TITLE: Local interpretation of machine learning models in remote sensing with SHAP: the case of global climate constraints on photosynthesis phenology

Adrià Descals[1,2], Aleixandre Verger[1,2,3], Gaofei Yin[4], Iolanda Filella[1,2], and Josep Peñuelas[1,2]

[1]CREAF, Cerdanyola del Vallès, Barcelona 08193, Catalonia, Spain

[2]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Barcelona 08193, Catalonia, Spain

[3]CIDE, CSIC-UV-GV, València 46113, Spain

[4]Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China

Abstract

Data-driven models using machine learning have been widely used in remote sensing applications such as the retrieval of biophysical variables and land cover classification. However, these models behave as a 'black box', meaning that the relationships between the input and predicted variables are hard to interpret. Recent regression models that downscale sun-induced fluorescence (SIF) with MODIS and weather variables are an example. The impact of weather variables on the predicted SIF in these models is unknown. The explanation of such weather-SIF relationships would aid in the understanding of climate-related constraints on photosynthesis phenology since SIF is a proxy of gross primary productivity. Here, we used SHapley Additive exPlanations (SHAP) –a novel technique based on game theory– for explaining the contribution of input variables to the individual predictions in a machine learning model. We explored the capabilities of this technique with a weather-SIF model. The regression model predicted ESA-TROPOSIF measurements from ERA5-Land air temperature, shortwave radiation, and vapor-pressure-deficit (VPD) data. The SHAP values of the model were estimated at the start and end

of the growing season for the entire globe. These values depicted the global constraints of the three climate variables on the photosynthetically active season and confirmed existing knowledge on the limiting factors of terrestrial photosynthesis with unprecedented spatial detail. Radiation was the limiting factor in tropical rainforest and VPD constrained the start and end of the growing season in tropical dryland ecosystems. In extra-tropical regions, temperature was the main limiting factor during the start of the growing season, but both temperature and radiation constrained photosynthesis at the end of the growing season. This technique may help future remote sensing studies that require the use of non-interpretable machine-learning regression models and explain how input variables contribute to the model prediction in a spatiotemporally explicit manner.

Keywords: SHapley Additive exPlanations, explainable machine learning, local interpretation, sun-induced fluorescence, vegetation phenology, climate constraints, photosynthesis dynamics.

1. INTRODUCTION

The field of vegetation phenology has gained attention recently, with the number of publications on phenology quintupling in the last two decades (Fu et al., 2020). The transition between the dormant and growing season and the climate factors determining it have been explained globally by models employing climate thresholds. Jolly et al. (2005) proposed the growing season index (GSI), which is calculated with cut-off functions on three weather variables: temperature, vapor-pressure-deficit (VPD), and day length. These cut-off functions represent thresholds that were subjectively defined by expert knowledge and are constant for the entire globe. The cut-off functions converts the climate factors into a GSI value; low values of tempereature, VPD, and day length lead to low values of GSI. The GSI shows seasonal changes throughout the year and aims to replicate a spectral index (e.g., normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI)) or a biophysical variable (e.g., leaf area index (LAI)).

Other studies used regression models to fit climate reanalysis datasets to vegetation indices or biophysical variables. Both standard machine learning regression –such as random forests (Li and Xiao, 2019)–, and deep learning (Ahmad et al., 2020) have been used given their ability to fit non-linear and non-parametric relationships between dependent and independent variables. This methodology predicts vegetation indices or biophysical variables, and the climate thresholds are, thus, defined empirically and more accurately than the cut-off functions (Jolly et al., 2005). However, an important flaw in machine learning models is the lack of interpretability. Contrarily to Jolly et al. (2005), the impact of the weather variables on the predicted outcome remains challenging in machine learning models. Recent regression models using vegetation indices and weather variables to downscale sun-induced fluorescence (weather-SIF) are an example. For instance, the GOSIF product (Li and Xiao, 2019) uses a machine learning regression model to fit SIF with weather variables and EVI. Another product is the SIFnet (Gensheimer et al., 2022), which downscales SIF measurements from TROPOspheric Monitoring Instrument (TROPOMI) using auxiliary data.

Feature importance in machine learning is a technique used to determine the importance of input variables in predicting the target variable of a model. One common method is to use tree-based models, such as Random Forest or Gradient Boosting, which provide a feature importance score (i.e. Gini index) based on how much each feature contributes to reducing the impurity of the tree nodes (Breiman, 2001). These techniques provide a score that reflects the overall (or global) importance of each input variable in the model. For example, in a weather-SIF model, the Gini index would rank the weather variables by their overall importance in the model. This overall importance, however, would neglect how relevant each weather variable is in a specific observation (i.e. the importance of the weather variables in a particular pixel and moment of the time series). In this context, local feature importance provides a more comprehensive and interpretable way of measuring feature importance. By calculating local feature importance, one can gain a better understanding of which weather variables are most important for those single predictions.

A state-of-the-art local interpretation method for model explainability is SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). SHAP has been used for understanding the risk of hypoxemia during anaesthesia (Lundberg et al., 2018), interpret the features that make an online product review helpful (Meng et al., 2020), understanding the pollutant removal mechanisms in wastewater treatment plants (Wang et al., 2022), or analysing large-scale biobank data for potential gene–gene and gene–environment interactions (Johnsen et al., 2021). This technique is, however, novel in remote sensing studies and it only started recently to be employed to improve understanding of spatial and temporal relationships of geospatial data modeling (Li et al., 2022; Zhan et al., 2022). In a weather-SIF model, the use of a technique such as SHAP for local feature importance could potentially provide insights into the weather factors that are constraining photosynthetic phenology. By explaining the contribution of input variables to the individual predictions in a machine learning model, SHAP has the potential to elucidate how the weather variables specifically influence the individual prediction of SIF in a spatiotemporally explicit manner.

The aim of this study was to demonstrate the capability of SHAP to explain the correlation between geospatial gridded data and model predictions in a machine learning model. We used the case of weather-SIF models (Li and Xiao, 2019) to determine the global constraints of weather variables on vegetation activity. To achieve the objective of the study, we sampled SIF measurements and temperature, shortwave radiation, and VPD in specific sites at the global scale. Then, we trained a machine learning model that predicted SIF from weather variables and applied the model for the entire globe. We estimated two phenological metrics, the start of season (SoS) and end of season (EoS), from the predicted SIF time series. The SHAP technique was used to describe the effect of weather variables on SIF at the timing of the SoS and the EoS and, thus, determine the climate constraints on vegetation phenology at the global scale. Finally, we discussed and validated the model interpretation with SHAP compared with the current understanding on photosynthesis dynamics.

2. DATA

2.1. TROPOSIF global sun-induced fluorescence dataset

We used the TROPOSIF L2B product (Guanter et al., 2021), which provides non-gridded SIF measurements derived from observations in the 743–758 nm and 665–785 nm part of the spectrum from the TROPOspheric Monitoring Instrument (TROPOMI) sensor. TROPOSIF provides SIF measurements for the entire land area of the globe at a spatial resolution of 3.5 km × 5.5 km at nadir. The observations were made by the TROPOMI sensor onboard Sentinel-5. The methodology that generates SIF uses a retrieval method that fits the top-of-atmosphere radiances with SIF training sets (Guanter et al., 2015). We used the data for all the product time coverage, which spans from May 2018 to April 2021. We used the SIF_745_corr, which represents SIF in the 743–758 nm window. SIF observations that presented a cloud cover greater than 50% were rejected. The TROPOSIF L2B product already masks observations with cloud cover greater than 80%, a view zenith angle greater than 60°, and a solar zenith angle greater than 70° since the SIF retrievals at these conditions are affected by directional effects (including shadow influence) and are not reliable (Joiner et al., 2020).

2.2. ERA5-Land hourly data

We used gridded climatic data from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis data version 5 (ERA5-Land) hourly dataset (Muñoz-Sabater et al., 2021). The ERA5-Land is a reanalysis dataset that covers a period from 1950 to present. The data were produced by a combination of modelled data with observations collected across the globe and improves upon the ERA-5 since it has higher spatial resolution (about 9 km) at the same temporal resolution (1 hour). We used the near-surface air temperature (2m temperature) and the surface solar radiation downwards –the solar shortwave radiation that reaches the surface of the Earth. We also estimated VPD –atmospheric demand for evapotranspiration– using the ERA5 near-surface air temperature and dew point temperature (2m dew point temperature) as described by Barkhordarian et al. (2019).

We chose air temperature, radiation, and VPD for two reasons. First, air temperature, shortwave radiation, and vapor-pressure deficit are important environmental factors that affect vegetation photosynthesis. Air temperature influences the rate of photosynthesis by affecting enzyme activity, the solubility of carbon dioxide in water, and the diffusion rate of gasses through plant tissues. Shortwave radiation, particularly photosynthetically active radiation (PAR), provides the energy required for photosynthesis. Vapor-pressure deficit affects the rate of transpiration, and therefore, the availability of water for photosynthesis. Second, we wanted to replicate the weather SIF-model used in a prior study (Li and Xiao, 2019), which also used air temperature, radiation, and VPD as input variables, and explain spatially and temporally the importance of these variables in the SIF predictions. These three variables were also used in the growing season index (Jolly et al., 2005). The growing season index is the result of a parametric model that also assessed the significance of air temperature, photoperiod as a proxy for radiation, and VPD as factors that explain global leaf phenology.

3. METHODS

3.1. Extraction of training pairs (TROPOSIF - ERA5) in BELMANIP2 sites

We collected pairs of TROPOSIF measurements and ERA5 observations as training data. These data were used to train a machine learning model that predicted SIF from temperature, shortwave radiation, and VPD. We extracted the TROPOSIF and ERA5 data from the BEnchmark Land Multisite ANalysis and Intercomparison of Products version 2 (BELMANIP2) sites for the period going from May 2018 to April 2021. The BELMANIP2 consists of a collection of 445 sites of homogeneous areas that include the most representative land covers of the world (Weiss et al., 2014). BELMANIP2 sites have been used to validate global satellite datasets, such as reflectance products (Franch et al., 2017), biophysical variables (Verger et al., 2014), or phenology metrics (Kandasamy and Fernandes, 2015). We excluded bare soil, cropland, and other non-natural or non-vegetated land covers, which resulted in 233 sites (see location map of the BELMANIP2 points in Fig. S1) including the following land covers: evergreen needleleaf forests (ENF), deciduous needleleaf forests (DNF), deciduous broadleaf forests (DBF), mixed forests (MX), closed shrublands (CSH), open shrublands (OSH), woody savannah (WSA), savannah (SAV), and grasslands (GRA). The land cover types were determined for each BELMANIP2 site with the 'LC_Type1' layer of the MCD12Q1v6 product derived from MODIS. The ERA5 data were extracted at hourly temporal resolution, and then aggregated daily. The TROPOSIF dataset provides daily non-gridded SIF measurements. We, thus, extracted the daily SIF observations that were located the closest to a BELMANIP2 site. SIF observations more than 5 km away from a BELMANIP2 site were rejected. A total of 140,969 pairs of data were generated from the 233 sites for the May 2018 - April 2021 period.

3.2. Weather – SIF model

We used Gradient Boosting regression (Friedman, 2001) to fit ERA5 data (air temperature, shortwave radiation, and VPD) to TROPOSIF observations. Gradient Boosting is an ensemble model that uses decision trees as weak learners, where decision trees are trained sequentially by correcting the errors of a previously trained decision tree. The performance of the decision trees is improved using a loss function. A loss function is a function that measures the difference between the predicted output and the actual output in a machine learning algorithm. The loss function is used to optimize the model by updating its parameters in a way that minimizes the loss. We used Gradient Boosting because it is a common machine learning model used by the research community, it can easily capture non-linear and non-parametric relationships, and has very fast training and deployment times (Bentéjac et al., 2021). We trained the Gradient Boosting model with 75% of the data and kept the remaining 25% for validation. The accuracy metrics that we reported are the mean error (ME: difference between predicted minus observed), root-mean-squared error (RMSE), and the coefficient of determination (R^2).

We performed hyperparameter tuning to find the optimal parameters of the Gradient Boosting regression model. Hyperparameters are parameters that are set before the training and control the learning process of the machine learning model (Yang and Shami, 2020). The hyperparameter tuning consisted of a random search for different combinations of hyperparameter values. The range of hyperparameters is depicted in Table S1. For each combination of parameters, the RMSE of the model was evaluated with a 4-fold partition (75% training and 25% validation). The accuracy of the Gradient Boosting model was tested using 100 different hyperparameter combinations, and the model with the lowest root-mean-squared error (RMSE) was selected.

3.3. Local interpretation with SHAP

The local interpretation of the Gradient Boosting was done with SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). SHAP is a state-of-the-art technique for machine learning explainability; it aims to explain the correlations between input and output variables in any machine learning model, in both regression and classification algorithms. SHAP is based on the Shapley values (Shapley, 1953) of game theory and is categorized as a local interpretation technique –it explains the contribution of the input variables to individual model predictions.

SHAP values represent the marginal effect of the input variables on the prior expectation of the model output. A negative SHAP value for a given input variable implies that the input variable has a decreasing effect on the model output, and vice versa; a positive SHAP value means that the input variable increases the model output. The greater the absolute value of SHAP, the greater the impact of the input variable on the predicted value. The model prediction is the prior expectation of the model plus the summatory of the SHAP values of all input variables. The mathematical formulation for SHAP is described in (Lundberg and Lee, 2017). In this study, we used the SHAP package in Python and used the approximation method for tree-based machine learning models (Lundberg et al., 2020). In the weather-SIF model, SHAP can potentially evaluate the importance of the weather observations on individual SIF predictions for any given date and location.

3.4. Land Surface Phenology metric estimation

Phenological metrics were extracted from the predicted SIF time series. To achieve this, we first predicted SIF at the global scale using the Gradient Boosting model. Then, we estimated two phenological metrics, the start of season (SoS) and end of season (EoS), from the predicted SIF time series. The SoS and EoS were extracted using the Maximum Separation (MS) method (Descals et al., 2020b). MS is a threshold-based method that can effectively estimate phenological metrics without the need of time series pre-processing prior to the phenology extraction. These types of time series pre-processing include smoothing and interpolation techniques that are applied to improve the robustness of the phenology estimates. However, these pre-processing steps may produce a time series that differs from the original, resulting in biases in the phenology estimates. The Maximum Separation method can be applied directly to the original time series.

As any threshold-based method, the MS required a threshold value to calculate the SoS and EoS from the SIF time series. For each pixel, we defined a dynamic threshold, which represented 20% of the amplitude plus the minimum SIF value in the time series. The MS runs a moving window that calculates the proportion of observations that are above the threshold before and after the central day of the moving window. We determined a moving window size of 120 days (including the days before and after the central day). The moving window is applied for every day of the time series. SoS and EoS are defined as the days of the year when the difference in proportions (before minus after) reaches the minimum and maximum during the year. The implementation of the MS method is available in Python and in Google Earth Engine (Descals et al., 2020b).

4. RESULTS

The combination of hyperparameters that lead to the lowest RMSE in the validation dataset is shown in Table S1. For these hyperparameters, RMSE was 0.21 mW m^{-2} sr^{-1} nm^{-1}, ME was -0.00 mW m^{-2} sr^{-1} nm^{-1}, and R^2 was 0.38. The accuracy metrics differed slightly depending on the land cover type (Fig. S2). The lowest accuracy was found in DBF (ME = 0.14 mW m^{-2} sr^{-1} nm^{-1}, RMSE = 0.35 mW m^{-2} sr^{-1} nm^{-1}), while the accuracy of the other land covers was close to the overall accuracy, with a minimal ME (ranging from -0.04 mW m^{-2} sr^{-1} nm^{-1} in OSH to 0.06 mW m^{-2} sr^{-1} nm^{-1} in MX) and similar RMSE (ranging from 0.15 mW m^{-2} sr^{-1} nm^{-1} in CSH to 0.23 mW m^{-2} sr^{-1} nm^{-1} in MX). The model saturated the predicted values to 0.5 mW m^{-2} sr^{-1} nm^{-1} in observations with high

SIF. Some observations above that value were underestimated, particularly in DBF and MX, but also in GRA and SAV. Overall, the model fitted the data without substantial biases for SIF observations below 0.5 mW m^{-2} sr^{-1} nm^{-1}, which were the bulk of SIF observations. The cross-site validation did not differ substantially from the overall statistics except for the mean error. RMSE was 0.21 mW m^{-2} sr^{-1} nm^{-1}, ME was -7.22 mW m^{-2} sr^{-1} nm^{-1}, and R^2 was 0.37. Cross-site validation consisted of a data partition in which the sites were partitioned in 4 folds. The accuracy metrics were evaluated using 3 folds (75% of sites) for training and 1 fold (25% of sites) for testing. Time series for two BELMANIP2 sites exemplify the predicted SIF compared to the TROPOSIF measurements (Fig. 1), and show that the model replicates the seasonality of the observed SIF, as also shown in the comparison between phenology metrics estimated with observed and predicted SIF (Fig. S3 and Fig. S4).

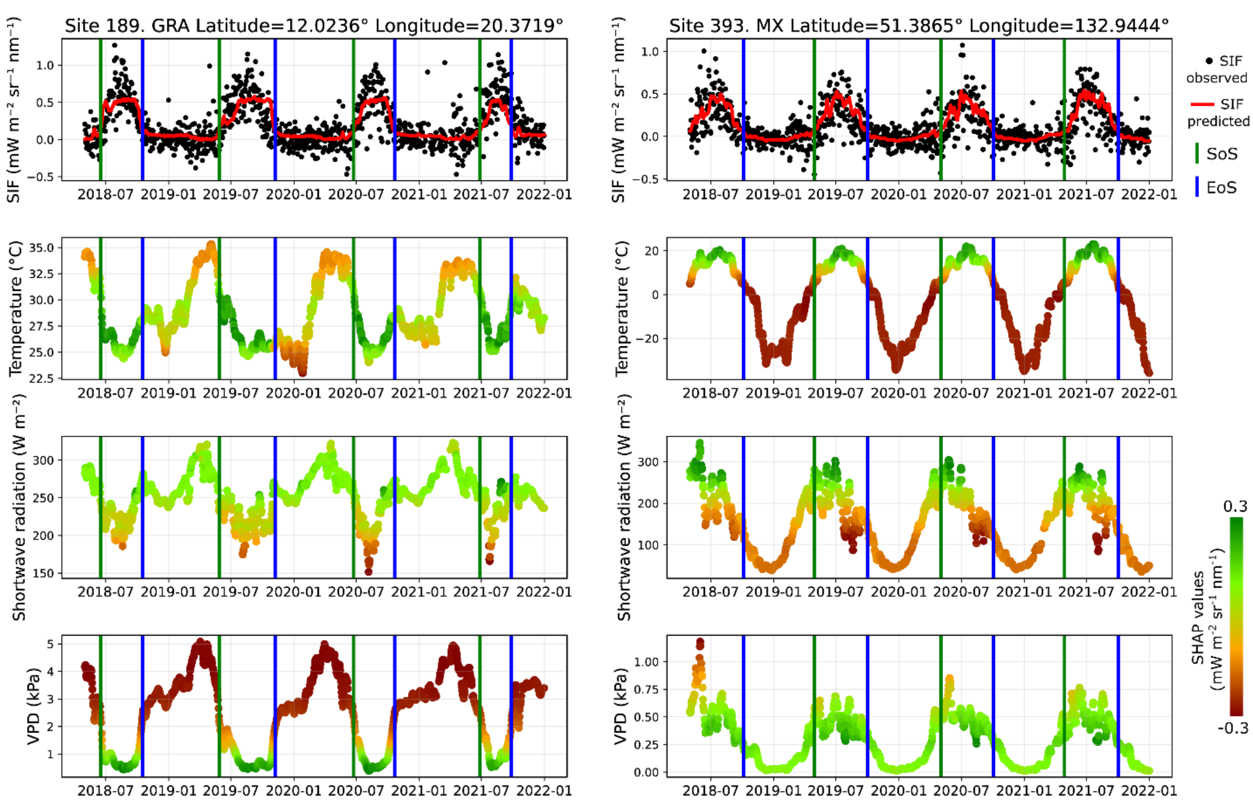

Figure 1. Time series of observed and predicted sun-induced fluorescence (SIF), air temperature, shortwave radiation, and vapor-pressure-deficit (VPD) in one grassland (GRA) and one mixed forest (MX) sites of the BELMANIP2 network. Vertical green lines depict the start of the growing season and blue lines represent the end of the growing season derived from predicted SIF time series. The observed SIF was extracted from the TROPOSIF dataset, while the predicted SIF was estimated with three climate variables using a machine learning regression model. Colours in the air temperature, shortwave radiation, and VPD time series depict the SHAP values. SHAP values indicate the impact of the input variables on the model mean SIF. Negative SHAP values mean that the input variable decreases the predicted SIF.

The time series of SHAP values represented the impact of the input climate variables on the predicted SIF. For instance, both low temperature and low shortwave radiation during winter were the most limiting factors in the BELMANIP2 site 393 (Fig. 1), located mid-latitude in a temperate climate. In BELMANIP2 site 189 –a site in a dryland ecosystem– the SHAP values indicate that seasonal changes in VPD determined the SIF seasonality, with the growing season occurring when VPD values decrease to their annual minimum. These SHAP time series show the seasonal climate constraints throughout the year, and the constraints can be extracted at the start and end of the growing season. For example, in the site covering a dryland ecosystem (site 189), the SHAP values at the end of the growing season 2019 were 0.18 mW m^{-2} sr^{-1} nm^{-1} for air temperature, 0.03 mW m^{-2} sr^{-1} nm^{-1} for shortwave radiation and -0.26 mW m^{-2} sr^{-1} nm^{-1} for VPD. The low SHAP value for VPD means that this variable had a negative contribution on the prior expectation of the SIF model (0.13 mW m^{-2} sr^{-1} nm^{-1}), indicating that VPD was constraining vegetation activity at that moment of the year. The predicted SIF at the end of season was 0.08

mW m^{-2} sr^{-1} nm^{-1}, which is the result of adding the SHAP values (0.18 + 0.03 – 0.26 mW m^{-2} sr^{-1} nm^{-1}) to the prior expectation of the model (0.13 mW m^{-2} sr^{-1} nm^{-1}) (Fig. S5).

SHAP values for the BELMANIP2 sites showed that high temperature and shortwave radiation had an overall positive impact on modelled SIF, while higher VPD had a negative impact (Fig. 2). VPD was the variable with the highest overall importance followed by temperature and shortwave radiation. The most extreme VPD values had an effect of approximately -0.3 and 0.3 mW m^{-2} sr^{-1} nm^{-1}, while the lowest and highest shortwave radiation had a lower effect, approximately -0.1 and 0.15 mW m^{-2} sr^{-1} nm^{-1}.

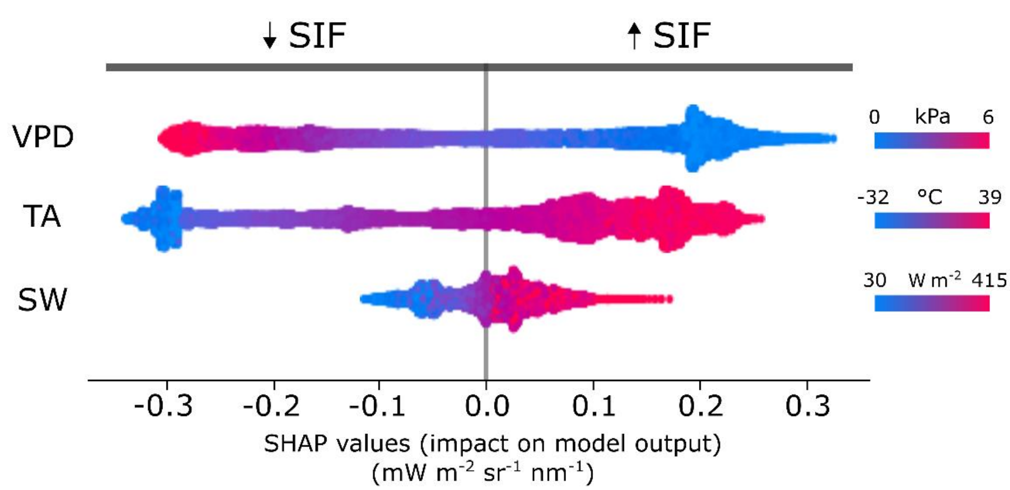

Figure 2. SHAP values of three input variables (air temperature (TA), shortwave radiation (SW), and vapor-pressure-deficit (VPD)) in a machine learning model that predicts sun-induced fluorescence (SIF). The SHAP values were estimated for 35,242 site-year observations of the BELMANIP2 network (25% of the total observations were kept for model validation). SHAP values indicate the contribution of the input climate variables on the mean SIF. Negative SHAP values mean that the input variable decreases the

predicted SIF and vice versa. The higher the absolute SHAP value, the higher the impact on predicted SIF. Color bars indicate the range of values (minimum to maximum) for each climate variable.

The impact of temperature, shortwave radiation, and VPD on predicted SIF differed spatially during the SoS and EoS. The maps of SHAP values at the SoS and EoS show that it was mostly in extratropical areas that temperature constrained SIF (Fig. 3 and Fig. S6). However, VPD was the highest constraint in tropical dryland ecosystems, while radiation was the limiting factor in tropical rainforests. The impact of temperature, shortwave radiation and VPD differed in some regions depending on whether it was the start or the end of the season. The most prominent difference was observed in extratropical regions. Temperature was the only factor that constrained SIF during the SoS in the Northern Hemisphere but both temperature and secondarily radiation constrained SIF in extratropical regions during the EoS.

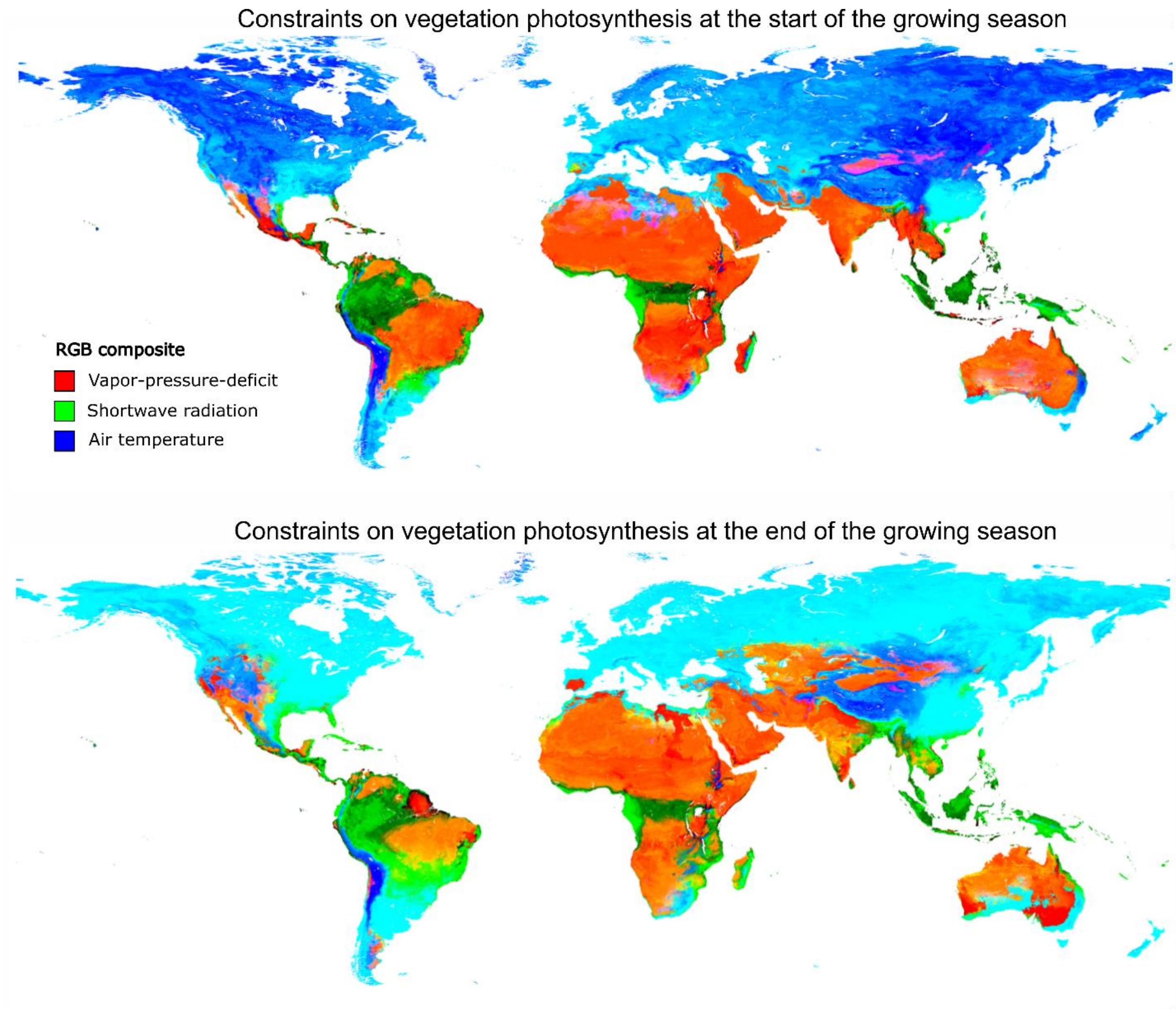

Figure 3. Maps of the climate constraints on photosynthesis at the start and end of the growing season. The maps represent an RGB composite of SHAP values. SHAP values were estimated from a machine learning model that fitted sun-induced fluorescence (SIF) with three climate variables: air temperature, shortwave radiation, and vapor-pressure-deficit. The SoS and EoS were estimated from daily averaged predicted SIF time series for the 2012-2021 period. The three climate variables were extracted from the ERA5-Land dataset and daily averaged for the 2012-2021 period. Low SHAP values indicate that the input variable decreases the average modelled SIF, suggesting the climate variable constrains photosynthesis.

The maps depict the inverse of the SHAP values for illustration purposes (higher values indicate a greater SIF constraint).

5. DISCUSSION

The results demonstrated the capabilities of SHAP in a case study that made use of geospatial climate data as input variables of a machine learning model. A weather-SIF model was trained on BELMANIP2 sites using ERA5-Land and TROPOSIF measurements. SHAP values showed the spatial and temporal impacts of the three climate variables –air temperature, shortwave radiation, and VPD– on SIF, indicating climate constraints on the photosynthesis dynamics.

The model showed good performance for replicating the vegetation seasonality even though it only considered weather variables as input. The model had a minimal bias, but an underestimation of high SIF values (SIF >0.5 mW m^{-2} sr^{-1} nm^{-1}) was apparent in DBF (Fig. S2). Despite this, the underestimation of high SIF values did not affect the phenology estimation, which was the purpose of this work. The underestimated SIF would be potentially corrected if a proxy for the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) was included in the model. Proxies for fAPAR are the NDVI and EVI. Should a spectral index be used, the model would have similar input variables as the GOSIF product (Li and Xiao, 2019), which includes a spectral index (EVI) and weather variables (air temperature, VPD, and photosynthetically active radiation (PAR)). If we included a spectral index, however, the model would be less explainable because part of the predicted SIF would be attributed to changes in NDVI, which would conceal the marginal contributions of weather variables on the predicted SIF. Besides that, our model explanation aimed to understand the impact on SIF purely attributed to climate variables and, thus, providing accurate SIF predictions were less important than reproducing the SIF seasonality. Another potential limitation of our experiment is that we only used concurrent SIF and weather variables, neglecting the forcing requirements involved during the pre-season. However, the forcing requirements are minimal in non-deciduous vegetation types (Descals et al., 2022, 2020a), and have negligible effect on other ecosystems of the temperate and cold biomes of the Northern Hemisphere. Non-deciduous vegetation in the temperate and cold biomes of the Northern Hemisphere includes temperate and boreal evergreen forests, and tundra shrublands and grasslands.

We used air temperature, shortwave radiation, and vapor-pressure deficit because these variables are key environmental factors that affect vegetation photosynthesis. However, vegetation photosynthesis is a complex process influenced by additional factors, such as soil water availability, carbon dioxide concentration, or nutrient availability. Thus, although air temperature, shortwave radiation, and vapor-pressure deficit are surrogates for predicting SIF, they do not capture the full complexity of photosynthesis. Moreover, the primary objective of the study was to demonstrate the potential of local interpretation models in remote sensing, utilizing a previously reported model that predicts SIF based on weather data at the global scale (Li and Xiao, 2019).

SHAP values confirmed previous finding on the spatial and temporal climate constraints on the vegetation activity. SHAP maps showed that VPD, which reflects atmospheric dryness, was the main factor limiting SIF in tropical dryland ecosystems at the start and end of the growing season.

Tropical dry land ecosystems include tropical desert, tropical monsoon, and tropical savannah climates. The growing season in these areas is sensitive to precipitation (Zhang et al., 2022), and high evaporative demand induces vegetation –mostly grasslands and sparse woody vegetation– into dormancy in the form of deep roots (Zhou et al., 2020). In tropical rainforests, however, both temperature and water are adequate for plants, and radiation was the only factor found to constrain SIF. This differs from the maps produced by Jolly et al. (2005), which depicted tropical rainforests (Amazon and Central Africa) without any climate limitations, potentially because a uniform threshold was used all over the Earth. However, previous studies do suggest that radiation is a limiting factor in this biome (Aguilos et al., 2018; Weber et al., 2009), which supports our finding. In extratropical areas, temperature was the main constraint at the start and end of the growing season, which is supported by extensive literature (Fu et al., 2020; Peñuelas et al., 2009; Piao et al., 2019). Moreover, although cold temperatures drive vegetation into dormancy, we also found a divergent constraint in terms of radiation. Overall, radiation was not limiting SIF during the SoS, but it did during the EoS. This is consistent with recent findings with point out at an increasing constraint of radiation on photosynthesis during the EoS (Descals et al., 2022; Zhang et al., 2020), which show that due to radiation constraints, rising temperatures will not increase autumn greening. Thus, the results obtained with SHAP are overall in line with the current understanding of global photosynthesis dynamics. However, it is worth mentioning that the resolution of the weather dataset (ERA5-Land) limits the detail and interpretability of the results in localized regions with a high diversity of climates. To depict regions with such diversity of climates a finer resolution dataset and a more nuanced selection of input variables would be required.

SHAP proved to be a useful technique for explaining the correlations between SIF and climate factors that were captured by the machine learning model. The explainability of the ML models with SHAP, on the other hand, must be considered with caution. SHAP values show researchers which correlations machine learning has found, but these correlations do not necessarily imply a causality between input and output variables (Heskes et al., 2020). Expert knowledge is required to determine whether the correlations are coherent with the reality of the problem, and further research is required to determine whether causality exists. In our case study, we validated our findings with literature that supported the results revealed by SHAP maps. The capability of SHAP to explain spatially and temporally the predictions from geospatial gridded time series might assist future remote sensing applications.

6. Author contributions

AD and JP conceived the research idea. AD and JP designed the study. AD performed the analyses and wrote the first version of the manuscript. AD, AV, GY, IF, and JP contributed to the interpretation of the results and to revisions of the manuscript.

7. Funding sources

This work represents a contribution to CSIC Thematic Interdisciplinary Platform TELEDETECT. This research was supported by the Spanish Government grant PID2019-110521GB-I00, the Fundación Ramón Areces grant ELEMENTAL-CLIMATE, and the Catalan Government grant SGR2017-1005.

8. Data availability statement

The data that support the findings of the study are openly available from TROPOSIF L2B dataset [https://doi.org/10.5270/esa-s5p_innovation-sif-20180501_20210320-v2.1-202104], ERA5-Land Hourly - ECMWF Climate Reanalysis [https://doi.org/10.24381/cds.e2161bac], and MODIS Land Cover Type MCD12Q1 at [https://doi.org/10.24381/cds.e2161bac].

9. Code availability statement

Code for training the SIF-weather model and explaining the model with SHAP is available at https://github.com/adriadescals/SHAP_PHENO_SIF

10. REFERENCES

Aguilos, M., Hérault, B., Burban, B., Wagner, F., Bonal, D., 2018. What drives long-term variations in carbon flux and balance in a tropical rainforest in French Guiana? Agricultural and Forest Meteorology 253, 114–123.

Ahmad, R., Yang, B., Ettlin, G., Berger, A., Rodríguez-Bocca, P., 2020. A machine-learning based ConvLSTM architecture for NDVI forecasting. International Transactions in Operational Research.

Bentéjac, C., Csörg\Ho, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review 54, 1937–1967.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Descals, A., Verger, A., Filella, I., Baldocchi, D., Janssens, I.A., Fu, Y.H., Piao, S., Peaucelle, M., Ciais, P., Peñuelas, J., 2020a. Soil thawing regulates the spring growth onset in tundra and alpine biomes. Science of the Total Environment 742, 140637.

Descals, A., Verger, A., Yin, G., Filella, I., Fu, Y.H., Piao, S., Janssens, I.A., Peñuelas, J., 2022. Radiation-constrained boundaries cause nonuniform responses of the carbon uptake phenology to climatic warming in the Northern Hemisphere. Global Change Biology.

Descals, A., Verger, A., Yin, G., Peñuelas, J., 2020b. A Threshold Method for Robust and Fast Estimation of Land-Surface Phenology Using Google Earth Engine. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 601–606.

Franch, B., Vermote, E.F., Roger, J.-C., Murphy, E., Becker-Reshef, I., Justice, C., Claverie, M., Nagol, J., Csiszar, I., Meyer, D., others, 2017. A 30+ year AVHRR land surface reflectance climate data record and its application to wheat yield monitoring. Remote Sensing 9, 296.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics 1189–1232.

Fu, Y., Li, X., Zhou, X., Geng, X., Guo, Y., Zhang, Y., 2020. Progress in plant phenology modeling under global climate change. Science China Earth Sciences 63, 1237–1247.

Gensheimer, J., Turner, A.J., Köhler, P., Frankenberg, C., Chen, J., 2022. A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet). Biogeosciences 19, 1777–1793.

Guanter, L., Aben, I., Tol, P., Krijger, J., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., Landgraf, J., 2015. Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. Atmospheric Measurement Techniques 8, 1337–1352.

Guanter, L., Bacour, C., Schneider, A., Aben, I., van Kempen, T.A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., others, 2021. The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission. Earth System Science Data 13, 5423–5440.

Heskes, T., Sijben, E., Bucur, I.G., Claassen, T., 2020. Causal shapley values: Exploiting causal knowledge to explain individual predictions of complex models. Advances in neural information processing systems 33, 4778–4789.

Johnsen, P.V., Riemer-Sørensen, S., DeWan, A.T., Cahill, M.E., Langaas, M., 2021. A new method for exploring gene–gene and gene–environment interactions in GWAS with tree ensemble methods and SHAP values. BMC bioinformatics 22, 1–29.

Joiner, J., Yoshida, Y., Köehler, P., Campbell, P., Frankenberg, C., van der Tol, C., Yang, P., Parazoo, N., Guanter, L., Sun, Y., 2020. Systematic orbital geometry-dependent variations in satellite solar-induced fluorescence (SIF) retrievals. Remote sensing 12, 2346.

Jolly, W.M., Nemani, R., Running, S.W., 2005. A generalized, bioclimatic index to predict foliar phenology in response to climate. Global Change Biology 11, 619–632.

Kandasamy, S., Fernandes, R., 2015. An approach for evaluating the impact of gaps and measurement errors on satellite land surface phenology algorithms: Application to 20 year NOAA AVHRR data over Canada. Remote Sensing of Environment 164, 114–129.

Li, W., Migliavacca, M., Forkel, M., Denissen, J., Reichstein, M., Yang, H., Duveiller, G., Weber, U., Orth, R., 2022. Widespread increasing vegetation sensitivity to soil moisture. Nature Communications 13, 1–9.

Li, X., Xiao, J., 2019. A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS, and reanalysis data. Remote Sensing 11, 517.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence 2, 56–67.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30.

Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., others, 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature biomedical engineering 2, 749–760.

Meng, Y., Yang, N., Qian, Z., Zhang, G., 2020. What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. Journal of Theoretical and Applied Electronic Commerce Research 16, 466–490.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., others, 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data Discussions 1–50.

Peñuelas, J., Rutishauser, T., Filella, I., 2009. Phenology feedbacks on climate change. Science 324, 887–888.

Piao, S., Liu, Q., Chen, A., Janssens, I.A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., Zhu, X., 2019. Plant phenology and global climate change: Current progresses and challenges. Global change biology 25, 1922–1940.

Shapley, L.S., 1953. Stochastic games. Proceedings of the national academy of sciences 39, 1095–1100.

Verger, A., Baret, F., Weiss, M., 2014. Near real-time vegetation monitoring at global scale. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 3473–3481.

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M., 2022. Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. Journal of Environmental Management 301, 113941.

Weber, U., Jung, M., Reichstein, M., Beer, C., Braakhekke, M., Lehsten, V., Ghent, D., Kaduk, J., Viovy, N., Ciais, P., others, 2009. The interannual variability of Africa's ecosystem productivity: a multi-model analysis. Biogeosciences 6, 285–295.

Weiss, M., Baret, F., Verger, A., 2014. BELMANIP2: Enhancement of the CEOS-BELMANIP ensemble of sites used for the validation of land products from medium resolution sensors., in: Fourth International Symposium on Recent Advances in Quantitative Remote Sensing.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316.

Zhan, C., Orth, R., Migliavacca, M., Zaehle, S., Reichstein, M., Engel, J., Rammig, A., Winkler, A.J., 2022. Emergence of the physiological effects of elevated CO2 on land–atmosphere exchange of carbon and water. Global Change Biology 28, 7313–7326.

Zhang, Y., Commane, R., Zhou, S., Williams, A.P., Gentine, P., 2020. Light limitation regulates the response of autumn terrestrial carbon uptake to warming. Nature Climate Change 10, 739–743.

Zhang, Y., Gentine, P., Luo, X., Lian, X., Liu, Y., Zhou, S., Michalak, A.M., Sun, W., Fisher, J.B., Piao, S., others, 2022. Increasing sensitivity of dryland vegetation greenness to precipitation due to rising atmospheric CO2. Nature communications 13, 4875.

Zhou, Y., Wigley, B.J., Case, M.F., Coetsee, C., Staver, A.C., 2020. Rooting depth as a key woody functional trait in savannas. New Phytologist 227, 1350–1361.