Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL104962

Key Points:
- Plant Functional Types (PFTs), as often used in land flux studies, are not easily empirically associated with site climate and/or flux regimes
- A broad selection of alternative vegetation/land cover classifications do not offer greater predictability
- The disconnect between PFTs and climate/flux regimes has implications for modeling and analysis of terrestrial systems

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Cranko Page,
joncrankopage@gmail.com

Citation:

Cranko Page, J., Abramowitz, G., De Kauwe, M. G., & Pitman, A. J. (2024). Are plant functional types fit for purpose? *Geophysical Research Letters*, *51*, e2023GL104962. https://doi.org/10.1029/2023GL104962

Received 13 JUN 2023
Accepted 20 NOV 2023

Are Plant Functional Types Fit for Purpose?

Jon Cranko Page[1,2], Gab Abramowitz[1,2], Martin. G. De Kauwe[3], and Andy J. Pitman[1,2]

[1]Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia, [2]ARC Centre of Excellence in Climate Extremes, Sydney, NSW, Australia, [3]School of Biological Sciences, University of Bristol, Bristol, UK

Abstract For over 40 years, Plant Functional Types (PFTs) have been used to discretize the ~400,000 species of terrestrial plants into "similar" classes. Within Earth System Models (ESMs), PFTs simplify terrestrial biosphere modeling in combination with soil information and other site characteristics. However, in flux analysis studies, PFT schemes are often implemented as the sole analytical lens to clarify complex behavior. This usage assumes that PFTs adequately enable a mapping between climate inputs and flux outputs. Here, we show that random forest models, trained using aggregated climate and flux measurements from 245 eddy-covariance sites, cannot accurately predict PFT groupings, regardless of the nature of the PFT scheme. Similarly, PFTs provide negligible benefit when using site climate to predict site flux regimes and vice versa. While use of PFT classifications is convenient, our results suggest they do not aid analytical skill, which has important implications for future terrestrial flux studies.

Plain Language Summary To understand how the land surface behaves, we often divide plants into a small number (20 or less) of "similar" groups, such as evergreen forests, or grasslands, known as Plant Functional Types (PFTs). The idea is that landscapes with similar large-scale characteristics will behave in the same way. In land surface models, these PFT groups determine how the simulated plants react to the climate in combination with soil information and other characteristics, yet analysis of observations often use PFT groups alone to try to explain variations in results between different experimental sites. We use machine learning to show that while PFTs might be visually compelling, they do not necessarily represent behavior groupings and might actually hide real world behavior if used for analysis. As such, we suggest that future studies instead try to look at more specific site characteristics when trying to explain analysis results.

1. Introduction

The Earth's land surface is highly complex and heterogeneous, home to nearly 400,000 plant species (WFO, 2023). Terrestrial ecosystems are responsible for the uptake of around one-third of anthropogenic greenhouse gas emissions (Friedlingstein et al., 2022), providing a vital mitigating effect on the pace of climate change. The global carbon sink is highly variable on the annual timescale (Le Quere et al., 2009); a variability almost entirely driven by fluctuations in terrestrial carbon uptake (Bousquet et al., 2000; Friedlingstein et al., 2022; Piao et al., 2020). Uncertainty in modeling the terrestrial carbon sink contributes significant uncertainty to future climate projections (Arora et al., 2020), and therefore substantial effort is being invested in improving earth system models (ESMs). How the extraordinarily diverse land surface is represented within the confines of ESMs is a topic of active debate, especially centered around the use of plant functional groups or types (PFTs).

PFTs were introduced to ESMs in the 1980s and are utilized to partition terrestrial vegetation into discrete classes that are assumed to behave similarly, hence reducing computational demands and model complexity (Bonan, 1996; Box, 1996; DeFries et al., 1995; Smith et al., 1997; Wilson & Henderson-Sellers, 1985). Over time, ESMs developed to represent the diverse landscapes contained within individual model grid cells as patches of different PFTs, which can sometimes compete with each other ecologically (Bonan et al., 2002; Fisher & Koven, 2020; Sitch et al., 2003). Not long after their introduction, PFTs were under scrutiny for their coarse discretization of plant traits into a finite, and often small, set of classes (Lavorel & Garnier, 2002; I. J. Wright et al., 2004; J. P. Wright et al., 2006). A substantial number of studies continue to argue for increased granularity in the representation of vegetation in ESMs (Alton, 2011; Pappas et al., 2016). This often takes the form of trait-based approaches (Scheiter et al., 2013; Van Bodegom et al., 2012; Yang et al., 2015), amid other strategies such as those based on optimality (Harrison et al., 2021) and evolutionary (Anderegg et al., 2022) principles.

The arguments for increased dimensionality in land surface representation tend to arise from an ecological and visual perspective. Importantly, a number of studies have demonstrated that attempts to simplify plant trait diversity into discrete classes do not readily group by PFTs (Konings & Gentine, 2017; Rogers, 2014; but see Lin et al., 2015). Moreover, it is possible that such increases in model complexity (driven, e.g., by a greater number of PFTs), while no doubt important for more accurately describing vegetation function, will not result in ESMs of higher fidelity when considering the representation of ecosystem fluxes (Lovenduski & Bonan, 2017). More PFTs, if requiring increased parameterization to define them, may drive an increase in parameter uncertainty, a major source of ESM uncertainty (Dietze et al., 2014; Raczka et al., 2018). It has already been suggested that ESMs are too complex (Haughton et al., 2016), and that empirical models with only meteorological forcings (i.e., no vegetation information) can outperform ESMs when modeling fluxes (Haughton et al., 2018; Prentice et al., 2015; Tan et al., 2023). In fact, modeling of the terrestrial carbon sink was not significantly improved between CMIP5 and CMIP6 for many aspects (Arora et al., 2020). Conversely, recent studies have not necessarily produced results consistent with PFTs being fully obsolete (Thomas et al., 2019). Laine et al. (2022) found PFTs to exhibit strong control on spatial variation of peatland fluxes, while a PFT approach for modeling global fluxes produced similar results to a more complicated ”environmental filtering” framework in a study by Famiglietti et al. (2023).

While most ESMs have taken steps beyond simple PFT representations, analysis of terrestrial fluxes still utilize PFTs as explanatory variables, or to assess and evaluate study findings (see, e.g., Cranko Page et al., 2022; Ma et al., 2020; Teckentrup et al., 2021). For instance, the land surface component of ESMs are often tested against observations from eddy-covariance flux towers (Best et al., 2015), with PFTs taken from the flux data (e.g., AmeriFlux, 2023; FLUXNET, 2023) used to group sites and assess performance. This frequently stems from the desire for results to be seen through a low-dimensional lens such that emergent, simple messages can be extracted from the data. PFTs provide a 1D or 2D lens with the assumption that their discrete categories provide explanatory power for often continuous analysis results. The International Geosphere-Biosphere Program (IGBP) classification (Strahler et al., 1999) is one of the most common PFT schemes used in such a manner. This is most likely due to a number of factors; the IGBP scheme is well-established, easy to access for flux sites (see FLUXNET, 2023), and conceptually simple to visualize and understand.

Despite the acknowledged limitations of PFTs, whether or not they are fit for purpose for model benchmarking and evaluation of studies is still relatively unexplored. If PFTs are not fit for purpose, then interpreting results in the context of PFTs is effectively applying a convoluting transformation to the results, potentially obfuscating more robust conclusions from the analysis. The key role of PFTs in ESMs is to parametrize vegetation variation, and so provide a mechanism by which climate forcings together with model states, such as soil type and moisture, are converted to carbon, water, and energy fluxes. There is, therefore, an unspoken assumption about PFTs that remains in the community: that PFTs adequately provide the bridge between climate inputs and flux outputs for terrestrial ecosystems.

To explore these issues we use data from 245 flux sites and examine the following hypotheses:

1. The IGBP classification is not suitable for mapping the climate experienced at a site to the fluxes measured.
2. The IGBP classification is outperformed by more complex PFT classification schemes.
3. Site climate and flux regimes are closely related, and contain enough mutual information such that they can reliably predict one another.

2. Methods

2.1. Data

2.1.1. Eddy-Covariance Data

Eddy-covariance flux data were taken from two different sources to provide a total of 245 sites following quality checks (Figures S1 and S2 in Supporting Information S1). The PLUMBER2 data set consists of half-hourly time series for 170 sites which are quality-checked and gap-filled prior to publication (A. M. Ukkola et al., 2022). The Integrated Carbon Observation System (ICOS) data set consists of hourly time series for 110 sites (ICOS RI, 2023). Where a site was present in both the PLUMBER2 and ICOS data sets, the PLUMBER2 data set was taken to ensure consistency over the greater number of sites. The sites were then quality-checked for any missing data, inconsistencies, or other anomalous data. Since this study used aggregated metrics of site observations, periods of bad data (as entire years) could be omitted from site timeseries while retaining the good years of data in the study.

2.1.2. Site Metrics

For each site, six observed variables were extracted. Three were the fluxes measured at each site: Net Ecosystem Exchange (NEE), latent heat flux (LE) and sensible heat flux (H). Three climate variables were also used: precipitation (PPT), air temperature (TAIR) and incoming short-wave radiation (SW). For each variable, four metrics were calculated to summarize the variable's behavior through time. The metrics are Daily Mean (DM), Daily Variability (DV), Seasonal Variability (SV), and Interannual Variability (IAV). DM is the mean value of the variable. The mean value for each day is calculated from the half-hourly or hourly data, and then a further mean is taken across the entire timeseries. DV is calculated as the mean of DV. For each day, the standard deviation in half-hourly or hourly value is calculated, and the mean of these is then taken across the time series. SV is calculated by first finding the mean value for each month in the time series. The standard deviation between the monthly mean values is calculated for each year, before the mean is taken over all years in the timeseries. IAV is calculated by finding the mean value for each year in the time series, and then taking the standard deviation of this across all years. Where IAV was not defined, namely sites with only a single year of data, it was gap-filled using 5-nearest neighbor imputation. The neighbors were defined by the Euclidean distance between the other input metrics and the mean IAV from the nearest five neighbors was taken as the IAV for the site with missing data. Note, this was only performed where the model or clustering algorithm could not handle missing data. These four metrics provide a summary for each variable, capturing the mean magnitude as well as variability on daily, seasonal, and interannual timescales. These four metrics for NEE, LE, and H are combined to make the 12 indicators used to define the flux regime of a site. The same metrics for PPT, TAIR, and SW are used to define the climate regime.

2.1.3. Plant Functional Types and Biome Classifications

Ten PFT or ”biome classification” (BC) schemes were used for all of the sites. The first was the IGBP classification, which is the PFT reported by FLUXNET together with the site data, and available on the flux network websites. Table S1 in Supporting Information S1 shows the names and site breakdown for the IGBP classification (see Strahler et al., 1999). The other nine BC schemes are taken from the Biome Inventory data set (Fischer et al., 2022b), and utilize different methods of partitioning the land surface into finite, discrete classes (Allen et al., 2020; Beck et al., 2018; Buchhorn et al., 2020; FAO, 2012; Friedl et al., 2010; Hengl et al., 2018; Higgins et al., 2016; Leemans, 1990; Zhang et al., 2017). The nine schemes were taken to cover a range of classification methodologies, resolutions, and ages. Table S2 in Supporting Information S1 details these classification schemes and why they were chosen for this analysis. Where extraction of a site's classification from the BC raster resulted in NA data (due to e.g., the pixel falling within the ocean for coarser maps), the data extraction was stepwise expanded by 5,000 m until a non-zero number of pixels within the extraction area had a non-NA value. The mode of the classifications for these pixels was then taken as the value for the site. Throughout this study, BC schemes will be referred to as PFT schemes due to their overlapping roles in partitioning terrestrial ecosystems.

2.2. Modeling

To explore the role of PFTs in bridging the climate and flux regimes of sites, sites were clustered by three different methodologies and these clusters were then predicted using Random forests (RFs) (see Figure S3 in Supporting Information S1 for a schematic). The first clustering methodology was the PFT classifications detailed above. In addition to this, the sites were also partitioned into metaclusters (MCs) using one of two clustering techniques; self-organizing maps with hierarchical clustering, and k-means clustering using the kmeans++ initialization method.

2.2.1. Self-Organizing Maps

For both climate and flux regimes, a self-organizing map (SOM) was fitted to the metric data from the 245 sites. A SOM is an unsupervised machine-learning framework that reduces the dimension of a data set to (usually) two dimensions while approximating the topology of the original data set (Kohonen, 1982, 2013). In each clustering run, 50 SOMs with different initial seeds were generated for the data. Each SOM's ability to effectively cluster the metric data was assessed on two measures of performance, the Kaski-Lagus error (Kaski & Lagus, 1996) and the percentage of variation explained. Out of the 50 SOMs, the best-performing SOM was utilized in the further modeling and analysis. The best SOM was then further clustered into the MCs using hierarchical clustering, namely agglomerative nesting (Kaufman & Rousseeuw, 1990a, 1990b). This was performed to reduce the number of discrete classes so that they matched the number of PFTs in the particular PFT scheme being tested.

2.2.2. k-Means Clustering

Alternative MCs for the climate and flux regimes were calculated using a k-means clustering (KM) approach (Jain et al., 1999; MacQueen, 1967). The clustering was initialized using the k-means++ algorithm (Arthur & Vassilvitskii, 2007). In the same vein as the SOMs, the KM was run 50 times and the best KM were selected where they minimized the Sum of Squared Errors (SSE). Where multiple iterations produced the same SSE, a KM model which also maximized the Between Sum of Squares over Total Sum of Squares (BSS/TSS) was randomly selected.

2.2.3. Random Forest Models

Random forests (RFs) were used to understand how much information the regimes of climate, flux, or vegetation contained about each other. They were used to predict the MCs of the sites, whether produced by SOMs or KMs, or the PFT classifications. Each RF was tested out-of-sample, with 66% of the sites randomly assigned to the training set, while maintaining the overall ratio of classes from the complete list of sites. Performance was then assessed against the 33% of sites assigned to the testing set. For the RFs, 100 models were generated (each with its own training/testing split) and the results were averaged or summed across these RFs, as required, to account for the stochasticity of the models and the training/testing split.

RFs were assessed using the accuracy and macro F1 score metrics, which fall between 0 and 1 (Manning et al., 2008; Opitz & Burst, 2019). The accuracy metric is simply the percentage of correct predictions. Meanwhile, the macro F1 score is a metric that balances correct predictions against false positives. This is preferred over accuracy for data sets with large class imbalances where each class can be considered of equal importance despite the differing sample sizes, since otherwise assigning every site to the most common class would provide a relatively large accuracy. This would be an issue when clustering eddy-covariance towers with their spatially correlated deployments (see Table S1 in Supporting Information S1). Performance within the RFs indicates how informative the site clusterings are; if RFs have good performance (accuracy and F1 score close to 1), then the clustering involved partitions the sites into well-defined clusters of climate/flux regimes or PFT classification. Low accuracy and F1 score would imply that the RF drivers do not have the predictive capacity for the predicted classes.

2.3. Analysis

Initialization of the models (both the metaclustering and the RFs) are dependent on the random seed used to generate them. To ensure that the results were robust, the SOM + RF model for 11 MCs of climate and flux regimes, predicted using the other regime, was tested using 2,000 different random seeds (Figure S4 in Supporting Information S1). Seeds were assessed on the Kaski-Lagus Error and percent of variation explained for SOMs, and on the mean macro F1 score for the RFs. The seed that performed best on average across these metrics was chosen and used for all model runs. The RFs were also insensitive to the number of trees (Figure S5 in Supporting Information S1).

3. Results

In general, the site IGBP classes are unable to be accurately predicted by RFs using the combined climate and flux regimes (Figure 1a), or each of the regimes independently (Figures 1b and 1c). Site climate regimes are the worst at predicting IGBP classes (median accuracy 0.33, median F1 0.27), while flux regimes are better (median accuracy 0.47, median F1 0.40), and the combination of both regimes as predictors provides the best RF performance (median accuracy 0.53, median F1 0.44). While some classes are predicted with relatively good accuracy, such as croplands (CRO) with an accuracy of 0.8 for both combined regimes (Figure 1a) and flux regimes (Figure 1c), others are notably poor (e.g., closed shrublands, CSH, with 0.0 or grasslands, GRA, with sub-0.50 accuracy for all three predictor sets). Overall, forest classes are generally among the most predictable, with mixed forests (MF) an outlier, although performance still never exceeds an accuracy of 0.7 (EBF, evergreen broadleaf forests, class predicted by climate and flux regimes, Figure 1a). Consistent misclassification is a feature for all three input schemes, with MF being assigned as ENF (evergreen needleleaf forests) and woody savannahs (WSA) assigned

10.1029/2023GL104962

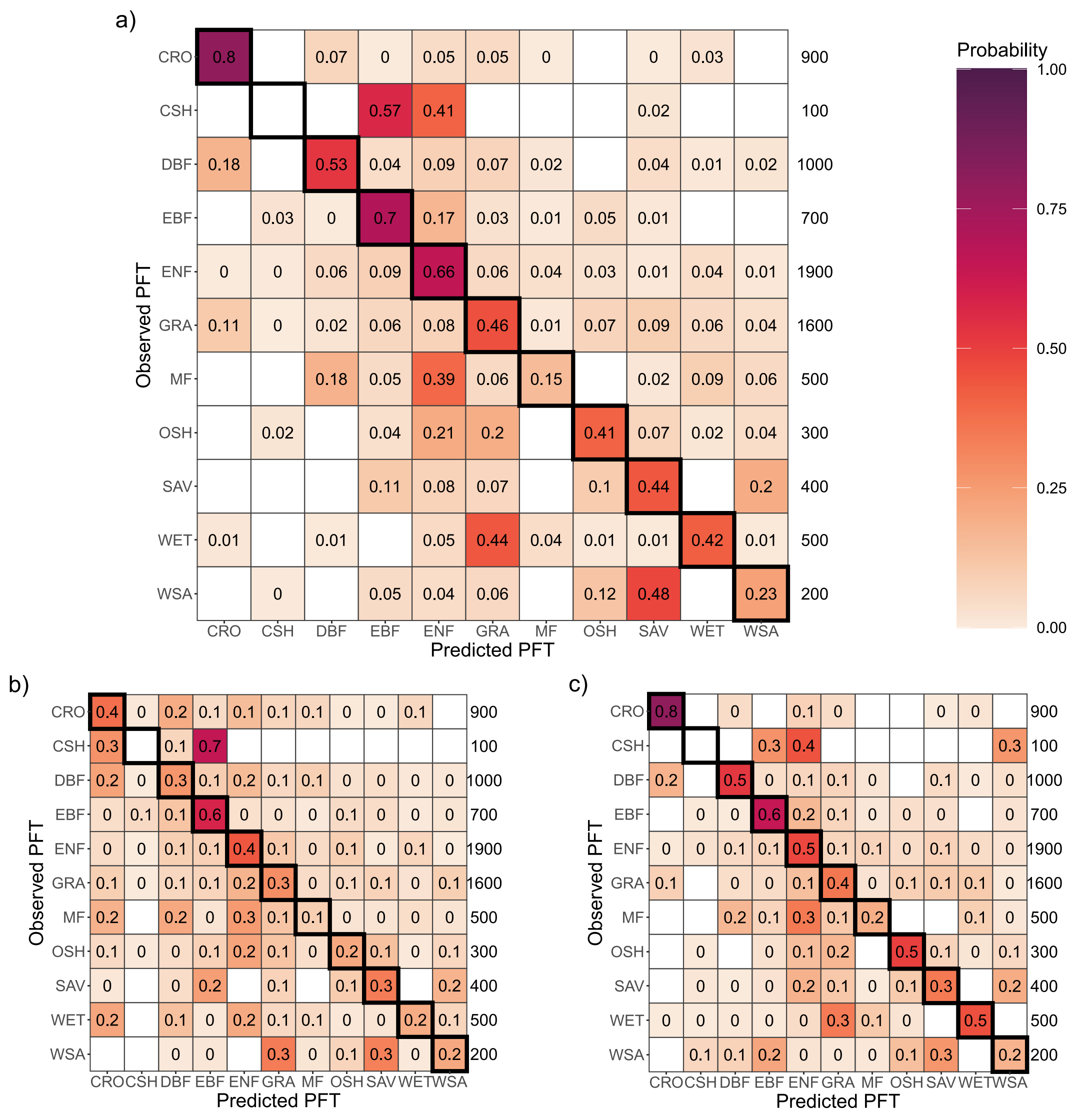

Figure 1. Probabilities of International Geosphere-Biosphere Program (IGBP) class prediction from Random forests (RFs). (a) shows the results when the predictors are the combined climate and flux regimes, (b) is the results for the climate regime as the predictor, and (c) for the flux regime. The *x*-axis is the predicted IGBP class from the RFs and the *y*-axis is the observed IGBP class. The numbers are the probability of each class being predicted for a given observed class (e.g., the probability of a CRO (cropland) site being predicted as DBF (deciduous broadleaf forests) from its combined climate and flux regimes is 0.07). Note that blank spaces indicate a zero probability of assignment, while a score of 0 is a probability greater than 0 rounded to two decimal figures. The numbers to the right of each plot indicate the number of times sites belonging to the observed IGBP class fall within the testing set - note these are multiples of 100 because 100 RFs are run and each testing set maintains the proportion of observed classes from the full data set (e.g., with 20% of the 245 sites observed as GRA, 20% of each testing set of 82 sites will also be GRA).

to savannahs (SAV) with probability exceeding 0.3. Similarly the RFs often disagree on site classification. For the combined regimes and flux regime alone, each class is assigned to any of eight classes on average, while this is nine classes for the climate regime alone. In other words, of all possible class-to-class assignments, only 27% never occur for the combined regime RFs, 26% for the flux regime RFs, and 15% for the climate regime alone.

While the IGBP class is poorly predicted by all three input schemes (i.e., climate regimes, flux regimes, and the combination of the two), the other nine PFT schemes tested are not substantially more predictable (all have median and mean F1 score of 0.61 or less, Figure 2). Interestingly, the majority of PFT schemes are better predicted by site climate regime than site flux regime - only the IGBP and Friedl et al. (2010) schemes are better predicted by the flux regime (median F1 score of 0.4 for flux vs. 0.28 for climate, and 0.45 vs. 0.38 respectively). This illustrates that

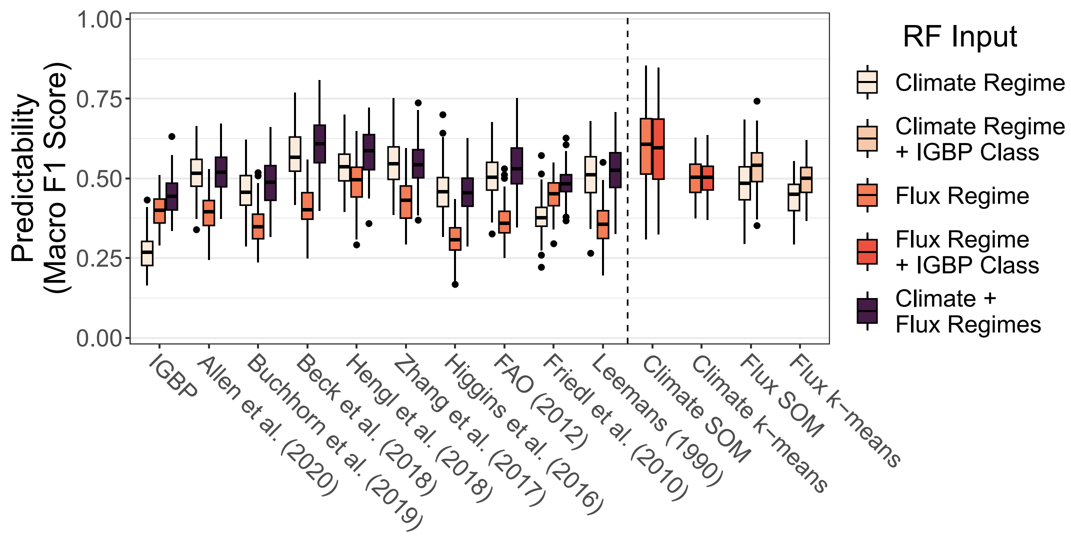

Figure 2. F1 scores for Random forests (RFs) predicting various classification schemes. The classification schemes are on the *x*-axis, and boxplots are grouped by classification scheme and RF input. Ten Plant Functional Type schemes and two metaclusterings (self-organizing map and k-means) of site climate and flux regimes are shown, separated by the vertical dashed line, with five different possible input schemes. The results for the metaclusterings are those with 11 MCs to match the number of International Geosphere-Biosphere Program classes. The boxplots are such that the main box covers the first to third quartiles with the middle line indicating the median. The whiskers extend to the furthest value no more than 1.5 times the interquartile range from the box, with dots indicating data points lying outside this range.

most PFT schemes are heavily influenced by the climate regime at a location. In fact, for the remaining eight PFT schemes, the addition of the flux regime as a predictor increases the median F1 score by 0.05 or less. Hence, the flux regime of a site contains negligible information about the site PFT that is not already encoded in the climate regime. Additionally, knowledge of the IGBP classification of a site provides only minor benefits when predicting regime MCs (Figure 2 and Figure S7 in Supporting Information S1). When classifying site climate regimes using site flux regimes as input to the RFs, the inclusion of site IGBP as a predictor has no or even negative impact on RF performance (median F1 score reductions of 0 and 0.01). Conversely, the site IGBP classification does provide some additional information when predicting flux MCs (median F1 score improvements of 0.05 and 0.06, an increase of 11%).

Comparing the PFT classification schemes to the regime MCs, seven of the PFT schemes are more accurately predicted than the flux MCs with the same number of classes by the RFs driven by site climate regimes, those with 12 or more clusters (top facet of Figure 3). The median F1 score is between 0.46 and 0.57 for these PFT schemes, while the MC median F1 scores fall between 0.23 and 0.44, with a mean difference of 0.2 for the SOM-derived MCs and 0.21 for the KM-derived MCs. In comparison, the flux regimes are able to predict the PFT schemes and climate MCs with similar performance (difference in median F1 scores for the same seven PFT schemes range from −0.10 to 0.04 for SOM-derived MCs and −0.12 to 0.05 for the KM-derived MCs). When the metaclustering model performance is considered, there is no clear preference for either the SOM + hierarchical clustering or the k-means clustering methods. Additionally, the performance of the MC prediction is similar for both directions of prediction (i.e., climate regime MCs predicted using flux regime RFs, and flux regime MCs predicted using climate regime RFs). The more recent PFT schemes are not more predictable than older schemes. Noticeably, increasing PFT scheme or metaclustering fidelity (i.e., increasing the number of clusters to which sites may be assigned) does not improve performance. Indeed, as the number of MCs increases, F1 scores decrease before plateauing at around 0.25. Meanwhile, performance for the PFT schemes is very consistent irrespective of the number of clusters the 245 sites are partitioned between.

4. Discussion

4.1. The IGBP Classification Does Not Map Between Climate and Flux Regimes

The IGBP classification is arguably the most commonly used PFT scheme (AmeriFlux, 2023; FLUXNET, 2023; OzFlux, 2023). However, it is now over 25 years old (Belward, 1996; Strahler et al., 1999), and many attempts have

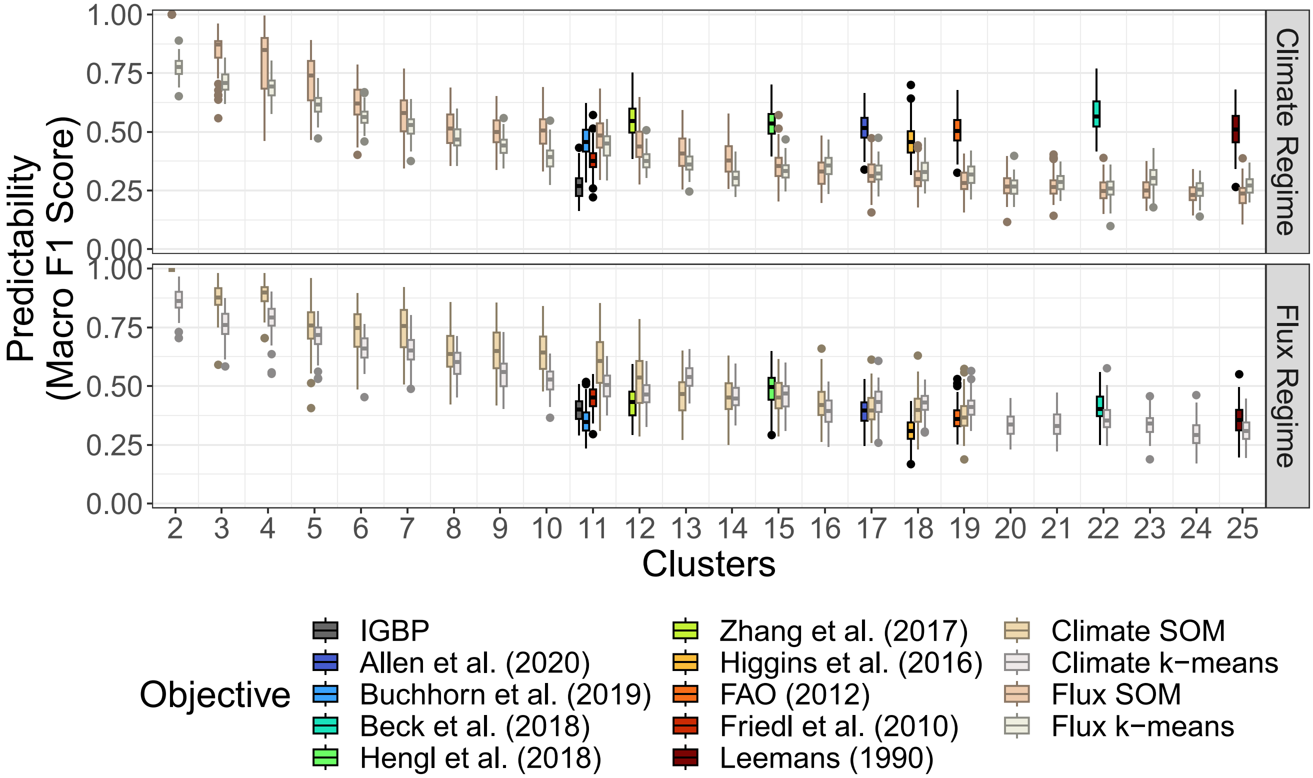

Figure 3. F1 scores for Random forests (RFs) predicting various classification schemes. The *x*-axis indicates the number of clusters that cover the 245 sites. Note that, for some Plant Functional Type (PFT) schemes, not every possible class is represented in the 245 sites. The top plot is where the climate regimes of the sites are the predictors, and the bottom plot is the flux regimes. Ten PFT schemes are colored and outlined in black. RF performance when predicting metaclusters of the alternate regime determined by either self-organizing map (SOM) + hierarchical clustering or k-means clustering for 2 to 25 clusters are shown in brown and gray respectively. Note that no SOM + hierarchical clustering was available for the climate regime for 20 or more clusters as at least one cluster was empty (contained 0 sites). The boxplots are such that the main box covers the first to third quartiles with the middle line indicating the median. The whiskers extend to the furthest value no more than 1.5 times the interquartile range from the box, with dots indicating data points lying outside this range.

been made to produce improved classifications (e.g., Beierkuhnlein & Fischer, 2021; Ellis & Ramankutty, 2008; Harper et al., 2022; Pillar, 1999; Poulter et al., 2015; van Bodegom et al., 2014). Our results indicate that reliance on the IGBP classifications is likely confounding analyses and associations between climate and flux at eddy-covariance sites. The IGBP class is the worst predicted of all PFT schemes in Figure 3, and also suffers in comparison to the regime MCs. While forests, a key driver of the land carbon sink, are among the most predictable sites, accuracy is never above 0.7, and forest classes are surprisingly outperformed by croplands for two of the three input schemes. When fitting RFs using all 24 climate and flux variables, the IGBP classification is outperformed by the other nine PFT schemes (Figure 2). Further, it provides little additional predictability when mapping climate to flux regimes and vice versa (Figure 2). As such, our first hypothesis can be confirmed and improving access to site classifications derived from alternative schemes is necessary.

4.2. Square Pegs or Round Holes?

A few potential reasons exist for why the IGBP classification might perform poorly in this study. First, there is a risk that the metrics used do not accurately capture the site behaviors. However, introducing further metrics (i.e., metrics of extremeness, autocorrelation, and skewness on a daily timescale) did not improve performance (Figure S6 in Supporting Information S1), indicating that this is unlikely to be the case.

Second, it is possible that a number of sites are incorrectly classified in the ”observations” - that is, the reported IGBP class at some sites is wrong, or that site disturbance has a strong influence on predictability. If incorrect classification were the case, then it would be expected that the sites in question would be consistently assigned to the ”correct” class by the RFs, as long as enough correctly observed sites are included in the training set. Here, only three sites appear to have potentially incorrect reported classes (Figure S8 in Supporting Information S1).

Two of these, PA-SPn and PA-SPs, are sites in Panama with a history of disturbance (Wolf, Eugster, Potvin, & Buchmann, 2011; Wolf, Eugster, Potvin, Turner, & Buchmann, 2011). This anthropogenic influence results in a vegetation composition not in sync with the sites' tropical climate, likely meaning the RFs lack exposure to this combination of climate and flux regimes, and hence struggle to correctly classify the DBF and GRA sites (note the overwhelming preference for EBF, a typical tropical PFT). The other site, BE-Lcr, exhibits a strong preference for classification as cropland as opposed to the reported DBF. Interestingly, satellite images (not shown; Google Maps, 2023) suggest that this site lies within an agricultural region and is surrounded by farmland, perhaps indicating that crops fall within the tower's footprint.

Another caveat is that the IGBP classification potentially has some indistinct classes. There is evidence to support this: all three of the CSH sites are incorrectly classified in 100% of testing set occurrences (Figure S9 in Supporting Information S1). As "closed shrublands," this class is plausibly too similar to other classes as a woody-plant-dominated ecosystem with a closed canopy. Notably, the key differentiation between CSH and the forest PFTs is plant height (Strahler et al., 1999). The classes differ in being either under (CSH) or above (forest PFTs) 2 m canopy height, a relatively arbitrary distinction. However, the RF models have no overriding preference when classifying the three sites (BE-Maa, IT-Noe, and US-KS2), either as a group or individually. This signifies that CSH cannot simply be treated as a "subset" of a different IGBP type, and also that the three sites cannot be individually re-assigned. Instead, this could be an issue related to the sample size, since only three CSH sites are included in this analysis, and so the RFs are only exposed to two CSH sites in each training set. Note that the second smallest class is WSA, with more than twice the sites, at seven sites, and only three of these sites have incorrect prediction rates of 90% or more.

4.3. The IGBP Scheme Is Not Substantially Outperformed by Other PFT Schemes

Many alternative schemes to the IGBP classification have been proposed (e.g., Harrison et al., 2010; Poulter et al., 2011, 2015; Wang & Price, 2007), including those used in this study. Our results indicate that the poor classification of climate and flux regimes by the IGBP scheme is not an issue unique to that particular discretization, although the IGBP scheme does exhibit the worst performance for the climate and combined regimes. Instead, a wide variety of PFT schemes, derived from differing approaches (Fischer et al., 2022b), are all predicted with a similar level of performance. As such, we suggest that our second hypothesis can be rejected. A more fruitful avenue would be for newly proposed schemes that take full advantage of the recent efforts to collate plant traits (Falster et al., 2021; Jucker et al., 2022; Kattge et al., 2011). Additional exploration of characteristics that drive site variability, not just traits but also other dimensions of diversity, will also provide much-needed evidence for new classifying schemes (Funk et al., 2017; Joswig et al., 2022; Migliavacca et al., 2021; Shiklomanov et al., 2020; Westerband et al., 2021). Our methodology could easily be applied to ensure that proposed schemes are capturing site variability as intended.

While many studies argue for increasing complexity in terrestrial classification (Butler et al., 2022; Wullschleger et al., 2014; Xu & Trugman, 2021), Figure 3 shows that improved model fidelity does not follow simply from higher granularity of PFTs. In fact, the PFT classifications are similarly predicted by climate and flux regimes irrespective of the number of categories being predicted. Such a result supports arguments against increasing the number of PFTs, which in turn would increase the parameter uncertainty in ESMs (Koven et al., 2020). As such, particular care must be spent ensuring that additional degrees of freedom when classifying the land surface must act along useful axes of heterogeneity (Díaz et al., 2016; Famiglietti et al., 2021; Migliavacca et al., 2021), axes that are directly constructed to aid predictability. These might well need to be established empirically.

4.4. Climate and Flux Regimes Contain Some Shared Information

There is shared information between climate and flux regimes, as shown through the F1 scores for SOM and KM MC classification in Figures 2 and 3. However, the scores vary widely between RF models and the median F1 scores for 11 MCs are between 0.45 and 0.61. This indicates a requirement for further ecosystem characteristics to model the connection between the land and the atmosphere, determining how input forcings to the land surface (i.e., climate) are transformed into outputs (i.e., carbon, water, and energy fluxes). Such information is likely to be related to vegetation, but may also describe site soil characteristics (De Long et al., 2019; Zhou et al., 2021), disturbance history (Amiro et al., 2010; Pugh et al., 2019), or factors such as hydrological functioning (Euskirchen et al., 2020; Griebel et al., 2020; Pérez-Ruiz et al., 2022) and topography (Hoover et al., 2021; Xie et al., 2021). Modern ESMs nearly always account for such characteristics, through for example, soil type maps and hydrological modules, but analysis of results may not consider these sources of information as explanatory factors. This does not support our hypothesis that the two regimes are closely related, and appears at odds with studies where simple empirical models of fluxes outperform ESMs when driven only by climate (Best et al., 2015).

4.5. Next Steps

To progress beyond PFTs and increase the fidelity of functional site classifications, the community is dependent on access to detailed information regarding each site's characteristics, including not just vegetation and soil composition, but also site hydrology and disturbance. While valuable efforts have been made in this area (Falster et al., 2021; Jucker et al., 2022; Kattge et al., 2011, 2020), owners of flux sites could dramatically improve the effectiveness of their data, and provide many benefits to the modeling community, by supplying additional site characteristics in easily accessible formats. Provision of the dominant, and/or the composition of, site species would likely have the greatest impact in linking flux and trait data. A concerted effort by flux networks to gather and distribute this data should be a priority, if the drive for increased granularity in terrestrial representation is to be realized.

In addition to motivating the transition away from existing PFT classifications, these results can also be used to identify future flux tower locations. With the geographic distribution of eddy-covariance towers being highly heterogeneous and weighted toward certain ecosystems, ensuring that the terrestrial sphere is suitably sampled is a key future requirement, especially as many under-sampled locations (e.g., tropics, drylands) are likely to be impacted by climate change (both in terms of functioning and distribution). Using membership of MCs as indication of the current sampling rates of the land surface by flux towers, it would likely be possible to use remote sensing products of climate and fluxes, or implement short-term flux measurements at proposed sites, to identify ecosystems that belong to those MCs of climate and flux regimes that lack representation. Placing new towers in these locations would have the additional benefit of ground-truthing *a-priori* assumptions about their regimes.

5. Conclusion

The desire to move away from PFTs within the land surface community is well articulated (Anderegg et al., 2022; Reichstein et al., 2014; Van Bodegom et al., 2012; Yang et al., 2015), yet studies assessing the suitability of PFTs *as currently used* are lacking (Alton, 2011; Groenendijk et al., 2011; Stoy et al., 2009). Here we have provided empirical evidence that PFTs are not providing sufficient information when attempting to bridge climate inputs and terrestrial fluxes. Our results highlight an urgent and fundamental issue for analysis of ESM outputs, especially as we move toward CMIP7. The continued move toward better representation of vegetation heterogeneity within the models must be replicated within external analysis, and progress beyond the limitations of PFTs before these confounding effects are further integrated into existing model intercomparison studies. Studies of atmosphere-biosphere interactions from a flux perspective must be careful when using PFTs as a lense through which to simplify results. More specific differentiation of vegetation and site characteristics during analysis would avoid the confounding effects introduced via PFTs.

Data Availability Statement

The PLUMBER2 flux data set used in this study is available from A. A. Ukkola (2020). Please view Table S3 in Supporting Information S1 for access to all ICOS data. The Biome Inventory data set is available from Fischer et al. (2022a). The code to reproduce the results from the manuscript is available from Cranko Page (2023). All analyses were performed using version 4.2.2 of the R software (R Core Team, 2020). RF modeling used the “ranger” package (v0.14.1; M. N. Wright & Ziegler, 2017), SOM creation used the “kohonen” (v3.0.11; Wehrens & Buydens, 2007; Wehrens & Kruisselbrink, 2018) and “aweSOM” packages (v1.3; Boelaert et al., 2022), and data was pre-processed with the “caret” package (v6.0-93; Kuhn & Max, 2008). Hierarchical clustering used the “agnes” function from the “cluster” package (v2.1.4; Maechler et al., 2022), while k-means clustering used the “KMeans_rcpp” function from the “ClusterR” package (v1.3.0; Mouselimis, 2022). Extraction of site PFT classifications from the Biome Inventory data set, which was first reprojected using “gdal” (v3.6.4; Rouault et al., 2023), was conducted using the “extract” function from the “raster” package (v3.6.14; Hijmans, 2023).

10.1029/2023GL104962

Acknowledgments
JCP, GA, MGDK, and AJP were supported by the Australian Research Council Centre of Excellence for Climate Extremes (CE170100023). MGDK was additionally supported by the ARC Discovery Grants (Grant DP190102025 and DP190101823), the NSW Research Attraction and Acceleration Program, and acknowledges funding from the UK Natural Environment Research Council (NE/W010003/1). We thank the National Computational Infrastructure at the Australian National University, an initiative of the Australian Government for access to supercomputer resources. The authors would also like to thank Daniel Falster at UNSW for his insightful comments and discussions, and the site PIs at ICOS for providing the data on the flux products. Open access publishing facilitated by University of New South Wales, as part of the Wiley - University of New South Wales agreement via the Council of Australian University Librarians.

References

Allen, J. R. M., Forrest, M., Hickler, T., Singarayer, J. S., Valdes, P. J., & Huntley, B. (2020). Global vegetation patterns of the past 140,000 years. *Journal of Biogeography*, *47*(10), 2073–2090. https://doi.org/10.1111/jbi.13930

Alton, P. B. (2011). How useful are plant functional types in global simulations of the carbon, water, and energy cycles? *Journal of Geophysical Research*, *116*(G1), G01030. https://doi.org/10.1029/2010JG001430

AmeriFlux. (2023). AmeriFlux site search. Retrieved from https://ameriflux.lbl.gov/sites/site-search/

Amiro, B. D., Barr, A. G., Barr, J. G., Black, T. A., Bracho, R., Brown, M., et al. (2010). Ecosystem carbon dioxide fluxes after disturbance in forests of North America. *Journal of Geophysical Research*, *115*(G4). https://doi.org/10.1029/2010JG001390

Anderegg, L. D. L., Griffith, D. M., Cavender-Bares, J., Riley, W. J., Berry, J. A., Dawson, T. E., & Still, C. J. (2022). Representing plant diversity in land models: An evolutionary approach to make "Functional Types" more functional. *Global Change Biology*, *28*(8), 2541–2554. https://doi.org/10.1111/gcb.16040

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173–4222. https://doi.org/10.5194/bg-17-4173-2020

Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The advantages of careful seeding. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, *8*, 1027–1035. https://doi.org/10.1145/1283383.1283494

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2018.214

Beierkuhnlein, C., & Fischer, J.-C. (2021). Global biomes and ecozones – Conceptual and spatial communalities and discrepancies. *Erdkunde*, *75*(4), 249–270. https://doi.org/10.3112/erdkunde.2021.04.01

Belward, A. S. (1996). *The IGBP-DIS global 1 km land cover data set "DISCover": Proposal and implementation plans: Report of the land recover working group of IGBP-DIS*. IGBP-DIS.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., et al. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, *16*(3), 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Boelaert, J., Ollion, E., Sodoge, J., Megdoud, M., Naji, O., Kote, A. L., et al. (2022). aweSOM: Interactive self-organizing maps. [Software]. https://cran.rstudio.com/web/packages/aweSOM/

Bonan, G. B. (1996). *A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide (Tech. Rep. Nos. NCAR/TN-417+STR)*. University Corporation for Atmospheric Research. Retrieved from https://opensky.ucar.edu/islandora/object/technotes%3A185/

Bonan, G. B., Levis, S., Kergoat, L., & Oleson, K. W. (2002). Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, *16*(2), 5-1. https://doi.org/10.1029/2000GB001360

Bousquet, P., Peylin, P., Ciais, P., Quéré, C. L., Friedlingstein, P., & Tans, P. P. (2000). Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science*, *290*(5495), 1342–1346. https://doi.org/10.1126/science.290.5495.1342

Box, E. O. (1996). Plant functional types and climate at the global scale. *Journal of Vegetation Science*, *7*(3), 309–320. https://doi.org/10.2307/3236274

Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., et al. (2020). Copernicus global land service: Land cover 100m: Collection 3: Epoch 2015: Globe [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.3939038

Butler, E. E., Wythers, K. R., Flores-Moreno, H., Ricciuto, D. M., Datta, A., Banerjee, A., et al. (2022). Increasing functional diversity in a global land surface model illustrates uncertainties related to parameter simplification. *Journal of Geophysical Research: Biogeosciences*, *127*(3), e2021JG006606. https://doi.org/10.1029/2021JG006606

Cranko Page, J. (2023). PFTFluxRegimes [Software]. Zenodo. https://doi.org/10.5281/ZENODO.8365252

Cranko Page, J., De Kauwe, M. G., Abramowitz, G., Cleverly, J., Hinko-Najera, N., Hovenden, M. J., et al. (2022). Examining the role of environmental memory in the predictability of carbon and water fluxes across Australian ecosystems. *Biogeosciences*, *19*(7), 1913–1932. https://doi.org/10.5194/bg-19-1913-2022

DeFries, R. S., Field, C. B., Fung, I., Justice, C. O., Los, S., Matson, P. A., et al. (1995). Mapping the land surface for global atmosphere-biosphere models: Toward continuous distributions of vegetation's functional properties. *Journal of Geophysical Research*, *100*(D10), 20867–20882. https://doi.org/10.1029/95JD01536

De Long, J. R., Jackson, B. G., Wilkinson, A., Pritchard, W. J., Oakley, S., Mason, K. E., et al. (2019). Relationships between plant traits, soil properties and carbon fluxes differ between monocultures and mixed communities in temperate grassland. *Journal of Ecology*, *107*(4), 1704–1719. https://doi.org/10.1111/1365-2745.13160

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., et al. (2016). The global spectrum of plant form and function. *Nature*, *529*(7585), 167–171. https://doi.org/10.1038/nature16489

Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., et al. (2014). A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 286–300. https://doi.org/10.1002/2013JG002392

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: Anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*(8), 439–447. https://doi.org/10.1890/070062

Euskirchen, E. S., Kane, E. S., Edgar, C. W., & Turetsky, M. R. (2020). When the source of flooding matters: Divergent responses in carbon fluxes in an Alaskan rich fen to two types of inundation. *Ecosystems*, *23*(6), 1138–1153. https://doi.org/10.1007/s10021-019-00460-z

Falster, D., Gallagher, R., Wenk, E. H., Wright, I. J., Indiarto, D., Andrew, S. C., et al. (2021). AusTraits, a curated plant trait database for the Australian flora. *Scientific Data*, *8*(1), 254. https://doi.org/10.1038/s41597-021-01006-6

Famiglietti, C. A., Smallman, T. L., Levine, P. A., Flack-Prain, S., Quetin, G. R., Meyer, V., et al. (2021). Optimal model complexity for terrestrial carbon cycle prediction. *Biogeosciences*, *18*(8), 2727–2754. https://doi.org/10.5194/bg-18-2727-2021

Famiglietti, C. A., Worden, M., Quetin, G. R., Smallman, T. L., Dayal, U., Bloom, A. A., et al. (2023). Global net biome CO_2 exchange predicted comparably well using parameter–environment relationships and plant functional types. *Global Change Biology*, *29*(8), 2256–2273. https://doi.org/10.1111/gcb.16574

FAO. (2012). *Global ecological zones for FAO forest reporting: 2010 update* (Tech. Rep.). United Nations Food and Agriculture Organisation. Retrieved from https://www.fao.org/3/ap861e/ap861e.pdf

Fischer, J.-C., Walentowitz, A., & Beierkuhnlein, C. (2022a). The biome inventory – Standardizing global biogeographical land units [Dataset]. Dryad, *31*(11), 2172–2183. https://doi.org/10.5061/DRYAD.HQBZKH1JM

Fischer, J.-C., Walentowitz, A., & Beierkuhnlein, C. (2022b). The biome inventory – Standardizing global biogeographical land units. *Global Ecology and Biogeography*, *31*(11), 2172–2183. https://doi.org/10.1111/geb.13574

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4). https://doi.org/10.1029/2018MS001453

FLUXNET (2023). List of FLUXNET 2015 sites. Retrieved from https://fluxnet.org/sites/site-list-and-pages/

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*(1), 168–182. https://doi.org/10.1016/j.rse.2009.08.016

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/essd-14-4811-2022

Funk, J. L., Larson, J. E., Ames, G. M., Butterfield, B. J., Cavender-Bares, J., Firn, J., et al. (2017). Revisiting the Holy Grail: Using plant functional traits to understand ecological processes. *Biological Reviews*, *92*(2), 1156–1173. https://doi.org/10.1111/brv.12275

Google Maps. (2023). Location of BE-Lcr. Retrieved from https://maps.app.goo.gl/xLDmAdCc3njS2Gdq6

Griebel, A., Bennett, L. T., Metzen, D., Pendall, E., Lane, P. N. J., & Arndt, S. K. (2020). Trading water for carbon: Maintaining photosynthesis at the cost of increased water loss during high temperatures in a temperate forest. *Journal of Geophysical Research: Biogeosciences*, *125*(1), e2019JG005239. https://doi.org/10.1029/2019JG005239

Groenendijk, M., Dolman, A. J., van der Molen, M. K., Leuning, R., Arneth, A., Delpierre, N., et al. (2011). Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data. *Agricultural and Forest Meteorology*, *151*(1), 22–38. https://doi.org/10.1016/j.agrformet.2010.08.013

Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., et al. (2022). A 29-year time series of annual 300-metre resolution plant functional type maps for climate models. *Earth System Science Data Discussions*, 1–37. https://doi.org/10.5194/essd-2022-296

Harrison, S. P., Cramer, W., Franklin, O., Prentice, I. C., Wang, H., Brännström, Å., et al. (2021). Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *New Phytologist*, *231*(6), 2125–2141. https://doi.org/10.1111/nph.17558

Harrison, S. P., Prentice, I. C., Barboni, D., Kohfeld, K. E., Ni, J., & Sutra, J.-P. (2010). Ecophysiological and bioclimatic foundations for a global plant functional classification. *Journal of Vegetation Science*, *21*(2), 300–317. https://doi.org/10.1111/j.1654-1103.2009.01144.x

Haughton, N., Abramowitz, G., & Pitman, A. J. (2018). On the predictability of land surface fluxes from meteorological variables. *Geoscientific Model Development*, *11*(1), 195–212. https://doi.org/10.5194/gmd-11-195-2018

Haughton, N., Abramowitz, G., Pitman, A. J., Or, D., Best, M. J., Johnson, H. R., et al. (2016). The plumbing of land surface models: Is poor performance a result of methodology or data quality? *Journal of Hydrometeorology*, *17*(6), 1705–1723. https://doi.org/10.1175/JHM-D-15-0171.1

Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., & Prentice, I. C. (2018). Global mapping of potential natural vegetation: An assessment of machine learning algorithms for estimating land potential. *PeerJ*, *6*, e5457. https://doi.org/10.7717/peerj.5457

Higgins, S. I., Buitenwerf, R., & Moncrieff, G. R. (2016). Defining functional biomes and monitoring their change globally. *Global Change Biology*, *22*(11), 3583–3593. https://doi.org/10.1111/gcb.13367

Hijmans, R. J. (2023). Raster: Geographic data analysis and modeling [Software]. https://CRAN.R-project.org/package=raster

Hoover, D. L., Lauenroth, W. K., Milchunas, D. G., Porensky, L. M., Augustine, D. J., & Derner, J. D. (2021). Sensitivity of productivity to precipitation amount and pattern varies by topographic position in a semiarid grassland. *Ecosphere*, *12*(2), e03376. https://doi.org/10.1002/ecs2.3376

ICOS, R. I. (2023). Ecosystem final quality (L2) product in ETC-Archive format – Release 2023-1 (Version 1.0) [Dataset]. ICOS ERIC – Carbon Portal. https://doi.org/10.18160/YDH2-VFYE

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Joswig, J. S., Wirth, C., Schuman, M. C., Kattge, J., Reu, B., Wright, I. J., et al. (2022). Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nature Ecology & Evolution*, *6*(1), 36–50. https://doi.org/10.1038/s41559-021-01616-8

Jucker, T., Fischer, F. J., Chave, J., Coomes, D. A., Caspersen, J., Ali, A., et al. (2022). Tallo: A global tree allometry and crown architecture database. *Global Change Biology*, *28*(17), 5254–5268. https://doi.org/10.1111/gcb.16302

Kaski, S., & Lagus, K. (1996). Comparing self-organizing maps. In C. von derMalsburg, W. vonSeelen, J. C. Vorbrüggen, & B. Sendhoff (Eds.), *Artificial neural networks — ICANN 96* (pp. 809–814). Springer. https://doi.org/10.1007/3-540-61510-5_136

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., et al. (2020). TRY plant trait database – Enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. https://doi.org/10.1111/gcb.14904

Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., et al. (2011). TRY – A global database of plant traits. *Global Change Biology*, *17*(9), 2905–2935. https://doi.org/10.1111/j.1365-2486.2011.02451.x

Kaufman, L., & Rousseeuw, P. J. (1990a). Agglomerative nesting (program AGNES). In *Finding groups in data* (pp. 199–252). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470316801.ch5

Kaufman, L., & Rousseeuw, P. J. (1990b). *Finding groups in data: An introduction to cluster analysis*. Wiley.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/BF00337288

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Konings, A. G., & Gentine, P. (2017). Global variations in ecosystem-scale isohydricity. *Global Change Biology*, *23*(2), 891–905. https://doi.org/10.1111/gcb.13389

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., et al. (2020). Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) at Barro Colorado Island, Panama. *Biogeosciences*, *17*(11), 3017–3044. https://doi.org/10.5194/bg-17-3017-2020

Kuhn, M., & Max (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Laine, A. M., Korrensalo, A., & Tuittila, E.-S. (2022). Plant functional traits play the second fiddle to plant functional types in explaining peatland CO2 and CH4 gas exchange. *Science of the Total Environment*, *834*, 155352. https://doi.org/10.1016/j.scitotenv.2022.155352

Lavorel, S., & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail. *Functional Ecology*, *16*(5), 545–556. https://doi.org/10.1046/j.1365-2435.2002.00664.x

Leemans, R. (1990). Possible changes in natural vegetation patterns due to global warming (IIASA Working Paper). WP-90-008. https://pure.iiasa.ac.at/id/eprint/3443/

Le Quere, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., & Woodward, F. I. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, *2*(12), 831–836. https://doi.org/10.1038/ngeo689

Lin, Y.-S., Medlyn, B. E., Duursma, R. A., Prentice, I. C., Wang, H., Baig, S., et al. (2015). Optimal stomatal behaviour around the world. *Nature Climate Change*, *5*(5), 459–464. https://doi.org/10.1038/nclimate2550

Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters*, *12*(4), 044020. https://doi.org/10.1088/1748-9326/aa66b8

Ma, X., Huete, A., Moore, C. E., Cleverly, J., Hutley, L. B., Beringer, J., et al. (2020). Spatiotemporal partitioning of savanna plant functional type productivity along NATT. *Remote Sensing of Environment*, *246*, 111855. https://doi.org/10.1016/j.rse.2020.111855

MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th Berkeley symposium on Mathematics and Statistics Probability* (pp. 281–297).

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). Cluster: Cluster analysis basics and extensions [Software]. https://CRAN.R-project.org/package=cluster

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Text classification and Naive Bayes. In *Introduction to information retrieval* (pp. 253–287). Cambridge University Press.

Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., et al. (2021). The three major axes of terrestrial ecosystem function. *Nature*, *598*(7881), 468–472. https://doi.org/10.1038/s41586-021-03939-9

Mouselimis, L. (2022). ClusterR: Gaussian mixture models, K-means, mini-batch-kmeans, K-medoids and affinity propagation clustering [Software]. https://CRAN.R-project.org/package=ClusterR

Opitz, J., & Burst, S. (2019). Macro F1 and macro F1. arXiv e-prints. https://doi.org/10.48550/ARXIV.1911.03347

OzFlux. (2023). OzFlux data portal. Retrieved from http://data.ozflux.org.au/portal/home

Pappas, C., Fatichi, S., & Burlando, P. (2016). Modeling terrestrial carbon and water dynamics across climatic gradients: Does plant trait diversity matter? *New Phytologist*, *209*(1), 137–151. https://doi.org/10.1111/nph.13590

Pérez-Ruiz, E. R., Vivoni, E. R., & Sala, O. E. (2022). Seasonal carryover of water and effects on carbon dynamics in a dryland ecosystem. *Ecosphere*, *13*(7), e4189. https://doi.org/10.1002/ecs2.4189

Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., et al. (2020). Interannual variation of terrestrial carbon cycle: Issues and perspectives. *Global Change Biology*, *26*(1), 300–318. https://doi.org/10.1111/gcb.14884

Pillar, V. D. (1999). On the identification of optimal plant functional types. *Journal of Vegetation Science*, *10*(5), 631–640. https://doi.org/10.2307/3237078

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., & Zimmermann, N. E. (2011). Plant functional type mapping for earth system models. *Geoscientific Model Development*, *4*(4), 993–1010. https://doi.org/10.5194/gmd-4-993-2011

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., et al. (2015). Plant functional type classification for earth system models: Results from the European Space Agency's Land Cover Climate Change Initiative. *Geoscientific Model Development*, *8*(7), 2315–2328. https://doi.org/10.5194/gmd-8-2315-2015

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y.-P. (2015). Reliable, robust and realistic: The three R's of next-generation land-surface modelling. *Atmospheric Chemistry and Physics*, *15*(10), 5987–6005. https://doi.org/10.5194/acp-15-5987-2015

Pugh, T. A. M., Arneth, A., Kautz, M., Poulter, B., & Smith, B. (2019). Important role of forest disturbances in the global biomass turnover and carbon sinks. *Nature Geoscience*, *12*(9), 730–735. https://doi.org/10.1038/s41561-019-0427-2

Raczka, B., Dietze, M. C., Serbin, S. P., & Davis, K. J. (2018). What limits predictive certainty of long-term carbon uptake? *Journal of Geophysical Research: Biogeosciences*, *123*(12), 3570–3588. https://doi.org/10.1029/2018JG004504

R Core Team. (2020). R: A language and environment for statistical computing [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., & Baldocchi, D. D. (2014). Linking plant and ecosystem functional biogeography. *Proceedings of the National Academy of Sciences*, *111*(38), 13697–13702. https://doi.org/10.1073/pnas.1216065111

Rogers, A. (2014). The use and misuse of Vc,max in Earth System Models. *Photosynthesis Research*, *119*(1), 15–29. https://doi.org/10.1007/s11120-013-9818-1

Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., et al. (2023). GDAL [Software]. Zenodo. https://doi.org/10.5281/ZENODO.5884351

Scheiter, S., Langan, L., & Higgins, S. I. (2013). Next-generation dynamic global vegetation models: Learning from community ecology. *New Phytologist*, *198*(3), 957–969. https://doi.org/10.1111/nph.12210

Shiklomanov, A. N., Cowdery, E. M., Bahn, M., Byun, C., Jansen, S., Kramer, K., et al. (2020). Does the leaf economic spectrum hold within plant functional types? A Bayesian multivariate trait meta-analysis. *Ecological Applications*, *30*(3), e02064. https://doi.org/10.1002/eap.2064

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., et al. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, *9*(2), 161–185. https://doi.org/10.1046/j.1365-2486.2003.00569.x

Smith, T. M., Smith, T. M., Smith, T. M., Shugart, H. H., Woodward, F. I., & Shugart, P. H. H. (1997). *Plant functional types: Their relevance to ecosystem properties and global change*. Cambridge University Press.

Ştefan, V., & Levin, S. (2018). Plotbiomes: R package for plotting Whittaker biomes with ggplot2 [Software]. Zenodo. https://doi.org/10.5281/ZENODO.7145245

Stoy, P. C., Richardson, A. D., Baldocchi, D. D., Katul, G. G., Stanovick, J., Mahecha, M. D., et al. (2009). Biosphere-atmosphere exchange of CO_2 in relation to climate: A cross-biome analysis across multiple time scales. *Biogeosciences*, *6*(10), 2297–2312. https://doi.org/10.5194/bg-6-2297-2009

Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E., & Moody, A. (1999). MODIS land cover product algorithm theoretical basis document (ATBD) version 5.0. Retrieved from http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf

Tan, S., Wang, H., Prentice, I. C., Yang, K., Nóbrega, R. L. B., Liu, X., et al. (2023). Towards a universal evapotranspiration model based on optimality principles. *Agricultural and Forest Meteorology*, *336*, 109478. https://doi.org/10.1016/j.agrformet.2023.109478

Teckentrup, L., De Kauwe, M. G., Pitman, A. J., Goll, D. S., Haverd, V., Jain, A. K., et al. (2021). Assessing the representation of the Australian carbon cycle in global vegetation models. *Biogeosciences*, *18*(20), 5639–5668. https://doi.org/10.5194/bg-18-5639-2021

Thomas, H. J. D., Myers-Smith, I. H., Bjorkman, A. D., Elmendorf, S. C., Blok, D., Cornelissen, J. H. C., et al., (2019). Traditional plant functional groups explain variation in economic but not size-related traits across the tundra biome. *Global Ecology and Biogeography*, *28*(2), 78–95. https://doi.org/10.1111/geb.12783

Ukkola, A. (2020). PLUMBER2: Forcing and evaluation datasets for a model intercomparison project for land surface models v1.0 [Dataset]. NCI Australia. https://doi.org/10.25914/5FDB0902607E1

Ukkola, A. M., Abramowitz, G., & De Kauwe, M. G. (2022). A flux tower dataset tailored for land model evaluation. *Earth System Science Data*, *14*(2), 449–461. https://doi.org/10.5194/essd-14-449-2022 van Bodegom, P. M., Douma, J. C., & Verheijen, L. M. (2014). A fully traits-based approach to modeling global vegetation distribution. *Proceedings of the National Academy of Sciences*, *111*(38), 13733–13738. https://doi.org/10.1073/pnas.1304551110

Van Bodegom, P. M., Douma, J. C., Witte, J. P. M., Ordoñez, J. C., Bartholomeus, R. P., & Aerts, R. (2012). Going beyond limitations of plant functional types when predicting global ecosystem–atmosphere fluxes: Exploring the merits of traits-based approaches. *Global Ecology and Biogeography*, *21*(6), 625–636. https://doi.org/10.1111/j.1466-8238.2011.00717.x

Wang, A., & Price, D. T. (2007). Estimating global distribution of boreal, temperate, and tropical tree plant functional types using clustering techniques. *Journal of Geophysical Research*, *112*(G1). https://doi.org/10.1029/2006JG000252

Wehrens, R., & Buydens, L. M. C. (2007). Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, *21*(5), 1–19. https://doi.org/10.18637/jss.v021.i05

Wehrens, R., & Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software*, *87*(7), 1–18. https://doi.org/10.18637/jss.v087.i07

Westerband, A. C., Funk, J. L., & Barton, K. E. (2021). Intraspecific trait variation in plants: A renewed focus on its role in ecological processes. *Annals of Botany*, *127*(4), 397–410. https://doi.org/10.1093/aob/mcab011

WFO. (2023). World flora online. http://worldfloraonline.org/

Wilson, M. F., & Henderson-Sellers, A. (1985). A global archive of land cover and soils data for use in general circulation climate models. *Journal of Climatology*, *5*(2), 119–143. https://doi.org/10.1002/joc.3370050202

Wolf, S., Eugster, W., Potvin, C., & Buchmann, N. (2011). Strong seasonal variations in net ecosystem CO_2 exchange of a tropical pasture and afforestation in Panama. *Agricultural and Forest Meteorology*, *151*(8), 1139–1151. https://doi.org/10.1016/j.agrformet.2011.04.002

Wolf, S., Eugster, W., Potvin, C., Turner, B. L., & Buchmann, N. (2011). Carbon sequestration potential of tropical pasture compared with afforestation in Panama. *Global Change Biology*, *17*(9), 2763–2780. https://doi.org/10.1111/j.1365-2486.2011.02460.x

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., et al. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827. https://doi.org/10.1038/nature02403

Wright, J. P., Naeem, S., Hector, A., Lehman, C., Reich, P. B., Schmid, B., & Tilman, D. (2006). Conventional functional classification schemes underestimate the relationship with ecosystem functioning. *Ecology Letters*, *9*(2), 111–120. https://doi.org/10.1111/j.1461-0248.2005.00850.x

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17. https://doi.org/10.18637/jss.v077.i01

Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., et al. (2014). Plant functional types in Earth system models: Past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Annals of Botany*, *114*(1), 1–16. https://doi.org/10.1093/aob/mcu077

Xie, X., Chen, J. M., Gong, P., & Li, A. (2021). Spatial scaling of gross primary productivity over sixteen mountainous watersheds using vegetation heterogeneity and surface topography. *Journal of Geophysical Research: Biogeosciences*, *126*(5), e2020JG005848. https://doi.org/10.1029/2020JG005848

Xu, X., & Trugman, A. T. (2021). Trait-based modeling of terrestrial ecosystems: Advances and challenges under global change. *Current Climate Change Reports*, *7*(1), 1–13. https://doi.org/10.1007/s40641-020-00168-6

Yang, Y., Zhu, Q., Peng, C., Wang, H., & Chen, H. (2015). From plant functional types to plant functional traits: A new paradigm in modelling global vegetation dynamics. *Progress in Physical Geography: Earth and Environment*, *39*(4), 514–535. https://doi.org/10.1177/0309133315582018

Zhang, X., Wu, S., Yan, X., & Chen, Z. (2017). A global classification of vegetation based on NDVI, rainfall and temperature. *International Journal of Climatology*, *37*(5), 2318–2324. https://doi.org/10.1002/joc.4847

Zhou, H., Shao, J., Liu, H., Du, Z., Zhou, L., Liu, R., et al. (2021). Relative importance of climatic variables, soil properties and plant traits to spatial variability in net CO_2 exchange across global forests and grasslands. *Agricultural and Forest Meteorology*, *307*, 108506. https://doi.org/10.1016/j.agrformet.2021.108506