Representativeness of FLUXNET sites across Latin America

Authors: Samuel Villarreal[1,2], Rodrigo Vargas[1]

[1]University of Delaware, Department of Plant and Soil Sciences. 531 South College Avenue. Newark, DE, 19716, USA.

[2]Universidad Autónoma de Querétaro, Facultad de ingeniería. Cerro de las Campanas S/N, Centro Universitario, 76010 Santiago de Querétaro, México.

Corresponding author: Rodrigo Vargas (rvargas@udel.edu)

Key Points:

- There is a need to increase spatial representativeness of terrestrial biophysical process across Latin America
- Enhancing representativeness can improve understanding of biophysical process and model benchmarking and parameterization.
- Need to promote participation to update and share information

Abstract

Environmental observatory networks (EONs) provide information to better understand, model and forecast the spatial and temporal dynamics of Earth's biophysical process. Consequently, representativeness analyses of EONs are important to provide insights for improving EONs' management, design, and interpretation of their value-added products (e.g., datasets, model predictions). We assessed the representativeness of registered FLUXNET sites (n=41, revised on September 2018) across Latin America (LA), a region of great importance for the global carbon and water cycles, which represents nearly 13% of the world's land surface area. Representativeness analyses were performed using a 0.05^{o} spatial grid for multiple environmental variables, gross primary productivity (GPP) and evapotranspiration (ET) across LA. Our results showed a potential spatial representativeness of 34% of the surface area for climate properties, 36% for terrain parameters, 34% for soil resources, and 45% when all aforementioned environmental variables were summarized into a principal component analysis. Furthermore, there was a 48% potential representativeness for GPP and 34% for ET. Unfortunately, data from these 41 sites is not all readily available for the scientific community, limiting synthesis studies and model benchmarking/parametrization. We discussed the need to enhance interoperability, promote the participation of active/inactive sites to share information with local, regional and international networks, and promote monitoring efforts across this region of the world to increase the accuracy of regional-to-global data-driven products.

Plain Language Summary

Environmental observatory networks (EONs) provide information to better understand and forecast the spatial and temporal dynamics of Earth's biophysical process. Consequently, representativeness analyses of EONs are important to provide insights for improving EONs' management, design, and interpretation of their value-added products (e.g., datasets, model predictions). We assessed the representativeness of registered FLUXNET sites across Latin America (LA), a region of great importance for the global carbon and water cycles, which represents nearly 13% of the world's land surface area. Representativeness analyses were performed for multiple environmental variables including gross primary productivity (GPP) and evapotranspiration (ET). Our results showed a potential spatial representativeness of 34% of the surface area for climate properties, 36% for terrain parameters, and 34% for soil resources. Furthermore, there was a 48% potential representativeness for GPP and 34% for ET. Unfortunately, data from these study sites is not all readily available for the scientific community, limiting synthesis studies and model benchmarking/parametrization. We discussed the need to enhance interoperability, promote the participation of active/inactive sites to share information with local, regional and international networks, and promote monitoring efforts across this region of the world to increase the accuracy of regional-to-global data-driven products.

1 Introduction

Environmental monitoring, especially long-term monitoring programs are a backbone component for environmental science and policy (Chabbi et al., 2017; Lovett et al., 2007). Environmental monitoring is fundamental to foster knowledge as it promotes creativity for scientific methodologies, generates invaluable data products, and provides baselines and information to address socio-environmental grand challenges (Lovett et al., 2007; Scholes et al., 2017). It is assumed that developing environmental observatory networks (EONs) would lead to successful environmental monitoring efforts as EONs are entities designed to provide insights to address complex regional-to-global socio-ecological problems through a coordinated effort (Chabbi et al., 2017; Keller et al., 2011; Scholes et al., 2017). Some key tasks lead by EONs include data collection, data sharing and synthesis activities that are useful for scientific discovery and making informed environmental policy or management decisions (Lovett et al., 2007; Scholes et al., 2017; Villarreal et al., 2018).

An example of an EON is FLUXNET, which represents a global network of study sites using the eddy-covariance method to measure the exchange of mass and energy between the land surface and the atmosphere (Baldocchi et al., 2001; Baldocchi, 2020). FLUXNET is a global ‘network of regional networks’ that promotes compilation, harmonization, standardization, archiving, and synthesis activities of eddy-covariance data. The FLUXNET network is present across multiple ecosystems around the world, so it is possible to generate knowledge of the interaction between terrestrial ecosystems and the atmosphere from regions to the global scale (Falge et al., 2002; J. B. Fisher et al., 2008; Keenan et al., 2014; Schwalm et al., 2017). However, FLUXNET sites are neither randomly or systematically distributed, so they under-represent certain regions and ecosystems across the world (Baldocchi, 2020; Kumar et al., 2016; Papale et al., 2015; Villarreal et al., 2019). Consequently, representativeness assessments of EONs are critical as they provide information for EONs design/growth, and insights for interpretations and implications of data-driven (or value-added) products (Sulkava et al., 2011; Villarreal et al., 2018). These assessments are relevant to increase EONs' applicability and to guide regional-to-global management and research efforts (Jongman et al., 2017; Lovett et al., 2007).

The representativeness of EONs has been mostly assessed using climate and vegetation parameters (Hargrove et al., 2003; Kumar et al., 2016; Sulkava et al., 2011). For example, through stratification of climate, vegetation and soil information some studies have assessed the representativeness of AmeriFlux and FLUXNET (Hargrove et al., 2003; Kumar et al., 2016), while recent studies have incorporated functional information from ecosystems (Alcaraz-Segura et al., 2017; Villarreal et al., 2018, 2019). A common approach to assess representativeness of EONs has been the estimation of minimum distances within a multivariate space (Hargrove et al., 2003; Sulkava et al., 2011). An alternative approach is the use of machine learning techniques, which estimate the spatial distribution of the environmental range monitored by the EON's study sites (i.e., nodes) across the spatial domain of the network (Villarreal et al., 2018, 2019).

We propose that it is possible to assess the representativeness of EONs based on concepts derived from species distributions models (SDMs). Briefly, SDMs define a geographic space that includes a set of environmental data layers, and then delineate an area within the geographic space that corresponds to environmental properties that are suitable for the presence of a certain species (Drew et al., 2011; Evans et al., 2011). We propose that this concept can be applied to assess the representativeness of EONs, since the goal is to delineate the spatial distribution of environmental factors across a geographic space that should be similar to the environmental range monitored by corresponding monitoring sites within an EON (Villarreal et al., 2018).

Here, we present a representativeness assessment of eddy-covariance sites registered with FLUXNET across Latin America (LA). LA is a region that is largely characterized by its wide ecosystem diversity along with a broad gradient of land-use and land-use-change types. LA includes nearly 13% of the global land surface area but only about 5% of all registered FLUXNET sites are located within this region (estimated for year 2018, but see methods for details). The density of registered FLUXNET sites in LA is very low when compared to regions such as the United States or Europe, and global data-driven products are forced to use information from FLUXNET sites outside LA to predict patterns in LA. Hence, a representativeness analysis is needed to better interpret the available information within LA and the output of regional-to-global data-driven products parameterized with FLUXNET data.

The overarching goal of this study is to provide an assessment of the representativeness of registered FLUXNET sites across LA to monitor environmental factors such as climate, topography and soil resources along with ecosystem process such as gross primary productivity (GPP) and evapotranspiration (ET). We asked four interrelated research questions: 1) What is the representativeness of FLUXNET sites across LA to characterize climate, topography and soil resources spatial variability? 2) What is the representativeness of FLUXNET sites to monitor GPP and ET patterns across LA? 3) How does representativeness of FLUXNET sites to monitor GPP and ET varies as spatial scale changes? and 4) How many more sites are needed to substantially improve representativeness of GPP and ET across this region? Finally, this study is based on publicly available information and open source software, so this framework can be applied anywhere across the world.

2 Data and Methods

2.1 FLUXNET registered sites

FLUXNET provides standardized data products through coordination among multiple regional eddy covariance networks across the globe (http://fluxnet.fluxdata.org). We used this online database to extract the geographical location of eddy-covariance sites across LA registered with FLUXNET. We identified 41 registered sites distributed across different ecosystems (revised on September 2018; Supplementary Material Table S1) and we considered these sites for further analyzes despite they are active or inactive and if they have provided data or not to the FLUXNET database. Consequently, this study is conservative and provides a potential representation of eddy-covariance sites across LA. We hope that this study will encourage principal investigators to register their sites and share data with FLUXNET to improve the representation of LA in regional and global studies.

We recognize that there are several challenges and assumptions for performing an accurate representativeness assessment of eddy-covariance sites across LA. First, the assumption of 41 registered FLUXNET sties for year 2018 does not mean that those sites are active nor that their data is or will be available for the scientific community. Second, there are unregistered eddy-covariance sites across LA and new sites are being installed or have become inactive. We clarify that our assessment is a potential representation because data from these 41 sites is not readily available for the scientific community. For example, the FLUXNET2015 dataset only includes 7 eddy covariance sites across LA (i.e., BR-Sa1, BR-Sa2, GF-Guy, Pa-SPn, PA-SPs, AR-SLu, AR-Vir), and the AmeriFlux network (revised August 2020) has 23 registered sites but only 8 of them share data with the network (Supplementary Material Table S2).

2.2 Environmental factors

A set of variables related to climate, terrain parameters, and soil resources variability were used to assess the representatives of environmental state factors, as they constrain the spatial patterns of ecosystem processes such as GPP and ET (Amundson, 1991; Chapin et al., 2002). We used 19 bioclimatic predictors to characterize climate conditions: mean annual conditions (i.e., annual mean temperature, annual precipitation), mean annual seasonal conditions (i.e., temperature seasonality) and intra-annual seasonal conditions (i.e., mean temperature of the driest quarter or precipitation of the wettest quarter) of temperature and precipitation (Hijmans et al., 2005). Terrain parameters were characterized by slope, elevation, topographic water capacity, and solar radiation index. Soil resources were characterized by soil organic carbon, soil nitrogen, soil phosphorus and soil water content. The bioclimatic predictors were downloaded from *worldclim.org* (accessed May 2018). Most terrain parameters and soil resources variables were downloaded from *worldgrids.org* (accessed May 2018), but soil organic carbon was downloaded from *www.fao.org* (accessed May 2018) and soil phosphorus from *data.nasa.gov* (accessed May 2018). Respectively, NASA-MODIS products MOD17A2 and MOD16A2 from 2001 to 2014 were used to characterize GPP and ET as previously done for assessment of the AmeriFlux and MexFlux networks (Villarreal et al., 2018, 2019). The statistic parameters used to characterize GPP and ET dynamics were the mean (GPP_mean, ET_mean) and the coefficient of variation (GPP_CV, ET_CV), since they have been used as proxies to represent ecosystem productivity and seasonality, respectively (Alcaraz-Segura et al., 2017; Villarreal et al., 2018, 2019).

2.2 Data harmonization and FLUXNET representativeness

All variables were standardized into a similar geographical system (GS), which consisted in harmonizing all variables into the same projection (i.e., WGS84) and transforming them into the same spatial resolution (i.e., 0.05^{o}). We selected 0.05^{o} as this resolution is largely used to represent environmental patterns at a regional scale (Chrysoulakis et al., 2003; Löw et al., 2011) and has been used to assess the representativeness of AmeriFlux, MexFlux and the National Ecological Observatory Network (NEON; Villarreal et al., 2018; Villarreal et al., 2019). In addition, all variables representing climate, terrain parameters and soil resources were reduced in dimensionality using a principal component analysis (PCA) to assess the representativeness of these combined environmental factors (using the first two principal components) from a multivariate approach. Representativeness were performed for all environmental parameters at 0.05^{0}, while GPP and ET representativeness were estimated at 0.05^{0}, 0.25^{0}, 0.50^{0} and 1.0^{0}, since global models of GPP and ET are usually estimated based on these spatial resolutions (Fisher and Koven, 2020).

Representativeness was estimated using random forest (RF) applied for SDMs. RF is a widely used technique in SDMs, especially for rare species that have few observations over a broad region (Cutler et al., 2007; Evans et al., 2011). We propose that the relative few numbers of eddy-covariance sites across the large geographic extent of LA is a similar case study. As a machine-learning technique, RF produces classifications trees from bootstrapping samples from a given dataset (i.e., training-data), while the observations that are not consider (out-of-bag data) are later used for predictions and model evaluation. First, classifications trees (CT) are built from sample bootstrapping by repeatedly partitioning the training-data into a binary-series of clusters (i.e., child-nodes) that split the data into more or less homogeneous child-nodes with respect to the response variable, this process continues with each child-node until stops (Marmion et al., 2009). Second, the grown trees are used to predict the out-of-bag observations. The class that is predicted of an observation is estimated by the majority vote of the out-of-bag predictions for that observation (Cutler et al., 2007; Evans et al., 2011; Marmion et al., 2009). Finally, RF produces a raster map that represents the relative similarity of each pixel to the sample points or presence data (Schmitt et al., 2017), which in this case corresponds to the geographic locations of FLUXNET sites across LA.

Model performance was assessed using True Skin Statistics (TSS), which corresponds to the sum of the model sensitivity (i.e., proportion of presence correctly predicted) and the specificity (i.e., proportion of absence correctly predicted) minus one. TSS ranges from -1 to 1, being -1 a predictability power worse than a random model, 0 indicates a random predictability, and 1 corresponds to a perfect model (Liu et al., 2011). Absence data points were generated by random selection, as randomly selected points usually produce reliable distribution models (Barbet-Massin et al., 2012). Also, the uncertainty of each model was assessed by repeating 10 times each model with only one iteration and calculate their mean and standard deviation, as it was performed in a previous study (Villarreal et al., 2019).

The optimal number of absence data and model repetition for each environmental set of variables (i.e., climate, terrain parameters, soil resources, GPP and ET) was selected based on the TSS by an iterative process. We selected the number of absence data and model repetition that had the higher TSS (Barbet-Massin et al., 2012). The number of absence and repetitions were different for each environmental set of variables. Data management and analysis were performed using the R programming language (R project for statistical computing; www.r-project.org) and the 'SSDM' library (Schmitt et al., 2017).

2.3 Improving the representativeness of GPP and ET across LA

A final goal was to provide insights about how many more sites are needed across LA to improve representativeness of GPP and ET. To this end, we used the constrained Latin hypercube sampling technique (cLHS; Minasny & McBratney, 2006). The cLHS is a multivariate statistical technique that ensures a full coverage of the range of the variables involved in the multivariate space. For this study we used the mean and standard deviation of GPP and ET as discussed earlier. The cLHS serves as an efficient sampling strategy and it has been previously used for EON's representativeness analysis (Villarreal et al., 2019). For this assessment we followed a sequential approach: (a) we started by adding additional sites across LA in increments of 10 sites until reaching 100 sites; then (b), additional sites were added in increments of 20 until reaching 200 sites across LA. We stopped at 200 potential sites across LA as an arbitrary number equivalent to the approximate total number of eddy-covariance sites registered in the AmeriFlux network for the conterminous United States. The assessment of representativeness by adding new sites was also performed using RF as described above.

3 Results

3.1 Distribution of FLUXNET sites across LA

There is a large diversity of terrestrial ecosystems across LA with 15 out of 16 possible International Geosphere-Biosphere Program (IGBP) categories (MODIS MCD12Q1, 2012), but the extensions of these categories are not evenly distributed. For example, 5 out of 16 categories incorporate 80% of LA land surface, where the largest categories are Evergreen Broadleaf Forest (34% of LA) and Savanna (19% of LA). The less extended categories are Evergreen Needle-Leaf Forest (0.05% of LA) and Closed Shrublands (0.04% of LA). The number of FLUXNET sites at each IGBP category is also not evenly distributed (Table 1). For example, Evergreen Broadleaf

Forest has the highest number sites (n = 19), followed by Woody Savanna (n = 8) and Open Shrublands (n = 5). Six IGBP categories do not have any sites registered in the FLUXNET network (Table 1).

3.2 Representativeness of environmental factors, GPP and ET

The representativeness of FLUXNET sites at 0.05^0 differed for each environmental factor (i.e., climate, terrain properties, soil resources, combined environmental properties) or ecosystem process assessed (i.e., GPP and ET). The tested variables with the highest spatial representativeness were GPP (48%), combined environmental factors (45%; derived from the PCA) and terrain parameters (36%); while climate, soil resources and ET had similar representativeness (34%; Table 2).

The representativeness between the IGBP categories was different among the distinct environmental factors. The categories with the highest representativeness among the different environmental variables assessed were Shrublands and Savannas, while forest ecosystems (Evergreen and Deciduous Broadleaf Forest) had similar values as those from managed ecosystems (e.g., Croplands; Supplementary Material Table S3).

After assessing the representativeness of the environmental factors and ecosystems processes targeted in this study, we identified the two most important variables for each representativeness model and proceed to find differences between represented and non-represented regions (Table 3). For the bioclimatic predictors (Figure 2a-b), precipitation seasonality above 120 mm and below 40 mm and the annual mean diurnal temperature range above 20 ^{0}C and below 6 ^{0}C were not represented (Figure 4c-d). For the terrain properties the majority of the IGBP classes were represented, while values 90>TWI>75 were not represented for TWI (topographic wetness index; Figure 2e-f)). For soil resources (Figure 2g-h) soil organic carbon >80 g/m^2 and soil nitrogen below <500 and above >2000 mg/m^2 were not represented (Figure 2i-j). For all the environmental drivers combined into a PCA, the components that had the highest influence on PCA were *terrain complexity* (21% of total variability) and soil nitrogen (6% of total variability). Assessing the representativeness on those PCA1 and PCA 2, PC1 was only represented within the range -0.08 to 0.16 (Figure 2k-l) while PC2 was represented within the range of -1 to 1.5 (Figure 2 m-n). In order to assessed differences between the areas represented and non-represented we compared the 0.95 mean confidence interval for all the variables assessed (Table 3).

The representativeness of GPP and ET at 0.05^0 is shown in Figure 3. For GPP, the representativeness of FLUXNET sites for GPP_mean was bias towards values above 4 g/day, while for GPP_CV only values above > 2 g/day were not represented. The mean for GPP_mean and GPP_CV across all LA was 4.3 g/day and 1.20 g/day, respectively. GPP_mean and GPP_CV monitored by the FLUXNET sites were 4.43 g/day and 1.24 g/day, respectively, while for the non-represented areas GPP_mean and GPP_CV were 4.21 and 1.16 g/day (table 3; Figure 3 a-d). The representativeness of ET, the ET_mean representativeness of FLUXNET sites was biased towards values above 2.5 mm/day, while for ET_CV values close to 2.0 mm/day were not represented. The mean for ET_mean and ET_CV across all LA were 2.51 mm/day and 0.71 mm/day, respectively. The ET_mean and ET_CV monitored by FLUXNET sites were 2.42 mm/day and 0.67 mm/day, respectively (Figure 3e-f). While for the non-represented areas ET_mean and ET_CV were 2.42 and 0.68 mm/day, respectively (Figure 3 g-h). In order to assessed differences between the areas represented and non-represented we compared the 0.95 mean confidence interval for all the variables assessed (Table 3). The representativeness of GPP and ET at 0.25^0, 0.50^0 and 1.0^0 slightly decreased as the spatial resolution became coarser. For example, GPP representativeness was 24%, 21% and 18% while ET had 22%, 19% and 16% at 0.25^0, 0.50^0 and 1.0^0, respectively. The decrease in representativeness was associated to a decrease in the number of effective sites (35, 31 and 30 effective sites) as the spatial resolution was coarser because multiple sites can be included within larger pixels.

The spatial representativeness for the two variable that had the largest influence for each environmental factor and for GPP and ET is shown in Figure 4 - 5, respectively. The variables with the largest influence for bioclimatic factor are precipitation seasonality and annual mean diurnal range (Figure 4 a-b). For terrain parameters are IGBP class and topography wetness index (Figure 4 c-d). For soil are soil organic carbon and soil nitrogen (Figure 4 e-f). While for the PCA are PC1 and PC2 (Figure 5 g-h). Figure 5 shows the spatial representativeness for GPP and ET; GPP_mean had a larger influence than GPP_CV (Figure 5 a-b), while ET had ET_CV as the variable with a larger influence than ET_mean (Figure 5 c-d).

3.3 Representativeness by adding potential new sites and uncertainty assessment

The overall representativeness across LA substantially increased by adding new study sites from 45% to 86% for GPP and from 42% to 80% for ET (Figure 6a, b; Table 4). Our analysis suggests that if 200 sites were to be installed across LA, then efforts should be focused across Grasslands, Savanna, Open Scrublands and Evergreen Broadleaf Forests (Table 4). The addition of new study sites progressively increased the predictive power for each model. This was indicated by the validation statistic (e.g., area under de curve; AUC) which increased from 0.003 to 0.018 and 0.003 to 0.017 for GPP and ET, respectively (Figure 6).

4 Discussion

4.1 Representativeness across land cover, climate, edaphic and topographic characteristics

Overall, the potential spatial representativeness across LA included 34% of the surface area for climate properties, 36% for terrain parameters, 34% for soil resources, and 45% when all environmental properties were combined in a multivariate space. Here we highlight the representativeness of shrublands, woody savannas, croplands and forested ecosystems (i.e., evergreen broadleaf forest and deciduous forest) as important IGBP classes across LA.

Shrublands and woody savannas had the highest representativeness for environmental properties (i.e., climate, edaphic, topographic, and combined environmental properties) across all IGBP classes (Supplementary Material Table S3). This is explained because those ecosystems have a relative high ratio of eddy-covariance sites per surface (Table 1), but their high representativeness also suggests that those ecosystems are likely less environmentally heterogeneous, arguably due to: a) limited water conditions/stress that trigger common strategies such as water-use efficiency (Biederman et al., 2016; Huxman et al., 2004; Ponce-Campos et al., 2013); and b) a high convergence on functional and structural properties such as shrublands and savannas between North and South America (Paruelo et al., 1998). The functional and structural similarities between the shrublands and savannas within North and South America could provide insights about why scattered sites could contribute to a relative high representativeness, even though most of these sites are located at the northern domain of LA (Figure 1). These results support the assumption that upscaling eddy-covariance data could be performed by using information from few sites representing some IGBP classes around the world (Jung et al., 2019), but we highlight that this assumption may not be widely applicable for all IGBP classes under different environmental conditions (Vargas et al., 2013). Despite the relative high representativeness of shrublands and woody savannas, changing climate conditions such as droughts (Biederman et al., 2016; Villarreal et al., 2016) along with land cover changes (due to land use change and disturbances) could decrease expected regional functional and structural similarities and consequently increase environmental heterogeneity across these IGBP classes.

The representativeness of forested ecosystems such as evergreen broadleaf forest and deciduous forest were consistently lower than shrublands and woody savannas (Supplementary Material Table S3). These results are relevant because more than 50% of the sites are located across evergreen broadleaf forest and deciduous forest (Table 1). These results suggest a larger heterogeneity in climatic, topographic and soil resources that ultimately influence the lower representativeness within forested ecosystems. This could be a result of a larger precipitation gradient across forested ecosystems than for shrublands and savannas (Chapin et al., 2002), an arguably more dynamic cycling of soil nutrients (Chapin et al., 1986; Vitousek, 1984), and larger variability in canopy phenology (Richardson et al., 2013). Improvements are needed for monitoring the large range of environmental properties of forested ecosystems across LA to properly assess and forecast the influence of global environmental change within FLUXNET sites (Baldocchi, 2020; Keenan et al., 2014). We recognize that multiple efforts on tropical wet forest have provided important information and have fostered our knowledge on carbon and water fluxes at natural, converted and afforested sites across LA (Andreae et al., 2002; Avissar et al., 2002; Keller et al., 2004). However, we highlight that the scientific community should be cautions when extrapolating information from a few forested sites within LA or assuming that other forested sites across the world accurately represent characteristics within LA.

Other ecosystems such as croplands and cropland/natural-vegetation mosaics have a relative similar representativeness (as forested ecosystems) among environmental properties despite a substantial smaller number of available sites (Table 1). This lower representativeness influences the development and benchmarking of terrestrial ecosystem models as they usually have lower performance across croplands (Schaefer et al., 2012). This limitation may hinder our capacity of forecast the influence of global environmental change on food production and food security within LA (Graesser et al., 2015; Ramankutty et al., 2002). Since warming and droughts are expected to cause large damage on agriculture in developing countries, a well monitored program that allows to collect information on the impact of global environmental change on food production within this region is imperative in order to design adaptation strategies (Graesser et al., 2015; Ramankutty et al., 2002).

Other IGBP classes such as mixed forest, grasslands and permanent wetlands have also been recognize for their important role in climate regulation and soil nutrient cycling (Baldocchi et al., 2000; Conant et al., 2001; Whiting & Chanton, 2001). Unfortunately, none of these IGBP classes were represented in our analysis (Table 1) and consequently highlight the potential bias of data-driven products when using information from sites outside LA to predict ecosystem processes within LA. This challenge is also observed for the recently compiled FLUXNET-CH4 dataset where there is minimal information form tropical wetlands (despite their importance for the global methane cycle; Poulter et al 2017) and very few sites from LA (Knox et al., 2019).

4.2 Representativeness for GPP and ET

There was a 48% potential representativeness for GPP and 34% for ET considering 41 eddy-covariance sites distributed across LA. These results compare with the representativeness by the MexFlux network of 3% for GPP and 5% for ET across Mexico (Villarreal et al., 2019), or the representativeness by AmeriFlux sites of 46% of the spatial functional heterogeneity (i.e., enhanced vegetation index dynamics) across the conterminous United Sates (Villarreal et al., 2018). Efforts to assess the representativeness of FLUXNET have focused on testing if measurements taken at specific locations can be extrapolated at explicit spatial and temporal extents (Chu et al., 2017; He et al., 2015; Yang et al., 2008), broadly concluding that representativeness largely depend on the heterogeneity of fine-scale ecosystem processes (Chen et al., 2011, 2012; He et al., 2015). Those studies highlight the challenges involved in representing regional-to-global information across heterogeneous regions (e.g., Mexico with large environmental gradients) and across larger regions even with high density of eddy covariance sites (e.g., United States and Europe; Sulkava et al., 2011; Villarreal et al., 2018).

Our results show that GPP_mean and ET_mean derived from areas represented by FLUXNET sites are substantially lower than the non-represented region across LA, while GPP_CV and ET_CV were relatively similar between both areas (Figures 3). We suggest that those results may represent the overall lower representativeness across forested sites. This could be possible because at a regional and global scale forested ecosystems largely influence the annual mean of GPP, while the interannual variability is mostly influenced by water-limited ecosystems (Ahlstrom 2014), which arguably are better represented across LA.

We highlight that our assessment of potential representativeness for GPP and ET across LA is conservative and should be seen as a current "best case" scenario if data from all sites is available for the scientific community. Consequently, most studies that have used available FLUXNET data (e.g., FLUXNET2015) for upscaling purposes, parameterize models, or performed data-syntheses that include LA are likely biased. The FLUXNET network has provided important information about how ecosystem metabolism responds to biophysical factors and key insights for improving terrestrial ecosystem process models (Baldocchi, 2020), however, its accuracy to represent these ecosystem processes at regional and global scales largely depends on the density, distribution and diversity of sites. The addition of new sites across unrepresented regions of the world will be able to test the current assumptions of biophysical drivers for ecosystem processes and the importance of these regions for regional-to-global water and carbon cycles.

Previous studies have assessed how the distribution and density of monitoring sites affect the prediction capacity of upscaling models to properly represent GPP and ET at larger spatial scales and to capture their inter-annual variability (Papale et al., 2015; Sulkava et al., 2011; Zhang et al., 2020). Those results suggest that regions and/or continents with relatively small number of monitoring sites (e.g., Latin-America, Africa) could led to predictions with large errors. This is of utmost importance for regions such as LA, since it is home of some of the largest rainforest across the globe (e.g., Amazon basin, Mayan rainforest) which are key to understand regional-to-global water and carbon cycles (Ahlström et al., 2016; Saatchi et al., 2011). Furthermore, water-limited ecosystems, which are important for the interannual variability of the global carbon cycle (Ahlström et al., 2016; Poulter et al., 2014), may have different GPP and ET dynamics as expected for other water-limited ecosystems around the world (Zhang et al., 2020).

Finally, there is an important challenge for monitoring and representing GPP and ET dynamics across LA. Most FLUXNET sites are distributed across natural protected ecosystems, but the landscapes across LA have high alpha and beta diversity (González-Caro1 et al., 2014) (Gallardo-Cruz et al., 2009), are subject to intense and rapid natural and anthropogenic disturbances (Pauchard & Barbosa, 2013), and consequently the ecological trajectories of these ecosystems are complex (Bustamante et al., 2016). Examples of challenges involved with the complexity of these landscapes include: a) the debate on the role of the Amazon basin as a carbon sink our source (Andreae et al., 2002; Avissar et al., 2002; Davidson & Artaxo, 2004); b) testing the consistency of the expected carbon use efficiency (the ratio between ecosystem respiration and GPP) for disturbed ecosystems (Baldocchi & Penuelas, 2019); or c) how old grown tropical forests respond to weather variability (Rojas-Robles et al., 2020).

4.3 Influence of spatial scale on the representativeness of GPP and ET

The FLUXNET database has been extensively used for development of terrestrial ecosystem process models and data-driven products (Baldocchi, 2020; Jung et al., 2019), but there are challenges associated with: a) the distribution of sites that provide ground truth data; and b) the mismatch between the footprint that is represented by the ground truth data and the footprint (i.e., pixel size) of the output from the models. Here, we discuss how the representativeness of GPP and ET is influenced by different spatial scales typically used for global estimations (i.e., 0.25, 0.50, 1.0 geographic degrees).

We found that as the pixel size increases (i.e., at coarser spatial resolutions) representativeness decrease to about 25% for GPP and ET across LA. Again, these results are conservative and assuming that all data from the 41 sites is available for the scientific community. One explanation about why representativeness decrease as pixel size increase is that there are study sites within close proximity that may fall within one pixel at coarser spatial resolutions. Consequently, many environmental properties are not represented by the fewer sites included within the coarser pixels; in other words, fewer pixels provide information within LA. Future studies, could focus on how the density of sites across LA influence modeling errors for GPP and ET (Papale et al., 2015) or upscaling data-driven products relevant for the regional (Guevara et al., 2018) and global carbon cycle (Jung et al., 2019; Warner et al., 2019) .

Consequently, an important issue is where to establish new study sites to improve the representativeness of EONs.

4.4 Improving Representativeness

LA incorporates one of the most ecological diverse regions in the world, having a large influence on the global carbon and water cycles, the Earth's climate systems and nutrient cycles (Balvanera et al., 2012). Our results show spatial representativeness gaps among assessed environmental properties (i.e., bioclimatic, terrain properties and soil resources) including GPP and ET (Figures 2-3) to identify under-represented regions from different IGBP classes (Supplementary Material Table S3). We highlight that larger monitoring efforts should focus on evergreen broadleaf forest, croplands, savanna and open shrublands, but including permanent wetlands and snow and ice classes are needed to cover the biome spectrum across LA (Table 4).

What would happen if 200 strategically located sites are installed across LA? Our results show that: a) representativeness would increase to 86% and 80% of GPP and ET, respectively (Figure 6a, b); and b) the correlation between represented regions and environmental properties monitored by these sites would increase (Figure 6c, d). Previous studies have highlighted the need to increase monitoring across underrepresented ecosystems within FLUXNET, such as water-limited ecosystems (Hargrove et al., 2003; Papale et al., 2015; Villarreal et al., 2018), and it is known that increasing the number of sites within a region of interest would result in lower uncertainty estimates (Papale et al., 2015; Sulkava et al., 2011; Villarreal et al., 2019). However, there is always a tradeoff between monitoring space and time to properly represent spatial and temporal variability. Our study focuses on maximizing spatial representativeness of long-term means of environmental properties, GPP and ET, but we highlight that temporal representativeness should also be considered (Chu et al., 2017). These assessments could be tested across LA as more sites with long-term eddy-covariance records are included and compared across our understanding of the temporal and spatial representativeness of the FLUXNET network.

4.5 Challenges and Opportunities

Our study is based using information from sites that were registered in FLUXNET by 2018. We highlight that many of these sites have not shared information with FLUXNET of AmeriFlux and consequently their data is not widely available for the scientific community to perform synthesis studies and data-driven products. By hence, our representativeness analysis must be taken as a conservative approach, especially for those IGBP classes with no registered eddy-covariance sites (Table 1). For this study, we assumed that if a site was registered with FLUXNET by 2018, then the eddy-covariance information either is available, or the principal investigator is willing to contribute with FLUXNET in the near future. An example are the sites located across Mexico (i.e., MexFlux) that are affiliated with FLUXNET but the data is not currently available for the wider scientific community (Villarreal et al., 2019). We are also aware that there are multiple unaffiliated sites across LA such as in semi-arid grasslands (Hinojo-Hinojo et al., 2016), croplands (Lewczuk et al., 2017), the Chacon region (Toledo & Figuerola, 2019), Andes (Callañaupa-Gutierrez et al., 2020), Patagonia (Valdés-Barrera et al., 2019), among other sites across LA. We hope that this study motivates principal investigators and regional networks (i.e., MexFlux, Brasflux, SULFLUX) to join and contribute to FLUXNET to build a stronger global network and increase our understanding of LA in the global water and carbon cycles.

Traditionally, establishment of eddy-covariance sites across LA has not been done under a national coordinated effort, and consequently, monitoring efforts have been performed by individual research groups or by local networks with clear questions focused on specific biomes (Roberti et al., 2012; Vargas et al., 2013). Furthermore, increasing the representativeness of FLUXNET across LA, and other areas of the world such as Africa, is difficult for several reasons. There are challenges related to limited economic resources (e.g., limited funding opportunities, increased costs due to importation taxes), human resources (e.g., fewer trained scientists to operate eddy-covariance sites and analyze data), and logistic issues (e.g., security, accessibility).

Our results bring attention on the possibilities that exist by coordinating optimized monitoring efforts to improve FLUXNET spatial representativeness. Although adding 200 new study sites is currently not possible, there are a few initial steps that could be done: a) a coordinated effort to archive, synthesized and analyze existing information across LA; b) collaborate on collection of new information across underrepresented sites with national, regional or partnerships outside LA; and c) promote a stronger culture of data sharing and proper recognition of the efforts of researchers across LA. We recognize that these efforts will require an increase of interoperability across existing networks and researchers within LA, and we must work as a community to reduce conceptual, technological, organizational and cultural barriers (Vargas et al., 2017). Finally, we warn about "Helicopter research" and call for quality and equality of partnerships where there should be a mutual benefit between local scientists (with local infrastructure and local knowledge) and foreign scientists performing research across LA. Although we fully advocate for data sharing and following FAIR data principles (Fisher & Koven, 2020) we strongly support that contributions from LA scientists should be recognized by the FLUXNET community especially when those efforts increase the representativeness of information for the regional-to-global water and carbon cycles.

5 Conclusions

The present study used machine learning methods to assess the representativeness of FLUXNET sites across LA as it has been applied on previous studies assessing the representativeness of EONs (Villarreal *et al* 2018; Villarreal *et al* 2019). This study provides an overall scope of FLUXNET representativeness gaps of climate, topographic, soil resources along with GPP and ET across LA, a region that plays an important role in the global dynamic of ecosystem processes and climate regulation. Throughout 41 sites affiliated to FLUXNET (revised in 2018 but see methods for details) our results show that the spatial representativeness of environmental properties such as climate, topography and soil resource is around 40% of LA, while for GPP and ET the spatial representativeness is 48% and 34%, respectively. The overall representativeness of FLUXNET could substantially increase if around 200 sites were strategically located across LA, specially within evergreen broadleaf forest, savanna and open shrublands; however, to coordinated such an effort there should be a higher degree of interoperability among scientists, research groups, and local-to-global networks. Nonetheless, the benefits from enhancing FLUXNET representativeness across LA would help to improve our knowledge on the impact of environmental change (i.e., climate change, land use / land cover changes, extreme climate events) which could support science-based environmental policies and environmental management decisions.

Acknowledgments

SV acknowledges support from a CONACyT doctoral fellowship. RV acknowledges support from NASA Carbon Monitoring System (80NSSC18K0173). We thank Trevor Keenan and the AmeriFlux community for discussing preliminarily results from this research. Original data to perform the analysis is publicly available (see methods for details), and output data will be archived at ORNL DAAC (pending solicitation).

References

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., et al. (2016). The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. *Science*, *348*(6237), 895–899. https://doi.org/10.1002/2015JA021022

Alcaraz-Segura, D., Lomba, A., Sousa-Silva, R., Nieto-Lugilde, D., Alves, P., Georges, D., et al. (2017). Potential of satellite-derived ecosystem functional attributes to anticipate species range shifts. *International Journal of Applied Earth Observation and Geoinformation*, *57*(May), 86–92. https://doi.org/10.1016/j.jag.2016.12.009

Amundson R., J. H. (1991). On a state factor model of ecosystems. *Biogeosciences*, *47*(8).

Andreae, M. O., Artaxo, P., Brandão, C., Carswell, F. E., Ciccioli, P., Da Costa, A. L., et al. (2002). Biogeochemical cycling of carbon, water, energy, trace gases, and aerosols in Amazonia: The LBA-EUSTACH experiments. *Journal of Geophysical Research D: Atmospheres*, *107*(20). https://doi.org/10.1029/2001JD000524

Avissar, R., Dias-Silva, P. L., Silva-Dias, M. A. F., & Nobre, C. (2002). The Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA): Insights and future research needs. *Journal of Geophysical Research*, *107*(D20), 8086. https://doi.org/10.1029/2002JD002704

Baldocchi, D., & Penuelas, J. (2019). The physics and ecology of mining carbon dioxide from the atmosphere by ecosystems. *Global Change Biology*, *25*(4), 1191–1197. https://doi.org/10.1111/gcb.14559

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET : A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide , Water Vapor , and Energy Flux Densities, (February).

Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biology*, (June 2019), 242–260. https://doi.org/10.1111/gcb.14807

Baldocchi, D. D., Finnigan, J., Wilson, K. B., U, K. T. P., & Falge, E. (2000). On measuring net ecosystem carbon exchange over tall vegetation on complex terrain. *Boundary-Layer Meteorology*, *96*(1–2), 257–291. https://doi.org/10.1023/A:1002497616547

Balvanera, P., Uriarte, M., Almeida-Leñero, L., Altesor, A., DeClerck, F., Gardner, T., et al. (2012). Ecosystem services research in Latin America: The state of the art. *Ecosystem Services*, *2*, 56–70. https://doi.org/10.1016/j.ecoser.2012.09.006

Barbet-Massin, M., Jiguet, F., Helene Albert, C., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution, 3*, 327–338. https://doi.org/doi: 10.1111/j.2041-210X.2011.00172.x

Biederman, J. A., Scott, R. L., Goulden, M. L., Vargas, R., Litvak, M. E., Kolb, T. E., et al. (2016). Terrestrial carbon balance in a drier world: The effects of water availability in southwestern North America. *Global Change Biology*, *22*(5), 1867–1879. https://doi.org/10.1111/gcb.13222

Bustamante, M. M. C., Roitman, I., Aide, T. M., Alencar, A., Anderson, L. O., Aragão, L., et al. (2016). Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity. *Global Change Biology*, *22*(1), 92–109. https://doi.org/10.1111/gcb.13087

Callañaupa-Gutierrez, M, S., Segura Cajachagua, H., Saavedra, M., Flores, J., Cuxart, J., & Silva, Y. (2020). Seasonal variability of evapotranspiration in the central Andes of Peru using eddy covariance techniques and empirical methods. In *EGU General Assembly Conference Abstracts* (p. 6158). AA(Instituto Geofísico del Perú, Lima, Peru). Retrieved from https://ui.adsabs.harvard.edu/abs/2020EGUGA..22.6158M

Chabbi, A., Loescher, H. W., Tye, M. R., & Hudnut, D. (2017). Integrated Experimental Research Infraestructure: A Paradigm Shift to Face an Uncertain World and Innovate for Societal Benefit. *CRC Press*.

Chapin, F. S., Vitousek, P. M., & Van Cleve, K. (1986). The Nature of Nutrient Limitation in Plant Communities. *The American Naturalist*, *127*(1), 48–58. https://doi.org/10.1086/284466

Chapin, F. S., Matson, P. a, & Mooney, H. a. (2002). *Principles of Terrestrial Ecosystem Ecology*. *System* (Vol. 21). https://doi.org/10.1007/978-1-4419-9504-9

Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., et al. (2011). Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling. *Agricultural and Forest Meteorology*, *151*(1), 87–100. https://doi.org/10.1016/j.agrformet.2010.09.005

Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Black, T. A., et al. (2012). Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program Network using remote sensing and footprint analysis. *Remote Sensing of Environment*, *124*, 742–755. https://doi.org/10.1016/j.rse.2012.06.007

Chrysoulakis, N., Spiliotopoulos, M., Domenikiotis, C., & Dalezios3, N. (2003). Towards monitoring of regional atmospheric instability through modis/aqua images, (November), 7–9.

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

Conant, R. T., Paustian, K., & Elliott, E. T. (2001). Grassland management and conversion into grassland: effect on soil carbon. *Ecological Applications*, *11*(2), 343–355. https://doi.org/10.1890/1051-0761(2001)011[0343:GMACIG]2.0.CO;2

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J.

(2007). Random Forests for Classification in Ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Davidson, E. A., & Artaxo, P. (2004). Globally significant changes in biological processes of the Amazon Basin: Results of the large-scale Biosphere-Atmosphere Experiment. *Global Change Biology*, *10*(5), 519–529. https://doi.org/10.1111/j.1529-8817.2003.00779.x

Drew, C. A., Wiersma, Y. F., & Huettmann, F. (2011). Predictive species and habitat modeling in landscape ecology: Concepts and applications. *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*, 1–313. https://doi.org/10.1007/978-1-4419-7390-0

Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeing Species Distribution and Change Using Random Forest. *In Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*, 139–159. https://doi.org/10.1007/978-1-4419-7390-0

Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., et al. (2002). Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agricultural and Forest Meteorology*, *113*(1–4), 53–74. https://doi.org/10.1016/S0168-1923(02)00102-8

Fisher, J. B., Tu, K. P., & Baldocchi, D. D. (2008). Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, *112*, 901–919. https://doi.org/10.1016/j.rse.2007.06.025

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. *Journal of Advances in Modeling Earth Systems*, *12*. https://doi.org/10.1029/2018MS001453

Gallardo-Cruz, J. A., Pérez-García, E. A., & Meave, J. A. (2009). β-Diversity and vegetation structure as influenced by slope aspect and altitude in a seasonally dry tropical landscape. *Landscape Ecology*, *24*(4), 473–482. https://doi.org/10.1007/s10980-009-9332-1

González-Caro1, S., María Natalia, U., Álvarez, E., Stevenson, P., & Swenson, N. (2014). Phylogenetic alpha and beta diversity in tropical tree assemblages along regional- scale environmental gradients in northwest South America. *Journal of Plant Ecology*, *7*(2), 145–153. https://doi.org/10.1093/jpe/rtt076

Graesser, J., Aide, T. M., Grau, H. R., & Ramankutty, N. (2015). Cropland/pastureland dynamics and the slowdown of deforestation in Latin America. *Environmental Research Letters*, *10*(3). https://doi.org/10.1088/1748-9326/10/3/034017

Guevara, M., Olmedo, G. F., Stell, E., Yigini, Y., Duarte, Y. A., Hernández, C. A., et al. (2018). No silver bullet for digital soil mapping : country-specific soil organic carbon estimates across Latin America. *Soil*, (June), 173–193.

Hargrove, W. W., Hoffman, F. M., & Law, B. E. (2003). New analysis reveals representativeness of the AmeriFlux network. *Eos, Transactions American Geophysical Union*, *84*(48), 529. https://doi.org/10.1029/2003EO480001

He, H., Zhang, L., Gao, Y., Ren, X., Zhang, L., Yu, G., & Wang, S. (2015). Regional representativeness assessment and improvement of eddy flux observations in China. *Science of the Total Environment*, *502*, 688–698. https://doi.org/10.1016/j.scitotenv.2014.09.073

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Huxman, T. E., Smith, M. D., Fay, P. A., Knapp, A. K., Shaw, M. R., Loik, M. E., et al. (2004). Convergence across biomes to a common rain-use efficiency, (April), 1–4. https://doi.org/10.1038/nature02597.1.

Jongman, R. H. G., Skidmore, A. K., (Sander) Mücher, C. A., Bunce, R. G. H., & Metzger, M. J. (2017). Global Terrestrial Ecosystem Observations: Why, Where, What and How? In M. Walters & R. J. Scholes (Eds.), *The GEO Handbook on Biodiversity Observation Networks* (pp. 19–38). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-27288-7_2

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-valls, G., et al. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, 1–14. https://doi.org/10.1038/s41597-019-0076-8

Keenan, T. F., Gray, J., Friedl, M. A., Toomey, M., Bohrer, G., Hollinger, D. Y., et al. (2014). Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change*, *4*(July). https://doi.org/10.1038/NCLIMATE2253

Keller, M., Alencar, A., Asner, G. P., Braswell, B., Bustamante, M., Davidson, E., et al. (2004). Ecological Research in the Large-Scale Biosphere– Atmosphere Experiment in Amazonia: Early Results. *Ecological Applications*, *14*(sp4), 3–16. https://doi.org/10.1890/03-6003

Keller, M., Schimel, D. S., Hargrove, W. W., & Hoffman, F. M. (2011). A continental Ecological Strategy Observatory for the National Network. *The Ecological Society of America*, *6*(5), 282–284.

Knox, S. H., Jackson, R. B., Poulter, B., McNicol, G., Etienne, F.-C., Zhang, Z., et al. (2019). Fluxnet-ch 4 synthesis activity. *American Meteorological Society*, (December). https://doi.org/10.1175/BAMS-D-18-0268.1

Kumar, J., Hoffman, F. M., Hargrove, W. W., & Collier, N. (2016). Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements. *Earth System Science Data Discussions*, (August), 1–25. https://doi.org/10.5194/essd-2016-36

Lewczuk, N. A., Posse, G., Richter, K., & Achkar, A. (2017). CO2 and N2O flux balance on soybean fields during growth and fallow periods in the Argentine Pampas—A study case. *Soil and Tillage Research*, *169*, 65–70. https://doi.org/10.1016/j.still.2017.01.017

Liu, C., White, M., & Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, *34*(2), 232–243. https://doi.org/10.1111/j.1600-0587.2010.06354.x

Lovett, G. M., Burns, D. a, Driscoll, C. T., Jenkins, J. C., Mitchell, M. J., Rustad, L., et al. (2007). Who needs environmental monitoring ? *The Ecological Society of America*, *5*(5), 253–260. https://doi.org/10.1890/1540-9295(2007)5[253:WNEM]2.0.CO;2

Löw, F., Michel, U., Dech, S., & Conrad, C. (2011). Development of a satellite-based multi-scale land use classification system for land and water management in Uzbekistan and Kazakhstan. *Proceedings SPIE 8181, Earth Resources and Environmental Remote Sensing/GIS Applications II, 81811K*, *8181*(0), 1–7. https://doi.org/doi:10.1117/12.898038

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*(1), 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, *32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., et al. (2015). Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. *Journal of Geophysical Research: Biogeoscience*, 1941–1957. https://doi.org/10.1002/2015JG002997.Received

Paruelo, J. M., Jobbagy, E. G., Sala, O. E., Lauenroth, W. K., & Burke, I. C. (1998). Functional and Structural Convergence of Temperate Grassland and Shrubland Ecosystems. *Ecological Applications*, *8*(1), 194–206.

Pauchard, A., & Barbosa, O. (2013). Regional Assessment of Latin America: Rapid Urban Development and Social Economic Inequity Threaten Biodiversity Hotspots BT. In T. Elmqvist, M. Fragkias, J. Goodness, B. Güneralp, P. J. Marcotullio, R. I. McDonald, et al. (Eds.), *Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities A Global Assessment* (pp. 589–608). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7088-1_28

Ponce Campos, G. E., Moran, M. S., Huete, A., Zhang, Y., Bresloff, C., Huxman, T. E., et al. (2013). Ecosystem resilience despite large-scale altered hydroclimatic conditions. *Nature*, *494*(7437), 349–52. https://doi.org/10.1038/nature11836

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., et al. (2014). Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, *509*(7502), 600–603. https://doi.org/10.1038/nature13376

Ramankutty, N., Foley, J. A., Norman, J., & Mcsweeney, K. (2002). The global distribution of cultivable lands : current patterns and sensitivity to possible climate change. *Global Ecology and Biogeography*, 377–392.

Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., & Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, *169*, 156–173. https://doi.org/10.1016/j.agrformet.2012.09.012

Roberti, D. R., Acevedo, O. C., & Moraes, O. L. L. (2012). A Brazilian network of carbon flux stations. *Eos*, *93*(21), 203. https://doi.org/10.1029/2012EO210005

Rojas-Robles, N. E., Garatuza-Payán, J., Álvarez-Yépiz, J. C., Sánchez-Mejía, Z. M., Vargas, R., & Yépez, E. A. (2020). Environmental Controls on Carbon and Water Fluxes in an Old-Growth Tropical Dry Forest. *Journal of Geophysical Research: Biogeosciences*, *125*(8), e2020JG005666. https://doi.org/doi:10.1029/2020JG005666

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*, 9899–9904. https://doi.org/10.1073/pnas.1019576108

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., et al. (2012). A model-data comparison of gross primary productivity : Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research: Biogeoscience*, *117*, 1–15. https://doi.org/10.1029/2012JG001960

Schmitt, S., Pouteau, R., Justeau, D., de Boissieu, F., & Birnbaum, P. (2017). ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models. *Methods in Ecology and Evolution*, *8*(12), 1795–1803. https://doi.org/10.1111/2041-210X.12841

Scholes, R. J., Gill, M. J., Costello, M. J., Sarantakos, G., & Walters, M. (2017). Working in Networks to Make Biodiversity Data More Available. In M. Walters & R. J. Scholes (Eds.), *The GEO Handbook on Biodiversity Observation Networks* (pp. 1–17). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-27288-7_1

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., et al. (2017). Global patterns of drought recovery. *Nature*, *548*(7666), 202–205. https://doi.org/10.1038/nature23021

Sulkava, M., Luyssaert, S., Zaehle, S., & Papale, D. (2011). Assessing and improving the representativeness of monitoring networks: The European flux tower network example. *Journal of Geophysical Research: Biogeosciences*, *116*(2), 1–14. https://doi.org/10.1029/2010JG001562

Toledo, E. L., & Figuerola, P. (2019). Conductores ambientales de la productividad de buffelgrass en el Chaco Árido de La Rioja (Argentina). *RADA*, (November).

Valdés-Barrera, A., Kutzbach, L., Celis-Diez, J. L., Armesto, J. J., Holl, D., & Perez-Quezada, J. F. (2019). Effects of disturbance on the carbon dioxide balance of an anthropogenic peatland in northern Patagonia. *Wetlands Ecology and Management*, *27*(5), 635–650. https://doi.org/10.1007/s11273-019-09682-3

Vargas, R., Sonnentag, O., Abramowitz, G., Carrara, A., Chen, J. M., Ciais, P., et al. (2013). Drought Influences the Accuracy of Simulated Ecosystem Fluxes: A Model-Data Meta-analysis for Mediterranean Oak Woodlands. *Ecosystems*, *16*, 749–764. https://doi.org/10.1007/s10021-013-9648-1

Vargas, R., Yépez, E. a., Andrade, J. L., Ángeles, G., Arredondo, T., Castellanos, a. E., et al. (2013). Progress and opportunities for monitoring greenhouse gases fluxes in Mexican ecosystems: The MexFlux network. *Atmosfera*, *26*(3), 325–336.

Vargas, R., Alcaraz-Segura, D., Birdsey, R., Brunsell, N. A., Cruz-Gaistardo, C. O., de Jong, B., et al. (2017). Enhancing interoperability to facilitate implementation of REDD+: case study of Mexico. *Carbon Management*, *8*(1), 57–65. https://doi.org/10.1080/17583004.2017.1285177

Villarreal, S., Vargas, R., Yepez, E. A., Acosta, J. S., Castro, A., Escoto-rodriguez, M., et al. (2016). Contrasting precipitation patterns influence complexity in evapotranspiration dynamics in water-limited shrublands. *Journal of Geophysical Research: Biogeoscience*, 1–15. https://doi.org/10.1002/2015JG003169.Received

Villarreal, S., Guevara, M., Alcaraz-Segura, D., Brunsell, N. A., Hayes, D., Loescher, H. W., & Vargas, R. (2018). Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States. *Agricultural and Forest Meteorology*, *262*(August), 423–433. https://doi.org/10.1016/j.agrformet.2018.07.016

Villarreal, S., Guevara, M., Alcaraz-segura, D., & Vargas, R. (2019). Optimizing an Environmental Observatory Network Design Using Publicly Optimizing an Environmental Observatory Network Design Using Publicly Available Data. *Journal of Geophysical Research: Biogeosciences*, (May). https://doi.org/10.1029/2018JG004714

Vitousek, P. M. (1984). Litterfall , Nutrient Cycling , and Nutrient Limitation in Tropical Forests Author. *Ecology*, *65*(1), 285–298.

Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E., & Vargas, R. (2019). Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale. *Global Biogeochemical Cycles*, *33*(12), 1733–1745. https://doi.org/10.1029/2019GB006264

Whiting, G. J., & Chanton, J. P. (2001). Greenhouse carbon balance of wetlands: Methane emission versus carbon sequestration. *Tellus, Series B: Chemical and Physical Meteorology*, *53*(5), 521–528. https://doi.org/10.1034/j.1600-0889.2001.530501.x

Yang, F., Zhu, a. X., Ichii, K., White, M. a., Hashimoto, H., & Nemani, R. R. (2008). Assessing the representativeness of the AmeriFlux network using MODIS and GOES data. *Journal of Geophysical Research: Biogeosciences*, *113*(4), 1–11.

https://doi.org/10.1029/2007JG000627

Zhang, L., Xiao, J., & Zhou, Y. (2020). Increased carbon uptake and water use efficiency in global semi-arid ecosystems Increased carbon uptake and water use efficiency in global semi-arid ecosystems. *Environmental Research Letters*, (February). https://doi.org/10.1088/1748-9326/ab68ec

Figures

Figure 1. Spatial distribution of different biomes across Latin America according to the International Geosphere-Biosphere Program (IGBP) and the location of the eddy-covariance sites affiliated with FLUXNET. See supplementary material table S2 for details about FLUXNET sites.

Figure 2. Contrast between the dynamic of the regions represented and non-represented by FLUXNET sites for the two most influential variables for each environmental parameter assessed; Bioclimatic (a-d), terrain parameters (e-f), soil properties (g-j), and the combined environmental factors into a principal component analysis (k-n).

Figure 3. Contrast between the dynamic of the regions represented and non-represented by FLUXNET sites for gross primary and evapotranspiration productivity their mean and coefficient of variation; (GPP; a-d) and (ET; e-h).

Figure 4. Spatial representativeness of FLUXNET sites for the different environmental parameters based on random forest models (a, c, e and g), along with the multivariate space represented by the first two principal components (b, d, f and h).

Figure 5. Spatial representativeness of FLUXNET sites for the different environmental parameters based on random forest models (a, c, e and g), along with the multivariate space represented by the first two principal components (b, d, f and h).

Figure 6. Increment of representativeness according with the addition of strategic location for potential new study-sites based on cHLS technique for GPP (a) and ET (b). Also, there is an increment in representativeness validation parameter (area under the curve; AUC) as more sites are added (c-d).

Figure 1

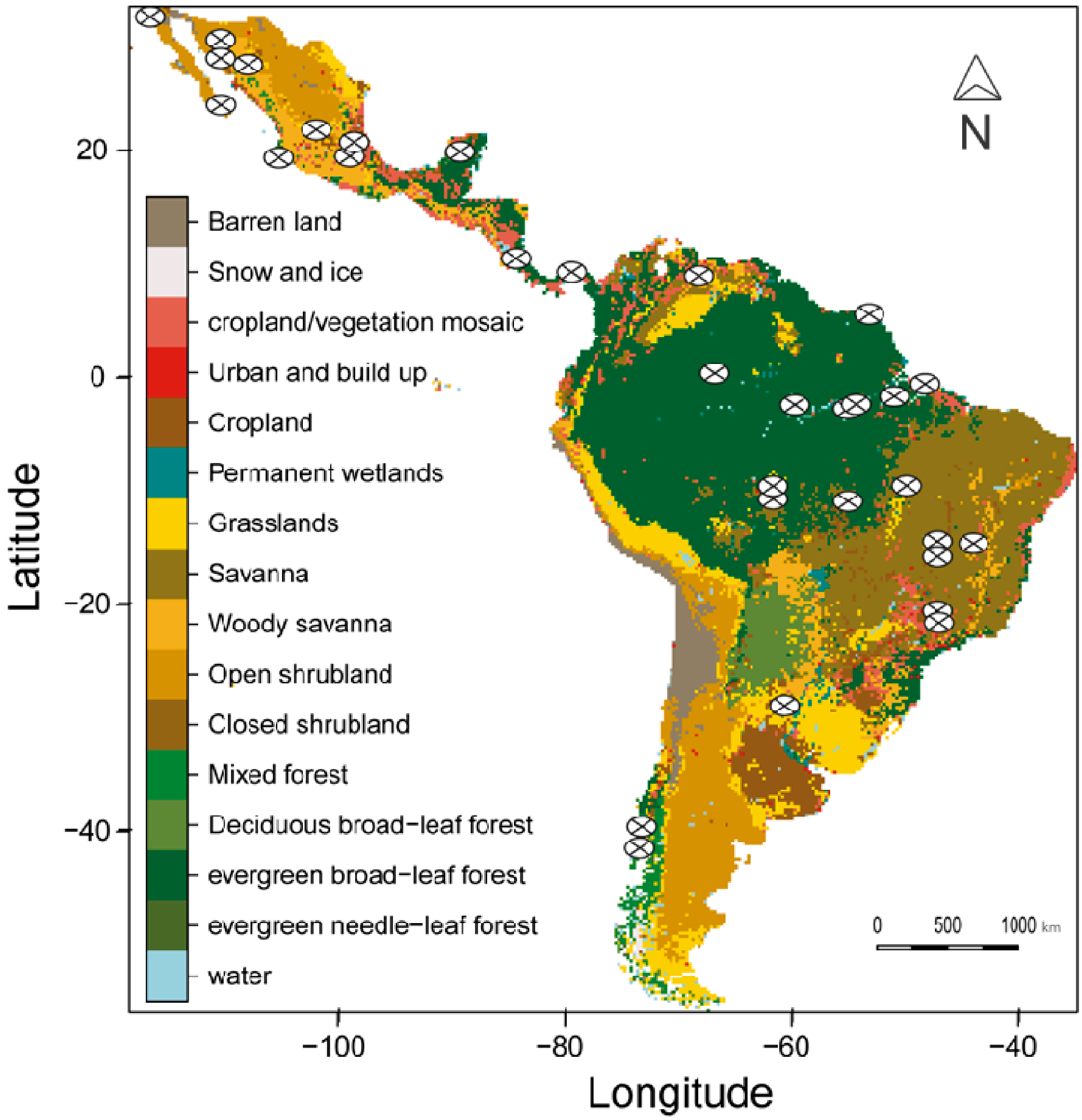
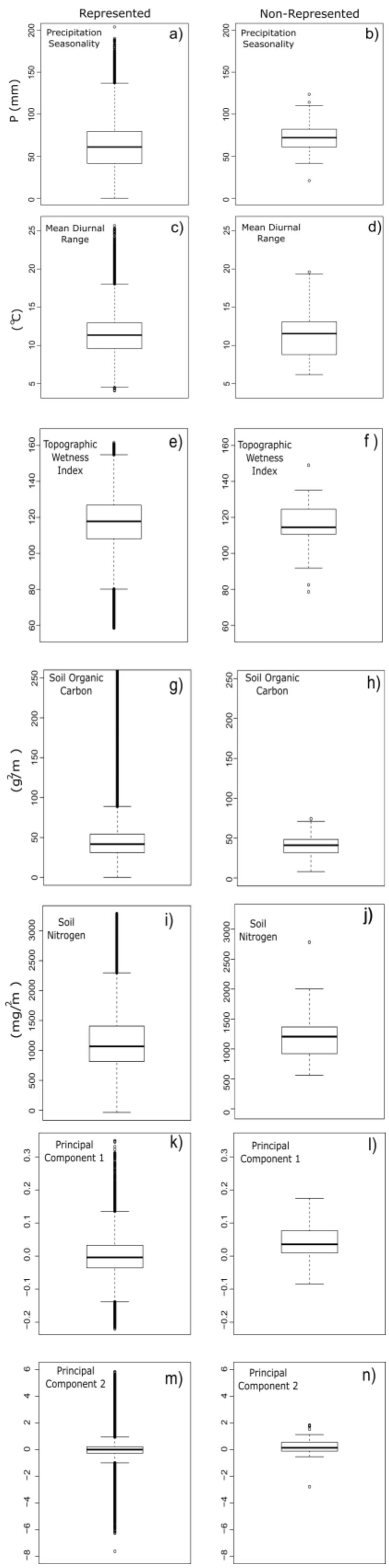
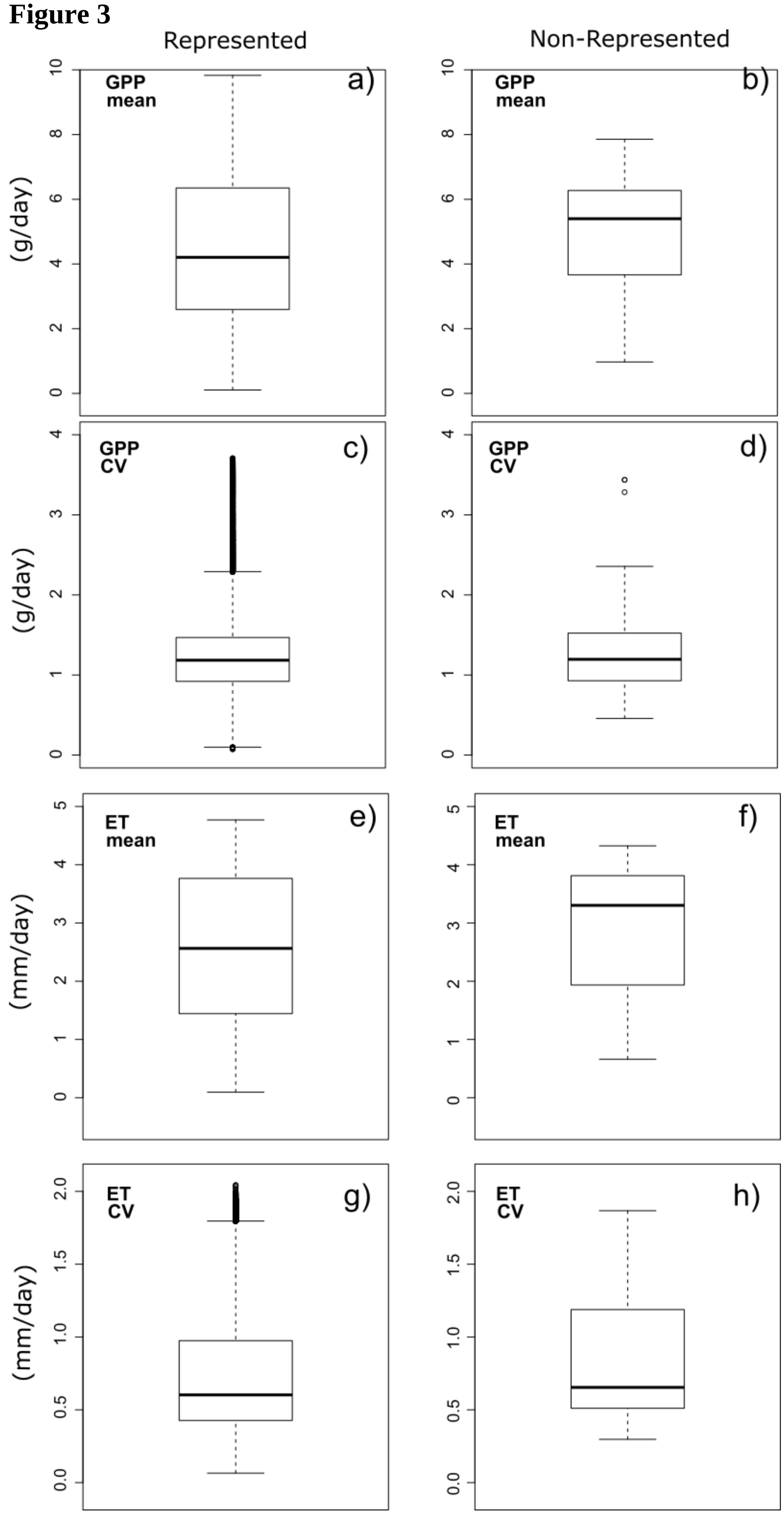

20
0
−20
−40
Latitude
−100
−80
−60
−40
Longitude
N
0
500
1000 km
Barren land
Snow and ice
cropland/vegetation mosaic
Urban and build up
Cropland
Permanent wetlands
Grasslands
Savanna
Woody savanna
Open shrubland
Closed shrubland
Mixed forest
Deciduous broad−leaf forest
evergreen broad−leaf forest
evergreen needle−leaf forest
water

Figure 2


Represented
Non-Represented
Precipitation Seasonality
a)
b)
P (mm)
Mean Diurnal Range
c)
d)
(°C)
Topographic Wetness Index
e)
f)
Soil Organic Carbon
g)
h)
(g/m²)
Soil Nitrogen
i)
j)
(mg/m²)
Principal Component 1
k)
l)
Principal Component 2
m)
n)

857 **Figure 3**



858

Figure 4

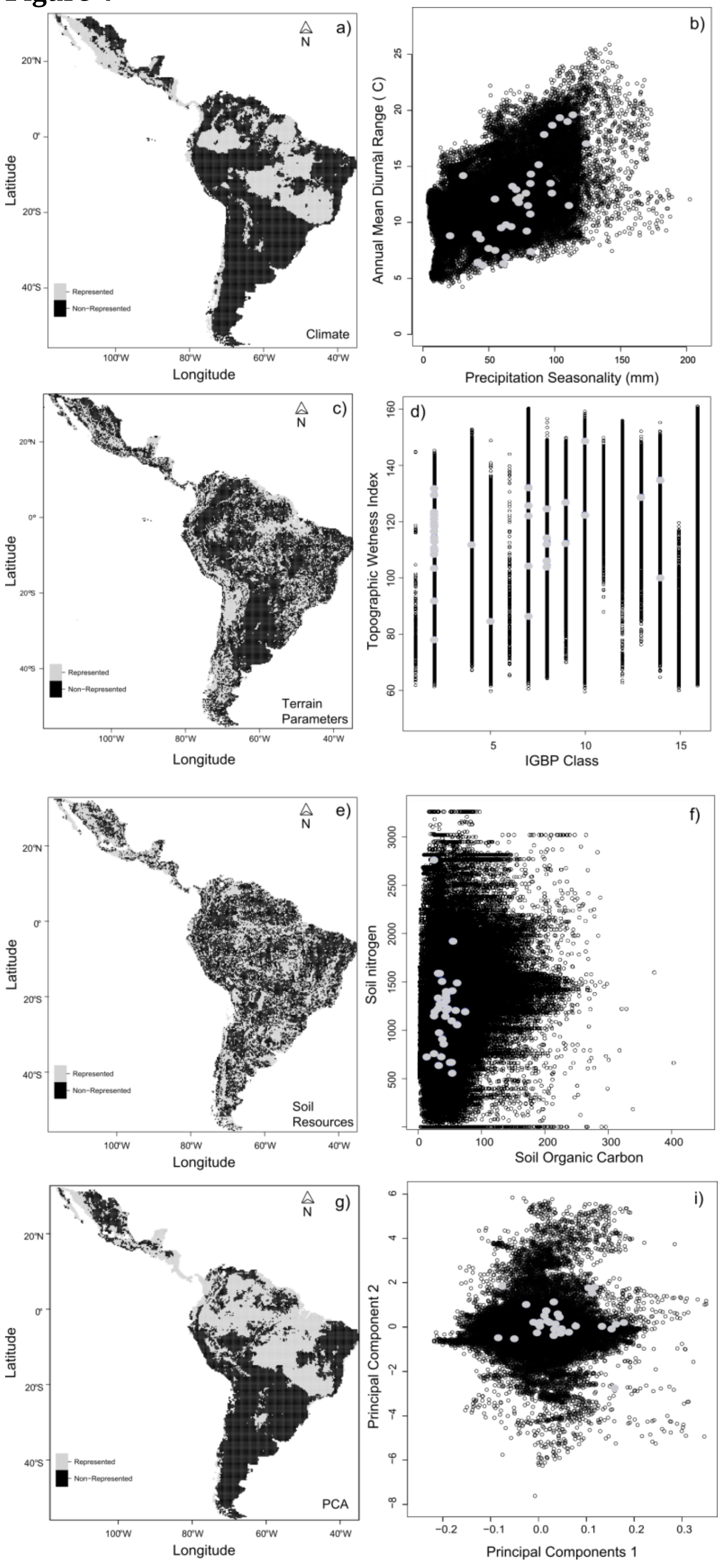

a)
N
Climate
Represented
Non−Represented
Latitude
Longitude
b)
Annual Mean Diurnal Range (C)
Precipitation Seasonality (mm)
c)
Terrain
Parameters
d)
Topographic Wetness Index
IGBP Class
e)
Soil
Resources
f)
Soil nitrogen
Soil Organic Carbon
g)
PCA
i)
Principal Component 2
Principal Components 1

Figure 5

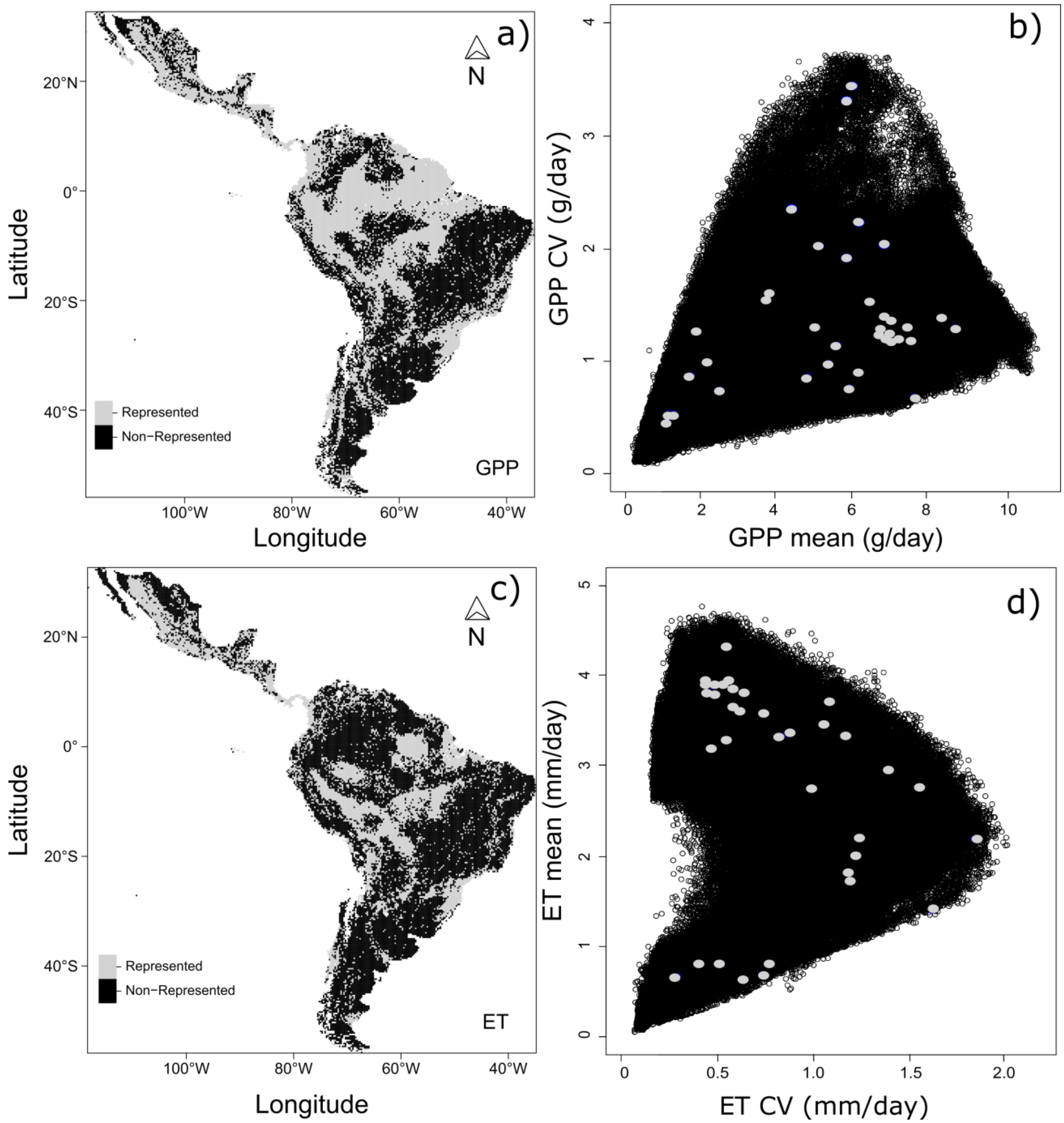
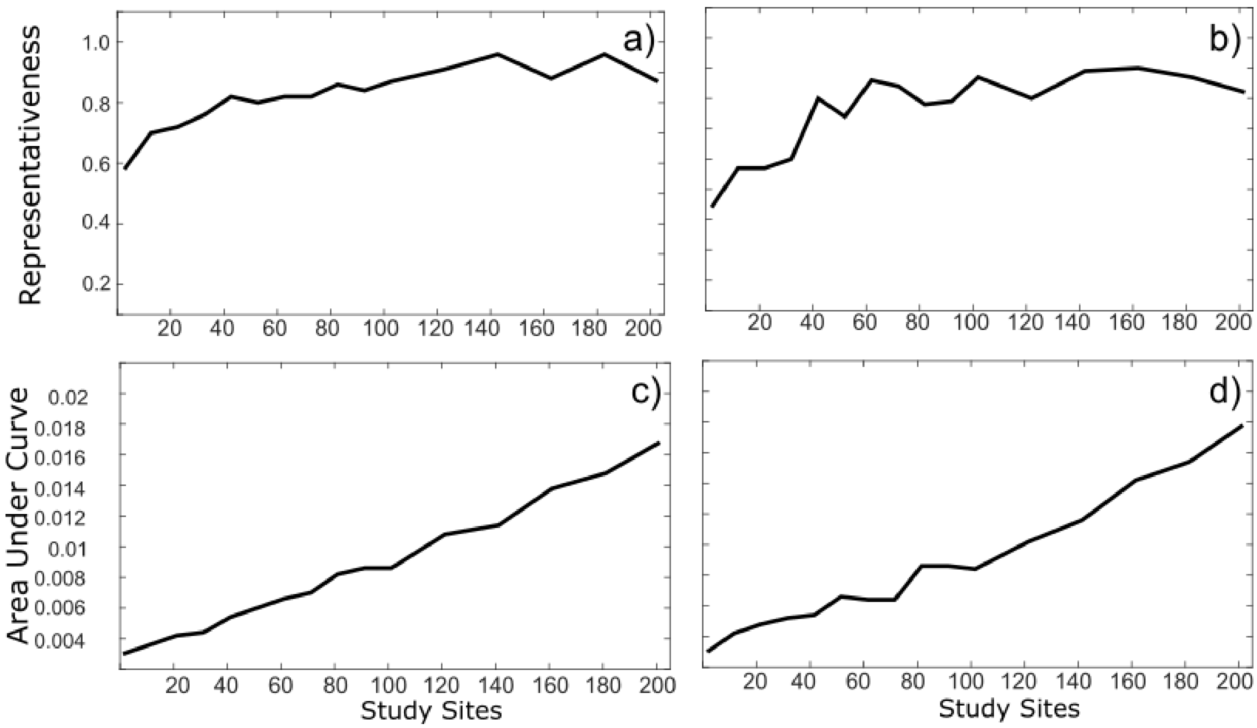

a)
N
20°N
0°
20°S
40°S
Latitude
Represented
Non−Represented
GPP
100°W
80°W
60°W
40°W
Longitude
b)
GPP CV (g/day)
GPP mean (g/day)
c)
ET
d)
ET mean (mm/day)
ET CV (mm/day)

Figure 6


a)
b)
c)
d)
Representativeness
1.0
0.8
0.6
0.4
0.2
Area Under Curve
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
20 40 60 80 100 120 140 160 180 200
Study Sites

Tables

Table 1. Surface area of each of the International Geosphere-Biosphere Project (IGBP) biome class across Latin America and the number of eddy-covariance sites at each class.

Table 2. Parameters used to characterize each model, their performance and the overall representativeness across Latin America.

Table 3. Comparison of the 0.95 mean confidence interval among each environmental variable assessed and the represented and non-represented areas.

Table 4. Number of potential-study sites based on cLHS for each International Geosphere-Biosphere Project biome class.

Table 1

IGBP Class	Area covered (%)	Number of eddy-covariance sites
Water	1.08	0
Evergreen needle-leaf forest	0.05	0
Evergreen broad-leaf forest	33.83	19
Deciduous needle-leaf forest	0.00	0
Deciduous broad-leaf forest	3.00	2
Mixed Forest	1.52	0
Closed Shrubland	0.04	1
Open Shrubland	12.64	5
Woody Savanna	6.00	8
Savanna	18.62	1
Grassland	9.37	0
Permanent Wetland	0.88	0
Cropland	4.96	1
Urban and Build-Up	0.32	1
Cropland/natural vegetation mosaic	4.65	3
Snow and Ice	0.21	0
Barren Land	2.83	0
Total	**100.00**	**41**

Table 2

Environmental Model	**Absence**	**Repetition**	**Model Performance (True Skill Statistic)**	**Threshold Binary Map**	**Representativeness Percent**	**SD**
Climate	100	5	0.49	0.35	34	0.47
Topography	100	3	0.24	0.33	36	0.48
Soil Resources	1000	7	0.22	0.02	34	0.47
Env. Prop.	100	5	0.51	0.31	45	0.50
GPP	10000	7	0.17	>0.01	48	0.50
ET	10000	7	0.08	0.05	34	0.47

Table 3

Environmental Variable	Latin America			Represented			Non-Represented		
	L.I.	H.I.	Mean	L.I.	H.I.	Mean	L.I.	H.I.	Mean
Precipitation Seasonality	60.85	61.28	61.06	69.71	70.41	70.06	56.18	56.71	56.44
Mean Temperature Range	11.58	11.63	11.60	11.55	11.60	11.57	11.63	11.71	11.67
Topographic Wetness Index (TWI)	115.70	115.85	115.78	109.23	109.47	109.35	119.28	119.45	119.37
Soil Organic Carbon	45.02	45.04	45.03	45.15	45.18	45.17	45.08	45.11	45.09
Soil Nitrogen	1109	1109	1109	1145	1145	1145	1085	1085	1085
PC1	-8.78	8.78	0.00	5.82	6.03	5.92	-4.41	-4.18	-4.30
PC2	-5.80	5.80	0.00	1.90	2.05	1.97	-1.51	-1.34	-1.43
GPP_mean	4.32	4.33	4.33	4.42	4.44	4.43	4.20	4.22	4.21
GPP_CV	1.20	1.20	1.20	1.23	1.24	1.24	1.16	1.17	1.16
ET_mean	2.51	2.51	2.51	2.59	2.60	2.59	2.42	2.43	2.42
ET_CV	0.71	0.72	0.71	0.75	0.75	0.75	0.67	0.68	0.68

Table 4

IGBP Class	Registered Sites	Including new Sites
Water	0	0
Evergreen needle-leaf forest	0	0
Evergreen broad-leaf forest	19	80
Deciduous needle-leaf forest	0	0
Deciduous broad-leaf forest	2	6
Mixed Forest	0	4
Closed Shrubland	1	1
Open Shrubland	5	30
Woody Savanna	8	15
Savanna	1	54
Grassland	0	19
Permanent Wetland	0	1
Cropland	1	11
Urban and Build-Up	1	2
Cropland/natural vegetation mosaic	3	9
Snow and Ice	0	1
Barren Land	0	6