from VPM and MOD17 for maize and soybean from 2008 to 2018. All regression models have p-value < 0.001.

Year	VPM							
	Maize				Soybean			
	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)
2008	0.33	0.926	-14.6	350.79	0.14	0.898	-1	82.22
2009	0.36	0.955	-13.42	362.46	0.15	0.92	-1	87.74
2010	0.29	0.943	-14.09	329.04	0.14	0.93	-3.05	88.48
2011	0.28	0.95	-15.84	341.11	0.13	0.919	-2.19	83.09
2012	0.25	0.891	-16.03	290.61	0.13	0.921	-2.23	81.25
2013	0.31	0.954	-12.26	359.62	0.14	0.934	-1.84	89.23
2014	0.32	0.965	-9.79	374.19	0.15	0.949	-2.94	103.68
2015	0.31	0.95	-15.65	380.38	0.15	0.914	-4.07	106.25
2016	0.32	0.945	-19.88	406.84	0.15	0.928	-3.72	112.72
2017	0.33	0.933	-19.56	388.7	0.15	0.914	-3.79	113.11
2018	0.33	0.948	-20.02	398.62	0.15	0.922	-4.34	122.18

Year	MOD17							
	Maize				Soybean			
	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)	slope	R^2	bias (10^3 ton)	RMSE (10^3 ton)
2008	0.59	0.907	-18.98	339.18	0.15	0.878	-2.45	74.54
2009	0.62	0.921	-16.83	357.01	0.16	0.877	-1.22	80.78
2010	0.55	0.919	-14.4	326.75	0.16	0.914	-2.56	82.46
2011	0.57	0.918	-14.46	338.97	0.16	0.862	-0.51	77.51
2012	0.53	0.82	-16.08	284.75	0.17	0.879	-1.95	75.1
2013	0.6	0.916	-10.58	357.09	0.17	0.904	-1.18	83.05
2014	0.65	0.932	-10.92	370.25	0.18	0.903	-1.57	96.47
2015	0.62	0.911	-16.75	375.58	0.18	0.889	-3.91	98.46
2016	0.6	0.925	-21.2	403.72	0.18	0.912	-3.4	104.84
2017	0.62	0.909	-21.68	384.69	0.17	0.879	-2.73	105.24
2018	0.66	0.91	-22.43	392.83	0.18	0.882	-4.08	112.8

4.3.3 *In-season relationships between cumulative GPP and NASS grain production over time in a year during 2008-2018*

In the CONUS, both maize and soybean are cultivated as single crop in a year at individual crop fields. Maize crops are usually planted in April through June and harvested in October and November. Soybean crops are usually planted in late April through June and harvested in September through November. For simplicity, we calculated cumulative GPP_{VPM} values (GPP_{VPM_acc}) of maize and soybean in a county from January 1st at 8-day interval, and then we established the simple linear regression models that used NASS grain production (Y, dependent variable) and GPP_{VPM_acc} (X, independent variable) over time (8-day interval) within a year (NASS grain production = a * GPP_{VPM_acc} + b). We calculated average R^2 value of all counties for each time step and reported the R^2 values over each time step in a year (Fig. 4.8). According to the R^2 curve, the model prediction skill increases over time and reaches 90% by the end of July (Fig. 8), which is approximately one to two months before the start of harvesting time for soybean and maize crops. The model prediction skill showed slight differences among individual years. For maize, the prediction skill was slightly lower in 2008, 2009, and 2012 than in other years, which could be explained by the warm spring and summer drought in 2012 and the underestimation of planted areas in 2008 and 2009 from the CDL dataset. Similarly, for soybean, the prediction skill was slightly lower in 2008 and 2009 than in other years, but it was relatively stable in the drought 2012. In comparison, GPP_{MOD17_acc} also showed very good prediction skills in most years for both maize and soybean, except for maize in 2012 (Fig. 4.8).

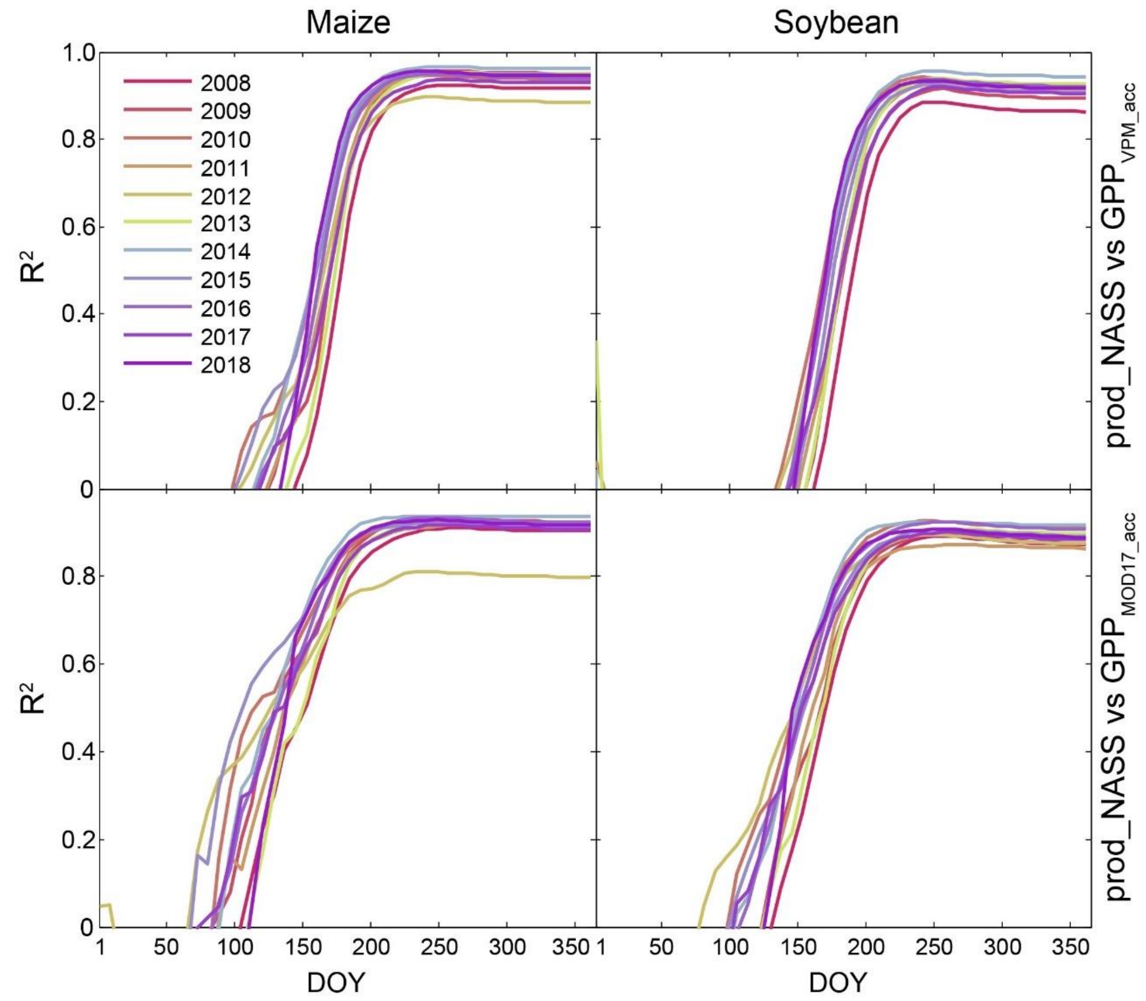

Figure 4.25 Linear regression models for county-level crop grain production from NASS statistics and accumulative total GPP estimates calculated from simulated mean GPP by VPM and MOD17 multiplying by CDL pixel-counting acreage over time (8-day interval) for maize and soybean from 2008-2018

We further selected five top maize and soybean production states (Iowa, Illinois, Nebraska, Minnesota, and Indiana) to explore the in-season relationship between GPP$_{VPM\_acc}$ (and GPP$_{MOD17\_acc}$) and grain production from the NASS dataset (Fig. 4.9). The GPP$_{VPM\_acc}$ and GPP$_{MOD17}$_acc are calculated from the planting dates providing by USDA/NASS (2010) at the state level. According to the R^2 curves (Fig. 4.9), the model prediction skill for individual states was slightly higher and less variable, in comparison of national data analysis (Fig. 4.8). Severe drought in 2012 has moderate effect on model prediction skill for maize grain production in Illinois, Nebraska, Minnesota and Indiana states, and for soybean grain production in Illinois and Nebraska (Fig. 4.9). Both $\text{GPP}_{\text{VPM\_acc}}$ and $\text{GPP}_{\text{MOD17\_acc}}$ have strong in-season relationships with NASS grain production in these five states (Fig. 4.9), which suggests that cumulative GPP is a useful variable for explanation and prediction of maize and soybean grain production at the county scale within maize and soybean crop growing seasons.

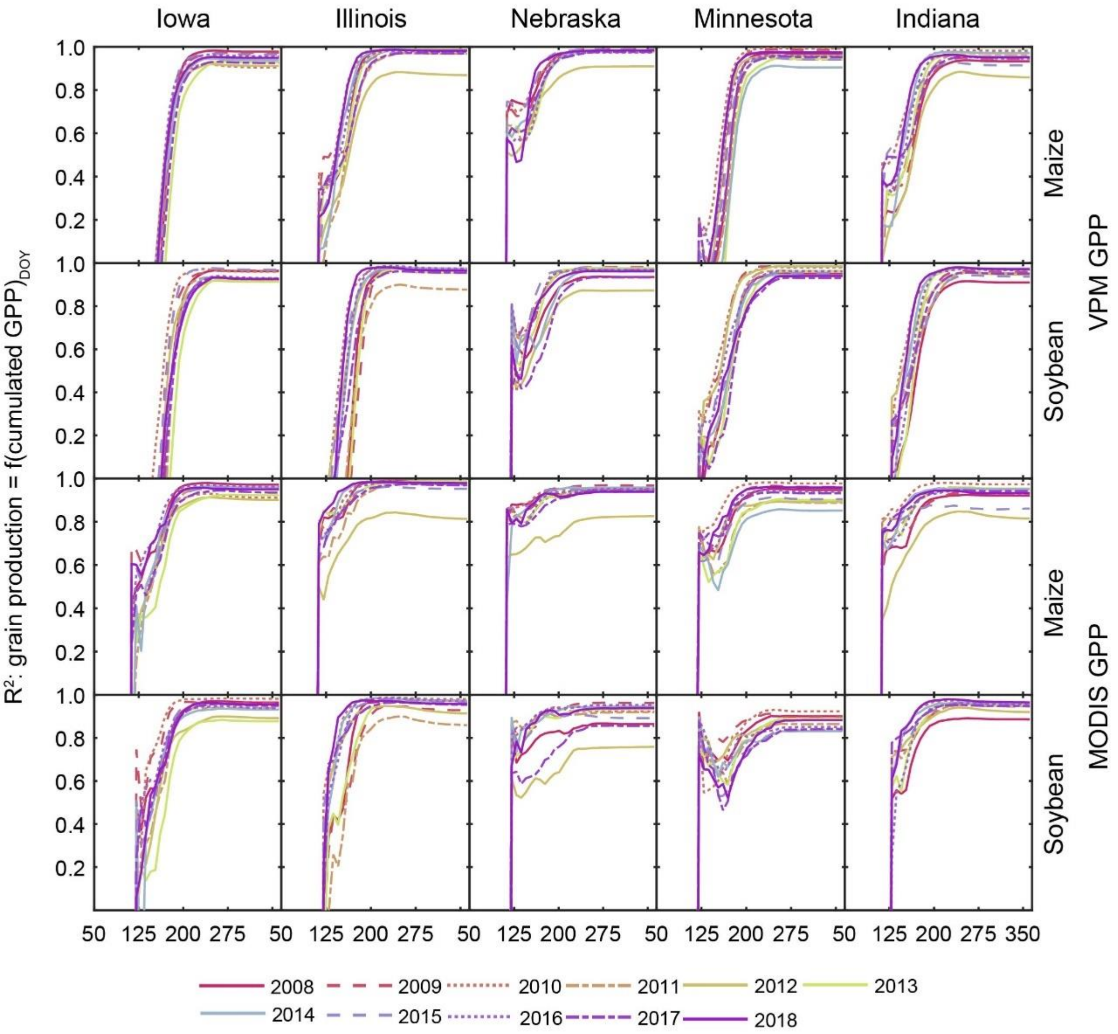

Figure 4.26 Linear regression models for county-level crop grain production from NASS statistics and accumulative total GPP estimates calculated from simulated

mean GPP by VPM and MOD17 multiplying by CDL pixel-counting acreage over time (8-day interval) for five top productive states from 2008-2018

4.4 Discussion

4.4.1 Maize and soybean planted and harvested areas from the CDL and NASS datasets

Satellite remote sensing has been widely used to identify and map cropland planted area in the CONUS (Cai et al. 2018; Massey et al. 2017; Wang et al. 2019a; Zhong et al. 2014). The CDL crop planted area datasets have an overall accuracy of 85% to 95% for major crop types in the CONUS, and up to 97% producer and user accuracy for maize and soybean. Such high classification accuracy was achieved by the machine learning image classification algorithm (Mueller et al. 2009) and large amounts of ground reference data used to train the algorithms. The training and validation ground reference data were sampled from USDA Farm Service Agency (FSA) Common Land Unit (CLU) database and its associated attributes reported by famers. Note that several global GPP data products, e.g., MOD17A2 (Running et al. 2004) have not incorporated the CDL dataset, thus they underestimate GPP of maize and other C_4 crop fields. Our previous study in the CONUS (Wu et al. 2018) and this study clearly show that the use the CDL cropland area dataset is essential for simulations of VPM and other data-driven models.

The spatial-temporal consistency of crop planted areas between the remote sensing approach (e.g., CDL) and the agricultural statistical approach (e.g., NASS) at administrative levels (e.g., county, state or province, nation) has been an important research topic among both agricultural and remote sensing communities (Cai et al. 2018; Wang et al. 2019a). Previous studies reported good agreement between the CDL and

NASS planted area data in 2009 (Boryan et al. 2011; USDA 2014). Our study also shows that the CDL crop planted area estimates had good spatial-temporal consistency with the NASS planted area estimates at county, state and national scales during 2008-2018. The NASS agricultural statistics uses stratification methods to classify land into different agricultural intensity groups or strata based on percent cultivation in a given land parcel, which provides the area sampling frames (Boryan and Yang 2017). In 2010, an automatic stratification method based on the CDL dataset was developed and used in several states (Boryan et al. 2014), which significantly improved stratification accuracies in intensively cropped areas and performed less well in non-agricultural areas as compared with the land cover map method. Recently, an integrated automated stratification and traditional manual hybrid stratification process was implemented in NASS area frame operations (Boryan and Yang 2017).

Our study demonstrates the potential of the CDL and NASS statistic datasets in understanding the changes of planted area, harvested area, and grain production of maize and soybean in the CONUS during 2008-2018. Over these years, maize and soybean planted areas in the CONUS were not affected by summer drought but did increase in response to international demand and grain price in late 2010s. However, maize harvested area and grain production in the CONUS was substantially reduced in 2012 with severe summer drought, particularly in the Midwest states. As the climate models predict larger climate variation and more frequent and severe drought in future years (Dai 2012; IPCC 2013; Trenberth et al. 2013), improvingthe resilience of maize and soybean crops to climate variation and change is a major challenge for the farmers and the society.

4.4.2 *Harvest Index − The relationships between GPP, AGB and NASS grain production of maize and soybean*

Gross primary production (GPP), net primary production (NPP) and aboveground biomass (AGB) are related to grain yield (ton ha^{-1}) and production (ton). The "Harvest Index" (HI) is widely used term (Hay 1995) and often defined across various scales from plants to fields, and county as the ratio between crop grain yield (ton ha^{-1}) and aboveground biomass (AGB), namely HI_{AGB}. Grain yield of individual plants is affected by two processes: (1) flowering and pollination, which affects grain number, and (2) grain-filling, which determines individual grain sizes. Many studies have shown that these two processes are highly sensitive to heat and drought stresses (Guan et al. 2016; Liu et al. 2008; Lobell et al. 2014). Many studies reported that HI_{AGB} values often vary substantially among individual crop types, for example, 0.25 – 0.58 for maize (Guan et al. 2016), and 0.30 – 0.44 for soybean (Johnson and Major 1979; Krisnawati and Adie 2015; Lobell et al. 2002; Monfreda et al. 2008), which could be attributed to large degree how and when maize and soybean plants were harvested and AGB was measured.

Harvest Index can also be defined as the ratio between NASS crop grain production and gross primary production, namely HI_{GPP}. In a study on croplands in Montana (He et al. 2018a), GPP data from the data-driven model during 2008-2015 and calibrated HI_{GPP_VPM} (0.44) were used to estimate maize grain production; The resultant GPP-derived maize production had a strong linear relationship with NASS maize production at the county scale ($R^2 = 0.82$). Our study shows that GPP_{VPM_Year} data during 2008-2018 were strongly correlated with NASS grain production (GP) data for maize ($GP = 0.31 * GPP_{VPM\_Year}$, $R^2 = 0.93$) and soybean ($GP= 0.14 * GPP_{VPM\_Year}$, $R^2 = 0.91$) at the county scale over the CONUS (Fig. 7). HI_{GPP} of maize, which is the slope of the simple linear regression model between GP and GPP_{VPM_Year} of maize at the county scale in individual years, varied from 0.25 in the severe drought year (2012) to 0.36 in the wet year (2009) (Table 5). HI_{GPP} of soybean varied from 0.12 in the drought year (2012) to 0.14 in the wet year (2009) (Table 5). The interannual variations of HI_{GPP} in this study at the CONUS scale could come from multiple sources, many studies have discussed the effects of environment, management and crop genetics (variety) (Erickson et al. 2017; Licht et al. 2019; Lobell and Azzari 2017). In this study, the environmental factors such as severe drought in 2012 have a strong effect on GPP and maize grain production. The differences of planted area between CDL and NASS datasets were larger in 2008 and 2009 than in other years (Table 4.1 and Table 4.2), which could lead to moderate variations of annual HI_{GPP} in those two years. It is well known that crop genetics (e.g., crop variety) affects crop grain yield and production, as some crop types and genotypes are more tolerant to drought and pathogens, and more sensitive to changing crop management, like narrow row spacing and more intensive managment . However, as HI_{GPP} values of maize and soybean exhibit relatively moderate interannual variations during “normal” years, it clearly indicates the potential of using HI_{GPP} and GPP_{VPM} data to estimate maize and soybean grain production over those “normal” years at the county scale. Additional efforts are needed to elucidate the relationships between GPP and NASS grain production at those individual farms used in the NASS crop surveys, which could further reduce the spatial-temporal variations of harvest index (HI_{GPP}) for maize and soybean crops.

4.4.3 *Explanation and prediction of maize and soybean grain production by GPP at the county scale*

Numerous studies have used vegetation indices to predict crop grain yields (Bolton and Friedl 2013; Zhao et al. 2015). A number of LUE models estimate daily GPP of croplands (He et al. 2018a; Yuan et al. 2007; Zhang et al. 2017). Several studies have used GPP data to estimate crop grain yields by assuming that yield is a function of GPP, autotrophic respiration, HI_{AGB} and the root to shoot ratio (Guan et al. 2016; Marshall et al. 2018; Yuan et al. 2016b). These studies then compared the resultant yield estimate with the yield data from the flux tower sites (Yuan et al. 2016b), and NASS yield data in the Midwest Corn-Belt (Guan et al. 2016) and the CONUS (Marshall et al. 2018). These studies include maize, soybean, and winter wheat, and reported moderately strong relationships between NASS grain yield data and modeled yield (R^2 ranging from 0.5 to 0.7) at the county and state scales (Guan et al. 2016; Marshall et al. 2018). Note that NASS crop grain yield data at the county scale were derived from the survey and sampling approach, thus, further studies are needed to compare yield data at individual farms or fields used in the NASS crop surveys.

In our study, we focus on the relationship between GPP and NASS grain production of maize and soybean in the CONUS at the county scale. To explore the potential of in-season forecasting, a linear regression model between cumulative GPP_{VPM_acc} over time at 8-day interval and annual NASS grain production at the county scale accounted for more than 80% of variation of NASS grain production of maize and soybean among all the counties in the CONUS by the end of June, and more than 90% by the end of July (Fig. 4.8 and Fig. 4.9). Peng et al. (2018) incorporate satellite derived EVI and climate forecast data in a crop model to forecast U.S. maize yield, they also report EVI improved the forecasting significantly in July and August. Therefore, the satellite-based information can play an important role in early crop yield and production forecast.

The capacity of in-season forecasting can be further improved in several aspects. First, the GPP_{VPM} simulation in this study was carried out at a moderate spatial resolution (500-m), and it could be improved by using high spatial resolution (e.g., 30-m Landsat, and 10-m Sentinel-2). Second, in-season forecasting of crop grain production requires in-season maps of crop type (e.g., maize, soybean), planted area and harvested area at high spatial resolutions (e.g., 30-m or 10-m). In this study, we use the crop type and planted area maps from the CDL dataset with a high spatial resolution of 30-m. However, the CDL dataset took time to generate and was often scheduled to release in the spring of next year (one-year delay), which cannot meet the demand for grain production prediction in the current year (in-season forecast mode). Numerous crop-type classification studies have used single image (Van Niel and McVicar 2004; Yang et al. 2011) or multiple images (Chang et al. 2007; Foerster et al. 2012) in the growing season. Recently, a few studies reported their efforts for high spatial resolution and in-season crop mapping (Cai et al. 2018; Wang et al. 2019a). It remains a major challenge for the remote sensing community to develop in-season maps of crop types, planted areas, and harvested areas in the CONUS.

4.5 Conclusion

Our study evaluated spatial-temporal dynamics of NASS crop statistical data (crop planted area, harvested area, grain production), satellite-based CDL crop planted area, and GPP estimates from the VPM model at the county, state and CONUS scales during

2008-2018. There are strong spatial-temporal consistencies between the crop planted area from the CDL dataset and crop planted area from the NASS crop statistics during 2008–2018 at the county and state scales, which supports the use of the CDL dataset for model applications (He et al. 2018; Marshall et al. 2018). For maize and soybean crops, the HI_{GPP} values, which is calculated as the ratio between NASS grain production and GPP at the county scale, have relatively small variations over years during 2008-2018, except the extreme drought year (2012). Cumulative GPP_{VPM} and GPP_{MOD17} over time at 8-day interval within the maize and soybean growing season, together with HI_{GPP}, were able to explain and predict grain production of maize and soybean at the county scale about 1-2 month ahead of crop harvest. The strong and robust linear relationships between GPP_{VPM_acc} and NASS grain production of maize and soybean in the CONUS at the county scale highlight the potential of GPP_{VPM_acc} data products in monitoring and forecasting maize and soybean grain production in the CONUS.

Chapter 5: Spatial-temporal dynamics of winter wheat croplands in the Contiguous United States: Consistency between agricultural statistical and satellite approaches

Abstract

Winter wheat is a major staple crop grown in the US, and the US is the third largest wheat exporters globally. Timely and reliable winter wheat production prediction in the US is important for regional and global food security. In this study, we evaluated the consistency between the agricultural statistical data and satellite-based data for winter wheat over the CONUS at county and national scales. First, we evaluated the planted area estimates from the Cropland Data Layer (CDL) and National Agricultural Statistics Service (NASS) during 2008-2018. There was strong spatial-temporal consistency between the CDL and NASS datasets for the planted acreage. However, both the CDL and NASS planted acreage showed a significant difference (>40%) from the NASS harvested area in the Southern Great Plains, where winter wheat is used as both a grain crop and forage crop. We then estimated gross primary production (GPP) of winter wheat at 8-day interval during 2008-2018 using the Vegetation Photosynthesis Model (VPM), which was driven by the CDL, climate, and satellite images. The total GPP_{VPM} had a moderate linear relationship with grain production of winter wheat, with R^2 value of 0.68 at county scale over the CONUS. When excluding those counties with a significant difference (over 40%) between planted and harvested area, the relationship is improved with the R^2 value of 0.85. Our findings suggest that the importance of reliable in-season

crop type classification and the potential of GPP_{VPM} in crop monitoring, which could provide useful data to decision makers, stakeholders and the public..

5.1 Introduction

Wheat is the most widely cultivated staple food crop in the world, accounting for approximately one sixth of the total arable land area, and wheat grain contributes about 20% of the total dietary calories worldwide (Mitchell and Mielke 2005). Wheat is also widely traded in the world food market, and approximately ~23% of the world's wheat grain production are traded internationally every year. The United States of America (USA) is the fifth largest wheat producer (after European Union, China, India, Russia) and the third largest wheat exporter (after Russia and European Union) (FAS/USDA 2020). Winter wheat dominates the USA wheat production, accounting for approximately 80% of the total wheat production (USDA 2014). The fluctuation of winter wheat production in the USA could have significant impacts on the global food security. Therefore, timely, reliable, and spatially specific information on winter wheat planted area, harvested area, and grain yield and production are critical for regional and global food security and international food trade.

The agricultural surveys and statistical reports by governmental agencies have been the major data sources for crop planted area, harvested area, grain yield, and grain production in a year at various administration levels (ADB 2015; Carfagna and Carfagna 2010; NASS/USDA 2012; NASS/WAOB/USDA 1999). The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) is responsible for in-season forecast and year-end estimation of grain production of most crops grown in the USA. Each year NASS conducts several surveys and extensive field observations to collect a variety of data needed to fulfill this task (NASS/USDA 2012; NASS/WAOB/USDA 1999). However, the data collection through agricultural surveys is not only time consuming and costly but also have long time lags and data gaps (De Groote and Traore 2005).

Satellite-based remote sensing has been widely used to identify and map crop planted area and harvested area since the early 1970s (Mulla 2013). Numerous studies have used multiple-temporal remote sensing data to characterize and map individual crop types (Massey et al. 2017; Skakun et al. 2017; Yang et al. 2019; Zhong et al. 2019) . The annual Cropland Data Layer (CDL) datasets, which were produced by NASS/USDA (Boryan et al. 2011) reported annual planted areas for all major crop types in the contiguous United States (CONUS) over the period of 2008-2018. The annual maps of pixel-based crop planted area from the CDL provide a supplementary acreage estimates to NASS acreage survey data, and it was reported that CDL-based crop planted area estimates were slightly lower than the NASS statistical estimates (Boryan et al. 2011). To date, no systematic comparative study was carried out to examine the spatial-temporal consistency of winter wheat planted area estimates between CDL and NASS datasets during 2008-2018 in the CONUS. No comparative study was conducted to investigate the differences between winter wheat planted areas (CDL, NASS) and harvested area (NASS) during 2008-2018 in the CONUS, neither.

Numerous efforts have been made to estimate crop yield and/or crop production by satellite remote sensing (Franch et al. 2015; Huang et al. 2016; Zhuo et al. 2019). Researchers use climate and satellite data to predict crop yield and/or production across a wide range of spatial scales and geographic regions. Both statistical models and process-based models have been used to estimate crop yields (Becker-Reshef et al. 2010; Cai et al. 2019; Wang et al. 2019b; Zhuo et al. 2019). Climate variables, such as air temperature, solar radiation and precipitation are common inputs to both statistical and process-based models (Gornott and Wechsung 2016; Zhuo et al. 2019). Though climate data describes the environmental condition that affects crop growth, it cannot directly detect the impact of both abiotic and biotic factors on crop growth (Hatfield et al. 2008; Mahlein et al. 2012). Thus, it requires more input rather than climate data alone for predicting crop yield accurately (Cai et al. 2019). Various remote sensing-based data using various spectral bands have been extensively used for monitoring crop growth and estimating crop yield (Becker-Reshef et al. 2010; Franch et al. 2015; He et al. 2018a). For example, vegetation indices (e.g. Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)) derived from optical bands have been used to estimate winter wheat yield and production by applying an empirical harvest index (HI) specific to different crop types in different countries (Becker-Reshef et al. 2010; Franch et al. 2015). Satellite-based gross primary production (GPP) also have been used to estimate crop yield and/or production (add reference]. GPP can be estimated by using a light use efficiency (LUE) model driven by both remote sensing images and climate data (Running et al. 2004; Wu et al. 2018; Zhang et al. 2017). Crop yield can then be estimated as the product of GPP and GPP-based HI (HIGPP) (He et al. 2018a). He et al. (2018a) modeled HIGPP from county-level MODIS GPP and grain production statistics for croplands in Montana during 2008-2015 and applied the HIGPP for yield prediction. To date, no study has been conducted to evaluate the spatial-temporal relationships between annual GPP and annual winter wheat production in the CONUS during 2008 - 2018. There is also no study to assess the potential of using GPP data for in-season forecasting of winter wheat production in the CONUS. In this study, we investigated the spatial-temporal dynamics of winter wheat planted area, harvested area, gross primary production, and grain production in CONUS at county and national scales during 2008-2018, which could shed new insights on vulnerability and resilience of winter wheat production in CONUS to climate and other driving factors. First, we analyzed the agricultural statistical data on winter wheat planted area, harvested area and grain production from the USDA NASS and satellite-based data on winter wheat planted area from the USDA Cropland Data Layer (CDL) dataset during 2008-2018. The analysis will quantify the spatial-temporal consistency of winter wheat planted area between NASS and CDL datasets, the spatial-temporal differences between winter wheat planted area and harvested area, and the relationships between winter wheat grain production and planted area or harvested area. Second, we analyzed the gross primary production (GPP_{VPM}) data from the satellite-based Vegetation Photosynthesis Model (VPM), which were already evaluated with GPP data from the in-situ cropland eddy flux tower sites (GPP_{EC}) (Wu et al. 2018), and grain production data from the NASS at county scale. The analysis will quantify the spatial-temporal relationships between winter wheat GPP_{VPM} and grain production (GP) from the agricultural survey statistics at county- and national-level under varying climate conditions (drought year versus normal years). The resultant relationship between GPP_{VPM} and grain production, which is called as harvest index (HI) and defined as HI_{GPP} (HI_{GPP} = GP / GPP), could be used for comparison with other HI value defined by net primary production (HI_{NPP}; HI_{NPP} = GP / NPP) and aboveground biomass (HI_{AGB}; HI_{AGB} = GP / AGB). Third, we calculated the relationships between cumulative GPP_{VPM} over time and annual grain production at county scale and explored the potential to monitor winter wheat grain production in the CONUS within the winter wheat growing season.

5.2 Materials and methods

5.2.1 *Study area*

The study area is the Contiguous United States (CONUS), which includes 48 states and 3,233 counties. Climate in the CONUS ranges from subtropical climate in the southern region (e.g., Florida) to temperate climate in the northern region. Winter wheat is primarily grown in the Great Plains, with Southern Great Plains (Kansas, Oklahoma, and Texas) being the largest producing zone, contributing more than 40% of the total wheat production in the US (Steiner et al. 2015). Winter wheat cultivation is largely rain-fed. Note that some winter wheat fields in the Southern Great Plains are treated as dual-purpose use fields, in other words, these fields could be used as a cool-season beef cattle grazing fields and/or grain production, dependent upon weather, market conditions and other factors (Edwards et al. 2011).

5.2.2 *Winter wheat planted area, harvested area, and grain production data during 2008-2018 from the USDA-NASS statistical dataset*

We used the official archive of annual county- and national- level statistics on crop planted area, harvested area, grain yield and production, which is available from the USDA National Agricultural Statistic Service (NASS) Quick Stats database (https://quickstats.nass.usda.gov/). The NASS crop statistics are based on data obtained from multiple frame-based sample surveys of farm operators, objective yield surveys, processors, and commercial storage firms, and so on. For example, the NASS winter wheat planted area and harvested area estimates come mostly from the Agricultural Survey in December and June. Farmers are asked by investigators about the acreage they planted and the acreage they intend to harvest (USDA 2014). The grain yield statistics were established based on both the Agricultural Yield Survey (AYS) and Objective Yield Survey (OYS). AYS is based on farmers' reported yield and OYS is based on model simulation driven by field biophysical crop observations. The grain production is estimated from the expected harvested area and yield at the survey reference date and predicted assuming normal conditions for the remainder of the season.

5.2.3 *Winter wheat planted area data from the USDA-NASS Cropland Data Layer dataset*

The annual CDL dataset is a remote sensing-based land cover product at 30-m spatial resolution, which includes more than 100 crop types in the CONUS since 2008 (Boryan et al. 2011). The CDL product is based on the machine learning method which utilizes both in-situ ground reference data and multi-temporal satellite imagery to locate and identify field crops. The major sources of training data used as in the supervised classification, includes the USDA's Farm Service Agency (FSA) Common Land Unit (CLU) data and the National Land Cover Dataset (NLCD). The CLU data were collected based on the reported crop types and crop acreage from producers to the FSA county offices in the growing season. The CDL classifier utilized multiple remote sensing images, including AWiFS, Landsat TM and ETM+, Deimos-1 and UK-DMC-2 and MODIS satellite data (Boryan et al. 2011). The CDL data has a spatial resolution of 30 meters in 2010-2018 and of 56 meters in 2008-2009. In 2018, the CDL data from 2008-2009 were reproduced to 30 meters to keep a consistent resolution. In this study, we used the 30-m CDL data throughout 2008-2018 to keep our analysis in a consistent spatial resolution.

The CDL data provide very detailed information about multiple crop distribution, with classification accuracy higher than 90% for major crops (maize, soybean and winter wheat) (Boryan et al. 2011). It has been widely applied in scientific research related to land use and land cover change, agricultural sustainability, and agricultural production decision-making. An interactive Web-based CDL data portal, CropScape, was developed to visualize and analyze the CDL data in a more effective and efficient way (Han et al. 2012). In this study, the acreage statistics was calculated from CropScape for all the crop categories at each county during 2008-2018. The annual total area for winter wheat at each counties calculated as the sum of both the single cropping and double cropping area, respectively. The national total areas for winter wheat in each year was then calculated by adding the annual total area over all the counties in the nation.

5.2.4 *Gross primary production (GPP) data from the Vegetation Photosynthesis Model (VPM)*

The VPM model is a light use efficiency model and estimates daily GPP as a product of the amount of PAR absorbed by chlorophyll in the canopy ($APAR_{chl}$) and the light use efficiency (Xiao et al. 2004a; Xiao et al. 2004b). A number of publications have reported the evaluation of GPP_{VPM} with GPP_{EC} estimates from the eddy flux tower sites, including maize (Dong et al. 2015; Kalfas et al. 2011), soybean (Jin et al. 2015; Wagle et al. 2015), winter wheat (Doughty et al. 2018; Yan et al. 2009) and paddy rice (Xin et al. 2017). All these publications reported strong agreement between GPP_{VPM} and GPP_{EC}, with a range of R^2 values from 0.70 to 0.98. The input data for simulations of the VPM model include meteorological data (shortwave radiation, air temperature), and satellite-based vegetation indices (EVI, LSWI). More detailed description of the VPM model can be found in previous publications (Wu et al. 2018; Xiao et al. 2004a; Xiao et al. 2004b; Zhang et al. 2017). In this study, we followed the same procedure to run VPM simulations as reported in a previous study (Wu et al. 2018). Different maximum light use efficiency (LUE_0) parameter values for C_3 and C_4 crops were used, which improves daily GPP estimation in croplands (Wu et al. 2018). The growing-season total GPP of winter wheat at each pixel were calculated by summarizing daily GPP between the USDA planted and harvested dates (NASS/WAOB/USDA 1999) . The mean GPP (GPP_{mean}) at each county during the growing season were area-weighted based on the area fraction for winter wheat derived from the CDL datasets in each 500-m pixel. The growing season total GPP (GPP_{total}) for winter wheat at each county was calculated as GPP_{mean} multiplied by the total area of all those pixels with winter wheat located in the county.

5.2.5 *Statistic metrics*

To explore the interannual changes of grain production, planted area, harvested area and grain yield of winter wheat from 2008-2018, we calculated the anomaly of each variable as the difference between the value in specific year and multi-year average from 2008-2018, and then normalized by the multi-year average. Simple linear regression models were used to characterize the relationship between grain production and GPP_{total}, and between grain production and cropping areas at county and national scales, at a minimum 0.05 significance threshold (p-value). Model performance was evaluated using the coefficient of determination (R^2), and bias (RMSE) between the modelled grain production estimates and the corresponding grain production statistics. The regression slope between grain production and GPP_{total}, representing the conversion coefficient from GPP_{total} to grain production, is derived as the GPP-based Harvest Index (HI_{GPP}).

Timely prediction of grain yield is one requirement in yield monitoring (Cai et al. 2019). To explore the in-season predictability from GPP_{VPM}, linear regression models were applied between grain production and accumulative GPP (GPP_{accum}) over time at 8-day temporal resolution at county-scale (see Equation 5.1).

$$Grain\ Production = a * \sum_{1}^{t}(GPP_t \times k) + b \qquad t = 1, 2, 3, \ldots, 46 \qquad (5.1)$$

Where t is the number of time steps in a year, which ranges from 1 to 46, as time series GPP has 46 data points in a year; k is the number of days in each time step, k is equal to 8 days when t ranges from 1 to 45, and k is 5 (non-leap year) or 6 (leap year) when t is 46. In this study, we simply used a calendar year schedule to run the statistical models for all counties, though winter wheat is usually planted in autumn. The GPP_{accum} during the cool time is usually low and can be negligible in affecting the modelling skills.

5.3 Results

5.3.1 *Spatial distribution and inter-annual changes of winter wheat planted area, harvested area, and grain production during 2008-2018 from the CDL and NASS statistics datasets*

At the national scale, winter wheat planted area, harvested area and grain production from NASS were the highest in 2008 but decreased substantially in 2017 and 2018 (Fig. 5.1a). Production showed a decreasing trend during 2008-2018, which is largely determined by the decrease of planted and harvested area (Fig. 5.1). Winter wheat yield was relatively stable during 2008-2013 but varied significantly in recent years, with the lowest value in the drought year 2014 and the largest value in 2016. Overall, the planted acreage derived from the CDL data showed a similar changing trend as NASS planted acreage but with a different magnitude. The NASS harvested area showed a different changing magnitude as the NASS planted acreage, as well as the CDL planted acreage. The harvested percentage of winter wheat varied from ~75% (2013, 2014, 2017, and 2018) to ~85% (2008, 2010, and 2012) during the 11 years (Fig. 5.1b).

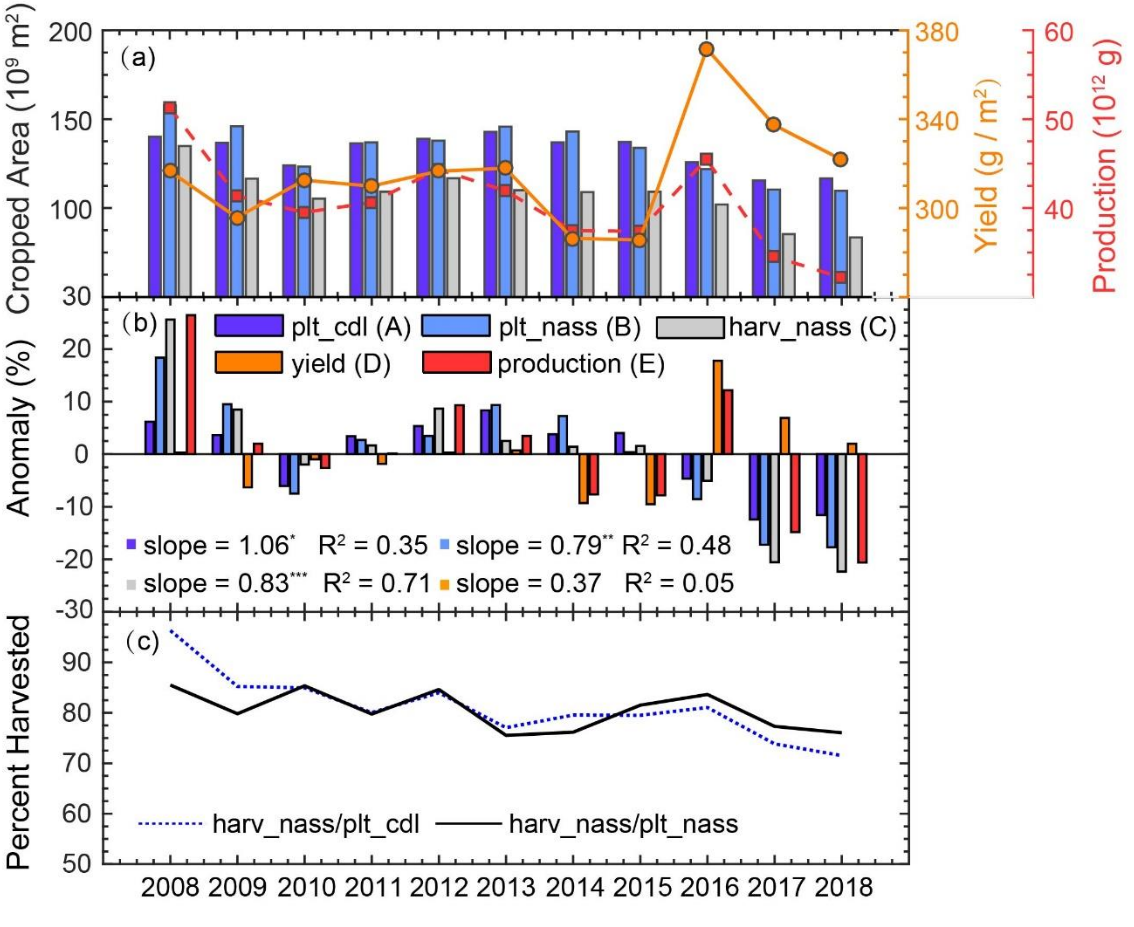

Figure 5.27 Interannual variability of (a) planted area from CDL and NASS statistics, harvested area, yield, production from NASS statistics; (b) harvested percentage of winter wheat derived from CDL and NASS planted acreage during 2008-2018.

At county scale, the CDL dataset in 2010 shows that winter wheat was cultivated in 2,853 counties, and many counties in the Southern Great Plains, known as the Winter Wheat Belt, had a large winter wheat planted area, (Fig. 5.2a). The NASS statistical dataset in 2010 reported winter wheat planted area and harvested area in 2,426 counties (Fig. 5.2b,c). The spatial distribution of CDL-based winter wheat planted area is consistent with that of NASS-based winter wheat planted area. The NASS statistical dataset in 2010 reported winter wheat yield and production in 1,263 counties (Fig. 5.2d,e). The spatial distribution of winter wheat yield in 2010 (Fig. 5.2d) is very different from that of winter wheat planted area, and many counties in the northwest region and California had the highest yield while most counties in the Southern Great Plains and the southeastern CONUS had the lowest yield.

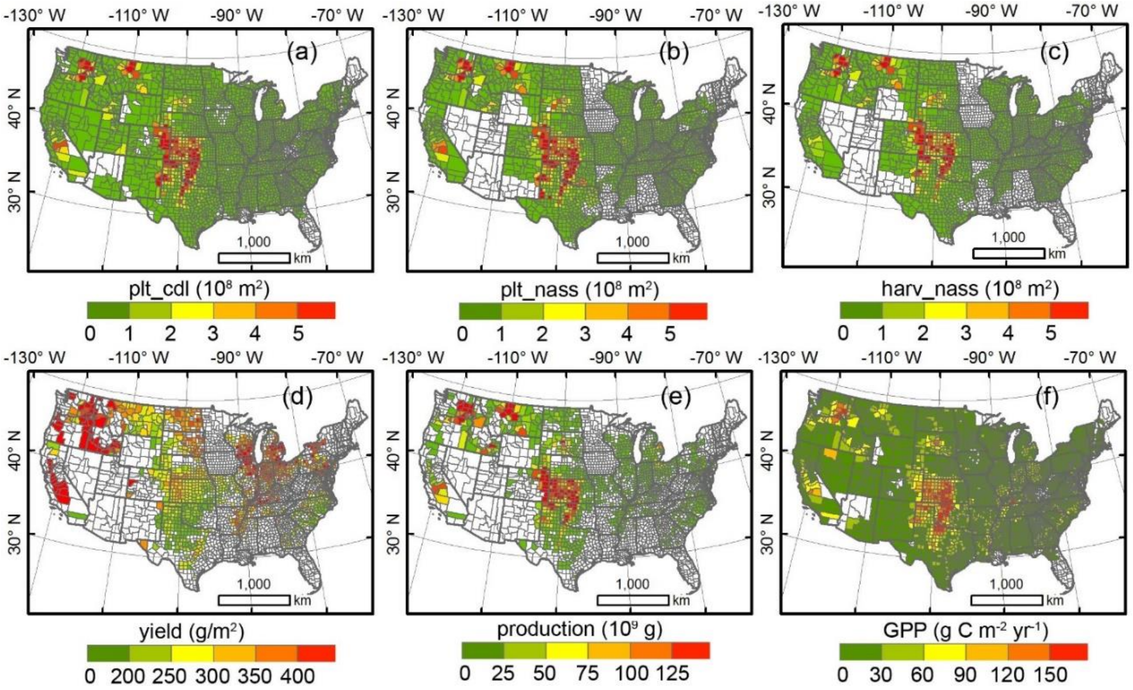

Figure 5.28 The spatial distribution of (a) winter wheat planted area from the CDL, (b) winter wheat planted area from the NASS statistics datasets, (c) harvested area from the NASS statistics, (d) grain yield and (e) production from the NASS statistics, and (f) annual averaged VPM GPP at county scale in the CONUS in 2010.

For all the counties having winter wheat planted area and harvested area from both CDL and NASS datasets during 2008-2018, the CDL-based winter wheat planted area agreed well with NASS-based winter wheat planted area at county scale ($R^2 = 0.98$, $p<0.001$) (Fig. 5.3a), with a relative small interannual variation as shown in the slope values (<10%) (Table 5.1). In many counties, winter wheat harvested areas were much smaller than winter wheat planted areas from the CDL dataset and the NASS dataset (Fig. 3b,c). Most of those counties with large discrepancy between winter wheat planted area and harvested area were distributed in the Southern Great Plains and California, where some of winter wheat fields were used as cool-season forage for beef cattle production (grazing or haying), which is often called as the dual purpose use of winter wheat fields. We further calculated the interannual trend of winter wheat planted areas from both CDL and NASS datasets during 2008-2018 (Fig. 5.4). There are few counties included from the NASS dataset because of the data missing over the continuous 11 years for many counties. Many counties plant less winter wheat than before. But the spatial distribution has some differences, the CDL data showed a decreasing planting of winter wheat in the Northern Great Plains and Kansas, and an increase of planting in Oklahoma and Texas. Compared to CDL acreage data, the NASS data showed a significant decrease in planted area in Oklahoma. This difference could be related to the data accuracy of the two datasets.

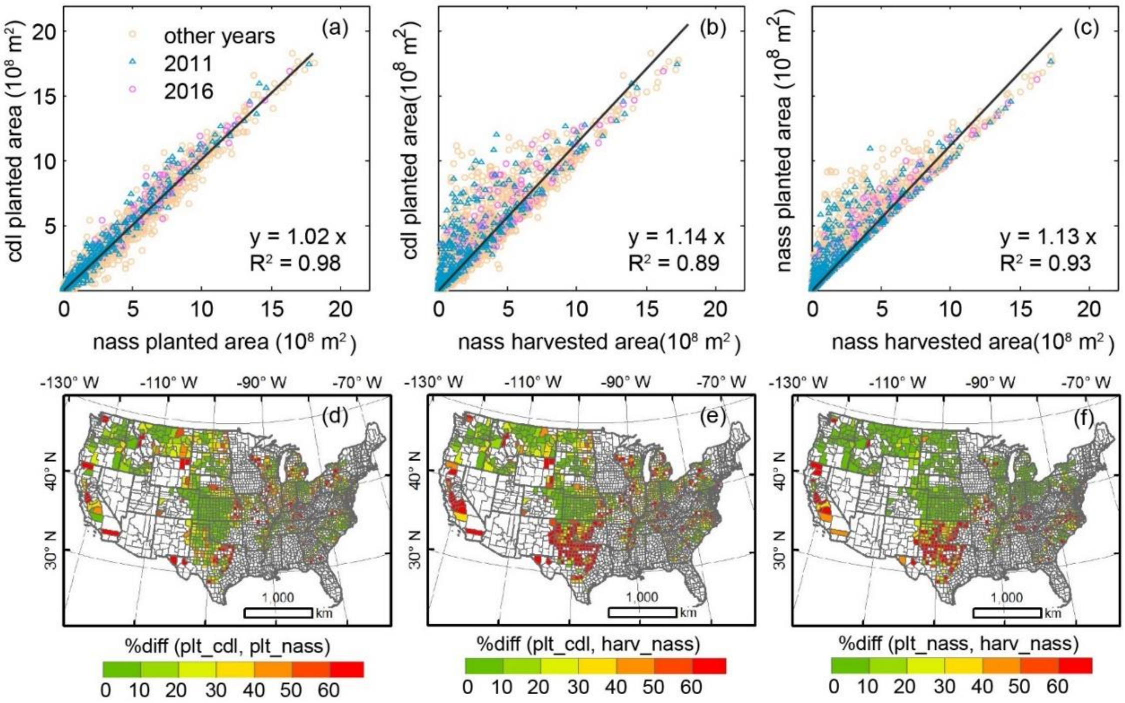

Figure 5.29 The comparisons between winter wheat planted areas from the CDL and NASS datasets and between winter wheat planted area and harvested area during 2008-2018 at county scale, and the spatial discrepancy between CDL and NASS planted area, CDL planted area and NASS harvested area, and NASS planted and harvested area in 2010. The 2011 is a typical drought year over the winter wheat belt, and 2016 is a wet year.

Table 5.12 Relationship between NASS and CDL cropping area for winter wheat from 2008-2018

	plt_cdl vs plt_nass		plt_cdl vs harv_nass		plt_nass vs harv_nass	
	slope	R^2	slope	R^2	slope	R^2
2008	0.960	0.973	1.070	0.941	1.111	0.958
2009	0.987	0.986	1.105	0.873	1.126	0.897
2010	1.016	0.986	1.121	0.953	1.106	0.971

2011	1.025	0.989	1.152	0.862	1.137	0.898
2012	1.025	0.986	1.101	0.909	1.086	0.949
2013	1.010	0.990	1.145	0.867	1.141	0.888
2014	0.973	0.984	1.134	0.841	1.179	0.882
2015	1.055	0.985	1.190	0.926	1.135	0.955
2016	1.060	0.985	1.166	0.917	1.113	0.959
2017	1.071	0.968	1.209	0.870	1.146	0.937
2018	1.091	0.977	1.224	0.839	1.142	0.902

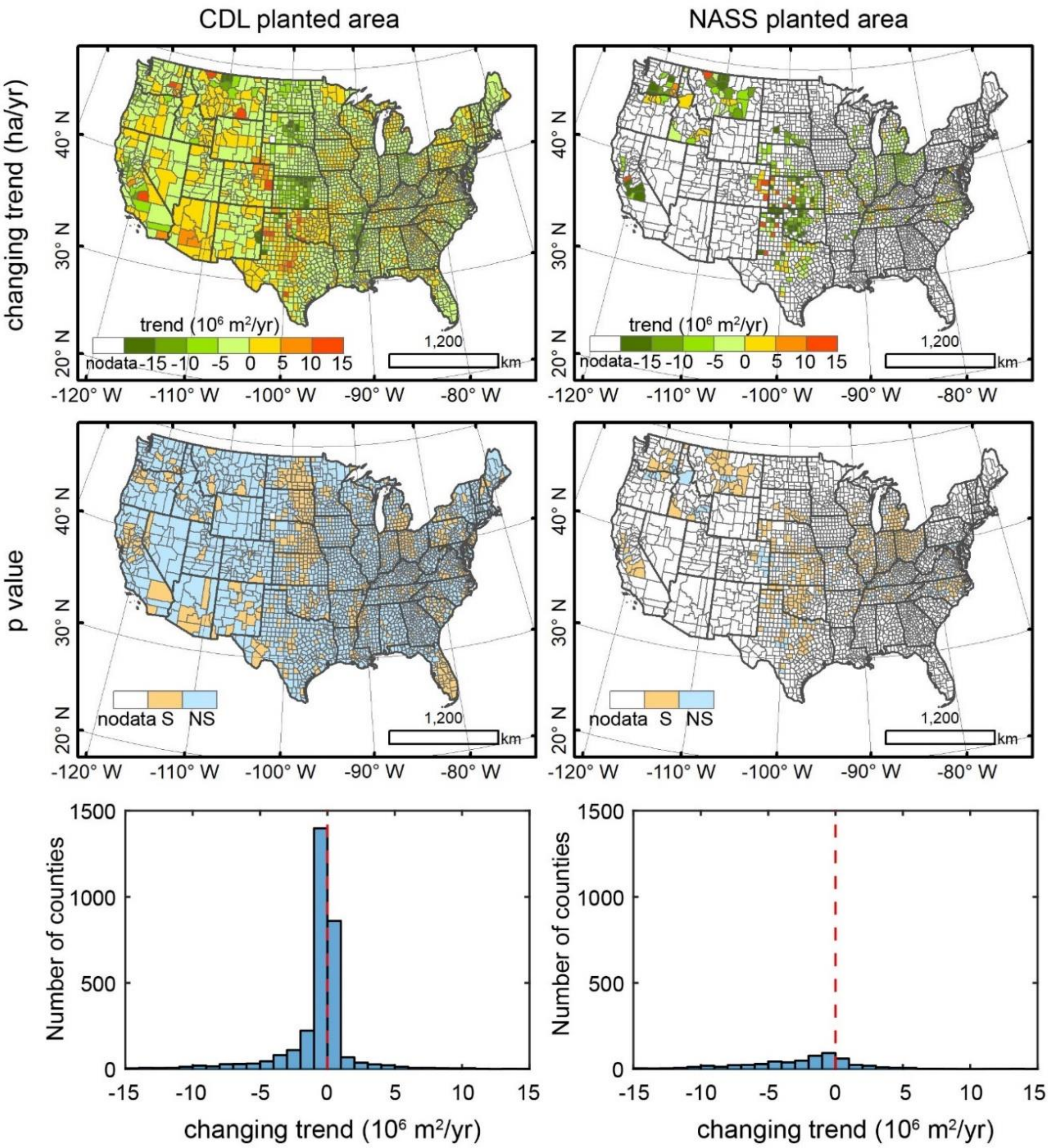

Figure 5.30 Trend of planted area for winter wheat from 2008-2018

For all the counties having winter wheat planted area, harvested area and grain production data during 2008-2018, the NASS-based grain production had a slightly stronger relationship with NASS-based harvested areas than do CDL-based planted area and NASS-based planted area (Fig. 5.5). Table 5.2 shows the interannual variations of the relationships (slope and R^2 values in the simple linear regression models) between grain production and planted area and harvested area, and the poorest relationship occurred in the spring drought year 2014.

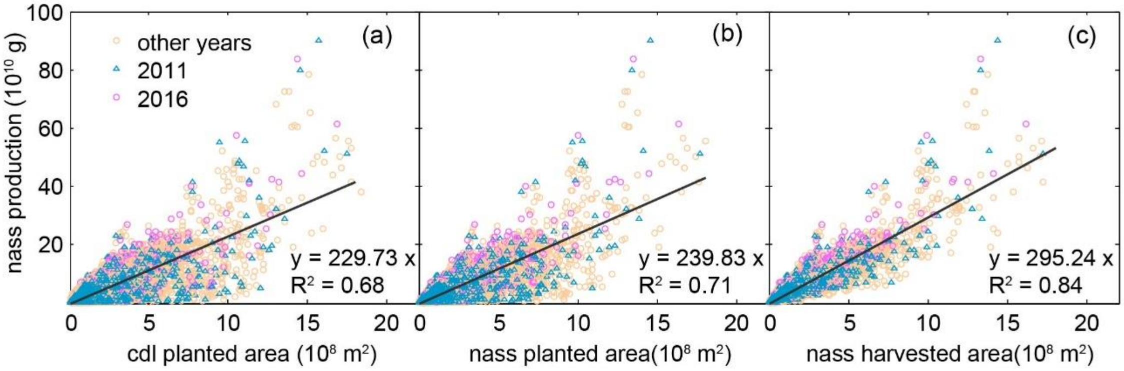

Figure 5.31 The relationships between winter wheat grain production from the NASS dataset and winter wheat cropping area from the CDL and NASS datasets during 2008-2018 at the county scale. The black solid line is the regression line for all the counties during 2008-2018.

Table 5.13 A summary statistics for the relationships between winter wheat grain production from the NASS dataset and winter wheat cropping areas from the CDL and NASS datasets during 2008-2018.

	plt_cdl		plt_nass		harv_nass	
	slope	R^2	slope	R^2	slope	R^2
2008	257.563	0.797	251.732	0.813	293.283	0.891
2009	214.095	0.662	215.726	0.692	273.397	0.863
2010	257.448	0.801	265.183	0.819	306.041	0.888
2011	215.925	0.601	225.973	0.631	290.309	0.795
2012	243.865	0.751	256.137	0.794	293.257	0.868
2013	221.878	0.610	227.598	0.632	296.460	0.817
2014	181.539	0.555	179.375	0.572	244.449	0.768
2015	197.584	0.723	213.490	0.761	256.620	0.852
2016	278.651	0.745	305.255	0.801	359.768	0.893
2017	244.456	0.675	275.462	0.748	343.499	0.876
2018	218.487	0.595	249.051	0.653	319.785	0.802

***5.3.2** The relationship between GPPVPM and grain production at county scale during 2008-2018: The harvest index as defined by GPP and grain production*

At the county scale, the NASS winter wheat grain production showed moderately linear relationships with annual GPP_{total} (Fig. 5.6a). When all county-year data in the CONUS were used in the simple linear regression model between grain production (GP) and GPP, GPP_{total} accounted for 68% of winter wheat grain production from the NASS dataset (Fig. 5.6a). The relationships between NASS winter wheat grain production and GPP_{VPM} varied by years, with R^2 values ranging from 0.59 in 2009 to 0.75 in 2016 (Table 5.3). The slope values in the simple linear regression models with no intercept ($GP = HI_{GPP} * GPP$) varied from 0.25 in 2012 to 0.33 in 2011 (Table 5.3).

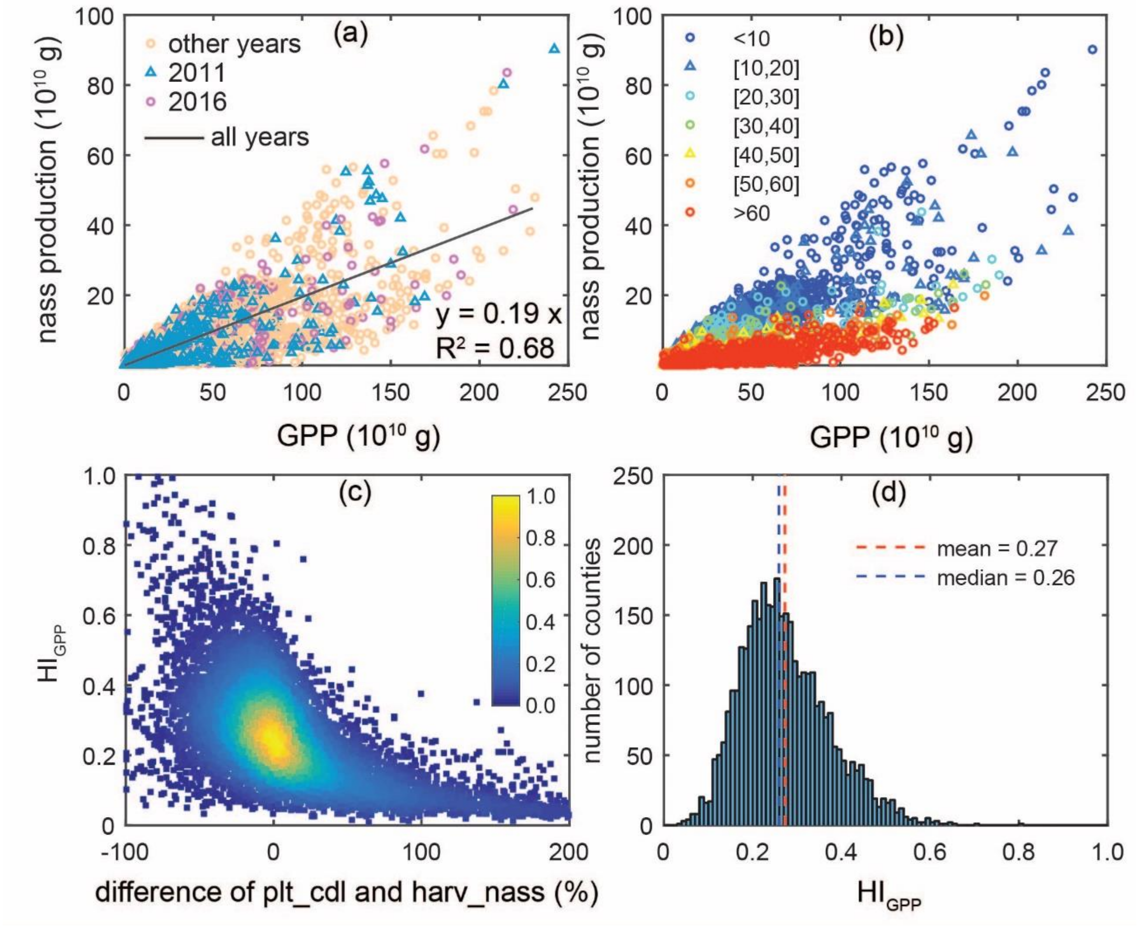

Figure 5.32 (a) Relationship between NASS production and cropping area for winter wheat from 2008-2018, the black solid line is the regression line for all the county-level production and cropping areas from 2008-2018. (b) Relationship between NASS production and cropping area for winter wheat from 2008-2018 considering the difference between CDL planted area (plt_cdl) and NASS harvested area (harv_nass). (c) Density plot of the relationship between HIGPP and difference of plt_cdl and harv_nass. (d) histogram of HIgpp for all county and years with a difference of plt_cdl and harv_nass less than <10%.

Table 5.14 Statistics of linear regression between NASS production and GPP_{VPM} for winter wheat during 2008-2018 (pvalue < 0.05)

Year	slope	R^2
2008	0.306	0.711
2009	0.298	0.591
2010	0.320	0.741
2011	0.343	0.688
2012	0.254	0.694
2013	0.290	0.692
2014	0.295	0.609
2015	0.229	0.660
2016	0.304	0.746
2017	0.306	0.713
2018	0.321	0.709

As the differences between CDL planted area and NASS harvested area varied by county and year, Fig. 5.6b shows that HI_{GPP}, the conversion coefficient from GPP_{total} to grain production, were significantly affected by the differences between CDL planted area and NASS harvested area, with a much higher value in those counties where the differences between planted and harvested area are less than 30%. The relationship between NASS winter wheat production and GPP_{total} is also stronger in those counties (Table 5.4).

Table 5.15 Statistics of linear regression between NASS production and GPP_{VPM} at counties by considering the relative differences between CDL planted area and NASS harvested area (pvalue < 0.05).

relative difference (%)	slope	R^2
[0,10)	0.27	0.87
[10,20)	0.22	0.83
[20,30)	0.17	0.79
[30,40)	0.14	0.79
[40,50)	0.11	0.81
[50,60)	0.1	0.83
>60	0.07	0.69

HI_{GPP} showed significant variation across counties and years, mostly ranging from 0-0.4 (Fig. 5.6c). This variation is largely contributed to the difference between CDL planted area and NASS harvested area (Fig. 5.6c). For those counties with similar CDL planted area and NASS harvested area (difference <10%), the HI_{GPP} is mainly concentrated in 0.2-0.4 (Fig. 5.6d).

5.3.3 *In-season forecasting of winter wheat grain production using cumulative GPP data*

In CONUS, winter wheat is usually planted in September to November, and harvested in June to August of next year. We assessed the potential of the simple linear regression model that used cumulative 8-day GPP over time (8-day time interval) to predict grain production of winter wheat at county scale over individual years (grain production = a * cumulated GPP + b). When using all the counties-year data in the CONUS, the model prediction skill increases over time and reaches in the range of 60% to 80% by the end of June (Fig. 5.7a). When using all the counties-year data in a state, the model prediction skills for those states located in the cold northern part of the CONUS, where there are little differences between the planted area and harvested area, reach 90% by the end of May (Fig. 5.7b,c). For those states located in the Southern Great Plains with big differences between CDL planted area and NASS harvested area, the model prediction skill varies over years, ranging from 70% to 90% (Fig. 5.7d,e). After excluding those counties with larger than 30% difference between CDL planted area and NASS harvested area, the model prediction skill for the CONUS increases and varies between 80% and 90% over years (Fig. 5.7f).

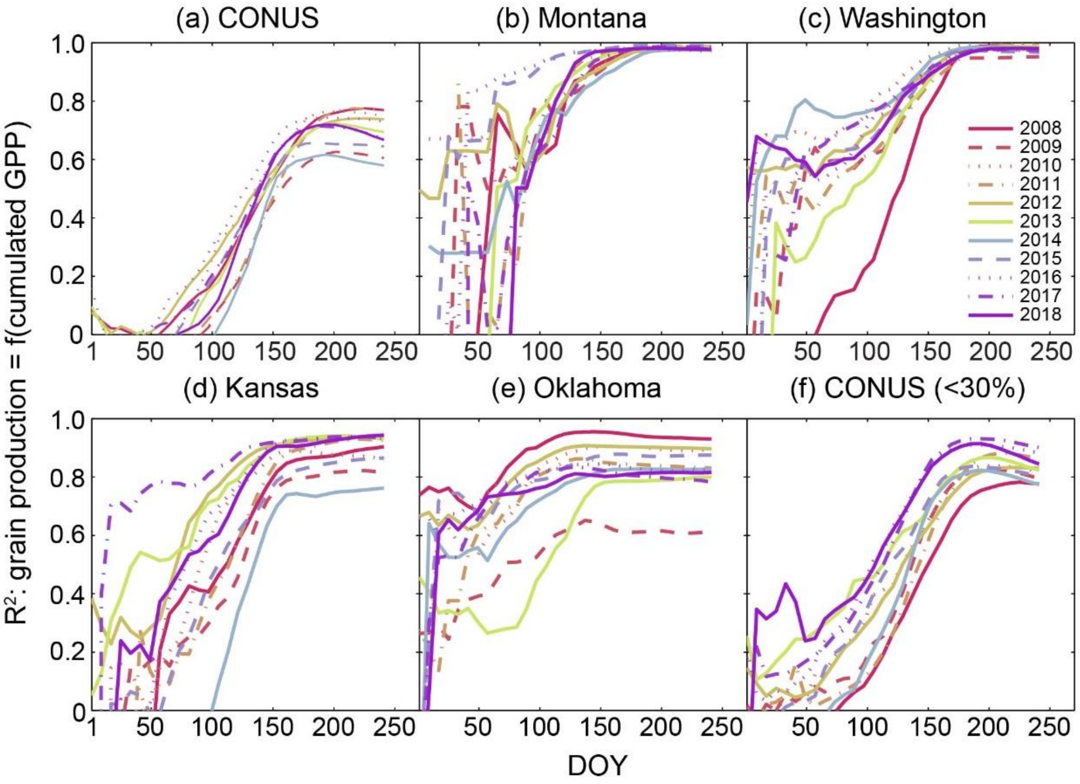

Figure 5.33 The prediction skill of the linear regression models that predict county-level crop grain production from NASS statistics by using accumulative GPP estimates over time (8-day interval) from the VPM model and CDL cropping area over years for winter wheat during 2008-2018 over counties in the (a) all counties in CONUS; (b) all counties in Montana; (c) all counties in Washington; (d) all counties in Kansas; (e) all counties in Oklahoma; (f) CONUS for all counties with less than 30% differences between CDL planted area and NASS harvested area.

5.4 Discussion

5.4.1 *Spatial-temporal consistency of winter wheat harvested area from NASS datasets and planted area from the NASS and CDL datasets*

Crop planted area and its spatial distribution are the first data layer for crop production monitoring and forecasting. In this study, the CDL winter wheat planted area estimates in the CONUS had a good spatial-temporal consistency with the NASS planted area estimates at county-level during 2008-2018 (R^2 = 0.98). This is consistent with the findings from several previous studies which compared these two different planted area estimates over some counties or states in a few years (Becker-Reshef et al. 2010; Boryan et al. 2011). This high spatial-temporal consistency can be largely attributed to the accuracy of the CDL and NASS statistical data, especially for major crops like maize, soybean and winter wheat (Boryan et al. 2011). The high accuracy and robustness of the CDL dataset for winter wheat planted area makes it reliable for crop monitoring.

Crop harvested area and its spatial distribution are directly related to crop production because some crop fields could not be harvested due to damage, failure, and other factors. In our study, we found the spatial-temporal consistency of winter wheat planted and harvested area in the CONUS varied substantially, especially in the Southern Great Plains and in some spring drought years (2011, 2013-2014). The Southern Great Plains is one of key regions for the Nation's wheat and beef production. In many counties of the Southern Great Plain region, winter wheat is used as a dual purpose graze-grain system, producing both a grain crop and a forage crop to supply beef cattle (Edwards et al. 2011; Hossain et al. 2004). Winter wheat intended for dual-purpose management in this region is usually sown two to four weeks earlier at higher seeding densities (1.5-2 times) than those used in grain-only system. Grazing is usually started 45-60 days after sowing and is terminated at the first hollow stem stage of growth (Edwards and Horn 2010). Most of those grazed winter wheat areas recovered and continued to grow for grain production after removing beef cattle, while some fields were kept as pasture (grazing-out) or baled for hay for beef cattle production (Steiner et al. 2018). Drought is an important factor that can lead to the loss of grain yield and abandonment of winter wheat acres in the Southern Great Plains. The winter wheat crop in this area normally breaks from dormancy in late February or early- to mid-March. A lack of precipitation or snow before the dormancy period could significantly reduce the growth of winter wheat crops (Steiner et al. 2018). Those could lead to the significant differences between planted and harvested area, especially over the Southern Great Plains.

Our study demonstrates the potential of the CDL and NASS datasets in understanding the changes of planted area, harvested area, and grain production of winter wheat in the CONUS during 2008-2018. Over these years, winter wheat production in the CONUS decreased in 2008-2010, then increased in 2010-2013, and decreased again during 2013-2018, which is mainly consistent with the fluctuation of planted/harvested area (Fig. 1). This can be largely explained by the international market and grain prices during this period. The winter wheat market in the U.S. have faced increasing competition from key competitors in the Black Sea, the European Union (EU), and South America. For example, Russia and Ukraine have displaced U.S. hard red winter wheat markets, while European soft winter wheat has challenged the U.S. soft wheat market in Middle Eastern and North African (Bond 2020).

5.4.2 *Harvest Index − The relationships between GPP and NASS grain production*

The “Harvest index” (HI) is an important term used in the study of crop production and is defined in three ways: (1) the ratio between crop grain production (yield) and aboveground biomass (AGB), namely HI_{AGB} (Prince et al. 2001); (2) the ratio between crop grain production (yield) and net primary production (NPP), namely HI_{NPP} (Guan et al. 2016; Lobell et al. 2002); and (3) the ratio between crop grain production (yield) and gross primary production (GPP), namely HI_{GPP} (He et al. 2018a). The three HIs definitions are different but related to each other, as GPP, NPP and AGB are related to each other. GPP represents the total carbon fixed by photosynthesis during the growing stage. NPP refers to the difference between GPP and autotrophic respiration and is often estimated as a sum of aboveground biomass and belowground biomass at the end of the plant growing season. AGB is the total biomass allocated to the leaf, stem, branches, and seeds. The grain production (yield) is part of AGB at harvest. Several studies have reported that there is a close relationship between canopy photosynthesis and yield at the seasonal scale (Yuan et al. 2016b; Zelitch 1982). Some previous studies utilized a fixed converting ratio from NPP to yield by applying a constant HI_{AGB} and a fraction of AGB to total biomass (Lobell et al. 2002; Prince et al. 2001). But The yield and photosynthesis relationship could be affected by wheat species, cultivars, and environmental factors (e.g., temperature, water, nitrogen) (Triboi and Triboi-Blondel 2002). For example, Guan et al. (2016) reported that the portion of photosynthesis (GPP) that turns to NPP for crops over the Midwest US varied with temperature based on NPP estimated from NASS yield and SIF-based GPP, ranging from 0.2-0.7 for the CONUS. Liu et al. (2008) also reported that plant roots tend to grow during the drought stress while shoot growth is decreased. Our results showed a mean HI_{GPP} of 0.27 over counties with the differences between CDL planted area and NASS harvested area less than 10%, but it also fluctuated over counties and years (Fig.5). This variation of GPP-yield relationship (HI_{GPP}) could not only result from above-mentioned environmental factors but also crop management and crop varieties. For example, the winter wheat yields in the Southern Great Plains are lower and more variable than other counties, because part of the crop is used as pasture when grazing is more profitable than harvesting. The rate of abandonment could vary in some climate extreme years as adverse weather reduces wheat yield. In addition, the fertilizer application, tillage, row spacing could also affect the yield both spatially and temporally (Vocke and Ali 2013). Further efforts are needed to quantify the responses of GPP and yield during climate extreme events.

5.4.3 *In-season forecasting of winter wheat grain production*

The LUE models provide a simple but efficient way to estimate biomass or yield (Yuan et al. 2016b). Unlike empirical models, they are semi-empirical and can be applied in both site or regional scale with calibration and validation (Marshall et al. 2018; Yuan et al. 2016b). In our study, we combined GPP estimated from the VPM_{gpp} and CDL crop area and evaluated the relationship of these estimates with official production statistics. The modelled crop grain production estimates correlated well with the county- and state-level NASS winter wheat grain production. Our current study focuses on county-level and state-level NASS reported crop production because the USDA NASS benchmark production data are readily available at those scales. The lack of field-scale crop yield records limit the validation of GPP_{VPM}-based estimates of grain production at the field scale. The GPP simulation in this study was executed at a moderate spatial resolution (500-m). For the winter wheat major producing region, which produced >40% of the total winter wheat, it is relatively close to the field scale because there are few mixed pixels in these counties and state. But for other regions, the mixed pixels problem will cause significant discrepancy in calculating wheat GPP and grain production. GPP products with high spatial resolution (e.g., 30-m Landsat, and 10-m Sentinel-2) and suitable temporal frequency (weekly) may overcome the above limitations for field-scale agricultural applications.

This study highlighted the importance of accurate and timely spatial classification of crop types in prediction of crop production, including both planted area and harvested area. Planted area provided the basic information at the start of growing season while harvested area is directly related to the grain production. For some major crops like maize and soybean, there could be no significant differences between planted area and harvested area in the CONUS. But our study showed there are significant differences between them for winter wheat over the CONUS, especially the Southern Great Plains, with a relative difference of planted to harvested area greater than 40%. Numerous studies tried to provide in-season crop types map as early as possible based on remote sensing data (Cai et al. 2018; Qiu et al. 2017), however, there is still a lack of attention about both planted and harvested area mapping. A separate classification of planted and harvested area can help to get a more accurate prediction of final production. Another issue is to have timely high resolution (e.g., 30 m or 10 m) crop classification map. The CDL data usually has six months lag time after harvest before releasing to the public, which precludes in-season production prediction. An emergence of some high-spatial resolution satellite observations, such as the Sentinel-2 (10 or 5-day revisit) and Sentinel-1 (12 or 6-day revisit), provides time-series image data at a finer spatial resolution (10-m) and weekly intervals, which offers an unprecedented opportunity to map different types of crops timely at a field scale, especially in those regions with complicated crop landscapes and frequent cloud cover.

5.5 Conclusion

We evaluated the spatial-temporal consistency among NASS crop statistical data (planted acreage, harvested acreage, grain production), satellite-based CDL planted acreage, GPP estimates from VPM and CDL planted area for winter wheat at county and national scales during 2008-2018. High consistency was found between CDL planted acreage and NASS planted acreage, but there is a significant difference between the planted acreage and harvested acreage. There is also moderate correlation between annual total GPP_{VPM} and grain production at county and state scales. Cumulative GPP_{VPM} at 8-day intervals within the wheat growing season, together with HI_{GPP}, was able to predict wheat grain production at county-level from early May to late June, varying in different states. The HI_{GPP}, which is calculated as the relationship slope between GPP and grain production, is fluctuate over counties and years, but more concentrated when excluding those counties with large differences of planted and harvested area, with an averaged value of 0.27. The results from this study highlight the importance to differentiate planted and harvested area in crop mapping. Our results also demonstrate the potential of accurate GPP_{VPM} products in monitoring grain production of winter wheat in the CONUS.

Chapter 6: Conclusions and perspectives

My research has advanced our knowledge in the field of remote sensing of vegetation in two main ways. First, we demonstrated the ability of SIF and hyperspectral-related vegetation indices to track the seasonality of GPP at the site level. Further, we analyzed the effect of drought on SIF and vegetation indices (Chapter 2). I also examined the combined impact of spring warming and summer drought on terrestrial carbon uptake at a regional scale by utilizing multiple datasets (Chapter 3). Then we applied the regional GPP products to estimate crop production, for maize/soybean (Chapter 4) and winter wheat (Chapter 5). As it is projected that an increase in the intensity and duration of climate extremes will happen in the future, it is not only important to understand annual carbon uptake and the regional carbon cycle, but also has important implications for monitoring food production to ensure food security and understanding how anthropogenic activity may help mitigate the impacts of climate extremes.

The results of my study in Chapter 2 and Chapter 3 have demonstrated the potential of SIF to track the seasonal dynamics of GPP and the drought impact at site and regional scales. In the future, it will be interesting to explore how to incorporate SIF into GPP models. However, more systematic work is needed for understanding the underlying mechanism for the relationship between GPP and SIF. There is a complex relationship between the carbon fixation during the Calvin cycle and SIF emitted from the leaf during the light reaction. It is necessary to better understand how this SIF-GPP relationship varies under environmental stress over time and for difference ecosystems. In addition, the in-situ or satellite observed top-of-canopy SIF is not identical to the total SIF emitted from the canopy, which is strongly affected by the direction of observation and the scattering and reabsorption related to the canopy structure. A combination of in-situ observations and radiative transfer models can help better understand the process.

For future regional application, an accurate and timely spatial classification of crop types is essential in prediction of crop production, including both planted area and harvested area. At state and county scales we found a strong correlation between VPM GPP and crop production for maize and soybean, but not for the dual-purpose winter wheat. There is still no timely high resolution (<=30m) map of annual crop type, let alone such in-season crop maps to identify planted and harvested area over years. Planted area provided the basic information at the start of growing season while harvested area is directly related to the grain production. A separate classification of crop types at the beginning of growing season and harvest time can help to accurately predict final production. Recently, emergence of high-spatial resolution satellite observations, such as the Sentinel-2 (10 or 5-day revisit) and Sentinel-1 (12 or 6-day revisit), provides time-series image data at a fine spatial resolution (10-m) and weekly intervals, which offers an unprecedented opportunity to map different types of crops timely at a field scale. Timely high-resolution crop planted area maps would allow high resolution GPP (e.g., 30-m Landsat, and 10-m Sentinel-2) for agricultural applications than the 500-m global GPP simulations currently available.

References

ADB, A.D.B. (2015). Results of the Methodological Studies for Agricultural and Rural Statistics. *Manila*

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N.C., Peylin, P., Piao, S.L., Sitch, S., Viovy, N., Wiltshire, A., & Zhao, M.S. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics, 53*, 785-818

Atzberger, C. (2013). Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs (vol 5, pg 949, 2013). *Remote Sensing, 5*, 4124-4124

Bacelar, E.L.V.A., Moutinho-Pereira, J.M., Gonçalves, B.M.C., Brito, C.V.Q., Gomes-Laranjo, J., Ferreira, H.M.F., & Correia, C.M. (2012). Water Use Strategies of Plants Under Drought Conditions, 145-170

Baldocchi, D., Falge, E., Gu, L.H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X.H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K.T.P., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., & Wofsy, S. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society, 82*, 2415-2434

Baldocchi, D.D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology, 9*, 479-492

Baldocchi, D.D., Xu, L.K., & Kiang, N. (2004). How plant functional-type, weather, seasonal drought, and soil physical properties alter water and energy fluxes of an oak-grass savanna and an annual grassland. *Agricultural and Forest Meteorology, 123*, 13-39

Becker-Reshef, I., Vermote, E., Lindeman, M., & Justice, C. (2010). A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sensing of Environment, 114*, 1312-1323

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rodenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., & Papale, D. (2010). Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science, 329,* 834-838

Bolton, D.K., & Friedl, M.A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology, 173,* 74-84

Bond, J.K. (2020). Wheat Outlook: May 2020. *ERS, May 2020*

Boryan, C., & Yang, Z. (2017). Integration of the cropland data layer based automatic stratification methods into the traditional area frame construction process. *Survey Research Methods, 11,* 289-306

Boryan, C., Yang, Z., Di, L., & Hunt, K. (2014). A New Automatic Stratification Method for U.S. Agricultural Area Sampling Frame Construction Based on the Cropland Data Layer. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7,* 4317-4327

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International, 26,* 341-358

Burke, M., & Lobell, D.B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc Natl Acad Sci U S A, 114,* 2189-2194

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology, 274,* 144-159

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment, 210,* 35-47

Carfagna, E., & Carfagna, A. (2010). Alternative Sampling Frames and Administrative Data. What is the Best Data Source for Agricultural Statistics? *Agricultural Survey Methods*

Castillioni, K., Wilcox, K., Jiang, L., Luo, Y., Jung, C.G., & Souza, L. (2020). Drought mildly reduces plant dominance in a temperate prairie ecosystem across years. *Ecology and Evolution, 10,* 6702-6713

Cavanaugh, M.L., Kurc, S.A., & Scott, R.L. (2011). Evapotranspiration partitioning in semiarid shrubland ecosystems: a two-site evaluation of soil moisture control on transpiration. *Ecohydrology, 4,* 671-681

Chang, J., Hansen, M.C., Pittman, K., Carroll, M., & DiMiceli, C. (2007). Corn and soybean mapping in the united states using MODN time-series data sets. *Agronomy Journal, 99,* 1654-1664

Chen, J.M. (1996). Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agricultural and Forest Meteorology, 80,* 135-163

Chen, T., van der Werf, G.R., Dolman, A.J., & Groenendijk, M. (2011). Evaluation of cropland maximum light use efficiency using eddy flux measurements in North America and Europe. *Geophysical Research Letters, 38,* n/a-n/a

Chen, X., Mo, X., Zhang, Y., Sun, Z., Liu, Y., Hu, S., & Liu, S. (2019). Drought detection and assessment with solar-induced chlorophyll fluorescence in summer maize growth period over North China Plain. *Ecological Indicators, 104*, 347-356

Chu, H.S., Chen, J.Q., Gottgens, J.F., Ouyang, Z.T., John, R., Czajkowski, K., & Becker, R. (2014). Net ecosystem methane and carbon dioxide exchanges in a Lake Erie coastal marsh and a nearby cropland. *Journal of Geophysical Research-Biogeosciences, 119,* 722-740

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A.D., Friedlingstein, P., Grunwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J.F., Sanz, M.J., Schulze, E.D., Vesala, T., & Valentini, R. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature, 437,* 529-533

Cohn, A.S., VanWey, L.K., Spera, S.A., & Mustard, J.F. (2016). Cropping frequency and area response to climate variability can exceed yield response. *Nature Climate Change*, *6*, 601-604

Collatz, G.J., Ribas-Carbo, M., & Berry, J.A. (1992). Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants. *Australian Journal of Plant Physiology, 19*, 519-538

Cook, B.D., Davis, K.J., Wang, W.G., Desai, A., Berger, B.W., Teclaw, R.M., Martin, J.G., Bolstad, P.V., Bakwin, P.S., Yi, C.X., & Heilman, W. (2004). Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA. *Agricultural and Forest Meteorology, 126,* 271-295

Dai, A. (2012). Increasing drought under global warming in observations and models. *Nature Climate Change, 3,* 52-58

De Groote, H., & Traore, O. (2005). The cost of accuracy in crop area estimation. *Agricultural Systems, 84*, 21-38

Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J.A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., Kang, M., Li, J., & Moya, I. (2020). Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops. *Remote Sensing of Environment, 241*, 111733

Disney, M., Lewis, P., & Saich, P. (2006). 3D modelling of forest canopy structure for remote sensing simulations in the optical and microwave domains. *Remote Sensing of Environment, 100*, 114-132

Dong, J.W., Xiao, X.M., Wagle, P., Zhang, G.L., Zhou, Y.T., Jin, C., Torn, M.S., Meyers, T.P., Suyker, A.E., Wang, J.B., Yan, H.M., Biradar, C., & Moore, B. (2015). Comparison of four EVI-based models for estimating gross primary production of maize and soybean croplands and tallgrass prairie under severe drought. *Remote Sensing of Environment, 162*, 154-168

Doraiswamy, P.C., Moulin, S., Cook, P.W., & Stern, A. (2003). Crop yield assessment from remote sensing. *Photogrammetric Engineering and Remote Sensing, 69*, 665-674

Doughty, R., Xiao, X., Wu, X., Zhang, Y., Bajgain, R., Zhou, Y., Qin, Y., Zou, Z., McCarthy, H., Friedman, J., Wagle, P., Basara, J., & Steiner, J. (2018). Responses of gross primary production of grasslands and croplands under drought, pluvial, and irrigation conditions during 2010–2016, Oklahoma, USA. *Agricultural Water Management, 204*, 47-59

Edwards, J.T., Carver, B.F., Horn, G.W., & Payton, M.E. (2011). Impact of Dual-Purpose Management on Wheat Grain Yield. *Crop Science, 51*, 2181-2185

Edwards, J.T., & Horn, G.W. (2010). First hollow stem: A critical growth stage for dual-purpose producers. *Fact Sheet PSS 2147. Okla. State Univ. Coop. Ext. Serv., Stillwater, OK.*

Erickson, B., Lowenberg-Deboer, J., & Bradford, J. (2017). 2017 precision agriculture dealership survey

Farquhar, G.D., Caemmerer, S.V., & Berry, J.A. (1980). A Biochemical Model of Photosynthetic CO2 Assimilation in Leaves of C3 Species. *Planta, 149*, 78-90

FAS/USDA (2020). World Agricultural Production

Fischer, M.L., Billesbach, D.P., Berry, J.A., Riley, W.J., & Torn, M.S. (2007). Spatiotemporal Variations in Growing Season Exchanges of CO2, H2O, and Sensible Heat in Agricultural Fields of the Southern Great Plains. *Earth Interactions, 11*, 1-21

Foerster, S., Kaden, K., Foerster, M., & Itzerott, S. (2012). Crop type mapping using spectral-temporal profiles and phenological information. *Computers and Electronics in Agriculture, 89*, 30-40

Foody, G.M., & Dash, J. (2007). Discriminating and mapping the C3 and C4 composition of grasslands in the northern Great Plains, USA. *Ecological Informatics, 2*, 89-93

Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., & Sobrino, J.A. (2015). Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. *Remote Sensing of Environment, 161*, 131-148

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M.D., Smith, P., van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J.G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S.I., Walz, A., Wattenbach, M., Zavala, M.A., & Zscheischler, J. (2015). Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Glob Chang Biol, 21*, 2861-2880

Frankenberg, C., Fisher, J.B., Worden, J., Badgley, G., Saatchi, S.S., Lee, J.-E., Toon, G.C., Butz, A., Jung, M., Kuze, A., & Yokota, T. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters, 38,* n/a-n/a

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment, 114*, 168-182

Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., Mücher, S., Tetrault, R., Moorthy, I., & McCallum, I. (2019). A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, *168*, 258-272

Gardiner (2016). U.S. Agriculture Exports: Recent Trends and Commodity Exposure to International Trade. *Economic Report, FCA – ORP – AEPT*

Gornott, C., & Wechsung, F. (2016). Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany. *Agricultural and Forest Meteorology, 217,* 89-100

Gough, C.M., Hardiman, B.S., Nave, L.E., Bohrer, G., Maurer, K.D., Vogel, C.S., Nadelhoffer, K.J., & Curtis, P.S. (2013). Sustained carbon uptake and storage following moderate disturbance in a Great Lakes forest. *Ecological Applications, 23*, 1202-1215

Gough, C.M., Vogel, C.S., Schmid, H.P., Su, H.B., & Curtis, P.S. (2008). Multi-year convergence of biometric and meteorological estimates of forest carbon storage. *Agricultural and Forest Meteorology, 148,* 158-170

Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., & Lobell, D.B. (2016). Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. *Glob Chang Biol, 22*, 716-726

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., & Griffis, T.J. (2014). Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proc Natl Acad Sci U S A, 111*, E1327-1333

Han, W.G., Yang, Z., Di, L.P., & Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture, 84,* 111-123

Hashimoto, H., Wang, W., Milesi, C., Xiong, J., Ganguly, S., Zhu, Z., & Nemani, R. (2013). Structural Uncertainty in Model-Simulated Trends of Global Gross Primary Production. *Remote Sensing, 5,* 1258-1273

Hatala, J.A., Detto, M., Sonnentag, O., Deverel, S.J., Verfaillie, J., & Baldocchi, D.D. (2012). Greenhouse gas (CO2, CH4, H2O) fluxes from drained and flooded agricultural peatlands in the Sacramento-San Joaquin Delta. *Agriculture Ecosystems & Environment, 150,* 1-18

Hatfield, J.L., Gitelson, A.A., Schepers, J.S., & Walthall, C.L. (2008). Application of spectral remote sensing for agronomic decisions. *Agronomy Journal, 100*, S117-S131

Hay, R.K.M. (1995). Harvest Index - a Review of Its Use in Plant-Breeding and Crop Physiology. *Annals of Applied Biology, 126*, 197-216

He, M., Kimball, J., Maneta, M.P., Maxwell, B., Moreno, A., Begueria, S., & Wu, X. (2018a). Regional crop gross primary productivity and yield estimation using fused Landsat-MODIS data. *Remote Sensing, 10,* 372

He, M.Z., Kimball, J.S., Maneta, M.P., Maxwell, B.D., Moreno, A., Begueria, S., & Wu, X.C. (2018b). Regional Crop Gross Primary Productivity and Yield Estimation Using Fused Landsat-MODIS Data. *Remote Sensing, 10*

He, W., Ju, W., Schwalm, C., Sippel, S., Wu, X., He, Q., Song, L., Zhang, C., Li, J., Sitch, S., Viovy, N., Friedlingstein, P., & Jain, A. (2018c). Terrestrial net carbon uptake over North America in 2011 and 2012. *Journal of Geophysical Research-Biogeosciences, 123*

Hilker, T., Lyapustin, A.I., Tucker, C.J., Hall, F.G., Myneni, R.B., Wang, Y.J., Bi, J., de Moura, Y.M., & Sellers, P.J. (2014). Vegetation dynamics and rainfall sensitivity of the Amazon. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 16041-16046

Hilton, T.W., Whelan, M.E., Zumkehr, A., Kulkarni, S., Berry, J.A., Baker, I.T., Montzka, S.A., Sweeney, C., Miller, B.R., & Elliott Campbell, J. (2017). Peak growing season gross uptake of carbon in North America is largest in the Midwest USA. *Nature Climate Change*, *7*, 450-454

Hoerling, M., Eischeid, J., Kumar, A., Leung, R., Mariotti, A., Mo, K., Schubert, S., & Seager, R. (2014). Causes and Predictability of the 2012 Great Plains Drought. *Bulletin of the American Meteorological Society, 95*, 269-282

Hossain, I., Epplin, F.M., Horn, G.W., & Krenzer, E.G. (2004). Wheat production and management practices used by Oklahoma grain and livestock producers. *Bull. 818. Okla. State Univ. Coop. Ext. Serv., Stillwater, OK.*

Huang, J.X., Sedano, F., Huang, Y.B., Ma, H.Y., Li, X.L., Liang, S.L., Tian, L.Y., Zhang, X.D., Fan, J.L., & Wu, W.B. (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agricultural and Forest Meteorology, 216*, 188-202

Huete, A.R., Liu, H.Q., Batchily, K., & vanLeeuwen, W. (1997). A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sensing of Environment, 59*, 440-451

Hurtt, G.C., Pacala, S.W., Moorcroft, P.R., Caspersen, J., Shevliakova, E., Houghton, R.A., & Moore, B., 3rd (2002). Projecting the future of the U.S. carbon sink. *Proc Natl Acad Sci U S A*, *99*, 1389-1394

Iizumi, T., & Ramankutty, N. (2015). How do weather and climate influence cropping area and intensity? *Global Food Security, 4*, 46-50

Illston, B.G., Basara, J.B., Fisher, D.K., Elliot, R., Fiebrich, C.A., Crawford, K.C., Humes, K., & Hunt, E. (2008). Mesoscale monitoring of soil moisture across a statewide network. *Journal of Atmospheric and Oceanic Technology, 25*, 167-182

IPCC (2013). Climate change 2013: The Physics Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovenmental Panel on Climate Change. *Tech. rep., IPCC, Cambridge University Press, New York, NY, USA, 2013*

Jin, C., Xiao, X., Merbold, L., Arneth, A., Veenendaal, E., & Kutsch, W.L. (2013). Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. *Remote Sensing of Environment*, *135*, 189-201

Jin, C., Xiao, X.M., Wagle, P., Griffis, T., Dong, J.W., Wu, C.Y., Qin, Y.W., & Cook, D.R. (2015). Effects of in-situ and reanalysis climate data on estimation of cropland gross primary production using the Vegetation Photosynthesis Model. *Agricultural and Forest Meteorology, 213*, 240-250

Johnson, D.M. (2016). A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *International Journal of Applied Earth Observation and Geoinformation, 52,* 65-81

Johnson, D.R., & Major, D.J. (1979). Harvest Index of Soybeans as Affected by Planting Date and Maturity Rating. *Agronomy Journal, 71,* 538-541

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A.P., Middleton, E.M., Huemmrich, K.F., Yoshida, Y., & Frankenberg, C. (2013). Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques*, *6,* 2803-2823

Kalfas, J.L., Xiao, X., Vanegas, D.X., Verma, S.B., & Suyker, A.E. (2011). Modeling gross primary production of irrigated and rain-fed maize using MODIS imagery and CO2 flux tower data. *Agricultural and Forest Meteorology, 151*, 1514-1528

Knutson, T.R., Zeng, F., & Wittenberg, A.T. (2013). The extreme March-May 2012 warm anomaly over the Eastern United States: Global context and multimodel trend analysis. *Bulletin of the American Meteorological Society, 94,* S13-S17

Krisnawati, A., & Adie, M. (2015). Variability of Biomass and Harvest Index from Several Soybean Genotypes as Renewable Energy Source. *Energy Procedia, 65,* 14-21

Lark, T.J., Mueller, R., Johnson, D.M., & Gibbs, H.K. (2017). Measuring land-use and land-cover change using the U.S. department of agriculture’s cropland data layer: Cautions and recommendations. *International Journal of Applied Earth Observation and Geoinformation, 62*, 224-235

Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P., & Wohlfahrt, G. (2010). Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biology, 16*, 187-208

Law, B.E., Turner, D., Campbell, J., Sun, O.J., Van Tuyl, S., Ritts, W.D., & Cohen, W.B. (2004). Disturbance and climate effects on carbon stocks and fluxes across Western Oregon USA. *Global Change Biology, 10*, 1429-1444

Law, B.E., Williams, M., Anthoni, P.M., Baldocchi, D., & Unsworth, M.H. (2000). Measuring and modelling seasonal variation of carbon dioxide and water vapour exchange of a pinus ponderosa forest subject to soil water deficit. *Global Change Biology, 6*, 613-630

Le Quéré, C., Raupach, M.R., Canadell, J.G., Marland, G., Le Quéré et al, C., Le Quéré et al, C., Raupach, M.R., Canadell, J.G., Marland, G., Bopp, L., Ciais, P., Conway, T.J., Doney, S.C., Feely, R.A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R.A., House, J.I., Huntingford, C., Levy, P.E., Lomas, M.R., Majkut, J., Metzl, N., Ometto, J.P., Peters, G.P., Prentice, I.C., Randerson, J.T., Running, S.W., Sarmiento, J.L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G.R., Woodward, F.I., & et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience, 2*, 831-836

Leng, G.Y., Zhang, X.S., Huang, M.Y., Yang, Q.C., Rafique, R., Asrar, G.R., & Leung, L.R. (2016). Simulating county-level crop yields in the Conterminous United States using the Community Land Model: The effects of optimizing irrigation and fertilization. *Journal of Advances in Modeling Earth Systems, 8*, 1912-1931

Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature, 529*, 84-87

Leuning, R., Cleugh, H.A., Zegelin, S.J., & Hughes, D. (2005). Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. *Agricultural and Forest Meteorology, 129*, 151-173

Li, J.D., Duan, Q.Y., Wang, Y.P., Gong, W., Gan, Y.J., & Wang, C. (2018). Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling. *International Journal of Climatology, 38*, E1016-E1031

Li, X.L., Liang, S.L., Yu, G.R., Yuan, W.P., Cheng, X., Xia, J.Z., Zhao, T.B., Feng, J.M., Ma, Z.G., Ma, M.G., Liu, S.M., Chen, J.Q., Shao, C.L., Li, S.G., Zhang, X.D., Zhang, Z.Q., Chen, S.P., Ohta, T., Varlagin, A., Miyata, A., Takagi, K., Saiqusa, N., &

Kato, T. (2013). Estimation of gross primary production over the terrestrial ecosystems in China. *Ecological Modelling, 261*, 80-92

Li, Z., Zhang, Q., Li, J., Yang, X., Wu, Y., Zhang, Z., Wang, S., Wang, H., & Zhang, Y. (2020). Solar-induced chlorophyll fluorescence and its link to canopy photosynthesis in maize from continuous ground measurements. *Remote Sensing of Environment, 236*, 111420

Licht, M.A., Parvej, M.R., & Wright, E.E. (2019). Corn Yield Response to Row Spacing and Plant Population in Iowa. *Crop, Forage & Turfgrass Management, 5*

Liu, X., Jin, J., Wang, G., & Herbert, S.J. (2008). Soybean yield physiology and development of high-yielding practices in Northeast China. *Field Crops Research, 105*, 157-171

Liu, X., Liu, L., Hu, J., Guo, J., & Du, S. (2020). Improving the potential of red SIF for estimating GPP by downscaling from the canopy level to the photosystem level. *Agricultural and Forest Meteorology, 281*, 107846

Lobell, D.B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research, 143*, 56-64

Lobell, D.B., & Azzari, G. (2017). Satellite detection of rising maize yield heterogeneity in the U.S. Midwest. *Environmental Research Letters, 12*, 014014

Lobell, D.B., Hicke, J.A., Asner, G.P., Field, C., Tucker, C., & Los, S. (2002). Satellite estimates of productivity and light use efficiency in United States agriculture, 1982-98. *Global Change Biology, 8*, 722-735

Lobell, D.B., & Ortiz-Monasterio, J.I. (2008). Satellite monitoring of yield responses to irrigation practices across thousands of fields. *Agronomy Journal, 100*, 1005-1012

Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., & Hammer, G. (2014). Greater sensitivity to drought accompanies maize yield increases in the U.S. Midwest. *Science, 344*, 516-519

Lobell, D.B., Sibley, A., & Ivan Ortiz-Monasterio, J. (2012). Extreme heat effects on wheat senescence in India. *Nature Climate Change, 2*, 186-189

Ma, S., Baldocchi, D.D., Xu, L., & Hehn, T. (2007). Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California. *Agricultural and Forest Meteorology, 147*, 157-171

Mahlein, A.K., Oerke, E.C., Steiner, U., & Dehne, H.W. (2012). Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology, 133*, 197-209

Marshall, M., Tu, K., & Brown, J. (2018). Optimizing a remote sensing production efficiency model for macro-scale GPP and yield estimation in agroecosystems. *Remote Sensing of Environment, 217*, 258-271

Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M., & Sánchez Meador, A.J. (2017). MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types. *Remote Sensing of Environment, 198*, 490-503

Matamala, R., Jastrow, J.D., Miller, R.M., & Garten, C.T. (2008). Temporal changes in C and N stocks of restored prairie: Implications for C sequestration strategies. *Ecological Applications, 18*, 1470-1488

Maxwell, K., & Johnson, G.N. (2000). Chlorophyll fluorescence - a practical guide. *Journal of Experimental Botany, 51*, 659-668

Meade, B., Puricelli, E., McBride, W., Valdes, C., Hoffman, L., Foreman, L., & Dohlman, E. (2016). Corn and soybean production costs and export competitiveness in Argentina, Brazil, and the United States. *EIB-154, U.S. Department of Agriculture, Economic Research Service*

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., & Moreno, J. (2009). Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. *Remote Sensing of Environment, 113*, 2037-2051

Miao, G., Guan, K., Suyker, A.E., Yang, X., Arkebauer, T.J., Walter‐Shea, E.A., Kimm, H., Hmimina, G.Y., Gamon, J.A., Franz, T.E., Frankenberg, C., Berry, J.A., & Wu, G. (2020). Varying Contributions of Drivers to the Relationship Between Canopy Photosynthesis and Far‐Red Sun‐Induced Fluorescence for Two Maize Sites at Different Temporal Scales. *Journal of Geophysical Research: Biogeosciences, 125*

Miao, G., Guan, K., Yang, X., Bernacchi, C.J., Berry, J.A., DeLucia, E.H., Wu, J., Moore, C.E., Meacham, K., Cai, Y., Peng, B., Kimm, H., & Masters, M.D. (2018). Sun-

Induced Chlorophyll Fluorescence, Photosynthesis, and Light Use Efficiency of a Soybean Field from Seasonally Continuous Measurements. *Journal of Geophysical Research: Biogeosciences, 123*, 610-623

Mitchell, D.O., & Mielke, M. (2005). Wheat: The global market, policies, and priorities

Monfreda, C., Ramankutty, N., & Foley, J.A. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles, 22*, n/a-n/a

Monteith, J.L. (1972). Solar-Radiation and Productivity in Tropical Ecosystems. *Journal of Applied Ecology, 9*, 747-766

Monteith, J.L. (1977). Climate and Efficiency of Crop Production in Britain. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 281*, 277-294

Mulla, D.J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering, 114*, 358-371

NASS/USDA (2012). The Yield Forecasting Program of NASS. *National Agricultural Statistics Service, U.S. Department of Agriculture, Staff Report SMB 06-01*

NASS/WAOB/USDA (1999). Understanding USDA crop forecasts. *National Agricultural Statistics Service and World Agricultural Outlook Board, U.S. Department of Agriculture, Miscellaneous Publication, No. 1554*

Ollinger, S.V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist, 189*, 375-394

Pacala, S., Hurtt, G., Baker, D.B., & Peylin, P. (2001). Consistent land- and atmosphere-based U.S. carbon sink estimates. *Science, 292*, 2316-2319

Pan, Y.D., & Schimel, D. (2016). Biogeochemistry: Synergy of a warm spring and dry summer. *Nature, 534*, 483-484

Parazoo, N.C., Barnes, E., Worden, J., Harper, A.B., Bowman, K.B., Frankenberg, C., Wolf, S., Litvak, M., & Keenan, T.F. (2015). Influence of ENSO and the NAO on terrestrial carbon uptake in the Texas-northern Mexico region. *Global Biogeochemical Cycles, 29*, 1247-1265

Peng, B., Guan, K., Pan, M., & Li, Y. (2018). Benefits of Seasonal Climate Prediction and Satellite Data for Forecasting U.S. Maize Yield. *Geophysical Research Letters, 45,* 9662-9671

Peters, W., Jacobson, A.R., Sweeney, C., Andrews, A.E., Conway, T.J., Masarie, K., Miller, J.B., Bruhwiler, L.M., Petron, G., Hirsch, A.I., Worthy, D.E., van der Werf, G.R., Randerson, J.T., Wennberg, P.O., Krol, M.C., & Tans, P.P. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc Natl Acad Sci U S A, 104,* 18925-18930

Peylin, P., Law, R.M., Gurney, K.R., Chevallier, F., Jacobson, A.R., Maki, T., Niwa, Y., Patra, P.K., Peters, W., Rayner, P.J., Rödenbeck, C., van der Laan-Luijkx, I.T., & Zhang, X. (2013). Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions. *Biogeosciences, 10,* 6699-6720

Porcar-Castell, A., Tyystjarvi, E., Atherton, J., van der Tol, C., Flexas, J., Pfundel, E.E., Moreno, J., Frankenberg, C., & Berry, J.A. (2014). Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. *Journal of Experimental Botany, 65,* 4065-4095

Potter, C., Klooster, S., Genovese, V., Hiatt, C., Boriah, S., Kumar, V., Mithal, V., & Garg, A. (2012). Terrestrial ecosystem carbon fluxes predicted from MODIS satellite data and large-scale disturbance modeling. *International Journal of Geosciences*

Potter, C., Randerson, J.T., Field, C., Matson, P., Vitousek, P., Mooney, H., & Klooster, S. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles, 7,* 811-841

Prince, S.D., Haskett, J., Steininger, M., Strand, H., & Wright, B. (2001). Net primary production of U.S. midwest croplands from agricultural harvest yield data. *Ecological Indicators, 11*, 1194-1205

Pritsolas, J., & Pearson, R. (2018). Critical Review of Supporting Literature on Land Use Change in the EPA’s Second Triennial Report to Congress. *Renewable Fuels Association*

Qiu, B., Luo, Y., Tang, Z., Chen, C., Lu, D., Huang, H., Chen, Y., Chen, N., & Xu, W. (2017). Winter wheat mapping combining variations before and after estimated heading dates. *Isprs Journal of Photogrammetry and Remote Sensing, 123*, 35-46

Ramankutty, N., Evan, A.T., Monfreda, C., & Foley, J.A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles, 22*, n/a-n/a

Rammig, A., Wiedermann, M., Donges, J.F., Babst, F., von Bloh, W., Frank, D., Thonicke, K., & Mahecha, M.D. (2014). Tree-ring responses to extreme climate events as benchmarks for terrestrial dynamic vegetation models. *Biogeosciences Discussions, 11*, 2537-2568

Randerson, J.T., Thompson, M.V., Malmstrom, C.M., Field, C.B., & Fung, I.Y. (1996). Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO2. *Global Biogeochemical Cycles, 10*, 585-602

Ray, D.K., Mueller, N.D., West, P.C., & Foley, J.A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *Plos One, 8*

Reeves, M.C., Zhao, M., & Running, S.W. (2005). Usefulness and limits of MODIS GPP for estimating wheat yield. *International Journal of Remote Sensing, 26*, 1403-1421

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., & Wattenbach, M. (2013). Climate extremes and the carbon cycle. *Nature, 500*, 287-295

Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., OgÉE, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., GrÜNwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Pilegaard, K., Pumpanen, J., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J.F., Sanz, M.J., Vesala, T., & Zhao, M. (2007). Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology, 13*, 634-651

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., & Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology, 11*, 1424-1439

Reichstein, M., Tenhunen, J.D., Roupsard, O., Ourcival, J.M., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G., & Valentini, R. (2002). Severe drought effects on ecosystem CO(2) and H(2)O fluxes at three Mediterranean evergreen sites: revision of current hypotheses? *Global Change Biology, 8*, 999-1017

Rippey, B.R. (2015). The U.S. drought of 2012. *Weather and Climate Extremes, 10, Part A*, 57-64

Running, S., Nemani, R., Heinsch, F.A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A contiguous satellite-derived measure of global terrestrial primary production. *Bioscience, 54*, 547-560

Sakamoto, T., Gitelson, A.A., & Arkebauer, T.J. (2014). Near real-time prediction of US corn yields based on time-series MODIS data. *Remote Sensing of Environment, 147,* 219-231

Schaefer, K., Collatz, G.J., Tans, P., Denning, A.S., Baker, I., Berry, J., Prihodko, L., Suits, N., & Philpott, A. (2008). Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model. *Journal of Geophysical Research, 113*

Schaefer, K., Schwalm, C.R., Williams, C., Arain, M.A., Barr, A., Chen, J.M., Davis, K.J., Dimitrov, D., Hilton, T.W., Hollinger, D.Y., Humphreys, E., Poulter, B., Raczka, B.M., Richardson, A.D., Sahoo, A., Thornton, P., Vargas, R., Verbeeck, H., Anderson, R., Baker, I., Black, T.A., Bolstad, P., Chen, J., Curtis, P.S., Desai, A.R., Dietze, M., Dragoni, D., Gough, C., Grant, R.F., Gu, L., Jain, A., Kucharik, C., Law, B., Liu, S., Lokipitiya, E., Margolis, H.A., Matamala, R., McCaughey, J.H., Monson, R., Munger, J.W., Oechel, W., Peng, C., Price, D.T., Ricciuto, D., Riley, W.J., Roulet, N., Tian, H., Tonitto, C., Torn, M., Weng, E., & Zhou, X. (2012). A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research: Biogeosciences, 117,* doi:10.1029/2007JG000603

Schmid, H.P., Grimmond, S.B., Cropley, F., Offerle, B., & Su, H. (2000). Measurements of CO2 and energy fluxes over a mixed hardwood forest in the mid-western United States. *Agricultural and Forest Meteorology, 103*, 357-374

Schneider, J.M., Fisher, D.K., Elliott, R.L., Brown, G.O., & Bahrmann, C.P. (2003). Spatiotemporal variations in soil water: First results from the ARM SGP CART network. *Journal of Hydrometeorology, 4*, 106-120

Schwalm, C.R., Williams, C.A., Schaefer, K., Anderson, R., Arain, M.A., Baker, I., Barr, A., Black, T.A., Chen, G., Chen, J.M., Ciais, P., Davis, K.J., Desai, A., Dietze, M., Dragoni, D., Fischer, M.L., Flanagan, L.B., Grant, R., Gu, L., Hollinger, D., Izaurralde, R.C., Kucharik, C., Lafleur, P., Law, B.E., Li, L., Li, Z., Liu, S., Lokupitiya, E., Luo, Y., Ma, S., Margolis, H., Matamala, R., McCaughey, H., Monson, R.K., Oechel, W.C., Peng, C., Poulter, B., Price, D.T., Riciutto, D.M., Riley, W., Sahoo, A.K., Sprintsin, M., Sun, J., Tian, H., Tonitto, C., Verbeeck, H., & Verma, S.B. (2010). A model-data intercomparison of CO2exchange across North America: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research, 115*

Scott, R.L., Biederman, J.A., Hamerlynck, E.P., & Barron-Gafford, G.A. (2015). The carbon balance pivot point of southwestern U.S. semiarid ecosystems: Insights from the 21st century drought. *Journal of Geophysical Research: Biogeosciences, 120*, 2612-2624

Scott, R.L., Hamerlynck, E.P., Jenerette, G.D., Moran, M.S., & Barron-Gafford, G.A. (2010). Carbon dioxide exchange in a semidesert grassland through drought-induced vegetation change. *Journal of Geophysical Research, 115*

Scott, R.L., Jenerette, G.D., Potts, D.L., & Huxman, T.E. (2009). Effects of seasonal drought on net carbon dioxide exchange from a woody-plant-encroached semiarid grassland. *Journal of Geophysical Research, 114*

Sims, D.A., Brzostek, E.R., Rahman, A.F., Dragoni, D., & Phillips, R.P. (2014). An improved approach for remotely sensing water stress impacts on forest C uptake. *Global Change Biology, 20*, 2856-2866

Sippel, S., Forkel, M., Rammig, A., Thonicke, K., Flach, M., Heimann, M., Otto, F.E.L., Reichstein, M., & Mahecha, M.D. (2017). Contrasting and interacting changes in simulated spring and summer carbon cycle extremes in European ecosystems. *Environmental Research Letters, 12*, 075006

Sippel, S., Zscheischler, J., & Reichstein, M. (2016). Ecosystem impacts of climate extremes crucially depend on the timing. *Proc Natl Acad Sci U S A, 113*, 5768-5770

Skakun, S., Franch, B., Vermote, E., Roger, J.-C., Becker-Reshef, I., Justice, C., & Kussul, N. (2017). Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. *Remote Sensing of Environment, 195*, 244-258

Song, C., Dannenberg, M.P., & Hwang, T. (2013). Optical remote sensing of terrestrial ecosystem primary productivity. *Progress in Physical Geography, 37*, 834-854

Steiner, J.L., Briske, D.D., Brown, D.P., & Rottler, C.M. (2018). Vulnerability of Southern Plains agriculture to climate change. *Climatic Change, 146*, 201-218

Steiner, J.L., Schneider, J.M., Pope, C., & Steele, R.F. (2015). Southern Plains Assessment of Vulnerability and Preliminary Adaptation and Mitigation Strategies for Farmers, Ranchers and Forest Land Owners. *T. Anderson, Ed., United States Department of Agriculture, 61 pp.*

Still, C.J., Berry, J.A., Collatz, G.J., & DeFries, R.S. (2003). Global distribution of C3 and C4 vegetation: carbon cycle implications. *Global Biogeochemical Cycles, 17*

Sturtevant, C., Ruddell, B.L., Knox, S.H., Verfaillie, J., Matthes, J.H., Oikawa, P.Y., & Baldocchi, D. (2016). Identifying scale-emergent, nonlinear, asynchronous processes of wetland methane exchange. *Journal of Geophysical Research-Biogeosciences, 121*, 188-204

Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J., Gu, L., Magney, T.S., Kohler, P., Evans, B., & Yuen, K. (2017). OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science, 358*

Sun, Y., Fu, R., Dickinson, R., Joiner, J., Frankenberg, C., Gu, L.H., Xia, Y.L., & Fernando, N. (2015). Drought onset mechanisms revealed by satellite solar-induced chlorophyll fluorescence: Insights from two contrasting extreme events. *Journal of Geophysical Research-Biogeosciences, 120*, 2427-2440

Suyker, A.E., Verma, S.B., Burba, G.G., & Arkebauer, T.J. (2005). Gross primary production and ecosystem respiration of irrigated maize and irrigated soybean during a growing season. *Agricultural and Forest Meteorology, 131*, 180-190

Tilman, D., Balzer, C., Hill, J., & Befort, B.L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108,* 20260-20264

Trenberth, K.E., Dai, A., van der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., & Sheffield, J. (2013). Global warming and changes in drought. *Nature Climate Change*, *4*, 17-22

Triboi, E., & Triboi-Blondel, A.M. (2002). Productivity and grain or seed composition: a new approach to an old problem - invited paper. *European Journal of Agronomy, 16,* 163-186

Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Running, S.W., Zhao, M., Costa, M.H., Kirschbaum, A.A., Ham, J.M., Saleska, S.R., & Ahl, D.E. (2006). Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment, 102,* 282-292

Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., & Munger, J.W. (2007). Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest. *Journal of Geophysical Research, 112,* doi: 10.1029/2006JG000293

USDA, N. (2014). USDA Crop Production 2014 Summary. *USDA, 1*

van der Laan-Luijkx, I.T., van der Velde, I.R., van der Veen, E., Tsuruta, A., Stanislawska, K., Babenhauserheide, A., Zhang, H.F., Liu, Y., He, W., Chen, H., Masarie, K.A., Krol, M.C., & Peters, W. (2017). The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: implementation and global carbon balance 2001–2015. *Geoscientific Model Development, 10,* 2785-2800

van der Molen, M.K., Dolman, A.J., Ciais, P., Eglin, T., Gobron, N., Law, B.E., Meir, P., Peters, W., Phillips, O.L., Reichstein, M., Chen, T., Dekker, S.C., Doubková, M., Friedl, M.A., Jung, M., van den Hurk, B.J.J.M., de Jeu, R.A.M., Kruijt, B., Ohta, T., Rebel, K.T., Plummer, S., Seneviratne, S.I., Sitch, S., Teuling, A.J., van der Werf, G.R., & Wang, G. (2011). Drought and ecosystem carbon cycling. *Agricultural and Forest Meteorology, 151,* 765-773

van der Velde, I.R., Miller, J.B., Schaefer, K., van der Werf, G.R., Krol, M.C., & Peters, W. (2014). Terrestrial cycling of 13CO2 by photosynthesis, respiration, and biomass burning in SiBCASA. *Biogeosciences, 11,* 6553-6571

van der Velde, M., Wriedt, G., & Bouraoui, F. (2010). Estimating irrigation use and effects on maize yield during the 2003 heatwave in France☆. *Agriculture, Ecosystems & Environment*, *135*, 90-97 van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Kasibhatla, P.S., & Arellano, A.F. (2006). Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics*, *6*, 3423-3441

van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Mu, M., Kasibhatla, P.S., Morton, D.C., DeFries, R.S., Jin, Y., & van Leeuwen, T.T. (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009). *Atmospheric Chemistry and Physics, 10*, 11707-11735

van Gorsel, E., Delpierre, N., Leuning, R., Black, A., Munger, J.W., Wofsy, S., Aubinet, M., Feigenwinter, C., Beringer, J., Bonal, D., Chen, B., Chen, J., Clement, R., Davis, K.J., Desai, A.R., Dragoni, D., Etzold, S., Grünwald, T., Gu, L., Heinesch, B., Hutyra, L.R., Jans, W.W.P., Kutsch, W., Law, B.E., Leclerc, M.Y., Mammarella, I., Montagnani, L., Noormets, A., Rebmann, C., & Wharton, S. (2009). Estimating nocturnal ecosystem respiration from the vertical turbulent flux and change in storage of CO2. *Agricultural and Forest Meteorology, 149*, 1919-1930

Van Niel, T.G., & McVicar, T.R. (2004). Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. *Computers and Electronics in Agriculture, 45*, 91-108

Verma, M., Friedl, M.A., Richardson, A.D., Kiely, G., Cescatti, A., Law, B.E., Wohlfahrt, G., Gielen, B., Roupsard, O., Moors, E.J., Toscano, P., Vaccari, F.P., Gianelle, D., Bohrer, G., Varlagin, A., Buchmann, N., van Gorsel, E., Montagnani, L., & Propastin, P. (2014). Remote sensing of annual terrestrial gross primary productivity from MODIS: an assessment using the FLUXNET La Thuile data set. *Biogeosciences, 11*, 2185-2200

Vocke, G., & Ali, M. (2013). U.S. wheat production practices, costs, and yields: variations across regions. *Department of Agriculture, Economic Research Service, E1B-116.*

von Buttlar, J., Zscheischler, J., Rammig, A., Sippel, S., Reichstein, M., Knohl, A., Jung, M., Menzer, O., Arain, M.A., Buchmann, N., Cescatti, A., Gianelle, D., Kieley, G., Law, B.E., Magliulo, V., Margolis, H., McCaughey, H., Merbold, L., Migliavacca, M., Montagnani, L., Oechel, W., Pavelka, M., Peichl, M., Rambal, S., Raschi, A., Scott, R.L., Vaccari, F.P., van Gorsel, E., Varlagin, A., Wohlfahrt, G., & Mahecha, M.D. (2017). Impacts of droughts and extreme temperature events on gross primary production and ecosystem respiration: a systematic assessment across ecosystems and climate zones. *Biogeosciences Discussions*, 1-39

Wagle, P., Xiao, X.M., & Suyker, A.E. (2015). Estimation and analysis of gross primary production of soybean under various management practices and drought conditions. *Isprs Journal of Photogrammetry and Remote Sensing, 99*, 70-83

Wagle, P., Zhang, Y.G., Jin, C., & Xiao, X.M. (2016). Comparison of solar-induced chlorophyll fluorescence, light-use efficiency, and process-based GPP models in maize. *Ecological Applications, 26*, 1211-1222

Wang, C., Hunt, E.R., Zhang, L., & Guo, H.D. (2013). Phenology-assisted classification of C3 and C4 grasses in the U.S. Great Plains and their climate dependency with MODIS time series. *Remote Sensing of Environment, 138*, 90-101

Wang, S., Azzari, G., & Lobell, D.B. (2019a). Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment, 222*, 303-317

Wang, Y.L., Xu, X.G., Huang, L.S., Yang, G.J., Fan, L.L., Wei, P.F., & Chen, G. (2019b). An Improved CASA Model for Estimating Winter Wheat Yield from Remote Sensing Images. *Remote Sensing, 11*

Wardlow, B.D., & Egbert, S.L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sensing of Environment, 112*, 1096-1116

Welp, L.R., Randerson, J.T., & Liu, H.P. (2007). The sensitivity of carbon fluxes to spring warming and summer drought depends on plant functional type in boreal forest ecosystems. *Agricultural and Forest Meteorology, 147*, 172-185

Wolf, S., Eugster, W., Ammann, C., Häni, M., Zielis, S., Hiller, R., Stieger, J., Imer, D., Merbold, L., & Buchmann, N. (2014). Corrigendum: Contrasting response of grassland versus forest carbon and water fluxes to spring drought in Switzerland (2013Environ. Res. Lett. 8 035007). *Environmental Research Letters, 9*, 089501

Wolf, S., Keenan, T.F., Fisher, J.B., Baldocchi, D.D., Desai, A.R., Richardson, A.D., Scott, R.L., Law, B.E., Litvak, M.E., Brunsell, N.A., Peters, W., & van der Laan-Luijkx, I.T. (2016). Warm spring reduced carbon cycle impact of the 2012 US summer drought. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 5880-5885

Wu, C., Gonsamo, A., Zhang, F., & Chen, J.M. (2014). The potential of the greenness and radiation (GR) model to interpret 8-day gross primary production of vegetation. *Isprs Journal of Photogrammetry and Remote Sensing, 88*, 69-79

Wu, G., Guan, K., Jiang, C., Peng, B., Kimm, H., Chen, M., Yang, X., Wang, S., Suyker, A.E., Bernacchi, C.J., Moore, C.E., Zeng, Y., Berry, J.A., & Cendrero-Mateo, M.P. (2020). Radiance-based NIRv as a proxy for GPP of corn and soybean. *Environmental Research Letters, 15,* 034009

Wu, X., Ju, W., Zhou, Y., He, M., Law, B., Black, T., Margolis, H., Cescatti, A., Gu, L., Montagnani, L., Noormets, A., Griffis, T., Pilegaard, K., Varlagin, A., Valentini, R., Blanken, P., Wang, S., Wang, H., Han, S., Yan, J., Li, Y., Zhou, B., & Liu, Y. (2015). Performance of Linear and Nonlinear Two-Leaf Light Use Efficiency Models at Different Temporal Scales. *Remote Sensing, 7*, 2238-2278

Wu, X.C., Xiao, X.M., Zhang, Y., He, W., Wolf, S., Chen, J.Q., He, M.Z., Gough, C.M., Qin, Y.W., Zhou, Y.L., Doughty, R., & Blanken, P.D. (2018). Spatiotemporal Consistency of Four Gross Primary Production Products and Solar-Induced Chlorophyll Fluorescence in Response to Climate Extremes Across CONUS in 2012. *Journal of Geophysical Research-Biogeosciences, 123*, 3140-3161

Xiao, J., Li, X., He, B., Arain, M.A., Beringer, J., Desai, A.R., Emmel, C., Hollinger, D.Y., Krasnova, A., Mammarella, I., Noe, S.M., Serrano Ortiz, P., Rey-Sanchez, C., Rocha, A.V., & Varlagin, A. (2019). Solar-induced chlorophyll fluorescence exhibits a universal relationship with gross primary productivity across a wide variety of biomes. *Glob Chang Biol*

Xiao, X.M., Hollinger, D., Aber, J., Goltz, M., Davidson, E.A., Zhang, Q.Y., & Moore, B. (2004a). Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sensing of Environment, 89*, 519-534

Xiao, X.M., Hollinger, D., Aber, J., Goltz, M., Davidson, E.A., Zhang, Q.Y., & Moore, B. (2004b). Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sensing of Environment, 89*, 519-534

Xiao, X.M., Zhang, Q.Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Berrien, M., & Ojima, D. (2004b). Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment, 91*, 256-270

Xin, F.F., Xiao, X.M., Zhao, B., Miyata, A., Baldocchi, D., Knox, S., Kang, M., Shim, K.M., Min, S., Chen, B.Q., Li, X.P., Wang, J., Dong, J.W., & Biradar, C. (2017). Modeling gross primary production of paddy rice cropland through analyses of data from CO2 eddy flux tower sites and MODIS images. *Remote Sensing of Environment, 190*, 42-55

Xin, Q., Broich, M., Suyker, A.E., Yu, L., & Gong, P. (2015). Multi-scale evaluation of light use efficiency in MODIS gross primary productivity for croplands in the Midwestern United States. *Agricultural and Forest Meteorology, 201*, 111-119

Xin, Q., Gong, P., Yu, C., Yu, L., Broich, M., Suyker, A., & Myneni, R. (2013). A Production Efficiency Model-Based Method for Satellite Estimates of Corn and Soybean Yields in the Midwestern US. *Remote Sensing, 5*, 5926-5943

Yan, H., Fu, Y.L., Xiao, X., He, Q.H., He, H.L., & Ediger, L. (2009). Modeling gross primary productivity for winter wheat–maize double cropping system using MODIS time series and CO2 eddy flux tower data. *Agriculture, Ecosystems & Environment, 129*, 391-400

Yang, C.H., Everitt, J.H., & Murden, D. (2011). Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture, 75*, 347-354

Yang, H.L., Yang, X., Zhang, Y.G., Heskel, M.A., Lu, X.L., Munger, J.W., Sun, S.C., & Tang, J.W. (2017). Chlorophyll fluorescence tracks seasonal variations of photosynthesis from leaf to canopy in a temperate forest. *Global Change Biology, 23*, 2874-2886

Yang, X., Tang, J., Mustard, J.F., Lee, J., Rossini, M., Joiner, J., Munger, J., Kornfeld, A., & Richardson, A. (2015). Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophysical Research Letters, 42*, 2977-2987

Yang, Y., Tao, B., Ren, W., Zourarakis, D.P., Masri, B.E., Sun, Z., & Tian, Q. (2019). An Improved Approach Considering Intraclass Variability for Mapping Winter Wheat Using Multitemporal MODIS EVI Images. *Remote Sensing, 11*, 1191

Yuan, W., Cai, W., Chen, Y., Liu, S., Dong, W., Zhang, H., Yu, G., Chen, Z., He, H., Guo, W., Liu, D., Liu, S., Xiang, W., Xie, Z., Zhao, Z., & Zhou, G. (2016a). Severe summer heatwave and drought strongly reduced carbon uptake in Southern China. *Sci Rep, 6*, 18813

Yuan, W., Cai, W., Xia, J., Chen, J., Liu, S., Dong, W., Merbold, L., Law, B., Arain, A., Beringer, J., Bernhofer, C., Black, A., Blanken, P.D., Cescatti, A., Chen, Y., Francois, L., Gianelle, D., Janssens, I.A., Jung, M., Kato, T., Kiely, G., Liu, D., Marcolla, B., Montagnani, L., Raschi, A., Roupsard, O., Varlagin, A., & Wohlfahrt, G. (2014). Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. *Agricultural and Forest Meteorology, 192-193,* 108-120

Yuan, W., Chen, Y., Xia, J., Dong, W., Magliulo, V., Moors, E., Olesen, J.E., & Zhang, H. (2016b). Estimating crop yield using a satellite-based light use efficiency model. *Ecological Indicators, 60,* 702-709

Yuan, W.P., Cai, W.W., Nguy-Robertson, A.L., Fang, H.J., Suyker, A.E., Chen, Y., Dong, W.J., Liu, S.G., & Zhang, H.C. (2015). Uncertainty in simulating gross primary production of cropland ecosystem from satellite-based models. *Agricultural and Forest Meteorology, 207,* 48-57

Yuan, W.P., Liu, S., Zhou, G.S., Zhou, G.Y., Tieszen, L.L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A.H., Goulden, M.L., Hollinger, D.Y., Hu, Y., Law, B.E., Stoy, P.C., Vesala, T., & Wofsy, S.C. (2007). Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agricultural and Forest Meteorology, 143,* 189-207

Zelitch, I. (1982). The close relationship between net photosynthesis and crop yield

Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., & Berry, J.A. (2019). A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence. *Remote Sensing of Environment, 232,* 111209

Zhang, Y., Song, C., Sun, G., Band, L.E., Noormets, A., & Zhang, Q. (2015). Understanding moisture stress on light use efficiency across terrestrial ecosystems based on global flux and remote-sensing data. *Journal of Geophysical Research: Biogeosciences, 120,* 2053-2066

Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., Joiner, J., Guanter, L., Zhang, Y., Zhang, G., Qin, Y., Wang, J., & Moore, B. (2016a). Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. *Remote Sensing of Environment, 183,* 154-169

Zhang, Y., Xiao, X., Wolf, S., Wu, J., Wu, X., & al., e. (2018). spatial-temporal convergence of maximum daily light-use efficiency based on radiation absorption by canopy chlorophyll. *Geophysical Research Letters*, *45*, 3508-3519

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., & Dong, J. (2017). A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. *Scientific Data, 4*, 170165

Zhang, Y., Xiao, X.M., Zhou, S., Ciais, P., McCarthy, H., & Luo, Y.Q. (2016b). Canopy and physiological controls of GPP during drought and heat wave. *Geophysical Research Letters, 43*, 3325-3333

Zhao, Y., Chen, X.P., Cui, Z.L., & Lobell, D.B. (2015). Using satellite remote sensing to understand maize yield gaps in the North China Plain. *Field Crops Research, 183*, 31-42

Zhao, Y., & Lobell, D.B. (2017). Assessing the heterogeneity and persistence of farmers' maize yield performance across the North China Plain. *Field Crops Research, 205*, 55-66

Zhong, L., Hu, L., Zhou, H., & Tao, X. (2019). Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US. *Remote Sensing of Environment*, *233*, 111411

Zhong, L.H., Gong, P., & Biging, G.S. (2014). Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment, 140*, 1-13

Zhu, Q., Zhuang, Q., Henze, D., Bowman, K., Chen, M., Liu, Y., He, Y., Matsueda, H., Machida, T., Sawa, Y., & Oechel, W. (2014). Constraining terrestrial ecosystem CO2 fluxes by integrating models of biogeochemistry and atmospheric transport and data of surface carbon fluxes and atmospheric CO2 concentrations. *Atmospheric Chemistry and Physics Discussions, 14*, 22587-22638

Zhuo, W., Huang, J.X., Li, L., Zhang, X.D., Ma, H.Y., Gao, X.R., Huang, H., Xu, B.D., & Xiao, X.M. (2019). Assimilating Soil Moisture Retrieved from Sentinel-1 and Sentinel-2 Data into WOFOST Model to Improve Winter Wheat Yield Estimation. *Remote Sensing, 11*

Zscheischler, J., Mahecha, M.D., von Buttlar, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J.T., Scholkopf, B., Seneviratne, S.I., Tomelleri, E., Zaehle, S., & Reichstein, M. (2014). A few extreme events dominate global interannual variability in gross primary production. *Environmental Research Letters, 9*, doi:10.1088/1748-9326/1089/1083/035001