***Fine resolution remote sensing spectra improves estimates of gross primary production of croplands***

**Authors**

Gabriela Shirkey[a,b*], Ranjeet John[c,d], Jiquan Chen[a,b,e], Kyla Dahlin[a], Michael Abraha[b,e,f], Pietro Sciusco[a,b], Cheyenne Lei[a,b], David E. Reed[g]

[a] Department of Geography, Environment & Spatial Sciences, Michigan State University, East Lansing, MI, 48823, USA

[b] Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI, 48823, USA

[c] Department of Biology, University of South Dakota, Vermillion, SD, 57069, USA

[d] Department of Sustainability and Environment, University of South Dakota, Vermillion, SD, 57069, USA

[e] Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, 48824, USA

[f] W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI, 49060, USA

[g] Environmental Science, University of Science and Arts of Oklahoma, Chickasha, OK, 73018, USA

*Corresponding author: shirkeyg@msu.edu

**Abstract**

Gross primary production (GPP) is a fundamental measure of the terrestrial carbon cycle critical to our understanding of ecosystem function under the changing climate and land use. Remote sensing enables access to continuous spatial coverage, but remains challenged in heterogeneous croplands. Coarse resolution products, like MOD17A (500 m), may aggregate fragmented land cover types commonly found in heavily managed landscapes and misrepresent their respective contribution to carbon production. Consequently, this study demonstrates the capability of fine-resolution imagery (20-30 m) and available red-edge vegetation indices to characterize GPP across seven Midwest cropping systems. Four sites were established on a 22-year-old USDA Conservation Reserve Program (CRP); and the other three on land conventionally farmed with corn-soybean-wheat rotation (AGR). We compare *in situ* GPP estimates from eddy-covariance towers with ten satellite models: eight variants of the vegetation photosynthesis models (VPM), of which five include a red-edge vegetation index, as well as conventional products Landsat CONUS GPP (30 m) and MOD17A2H V6 (500 m). Daily and cumulative fine-resolution imagery integrated within VPM generally agreed with tower-based GPP in heterogeneous landscapes more than those from MODIS 500 m VPM or conventional GPP products from MOD17AH V6 or Landsat 8 CONUS. Replacing EVI2 with red-edge indices NDRE2, NDRE1, and MTCI in Sentinel 2 VPMs notably improved explanation of variance and estimation of cumulative GPP. While existing methods using MODIS- and Landsat-derived GPP are important baselines for regional and global studies, future research may benefit from the higher spatial, temporal, and radiometric resolution.

**Keywords**: LUE; GPP; Carbon; Croplands; Remote sensing; Ecosystem function

**Abbreviations:**

EC: Eddy covariance

CRP: Conservation Reserve Program

GPP: Gross primary production

$GPP_{Tower}$: GPP estimate from eddy covariance towers

$GPP_{MODIS}$: MOD17AH V6 GPP

$GPP_{CONUS}$: Landsat CONUS GPP

$GPP_{VPM\text{-}MODIS}$: Vegetation photosynthesis model with MODIS data (500 m)

$GPP_{VPM\text{-}LS8}$: Vegetation photosynthesis model with Landsat 8 data (30 m)

$GPP_{VPM\text{-}S2}$: Vegetation photosynthesis model with Sentinel-2 data (20 m)

NLCD: National Land Cover Database

LUE: Light use efficiency

LSWI: Land surface water index

PPFD: Photosynthetic photon flux density

VI: Vegetation Index

VPM: Vegetation photosynthesis model

$VPM_{S2\text{-}CIg}$: Vegetation photosynthesis model with Sentinel-2 (20 m) green chlorophyll index (CIg)

$VPM_{S2\text{-}Clr}$: Vegetation photosynthesis model with Sentinel-2 (20 m) red-edge chlorophyll index (Clr)

$VPM_{S2\text{-}NDRE1}$: Vegetation photosynthesis model with Sentinel-2 (20 m) normalized difference red-edge index 1 (NDRE1)

$VPM_{S2\text{-} NDRE2}$: Vegetation photosynthesis model with Sentinel-2 (20 m) normalized difference red-edge index 2 (NDRE2)

$VPM_{S2\text{-}MTCI}$: Vegetation photosynthesis model with Sentinel-2 (20 m) MERIS terrestrial red-edge chlorophyll index (MTCI)

## 1. Introduction

Rising demands for food, biofuel, and other commodities across the globe are driving increases in cropland cover and productivity (Godfray et al., 2010; Potapov et al., 2022; Tilman et al., 2011). This intensity increases greenhouse gas (GHG) emissions and threatens ecosystem health through fragmentation and loss of habitat (Houghton et al., 2012; USGCRP, 2018; Zabel et al., 2019). Cropland and managed grasslands are the dominant land cover types of many industrial, newly industrializing as well as developing countries (Bondeau et al., 2007; Foley et al., 2005), totaling 38% of the global land surface (Ramankutty et al., 2008) and ~30% of global net primary production appropriated by humans (Haberl et al., 2007). Given the association between cropland intensification, rising GHG emissions, and threat to biodiversity and ecosystem functions, the United Nations Sustainable Development Goals (SDGs) call for economies to become carbon (C) neutral by 2030 as well as to prioritize food security and ecological resource protection (United Nations, 2015). To aid our understanding of how ecosystem functions respond to changing climate, particularly where cropland is dominant, accurate C estimations are essential.

Terrestrial GPP is the major driver of land C sequestration and vital to the global C balance, but is highly variable in croplands due to land management practices (e.g., crop rotation, irrigation, abandonment, etc.). As intensification continues, croplands will also experience an increase in terrestrial gross primary production (GPP), the amount of carbon dioxide 'fixed' as organic material through photosynthesis. In addition to physical influences and disturbances (i.e., climate, geomorphology, land cover change, wildfires, floods), the magnitude and dynamics of GPP are also driven by anthropogenic activities that alter land use and land cover dynamics, as well as biogeochemical cycles (Abraha et al., 2018; Anav et al., 2015; Hibbard et al., 2017; Lei et al., 2021; Piao et al., 2009; Sciusco et al., 2020). Therefore, it is challenging to generate specific C balance estimates within croplands (Gelybó et al., 2013).

While GPP cannot be directly measured, it can be modeled using the eddy-covariance (EC) method, which partitions net ecosystem exchange (NEE) into GPP and ecosystem respiration (Aubinet et al., 2012;

Baldocchi et al., 2012; Lasslop et al., 2010; Papale et al., 2006; Reichstein et al., 2005). Eddy covariance field-scale measures of C, water and energy cycles have provided detailed knowledge on cropland and grassland contributions to GHG exchanges, C budgets and opportunities for natural climate solutions (Abraha et al., 2019; Chen et al., 2018; Hemes et al., 2021; Shao et al., 2017). At regional to global scales, many studies have scaled EC tower observations using data-driven, process based models (Beer et al., 2010; Jung et al., 2009) and found meteorological data have little impact on upscaled GPP with high-quality satellite data (Joiner & Yoshida, 2020). Measures are scaled by evaluating the relationships between tower-based GPP estimates and satellite-based, gridded and reanalysis data of climate, meteorological, and surface-reflectance estimates to constrain and calibrate models that monitor vegetation health and yield (Cai et al., 2021; Kumar & Mutanga, 2017; Lin et al., 2019; Wolanin et al., 2019; Xiao et al., 2011). Scaling and extrapolation to regional or global representativeness should be exercised with caution as it can increase uncertainty (Beer et al., 2010; Chu et al., 2017). This can be understood as the Modifiable Areal Unit Problem (MAUP) that includes (1) the "scale problem", when areal data is aggregated into several sets of larger units; and (2) the "zoning problem", when a given set of areal units are recombined into zones that are of the same size but located differently. Both problems result in variation in data values and subsequently different conclusions (Jelinski & Wu, 1996).

Similarly, the choice of model and spatial resolution may either inflate or underestimate GPP in heterogeneous croplands. Model comparison is necessary, as it identifies variations that could help identify shortcomings and areas for future improvement (Morales et al., 2005). Comparison is also a prerequisite for analyzing spatiotemporal biosphere-atmosphere fluxes as it reveals effects from different model structures (i.e., structural uncertainty) (Wang et al., 2011; Zhao et al., 2012), parameter values, meteorological input data, and vegetation and soil C pools (Anav, 2015). Therefore, examination of various GPP models and their spatial and temporal variations in croplands is necessary to advance our understanding of land management and land use effects on the global C budget.

Integration of EC and remote sensing methods have greatly advanced our ability to estimate GPP. However, due to the intense fragmentation, there can be a mismatch between small patches and conventional remote sensing spatial resolution  (Ustin & Middleton, 2021). For example, global products, like the highly utilized 8-day Moderate-Resolution Imaging Spectroradiometer (MODIS) MOD17A2/A3 and MYD17A2/A3 GPP products (1 km–500 m), can be challenging if used in the context of land cover areas with complex vegetation or mixed pixels (Running & Zhao, 2015). In fact, coarse remote sensing models may aggregate nearby land cover patches within the same estimate of land cover GPP productivity, introducing a mischaracterization of landscape processes (Reeves et al., 2005; F. Zhang et al., 2012). To estimate GPP within fragmented landscapes under various management practices, remote sensing offers several approaches to estimate GPP using measurements of optical parameters directly related to vegetation activity (Damm et al., 2015; Myneni & Ross, 2012). Advancements in optical sensors, such as those carried aboard Landsat-8 (2013–now) and -9 (2021–now), offer 30 m spatial resolution whereas Sentinel-2 A and B (2015/2017–now) offer 10-20 m spatial resolution and narrow red-edge bands—enabling phenology studies and parametrization at a much higher resolution than previously possible (Li & Roy, 2017).

Of the primary remote-sensing based models, the most common are light use efficiency (LUE) based estimates that are built on function convergence theory (Field et al., 1995; Monteith, 1972, 1977), which states that plant canopies will harvest the most light to fix C given the constraints from the environment (Goetz et al., 2000). Following this framework are the production efficiency models (PEMs), where GPP is estimated as a product of the fraction of the photosynthetically active radiation (*f*PAR) absorbed by the canopy (e.g., Goetz et al., 1999; Ruimy et al., 1999; Running et al., 2004). For example, the Landsat conterminous United States (CONUS) GPP product captures fine spatial scale (30 m) variability in GPP production with biome-specific inputs and provides ready-to-use product covering croplands, forests, grasslands and shrublands (Robinson et al., 2018). The vegetation photosynthesis model (VPM) similarly estimates GPP in various ecosystems, and its performance aligns well with EC GPP (John et al., 2013; Li et al., 2007; Wagle et al., 2015; Xiao et al., 2004a; Xiao et al., 2004b; Zhang et al., 2016).

Further, many remote sensing-based GPP models, such as VPM, rely on vegetation indices (VI) as input variables that serve as a proxy of *f*PAR and associated nutrient and absorption characteristics. Red-edge bands offered from the Sentinel-2A and B satellites offer additional VIs capable of estimating GPP, as vegetation red-edge (680-780 nm) captures the absorption of chlorophyll at 680 nm and higher absorption at 780 nm, detecting both moderate-to-high values (Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, 1965; Gitelson & Merzlyak, 1996; Horler et al., 1983). This is significant as chlorophyll has demonstrated a high sensitivity to seasonal changes and a strong relationship to GPP in croplands (Clevers & Gitelson, 2013; Lin et al., 2019; Wu et al., 2008) . In addition, fine spatial resolution of the Sentinel-2 data provides temporally detailed information for characterizing spatially heterogeneous GPP best in croplands and grasslands compared to forest sites (Lin et al., 2019). Across grassland sites in southeast Australia, Sentinel-2 red-edge data estimates of GPP agreed well with EC GPP ($R^2$ = 0.77 and RMSE = 0.81 g C $m^{-2}$ $day^{-1}$) (Lin et al., 2019). Sentinel-2 and Landsat 8 data have also been used to estimate a neural network GPP model on five crop fields (four in the USA and one in Germany) ($R^2$ = 0.92 and RMSE = 1.38 g C $m^{-2}$ $day^{-1}$) (Wolanin et al., 2019). EVI2-derived GPP from MODIS (500m, 250m) and Sentinel-2 (10m) and EC-derived were evaluated in eight sites in the Nordic region ($R^2$ 0.69-0.91 and RMSE 0.49-2.19 g C $m^{-2}$ $day^{-1}$) (Cai et al., 2021). Few studies, however, cross-compare product resolutions in VPM to investigate changes across scales within the same cover type; or have tested red-edge VIs. More commonly, VPM is cross-evaluated with other GPP products, such as MOD17, a temperature and greenness model, a greenness and radiation model, and the EC-LUE model (F. Li et al., 2013; Chaoyang Wu et al., 2011). Therefore, red-edge VIs from Sentinel-2 integrated into the VPM may enhance our ability to estimate GPP in heterogeneous croplands (Chen et al., 2011; Turner et al., 2003).

In this study, we evaluate whether GPP estimates derived using higher spatial resolution of satellite data is advantageous to conventional remote sensing products in managed croplands. We ask the following questions: (1) Can fine resolution GPP products built with red-edge VIs effectively capture significant differences at field-scale? (2) Are they significantly different from the conventionally used models—MOD17A2H V6 (500m) and Landsat-8 CONUS (30m)? and (3) How consistent are GPP anomalies across models within each site? While coarse resolution GPP products are reasonable for studies of large spatial extents, like global and regional (Running & Zhao, 2015), local-scale estimates of

GPP are needed for local-scale management, estimates of C sequestration, and for C accounting. We generate site-specific LUE coefficients and model GPP, utilizing the VPM across MODIS (500 m), Landsat-8 (30 m), and Sentinel-2 (20 m) resolutions. By comparing multiple possible approaches to estimating GPP, we show which products are the most accurate in our managed cropping systems.

## Methods

### *2.1. Study sites*

Our study sites are located within the northeast portion of the US Midwest Corn Belt in southwest Michigan, USA, at the Great Lakes Bioenergy Research Center (GLBRC) of the W. K. Kellogg Biological Station (KBS) Long-Term Ecological Research (LTER) sites (42°24′ N, 85°24′ W, 288m a.s.l.; Figure 1, Table S1). The sites are in a humid continental temperate climate with mean annual air temperature 9.9 °C and mean total annual precipitation 1027 mm (Michigan State Climatologist's Office, 2013). Soils are Typic Hapludalfs, well-drained sandy loams (Bhardwaj et al., 2011; Thoen, 1990). From May through September, roughly representing the growing season, mean air temperature and total precipitation are 19.7°C and 523 mm, respectively, with highest temperatures in July (Abraha et al., 2018). Our study period spans March through November (DOY 60–334), including the growing season as well as its onset and offset, for years 2018 and 2019. Precipitation, air temperature, and photosynthetic photon flux density (PPFD) a nearby meteorological stations (http://lter.kbs.msu.edu/datatables, accessed June 2020). Seasonal dynamics of GPP are driven by PPFD and temperature in these temperate croplands, where GPP lowers to near-zero in the winter season — DOY 335-59 (December through February) — due to near absence of photosynthetic activity caused by snow cover, harvest as well as low PPFD and temperatures.

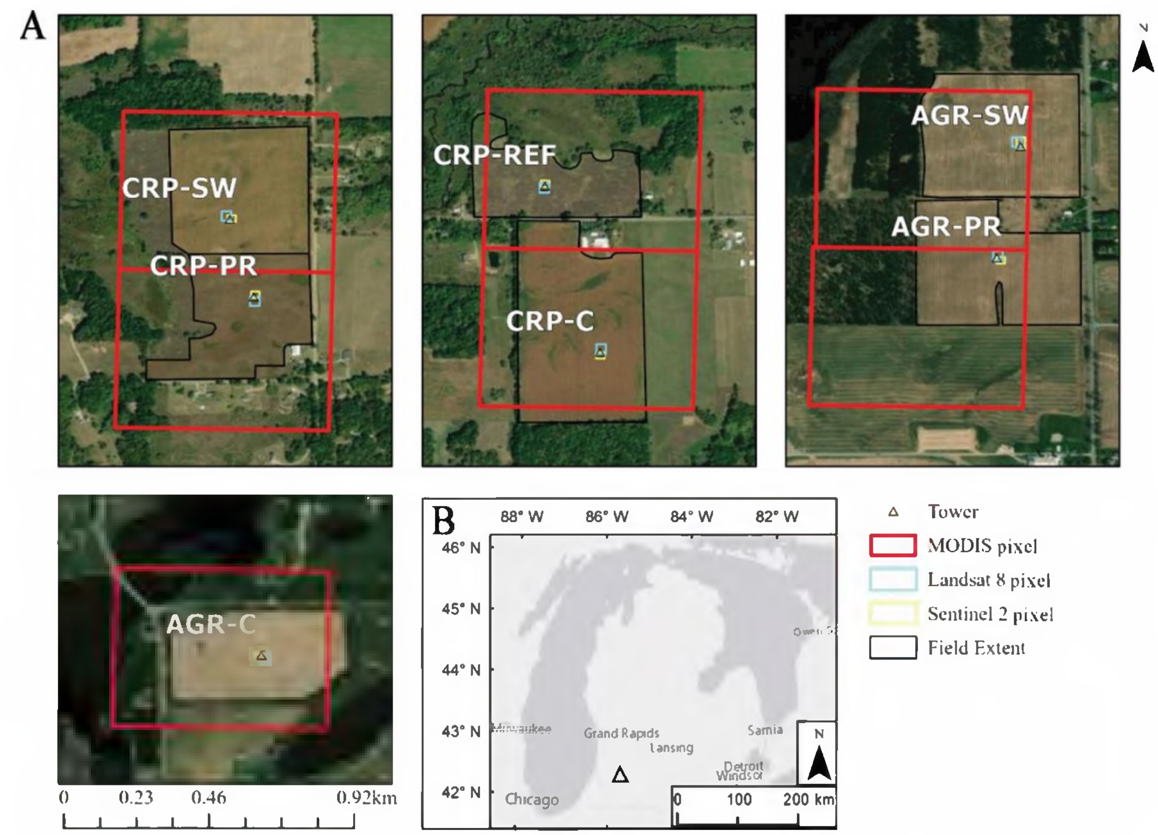


Figure 1. Location of eddy-covariance (EC) flux towers used in this study, where (A) are individual field extents and individual pixels for MODIS, Landsat-8, and Sentinel-2; and (B) is the location the towers at Kellogg Biological Station, Michigan, USA.

We consider seven study sites that are named according to their land cover history prior to 2009 and present land cover after land use conversion (i.e., names are interpreted as HISTORIC-PRESENT, Fig. 1). Two distinct land use histories—agriculturally cultivated land (i.e., AGR-) and Conservation Reserve Program grassland (i.e., CRP-)—were used. In one group, three fields were managed as CRP grasslands for 22 years with smooth bromegrass (*Bromus inermis Leyss*)—a cool season C3 grass of Eurasian origin—as the dominant vegetation (Abraha et al., 2016). The second set of three fields included conventionally-tilled corn-soybean rotations (AGR) cultivated as such for decades prior to this study. Both groups were converted to their present land cover types in 2009. Therefore, CRP sites include: (1) no-till corn (CRP-C); (2) restored prairie (CRP-PR); and (3) switchgrass (CRP-SW); and AGR sites include (4) no-till corn (AGR-C); (5) switchgrass (AGR-SW); and (6) restored prairie (AGR-PR); and (7) historically preserved CRP land (CRP-REF) (Fig. 1, Table S1). Upon conversion, the former CRP fields held significantly higher soil organic C and nitrogen (N) concentrations than the former AGR fields within its top 0.25m of soil (Abraha et al., 2018b; Zenone et al., 2011). The fields restored to prairie were planted with a mixture of 19 species (Abraha et al., 2016). During the study period, planting dates for AGR-C was May 7, 2018 (DOY 127) and May 11, 2019 (DOY 131); whereas for CRP-C it occurred on May 2, 2018 (DOY 122) and May 6, 2019 (DOY 126).

### *2.2. Eddy covariance*

All EC systems included a LI-7500 open-path infrared gas analyzer (IRGA, LI-COR Bioscience, Lincoln, NE) for $CO_2$ and water ($H_2O$) concentration and a CSAT3 three-dimensional sonic anemometer (Campbell Scientific Inc. CSI, Logan, UT) for wind speed and direction measurements. Half-hourly meteorological measurements of incoming and outgoing radiation (CNR1, Kipp & Zonen, Delft, The Netherlands) and air temperature and relative humidity (HMP45C, CSI) were also measured at each site. All EC instruments are mounted 1.5–2.0 m above the vegetation and logged at 10Hz using a Campbell CR5000 datalogger. Half-hourly fluxes were processed in EdiRe for screening out-of-range data due to bad weather, sensors, and/or logger malfunction as well as de-spiking. For full data quality control details, please see Abraha *et al.* (2015).

Gapfilling and flux partitioning was completed in the standardized FLUXNET gap-filling algorithm from REddyProc (Wutzler et al., 2018). Gap-filling included a Ustar correction with thresholds estimated using the Moving Point Test (Papale et al., 2006), bootstrap uncertainty within the year, and flux partitioning by daytime (Lasslop et al., 2010). We used quality control flags ("*fqc*") of 0-3 in this study, where least reliable (i.e., *fqc*=3) estimates comprised less than 0.54% of any site-year, and values outside of three standard deviations were linearly interpolated with the package "*seismicRoll*" (Callahan et al., 2020) in RStudio 1.3.1056 (R Core Team, 2019). We present GPP uncertainty across aggregated values due to estimation of the Ustar threshold, as well as the percent NEE gap-filled prior to partitioning.

### *2.3. Satellite products and indices*

We obtained GPP (kg C $m^{-2}$) from the MODIS MOD17A2H V6 product (8-day revisit time and 500 m resolution; hereafter $GPP_{MODIS}$) and the Landsat 8 CONUS product (16-day revisit time and 30 m resolution; hereafter $GPP_{CONUS}$) (Robinson et al., 2018). Both $GPP_{MODIS}$ and $GPP_{CONUS}$ were retrieved from Google Earth Engine (GEE) platform (Gorelick et al., 2017) using point sampling to select the nearest pixel to the site's tower location. We considered only pixels nearby each tower, which brought us to consider 1 (500x500 m) MODIS  pixel and 3x3 Landsat-8 (30x30 m) and Sentinel-2 (20x20 m) pixels. The models used to calculate $GPP_{MODIS}$ and $GPP_{CONUS}$ are based on the LUE model (Running et al., 2004). However $GPP_{MODIS}$ retrieves climate, land cover, *f*PAR and LAI parameters from GMAO/NASA (0.5°), MOD12Q1 (500 m), and MOD15A2H (500 m), respectively, whereas $GPP_{CONUS}$ retrieves these parameters from Idaho Metdata (4 km), National Land Cover Database (NLCD; 30 m), and MOD09Q1 (250 m), respectively. To derive daily estimates, composite images $GPP_{MODIS}$ and $GPP_{CONUS}$ were divided by 8 and 16, respectively, and multiplied by 1000 to convert from kg C to g C, with final GPP units being expressed as g C $m^{-2}$ $d^{-1}$.

For VPM (Section 2.4), we used surface reflectance from MODIS, Landsat-8 and Sentinel-2 (acquisition details below) to calculate vegetation indices (VIs). The VIs include (1) the enhanced vegetation index 2 (EVI2) (Jiang et al., 2008) to account for moisture sensitivity; (2) the land surface water index (LSWI) (Xiao et al., 2004b), which is based on the shortwave-infrared (SWIR) and represents vegetation water content and soil moisture. In place of EVI2, we also test VIs including (3) the green Chlorophyll Index (CIg) and red-edge (4) Chlorophyll Index (CIr) (Gitelson et al., 2003, 2006); the (5) normal deviation index of the red edge 1 (NDRE1) (Sims & Gamon, 2002) and (6) normal deviation index of the red edge 2 (NDRE2) (Barnes et al., 2000); as well as the (7) medium-resolution imaging spectrometer, MERIS, terrestrial chlorophyll index (MTCI) (Dash & Curran, 2004). Surface reflectance and land surface temperature layers were quality checked and linearly interpolated for a representative time series.

The MODIS MOD09A1 v006 product provides surface reflectance at 500 m resolution every 8 days and it was used to calculate VIs using red (620–670 nm), near-infrared (NIR; 841–875 nm) and SWIR

(1628–1652 nm) bands. MODIS Terra has an overpass at 10:30 AM local time. Data was acquired using the USGS AppEEARS online tool (https://lpdaac.usgs.gov/tools/appeears/, accessed January 2021) and screened for cloud cover and artefacts using QA/QC bits and 500m state flags, as instructed by the MODIS User Guide Tables 10 and 13, to select the best quality data (Vermote et al., 2015). Gaps due to poor quality were linearly interpolated. USGS Landsat 8 surface reflectance (Tier 1) provided 30 m resolution imagery every 16 days to calculate VIs EVI2 (Eq. 4) and LSWI (Eq. 6) using red (636-673 nm), NIR (851-879 nm), and SWIR (1566-1651 nm). As for $GPP_{LS8\text{-}VPM}$, we acquired Landsat 8 data using GEE, and we used the pixel quality band "QA_PIXEL" to identify cloud and cloud shadow pixels

The Sentinel-2 is a constellation of two polar-orbiting satellites in the same sun-synchronous orbit. Surface reflectance over the study area provides a high revisit time of 10 days at the equator for a single and 5 days when 2 satellites under cloud-free conditions, which results in 2-3 days at mid-latitudes. Overpass for Sentinel-2 is 10:30 AM local time and is a compromise for illumination and least potential cloud cover, similar to the overpass time of Landsat and MODIS. Sentinel-2A spatial resolution is offered at 10, 20, and 60 m with a total of 12 multispectral bands; of which, three are red edge bands. Bands used (and their center wavelength) for EVI2 and LSWI include NIR (B8, 842 nm; 20 m spatial resolution), red (665; 10 m spatial resolution), and SWIR (1610 nm; 20 m spatial resolution), respectively. For red-edge indices (Eqs. 5-9), we also included the following: B3 (green, 560 nm), B5 (red-edge, 705 nm), B6 (red-edge, 740 nm), and B7 (red-edge, 783 nm).

The red band was resampled to 20 m resolution to match that of NIR and SWIR. Images were obtained from the Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home, accessed February 2021) of the European Space Agency. We downloaded images as level 2A (i.e., surface reflectance) over the study area. Where level 2A was not available, we downloaded level 1C top-of-atmosphere (TOA) images that were then atmospherically corrected to obtain surface reflectance by using the default settings of the Sen2Cor (v. 2.5.5) algorithm (Müller-Wilm et al., 2018). We performed the cloud mask in RStudio by using the cloud mask probability band "*MSK_CLDPRB*", to identify cloud pixels, and the scene classification map band "*SCL*", to identify water pixels. We then used a NIR thresholds to identify potential cloud shadow pixels (for more info, see https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless). We employed ArcMap (v. 10.6) to rescale the surface reflectance to 0–1.

Lastly, to understand how heterogeneous systems can benefit from fine-resolution imagery, we estimate the composition (30 m) of land cover type within each of the remote sensing pixels employed to estimate GPP, described above, within ArcGIS Pro (v. 2.9). We acquired land cover from the USGS National Land Cover Database 2019 via GEE (Dewitz & Survey, 2021). Land cover estimates included cropland, water, wetland, grassland, wetlands, developed and forest; where grassland includes pasture, hay, grassland, shrub/scrub, wetlands include woody wetlands and emergent herbaceous wetlands, developed includes open space, and low, middle and high intensity developed areas, and forest includes evergreen, deciduous and mixed forests.

### *2.4. Vegetation photosynthesis model (VPM)*

The VPM model is built similarly to the $GPP_{MODIS}$ equation (Xiao et al., 2004a; Xiao et al., 2004b), however the difference lies in the creation of LUE ($\varepsilon_g$, Eq. 2) from remote sensing and meteorological inputs rather than the use of a look up table, where:

$$VPM = \varepsilon_g \times (f\text{PAR}) \times (\text{PAR}) \,, \quad (1)$$

$$\varepsilon_g = \varepsilon_{max} \times T\mathit{scalar} \times W\mathit{scalar} \times P\mathit{scalar} \quad (2)$$

Here, *VPM* represents Sentinel-2, Landsat-8 and MODIS VPMs, hereafter $GPP_{VPM\text{-}S2}$ , $GPP_{VPM\text{-}LS8}$, and $GPP_{VPM\text{-}MODIS}$, respectively; $f$PAR is the fraction of photosynthetically active radiation absorbed by chlorophyll, PAR is photosynthetically active radiation (μmol $m^{-2}$ $s^{-1}$) acquired from nearby a weather station (http://lter.kbs.msu.edu/datatables, accessed June 2020), $\varepsilon_g$ is the LUE — the rate of $CO_2$ uptake (μmol $CO_2$ $PAR^{-1}$). The value of $\varepsilon_{max}$ is maximum LUE estimated from a nonlinear hyperbolic Michaelis–Menten model (Wang et al., 2010), and *Tscalar*, *Wscalar* and *Pscalar* are the scaling regulators for the effects of air temperature, water and leaf phenology, respectively, on the vegetation.

Common in LUE models, including PEMs, is the application of *f*PAR as a function of the normalized difference vegetation index (NDVI) (Tucker, 1979). It is well acknowledged that NDVI is constrained by sensitivity to soil moisture and saturates at high leaf densities (Huete et al., 2002). To address this, VPM applies EVI as a function of *f*PAR for an enhanced characterization of vegetation at the global scale (Huete et al., 2006; Jiang et al., 2008; Xiao et al., 2004a). To calculate $f$PAR, EVI can act as a linear function and the coefficient $a$ is set to 1.0 (Xiao et al., 2005; Xiao et al., 2004b). In this study, we apply EVI2 to avoid high signal to noise ratios from atmospheric interference (e.g., aerosol or residual clouds) common in blue band wavelengths (Jiang et al., 2008).

$$fPAR = a \times (EVI2) \tag{3}$$

$$EVI2 = 2.5\,\frac{NIR-RED}{NIR+2.4RED+1} \tag{4}$$

To evaluate the potential for red-edge bands available from Sentinel-2 to advance the VPM's applications, we chose to replace EVI2 with one of five red-edge VIs, CIg, CIr, NDRE1, NDRE2 and MTCI, calculated as:

$$CIg = \frac{B7}{B3} - 1 \tag{5}$$

$$CIr = \frac{B7}{B5} - 1 \tag{6}$$

$$NDRE1 = \frac{B6-B5}{B6+B5} \tag{7}$$

$$NDRE2 = \frac{B8-B5}{B8+B5} \tag{8}$$

$$MTCI = \frac{B6-B5}{B5-B4} \tag{9}$$

where the center of each Sentinel-2 band is as follows: B3 (560 nm), B4 (665 nm), B5 (705 nm), B6 (740 nm), B7 (783 nm), B8 (842 nm).

Down regulation scalars *Wscalar, Tscalar, Pscalar* demonstrate the effects of water, temperature, and leaf phenology respectively on the vegetation's LUE. *Wscalar* is estimated as:

$$Wscalar = \frac{1+LSWI}{1+(LSWI)max} \tag{10}$$

$$LSWI = \frac{NIR-SWIR}{NIR+SWIR} \tag{11}$$

where *(LSWI)max* is the maximum LSWI during the growing season. *Tscalar* measures the sensitivity of photosynthesis to temperature, calculated at each time step using the equation developed for the Terrestrial Ecosystem Model (Raich et al., 1991):

$$Tscalar = \frac{(T - Tmin)(T - Tmax)}{[(T - Tmin)(T - Tmax)] - (T - Topt)^2} \quad (12)$$

where *Tmin, Tmax,* and *Topt* are the photosynthesis minimum, maximum, and optimal temperatures (°C), respectively (Raich et al., 1991) (Table S2). If air temperature falls below *Tmin*, *Tscalar* is set to zero. *Pscalar* accounts for the effects of leaf phenology on photosynthesis at the canopy level. Calculation of *Pscalar* is dependent on the life expectancy of the leaves. *Pscalar* has two phases when a canopy is dominated by leaves with a life expectancy of one year (i.e., growing season) without replacement. From bud burst to full leaf expansion, *Pscalar* is calculated as:

$$Pscalar = \frac{1 + LSWI}{2} \quad (13)$$

whereas following expansion, the *Pscalar* is set to 1 with no alteration for senescence. Grassland systems such as prairie and switchgrass are set to 1 throughout the study period (Wang et al., 2010; Xiao et al., 2004a).

*2.5. Statistical analysis and uncertainty*

To understand how tower GPP estimates relate to either NDVI or EVI2, we performed sensitivity tests of both indices to $GPP_{Tower}$ acquired from MODIS, Landsat-8 and Sentinel-2 for each site-year using a procedure outlined in Gitelson (2004):

$$S = [d(EVI2)/d(NDVI) \bullet [\Delta(EVI2)/ \Delta (NDVI)]^{-1} \quad (14)$$

where *d(EVI2)* and *d(NDVI)* are the first derivatives of the indices with respect to $GPP_{Tower}$ and *Δ(EVI2)* and *Δ(NDVI)* are the differences between the maximum and minimum index, respectively. The function *S* tracks the sensitivity of EVI2 and NDVI to changes in $GPP_{Tower}$. Values of $S$ <1 can be interpreted where NDVI is more sensitive than EVI2 to $GPP_{Tower}$, and values $S$ >1 as indicate that EVI2 was more sensitive than NDVI to $GPP_{Tower}$. When $S = 1$, NDVI and EVI2 are assumed to be equally sensitive. We acknowledge that *S* does not account for estimate errors of d(EVI2)/d(NDVI), which may bias sensitivity evaluations

We evaluated seasonal dynamics of PPFD, air temperature, precipitation, as well as EVI2 and NDVI from MODIS, Landsat-8 and Sentinel-2 in a time series alongside $GPP_{Tower}$ for each site-year. A comparison of GPP sums during the study period (March–November) and growing season (June, July, August) evaluates differences between $GPP_{MODIS}$, $GPP_{VPM-MODIS}$, $GPP_{VPM-LS8}$, $GPP_{VPM-S2}$, $GPP_{CONUS}$, and $GPP_{Tower}$. Days without estimates from the VPM model or other products (i.e., days in-between acquisitions) were linearly interpolated within the R package "*zoo*" to generate cumulative GPP estimates (Zeileis & Grothendieck, 2005).

Three metrics were used to evaluate the performance of GPP satellite estimates in comparison with $GPP_{Tower}$, including the coefficient of determination (adjusted $R^2$, hereafter $R^2$), root mean square error (RMSE), and Spearman's Rho (ρ), which is a non-parametric test that estimates the model's ability to increase or decrease in a similar trend to observed values. Estimates closer to 1 indicate a positive relationship and those closer to -1 indicate a negative relationship. In the linear models, we only included original acquisition days (i.e., days corresponding to satellite acquisitions) that paired tower estimates. To assess model implications on GPP estimates, and by proxy the resolution implications, we tested for significant difference in GPP models among sites with the Kruskal-Wallis test and Dunn post-hoc test in the R packages "*stats*" and "*dunn.test*" (Dinno, 2017; Dunn, 1964; Kruskal & Wallis, 1952; R Core Team, 2019). The Kruskal-Wallis test extends from the Wilcoxon Rank test that is used for two samples (Vargha & Delaney, 1998), and determines if there is a significant difference ($p$-value $<0.05$) in the median GPP estimate between models. It replaces a one-way analysis of variance (ANOVA) when data is not normally distributed. The result of the Kruskal-Wallis is $H$, which is interpreted as chi-square; and z is result of the Dunn's Test for multiple comparisons.

Since our study area has strong seasonal changes of temperate zones, our data and predictions violate the statistical assumptions that they are independent and identically distributed. We address this concern of temporal autocorrelation in a second regression analysis by removing interannual and seasonal variation from each time series. We estimated zero-centered daily GPP anomalies and evaluated how these anomalies vary by GPP model and site-year. To generate average GPP seasonality (GPPS) on a daily time step (t) for each site (x) we averaged the daily GPP estimates from the different approaches for each year then smoothed the result with a Gaussian blur of 15 days to remove noise using the R package "smoother" (Hamilton, 2015). To remove interannual differences, we calculated $GPP_{x,yr}$ as the site-year annual mean of all GPP models. GPP anomalies (GPPA) were thus calculated as:

$$GPPA_{x,t} = GPP_{x,t} - GPPS_{x,t} - GPP_{x,yr} \quad (15)$$

Therefore, when an anomaly estimate is near-zero it has a small difference from the average, zero-centered seasonal pattern. Once we calculated daily GPPA (Eq. 15), we only included estimates that coincide with model acquisition dates to avoid inflation in our analysis. In the linear regression of anomalies, models agreeing well with $GPP_{Tower}$ will express similar values (i.e., differences from the mean) with $GPP_{Tower}$. In the linear regression of anomalies, models agreeing well with $GPP_{Tower}$ will express similar values (i.e., differences from the mean) with $GPP_{Tower}$.

## 3. Results

### *3.1. Seasonal changes of climate, vegetation indices and tower GPP*

Seasonal changes in air temperature, precipitation and PPFD at the LTER/KBS (i.e., study area) revealed that 2018 was on average warmer and drier than 2019 during the study period (March-November) (Fig. S1). For the study area in 2018, there was an average air temperature of 10.59 °C and a cumulative 796 mm of precipitation; whereas 2019 had an average air temperature of 9.25 °C and cumulative 896 mm of precipitation. We found $GPP_{Tower}$ increased sharply in May of both years at in all site-years (Fig. 2) due to the temperature increase, where the study area's monthly average air temperature from April to May increased from 4.49 °C to 18.18 °C in 2018, and 8.47 °C to 13.97 °C in 2019. We also found the study area in 2019 had notably higher cumulative monthly and average daily precipitation in spring months reaching 114(2.8), 92(2.97), and 173(5.77) mm in April, May and June; whereas 2018 had 63(2.1), 220(7.10), and 80(2.67) mm, respectively. $GPP_{Tower}$ uncertainty due to Ustar filtering for all site-years was < 3% (0.81–2.97%), with <28% (16.16–27.51%) of NEE identified for gapfilling (Table S4).

We found that MODIS 500 m pixels do not well represent each study site and include large aggregations of neighboring land covers (Table S3). One MODIS pixel including a tower may overlap two fields or nearby forest and marshland (Fig. 1). Conversely, the resolution of Sentinel-2 and Landsat 8 (20 m and 30 m, respectively) results in homogeneous pixels at each of the seven sites. Therefore, reflectance and vegetation indices from Landsat 8 and Sentinel 2 are more likely to represent the land cover of interest and minimize influence from neighboring vegetation. Monthly variability in $GPP_{Tower}$ during the growing season coincided well with the variations in precipitation, temperature, PPFD and EVI2/red-edge VIs. The $GPP_{Tower}$ during the growing season peaked in late July (DOY 185–217), which closely coincides with peak PPFD and temperature in the study area (Fig. S1). Peak dates of daily GPP at AGR-C and CRP-C from 2018 were delayed by approximately 20 days in 2019; whereas AGR-PR experienced a 15-day delay, and remaining sites peaked within 11 days (Fig 2).

The interannual seasonal dynamics of EVI2 differs in amplitude across sites and between satellites (Fig. 2). Maximum EVI2 for Sentinel-2 across sites ranged 0.65–0.86, whereas Landsat-8 and MODIS ranged 0.55–0.80 and 0.59–0.68, respectively. Sentinel-2 best captured the onset, offset, and volatility of the growing season. MODIS and, to a lesser extent, Landsat-8 EVI2 trends often exhibited lower estimates near the growing season peak. Notably, MODIS EVI2 increased before $GPP_{Tower}$ in the onset of the growing season and lags in the offset, particularly in AGR-C, CRP-C, AGR-PR and CRP-REF. Interannual seasonal dynamics of red-edge VIs capture peak growing season GPP well, particularly in corn systems, and reach higher peaks than EVI2 in CRP-PR and CRP-REF sites (Fig. 3). Red-edge VIs also demonstrate a similar trend as GPP during spring and fall in all sites.

MODIS EVI2 is more sensitive to variations in $GPP_{Tower}$; whereas for Landsat-8 and Sentinel-2, EVI2 and NDVI have similar sensitivity (i.e., 0.00 ± 0.10) (Table 1). MODIS EVI2 is more sensitive to $GPP_{Tower}$ in all site years except CRP-SW in 2018. We note that historical cropland sites AGR-C, AGR-

PR and AGR-SW as well as CRP-REF and CRP-C have higher sensitivities to MODIS EVI2. For Landsat-8, AGR-C, AGR-PR, CRP-PR and CRP-REF exhibit sensitivities to both NDVI and EVI2 in different years, with CRP-C, AGR-SW and CRP-SW demonstrating higher sensitivities to NDVI in both years. Similarly, Sentinel-2 saw sensitivities change between years, but exhibited slightly higher sensitivity to NDVI in AGR-C, AGR-PR and CRP-SW. Overall, we found Landsat-8 sensitivities remained within ±0.10 of 1.00 (i.e., equal sensitivity) for 9:14 (i.e., 9 out of 14) site-years, respectively; whereas Sentinel-2 exhibited sensitivities ±0.10 of 1.00 within 12:14 site-years.

Table 1. The relative sensitivity of EVI2 to NDVI. Values of $S < 1$ indicate that NDVI is more sensitive than EVI2, sensitivities are considered to be equal when $S = 1$, and values of $S > 1$ indicate EVI2 having a greater sensitivity than NDVI.

| Site | MODIS | | Landsat-8 | | Sentinel-2 | |
|---|---|---|---|---|---|---|
| | *2018* | *2019* | *2018* | *2019* | *2018* | *2019* |
| AGR-C | 1.31 | 1.33 | 1.08 | 0.99 | 0.90 | 0.96 |
| AGR-PR | 1.32 | 1.26 | 0.94 | 1.00 | 0.77 | 0.94 |
| AGR-SW | 1.30 | 1.22 | 0.94 | 0.92 | 0.99 | 1.01 |
| CRP-C | 1.11 | 1.13 | 0.77 | 0.99 | 0.78 | 1.00 |
| CRP-PR | 1.07 | 1.18 | 1.04 | 0.86 | 1.04 | 0.93 |
| CRP-REF | 1.20 | 1.21 | 1.22 | 0.81 | 1.01 | 0.91 |
| CRP-SW | 0.77 | 1.08 | 0.97 | 0.57 | 0.95 | 0.94 |

Differences between sensitivities of EVI2 and red-edge VIs to $GPP_{Tower}$ vary (Table 2). In most cases, NDRE1 is near similar in sensitivity to EVI2 in all sites except CRP-C, where NDRE1 is more sensitive. Between NDRE2 and EVI2, most sites had near-equal sensitivities, except for AGR-SW 2018 where EVI2 has higher sensitivity. Both Clg and Clr show a lower sensitivity than EVI2 in all site-years except in CRP-C. Lastly, sensitivities of MTCI and EVI2 were near equal in all site years except AGR-SW 2018, where EVI2 has higher sensitivity. Overall, NDRE1 and NDRE2 have 8:14, Clg and Clr have 2:14, and MTCI 5:14 site years with higher sensitivity than EVI2 to $GPP_{Tower}$.

Table 2. The relative sensitivity of EVI2 to Sentinel-2 red-edge bands NDRE1, NDRE2, Clg, Clr, and MTCI. Values of S < 1 indicate that the red-edge index is more sensitive than EVI2, sensitivities are considered to be equal when S = 1, and values of S > 1 indicate EVI2 having a greater sensitivity than the respective red-edge index.

| Site | NDRE1 | | NDRE2 | | Clg | | Clr | | MTCI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *2018* | *2019* | *2018* | *2019* | *2018* | *2019* | *2018* | *2019* | *2018* | *2019* |
| AGR-C | 0.88 | 0.92 | 0.87 | 0.93 | 1.19 | 1.13 | 1.14 | 1.15 | 0.98 | 1.05 |
| AGR-PR | 1.08 | 1.06 | 1.08 | 1.09 | 1.43 | 1.27 | 1.34 | 1.25 | 1.05 | 1.06 |
| AGR-SW | 1.32 | 0.95 | 1.40 | 1.02 | 1.61 | 1.14 | 1.59 | 1.12 | 1.26 | 1.00 |
| CRP-C | 0.73 | 0.74 | 0.74 | 0.75 | 0.94 | 0.89 | 0.97 | 0.89 | 0.90 | 0.86 |
| CRP-PR | 0.96 | 0.93 | 0.93 | 0.93 | 1.26 | 1.08 | 1.16 | 1.08 | 1.05 | 0.99 |
| CRP-REF | 1.00 | 1.00 | 1.12 | 1.02 | 1.34 | 1.20 | 1.14 | 1.09 | 0.93 | 1.12 |
| CRP-SW | 0.96 | 0.91 | 0.96 | 0.94 | 1.12 | 1.13 | 1.26 | 1.18 | 1.13 | 1.04 |

In both years, $GPP_{VPM\text{-}S2}$ explains more variability and is statistically significant in the linear regression analysis with $GPP_{Tower}$ during the study period (Table S5). $GPP_{VPM\text{-}S2}$ demonstrates visibly higher peaks in the growing season than other models, but occasionally over estimates in 2018 (AGR-C, AGR-PR, CRP-C, CRP-PR, CRP-REF) and in 2019 (CRP-C, CRP-REF). MODIS products generally underestimate these amplitudes (Figs. 4, 5). MODIS products largely underestimate corn and switchgrass systems where $GPP_{VPM\text{-}S2}$ captured GPP dynamics. In addition, VPMs coincide with $GPP_{Tower}$ peaks and variations more than $GPP_{MODIS}$ and $GPP_{CONUS}$, particularly in corn systems. Average daily $GPP_{Tower}$ is higher in 2018 compared to 2019; where in 2018, the most productive sites (CRP-SW, AGR-C, and AGR-PR) reached 5.66–6.27 g C $m^{-2}$ $d^{-1}$ compared to the most productive sites in 2019 (CRP-PR, CRP-C, and CRP-SW) with a range of 5.73–5.78 g C $m^{-2}$ $d^{-1}$. Corn systems have the highest daily productivity in both years but experienced the greatest shift in peak dates between 2018 and 2019. In both years, the highest daily sum recorded were in sites CRP-C, AGR-C, and CRP-SW while the lowest was observed in CRP-REF.

When exchanging EVI2 for a red-edge VI in the Sentinel-2 VPM, there is a significant improvement across site-years. Particularly, NDRE1 and NDRE2 improve the Sentinel-2 VPM in eight out of 14 site-years compared to other red-edge VIs. In 2018, NDRE2 improves AGR-C, CRP-C, and CRP-SW by improving explanation of variation by 8%, 11% and 4%, respectively; whereas in 2019, it improves AGR-C, AGR-PR, AGR-SW, CRP-C, and CRP-SW by 7%, 4%, 3%, 16% and 4%, respectively (Table

S5). NDRE1 also improves AGR-C in both years and CRP-C in 2018 by the same explanation of variance as NDRE2. While $GPP_{VPM\text{-}LS8}$ is better than $GPP_{VPM\text{-}S2}$ in both CRP-C site-years, but with NDRE2 the VPM improves by 11% and 16% in 2018 and 2019, respectively. Red-edge VIs NDRE1, Clr and Clg do not improve the Sentinel-2 VPM beyond that of NDRE2. While MTCI does improve the Sentinel-2 VPM in CRP-REF and explains 4% more variation and is the leading GPP model for both site-years, it still overestimates during the peak growing like $GPP_{VPM\text{-}S2}$ and VPM $_{VPM\text{-}Clg}$ (Fig. 5)

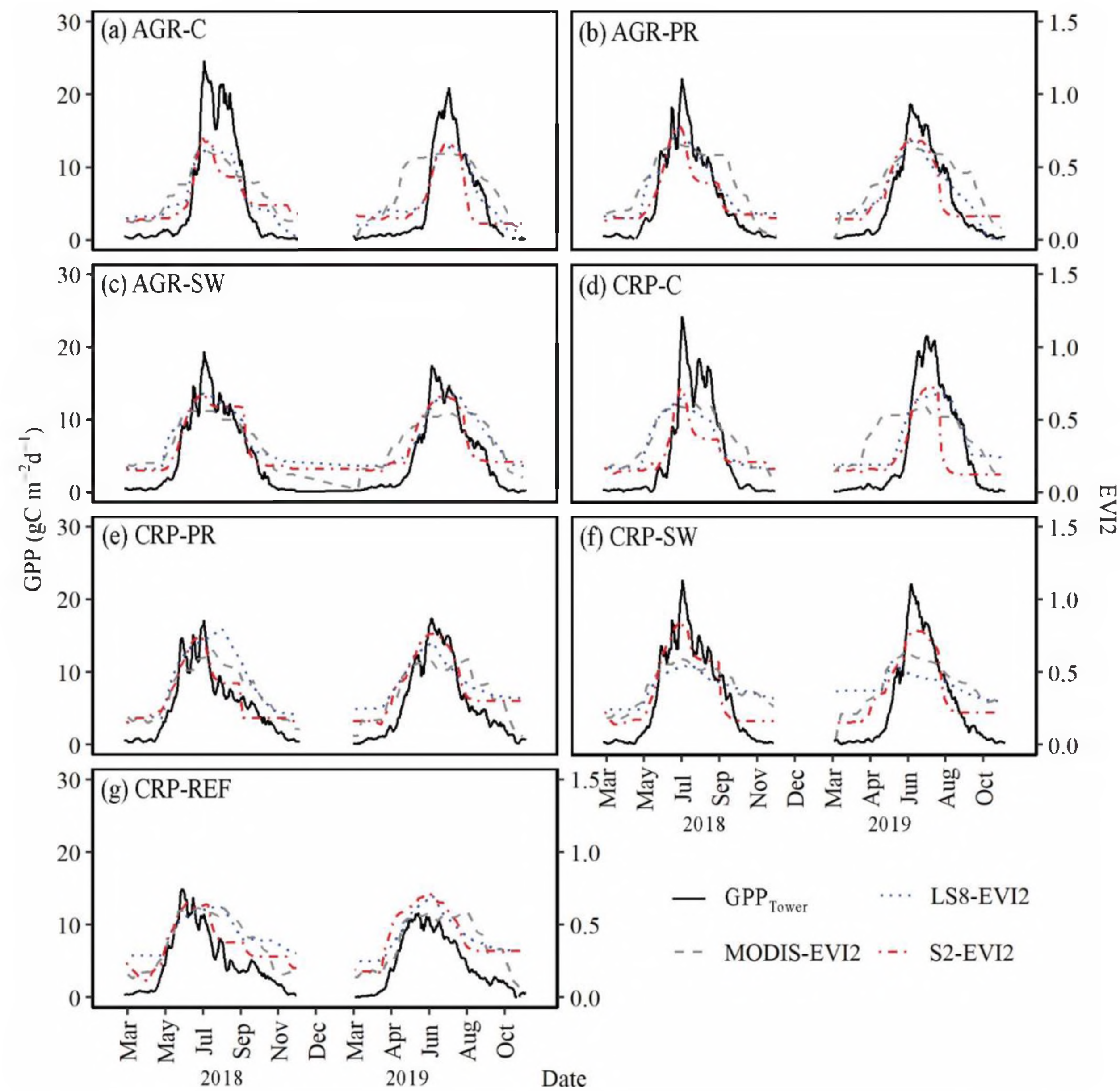

Figure 2. Daily $GPP_{Tower}$ estimates (g C $m^{-2}$ $d^{-1}$) as well as MODIS, Landsat-8, and Sentinel-2 EVI2 at (a) AGR-C, (b) AGR-PR, (c) AGR-SW, (d) CRP-C, (e) CRP-PR, (f),CRP-SW, and (g) CRP-REF sites 2018–2019.

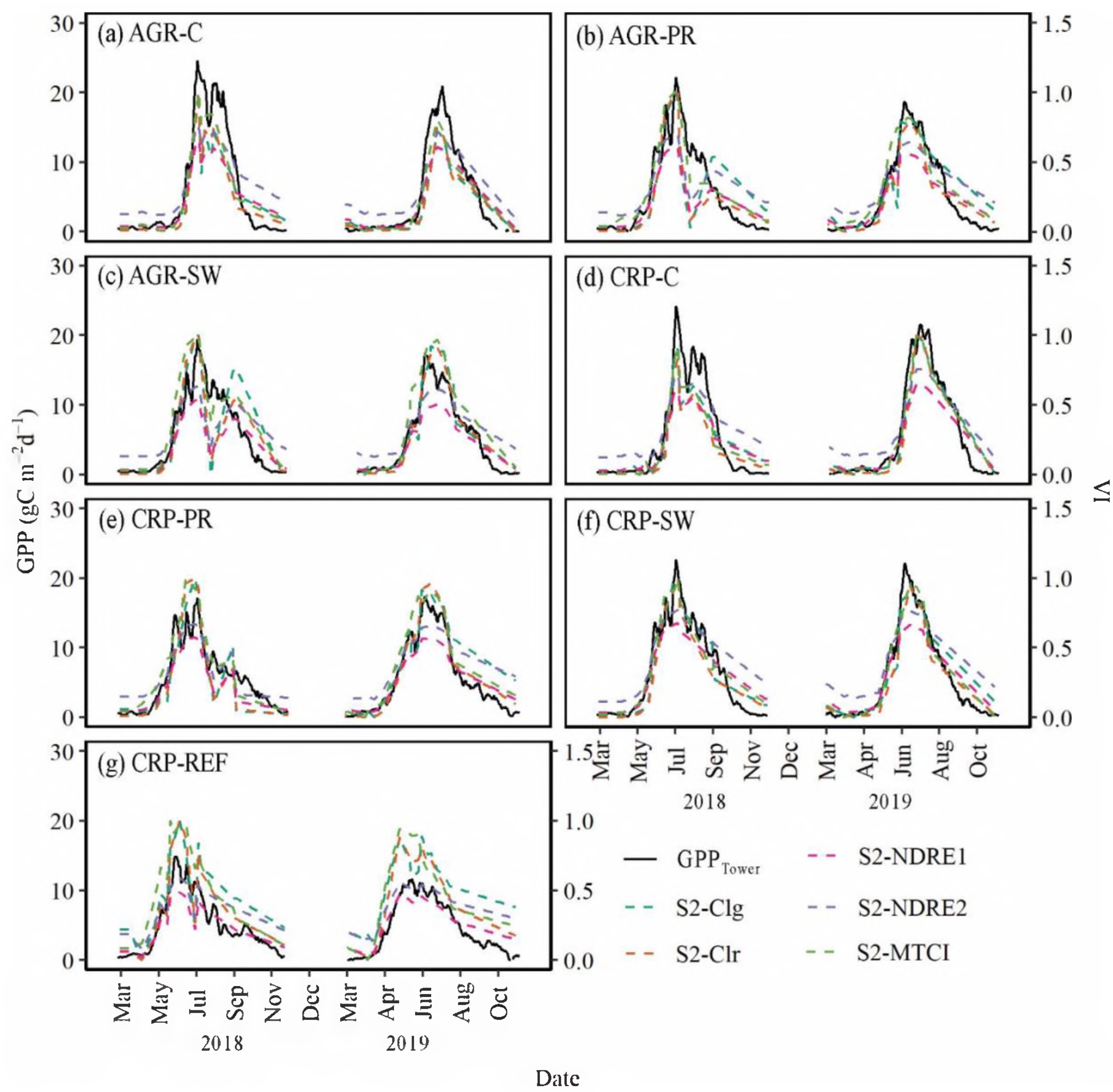


Figure 3. Daily $GPP_{Tower}$ estimates (g C $m^{-2}$ $d^{-1}$) as well as Sentinel-2 red edge vegetation indices Clg, Clr, NDRE1, NDRE2, and MTCI at (a) AGR-C, (b) AGR-PR, (c) AGR-SW, (d) CRP-C, (e) CRP-PR, (f),CRP-SW, and (g) CRP-REF sites 2018–2019.

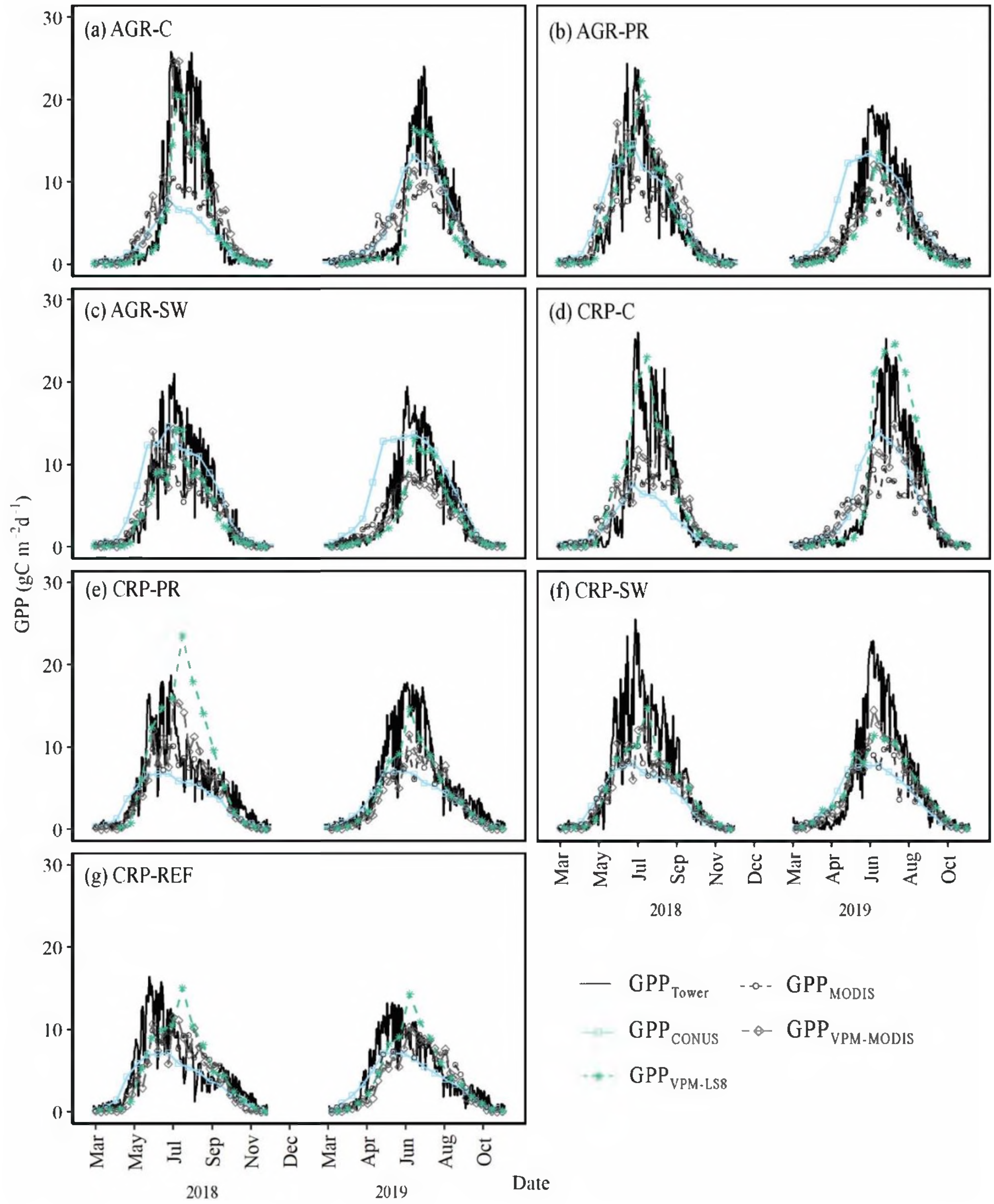


Figure 4. Temporal changes in GPP$_{Tower}$ , conventional and VPMs including CONUS and MODIS resolutions 2018–2019 for the seven study sites: (a) AGR-C, (b) AGR-PR, (c) AGR-SW, (d) CRP-C, (e) CRP-PR, (f), CRP-SW, and (g) CRP-REF.

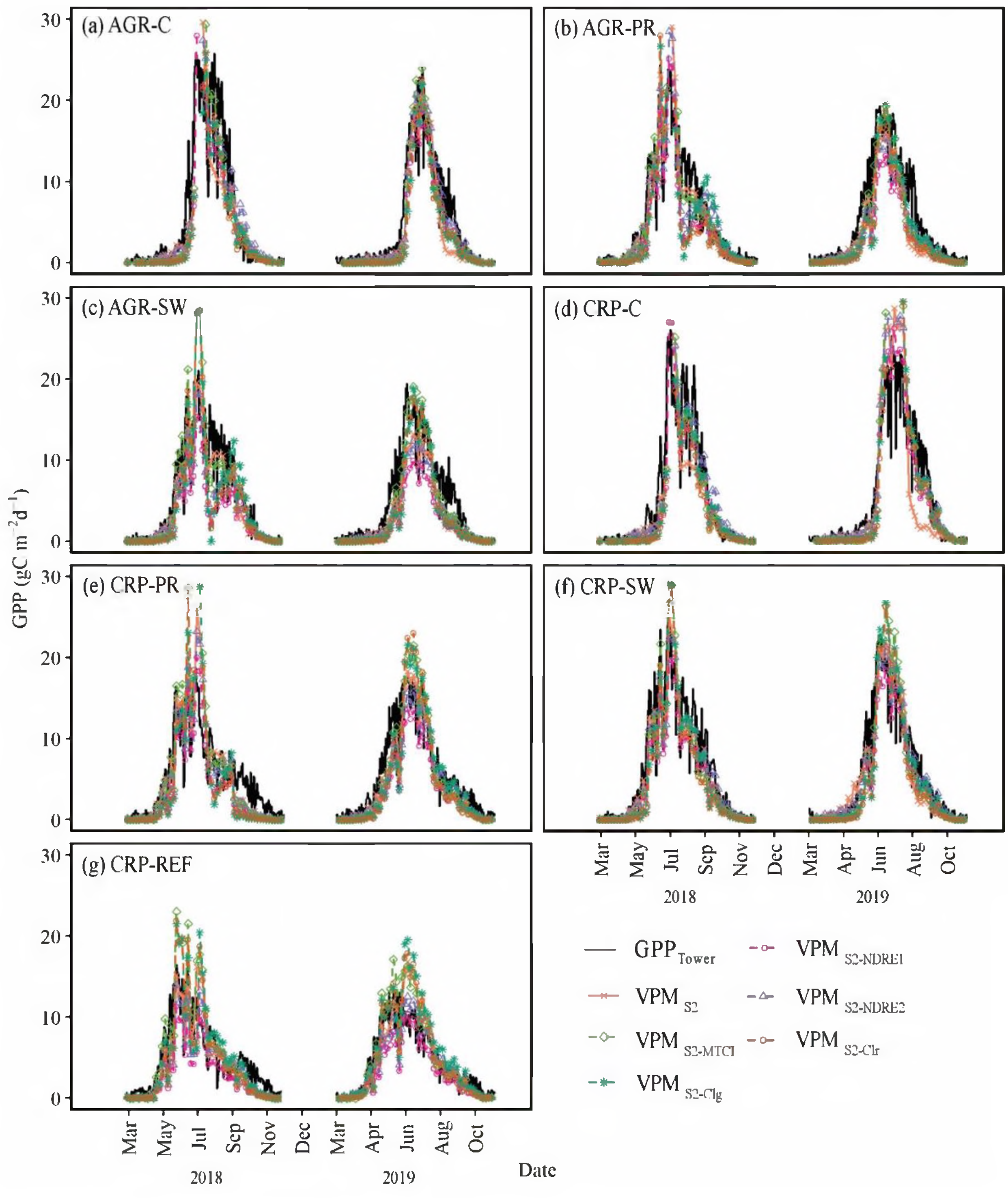


Figure 5. Temporal changes in GPP$_{Tower}$ and Sentinel-2 VPM RS models 2018–2019 for the seven study sites: (a) AGR-C, (b) AGR-PR, (c) AGR-SW, (d) CRP-C, (e) CRP-PR, (f) CRP-SW, and (g) CRP-REF.

During the study period, $GPP_{VPM\text{-}S2}$ estimated 5:14 site-year sums at ±10% that of $GPP_{Tower}$ sums, whereas $GPP_{VPM\text{-}LS8}$ had 3:14, $GPP_{VPM\text{-}MODIS}$ 2:14, $GPP_{MODIS}$ had 0:14, and $GPP_{CONUS}$ 3:14 (Table 3). When using red-edge VIs, $VPM_{S2\text{-}NDRE1}$ models estimated 1:14, $VPM_{S2\text{-}NDRE2}$ 5:14, $VPM_{S2\text{-}MTCI}$ 6:14, $VPM_{S2\text{-}CIg}$ 4:14, and $VPM_{S2\text{-}CIr}$ had 2:14 site-year sums at ±10% that of $GPP_{Tower}$. Overall, Sentinel-2 VPMs were closer to the study-period sums of $GPP_{Tower}$ than other models. Cumulative satellite GPP estimates by site-year had difference of ~9–800 g C $m^{-2}$ from $GPP_{Tower}$, with an average difference of 229.69 g C $m^{-2}$. Models that had a site within ±10% of $GPP_{Tower}$ in both 2018 and 2019 included $GPP_{VPM\text{-}LS8}$, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}CIr}$ for sites CRP-REF, CRP-C, and CRP-REF, respectively. Model $VPM_{S2\text{-}MTCI}$ remained within ±10% of $GPP_{Tower}$ more often than other models including by site-year and cumulative annual GPP during the study period. $GPP_{MODIS}$ and $VPM_{S2\text{-}NDRE1}$ underestimated all site-years, but other models overestimated occasionally, including $GPP_{VPM\text{-}LS8}$ (5:14), $GPP_{VPM\text{-}MODIS}$ (2:14),$GPP_{CONUS}$ (4:14), $VPM_{S2\text{-}NDRE2}$ (1:14), $VPM_{S2\text{-}MTCI}$ (5:14), $VPM_{S2\text{-}CIg}$ (2:14), and $VPM_{S2\text{-}CIr}$ (2:14).

Cumulative GPP for the peak growing season (June, July, and August) indicate that $VPM_{S2\text{-}NDRE2}$ and $VPM_{S2\text{-}MTCI}$ best matched $GPP_{Tower}$, with 8:14 site-years within ±10% tower sums (Table 4). Non-red-edge model $GPP_{VPM\text{-}S2}$ closely followed with 7:14 site-years. When estimated by $GPP_{VPM\text{-}LS8}$ and $GPP_{VPM\text{-}S2}$ in 2018 and by $GPP_{VPM\text{-}S2}$ in 2019, cumulative GPP of all sites in the study area was within ±10% of that estimated by $GPP_{Tower}$. When considering red-edge models, however, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}CIg}$, and $VPM_{S2\text{-}CIr}$ all estimated both 2018 and 2019 cumulative GPP within ±10% tower sums. However, $VPM_{S2\text{-}NDRE1}$, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}MTCI}$, $VPM_{S2\text{-}CIg}$, and $VPM_{S2\text{-}CIr}$ overestimated 1:14, 3:14, 9:14, 6:14, and 5:14 site-years, respectively. Compared to other models, $VPM_{S2\text{-}NDRE2}$ reliably estimated peak growing season cumulative GPP at individual and collective fields.

Table 3. March–November cumulative GPP (g C m$^{-2}$) as estimated from $GPP_{Tower}$, conventional products $GPP_{MODIS}$ and $GPP_{CONUS}$, and VPM models $GPP_{VPM\text{-}MODIS}$, $GPP_{VPM\text{-}LS8}$, $GPP_{VPM\text{-}S2}$, $VPM_{S2\text{-}NDRE1}$, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}MTCI}$, $VPM_{S2\text{-}Clg}$, $VPM_{S2\text{-}Clr}$. Values in bold indicate ±10% of total $GPP_{Tower}$.

| Year | SITE | $GPP_{Tower}$ | $GPP_{MODIS}$ | $GPP_{CONUS}$ | | $GPP_{VPM\text{-}}$ | | | | $VPM_{S2\text{-}}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *MODIS* | *LS8* | *S2* | *NDRE1* | *NDRE2* | *MTCI* | *Clg* | *Clr* |
| 2018 | AGR-C | 1598.83 | 1092.38 | 797.32 | **1649.99** | 1254.12 | 1340.96 | 1226.33 | **1577.21** | **1597.10** | 1369.89 | 1335.42 |
| | AGR-PR | 1554.91 | 1165.29 | **1649.75** | 1802.58 | **1612.87** | **1529.12** | 1124.71 | **1476.40** | 1786.66 | **1503.21** | 1345.71 |
| | AGR-SW | 1501.52 | 1108.52 | 1729.91 | 1241.52 | 1161.16 | 1276.53 | 822.97 | 1097.84 | **1522.17** | 1305.52 | 1218.87 |
| | CRP-C | 1417.42 | 1122.50 | 776.11 | 1206.72 | 1701.10 | 1184.19 | 1146.05 | **1484.45** | **1346.03** | **1340.52** | 1129.04 |
| | CRP-PR | 1469.99 | 1147.00 | 967.75 | **1349.28** | 1979.34 | **1439.21** | 1001.07 | 1289.68 | **1592.61** | 1270.01 | 1300.73 |
| | CRP-REF | 1327.21 | 1173.51 | 977.53 | 1088.91 | **1341.41** | 1122.34 | 741.01 | 1008.62 | **1425.93** | **1343.52** | **1216.46** |
| | CRP-SW | 1725.24 | 1171.74 | 1009.05 | 1152.18 | 1326.74 | 1464.69 | 1243.94 | 1519.22 | 1495.99 | 1443.48 | 1293.13 |
| | *Total* | *10595.12* | *7980.93* | *7907.42* | *9491.18* | ***10376.75*** | *9357.03* | *7306.07* | *9453.42* | ***10766.50*** | ***9576.15*** | *8839.37* |
| 2019 | AGR-C | 1340.88 | 1084.37 | **1331.38** | 1120.08 | 1075.59 | 1043.42 | 993.81 | **1242.04** | 1111.09 | 944.38 | 948.55 |
| | AGR-PR | 1465.36 | 1128.72 | 1717.36 | 1032.44 | 853.29 | 1013.46 | 794.34 | 1002.62 | 1109.12 | 1079.95 | 907.22 |
| | AGR-SW | 1366.86 | 1091.51 | 1838.63 | 795.20 | 1019.06 | 899.23 | 635.01 | 847.33 | 1141.78 | 968.39 | 942.40 |
| | CRP-C | 1596.14 | 1082.76 | **1437.37** | 1305.52 | 2031.73 | **1456.51** | **1446.59** | 1777.44 | 1888.23 | 1844.54 | 1773.22 |
| | CRP-PR | 1574.03 | 1077.00 | 975.83 | 1004.45 | 1233.13 | 1314.00 | 965.03 | 1225.00 | 1401.90 | 1407.05 | 1324.88 |
| | CRP-REF | 1265.02 | 1118.37 | 986.77 | 1010.13 | **1227.60** | **1328.91** | 846.67 | 1109.47 | 1420.46 | 1521.20 | **1357.44** |
| | CRP-SW | 1567.16 | 1128.18 | 1025.81 | 1257.41 | 1341.75 | **1453.05** | 1195.85 | **1488.37** | **1443.98** | **1436.62** | 1260.02 |
| | *Total* | *10175.45* | *7710.90* | ***9313.14*** | *7525.22* | *8782.16* | *8508.59* | *6877.31* | *8692.26* | ***9516.56*** | ***9202.13*** | *8513.73* |

Table 4. June–August cumulative GPP (g C m$^{-2}$) as estimated from $GPP_{Tower}$, conventional products $GPP_{MODIS}$ and $GPP_{CONUS}$, and VPM models $GPP_{VPM\text{-}MODIS}$, $GPP_{VPM\text{-}LS8}$, $GPP_{VPM\text{-}S2}$, $VPM_{S2\text{-}NDRE1}$, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}MTCI}$, $VPM_{S2\text{-}Clg}$, $VPM_{S2\text{-}Clr}$. Values in bold indicate ±10% of total $GPP_{Tower}$.

| Year | SITE | $GPP_{Tower}$ | $GPP_{MODIS}$ | $GPP_{CONUS}$ | $GPP_{VPM\text{-}}$ | | | $VPM_{S2\text{-}}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *MODIS* | *LS8* | *S2* | *NDRE1* | *NDRE2* | *MTCI* | *Clg* | *Clr* |
| 2018 | AGR-C | 1391.29 | 738.58 | 554.12 | **1275.75** | 1087.65 | 1131.24 | 1065.48 | **1311.99** | **1460.64** | 1225.44 | 1237.59 |
| | AGR-PR | 1184.43 | 724.16 | 1039.64 | **1272.96** | **1278.57** | **1261.86** | 902.85 | **1121.71** | 1505.76 | **1192.06** | **1173.15** |
| | AGR-SW | 1128.22 | 709.44 | **1099.14** | 908.67 | 892.78 | **1022.95** | 629.49 | 802.30 | **1221.38** | 993.42 | 983.67 |
| | CRP-C | 1209.67 | 722.41 | 539.37 | 880.98 | 1392.21 | 1005.68 | 998.37 | **1232.87** | **1231.34** | **1185.96** | 1042.77 |
| | CRP-PR | 904.32 | 732.37 | 540.12 | **985.65** | 1513.78 | 1117.23 | 807.45 | 1010.78 | 1293.86 | 1115.91 | 1144.93 |
| | CRP-REF | 729.02 | **763.76** | 534.06 | **776.52** | 952.55 | **785.15** | 520.92 | **685.25** | 986.70 | 928.80 | 876.51 |
| | CRP-SW | 1277.66 | 726.96 | 622.25 | 799.60 | 915.50 | **1211.05** | 1008.55 | **1191.43** | **1259.59** | **1280.42** | 1146.39 |
| | *Total* | *7824.61* | *5117.68* | *4928.71* | *6900.13* | ***8033.04*** | ***7535.16*** | *5933.11* | ***7356.32*** | *8959.27* | ***7922.01*** | ***7605.01*** |
| 2019 | AGR-C | 1054.59 | 721.68 | **975.09** | 816.75 | 904.75 | **959.19** | 844.38 | **1034.79** | **976.78** | 827.65 | 826.75 |
| | AGR-PR | 1166.85 | 711.91 | **1067.32** | 765.72 | 694.01 | 883.87 | 689.71 | 830.63 | 980.70 | 937.14 | 825.34 |
| | AGR-SW | 1043.18 | 710.13 | 1148.15 | 575.12 | 788.09 | 771.31 | 549.22 | 694.11 | **1008.74** | 874.31 | 851.93 |
| | CRP-C | 1198.79 | 677.43 | 1052.35 | 887.20 | 1542.26 | 1344.60 | **1218.03** | 1466.30 | 1630.16 | 1567.92 | 1519.85 |
| | CRP-PR | 1092.92 | 699.95 | 561.12 | 724.21 | 883.75 | **1047.35** | 780.58 | 943.20 | **1170.22** | **1143.78** | **1148.72** |
| | CRP-REF | 770.39 | **722.25** | 545.17 | **723.68** | 878.22 | 964.54 | 605.65 | **767.94** | 991.55 | 1054.68 | 985.61 |
| | CRP-SW | 1263.43 | 735.01 | 637.43 | 911.20 | 889.86 | **1226.98** | 1024.97 | **1214.15** | **1293.07** | **1245.41** | **1138.83** |
| | *Total* | *7590.15* | *4978.37* | *5986.63* | *5403.88* | *6580.95* | ***7197.84*** | *5712.54* | ***6951.12*** | ***8051.22*** | ***7650.90*** | ***7297.05*** |

In all site years, the finer resolution $GPP_{VPM\text{-}S2}$ and $GPP_{VPM\text{-}LS8}$ out-performed $GPP_{MODIS}$, $GPP_{CONUS}$ and $GPP_{VPM\text{-}MODIS}$ (Fig. 6, Table S5) and agreed the best with $GPP_{Tower}$. Each model had a significant ($p$<0.05) and strong positive trend with $GPP_{Tower}$ in 2018 and 2019. The largest variation in model estimates were found in corn systems for both years and prairie systems in 2018. $GPP_{VPM\text{-}MODIS}$, $GPP_{MODIS}$ and $GPP_{CONUS}$ models generally underestimated; and $GPP_{VPM\text{-}S2}$ and $GPP_{VPM\text{-}LS8}$ models aligned best with the 1:1 slope, with the exception of $GPP_{VPM\text{-}S2}$ and $GPP_{VPM\text{-}LS8}$ overestimation of CRP-C 2019 and CRP-PR 2018. In CRP-REF, all models were in close agreement with $GPP_{Tower}$. In both years, $GPP_{MODIS}$ and $GPP_{CONUS}$ had the highest RMSE in corn and switchgrass systems, as well as AGR-PR. In all sites, VPM models had lower RMSE than conventional products $GPP_{MODIS}$ and $GPP_{CONUS}$ with the exception of CRP-REF (both years) and CRP-PR (2018) (Fig. 8). Compared to $GPP_{VPM\text{-}S2}$, RMSE at corn sites was lower for $GPP_{VPM\text{-}LS8}$ for both years and lower for $GPP_{VPM\text{-}MODIS}$ in 3:4 site-years.

When considering enhancements from red-edge VIs in VPM, the NDRE1 and NDRE2 VIs increase explanation of variability in eight out of fourteen site-years (Fig. 7). While RMSE values of red-edge VPMs were often higher in 2018 than that of the EVI2-based $GPP_{VPM\text{-}S2}$, they were near equal in 2019 (Table S5, Fig.8). Sites that benefitted in both years from red-edge VPMs included AGR-C, CRP-C, and CRP-SW; whereas AGR-PR and AGR-SW only saw benefits in 2019. While both NDRE1 and NDRE2 improve AGR-C in both years and CRP-C in 2018 by the same explanation of variance, NDRE1 has a lower RMSE in all three site-years and a closer 1:1 slope in two of three site years.

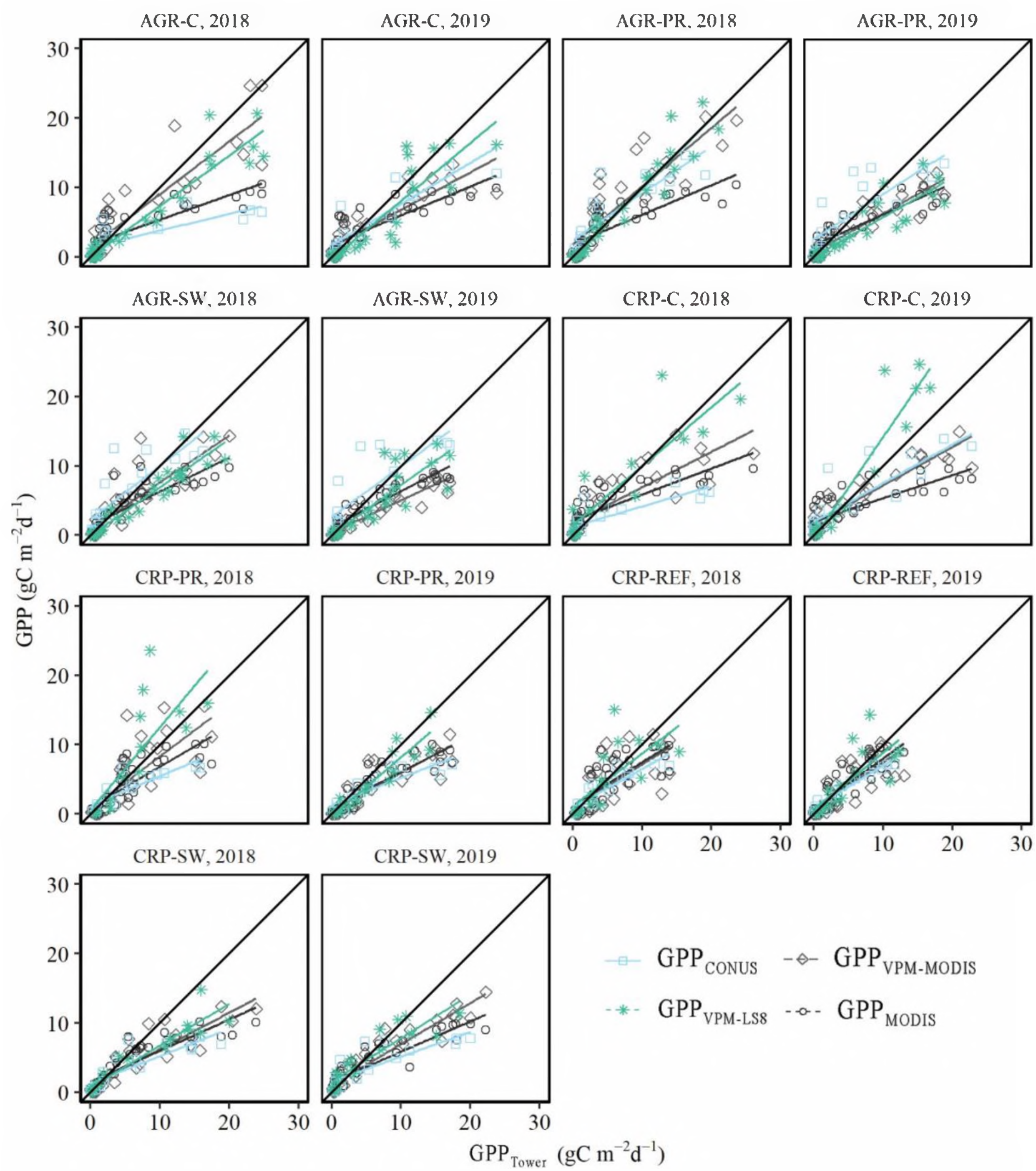


Figure 6. Comparison of daily GPP$_{tower}$ with daily $_{LS8}$, GPP$_{VPM-MODIS}$, GPP$_{MODIS}$, and GPP$_{CONUS}$ by site-year. Solid black linedepicts a 1:1 relationship.

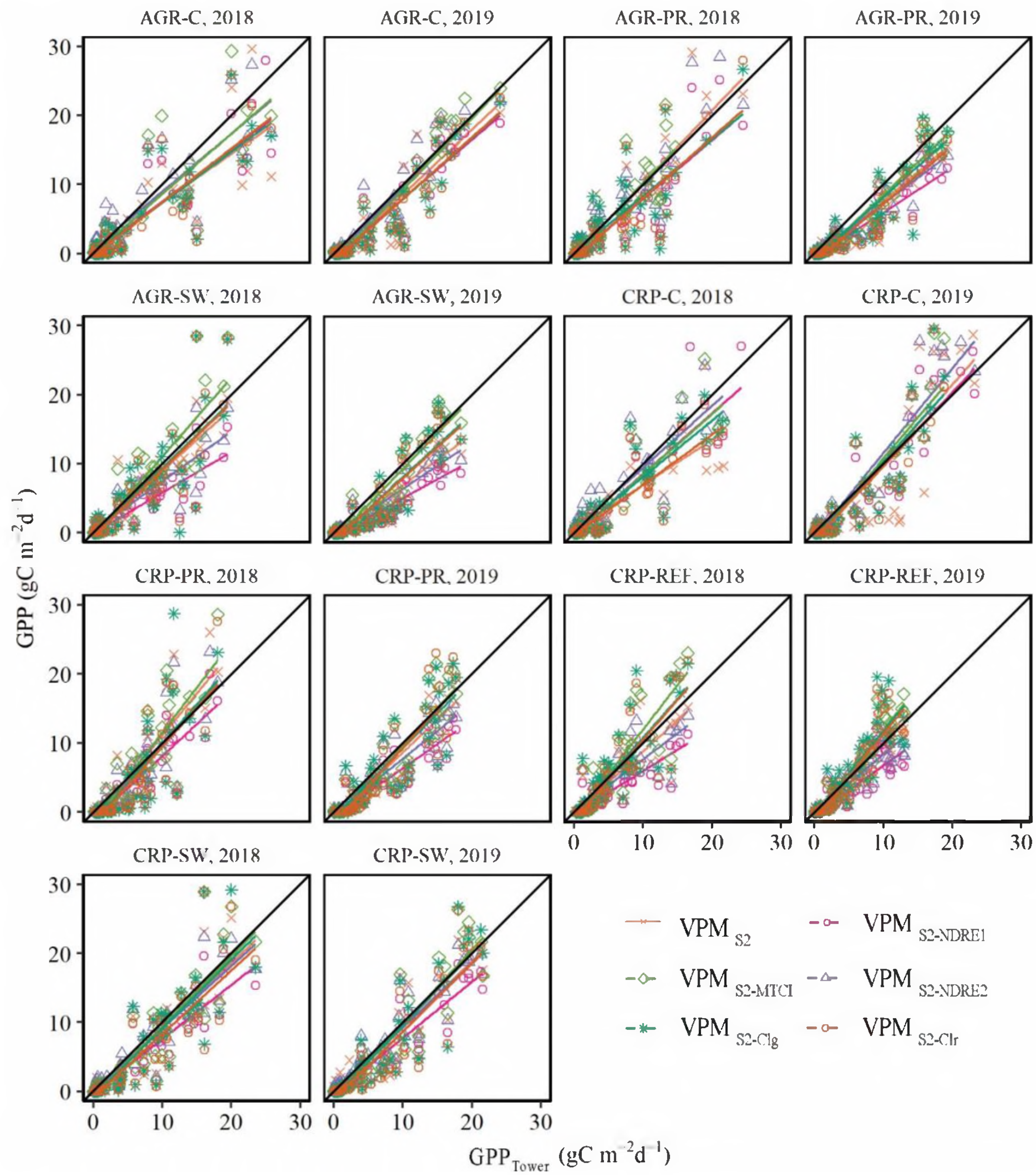


Figure 7. Comparison of daily $GPP_{tower}$ with daily $VPM_{S2\text{-}MTCI}$, $VPM_{S2\text{-}Clg}$, $VPM_{S2\text{-}Clr}$, $VPM_{S2\text{-}NDRE1}$, and $VPM_{S2\text{-}NDRE2}$ by site-year. Solid black line depicts a 1:1 relationship.

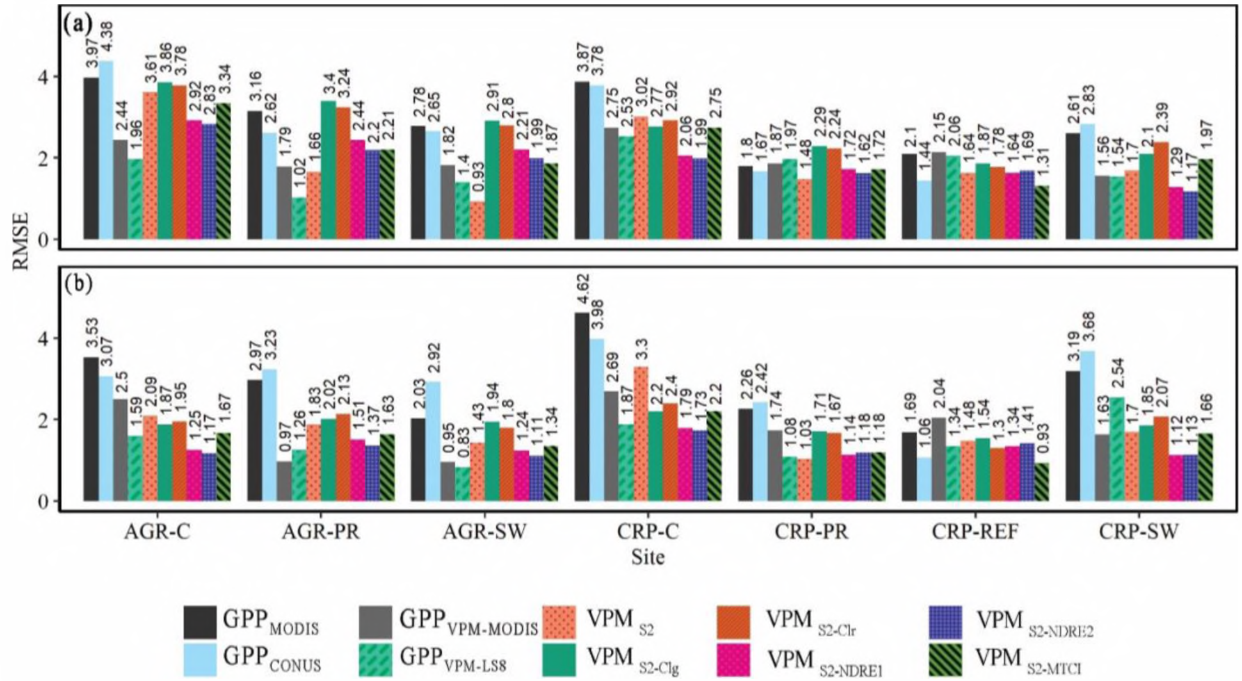


Figure 8. Comparison model RMSE (g C $m^{-2}$ $d^{-1}$) of daily $GPP_{Tower}$ with daily remote sensing GPP models across the seven land cover types in (a) 2018 and (b) 2019.

GPP estimates are significantly different between models at all sites, except CRP-PR, according to the Kruskal-Wallis rank sum test ($p < 0.05$) (Fig. 9). A pair-wise post-hoc Dunn test demonstrated that in site AGR-C, significant differences were found between pairs $GPP_{MODIS}$:$VPM_{S2-Clr}$ and $GPP_{MODIS}$:$VPM_{S2-Clg}$ ($z$= 3.92, $p$=0.004; $z$= 3.66, $p$=0.01, respectively); while CRP-C had differences between $GPP_{MODIS}$:$VPM_{S2-Clr}$ ($z$= 3.62, $p$=0.01). In sites CRP-PR, CRP-REF, CRP-SW, AGR-PR and AGR-SW, there were no significant ($p < 0.05$) differences between model pairs.

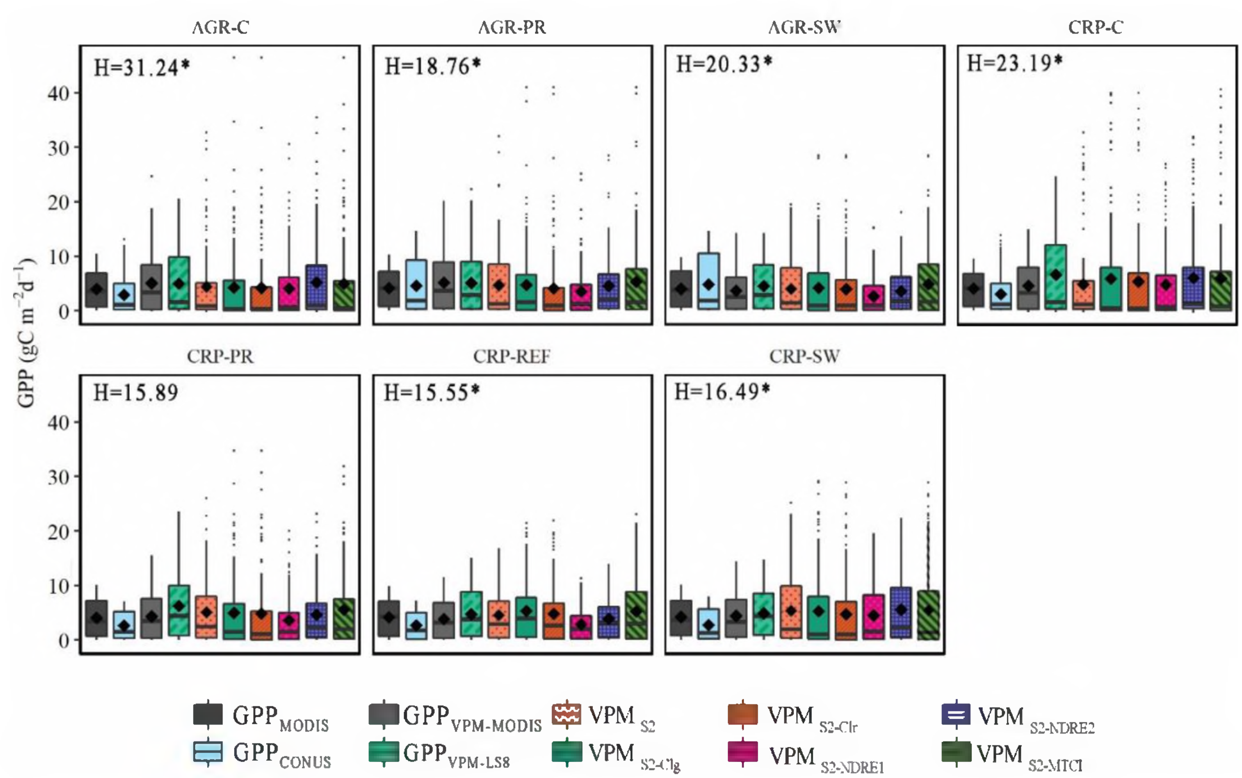


Figure 9. Box-plot comparisons of GPP models by land cover type during 2018–2019. Inside the boxplot, a black diamond indicates the mean, error bars are mean standard error, and a black horizontal line depicts the median; outside the boxplot, whiskers indicate the maximum and minimum values and points indicate outliers. Results of the Kruskal-Wallis include *H*, which is interpreted as chi-square. A significant *p*-value <0.05 is indicated with an *.

### *3.4. GPP anomaly estimates*

We evaluated anomalies generated from each GPP model from seasonal means and found large anomalies existed in the peak growing seasons (June-August) (Figs. 10,11). $GPP_{Tower}$ anomalies in regression analysis demonstrated that $GPP_{VPM\text{-}S2}$ exhibited the highest positive trend out of conventional models, with a significant relationship ($p$<0.05) in switchgrass and prairie systems but was second to $GPP_{VPM\text{-}LS8}$ at the corn systems. CRP-REF anomalies did not match well with any model, evidenced by insignificant, positive trends (Fig. S2, S3 Table S6). In red-edge VPMs, we found that most anomalies occurred during peak growing season due to models $VPM_{S2\text{-}MTCI}$, $VPM_{S2\text{-}Clr}$, and $VPM_{S2\text{-}Clg}$, which overestimated GPP in

2018 site-years and in CRP-C 2019. Generally, $VPM_{S2\text{-}NDRE1}$ and $VPM_{S2\text{-}NDRE2}$ did not overestimate, with the exception of CRP-C 2019, and had more outliers that underestimated.

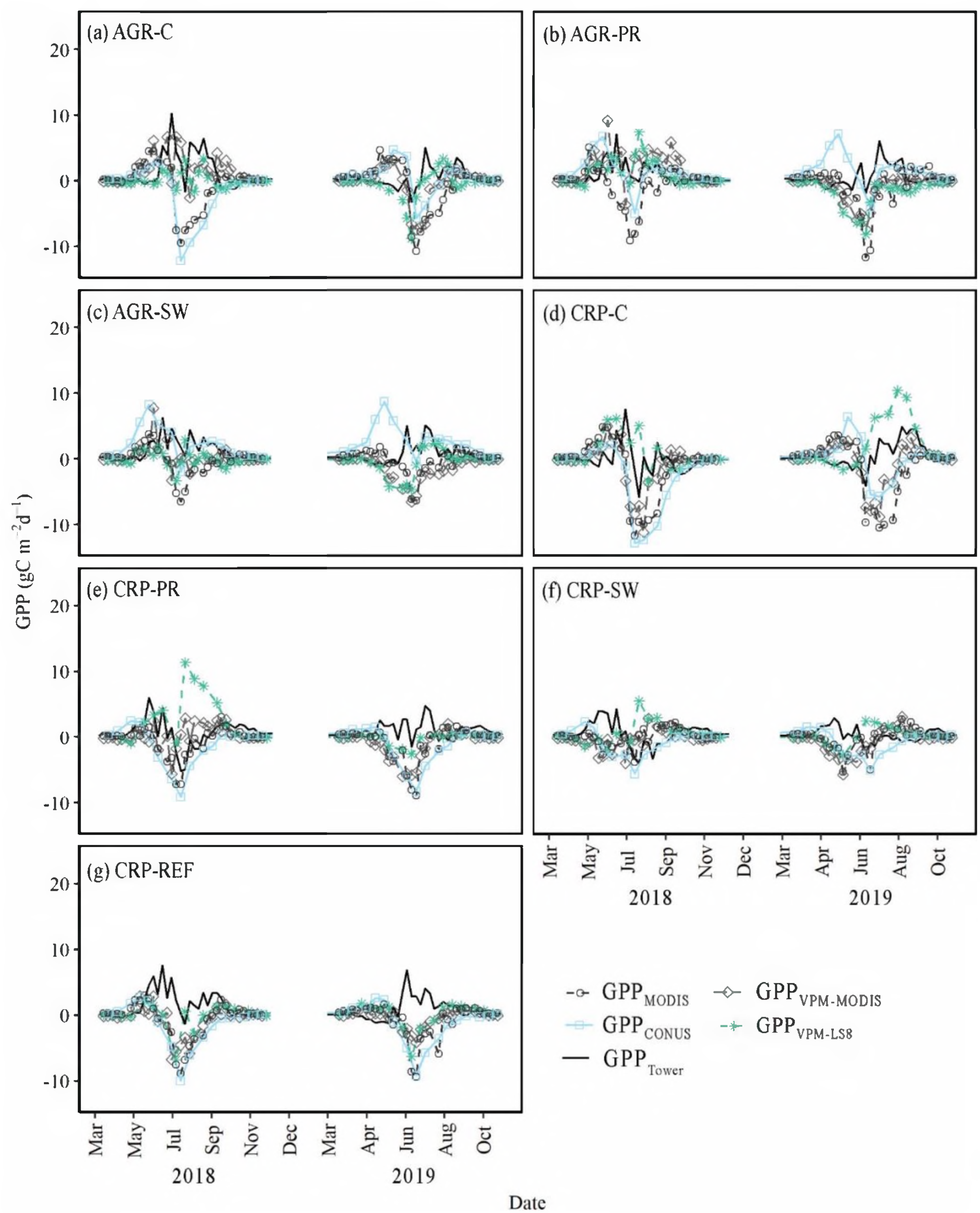


Figure 10. Anomalies of GPP (g C $m^{-2}$ $d^{-1}$) from $GPP_{MODIS}$, $GPP_{CONUS}$, $GPP_{VPM-MODIS}$ and $GPP_{VPM-LS8}$ over time for the seven study sites: (a) AGR-C, (b) AGR-PR, (c) AGR-SW, (d) CRP-C, (e) CRP-PR, (f) CRP-REF, and (g) CRP-SW.

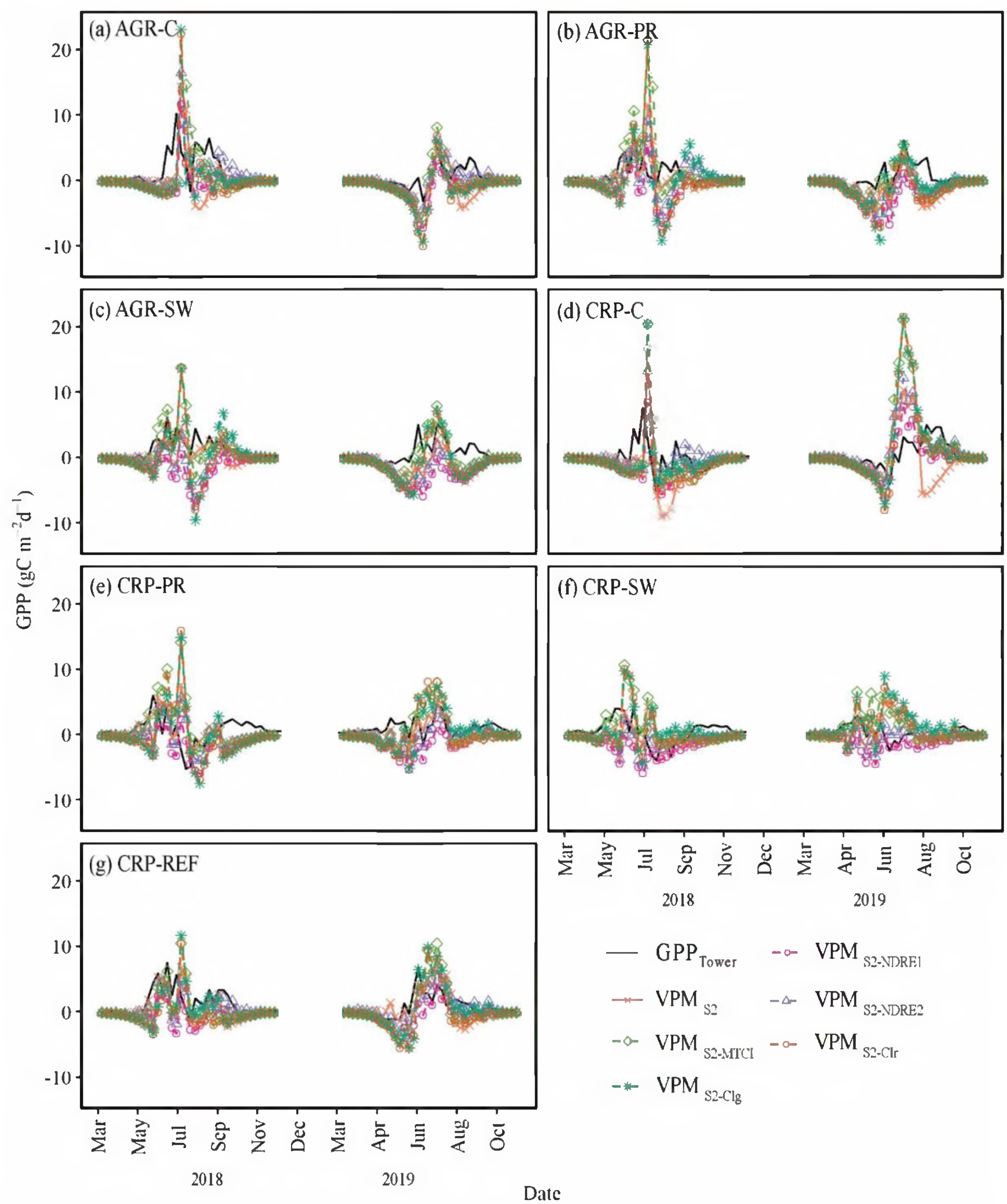


Figure 11. Anomalies of GPP (g C $m^{-2}$ $d^{-1}$) from $GPP_{VPM\text{-}S2}$, $VPM_{S2\text{-}Clg}$, $VPM_{S2\text{-}Clr}$, $VPM_{S2\text{-}NDRE1}$, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}MTCI}$ over time for the seven study sites: (a) AGR-C, (b) AGR-PR, (c) AGR-SW, (d) CRP-C, (e) CRP-PR, (f) CRP-REF, and (g) CRP-SW.

Anomalies exhibited both positive and negative trends compared to $GPP_{Tower}$, with fine-resolution VPMs outperforming conventional models. Red-edge VPMs had strong, positive trend at the exception of sites AGR-SW ($VPM_{S2\text{-}NDRE1}$) and CRP-REF ($VPM_{S2\text{-}Clg}$, $VPM_{S2\text{-}NDRE2}$). $GPP_{VPM\text{-}LS8}$ exhibited the strongest, positive trend and the lowest RMSE in corn sites between conventional models and $GPP_{VPM\text{-}S2}$; whereas $GPP_{VPM\text{-}S2}$ exhibited this for remaining sites, except CRP-REF (Fig. 12). In red-edge models, the lowest RMSE was $VPM_{S2\text{-}MTCI}$ in AGR-C, AGR-PR, and AGR-SW; and variable in remaining sites. Sites, AGR-C, CRP-C and CRP-SW tend to have higher RMSEs. Conventional $GPP_{CONUS}$ and $GPP_{MODIS}$ had negative trends, except for $GPP_{MODIS}$ in CRP-PR (ρ=0.30) and $GPP_{VPM\text{-}S2}$ in AGR-SW (ρ=0.02) and CRP-REF (ρ=0.40)) Similarly, $GPP_{VPM\text{-}MODIS}$ had a negative or zero trend in all sites except for AGR-PR (ρ=0.20).

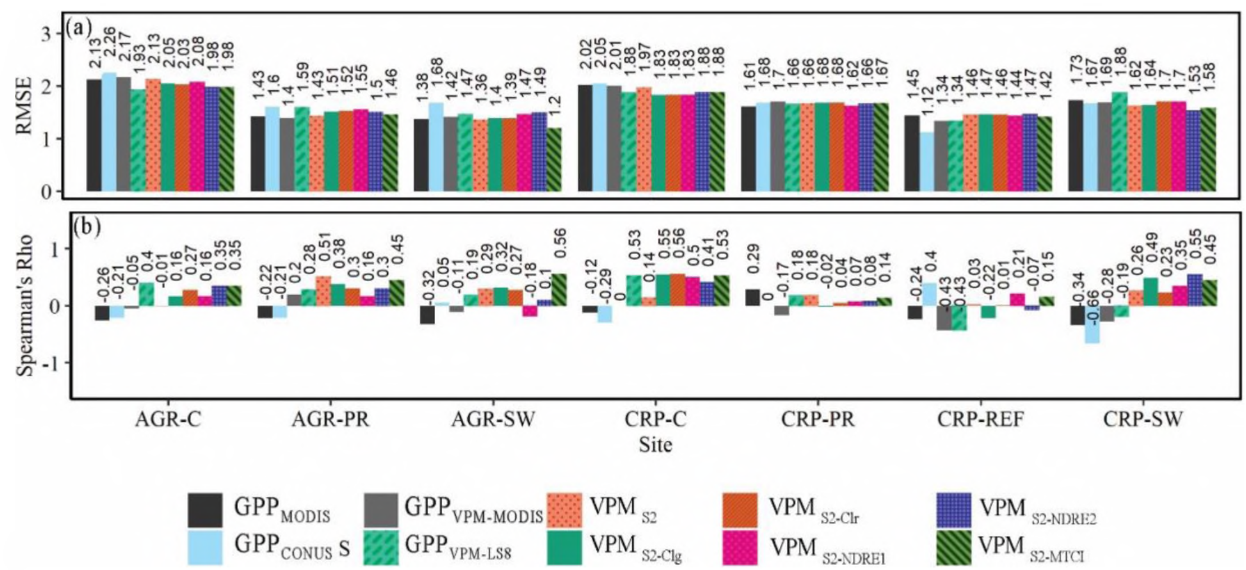


Figure 12. Comparison of anomaly model RMSE (g C $m^{-2}$ $d^{-1}$) and Spearman's Rho (ρ) coefficients of daily $GPP_{tower}$ with daily GPP from all remote sensing models across the seven land cover types 2018–2019.

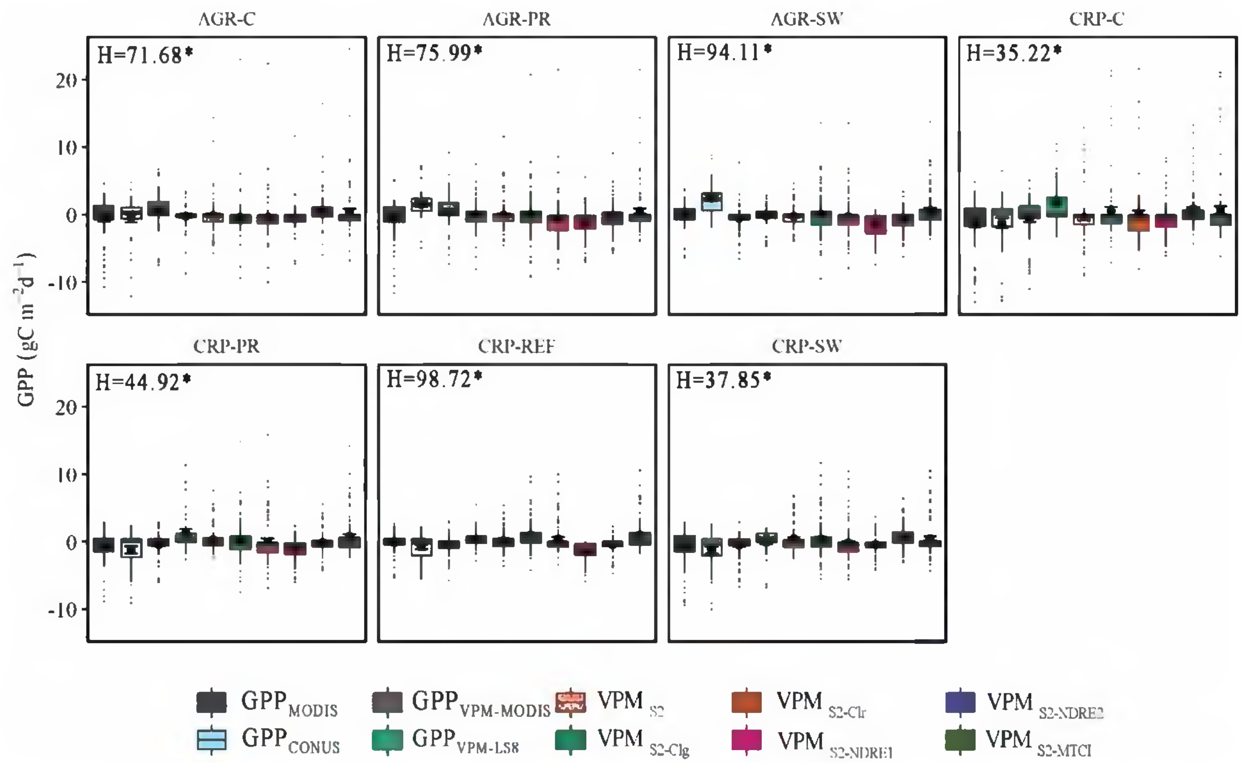


Figure 13. Box-plot comparisons of GPP (g C m$^{-2}$ d$^{-1}$) anomalies by model at seven land cover type during 2018–2019. Inside the boxplot, a black circle indicates the mean, error bars are mean standard error, and a black horizontal line depicts the median; outside the boxplot, whiskers indicate the maximum and minimum values and points indicate outliers. Results of the Kruskal-Wallis include *H*, which is interpreted as chi-square, and significance *p*-value <<0.05 is indicated with an asterisk (*).

Significant differences exist between anomaly GPP models at each site, according to the Kruskal-Wallis rank test (Fig. 13). The site with greatest variance from the mean was CRP-C. From the pairwise comparison Dunn test (Table 5), we also observed that a significant difference in anomaly medians between $GPP_{Tower}$ and $GPP_{VPM\text{-}S2}$ exist in five sites, including AGR-C, AGR-PR, AGR-SW, CRP-PR, and CRP-SW. Significant differences also existed between $GPP_{Tower}$ and $GPP_{VPM\text{-}LS8}$ at AGR-PR and AGR-SW, as well as between $GPP_{VPM\text{-}LS8}$ and $GPP_{VPM\text{-}S2}$ in CRP-C. The fewest differences between red-edge VPMs and $GPP_{Tower}$ were with $VPM_{S2\text{-}NDRE2}$ (AGR-PR, AGR-SW, CRP-PR, CRP-REF) and $VPM_{S2\text{-}MTCI}$ (AGR-C, AGR-SW, CRP-PR, CRP-REF); and the highest was with $VPM_{S2\text{-}Clr}$, which was significantly different in seven sites. $VPM_{S2\text{-}NDRE2}$ and $VPM_{S2\text{-}MTCI}$ also had the fewest differences between other models.

Table 5. Dunn test pairwise comparison of significant differences ($p$<0.05) between models at each site 2018–2019 for GPP anomalies.

| | GPP CONUS | GPP MODIS | GPP VPM-MODIS | GPP VPM-LS8 | GPP VPM-S2 | VPM S2-Clg | VPM S2-Clr | VPM S2-NDRE1 | VPM S2-NDRE2 | VPM S2-MTCI | GPP Tower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GPP CONUS | - | - | - | - | - | - | - | - | - | - | - |
| GPP MODIS | □□ | - | - | - | - | - | - | - | - | - | - |
| GPP VPM-MODIS | □□ | | - | - | - | - | - | - | - | - | - |
| GPP VPM-LS8 | ⁕□ | | | - | - | - | - | - | - | - | - |
| GPP VPM-S2 | ⁕□ | | ▲▲ | Δ | - | - | - | - | - | - | - |
| VPM S2-Clg | ⁕□ | ▲ | ▲♦ | | | - | - | - | - | - | - |
| VPM S2-Clr | ▲⁕□ | ▲⁕ | ▲⁕ | Δ■○ | | | - | - | - | - | - |
| VPM S2-NDRE1 | ⁕□♦ | ▲⁕□■ ♦ | ▲⁕□■ ♦ | □ Δ■♦ | □■♦ | □■♦ | ♦ | - | - | - | - |
| VPM S2-NDRE2 | ⁕□ | | | | | ▲♦ | ▲○ | ▲■♦○ | - | - | - |
| VPM S2-MTCI | ⁕□ | | ▲ | | | | | □■♦ | | - | - |
| GPP Tower | ■○■○ | ■♦○■ ♦○ | □■○□ ■○ | ⁕□ | ▲⁕□■ ○▲⁕□ ■○ | ▲⁕□■ ○ | ▲⁕□ Δ■♦○ | ▲⁕□■ ♦○ | ⁕□■♦ | ▲□■○ | - |

Sites: □: AGR-SW, ○: CRP-SW, ♦: CRP-REF, ■: CRP-PR, ⁕: AGR-PR, ▲: AGR-C, Δ: CRP-C

## 4. Discussion

While VPM developed using MODIS products still provides a valuable product that is widely available spatially and temporally, complex and heterogeneous land cover types such as managed agricultural-prairie landscapes benefit from the use of finer spatial resolution imagery (Chen et al., 2019). Fine spatial resolution reflectance indices from Sentinel-2 and Landsat-8 increased the accuracy of VPM models in our study. Particularly, when red-edge VIs replace EVI2 in Sentinel 2 VPMs, we found improvements in model validation, cumulative GPP estimates, and fewer differences between $GPP_{Tower}$ medians than that of $GPP_{VPM\text{-}S2}$.

Sensitivity of VIs EVI2 and NDVI to $GPP_{Tower}$ differed greatly between MODIS (500 m) and the finer resolutions of Landsat-8 (30 m) or Sentinel-2 (20 m). If selecting between the two in agricultural-prairie systems, it is prudent to use EVI2. For finer resolution VPMs, NDVI may be suitable upon further study. MODIS had high sensitivity to EVI2 in 13:14 site years than NDVI, of which only 2:14 site-years had sensitivity ±0.10 of 1.00 (i.e., near equal sensitivity). We find this supports similar research on MODIS LUE-based GPP models, where the ability to capture GPP variations is closely tied to the accuracy of *f*PAR and that 8-day MODIS data do not consistently capture fall and spring's rapid changes in phenology, likely introducing error to annual GPP estimates (Verma et al., 2014). Conversely, near-equal sensitivity was apparent in Landsat-8 and Sentinel-2, with 9:14 and 12:14 site-years with sensitivities ±0.10 of 1, respectively. Given EVI2 and NDVI uses the same two bands (i.e., NIR, Red), the differences between satellite products could arise from differences in radiometric resolution (i.e., bandwidth), spatial resolution and sampling frequency. In fact, the wavelength ranges of MODIS, Landsat-8, and Sentinel-2 red bands (nm) are 620–670, 636-673, 650-680, respectively; while the NIR bands are 841–875, 851-879, and 855-875, respectively. These slight differences in bandwidth, along with differences in sampling dates and spatial resolution from Landsat-8 and Sentinel-2, may have resulted in substantial differences in GPP estimates. We found that NDRE1 and NDRE2 were slightly more sensitive than EVI2 to $GPP_{Tower}$, with 8:14 site years, that MTCI was near-equal sensitive, and that EVI2 was generally more sensitive to $GPP_{Tower}$ than Clr and Clg. Both sensitivities of pairs (1) NDRE1 and NDRE2; and (2) Clr and Clg were similar, respectively, as the equations are similar and the difference within each pair is minimal (Eqs. 4-7).

GPP estimates in our study area, and many other Midwestern cropland regions, are notably underestimated by MODIS products, likely due to mixed pixels (Wang et al., 2015; Zhang et al., 2016). We found that land cover (NLCD, 30 m) within a single MODIS 500 m pixel overlapped cropland, developed areas, forests, grasslands and wetlands (Table S3). Our results demonstrated that $GPP_{VPM\text{-}MODIS}$ underestimated, particularly in the peak growing season, at all sites, more than other GPP models. The least to underestimate cumulative GPP includes $VPM_{S2\text{-}MTCI}$ during the study period (9:14) and peak growing season (5:14), and $VPM_{S2\text{-}CIg}$ in the peak growing season (8:14). When comparing conventional and non-red-edge VPMs, finer resolution VPM models are closer to daily and cumulative $GPP_{Tower}$, with $GPP_{VPM\text{-}LS8}$ capturing the variation in corn systems best and $GPP_{VPM\text{-}S2}$ best capturing grassland systems. Additionally, a heavy rainfall in the spring of 2019 (wet year) may have affected GPP production in some sites. Peak growing season (June-August) is also best reflected in $GPP_{VPM\text{-}S2}$ compared to other conventional GPP products and $GPP_{VPM\text{-}LS8}$. While over- and underestimation can interfere with scaled-up estimates (Jelinski & Wu, 1996) we found finer resolution (30 m and 20 m) GPP products demonstrated the capacity to improve GPP estimates across various corn and grassland systems.

Our anomaly analysis of covariance further enhanced our ability to evaluate interannual variation and identify significant differences between model estimates. In a similar study, covariance between interannual anomalies in MODIS products did not significantly correlate with $GPP_{Tower}$ in croplands; however, few MODIS products except VPM and MOD17A did explain substantial variance in grasslands because they include finer meteorological inputs and account for rapid development and senescence (Verma et al., 2014). Our results reflect this, as $GPP_{MODIS}$ and $GPP_{VPM\text{-}MODIS}$ did not significantly correlate with $GPP_{Tower}$ anomalies. We found significant differences in medians between $GPP_{VPM\text{-}S2,}$ $GPP_{VPM\text{-}LS8}$ and $GPP_{Tower}$ anomalies existed, indicating that one model simply over- or underestimated more often than its counterpart. While significant differences between medians in high-resolution and red-edge VPMs and $GPP_{Tower}$ exist, we do not believe this undermines their demonstrated accuracy in regression analysis and in seasonal summations. Particularly, anomalies of $GPP_{Tower}$ also have significant differences from $GPP_{MODIS}$ and $GPP_{VPM\text{-}MODIS}$ medians at three sites, and significant differences with $GPP_{VPM\text{-}LS8}$ and $GPP_{CONUS}$ at two sites; whereas it has significant differences with $VPM_{S2\text{-}NDRE2}$ and $VPM_{S2\text{-}MTCI}$ at four sites. Understanding that MODIS products largely underestimate GPP (Tables 3, 4) and aggregate nearby land covers, we recommend Landsat-8 and Sentinel-2 GPP products. More so, Sentinel-2 VPMs demonstrate greater ability than Landsat-8 products to remain within ±0.10% of both cumulative study period and peak growing season $GPP_{Tower}$; with red-edge $VPM_{S2\text{-}NDRE2}$ and $VPM_{S2\text{-}MTCI}$ equal to or out performing $GPP_{VPM\text{-}S2}$, respectively.

From both regression analyses in this study, $GPP_{VPM\text{-}LS8}$ still agreed strongest with corn systems compared to $GPP_{VPM\text{-}S2}$, which performed better in grassland systems with its largest anomalies during the peak growing season. However, when incorporating NDRE2 into the Sentinel-2 VPM, it could outperform $GPP_{VPM\text{-}LS8}$ in CRP-C site-years; demonstrating a potential to use red-edge VI with high-resolution imagery in both corn and grassland covers. The only site years where $GPP_{VPM\text{-}S2}$ still outperforms all other models, including red-edge VI VPMs, was in AGR-SW 2018 and in CRP-PR 2018 and 2019, where there are narrow differences (Table S5). We conclude that red-edge VIs, particularly NDRE2, may significantly improve the VPM's ability to estimate variations in GPP when used as an alternative to EVI2.

While our study area benefitted from finer resolution models, this may not stand true in all landscapes and elsewhere. In Nordic eddy covariance flux measurement sites, modelled GPP with linear regression and EVI2 and various environmental inputs detected a minimal difference with a consistent estimate across MODIS (500 m and 250 m) and Sentinel-2 (10 m) resolutions (Cai et al., 2021). An additional consideration for future studies is GPP production from cover crops, which is a common practice that may influence variability in annual estimates. Ultimately, the choice of GPP product depends on the intended application. Here, we advocate for fine-resolution imagery and the consideration of red-edge in GPP models to capture details at a local-scale that reflects land management and activities in heterogeneous cropland. However, Landsat provides data since 1972 and offers great historical detail far beyond what Sentinel 2 can offer, and may be more suitable for investigations of long-term change. Additionally, further consideration may be placed on temporal resolution, which imparts its own effect on aggregation of disturbance or land management useful for scaling investigations. Differences between Landsat and MODIS data lies in the acquisition and data retrieval, where Landsat is instantaneous and at higher risk of acquiring poor *f*PAR or LAI due to atmospheric effects and cloud cover and MODIS is a composite taking the best image from an 8-day span (Robinson et al., 2018). Future investigations on resolution and GPP estimates may consider utilizing the newly released MOD17A3HGF v061product, which may provide different results due to its updated protocol that cleans poor-quality inputs from 8-day LAI/fPAR based on pixel quality control labels. Additionally, the MODIS GPP product FluxSat v2.0 offers daily estimates of GPP using FLUXNET eddy covariance tower site data and coincident satellite data (Joiner & Yoshida, 2021).

While EC methods provide direct and suitable estimates of $CO_2$ fluxes at the local scale useful to both calibration and validation of remote sensing GPP models, we acknowledge they are also subject to error and uncertainty that are important to validation of remote sensing models and interannual analysis (Wang et al., 2015). Recent studies show that the flux tower footprint, used in validation and site-specific measurements, often extends beyond the target ecosystem, depending on time and atmospheric conditions (e.g., wind speed and direction) (Chu et al., 2021; Giannico et al., 2018). Consequently, in highly heterogeneous landscapes, multiple EC towers may be required to capture spatial representativeness necessary for validating global scale model grids (Wang et al., 2015). Our results support this, as $GPP_{MODIS}$ and $GPP_{VPM\text{-}MODIS}$ underestimated cumulative GPP as well as daily estimates during the study period and growing season (June, July, August).

Evaluation and monitoring of GPP with Landsat-8 and Sentinel-2 reveals how terrestrial C responds to land management, climate mitigation policies, and disturbance in heterogeneous cropland systems. It also supports cost-effective land management programs and increases the understanding of anthropogenic disturbances to ecosystem functions. Both Landsat-8 and Sentinel-2 are available freely online and easily accessible via Google Earth Engine, greatly improving their employability in policy and stakeholder programs. For example, the economic benefit of management and incentive programs attract farmers to convert low-producing corn for ethanol to perennial grasses, such as switchgrass, produce co-benefits, such as C sequestration (Kreig et al., 2021). Future applications with red-edge imagery from Sentinel 2 will benefit from high spatial and temporal resolution data, paving a way towards near real-time monitoring of GPP.

## 5. Conclusion

Fine-resolution (30 m and 20 m) satellite imagery and red-edge VIs integrated within VPM generally agree with daily and cumulative $GPP_{Tower}$ in field sites more so than coarse resolution imagery in VPM or conventional GPP products (e.g., $GPP_{MODIS}$ or $GPP_{CONUS}$) do. A substitution of a red-edge VI for EVI2 in the Sentinel 2 VPMs demonstrated improved explanations of variation and cumulative GPP estimates, compared to EVI2-based $GPP_{VPM\text{-}S2}$.

We found that vegetation indices of EVI2 and NDVI express different sensitivities by satellite origin, where MODIS-derived EVI2 had higher sensitivity than NDVI to $GPP_{Tower}$ in all but one site; and Landsat-8 and Sentinel-2 EVI2 and NDVI had near equal sensitivity in most site-years. Compared to EVI2, red-edge VIs NDRE1 and NDRE2 were slightly more sensitive to $GPP_{Tower}$. Seasonal GPP amplitude and growing season peaks are best captured by Sentinel-2 VPMs, followed by $GPP_{VPM\text{-}LS8}$, whereas conventional products underestimate growing season peaks. Overall, Sentinel-2 VPMs demonstrate greater ability than Landsat-8 and MODIS products to remain within ±0.10% of both cumulative study period and peak growing season $GPP_{Tower}$; with red-edge $VPM_{S2\text{-}NDRE2}$ and $VPM_{S2\text{-}MTCI}$ equal to or out performing $GPP_{VPM\text{-}S2}$, respectively. Red-edge Sentinel 2 VPMs collectively outperformed conventional GPP models and Landsat 8 products, when considering cumulative GPP estimates, model validations and significant differences between anomaly medians. We conclude that red-edge VIs, particularly NDRE2, may significantly improve our ability to estimate variations in GPP when used as an alternative to EVI2 in GPP models.

As many croplands are composed of areas less than 500 m, MODIS derived scalars may be composed of a mix of land cover types and therefore incorrectly estimate GPP. We demonstrated the capability of using $GPP_{VPM\text{-}LS8}$, $GPP_{VPM\text{-}S2}$ and red-edge $VPM_{S2\text{-}Clr}$, $VPM_{S2\text{-}Clg}$, $VPM_{S2\text{-}NDRE1}$, $VPM_{S2\text{-}NDRE2}$, $VPM_{S2\text{-}MTCI}$ in highly heterogeneous cropland, including corn, switchgrass, and restored prairie systems, in both historical cropland  and recently converted (i.e., 2009) CRP land. We found that our fine resolution GPP products (30 m and 20 m), and particularly red-edge Sentinel 2 VPMs, agreed best with $GPP_{Tower}$ and are significantly different than MODIS products in multiple cropland sites with differing land use history. While existing methods using MODIS-derived GPP models serve as an important baseline for studies with large spatial extents, future endeavors to estimate GPP in managed landscapes with greater frequency and improved accuracy are accessible and affordable at 30 m and 20 m resolutions.

**Acknowledgements**

Support for this research was provided by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Awards DE-SC0018409), by the National Science Foundation Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station, and by Michigan State University AgBioResearch. Authors are grateful for expert knowledge from Dr. Asko Normets, regarding light use efficiency and eddy covariance data; and Dr. Xiangming Xiao, regarding the vegetation photosynthesis model. Support for this research was made possible by members of the Landscape Ecology and Ecosystem Science Lab, particularly Terenzio Zenone and Michael Abraha who constructed and maintain the research sites. We also acknowledge the significant support and mentorship provided through the FLUXNET ECN community.

**Funding**

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1848739).

**References**

Abraha, M., Chen, J., Chu, H., Zenone, T., John, R., Su, Y. J., Hamilton, S. K. & Robertson, G. P. (2015). Evapotranspiration of annual and perennial biofuel crops in a variable climate. *Global Change Biology Bioenergy*, *7*(6), 1344–1356. https://doi.org/10.1111/gcbb.12239

Abraha, M., Gelfand, I., Hamilton, S. K., Chen, J. & Robertson, G. P. (2018). Legacy effects of land use on soil nitrous oxide emissions in annual crop and perennial grassland ecosystems. *Ecological Applications*, *28*(5), 1362–1369. https://doi.org/10.1002/eap.1745

Abraha, M., Gelfand, I., Hamilton, S. K., Chen, J. & Robertson, G. P. (2019). Carbon debt of field-scale conservation reserve program grasslands converted to annual and perennial bioenergy crops. *Environmental Research Letters*, *14*(2). https://doi.org/10.1088/1748-9326/aafc10

Abraha, M., Gelfand, I., Hamilton, S. K., Shao, C., Su, Y. J., Robertson, G. P. & Chen, J. (2016). Ecosystem water-use efficiency of annual corn and perennial grasslands: contributions from land-use history and species composition. *Ecosystems*, *19*(6), 1001–1012. https://doi.org/10.1007/s10021-016-9981-2

Abraha, M., Hamilton, S. K., Chen, J. & Robertson, G. P. (2018). Ecosystem carbon exchange on conversion of Conservation Reserve Program grasslands to annual and perennial cropping systems. *Agricultural and Forest Meteorology*, *253–254*(February), 151–160. https://doi.org/10.1016/j.agrformet.2018.02.016

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. *Reviews of Geophysics*, *53*(3), 785–818. https://doi.org/10.1002/2015RG000483

Aubinet, M., Vesala, T. & Papale, D. (Eds.). (2012). *Eddy covariance: a practical guide to measurement and data analysis* (1st ed.). Springer Dordrecht. https://doi.org/10.1007/978-94-007-2351-1

Baldocchi, D., Reichstein, M., Papale, D., Koteen, L., Vargas, R., Agarwal, D. & Cook, R. (2012). The role of trace gas flux networks in the biogeosciences. *Eos*, *93*(23), 217–218. https://doi.org/10.1029/2012EO230001

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T. & Lascano, R. J. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. *In Proceedings of the Fifth International Conference on Precision Agriculture*, (Vol. 1619, 6).

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, 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., … Papale, D. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Bhardwaj, A. K., Zenone, T., Jasrotia, P., Robertson, G. P., Chen, J. & Hamilton, S. K. (2011). Water and energy footprints of bioenergy crop production on marginal lands. *Global Change Biology Bioenergy*, *3*(3), 208–222. https://doi.org/10.1111/j.1757-1707.2010.01074.x

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-campen, H., Müller, C., Reichstein, M. & Smith, B. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, *13*(3), 679–706. https://doi.org/10.1111/j.1365-2486.2006.01305.x

Cai, Z., Junttila, S., Holst, J., Jin, H., Ardö, J., Ibrom, A., Peichl, M., Mölder, M., Jönsson, P., Rinne, J., Karamihalaki, M. & Eklundh, L. (2021). Modelling daily gross primary productivity with sentinel-2 data in the nordic region–comparison with data from modis. *Remote Sensing*, *13*(3), 1–18. https://doi.org/10.3390/rs13030469

Callahan, J., Casey, R., Templeton, M. & Sharer, G. (2020). *seismicRoll: Fast Rolling Functions for Seismology using "Rcpp."* R package version 1.1.4. https://cran.r-project.org/package=seismicRoll

Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., Black, T. A., Arain, M. A., Bourque, C. P. A., Flanagan, L. B. & Lafleur, P. . (2011). Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling. *Agricultural and Forest Meteorology*, *151*(1), 87–100.

Chen, J., Sciusco, P., Ouyang, Z., Zhang, R., Henebry, G. M., John, R. & Roy, D. P. (2019). Linear downscaling from MODIS to landsat: connecting landscape composition with ecosystem functions. *Landscape Ecology*, *34*(12), 2917–2934. https://doi.org/10.1007/s10980-019-00928-2

Chen, M., Griffis, T. J., Baker, J. M., Wood, J. D., Meyers, T. & Suyker, A. (2018). Comparing crop growth and carbon budgets simulated across AmeriFlux agricultural sites using the Community Land Model (CLM). *Agricultural and Forest Meteorology*, *256–257*, 315–333. https://doi.org/10.1016/J.AGRFORMET.2018.03.012

Chu, H., Baldocchi, D. D., John, R., Wolf, S. & Reichstein, M. (2017). Fluxes all of the time? A primer on the temporal representativeness of FLUXNET. *Journal of Geophysical Research: Biogeosciences*, *122*(2), 289–307. https://doi.org/10.1002/2016JG003576

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., Arkebauer, T. J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Bracho, R., Brown, S., … Zona, D. (2021). Representativeness of eddy-covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology*, *301–302*, 108350. https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108350

Clevers, J. G. & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, *23*, 344–351.

Damm, A., Guanter, L., Paul-Limoges, E., van der Tol, C., Hueni, A., Buchmann, N., Eugster, W., Ammann, C. & Schaepman, M. E. (2015). Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. *Remote Sensing of Environment*, *166*, 91–105.

https://doi.org/10.1016/j.rse.2015.06.004

Dash, J. & Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, *25–23*, 5403–5413. https://doi.org/10.1080/0143116042000274015

Dewitz, J. & Survey, U. S. G. (2021). *National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021): U.S. Geological Survey data release*. https://doi.org/10.5066/P9KZCM54

Dinno, A. (2017). *dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums*. R Package Version 1.3.5. https://cran.r-project.org/package=dunn.test

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, *6*(3), 241–252. https://doi.org/10.1080/00401706.1964.10490181

Field, C. B., Randerson, J. T. & Malmström, C. M. (1995). Global net primary production: Combining ecology and remote sensing. *Remote Sensing of Environment*, *51*(1), 74–88. https://doi.org/10.1016/0034-4257(94)00066-V

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N. & Snyder, P. K. (2005). Global consequences of land use. *Science*, *309*(5734), 570–574.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral properties of plants. *Applied Optics*, *4*(1), 11–20.

Gelybó, G., Barcza, Z., Kern, A. & Kljun, N. (2013). Effect of spatial heterogeneity on the validation of remote sensing based GPP estimations. *Agricultural and Forest Meteorology*, *174–175*, 43–53. https://doi.org/10.1016/j.agrformet.2013.02.003

Giannico, V., Chen, J., Shao, C., Ouyang, Z., John, R. & Lafortezza, R. (2018). Contributions of landscape heterogeneity within the footprint of eddy-covariance towers to flux measurements. *Agricultural and Forest Meteorology*, *260–261*, 144–153. https://doi.org/10.1016/j.agrformet.2018.06.004

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, *161*(2), 165–173.

Gitelson, A. A. & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra : algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, *148*(3–4), 494–500. https://doi.org/10.1016/S0176-1617(96)80284-7

Gitelson, A. A., Verma, S. B., Keydan, G., Leavitt, B., Arkebauer, T. J., Burba, G. G., Suyker, A. E., Rundquist, D. C., Keydan, G. & Leavitt, B. (2003). *Novel technique for remote estimation of CO 2 flux in maize*. *30*(9), 2–5. https://doi.org/10.1029/2002GL016543

Gitelson, A. A., Viña, A., Verma, S. B., Rundquist, D. C., Arkebauer, T. J., Keydan, G., Leavitt, B., Ciganda, V., Burba, G. G. & Suyker, A. E. (2006). Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity. *Journal of Geophysical Research Atmospheres*, *111*(8), 1–13. https://doi.org/10.1029/2005JD006017

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, *327*(5967), 812–818. https://doi.org/10.1126/science.1185383

Goetz, S. J., Prince, S. D., Goward, S. N., Thawley, M. M. & Small, J. (1999). Satellite remote sensing of primary production: An improved production efficiency modeling approach. *Ecological Modelling*, *122*(3), 239–255. https://doi.org/10.1016/S0304-3800(99)00140-4

Goetz, S. J., Prince, S. D., Small, J. & Gleason, A. C. R. (2000). Interannual variability of global terrestrial primary production: Results of a model driven with satellite observations. *Journal of Geophysical Research: Atmospheres*, *105*(D15), 20077–20091. https://doi.org/10.1029/2000JD900274

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

Haberl, H., Erb, K. H., Krausmann, F., Gaube, V., Bondeau, A., Plutzar, C., Gingrich, S., Lucht, W. & Fischer-Kowalski, M. (2007). Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(31), 12942–12947. https://doi.org/10.1073/pnas.0704243104

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 and Technology*, *55*(6), 3494–3504. https://doi.org/10.1021/acs.est.0c06421

Hibbard, K. A., Hoffman, F. M., Huntzinger, D. & West, T. O. (2017). Changes in Land Cover and Terrestrial Biogeochemistry. In  and T. K. M. Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart (Ed.), *Climate science special report: fourth national climate assessment, volume I* (pp. 277–302). U.S. Global Change Research Program. https://doi.org/10.7930/J0416V6X.U.S.

Horler, D. N. H., Dockray, M. & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, *4*(2), 273–288.

Houghton, R. A., House, J. I., Pongratz, J., Werf, G. R. Van Der, Defries, R. S., Hansen, M. C. & Qu, C. Le. (2012). *Carbon emissions from land use and land-cover change*. *4*, 5125–5142. https://doi.org/10.5194/bg-9-5125-2012

Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao, X. & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyra, L. R., Yang, W., Nemani, R. R. & Myneni, R. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, *33*(6), 2–5. https://doi.org/10.1029/2005GL025583

Jelinski, D. E. & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, *11*(3), 129–140.

Jiang, Z., Huete, A. R., Didan, K. & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, *112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

John, R., Chen, J., Noormets, A., Xiao, X., Xu, J., Lu, N. & Chen, S. (2013). Modelling gross primary production in semi-arid Inner Mongolia using MODIS imagery and eddy covariance data. *International Journal of Remote Sensing*, *34*(8), 2829–2857. https://doi.org/10.1080/01431161.2012.746483

Joiner, J. & Yoshida, Y. (2020). Satellite-based reflectances capture large fraction of variability in global gross primary production (GPP) at weekly time scales. *Agricultural and Forest Meteorology*, *291*, 108092. https://doi.org/10.1016/j.agrformet.2020.108092

Joiner, J. & Yoshida, Y. (2021). *Global MODIS and FLUXNET-derived daily gross primary production, V2*. ORNL DAAC. https://doi.org/doi.org/10.3334/ORNLDAAC/1835

Jung, M., Reichstein, M. & Bondeau, A. (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, *6*(10), 2001–2013. https://doi.org/10.5194/bg-6-2001-2009

Kreig, J. A. F., Parish, E. & Jager, H. I. (2021). Growing grasses in unprofitable areas of US Midwest croplands could increase species richness. *Biological Conservation*, *261*, 109289. https://doi.org/10.1016/j.biocon.2021.109289

Kruskal, W. H. & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Kumar, L. & Mutanga, O. (2017). Remote sensing of above-ground biomass. *Remote Sensing*, *9*(9), 1–8. https://doi.org/10.3390/rs9090935

Lasslop, G., Reichstein, M., Papale, D., Richardson, A., 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*(1), 187–

208. https://doi.org/10.1111/j.1365-2486.2009.02041.x
Lei, C., Abraha, M., Chen, J. & Su, Y. (2021). Long-term variability of root production in bioenergy crops from ingrowth core measurements. *Plant Ecology*, *14*(5), 757–770. https://doi.org/10.1093/jpe/rtab018
Li, F., Wang, X., Zhao, J., Zhang, X. & Zhao, Q. (2013). A method for estimating the gross primary production of alpine meadows using MODIS and climate data in China. *International Journal of Remote Sensing*, *34*(23), 8280–8300. https://doi.org/10.1080/01431161.2013.834394
Li, J. & Roy, D. P. (2017). A global analysis of Sentinel-2a, Sentinel-2b and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sensing*, *9*(9). https://doi.org/10.3390/rs9090902
Li, Z., Yu, G., Xiao, X., Li, Y., Zhao, X., Ren, C., Zhang, L. & Fu, Y. (2007). Modeling gross primary production of alpine ecosystems in the Tibetan Plateau using MODIS images and climate data. *Remote Sensing of Environment*, *107*(3), 510–519. https://doi.org/10.1016/j.rse.2006.10.003
Lin, S., Li, J., Liu, Q., Li, L., Zhao, J. & Yu, W. (2019). Evaluating the effectiveness of using vegetation indices based on red-edge reflectance from Sentinel-2 to estimate gross primary productivity. *Remote Sensing*, *11*(11). https://doi.org/10.3390/rs11111303
Michigan State Climatologist's Office. (2013). *Gull Lake (3504)*. http://climate.geo.msu.edu/climate_mi/stations/3504/. Accessed November 22, 2021.
Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, *9*(3), 747–766.
Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *281*(980), 277–294.
Morales, P., Sykes, M. T., Prentice, I. C., Smith, P., Smith, B., Bugmann, H., Zierl, B., Friedlingstein, P., Viovy, N., Sabaté, S., Sánchez, A., Pla, E., Gracia, C. A., Sitch, S., Arneth, A. & Ogee, J. (2005). Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. *Global Change Biology*, *11*(12), 2211–2233. https://doi.org/10.1111/j.1365-2486.2005.01036.x
Müller-Wilm, U., Devignot, O. & Pessiot, L. (2018). *S2 MPC: Sen2Cor configuration and user manual*. http://step.esa.int/thirdparties/sen2cor/2.9.0/docs/S2-PDGS-MPC-L2A-SUM-V2.9.0.pdf
Myneni, R. B. & Ross, J. (2012). *Photon-Vegetation interactions: applications in optical remote sensing and plant ecology*. Springer Science & Business Media.
Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T. & Yakir, D. (2006). Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty

estimation. *Biogeosciences*, *3*(4), 571–583. https://doi.org/10.5194/bg-3-571-2006

Piao, S., Ciais, P., Friedlingstein, P., De Noblet-Ducoudré, N., Cadule, P., Viovy, N. & Wang, T. (2009). Spatiotemporal patterns of terrestrial carbon cycle during the 20th century. *Global Biogeochemical Cycles*, *23*(4), 1–16. https://doi.org/10.1029/2008GB003339

Potapov, P., Turubanova, S., Hansen, M. C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q. & Cortez, J. (2022). Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nature Food*, *3*(1), 19–28. https://doi.org/10.1038/s43016-021-00429-z

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Raich, A. J. W., Rastetter, E. B., Melillo, J. M., Kicklighter, D. W., Steudler, P. A., Peterson, J., Grace, A. L., Iii, B. M. & Vörösmarty, C. J. (1991). Potential net primary productivity in South America: application of a global model. *Ecological Applications*, *1*(4), 399–429.

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*(GB1003). https://doi.org/10.1029/2007GB002952

Reeves, M. C., Zhao, M. & Running, S. W. (2005). Usefulness and limits on MODIS GPP for estimating wheat yield. *International Journal of Remote Sensing*, *26*(7), 1403–1421. https://doi.org/10.1080/01431160512331326567

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, 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

Robinson, N. P., Allred, B. W., Smith, W. K., Jones, M. O., Moreno, A., Erickson, T. A., Naugle, D. E. & Running, S. W. (2018). Terrestrial primary production for the conterminous United States derived from Landsat 30 m and MODIS 250 m. *Remote Sensing in Ecology and Conservation*, *4*(3), 264–280. https://doi.org/10.1002/rse2.74

Ruimy, A., Kergoat, L. & Bondeau, A. (1999). Comparing global models of terrestrial net primary productivity (NPP): Analysis of differences in light absorption and light-use efficiency. *Global Change Biology*, *5*(S1), 56–64. https://doi.org/10.1046/j.1365-2486.1999.00007.x

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), 547–560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Running, S. W. & Zhao, M. (2015). *Daily GPP and annual NPP (MOD17A2/A3) products NASA Earth Observing System MODIS land algorithm. MOD17 user's guide*. The Numerical Terradynamic Simulation Group. http://www.ntsg.umt.edu/sites/ntsg.umt.edu/files/modis/MOD17UsersGuide2015_v3.pdf

Sciusco, P., Chen, J., Abraha, M., Lei, C., Robertson, G. P., Zhang, R., Lafortezza, R., Shirkey, G., Ouyang, Z., Zhang, R. & John, R. (2020). Spatiotemporal variations of albedo due to land use: contributions to global warming impacts in managed agricultural landscapes. *Landscape Ecology*, *35*(6), 1385–1402. https://doi.org/10.1007/s10980-020-01022-8

Shao, C., Chen, J., Chu, H., Lafortezza, R., Dong, G., Abraha, M., Batkhishig, O., John, R., Ouyang, Z., Zhang, Y. & Qi, J. (2017). Grassland productivity and carbon sequestration in Mongolian grasslands: The underlying mechanisms and nomadic implications. *Environmental Research*, *159*. https://doi.org/10.1016/j.envres.2017.08.001

Sims, D. A. & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species , leaf structures and developmental stages. *Remote Sensing of Environment*, *81*(2–3), 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

Thoen, G. (1990). *Soil survey of Barry County, Michigan*. USDA Soil Conservation Service, Michigan State University Agricultural Experiment Station, and Michigan Technological University.

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*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Zhao, M., Running, S. W., Wofsy, S. C., Urbanski, S., Dunn, A. L. & Munger, J. W. (2003). Scaling Gross Primary Production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remote Sensing of Environment*, *88*(3), 256–270. https://doi.org/10.1016/j.rse.2003.06.005

United Nations. (2015). *Transforming our world: the 2030 agenda for sustainable development*. https://doi.org/10.1163/15718093-12341365

USGCRP. (2018). Impacts, risks, and adaptation in the United States, Volume II. In D. R. Reidmiller, C. W. Avery, K. E. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. Maycock & B. C. Stewart (Eds.), *Fourth National Climate Assessment: Vol. II*. U.S. Global Change Research Program. https://doi.org/10.7930/NCA4.2018

Ustin, S. L. & Middleton, E. M. (2021). Current and near-term advances in Earth observation for ecological applications. *Ecological Processes*, *10*(1), 1–57. https://doi.org/10.1186/s13717-020-00255-4

Vargha, A. & Delaney, H. D. (1998). The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, *23*(2), 170–192. https://doi.org/10.3102/10769986023002170

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*(8), 2185–2200. https://doi.org/10.5194/bg-11-2185-2014

Vermote, E., Roger, J. C. & Ray, J. P. (2015). MODIS surface reflectance user's guide: collection 6. In *NASA EOSDIS Land Processes DAAC*. http://modis-sr.ltdri.org

Wagle, P., Xiao, X. & 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*. https://doi.org/10.1016/j.isprsjprs.2014.10.009

Wang, J., Zhuang, J., Wang, W., Liu, S. & Xu, Z. (2015). Assessment of uncertainties in eddy covariance flux measurement based on intensive flux matrix of HiWATER-MUSOEXE. *IEEE Geoscience and Remote Sensing Letters*, *12*(2), 259–263. https://doi.org/10.1109/LGRS.2014.2334703

Wang, W., Dungan, J., Hashimoto, H., Michaelis, A. R., Milesi, C., Ichii, K. & Nemani, R. R. (2011). Diagnosing and assessing uncertainties of terrestrial ecosystem models in a multimodel ensemble experiment: 1. Primary production. *Global Change Biology*, *17*(3), 1350–1366.

Wang, Z., Xiao, X. & Yan, X. (2010). Modeling gross primary production of maize cropland and degraded grassland in northeastern China. *Agricultural and Forest Meteorology*, *150*(9), 1160–1167. https://doi.org/10.1016/j.agrformet.2010.04.015

Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y. & Guanter, L. (2019). Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. *Remote Sensing of Environment*, *225*, 441–457. https://doi.org/10.1016/j.rse.2019.03.002

Wu, C., Niu, Z., Tang, Q. & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, *148*(8–9), 1230–1241.

Wu, Chaoyang, Chen, J. M. & Huang, N. (2011). Predicting gross primary production from the enhanced vegetation index and photosynthetically active radiation: Evaluation and calibration. *Remote Sensing of Environment*, *115*(12), 3424–3435. https://doi.org/10.1016/j.rse.2011.08.006

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O. &

Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Xiao, J., Zhuang, Q., Law, B. E., Baldocchi, D. D., Chen, J., Richardson, A. D., Melillo, J. M., Davis, K. J., Hollinger, D. Y., Wharton, S., Oren, R., Noormets, A., Fischer, M. L., Verma, S. B., Cook, D. R., Sun, G., McNulty, S., Wofsy, S. C., Bolstad, P. V., … Torn, M. S. (2011). Assessing net ecosystem carbon exchange of U.S. terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations. *Agricultural and Forest Meteorology*, *151*(1), 60–69. https://doi.org/10.1016/j.agrformet.2010.09.002

Xiao, X., Hollinger, D., Aber, J., Goltz, M., Davidson, E. A., Qingyuan, Z. & Berrien, M. I. (2004). Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sensing of Environment*, *89*, 519–534. www.elsevier.codloeate/rse

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B. & Ojima, D. (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

Xiao, X., Zhang, Q., Saleska, S., Hutyra, L., De Camargo, P., Wofsy, S., Frolking, S., Boles, S., Keller, M. & Moore, B. (2005). Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment*, *94*(1), 105–122. https://doi.org/10.1016/j.rse.2004.08.015

Zabel, F., Delzeit, R., Schneider, J. M., Seppelt, R., Mauser, W. & Václavík, T. (2019). Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-10775-z

Zeileis, A. & Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, *14*(6), 1–27. https://doi.org/10.18637/jss.v014.i06

Zenone, T., Chen, J., Deal, M. W., Wilske, B., Jasrotia, P., Xu, J., Bhardwaj, A. K., Hamilton, S. K. & Philip Robertson, G. (2011). CO2 fluxes of transitional bioenergy crops: Effect of land conversion during the first year of cultivation. *Global Change Biology Bioenergy*, *3*(5), 401–412. https://doi.org/10.1111/j.1757-1707.2011.01098.x

Zhang, F., Chen, J. M., Chen, J., Gough, C. M., Martin, T. A. & Dragoni, D. (2012). Evaluating spatial and temporal patterns of MODIS GPP over the conterminous U.S. against flux measurements and a process model. *Remote Sensing of Environment*, *124*, 717–729. https://doi.org/10.1016/j.rse.2012.06.023

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. (2016). Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. *Remote Sensing of Environment*, *183*, 154–169. https://doi.org/10.1016/j.rse.2016.05.015

Zhao, Y., Ciais, P., Peylin, P., Viovy, N., Longdoz, B., Bonnefond, J. M., Rambal, S., Klumpp, K., Olioso, A., Cellier, P., Maignan, F., Eglin, T. & Calvet, J. C. (2012). How errors on meteorological variables impact simulated ecosystem fluxes: A case study for six French sites. *Biogeosciences*, *9*(7), 2537–2564. https://doi.org/10.5194/bg-9-2537-2012