



## Article

# Exploring the Individualized Effect of Climatic Drivers on MODIS Net Primary Productivity through an Explainable Machine Learning Framework

Luyi Li <sup>1</sup>, Zhenzhong Zeng <sup>2</sup>, Guo Zhang <sup>3,4</sup> , Kai Duan <sup>1</sup>, Bingjun Liu <sup>1</sup> and Xitian Cai <sup>1,5,\*</sup>

<sup>1</sup> Center for Water Resources and Environment, School of Civil Engineering, Sun Yat-sen University, Guangzhou 510275, China

<sup>2</sup> School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

<sup>3</sup> CMA Earth System Modeling and Prediction Centre, China Meteorological Administration, Beijing 100081, China

<sup>4</sup> State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

<sup>5</sup> Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

\* Correspondence: caixitian@mail.sysu.edu.cn

**Abstract:** Along with the development of remote sensing technology, the spatial–temporal variability of vegetation productivity has been well observed. However, the drivers controlling the variation in vegetation under various climate gradients remain poorly understood. Identifying and quantifying the independent effects of driving factors on a natural process is challenging. In this study, we adopted a potent machine learning (ML) model and an ML interpretation technique with high fidelity to disentangle the effects of climatic variables on the long-term averaged net primary productivity (NPP) across the Amazon rainforests. Specifically, the eXtreme Gradient Boosting (XGBoost) model was employed to model the Moderate-resolution Imaging Spectroradiometer (MODIS) NPP data, and the Shapley additive explanation (SHAP) method was introduced to account for nonlinear relationships between variables identified by the model. Results showed that the dominant driver of NPP across the Amazon forests varied in different regions, with temperature dominating the most considerable portion of the ecoregion with a high importance score. In addition, light augmentation, increased CO<sub>2</sub> concentration, and decreased precipitation positively contributed to Amazonia NPP. The wind speed for most vegetated areas was under the optimum, which benefits NPP, while sustained high wind speed would bring substantial NPP loss. We also found a non-monotonic response of Amazonia NPP to VPD and attributed this relationship to the moisture load in Amazon forests. Our application of the explainable machine learning framework to identify the underlying physical mechanism behind NPP could be a reference for identifying relationships between components in natural processes.

**Keywords:** net primary productivity; MODIS; vegetation response; climatic drivers; machine learning; Amazon



**Citation:** Li, L.; Zeng, Z.; Zhang, G.; Duan, K.; Liu, B.; Cai, X. Exploring the Individualized Effect of Climatic Drivers on MODIS Net Primary Productivity through an Explainable Machine Learning Framework.

*Remote Sens.* **2022**, *14*, 4401. <https://doi.org/10.3390/rs14174401>

Academic Editors: Xiaobin Guan, Zhaoying Zhang, Xinyao Xie and Xing Li

Received: 25 July 2022

Accepted: 31 August 2022

Published: 4 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

To grapple with climate change, mitigating the increasing carbon dioxide (CO<sub>2</sub>) concentration in the atmosphere becomes of great urgency [1,2]. Terrestrial ecosystems greatly benefit climate change alleviation since they absorb about one-third of CO<sub>2</sub> released by fossil fuel emissions and land use [2]. Therefore, it is important to broaden our understanding of the spatial and temporal dynamics of terrestrial carbon fluxes for model simulations to optimize policymakers' decisions on forest carbon sequestration measures. Vegetation productivity is the driving force in the terrestrial carbon cycle and a staple regulator of the natural carbon sequestration process of the terrestrial ecosystem. The Amazon rainforest is the largest intact tropical forest worldwide [3] and one of the regions with the most vigorous carbon exchange. It accounts for 50% of carbon stocks in tropical forests [4]. Thus,

exploring the Amazonia vegetation productivity and the driving mechanism behind it is essential.

Exploring the dominant factors and their driving pattern of Amazonia NPP is important in identifying the future direction and measures to improve productivity and address climate change. Previous studies have shown that under natural conditions, climatic factors are usually the major factors affecting the spatial variability of vegetation growth [5–8]. Generally, temperature, precipitation, and radiation are the dominating climatic factors of ecosystem carbon exchange [9–11], but at the regional level, the specific dominant environmental factors may vary [12–15]. For example, radiation dominated the productivity increase in 1982–1999 in the Amazon rainforest due to the reduced cloud cover and the improved light conditions [12]. CO<sub>2</sub> concentration is argued to dominate the changes in vegetation productivity worldwide [16], and VPD also appears to have played a significant role [17]. However, in recent years, studies have shown that the CO<sub>2</sub> fertilization effects in South America have increased [18], and the VPD has increased in many places worldwide [19]. These changes have prompted us to explore the dominant driving factors behind NPP and how it is affected by changes in various climate factors.

In recent years, the effects of atmospheric dryness and soil water dryness on vegetation have been studied in the context of global warming, sparking debate on the dominant factors and driving mechanisms of plant water stress. They present different responses of vegetation to atmospheric dryness in humid ecosystems. For example, Chen et al. and Green et al. [20,21] found that Amazonia vegetation growth responded positively to VPD in seasonal dynamics. Chen et al. [22] also suggested a positive correlation between vegetation productivity and VPD in humid ecosystems on daily data exploration. However, Yuan et al. [17] showed a negative correlation between VPD and vegetation productivity in the southern hemisphere on a longer interannual scale.

Explainable machine learning (XML) techniques may provide a valuable tool to reveal the complex relationship between physical quantities. XML has irresistible advantages that combine the robust data fitting ability of machine learning without shouldering empirical assumptions of biophysical mechanisms [23] while making complex relationships transparent and interpretable [24]. At present, many XML technologies have been developed. Some XML models' structures are relatively simple and inherently interpretable such as Cubist. On the contrary, some models' structures are complex. Still, they can be interpreted by model-specific interpretation techniques such as knowledge distillation or by model-agnostic techniques such as Local Interpretable Model-agnostic Explanations (LIME) or Shapley Additive exPlanations (SHAP). So far, XML has had broad applications in Earth science [25–27].

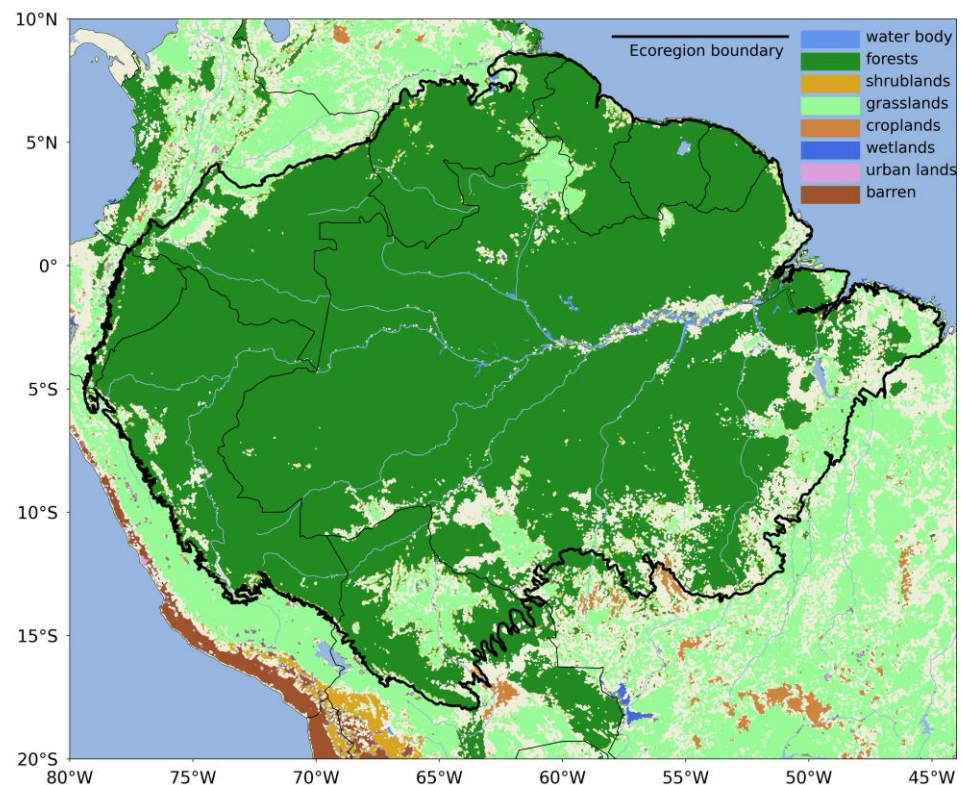
In this study, we adopted net primary productivity (NPP) to characterize Amazonia vegetation productivity since it indicates the rate at which carbon accumulates in plants after considering the losses from plant respiration and other metabolic processes [28,29]. We employed a machine learning method fed with climate data to establish a prediction model with high accuracy to model the continuous long-term averaged MODIS NPP in the Amazon ecoregion. We then adopted a high-fidelity model interpretation method to attribute NPP to the individualized effect of various driving factors. Our goals were (1) to determine the main drivers of NPP spatial variability in the Amazon region and (2) to reveal the individualized patterns of how each driver impacts NPP in the Amazon ecoregion. Our findings could benefit ecosystem model development through the improvements in determining model drivers and their influence patterns on carbon exchange.

## 2. Methodology

### 2.1. Study Area

The spatial extent of the Amazon ecoregion (Figure 1) is based on the Terrestrial Ecoregions of the World [30]. The vegetation of the Amazon is a mostly moist old-growth forest, and the broadleaf forest is the prevailing plant function type across the Amazon Rainforest. The Amazon ecoregion is located at an elevation below 1000 m [31], bounded

by the Andes Mountains to the west, the Lanos Mountains and the Atlantic Ocean to the north and east, and the Cerrado (the vast tropical savanna ecoregion in Brazil), and dry forests to the east and south [32], with its annual mean temperature above 20 °C and mean annual precipitation more than 2000 mm in the period ranging from 2001 to 2020.



**Figure 1.** Amazon ecoregion based on the definition from the Terrestrial Ecoregions of the World.

## 2.2. Data

We adopted Moderate Resolution Imaging Spectroradiometer (MODIS) NPP (MOD17A3H, version 6.0) [33] for analysis of vegetation productivity. The product is a cumulative 8-day composite of values with 500 m pixel size and was derived from a light use efficiency photosynthesis model fed with MODIS fPAR (fraction of absorbed photosynthetically active radiation). Taking vegetation maintenance and respiration into account [13,34], it accounted for respiratory costs and environmental stress on vegetation growth separately [35] by a Biome Parameter Lookup Table (BPLUT). MODIS NPP shows consistency ( $R^2 = 0.77$ ) with an extensive worldwide NPP dataset assembled from more than 1000 field-observed data [34] and also shows close agreement with the Bigfoot NPP products, which were measured at nine different ecosystem flux towers throughout the world [36].

To explore the relationship between NPP and its potential drivers, we considered the mean annual values of the following environmental variables in light of previous studies [15,17,19,37,38]: daytime land surface temperature (TS; °C), precipitation (P; mm), downward surface shortwave radiation (SR;  $W m^{-2}$ ), wind speed (WS;  $m s^{-1}$ ), soil moisture content of 0–10 cm (SM10;  $m^3 m^{-3}$ ), soil moisture content of 100–200 cm (SM200;  $m^3 m^{-3}$ ), vapor pressure deficit (VPD; kPa), and atmospheric carbon dioxide concentration ( $CO_2$ ;  $mol m^{-2} s^{-1}$ ). In addition, three factors depicting soil characteristics were also considered, including total Nitrogen (TN; % of weight), total phosphorus (TP; % of weight), and clay content (CLAY; % of weight), in the soil layer to the depth of 2.3 m.

Here, TS is derived from a monthly MODIS product with a  $0.05^\circ \times 0.05^\circ$  spatial resolution (MOD11C3, version 6.0) [39]. We chose land surface temperature as an explainable variable instead of air temperature because the former more approaches the canopy temperature [40]. SR, P, and WS were extracted from the TerraClimate (Table 1) product

that provides a 1/24-degree gridded monthly meteorological dataset. The TerraClimate dataset was derived from a combination of high-spatial-resolution climatological normal from the WorldClim dataset, the time-varying data from CRU Ts4.0, and the Japanese 55-year Reanalysis (JRA55) by using climatically aided interpolation [41]. We adopted this data product mainly due to its high spatial resolution and verified accuracy on a global scale. SM10 and SM200 were collected from FLDAS Noah Land Surface Model L4 Global Monthly  $0.1^\circ \times 0.1^\circ$  reanalysis dataset for its high resolution and long available period compatible with our study [42]. VPD is an indicator illustrating atmospheric dryness and is defined as the difference between the saturation vapor pressure at air temperature and the actual vapor pressure of the air. VPD was calculated from air temperature and specific humidity data from the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) data in Equation (1) [43,44]. CO<sub>2</sub> was extracted from the CarbonTracker dataset (CT2019B), which estimates the land biosphere net CO<sub>2</sub> fluxes on the global  $1^\circ \times 1^\circ$  grid [45]. TN, TP, and CLAY were obtained from the gridded Global Soil Dataset for use in Earth System Models (GSDE) dataset with 30 arc-seconds resolution [46].

$$VPD = 0.611 \times \exp \frac{17.27 \times TA}{TA} \left( 1 - \frac{RH}{100} \right) \quad (1)$$

where  $TA$  is the air temperature (K) and  $RH$  is the relative humidity ( $\text{kg kg}^{-1}$ ).

**Table 1.** Summary of data products used in this study.

Dataset	Data	Unit	Spatial Resolution	Temporal Resolution	Reference	Data Acquisition
MODIS (Version 6.0)	Net primary productivity (MOD17A3H)	$\text{g C m}^{-2}$	500 m in a Sinusoidal projection	yearly	[33]	<a href="https://e4ftl01.cr.usgs.gov/">https://e4ftl01.cr.usgs.gov/</a> (accessed on 19 March 2022)
	Daytime land surface temperature (MOD11C3)	$^\circ\text{C}$	$0.05^\circ$	monthly	[39]	
	Land cover type (MCD12C1)	–	$0.05^\circ$	yearly	[47]	
TerraClimate	Downward surface shortwave radiation	$\text{W m}^{-2}$	$1/24^\circ$	monthly	[41]	<a href="https://www.climatologylab.org/terraclimate.html">https://www.climatologylab.org/terraclimate.html</a> (accessed on 19 June 2022)
	Precipitation	mm				
	Wind speed	$\text{m s}^{-1}$				
FLDAS (Noah Land Surface Model L4)	Soil moisture content of 0–10 cm	$\text{m}^3 \text{m}^{-3}$	$0.1^\circ$	monthly	[42]	<a href="https://ldas.gsfc.nasa.gov/FLDAS/">https://ldas.gsfc.nasa.gov/FLDAS/</a> (accessed on 6 July 2022)
	Soil moisture content of 100–200 cm					
	Air temperature	K				
	Specific humidity	$\text{kg kg}^{-1}$				
CarbonTracker CT2019B	Land biosphere net CO <sub>2</sub> fluxes	$\text{mol m}^{-2} \text{s}^{-1}$	$1^\circ \times 1^\circ$	monthly	[45]	<a href="https://gml.noaa.gov/ccgg/carbontracker/CT2019B/">https://gml.noaa.gov/ccgg/carbontracker/CT2019B/</a> (accessed on 5 July 2022)
GSDE	Total Nitrogen	% of weight	$30''$	–	[46]	<a href="http://globalchange.bnu.edu.cn/research/soilw">http://globalchange.bnu.edu.cn/research/soilw</a> (accessed on 6 July 2022)
	Total phosphorus					
	Clay content					

In addition, we applied land cover data to filter pixels that had plant function types changed to exclude the effects of land use change. Land cover data were extracted from the MODIS dataset (MCD12C1, version 6.0) [47], where we employed the International Geosphere-Biosphere Program (IGBP) classification layer to determine the land cover type. Moreover, we used the quality layer for all the applied MODIS data to filter pixels with poor quality. All the selected driving datasets were reprojected to the World Geodetic System 1984 and resampled to  $0.05^\circ \times 0.05^\circ$  spatial resolution using the bilinear resampling method (for downsampling) and the average composite method (for upsampling). To eliminate the inter-annual fluctuation and avoid distractions from short-term phenomena, we averaged all data over 20 years from 2001 to 2020, except for CO<sub>2</sub>, of which the available period only ranged from 2001 to 2018. After filtering the raw data according to the conditions applicable for all variables, there were 171,819 samples for model inputs.

### 2.3. Machine Learning Model

Tree-based models consistently outperform neural networks in tackling tabular-style datasets whose features lack strong multiscale structures and have individual meanings [48]. Therefore, we developed our prediction model by adopting the tree-based eXtreme Gradient Boosting (XGBoost) algorithm [49]. XGBoost is an enhanced version of GBDT (Gradient Boosting Decision Tree) with its core algorithm based on the idea of “boosting”—a stepwise forward additive strategy. Specifically, the process of “boosting” in XGBoost is explained below. As a powerful estimator (ensembled estimator), XGBoost executes its output prediction by aggregating all the predictions of a set of the weak estimators (single estimator, usually decision trees) that are orderly constructed in the fitting process. The first estimator constructed is fitted to the whole data during the process. A later estimator is then built and fitted to the residuals of the previous estimator prediction. The significant improvement in XGBoost compared to GBDT draws on the regularization term added in the target function, the more precise loss function, and a set of related regularization work which can reduce overfitting and errors, such as shrinkage and column subsampling [49].

We used the XGBoost package and the Scikit-learn library to implement the model construction in the Python 3.8 environment. Among the 171,819 available samples for model construction, 120,273 for training and 51,546 for testing were randomly sampled. Hyperparameter tuning before model training was performed using a 10-fold cross-validation method based on the training set. The performance of the model trained after each round of parameter update was evaluated with the averaged root mean square error (RMSE) of the 10 cross-validation results.

The model hyperparameters were optimized by a stochastic hill-climbing algorithm, a stochastic local search optimization algorithm. Specifically, the hyperparameter adjustment range for each parameter was set first, thus defining the parameter searching space. Then, the initial search value and step size for each hyperparameter were set. When the hyperparameter tuning started, the parameter stepped forward from the initial value in the search space until the local optimum point (the point corresponding to the minimum RMSE) was met, which is the so-called “hill climbing” process. The search for each parameter was performed sequentially. Once the optimal value of the last parameter is obtained, it is fixed, the parameter combination is updated, and then the search traversal of the latter parameter is performed. We mainly tuned nine hyperparameters and confirmed the optimized hyperparameters as follows: `n_estimators = 700`, `learning_rate = 0.15`, `max_depth = 10`, `colsample_bylevel = 0.9`, `colsample_bynode = 0.9`, `colsample_bytree = 0.9`, `reg_alpha = 1000`, `reg_lambda = 150`, and `gamma = 0`.

We also used another two tree-based ML algorithms—Random Forest and Cubist—for modeling as a comparison. The former is also an ensemble tree algorithm based on the bagging concept. On our dataset, the results of Random Forest after hyperparameter optimization are not as good as XGBoost. The essence of Cubist is a piecewise linear decision tree [50,51]. Cubist has a simple model structure with inherent interpretability. The main principle is to divide the whole response variables into subsets by different rules

expressed by linear formula between the response variables and their explanatory variable. Our test results show that on our data, at least establishing 500 linear regression models ( $n\_rules = 500$ ) can achieve a fitting score comparable to XGBoost. However, excessive rule establishment sacrifices the interpretability of the model. Therefore, we finally adopted XGBoost as the fitting model.

### 2.4. Model Interpretation

Many machine learning interpretation techniques with high fidelity have been developed [52–55]. In this study, we adopted the SHapley Additive exPlanation (SHAP) method for its mature application technology and series of good qualities such as additive property and fairness in attribution [56,57]. SHAP has its solid theoretical foundation—Shapley value [58], which came from the coalitional game theory [59,60]. A SHAP value is the amount of mathematical deformation of the Shapley value within the SHAP framework. SHAP quantifies the contributions of each sample to the predicted value and decomposes the contributions into piece contributions from each feature. When SHAP begins to interpret a prediction model, it first obtains the “base value” of the model, which indicates the value that the model would have predicted if no knowledge of the features was provided for the current output (the average of the predicted values by the model). For each sample, our prediction model predicts the corresponding predicted value. Then, SHAP translates the difference between the base value and the predicted value into the sum of the attribution value for each input feature, where the attribution values are the SHAP values [61].

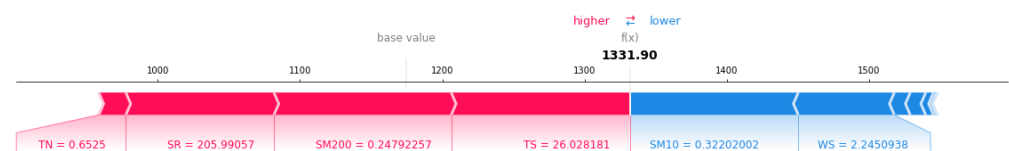
The explanation could be specified as:

$$g(z') = \varphi_0 + \sum_{j=1}^M \varphi_j z'_j \tag{2}$$

$$\varphi_i = \sum_{S \subseteq \{N/\{i\}\}} \frac{|S|!(M-|S|-1)!}{M!} (f_x(S \cup \{i\}) - f_x(S)) \tag{3}$$

where  $g$  is the explanation model,  $\varphi_0$  is the average of the prediction, and  $\varphi_i \in \mathbb{R}$  is the feature attribution for a feature  $i$ .  $z' \in \{0, 1\}^M$  is the coalition vector,  $M$  is the number of the features input,  $N$  is the set of all input features,  $S$  is the set of non-zero indexes, and  $f_x$  is the expected value of the function conditioned on a subset of the input features.

Figure 2 visualizes the decomposition of an instance prediction into each feature. Based on the feature attributions by SHAP, we could do some further analyses such as global evaluation of feature importance, exploring the relationship between the value of a feature and the impact on the prediction and the interaction effect between features.



**Figure 2.** Decomposed SHAP values for the prediction of an example individual pixel. The base value indicates the averaged output of the predictions.  $f(x)$  indicates the specific prediction of this sample. Features that increase the value of the prediction are shown in red; those that lower the prediction value are shown in blue.

The author of SHAP has also developed an efficient implementation of SHAP on tree-based models, TreeSHAP [61], where it provides the concept of a SHAP interaction value to assume the dependence relationship of features. A SHAP interaction value characterizes the magnitude of the interaction effect of paired features. According to the same principle of SHAP value, the sum of the SHAP interaction values of the feature with all other variables (including the interaction with itself) constitutes the SHAP value of this feature [61].

For SHAP interaction values, the explanation could be specified as:

$$\varphi_i = \sum \varphi_{i,j} \tag{4}$$

$$\varphi_{i,j} = \sum_{S \subseteq \setminus \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!} \delta_{ij}(S) \tag{5}$$

$$\delta_{ij}(S) = f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S) \tag{6}$$

when  $i \neq j$ . The SHAP interaction value between feature  $i$  and feature  $j$  is equally split between each feature. Thus  $\varphi_{i,j} = \varphi_{j,i}$  and the total interaction effect is  $\varphi_{i,j} + \varphi_{j,i}$ .

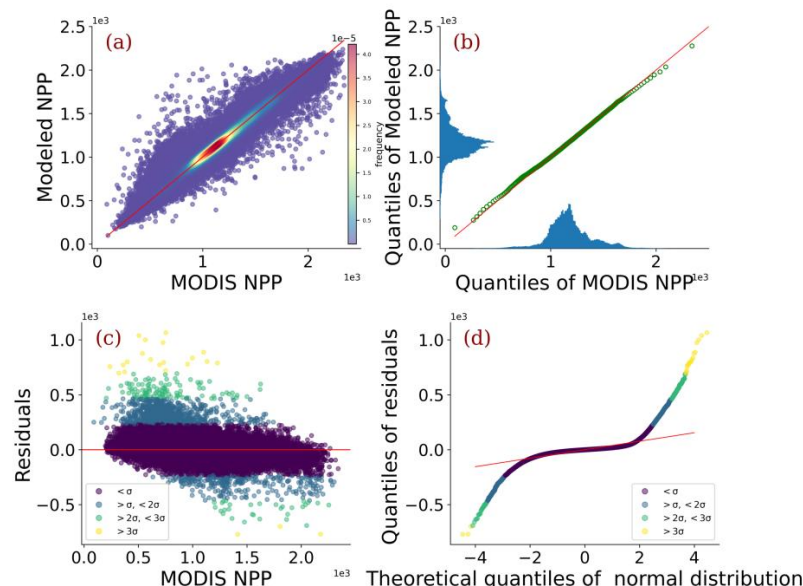
The main effects for a prediction can be defined as the interaction of the feature with itself, which can be calculated as:

$$\varphi_{i,i} = \varphi_i - \sum_{i \neq j} \Phi_{i,j} \tag{7}$$

### 3. Results

#### 3.1. Model Evaluation

Model evaluation was based on a testing subset. The constructed model obtained a coefficient of determination ( $R^2$ ) of 0.923, a mean absolute error (MAE) of 30.85, and a root mean square error (RMSE) of 64.28 on the testing subset (mean value and median value of the testing data were 1152.12  $\text{g}\cdot\text{C}/\text{m}^2$  and 1133.47  $\text{g}\cdot\text{C}/\text{m}^2$ , respectively). Figure 3a compares the modeled NPP with the MODIS NPP. Figure 3b is a Quantile–Quantile plot (QQ plot) embedded by frequency histograms, which compares their overall distributions. We can see that there are some large deviations at the beginning and the end of the fitted data distribution, while the vast majority of points are consistently fitted.



**Figure 3.** Model evaluation plots based on the testing data: (a) The scatter plot depicts a detailed comparison of the two data. (b) A quantile–quantile plot of the modeled and MODIS NPP laced with the histogram of distribution, respectively. (c) Residuals versus the observed NPP. (d) A comparison plot of model residual distribution and a normal distribution.  $\sigma$  indicates the standard deviation of the testing data of MODIS NPP.

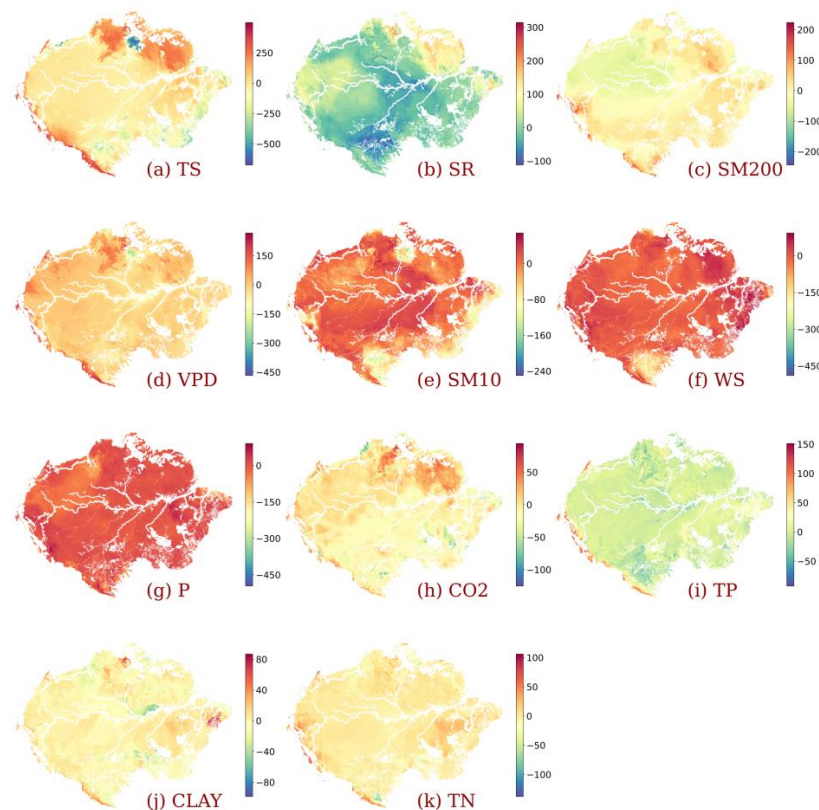
Residual analysis was further performed. Figure 3c shows a scatter plot of residuals ( $y$ -axis) versus the observed ( $x$ -axis) values of the dependent variable. For a well-performed model, we expect a symmetric spread of points around the horizontal line at zero, indicating random deviations of predictions from the observed values. The residuals are positive

for the small observed values of the dependent variable, while for large values, they are negative. Especially the outlier dots that have exceeded the threshold of three standard deviations, mostly distributed at both ends of the distribution. Thus, the plot suggests that the predictions inclined to the average. Figure 3d shows a comparison between the QQ plot of the residual distribution and a set of normally distributed data with zero mean. Theoretically, the residual should be random and unpredictable. One of the manifestations of this randomness could be that the bias fits well with the normal distribution. As seen from the Figure 3d, most of the residuals predicted by the model are close to the normal distribution, but the residual values of the beginning and the end have a relatively large deviation.

Overall, the constructed model predicted NPP was consistent with most of the testing MODIS NPP with evaluation metrics indicating good performance, but in a local view, the model overestimated small values and underestimated large values. In particular, the deviations were significant for the extreme values. Only samples that have residuals less than three times the standard deviation of the testing data joined the SHAP interpretation.

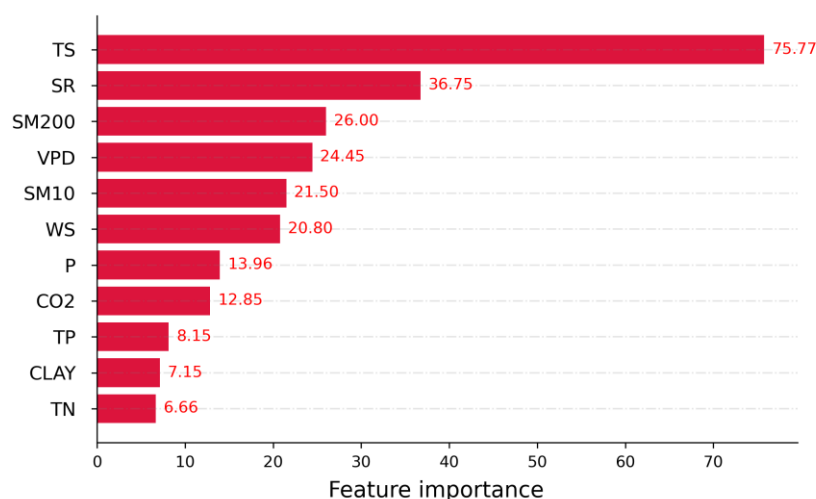
### 3.2. Relative Importance of Drivers of NPP

SHAP attribution was conducted (Figure 4). Feature importance was quantified as each feature's mean absolute SHAP values for all pixels. Then, the features were ranked according to these importance indexes, as shown in Figure 5. TS appeared to be the most crucial driver for the long-term averaged NPP distribution, followed by SR, whose mean SHAP value was less than half of TS's. SM200, VPD, SM10, WS, P, and CO<sub>2</sub> orderly came next, with their feature importance gradually decreasing and being 26.00 g·C/m<sup>2</sup>, 24.45 g·C/m<sup>2</sup>, 21.50 g·C/m<sup>2</sup>, 20.80 g·C/m<sup>2</sup>, 13.96 g·C/m<sup>2</sup>, and 12.85 g·C/m<sup>2</sup>, respectively. The remaining three factors depicting the soil characteristic, TP, CLAY, and TN, ranked in the lowest stream with their mean absolute SHAP values less than 10 g·C/m<sup>2</sup>.



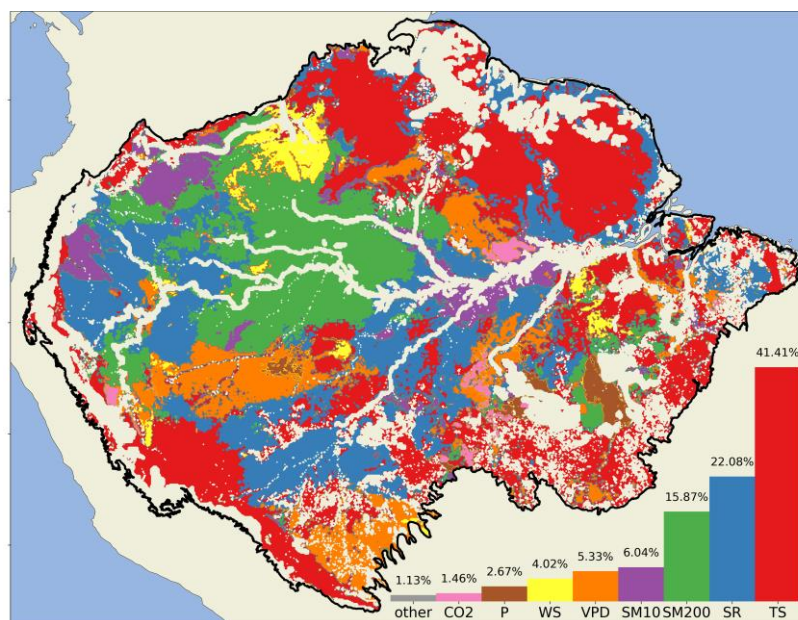
**Figure 4.** Spatial distribution of SHAP values for each  $0.05^\circ \times 0.05^\circ$  sample of (a) TS, (b) SR, (c) VPD, (d) SM200, (e) SM10, (f) WS, (g) CO<sub>2</sub>, (h) P, (i) CLAY, (j) TN, and (k) TP. Units of SHAP values: g C/m<sup>2</sup>.





**Figure 5.** Feature importance plot. The feature importance was quantified as the average of absolute SHAP values (mean  $|\text{SHAP values}|$ ) of all  $0.05^\circ \times 0.05^\circ$  pixels. Units of SHAP values:  $\text{g C}/\text{m}^2$ .

In addition to measuring the magnitude of the force that a feature impacted on the NPP output, the primary drivers of each  $0.05$ -degree pixel were also identified and mapped in Figure 6. The dominant driver of NPP varied widely across the ecoregion. Therefore, we calculated the percentage of the spatial area dominated by the drivers at all grid cells and ranked the drivers in light of this count. TS was the most significant impact factor of NPP across variable areas. Radiation also has a considerable dominant sphere. Moisture condition has great clout in the west of the ecoregion where precipitation is abundant.

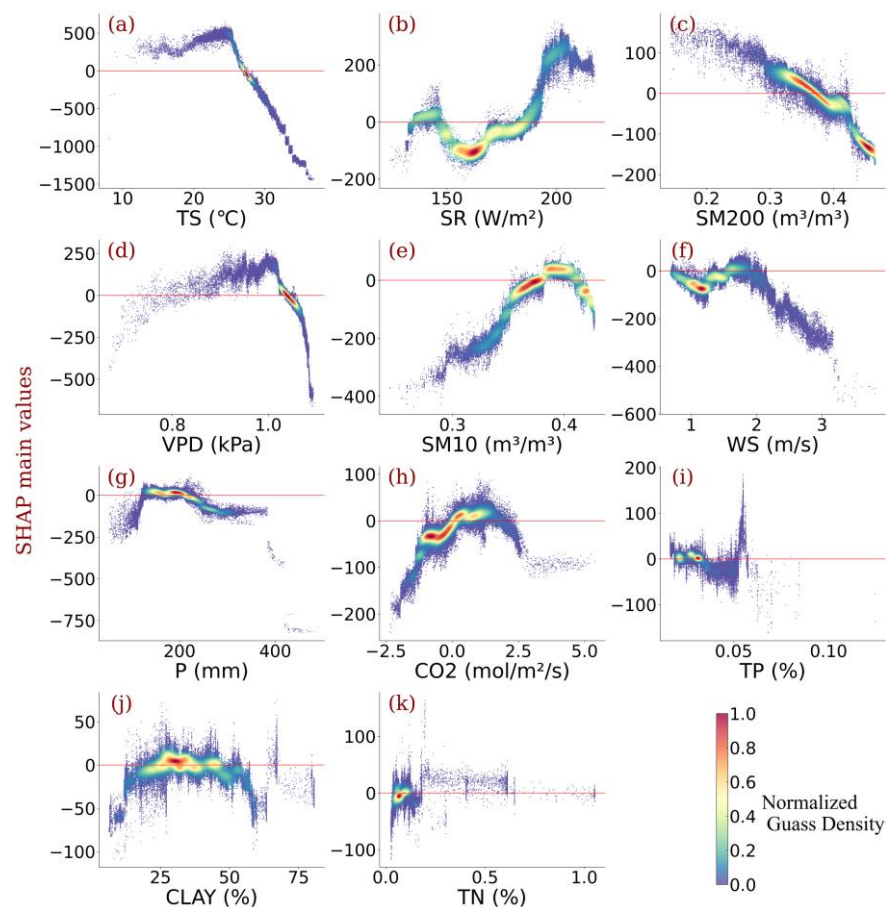


**Figure 6.** Spatial variability of primary drivers for each pixel with a resolution of  $0.05^\circ \times 0.05^\circ$ . The heights of the bars indicate the percentage of the cells occupied by the corresponding drivers. Soil indicates the combination of CLAY, TN, and TP.

Overall, the results indicate that TS was the major driving factor of the spatial variability of NPP in the Amazon ecoregion, both in terms of force and dominant space. In a more local analysis, the dominant driver varied across the Amazon rainforest. SR and moisture conditions (VPD, SM200, and SM10) also contributed to the spatial pattern of Amazonia NPP. WS and  $\text{CO}_2$  played a minor role. Precipitation almost made a negligible effect.

### 3.3. Impact of Individualized Climatic Drivers on Spatial Variability of NPP

We plotted the dependencies between each factor's values and its corresponding SHAP values for all samples to explore how these drivers affect the final model outputs (Figure 7). The plots show the main effects of individual variables on the NPP, eliminating the interactions between variables (SHAP dependence plots confounded interaction effect are shown in Figure A1). The zero line of SHAP value (horizontal red line) distinguishes the positive and negative driving force exerted on NPP compared to the multi-year average forecast (base value). Overall, almost all of the factors influenced NPP non-monotonically, suggesting that the influence of the environmental factors is regionally variable. Only SM200 exerted a relatively linear and monotonical effect on NPP that enhanced SM200 directly decreased NPP. NPP had a similar non-monotonic relationship with TS, VPD, and P. With the enhancement of the three factors, NPP reversed its original increasing trajectory after reaching a certain level. Areas with extreme values of P even experienced substantial NPP loss. The effects of SR, SM10, and CO<sub>2</sub> on NPP were also similar, except for SR less than 150 W/m<sup>2</sup>, which directly increases the NPP gain, but the threshold effects all occurred until they reached a considerably large value. Most areas of low WS (WS < 2 m/s) failed to contribute NPP to the average level (SHAP value = 0), and high WS directly impairs NPP. The effect of CLAY on NPP mostly fluctuated in a small scope along the gradient of CLAY but exerted a negative impact on NPP when CLAY is relatively low or when CLAY is relatively high. Samples of high TN and high TP were scarce, but it can be seen that high TP brought NPP loss and high TN brought NPP gain.



**Figure 7.** SHAP dependence plots depicting the SHAP main values along the gradient of (a) TS, (b) SR, (c) SM200, (d) VPD, (e) SM10, (f) WS, (g) P, (h) CO<sub>2</sub>, (i) TP, (j) CLAY, and (k) TN. The lateral axis indicates the gradient of variable values. The vertical axis indicates the magnitude of the SHAP main value. Positive SHAP values indicate the positive force on NPP output while negative SHAP values indicate the opposite. The colors indicate the density. SHAP main values units: g C/m<sup>2</sup>.

## 4. Discussion

### 4.1. Dominant Drivers of Spatial Variability of Amazonia NPP

In a humid forest ecosystem, vegetation productivity is generally dominated by radiation or temperature [62,63]. Our SHAP analysis suggests that the temperature, dwarfing other limitation elements, is the most potent driver of the spatial heterogeneity of Amazonia NPP in terms of strength and space dominance (Figures 5 and 6). This is reasonable because the temperature is the greatest limiting factor for the photosynthesis rate in well-lit, warm leaves. To be specific, the photosynthetic efficiency of the canopy mainly depends on the gas exchange rate of the sunlit leaves rather than the gas flux of the underlit leaves [64]. Moreover, most of the Amazon forests are tall, old-growth forests with large canopies, reasonable light-receiving structures, with abundant water supply. Additionally, in recent decades, as the dry season has extended, cloud cover has decreased, and radiation has increased [12,65], thereby canopy leaves have received more sunlight. Thus, the temperature became the dominant limiting factor.

The response of NPP to temperature was in tune with that in the leaf scale [66–68]. It was worth noting that the temperatures of most of the vegetation had exceeded the optima for photosynthesis, after which NPP showed negative feedback to increasing temperature (Figure 7a) [40,64]. Considering the current upward trajectory of temperature and such a high sensibility of NPP, it should be noted that the nucleus density center may move down and the Amazonia vegetation productivity is likely to wind down soon.

Solar radiation also contributed to the spatial pattern of Amazonia NPP. Most of the Amazonia vegetation complied with the fact that enhanced light intensity directly promotes increased photosynthesis [69,70] (Figure 7b). However, in our SHAP analysis, the west of the Amazon ecoregion covered with high precipitation made an exception for low radiation to play a relatively positive impact on NPP (Figures 7b, 4b, and A2b). This may be because under the low radiation conditions caused by the obstruction of thick clouds brought by the large water vapor in the atmosphere [69,71]; although, the vegetation receives less direct beam and total irradiance, the diffuse irradiance that penetrates canopies more efficiently than direct beam irradiance would be increased, thereby increasing the area of photosynthetic activity [72,73]. Therefore, the effect of radiation on vegetation productivity may also depend on the trade-off between total irradiance, diffuse irradiance, and radiation utilization efficiency [72]. More positive effects of solar radiation enhancement can be obtained for well-structured stand and canopy structures.

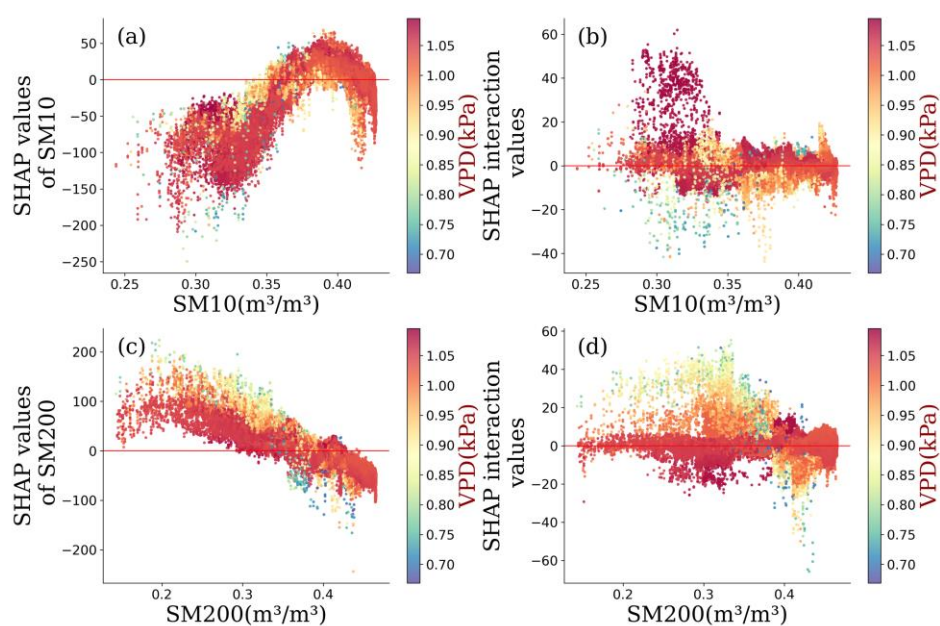
Our SHAP analysis suggested an optimal wind speed of about 1.8 m/s (Figure 7f). The mechanisms of how wind speed influences vegetation are complex. Wind movement adjusted the photosynthesis of plants by affecting the leaf structure of the canopy [74,75]. For example, an appropriate increase in wind speed can increase the light receiving area of the blade and the radiation absorption efficiency. However, the lower wind speed had more respiratory CO<sub>2</sub> recycled [76]. Generally, wind speed greater than 2 m/s can be called strong wind. Sustained strong winds will hinder local vegetation leaf reproduction and growth [77]. Moreover, high wind speed promoted the speed of fire spread [78]. It is notable that the Amazon Rainforest suffered frequent fires in recent years in places with high wind speeds.

In addition, an elevated CO<sub>2</sub> concentration promoted Amazonia NPP (Figure 7h), but it also had a threshold. A previous study reported that the CO<sub>2</sub> fertilization effect in Southern America has increased [18]. This is beneficial to the NPP growth in Amazon forests. However, the increasing CO<sub>2</sub> was always accompanied by increasing temperature and other variables changes, but the effect of CO<sub>2</sub> on Amazonia NPP seemed relatively limited (Figures 5 and 6).

### 4.2. Amazonia Vegetation Responds to Moisture

Our study demonstrated a non-monotonic response of Amazonian vegetation to VPD, with NPP initially being slightly promoted and then strongly suppressed in response to increasing drying conditions (Figure 7d). Excluding phenological and seasonal dynamic

factors such as leaf turnover [21], the moisture load of the humid rainforest itself can better explain the long-term vegetation–VPD response mechanism of the Amazon rainforest. On the one hand, the interaction of VPD with almost all magnitudes of shallow soil water content shows a relatively more positive effect on NPP than that of low VPD (Figure 8b). On the other hand, high VPD showed a suppressive impact on NPP in regions with low deep soil moisture content but relaxed its negative effect in areas with high deep soil moisture content (Figure 8c,d), indicating that the vegetation–VPD relationship may be regulated by soil moisture. Amazon rainforests have a strong water-holding capacity and deep root systems, which determines its water use strategy, in that the physiological activities of vegetation rely more on stable and abundant deep soil water. However, excessive water load may stress vegetation photosynthesis by impairing enzyme activity, chlorophyll, and soluble protein levels [79]. Therefore, the reduction in deep soil moisture could alleviate the pressure caused by the oversupply of moisture (Figure 7c).



**Figure 8.** Coupling effect and interaction effect between soil moisture content and vapor pressure deficit. Units of SHAP and SHAP interaction values:  $g\ C/m^2$ : (a) Coupling effect between SM10 and VPD. (b) Interaction effect between SM10 and VPD. (c) Coupling effect between SM200 and VPD. (d) Interaction effect between SM200 and VPD.

Furthermore, shallow soil water could act as a backup reservoir when water supplements become urgent due to its strong resilience. Therefore, enhanced shallow soil water content conducted NPP gain but also brought stress on NPP when it exceeded a certain level (Figure 7e). Similarly, further consistent precipitation increases may also impair the Amazonia NPP (Figure 7g).

In this context, appropriately increasing dryness may help mitigate the water overload by promoting leaf stomata openness and evapotranspiration. Moreover, as dryness increases, Stomatal stomata gradually become more adaptable and less sensitive to VPD [80,81], thereby enhancing photosynthesis ability. However, when VPD increases to a certain extent, the water loss caused by intensified transpiration exceeds the benefit of photosynthesis. At the same time, the reduction in the temperature by evaporative cooling also increases the cost of photosynthesis. Therefore, over-high VPD shifts its positive effect toward the suppressive effect on vegetation productivity [17,82,83]. Generally, the optimal VPD for plants ranges from 0.3 to 1.0 kPa [84,85], and water stress occurs in plants when VPD is over 1 kPa [86], which is consistent with our findings (Figure 7d). It is worth noting that most of the Amazonia vegetation is under or exceeding optimal VPD conditions, indicating that further increments of VPD may reduce vegetation productivity.

#### 4.3. Benefits and Uncertainty of Explainable Machine Learning

Machine learning has profound applications due to its notable advantages, among which the most notable is the powerful ability to explore complex relationships behind data. However, data-driven methods are intrinsically constrained by data and algorithms [87]. Moreover, machine learning fails to handle problems of extreme values well. Thus, carefully screening and processing for data quality and integrity is a crucial step to reducing uncertainty in ML model construction.

A distinct drawback of SHAP values is that they are provided as the property of the additive contribution of explanatory variables [56]. In other words, SHAP assumes that the variables are independent of each other, which would make the model mixed with high interaction less credible. However, the situation that the characteristics are entirely independent rarely exists. Especially in Earth Science, various environmental factors have coupling or interaction effects. Despite TreeSHAP [61] assuming less feature independence and explaining some feature dependencies, it still fails to fully explain [57], because, for example, SHAP interaction value is also endowed with the attribute of additive contribution. We carefully observed correlations between features and made tradeoffs before model input. Moreover, the model performance and consistency of the model results with known physical mechanisms (Figure 3) lend us confidence in the constructed model.

Note that XML does not capture the complete behavior of a physical system but rather provides a rough approximation of the system's behavior [24]. Therefore, the model's credibility relies more on the rigorous and comprehensive selection of features during model construction to approach the actual situation as closely as possible. Despite all the uncertainty and drawbacks, XML remains a valuable method boasting a large potential to comprehend the recondite functionality behind phenomena in Earth Science, not only for prediction [25,88,89] but also for contrastive explanations [15,90].

#### 5. Conclusions

As NPP is driven by multiple climatic variables, we adopted an explainable machine learning technique to understand how the long-term averaged MODIS NPP responds to the climate variables in the Amazon rainforest. The specific operation included: first, NPP was learned through a potent machine learning model; then, the model was fed into the SHAP framework to explain the mechanisms. Conclusions based on the SHAP analyses are as follows:

1. Relative contributions of each driver were identified, showing that the temperature outperformed other climatic variables in contributing to Amazonia NPP variability. Radiation and vapor pressure deficit also made a considerable contribution. Wind speed, CO<sub>2</sub> concentration, and precipitation were also responsible.
2. Individualized feature attribution was detected. In most areas of Amazon forests, the temperature exceeded the optimal value for NPP growth. Generally, elevated radiation and increased CO<sub>2</sub> concentration promote NPP gain monotonically, while high precipitation impairs NPP. In addition, for most vegetation, the wind speed did not reach the optimum value that benefits NPP, and sustained high wind speed would bring substantial NPP loss.
3. Amazonia NPP responded to VPD non-monotonically. Considering the distinct response of NPP to soil water content under different layers, the relationship between NPP and VPD was highly connected to the water use policy and moisture overload conditions in Amazon forests. Further increases in VPD largely impaired NPP despite the moisture overload conditions in Amazon forests.

Continuous and high-covered remote sensing data allow us to make large-scale and high-precision vegetation analyses, and the explainable machine learning technology enables us to better understand the response mechanism of MODIS NPP to its climatic factors. Applying a combination of remote sensing data and explainable machine learning in Earth Science could serve as an effective reference for comprehending natural dynamics in the ecosystem.

**Author Contributions:** Conceptualization, L.L. and X.C.; data curation, L.L.; methodology, L.L.; Formal analysis, L.L. and X.C.; writing—original draft, L.L.; writing—review and editing, Z.Z., G.Z., K.D., B.L. and X.C.; Supervision, X.C.; Funding acquisition, X.C. and K.D. All authors have read and agreed to the published version of the manuscript.

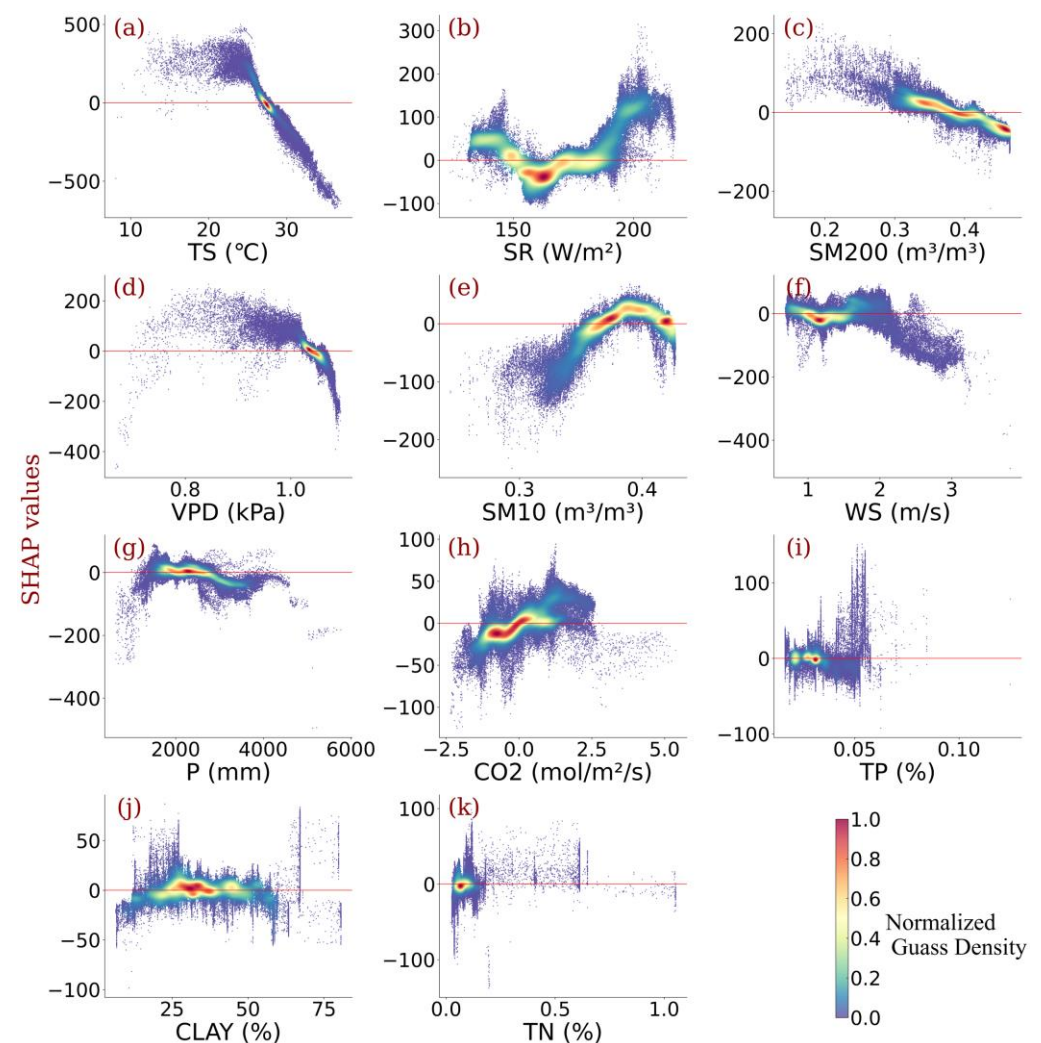
**Funding:** This research was funded by the National Key Research and Development Program of China (2021YFC3200205), the Natural Science Foundation of Guangdong Province, China (Grant No. 2022A1515010676), and the National Natural Science Foundation of China (51909285).

**Data Availability Statement:** Sources of all the data used in this study are listed in Table 1.

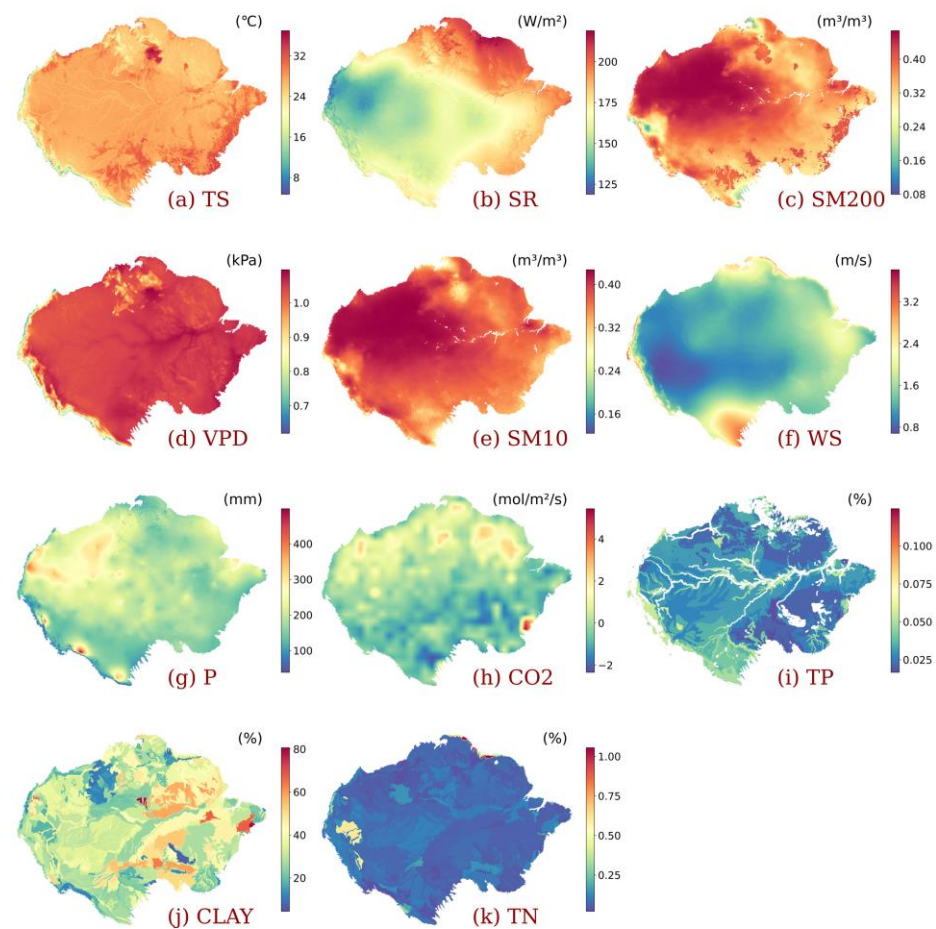
**Acknowledgments:** The authors would like to acknowledge the valuable comments on the manuscript provided by the editors and the two anonymous reviewers.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A



**Figure A1.** SHAP dependence plots depicting the SHAP values along the gradient of (a) TS, (b) SR, (c) VPD, (d) SM200, (e) SM10, (f) WS, (g) CO<sub>2</sub>, (h) P, (i) CLAY, (j) TN, and (k) TP. The lateral axis indicates the gradient of variable values. The vertical axis indicates the magnitude of the SHAP main value. Positive SHAP values indicate the positive force on NPP output while negative SHAP values indicate the opposite. The colors indicate the density. SHAP values units: g C/m<sup>2</sup>.



**Figure A2.** Spatial distribution of values for each  $0.05^\circ \times 0.05^\circ$  sample of (a) TS, (b) SR, (c) SM200, (d) VPD, (e) SM10, (f) WS, (g) P, (h) CO<sub>2</sub>, (i) TP, (j) CLAY, and (k) TN.

## References

1. Bonan, G.B.; Doney, S.C. Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science* **2018**, *359*, 533. [[CrossRef](#)]
2. Lovenduski, N.S.; Bonan, G.B. Reducing uncertainty in projections of terrestrial carbon uptake. *Environ. Res. Lett.* **2017**, *12*. [[CrossRef](#)]
3. Hansen Matthew, C.; Stehman Stephen, V.; Potapov Peter, V. Quantification of global gross forest cover loss. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8650–8655. [[CrossRef](#)]
4. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A Large and Persistent Carbon Sink in the World's Forests. *Science* **2011**, *333*, 988–993. [[CrossRef](#)]
5. Field, C.B.; Behrenfeld, M.J.; Randerson, J.T.; Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **1998**, *281*, 237–240. [[CrossRef](#)]
6. Chu, C.; Bartlett, M.; Wang, Y.; He, F.; Weiner, J.; Chave, J.; Sack, L. Does climate directly influence NPP globally? *Glob. Chang. Biol.* **2016**, *22*, 12–24. [[CrossRef](#)]
7. De Jong, R.; Schaepman, M.E.; Furrer, R.; De Bruin, S.; Verburg, P.H. Spatial relationship between climatologies and changes in global vegetation activity. *Glob. Chang. Biol.* **2013**, *19*, 1953–1964. [[CrossRef](#)]
8. Chu, C.; Lutz, J.A.; Král, K.; Vrška, T.; Yin, X.; Myers, J.A.; Abiem, I.; Alonso, A.; Bourg, N.; Burslem, D.F.; et al. Direct and indirect effects of climate on richness drive the latitudinal diversity gradient in forest trees. *Ecol. Lett.* **2019**, *22*, 245–255. [[CrossRef](#)]
9. Craine, J.M.; Nippert, J.B.; Elmore, A.J.; Skibbe, A.M.; Hutchinson, S.L.; Brunsell, N.A. Timing of climate variability and grassland productivity. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3401–3405. [[CrossRef](#)]
10. Beer, C.; Reichstein, M.; Tomelleri, E.; Ciais, P.; Jung, M.; Carvalhais, N.; Rödenbeck, C.; Arain, M.A.; Baldocchi, D.; Bonan, G.B.; et al. Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science* **2010**, *329*, 834–838. [[CrossRef](#)]
11. Bonan, G.B. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* **2008**, *320*, 1444–1449. [[CrossRef](#)]

12. Nemani, R.R.; Keeling, C.D.; Hashimoto, H.; Jolly, W.M.; Piper, S.C.; Tucker, C.J.; Myneni, R.B.; Running, S.W. Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999. *Science* **2003**, *300*, 1560–1563. [[CrossRef](#)] [[PubMed](#)]
13. Running, S.W.; Nemani, R.R.; Heinsch, F.A.; Zhao, M.S.; Reeves, M.; Hashimoto, H. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *Bioscience* **2004**, *54*, 547–560. [[CrossRef](#)]
14. Churkina, G.; Running, S.W. Contrasting Climatic Controls on the Estimated Productivity of Global Terrestrial Biomes. *Ecosystems* **1998**, *1*, 206–215. [[CrossRef](#)]
15. Zhou, H.; Shao, J.; Liu, H.; Du, Z.; Zhou, L.; Liu, R.; Bernhofer, C.; Grünwald, T.; Dušek, J.; Montagnani, L.; et al. Relative importance of climatic variables, soil properties and plant traits to spatial variability in net CO<sub>2</sub> exchange across global forests and grasslands. *Agric. For. Meteorol.* **2021**, *307*. [[CrossRef](#)]
16. Li, P.; Peng, C.; Wang, M.; Li, W.; Zhao, P.; Wang, K.; Yang, Y.; Zhu, Q. Quantification of the response of global terrestrial net primary production to multifactor global change. *Ecol. Indic.* **2017**, *76*, 245–255. [[CrossRef](#)]
17. Yuan, W.; Zheng, Y.; Piao, S.; Ciais, P.; Lombardozzi, D.; Wang, Y.; Ryu, Y.; Chen, G.; Dong, W.; Hu, Z.; et al. Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.* **2019**, *5*, eaax1396. [[CrossRef](#)] [[PubMed](#)]
18. Wang, S.; Zhang, Y.; Ju, W.; Chen, J.M.; Ciais, P.; Cescatti, A.; Sardans, J.; Janssens, I.A.; Wu, M.; Berry, J.A.; et al. Recent global decline of CO<sub>2</sub> fertilization effects on vegetation photosynthesis. *Science* **2021**, *371*. [[CrossRef](#)]
19. Song, Y.; Jiao, W.Z.; Wang, J.; Wang, L.X. Increased Global Vegetation Productivity Despite Rising Atmospheric Dryness Over the Last Two Decades. *Earths Future* **2022**, *10*. [[CrossRef](#)]
20. Chen, R.N.; Liu, L.Y.; Liu, X.J. The Negative Impact of Excessive Moisture Contributes to the Seasonal Dynamics of Photosynthesis in Amazon Moist Forests. *Earths Future* **2022**, *10*. [[CrossRef](#)]
21. Green, J.K.; Berry, J.; Ciais, P.; Zhang, Y.; Gentine, P. Amazon rainforest photosynthesis increases in response to atmospheric dryness. *Sci. Adv.* **2020**, *6*, eabb7232. [[CrossRef](#)] [[PubMed](#)]
22. Chen, N.; Song, C.; Xu, X.; Wang, X.; Cong, N.; Jiang, P.; Zu, J.; Sun, L.; Song, Y.; Zuo, Y.; et al. Divergent impacts of atmospheric water demand on gross primary productivity in three typical ecosystems in China. *Agric. For. Meteorol.* **2021**, *307*, 108527. [[CrossRef](#)]
23. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)] [[PubMed](#)]
24. Mittelstadt, B.; Russell, C.; Wachter, S. Explaining Explanations in AI. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; 2019; pp. 279–288.
25. Besnard, S.; Koirala, S.; Santoro, M.; Weber, U.; Nelson, J.; Gütter, J.; Herault, B.; Kassi, J.; N’Guessan, A.; Neigh, C.; et al. Mapping global forest age from forest inventories, biomass and climate data. *Earth Syst. Sci. Data* **2021**, *13*, 4881–4896.
26. Stirnberg, R.; Cermak, J.; Kotthaus, S.; Haeffelin, M.; Andersen, H.; Fuchs, J.; Kim, M.; Petit, J.-E.; Favez, O. Meteorology-driven variability of air pollution (PM<sub>1</sub>) revealed with explainable machine learning. *Atmos. Chem. Phys.* **2021**, *21*, 3919–3948. [[CrossRef](#)]
27. Wang, H.; Yan, S.J.; Ciais, P.; Wigneron, J.P.; Liu, L.B.; Li, Y.; Fu, Z.; Ma, H.L.; Liang, Z.; Wei, F.L.; et al. Exploring complex water stress-gross primary production relationships: Impact of climatic drivers, main effects, and interactive effects. *Glob. Chang. Biol.* **2022**, *28*, 4110–4123. [[CrossRef](#)]
28. Boisvenue, C.; Running, S.W. Impacts of climate change on natural forest productivity—Evidence since the middle of the 20th century. *Glob. Chang. Biol.* **2006**, *12*, 862–882. [[CrossRef](#)]
29. Vitousek, P.M.; Ehrlich, P.R.; Ehrlich, A.H.; Matson, P.A. Net Primary Production: Original Calculations. *Science* **1987**, *235*, 730. [[CrossRef](#)]
30. Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D’Amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; et al. Terrestrial ecoregions of the worlds: A new map of life on Earth. *Bioscience* **2001**, *51*, 933–938. [[CrossRef](#)]
31. Rödiger, E.; Cuntz, M.; Heinke, J.; Rammig, A.; Huth, A. Spatial heterogeneity of biomass and forest structure of the Amazon rain forest: Linking remote sensing, forest modelling and field inventory. *Glob. Ecol. Biogeogr.* **2017**, *26*, 1292–1302. [[CrossRef](#)]
32. Bauer, L.; Knapp, N.; Fischer, R. Mapping Amazon Forest Productivity by Fusing GEDI Lidar Waveforms with an Individual-Based Forest Model. *Remote Sens.* **2021**, *13*, 4540. [[CrossRef](#)]
33. Running, S.; Mu, Q.; Zhao, M. MOD17A3H MODIS/Terra Net Primary Production Yearly L4 Global 500m SIN Grid V006; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2015. [[CrossRef](#)]
34. Zhao, M.; Heinsch, F.A.; Nemani, R.R.; Running, S.W. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* **2005**, *95*, 164–176. [[CrossRef](#)]
35. Fensholt, R.; Sandholt, I.; Rasmussen, M.S.; Stisen, S.; Diouf, A. Evaluation of satellite based primary production modelling in the semi-arid Sahel. *Remote Sens. Environ.* **2006**, *105*, 173–188. [[CrossRef](#)]
36. Turner, D.P.; Ritts, W.D.; Cohen, W.B.; Gower, S.T.; Running, S.W.; Zhao, M.; Costa, M.H.; Kirschbaum, A.A.; Ham, J.M.; Saleska, S.R.; et al. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sens. Environ.* **2006**, *102*, 282–292. [[CrossRef](#)]
37. Fu, Z.; Ciais, P.; Prentice, I.C.; Gentine, P.; Makowski, D.; Bastos, A.; Luo, X.; Green, J.K.; Stoy, P.C.; Yang, H.; et al. Atmospheric dryness reduces photosynthesis along a large range of soil water deficits. *Nat. Commun.* **2022**, *13*, 989. [[CrossRef](#)]
38. Zhu, Z.; Piao, S.; Myneni, R.B.; Huang, M.; Zeng, Z.; Canadell, J.G.; Ciais, P.; Sitch, S.; Friedlingstein, P.; Arneeth, A.; et al. Greening of the Earth and its drivers. *Nat. Clim. Chang.* **2016**, *6*, 791. [[CrossRef](#)]



39. Wan, Z.; Hook, S.; Hulley, G. *MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V061*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2021. [[CrossRef](#)]
40. Huang, M.; Piao, S.; Ciais, P.; Peñuelas, J.; Wang, X.; Keenan, T.F.; Peng, S.; Berry, J.A.; Wang, K.; Mao, J.; et al. Air temperature optima of vegetation productivity across global biomes. *Nat. Ecol. Evol.* **2019**, *3*, 772–779. [[CrossRef](#)]
41. Abatzoglou, J.T.; Dobrowski, S.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [[CrossRef](#)]
42. McNally, A. *FLDAS Noah Land Surface Model L4 Global Monthly 0.1 × 0.1 Degree (MERRA-2 and CHIRPS)*; Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA, 2018. [[CrossRef](#)]
43. Howell, T.A.; Dusek, D.A. Comparison of Vapor-Pressure-Deficit Calculation Methods—Southern High Plains. *J. Irrig. Drain. Eng.* **2001**, *127*, 329–330. [[CrossRef](#)]
44. Zhang, T.; Zhang, Y.; Xu, M.; Zhu, J.; Chen, N.; Jiang, Y.; Huang, K.; Zu, J.; Liu, Y.; Yu, G. Water availability is more important than temperature in driving the carbon fluxes of an alpine meadow on the Tibetan Plateau. *Agric. For. Meteorol.* **2018**, *256*, 22–31. [[CrossRef](#)]
45. Jacobson, A.R.; Schuldt, K.N.; Miller, J.B.; Oda, T.; Tans, P.; Arlyn, A.; Mund, J.; Ott, L.; Collatz, G.J.; Aalto, T.; et al. *CarbonTracker CT2019B*; NOAA Earth System Research Laboratory, Global Monitoring Division: Bondville, IL, USA, 2020. [[CrossRef](#)]
46. Shangquan, W.; Dai, Y.; Duan, Q.; Liu, B.; Yuan, H. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **2014**, *6*, 249–263. [[CrossRef](#)]
47. Friedl, M.; Sulla-Menashe, D. *MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2015. [[CrossRef](#)]
48. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)] [[PubMed](#)]
49. Chen, T.Q.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining (Kdd'16), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
50. Rossel, R.A.V.; Lee, J.; Behrens, T.; Luo, Z.; Baldock, J.; Richards, A. Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls. *Nat. Geosci.* **2019**, *12*, 547. [[CrossRef](#)]
51. Quinlan, J.R. Learning with Continuous Classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; Volume 10, pp. 475–476.
52. Barnes, E.A.; Toms, B.; Hurrell, J.W.; Ebert-Uphoff, I.; Anderson, C.; Anderson, D. Indicator Patterns of Forced Change Learned by an Artificial Neural Network. *J. Adv. Model. Earth Syst.* **2020**, *12*. [[CrossRef](#)]
53. Molina, M.J.; Gagne, D.J.; Prein, A.F. A Benchmark to Test Generalization Capabilities of Deep Learning Methods to Classify Severe Convective Storms in a Changing Climate. *Earth Space Sci.* **2021**, *8*. [[CrossRef](#)]
54. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. [[CrossRef](#)]
55. Rasp, S.; Thuerey, N. Data-Driven Medium-Range Weather Prediction with a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench. *J. Adv. Model. Earth Syst.* **2021**, *13*. [[CrossRef](#)]
56. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
57. Aas, K.; Jullum, M.; Løland, A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artif. Intell.* **2021**, *298*. [[CrossRef](#)]
58. Shapley, L.S. Stochastic Games. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 1095–1100. [[CrossRef](#)]
59. Strumbelj, E.; Kononenko, I. An Efficient Explanation of Individual Classifications using Game Theory. *J. Mach. Learn. Res.* **2010**, *11*, 1–18. [[CrossRef](#)]
60. Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [[CrossRef](#)]
61. Lundberg, S.M.; Erion, G.G.; Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**, arXiv:1802.03888.
62. Lli, H.; Zhang, F.; Li, Y.; Wang, J.; Zhang, L.; Zhao, L.; Cao, G.; Zhao, X.; Du, M. Seasonal and inter-annual variations in CO<sub>2</sub> fluxes over 10 years in an alpine shrubland on the Qinghai-Tibetan Plateau, China. *Agric. For. Meteorol.* **2016**, *228*, 95–103. [[CrossRef](#)]
63. Xu, M.J.; Wang, H.M.; Wen, X.F.; Zhang, T.; Di, Y.B.; Wang, Y.D.; Wang, J.L.; Cheng, C.P.; Zhang, W.J. The full annual carbon balance of a subtropical coniferous plantation is highly sensitive to autumn precipitation. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)]
64. Doughty, C.E.; Goulden, M.L. Are tropical forests near a high temperature threshold? *J. Geophys. Res. Biogeogr.* **2008**, *113*. [[CrossRef](#)]
65. Phillips, O.L.; Aragao, L.E.O.C.; Lewis, S.L.; Fisher, J.B.; Lloyd, J.; Lopez-Gonzalez, G.; Malhi, Y.; Monteagudo, A.; Peacock, J.; Quesada, C.A.; et al. Drought Sensitivity of the Amazon Rainforest. *Science* **2009**, *323*, 1344–1347. [[CrossRef](#)]
66. Medlyn, B.E.; Dreyer, E.; Ellsworth, D.; Forstreuter, M.; Harley, P.C.; Kirschbaum, M.U.F.; LE Roux, X.; Montpied, P.; Strassmeyer, J.; Walcroft, A.; et al. Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data. *Plant Cell Environ.* **2002**, *25*, 1167–1179. [[CrossRef](#)]
67. Lloyd, J.; Farquhar, G.D. Effects of rising temperatures and [CO<sub>2</sub>] on the physiology of tropical forest trees. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 1811–1817. [[CrossRef](#)]

68. Kattge, J.; Knorr, W. Temperature acclimation in a biochemical model of photosynthesis: A reanalysis of data from 36 species. *Plant Cell Environ.* **2007**, *30*, 1176–1190. [[CrossRef](#)]
69. Graham, E.A.; Mulkey, S.S.; Kitajima, K.; Phillips, N.G.; Wright, S.J. Cloud cover limits net CO<sub>2</sub> uptake and growth of a rainforest tree during tropical rainy seasons. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 572–576. [[CrossRef](#)] [[PubMed](#)]
70. Eiserhardt, W.L.; Svenning, J.-C.; Kissling, W.D.; Balslev, H. Geographical ecology of the palms (Arecaceae): Determinants of diversity and distributions across spatial scales. *Ann. Bot.* **2011**, *108*, 1391–1416. [[CrossRef](#)] [[PubMed](#)]
71. Bertani, G.; Wagner, F.H.; Anderson, L.O.; Aragão, L.E.O.C. Chlorophyll Fluorescence Data Reveals Climate-Related Photosynthesis Seasonality in Amazonian Forests. *Remote Sens.* **2017**, *9*, 1275. [[CrossRef](#)]
72. Durand, M.; Murchie, E.H.; Lindfors, A.V.; Urban, O.; Aphalo, P.J.; Robson, T.M. Diffuse solar radiation and canopy photosynthesis in a changing environment. *Agric. For. Meteorol.* **2021**, *311*, 108684. [[CrossRef](#)]
73. Roderick, M.L.; Farquhar, G.; Berry, S.L.; Noble, I.R. On the direct effect of clouds and atmospheric particles on the productivity and structure of vegetation. *Oecologia* **2001**, *129*, 21–30. [[CrossRef](#)]
74. Burgess, A.J.; Retkute, R.; Preston, S.P.; Jensen, O.E.; Pound, M.P.; Pridmore, T.P.; Murchie, E.H. The 4-Dimensional Plant: Effects of Wind-Induced Canopy Movement on Light Fluctuations and Photosynthesis. *Front. Plant Sci.* **2016**, *7*, 1392. [[CrossRef](#)]
75. Burgess, A.J.; Gibbs, J.A.; Murchie, E.H. A canopy conundrum: Can wind-induced movement help to increase crop productivity by relieving photosynthetic limitations? *J. Exp. Bot.* **2019**, *70*, 2371–2380. [[CrossRef](#)]
76. Sternberg, L.D.S.; Moreira, M.Z.; Martinelli, L.A.; Victoria, R.L.; Barbosa, E.M.; Bonates, L.C.; Nepstad, D.C. Carbon dioxide recycling in two Amazonian tropical forests. *Agric. For. Meteorol.* **1997**, *88*, 259–268. [[CrossRef](#)]
77. de Langre, E. Effects of Wind on Plants. *Annu. Rev. Fluid Mech.* **2008**, *40*, 141–168. [[CrossRef](#)]
78. Sullivan, A.L. Wildland surface fire spread modelling, 19902007. 2: Empirical and quasi-empirical models. *Int. J. Wildland Fire* **2009**, *18*, 369–386. [[CrossRef](#)]
79. Yordanova, R.Y.; Popova, L.P. Flooding-induced changes in photosynthesis and oxidative status in maize plants. *Acta Physiol. Plant.* **2007**, *29*, 535–541. [[CrossRef](#)]
80. Grossiord, C.; Buckley, T.N.; Cernusak, L.A.; Novick, K.A.; Poulter, B.; Siegwolf, R.T.W.; Sperry, J.S.; McDowell, N.G. Plant responses to rising vapor pressure deficit. *New Phytol.* **2020**, *226*, 1550–1566. [[CrossRef](#)] [[PubMed](#)]
81. López, J.; Way, D.A.; Sadok, W. Systemic effects of rising atmospheric vapor pressure deficit on plant physiology and productivity. *Glob. Chang. Biol.* **2021**, *27*, 1704–1720. [[CrossRef](#)] [[PubMed](#)]
82. He, B.; Chen, C.; Lin, S.; Yuan, W.; Chen, H.W.; Chen, D.; Zhang, Y.; Guo, L.; Zhao, X.; Liu, X.; et al. Worldwide impacts of atmospheric vapor pressure deficit on the interannual variability of terrestrial carbon sinks. *Natl. Sci. Rev.* **2022**, *9*, nwab150. [[CrossRef](#)] [[PubMed](#)]
83. Novick, K.A.; Ficklin, D.; Stoy, P.C.; Williams, C.A.; Bohrer, G.; Oishi, A.; Papuga, S.; Blanken, P.D.; Noormets, A.; Sulman, B.; et al. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nat. Clim. Chang.* **2016**, *6*, 1023–1027. [[CrossRef](#)]
84. Li, Q.; Wei, M.; Li, Y.; Feng, G.; Wang, Y.; Li, S.; Zhang, D. Effects of soil moisture on water transport, photosynthetic carbon gain and water use efficiency in tomato are influenced by evaporative demand. *Agric. Water Manag.* **2019**, *226*. [[CrossRef](#)]
85. Yu, X.; Zhao, M.; Wang, X.; Jiao, X.; Song, X.; Li, J. Reducing vapor pressure deficit improves calcium absorption by optimizing plant structure, stomatal morphology, and aquaporins in tomatoes. *Environ. Exp. Bot.* **2022**, *195*, 104786. [[CrossRef](#)]
86. Shamshiri, R.; Che Man, H.; Zakaria, A.J.; Beveren, P.V.; Wan Ismail, W.I.; Ahmad, D. Membership function model for defining optimality of vapor pressure deficit in closed-field cultivation of tomato. In Proceedings of the III International Conference on Agricultural and Food Engineering, Kuala Lumpur, Malaysia, 21 March 2017; pp. 281–290.
87. Yala, A.; Lehman, C.; Schuster, T.; Portnoi, T.; Barzilay, R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. *Radiology* **2019**, *292*, 60–66. [[CrossRef](#)]
88. Silva, S.J.; Keller, C.A.; Hardin, J. Using an Explainable Machine Learning Approach to Characterize Earth System Model Errors: Application of SHAP Analysis to Modeling Lightning Flash Occurrence. *J. Adv. Model. Earth Syst.* **2022**, *14*. [[CrossRef](#)]
89. García, M.V.; Aznarte, J.L. Shapley additive explanations for NO<sub>2</sub> forecasting. *Ecol. Inform.* **2020**, *56*. [[CrossRef](#)]
90. Cheng, S.; Cheng, L.; Qin, S.; Zhang, L.; Liu, P.; Liu, L.; Xu, Z.; Wang, Q. Improved Understanding of How Catchment Properties Control Hydrological Partitioning Through Machine Learning. *Water Resour. Res.* **2022**, *58*. [[CrossRef](#)]