9-2023

Estimation of Gross Primary Productivity of Rice in Arkansas Using the Vegetation Photosynthesis Model

Riasad Bin Mahbub
University of Arkansas-Fayetteville

Estimation of Gross Primary Productivity of Rice in Arkansas Using the Vegetation
Photosynthesis Model

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Environmental Dynamics

by

Riasad Bin Mahbub
North South University
Bachelor of Science in Environmental Science, 2018

August 2023
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____

Benjamin Runkle, Ph.D.
Thesis Director

_____   _____

Kusum Naithani, Ph.D.   Jason Tullis, Ph.D.
Committee Member   Committee Member

Abstract

An estimate of the gross primary productivity (GPP) of rice fields can be instrumental to understand their harvest yield and to fulfill an array of agricultural monitoring needs. One of the most common satellite-based models to estimate GPP is the vegetation photosynthesis model (VPM). In this study, we use the VPM model for rice cropland in Arkansas and validate our findings against 16 site-years in-situ data (eddy covariance (EC)). At the site scale, results validated against 16 site-years have shown that the VPM with site information (R^2 = 0.71, MAE = 2.90 g C m^{-2}day^{-1}, and RMSE = 4.04 g C m^{-2}day^{-1}) outperforms VPM based on spatial information (R^2 = 0.59, MAE = 4.9 g C m^{-2}day^{-1}, and RMSE = 3.48 g C m^{-2}day^{-1}). At the state scale, in the timeframe between 2008 to 2020, the mean photosynthetic carbon uptake of Arkansas rice fields was 1563.81± 129.09 g C m^{-2} season^{-1}. The spatial distribution of GPP has shown that rice fields located between 33.5° N and 34.5° N have higher GPP values (1840.40 ± 8.34 g C m^{-2} season^{-1}) than other rice regions of Arkansas. At the county-scale, GPP has shown an R^2 value of 0.07 against reported yield obtained from an agricultural survey. This GPP dataset will help to identify its underlying meteorological and soil factors, derive a relationship with yield, and investigate crop responses to a changing climate.

Table of Contents

1. Introduction

Knowledge about gross primary productivity (GPP), defined as the photosynthetic carbon uptake by terrestrial plants at the ecosystem scale, can provide a useful framework to find regions that need urgent crop management. Continuous spatiotemporal GPP data helps to predict crop yield (Xin et al., 2020), investigate crop response to changing climate (Flach et al., 2018; Y. Wang et al., 2020), identify the factors driving site carbon dynamics (Sun et al., 2018), and render a benchmark for validating prediction output from different environmental models (Schaefer et al., 2008). Additionally, this data helps to understand plant productivity across different land use management practices, improve agricultural statistical data, and develop better carbon budgets and crop stewardship. Given its crucial role in food security and the carbon cycle, it is important to have a clear understanding of GPP's spatiotemporal patterns, especially for the top rice-producing state of the United States (USA), Arkansas (Hardke et al., 2022). Arkansas contributes 47.5% of total rice production in the United States (Hardke et al., 2022). Rice has an economic impact of over $1.7 billion in Arkansas (Alhassan et al., 2019). There have been limited regional and local studies on the quantification of the GPP of rice in this region (Runkle et al., 2017; Reba et al., 2019a). Developing a quantification technique for the state's- rice GPP can be a useful way to overcome the constraints of data scarcity.

Geospatial maps can be a powerful tool to understand processes from large sets of data (Robinson et al., 2017). They can be used to make better climate-smart decisions which farmers can rely on. Looking into large spatial data, different analysis can be performed on the relationship between water use or rice irrigation system and gross primary productivity (Xie et al., 2021; Xin et al., 2020). A spatially organized dataset is a foundational need on which to research how

1

different agronomic practices viz., landform formation and management or cultivar use have any relationship with GPP.

To estimate GPP for a large areal coverage (i.e., at the regional or global scale), remotely sensed data have been used in various satellite-based model frameworks. Due to the simple structure and availability of data, one of the widely used methods for measuring GPP using satellite images is the vegetation photosynthesis model (VPM). The VPM is based upon the light use efficiency concept and has been widely applied to estimate the GPP at global and different ecosystem scales (Running et al., 2004; Xiao, 2004; Jin et al., 2013; Wang et al., 2016; Xin et al., 2017; Zhang et al., 2017). The model estimates 8-day mean GPP as the product of light use efficiency (LUE (ε_g)) and absorbed photosynthetically active radiation by the canopy ($APAR_{canopy}$). GPP (GPP_{EC}) data derived from eddy covariance (EC) stations have been used as reasonable validating ground-based measurements to validate and further calibrate the satellite-based model to a site-specific regional model.

Our overarching goal in this study is to estimate the trends and patterns of GPP of rice in Arkansas. To check the effectiveness of the model we are comparing the VPM_{site} (based on site information) and $VPM_{spatial}$ (based on satellite and spatial information) at both site and state scales. The results from this study will tackle the limitations of state-scale rice GPP data scarcity and aid us to better understand the carbon dynamics and driving mechanisms of GPP of rice paddies in Arkansas.

2. Materials and Methods

2.1 Workflow

Our work is divided into two scales: (i) site-scale estimation of GPP and (ii) state-scale estimation of GPP. In the first scale, we use the VPM to estimate GPP at site-scale and validate our results against GPP measured from EC towers. Root Mean Squared Error (RMSE), R^2, and Mean Absolute Error (MAE) were used to assess the two models' predictive performance. We use $VPM_{spatial}$ to derive an estimate for the state-scale analysis.

After validating our results against EC data, we estimate state-scale rice GPP using satellite images and satellite climatological data. The satellite image analysis and statistical analysis were carried out in Google Earth Engine (GEE) (Gorelick et al., 2017), and data visualizations were carried out in R studio platform (RStudio Team, 2020; Wickham, 2016). In this study, the temporal domain of all data was aggregated to 8-day bin, and the spatial domain of all data was aggregated to 500 m resolution, which was consistent with the MODIS image collection date.

Abbreviation	
DOP	Day of planting
DOH	Day of harvest
VPM_{site}	VPM based on site information
$VPM_{spatial}$	VPM based on satellite climatological data and modeled DOP and DOH
GPP_{EC}	GPP data from the eddy covariance stations

2.2 Study Sites

In this study, we brought together the GPP data collected over a selection of EC towers employed in rice fields in Arkansas. This dataset comprises 16 total seasons from among 10 rice fields. The sites are located in the Northeastern and central part of Arkansas (Fig. 1). Table 2 summarizes the planting and harvesting information of rice plants, data availability and

environmental characteristics including mean annual temperature, precipitation and soil classification. Rice is generally planted in April and harvested between the end of August or early September. For additional site information and a description of the instruments we refer to our previous research for US-HRA and US-HRC (Reavis et al., 2021; Runkle et al., 2019; Suvočarev et al., 2019); for US-BDA and US-BDC (Reba et al., 2019b) and for US-OF1, US-OF2, US-OF3, US-OF4, US-OF5 and US-OF6 sites (Massey et al., 2022).
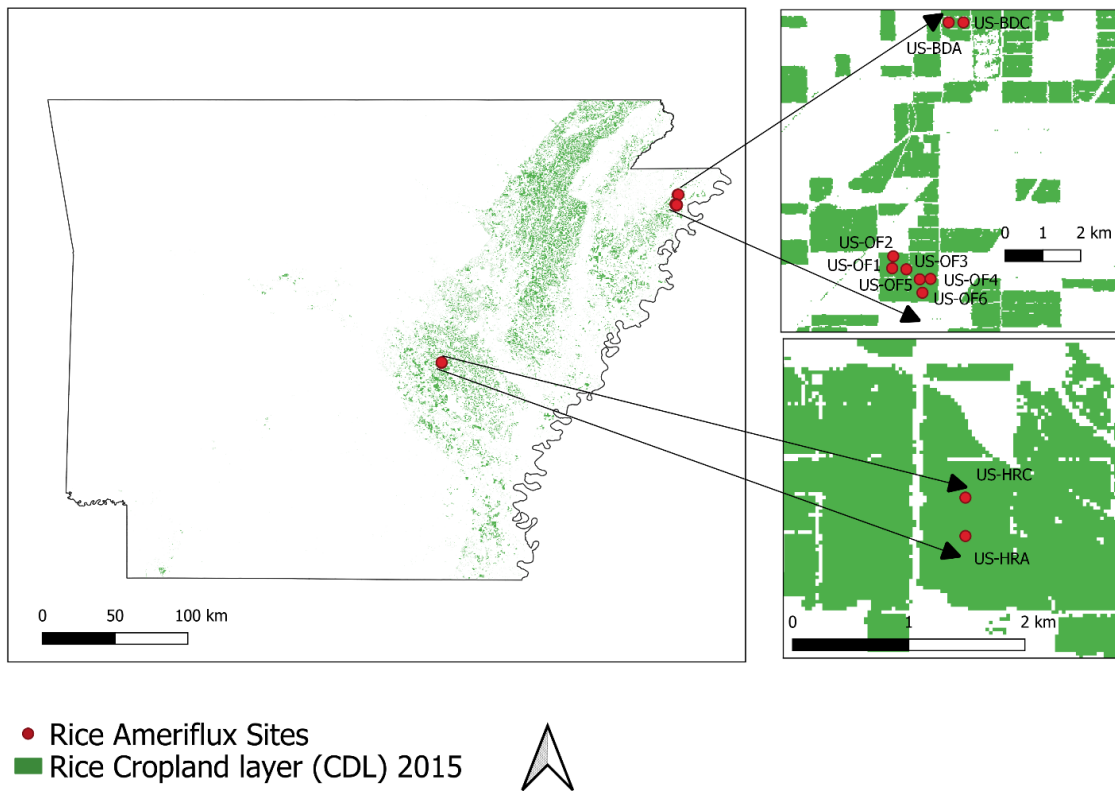

Figure 1: Regional map of the site locations colored by red dots. Eight sites are located in Northeastern Arkansas and two sites (US-HRA and US-HRC) are in central Arkansas. The base rice map in green is derived from the Cropland Data Layer (CDL) for the year 2015 (USDA National Agricultural Statistics Service, 2017)

Table 1: Description of the sites. MAT and MAP data were taken from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) dataset (Daly et al., 2008, 2015).

Site Id	Management	Planting date	Harvesting date	Lat (ºN)	Lon (ºE)	MAT (ºC)	MAP (mm)
US-HRA	DF	04-08-2015	08-19-2015	34.585	-91.751	17.6	1422.3
	AWD	04-25-2016	09-11-2016				
	DF	04-10-2017	08-26-2017				
US-HRC	AWD	04-07-2015	08-19-2015	34.588	-91.751	17.6	1422.3
	AWD	04-23-2016	09-13-2016				
	DF	04-09-2017	08-27-2017				
US-OF1	AWD	04-01-2017	09-06-2017	35.7370	-90.049	16.9	1325.9
US-OF2	MIRI	04-01-2017	09-06-2017	35.740	-90.048	16.9	1325.9
US-OF3	FIR	04-01-2017	09-18-2017	35.7372	-90.049	16.9	1325.9
US-OF4	FIR	04-09-2018	08-28-2018	35.734	-90.038	16.9	1775.3
US-OF5	MIRI	04-09-2018	08-28-2018	35.733	-90.040	16.9	1775.3
US-OF6	AWD	04-09-2018	08-28-2018	35.730	-90.040	16.9	1775.3
US-BdA	AWD	04-09-2015	08-27-2015	35.809	-90.327	16.7	1372.2
	DF	04-09-2016	08-27-2016				
US-BdC	DF	04-09-2015	08-27-2015	35.809	-90.028	16.7	1372.2
	DF	04-09-2016	08-27-2016				

DF = Delayed flooding, AWD = Alternate wetting and drying, MIRI = multiple-inlet rice irrigation, FIR = Furrow irrigated rice, MAT = Mean annual temperature, MAP = Mean annual precipitation. DOP and DOH are the planting and harvesting date (MM-DD-YYYY).

2.3 Site-scale data:

Site-scale information includes CO_2 flux and climatological data. Data collected from field instruments gives us the opportunity to test the accuracy and calibrate satellite-based models at the site scale and renders its applicability at the state scale.

2.3.1 CO_2 flux measurements

EC data were collected across the 16 field-seasons in the 10 sites. In the pre-processing stage, the high-frequency raw data were corrected based on transducer shadowing (Horst et al., 2015) and converted to half-hourly fluxes of CO_2 using EddyPro software (LI-COR, Inc., 2021). After processing the raw data on EddyPro software, the 30-min fluxes were corrected based on quality flags and wind direction. On the post-processing stage, the u^* threshold correction was applied and GPP and respiration (Reco) were partitioned from the net ecosystem exchange of CO_2 following Reichstein et al. (2005). For better understanding about the flux data correction and gap-filling method we refer to (Tajfar et al., 2023 under preparation).

2.3.2 Site climatological data

Air temperature data was recorded using humidity and temperature probe (HMP 60, Vaisala, Vantaa, Finland). Photosynthetically active radiation (PAR) data was recorded using LI-190 SL quantum sensor (LI-COR, Lincoln, NE, US). The half-hourly averaged temperature and PAR data from 1 Hz data were averaged to an 8-day period to be consistent with the MODIS 8-day composite. Site daily downward shortwave radiation (site DSWR) were measured using a 4-component net radiometer (CNR4 radiometer, Kipp and Zonen, Netherlands).

2.4 Satellite data

2.4.1 Satellite reflectance data

The land surface reflectance was taken at 8-day intervals at 500-meter resolution from the MODIS data (MOD09A1 Version 6). Poor-quality data were masked out for pixels affected by cloud, cloud shadow, and aerosol interference using the quality assurance (QA) layer. The bands (Near Infrared (NIR) (841-876 nm), Red (620-670 nm), Blue (459-479 nm) and Short-wave Infrared (SWIR) (1628-1652)) were scaled using a scaling factor 0.0001 and were used to estimate the vegetation indices, enhanced vegetation index (EVI) and land surface water index (LSWI) using equations (1) and (2) respectively. The missing values were gap-filled and smoothed using the traditional smoothing algorithm of the Savitzky-Golay filter (Savitzky & Golay, 1964; Chen et al., 2021).

The MODIS pixels for the 10 rice fields were selected based on the targeted flux footprint from each EC tower, given that some pixels did not cover the total footprint area or covered some additional area apart from the area of interest.

$$EVI = 2.5 \times \frac{NIR - Red}{NIR + 6 \ \times Red - 7.5 \ \times Blue + 1} \tag{1}$$

$$LSWI \ = \frac{NIR - SWIR}{NIR + SWIR} \tag{2}$$

2.4.2 Satellite Climatological Data

Climate data for deriving a state-scale estimate (temperature and DSWR) were derived from the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS) dataset (Kalnay et al., 1996). NCEP provides data at 6 hourly temporal intervals, and to keep the data consistent with MODIS data, it was aggregated to an 8-day bin period and resampled over the

7

MODIS 500m resolution. We use the average of temperature at 2m and maximum temperature at 2m to estimate the mean temperature (Zhang et al., 2017). We used a conversion factor of 0.9 to reduce the bias of the satellite DSWR data, determined from the scatter plot of satellite DSWR and site DSWR data (see section 2.3.2). The downward shortwave radiation (DSWR) (Wm^{-2}) was converted to photosynthetic active radiation (PAR) ($\mu mol\ m^{-2}\ s^{-1}$) using a conversion factor of 2.02 following (Mavi & Tupper, 2004).

2.5 Model Description

2.5.1 Vegetation Photosynthesis Model (VPM)

Building upon the concept of light use efficiency, the VPM framework provides GPP data as the product of LUE and $APAR_{canopy}$ (Eq. 3). LUE was estimated using equation 8 multiplying potential ε_0 with Ts and Ws. ($GPP_{VPM} = \varepsilon_g \times APAR_{canopy}$), where $APAR_{canopy}$ is the product of photosynthetically active radiation (PAR) and the fraction of absorbed photosynthetic active radiation ($fPAR_{canopy}$). $fPAR_{canopy}$ is measured as a linear function of EVI (Eq. 6). In this model, the ε_g is measured as the product of apparent quantum yield or maximum light use efficiency (ε_0), T_{scalar}, and W_{scalar} (Eq. 8), which are the downward-regulation scalars for the influence of temperature and water on light use efficiency, respectively. T_{scalar} is calculated using the parameter optimum temperature (T_{opt}), the variable mean temperature (T) and the constants of minimum temperature (T_{min}) (0 ℃) and maximum temperature (T_{max}) (48 ℃) (Xin et al., 2017) (Eq. 4). W_{scalar} is the measure of the effect of drought stress on plants is measured using LSWI index (Eq. 5). In the VPM, we used rice-specific constant values for ε_0 (0.05 mol CO_2 mol^{-1} PPFD) and T_{opt} (30 ℃) (Xin et al., 2017; Huang et al., 2021). The method to estimate state-scale PAR was described in section 2.3.

$$GPP_{VPM} = \mathcal{E}_g \times APAR_{canopy} \qquad (3)$$

$$T_s = \frac{(T - T_{min})(T - T_{max})}{(T - T_{min})(T - T_{max}) - \left(T - T_{opt}\right)^2} \qquad (4)$$

$$W_{scalar} = \frac{1 + LSWI}{1 + LSWI_{max}} \qquad (5)$$

$$fPAR_{chl} = (EVI - 0.1) \times 1.25 \qquad (6)$$

$$APAR_{canopy} = fPAR_{canopy} \times PAR \qquad (7)$$

$$\mathcal{E}_g = W_{scalar} \times T_{scalar} \times \mathcal{E}_o \qquad (8)$$

2.5.2 Estimating state-scale gross primary productivity of rice

At the state scale, we implement VPM using satellite reflectance data (section 2.3.2) and satellite climatological data (section 2.3.2).

We derived rice planted areas for the state of Arkansas for each year from 2008 to 2020 from the CropScape cropland data layer website (USDA National Agricultural Statistics Service Cropland Data Layer, 2022). We run the VPM within the geographic boundaries of the rice planted areas in Arkansas using the satellite data for each year in the GEE platform. Later we retrieve the images into the RStudio platform for further geospatial analysis and visualization.

To understand the relationship between GPP and yield we plot the yearly GPP data against yield of 26 counties where rice is grown. The GPP data which was calculated at 500m resolution was aggregated to county-scale resolution. The county yield data were obtained from USDA National Agricultural Statistics Service, Quick Stats website (USDA National Agricultural Statistics Service, 2017).

3. Results

3.1 Site-scale performance of the two models

3.1.1 Model

In all site-years, VPM explained 58-70% of the seasonal variation in the GPP derived from EC stations. Across 16 site-years data, the VPM_{site} model has an overall RMSE = 4.05 g C m^{-2} day^{-1}, R^2 = 0.7, MAE = 3.5 g C m^{-2}day^{-1}, making it the better of the two models. The $VPM_{spatial}$ model has performance of RMSE = 4.92 g C m^{-2} day^{-1}, R^2 = 0.58, MAE = 3.50 g C m^{-2} day^{-1}. In that evaluation, we find GPP estimated using the VPM_{site} outperforms GPP estimated using the $VPM_{spatial}$ model. At site-scale, across 16 site-years data, the parameters T_s and W_s had mean values of 0.98 (range: 0.86-1, standard deviation: 0.02) and 0.83 (range: 0.53-1, standard deviation: 0.14) respectively. We use the Taylor diagram to evaluate the performance of the two models (Fig. S1). The Taylor diagram shows that the VPM models are able to capture more than 70% of the correlation. The variation of the GPP data generated by VPM is less than the EC data. The standard deviation of GPP data from 16 site-years is 7 g C m^{-2} day^{-1} whereas the standard deviation of the VPM models is around 6 g C m^{-2} day^{-1}. The centered root means square errors for VPM_{site} and $VPM_{spatial}$ are 3.8 g C m^{-2} day^{-1} and 4.6 g C m^{-2} day^{-1} respectively. The residual plots of the models show that there is a higher degree of residuals between 50-100 days after planting (Fig. S2). This highlights that the models are not able to capture the magnitude at the mid-season period of growing season. The percent residual plot (Fig. S2 (B and D)) normalizes the residual by dividing the residual by the observed values. The percent residual plot shows a higher percentage of residual at the start of the season.
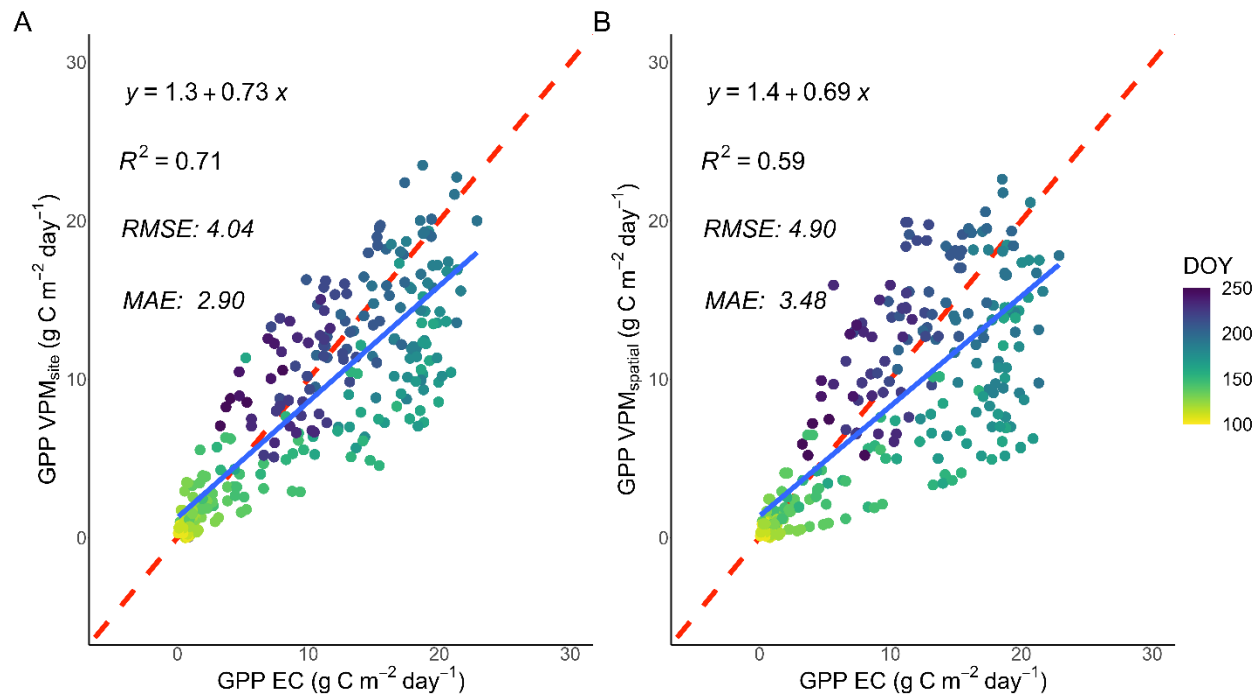
Figure 2: Comparison of two GPP estimates using two different data sources. EC data (GPP EC) on the x axis has been used as the observed value here and is used as a comparison to: (a) GPP from the VPM using site information and observed DOP and DOH and GPP EC, (b) GPP from the VPM using satellite information. The points are colored based on the day of the year. The red dashed line denotes the 1:1 relationship and the blue line denotes the fitted regression line.

3.1.2 Temporal pattern of GPP across the growing season.

The temporal nature of GPP from the two models and EC stations is broadly consistent (Fig. 3). However, there is a disagreement between the GPP_{EC} and the VPM models on the timing of the peak. A later peak and rise are observed in the case of the VPM_{site} and the $VPM_{spatial}$. Both VPM_{site} and $VPM_{spatial}$ seem to underestimate the magnitude of GPP, and they peak later than the observed values. At the start of the growing season, the measured GPP values from EC stations are closer to 0 g C $m^{-2}day^{-1}$, afterwards they reach the peak value of approx. 22 g C $m^{-2}day^{-1}$ around 80 days after planting and drop to 0 g C $m^{-2}day^{-1}$ after 150 days of planting.
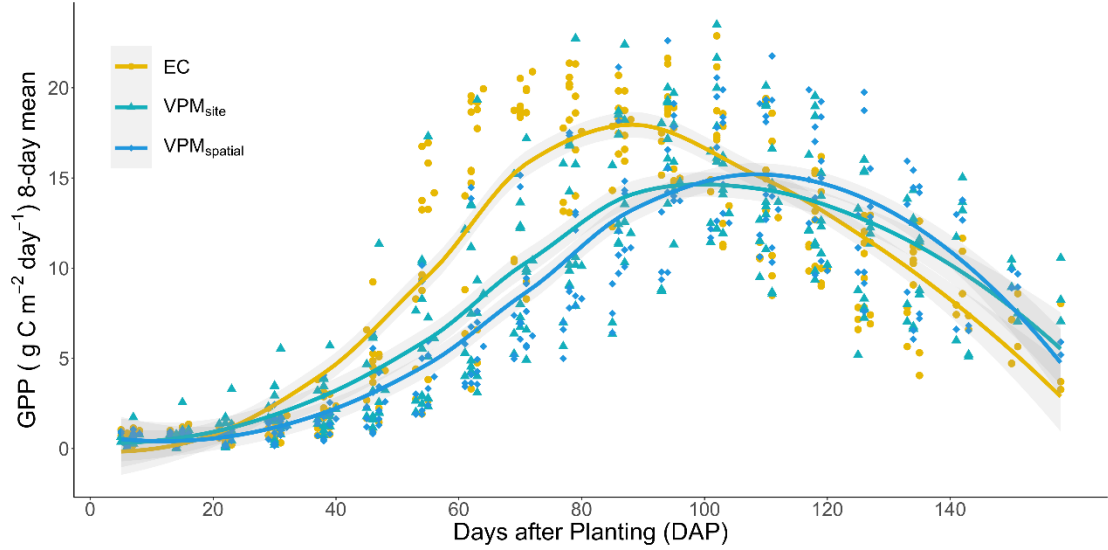
Figure 3: Temporal pattern of GPP derived from two models and EC tower for all 16 site-years. The color and shape of the dots represent the source or model of the data. The regression line of each model or data source has been derived using loess fitting and used to aid in identifying the seasonal pattern of GPP. The grey shaded area is the 95% confidence level interval for the predictions from the loess fitting.

3.2 GPP estimated at state scale

3.2.1 Spatial variability of GPP of rice layer in Arkansas

We implemented the VPM$_{spatial}$ model at the state-scale covering all the rice fields in Arkansas. Fig 4 shows the spatial distribution of 13 years averaged cumulative GPP across the whole rice layer in Arkansas. By calculating the latitudinal average of GPP using GAM regression, we see differences in GPP across the state. The GPP values are higher between 33.5° N and 34.5° N (1840.40. ± 8.34 g C m^{-2} season^{-1}) latitude. In terms of GPP, these regions can be attributed as the most productive rice growing regions in Arkansas. On the contrary, lower GPP values (1433.61 ± 4.07 g C m^{-2} season^{-1}) were observed around 36° N. From 36° N, the GPP value increases with the decrease in latitude however the value decreases after the peak around 34° N. We observe low GPP values, less than 800 g C m^{-2} season^{-1}, around the riverine regions of Arkansas. These values

may derive from mixed pixel effects where riverine regions share the same pixel with rice or are mislabeled entirely.
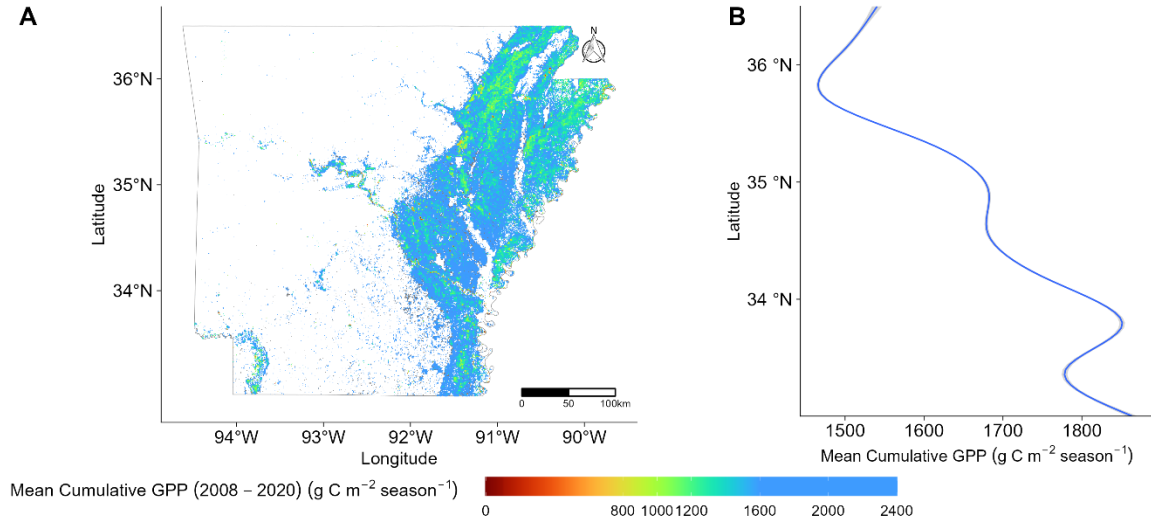

Figure 4: Spatial profile of mean annual GPP across the state from the VPM$_{spatial}$ model. (A) Map of the mean cumulative GPP (2008-2020) of the rice layer in Arkansas. (B) Mean latitudinal distribution of mean cumulative GPP (2008-2020), created with a GAM regression to demonstrate the spatial dynamics of the mean cumulative GPP, where the shaded line defines the 95% confidence level interval for the predictions from the loess fitting.

3.2.2 Annual variability of GPP of rice layer in Arkansas

We averaged all the pixels of the rice layer for each year (Figure 5**:** Inter-seasonal variability of mean cumulative GPP in the rice growing season in Arkansas.). We found that while there has been an apparent increase in the average annual GPP of rice in Arkansas, the increase is not statistically significant ($p > 0.01$). Moreover, the Mann Kendall test failed to confirm the significance of time series change of GPP ($p = 0.28$). In the timeframe between 2008 to 2020, the mean photosynthetic carbon uptake of Arkansas rice fields was 1563.81 ± 129.09 g Cm^{-2} season^{-1}. The lowest GPP was observed in the year 2011 (1317.63 ± 391.25 g C m^{-2} season^{-1}) followed by 2015 (1375.08 ± 366.06 g C m^{-2} season^{-1}), and 2010 (1454.72 ± 403.55 g C m^{-2} season^{-1}). On the

13

other hand, years 2012 (1776.09 $\pm$ 377.32 g C m^{-2} season^{-1}), 2017 (1712.35 $\pm$ 339.14 g C m^{-2} season^{-1}), and 2020 (1654.91 $\pm$ 384.84 g C m^{-2} season^{-1}) showed higher magnitude of GPP.
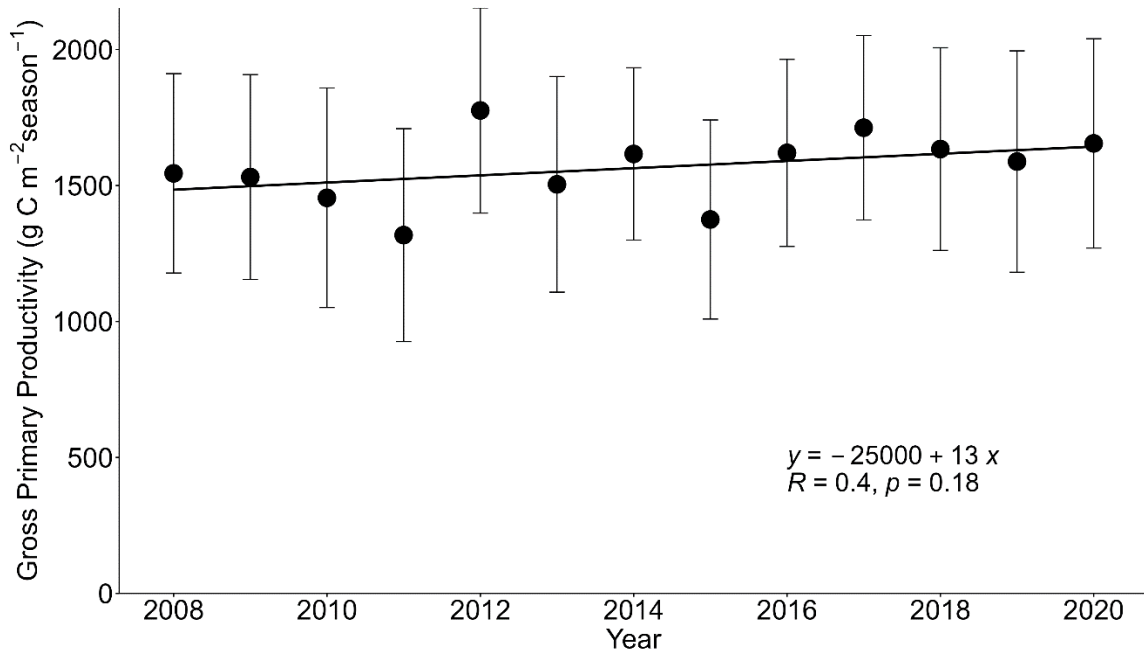

Figure 5: Inter-seasonal variability of mean cumulative GPP in the rice growing season in Arkansas. Error bars reflect standard deviation of the mean cumulative GPP across the spatial extent of rice croplands.

3.3 GPP Harvest Relationship

In order to test the relationship between GPP and rice yield we plot the mean cumulative GPP against the yield at the county-scale from the year 2008 to 2020 (Figure 6: County-scale relationship between mean cumulative GPP and reported yield. Each scatter represents a county-year.). Overall, across 13 years of data, the GPP-yield relationship shows a positive and significant correlation based on the slope (0.092), RMSE (61 g m^{-2} season^{-1}), and R^2 value (0.07) (p <0.001). The yield and GPP values were higher in the later years than the yield and GPP values from earlier years. The yield values ranged from 600-900 g m^{-2} season^{-1}.
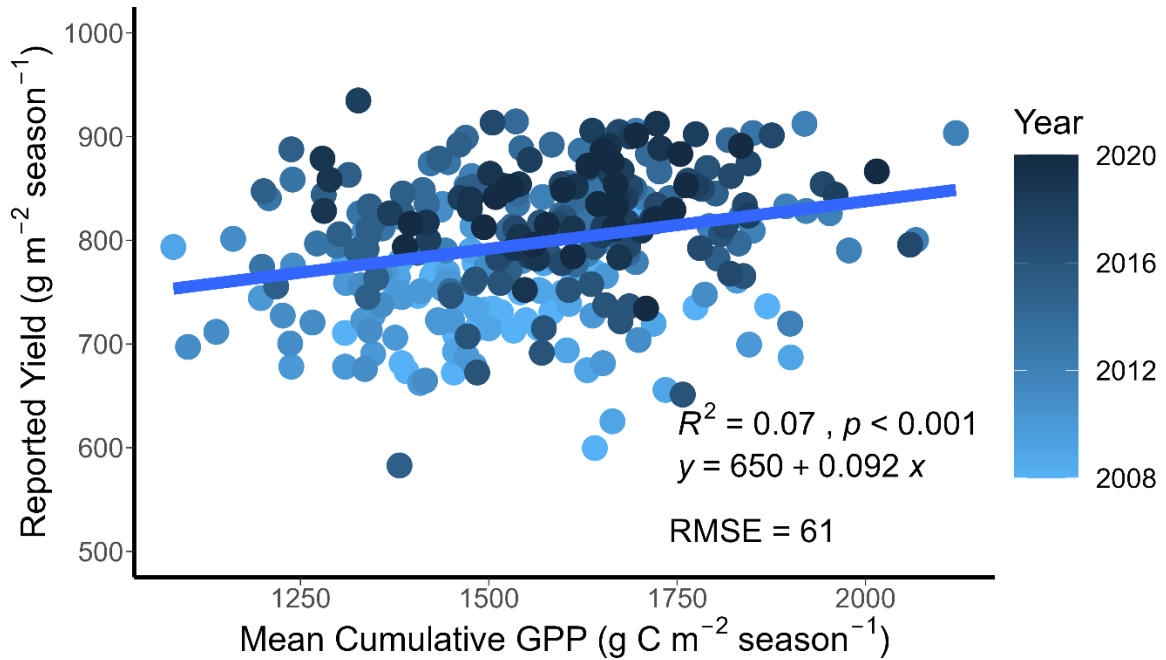
14

Figure 6: County-scale relationship between mean cumulative GPP and reported yield. Each scatter represents a county-year.

4. Discussion

4.1 VPM model based on site information performs better than satellite information

Statistical metrics show the superiority of VPM_{site} over $VPM_{spatial}$. VPM_{site} has higher R^2, lower RMSE and MAE than the $VPM_{spatial}$. Moreover, VPM_{site} better captures the seasonal variation. These results indicate that site information performs better than satellite information. The models capture more than 70% of the correlation. These models can show guidance on gap filling missing GPP values at annual scale.

4.2 Uncertainties of the model and future work and improvements

One of the sources of error working with satellite data, especially with MODIS's 500-meter resolution, is the mixed pixel effect (Bandaru et al., 2013). Rice field pixels are influenced by the surrounding land cover. The mixed-pixel effect can be reduced through selecting a homogenous

15

land cover (Chen et al., 2018). The mixed pixel effect might be less in the case of rice in Arkansas since most of the rice fields are concentrated in eastern regions and adjacent to each other. However, surrounding dense vegetation and riverine regions added some negligible errors in the state-scale estimation.

Several sources of error may have affected the results of this study. The spatial GPP data was derived from multiple spatial data sources. Satellite-based climatological datasets are mostly derived from weather station datasets and may not capture fine-scale variation for some variables, such as shortwave radiation (Cui et al., 2012). Eddy covariance data is a widely used technique, but there may be some sources of error in the data generated from the sensors. There may be some degree of noise due to the precision of calibration in the sensors (Baldocchi, 2003). Some of the main sources of EC system are (i) lack of adequate fetch, (ii) flow distortion due to other sensors and (iii) buoyancy effect (Chi et al., 2016; Waldo et al., 2016). The first error has minimum effect in the zero grade rice field if the direction of wind covers the fetch. To avoid the second error all other meteorological sensors were set parallel to the EC system rather than in the area of fetch. Lastly, the WPL correction was applied through EddyPro software to reduce the buoyancy effect. However, there can be uncertainties from gap filling the missing and filtered out low quality observations. In future this issue can be addressed through understanding the level of uncertainties using Monte Carlo approach (Richardson & Hollinger, 2007). Lastly, the residual plots indicate that there is a need to consider phenological-based parameters in the model to improve accuracy.

4.3 State-scale GPP and potential application of this dataset

Having explained the better accuracy of the VPM at site-scales, we use the VPM (VPM$_{spatial}$) to derive a state-scale estimation of rice plants in Arkansas. The spatial pattern of GPP shows that rice grown in regions between 33.5° N and 34.5° N latitude have higher magnitude than

other regions. The higher magnitude of GPP can be attributed to factors like elevated precipitation, medium temperature and radiation ranging from 500-600 Wm^{-2} (Anav et al., 2015). The lower magnitude of GPP in the years 2010 and 2011 may be driven by high nighttime temperatures. Arkansas rice fields have observed higher nighttime temperatures and reduced yield in the years 2010 and 2011 compared to other years (Nalley et al., 2016). Higher nighttime temperature has been linked with lower rice yield (Peng et al., 2004). Higher nighttime temperature causes higher respiration rate and weak membrane stability which ultimately leads to lower yield (Mohammed & Tarpley, 2009). A similar pattern has been seen in our study where the cumulative GPP in the year 2010 and 2011 were lower, suggesting that the processes that control yield also control GPP and GPP can be a powerful tool to predict and understand yield dynamics. Studies have also found that higher nighttime temperature caused lower quality of rice grain and increased the level of chalk in the year 2010 (Lanning et al., 2011).

GPP has a positive and significant relationship with yield and can explain 7% of variation of yield at county scale, but there are some years (2009, 2018, and 2019) that showed weak positive relationships. This different relationship across different years may suggest that rice plants have shown either different autotrophic respiration leading to different carbon use efficiency levels or have different harvest index values. GPP alone cannot explain these factors on a county-scale and a full explanation might require other agronomic and climatological variables. Overall, from 2008 to 2020 on a state-wide basis, there has been an increase of 14% and 11% in both mean cumulative GPP and yield respectively. This increase in yield in US rice production has been mostly driven by precision land-leveling and hybrid rice varieties, and regions like Arkansas have achieved higher yield magnitude by adopting hybrid and herbicide-resistant varieties of rice (Espe et al., 2016).

Our study can benefit the implementation of nature-based climate solutions (NbCS) and greenhouse gas budget accounting. First, the method of developing $VPM_{spatial}$ can be an example of a practical deployment of NbCS which is the upscale component (deriving flux measurements across a larger spatial scale) (Hemes et al., 2021). The GPP dataset will ultimately help to achieve one of the major goals of NbCS, i.e., to derive baseline GPP datasets, which focuses on harnessing the power of photosynthesis to sequester the carbon back into the soil (Runkle, 2022). This baseline dataset can provide answers to the questions in identifying the specific agronomic practices (fertilization or irrigation techniques) or climatological factors that drive the spatial variation of GPP. In terms of greenhouse budgeting, this improved dataset will enable us to derive an estimate of net biome productivity (NBP) which can be derived through subtracting autotrophic and heterotrophic respiration, crop yield, and other terms such as dissolved carbon in water, fire-based emissions, and carbon introduced through fertilizer applications from the GPP. Through incorporating net CH_4 and NO_2 emissions this data can contribute to a holistic assessment of the greenhouse gas budget of Arkansas's rice production region. Policy makers can then make better decisions to support optimal irrigation and fertilization techniques for sequestering long term carbon in the rice ecosystem.

Additionally, the state-scale GPP dataset can be used with other models to estimate the state-scale methane budgets of rice fields in Arkansas and gain better understanding about the biogeochemistry of the rice system at a larger spatial scale. GPP is one of the main variables in predicting methane because 50% of the GPP in the growing season gets assimilated to the soil labile carbon pool which further used as a substrate by the methanogen to produce methane (Pendall et al., 2004; Oikawa et al., 2017). Thus, this improved dataset can act as a baseline to derive an improved estimate of the state's methane budget.

5. Conclusion

We aimed to derive a state-scale GPP estimation of rice in Arkansas. To develop this dataset, we use VPM model at site and state scle. At the site scale, we found that the VPM model with site information (R^2 = 0.71, MAE = 2.90 g C m^{-2} day^{-1}, and RMSE = 4.04 g C m^{-2} day^{-1}) outperformed the VPM model without site information (R^2 = 0.59, MAE = 4.9 g C m^{-2}day^{-1}, and RMSE = 3.48 g C m^{-2}day^{-1}). At the state scale, we found that the mean photosynthetic carbon uptake of Arkansas rice fields was 1563.81± 129.09 g C m^{-2}season^{-1}. The spatial distribution of GPP showed that rice fields located between 33.5° N and 34.5° N have higher GPP values (1840.40. ± 8.34 g C m^{-2} season^{-1}) than other rice regions of Arkansas. At the county-scale, we found that GPP has an R^2 value of 0.07 against reported yield obtained from an agricultural survey. This dataset aids us to more understand greenhouse gas budgets to make better climate smart decisions at larger spatial scale. However, like most other models there is an opportunity for improvement and the accuracy of this model can be improved through incorporating phenological information.

6. Acknowledgements

7. References

Alhassan, M., Lawrence, C. B., Richardson, S., & Pindilli, E. J. (2019). *The Mississippi Alluvial Plain aquifers—An engine for economic activity* (No. 2327–6932). US Geological Survey.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., & Zhao, M. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review: GPP Spatiotemporal Patterns. *Reviews of Geophysics*, *53*(3), 785–818. https://doi.org/10.1002/2015RG000483

Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future: CARBON BALANCE and EDDY COVARIANCE. *Global Change Biology*, *9*(4), 479–492. https://doi.org/10.1046/j.1365-2486.2003.00629.x

Bandaru, V., West, T. O., Ricciuto, D. M., & César Izaurralde, R. (2013). Estimating crop net primary production using national inventory data and MODIS-derived parameters. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 61–71. https://doi.org/10.1016/j.isprsjprs.2013.03.005

Chen, X., Wang, D., Chen, J., Wang, C., & Shen, M. (2018). The mixed pixel effect in land surface phenology: A simulation study. *Remote Sensing of Environment*, *211*, 338–344. https://doi.org/10.1016/j.rse.2018.04.030

Chen, Y., Cao, R., Chen, J., Liu, L., & Matsushita, B. (2021). A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter. *ISPRS Journal of Photogrammetry and Remote Sensing*, *180*, 174–190. https://doi.org/10.1016/j.isprsjprs.2021.08.015

Chi, J., Waldo, S., Pressley, S., O'Keeffe, P., Huggins, D., Stöckle, C., Pan, W. L., Brooks, E., & Lamb, B. (2016). Assessing carbon and water dynamics of no-till and conventional tillage cropping systems in the inland Pacific Northwest US using the eddy covariance method. *Agricultural and Forest Meteorology*, *218–219*, 37–49. https://doi.org/10.1016/j.agrformet.2015.11.019

Cui, B., Toth, Z., Zhu, Y., & Hou, D. (2012). Bias Correction for Global Ensemble Forecast. *Weather and Forecasting*, *27*(2), 396–410. https://doi.org/10.1175/WAF-D-11-00011.1

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Daly, C., Smith, J. I., & Olson, K. V. (2015). Mapping Atmospheric Moisture Climatologies across the Conterminous United States. *PLOS ONE*, *10*(10), e0141140. https://doi.org/10.1371/journal.pone.0141140

Espe, M. B., Cassman, K. G., Yang, H., Guilpart, N., Grassini, P., Van Wart, J., Anders, M., Beighley, D., Harrell, D., Linscombe, S., McKenzie, K., Mutters, R., Wilson, L. T., & Linquist, B. A. (2016). Yield gap analysis of US rice production systems shows

opportunities for improvement. *Field Crops Research*, *196*, 276–283. https://doi.org/10.1016/j.fcr.2016.07.011

Flach, M., Sippel, S., Gans, F., Bastos, A., Brenning, A., Reichstein, M., & Mahecha, M. D. (2018). Contrasting biosphere responses to hydrometeorological extremes: Revisiting the 2010 western Russian heatwave. *Biogeosciences*, *15*(20), 6067–6085. https://doi.org/10.5194/bg-15-6067-2018

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hardke, J., Sha, X., & Bateman, N. (2022). *B.R. Wells Arkansas Rice Research Studies 2021*.

Hemes, K. S., Runkle, B. R. K., Novick, K. A., Baldocchi, D. D., & Field, C. B. (2021). An Ecosystem-Scale Flux Measurement Strategy to Assess Natural Climate Solutions. *Environmental Science & Technology*, *55*(6), 3494–3504. https://doi.org/10.1021/acs.est.0c06421

Horst, T. W., Semmer, S. R., & Maclean, G. (2015). Correction of a Non-orthogonal, Three-Component Sonic Anemometer for Flow Distortion by Transducer Shadowing. *Boundary-Layer Meteorology*, *155*(3), 371–395. https://doi.org/10.1007/s10546-015-0010-3

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. https://doi.org/10.1016/j.rse.2013.03.033

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., & Woollen, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472.

Lanning, S. B., Siebenmorgen, T. J., Counce, P. A., Ambardekar, A. A., & Mauromoustakos, A. (2011). Extreme nighttime air temperatures in 2010 impact rice chalkiness and milling quality. *Field Crops Research*, *124*(1), 132–136. https://doi.org/10.1016/j.fcr.2011.06.012

LI-COR, Inc. (2021). *EddyPro® version 7.0 Help and User's Guide*.

Massey, J. H., Reba, M. L., Adviento-Borbe, M. A., Chiu, Y. L., & Payne, G. K. (2022). Direct comparisons of four irrigation systems on a commercial rice farm: Irrigation water use efficiencies and water dynamics. *Agricultural Water Management*, *266*, 107606. https://doi.org/10.1016/j.agwat.2022.107606

Mavi, H. S., & Tupper, G. J. (2004). *Agrometeorology: Principles and applications of climate studies in agriculture*. CRC Press.

Mohammed, A.-R., & Tarpley, L. (2009). Impact of High Nighttime Temperature on Respiration, Membrane Stability, Antioxidant Capacity, and Yield of Rice Plants. *Crop Science*, *49*(1), 313–322. https://doi.org/10.2135/cropsci2008.03.0161

Nalley, L., Tack, J., Barkley, A., Jagadish, K., & Brye, K. (2016). Quantifying the Agronomic and Economic Performance of Hybrid and Conventional Rice Varieties. *Agronomy Journal*, *108*(4), 1514–1523. https://doi.org/10.2134/agronj2015.0526

Oikawa, P. Y., Jenerette, G. D., Knox, S. H., Sturtevant, C., Verfaillie, J., Dronova, I., Poindexter, C. M., Eichelmann, E., & Baldocchi, D. D. (2017). Evaluation of a hierarchy of models reveals importance of substrate limitation for predicting carbon dioxide and methane exchange in restored wetlands: Model for Wetland Greenhouse Gas Fluxes. *Journal of Geophysical Research: Biogeosciences*, *122*(1), 145–167. https://doi.org/10.1002/2016JG003438

Pendall, E., Mosier, A. R., & Morgan, J. A. (2004). Rhizodeposition stimulated by elevated CO_2 in a semiarid grassland. *New Phytologist*, *162*(2), 447–458. https://doi.org/10.1111/j.1469-8137.2004.01054.x

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., & Cassman, K. G. (2004). Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences*, *101*(27), 9971–9975. https://doi.org/10.1073/pnas.0403720101

Reavis, C. W., Suvočarev, K., Reba, M. L., & Runkle, B. R. K. (2021). Impacts of alternate wetting and drying and delayed flood rice irrigation on growing season evapotranspiration. *Journal of Hydrology*, *596*, 126080. https://doi.org/10.1016/j.jhydrol.2021.126080

Reba, M. L., Fong, B. N., & Rijal, I. (2019a). Fallow season CO2 and CH4 fluxes from US mid-south rice-waterfowl habitats. *Agricultural and Forest Meteorology*, *279*, 107709. https://doi.org/10.1016/j.agrformet.2019.107709

Reba, M. L., Fong, B. N., & Rijal, I. (2019b). Fallow season CO2 and CH4 fluxes from US mid-south rice-waterfowl habitats. *Agricultural and Forest Meteorology*, *279*, 107709. https://doi.org/10.1016/j.agrformet.2019.107709

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., … Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Global Change Biology*, *11*(9), 1424–1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x

Richardson, A. D., & Hollinger, D. Y. (2007). A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO2 flux record. *Agricultural and Forest Meteorology*, *147*(3–4), 199–208. https://doi.org/10.1016/j.agrformet.2007.06.004

Robinson, A. C., Demšar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., Kraak, M.-J., Camboim, S. P., & Sluter, C. R. (2017). Geospatial big data and cartography: Research challenges and opportunities for making maps that matter. *International Journal of Cartography*, *3*(sup1), 32–60. https://doi.org/10.1080/23729333.2016.1278151

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA. URL http://www.rstudio.com/.

Runkle, B. R. K. (2022). Review: Biological engineering for nature-based climate solutions. *Journal of Biological Engineering*, *16*(1), 7. https://doi.org/10.1186/s13036-022-00287-8

Runkle, B. R. K., Rigby, J. R., Reba, M. L., Anapalli, S. S., Bhattacharjee, J., Krauss, K. W., Liang, L., Locke, M. A., Novick, K. A., Sui, R., Suvočarev, K., & White, P. M. (2017). Delta-Flux: An Eddy Covariance Network for a Climate-Smart Lower Mississippi Basin. *Agricultural & Environmental Letters*, *2*(1). https://doi.org/10.2134/ael2017.01.0003

Runkle, B. R. K., Suvočarev, K., Reba, M. L., Reavis, C. W., Smith, S. F., Chiu, Y.-L., & Fong, B. (2019). Methane Emission Reductions from the Alternate Wetting and Drying of Rice Fields Detected Using the Eddy Covariance Method. *Environmental Science & Technology*, *53*(2), 671–681. https://doi.org/10.1021/acs.est.8b05535

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *BioScience*, *54*(6), 547. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Savitzky, Abraham., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

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: Biogeosciences*, *113*(G3).

Sun, Z., Wang, X., Yamamoto, H., Tani, H., Zhong, G., Yin, S., & Guo, E. (2018). Spatial pattern of GPP variations in terrestrial ecosystems and its drivers: Climatic factors, CO2 concentration and land-cover change, 1982–2015. *Ecological Informatics*, *46*, 156–165. https://doi.org/10.1016/j.ecoinf.2018.06.006

Suvočarev, K., Castellví, F., Reba, M. L., & Runkle, B. R. K. (2019). Surface renewal measurements of H, λE and CO2 fluxes over two different agricultural systems. *Agricultural and Forest Meteorology*, *279*, 107763. https://doi.org/10.1016/j.agrformet.2019.107763

USDA National Agricultural Statistics Service. (2017). *NASS - Quick Stats. USDA National Agricultural Statistics Service.* https://data.nal.usda.gov/dataset/nass-quick-stats

USDA National Agricultural Statistics Service Cropland Data Layer. (2022). *Published crop-specific data layer [Online].*

Waldo, S., Chi, J., Pressley, S. N., O'Keeffe, P., Pan, W. L., Brooks, E. S., Huggins, D. R., Stöckle, C. O., & Lamb, B. K. (2016). Assessing carbon dynamics at high and low rainfall agricultural sites in the inland Pacific Northwest US using the eddy covariance method. *Agricultural and Forest Meteorology*, *218–219*, 25–36. https://doi.org/10.1016/j.agrformet.2015.11.018

Wang, J., Xiao, X., Wagle, P., Ma, S., Baldocchi, D., Carrara, A., Zhang, Y., Dong, J., & Qin, Y. (2016). Canopy and climate controls of gross primary production of Mediterranean-type deciduous and evergreen oak savannas. *Agricultural and Forest Meteorology*, *226–227*, 132–147. https://doi.org/10.1016/j.agrformet.2016.05.020

Wang, Y., Sperry, J. S., Anderegg, W. R. L., Venturas, M. D., & Trugman, A. T. (2020). A theoretical and empirical assessment of stomatal optimization modeling. *New Phytologist*, *227*(2), 311–325. https://doi.org/10.1111/nph.16572

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Xiao, X. (2004). Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, *91*(2), 256–270. https://doi.org/10.1016/j.rse.2004.03.010

Xie, Y., Gibbs, H. K., & Lark, T. J. (2021). Landsat-based Irrigation Dataset (LANID): 30 m resolution maps of irrigation distribution, frequency, and change for the US, 1997–2017. *Earth System Science Data*, *13*(12), 5689–5710. https://doi.org/10.5194/essd-13-5689-2021

Xin, F., Xiao, X., Dong, J., Zhang, G., Zhang, Y., Wu, X., Li, X., Zou, Z., Ma, J., Du, G., Doughty, R. B., Zhao, B., & Li, B. (2020). Large increases of paddy rice area, gross primary production, and grain production in Northeast China during 2000–2017. *Science of The Total Environment*, *711*, 135183. https://doi.org/10.1016/j.scitotenv.2019.135183

Xin, F., Xiao, X., Zhao, B., Miyata, A., Baldocchi, D., Knox, S., Kang, M., Shim, K., Min, S., Chen, B., Li, X., Wang, J., Dong, J., & 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. https://doi.org/10.1016/j.rse.2016.11.025

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*(1), 170165. https://doi.org/10.1038/sdata.2017.165
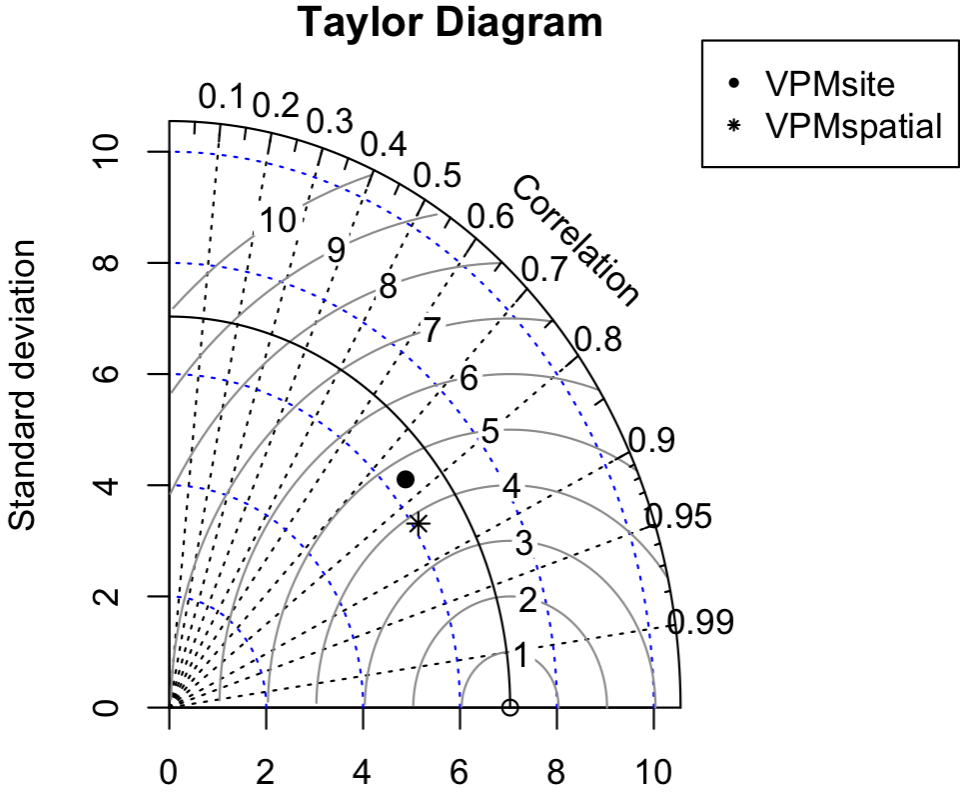
8. Appendix

Fig. S1: Tailor diagram of the performances of the models VPM$_{site}$ (highlighted as bullet point) and VPM$_{spatial.}$(highlighted as asterisk). The contour lines show the centered root mean squared errors.
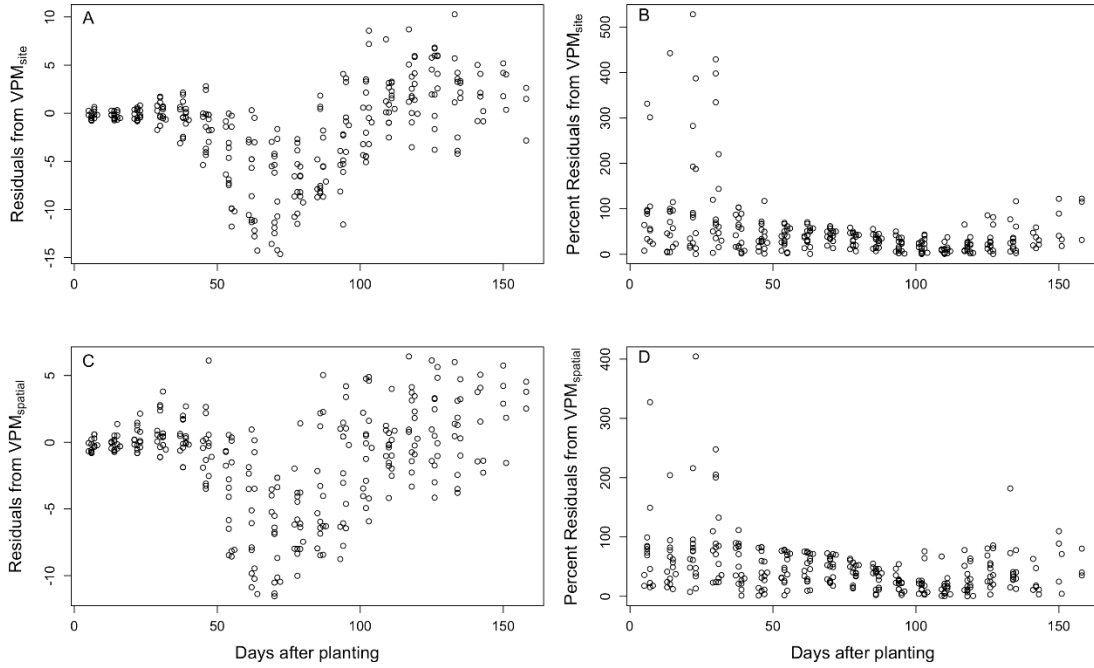
Fig. S2: Residual plots of the models VPM$_{site}$ (A) and VPM$_{spatial}$ (C). Percent residual plots of the models VPM$_{site}$ (B) and VPM$_{spatial}$ (D) (absolute values were taken to calculate the percent residual plots).