NOVEL APPROACH TO ASSESS THE REPRESENTATIVENESS AND METHODOLOGIES TO DESIGN AN ENVIRONMENTAL OBSERVATORY NETWORK

by

Samuel Villarreal

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Water Science and Policy

Spring 2019

NOVEL APPROACH TO ASSESS THE REPRESENTATIVENESS AND METHODOLOGIES TO DESIGN AN ENVIRONMENTAL OBSERVATORY NETWORK

by

Author

Approved: ____

Erik H. Ervin, Ph.D.
Chair of the Department of Plant and Soil Sciences

Approved: ____

Mark Rieger, Ph.D.
Dean of the College of Agriculture and Natural Resources

Approved: ____

Douglas J. Doren, Ph.D.
Interim Vice Provost for the Office of Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Rodrigo Vargas, Ph.D.
Professor in charge of dissertation (Adviser)

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Daniel L. Leathers, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Sara Rauscher, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Domingo Alcaraz-Segura, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

This work would not have been possible without financial support from the Consejo Nacional de Ciencias y Tecnologia (CONACYT), Mexico. Support from the United States Department of Agriculture (2014-67003-22070), support from the National Science Foundation (EF-102980 and 1655294), and support from NASA Carbon Monitoring Systems (80NSSC18K0173).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Environmental monitoring, especially long-term monitoring programs are a backbone component for environmental science and policy. Where environmental observatory networks (EONs) are entities that coordinates environmental monitoring efforts and provides useful information that helps to develop knowledge at a regional to global scale. However, due to intrinsic environmental variability and EONs organizational structure, the ability of EONs to properly represents environmental dynamic is prompt to poorly represents certain regions or ecosystems. This dissertation focuses on developed a different approach to assess EONs representativeness and design, by using a time-varying land-cover surface classification that characterize ecosystem functional heterogeneity based on carbon uptake dynamics (i.e., ecosystem functional types; EFTs), and by using machine learning techniques (i.e., maxent, random forest) to assess EONs representativeness. This study is divided into three main objectives; A) assess the representativeness of AmeriFlux and the National Ecological Observatory Network (NEON) to monitor the spatial and temporal variability of EFTs across the conterminous Unites States; B) propose a flexible framework to optimize the design of an EON using a publicly available data in a high-diverse country (i.e., Mexico; and C) Assess the representativeness of ecosystem states factors (i.e., climate, topography and soil resources) along with ecosystem processes (i.e., gross primary productivity and evapotranspiration) of FLUXNET eddy-covariance sites in Latin America. Results indicate that this dissertation provides valuable information for EONs management as

identifies spatial information gaps and could guide an optimal EONs design. Also, is based on a reproducible framework using publicly available information and it could be applied anywhere in the world.

Chapter 1

INTRODUCTION

1.1 Environmental Observatory Networks

The development of human civilization has been related to an escalated alteration and transformation of Earth and its natural resources, this transformation has reached a peak that nowadays human activities are one of the most dominant forces shaping the Earth surface (Steffen *et al* 2007, Lewis and Maslin 2015). This human-induced transformation is characterized by changes in atmosphere chemistry, land-cover change, earth's global temperature, among others. In order to account for the impact and consequences of such transformation, environmental monitoring provides information of physical, chemical, and/or biological variables which are designed to answer clear and specific questions on global change relate it issues (Lovett *et al* 2007, Scholes *et al* 2017). By hence, environmental monitoring is a backbone component for environmental science and for environmental policy (Ciais *et al* 2014, Vaughan *et al* 2001).

The complexity of global change makes environmental monitoring a challenging task due to the broad spectrum of its impact and scope (Running et al 1999, Turner *et al* 1990) F). Arguably, a successful strategy comes from collaborative efforts which can be organized under environmental observatory networks (EON's). Briefly, EONs can be defined as a relatively loose affiliation of organizations that agree to create value by collaborating towards a common purpose while retaining their individual mandates, resources and management (Scholes 2017, 2012). The benefits that EON's provides are the collection and dissemination of environmental data along with efforts towards the standardization of protocols, data sharing and synthesis activities. Furthermore, EONs have provided value added products that include databases, maps, conceptual models, software/analytical tools for ecological modeling, and virtual communities of practices. These products have been useful for the scientific community and policy maker to assess knowledge gaps and expands the frontiers of ecological understanding (Novick *et al* 2017, Kampe *et al* 2010a, Villarreal *et al* 2018).

EON's are usually organized under a "bottom-up" or a "top-down" approach, both having their own advantage and disadvantages. For example, a bottom-up structure provides a more devolved framework which allows a broader environmental scope as different organizations retain their individual objectives, resources and management while sharing a common purpose or goal (Scholes *et al* 2017, Novick *et al* 2017), however, this approach is also prone to make redundant observations while under-representing other regions. A "top-down" approach is a more centralized coordinated effort having well defined and clear specific goals, which allows a better optimization of resources, but its scope is usually narrower compare to the bottom-up approach (Scholes *et al* 2017, Kampe *et al* 2010b). A common example of bottom-up approach are AmeriFlux and FLUXNET. AmeriFlux is a network of PI-managed sites measuring carbon, water, and energy fluxes representing major climate and ecological biomes fluxes across the Americas (Novick *et al* 2017). FLUXNET is a global network that clustered/gathered regional networks with the purpose to compile, archive and distribute data from the major climate and ecological biomes across the world to the scientific community (Baldocchi et al 2001, Williams *et al* 2009). The National Ecological Observatory Network (NEON) is organized under a "top-down" approach, designed for discovering, understanding and forecasting of ecosystem processes at a continental scale (Kampe *et al* 2010b, Keller *et al* 2011). NEON study-sites are distributed across 20 sampling domains which were delineated based on ecoclimatic variables spatially clustered to grouped the same fraction of the total ecoclimatic variance with the purpose to better represents the main ecoclimatic characteristics of the United States (Schimel et al 2007).

The aforementioned EONs have monitored the exchange of matter (e.g., CO2, H20, CH4) and energy (e.g., surface energy balance) between terrestrial ecosystems and the atmosphere using the eddy covariance (EC) technique, this method has the potential to monitor how ecosystems respond to a wide spectrum of different climate regimes if EC study-sites are deployed under a coordinated network of sites (Baldocchi 2003, Baldocchi et al 2001). For example, AmeriFlux through a network of more than 260 EC study-sites across the Americas provides information of ecosystem carbon, water, and energy fluxes across a wide range of climate regimes and its data collection making this network an exceptional tool to assess the ecosystem response to slowly evolving changes in climate and land cover, along with extreme or rare events such as droughts, floods, wildfires, among others (Novick., 2017).

1.2 Representativeness of environmental observatory networks

Representativeness studies are of prime importance in order to increase EONs utility by providing information to discern when, where and at what frequency EONs have monitoring or should monitor ecological processes, along with information to determine whether to maintain/remove current study-sites (Villarreal et al 2018). By hence, representativeness assessments are of prime importance for a proper design and management of EON's. Traditionally, these studies mainly used information of climate and plant functional type, which have been used to determine that the number of study sites to properly monitored a certain ecosystem process at specific scale (He et al 2015, Chen et al 2011), suggest new arrangements of study sites (Sulkava et al 2011), and identified what type of ecosystems are under/over represented (Hargrove et al 2003, Kumar et al 2016). However, due to the utility and relevance of EONs there is a pressing need to design different scientific approaches to assess the representativeness of EONs for current and near-future applications (Lovett et al 2007, Jongman et al 2017).

The novelty of this study is the addition of functional properties at ecosystem scale to assess the representativeness of EONs (Alcaraz et al 2006, Villarreal et al 2018), since recent studies discuss that ecosystem properties related to carbon and water fluxes are insufficiently explain by climate controls and classical plant functional types (Petrakis et al 2017, Reichstein et al 2014, Violle et al 2014). We used ecosystem functional types (EFTs) to characterize the amount and timing of carbon exchange between the ecosystem and the atmosphere (Alcaraz-Segura et al 2013, Alcaraz et al 2006, Alcaraz-Segura et al 2017). The EFT concept is analogous to Plant Functional Type (PFT) concept but defined at a higher level of biological organization. As species can be grouped into plant functional types based on common species traits, ecosystems can be grouped into ecosystem functional types based on their similar ecosystem functioning. In practice, EFT is a time-varying land surface classification based on remote sensing vegetation indexes (i.e., MODIS-EVI) that are used to represent the spatial patterns and temporal variability of key ecosystem functional traits (i.e., productivity, seasonality and phenology) without prior knowledge of vegetation type or canopy architecture (Alcaraz-Segura et al., 2017; 2013; Cabello et al., 2013). Therefore, the ecosystem functional characterization obtained with EFTs can infer information on vegetation structure and composition (e.g., canopy architecture, vegetation type, PFT) because they constitute complementary dimensions of biodiversity complexity (Noss 1990, Pettorelli et al 2016), and it can be integrated into representativeness studies of EONs (Villarreal et al 2018).

In this study we assess the representativeness of EON's based on concepts derived from species distributions models (SDM). The general idea of SDM is to 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 to the presence of a certain specie (Drew et al 2011, Evans et al 2011). We propose that this concept can be applied to assess EON's representativeness, since the goal is to delineate the spatial distribution of environmental properties across a geographic space that should be similar to the environmental range monitored by corresponding EON's study sites (Villarreal et al 2018).

1.3 Overview of research

The research presented in this dissertation proposed different methodologies to assess the representativeness of EON's and to suggest an analytical framework to optimally design an EON's. Chapter 2 presents the assessment of AmeriFlux and NEON to represents the spatial and temporal characteristics of EFT's across the conterminous united states (CONUS) during the years 2001-2014. Chapter 3 presents a proposed analytical framework to optimally design an EON that monitors GPP and ET across a large biodiverse and heterogeneous country like Mexico. Chapter 4 presents the assessment of FLUXNET representativeness across Latin America (LA) to monitor environmental properties (i.e., bioclimatic predictors, terrain properties and soil resources) along with ecosystem processes (i.e, GPP and ET), and suggesting the addition of potential study-sites to increase the network representativeness.

Chapter 2 (published 2018 in Agricultural and Forest Meteorology)

Presents a study performed across CONUS to assess the spatial and temporal representativeness of EFTs by AmeriFlux and NEON, and their combined core sites (i.e., sites with long-term support). This study investigates: a) what are the different EFTs categories represented by each network, b) What is the EFT inter-annual variability (EFTint; number of unique EFTs per pixel during 2001-2014) representativeness by each network and their combined core sites, and c) What is the spatial representativeness of EFT categories and EFTint based on a maximum entropy approach (i.e., spatial functional heterogeneity) by each network and their combined core sites.

Chapter 3 (In review for publication in Journal of Geophysical Research Biogeoscience).

Here, we test a flexible framework to optimize the design of an environmental observatory network (EON) using publicly available data for Mexico. We address three pervasive challenges for designing EONs: 1) How to classify ecological heterogeneity to determine multi-scale sampling domains; 2) How to set geographic priorities to maximize the representativeness of new study sites; and 3) How to assess the representativeness of new study sites. We used unsupervised classification methods (i.e., factorial and cluster analysis) to spatially delineate ecologically similar sampling domains. Then, we identified the most representative sites within each domain using a conditioned Latin Hypercube-based sampling strategy. Finally, we demonstrated the applicability of this approach by assessing the spatial representativeness of the eddy covariance network in Mexico (i.e., MexFlux).

Chapter 4 (In preparation to submit it in Environmental Research Letters).

The representativeness of FLUXNET sites across LA was assessed in order to address: a) What is the representativeness of FLUXNET-LA study-sites to monitor environmental properties such as bioclimate predictors, terrain properties and soil resources; b) What is the representativeness of FLUXNET-LA study- sites to monitor GPP and ET?; and c) Where new study-sites should be located in order to increase FLUXNET-LA representativeness?. These representativeness analyses were performed based on concepts derived from species distribution models (SDM), since the goal was to delineate the spatial distribution of environmental properties across a geographic space that should be similar to the environmental range monitored by corresponding FLUXNET sites. The proposed representativeness framework is based on publicly available information and open source software and it can be applied to any other region across the world.

Taken these three studies together they provide a more comprehensive picture on EON's representativeness and design addressing some of their most pervasive issues. The running theme of this dissertation is the addition of functional properties on EON's studies, since they provide knowledge of the temporal spatial patterns of ecosystem functioning at the regional scale and provides a proper background to assess the effects of environmental changes on ecosystems processes (Vitousek et al., 1997. Gitay & Noble 1997. Alcaraz et al., 2006). Along with the use of species distribution models (SDMs) to assess the representativeness of EONs since they provide a quantitative assessment on the spatial environmental range potentially monitored by EONs (Elith and Graham 2009, Elith and Leathwick 2009).

BIBLIOGRAPHY

Alcaraz-Segura D, Lomba A, Sousa-Silva R, Nieto-Lugilde D, Alves P, Georges D, Vicente J R and Honrado J P 2017 Potential of satellite-derived ecosystem functional attributes to anticipate species range shifts *Int. J. Appl. Earth Obs. Geoinf.* **57** 86–92 Online: http://dx.doi.org/10.1016/j.jag.2016.12.009

Alcaraz-Segura D, Paruelo J M, Epstein H E and Cabello J 2013 Environmental and human controls of ecosystem functional diversity in temperate South America *Remote Sens.* **5** 127–54

Alcaraz D, Paruelo J and Cabello J 2006 Identification of current ecosystem functional types in the Iberian Peninsula *Glob. Ecol. Biogeogr.* **15** 200–12 Online: http://doi.wiley.com/10.1111/j.1466-822X.2006.00215.x

Baldocchi, D., Falge, E., Gu, L., & Olson R 2001 FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities *Bull. Am. Meteorol. Soc.* **82** 2415

Baldocchi D D 2003 Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, Present and Future *Glob. Chang. Biol.* **9** 1–14

Chen B, Coops N C, Fu D, Margolis H A, Amiro B D, Barr A G, Black T A, Arain M A, Bourque C P A, Flanagan L B, Lafleur P M, McCaughey J H and Wofsy S C 2011 Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling *Agric. For. Meteorol.* **151** 87–100 Online: http://dx.doi.org/10.1016/j.agrformet.2010.09.005

Ciais P, Dolman A J, Bombelli A, Duren R, Peregon A, Rayner P J, Miller C, Gobron N, Kinderman G, Marland G, Gruber N, Chevallier F, Andres R J, Balsamo G, Bopp L, Bréon F M, Broquet G, Dargaville R, Battin T J, Borges A, Bovensmann H, Buchwitz M, Butler J, Canadell J G, Cook R B, Defries R, Engelen R, Gurney K R, Heinze C, Heimann M, Held A, Henry M, Law B, Luyssaert S, Miller J, Moriyama T, Moulin C, Myneni R B, Nussli C, Obersteiner M, Ojima D, Pan Y, Paris J D, Piao S L, Poulter B, Plummer S, Quegan S, Raymond P, Reichstein M, Rivier L, Sabine C, Schimel D, Tarasova O, Valentini R, Wang R, Van Der Werf G, Wickland D, Williams M and Zehner C 2014 Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system *Biogeosciences* **11** 3547–602

Drew C A, Wiersma Y F and Huettmann F 2011 Predictive species and habitat modeling in landscape ecology: Concepts and applications *Predict. Species Habitat Model. Landsc. Ecol. Concepts Appl.* 1–313

Elith J and Graham C H 2009 Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models *Ecography (Cop.).* **32** 66–77

Elith J and Leathwick J R 2009 Species Distribution Models: Ecological Explanation and Prediction Across Space and Time *Annu. Rev. Ecol. Evol. Syst.* **40** 677–97 Online: http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.110308.120159

Evans J S, Murphy M A, Holden Z A and Cushman S A 2011 Modeing Species Distribution and Change Using Random Forest *Predict. Species Habitat Model. Landsc. Ecol. Concepts Appl.* 139–59

Hargrove W W, Hoffman F M and Law B E 2003 New analysis reveals representativeness of the AmeriFlux network *Eos, Trans. Am. Geophys. Union* **84** 529

He H, Zhang L, Gao Y, Ren X, Zhang L, Yu G and Wang S 2015 Regional representativeness assessment and improvement of eddy flux observations in China *Sci. Total Environ.* **502** 688–98 Online: http://dx.doi.org/10.1016/j.scitotenv.2014.09.073

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

Kampe T U, Johnson B R, Keller M, Kampe T U, Johnson B R and Kuester M 2010a NEON : the first continental- scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure canopy biochemistry and structure *J. Appl. Remote Sens.* **4:043510**

Kampe T U, Johnson B R, Kuester M and Keller M 2010b NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure *J. Appl. Remote Sens.* **4** 043510

Keller M, Schimel D S, Hargrove W W and Hoffman F M 2011 A continental Ecological Strategy Observatory for the National Network *Ecol. Soc. Am.* **6** 282–4

Kumar J, Hoffman F M, Hargrove W W and Collier N 2016 Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements *Earth Syst. Sci. Data Discuss.* 1–25 Online: http://www.earth-syst-sci-data-discuss.net/essd-2016-36/

Lewis S L and Maslin M A 2015 Defining the Anthropocene *Nature* **519** 171–80 Online: http://dx.doi.org/10.1038/nature14258

Lovett G M, Burns D a, Driscoll C T, Jenkins J C, Mitchell M J, Rustad L, Shanley J B, Likens G E and Haeuber R 2007 Who needs environmental monitoring ? *Ecol. Soc. Am.* **5** 253–60

Noss R 1990 Indicators for monitoring biodiversity: A hierarchical approach *Conserv. Biol.* **4** 355–64 Online: http://doi.wiley.com/10.1111/j.1523-1739.1990.tb00309.x

Novick K A, Biederman J A, Desai A R, Litvak M E, Moore D J P, Scott R L and Torn M S 2017 The AmeriFlux network: A coalition of the willing *Agric. For. Meteorol.*

Petrakis S, Barba J, Bond-Lamberty B and Vargas R 2017 Using greenhouse gas fluxes to define soil functional types *Plant Soil* 868–74

Pettorelli N, Wegmann M, Skidmore A, Mücher S, Dawson T P, Fernandez M, Lucas R, Schaepman M E, Wang T, O'Connor B, Jongman R H G, Kempeneers P, Sonnenschein R, Leidner A K, Böhm M, He K S, Nagendra H, Dubois G, Fatoyinbo T, Hansen M C, Paganini M, de Klerk H M, Asner G P, Kerr J T, Estes A B, Schmeller D S, Heiden U, Rocchini D, Pereira H M, Turak E, Fernandez N, Lausch A, Cho M A, Alcaraz-Segura D, McGeoch M A, Turner W, Mueller A, St-Louis V, Penner J, Vihervaara P, Belward A, Reyers B and Geller G N 2016 Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions *Remote Sens. Ecol. Conserv.* **2** 122–31

Reichstein M, Bahn M, Mahecha M D, Kattge J and Baldocchi D D 2014 Linking plant and ecosystem functional biogeography *Proc. Natl. Acad. Sci.* **111**

Running S W, Baldocchi D D, Turner D P, Gower S T, Bakwin P S and Hibbard K A 1999 A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data *Remote Sens. Environ.* **70** 108–27

Schimel D, Hargrove W, Hoffman F and MacMahon J 2007 NEON: A hierarchically designed national ecological network *Front. Ecol. Environ.* **5** 59

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

Scholes R J, Walters M, Turak E, Saarenmaa H, Heip C H R, Tuama É Ó, Faith D P, Mooney H A, Ferrier S, Jongman R H G, Harrison I J, Yahara T, Pereira H M, Larigauderie A and Geller G 2012 Building a global observing system for biodiversity *Curr. Opin. Environ. Sustain.* **4** 139–46

Steffen W, Crutzen P J and McNeill J R 2007 The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature *AMBIO A J. Hum. Environ.* **36** 614–21 Online: http://www.bioone.org/doi/abs/10.1579/0044-7447%282007%2936%5B614%3ATAAHNO%5D2.0.CO%3B2

Sulkava M, Luyssaert S, Zaehle S and Papale D 2011 Assessing and improving the representativeness of monitoring networks: The European flux tower network example *J. Geophys. Res. Biogeosciences* **116** 1–14

Turner II B.L., Kasperson R.E., Meyer W.B., Dow K.M., Golding M., Kasperson J. X., Mitchell R. C. R S J 1990 Two types of global environmental change: Definitional and spatial-scale issues in their human dimenssions *Glob. Environ. Chang.* **1** 14–22

Vaughan H, Brydges T, Fenech A and Lumb A 2001 Monitoring long-term ecological changes through the ecological monitoring and assessment network: Science-based and policy relevant *Environ. Monit. Assess.* **67** 3–28

Villarreal S, Guevara M, Alcaraz-Segura D, Brunsell N A, Hayes D, Loescher H W and Vargas R 2018 Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States *Agric. For. Meteorol.* **262** 423–33 Online: https://doi.org/10.1016/j.agrformet.2018.07.016

Violle C, Reich P B, Pacala S W, Enquist B J and Kattge J 2014 The emergence and promise of functional biogeography *Proc. Natl. Acad. Sci.* **111**

Williams M, Richardson a. D, Reichstein M, Stoy P C, Peylin P, Verbeeck H, Carvalhais N, Jung M, Hollinger D Y, Kattge J, Leuning R, Luo Y, Tomelleri E, Trudinger C and Wang Y-P 2009 Improving land surface models with FLUXNET data *Biogeosciences Discuss.* **6** 2785–835

Chapter 2

ECOSYSTEM FUNCTIONAL DIVERSITY AND THE REPRESENTATIVENESS OF ENVIRONMENTAL NETWORKS ACROSS THE CONTERMINOUS UNITED STATES

As *published* in Agricultural and Forest Meteorology. 2018.

DOI: 10.1016/j.agrformet.2018.07.016

Samuel Villarreal[a], Mario Guevara[a], Domingo Alcaraz-Segura[b,c,d], Nathaniel A. Brunsell[e], Daniel Hayes[f], Henry W. Loescher[g,h], Rodrigo Vargas[a,*]

Authors:

Affiliations:

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

[b] University of Granada, Department of Botany, Av. de Fuentenueva, s/n. 18701, Granada, Spain

[c] Andalusian Center for the Assessment and Monitoring of Global Change (CAESCG) – University of Almería, Crta. San Urbano, 04120, Almería, Spain

[d] iecolab, Interuniversitary Institute for Earth System Research (IISTA) – University of Granada, Av. del Mediterráneo, 18006, Granada, Spain

[e] University of Kansas, Department of Geography and Atmospheric Science, 1475 Jayhawk Blvd, Lawrence, KS, 66045-7316, USA

[f] University of Maine, School of Forest Resources, 5755 Nutting Hall, Orono, ME, 04469-5755, USA

[g] Battelle – National Ecological Observatory Network (NEON), 1685 38st, Suite 100, Boulder, CO, 80301, USA

[h] Institute of Alpine and Arctic Research (INSTAAR), University of Colorado, Boulder, CO, 20301, USA

Abstract

Environmental observatory networks (EONs) are coordinated efforts to provide knowledge that ultimately delivers transformational ecological science from regional to global scales. We used ecosystem functional types (EFTs), a time varying land surface classification, as an alternative way to characterize ecosystem functional heterogeneity based on carbon uptake dynamics. We assessed the representativeness of the eddy-covariance sites of AmeriFlux and NEON, and their combined core sites (i.e., sites with long-term support) across the conterminous United States (CONUS) based on: a) the number of different EFT categories (EFT_{mode}) represented by each network, b) representativeness of the EFT inter-annual variability (EFT_{int}; number of unique EFTs within each pixel during years 2001–2014), and c) the spatial representation of EFT_{mode} and EFT_{int} based on a maximum entropy approach (i.e., spatial functional heterogeneity). AmeriFlux represents 50% of all possible EFT categories, includes most of EFT_{int} values (9 out of 14), and represents 55% of the spatial functional heterogeneity across CONUS. NEON represents 23% of all possible EFT categories, 7 out of 14 possible EFTint values, and 23% of the spatial functional heterogeneity across CONUS. The combined effort of AmeriFlux and NEON core sites represents 33% of all possible EFT categories, 7 out of 14 possible EFT_{int} values, and 46% of the spatial functional heterogeneity across CONUS. We used the NEON ecoclimatic domains to summarize our results within a geographical context. The least represented NEON ecoclimatic domains were Desert Southwest, Southern Rockies and Colorado Plateau, Great Basin, Northern Plains, and Central Plains. Our results provide insights about the potential of AmeriFlux to address questions regarding

decadal and inter-annual variability of ecosystem functional heterogeneity across CONUS.

Highlights

- Alternative approach to assess the representativeness of environmental networks
- Ecosystem Functional Types characterize ecosystem functional heterogeneity
- The AmeriFlux network has unique spatial and temporal information
- Network collaboration enhances long-term representativeness

Keywords

Environmental networks, complex terrain, temporal representativeness, spatial representativeness, ecosystem functionality, functional diversity.

Abbreviations

Conterminous United States (CONUS), Ecosystem Functional Types (EFT), Ecosystem Functional Types inter-annual variability (EFT_{int}), Environmental observatory networks (EONs), National Ecological Observatory Network (NEON).

2.1 Introduction

Environmental observatory networks (EONs) are organizations that are affiliated in a flexible way that agree to join efforts towards a common purpose while retaining their individual objectives, resources, and management. It has been discussed that EONs are the proper structure to address complex, global and socially imperative issues (Scholes et al., 2017). EONs promote collection and dissemination of environmental data along with efforts towards standardization of protocols, data sharing and synthesis activities. Furthermore, EONs have provided value added products that include databases, maps, conceptual models, software/analytical tools for ecological modeling, and virtual communities of practice. These products have been useful for the scientific community and policy makers to assess knowledge gaps and expand the frontiers of ecological understanding (Ciais et al., 2014; Running et al., 1999). Examples of EONs include: AmeriFlux, National Ecological Observatory Network (NEON), FLUXNET, Integrated Carbon Observation System (ICOS), the Spectral Network (SpecNET), Long Term Ecological Research Network, among others (Peters et al., 2014).

Among different research efforts, the aforementioned EONs have monitored the exchange of matter (e.g., H_2O, CO_2, CH_4) and energy (e.g., heat and solar radiation) between terrestrial ecosystems and the atmosphere to better understand biosphere-atmosphere interactions (Baldocchi et al., 2001; Baldocchi et al., 2012; Law, 2005). Consequently, representativeness studies are of prime importance to discern when, where, and at what frequency EONs have been measuring or should measure ecological processes (Baldocchi et al., 2012b; Jongman et al., 2017; Vaughan et al., 2001; Vos et al., 2000). These assessments inform EONs on how to increase their utility, so the generated information could be applicable at regional and/or global scales (Ciais et al., 2014; Jongman et al., 2017; Schimel and Keller, 2015). Thus, there is a pressing need to design different scientific approaches to assess the representativeness of EONs for current and near-future applications (Lovett et al., 2007; Jongman et al., 2017).

A spatial and temporal representativeness analyses would inform where to establish new study sites and the basis to determine whether to maintain/remove current sites across networks. Thus, these analyses provide insights to improve management decisions and optimize network operability and interoperability (Vargas et al., 2017; Jongman et al., 2017). Previous studies have analyzed the spatial representativeness of national eddy-covariance networks (i.e., Canadian Carbon Program, ChinaFlux) and have concluded that the degree of fine-scale ecosystem processes across landscapes determine the number of study sites needed within a network to properly monitored those processes (Chen et al., 2012, 2011; He et al., 2015). Other studies have used cluster-based approaches to delineate spatial sampling domains and assess the spatial representativeness of EONs, and suggested arrangements of study sites of EONs such as CarboEurope-IP (Sulkava et al., 2011) and FLUXNET (Kumar et al., 2016). Representativeness studies across the conterminous United States (CONUS) have concluded that arid and semiarid ecosystems, as well as elevational changes, were under-represented by AmeriFlux during the first decade of the 2000’s (Hargrove et al., 2003; Yang et al., 2008).

In general, studies on EONs representativeness have used information regarding the spatial heterogeneity of mean climate conditions and plant functional types (PFTs) composition to represent the dynamics of ecosystem processes (i.e., carbon uptake; Hargrove et al., 2003; Kumar et al., 2016), along with ecosystem productivity and seasonality (Cramer et al., 2001; Falge et al., 2002). However, recent studies have discussed that the variability of ecological processes at the ecosystem level is insufficiently explained by using the PFTs approach (Bond-Lamberty et al., 2016; Petchey and Gaston, 2006; Petrakis et al., 2017; Reichstein et al., 2014; Wright et al., 2006).

Arguably, ecosystem functionality could complement the evaluation of the representativeness of EONs by incorporating several aspects: First, information on ecosystem functionality complements descriptions based solely on climate or vegetation structure; for example, by complementing climate drivers information with information on canopy productivity, and the temporal patterns of seasonality or phenology (Valentini et al., 1999, Alcaraz-Segura et al., 2006; 2017). Second, the inertia of ecosystem structural attributes may delay the quantification of ecosystem responses to environmental changes, while ecosystem processes (i.e., exchange of energy and matter of an ecosystem) have a faster quantifiable response (Milchunas and Lauenroth 1995; Mouillot et al., 2013). Third,, ecosystem function offers an integrative response to environmental drivers and changes (Nagendra et al., 2013; Vaz et al., 2015). Last, functional attributes allow the qualitative and quantitative assessment of ecosystem services (Costanza et al., 1997).

We explored the applicability of Ecosystem Functional Types (EFT) (Alcaraz et al., 2006) as an alternative way to characterize ecosystem functional heterogeneity (Alcaraz-Segura et al., 2013) and assess the representativeness of eddy covariance sites across AmeriFlux and NEON. EFTs have been conceptually defined as groups of ecosystems or patches of the land surface that share similar dynamics of matter and energy exchanges between the biota and the physical environment (Alcaraz et al., 2006; Paruelo et al., 2001). The EFT concept is analogous to the Plant Functional Type (PFT) concept but defined at a higher level of biological organization. As species can be grouped into plant functional types based on common species traits, ecosystems can be grouped into ecosystem functional types based on their similar ecosystem functioning. In practice, EFT is a time-varying land surface classification based on remote sensing vegetation indexes (i.e., MODIS-EVI) that are used to represent the spatial patterns and temporal variability of key ecosystem functional traits (i.e., productivity, seasonality and phenology) without prior knowledge of vegetation type or canopy architecture (Alcaraz-Segura et al., 2017, 2013; Cabello et al., 2013). Therefore, the ecosystem functional characterization obtained with EFTs can be infer information on vegetation structure and composition (e.g., canopy architecture, vegetation type, PFT), because they constitute complementary dimensions of biodiversity complexity (Noss, 1990; Pettorelli et al., 2016).

The overarching goal of this study was to assess the representativeness of AmeriFlux and NEON based on ecosystem functional diversity characterized by EFTs across CONUS. These networks monitor a wide range of ecosystem types (Novick et al., 2017; Schimel et al., 2007), and recently have joined forces to have a long-term monitoring plan to support core sites. Data from both AmeriFlux and NEON support governmental and intergovernmental programs and reports, such as the North American Carbon Program (NACP), State of the Carbon Cycle Report (SoCCR), the UN Intergovernmental Panel on Climate Change (IPCC), and multiple regional to global syntheses activities. We assess the representativeness by analyzing the categorical, temporal, and spatial representation of EFTs across CONUS. Specifically, we quantify the representativeness of (a) the historical AmeriFlux archive (i.e., all sites active and inactive within the AmeriFlux network), (b) core and relocatable NEON sites, and (c) the joint effort of AmeriFlux and NEON active core sites. In light of the 20th anniversary of the AmeriFlux network, we asked three interrelated research questions: What are the spatial and temporal patterns of EFTs across CONUS? How do the historical AmeriFlux archive and planned NEON sites represent spatial and temporal patterns of EFTs across CONUS? and What is the representativeness of the joint effort of AmeriFlux and NEON core sites? We used the 17 NEON ecoclimatic domains across CONUS as geographical categories to organize and summarize the results of this study. Our EFT-based approach provides an alternative framework to previous assessments of the representativeness of EON's (Hargrove et al., 2003; Yang et al., 2008; Chen et al., 2012), it is explicitly based on ecosystem functional attributes derived from publicly available data, provides insights for the design, improvement, and growth of EONs, and it is applicable to other EONs around the world.

2.2 Materials and methods

2.2.1 Environmental Observatory Networks

AmeriFlux is an integrated "bottom-up" effort from principal investigators (PIs) to coordinate eddy covariance measurements across the most common ecosystems in the United States and the Americas (Keller et al., 2011; Law, 2005; Novick et al., 2017). The historical AmeriFlux archive represents the total wealth of information collected by all active and inactive study sites registered since the establishment of AmeriFlux. The historical AmeriFlux archive has a total of 207 registered study sites across the CONUS and 46 of those sites are currently considered to be core sites. Those core sites have received direct support and funding from the

AmeriFlux Management Program (AMP) and are more likely to remain active (i.e., long-term, >10 years) than independently funded sites (AMP 2017). The number of active sites within the AmeriFlux archive has varied through time due to multiple factors (e.g., available funding, human resources, project timelines).

The National Ecological Observatory Network (NEON) is an ecological observatory platform that is organized under a "top-down" approach, which is designed for discovering, understanding and forecasting of ecosystem processes at a continental scale (Kampe et al., 2010, Schimel et al., 2011). NEON observations are distributed across 20 ecoclimatic domains (i.e., NEON domains), which act as spatial sampling domains and represent regions of distinct landforms, vegetation, climate and ecosystem dynamics (Keller et al., 2011; Schimel et al., 2007). NEON domains are derived from ecoclimatic variables that are clustered based on a multivariate statistical approach, the clusters are formed in a way that each of them grouped the same fraction of the total ecoclimatic variance (Hargrove and Hoffman. 1999; Keller et al., 2011). Each NEON domain is represented by one core wild land site (total 20 observatory sites, 17 within CONUS) and additional relocatable sites (39 within CONUS) to represent the ecoclimatic properties and gradients within and among NEON domains (Keller et al., 2011; Schimel et al., 2011), or address grand challenge areas as described by the National Academy of Sciences (NRC, 2001, 2003, 2007). Throughout this study, we used the 17 NEON domains across CONUS to organize and summarize our results.

Finally, AmeriFlux and NEON have a unique opportunity for long-term monitoring by joining efforts from core sites. AmeriFlux has selected 46 core sites while NEON has designed 17 core observatory sites across CONUS (n = 68 of

AmeriFlux plus NEON core sites). Thus, there is a need to provide information of potential representativeness as these networks join long-term monitoring efforts.

2.2.2 Terrain complexity

We used terrain complexity as a static topographic metric derived from a digital elevation model. Complex topography is an important limitation for the eddy-covariance technique as it influences the assumption of horizontal homogeneity required for a proper estimation of biosphere-atmosphere fluxes (Göckede et al., 2004). Terrain complexity was derived from a publicly available digital elevation model consisting of a 30-arc second resolution global topographic/bathymetric grid (Becker et al., 2009). Terrain complexity was defined by calculating ±1 standard deviation of the terrain altitude within areas of approximately 0.05^{o} x 0.05^{o}. We used this resolution to represent the major topographic characteristics of CONUS as this resolution is widely used in country-scale or regional studies (Löw et al., 2005; Piao et al., 2015; Chrysoulakis et al., 2003). We used this metric to describe the mean terrain complexity for each one of the NEON domains across CONUS.

2.2.3 Ecosystem Functional Types

The basis of the concept of EFTs assumes that by using time-series of satellite images it is possible to identify and map areas with similar ecosystem functional characteristics (Alcaraz-Segura et al., 2017; Alcaraz et al., 2006; Paruelo et al., 2001). Spectral indices derived from satellite images can provide information about key ecosystem functional aspects such as primary production, evapotranspiration, surface temperature and albedo (Fernandez et al., 2010; Lee et al., 2013; Paruelo et al., 1997).

In this work, we used a holistic approach to assess ecosystem functioning (Jax 2010). The regional ecosystem functional heterogeneity was characterized by means of EFTs derived from the seasonal dynamics of the Enhanced Vegetation Index (EVI), as a surrogate of carbon gain dynamics (Alcaraz-Segura et al., 2013). We used the 2001-2014 time-series of satellite images of the EVI from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) product MOD13C2 with a spatial resolution of 0.05º x 0.05º across CONUS. We used this resolution to characterize the patterns at the country scale as done in other studies (Löw et al., 2005; Trieb et al., 2002; Chrysoulakis et al., 2003). EFTs were derived from three meaningful metrics of the EVI seasonal curve related to the dynamics of terrestrial carbon gains: a) annual mean (EVI_Mean) as an estimator of primary production; b) EVI seasonal coefficient of variation (EVI_sCV) as a descriptor of seasonality; and c) the month of the annual maximum EVI value (DMAX) as an indicator of phenology. Those three metrics represent more than 80% of variance in the annual EVI time series (Alcaraz et al., 2006; Paruelo et al., 2001). The range of values of each EVI metric was divided into four intervals, giving a potential number of 64 EFTs (i.e., 4 x 4 x 4 = 64; Alcaraz-Segura et al., 2013). To obtain the intervals for EVI_mean and EVI_sCV, we extracted the first, second, and third quartiles for each year, and then calculated the inter-annual mean of each quartile. For EVI_DMAX, the four intervals agreed with the four seasons of the year (Alcaraz-Segura et al., 2013).

We labeled all 64 EFT categories using a previously published nomenclature, where two letters and one number describe each category (Alcaraz-Segura et al., 2017; Paruelo et al., 2001). Therefore, each EFT category is a summary of the information contained in the three EVI metrics for each 0.05º x 0.05º grid pixel. The first letter (capitalized) represents the EVI_Mean, which ranged from *A* (low primary productivity) to *D* (high primary productivity). The second letter represents EVI_sCV, which ranged from *a* (high seasonality) to *d* (low seasonality; Alcaraz-Segura et al., 2013). The third position is a number that represents DMAX, which indicates the phenology stage during the year (*1-4*: spring, summer, autumn, and winter, respectively, Alcaraz-Segura et al., 2013). For example, *Aa1* represents an EFT category of low productivity, high seasonality and with a growing season with a spring maximum. In contrast, *Dd2* represents an EFT with a high productivity, low seasonality and a growing season with summer maximum.

2.2.4 Network representativeness analyses

2.2.4.1 Categorical representativeness

To summarize the spatial heterogeneity of ecosystem functioning of the 2001-2014 period, we calculated the mode of the annual EFT maps. We refer to this as the EFT_{mode} and consequently it corresponds to the most dominant EFT for each pixel during the 14-year period. The categorical representativeness analysis evaluated whether each one of the EFT_{mode} categories found across CONUS was represented by: (a) the historical AmeriFlux archive; (b) NEON sites; or (c) AmeriFlux and NEON core sites. In addition, we analyzed how the AmeriFlux network has represented the EFT categories as sites have been added or became inactive in the network throughout the 2001–2014 period.

2.2.4.2 Temporal representativeness

EFT categories can change through time in a particular pixel as they represent annual dynamics of terrestrial carbon gains within each pixel across CONUS. Thus, we used the number of different EFTs occurring within each pixel throughout the 2001-2014 period as an indicator of the inter-annual variability in ecosystem functioning (EFT_{int}). For example, if a pixel displayed three EFT categories from 2001 to 2014, then EFT_{int} was 3; despite if one EFT was more abundant than the other two. An EFT_{int} of 14 meant that every year was different to the rest in terms of the EFT categories. The temporal representativeness analysis evaluated whether different values of inter-annual variability (EFT_{int}) were covered by: (a) the historical AmeriFlux archive; (b) NEON sites; or (c) AmeriFlux and NEON core sites.

2.2.4.3 Spatial representativeness

We assessed the spatial representativeness of the network using a probability distribution technique based on maximum entropy distribution (Phillips et al., 2004; 2006). We used this approach to express the suitability of the study sites to monitor the range of ecosystem functional heterogeneity across CONUS. The maximum entropy approach (Maxent) is largely used in estimating the relationship between spatial observations (i.e., site locations) and environmental or spatial properties (i.e., EFT_{mode} and EFT_{int}) associated with those locations across a well-defined geographic region (i.e., CONUS). Entropy can be seen as a measure of dispersedness, while the maximum entropy approach maximizes the entropy distribution of a set of environmental properties within a geographic space (Elith et al., 2011). Here, we performed a Maxent analysis for: (a) the historical AmeriFlux archive; (b) NEON sites; and (c) AmeriFlux and NEON core sites, to represent the heterogeneity of EFT_{mode} and EFT_{int} (i.e., environmental properties) across CONUS. The randomness of the Maxent model was tested using the area under the curve (AUC) of the training data (i.e., EFT_{mode} and EFT_{int}) and that of a random prediction as recommended

(Fielding and Bell, 2016; Hijmans, 2012; Liu et al., 2011; Phillips et al., 2004). A random classification has a typical value for the area under the curve equal to 0.5, while a non-random classification (i.e., distinction between potential presence and absence) has values closer to 1. The final result derived from our Maxent model are expressed using a Kappa index derived from cross-validation, where Kappa index of 1 indicates areas with characteristics that are more likely to be monitored by the study sites (i.e., sampling locations). We reported spatial representativeness results as the percentage ratio of those pixels with a Kappa index equal to 1 divided by the total number of pixels across CONUS. See Appendix A for more detail.

2.3 Results

2.3.1 Categorical representativeness

The EFT_{mode} for the 2001-2014 period across CONUS (Figure 2.1, Table 2.1) showed that ecosystems with high productivity were located in the NEON ecoclimatic domains of Northeast, Mid-Atlantic, Southeast, Appalachians and Cumberland Plateau, and the Ozark Complex. In contrast, ecosystems with low productivity were found at the Great Basin, Desert Southwest and Southern Rockies, and Colorado Plateau. The ecosystems with the highest seasonality were common in the Great Lakes, Prairie Peninsula, Northern Plains, Northeast, and Appalachians and Cumberland Plateau; while ecosystems with the lowest seasonality occurred in the South East, Central Plains, and Southern Plains. Most ecoclimatic domains were dominated by ecosystems with growing season with summer maxima, except for the Great Basin, Pacific Southwest and Desert Southwest where the growing season maxima was reached during spring.

The historical AmeriFlux archive covered 31 out of the 64 possible EFT_{mode} categories (Figure 2.2). In contrast, NEON sites only represented 16 EFT_{mode} categories (Figure 2.2c), and the combined efforts of the AmeriFlux and NEON core sites represented 21 EFT_{mode} categories (Figure 2.2d). The frequency distribution of the number of sites (across AmeriFlux and NEON networks) did not follow the frequency distribution of EFT_{mode} categories (Figure 2.2a). In other words, the most abundant EFT_{mode} categories across CONUS did not have the largest number of monitoring sites.

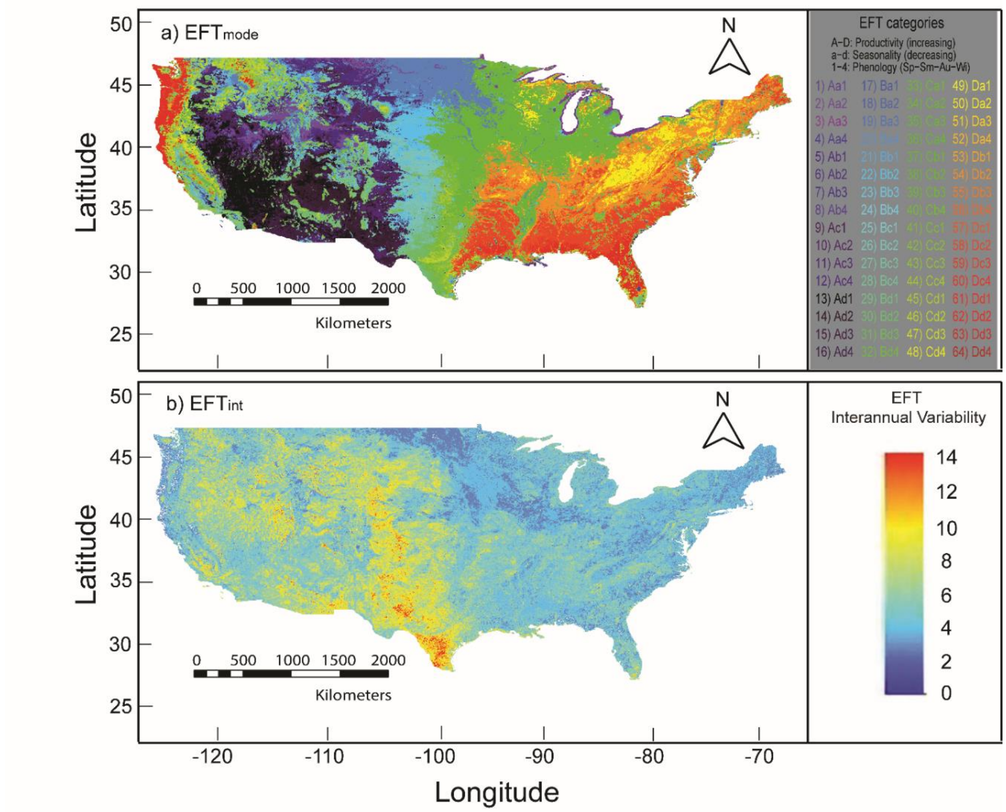


Figure **2.1.** Spatial distribution and inter-annual variability of Ecosystem Functional Types (EFTs) across the conterminous United States (CONUS) during the 2001-2014 period: a) spatial patterns of the mode of EFTs for the 2001-2014 period (EFT_{mode}); and b) inter-annual variability of EFTs (EFT_{int}; i.e., number of unique EFTs that occurred in the 14-year period), where red areas represent high inter-annual variability and blue areas low inter-annual variability. EFTs were calculated from Moderate Resolution Imaging Spectroradiometer Enhanced Vegetation Index (MODIS-EVI). Capital letters correspond to the EVI annual mean (EVI_Mean) level, ranging from *A* to *D* for low to high productivity. Small letters show the seasonal coefficient of variation (EVI_sCV), ranging from *a* to *d* for high to low seasonality for carbon uptake. The numbers indicate the season of maximum EVI (DMAX): (1) spring, (2) summer, (3) autumn, (4) winter. The map uses the 0.05ºx0.05º Global Climate Modeling Grid in geographic projection with Datum WGS84.

Table **2.1.** Dominant Ecosystem Functional Types (EFT_{mode}) and mean terrain complexity for each NEON ecoclimatic domain across the conterminous United States.

NEON ecoclimatic domain	**Two most dominant EFT_{mode}**	**Mean EFT inter-annual variability (EFT_{int})**	**Mean terrain complexity**
Northeast	Db2, Da2	3.3	63.1
Mid Atlantic	Db2,Dc2	3.8	30.1
South East	Dc2,Dd2	3.9	8.1
Atlantic Neotropical	Dd2,Cd2	4.9	1.7
Great Lakes	Ca2,Cb2	3.5	13.9
Appalachian and Cumberland Plateau	Db2,Da2	3.7	54.4
Prairie Peninsula	Ca2,Cb2	3.5	15.1
Ozark Complex	Db2,Dc2	4.2	22.7
Northern Plains	Ba2,Bb2	4.7	35.1
Central Plains	Bc2,Bb2	6.6	25.1
Southern Plains	Cc1,Cc2	6.7	20.1
Northern Rockies	Cd2,Bd2	5.1	207.6
Great Basin	Ad1,Ac1	5.6	142.6
Southern Rockies and Colorado Plateau	Ad2,Bd2	5.2	147.0
Desert Southwest	Ad1, Ad2	6.0	120.1
Pacific Northwest	Dd2,Cd2	3.5	194.2
Pacific Southwest	Bd1,Cd1	4.7	168.0

Note: The EFT_{mode} represents the summary of the spatial heterogeneity of ecosystem functioning of the 2001-2014 period. EFT_{int} represents the average number of different EFT that occurred within an ecocolimatic domain during the 2001-2014 period. Terrain complexity was defined by calculating the ±1 standard deviation of the terrain altitude within areas of approximately 0.05° x 0.05°.

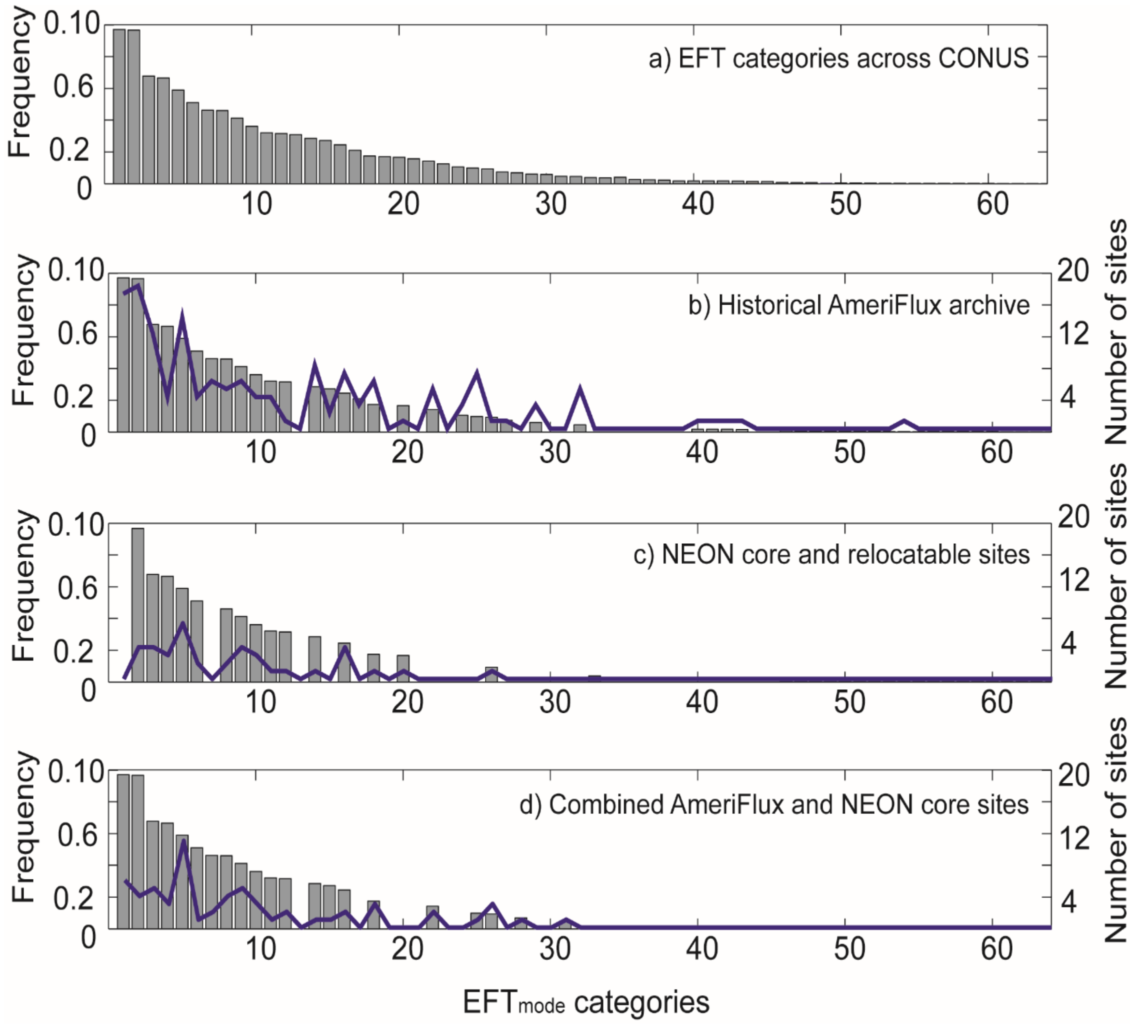


Figure 2.2. Categorical representativeness of the Ecosystem Functional Type mode (EFT_{mode}) of the 2001-2014 period across CONUS. EFT_{mode} corresponds to the most dominant EFT for the 14-year period at each pixel. a) Frequency of EFT_{mode} across CONUS; b) EFT_{mode} categories represented by the historical AmeriFlux archive; c) EFT_{mode} categories represented by NEON sites; and d) EFT_{mode} categories represented by current AmeriFlux and NEON core sites. Grey bars indicate represented frequency of EFT_{mode} categories, and lines represent the number of study sites in each EFT_{mode} category. X-axes indicate represented EFT_{mode} categories (in b, c, d) sorted by frequency (as in a) across CONUS.

Year-specific categorical representativeness of AmeriFlux changed through time as eddy covariance sites have been added or became inactive from the network (Table 2.2). Despite the sustained increase in the number of eddy covariance sites across the years, the number and EFT_{mode} categories represented by AmeriFlux have remained relatively constant since 2007. The most common EFT_{mode} categories represented by AmeriFlux are *Ca2* (i.e., ecosystems with medium high productivity, very high seasonality, and summer maximum) and *Db2* (i.e., ecosystems with very high productivity, low seasonality, and summer maximum).

Table 2.2. Changes in categorical representativeness of the AmeriFlux network in terms of number of EFT_{mode} categories represented by active sites for each year between 2001 and 2014.

Year	**Number of sites**	**Number of EFT_{mode} categories**	**Most represented EFT_{mode} categories**
2001	37	16	Dd2 (6), Ca2(5)
2002	46	17	Dd2 (7), Ca2(7)
2003	50	19	Db2(8), Dd2(7)
2004	77	24	Db2 (13), Ca2(10)
2005	78	24	Ca2(11), Db2(10)
2006	81	26	Ca2(13), Db2(10)
2007	99	28	Ca2(14), Db2(11)
2008	98	29	Ca2(14), Db2(9)
2009	108	29	Ca2(20), Db2(9)
2010	107	30	Ca2(20), Db2(9)
2011	111	30	Ca2(22), Db2(9)

Table 2.2 Continued

Year	**Number of sites**	**Number of EFT_{mode} categories**	**Most represented EFT_{mode} categories**
2012	117	31	Ca2(21), Db2(10)
2013	124	31	Ca2(21), Db2(11)
2014	131	31	Ca2(20), Db2(11)

Note: numbers in parenthesis under "Most represented EFT_{mode} categories" represent number of active sites for each EFT_{mode} category. For example, during year 2014 there were 20 active sites in the AmeriFlux network for EFT_{mode} category *Ca2*.

2.3.2 Temporal representativeness

We mapped the patterns of EFT_{int} (i.e., number of different EFT that occurred in a pixel during the 2001-2014 period) across CONUS (Figure 2.1b). The highest EFT_{int} was found in the Southern and Central Plains, while the lowest variability was found in the Great Lakes, Prairie Peninsula, Pacific Northwest, and Northeast (Table 2.1). Across CONUS, the most common values of EFT_{int} were between 3 and 5, EFT_{int} values <3 or >9 were less common, and the maximum value of EFT_{int} was 14 (Figure 2.3a).

The historical AmeriFlux archive included information of sites with EFT_{int} values between 1 and 9, and where values between 3 and 5 were also the most common (Figure 2.3b). Nearly 80% of study sites within the AmeriFlux network were located at EFT_{int} values between 3 and 6; 7% at EFT_{int} values ≤2; and 12% at EFT_{int} values ≥7. Across NEON sites, EFT_{int} values 3, 4 and 7 were the most common (Figure 2.3c). The combined effort of AmeriFlux and NEON core sites did not include EFT_{int} values <3 or >9, despite the fact that EFT_{int} value 2 is relatively abundant across CONUS (Figure 2.3a).

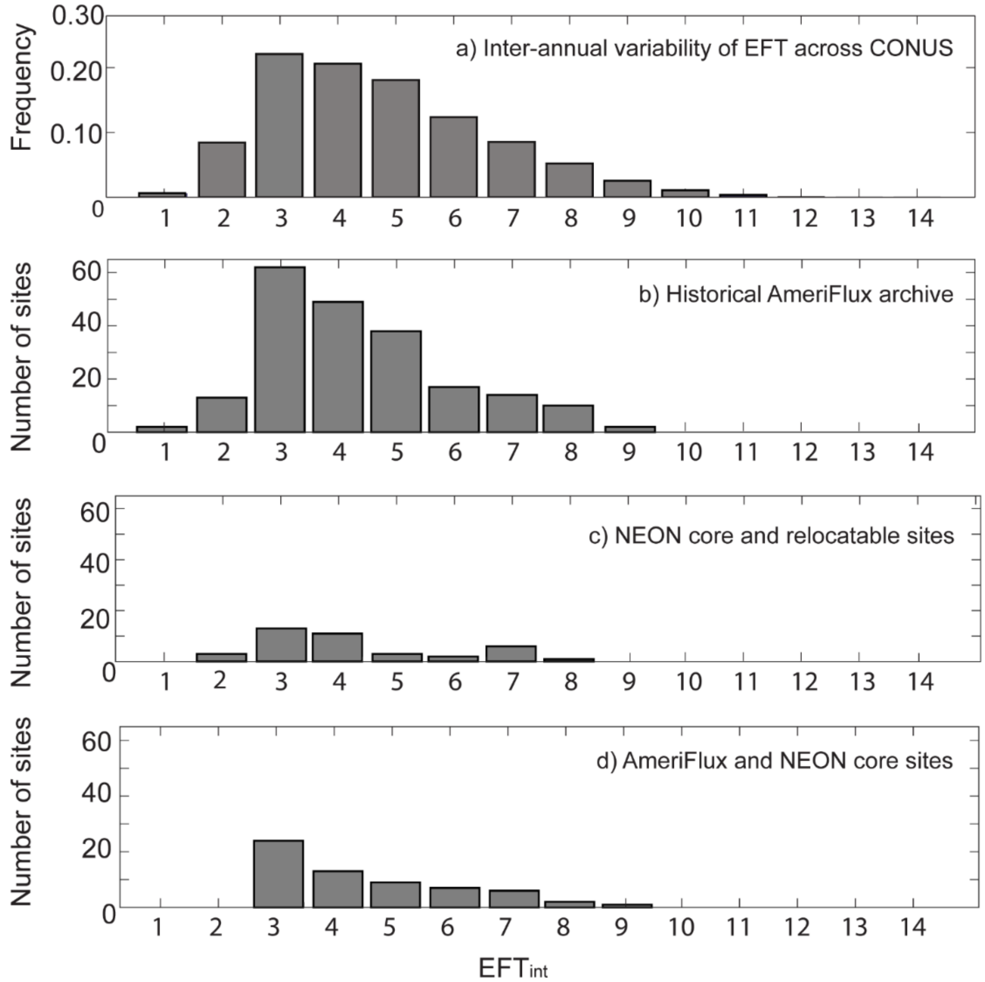


Figure 2.3. Representativeness of the inter-annual variability of EFTs during the 2001-2014 period across CONUS. Inter-annual variability of EFTs is expressed as the number of EFTs occurring in each 0.05^{o} pixel during the 14-year period (EFT_{int}). (a) Histogram distribution of EFT_{int} values across CONUS. Number of sites representing each value of EFT_{int} for (b) historical AmeriFlux archive; (c) NEON sites; and (d) current AmeriFlux and NEON core sites. X-axes represent EFT_{int} values for each panel.

Individual monitoring sites within the historical AmeriFlux archive had between 1 and 28 years of available eddy-covariance information (Figure 2.4). Most study sites (62) were located at an EFT_{int} value of 3 (Figure 2.3b, 4), and 42 of these sites had >3 years of available information (Figure 2.4). Sites at EFT_{int} values of 4 (49 sites) and 5 (38 sites) were also common, with 29 and 24 sites having more than 4 and 5 years of available information, respectively. Only one site with >9 years of information was located at an EFT_{int} value of 9 (Figure 2.4).

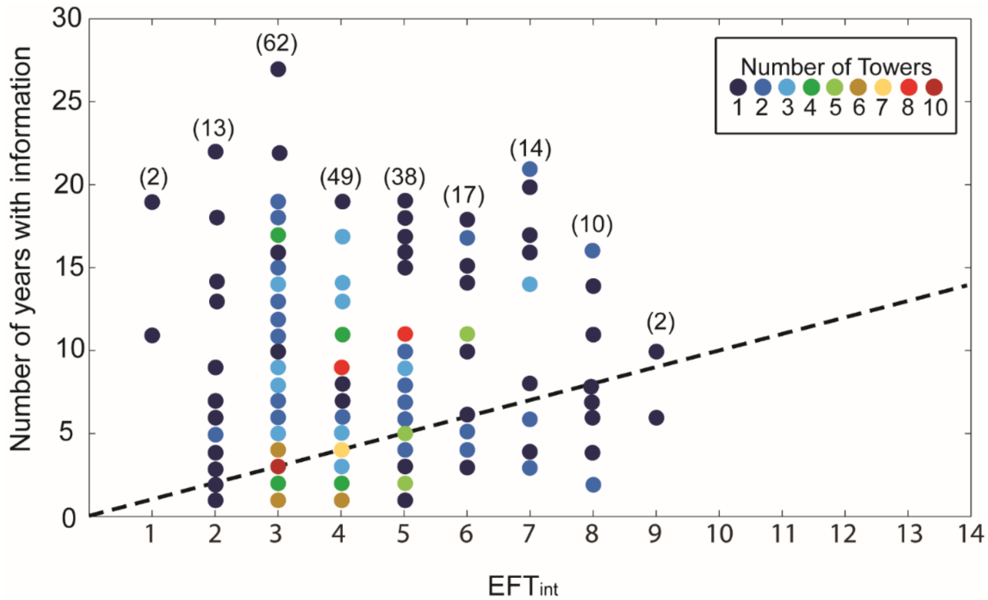


Figure 2.4. Representativeness of the inter-annual variability of EFTs (i.e., EFT_{int}) and number of years with information in the historical AmeriFlux archive. The X-axis represents values of EFT_{int}, and the Y-axis represents the number of years with eddy covariance information per site available in the historical AmeriFlux archive. Colors represent the number of sites that report a specific number of years with eddy covariance information for each value of EFT_{int}. Numbers in parenthesis indicate the number of total study sites available for each EFT_{int} value within the historical AmeriFlux archive. For example, there is a total of 2 sites in the historical AmeriFlux archive (number in parenthesis) with an EFT_{int} value of 9 (see X-axis), where one single site (color or the circle [dark blue]) has 6 years of information (see Y-axis) and a second single site (color of the circle [dark blue]) has 10 years of information (see Y-axis). The dashed line represents the threshold where the number of years of available information is equal to EFT_{int}.

2.3.3 Spatial Representativeness

The maximum entropy analysis provides information of the representativeness of AmeriFlux, NEON and the combine core sites to monitor the spatial variability of EFT_{mode} and EFT_{int} (Elith et al., 2011). The overall spatial representativeness is expressed as the ratio of all pixels with a Kappa index equal to 1 divided by the total number of pixels across CONUS. This resulted in a spatial representativeness of 55% by the historical AmeriFlux archive, 23% by NEON sites, and 46% by the combined AmeriFlux and NEON core sites of the CONUS surface (Table 2.3, Figure A-1). The most represented ecoclimatic domains by the historical AmeriFlux archive were Great Lakes, Prairie Peninsula, Northeast, and Appalachians and Cumberland Plateau whereas the least represented were Desert Southwest, Northern Plains and Great Basin (Table 2.3, Figure A-1). NEON sites had high spatial representation across the Northeast, Appalachians and Cumberland Plateau and Mid Atlantic domains, whereas the least represented domains were Desert Southwest, Northern Plains, and Southern Rockies and Colorado Plateau (Table 2.3). The most represented ecoclimatic domains by the combined effort of AmeriFlux and NEON core sites were Pacific Northwest, Northeast and Mid-Atlantic, whereas the least represented were Desert Southwest, Northern Plains, and Southern Rockies and Colorado Plateau (Table 2.3, Figure A-1).

Maxent model was tested using the area under the curve (AUC). The AUC for the historical AmeriFlux archive (0.65), NEON sites (0.59), or AmeriFlux and NEON core sites (0.63) were always higher than the AUC of a random prediction (0.5); thus, supporting the applicability of the maximum entropy analysis. The relative contribution of each variable to the maximum entropy analyses was 88% for EFT_{mode} and 12% for EFT_{int}.

Table 2.3. Spatial representativeness for each network and the combined core sites, based on the ratio of the cross-validation results derived from the maximum entropy analysis. The percentage correspond to the ratio of those pixels with a kappa index equal to 1 divided by the total number of pixels across each NEON domain.

NEON ecoclimatic domain	Historical AmeriFlux archive (%)	Core and relocatable NEON sites (%)	AmeriFlux and NEON core sites (%)
Northeast	89	85	84
Mid Atlantic	58	68	64
South East	55	54	58
Atlantic Neotropical	79	25	36
Great Lakes	92	30	46
Appalachians and Cumberland Plateau	89	70	58
Prairie Peninsula	91	10	24
Ozark Complex	63	52	44
Northern Plains	12	1	5
Central Plains	39	1	31
Southern Plains	55	8	40
Northern Rockies	48	12	40
Great Basin	20	4	17
Southern Rockies and Colorado Plateau	26	1	16
Desert Southwest	6	1	6
Pacific Northwest	83	39	87
Pacific Southwest	73	23	63

Note: numbers indicate percent area represented within a NEON ecoclimatic domain.

2.4 Discussion

2.4.1 Categorical representativeness

Our results demonstrate how the characterization of ecosystem functional heterogeneity made by EFTs at the regional scale can be applied to assess the representativeness of EONs. EFT_{mode} showed a contrasting pattern of carbon gain dynamics across CONUS. The ecoclimatic domains located in temperate humid conditions (Bailey, 1983) such as Northeast, Appalachian and Cumberland Plateau, Mid Atlantic, South East, Atlantic Neotropical and the Pacific Northwest showed high productivity, low seasonality, and had a growing season with summer maxima (Figure 2.1a; Table 2.1). In contrast, ecoclimatic domains located in grasslands and open-shrublands under dry conditions such as Great Basin, Desert Southwest and the Southern Rockies and Colorado Plateau showed the lowest productivity, low seasonality, and the growing season was tightly coupled with water availability (e.g., spring in Mediterranean regions, summer across the North American Monsoon region).

Our results demonstrate that ecosystem functional heterogeneity is well represented by the historical AmeriFlux archive, which included nearly 50% of all possible EFT_{mode} categories across CONUS. The AmeriFlux network, as a bottom-up community effort, has experienced the removal and addition of eddy covariance sites over the last two decades. Thus, at any given year, some EFT_{mode} categories could have been added or removed based on the location of active eddy covariance sites. The network has constantly increased the number of active sites across years, but the number of EFT_{mode} categories has remained relatively constant (~30 categories) since 2007. Furthermore, ecosystems with very high or medium high productivity, very high or low seasonality, and with growing seasons with summer maximum (*Ca2*, *Db2*) have been the most commonly monitored since 2005, likely due to the large interest on large terrestrial carbon sinks (Running et al., 1999). It is likely that AmeriFlux will continue providing information from these and other EFT_{mode} categories as researchers address unexplored ecological questions across ecosystems.

The long-term perspective of AmeriFlux and NEON core sites will provide information of the 12 most dominant EFT_{mode} categories across CONUS (33% of all possible EFT_{mode} categories). That said, the probability distribution of these core sites did not follow the probability distribution of EFT_{mode} categories across CONUS (Figure 2.2). This means that the most abundant EFT_{mode} categories do not necessarily have the largest number of study sites. Looking forward, these results open questions about network design, such as: a) Should new monitoring sites emphasize research on ecosystems within EFT's with the most frequency of occurrence (i.e., *Ca2* and *Db2*)? or b) Should new monitoring sites aim to represent the probability density distribution of ecosystem functional heterogeneity across CONUS? Long-term monitoring core sites are and will continue to be limited due to the financial and pragmatic requirements for their operation, but the joint effort by AmeriFlux and NEON provides an exciting and unique opportunity for decadal-scale information that otherwise would not be available.

2.4.2 Temporal representativeness

The inter-annual variability of EFTs showed contrasting patterns across CONUS. We postulate that NEON ecoclimatic domains with lower EFT_{int} values are typified by forested ecosystems in temperate humid regions, which are mainly constrained by temperature, light and nutrient cycling (Allen and Chapman, 2001;

Nemani et al., 2000; Vargas et al., 2010). In contrast, ecoclimatic domains with high EFT_{int} values are represented by grasslands and shrublands across water-limited regions, and are sensitive to changes in timing and magnitude of precipitation that substantially influence carbon gain dynamics (Arredondo et al., 2016; Schwinning et al., 2004; Vargas et al., 2010). Quantifying EFT_{int} values is important as recent studies have highlighted the need of long-term flux data records to describe the inter-annual variability of carbon uptake (Novick et al., 2017; Zscheischler et al., 2016).

We highlight that EFT_{int} represents the number of changes in EFT categories within a single pixel. This does not necessarily mean that changes in EFTs are changes in vegetation structure or composition (e.g., changes from a forest to a grassland). Changes in EFT categories could be the result of ecosystem structural changes such as those imposed by land-use change (e.g., deforestation), but also the result of more subtle changes. For example, a pixel could represent a grassland throughout our study period (i.e., 2001-2014), but displayed a EFT_{int} value of 5. This means that the plant functional type and vegetation structure was the same (i.e., grasslands) throughout the study period, but there were changes in terms of productivity, seasonality and phenology that resulted in different EFT categories. This could happen for instance, as a result of droughts, floods or fires. In addition, the same EFT_{int} value of 5 could be present in grasslands, shrublands, or evergreen forests, but it only indicates unique changes in EFT categories throughout the study period. Thus, site-specific interpretation of our results should take into consideration the underlying plant functional type and history (land use or weather) at a location of interest. Overall, our results highlight the importance of network representativeness to understand how changes in biophysical forcing factors could influence ecosystem functional heterogeneity across regions and the whole CONUS. We recognize that this approach requires further development and research, but also acknowledge that the addition of EFT information has already improved the performance of regional climate (Lee et al., 2013) and biodiversity models (Alcaraz-Segura et al., 2017, 2013).

The historical AmeriFlux archive has a good representation of the inter-annual variability of EFTs across CONUS. Most eddy covariance sites within AmeriFlux have EFT_{int} values between 3 and 6, which are also the most common values across CONUS. In contrast, NEON lacks representation of EFT_{int} values <3 and has a higher representation of sites with an EFT_{int} value of 7. The long-term perspective of AmeriFlux and NEON core sites will provide information of EFT_{int} values between 3 and 9. Both the historical AmeriFlux archive and NEON do not have sites at EFT_{int} values >9, regardless there are pixels with EFT_{int} values up to 14 across CONUS. We recognize that areas with high EFT_{int} values are rare, and properly monitoring their long-term carbon dynamics will require decades due to their high inter-annual variability. Long-term monitoring of ecosystems with low EFT_{int} values could provide information about ecosystem resiliency from weather variability and disturbances; while monitoring ecosystems with high EFT_{int} values could provide information from the most sensitive ecosystems in terms of carbon uptake dynamics.

Many AmeriFlux study sites have more years with site-specific measurements than the annual temporal variability of EFT (EFTint) associated to the location of those sites (Figure 2.4). For example, the network has information of 62 sites located at an EFT_{int} value of 3, but >40 sites have over 3 years of site-specific measurements. On one end of the spectrum, there are 2 sites with over 10 years of site-specific measurements at the EFT_{int} value of 1, where questions about ecosystem stability and resiliency could be asked. On the other end of the spectrum, there are 4 sites (out of 10) at the EFT_{int} value of 8 with over 10 years of site-specific measurements, where we can ask questions about sensitivity and variability of ecosystem processes. Overall, our results support that the AmeriFlux network has unique information to address questions regarding inter-annual variability of carbon gain dynamics, ecosystem stability and resiliency across the CONUS.

2.4.3 Spatial representativeness

Our results show that the historical AmeriFlux archive includes information of ecosystem functional heterogeneity for 55% of the CONUS. This contrast with the 23% of the CONUS represented by NEON sites, but the sites in this network are fewer and with a long-term perspective than the wide bottom-up effort of AmeriFlux. It is important to mention that the combined effort of AmeriFlux and NEON core sites represents 46% of CONUS surface, demonstrating that few but strategically located sites could represent a large proportion of the continental ecosystem functional heterogeneity.

In general, AmeriFlux and NEON (individually) do not properly represent ecosystems dominated by grasslands and shrublands across water-limited ecosystems. These results are in accordance with previous studies that identified an overall high representativeness of temperate forested ecosystems by the AmeriFlux network (Hargrove et al., 2003; Yang et al., 2007), but to our knowledge no assessment has been done for the NEON eddy covariance sites. Historically, there has been a (bias) better representation of ecosystems with larger potential to uptake and store carbon, likely due to the large interest on quantifying and characterizing the processes that control large terrestrial carbon sinks (Cramer et al., 2001; Luo et al., 2007; Running et al., 1999). We highlight that these forested lands are of critical importance for the regional carbon budget of North America (Hayes et al., 2012) and the world (Pan et al., 2011). That said, there is an increasing interest to improve the representation of water-limited ecosystems in ecosystem processes-based models (Biederman et al., 2016; Vargas et al., 2013) as is important to understand how their inter-annual variability contributes to the regional-to-global carbon balance (Ahlström et al., 2016; Biederman et al., 2017; Poulter et al., 2014)

Our results provide evidence that there is a lack of representation by the historical AmeriFlux archive and NEON sites across the Desert Southwest, Southern Rockies and Colorado Plateau, Great Basin, Northern Plains, and Central Plains ecoclimatic domains. These regions have been recognized to have wide range of bioclimatic drivers (Gilmanov et al., 2005; Zhang et al., 2010) and anthropogenic activities such as land-use-change (Chuluun and Ojima, 2002). Thus, research in these regions represent an opportunity to better understand socio-ecological processes and the nexus of food, energy, and water systems (Bazilian et al., 2011).

The combined effort of the AmeriFlux and NEON core sites lacks representation of the Prairie Peninsula ecoclimatic domain. These core sites have good representation of the CONUS surface (46%) and almost represent the same ecoclimatic domains as the historical AmeriFlux archive and NEON sites (Table 2.3, Figure A-1). The Midwest corn belt of the United States produces over 35% of the global corn production and is part of the Prairie Peninsula (Graham et al., 2007; Ort and Long, 2014). Hence, the ecosystem functional heterogeneity of this region represents a network limitation when long-term carbon dynamics for agro-ecosystems are considered for the spatial representativeness of the CONUS. The lower representation at this and other ecoclimatic domains brings attention to the limitations to cover a heterogeneous landscape with few core sites.

Finally, complex topography creates ecological niches that could influence carbon dynamics across topographic gradients and landscapes (Katul et al., 2006; Swanson et al., 1988). For example, it has been estimated that nearly 70% of the carbon uptake across the western CONUS occurs at high elevation, with about 50-85% taking place on complex terrain (Schimel et al., 2002). Unfortunately, complex topography is a large limitation for implementation of the eddy covariance technique as it is often violates assumptions for the technique for annual carbon budgets and promotes advection processes (Göckede et al., 2004). Congruently, three of the least represented NEON ecoclimatic domains are also characterized by complex topography (Southern Rockies and Colorado Plateau, Great Basin, Desert Southwest; Table 2.1). These results support previous reports that suggest AmeriFlux lacks representation of the western mountain ranges of the CONUS (Hargrove et al. 2003). We argue that monitoring ecosystem functional heterogeneity across complex terrain represents a final frontier for AmeriFlux and NEON networks that could limit an accurate spatial inference of carbon dynamics across CONUS.

2.4.4 Considerations and network inferences

We provide an alternative approach to assess representativeness of EON's based on metrics of ecosystem functional heterogeneity that complements previous assessments based on vegetation climatic or structural features. We interpret EVI dynamics as a surrogate for ecosystem carbon gain dynamics. There are known limitations when using EVI, especially in evergreen and water-limited ecosystems, that could influence the assumption that EVI is closely related to carbon gain dynamics (Ha et al., 2015; Sims et al., 2014). Hence, our EFT classification may inherit the intrinsic limitations of EVI and consequently there could be area-specific biases; for example: a) areas with apparent low inter-annual variability (i.e., EFT_{int}) could actually have larger inter-annual variability (e.g., evergreen forests); and b) areas with apparent low seasonality could in fact have larger seasonality (e.g., grasslands and shrublands). That said, there is strong evidence that EVI is still the best predictor for describing carbon gain dynamics at the continental scale (Rahman et al., 2005; Sims et al., 2006). Arguably, there is no definitive and universal definition for PFTs where there could be different criteria to develop classifications (Ustin et al., 2004). Similarly, there could be different criteria to develop classifications of EFTs (e.g., carbon gains, water balance, energy balance). We propose that the current limitations for calculation of EFTs based on EVI as a surrogate of carbon gain could be addressed with long-term remote sensing information on solar induce chlorophyll fluorescence (Joiner et al., 2011), or new pigment indexes sensitive to seasonality of evergreen conifers (Gamon et al., 2016).

The historical AmeriFlux archive is unique for representing regional biosphere-atmosphere interactions (focused in CONUS) and is only rivaled by information from European networks. The number of active sites has consistently grown every year, but the number of sites sharing data has decreased since 2005 (Novick et al 2017). This means that our representative analysis of the historical AmeriFlux archive is only applicable if all sites share the available data (Figure 2.4). Our analyses of the combined effort of AmeriFlux and NEON core sites likely represent the long-term representativeness of carbon uptake dynamics across CONUS, as data from these sites is available and funding for operation is less uncertain. New representativeness analyses could be based on other ecosystem functional processes such as ecosystem CO_2 losses to the atmosphere (i.e., ecosystem respiration), energy balance, water fluxes, or dynamics of non-CO_2 gases. Finally, an aspirational goal of AmeriFlux is to provide a collaborative and networking platform for all eddy-covariance sites across the Americas. This effort has fundamental benefits because understanding of global environmental challenges is only reached through international programmatic and scientific collaborations (Vargas et al., 2012); therefore, there are open research questions for the potential representativeness of the joint efforts of all regional networks across the Americas.

2.4.5 Conclusions

We used EFTs as an alternative approach to assess the representativeness of AmeriFlux and NEON to monitor ecosystem functional heterogeneity across CONUS. This analysis complements previous studies based on climatic or vegetation structural characteristics (Hargrove et al., 2003; Yang et al., 2008), and addresses the interests for considering alternative information on ecosystem functionality (Bond-Lamberty et al. 2016; Petrakis et al. 2017; Petchey and Gaston 2006; Reichstein et al., 2014; Valentini et al., 1999; Wright et al. 2006). Throughout its 20-year history of biosphere-atmosphere flux observations, the AmeriFlux network provides representation of ecosystem functional heterogeneity for 55% of CONUS. The joint effort of AmeriFlux and NEON core sites provides a long-term opportunity for representation of ecosystem functional heterogeneity for 46% of CONUS. The historical AmeriFlux archive also provides unique information about temporal variability of ecosystem functional heterogeneity due to decadal monitoring efforts at multiple study sites. Overall, representation could be enhanced across the Desert

Southwest, Southern Rockies and Colorado Plateau, Great Basin, Northern Plains, and Central Plains of the NEON ecoclimatic domains. Most of these regions are characterized by complex terrain and therefore represent a scientific and methodological challenge to measure biosphere-atmosphere fluxes. This study provides insights for EONs design and improvement, is based on publicly available data, and is applicable to other networks around the world.

Acknowledgments

This work is part of the North American Carbon Program, and the authors thank all the contributors toward developing the AmeriFlux and NEON observational design to make this study possible. SV acknowledges Consejo Nacional de Ciencia y Tecnologia (CONACyT) for a PhD scholarship. RV, NAB, DJH acknowledge support from US Department of Agriculture (2014-67003-22070). RV and HWL acknowledge support from the National Science Foundation (EF-102980 and 1652594). DAS acknowledges support from ERDF and the Spanish MINECO (grant JC2015-00316 and project CGL2014-61610-EXP). NEON is a project sponsored by the National Science Foundation and managed under cooperative support agreement (EF-1029808) to Battelle Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our sponsoring agencies or the environmental networks. This paper would not have taken shape if it were not for meaningful engagement with community members and our mentors. We thank two anonymous reviewers and Margaret Torn for insightful comments to improve this manuscript.

REFERENCES

Ahlström, A., Raupach, M.R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J.G., Friedlingstein, P., Jain, A.K., Kato, E., Poulter, B., Sitch, S., Stocker, B.D., Viovy, N., Wang, Y.P., Wiltshire, A., Zaehle, S., Zeng, N., 2016. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science (80-.). 348, 895–899. doi:10.1002/2015JA021022

Alcaraz-Segura, D., Lomba, A., Sousa-Silva, R., Nieto-Lugilde, D., Alves, P., Georges, D., Vicente, J.R., Honrado, J.P., 2017. Potential of satellite-derived ecosystem functional attributes to anticipate species range shifts. Int. J. Appl. Earth Obs. Geoinf. 57, 86–92. doi:10.1016/j.jag.2016.12.009

Alcaraz-Segura, D., Paruelo, J.M., Epstein, H.E., Cabello, J., 2013. Environmental and human controls of ecosystem functional diversity in temperate South America. Remote Sens. 5, 127–154. doi:10.3390/rs5010127

Alcaraz, D., Paruelo, J., Cabello, J., 2006. Identification of current ecosystem functional types in the Iberian Peninsula. Glob. Ecol. Biogeogr. 15, 200–212. doi:10.1111/j.1466-822X.2006.00215.x

Allen, A., Chapman, D., 2001. Impacts of afforestation on groundwater resources and quality. Hydrogeol. J. 9, 390–400. doi:10.1007/s100400100148

AMP (2017) About AmeriFlux Core Sites. What are the AmeriFlux core sites? Retrieved from http://www.ameriflux.lbl.goc/sites/ameriflux-core-sites/

Arredondo, T., Garcia-Moya, E., Huber-Sannwald, E., Loescher, H.W., Delgado-Balbuena, J., Luna-Luna, M., 2016. Drought manipulation and its direct and legacy effects on productivity of a monodominant and mixed-species semi-arid grassland. Agric. For. Meteorol. 223, 132–140. doi:10.1016/j.agrformet.2016.03.011

Bailey, R.G., 1983. Delineation of Ecosystem Regions. Environ. Manage. 7, 365–373.

Baldocchi, D., Falge, E., Gu, L., & Olson, R., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82, 2415. doi:http://dx.doi.org/10.1108/17506200710779521

Baldocchi D., M. Reichstein, D. Papale, L. Koteen, R. Vargas, Agarwal, D., Cook, R., 2012. The Role of Trace Gas Flux\rNetworks in the Biogeosciences. Eos (Washington. DC). Vol. 93, 217–224.

Bazilian, M., Rogner, H., Howells, M., Hermann, S., Arent, D., Gielen, D., Steduto, P., Mueller, A., Komor, P., Tol, R.S.J., Yumkella, K.K., 2011. Considering the energy, water and food nexus: Towards an integrated modelling approach. Energy Policy 39, 7896–7906. doi:10.1016/j.enpol.2011.09.039

Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., Weatherall, P., 2009. Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. Mar. Geod. 32, 355–371. doi:10.1080/01490410903297766

Biederman, J.A., Scott, R.L., Goulden, M.L., Vargas, R., Litvak, M.E., Kolb, T.E., Yepez, E.A., Oechel, W.C., Blanken, P.D., Bell, T.W., Garatuza-Payan, J., Maurer, G.E., Dore, S., Burns, S.P., 2016. Terrestrial carbon balance in a drier world: The effects of water availability in southwestern North America. Glob. Chang. Biol. 22, 1867–1879. doi:10.1111/gcb.13222

Bond-Lamberty, B., Epron, D., Harden, J., Harmon, M.E., Hoffman, F., Kumar, J., McGuire, A.D., Vargas, R., 2016. Estimating Heterotrophic respiration at large scales: challenges, approaches, and next steps. Ecosphere 7, 1–13. doi:10.1002/ecs2.1380

Cabello, J., Lourenco, P., Reyes, A., Alcaraz-Segura, D., 2013. Ecosystem Services Assessment of National Parks Networks for Functional Diversity and Carbon Conservation Strategies Using Remote Sensing. Earth Obs. Ecosyst. Serv. 179–200.

Chen, B., Coops, N.C., Fu, D., Margolis, H.A., Amiro, B.D., Barr, A.G., Black, T.A., Arain, M.A., Bourque, C.P.A., Flanagan, L.B., Lafleur, P.M., McCaughey, J.H., Wofsy, S.C., 2011. Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling. Agric. For. Meteorol. 151, 87–100. doi:10.1016/j.agrformet.2010.09.005

Chen, B., Coops, N.C., Fu, D., Margolis, H.A., Amiro, B.D., Black, T.A., Arain, M.A., Barr, A.G., Bourque, C.P.A., Flanagan, L.B., Lafleur, P.M., McCaughey, J.H., Wofsy, S.C., 2012. Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program Network using remote sensing and footprint analysis. Remote Sens. Environ. 124, 742–755. doi:10.1016/j.rse.2012.06.007

Chuluun, T., Ojima, D., 2002. Land use change and carbon cycle in and and semi-arid lands of East and Central Asia. Sci. China Ser. C-Life Sci. 45, 48.

Chrysoulakis1, N., Spiliotopoulos, M., Domenikiotis, C., Dalezios3, N., 2003. Towards monitoring of regional atmospheric instability through modis/aqua images. In: Dalezios, N.R., Dobesch, H. (Eds.), Proceedings of the International Symposium on GIS and Remote Sensing: Environmental Applications, Colos, Greece, November 2009, pp. 155e166.

Ciais, P., Dolman, A.J., Bombelli, A., Duren, R., Peregon, A., Rayner, P.J., Miller, C., Gobron, N., Kinderman, G., Marland, G., Gruber, N., Chevallier, F., Andres, R.J., Balsamo, G., Bopp, L., Bréon, F.M., Broquet, G., Dargaville, R., Battin, T.J., Borges, A., Bovensmann, H., Buchwitz, M., Butler, J., Canadell, J.G., Cook, R.B., Defries, R., Engelen, R., Gurney, K.R., Heinze, C., Heimann, M., Held, A., Henry, M., Law, B., Luyssaert, S., Miller, J., Moriyama, T., Moulin, C., Myneni, R.B., Nussli, C., Obersteiner, M., Ojima, D., Pan, Y., Paris, J.D., Piao, S.L., Poulter, B., Plummer, S., Quegan, S., Raymond, P., Reichstein, M., Rivier, L., Sabine, C., Schimel, D., Tarasova, O., Valentini, R., Wang, R., Van Der Werf, G., Wickland, D., Williams, M., Zehner, C., 2014. Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. Biogeosciences 11, 3547–3602. doi:10.5194/bg-11-3547-2014

Costanza, R., d'Arge, R., de Groot, R.S., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260

Cramer, W., Bondeau, A., Woodward, F.I., Prentice, I.C., Betts, R.A., Brovkin, V., Cox, P.M., Fisher, V., Foley, J.A., Friend, A.D., Kucharik, C., Lomas, M.R., Ramankutty, N., Sitch, S., Smith, B., White, A., Young-Molling, C., 2001. Global response of terrestrial ecosystem structure and function to CO2 and climate change: Results from six dynamic global vegetation models. Glob. Chang. Biol. 7, 357–373. doi:10.1046/j.1365-2486.2001.00383.x

Díaz, S., Purvis, A., Cornelissen, J.H.C., Mace, G.M., Donoghue, M.J., Ewers, R.M., Jordano, P., Pearse, W.D., 2013. Functional traits, the phylogeny of function, and ecosystem service vulnerability. Ecol. Evol. 3, 2958–2975. doi:10.1002/ece3.601

Elith, J., Phillips, S.J., Hastie, T., Dudı, M., 2011. A statistical explanation of MaxEnt for

ecologist. Divers. Distrib. 17, 43–57. doi:10.1111/j.1472-4642.2010.00725.xFalge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., Bernhofer, C., Burba, G., Clement, R., Davis, K.J., Elbers, J.A., Goldstein, A.H., Grelle, A., Granier, A., Gumundsson, J., Hollinger, D., Kowalski, A.S., Katul, G., Law, B.E., Malhi, Y., Meyers, T., Monson, R.K., Munger, J.W., Oechel, W., Paw U, K.T., Pilegaard, K., Rannik, ??llar, Rebmann, C., Suyker, A., Valentini, R., Wilson, K., Wofsy, S., 2002. Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. Agric. For. Meteorol. 113, 53–74. doi:10.1016/S0168-1923(02)00102-8

Fernández N, Paruelo JM, Delibes M. 2010. Ecosystem functioning of protected and altered Mediterranean environments: A remote sensing classification in Doñana, Spain. Remote Sensing of Environment 114: 211–220

Fielding, A.H., Bell, J.F., 2016. A review of methods for the assessment of prediction errors in conservation presence / absence models. Environ. Conserv. 24, 38–49.

Gamon, J.A., Huemmrich, K.F., Wong, C.Y.S., Ensminger, I., Garrity, S., Hollinger, D.Y., Noormets, A., Peñuelas, J., 2016. A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. Proc. Natl. Acad. Sci. 113, 13087–13092. doi:10.1073/pnas.1606162113

Gilmanov, T.G., Tieszen, L.L., Wylie, B.K., Flanagan, L.B., Frank, A.B., Haferkamp, M.R., Meyers, T.P., Morgan, J.A., 2005. Integration of CO2 flux and remotely-sensed data for primary production and ecosystem respiration analyses in the Northern Great Plains: Potential for quantitative spatial extrapolation. Glob. Ecol. Biogeogr. 14, 271–292. doi:10.1111/j.1466-822X.2005.00151.x

Göckede, M., Rebmann, C., Foken, T., 2004. A combination of quality assessment tools for eddy covariance measurements with footprint modelling for the characterisation of complex sites. Agric. For. Meteorol. 127, 175–188. doi:10.1016/j.agrformet.2004.07.012

Graham, R.L., Nelson, R., Sheehan, J., Perlack, R.D., Wright, L.L., 2007. Current and potential U.S. corn stover supplies. Agron. J. 99, 1–11. doi:10.2134/agronj2005.0222

Ha, W., Kolb, T.E., Springer, A.E., Dore, S., O'Donnell, F.C., Martinez Morales, R., Masek Lopez, S., Koch, G.W., 2015. Evapotranspiration comparisons between eddy covariance measurements and meteorological and remote-sensing-based models in disturbed ponderosa pine forests. Ecohydrology 8, 1335–1350. doi:10.1002/eco.1586

Hargrove, W.W., Hoffman, F.M., Law, B.E., 2003. New analysis reveals representativeness of the AmeriFlux network. Eos, Trans. Am. Geophys. Union 84, 529. doi:10.1029/2003EO480001

Hayes, D.J., Turner, D.P., Stinson, G., McGuire, A.D., Wei, Y., West, T.O., Heath, L.S., de Jong, B., McConkey, B.G., Birdsey, R.A., Kurz, W.A., Jacobson, A.R., Huntzinger, D.N., Pan, Y., Post, W. Mac, Cook, R.B., 2012. Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory-based data. Glob. Chang. Biol. 18, 1282–1299. doi:10.1111/j.1365-2486.2011.02627.x

Hijmans, R.J., 2012. Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. Ecology 93, 679–688. doi:10.1890/11-0826.1

Jax, K., 2010. Ecosystem Functioning 272. doi:10.1017/CBO9780511781216

Joiner, J., Yoshida, Y., Vasilkov, A.P., Yoshida, Y., Corp, L.A., Middleton, E.M., 2011. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. Biogeosciences 8, 637–651. doi:10.5194/bg-8-637-2011

Jongman, R., Skidmore A., , Mücher C., Bunce R., Metzger., 2017. Global Terrestrial Ecosystem Observation: Why, Where, What and How?, in: Michele Walters, R.G.S. (Ed.), The GEO Handbook on Biodiversity Observataion Network. Springer International Publishing, pp. 19–38. doi:10.1007/978-3-319-27288-7

Kampe, T.U., Johnson, B.R., Kuester, M., Keller, M., 2010b. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. J. Appl. Remote Sens. 4, 43510. doi:10.1117/1.3361375

Katul, G.G., Finnigan, J.J., Poggi, D., Leuning, R., Belcher, S.E., 2006. The influence of hilly terrain on canopy-atmosphere carbon dioxide exchange. Boundary-Layer Meteorol. 118, 189–216. doi:10.1007/s10546-005-6436-2

Keller, M., Schimel, D.S., Hargrove, W.W., Hoffman, F.M., 2011. A continental Ecological strategy Observatory for the National Network. America (NY). 6, 282–284.

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 Syst. Sci. Data Discuss. 1–25. doi:10.5194/essd-2016-36

Law, B., 2005. Carbon dynamics in response to climate and disturbance: Recent progress from multi-scale measurements and modeling in AmeriFlux. Plant Responses to Air Pollut. Glob. Chang. 205–213.

Lee, S.J., Berbery, E.H., Alcaraz-Segura, D., 2013. The impact of ecosystem functional type changes on the La Plata Basin climate. Adv. Atmos. Sci. 30, 1387–1405. doi:10.1007/s00376-012-2149-x

Liu, C., White, M., Newell, G., 2011. Measuring and comparing the accuracy of species distribution models with presence-absence data. Ecography (Cop.). 34, 232–243. doi: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., Shanley, J.B., Likens, G.E., Haeuber, R., 2007. Who needs environmental monitoring ? In a nutshell : Ecol. Soc. Am. 5, 253–260. doi: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. Proc. SPIE 8181, Earth Resour. Environ. Remote Sensing/GIS Appl. II, 81811K 8181, 1–7. doi:doi:10.1117/12.898038

Luo, H., Oechel, W.C., Hastings, S.J., Zulueta, R., Qian, Y., Kwon, H., 2007. Mature semiarid chaparral ecosystems can be a significant sink for atmospheric carbon dioxide. Glob. Chang. Biol. 13, 386–396. doi:10.1111/j.1365-2486.2006.01299.x

Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J., Thuiller, W., 2013. Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. PLoS Biol. 11. doi:10.1371/journal.pbio.1001569

Milchunas, D.G., Lauenroth W.K.,Inertia in plant community structure: state changes after cessation of nutrient-enrichment stress. Ecol. Appl., 5 (1995), pp. 452-458

Nagendra, H.S. Sudhira, Madhusudan Katti, and Maria Schewenius (2013) Sub-regional Assessment of India: Effects of Urbanization on Land Use, Biodiversity and Ecosystem Services. In the book Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities: A Global Assessment.

National Research Council (NRC). 2001. Grand challenges in environmental sciences. Washington, DC: National Academy Press.

National Research Council (NRC). 2005. The owner's role in project risk management. The National Academies Press, Washington, DC

National Research Council (NRC). 2007. Earth science and applications from space: national imperatives for the next decade and beyond. The National Academies Press, Washington

Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B., Running, S.W., 2000. Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999.

Noss, R., 1990. Indicators for monitoring biodiversity: A hierarchical approach. Conserv. Biol. 4, 355–364. doi:10.1111/j.1523-1739.1990.tb00309.x

Novick, K.A., Biederman, J.A., Desai, A.R., Litvak, M.E., Moore, D.J.P., Scott, R.L., Torn, M.S., 2017. The AmeriFlux network: A coalition of the willing. Agric. For. Meteorol. doi:10.1016/j.agrformet.2017.10.009

Ort, D.R., Long, S.P., 2014. Limits on Yields in the Corn Belt. Science (80-.). 344, 484–485. doi:10.1126/science.1253884

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A large and persistent carbon sink in the world's forests. Science 333, 988–93. doi:10.1126/science.1201609

Paruelo, J.M., Epstein, H.E., Lauenroth, W.K., Burke, I.C., 1997. ANPP Estimates from NDVI for the Central Grassland Region of the United States ANPP ESTIMATES FROM NDVI FOR THE CENTRAL GRASSLAND REGION OF THE UNITED STATES 78, 953–958.

Paruelo, J.M., Jobbagy, E.G., Sala, O.E., 2001. Current distribution of ecosystem functional types in temperate South America. Ecosystems 4, 683–698. doi:10.1007/s10021-001-0037-9

Petchey, O.L., Gaston, K.J., 2006. Functional diversity: Back to basics and looking forward. Ecol. Lett. 9, 741–758. doi:10.1111/j.1461-0248.2006.00924.x

Peters, D., Loescher, H., Al, P.E.T., 2014. Taking the pulse of a continent : expanding site-based research infrastructure for regional- to continental-scale ecology. Ecosphere 5, 1–23. doi:http://dx.doi.org/10.1890/ES13-00295.1

Petrakis, S., Barba, J., Bond-Lamberty, B., Vargas, R., 2017. Using greenhouse gas fluxes to define soil functional types. Plant Soil 868–874. doi:10.1007/s10745-006-9094-1

Pettorelli, N., Wegmann, M., Skidmore, A., Mücher, S., Dawson, T.P., Fernandez, M., Lucas, R., Schaepman, M.E., Wang, T., O'Connor, B., Jongman, R.H.G., Kempeneers, P., Sonnenschein, R., Leidner, A.K., Böhm, M., He, K.S., Nagendra, H., Dubois, G., Fatoyinbo, T., Hansen, M.C., Paganini, M., de Klerk, H.M., Asner, G.P., Kerr, J.T., Estes, A.B., Schmeller, D.S., Heiden, U., Rocchini, D., Pereira, H.M., Turak, E., Fernandez, N., Lausch, A., Cho, M.A., Alcaraz-Segura, D., McGeoch, M.A., Turner, W., Mueller, A., St-Louis, V., Penner, J., Vihervaara, P., Belward, A., Reyers, B., Geller, G.N., 2016. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. Remote Sens. Ecol. Conserv. 2, 122–131. doi:10.1002/rse2.15

Phillips, S., Dudík, M., Schapire, R., 2004. A maximum entropy approach to species distribution modeling. Proc. twenty-first … 655–662. doi:10.1145/1015330.1015412

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., Liu, R., Mao, J., Myneni, R.B., Peng, S., Poulter, B., Shi, X., Xiao, Z., Zeng, N., Zeng, Z., Wang, Y., 2015. Detection and attribution of vegetation greening trend in China over the last 30 years. Glob. Chang. Biol. 21, 1601–1609. doi:10.1111/gcb.12795

Poulter, B., Frank, D., Ciais, P., Myneni, R.B., Andela, N., Bi, J., Broquet, G., Canadell, J.G., Chevallier, F., Liu, Y.Y., Running, S.W., Sitch, S., van der Werf, G.R., 2014. Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. Nature 509, 600–603. doi:10.1038/nature13376

Rahman, A.F., Sims, D.A., Cordova, V.D., El-Masri, B.Z., 2005. Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem C fluxes. Geophys. Res. Lett. 32, 1–4. doi:10.1029/2005GL024127

Reichstein, M., Bahn, M., Mahecha, M.D., Kattge, J., Baldocchi, D.D., 2014. Linking plant and ecosystem functional biogeography. Proc. Natl. Acad. Sci. 111. doi:10.1073/pnas.1216065111

Running, S.W., Baldocchi, D.D., Turner, D.P., Gower, S.T., Bakwin, P.S., Hibbard, K.A., 1999. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. Remote Sens. Environ. 70, 108–127. doi:10.1016/S0034-4257(99)00061-9

Schimel, D., Hargrove, W., Hoffman, F., MacMahon, J., 2007. NEON: A hierarchically designed national ecological network. Front. Ecol. Environ. 5, 59. doi:10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2

Schimel, D., Keller, M., 2015. Big questions, big science: meeting the challenges of global ecology. Oecologia 177, 925–934. doi:10.1007/s00442-015-3236-3

Schimel, D., Kittel, T.G.F., Running, S., Monson, R., Turnipseed, A., Anderson, D., 2002. Carbon Sequestration Studied in Western U.S. Mountains. Eos (Washington. DC). 83, 445–456. doi:10.1029/2002EO000314

Scholes R., Gill M., Costello M., Sarantakos G., Walters M., 2017. Working in Networks to Make Biodiversity Data More Available in: Michele Walters, R.G.S. (Ed.), The GEO Handbook on Biodiversity Observataion Network. Springer International Publishing, pp. 19–38. doi:10.1007/978-3-319-27288-7

Schwinning, S., Sala, O.E., Loik, M.E., Ehleringer, J.R., 2004. Thresholds, memory, and seasonality: Understanding pulse dynamics in arid/semi-arid ecosystems. Oecologia 141, 191–193. doi:10.1007/s00442-004-1683-3

Sims, D.A., Brzostek, E.R., Rahman, A.F., Dragoni, D., Phillips, R.P., 2014. An improved approach for remotely sensing water stress impacts on forest C uptake. Glob. Chang. Biol. 20, 2856–2866. doi:10.1111/gcb.12537

Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L., Monson, R.K., Oechel, W.C., Schmid, H.P., Wofsy, S.C., Xu, L., 2006. On the use of MODIS EVI to assess gross primary productivity of North American ecosystems. J. Geophys. Res. Biogeosciences 111, 1–16. doi:10.1029/2006JG000162

Sulkava, M., Luyssaert, S., Zaehle, S., Papale, D., 2011. Assessing and improving the representativeness of monitoring networks: The European flux tower network example. J. Geophys. Res. Biogeosciences 116, 1–14. doi:10.1029/2010JG001562

Swanson, F.J., Kratz, T.K., Caine, N., Woodmansee, R.G., 1988. Landform Effects on Ecosystem Patterns and Processes. Bioscience 38, 92–98. doi:10.2307/1310614

Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., Green, R.O., 2004. Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. Bioscience 54, 523. doi:10.1641/0006-3568(2004)054[0523:UISTSE]2.0.CO;2

Valentini, R., Baldocchi, D.D. & Tenhunen, J.D. (1999) Ecological controls on land–surface atmospheric interactions. Integrating hydrology, ecosystem dynamics and biogeochemistry in complex landscapes (ed. by J.D.Tenhunen and P.Kabat), pp. 105–116. John Wiley & Sons, Berlin.

Vargas, R., Alcaraz-Segura, D., Birdsey, R., Brunsell, N.A., Cruz-Gaistardo, C.O., de Jong, B., Etchevers, J., Guevara, M., Hayes, D.J., Johnson, K., Loescher, H.W., Paz, F., Ryu, Y., Sanchez-Mejia, Z., Toledo-Gutierrez, K.P., 2017. Enhancing interoperability to facilitate implementation of REDD+: case study of Mexico. Carbon Manag. 8, 57–65. doi:10.1080/17583004.2017.1285177

Vargas, R., Baldocchi, D.D., Allen, M.F., Bahn, M., Black, T.A., Collins, S.L., Yuste, J.C., Hirano, T., Jassal, R.S., Pumpanen, J., Tang, J., 2010. Looking deeper into the soil: Biophysical controls and seasonal lags of soil CO2 production and efflux. Ecol. Appl. 20, 1569–1582. doi:10.1890/09-0693.1

Vargas, R., Loescher, H.W., Arredondo, T., Huber-Sannwald, E., Lara-Lara, R., Yépez, E. a., 2012. Opportunities for advancing carbon cycle science in Mexico: Toward a continental scale understanding. Environ. Sci. Policy 21, 84–93. doi:10.1016/j.envsci.2012.04.003

Vargas, R., Sonnentag, O., Abramowitz, G., Carrara, A., Chen, J.M., Ciais, P., Correia, A., Keenan, T.F., Kobayashi, H., Ourcival, J.M., Papale, D., Pearson, D., Pereira, J.S., Piao, S., Rambal, S., Baldocchi, D.D., 2013. Drought Influences the Accuracy of Simulated Ecosystem Fluxes: A Model-Data Meta-analysis for Mediterranean Oak Woodlands. Ecosystems 16, 749–764. doi:10.1007/s10021-013-9648-1

Vaughan, H., Brydges, T., Fenech, A., Lumb, A., 2001. Monitoring long-term ecological changes through the ecological monitoring and assessment network: Science-based and policy relevant. Environ. Monit. Assess. 67, 3–28. doi:10.1023/A:1006423432114

Vaz, A.S., Marcos, B., Goncalves, J., Monteiro, A., Alves, P., Civantos, E., Lucas, R., Mairota, P., Garcia-Robles, J., Alonso, J., Blonda, P., Lomba, A., Honrado, J.P., 2015. Can we predict habitat quality from space? A multi-indicator assessment based on an automated knowledge-driven system. Int. J. Appl. Earth Obs. Geoinf. 37, 106–113. doi:10.1016/j.jag.2014.10.014

Vos, P., Meelis, E., W., T.K., 2000. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. Environ. Monit. Assess. 61, 317–344.

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. J. Geophys. Res. Biogeosciences 113, 1–11. doi:10.1029/2007JG000627

Wright JP, Naeem S, Hector A, Lehman C, Reich PB, Schmid B, Tilman D. 2006. Conventional functional classification schemes underestimate the relationship with ecosystem functioning. Ecology Letters, 9(2), 111-120.

Zhang, L., Wylie, B.K., Ji, L., Gilmanov, T.G., Tieszen, L.L., 2010. Climate-Driven Interannual Variability in Net Ecosystem Exchange in the Northern Great Plains Grasslands. Rangel. Ecol. Manag. 63, 40–50.

Zscheischler, J., Fatichi, S., Wolf, S., Blanken, P.D., Bohrer, G., Keenan, T.F., Clark, K., Desai, A.R., Hollinger, D., Keenan, T., Novick, K.A., Seneviratne, S.I., 2016. Short-term favorable weather conditions are an important control of interannual variability in carbon and water fluxes. J. Geophys. Res. Biogeoscience. doi:10.1002/2016JG003503

Chapter 3

OPTIMIZING AN ENVIRONMENTAL OBSERVATORY NETWORK DESIGN USING PUBLICLY AVAILABLE DATA

List of Authors:

Samuel Villarreal[1], Mario Guevara[1], Domingo Alcaraz-Segura[2,3,4], Rodrigo Vargas[1].

Author affiliations:

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

[2]University of Granada, Department of Botany. Av. de Fuentenueva, s/n. 18701. Granada, Spain.

[3]Andalusian Center for the Assessment and Monitoring of Global Change (CAESCG) – University of Almería, Crta. San Urbano, 04120, Almería, Spain.

[4]iecolab. Interuniversitary Institute for Earth System Research (IISTA) – University of Granada, Av. del Mediterráneo, 18006, Granada, Spain.

Key Points:

- Alternative approach to inform the design of an environmental observatory network
- This framework provides insights about network representativeness
- This study identified spatial information gaps across a megadiverse country

Abstract

There is an increasing need to optimize resources for large-scale environmental monitoring efforts, especially in developing countries. Here, we test a flexible framework to optimize the design of an environmental observatory network (EON) using publicly available data for Mexico. This country represents a challenge for designing EONs because of its megadiversity and large climate and ecological heterogeneity. We address three pervasive challenges for designing EONs: 1) How to characterize and delineate ecologically similar areas; 2) How to set geographic priorities to establish new representative study sites; and 3) How to assess the representativeness of current and potential new study sites. We used unsupervised classification methods (i.e., factorial and cluster analysis) to spatially delineate ecologically similar sampling domains. Then, we identified the most representative sites within each domain using a conditioned Latin Hypercube-based sampling strategy. Finally, we demonstrated the applicability of this approach by assessing the spatial representativeness of the eddy covariance network in Mexico (i.e., MexFlux). We found that at least 84 distributed sampling sites are needed to represent more than 45% of the spatial heterogeneity of gross primary productivity (GPP) and evapotranspiration (ET) at the national-level. The current array of MexFlux only represents 3% of GPP and 5% of ET spatial variability at the national-level, while the same number of sites organized under an optimal framework nearly doubled these estimates. We conclude that our framework is an alternative approach to identify spatial information gaps and to guide EONs design. It is based on a data-driven approach and publicly available sources of information, so it could be applied anywhere in the world.

3.1 Introduction

Monitoring global environmental changes (e.g., land cover and land use, chemical composition of the atmosphere, climate) is crucial to guide mitigation and adaptation policies across the world (Lovett et al., 2007; Menoni et al., 2012). This has motivated the scientific community to study ecosystem processes and ecological interactions at multiple spatial and temporal scales. Addressing these grand challenges requires a coordinated effort of environmental monitoring of multiple physical, chemical and biological variables (Lovett et al., 2007). Such efforts are usually coordinated by environmental observatory networks (EONs), which can be defined as organizations whose members are affiliated in a flexible way to share procedures and to optimize the network design by avoiding duplicated measurements or redundant study sites (Sulkava et al., 2011). Thus, there is a pressing need to propose alternative scientific approaches to design representative EONs and optimize their performance across the world.

An optimized design for an EON is largely constrained by the number and spatial configuration of the study sites. The question of where to establish new study sites requires a "sampling frame" that strategically chooses locations that are representative of similar ecological areas according to specific scientific questions (e.g., similar areas in terms of ecosystem functioning related to gross primary production (GPP) and evapotranspiration (ET) dynamics) at a spatial scale of interest (Jongman et al., 2017). Ideally, these ecologically similar areas (or domains) should be identified based on environmental patterns that drive the ecological processes of interest according to an explicit ecological conceptual model (Metzger et al., 2005, 2013). For example, the National Ecological Observatory Network (NEON) in the United States defined sampling domains based on bioclimatic predictors, and for each domain identified strategic locations that better represent the bioclimatic characteristics of the domain (Keller et al., 2008; Schimel et al., 2007). Ultimately, a network design should also consider how to aggregate observations and how to prioritize new sites to maximize its effectiveness towards specific monitoring goals (Jongman et al., 2017; Lovett et al., 2007; Scholes et al., 2012; Villarreal et al., 2018). The selection of study sites depends on whether EONs are organized under a "bottom-up" or a "top-down" framework. A "bottom-up" framework is characterized by aggregating individual study-sites based on principal investigators interests within the network (Scholes et al., 2012). This devolved framework allows for a flexible coordination of monitoring efforts that could embrace a wider range of applications; however, it could have a low degree of interoperability that may limit EONs development (Vargas et al., 2017). A "top-down" framework is based on a more centrally-directed effort, which requires clearly defined common goals for different research groups. Its implementation requires a high degree of interoperability (Vargas et al., 2017) and its application range may be limited depending on the EONs' goals. Both approaches are helpful to address complex and interconnected issues that require efforts that cannot be undertaken by individual groups working alone. We highlight that "bottom-up" initiatives can also be supported by a "top-down" framework, which can provide guidelines towards an optimal network design, a high degree of interoperability, and wider applications (Holzer et al., 2018).

FLUXNET and associated regional networks (e.g., AmeriFlux, AsiaFlux, MexFlux) are monitoring net ecosystem exchange (NEE) and evapotranspiration (ET) worldwide, and have improved our understanding of terrestrial biogeochemistry and land-atmosphere interactions (Baldocchi et al., 2001). Most of these EONs are organized under a "bottom-up" approach, which is prone to over/under-represent certain ecosystems, land cover types or specific ecological conditions (Hargrove et al., 2003; Kumar et al., 2016; Sulkava et al., 2011; Villarreal et al., 2018). The under-representativeness of these EONs is also evident among countries, where developed countries are better represented than developing countries (Kumar et al., 2016). Under-represented regions hamper our understanding of land-atmosphere interactions at local, regional and global scales. For example, synthesis activities of carbon cycling across North America are mainly based on information collected from United States and Canada, being Mexico the most under-represented country (Hayes et al., 2012; Huntzinger et al., 2012; King et al., 2015). However, Mexico is a megadiverse country and has large climate and topographic heterogeneity (Mittermeier & Mittermeier, 1992), which influences ecosystem carbon dynamics in different ways than in the rest of North America (Hayes et al., 2012; King et al., 2015) (Hayes et al., 2012; King et al., 2015). The relevance of improving the representativeness of carbon and water fluxes study sites across Mexico is an opportunity to build a more comprehensive understanding of these biogeochemical cycles at a full continental-scale (Vargas et al., 2012, 2013).

Our goal in this study is to provide a framework using publicly available data to identify potential sampling locations or study sites that could maximize the representativeness of an EON to capture targeted ecosystem processes (i.e., carbon and water fluxes) at the national-level. The proposed framework addresses three main issues of EONs: 1) How to characterize and delineate ecologically similar areas; 2) How to set geographic priorities to establish new representative study sites; and 3) How to assess the representativeness of current and potential new study sites.

This framework is conceptually applicable to any country/region around the world. We tested this framework across Mexico using the regional eddy covariance network as reference (i.e., MexFlux; Vargas et al., 2013). The information from this study provides insights for interpretation of synthesis studies derived from current available sites, and guidance for network development using "top-down" information. This framework is based on publicly available information and it can provide insights towards evaluation of current EONs' design and future optimizations worldwide.

3.2 Material and Methods

This study presents a stepwise data-driven approach to provide insights to improve an EON design for monitoring national-scale carbon (i.e., GPP monitoring) and water (i.e., ET monitoring) fluxes. We organize a workflow considering three main tasks with subsequent steps (Table 1): Task A (Table 1, steps 1-4): provides the ecological conceptual framework (i.e., state system model) where environmental covariates that act as drivers of the target ecosystem fluxes (i.e., GPP and ET) were pre-selected to subsequently select only those variables with a spatial distribution pattern; Task B (Table 1, step 5-6): delineation of environmental similar areas (i.e., ESA's) based on the previously selected variables to be use as sampling domains, and selection of potential study sites that represents the most dominant environmental characteristics of each ESA; Task C (Table 1, step 7) assess the representativeness of those potential study sites and the current MexFlux sites to depict GPP and ET dynamics at the national-level using a machine-learning approach (i.e., random forest). We tested this approach across Mexico aiming for improvement in the spatial representativeness for monitoring GPP and ET across highly diverse ecological landscapes. In the following section, we explain the details and statistical techniques for this workflow.

Table 3.1. Workflow of the proposed framework to inform the design of an Environmental Observatory Network (EON). The framework is divided into three main tasks with subsequent steps: A) data selection and preparation (steps 1 to 4); B) sampling framework (steps 5 and 6); and C) network representativeness analysis (step 7).

Workflow			
Task A	Step 1	State system model	To select candidate environmental variables that could be drivers of the targeted ecosystem processes, e.g., gross primary productivity (GPP) and evapotranspiration (ET). These environmental variables are related to climate, topography, parent material, soil properties, ecosystem functional attributes, ecosystem disturbances, etc.
	Step 2	Variable selection	To maximize the spatial variance accounted by different candidate environmental variables by selecting them based on their spatial patterns and structure. We selected variables with a nugget/sill > 0.75 using information derived from semivariograms.
	Step 3	Data harmonization and standardization	To make variables comparable by harmonizing the final set of selected variables (from Step 2): a) into a similar spatial resolution (e.g., 0.05º); and b) by standardizing values (i.e., $-1 < \sigma < 1$, $\mu=0$).
	Step 4	Data aggregation	To avoid redundancy in the final dataset by linearly combining all variables using principal component analysis (PCA).

Table 3.1 Continued

	Workflow		
Task B	Step 5	Delineation and validation of sampling domains (i.e., Ecologically Similar Areas)	To identify a hierarchy of Ecologically Similar Areas (ESAs) by: a) an initial delineation (e.g., using K-means clustering) and validation (e.g., using Silhouette Analysis; i.e., 7 general ESAs in this study); and b) sub-division of general ESAs into subclasses (e.g., using K-means clustering and iterative Silhouette Analysis; 28 sub-ESA in this study).
	Step 6	Selection of potential study-sites	To select potential study-sites by using, e.g., a constrained Latin Hypercube (cHLS) analysis. In this study we select: a) 1 for each general ESA; b) 1 for each sub-ESA; c) 3 for each sub-ESA.
Task C	Step 7	Representativeness Assessment	To perform a representativeness analysis of potential study-sites (identified in step 6) to monitor targeted ecosystem processes (i.e., GPP and ET) by using: a) Random Forest Analysis; and by b) describing the representation of the ecosystem functional types (EFTs) monitored by these study-sites.

3.2.1 State system model and variable selection (steps 1 & 2)

Based on previous studies (Amundson R., 1991; Chapin et al., 2002), it is assumed that GPP and ET dynamics are mostly influenced by abiotic factors (e.g., climate, topography, soil resources supply) and major functional traits (e.g., plant and ecosystem functional types). In step 1 (table 1), we characterized climate properties based on 19 bioclimatic predictors (Hijmans et al., 2005). To characterize topography and parent material, we selected 15 environmental predictors from global data sharing initiatives such as www.worldgrids.org (Table B-1). To characterize the spatial heterogeneity of soil resource supply, we selected soil organic carbon (SOC) and the age of the most superficial rock strata (i.e., parent material; (Anderson 1988; Jobbágy

& Jackson, 2000)). Table B-1 provides a full list of all the pre-selected and selected variables.

Ecosystem functional traits can be represented by a land surface classification known as Ecosystem Functional Types (EFTs), conceptually defined as patches of the land surface that exchange matter and energy with the atmosphere in a similar way (Alcaraz-Segura et al., 2006, 2013; Paruelo et al., 2001). In practice, EFTs are a yearly varying land surface classification based on satellite-derived key ecosystem functional attributes (i.e., descriptors of primary productivity, seasonality and phenology of carbon gains) (Alcaraz-Segura et al., 2017; Lee et al., 2013). These attributes are obtained from annual curves of spectral vegetation indices. In this study, we used the Enhanced Vegetation Index (EVI) annual curve (i.e., MODIS-EVI MOD13C2 product for the 2001-2014 period) and derived: a) the EVI annual mean (EVI_Mean) as surrogate of primary production; b) the EVI seasonal coefficient of variation (EVI_sCV) as descriptor of seasonality; and c) the month of the EVI annual maximum (MMAX) as an indicator of phenology. The range of values of each EVI metric was divided into four intervals, giving a potential number of 64 EFTs (i.e., 4 x 4 x 4 = 64; (Alcaraz-Segura et al., 2013)). A detail explanation on the computation of each EVI metric can be found in previous studies (Alcaraz-Segura et al., 2013, 2017). In addition, we computed the number of unique EFT categories that occurred in each pixel between 2001 and 2014 to assess the inter-annual variability of EFTs (EFT_{int}). The same technique was used to estimate inter-annual variability in land cover categories (LC_{int}; MODIS MCD12C1 product). Both, EFT_{int} and LC_{int} were used as descriptors of variability in ecosystems functioning (EFTs) and structure (land cover).

In step 2 (table 3.1), data selection was based on semi-variograms, where we selected the variables with medium to strong spatial autocorrelation and excluded those variables with a nugget/sill ratio equal or higher than 0.75. In this case, it is assumed that those variables do not capture differences in spatial environmental features at distinct spatial locations and are more likely to be randomly distributed across our region of interest (Cambardella et al., 1994; Cruz-Cárdenas et al., 2014). The final number of selected variables was 27 (Table C-1).

3.2.2 Data harmonization, standardization, and aggregation (steps 3 and 4)

In step 3 (Table 3.1), data harmonization consisted in resampling using bilinear interpolation the 27 selected variables to the same spatial-resolution (i.e., 0.05^{o}) and set into the same geographical projection (WGS84). This spatial resolution is largely used to represent environmental patterns at national-level (Chrysoulakis et al., 2003; Löw et al., 2011; Villarreal et al., 2018). In step 4 (Table 1), the selected variables were standardized by centering the mean to 0 and scaling the standard deviation within -1 to 1 range.

3.2.3 Delineation of ecologically similar areas and selection of potential sampling sites (steps 5 and 6)

In step 5 (Table 3.1), the standardized variables were orthogonally decomposed by using principal component analysis (PCA) to avoid variable redundancy in the subsequent cluster analysis. Then, we delineated ecologically similar areas (ESA) to define spatial sampling domains (Hargrove & Hoffman, 2005; Metzger et al., 2005). ESA delineation was performed by applying hierarchical K-means cluster partitioning using all the principal components.

Since K-means requires to previously defined number of clusters, we characterized the spatial heterogeneity of our environmental dataset using the number of the distinct hierarchical ecoregions previously defined and proposed by the CEC (*Commission for Environmental Cooperation*) and CONABIO (*National Commission for the Knowledge and use of Biodiversity*) for North America and Mexico (CEC, 1997). The first hierarchical level corresponds to the broader ecoregions across North America, while the second and third hierarchical levels are further subdivisions of those broader ecoregions. The fourth hierarchical level is the subdivision of the third level and is only defined for Mexico. The number of ecoregions found in Mexico within the distinct hierarchical levels are 7, 21, 51 and 141, where we initially prescribed those ecoregions (i.e., ESAs) into our K-means analysis. Subsequently, we assessed which of those number of ESAs provided a better cluster tightness (i.e., cohesion or object similarity within the same cluster) and separation (i.e., dissimilarity among clusters) using a Silhouette analysis and the Silhouette index (Si; (Rousseeuw, 1987)). Based on these results, we identified 7 ESAs to characterize the spatial heterogeneity of our environmental dataset at the national scale. Following the hierarchical approach of the CEC and CONABIO, we subdivided these 7 general ESAs to further separate their internal environmental heterogeneity. We applied independent K-means for each one of the ESAs and tested the results using a Silhouette analysis through an iterative process. This iterative process started with a number of groups (n) equal to 2 and stopped when the Si for n groups was higher than the Si for $n+1$ groups.

In step 6 (Table 3.1), we selected potential sampling locations using the conditioned Latin Hypercube sampling technique (cLHS). cLHS is a stratified random procedure that samples a dataset (i.e., principal components in our study) from their multivariate distribution by forming a Latin Hypercube, where the multivariate distribution of the dataset is maximally stratified (Minasny & McBratney, 2006). This provides a full coverage of the range of each variable and it serves as an efficient sampling strategy (Minasny & McBratney, 2006). We selected sites according to the following guidelines: a) selection of one potential sampling site for each one of the 7 main ESAs; b) selection of two potential sampling sites for each one of the 7 ESAs as a way to test an ideal distribution of 14 sample sites (the current number of MexFlux sites is 14); c) selection of one site for each sub-ESA to account for internal ESA heterogeneity; and d) selection of three sites for each sub-ESA to potentially maximize representativeness within each sub-ESA.

3.2.4 Representativeness of proposed and actual study sites (step 7)

In step 7 (Table 3.1). We assessed the representativeness of the proposed study sites and of the current distribution of MexFlux sites for monitoring GPP and ET based on a random forest model (RF) used for species distribution (SDM). It is recommended to apply RF when there are few observations over a broad region (Evans et al 2011; Cutler et al 2007). RF creates nodes that are organized into pairs which forms branches of classification trees (CT), then, RF through "bootstrap aggregation" predicts multiple CT and average them to create the model output (Cutler et al., 2007; Evans et al., 2011). 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 the potential study sites or the current MexFlux sites.

First, RF was used to assess the spatial representativeness of GPP and ET characterized by their annual mean and annual coefficient of variation (GPP_mean, ET_mean, GPP_cv and ET_cv), as these metrics are known to capture most of the variability of the seasonal dynamics in a time-series of vegetation indexes derived from satellite information. The annual mean and annual coefficient of variation of GPP and ET and are used as surrogate of productivity and seasonality, respectively (Alcaraz-Segura et al., 2006, 2009, 2013). Second, we assessed the representativeness of MexFlux and the potential study sites to monitor the diversity of EFT categories. This was performed by identifying the distinct EFT categories occurring at each sampling sites and by summing their frequency of occurrence. For all cases, the model used was repeated 5 times and 5000 absence points were randomly selected as recommended (Barbet-Massin et al., 2012).

3.3 Results

3.3.1 Data selection and cluster partitioning

After selecting the initial data, the variables discarded and selected for further analysis are listed in Table B-1 (task A steps 1-3, Table 3.1). From the variables selected, the first two axes of the PCA (task A step 4, Table 1) captured 35% and 19%, respectively. PC1 had higher correlation with bioclimatic variables (top 5 ranging from 0.23 to 0.27) while PC2 was correlated with soil organic carbon (SOC), bioclimatic predictors and topographic variables (top 5 ranging from 0.28 to 0.45; Table C-1)**.**

ESA delineation (i.e., cluster partitioning) and validation (i.e., cluster tightness and separation; task B steps 5, Table 3.1) was consistent with the seven groups previously defined by CEC and CONABIO on its broader classification level (Table D-1; Figure 3.1). Based on S(i) results and using the sub-ESA delineation, we selected: a) six clusters for ESA1 and ESA2, respectively; b) three clusters for each ESA3, ESA5 and ESA7; and c) four clusters for ESA4 and two for ESA6 (Table D-1).

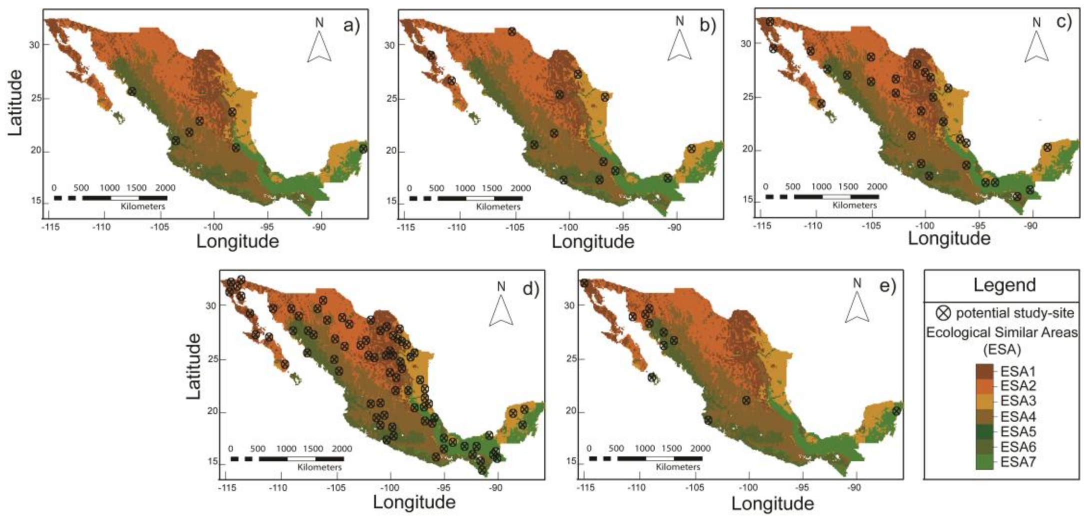


Figure 3.1. Spatial distribution of the seven general ecological similar areas (ESAs) for Mexico along with the spatial location of the proposed study sites (increasing number a to d): a) 7 general ESAs and 7 proposed study sites; b) 7 general ESAs and 14 proposed study sites; c) 7 general ESAs and 28 proposed study sites; d) 7 general ESAs and 84 proposed study sites and; d) 7 general ESAs and the distribution of the 14 current MexFlux sites.

ESAs showed environmental differences among each other (Table 3.2). For example, ESA1 and ESA2 (located at the northern part of Mexico) were characterized by low annual temperature and the lowest annual total precipitation, and whose dominant EFTs were characterized by low-productive and low-seasonal ecosystems with a maximum greening during autumn (Ad3 and Ac3, respectively; Figures 3.1 and 3.2). These two ESAs had the highest EFT_{int} (Table 3.2). In contrast, ESA7 (which was the warmest and wettest) was associated with high-productive and low-seasonal ecosystems with a maximum greening during summer (Dd2) and the lowest EFT_{int}

among all ESAs (Figures 3.1 and 3.2, Table 3.2). The spatial comparison among ESAs and the first hierarchical ecoregions proposed by CEC and CONABIO showed spatial matching between ESA1 and ESA2 with Deserts of North America and Southern Semiarid Highlands, while ESA7 spatially matched the Tropical Humid Forests (Table 3.2).

Table 3.2. Spatial similarity between the proposed ecologically similar areas (ESAs) and the CONABIO-CEC ecoregions (see methods for details). Similarity is expressed as the percentage of coincident surface area. Mean temperature, total annual precipitation, most dominant Ecosystem Functional Type (EFT) and the spatial mean of the inter-annual variability in EFT (EFT_{int}) for each of the 7 general ESAs is reported.

Ecologically Similar Area (ESA)	**Dominant Ecoregion**	**Temperature (ºC)**	**Precipitation (mm)**	**Dominant EFT**	**EFT_{int}**
ESA1	Deserts of North America (88%)	19	333	Ad3	6
ESA2	Deserts of North America (63%) & Southern Semiarid Highlands (32%)	18	340	Ac3	6
ESA3	Great Plains (52%) & Tropical Humid Forest (37%)	24	856	Dc2	6
ESA4	*Temperate Sierras* (62%) & Southern Semiarid Highlands (28%)	16	774	Ca2	4
ESA5	*Temperate Sierras* (62%) & Southern Semiarid Highlands (28%)	23	679	Ca2	5
ESA6	Tropical Dry Forest (73%)	24	937	Ca2	4
ESA7	Tropical Humid Forest (92%)	24	1866	Dd2	4

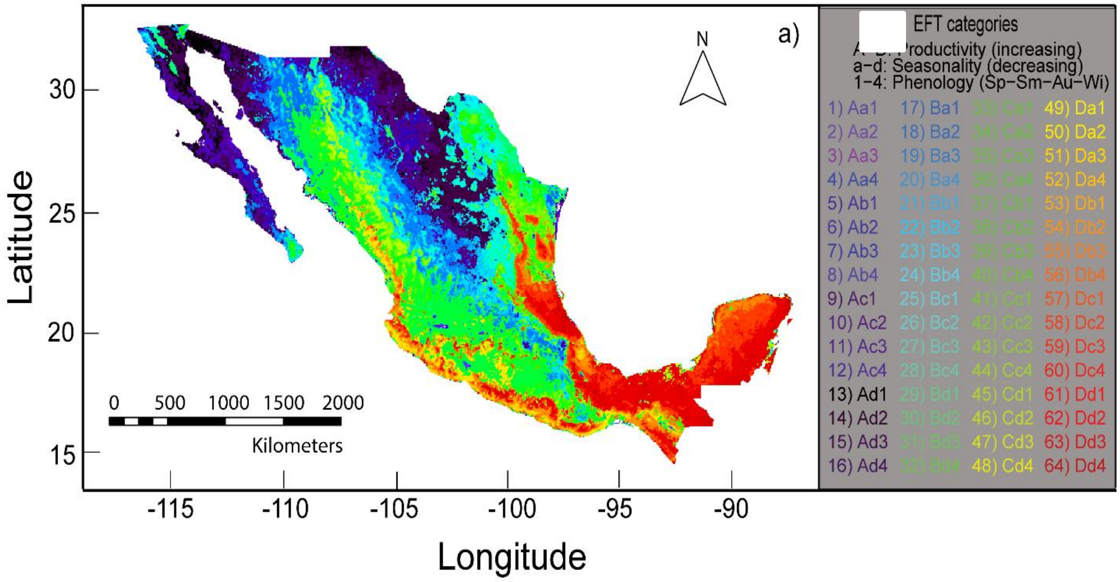


Figure 3.2. Spatial distribution of carbon uptake patterns as expressed by Ecosystem Functional Types (EFTs) across Mexico. The map shows the dominant EFT for the 2001-2014. EFTs were identified based on Moderate Resolution Imaging Spectroradiometer Enhanced Vegetation Index (MODIS-EVI) dynamics (0.05° pixel). Capital letters correspond to the EVI annual mean (EVI_Mean) level, ranging from A to D for low to high productivity. Small letters show the seasonal coefficient of variation (EVI_sCV), ranging from a to d for high to low seasonality of carbon gains. The numbers indicate the season of maximum EVI (MMAX): (1) spring (Sp), (2) summer (Sm), (3) autumn (Au), (4) winter (Wi).

3.3.2 Informing site selection and representativeness evaluation

he points selected as potential sampling sites correspond to locations with the most common environmental parameters (i.e., where the principal components spatially converge (Figure 1 and Figure 3, task b step 6, Table 1). Using this approach, we identified locations (i.e., study sites) within each ESAs ranging from 7 sites to 84 sites (Figure 1a-d). Furthermore, Figure 3a and 3b show the pair-distribution of PC1 vs PC2 and the distribution of the locations selected for each ESA (*n=7*) that correspond to the most frequent values of PC1 and PC2. The distribution range of the selected locations increased as each of the sub-ESA were sampled (from *n=14* to *n=84*, Figure 3 b-d), which allowed to represent a wider range of environmental characteristics (Figure 3). Our results show that the multivariate space of environmental variables across Mexico is difficult to represent even with 84 potential study sites (Figure 3d).

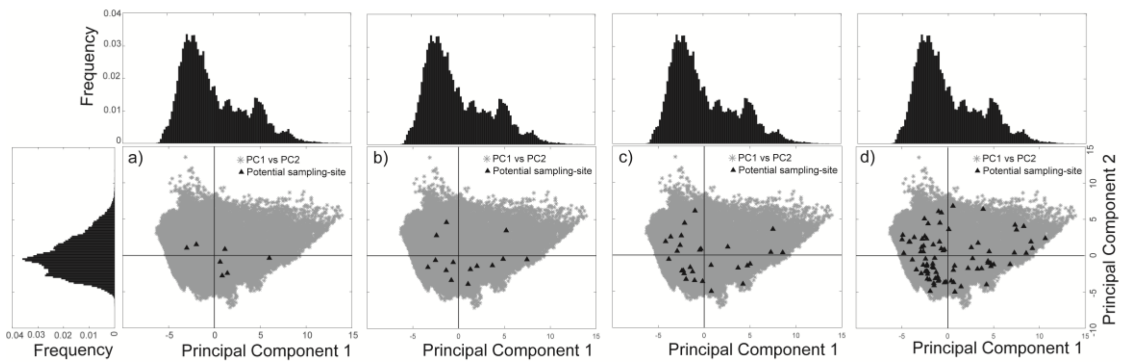


Figure 3.3. Distribution of proposed study sites across the multivariate space represented by the first two principal components (PC1, PC2) of a principal component analysis: a) distribution of 7 proposed study sites; b) distribution of 14 proposed study sites and c) distribution of 28 proposed study sites; d) distribution of 84 proposed study sites.

We assessed the representativeness of the potential study sites and current MexFlux sites using a RF approach (task C step 7, Table 3.1). Overall, the representativeness of GPP and ET was relatively similar for the potential study sites and increased with a higher number of sites (Table 3.3). Also, the representativeness of the potential study sites was in general higher than the MexFlux sites (Table 3.3). The spatial distribution of the represented/non-represented areas between GPP and ET was similar according to the same number of potential sampling sites, and it was the same for the MexFlux sites (Figure 3.1, 3.4-3.5).

Table 3.3. Spatial representativeness of gross primary productivity (GPP), evapotranspiration (ET) and Ecosystem Functional Types (EFT) for: 7, 14, 28 and 84 potential study sites, and the 14 current MexFlux sites. Percentages refer to the area of Mexico whose GPP and ET would be represented by the corresponding number and configuration of sites. EFT representativeness is reported as the number of categories represented divided by the total number of categories (i.e., 64), and by the frequency sum as percentage of all the distinct categories monitored.

	Proposed Sites (n=7)	**Proposed Sites (n=14)**	**Proposed Sites (n=28)**	**Proposed Sites (n=84)**	**MexFlux Sites (n=14)**
GPP	4%	8%	20%	45%	3%
ET	4%	8%	18%	49%	5%
EFT	7/64 (35%)	13/64(47%)	16/64(61%)	31/64(91%)	8/64(32%)

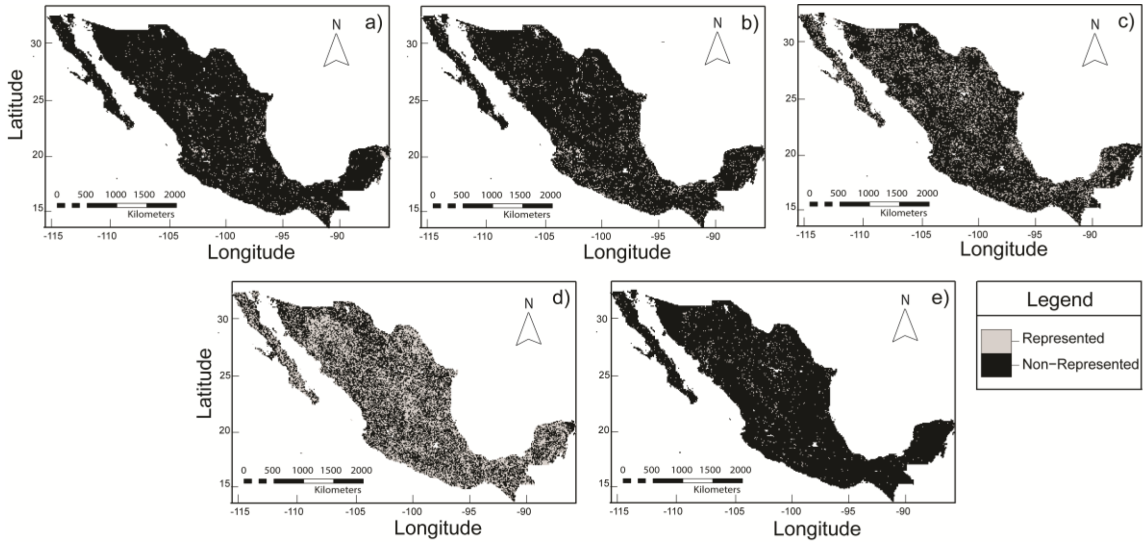


Figure 3.4. Spatial representativeness of gross primary productivity (GPP) based on a Random Forest approach for: a) 7 proposed study sites, b) 14 proposed study-sites, c) 28 proposed study-sites, d) 84 proposed study-sites; and e) the 14 current MexFlux sites. Black areas indicate regions not represented and grey areas represented regions by each scenario of study sites.

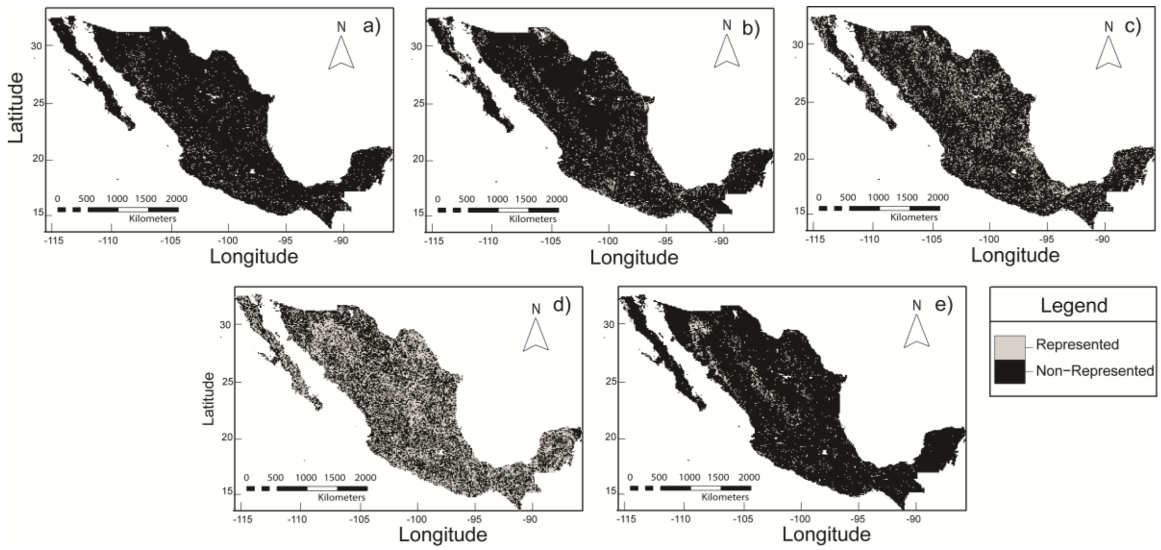


Figure 3.5. Spatial representativeness of evapotranspiration (ET) based on a Random Forest approach for: a) 7 proposed study sites, b) 14 proposed study-sites, c) 28 proposed study-sites, d) 84 proposed study-sites; and e) the 14 current MexFlux sites. Black areas indicate regions not represented and grey areas represented regions by each scenario of study sites.

The representativeness of EFT diversity was also higher for the potential study sites than the current MexFlux sites (Table 3.3), even when both consistently monitor some of the most frequent EFT categories (Figure 3.6). In the case of the potential study sites, the tendency to monitor the most frequent EFT categories was consistent for the different number of potential study sites (Figure 3.6b to 3.6f).

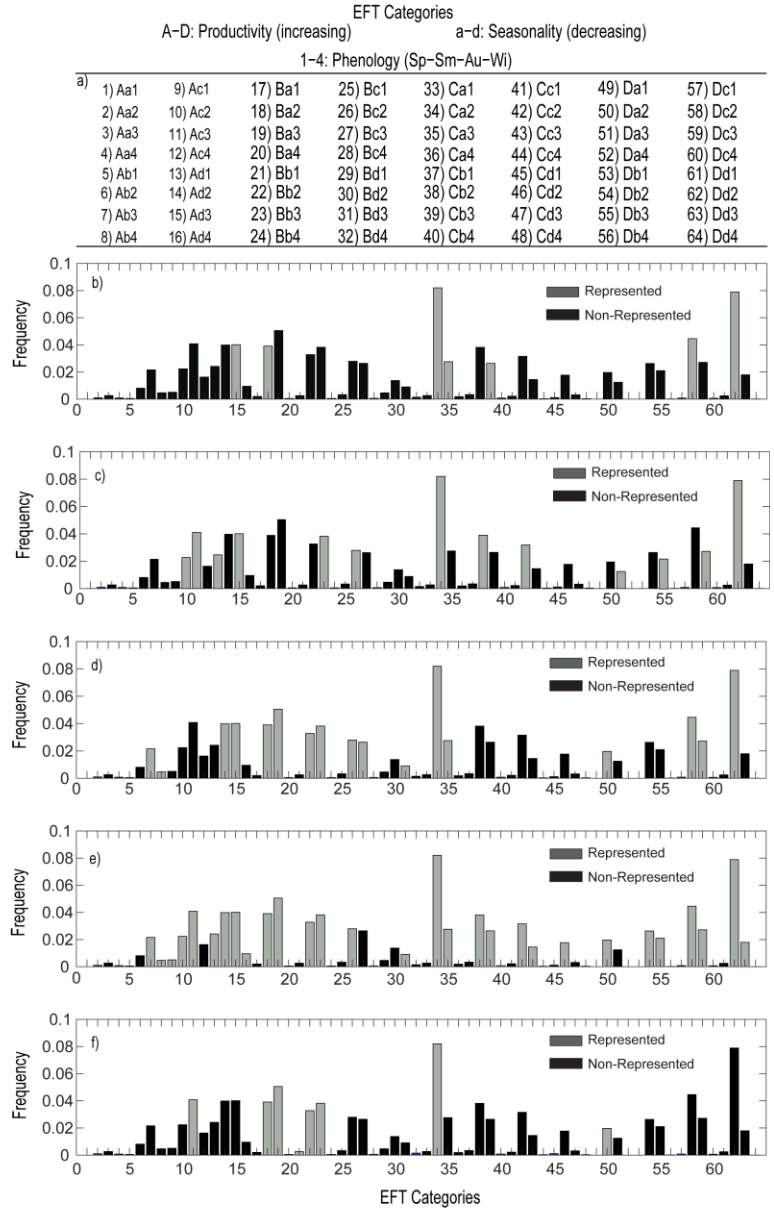


Figure 3.6. Representativeness in terms of ecosystem functional heterogeneity expressed by the unique Ecological Functional Type (EFT) categories potentially monitored by the proposed study sites and by the MexFlux sites: a) Ecosystem Functional Type legend b) The 7 potential sampling sites represent 7 out of 64 categories (11%), c) The 14 potential sampling sites represent 13 out of 64 categories (20%), c) the 28 proposed study sites represent 17 out of 64 categories (27%), d) the 84 proposed study sites represent 31 out of the 64 categories (48%), e) the current MexFlux sites represent 8 out of the 64 EFT categories (12%).

3.4 Discussion

An optimal design for an EON at a national-level requires a stratified sampling strategy that integrates different environmental variables and delineates distinct ecological sampling domains (Graef et al., 2005; Metzger et al., 2013). Representing such ecosystem heterogeneity in megadiverse countries, such as Mexico, makes an optimal EONs design particularly challenging (Hampton et al., 2013; Kelling et al., 2009; Vos et al., 2000). We argue that a first step could be to delineate the national-scale ecosystem heterogeneity, which can be delineated based on expert opinion or using quantitative models (Hargrove & Hoffman, 2005). The sampling domains, e.g., the Ecologically Similar Areas (ESAs) proposed in this study, aim to quantitatively classify the national-level (i.e., Mexico) ecological heterogeneity based on ecological covariates (i.e., bioclimate, topography, soil, ecosystem functional attributes) related to ecosystem processes of interest (i.e., GPP and ET).

Our data-driven approach to delineate ESAs reduces human subjectivity and provides a flexible, analytical and repeatable framework. This quantitative approach is comparable to the delineation of the NEON domains, which were derived from eco-climatic properties in the United States (Keller et al., 2011; Schimel & Keller, 2015). However, to improve our understanding of the land-atmosphere interactions, we propose that it is also needed to include information about ecosystem functionality (Bond-Lamberty et al., 2016; Petrakis et al., 2017; Reichstein et al., 2014). The addition of information on ecosystem functional attributes has enhanced biodiversity models (Alcaraz-Segura et al., 2013, 2017; Armas et al., 2017) and is useful for network design and assessment (Villarreal et al., 2018). We argue that including information related to environmental state factors along with ecosystem functional heterogeneity could complement bioclimatic information to define ESAs, since ecosystem-level observations suggest that climate controls are insufficient to explain land-atmosphere interactions (Reichstein et al., 2014). The spatial heterogeneity of ecosystem functionality and the inter-annual variability of carbon uptake dynamics (as is expressed by EFTs and EFT_{int}), complement the traditional ecoregionalization approach based only on climate, topographic and edaphic information (Keller et al., 2011; Kumar et al., 2011; Metzger et al., 2013). Under a long-term vision, delineation of domains or ecoregions could incorporate information about different climate or ecological scenarios to account for potential environmental changes that may influence an EON's design.

The ESAs proposed in our study could be used to complement and inform expert opinion based on ecological classification systems already available for Mexico (CEC, 1997). For example, the CEC-CONABIO ecoregions of North American Deserts correspond to an ecological region of desert and steppe climate, where the most dominant vegetation types are low growing shrubs and grasses (CEC, 1997). This information is complemented by the functional properties derived from EFTs, which correspond to ecosystems of low productivity, low seasonality and a relative high EFT_{int} (Table 3). Contrasting with North American Deserts, the Tropical Humid Forests are largely composed by evergreen and semideciduous forests, with relatively high annual temperature and precipitation higher than 1000 mm (CEC 1997). The functional characteristics for Tropical Humid Forests correspond to be highly productive with low seasonality EFTs and a relatively low EFT_{int} (Table 3.3). We postulate that EFT complements the traditional approach of ecoregion delimitation as it includes information based on spatial ecosystem functional heterogeneity derived from carbon uptake dynamics (Villarreal et al., 2018).

Our approach for delineation of ESAs adds some advantages over existing ecoregionalization efforts used for EON design since: a) it formally includes the targeted ecosystem processes into the EON design; b) provides a spatial sampling domain that allows the comparison of ecological responses across contrasting environments; c) it represents spatial domains with relatively homogeneous environmental characteristics to select and/or evaluate current and potential study sites (Metzger et al., 2005, 2013); and d) it can be applied to any other region/country around the world as it is based on publicly available data and is built on a flexible framework.

The conceptual design of EONs in developing countries is challenging due to limited country-specific information. However, network coordination is critical for developing countries due to limitation of human and economic resources for monitoring efforts at the national-level (Vargas et al., 2012, 2017). For the case of Mexico, monitoring of ecosystem scale carbon and water fluxes has not been a nationally coordinated effort with the goal to maximize regional representativeness. Instead, the growth of MexFlux has been a principal investigator's effort based on available individual grants, local expertise, and site-security and accessibility. Sites have been selected based on the interest of individual research groups (Vargas et al., 2013) as it is the case for most regional networks within FLUXNET. This approach has resulted in higher density of sites across the northwest of Mexico (Figure 1d) at ecosystems characterized by (see Figure 3.2) medium-low productivity (EFTs named as "B"), high to medium-high seasonality (EFTs named as "a" and "b") and a phenology of summer or autumn greening peaks (EFTs named "2" and "3"). On one side, these study-sites have contributed to synthesis studies to better understand the role of water-limited ecosystems across North America (Biederman et al., 2016; Villarreal et al., 2016). On the other side, the lack of coordination on the distribution of MexFlux sites has resulted in a lack of representativeness for monitoring GPP and ET at the national-level (Figures 3.4d and 3.5d, Table 3.3). This problem is not unique of MexFlux as even in a highly dense network like AmeriFlux there are large underrepresented areas within the region of interest (Villarreal et al., 2018).

Our results show that a coordinated organization of study sites could reach greater representativeness to monitor GPP, ET and EFT diversity than the distribution of the 14 current MexFlux sites (Figure 3.4, 3.5 and 3.6; Table 3.3). For example, a national coordinated array of 14 sites could represent 8% of the spatial variability in GPP and ET, respectively, which is nearly double than the current representativeness of MexFlux. (Figures 3.4a, 3.5a, Table 3.3). However, even though the representativeness of a high number of strategically selected sampling points (i.e., 84 study sites) could be higher than the current distribution of MexFlux sites, a substantial percentage of GPP and ET variability will remain under-represented across Mexico (Table 3.3). In general, the regions that are most difficult to represent are mountain areas across the Pacific coast characterized by medium-high productivity, high and medium-high seasonality, and a growing season peak in summer (Figure 3.2, 3.4c, 3.5c). This representativeness challenge is the result of the great diversity of ecosystems across Mexico due to anthropogenic activities, complex physiography, heterogeneous geology and diverse climate regions (Espinosa et al., 2008). Furthermore, these regions are even more challenging for MexFlux due to limitations of the eddy covariance technique in non-flat terrains (Hammerle et al., 2007; Hiller et al., 2008; Su et al., 2004).

Our framework relies on a data-driven approach to inform EON design, evaluation and improvement, but we recognize that the full implementation of the proposed sites for MexFlux may not be possible. We highlight that a major challenge for EON design is to find a balance between maximizing representativeness of the network and available human and economic resources. This challenge increases when there are limited human resources, funding opportunities and a low degree of interoperability among researchers, governments and society (Vargas et al., 2017). Furthermore, we recognize that there exists inherent uncertainty associated with each ecological covariate used in our framework, but we could not incorporate a formal assessment of uncertainty propagation since most of these publicly available products do not include a spatially explicit assessment of uncertainty. Even though, these globally available covariates were used as value-added products to give our framework a wider applicability.

We identify several challenges to increase the representativeness of MexFlux for monitoring GPP and ET across Mexico. First, our framework suggests new locations for potential study sites, but accessibility, security, topographic conditions, and proximity to a research center with expert knowledge are limiting factors. Second, the large number of study sites needed to represent the ecosystem functional heterogeneity across Mexico is beyond any current funding opportunities and will remain highly unlikely for the future. That said, our approach provides guidelines for researchers and organizations to evaluate current network performance and support informed decisions about network growth. The low current representativeness of MexFlux highlights the need to move towards a national strategy to spatially maximize the representativeness of the complex ecological functional heterogeneity of the carbon cycle across this megadiverse country.

3.5 Conclusions

Mexico is a megadiverse country whose complexity in ecological diversity represents both an opportunity and a challenge to increase our understanding of regional-to-continental land-atmosphere interactions. Our results show that few strategically located sites could result in higher representativeness of national-scale GPP and ET than the current array of MexFlux sites. However, our results suggest that over 80 sites are required to substantially improve the representation of GPP and ET at the national-level. The great ecological diversity across Mexico represents a challenge to EONs design since even minimum levels of representativeness will require large investments for infrastructure, maintenance, human resources and an overall increase in interoperability.

This study proposes a framework (based on publicly available data) to test and improve the design and representativeness of EONs to monitor GPP and ET at the regional-scale. The proposed framework could be applied to any region of the world, with particular interest for developing countries currently lacking detailed information of GPP and ET. For example, Latin America includes a large ecosystem diversity along with large gradients of land-uses and land-use-change, but there are only 41 eddy covariance sites affiliated with FLUXNET. The implementation of alternative assessment approaches can provide insights for decision-making to inform EON designs and improve the understanding of regional carbon and water cycles while maximizing available human and economic resources.

Acknowledgements

SV and MG acknowledge support from CONACyT doctoral fellowships. RV acknowledges support from NASA Carbon Monitoring Systems (80NSSC18K0173). DAS acknowledges support from JC2015-00316 grant, and the projects: CGL2014-61610-EXP, ECOPOTENTIAL (H2020 grant agreement 641762), and NASA-GEOBON award 80NSSC18K0446. This study is part of the synergy efforts from CarboNA, North American Carbon Program, and the Mexican Carbon Program. 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

Alcaraz-Segura, D., Cabello, J., & Paruelo, J. (2009). Baseline characterization of major Iberian vegetation types based on the NDVI dynamics. *Plant Ecology*, *202*(1), 13–29. https://doi.org/10.1007/s11258-008-9555-2

Alcaraz-Segura, D., Paruelo, J. M., Epstein, H. E., & Cabello, J. (2013). Environmental and human controls of ecosystem functional diversity in temperate South America. *Remote Sensing*, *5*(1), 127–154. https://doi.org/10.3390/rs5010127

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

Alcaraz, D., Paruelo, J., & Cabello, J. (2006). Identification of current ecosystem functional types in the Iberian Peninsula. *Global Ecology and Biogeography*, *15*(2), 200–212. https://doi.org/10.1111/j.1466-822X.2006.00215.x

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

Anderson D. W. (1988). The Effect of Parent Material and Soil Development on Nutrient Cycling in Temperate Ecosystems. *Biochemistry*, *5*(1), 71–97.

Armas, D., Guevara, M., Alcaraz-Segura, D., Vargas, R., Soriano-Luna, Á., Durante, P., & Oyonarte, C. (2017). Digital map of the organic carbon profile in the soils of Andalusia, Spain. *Ecosistemas*, *26*(3), 80–88. https://doi.org/10.7818/ECOS.2017.26-3.10

Baldocchi, D., Falge, E., Gu, L., & Olson, R. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

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

Bond-Lamberty, B., Epron, D., Harden, J., Harmon, M. E., Hoffman, F., Kumar, J., et al. (2016). Estimating Heterotrophic respiration at large scales: challenges, approaches, and next steps. *Ecosphere*, *7*(June), 1–13. https://doi.org/10.1002/ecs2.1380

Cambardella, C., Moorman, T., Novak, J., Parkin, T., Karlen, D., Turco, R., & Konopa, A. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Science Society of America Journal*, *58*(January 1994), 1501–1511. https://doi.org/10.1017/CBO9781107415324.004

CEC. (1997). *Ecological regions of North America: toward a common perspective*.

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

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

Cruz-Cárdenas, G., López-Mata, L., Ortiz-Solorio, C. A., Villaseñor, J. L., Ortiz, E., Silva, J. T., & Estrada-Godoy, F. (2014). Interpolation of mexican soil properties at a scale of 1:1,000,000. *Geoderma*, *213*(July 2016), 29–35. https://doi.org/10.1016/j.geoderma.2013.07.014

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

Espinosa D. O., Ocegueda S., Aguilar C., Flores O., Llorente-Bousquets J., V. B. (2008). Primera Parte: La perspectiva biogeográfica y ecosistémica. *Capital Natural de México, Vol. I : Conocimiento Actual de La Biodiversidad.*, 33–65.

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

Graef, F., Schmidt, G., Schröder, W., & Stachow, U. (2005). Determining ecoregions for environmental and GMO monitoring networks. *Environmental Monitoring and Assessment*, *108*(1–3), 189–203. https://doi.org/10.1007/s10661-005-3966-5

Hammerle, A., Haslwanter, A., Schmitt, M., Bahn, M., Tappeiner, U., Cernusca, A., & Wohlfahrt, G. (2007). Eddy covariance measurements of carbon dioxide, latent and sensible energy fluxes above a meadow on a mountain slope. *Boundary-Layer Meteorology*, *122*(2), 397–416. https://doi.org/10.1007/s10546-006-9109-x

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., et al. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162. https://doi.org/10.1890/120103

Hargrove, W. W., & Hoffman, F. M. (2005). Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental Management*, *34*. https://doi.org/10.1007/s00267-003-1084-0

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

Hayes, D. J., Turner, D. P., Stinson, G., McGuire, A. D., Wei, Y., West, T. O., et al. (2012). Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory-based data. *Global Change Biology*, *18*(4), 1282–1299. https://doi.org/10.1111/j.1365-2486.2011.02627.x

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

Hiller, R., Zeeman, M. J., & Eugster, W. (2008). Eddy-covariance flux measurements in the complex terrain of an Alpine Valley in Switzerland. *Boundary-Layer Meteorology*, *127*(3), 449–467. https://doi.org/10.1007/s10546-008-9267-0

Holzer, J. M., Adamescu, M. C., Bonet-García, F. J., Díaz-Delgado, R., Dick, J., Grove, J. M., et al. (2018). Negotiating local versus global needs in the International Long Term Ecological Research Network's socio-ecological research agenda. *Environmental Research Letters*, *13*(10). https://doi.org/10.1088/1748-9326/aadec8

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., et al. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, *232*, 144–157. https://doi.org/10.1016/j.ecolmodel.2012.02.004

Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, *10*(2), 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

Jongman, R., Skidmore, A., Mücher, C., Bunce, R., Metzger, R. (2017). Global Terrestrial Ecosystem Observation: Why, Where, What and How? In R. G. S. Michele Walters (Ed.), *The GEO handbook on Biodiversity Observataion Network* (pp. 19–38). Springer International Publishing. https://doi.org/10.1007/978-3-319-27288-7

Keller, M., Schimel, D. S., Hargrove, W. W., & Hoffman, F. M. (2008). A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment*, *6*(5), 282–284. https://doi.org/10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2

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.

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive Science: A New Paradigm for Biodiversity Studies. *BioScience*, *59*(7), 613–620. https://doi.org/10.1525/bio.2009.59.7.12

King, A. W., Andres, R. J., Davis, K. J., Hafer, M., Hayes, D. J., Huntzinger, D. N., et al. (2015). North America's net terrestrial CO2exchange with the atmosphere 1990-2009. *Biogeosciences*, *12*(2), 399–414. https://doi.org/10.5194/bg-12-399-2015

Kumar, J., Mills, R. T., Hoffman, F. M., & Hargrove, W. W. (2011). Parallel k-means clustering for quantitative ecoregion delineation using large data sets. *Procedia Computer Science*, *4*, 1602–1611. https://doi.org/10.1016/j.procs.2011.04.173

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

Lee, S. J., Berbery, E. H., & Alcaraz-Segura, D. (2013). The impact of ecosystem functional type changes on the La Plata Basin climate. *Advances in Atmospheric Sciences*, *30*(5), 1387–1405. https://doi.org/10.1007/s00376-012-2149-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

Menoni, S., Molinari, D., Parker, D., Ballio, F., & Tapsell, S. (2012). Assessing multifaceted vulnerability and resilience in order to design risk-mitigation strategies. *Natural Hazards*, *64*(3), 2057–2082. https://doi.org/10.1007/s11069-012-0134-4

Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Mücher, C. A., & Watkins, J. W. (2005). A climatic stratification of the environment of Europe. *Global Ecology and Biogeography*, *14*(6), 549–563. https://doi.org/10.1111/j.1466-822X.2005.00190.x

Metzger, M. J., Brus, D. J., Bunce, R. G. H., Carey, P. D., Gonçalves, J., Honrado, J. P., et al. (2013). Environmental stratifications as the basis for national, European and global ecological monitoring. *Ecological Indicators*, *33*, 26–35. https://doi.org/10.1016/j.ecolind.2012.11.009

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

Mittermeier, R., & Mittermeier, C. De. (1992). La importancia de la diversidad biologica de México. *Comisión Nacional Para El Conocimiento y Uso de La Biodiversidad*. Retrieved from http://www.eduteka.org/gestorp/recursos/docs/5515-2012-10-12-7891.pdf

Paruelo, J. M., Jobbagy, E. G., & Sala, O. E. (2001). Current distribution of ecosystem functional types in temperate South America. *Ecosystems*, *4*(7), 683–698. https://doi.org/10.1007/s10021-001-0037-9

Petrakis, S., Barba, J., Bond-Lamberty, B., & Vargas, R. (2017). Using greenhouse gas fluxes to define soil functional types. *Plant and Soil*, (December), 868–874. https://doi.org/10.1007/s10745-006-9094-1

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). https://doi.org/10.1073/pnas.1216065111

Rousseeuw, P. J. (1987). Silhouttes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schimel, D., & Keller, M. (2015). Big questions, big science: meeting the challenges of global ecology. *Oecologia*, *177*(4), 925–934. https://doi.org/10.1007/s00442-015-3236-3

Schimel, D., Hargrove, W., Hoffman, F., & MacMahon, J. (2007). NEON: A hierarchically designed national ecological network. *Frontiers in Ecology and the Environment*, *5*(2), 59. https://doi.org/10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2

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., Walters, M., Turak, E., Saarenmaa, H., Heip, C. H. R., Tuama, É. Ó., et al. (2012). Building a global observing system for biodiversity. *Current Opinion in Environmental Sustainability*, *4*(1), 139–146. https://doi.org/10.1016/j.cosust.2011.12.005

Su, H., Schmid, H. P., Grimmond, C. S. B., Vogel, S., & Oliphant, A. J. (2004). Spectral characteristics and correction of long-term eddy-covariance measurements over two mixed hardwood forests in non-flat terrain. *Boundary-Layer Meteorology*, *110*(2), 213–253.

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

Vargas, R., Loescher, H. W., Arredondo, T., Huber-Sannwald, E., Lara-Lara, R., & Yépez, E. a. (2012). Opportunities for advancing carbon cycle science in Mexico: Toward a continental scale understanding. *Environmental Science and Policy*, *21*, 84–93. https://doi.org/10.1016/j.envsci.2012.04.003

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

Vos, P., Meelis, E., & W., T. K. (2000). A framework for the design of ecological monitoring programs as a tool for environmental and nature management. *Environmental Monitoring and Assessment*, *61*(3), 317–344. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed5&NEWS=N&AN=2000125369

Chapter 4

REPRESENTATIVENESS OF FLUXNET SITES ACROSS LATIN AMERICA

In *preparation* for Environmental Research letter.

Authors:

Samuel Villarreal[1], Trevor Keenan[2,3], Rodrigo Vargas[1]

Affiliations:

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

[2] Earth Science Division, Lawrence Berkeley National Lab, Berkeley, CA 94709, USA.

[3]Department of Environmental Science, Policy and Management. UC Berkeley, Berkeley CA 94709, USA.

Abstract

Environmental observatory networks (EONs) provide information that help us understand, model and forecast the spatial and temporal patterns of Earth's biophysical process. The role of EONs includes data collection, data sharing, synthesis activities, and building communities of practice. Consequently, representativeness analyses are important since they provide insights to improve EON's management and network design and interpretation of data-driven products. We assessed the representativeness of FLUXNET sites across Latin America (LA), a region of great importance for the global carbon and water cycles, which represents nearly 13% of the land surface area. Representativeness analyses were performed based on concepts derived from species distribution models (SDM), since the goal was to delineate the spatial distribution of environmental properties across a geographic space that should be similar to the environmental range monitored by corresponding FLUXNET sites. Our results show a spatial representativeness of LA surface area of 0.34 % for climate properties, 0.36 % for terrain parameters, 0.34% for soil resources, 0.45% of all previous environmental drivers added into a principal component analysis, 0.48% for gross primary productivity and 0.34% for evapotranspiration. We discussed the need to enhance interoperability across and promote the participation of active or inactive sites to register and share information with local, regional and international networks non-affiliated study sites to FLUXNET in order to increase the utility of EON's across LA. The proposed representativeness framework is based on publicly available information and open source software and it can be applied to any other region across the world.

Keywords

machine learning, representativeness, Latin America, FLUXNET, AmeriFlux

4.1 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 answer challenging scientific questions (Lovett et al., 2007; Scholes et al., 2017). In theory, 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 task lead by EONs include data collection, data sharing, and synthesis activities. Furthermore, EONs provide a wide range of value-added products such as databases, conceptual models and synthesis reports that help decisions makers to make informed environmental policies or management actions (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 micrometeorological tower 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). FLUXNET is a global ‘network of regional networks’ that promotes the compilation, harmonization, standardization, and archiving of eddy-covariance data for the broader scientific community. As the FLUXNET archive includes multiple ecosystems across the world, it is possible to generate knowledge of the interaction between terrestrial ecosystems and the atmosphere at a global scale (Falge et al., 2002; Fisher et al., 2008; Schwalm et al., 2017). However, FLUXNET sites are not evenly distributed and they under-represent certain regions or ecosystems across the world (Kumar et al., 2016). 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). Consequently, these assessments are relevant to increase EON's 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) . A common approach to assess representativeness of EONs has been the estimation of minimum distances within a multivariate space (Hargrove et al., 2003; Kumar et al., 2016; 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; Villarreal et al in review).

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 to the presence of a certain species (Drew et al., 2011; Evans et al., 2011). This concept can be applied to assess the representativeness of EONs, since the goal is to delineate the spatial distribution of environmental properties across a geographic space that should be similar to the environmental range monitored by corresponding EON's study sites (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 covers. LA includes nearly 13% of the global land surface area, but only 5% off all registered FLUXNET sites are located within this region. It is clear that the density of registered FLUXNET sites in LA is low when compared to regions such as the United States or Europe, so a representativeness analysis is needed to better interpret data-driven products parameterized with FLUXNET data from this region.

The overarching goal of this study is to provide an assessment of the current representativeness of registered FLUXNET sites across LA to monitor environmental properties such as climate, topography and soil resources along with ecosystem process such as gross primary productivity (GPP) and evapotranspiration (ET). We asked three interrelated research questions: 1) What is the representativeness of FLUXNET sites across LA to characterize climate, topography and soil resources variability? 2) What is the representativeness of FLUXNET sites to represent GPP and ET patterns across LA? and 3) How many more sites are needed to 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 in the world.

4.2 Data and Methods

4.2.1 FLUXNET registered sites

FLUXNET provides standardized data products through coordination among various regional flux networks across the globe (http://fluxnet.fluxdata.org). We used the online database to extract the geographical location of eddy-covariance sites across LA affiliated to FLUXNET. Currently, there are 41 registered sites distributed across different ecosystems (September/2018), but we recognize that there are several unregistered eddy-covariance sites across LA. For this study, we only considered sites affiliated with FLUXNET despite if they are active or inactive and if they have provided data or not to the FLUXNET database. Consequently, this study provides a conservative representativeness of eddy-covariance sites across LA, and we hope that it will encourage principal investigators to register their sites and share data with FLUXNET to improve the representation of LA in regional and global studies.

4.2.2 Environmental properties

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 may constrained the spatial patterns of ecosystem processes such as GPP and ET (Amundson and Jenny 1991, Chapin *et al* 2002). We used 19 bioclimatic predictors to characterize climate conditions since it capture 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 network (Villarreal *et al* 2018). The statistic parameters used to characterize GPP and ET dynamics were the mean and the coefficient of variation, since they have been used as proxies for ecosystem productivity and seasonality (Alcaraz-Segura *et al* 2017; Villarreal et al 2018).

4.2.3 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°). We selected 0.05° 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 and NEON (Villareal *et al* 2018). 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 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 argue that the relative few eddy-covariance sites across the large geographic extent of LA is a similar case. 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. Then, each classifications trees (CT) are built from sample bootstrapping by repeatedly partitioning the training-data into a binary-series of clusters (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-nodes until is stopped (Marmion et al., 2009). The grown trees are used to predict the out-of-bag observations. Then, 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).

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) using the 'SSDM' library (Schmitt et al., 2017).

4.2.4 Improving the representativeness of GPP and ET

Our 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 and 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 (i.e., mean and std of GPP and ET). The cLHS serves as an efficient sampling strategy and it has been used in EON's representativeness analysis (Villarreal *et al* review). For this assessment we followed a sequential approach: (a) first we started by adding additional sites across LA in increments of 10 sites until reaching 100 sites; and 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 similar to the total number of eddy-covariance sites registered in the AmeriFlux network for the conterminous United States, which represents about 40% of LA surface area (Villarreal *et al*., 2018). The assessment of representativeness by adding new FLUXNET sites was also performed using RF as described above.

4.3 Results

4.3.1 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 current number of FLUXNET sites at each IGBP category is also not evenly distributed (Figure 4.1 and Table 4.1). For example, Evergreen Broadleaf Forest has the most sites (n = 19), followed by Woody Savanna (n = 8) and Open Shrublands (n = 5). Six IGBP categories across LA do not have any FLUXNET site (Table 4.1).

Table 4.1. Surface area of each of the International Geosphere-Biosphere Project biome class across Latin America and the number of eddy-covariance sites at each class.

IGBP Class	**Area covered (%)**	**Number of eddy-covariance towers**
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**

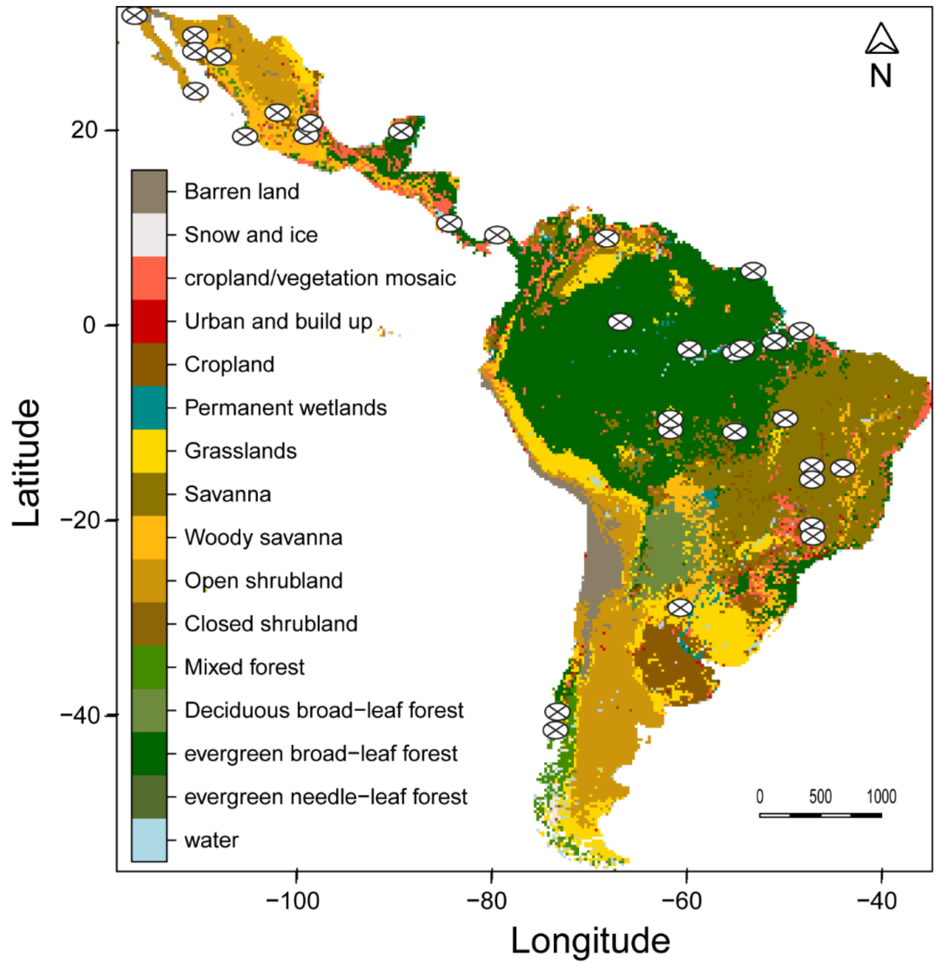


Figure 4.1. Spatial distribution of the different biomes across Latin America according to the International Geosphere-Biosphere Program (IGBP) and the location of the eddy covariance sites affiliated to FLUXNET.

4.3.2 Representativeness of environmental properties and ecosystem processes.

The representativeness of FLUXNET sites differ for each environmental characteristic assessed (i.e., climate, terrain properties, soil resources, combined environmental properties, GPP and ET). The tested variables with the highest spatial representativeness were GPP (0.48%), combined environmental properties (0.45%) and terrain parameters (0.36%), while climate, soil resources and ET had the same representativeness (0.34%; Table 4.2). The representativeness between the IGBP categories is different among the distinct environmental variables assessed. The categories having at least one study site with the higher representativeness among all the different environmental variables assessed (i.e., climatic, topographic, soil resources, combined environmental properties, GPP and ET) were shrublands and savannas, while forest ecosystems (Evergreen and deciduous broadleaf forest) have similar ranges than transform ecosystems (croplands, cropland natural vegetation mosaic; Table E-1).

Table 4.2. Parameters used to characterize each model, their performance and their overall representativeness across Latin America.

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

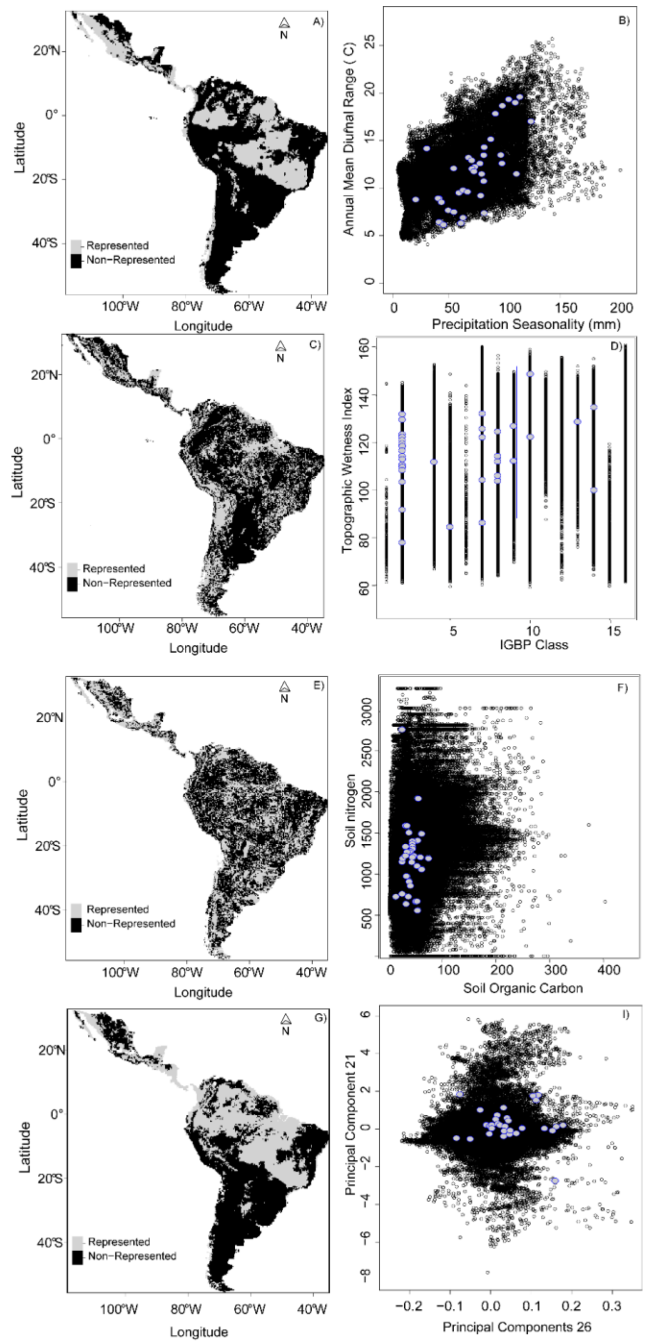


Figure 4.2. Spatial representativeness for the distinct environmental parameters based on random forest models (A,C, E and G), along with the distribution of the FLUXNET sites across the multivariate space represented by the two most influential variables for each model (B, D, F and H). Black areas indicate not-represented regions while gray areas are represented regions.

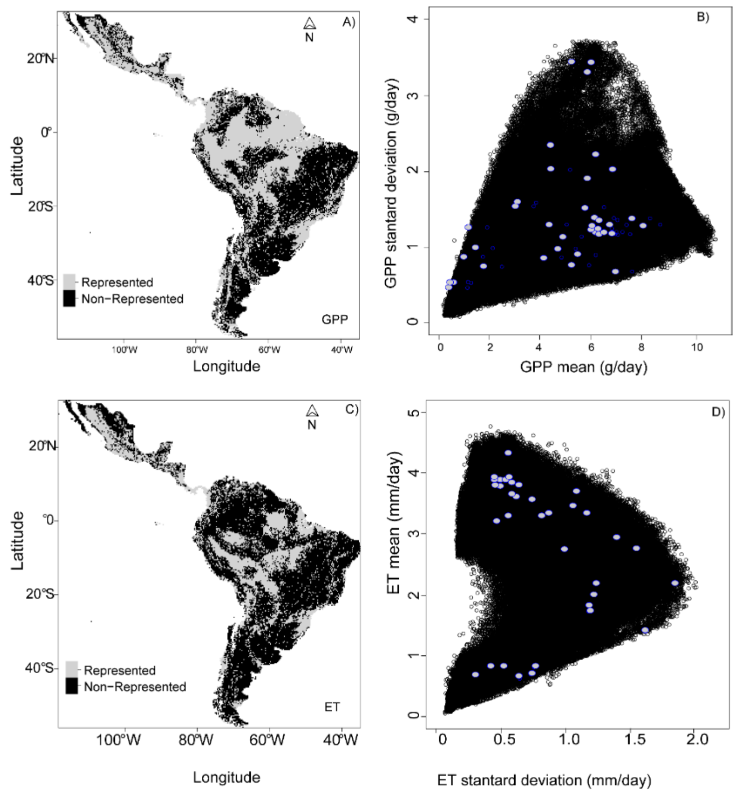


Figure 4.3. Spatial representativeness of gross GPP and ET based on random forest models (A,C), along with the distribution of the FLUXNET sites across the multivariate space represented by the two most influential variables for each model (B, D). Black areas indicate not-represented regions while gray areas are represented regions.

We compared the distribution between the represented and non-represented regions of the two most important variables for each representative model (Figures 4.2 and 4.3). For the bioclimatic predictor (Figure 4.2A-B), precipitation seasonality above 120mm and below 40mm and the annual mean diurnal temperature range above 20 ^{0}C and below 6 ^{0}C are not represented (Figure 4.4 A-B). For terrain properties

(Figure 4.2 C-D) the majority of the IGBP classes are represented while values 90>TWI>75 are not represented for TWI (Figure 4.4 C-D). For soil resources (Figure 4.2 E-F) soil organic carbon >80 g/m^2 and soil nitrogen below <500 and above >2000 mg/m^2 are not represented (Figure 4.4 E-F). For all the environmental drivers combined in a PCA the components that had the highest influence on the PCA representativeness map were PC26 and PC21 (Figure 4.2 G-H), PC26 is only represented within the range -0.08 to 0.16 while PC21 is represented within the range of -1 to 1.5 (Figure 4.4 G-H).

The representativeness of GPP and ET is likely to be spatially similar (Figure 4.3). For GPP (Figure 4.3 A-B), the representativeness of FLUXNET sites for GPP_mean tend to be bias towards values above the GPP_mean median for LA, while for GPP_CV only those values above > 2 g/day are not well represented (Figure 4.5A). The representativeness of ET is also similar (Figure 4.3 C-D), the ET_mean representativeness of FLUXNET sites is bias towards values above the ET_mean median for LA, while for ET_CV values close to 2.0 mm/day are not represented (Figure 4.5 C-D).

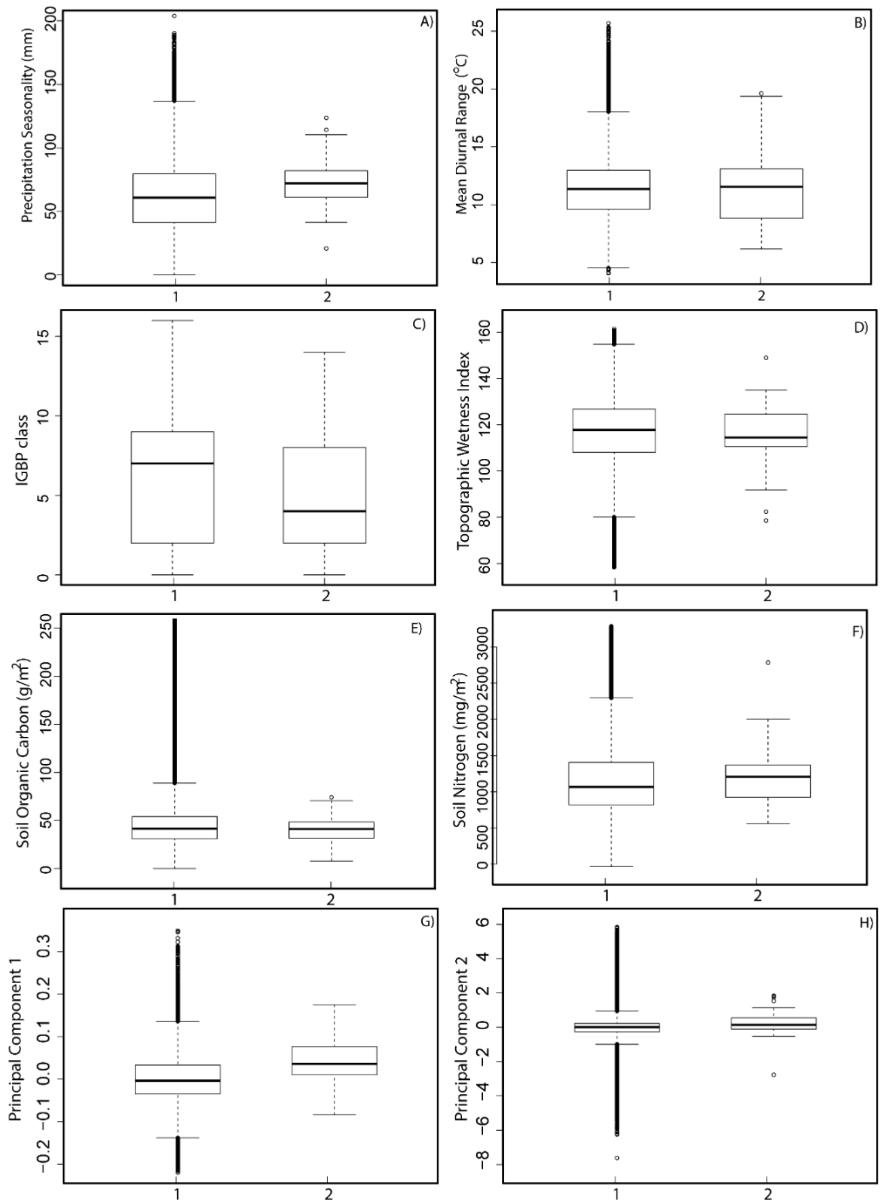


Figure 4.4. Distribution of the two most influential variables for each environmental parameter (A, C, E, G; number 1), along with the distribution of the sample monitored by the FLUXNET sites for each (B, D, G and H; number 2).

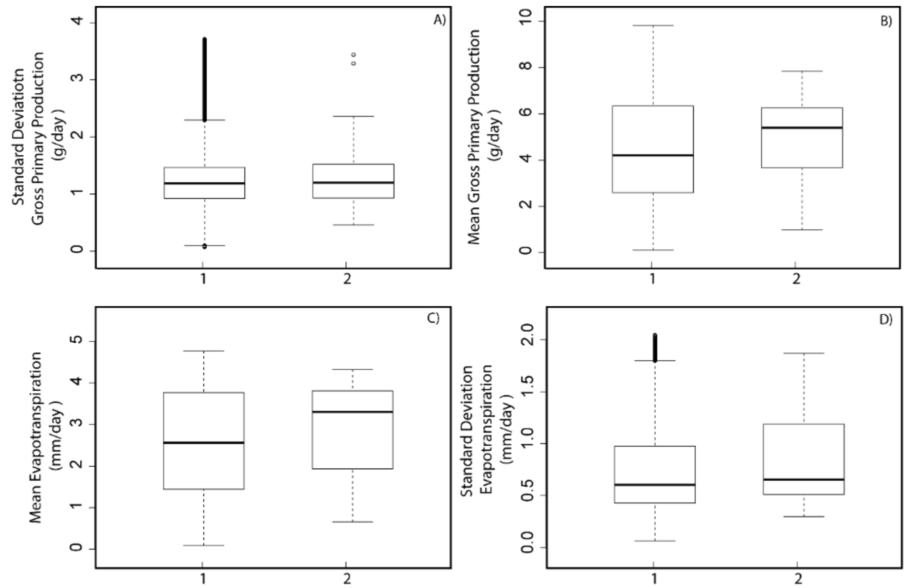


Figure 4.5. Distribution of the two most influential variables for GPP (A and C; number 1) and ET (E and G; number 1), along with the distribution of the sample monitored by FLUXNET for GPP (B and D; number 2) and for ET (F and H; number 2).

4.3.3 Representativeness by adding potential sampling sites

The overall representativeness of GPP and ET across LA slightly increased by adding new study-sites (Figure 4.6), despite that those sites represent the most dominant properties of GPP and ET and are mainly located at the larger IGBP classes (table 4.3). Also, the addition of new study-sites progressively increased the predictive power for each model (Figure 4.6), this is the correlation between the current and the additional sites with the represented regions.

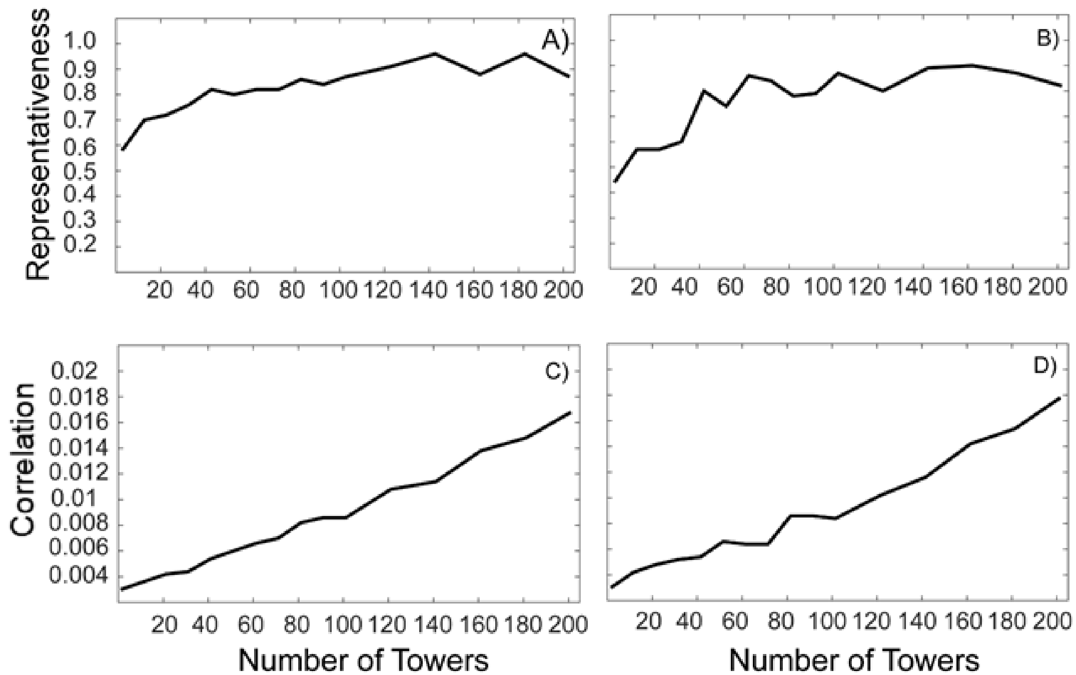


Figure 4.6. Representativeness according with the addition of potential study-sites based on the cHLS technique for GPP (A) and ET (B), along with their predictability increase (C and D).

Table 4.3. Number of potential-study sites based on cLHS for each International Geosphere-Biosphere Project biome class.

IGBP Class	Current number of Towers	Including new Study 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

4.4 Discussion

4.4.1 FLUXNET representativeness for environmental parameters and ecosystem processes.

Our results show that shrublands and woody savannas had the highest representativeness among all the different environmental properties assessed (i.e, bioclimatic predictors, terrain parameters, soil resources and combined environmental properties), along with ET and among the highest for GPP (Table E-1). Despite only having a combined 36% of the registered FLUXNET sites across LA (open shrublands, close shrublands and woody savannas; Table 4.1). These results suggest that these ecosystems are environmentally less heterogeneous than others, arguably due to a semi-arid permanent water stress conditions that triggers common strategies on resources use such as water-use efficiency (Biederman et al 2016, Huxman et al 2004, Ponce Campos 2013). Previous studies indicate a high convergence on structural and functional properties for dryland ecosystems such as shrublands and savannas between North and South America (Paruelo 1998), which could provide insights about why few sites could contribute to a relative high representativeness even though most of these sites are located at the Northern domain of LA (Figure 4.1). Despite the high representativeness of shrublands and savannas by a low number of eddy-covariance sites, the undergoing drought across these water-limited ecosystems highlights the need to support monitoring programs since these ecosystems are especially sensitive to changes in precipitation and temperature (Biederman 2016;2017; Villarreal 2017).

The representativeness of forested ecosystems such as evergreen broadleaf forest and deciduous forest were consistently lower than shrublands and woody savannas for all the distinct environmental variables assessed including GPP and ET (S1), despite that more than 50% of the monitoring efforts have focused across these ecosystems, especially on tropical forests (Table 1). These results suggest a larger variability in climatic, topographic and soil resources cycling. For example, the precipitation gradient is larger for forested ecosystems than for shrublands and savannas (Chapin *et al* 2002). Soil nutrient cycling in forest ecosystems is more dynamic and variable than water-limited ecosystems (Vitousek, 1984; Vitousek and Sanford, 1986). The complex natural dynamic of GPP and ET along with change in land-use (i.e., deforestation and agriculture) and climate variability, offset carbon sequestration of undisturbed forest at the point that there is no consensus if the Amazonia basin is a sink or source of carbon (Andreae et al., 2002; Avissar et al., 2002; Grace et al., 1995). These results highlight the challenges that leads to represents the broad spectrum of the environmental properties assessed across forested ecosystems in LA. However, the current efforts on tropical wet forest have supplied a wide variety of environmental information (e.g., atmospheric chemistry, land-use land-cover change), and have fostered our knowledge on carbon and water fluxes at natural, converted and afforested sites across broadleaf evergreen forest in LA (Andreae et al., 2002; Avissar et al., 2002; Keller et al., 2004). However, our results highlight the need to improve the representativeness of environmental variables (i.e., climatic, terrain parameters and soil resources) and ecosystem process across these ecosystems.

Other ecosystems such as croplands and crop-land natural vegetation mosaics have a relative similar representativeness among all the environmental variables assessed including GPP and ET, despite a substantial lower number of FLUXNET sites (Table 4.1). Arguably this is due by the low number of towers and the wide range of environmental conditions where this land-covers are located. The undergoing agricultural expansion particularly on forested ecosystems highlight the need to enhance GPP and ET representativeness across these ecosystems, since the land-use change of tropical forest to croplands reduces carbon uptake and it increase its seasonality, along with a potential increase in the intensity of precipitation seasonality and evapotranspiration (Andreoli and Kayano, 2005; Eva et al., 2004; Graesser et al., 2015).

Other ecosystems such as Mixed Forest, Grasslands and Permanent Wetlands also plays a key role in climate regulation, soil nutrient cycle and ecosystem processes such as GPP and ET (Baldocchi et al., 2000; Conant et al., 2001; Whiting and Chanton, 2001). However, the representativeness of those ecosystems is based on the approximation of eddy-covariance sites located at different IGBP classes because any of the aforementioned ecosystems has a study site affiliated to FLUXNET (Table 4.1). By hence, the interpretation of those results should be performed cautiously. However, there are some study sites located at grasslands and Permanent Wetlands and among other ecosystems that haven't been affiliated to FLUXNET (Delgado-Balbuena et al., 2013; Hinojo-Hinojo et al., 2018; Tonti et al., 2018). We hope that these results encourage affiliation of eddy-covariance across LA to FLUXNET in order to provide a more comprehensive view about the representativeness of eddy-covariance across LA.

4.4.2 Improving Representativeness

LA comprehend one of the most widely 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 the distinct environmental properties assessed (i.e, bioclimatic, terrain properties and soil resources) including GPP and ET (Figures 4.2-4.3), these gaps depicts the under-represented regions of the different IGBP classes (S4.1). For the environmental properties assessed, conditions located at the low or high end of the distribution across LA are usually under-represented but their medians are likely to be similar (Figure 4.4). These conditions are also similar for GPP and ET except that the median of the represented regions is higher than the distribution of GPP_mean and ET_mean across LA (Figure 4.5), this is arguably due by the larger efforts for monitoring tropical forest (Table 4.1).

In order to complement the current monitoring efforts across LA this study proposed a coordinated approach to improve FLUXNET representativeness of ecosystem fluxes such GPP and ET under a top-down approach, as recent studies have discussed the efficiency of top-down EON design to efficiently enhance spatial representativeness (Villarreal *et al* 2018; Villarreal in Review). Our results add two main advantages over the existing FLUXNET representativeness on GPP and ET dynamic: a) increased the spatial representativeness of GPP and ET as potential study-sites were strategically located (Figure 4.6A); b) increased the predictive capacity of FLUXNET as the correlation between the represented regions with the properties monitored by the potential and current study-sites also increased (Figure 6B). While looking at these results we have identified the need to add more study sites to increased FLUXNET representativeness and its validations, however, the addition of more sites brings EON's challenges related to its design and effort coordination, specially within a region of limited economic, logistic and human resources.

4.4.3 Opportunities and Challenges

Traditionally, monitoring carbon and water fluxes across LA has not been done under a national coordinated effort that could maximizes regional representativeness at a country level (Roberti et al., 2012; Vargas et al., 2013), monitoring efforts have been performed by individual research groups or by local networks with clear and specific questions on specific land covers (i.e., SulFlux; BrasFlux; Roberti *et al* 2012). However, recent studies have identified key aspects and challenges to coordinate monitoring efforts (Sierra et al., 2017; Villarreal et al., 2018), these are: a) Coordinated effort towards archive, synthesis and analyze existing information; b) Coordinated effort towards the collection of new information; c) The creation of platforms for data sharing, scientific discussion and potential actions among researchers and decision makers.

FLUXNET is a global network that owns an infrastructure able to archive and synthesizes the vast amount of previously collected information and it makes easier to scientists the assessment of current archived information, storage of new information, and sharing information and ideas (Baldochhi *et al* 2001). The primary functions of FLUXNET are toward the advancement of knowledge mostly on carbon, water, and energy fluxes from the local to global scale (Baldochhi *et al* 2001), by hence, FLUXNET can serve as a platform to coordinate monitoring efforts within LA. However, the poor interoperability across sites and local networks within LA represents a challenge. In order to overcome this challenge, we suggest the coordination of current monitoring efforts by implementing an interoperability framework to overcome barriers such as standardization of sampling protocols, data standards, duplications of efforts and poor data sharing among others (Vargas et al

2017). The coordination of these actions can be supported under the FLUXNET platform.

4.4.4 Limitation and consideration

Our study is based using location from currently affiliated to FLUXNET sites, but we are aware that there are multiple unaffiliated sites across LA. Just to mention a few examples, there have been research efforts across permanent wetlands (Tonti et al., 2018), semi-arid grasslands (Hinojo-Hinojo et al., 2016) and croplands (Lewczuk et al., 2017). Furthermore, we recognize that a site that is affiliated with FLUXNET may be active or inactive and may have (or not) contributed with information to the FLUXNET database. By hence, our representativeness results must be taken as a conservative approach, especially for those IGBP classes with no eddy-covariance site registered at FLUXNET (Table 4.1). We assume that if a site is currently registered with FLUXNET the eddy-covariance information either is available, or the principal investigator is willing to contribute in the near future. A clear 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 in review). This study could motivate principal investigator and regional networks (i.e., MexFlux, Brasflux, SULFLUX) to join and collaborate with FLUXNET and contribute with their data and knowledge to build a stronger network and increase our understanding of the Earth System.

The overarching aim of this study is to provide an overview of the representativeness of FLUXNET across LA and can be taken as a benchmark to encourage the scientific community to affiliate to AmeriFlux and FLUXNET with the purpose to have a common platform to boost the scientific research in this region, these actions could serves FLUXNET's goal of supporting synthesis, discussion and communication activities and to promote workshops and scientists visits (Baldocchi *et al* 2001). The enhancement of interoperability across sites and regional networks within LA could help to address questions such as; which ecosystems required more study sites? Are there some redundancy of efforts? How to optimize current monitoring efforts across LA?. Thus, by improving monitoring efforts within LA more pervasive questions such as the spatial and yearly variability of carbon and water fluxes across LA? How land-cover and land-use changes affects carbon and water fluxes across different ecosystems within LA? And, how and which biophysical factors and anthropogenic activities impact on water and carbon fluxes within LA?

4.5 Conclusions

The present study provides insights about the 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. Our results identified the ecosystems with the higher/lower representativeness and provided information about which ecosystems should be monitor in order to increase the representativeness of the region. This study also brings up the discussion about the necessity to add the study sites that are not affiliated to FLUXNET and the need to enhance interoperability among the different research groups in LA.

Acknowledgements

SV acknowledge support from CONACyT doctoral fellowships. RV acknowledges support from NASA Carbon Monitoring Systems (80NSSC18K0173). Original data to perform the analysis is publicly available (see methods for details), and output data will be archived at ORNL DAAC (pending solicitation).

REFERENCE

Alcaraz-Segura, D., Lomba, A., Sousa-Silva, R., Nieto-Lugilde, D., Alves, P., Georges, D., Vicente, J.R., Honrado, J.P., 2017. Potential of satellite-derived ecosystem functional attributes to anticipate species range shifts. Int. J. Appl. Earth Obs. Geoinf. 57, 86–92. doi:10.1016/j.jag.2016.12.009

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

Andreae, M.O., Artaxo, P., Brandão, C., Carswell, F.E., Ciccioli, P., Da Costa, A.L., Gulf, A.D., Esteves, J.L., Gash, J.H.C., Grace, J., Kabat, P., Lelieveld, J., Malhi, Y., Manzi, A.O., Meixner, F.X., Nobre, A.D., Nobre, C., Ruivo, M.D.L.P., Silva-Dias, M.A., Stefani, P., Valentini, R., Von Jouanne, J., Waterloo, M.J., 2002. Biogeochemical cycling of carbon, water, energy, trace gases, and aerosols in Amazonia: The LBA-EUSTACH experiments. J. Geophys. Res. D Atmos. 107. doi:10.1029/2001JD000524

Andreoli, R. V., Kayano, M.T., 2005. ENSO-related rainfall anomalies in South America and associated circulation features during warm and cold Pacific decadal oscillation regimes. Int. J. Climatol. 25, 2017–2030. doi:10.1002/joc.1222

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. J. Geophys. Res. 107, 8086. doi:10.1029/2002JD002704

Baldocchi, D., Falge, E., Gu, L., & Olson, R., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82, 2415. doi:http://dx.doi.org/10.1108/17506200710779521

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 Meteorol. 96, 257–291. doi:10.1023/A:1002497616547

Balvanera, P., Uriarte, M., Almeida-Leñero, L., Altesor, A., DeClerck, F., Gardner, T., Hall, J., Lara, A., Laterra, P., Peña-Claros, M., Silva Matos, D.M., Vogl, A.L., Romero-Duque, L.P., Arreola, L.F., Caro-Borrero, Á.P., Gallego, F., Jain, M., Little, C., de Oliveira Xavier, R., Paruelo, J.M., Peinado, J.E., Poorter, L., Ascarrunz, N., Correa, F., Cunha-Santino, M.B., Hernández-Sánchez, A.P., Vallejos, M., 2012. Ecosystem services research in Latin America: The state of the art. Ecosyst. Serv. 2, 56–70. doi: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 Ecol. Evol. 3, 327–338. doi:doi: 10.1111/j.2041-210X.2011.00172.x

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., Matson, P. a, Mooney, H. a, 2002. Principles of Terrestrial Ecosystem Ecology, System. doi:10.1007/978-1-4419-9504-9

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

Conant, R.T., Paustian, K., Elliott, E.T., 2001. Grassland management and conversion into grassland: effect on soil carbon. Ecol. Appl. 11, 343–355. doi: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, 2783–2792. doi:10.1890/07-0539.1

Delgado-Balbuena, J., Arredondo, J.T., Loescher, H.W., Huber-Sannwald, E., Chavez-Aguilar, G., Luna-Luna, M., Barretero-Hernandez, R., 2013. Differences in plant cover and species composition of semiarid grassland communities of central Mexico and its effects on net ecosystem exchange. Biogeosciences 10, 4673–4690. doi:10.5194/bg-10-4673-2013

Drew, C.A., Wiersma, Y.F., Huettmann, F., 2011. Predictive species and habitat modeling in landscape ecology: Concepts and applications. Predict. Species Habitat Model. Landsc. Ecol. Concepts Appl. 1–313. doi:10.1007/978-1-4419-7390-0

Eva, H.D., Belward, A.S., De Miranda, E.E., Di Bella, C.M., Gond, V., Huber, O., Jones, S., Sgrenzaroli, M., Fritz, S., 2004. A land cover map of South America. Glob. Chang. Biol. 10, 731–744. doi:10.1111/j.1529-8817.2003.00774.x

Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A., 2011. Modeing Species Distribution and Change Using Random Forest. Predict. Species Habitat Model. Landsc. Ecol. Concepts Appl. 139–159. doi:10.1007/978-1-4419-7390-0

Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., Bernhofer, C., Burba, G., Clement, R., Davis, K.J., Elbers, J.A., Goldstein, A.H., Grelle, A., Granier, A., Gumundsson, J., Hollinger, D., Kowalski, A.S., Katul, G., Law, B.E., Malhi, Y., Meyers, T., Monson, R.K., Munger, J.W., Oechel, W., Paw U, K.T., Pilegaard, K., Rannik, ??llar, Rebmann, C., Suyker, A., Valentini, R., Wilson, K., Wofsy, S., 2002. Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. Agric. For. Meteorol. 113, 53–74. doi: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 Sens. Environ. 112, 901–919. doi:10.1016/j.rse.2007.06.025

Grace, J., Lloyd, J., Mcintyre, J., Miranda, A., Meir, P., Miranda, H., Moncrieff, J., Massheder, J., Wright, I., Gash, J., 1995. Fluxes of carbon dioxide and water vapour over an undisturbed tropical forest in south-west Amazonia. Glob. Chang. Biol. 1, 1–12. doi:10.1071/PP97120

Graesser, J., Aide, T.M., Grau, H.R., Ramankutty, N., 2015. Cropland/pastureland dynamics and the slowdown of deforestation in Latin America. Environ. Res. Lett. 10. doi:10.1088/1748-9326/10/3/034017

Hargrove, W.W., Hoffman, F.M., Law, B.E., 2003. New analysis reveals representativeness of the AmeriFlux network. Eos, Trans. Am. Geophys. Union 84, 529. doi:10.1029/2003EO480001

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. Int. J. Climatol. 25, 1965–1978. doi:10.1002/joc.1276

Hinojo-Hinojo, C., Castellanos, A.E., Llano-Sotelo, J., Peñuelas, J., Vargas, R., Romo-Leon, J.R., 2018. High Vcmax, Jmaxand photosynthetic rates of Sonoran Desert species: Using nitrogen and specific leaf area traits as predictors in biochemical models. J. Arid Environ. 156, 1–8. doi:10.1016/j.jaridenv.2018.04.006

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: Walters, M., Scholes, R.J. (Eds.), The GEO Handbook on Biodiversity Observation Networks. Springer International Publishing, Cham, pp. 19–38. doi:10.1007/978-3-319-27288-7_2

Keller, M., Alencar, A., Asner, G.P., Braswell, B., Bustamante, M., Davidson, E., Feldpausch, T., Fernandes, E., Goulden, M., Kabat, P., Kruijt, B., Luizão, F., Miller, S., Markewitz, D., Nobre, A.D., Nobre, C. a., Priante Filho, N., da Rocha, H., Silva Dias, P., von Randow, C., Vourlitis, G.L., 2004. Ecological Research in the Large-Scale Biosphere– Atmosphere Experiment in Amazonia: Early Results. Ecol. Appl. 14, 3–16. doi: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. Ecol. Soc. Am. 6, 282–284.

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 Syst. Sci. Data Discuss. 1–25. doi: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 Tillage Res. 169, 65–70. doi: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 (Cop.). 34, 232–243. doi: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., Shanley, J.B., Likens, G.E., Haeuber, R., 2007. Who needs environmental monitoring ? Ecol. Soc. Am. 5, 253–260. doi: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. Proc. SPIE 8181, Earth Resour. Environ. Remote Sensing/GIS Appl. II, 81811K 8181, 1–7. doi: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. Divers. Distrib. 15, 59–69. doi: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. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

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

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 Ecol. Evol. 8, 1795–1803. doi: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: Walters, M., Scholes, R.J. (Eds.), The GEO Handbook on Biodiversity Observation Networks. Springer International Publishing, Cham, pp. 1–17. doi: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., Litvak, M., Ogle, K., Shaw, J.D., Wolf, A., Huntzinger, D.N., Schaefer, K., Cook, R., Wei, Y., Fang, Y., Hayes, D., Huang, M., Jain, A., Tian, H., 2017. Global patterns of drought recovery. Nature 548, 202–205. doi:10.1038/nature23021

Sierra, C.A., Mahecha, M., Poveda, G., Álvarez-Dávila, E., Gutierrez-Velez, V.H., Reu, B., Feilhauer, H., Anáya, J., Armenteras, D., Benavides, A.M., Buendia, C., Duque, Á., Estupiñan-Suarez, L.M., González, C., Gonzalez-Caro, S., Jimenez, R., Kraemer, G., Londoño, M.C., Orrego, S.A., Posada, J.M., Ruiz-Carrascal, D., Skowronek, S., 2017. Monitoring ecological change during rapid socio-economic and political transitions: Colombian ecosystems in the post-conflict era. Environ. Sci. Policy 76, 40–49. doi:10.1016/j.envsci.2017.06.011

Sulkava, M., Luyssaert, S., Zaehle, S., Papale, D., 2011. Assessing and improving the representativeness of monitoring networks: The European flux tower network example. J. Geophys. Res. Biogeosciences 116, 1–14. doi:10.1029/2010JG001562

Tonti, N.E., Gassmann, M.I., Pérez, C.F., 2018. First results of energy and mass exchange in a salt marsh on southeastern South America. Agric. For. Meteorol. 263, 59–68. doi:10.1016/j.agrformet.2018.08.001

Vargas, R., Yépez, E. a., Andrade, J.L., Ángeles, G., Arredondo, T., Castellanos, a. E., Delgado-Balbuena, J., Garatuza-Payán, J., González Del Castillo, E., Oechel, W., Rodríguez, J.C., Sánchez-Azofeifa, a., Velasco, E., Vivoni, E.R., Watts, C., 2013. Progress and opportunities for monitoring greenhouse gases fluxes in Mexican ecosystems: The MexFlux network. Atmosfera 26, 325–336.

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. Agric. For. Meteorol. 262, 423–433. doi:10.1016/j.agrformet.2018.07.016

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

Vitousek, P.M., Sanford, R.L., 1986. Nutrient Cycling in Moist Tropical Forest. Annu. Rev. Ecol. Syst. 17, 137–167. doi:10.1146/annurev.es.17.110186.001033

Whiting, G.J., Chanton, J.P., 2001. Greenhouse carbon balance of wetlands: Methane emission versus carbon sequestration. Tellus, Ser. B Chem. Phys. Meteorol. 53, 521–528. doi:10.1034/j.1600-0889.2001.530501.x

Chapter 5

CONCLUSIONS

This research comprehends different aspects on EONs management, going from their representativeness to its designs, which comprehends the quantitative delineation of spatial sampling domains, selecting locations for potential new study-sites, and the need for interoperability among the institutions involved.

5.1 Key conclusions

The key conclusions from this research are summarized below:

- EFTs provide an alternative approach to assess the representativeness of EONs, as this analysis complements previous studies which are only based on climatic or vegetation structural characteristics, also, EFTs addresses the interests for considering alternative information on ecosystem functionality. This study can also provide insights for EONs design and improvement.

- We provided a transparent methodology to optimally design an EON, our results show that strategically located study-sites could results in a higher representativeness of GPP and ET at a national-scale that study-sites selected under a bottom-up approach. However, in order to properly monitor GPP and ET across Mexico a higher degree of interoperability should be achieve among the different institutions involved.

- These results provided insights about the environmental properties (climate, terrain properties, soil resources) along with GPP and ET that are represented and under-represented by FLUXNET across LA, along with the location of strategically located sites to improve FLUXNET representativeness. This study can encourage the affiliation of existing eddy-covariance study sites to FLUXNET, in order to increase EONs representativeness.

5.2 Future directions

The findings presented in this dissertation provides insights about key knowledge gaps in EONs representativeness and designs. However, questions inevitably remain after these knowledge gaps are partially addressed. The representativeness assessment of AmeriFlux and NEON based on the spatial and temporal information derived from EFT lacks the assessment of AmeriFlux active study-sites at an annual scale, since AmeriFlux study/sites are subject to be shut-down depending on funding availability. The proposed framework for EONs designs in order to be implemented it requires a more detailed studies, which would likely consider the spatial distribution of the study sites with other spatial information such as road maps, topography, population centers, research centers along with alternatives that are logistically suitable to install an eddy-covariance study-site. The location of FLUXNET study sites used to assess its represents across LA not necessarily represents the current wealth of information of the network, since not all study-sites have provided information to FLUXNET and not all of them is currently active. Also, the addition of information relate it to natural and human-induced disturbance should be included in order to have a more detailed representativeness picture for the study-sites affiliated to FLUXNET across LA.

5.3 Final thoughts

The "take home message" of this dissertation is that the addition of the functional dimension at ecosystem scale can complement current studies based only on climate and plant functional types information. Also, the used of species distribution model are suitable to assess the representativeness of EONs since both challenges requires to 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 capture by presence data or study-sites locations.

Appendix A.1

SPATIAL REPRESENTANTIVENESS OF AMERIFLUX, NEON AND COMBINED CORE SITES

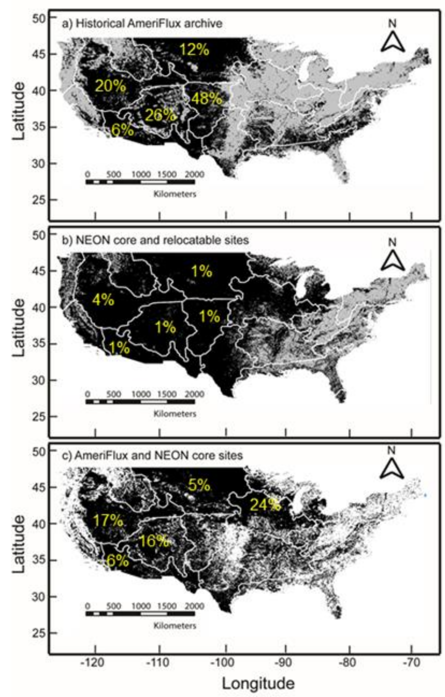


Figure A.1. Spatial representativeness by each network based on the cross-validation results derived from the Maximum entropy analysis having has environmental covariates ecosystem functional types (i.e., EFTmode) and ecosystem functional types inter-annual variability (i.e., EFTint) combined across CONUS for the 2001-2014 period. The spatial representativeness is estimated based on the ratio of those pixel with a Kappa index equal to 1 divided by the total pixel across each NEON ecoclimatic domain. Represented areas are marked as grey while black areas correspond to non-represented areas. The percentage on each map correspond to the five least represented NEON domains for: (a) historical AmeriFlux archive; (b) NEON sites; and (c) current AmeriFlux and NEON core sites. Represented areas are marked as grey and non-represented areas are marked as black

Appendix B.1

ENVIRONMENTAL DRIVERS OF GROSS PRIMARY PRODUCTIVITY AND EVAPOTRANSPIRATION

Table B.1. Environmental variables selected to represent the environmental factors that constrain the dynamic of gross primary productivity (GPP) and evapotranspiration (ET). Those parameters were selected/discarded based on their nugget/sill ratio obtained from semi-variograms.

Environmental Variable	**Nugget/Sill**	**Spatial class**
Bioclimatic predictors		
Annual Mean Temperature	0.61/15 = 0.04	S
Mean Diurnal Range (Mean of monthly (max temp – min temp))	0.01/5.4 < 0.01	S
Isothermality	0.01/77 < 0.01	S
Temperature seasonality	338/114695 < 0.01	S
Maximum temperature of warmest month	1.9 / 13 = 0.15	S
Minimum temperature of coldest month	0.01/48 < 0.01	S
Temperature annual range	0.77 / 33225 < 0.01	S
Mean temperature of wettests quarter	0.24 / 19 < 0.01	S
Mean temperature of driest quarter	0.01/98 < 0.01	S
Mean temperature of warmest month	0.78 / 14 = 0.05	S
Mean temperature of coldest month	0.32 / 30 = 0.01	S
Annual Precipitation	4326/12359090 < 0.01	S
Precipitation of wettest month	0.01/23453 < 0.01	S
Precipitation of driest month	1.4 / 102 = 0.01	S
Precipitation seasonality (coefficient of variation)	2.2 / 423 < 0.01	S
Precipitation of wettest month	0 / 216374 < 0.01	S
Precipitation of driest month	0.01 / 1286 < 0.01	S
Precipitation of warmest month	1585 / 7939862 < 0.01	S
Precipitation of coldest month	22 / 3865 < 0.01	S
Functional parameters		
Ecosystem Functional Productivity	0.09 / 1.2 = 0.07	S
Ecosystem Functional Seasonality	0.09 / 1.2 = 0.07	S
Ecosystem functional type inter-annual	1.1 / 4.6 = 0.23	S

variability		
Land Cover inter-annual variability	0.26 / 0.42 = 0.61	M
Terrain parameters		
Digital Elevation Model	12313 / 2187271 < 0.01	S
Digital Elevation Model	12313 / 2187271 < 0.01	S
Analytical Hillshading	0.06 / 0.26 = 0.23	S
Slope	0.07 = 0.14	S
Aspect	1.6 / 3.2 = 0.5	M
Cross-Sectional Curve	318894588/338415554 = 0.94	W
Longitudinal Index	382376203/419428126 = 0.91	W
Convergence Index	219 / 219 = 1	W
Closed depression	456 / 456 = 1	W
Flow accumulation	0.28 / 0.57 = 0.49	M
Topographic wetness index	2.1 / 12 = 0.17	S
LS factor	9.7 / 14 = 0.69	M
Channel Network Base level	0.1 / 391928 < 0.01	S
Vertical distance to channel network	12503 / 185186 = 0.06	S
Valley depth	0.01 / 464727 < 0.01	S
Relative slope position	0.01 / 0.07 = 0.14	S
Soil Parameters		
Soil Organic Carbon	85 / 426 = 0.20	S
Parent material age	151 / 757 = 0.20	S

Appendix C.1

PRINCIPAL COMPONENT ANALYSIS OF THE ENVIRONMENTAL DRIVERS SELECTED

Table C.1. The first two principal components comprehend 35% and 19% of the variability of all the environmental variables that were selected. The following table represents the correlation matrix between principal components PC1 and PC2 with those selected environmental variables.

Eigenvector	Principal Component 1	Principal Component 2
Annual Mean Temperature	0.2397	-0.1794
Mean Temperature of Warmest Quarter	0.2117	-0.2150
Mean Temperature of the Coldest Quarter	0.2583	0.0360
Annual Precipitation	0.2258	0.1890
Precipitation of the Wettest Month	0.2095	0.1919
Precipitation of the Driest Month	0.1912	0.1034
Precipitation Seasonality	-0.0588	0.0445
Precipitation of the Wettest Quarter	0.1989	0.2009
Precipitation of the Driest Quarter	0.1895	0.1001
Precipitation of Warmest Quarter	0.2181	0.1885
Precipitation of Coldest Quarter	0.1636	0.0923
Annual Mean Diurnal Range	-0.2306	-0.0671
Isothermality	0.098	0.2297
Temperature Seasonality	-0.1623	-0.2383
Maximum Temperature	0.0907	-0.3321
Minimum Temperature	0.2696	0.0242

Coldest Month		
Annual Temperature Range	-0.2177	-0.1872
Mean Temperature of the Wettest Month	0.1421	-0.3146
Mean Temperature of the Driest Month	0.2564	-0.0873
Ecosystem Functional Type-Productivity	0.2151	0.1598
Ecosystem Functional Type-Seasonality	-0.0777	-0.1696
Ecosystem Functional Type-Interannual Variability	-0.0424	-0.0048
Land Cover Inter-Annual Variability	0.0401	0.0403
Soil Organic Carbon	0.1566	0.4522
Rock Strata Age	-0.0046	-0.0070
Digital Elevation Model	-0.1959	0.2459
Topographic Wetness Factor	0.0644	0.1813
LS Factor	0.0341	-0.1028
Channel Network Base Level	-0.1951	0.1039
Vertical Distance to Channel Network	-0.0939	0.3051
Valley Depth	0.2172	-0.1290
Relative Slope Position	-0.1571	0.2826
Analytical Hillshading	-0.0030	0.0076
Slope	-0.0158	0.0658
Aspect	-0.0022	0.0202
Flow Accumulation	0.0155	-0.0821

Appendix D.1

HIERARCHICAL CLASSIFICATION OF ECOLOGICAL SIMILAR AREAS

Table D.1. The hierarchical delineation of the general ESAs and sub-ESAs was based on the silhouette scores (S(i)). S(i) provides a measure of the cohesion and separation of the elements within and among clusters, respectively. The first delineation consisted in delineating seven, twenty-one, fifty-one and one hundred and forty-one clusters using K-means and selecting the higher S(i) scores, those groups were selected as they are the groups already defined by the CEC-CONABIO classification, we selected 7 general ESA's as they had the highest S(i). The second delineation was performed through an iterative process for each ESA. This iterative process started with a number of groups (n) equal to 2 and stopped when the Si for n classes was higher than the Si for n+1 classes.

<table>
<tr><td colspan="4">1st Delineation</td></tr>
<tr><td>N=7</td><td>N=21</td><td>N=51</td><td>N=141</td></tr>
<tr><td>S(i) = 0.20</td><td>S(i) = 0.17</td><td>S(i) = 0.16</td><td>S(i) = 0.17</td></tr>
</table>

<table>
<tr><td colspan="7">2nd Delineation</td></tr>
<tr><td>ESA1</td><td>ESA2</td><td>ESA3</td><td>ESA4</td><td>ESA5</td><td>ESA6</td><td>ESA7</td></tr>
<tr><td>N=6</td><td>N=6</td><td>N=3</td><td>N=4</td><td>N=3</td><td>N=2</td><td>N=3</td></tr>
<tr><td>S(i)=0.23</td><td>S(i)=0.21</td><td>S(i)=0.23</td><td>S(i)=0.19</td><td>S(i)=0.25</td><td>S(i)=0.27</td><td>S(i)=0.23</td></tr>
</table>

Appendix E.1

SPATIAL REPRESENTATIVENESS OF DIFFERENT LAND COVERS ACROSS LATIN AMERIC

Table E.1. Representativeness for all the variables assessed at each IGBP class.

Bioclimatic		
IGBP class	**Mean**	**SD**
Evergreen needleleaf forest	0.49	0.48
Evergreen broadleaf forest	0.44	0.47
Deciduous Needleleaf Forest	NA	NA
Deciduous Broadlead Forest	0.15	0.33
Mixed Forest	0.51	0.48
Closed Shrubland	0.57	0.45
Open Shrubland	0.18	0.37
Woody Savana	0.48	0.47
Savana	0.43	0.47
Grasslands	0.09	0.26
Permanent Wetland	0.25	0.40
Cropland	0.21	0.39
Urban and Build-Up	0.29	0.44
Cropland/Natural vegetation mosaic	0.36	0.45
Snow and Ice	0.04	0.18
Barren Land	0.11	0.30

Terrain parameters		
IGBP class	**Mean**	**SD**
Evergreen needleleaf forest	0.62	0.41
Evergreen broadleaf forest	0.40	0.43
Deciduous Needleleaf Forest	NA	NA
Deciduous Broadlead Forest	0.20	0.37
Mixed Forest	0.75	0.35
Closed Shrubland	0.60	0.39
Open Shrubland	0.42	0.42
Woody Savana	0.37	0.40
Savana	0.28	0.37
Grasslands	0.26	0.39
Permanent Wetland	0.31	0.43
Cropland	0.15	0.31
Urban and Build-Up	0.40	0.42
Cropland/Natural vegetation mosaic	0.32	0.39
Snow and Ice	0.33	0.37
Barren Land	0.62	0.44

Soil Resources		
IGBP class	**Mean**	**SD**
Evergreen needleleaf forest	0.46	0.41
Evergreen broadleaf forest	0.38	0.35
Deciduous Needleleaf Forest	NA	NA
Deciduous Broadlead Forest	0.40	0.39
Mixed Forest	0.52	0.38
Closed Shrubland	0.72	0.35
Open Shrubland	0.45	0.37
Woody Savana	0.55	0.38
Savana	0.36	0.35
Grasslands	0.47	0.38
Permanent Wetland	0.53	0.38
Cropland	0.51	0.36
Urban and Build-Up	0.52	0.36
Cropland/Natural vegetation mosaic	0.54	0.37
Snow and Ice	0.57	0.38
Barren Land	0.58	0.40

Combined Environmental Parameters (PCA)		
IGBP class	**Mean**	**SD**
Evergreen needleleaf forest	0.92	0.26
Evergreen broadleaf forest	0.39	0.45
Deciduous Needleleaf Forest	NA	NA
Deciduous Broadlead Forest	0.47	0.47
Mixed Forest	0.88	0.28
Closed Shrubland	0.90	0.23
Open Shrubland	0.50	0.45
Woody Savana	0.57	0.46
Savana	0.37	0.45
Grasslands	0.44	0.46
Permanent Wetland	0.38	0.45
Cropland	0.49	0.46
Urban and Build-Up	0.47	0.45
Cropland/Natural vegetation mosaic	0.57	0.45
Snow and Ice	0.95	0.16
Barren Land	0.45	0.36

GPP		
IGBP class	**Mean**	**SD**
Evergreen needleleaf forest	0.54	0.49
Evergreen broadleaf forest	0.50	0.49
Deciduous Needleleaf Forest	NA	NA
Deciduous Broadlead Forest	0.51	0.49
Mixed Forest	0.70	0.45
Closed Shrubland	0.67	0.46
Open Shrubland	0.67	0.46
Woody Savana	0.52	0.49
Savana	0.43	0.49
Grasslands	0.46	0.49
Permanent Wetland	0.59	0.49
Cropland	0.51	0.49
Urban and Build-Up	0.61	0.49
Cropland/Natural vegetation mosaic	0.74	0.43
Snow and Ice	0.77	0.42
Barren Land	0.22	0.42

ET		
IGBP class	**Mean**	**SD**
Evergreen needleleaf forest	0.12	0.32
Evergreen broadleaf forest	0.58	0.49
Deciduous Needleleaf Forest	NA	NA
Deciduous Broadlead Forest	0.25	0.44
Mixed Forest	0.32	0.46
Closed Shrubland	0.06	0.25
Open Shrubland	0.24	0.42
Woody Savana	0.99	0.01
Savana	0.11	0.31
Grasslands	0.13	0.34
Permanent Wetland	0.08	0.27
Cropland	0.04	0.20
Urban and Build-Up	0.11	0.31
Cropland/Natural vegetation mosaic	0.57	0.49
Snow and Ice	0.06	0.23
Barren Land	0.09	0.28

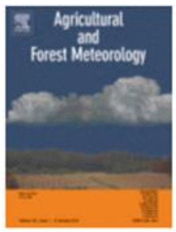
Appendix F.1

Reprint Permission Letter

3/25/2019 Rightslink® by Copyright Clearance Center

Copyright Clearance Center RightsLink®

Home | Account Info | Help



Title: Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States

Author: Samuel Villarreal,Mario Guevara,Domingo Alcaraz-Segura,Nathaniel A. Brunsell,Daniel Hayes,Henry W. Loescher,Rodrigo Vargas

Publication: Agricultural and Forest Meteorology

Publisher: Elsevier

Date: 15 November 2018

Logged in as:
Samuel Villarreal

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK | CLOSE WINDOW