

Lawrence Berkeley National Laboratory

LBL Publications

Title

Diverging climate response of corn yield and carbon use efficiency across the U.S.

Permalink

<https://escholarship.org/uc/item/0tw2k914>

Journal

Environmental Research Letters, 18(6)

ISSN

1748-9318

Authors

Yu, Shuo

Falco, Nicola

Patel, Nivedita

et al.

Publication Date

2023-06-01

DOI

10.1088/1748-9326/acd5e4

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

LETTER • **OPEN ACCESS**


Diverging climate response of corn yield and carbon use efficiency across the U.S.

To cite this article: Shuo Yu *et al* 2023 *Environ. Res. Lett.* **18** 064049

View the [article online](#) for updates and enhancements.

You may also like

- [CO₂ fertilization enhances vegetation productivity and reduces ecological drought in India](#)
Nivedita Dubey and Subimal Ghosh
- [Hearing random matrices and random waves](#)
M V Berry and Pragya Shukla
- [Self-averaging characteristics of spectral fluctuations](#)
Petr Braun and Fritz Haake



The Breath Biopsy® Guide
Fourth edition

FREE

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Diverging climate response of corn yield and carbon use efficiency across the U.S.

OPEN ACCESS

RECEIVED

4 November 2022

REVISED

2 May 2023

ACCEPTED FOR PUBLICATION

16 May 2023

PUBLISHED

9 June 2023

Shuo Yu^{1,*}, Nicola Falco² , Nivedita Patel³, Yuxin Wu² and Haruko Wainwright^{2,4}¹ Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA 94720, United States of America² Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America³ New York University, New York, NY 10012, United States of America⁴ Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America

* Author to whom any correspondence should be addressed.

E-mail: shuoy@berkeley.edu**Keywords:** carbon use efficiency, yield, climate change, machine learning, remote sensing

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Abstract**

In this paper, we developed an open-source package to analyze the overall trend and responses of both carbon use efficiency (CUE) and corn yield to climate factors for the contiguous United States. Our algorithm enables automatic retrieval of remote sensing data through the Google Earth Engine (GEE) and U.S. Department of Agriculture (USDA) agricultural production data at the county level through application programming interface (API). Firstly, we integrated satellite products of net primary productivity and gross primary productivity based on the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, and climatic variables from the European Centre for Medium-Range Weather Forecasts. Secondly, we calculated CUE and commonly used climate metrics. Thirdly, we investigated the spatial heterogeneity of these variables. We applied a random forest algorithm to identify the key climate drivers of CUE and crop yield, and estimated the responses of CUE and yield to climate variability using the spatial moving window regression across the U.S. Our results show that growing degree days (GDD) has the highest predictive power for both CUE and yield, while extreme degree days (EDD) is the least important explanatory variable. Moreover, we observed that in most areas of the U.S., yield increases or stays the same with higher GDD and precipitation. However, CUE decreases with higher GDD in the north and shows more mixed and fragmented interactions in the south. Notably, there are some exceptions where yield is negatively correlated with precipitation in the Missouri and Mississippi River Valleys. As global warming continues, we anticipate a decrease in CUE throughout the vast northern part of the country, despite the possibility of yield remaining stable or increasing.

1. Introduction

Atmospheric carbon dioxide concentrations increased from ~320 ppm in 1960th to over 420 ppm in 2022. (NOAA Global Monitoring Lab. 2022) This increase reflects only about half of the CO₂ emissions from human activities; the other half has been sequestered in the oceans and on land (Battin *et al* 2009). According to the World Bank collection of development indicators, while agricultural land has declined since 1949, it still accounted for 44% of the total land in the United States (U.S.) in 2018. Agriculture contributes to ~11% of total U.S. emission in 2020 (USDA 2020), and is an important way

for humans to affect carbon emissions by influencing land use, and developing low carbon intensity (CI) agronomic practices. What's more, the U.S. government has been actively advocating sustainable development in recent years. For example, the California Air Resources Board's low carbon fuel standard program adopted a life cycle analysis (LCA) technique to calculate the CI of biofuels, issuing credits to those that have lower CI than baseline gasoline or diesel. The program thus provides incentives to adopt low-CI agronomic practices, since LCA records the affiliated greenhouse gas emissions for both the production of feedstock and fuel conversion stages. However, such carbon management practices must

be considered in conjunction with agricultural productivity and yield, which is the primary source of profit for farmers.

The challenge has been that land–atmosphere carbon exchanges, as well as crop yields, are spatially and temporally heterogeneous, affected not only by agricultural practices but also by climate and other factors. Previously, studies have been conducted to quantify the variation in CI in agricultural activities (West and Marland 2002a, 2002b, Njakou Djomo and Ceulemans 2012, Liu *et al* 2020, Sun *et al* 2020). Liu *et al* (2020) conducted a study that presents the contrast in CI resulting from different land-management practices throughout the cradle-to-farm-gate activities. Most studies employ experimental data or adopt process-based simulation models, which fail to include regional and temporal climatic variabilities across the U.S. While some studies use simulation models accounting for climate, soil, and management heterogeneity (Ojeda *et al* 2022), we need to investigate the regional responses of yield and CUE climatic variability empirically. In addition, most studies do not account for carbon fluxes from ecosystem respiration, which is a large part of the carbon budget. We have a need to characterize the heterogeneity of both carbon fluxes and yield across the continental scale. In addition, significant interest exists in how these metrics would evolve in the future under the impact of climate change, since these metrics have a deep effect on farmers, consumers, and society at large.

There is extensive literature on the variability of carbon sequestration and carbon fluxes—i.e. atmospheric–land carbon exchanges—driven by climate variability, soil, and other factors (Dragoni *et al* 2011, Njakou Djomo and Ceulemans 2012, Keenan *et al* 2014, Yu *et al* 2021). In addition, carbon fluxes have been extensively monitored across the U.S. using remote sensing spaceborne imagery, such as data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the two NASA's satellites (Terra and Aqua). The ability to retrieve plant phenological properties and photosynthetic activity based on vegetation indices and light use efficiency models have allowed the continuous estimation of carbon fluxes in terms of net primary production (NPP) and gross primary production (GPP) at a continental scale (Xiao *et al* 2019, Della Nave *et al* 2022). Several papers have also investigated the relationships between carbon fluxes and climate using satellite data (Zhang *et al* 2009, Kwon and Larsen 2013, He *et al* 2018). He *et al* (2018) quantified the responses of carbon use efficiency (CUE), defined as the ratio of NPP to GPP, to climatic factors based on MODIS satellite data and five different process-based carbon cycle models. The authors concluded that MODIS CUE tends to decline more sharply in response to increasing temperature, especially in warm and dry conditions, compared to

other process-based models, and both the satellite-based dataset and results from process-based models remain relatively stable with enhanced precipitation. However, the study did not consider the potential differential effects from crop specific biological traits.

At the same time, the spatiotemporal variability of crop yield has been analyzed across the U.S. as a function of climate and other vectors. Some of these analyses adopt simulation models that account for climate, soil, and management heterogeneity and explain the mechanisms behind the phenomenon, but are limited by the scope of experimental data. It is also hard to define the right set of variables to include in these types of models (Wong and Asseng 2006, Lobell *et al* 2013, Rosenzweig *et al* 2014, Teixeira *et al* 2021, Ojeda *et al* 2021, Dokoohaki *et al* 2022; etc). For instance, Lobell *et al* (2013) construct the APSIM crop model to explain why extreme heat is more important than precipitation. Others carry out empirical analysis on aggregated data, but ignore field-level heterogeneity and do not account for different management, farming practices, soil quality, or other factors (Schlenker and Roberts 2009, Kawasaki and Shinsuke 2016, Kukal and Irmak 2018; etc). For example, Schlenker and Roberts (2009) have documented yields increase with temperature up to some specific threshold for each crop, above which the heat becomes harmful, with regressions of county-level average yields on functions of weather variables. Some agricultural economists also focus on farm profits or farmland value (Mendelsohn *et al* 1994, Schlenker *et al* 2006, Deschênes and Greenstone 2012, Ortiz-Bobea 2020). If land markets are operating properly, prices will reflect the present discounted value of land rents into the infinite future, which accounts for the full range of farmer adaptation. However, this approach may be susceptible to the omitted variables bias, owing to unmeasured characteristics such as soil quality and the option value to convert to a new use.

Our continental U.S. (CONUS) scale study focuses on investigating the co-variability among carbon fluxes, yield, and climate variability, across a range of weather, soil, and management conditions. In particular, we aim to quantify the spatiotemporal variability of the county-level crop yield and CUE, which is considered as the potential of carbon sequestration for cropland (He *et al* 2018). We focus on corn in this paper since corn is one of the most important staple foods and grown widely in the U.S. Production of corn in the U.S. accounts for $\frac{1}{3}$ of world production, with approximately 85.4 million acres of land devoted for corn harvesting in 2021. (USDA NASS 2017) We develop a machine-learning pipeline through the Google Earth Engine (GEE)—for automatic data retrieval of public remote sensing products to facilitate this large-scale data analysis study. We developed a Python package to

automatically retrieve the geospatial and remote sensing data layers (including satellite-based NPP and GPP), extract useful information, form different measurements, and produce map visualizations of the variables.

We then applied machine-learning methods to identify the key drivers for CUE and crop yield, as well as their responses to temperature and precipitation changes. Lastly, we analyzed a range of climate time series datasets including temperature, precipitation, and their products such as growing degree days (GDD). Based on the results, we categorized the counties according to correlations and explored the distribution of the relationships. The co-variability patterns among yield, CUE, and climatic factors could bolster our understanding of the effect of current and future climate change on agriculture and carbon budgets. Literature is lacking studies that focus on better understanding the relationship between CUE, which is directly related to CI, and crop yield, which is the main profit-making mechanism for farmers, at regional scales. Understanding such co-variability would fill this gap in literature. Moreover, by comparing the responses of CUE and yield to climate variables, we are able to show the areas that have a divergence in CUE and yield trend and thus should be focused by policymakers. These areas have the highest potential to benefit from conservative or regenerative agricultural practices, which should be subsidized and incentivized by the government.

2. Spatial data layers

Our county-level yield data are downloaded from the USDA National Agricultural Statistics Service (USDA NASS) Quick Stats Database. It spans most U.S. counties from 1950 to 2020. NASS collects data for county-level yields through the December Agricultural Survey and the County Agricultural Production Survey each year, where farm operators report acreage, yield, and production for crops such as corn, soybean, and hay via mail, telephone, personal interview, or electronic reporting. Data accuracy is ensured through strict guidelines, including double-checking of questionable results with operators by experienced NASS statisticians, followed by a review by the NASS Agricultural Statistics Board to check for consistency, adherence to publication standards, and accuracy, before the data are published. Minimum response requirements of at least 30 producers or 25% of harvested acreage in a county are necessary for data publication, and data from other counties in the same district may be withheld if one county lacks sufficient responses. The NASS county data are used to support various Risk Management Agency crop insurance programs and Farm Service Agency (FSA) farm support programs. (Johanns and Greg 2020) The unit of corn yield is bushels per acre, where 1 bushel is 0.0254 metric ton for corn in the U.S. In this paper,

we use 19 years of data, from 2001 to 2019, to match the NPP data described below. We downloaded the data through Python API.

GEE provides a catalog of satellite imagery and geospatial datasets that enable planetary-scale interpretation capacities. We utilized two publicly available datasets, namely ERA5 Daily aggregates (precipitation and temperature) and MODIS Gross Primary Production/Net Primary Production. The GEE API, along with the code editor, was used to download data by filtering the required regions and time.

The MODIS GPP CONUS products (Robinson *et al* 2018) were estimates using MODIS Surface Reflectance for CONUS and the MODIS NPP CONUS dataset. They are both calculated using the MOD17 algorithm. Primary Production products present a measure of the growth of terrestrial vegetation. The production is then determined by the computation of a daily net photosynthesis value, which is ultimately combined over an eight-day interval for a year. The spatial resolution of MODIS data is 250 meters.

The land-use data adopted in this paper were obtained from the Cropland Data Layer (CDL) of the USDA NASS. The CDL, referred to as 'census by satellite' by the USDA, aims to accurately geo-locate major program crops annually in the U.S. The program takes Resourcesat-1 AWiFS imagery, agriculture ground truth from Farm Service Agency and NASS June Ag Survey, nonagricultural ground truth from U.S. Geological Survey National Land Cover Database (USGS NLCD), and other ancillary data as inputs. It then categorizes each pixel to over 250 crops and other land uses. The spatial resolution of USDA CDL data is 30 meters. (It is publicly available through GEE or the CropScape ((USDA-NASS) 2020) platform.)

The European Centre for Medium-Range Weather Forecasts (ECMWF) produces global statistical weather predictions and additional data. Specifically, it provides the ERA5 Daily Aggregates dataset (Copernicus Climate Change Service (C3S) 2017), which is a fifth-generation climate reanalysis dataset produced by ECMWF along with the Copernicus Climate Change Service.

ERA5 DAILY presents aggregated values for each day for the following seven ERA5 climate parameters gridded at 1 km resolution: 2 m air temperature, 2 m dew point temperature, total precipitation, mean sea level pressure, surface pressure, 10 m u-component of wind, and 10 m v-component of wind. It also presents daily minimum and maximum air temperature at 2 m that have been calculated based on the hourly 2 m air temperature data.

3. Methodology

The framework and ML pipeline are shown in figure 1. The framework integrates the GEE for data retrieval and processing (the details described below),

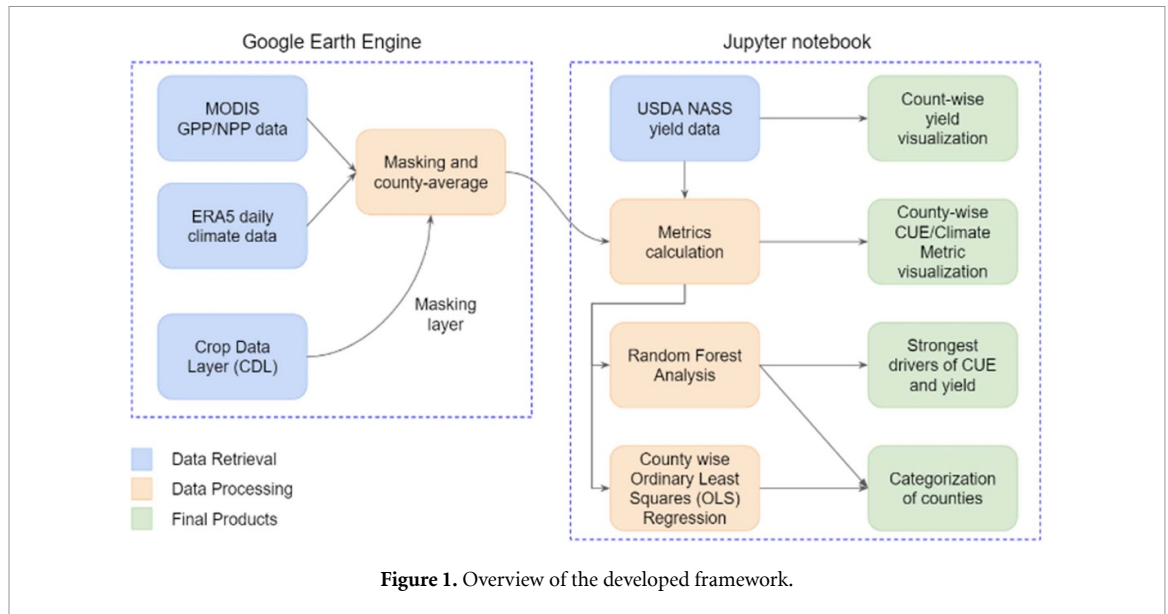


Figure 1. Overview of the developed framework.

and a series of Python-based codes as a workflow to perform the coupled analysis of climate, carbon fluxes, and yield.

3.1. Data retrieval: crop masking and county-level aggregation

First, we create the annual crop mask layer from CDL by assigning pixel values to 1 where the pixel value of the CDL image ‘cropland’ band is equal to the target crop and 0 otherwise. We then apply this mask layer to the variable images (e.g. GPP, NPP, the average temperature in each day, etc) through the `ee.Image.updateMask` function, so that the image values at the target crop pixels are kept, while other pixels are equal to NaN.

The algorithm then takes a collection of county-boundary polygons at the CONUS scale as input. Using these polygons, we apply a mean reducer (the `ee.Image.reduceRegions` function), which computes the average of non-NaN values in each masked image over the area of each county. This process results in the collection of county averages of each image for a selected crop.

Looping this masking and reducer step through all the image collections, we obtain the county averages of the variables for a selected crop, and store them in a feature collection. Finally, we download the collections as CSV files for easier integration with Jupyter notebooks.

3.2. Metric computation

Our metric computation focuses on crop yield and CUE. As a unit-free index, CUE can represent the efficiency of plants to sequester carbon from the atmosphere and is increasingly recognized as an important parameter shaping ecosystem carbon storage. We derive pixel-wise annual CUE data with MODIS GPP and NPP products, and then aggregate to county level by averaging data only on those grids labeled as corn.

The spatial heterogeneity of the corn CUE is shown in figure 2(a).

We use several climate metrics to depict local climate conditions that are important to agricultural production. GDD is used broadly to estimate the growth and development of plants. Following Schlenker and Roberts (2009), GDD is defined as follows

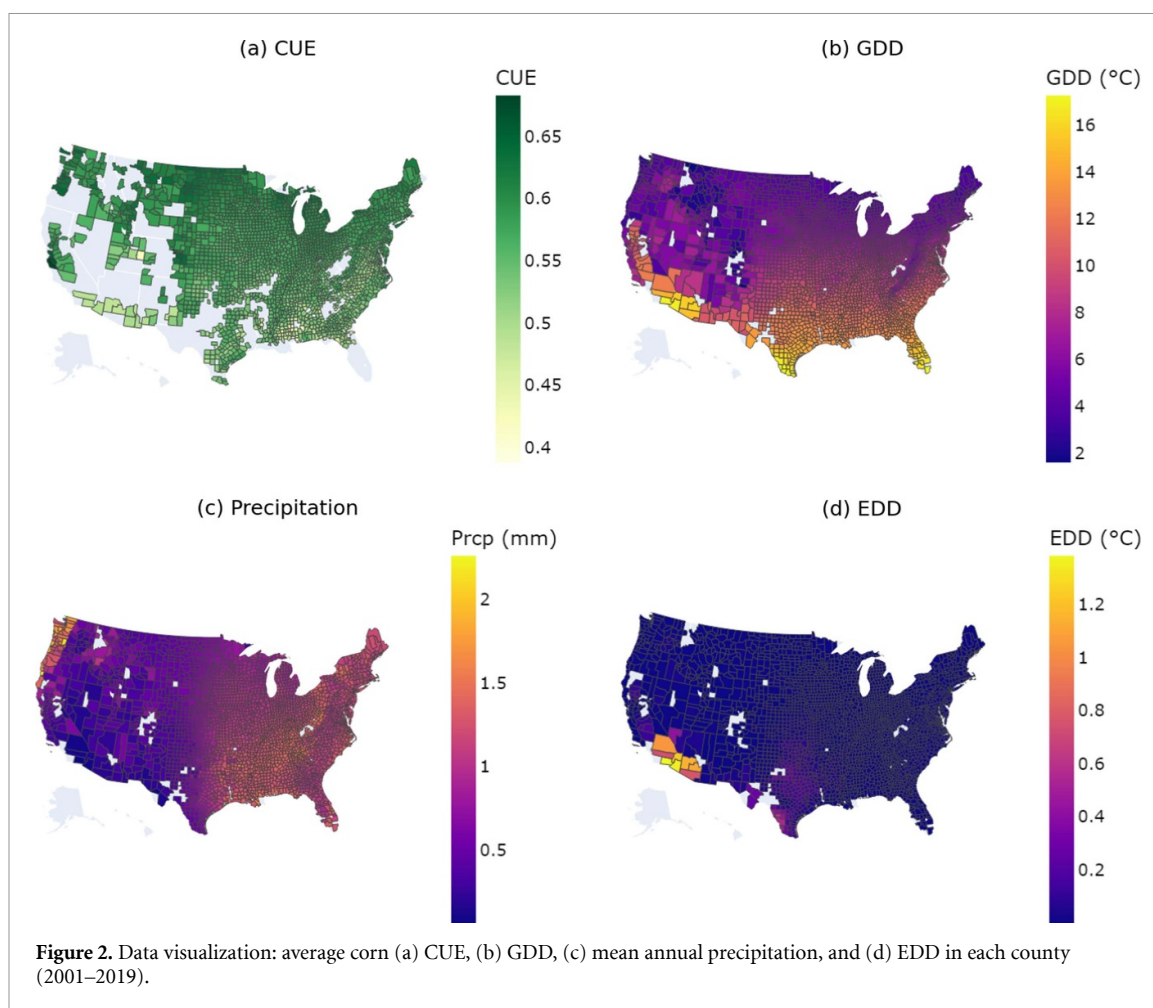
$$\text{GDD} = \frac{1}{T} \sum_{t=1}^T (\text{Temp}_t - \text{Temp}_{\text{base}}, 0),$$

where T is the number of days of the growing season, Temp_t is the average temperature on day t , and $\text{Temp}_{\text{base}}$ is a base temperature that might be different for each crop. The basic concept is that development will only occur if the temperature exceeds some minimum development threshold, or base temperature. This paper uses 8°C as the base temperature for both corn and soybeans (Schlenker *et al* 2006, Schlenker and Roberts 2009, Ortiz-Bobea 2020).

Another temperature metric we adopt is extreme degree days (EDD). EDD is broadly used to represent stress from extreme weather, since extreme temperatures higher than some thresholds are harmful to plant growth. Similar to GDD, it is defined as follows:

$$\text{EDD} = \frac{1}{T} \sum_{t=1}^T (\text{Temp}_t - \text{HighTemp}_{\text{base}}, 0),$$

where $\text{HighTemp}_{\text{base}}$ is the base temperature to define the extreme temperature, which is different for each crop. There are variations in the optimal temperature for maize growth (Schlenker and Roberts 2009, Wang *et al* 2018). We choose 30°C as base temperature in EDD calculation following Ortiz-Bobea’s (2020) approach to assess climate change’s impact on US agriculture profit using county-level data.



Another important prediction factor for agricultural production is water resources. Within our study, we use the countywide average daily precipitation for each year. We acknowledge that in the regions with one crop a year, using GDD, EDD, and precipitation metrics only during the growing season would improve the prediction accuracy of yields. We used annual metrics because in some regions, multiple growing seasons may occur within a year, making the use of annual measurements more appropriate. And for consistency, we used the same metrics for all the counties. Additionally, our interest in the carbon budget extends beyond the growing season, encompassing the entire year, including non-growing periods. As such, using annual measurements is considered to provide a more comprehensive understanding of the yield and CUE evolution in the context of climate change in our study.

3.3. Random forest (RF) regression

RF is a supervised learning algorithm used in classification or regression (Breiman 2001). RF is an ensemble method that first constructs several decision trees and outputs the average of the response from each tree as the final prediction.

In this study, recognizing the nonlinear relationship between yield/CUE and climate variables, we conduct a RF regression of yield and CUE, respectively, on the same set of variables—annual GDD, EDD, average precipitation, and the year of observation. We included years as a control variable in our model due to two reasons. First, there has been an upward trend in yield since the Green Revolution in the 1950s (Bobenrieth *et al* 2021). By including years as a control variable, we account for this technology trend in our analysis. Second, we included years to account for unobserved shocks that could impact the entire country at the same time, such as unpredicted events like pandemics. By controlling for the year, we can reduce the residual variance and increase the precision of the estimates. We did not include soil properties in our analysis, since it is highly variable within each county. We expected that if we have finer-scale yield data within each county (Falco *et al* 2021), it would be powerful to include the soil data into the random forest regression (Chaney *et al* 2016). One advantage of RF regression is that it can fully capture the nonlinearity of the different relationships. In our study, we made use of observations from all counties of CONUS from 2001 to 2019. The RF algorithm was

implemented using the scikit-learn library for Python language.

In our analysis, 60% of the data were randomly selected and used as the training set, with the remaining 40% used as the testing set. With the training set, we used ten-fold cross-validation to tune two parameters, the number of trees and the maximum depth of each tree. We used the optimal parameters and the training data to fit a random forest model, which was then used to predict our metrics over the testing data. The performance of the model was assessed in terms of R -squared (R^2) and root mean square error (RMSE).

In order to understand which feature has a higher impact on the estimated variable, we investigated the feature importance using the Gini importance index. Gini index is defined as $1 - \sum_{i=1}^C P_i^2$, where P_i is the probability of choosing an item with label i and C is the number of classes. It measures how often a randomly chosen element of a set would be incorrectly labeled if it was labeled randomly and independently according to the distribution of labels in the set. It reaches its minimum (zero) when all cases in the node fall into a single target category. Gini index is typically used as an impurity measurement in Classification and Regression Trees algorithms.

3.4. Spatial moving window regression

We employed the spatial moving window regression technique to estimate the responses of yield and CUE to climate variability. The spatial moving window increases the number of the data size, by considering one centered county and all counties bordering this county. We then fit an ordinary least squares (OLS) regression of yield and CUE, respectively, on the set of weather variables, including GDD, EDD, and annual average precipitation, within each window and recorded the resulting model parameters on the centered county. Counties with less than 4 bordering counties were excluded from the analysis to keep the size of the windows consistent. We counted on the spatial autocorrelation to produce localized results rather than a single, global regression model. This approach allowed us to better capture the spatial heterogeneity of the responses of yield and CUE to climatic factors, and to identify local patterns and trends that might be missed by a global analysis. With such regressions, we can detect whether the CUE and yield follow the same direction or diverge in each county, and therefore categorize each county based on the sign of correlation among variables. Alternative machine-learning methods (e.g. RF, gradient boosting, long short-term memory) for yield prediction, which allow for more variables than observations, focus more on forecasting precision rather than explanation of the phenomena. The effect of a single feature is intricate and hard to capture in those models. We adopted a simple OLS regression to obtain coefficients that are crucial for explanation. Further, the estimation should give us

consistent estimations, since weather variations are generally assumed to be exogenous. There is a weak positive correlation between GDD and precipitation, with a Pearson r coefficient at 0.226. Partial regression coefficients are essentially the same as those from multivariate regression according to the Frisch–Waugh theorem (Greene 2003). Therefore, the coefficient from multivariate moving window regression gives us the effect of GDD and precipitation on yield or CUE keeping the other factors constant, respectively. The correlation between GDD and precipitation should not affect the estimation of the corresponding effects on yield and CUE.

4. Results

4.1. Metric visualization

The metrics computed in our analysis are shown in figures 2(a)–(d). Figure 2(a) shows that the CUE is higher in the north than in the south as a whole, with the highest CUE in the north-central part. The CUE is relatively low on the southeast coast, but is prominently higher on the central west coast. Similar to the CUE layout but in the opposite direction, the average GDD gradually grows higher from north to south as in figure 2(b), with the highest on the southeast coast. Affected by westerly winds and ocean currents, the west coast is dominated by oceanic and Mediterranean climates, with mild weather and high GDD. Figure 2(c) displays that the wetter portions are in the northwest, which is characterized by the Pacific Coastal Ranges, the Willamette Valley, and the eastern part of the contiguous United States east of the 98th meridian. The drier areas are in the central-west regions, which include the Desert Southwest, Great Basin, valleys of northeast Arizona, eastern Utah, and central Wyoming. As for EDD, figure 2(d) shows that the extremely high temperature is concentrated in the Desert Southwest area.

4.2. Most predictive factors of CUE and yield

The RF regression results (figures 3(a-1) and (b-1)) show that the R -squared of the yield regression reaches 57.44% with an RMSE at 25.47, while the R -squared of the CUE regression achieves 55.52% with an RMSE at 0.03. These figures also show that the predicted values are more concentrated near the mean than the actual values. Especially for the low CUE points, their estimated values are obviously biased upwards, clustering around 0.5. This finding might result from low explanatory power below some threshold, calling for other variables not observed in our data to elucidate.

The most important factor in deciding both the yield and CUE based on the Gini importance index is GDD, according to figures 3(a-2) and (b-2). The determinants of yield are relatively dispersed, among which GDD occupies more than 30% of the importance, ranking first. Among the determinants of CUE,

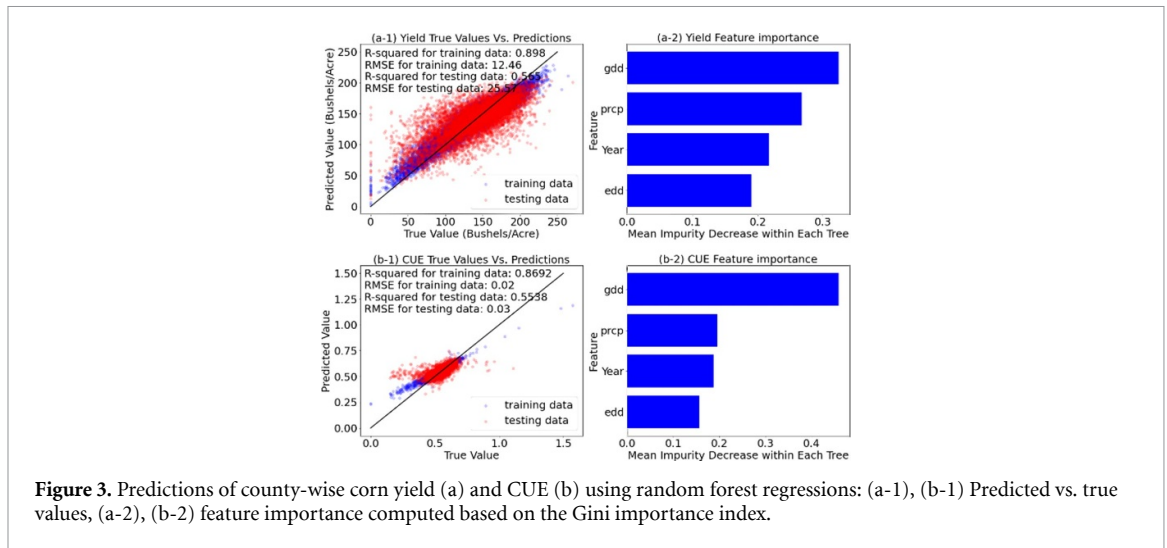


Figure 3. Predictions of county-wise corn yield (a) and CUE (b) using random forest regressions: (a-1), (b-1) Predicted vs. true values, (a-2), (b-2) feature importance computed based on the Gini importance index.

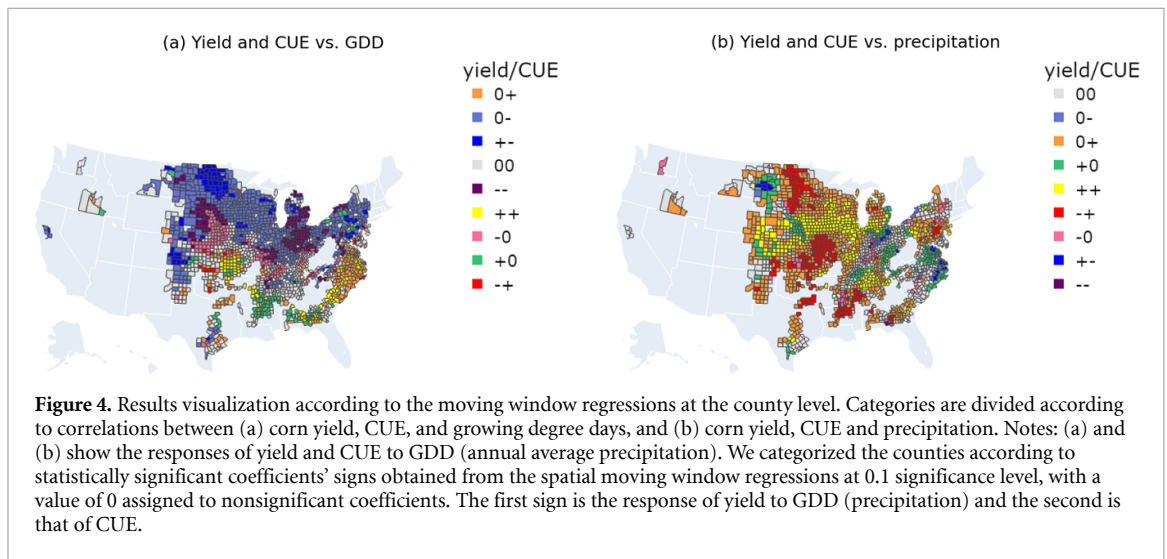


Figure 4. Results visualization according to the moving window regressions at the county level. Categories are divided according to correlations between (a) corn yield, CUE, and growing degree days, and (b) corn yield, CUE and precipitation. Notes: (a) and (b) show the responses of yield and CUE to GDD (annual average precipitation). We categorized the counties according to statistically significant coefficients' signs obtained from the spatial moving window regressions at 0.1 significance level, with a value of 0 assigned to nonsignificant coefficients. The first sign is the response of yield to GDD (precipitation) and the second is that of CUE.

GDD has a more obvious advantage, occupying a dominant position with more than 45% explanatory power. EDD is the least informative predictor for both CUE and yield.

4.3. Responses to climate variability

Figures 4(a) and (b) show the responses of yield and CUE to GDD and annual average precipitation respectively. We categorized the counties according to statistically significant coefficients' signs obtained from the spatial moving window regressions at 0.1 significance level, with a value of 0 assigned to nonsignificant coefficients.

In figure 4(a), almost the entire north is in areas of different shades of blue, where there are positive or no correlations between yield and GDD, but CUE is negatively correlated with GDD. However, in the southern part of the CONUS, especially in the southeast, the correlation between CUE and GDD becomes positive. Figure 4(b) shows that in the majority of the CONUS, CUE goes in the same direction as average precipitation. However, in the Middle and Lower

Mississippi Valley and the Missouri River Valley, the response of yield to precipitation is negative.

5. Discussion

Our objective is to establish a framework algorithm and provide an open-source package to analyze the overall trend and responses of both CUE and corn yield as a function of various climate factors. Our algorithm, taking advantage of automatic retrieval of publicly available data, allows for depiction of the co-variability among carbon fluxes, yield, and climate variability at the CONUS scale. Pixel masking of specific crops allows us to compare the yield and CUE—downloaded from different information sources—in a consistent manner. We use all data at the county level for consistency and comparability, since the yield data is publicly available at the county level. We recognize that employing environmental clusters instead of government divisions would reduce heterogeneity in environmental factors within each spatial unit, potentially improving

prediction accuracy. We anticipate that finer-scale yield data within each county would enable more robust regressions at the environmental cluster level.

Our RF regression reveals that GDD has the highest explanatory power for both CUE and yield. The order of feature importance for prediction of yield is the same as that for CUE.

Our county-level investigation illustrates that yield increases with greater GDD in 14.5% counties, with most counties staying the same, and increases with greater precipitation in 27.8% of the counties. The increases in yield are anticipated in the Northern part of the country. Higher GDD is related to the longer growing season, and higher precipitation will benefit crop growing by facilitating delivery of nutrients throughout the plant, especially in arid and semi-arid areas. The interactions are more mixed and fragmented in the south.

In the red, pink, and purple regions of figure 4(a), mainly in the mid-latitude regions, the yield decreases with increased GDD. These facts are partially consistent with the nonlinear relationships between yield and climate factors highlighted in Schlenker and Roberts (2009) and Roberts *et al* (2013). The regions in figure 4(b) where the correlation between yield and precipitation is negative are generally wet regions, such as the Middle and Lower Mississippi Valley. In areas with excessive precipitation and poor soil drainage, persisting saturation will cause large damage to crops. For corn, saturated soils inhibit root growth, leaf area expansion, and photosynthesis because of anoxic conditions and cooler soil temperatures. It also increases the incidence of moisture-loving diseases. This observation is consistent with, for example, Falco *et al* (2021).

The positive correlation between soil respiration and temperature is well documented (Lloyd and Taylor 1994, Eliasson *et al* 2005, Fang *et al* 2005, Hartley *et al* 2007). Additionally Steinweg *et al* (2008) suggested that the CUE of soil heterotrophs declines by approximately $0.009\text{ }^{\circ}\text{C}^{-1}$. Our analysis shows that CUE and GDD are negatively correlated in the vast northern areas (figure 4(a)), which is in line with the literature. However, according to our analysis, the relationship is reversed in the south. Hence, we conjecture that the correlation between CUE and temperature is U-shaped, changing from negative to positive over some threshold temperature (For example, Wetterstedt and Ågren 2011, Tucker *et al* 2013).

We also conjecture that management differences across regions have a great impact on yield. Yield is closely related to profit and is thus the main goal of agricultural production for farmers. Therefore, farmers tend to impose more management interventions to improve yields regardless of CUE. However, the lack of open data on irrigation practices (as well as fertilizer and pesticides application) have limited our

ability to incorporate farm management in our analysis. For instance, data on irrigated areas are only available for a few selected survey years on USDA NASS, and thus the variation is too sparse to be included in our time-series regressions. Human intervention has the potential to significantly affect carbon stock in the soil as well as yields (e.g. Paustian *et al* 2019). The absence of these variables may lead to omitted variable bias in our estimation. So, the correlation between yield and weather is relatively weak and versatile. After 2017, county-level data on cover crop and tillage practices is available (every five years) so that these datasets can be included in the future. With more information, future field level studies can make up for this lack of management detail.

Figure 4 implies that, with the progress of global warming, in the majority of the southern U.S., in the vast northern part of the country, CUE will go down, even though yield might stay the same or increase. In that case, the carbon emission problem will worsen, even though farmers may benefit from higher profits. We would recommend policy makers to design subsidies or regulations to incentivize farmers in these areas to take up conservative agricultural practices that help agricultural producers improve their environmental performance with respect to air quality and greenhouse gas emissions (e.g. nutrient management, conservation tillage, cover crops, and etc). It is important to note that in that case we should be wary of structural changes due to temperature rise and the increase in extreme weather. This work lays the path for several future research topics. In particular, by combining our results of CUE and yield responses with different scientific global warming scenarios predicted by climatologists, we can facilitate the prediction of climate change impacts on CUE and yield under different counterfactuals. As shown in sections 4 and 5, temperature plays a significant role in determining both research objectives. Further research featuring field-level data that include detailed climate and soil monitoring factors, as well as farm management variables, will be critical for capturing the causal effects between CUE, yield, climate, soil, and agricultural practices, so as to facilitate adaptation to climate change. Additionally, the inclusion of evapotranspiration would allow us to better understand the channel through which precipitation affects CUE and yield.

Data availability statement

The codes used to analyze the data are available in the Python package: py4openag (<https://pypi.org/project/py4openag/>).

The data that support the findings of this study and the codes used to analyze the data are openly available at the following URL/DOI: <https://github.com/shuoy528/erl-div-clim-resp.git>.

Acknowledgments

The information, data, or work presented herein was funded by the SMARTFARM (Systems for Monitoring and Analytics for Renewable Transportation Fuels from Agricultural Resources and Management) program of the Advanced Research Projects Agency – Energy (ARPA-E), U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORCID iD

Nicola Falco  <https://orcid.org/0000-0003-3307-6098>

References

- Battin T J, Luysaert S, Kaplan L A, Aufdenkampe A K, Richter A and Tranvik L J 2009 The boundless carbon cycle *Nat. Geosci.* **2** 598–600
- Bobenrieth E S, Bobenrieth J R, Guerra E A, Wright B D and Zeng D 2021 Putting the empirical commodity storage model back on track: crucial implications of a “negligible” trend *Am. J. Agric. Econ.* **103** 1034–57
- Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
- Chaney N W, Wood E F, McBratney A B, Hempel J W, Nauman T W, Brungard C W and Odgers N P 2016 POLARIS: a 30-meter probabilistic soil series map of the contiguous United States *Geoderma* **274** 54–67
- Copernicus Climate Change Service (C3S) 2017 ERA5: fifth generation of ECMWF atmospheric reanalyses of the global climate *Copernicus Climate Change Service Climate Data Store (CDS)* (available at: <https://cds.climate.copernicus.eu/cdsapp#!/home>)
- Della Nave F N, Ojeda J J, Irisarri J G N, Pembleton K, Oyarzabal M and Oesterheld M 2022 Calibrating APSIM for forage sorghum using remote sensing and field data under sub-optimal growth conditions *Agric. Syst.* **201** 103459
- Deschênes O and Greenstone M 2012 The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: reply *Am. Econ. Rev.* **102** 3761–73
- Dokoohaki H, Rai T, Kivi M, Lewis P, Gómez-Dans J L and Yin F 2022 Linking remote sensing with APSIM through emulation and Bayesian optimization to improve yield prediction *Remote Sens.* **14** 5389
- Dragoni D, Schmid H P, Wayson C A, Potter H, Grimmond C S B and Randolph J C 2011 Evidence of increased net ecosystem productivity associated with a longer vegetated season in a deciduous forest in south-central Indiana, USA *Glob. Change Biol.* **17** 886–97
- Eliasson P E, McMurtrie R E, Pepper D A, Stromgren M, Linder S and Agren G I 2005 The response of heterotrophic CO₂ flux to soil warming *Global Change Biol.* **11** 167–81
- Falco N et al 2021 Influence of soil heterogeneity on soybean plant development and crop yield evaluated using time-series of UAV and ground-based geophysical imagery *Sci. Rep.* **11** 1–17
- Fang C, Smith P, Moncrieff J B and Smith J U 2005 Similar response of labile and resistant soil organic matter pools to changes in temperature *Nature* **433** 57–59
- Greene W H 2003 *Least Squares Econometric Analysis* 5th edn (Upper Saddle River, NJ: Prentice Hall)
- Hartley I P, Heinemeyer A and Ineson P 2007 Effects of three years of soil warming and shading on the rate of soil respiration: substrate availability and not thermal acclimation mediates observed response *Global Change Biol.* **13** 1761–70
- He Y, Piao S, Li X, Chen A and Qin D 2018 Global patterns of vegetation carbon use efficiency and their climate drivers deduced from MODIS satellite data and process-based models *Agric. For. Meteorol.* **256** 150–8
- Johanns A and Greg T 2020 Collection of county yield data, how does NASS do it? (available at: www.extension.iastate.edu/agdm/newsletters/nl2020/mar20.pdf)
- Kawasaki K and Shinsuke U 2016 Quality matters more than quantity: asymmetric temperature effects on crop yield and quality grade *Am. J. Agric. Econ.* **98** 1195–209
- Keenan T F et al 2014 Net carbon uptake has increased through warming-induced changes in temperate forest phenology *Nat. Clim. Change* **4** 598–604
- Kukul M S and Irmak S 2018 US agro-climate in 20th century: growing degree days, first and last frost, growing season length, and impacts on crop yields *Sci. Rep.* **8** 1–14
- Kwon Y and Larsen C P 2013 Effects of forest type and environmental factors on forest carbon use efficiency assessed using MODIS and FIA data across the eastern USA *Int. J. Remote Sens.* **34** 8425–48
- Liu X, Kwon H, Northrup D and Wang M 2020 Shifting agricultural practices to produce sustainable, low carbon intensity feedstocks for biofuel production *Environ. Res. Lett.* **15** 084014
- Lloyd J and Taylor J A 1994 On the temperature dependence of soil respiration *Funct. Ecol.* **8** 315
- Lobell D B, Hammer G L, McLean G, Messina C, Roberts M J and Schlenker W 2013 The critical role of extreme heat for maize production in the United States *Nat. Clim. Change* **3** 497–501
- Mendelsohn R, Nordhaus W D and Shaw D 1994 The impact of global warming on agriculture: a Ricardian analysis *Am. Econ. Rev.* pp 753–71
- National Agricultural Statistics Service 2017 NASS - Quick Stats National Agricultural Statistics Service (available at: <https://data.nal.usda.gov/dataset/nass-quick-stats>) (Accessed September 2022)
- Njakou Djomo S and Ceulemans R 2012 A comparative analysis of the carbon intensity of biofuels caused by land use changes *Gcb Bioenergy* **4** 392–407
- NOAA Global Monitoring Laboratory 2022 Trends in atmospheric carbon dioxide (available at: <https://gml.noaa.gov/ccgg/trends/>)
- Ojeda J J et al 2022 Quantifying the effects of varietal types × management on the spatial variability of sorghum biomass across US environments *GCB Bioenergy* **14** 411–33
- Ojeda J J, Huth N, Holzworth D, Raymundo R, Zyskowski R F, Sinton S M, Michel A J and Brown H E 2021 Assessing errors during simulation configuration in crop models—a global case study using APSIM-Potato *Ecol. Modelling* **458** 109703
- Ortiz-Bobea A 2020 The role of nonfarm influences in Ricardian estimates of climate change impacts on US agriculture *Am. J. Agric. Econ.* **102** 934–59
- Paustian K, Collins H P and Paul E A 2019 Management controls on soil carbon *Soil Organic Matter in Temperate Agroecosystems* (Boca Raton, FL: CRC Press) pp 15–49
- Roberts M J, Schlenker W and Eyer J 2013 Agronomic weather measures in econometric models of crop yield with implications for climate change *Am. J. Agric. Econ.* **95** 236–43
- Robinson N P, Allred B W, Smith W K, Jones M O, Moreno A, Erickson T A, Naugle D E and Running S W 2018 Terrestrial primary production for the conterminous United States derived from Landsat 30 m and MODIS 250 m *Remote Sens. Ecol. Conserv.* **4** 264–80

- Rosenzweig C *et al* 2014 Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison *Proc. Natl Acad. Sci.* **111** 3268–73
- Schlenker W, Hanemann W M and Fisher A C 2006 The impact of global warming on U.S. agriculture: an econometric analysis of optimal growing conditions *Rev. Econ. Stat.* **88** 113–25
- Schlenker W and Roberts M J 2009 Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change *Proc. Natl Acad. Sci.* **106** 15594–8
- Steinweg J Megan, Plante A F, Conant R T, Paul E A and Tanaka D L 2008 Patterns of substrate utilization during long-term incubations at different temperatures *Soil Biol. Biochem.* **40** 2722–8
- Sun W, Canadell J G, Yu L, Yu L, Zhang W, Smith P, Fischer T and Huang Y 2020 Climate drives global soil carbon sequestration and crop yield changes under conservation agriculture *Glob. Change Biol.* **26** 3325–35
- Teixeira E *et al* 2021 Understanding spatial and temporal variability of N leaching reduction by winter cover crops under climate change *Sci. Total Environ.* **771** 144770
- Tucker C L, Bell J, Pendall E and Ogle K 2013 Does declining carbon-use efficiency explain thermal acclimation of soil respiration with warming? *Glob. Change Biol.* **19** 252–63
- USDA National Agricultural Statistics Service (USDA-NASS) 2020 Cropland Data Layer. Published Crop-Specific Data Layer (Washington, DC: USDA-NASS) (available at: <https://nassgeodata.gmu.edu/CropScape/>) (Accessed January 2022)
- Wang N, Wang E, Wang J, Zhang J, Zheng B, Huang Y and Tan M 2018 Modelling maize phenology, biomass growth and yield under contrasting temperature conditions *Agric. For. Meteorol.* **250–251** 319–29
- West T O and Marland G 2002a Net carbon flux from agricultural ecosystems: methodology for full carbon cycle analyses *Environ. Pollut.* **116** 439–44
- West T O and Marland G 2002b A synthesis of carbon sequestration, carbon emissions, and net carbon flux in agriculture: comparing tillage practices in the United States *Agric. Ecosyst. Environ.* **91** 217–32
- Wetterstedt J Å M and Ågren G I 2011 Quality or decomposer efficiency—which is most important in the temperature response of litter decomposition? A modelling study using the GLUE methodology *Biogeosciences* **8** 477–87
- Wong M T F and Asseng S 2006 Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscaling a crop model *Plant Soil* **283** 203–15
- Xiao J *et al* 2019 Remote sensing of the terrestrial carbon cycle: a review of advances over 50 years *Remote Sens. Environ.* **233** 111383
- Yu Z, Griffis T J and Baker J M 2021 Warming temperatures lead to reduced summer carbon sequestration in the US Corn Belt *Commun. Earth Environ.* **2** 1–11
- Zhang Y, Xu M, Chen H and Adams J 2009 Global pattern of NPP to GPP ratio derived from MODIS data: effects of ecosystem type, geographical location and climate *Glob. Ecol. Biogeogr.* **18** 280–90