

VEGETATION DYNAMICS IN NORTHERN
SOUTH AMERICA ON DIFFERENT TIME
SCALES

Dissertation
(kumulativ)

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt dem Rat der Chemisch-Geowissenschaftlichen Fakultät der
Friedrich-Schiller-Universität Jena
von Lina M. Estupinan Suarez, MSc
geboren am 11. September 1986 in Kolumbien

Gutachter:

1. _____

2. _____

Tag der Verteidigung: _____

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.

Ort,

Datum,

Unterschrift der Verfasserin/des Verfassers

Dedictory

To my mom Nidya, my siblings Milena and Juan Pablo who have supported me in all my outings and are always by my side. To my niece Julieta who I have watched to grow up from abroad. And to my dad, Antonio, who accompanies me everywhere.

Summary

Before satellite imagery was freely accessible, the assessment of vegetation activity at regional and global scales was unthinkable. Today, we can access more than three decades of data records associated with vegetation activity at fine spatial and temporal resolutions. New data products are also being released that are improving rapidly our ability to analyze ecosystem properties at multiple scales. All these satellite data products and model outputs associated with vegetation activity contain information of multiple processes that occur at different time scales. However, there are challenges in studying certain ecosystem processes that can be obscured by the dominance of a few processes that overwhelm the overall signal. In addition, handling efficiently these data sets implies certain technological challenges, especially when working across time in a geographical domain.

The overarching goal of this doctoral thesis was to understand the dynamics of vegetation activity occurring across time scales globally and in a regional context. To achieve this, I took advantage of open data sets, novel mathematical approaches for time series analyses, and state-of-the-art technology to effectively manipulate and analyze time series data. Specifically, I disentangled the longest records of vegetation greenness (> 30 years) in tandem with climate variables at 0.05° for a global scale analysis. Later, I focused my analysis on a particular region, northern South America (NSA), to evaluate vegetation activity at seasonal and interannual scales using moderate spatial resolution (0.0083°). NSA is dominated by natural land cover classes such as broadleaved evergreen forest (71.8 %), grasslands (7.1 %) and shrublands (4.0 %). In addition, it plays a key role in the water and carbon cycle at regional and global scales.

Two main approaches were used in this research; time series decomposition and dimensionality reduction analysis. Time series decomposition was carried out through the Fast Fourier Transformation (FFT), which transforms the data from the time domain to the frequency domain. It was used in Chapter 3 to disentangle time series of vegetation greenness and climate into short-term, seasonal cycle and longer-term oscillation after linear detrending. In addition, FFT was used in Chapter 4 to estimate the contribution of annual and semiannual oscillations to the variability of vegetation activity, and to compute a seasonality ratio (*SR*) that informs on whether vegetation activity for each pixel has one or two peaks during a year.

Dimensionality reduction analysis was carried out pixel-wise in Chapter 4 and over

the entire domain of NSA in Chapter 5. Specifically, I used Principal Component Analysis (PCA) to reduce a data set of vegetation variables acquired by satellites or reported by models. The resulting leading component (i.e., PC1) was used as a proxy of vegetation activity and as an input variable to assess modes of seasonality pixel-wise (Chapter 4). Furthermore, a variant of PCA called Global PCA was applied to reduce the space and dimension of the variables. Similarly, the resulting leading component was used to assess vegetation activity during different ENSO phases (Chapter 5). All analyses were contextualized within an ecological context (e.g., biotic units, ecoregions).

Results from my first analysis showed the regional dominance of spatio-temporal processes occurring beyond the seasonal cycle (Chapter 3). Thus, 27 % of NDVI variability was dominated either by short- (18 %) or longer-term oscillations (9 %) that are predominantly found in broadleaved evergreen forest, shrublands, and herbaceous covers, respectively. Also, the dominant co-variability map between NDVI and climate highlighted regional differences with a strong seasonal cycle of NDVI and temperature in the northern latitudes. Long-term oscillations were mostly found in the southern latitudes, and more complex vegetation climate interactions were found in tropical South America and South East Asia.

For the regional-level analysis, I contributed to the development of the Regional Earth System Data Lab (RegESDL). The RegESDL is a data infrastructure at moderate spatial resolution (0.0083°) for NSA (Chapter 4). It relies on two cores; analysis ready data cubes, and the Earth System Data Lab software. It efficiently supports multivariate statistics and time series analysis because of its data storage configuration and rapid access to any data cube dimension (e.g., time, space, variables). The RegESDL integrates data sets from Earth science, biodiversity and ecosystems from global and national sources, and contributes to inter-disciplinary research. It is an open source tool ready for big data analytics, and to facilitate the ongoing challenges of increasing satellite products and model output.

As a first implementation of the RegESDL, I assessed modes of seasonality in vegetation activity in NSA (Chapter 4). The vegetation activity proxy from PCA had a different performance at capturing data variability between land cover classes (~ 0.7 in grasslands and herbaceous cover, and ~ 0.33 in broadleaved evergreen forest). The *SR* shows heterogeneous seasonal patterns of vegetation in NSA. A key finding was that peaks of seasonality observed with remote sensing do not always correspond to climate seasonality driven by precipitation. Also, the variety of seasonality trajectories and timing underline the heterogeneity of ecosystems in this biodiverse region. The obtained results suggest that different environmental conditions drive vegetation seasonality, and local drivers have to be further investigated.

In my last study, I compared the vegetation activity in NSA during the ENSO phases

(Chapter 5). I observed contrasting variability of vegetation activity during the same ENSO phase on the southern Pacific coast in comparison with the northern inland savannas and the Caribbean coast. Also, I found that the strongest correlations between the vegetation activity and MEI are in the drier ecoregions during El Niño events. This finding highlights the role of arid and semiarid ecosystems in NSA, and their vulnerability during the ENSO cold phase.

In conclusion, assessing vegetation-climate dynamics at different temporal scales facilitates the observation and understanding of processes that are often obscured by one or few dominant processes. On the one hand, the global analysis showed the dominant seasonality of vegetation and temperature in northern latitudes in comparison with the heterogeneous patterns of the tropics, and the remarkable longer-term oscillations in the southern hemisphere. On the other hand, the regional analysis showed the complex and diverse land-atmosphere interactions in NSA when assessing seasonality and interannual variability of vegetation activity associated with ENSO. Overall, disentangling these processes and assessing them separately allows one to formulate new hypotheses of mechanisms in ecosystem functioning, reveal hidden patterns of climate-vegetation interactions, and inform about vegetation dynamics relevant for ecosystems conservation and management.

Zusammenfassung

Bevor Satellitenbilder frei zugänglich waren, war die Bewertung der Vegetationsaktivität auf regionaler und globaler Ebene undenkbar. Heute können wir auf mehr als drei Jahrzehnte an Datensätzen zur Vegetationsaktivität mit hoher räumlicher und zeitlicher Auflösung zugreifen. Zusätzlich werden neue Datenprodukte veröffentlicht, die unsere Fähigkeit zur Analyse von Ökosystemeigenschaften auf verschiedenen Ebenen rasch verbessern. Die Gesamtheit der Satellitendaten und Modelldaten zur Vegetationsaktivität enthält Informationen über verschiedene Prozesse, die auf unterschiedlichen Zeitskalen ablaufen. Dabei gibt es Herausforderungen bei der Untersuchung bestimmter Ökosystemprozesse, die durch die Dominanz einiger weniger Prozesse, die das Gesamtsignal überlagern, verdeckt werden können. Darüber hinaus bringt die effiziente Handhabung dieser Datensätze technologische Herausforderungen mit sich, insbesondere wenn in einem geografischen Gebiet über einen längeren Zeitraum gearbeitet wird. Das übergreifende Ziel dieser Doktorarbeit war es, die Dynamik der Vegetationsaktivität über Zeitskalen hinweg global und in einem regionalen Kontext zu verstehen. Um dies zu erreichen, nutzte ich offene Datensätze, neuartige mathematische Ansätze für Zeitreihenanalysen und modernste Technologien zur effektiven Bearbeitung und Analyse von Zeitreihendaten. Insbesondere habe ich die längsten Aufzeichnungen von Vegetationsindices (> 30 Jahre) in Verbindung mit Klimavariablen auf einer räumlichen Auflösung von $0,05^\circ$ für eine Analyse auf globaler Ebene entschlüsselt. Später konzentrierte ich meine Analyse auf eine bestimmte Region, das nördliche Südamerika (NSA), um die Vegetationsaktivität auf saisonaler und interannualer Ebene mit einer moderaten räumlichen Auflösung ($0,0083^\circ$) zu bewerten. Das NSA wird von natürlichen Bodenbedeckungsklassen wie immergrünen Laubwäldern (71,8 %), Grasland (7,1 %) und Buschland (4,0 %) dominiert. Darüber hinaus spielt es eine Schlüsselrolle im Wasser- und Kohlenstoffkreislauf auf regionaler und globaler Skala.

In dieser Untersuchung wurden zwei Hauptansätze verwendet: Zeitreihenzerlegung und Dimensionalitätsreduktionsanalyse. Die Zeitreihenzerlegung wurde mit Hilfe der schnellen Fourier-Transformation (FFT) durchgeführt, die die Daten vom Zeit- in den Frequenzraum transformiert. Sie wurde in Kapitel 3 verwendet, um die Zeitreihen der Vegetation und des Klimas in kurzfristige, saisonale Zyklen und längerfristige Oszillationen nach Abzug eines linearen Trends zu zerlegen. Darüber hinaus wurde die FFT in Kapitel 4 verwendet, um den Beitrag jährlicher und halbjährlicher Oszillationen

zur Variabilität der Vegetationsaktivität abzuschätzen und um ein Saisonalitätsverhältnis (SV) zu berechnen, das Aufschluss darüber gibt, ob die Vegetationsaktivität für jeden Pixel einen oder zwei Maxima im Jahr aufweist. Die Dimensionalitätsreduktionsanalyse wurde in Kapitel 4 pixelweise und in Kapitel 5 für den gesamten Bereich des NSA durchgeführt. Konkret verwendete ich die Hauptkomponentenanalyse (PCA), um einen Datensatz von Vegetationsvariablen zu reduzieren, die von Satelliten erfasst oder von Modellen gemeldet wurden. Die sich daraus ergebende Hauptkomponente (d. h. PC1) wurde als Proxy für die Vegetationsaktivität und als Eingangsvariable für die Bewertung der saisonalen Modalitäten auf Pixelbasis verwendet (Kapitel 4). Darüber hinaus wurde eine Variante der PCA, die globale PCA, angewandt, um räumliche Dimensionen sowie die verschiedenen Variablen selbst zu reduzieren. Die sich daraus ergebende Hauptkomponente wurde verwendet, um die Vegetationsaktivität während verschiedener ENSO-Phasen zu bewerten (Kapitel 5). Alle Analysen wurden in einen ökologischen Kontext eingeordnet (z. B. biotische Einheiten, Ökoregionen). Die Ergebnisse meiner ersten Analyse zeigten die regionale Dominanz von raum-zeitlichen Prozessen, die außerhalb des saisonalen Zyklus stattfinden (Kapitel 3). So wurden 27 % der NDVI-Variabilität entweder von kurz- (18 %) oder längerfristigen Oszillationen (9 %) dominiert, die vor allem in immergrünen Laubwäldern, Gebüschern bzw. krautigen Beständen zu finden sind. Auch die Karte der vorherrschenden Kovariabilität zwischen NDVI und Klima zeigte regionale Unterschiede mit einem starken saisonalen Zyklus von NDVI und Temperatur in den nördlichen Breitengraden. Langfristige Oszillationen wurden hauptsächlich in den südlichen Breitengraden festgestellt, und komplexere Wechselwirkungen zwischen Vegetation und Klima wurden im tropischen Südamerika und in Südostasien gefunden.

Für die Analyse auf regionaler Ebene war ich an der Entwicklung des Regional Earth System Data Lab (RegESDL) beteiligt. Das RegESDL ist eine Dateninfrastruktur mit moderater räumlicher Auflösung (0,0083°) für NSA (Kapitel 4). Es stützt sich auf zwei Kerne: analysefähige Datenwürfel und die Earth System Data Lab-Software. Es unterstützt effiziente multivariate Statistiken und Zeitreihenanalysen aufgrund seiner Datenspeicherkonfiguration und des schnellen Zugriffs auf jede beliebige Datendimension (z. B. Zeit, Raum, Variablen). Die RegESDL integriert Datensätze aus den Bereichen Geowissenschaften, Biodiversität und Ökosystemforschung aus globalen und nationalen Quellen und trägt zur interdisziplinären Forschung bei. Es handelt sich um ein Open-Source-Tool für Big-Data-Analysen, welches die laufenden Herausforderungen der zunehmenden Zahl von Satellitenprodukten und Modellausgaben erleichtern soll. Als erste Anwendung von RegESDL habe ich die Arten der Saisonalität

der Vegetationsaktivität im NSA bewertet (Kapitel 4). Das Maß für die Vegetationsaktivität aus der PCA hatte eine unterschiedliche Eignung bei der Erfassung der Datenvariabilität zwischen den Landbedeckungsklassen ($\sim 0,7$ bei Grasland und krautiger Bedeckung und $\sim 0,33$ bei immergrünem Laubwald). Die SV zeigt heterogene saisonale Muster der Vegetation im NSA. Eine wichtige Erkenntnis war, dass die mit Fernerkundung beobachteten Maxima der Saisonalität nicht immer mit der klimatischen Saisonalität übereinstimmen, die durch den Niederschlag bestimmt wird. Auch die Vielfalt der saisonalen Verläufe und Zeitpunkte unterstreicht die Heterogenität der Ökosysteme in dieser artenreichen Region. Die Ergebnisse deuten darauf hin, dass die Saisonalität der Vegetation von verschiedenen Umweltbedingungen abhängt, und dass lokale Faktoren weiter untersucht werden müssen.

In meiner letzten Studie habe ich die Vegetationsaktivität im NSA während verschiedener ENSO-Phasen verglichen (Kapitel 5). Ich stellte fest, dass sich die Variabilität der Vegetationsaktivität während derselben ENSO-Phase an der südlichen Pazifikküste im Vergleich zu den nördlichen Binnensavannen und der Karibikküste unterscheidet. Außerdem stellte ich fest, dass die stärksten Korrelationen zwischen der Vegetationsaktivität und der MEI in den trockeneren Ökoregionen während El-Nino-Ereignissen bestehen. Diese Erkenntnis unterstreicht die Rolle der ariden und semiariden Ökosysteme im NSA und ihre Anfälligkeit während der ENSO-Kältephase.

Zusammenfassend lässt sich sagen, dass die Bewertung der Vegetations-Klimadynamik auf verschiedenen zeitlichen Skalen die Beobachtung und das Verständnis von Prozessen erleichtert, die oft durch einen oder wenige dominante Prozesse überlagert werden. Einerseits zeigte die globale Analyse die dominante Saisonalität von Vegetation und Temperatur in den nördlichen Breiten im Vergleich zu den heterogenen Mustern in den Tropen sowie die auffälligen längerfristigen Oszillationen in der südlichen Hemisphäre. Andererseits zeigte die regionale Analyse die komplexen und vielfältigen Wechselwirkungen zwischen Land und Atmosphäre im NSA bei der Bewertung der Saisonalität und der interannuellen Variabilität der Vegetationsaktivität in Verbindung mit ENSO. Insgesamt ermöglicht die Entflechtung dieser Prozesse und ihre getrennte Betrachtung die Formulierung neuer Hypothesen über Mechanismen in der Funktionsweise von Ökosystemen, die Aufdeckung verborgener Muster von Wechselwirkungen zwischen Klima und Vegetation sowie Informationen über die Vegetationsdynamik, die für die Erhaltung und Bewirtschaftung von Ökosystemen relevant sind.

Contents

Summary	vii
Zusammenfassung	xi
1 Introduction	1
1.1 Motivation	1
1.2 Research questions	3
1.2.1 RQ1. Vegetation-climate variability across time scales	3
1.2.2 RQ2. Deployment of the Regional ESDL	4
1.2.3 RQ3. Modes of seasonality on vegetation activity in NSA	4
1.2.4 RQ4. Imprints of ENSO on vegetation activity in NSA	5
1.3 Thesis structure	5
2 Background	7
2.1 Methods for studying temporal dynamics	7
2.1.1 Vegetation variables derived from space	7
2.1.2 Time series decomposition	9
2.1.3 Principal component analysis	10
2.1.4 Data cubes and the Earth system data lab	12
2.2 Regional study: Northern South America	13
2.2.1 Climate	13
2.2.1.1 Annual cycle	13
2.2.1.2 Interannual variability	14
2.2.2 Vegetation	16
2.2.3 Vegetation-climate interactions	17
3 Towards a global understanding of vegetation–climate dynamics at multiple timescales	19
4 A regional Earth system data lab for understanding ecosystem dynamics: An example from tropical South America	39

5	Spatial patterns of vegetation activity related to ENSO in northern South America	61
5.1	Introduction	63
5.2	Methods	65
5.2.1	Study area	65
5.2.2	Vegetation variables	66
5.2.3	Global Principal Component Analysis	67
5.2.4	Time-lagged correlations between ENSO phases and the vegetation activity component	67
5.2.5	Hotspots of correlations between ENSO and vegetation activity by ecoregions	68
5.3	Results	68
5.3.1	Global PCA	68
5.3.2	Vegetation activity during different ENSO phases along lags	69
5.3.3	Hotspots of ENSO impacts on vegetation activity	71
5.4	Discussion	74
5.4.1	The vegetation activity component from the Global PCA	74
5.4.2	Lagged effects of ENSO in vegetation activity in NSA	75
5.4.3	ENSO hotspots by ecoregions	76
5.4.4	Limitations and opportunities	78
5.5	Conclusion	78
5.6	Acknowledgments	79
6	Discussion	81
6.1	Dominant modes of variability in NDVI and climate variables across time scales	81
6.2	Analysis of vegetation activity in northern South America	83
6.2.1	Diverse modes of seasonality of vegetation activity in NSA	83
6.2.2	Heterogeneous spatio-temporal activity of vegetation during ENSO phases in NSA	85
6.3	Opportunities from big data infrastructures	87
6.4	New avenues to study vegetation activity across time scales	88
7	Concluding remarks	91
	Bibliography	93
	Declaration on authorship	105
	Acknowledgments	109

A	Supplementary figures	111
B	Supplement manuscript 1	115
C	Supplement manuscript 2	141
D	Supplement manuscript 3	153

List of Figures

2.1	Time series decomposition of temperature of air at 2 m (T) using the Fast Fourier Transformation. Lat. $5.75^{\circ}N$, Lon. 70.25° . Decomposed oscillations are unitless (Data source: ERA5)	11
2.2	Workflow of the Global PCA. Lat: Latitude. Lon: Longitude. M: Number of pixels. N: Number of variables. t: time steps.	12
2.3	Data visualization of the Regional Earth System Data Lab. Example of three variables represented in the spatio-temporal domain. (Source: Mahecha, 2019)	13
2.4	Mean monthly values of precipitation (P, gray bars) and temperature of air at 2 m (T, yellow lines) from 1981 to 2010 at two different gauging stations in northern South America. Left: Unimodal precipitation regime Lat. 2.37° , Lon. -74.64° . Right: Bimodal precipitation regime Lat. 5.26° , Lon. -74.73° (Data source: IDEAM 2021).	14
2.5	Time series of the Multivariate ENSO index (MEI).	15
5.1	Map of northern South America. Grey polygons are selected ecoregions from Olson et al. (2001). The numbers indicate the ecoregions name. The black lines are the country borders.	69
5.2	Map of correlation anomalies. The color map represents the difference in lagged Spearman correlations between MEI and the leading vegetation activity component (VAC) for (a) the El Niño and Neutral phase, and (b) the La Niña and Neutral phase. The black lines are country borders.	72
5.3	Median values of the lagged Spearman correlation between VAC and MEI by ecoregions. Ecoregions are in rows. From left to right: first column shows the ecoregion location. Second column represents the median correlations within the ecoregion by ENSO phases. The ribbon correspond to the standard deviation. Data distribution from El Niño, La Niña and Neutral phases by lags is showed in columns three to five respectively. Showed ecoregions have the highest correlation and an area larger than 50000 km^2	73

A.1	Lagged Pearson correlation between the Multivariate ENSO Index and Isomap (PCA) components pixelwise. The showed value correspond to the highest lagged correlation between MEI and respective components from (a) Isomap and (b) PCA. (c) Difference between Isomap and PCA lagged correlations.	113
A.2	Time series of the Multivariate ENSO Index (MEI) and a selected output component from multidimensionality regression analysis by watersheds. The assessed variable was global primary productivity from BESS. MEI has been standardized to $\mu = 0$ and $\sigma = 1$. The showed component was selected because it has the highest absolute correlation with MEI. Watersheds are from HydroBASINS (level 4), and are in the rows. The dimensionality reduction methods are in the columns. Time is in the x-axis. Each subplot has information about the selected components, correlation value and respective lag in the legend.	114
D.1	Density plots of vegetation variables (y-axis) against PC ₁ (x-axis) for homogeneous land covers.	156
D.2	Absolute values of Spearman correlation (y-axis) versus precipitation of the driest month (x-axis) for all ecoregions in the study area.	156
D.3	Median values of the lagged Spearman correlation between MEI and precipitation (source: Global Precipitation Climatology Project) by ecoregions. Ecoregions are in rows. From left to right: first column shows the ecoregion location. Second column is the median correlation values within the ecoregion by ENSO phases. The ribbon correspond to the standard deviation. Data distribution from El Niño, La Niña and Neutral phases by lags is shown in columns three to five respectively. Shown ecoregions have the highest correlation and an area larger than 50000 km ²	157
D.4	Median values of the lagged Spearman correlation between MEI and air temperature at 2 m above the ground (source: source: ERA-Interim 80 km) by ecoregions. Ecoregions are in rows. From left to right: first column shows the ecoregion location. Second column is the median correlation values within the ecoregion by ENSO phases. The ribbon correspond to the standard deviation. Data distribution from El Niño, La Niña and Neutral phases by lags is shown in columns three to five respectively. Shown ecoregions have the highest correlation and an area larger than 50000 km ²	158

List of Tables

2.1	Vegetation variables data set. ρ : Surface reflectance. C_1 and C_2 : Coefficients of the aerosol resistance. G : Gain factor. L : Canopy background adjustment. NIR : Near-infrared.	9
2.2	Climatic impacts of El Niño and La Niña in northern South America. *: Anomalies not reported. Source: NOAA (2021a).	16
5.1	Climatic impacts of El Niño and La Niña in northern South America. *: Anomalies not reported. Source: NOAA (2021a).	66
5.2	Loadings from the Global PCA by (normalized) variable. EVI: Enhanced vegetation index. FPAR: Fraction of absorbed photosynthetically active radiation. GPP: Gross primary productivity. NDVI: Normalized difference vegetation index. VAC: Vegetation activity component. PC: Principal component.	70
D.1	Explained fraction of variance by principal components (PC) from the Global PCA and the land cover PCA.	155
D.2	First principal component loadings from the Global PCA and the land cover PCA.	155

List of Abbreviations

ABI	Advanced Baseline Imager
ARDC	Analysis Ready Data Cubes
AVHRR	Advanced Very High Resolution Radiometer
CSIRO	Commonwealth Scientific and Industrial Research Organization
CO₂	Carbon Dioxide
ENSO	El Niño Southern Oscillation
EMD	Empirical Mode Decomposition
ESDL	Earth System Data Lab
EVI	Enhanced Vegetation Index
FFT	Fast Fourier Transformation
FPAR	Fraction of Absorbed Photosynthetically Active Radiation
GIMMS	Global Inventory Modeling and Mapping Studies
GPP	Gross Primary Productivity
ITCZ	Intertropical Convergence Zone
IPCC	Intergovernmental Panel on Climate Change
MEI	Multivariate Enso Index
MODIS	MODerate Resolution Imaging Spectroradiometer
MSC	Mean Seasonal Cycle
NIR	Near-InfraRed
NIR_v	Near-InfraRed Reflectance of Terrestrial Vegetation
NDVI	Normalized Difference Vegetation Index
NSA	Northern South America
PCA	Principal Component Analysis
RegESDL	Regional Earth System Data Lab
SIF	Sun-Induced Chlorophyll Fluorescence
SR	Seasonality Ratio
TS	Time Series
VAC	Vegetation Activity Component

Chapter 1

Introduction

1.1 Motivation

Undoubtedly, with more than five decades of instrumental records, different lines of evidence have shown that the Earth system is changing due to anthropogenic forcing. Since the 1960s, studies have revealed that changes in atmospheric carbon dioxide (CO₂) concentration have a direct impact on climate. Specifically, Manabe and Wetherald (1967) reported how doubling atmospheric CO₂ has the effect of increasing the global surface temperature by 2.3 °C. A decade later, the global positive trend of atmospheric CO₂ concentration was clearly observed on Mauna Loa Observatory's records, with another important remark, a strong annual seasonality (Keeling et al., 1976). And a few years after, the first scientific report by climate experts, explaining the consequences of burning fossil fuels in the Earth's system, was published (Charney et al., 1979). Today, it is known that increasing global temperatures have an effect on carbon uptake by vegetation, extreme weather, ocean acidification, among others (Arias et al., 2021a). Undeniably, living on a warming planet is one of the greatest challenges humanity will face in this century.

Increasing the understanding of the Earth system has become of pivotal importance. Yet, an important task is to disentangle processes occurring at different time scales in the biosphere and atmosphere. An outstanding example is the work by Keeling et al. (1976), who distinguished the two main components (processes) of the atmospheric CO₂ records from Mauna Loa Observatory. They determined the positive trend of the gas concentration and the annual variability driven by the vegetation seasonality from the northern hemisphere. Disentangling multiple processes allows for a more comprehensive perspective of mechanisms in the Earth's system over time. One approach to achieve this is through time series analysis of data recorded systematically at regular time intervals.

Currently, global data of vegetation and climate variables, acquired from satellite retrievals or models, is available for time series analysis. This has facilitated the study of vegetation-climate interactions, which is essential for a better understanding of the

terrestrial carbon cycle. In general, vegetation fixes carbon by photosynthesis and releases it by respiration. Carbon uptake by vegetation is modulated by different factors such as vegetation type, water availability, radiation, among others (Bonan, 2015; Jarvis, 1976; Sellers et al., 1997). Importantly, processes involved in these interactions occur at different time scales; from seconds (e.g. stomatal aperture) to decades (interannual climate variability e.g. El Niño Southern Oscillation) (Braswell et al., 2005; Mahecha et al., 2010; Stoy et al., 2009).

In addition, regional analysis is becoming the focus of attention in recent years. As it is explained in the sixth assessment report of the Intergovernmental Panel on Climate Change (IPCC), regional climate can be moderated or amplified by local forcing, soil moisture feedbacks, land use/cover change, decadal/multi-decadal natural variability, among others (Arias et al., 2021a). Moreover, it is known that changes in the Earth system will not be evenly distributed in time or space. In consequence, understanding the past and present climate-vegetation interactions within a regional context is of prominent relevance. Such studies could be used to benchmark early warning systems, extreme events detection and, hopefully, facilitate a better adaptation to climate change.

One of the regions of greatest interest is the tropics. On one hand, the tropical rainforest (e.g., the Amazon) is considered one of the largest terrestrial carbon sinks (Baccini et al., 2017; Pan et al., 2011). On the other hand, tropical ecosystems are counted as a major source of uncertainty in processed-based and data-driven models (Jiang and Ryu, 2016; Jung et al., 2020). This is mostly explained by two different factors; (i) the low availability and quality of satellite retrievals from optical sensors because of frequent cloud cover; and, (ii) limited ground data for models' calibration and validation.

This doctoral research investigates the variability of vegetation activity across time scales, through time series analysis of biosphere-atmosphere variables with a specific emphasis in northern South America (NSA). Firstly, I evaluated the co-variability of vegetation and climate globally using the longest records of the Normalized Difference Vegetation Index (NDVI), which is a proxy of vegetation greenness strongly related to productivity (Chapter 3). Here, the main goal was to disentangle climate-vegetation interactions occurring at different time scales over 30 years at coarse spatial resolution (0.5°).

Secondly, I used moderate spatial resolution (0.0083°) over 14 years to evaluate time series from a set of vegetation variables in NSA. I started deploying a Regional Earth System Data Lab (RegESDL), which facilitates pixel-wise analyses over the time dimension (Chapter 4). My first focus was a comprehensive characterization of the modes of seasonality of vegetation activity (Chapter 4). A second focus was to understand the

interannual variability looking at the imprints of El Niño Southern Oscillation (ENSO) on vegetation activity regionally (Chapter 5). In the regional studies, my approach was to combine a set of vegetation variables and to use the leading component from a dimensionality-reduction analysis as a proxy of vegetation activity (see section 2.1.1). Each of these studies are briefly introduced from sections 1.2.1 to 1.2.4 and extensively explained in the following chapters.

1.2 Research questions

The overarching goal of my thesis was *to understand dynamics of vegetation activity occurring across time scales globally and in a regional context*. The research questions (RQ) addressed in this research were:

1. **RQ1:** What are the modes of variability in vegetation greenness, and what is its co-variability with climate across time scales globally?
2. **RQ2:** How to carry out computationally efficient time series analysis in NSA with moderate spatial resolution?
3. **RQ3:** What are the modes of seasonality of vegetation activity in NSA?
4. **RQ4:** Where are the hotspots of vegetation activity observed during the ENSO phases in NSA?

1.2.1 RQ1. Vegetation-climate variability across time scales

The dominant annual seasonality of photosynthesis obscures climate-biosphere interactions occurring at other time scales. Revealing these patterns is key for having a differentiated understanding of ecosystems processes that will be affected by and will respond differently to changes in climate. In my first study, I investigated vegetation-climate co-variability using the longest records of global surface reflectance by satellites associated with vegetation and climate variables. The main research question addressed in this study was; *What are the modes of variability in vegetation greenness, and what is its co-variability with climate across time scales globally?*

Specifically, the aim was to characterize the co-variability between the activity of the terrestrial biosphere and climate, spatially and by land cover classes, and to propose a new classification scheme for vegetation taking into account the temporal component. To do this, I used time series of NDVI in combination with air temperature and precipitation. In general, NDVI estimates vegetation greenness, which is

strongly related to productivity. After a linear detrending, time series were decomposed into the short-term, seasonal cycle and longer-term oscillations. In particular, it was observed that the dominant modes of variability for NDVI, precipitation and temperature are different among all variables (Chapter 3). Also, they certainly vary across the spatial domain by land cover classes. Furthermore, one of the key findings was the remarkable heterogeneity of the dominant climate-vegetation co-variability in NSA. This highlights the regional complexity in terms of the vegetation-climate interactions across time scales. Based on this, the following analyses were narrowed to assess the seasonal and interannual variability of vegetation activity at higher spatial resolution in the region.

1.2.2 RQ2. Deployment of the Regional ESDL

This work has involved a considerable amount of technical development, especially at the regional scale. This is because computer access to time series is often limited by the data storage configuration. Traditionally, spatio-temporal data layers are stored on a latitude-longitude grid. This configuration is not suitable for rapid data sampling along the time dimension, which is key for efficient time series analyses. To overcome this constraint, I addressed the second research question; *How to carry out computationally efficient time series analysis in NSA with moderate spatial resolution?* As a main result, I developed the “Regional Earth System Data Lab” (RegESDL) using moderate spatial resolution. The RegESDL significantly improves computational timing for time series analysis and facilitates multivariate statistics. In summary, the implementation of computationally intensive methods is easily and effectively implemented at 0.0083° (~ 1 km at the equator) over the time domain, but also for any other dimension in NSA.

1.2.3 RQ3. Modes of seasonality on vegetation activity in NSA

Vegetation cover in NSA is largely dominated by broadleaf evergreen forest (71.8 %), grasslands (7.1 %), and shrublands (4.0 %) (ESA, 2017). Ecosystems with a strong seasonality are wetlands, savannas and the tropical dry forest, however seasonality remains a controversial topic for evergreen tropical forests. Regionally, from the climate perspective, it is known that precipitation and radiation are one of the main drivers of vegetation activity (González-M et al., 2021; Junk, Bayley, and Sparks, 1989; Nemani et al., 2003). Particularly, precipitation has two main annual regimes; unimodal (one peak) and bimodal (two peaks), which are strongly bound to geographical location (Knoben, Woods, and Freer, 2018; Poveda, Waylen, and Pulwarty, 2006). Until now, it is unknown whether these modes of annual seasonality in precipitation are also observed in vegetation activity.

In my second study, I investigated *the modes of seasonality of vegetation activity in NSA*. To achieve this, I assessed different vegetation variables; the Enhanced Vegetation Index (EVI), Fraction of Absorbed Photosynthetically Active Radiation (FPAR), modeled Gross Primary Productivity (GPP) and NDVI (Chapter 4). Each variable carries different information on characteristics of vegetation, but it is unclear what is the best proxy for assessing seasonality of productivity of photosynthesis in the region. Therefore, at pursuing to capture the majority of information associated to seasonality of vegetation activity, I implemented a dimensionality reduction analysis pixel-wise. The resulting leading component was used as a proxy of vegetation activity. Next, I estimated the ratio of annual and semiannual oscillations using time series decomposition methods. The proposed seasonality ratio exposes the different modes of annual seasonality in tropical ecosystems that are driven by different environmental variables.

1.2.4 RQ4. Imprints of ENSO on vegetation activity in NSA

My last study is related to the interannual variability of vegetation activity in the region. ENSO is the dominant mode of interannual variability globally (Arias et al., 2021a), and it has a strong impact on precipitation in NSA which in turn impacts vegetation. Studies of ENSO impacts on vegetation are usually carried out for individual events (e.g. La Niña 2010-2011, El Niño 2015-2016) or ecosystems type. However, there is limited understanding of the main response of vegetation activity to several ENSO events. ENSO events cannot be easily generalized due to (i) the internal variability of ENSO itself, (ii) local forcing which might overrule its impacts. To tackle these issues, my last study sought *the identification of hotspots of vegetation activity observed during the ENSO phases in NSA*. My results showed contrasting spatio-temporal responses along the region. In general, a stronger correlation of vegetation activity is observed during El Niño (ENSO warm phase) than during La Niña (ENSO cold phase). Moreover, the most vulnerable ecoregions to El Niño are arid and semiarid ecosystems (Chapter 5).

1.3 Thesis structure

In the next chapter (Chapter 2) I expand the description of the scientific background and concepts that are key for my research. The analysis of global modes of variability of vegetation greenness and climate is presented in Chapter 3. In Chapter 4 and 5, I investigate the seasonal cycle and the interannual variability of vegetation activity, respectively, in NSA. The main findings are contextualized and discussed in Chapter 6.

Chapter 2

Background

2.1 Methods for studying temporal dynamics

2.1.1 Vegetation variables derived from space

Vegetation variables estimated from satellite retrievals or models provided the opportunity to monitor ecosystems globally. These variables are mostly based on optical sensors that capture surface reflectances. The main principle here is that plants use differentiated regions of the electromagnetic spectrum to convert light into energy. Specifically, photosynthetic pigments absorb red wavelengths for photosynthesis, whereas near-infrared is highly reflected. The difference between both reflectances is the basis for the computation of vegetation indices (Huete et al., 1997; Tucker and Sellers, 1986). Vegetation indices were the first global measurements used to assess vegetation activity since Earth observation satellites were launched (Ryu, Berry, and Baldocchi, 2019). Currently, they are considered surrogates of productivity, and usually referred to as vegetation "greenness". The main reason is that vegetation indices do not necessarily and immediately reflect the impact of environmental conditions on photosynthesis; they are nonlinear with respect to biomass, and saturate in dense canopies (Camps-Valls et al., 2021; Huete et al., 1997; Huete et al., 2002).

In the last decades, significant advances on global estimates of photosynthesis were possible. The integration of frameworks from remote sensing, plant physiology, climate modeling, and micrometeorology paved the way for the generation of GPP products (Ryu, Berry, and Baldocchi, 2019). Different methods have been developed to estimate GPP such as; (i) semi-empirical models based on light use efficiency models in tandem with differentiated coefficients for specific biomes (Running, Mu, and Zhao, 2015); (ii) process-based models that aim to represent atmosphere-vegetation-soil fluxes (Jiang and Ryu, 2016); (iii) and, data-driven models based on up-scaling of carbon-water fluxes, provided by a network of eddy covariance towers, using machine learning methods (Jung et al., 2011, 2020; Tramontana et al., 2016).

An important attribute of global vegetation products is the data sampling frequency. These vegetation products rely on satellite retrievals either for direct computations, model calibration or algorithm training (Huete et al., 1997; Jiang and Ryu, 2016; Jung et al., 2020; Running, Mu, and Zhao, 2015; Ryu, Berry, and Baldocchi, 2019; Tucker and Sellers, 1986). Consequently, their frequency is strongly tied to the frequency at which satellites revisit every point on Earth's surface. In general, retrievals by Earth observation satellites are provided at regular time intervals. These systematic measurements provide regular observations over time, and guarantee the integrity of time series for each data product.

The longest satellite retrievals used to compute vegetation variables are from the Advanced Very High Resolution Radiometer (AVHRR). Particularly, the NDVI data set based on AVHRR, and produced by the Global Inventory Modeling and Mapping Studies (GIMMS), provides more than three decades of data at 0.0833° spatial resolution, with a bi-weekly temporal resolution (Pinzon and Tucker, 2014). Since the year 2000 until present, higher spatial resolution data has been acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS). It offers a larger set of vegetation variables such as vegetation indices, GPP and FPAR. Moreover, MODIS products are processed at different spatio-temporal resolutions; from daily imagery to monthly composites, starting from 250 m to 0.05° depending on the specific data product.

In this doctoral research, I used time series of four vegetation variables, i.e., NDVI, EVI, FPAR and GPP, summarized in Table 2.1. NDVI is the normalized difference between the near-infrared and red surface reflectances (Tucker and Sellers, 1986). NDVI is probably the vegetation variable most used from satellite retrievals. Nevertheless, it is limited in very dense vegetated areas such as tropical forests where it saturates. EVI is another vegetation index, but its computation includes the blue band, and aerosols coefficients to account for atmospheric and background noise, and to reduce saturation (Huete et al., 1997; Huete et al., 2002). FPAR refers to the energy absorption by the vegetation canopy within the 0.4 - 0.7 μm wavelength spectrum (Knyazikhin, 1999; Myneni et al., 1997). GPP is gross primary production, i.e. carbon up-take by photosynthesis. GPP products include environmental drivers as constraints on the carbon uptake by terrestrial vegetation.

Finally, the quality of these variables is directly related to the atmospheric conditions at the time of measurement. They all have been either acquired or derived from passive remote sensors that are highly affected by clouds and haze (Estupinan-Suarez, Leon, and Sarmiento Pinzon, 2017; Hilker et al., 2012). This is one of the major challenges of working in NSA, data here is constrained due to the presence of clouds associated to the Intertropical Convergence Zone (ITCZ), and the transfer of moisture from the Amazon to the Andes (Eltahir and Bras, 1994; Poveda, Waylen, and Pulwarty,

TABLE 2.1: Vegetation variables data set. ρ : Surface reflectance. C_1 and C_2 : Coefficients of the aerosol resistance. G : Gain factor. L : Canopy background adjustment. NIR : Near-infrared.

Index abbreviation	Computation	Data span	Product
NDVI	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	2001 to present	MOD13A2
EVI	$G \times \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + C_1 \times \rho_{Red} - C_2 \times \rho_{Blue} + L}$	2001 to present	MOD13A2
FPAR	Radiative transfer algorithm	2001 to present	MOD15A2
GPP	Process-based models	2001 to 2014	BESS GPP

2006; Zemp et al., 2014). In spite of this, all variables carry valuable information related to vegetation processes such as canopy greening, productivity or energy absorption. For the regional analyses, my approach was to merge these datasets of vegetation variables to extract the largest possible amount of information using Principal Component Analysis (PCA) (see section 2.1.3). Within the scope of this thesis, I refer to "*vegetation activity*" as the leading component obtained from the PCA, which also carries the largest data variability. In addition, I use "*vegetation dynamics*" to refer indistinguishably to the variability of any vegetation variable used in this research either as input or output.

2.1.2 Time series decomposition

Time series refers to a set of observations x_t , each one being recorded at a specific time t within regular intervals (Brockwell and Davis, 2002). In consequence, the time difference between contiguous data points is equal. Importantly, time series carry different characteristics or features related to intrinsic frequency variations (Braswell et al., 2005; Defriez et al., 2016; Mahecha et al., 2010). Nevertheless, these signals are obscured depending on the signals' amplitude and the signal to noise ratio. Stronger signals such as vegetation seasonality might obscure the observation of processes occurring at other time scales. To tackle this, time series decomposition methods have been developed to disentangle original signals into different components.

Time series analyses are carried out at two different domains; time and frequency. The time domain relies on the adjacency of points and dependency of the current value on the past value. The frequency domain assumes that periodic or sinusoidal variation is found on most data sets (Brockwell and Davis, 2002). As a result, data is partitioned in different frequencies, and then the associated variance is estimated for each.

Data adaptive models such as the Empirical Mode Decomposition (EMD) works directly in the time domain. They are recognized for capturing nonlinear processes which occur in natural systems (Huang and Wu, 2008; Huang et al., 1998). However, different challenges emerge when data adaptive approaches are implemented within a geographical domain. Data adaptive methods can result in fuzzy spatial patterns because outcomes from neighbor pixels might differ significantly. We faced this limitation when carrying out global analysis with EMD, or assessing complex ecological and topographical regions such as the tropics (see Chapter 3). In contrast, other methods based on sinusoidal assumptions such as Fast Fourier showed more homogeneous and consistent results spatially.

The Discrete Fast Fourier Transform is an algorithm that transforms the time domain into a frequency domain based on sinusoidal functions (Brockwell and Davis, 2002). Frequencies containing the highest variance are identified using a power spectrum. In particular, for my doctoral research, time series were decomposed into three components after a linear detrending. The short-term signal contains periods < 0.9 years, except for two seasonal harmonics. The seasonal cycle contains oscillations with periods from 0.9 to 1.1, plus two harmonics with periods of 0.5 and 0.33 years (i.e., semiannual and quadrennial harmonics). The longer-term signal enclose periods > 1.1 years. Figure 2.1 shows an example of the air temperature components after the Fast Fourier Transformation in a tropical savanna. On the upper panel is the original signal spanning from 2001 to 2010 with a strong seasonality, although the maximum variability is $\sim 0.5^\circ$. The linear trend and decomposed oscillations are in the middle and bottom panel.

2.1.3 Principal component analysis

In general, dimensionality reduction analysis reveals hidden intrinsic features in high-dimensional data sets. A well established dimensionality reduction method is PCA which uses recursively linear functions. PCA reconstructs new variables, referred to as principal components, from the initial variables in an orthogonal space. The objective is to maximize the information contained in the first components, so following analysis can be focused exclusively on a subset of principal components. In consequence, the dimensions of the data set can be reduced without losing as much information as possible (Jolliffe and Cadima, 2016; Pearson, 1901).

I used PCA to assess a set of vegetation variables at seasonal (Chapter 4) and interannual time scales in NSA (Chapter 5). The main reasons to use PCA were; (i) the variance explained by each principal component is easily calculated; (ii) the contribution of each variable to each principal component is known, and therefore variables containing the most information are identified.

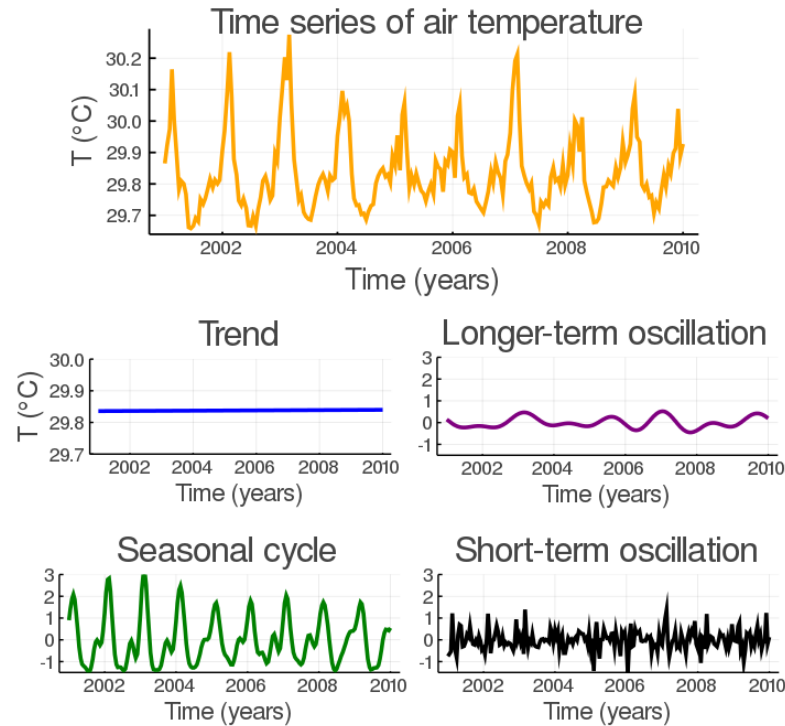


FIGURE 2.1: Time series decomposition of temperature of air at 2 m (T) using the Fast Fourier Transformation. Lat. $5.75^{\circ}N$, Lon. 70.25° . Decomposed oscillations are unitless (Data source: ERA5)

In both regional analyses presented in this dissertation, the first principal component (PC1) was used as a proxy of vegetation activity. Specifically, the seasonality analysis carried out in Chapter 4 used PC1, the leading component, as input to estimate annual and semiannual oscillations pixel-wise. In Chapter 5, I implemented a variant of PCA called Global PCA (Kraemer et al., 2020) to assess vegetation activity during the ENSO phases. Using the Global PCA approach, data was reduced over the spatial and variable dimensions.

The Global PCA steps, which are similar to a standard PCA, are explained below. First of all, data was scaled to zero mean and unit variance because the highest explained variance could be driven by the variables magnitude and not by intrinsic data variability (Kraemer, Reichstein, and Mahecha, 2018). The next steps are summarized in Figure 2.2 and are as follow; (i) create a matrix per time step including all land pixels from all variables; (ii) compute a covariance matrix per time step; (iii) combine the covariance matrix per time step into a global covariance matrix by element-wise mean; (iv) calculate eigenvectors from the global covariance matrix; and, (v) project data pixel-wise to the new PCA space. To do this, the eigenvector matrix is transposed (i.e. projection matrix), and multiplied with the transpose input matrix. An important remark is that the selection of the most informative eigenvectors is based on their

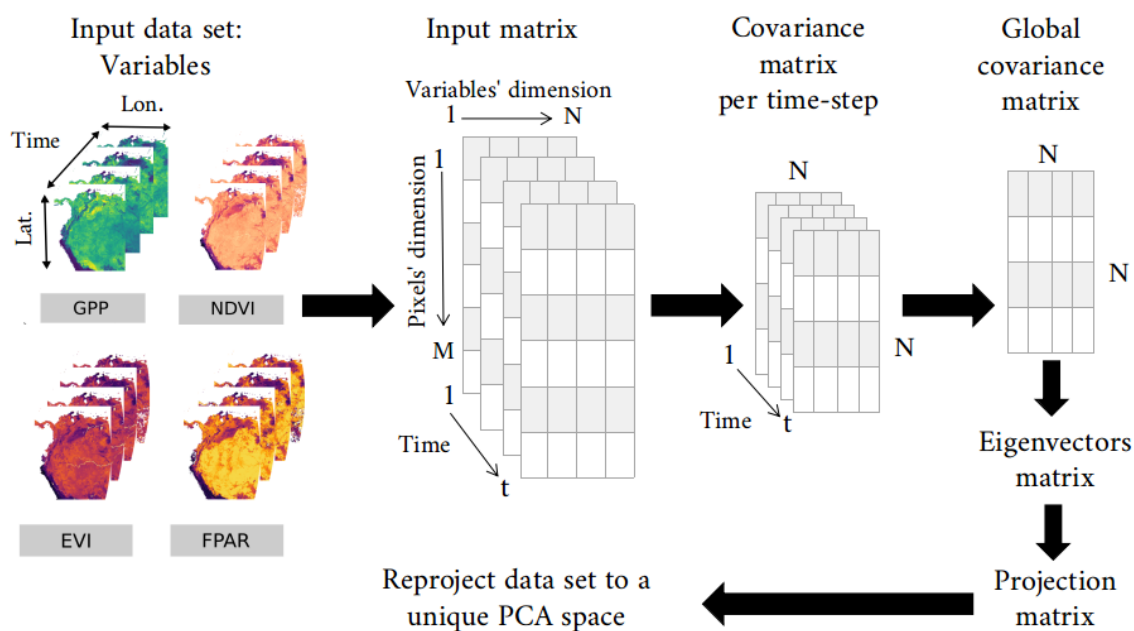


FIGURE 2.2: Workflow of the Global PCA. Lat: Latitude. Lon: Longitude.
M: Number of pixels. N: Number of variables. t: time steps.

scalar values (i.e., eigenvalues). In other words, the higher the eigenvalue, the more relevant the principal component.

2.1.4 Data cubes and the Earth system data lab

Timely and efficient access to data is a priority for up-to-date ecosystem assessment. This is key for investigating ecosystem responses to climate change, developing early warning systems, and monitoring programs. Today, this can be achieved with the increasing volume of satellite products and models, but also requires rapid advances in technology. Therefore, access to big data and processing capabilities can be a major challenge for researchers, especially for those in regions where funding for long-term projects such as data infrastructures are rarely absent. In addition to funding and infrastructure limitations, some of the current challenges posed by big data are: redundancy in data storage and preprocessing, increased computing power, interoperability, among others.

Data Cubes have become a worldwide data management strategy to handle big data. They are basically a multidimensional matrix that allows one to conveniently store geographic layers over time. For example, the Open Data Cube project, led by the Commonwealth Scientific and Industrial Research Organization (CSIRO), began by focusing on satellite products such as MODIS and Landsat (Lewis et al., 2017). Today,

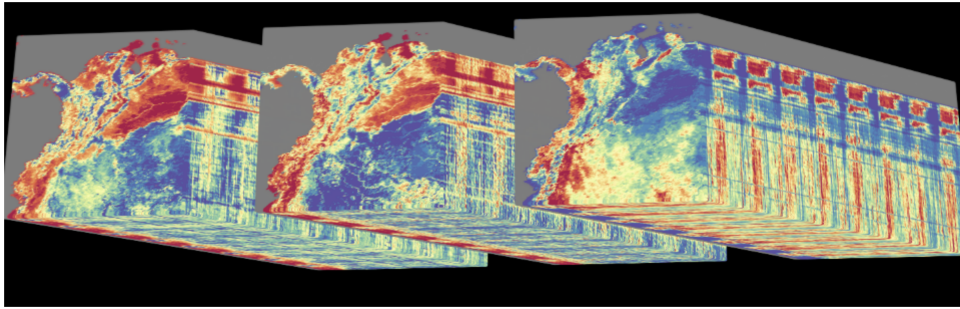


FIGURE 2.3: Data visualization of the Regional Earth System Data Lab. Example of three variables represented in the spatio-temporal domain. (Source: Mahecha, 2019)

the best-known platform is the Google Earth Engine, which stores geographic data from multiple sources and enables cloud computing. However, in these approaches, data is stored in a latitude-longitude grid that renders access to the temporal dimension inefficiently (Gorelick et al., 2017; ODC, 2021).

The Earth System Data Lab (ESDL) is a big data processing technology based on Data cubes. It was developed at the Max Planck Institute for Biogeochemistry by Mahecha et al. (2020). It relies on two cores; (i) analysis ready data cubes (ARDCs); and (ii) the ESDL software described by Gans (<https://esa-esdl.github.io/ESDL.jl/latest/>). ARDCs provide a harmonized spatio-temporal grid, and allows efficient access to the time dimension, which is key for time series analysis. The ESDL software supports the implementation of user-defined functions that are carried out efficiently to any cube dimension based on the user preference. I implemented a higher spatial resolution ESDL at 0.0083° 8-daily for NSA called the RegESDL (Figure 2.3), it facilitates big data analysis over the time dimension in the NSA region.

2.2 Regional study: Northern South America

2.2.1 Climate

2.2.1.1 Annual cycle

The annual hydro-climatological cycle in NSA is dominated by the migration of the ITCZ, and its interaction with the regional and local atmospheric circulation (Poveda, Waylen, and Pulwarty, 2006). Additionally, the Andes and the Amazon play a key role in the climate of this region. The Andes enhances wind convection due to orographic lifting. This interaction is easily observed between the Choco low-level jet and the western flank of the Andes, where annual precipitation ranges between 8000 to 10000 mm (Poveda and Mesa, 2000; Yepes et al., 2019). On the other hand, precipitation in the

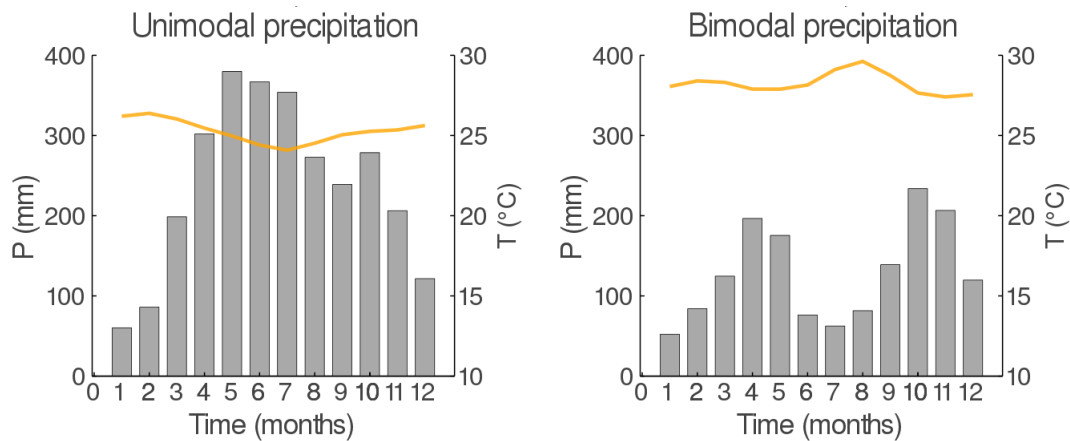


FIGURE 2.4: Mean monthly values of precipitation (P, gray bars) and temperature of air at 2 m (T, yellow lines) from 1981 to 2010 at two different gauging stations in northern South America. Left: Unimodal precipitation regime Lat. 2.37°, Lon. -74.64°. Right: Bimodal precipitation regime Lat. 5.26°, Lon. -74.73° (Data source: IDEAM 2021).

Amazon is the combination of multiple factors. It results from strong land-atmosphere interactions such as precipitation recycling (Zemp et al., 2017a), moisture transport from the Atlantic, and the ITCZ migration (Poveda, Waylen, and Pulwarty, 2006).

There are two main modes of precipitation in the region that result from the interaction between climate and topography. Unimodal precipitation regimes refer to one dry season and one wet season within the year (i.e., one peak of rainfall). Whereas bimodal precipitation refers to two dry seasons and two wet seasons (double peak of rainfall) (Knoben, Woods, and Freer, 2018). Compared to the strong seasonality of precipitation, temperature is mostly constant throughout the year. Figure 2.4 shows that the monthly mean values of temperature differ by only a few degrees (~ 3 °C) at two selected gauging stations. In the same figure, we observe a unimodal precipitation regime with a peak of rainfall between May and August, and a bimodal regime with a first peak in May-June and a second peak in October-November. Little has been explored about what is the response of vegetation activity to these different precipitation regimes. In this dissertation, I assessed modes of seasonality on vegetation activity to understand whether they can be related to precipitation regimes.

2.2.1.2 Interannual variability

ENSO is the dominant mode of interannual variability of climate globally (Arias et al., 2021a). It is a coupled climate phenomena between the atmosphere and the Pacific ocean. ENSO has three phases; El Niño (the warm phase), La Niña (the cold phase)

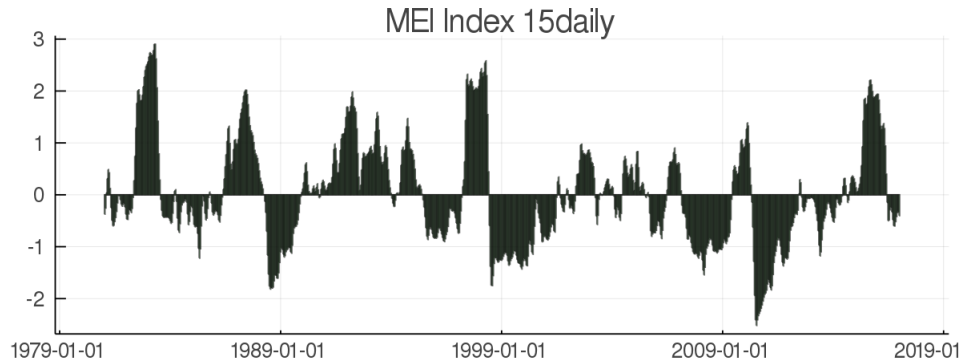


FIGURE 2.5: Time series of the Multivariate ENSO index (MEI).

and the Neutral phase. During the same phase, positive and negative changes in temperature and precipitation occur worldwide.

Different ENSO indices have been proposed based on predefined regions over the Pacific ocean, and selected atmospheric and oceanic variables. Specifically, the Multivariate ENSO Index Version 2.0 (MEI) is estimated from sea level pressure, sea surface temperature, surface zonal winds, surface meridional winds, and outgoing longwave radiation between 30°S – 30°N and 100°E – 70°W . MEI is the resulting leading principal component calculated of an empirical orthogonal function. MEI values are classified as follows: ≤ -0.5 correspond to La Niña whereas values ≥ 0.5 to El Niño, values from -0.5 to 0.5 correspond to the Neutral phase (NOAA, 2021b). Because MEI is computed from a set of variables, it is less sensitive to perturbations from a single variable. Another advantage of working with MEI is that the index does not carry signals of seasonality which facilitates time series analysis at interannual scales. Figure 2.5 shows MEI time series from 1980 to 2019.

It is known that ENSO causes anomalies in precipitation and temperature, which in turn affect vegetation activity. The two strongest ENSO events in the last decades were La Niña 2010–2011 and El Niño 2015–2016. During La Niña 2010–2011, it was observed that; (i) 40 % of the variance in global net primary productivity was explained by this event (Bastos et al., 2013); and, (ii) 5 % increased in methane emissions by wetlands (Pandey et al., 2017). During El Niño 2015–2016, Liu et al. (2017) reported that the tropics released 2.5 ± 0.34 gigatons more carbon than during other periods. Their results also showed heterogeneous climate forcing along the tropics, accompanied by the superposition of regional conditions such as soil and vegetation structure, and previous disturbance. In this sense, Liu et al. (2017) highlighted that changes in vegetation during El Niño must not be generalized to a single response. In this sense, the understanding of changes in terrestrial vegetation as affected by ENSO require more specific analyses at local and regional scales.

TABLE 2.2: Climatic impacts of El Niño and La Niña in northern South America. *: Anomalies not reported. Source: NOAA (2021a).

Region	El Niño		La Niña	
	Austral summer	Austral winter	Austral summer	Austral winter
Pacific coast	Warm and wet	Warm	Cool and dry	Cool
Caribbean coast	*	Warm and dry	*	Cool and wet
Amazon	Dry	*	Wet	*

In general, previous assessments of vegetation activity during El Niño and La Niña have been carried out for individual events. These analysis have contributed to explain annual changes in vegetation greenness and productivity (Bastos et al., 2013; Liu et al., 2017; Luo et al., 2018; Pandey et al., 2017; Patra et al., 2017; Schaik et al., 2018). Nevertheless, they are limited for detecting hotspots of ENSO effects over time. An exemption is the study by Kogan and Guo (2017), who analyzed both ENSO phases from 1982 to 2016 using vegetation health indices. Regions such as northern Brazil, southern Africa, eastern Australia, among others, were identified with moderate-to-strong stress (healthy) during El Niño (La Niña). However, their results were aggregated at the country level and do not elucidate differences between ecosystems. Yet, there is limited understanding on the differentiated responses by vegetation activity when assessing simultaneously different phases of ENSO using time series.

In NSA different hydro-climatological studies have reported changes in precipitation during ENSO phases (Bolaños et al., 2021; Cai et al., 2020; Poveda, Waylen, and Pulwarty, 2006; Poveda et al., 2001; Waylen and Poveda, 2002). In general, each ENSO phase does not manifest homogeneously across the region, but rather at different intensities and with contrasting patterns. Table 2.2 summarizes the climatic impacts of ENSO. An important remark is the opposite changes in the Caribbean and Pacific coast during the same ENSO phase.

2.2.2 Vegetation

NSA is dominated by natural ecosystems such as rainforests and savannas. They both are key for understanding regional and global dynamics in the water and carbon cycles at annual and interannual time scales (Holmgren et al., 2013). The Amazon forest, for example, is the largest terrestrial carbon sink (Baccini et al., 2017; Pan et al., 2011), and dominates water and moisture cycling in the continent (Poveda, Waylen, and Pulwarty, 2006; Zemp et al., 2017a). Further, savannas worldwide have been considered as a main driver of carbon interannual variability (Ahlström et al., 2015). In addition, wetlands,

including flooded savannas and swamp forest, are significant contributors of methane emissions (Bloom et al., 2017) that increases during La Niña (Pandey et al., 2017).

Understanding vegetation activity across time scales is still in an early phase. In NSA, seasonality of vegetation activity is mainly observed in water-limited ecosystems such as savannas, dry forest, and partially wetlands. In tropical rainforests, seasonality is still debated although data from tree rings (Giraldo et al., 2020), sun-induced chlorophyll fluorescence (SIF) (Schaik et al., 2018), and geostationary satellites (Hashimoto et al., 2021) suggest some annual cyclicity. Furthermore, the assessment of vegetation activity at other time scales such as interannual variability has been focused on the impacts of ENSO as explained in section 2.2.1.2 and 2.2.3.

In addition to this natural variability, ecosystems are facing rapid changes due to direct anthropogenic drivers such as selective logging, or extensive land cover change from forest to pastures, soy and palm plantations (Anaya et al., 2020; Armenteras et al., 2019; Barona et al., 2010; Clerici et al., 2020; Seymour and Harris, 2019; Song et al., 2018). Among the tropics, South America is the continent with the highest rate of forest loss, doubling the rates reported for Africa and South East Asia (Baccini et al., 2017). Clearly, South America is facing massive environmental pressure, threatening not only carbon and water ecosystem services, but also increasing its vulnerability to extreme events (Brando et al., 2019; Erfanian, Wang, and Fomenko, 2017). A deeper understanding in the region is needed to improve ecosystems conservation and management, develop early warning systems, and effective decision making.

2.2.3 Vegetation-climate interactions

At the annual scale, the main climatic drivers in tropical ecosystems are precipitation and solar radiation. In water limited ecosystems such as savannas, and tropical dry forest, plant seasonality is driven by precipitation (González-M et al., 2021; Siyum, 2020; Stan and Sanchez-Azofeifa, 2019). Similarly, the ecological functioning of most wetlands is driven by flood pulses controlled by the precipitation regime (Junk, Bayley, and Sparks, 1989; Junk et al., 2014). Regarding tropical forests, Nemani et al. (2003) reported solar radiation as the main driver of net primary production, but recently (Hashimoto et al., 2021) reinforced the role of precipitation.

At the interannual scale, anomalies of precipitation and temperature are mostly related to ENSO, which in turn also affects vegetation. Initially, Zeng, Mariotti, and Wetzel (2005) provided a global understanding of processes occurring in the tropics during El Niño. They generalized a dominant decline of precipitation that causes lower soil moisture and less net primary productivity. Complementary, higher temperatures

boost heterotrophic respiration. In combination, all these factors contribute to less carbon uptake by vegetation and higher carbon release by plants and soils to the atmosphere. Nevertheless, this simplified perspective is not applicable entirely to all tropical ecosystems, because sensitivities to the climate forcing differ by vegetation types. For example, Wang, Zeng, and Wang (2016) reported different sensitivities to temperature in the tropics, being stronger in the Amazon and weaker in African savannas and South Asian forests. Regarding precipitation, African savannas have stronger sensitivities than the Amazon that could be associated to a stronger soil moisture deficit in arid and semi-arid ecosystems.

Currently, we have a limited understanding of the diverse response of vegetation activity to climate forcing in tropical regions. This knowledge is pivotal when considering the urgent need of forecasting, and modeling ecosystems in a warming planet. In this sense, models are not considering the different variability across time scales. As well, policies for ecosystems conservation, and management take place at local and regional scales. Therefore analyses at higher spatio-temporal resolutions are critically needed to narrow-down the heterogeneous impacts of extreme weather on vegetation activity.

Chapter 3

Towards a global understanding of vegetation–climate dynamics at multiple timescales

First manuscript

Authors: Nora Linscheid[†], Lina M. Estupinan-Suarez[†], Alexander Brenning, Nuno Carvalhais, Felix Cremer, Fabian Gans, Anja Rammig, Markus Reichstein, Carlos A. Sierra, Miguel D. Mahecha

[†] These authors contributed equally to this work.

Status: Published

Journal: Biogeosciences, 17, 945–962, 2020. doi: 10.5194/bg-17-945-2020



Towards a global understanding of vegetation–climate dynamics at multiple timescales

Nora Linscheid^{1,2,★}, Lina M. Estupinan-Suarez^{1,3,★}, Alexander Brenning^{3,4}, Nuno Carvalhais^{1,5}, Felix Cremer^{3,6}, Fabian Gans¹, Anja Rammig², Markus Reichstein^{1,4,7}, Carlos A. Sierra¹, and Miguel D. Mahecha^{1,4,7}

¹Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany

²TUM School of Life Sciences Weihenstephan, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany

³Department of Geography, Friedrich Schiller University Jena, Löbdergraben 32, 07743 Jena, Germany

⁴Michael Stifel Center Jena for Data-Driven and Simulation Science, Ernst-Abbe-Platz 2, 07743 Jena, Germany

⁵Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT Universidade Nova de Lisboa, Caparica, Portugal

⁶Institute for Data Science, German Aerospace Center (DLR), 07745 Jena, Germany

⁷German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

★These authors contributed equally to this work.

Correspondence: Nora Linscheid (nlinsch@bgc-jena.mpg.de) and Lina M. Estupinan-Suarez (lestup@bgc-jena.mpg.de)

Received: 15 August 2019 – Discussion started: 17 September 2019

Revised: 23 December 2019 – Accepted: 14 January 2020 – Published: 24 February 2020

Abstract. Climate variables carry signatures of variability at multiple timescales. How these modes of variability are reflected in the state of the terrestrial biosphere is still not quantified or discussed at the global scale. Here, we set out to gain a global understanding of the relevance of different modes of variability in vegetation greenness and its covariability with climate. We used > 30 years of remote sensing records of the normalized difference vegetation index (NDVI) to characterize biosphere variability across timescales from submonthly oscillations to decadal trends using discrete Fourier decomposition. Climate data of air temperature (T_{air}) and precipitation (Prec) were used to characterize atmosphere–biosphere covariability at each timescale.

Our results show that short-term (intra-annual) and longer-term (interannual and longer) modes of variability make regionally highly important contributions to NDVI variability: short-term oscillations focus in the tropics where they shape 27 % of NDVI variability. Longer-term oscillations shape 9 % of NDVI variability, dominantly in semiarid shrublands. Assessing dominant timescales of vegetation–climate covariation, a natural surface classification emerges which captures patterns not represented by conventional classifications,

especially in the tropics. Finally, we find that correlations between variables can differ and even invert signs across timescales. For southern Africa for example, correlation between NDVI and T_{air} is positive for the seasonal signal but negative for short-term and longer-term oscillations, indicating that both short- and long-term temperature anomalies can induce stress on vegetation dynamics. Such contrasting correlations between timescales exist for 15 % of vegetated areas for NDVI with T_{air} and 27 % with Prec, indicating global relevance of scale-specific climate sensitivities.

Our analysis provides a detailed picture of vegetation–climate covariability globally, characterizing ecosystems by their intrinsic modes of temporal variability. We find that (i) correlations of NDVI with climate can differ between scales, (ii) nondominant subsignals in climate variables may dominate the biospheric response, and (iii) possible links may exist between short-term and longer-term scales. These heterogeneous ecosystem responses on different timescales may depend on climate zone and vegetation type, and they are to date not well understood and do not always correspond to transitions in dominant vegetation types. These scale de-

dependencies can be a benchmark for vegetation model evaluation and for comparing remote sensing products.

1 Introduction

Ecosystems and climate interact on multiple spatial and temporal scales. For example, the main driver of photosynthesis during the daily cycle typically is light availability, assuming no other resource limitation. At annual timescales, temperature can limit growth and development during certain phases of the year, particularly in the extratropics. While climate variability is traditionally very well characterized across timescales (e.g., Viles, 2003; Cao et al., 2012; Bala et al., 2010; Hannachi et al., 2017), it is less well known how the biosphere responds to variations in climate on different scales. Understanding the implications of such timescale dependencies of climate–vegetation interactions is challenging due to the variety of interwoven processes. These dependencies range from short-term climate extremes and biotic stress (e.g., insect outbreaks) to seasonal dynamics in climate-driven phenology and long-term dynamics that can again either reflect intrinsic ecosystem dynamics (e.g., vegetation successional dynamics) or climate-change- or land-use-induced process alterations. Investigating vegetation–climate dynamics globally across multiple timescales requires long-term observation on relevant vegetation dynamics and climate variables in combination with a method to separate ecosystem variability at different timescales.

The assessment of ecosystem variability, e.g., in responses to climate at the global scale, has only become feasible in the last decades. Long-term Earth observations (EOs) are now allowing us to assess ecosystem states consistently over more than 30 years. Vegetation indices such as the normalized difference vegetation index (NDVI) have often been interpreted as proxies for vegetation activity (Zeng et al., 2013; De Keersmaecker et al., 2015; Hawinkel et al., 2015; Kogan and Guo, 2017; Pan et al., 2018), despite well-known limitations of only reflecting vegetation greenness. While novel EOs may be more closely related to actual rates of photosynthesis (Sun-induced fluorescence, SIF; Guanter et al., 2007), NDVI from the Advanced Very High Resolution Radiometer (AVHRR) has the advantage of offering the longest updated records of vegetation remote sensing data every 15 d (d stands for day). In tandem with climate time series from the same period, this record provides a solid basis to globally assess biosphere–atmosphere interactions across timescales ranging from weeks to decades.

Temporal biosphere dynamics carry the imprint of different drivers across timescales, yet EOs can only record one integrated signal over time. This signal reflects a mixture of processes acting on different scales, which cannot be observed independently (Mahecha et al., 2007; Defriez and Reuman, 2017; Pan et al., 2018). Therefore, short-term

and long-term processes can be obscured by the dominant influence of the annual cycle (Braswell et al., 2005; Mahecha et al., 2010c). In order to study relevant ecosystem–climate interactions across temporal scales, information contained for each timescale thus first needs to be extracted from this integrated signal. Time series decomposition allows us to extract different frequencies such as annual, intra-annual, and interannual oscillations from vegetation and climate time series. Such approaches have proven useful, e.g., to characterize at what scales vegetation responses are dampened or amplified in comparison with their climate forcing (Stoy et al., 2009), how ecosystem variability is confined by hydrometeorological variability (Pappas et al., 2017), what scales of variability need to be considered to relate forcing variables and vegetation state comprehensively (Katul et al., 2001; Braswell et al., 2005), or to remove confounding effects from processes acting on longer timescales than the process in question (Mahecha et al., 2010b). However, to date most studies employing time series decomposition to study vegetation dynamics have focused on disentangling timescales from minutes to a few years based on flux data (Stoy et al., 2009; Katul et al., 2001; Mahecha et al., 2007, 2010c). Studies investigating long-term vegetation records by time series decomposition do exist but focus only on a specific region (Martínez and Gilabert, 2009; Canisius et al., 2007; Hawinkel et al., 2015) or do not provide cointerpretation with climate signals (Pan et al., 2018). Earth observation time series of vegetation and climate covering more than 30 years now allow us to characterize the timescale-resolved variability in the biosphere and its relation to climate globally across several decades. Additionally, the global coverage of these records allows one to attain a broader understanding in climate space and across vegetation types, which to date is equally lacking.

In this study, we set out to gain a global understanding of the relevance of the different modes of variability in vegetation greenness and its covariability with climate at timescales from submonthly oscillations to long-term trends. These timescale-specific vegetation–climate co-oscillations are expected to serve as a reference benchmark for comparing remote sensing products and terrestrial biosphere models. Specifically, we aim to (i) characterize variability of biosphere and climate time series explicitly on multiple timescales; (ii) understand spatial patterns of this scale-resolved variability and covariability globally; (iii) assess whether characteristic timescale-specific dynamics in the biosphere and climate relate to established climate classifications or land cover; and (vi) assess differences in correlations of biosphere with climate on short-term, seasonal, and longer-term timescales.

2 Methods

The code to produce all primary figures is made available as supplementary notebook (<https://doi.org/10.5281/zenodo.3611262>, Linscheid et al., 2019).

2.1 Data

A global gridded dataset of AVHRR NDVI was retrieved from the Global Inventory Monitoring and Modeling System (GIMMS, Pinzon and Tucker, 2014) at 15 d temporal and 0.083° spatial resolutions (GIMMS NDVI v3.1). Original data were aggregated to 0.5° by taking the mean of the corresponding 0.083° pixels. Corresponding records of air temperature (T_{air}) from the European Centre for Medium-Range Weather Forecasts (ERA-Interim v4, Dee et al., 2011) and precipitation (Prec) from the Multi-Source Weighted-Ensemble Precipitation (MSWEP, Beck et al., 2019) were aggregated to match temporal resolution by summation (Prec) or averaging (T_{air}). Spatial resolution of T_{air} was preserved (0.5°), while MSWEP values were averaged for spatial resampling (0.083 to 0.5°). Spatial and temporal resolution were fixed based on the coarsest resolution among the input datasets to ensure conservative results. The time period considered was from 1 January 1982 to 31 December 2015.

2.2 Preprocessing

Gaps in NDVI time series were filled with values from the mean seasonal cycle computed separately for each grid cell. Missing values were mostly present at high northern latitudes (Fig. S1 in the Supplement). Each time series (for each pixel) was normalized to zero mean and unit variance prior to performing fast Fourier transformation (FFT). For further analysis, the gap-filled data were discarded. Normalization, gap-filling, and FFT were performed in the Earth System Data Lab (<https://www.earthsystemdata.net/>, last access: 1 August 2019, Mahecha et al., 2019), using the implementation based on the programming language Julia. Analyses were performed on a latitude–longitude grid due to software and data considerations. In all spatial analyses on the latitude–longitude grid, the difference in size of grid cells between high latitudes and the Equator was accounted for through weighting values by grid cell size. Similarly, in all analyses that involved sampling of data points, the sampling frequency was weighted by grid cell size.

2.3 Time series decomposition

All pixel time series were first detrended using a linear model. We then used discrete FFT to decompose the detrended time series into underlying harmonic functions at different frequencies (Brockwell and Davis, 2006). The resulting Fourier spectra (Fig. S2) were reconstructed by inverse FFT into binned scale-specific subsignals for short-term, sea-

sonal, and longer-term oscillations: the seasonal signal was reconstructed from the Fourier spectrum at periods of 0.9–1.1 years, plus semiannual and 4-monthly harmonics (i.e., 0.5- and 0.33-year periods). The short-term signal was reconstructed from the Fourier spectrum of all periods < 0.9 year, except the two seasonal harmonics, representing interannual oscillations that are not directly linked to periods of seasonality. The longer-term signal was reconstructed from all remaining periods > 1.1 year, representing interannual and longer timescales. The subsignal binning was centered on the definition of the seasonal/annual bin similarly to Mahecha et al. (2010a) and Fürst (2009). The bin ranges were slightly adapted due to the FFT approach, which yields signals of different frequencies compared to the approach chosen by Mahecha et al. (2010a). To identify emerging features occurring at different latitudinal bands, mean values weighted by pixel area were calculated in the tropics (23.5° N to 23.5° S), extratropics (above 23.5° N and below 23.5° S), and globally.

2.4 Variance per timescale and co-oscillation regimes

For each timescale-specific signal, we calculated the proportion of variance of the original signal explained for each variable per grid cell. Each pixel of the global land surface was then classified into oscillation regimes depending on which scale explained the largest amount of variance in each variable (abbreviations: S – short term, A – seasonal, L – longer term, T – trend). For example, if the variance was dominated by the seasonal subsignal in NDVI and T_{air} , and by the short-term scale in Prec, this pixel would be classified as AAS (in the order of NDVI, T_{air} , and Prec). Theoretically, the superimposition yields 64 (4^3) possible combinations, of which only 26 occurred. For simplicity, our analysis was focused on the 11 most abundant oscillation regimes (99.7 % of pixels).

In order to complement static/traditional classifications, we compared our oscillation regimes with the Global Land Cover map project coordinated by the Joint Research Center (GLC2000, Bartholomé and Belward, 2005) and climate zones from the updated Köppen–Geiger global classification (Kottek et al., 2006, see Fig. S1). Only those pixels that contained data from all three data streams (Köppen–Geiger classes A–D, GLC2000, and our oscillation regimes) were considered in this analysis. Nonvegetated and nonnatural areas as defined by GLC2000 were disregarded for this analysis and onward (Table S1 in the Supplement). The final land surface assessed was 75 871 486 km², corresponding to 70 % of the vegetated GLC2000 area (Fig. S1). For the same area, we calculated the V measure (V), a spatial association index based on homogeneity and complementarity criteria proposed specifically for thematic map comparison (Nowosad and Stepinski, 2018). The index ranges from 0 to 1, with 1 being a perfect association, and was used to provide an overall comparison between the co-oscillation regime map with Köppen–Geiger and GLC2000 maps.

To assess the influence of gap-filling performed in the original GIMMS NDVI data due to influence of cloud cover or snow, we excluded time points that were retrieved by splines or mean seasonal cycle due to the lack of direct observation in NDVI (Pinzon and Tucker, 2014) at five different quality flag thresholds in our classification of oscillation regimes. Quality flags were aggregated from 0.083 to 0.5° by calculating the fraction of direct observations per 0.5° pixel at each time step. Subsequently, the dominant classification was repeated, excluding time steps with less than 30 %, 50 %, 70 %, 90 %, and 95 % direct observations for each grid cell. Furthermore, we repeated the time series decomposition method for NDVI and the enhanced vegetation index (EVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS). The vegetation indices product MOD13C1.006 is provided by NASA EOSDIS LP DAAC at 0.05°. Data were aggregated spatially by averaging valid pixels to 0.5° for the overlapping period with GIMMS NDVI (2001–2015). A comparison of the dominant oscillation regimes between products was carried out at a pixel basis.

2.5 Correlations between variables at each timescale

We correlated timescale-specific subsignals of NDVI, T_{air} , and Prec using Pearson's correlation coefficient, Spearman correlation, and partial correlation. For this analysis, all time points with NDVI < 0.2 were masked in order to consider only data points corresponding to active vegetation (Fig. S1). NDVI was lagged one time step (15 d) behind Prec in order to allow for the response time of vegetation to changes in water availability. Due to the 15 d temporal resolution of the data, a response time of up to 15 d is intrinsically included in our analyses. Each time lag is therefore an additional 15 d, and shorter responses cannot be assessed. We compared six different lags (from 15 to 90 d, Fig. S3). When correlating NDVI and precipitation instantaneously, we found almost exclusively negative correlations for the short-term scale. A lag of one time step was sufficient to arrive at expected positive correlations between NDVI and precipitation, while increasing the lag time did not substantially improve or alter the results. We thus chose to globally use a lag of one time step (representing a 15–30 d response time) between precipitation and NDVI across all scales. Globally, temperature appeared to be most strongly correlated to NDVI instantaneously (not lagged); thus, no time lag was introduced between air temperature and NDVI. Recent studies assessing time lags and memory effects between vegetation and climate also indicate that time lags of around 1 month generally carry most of the explanatory power for predicting vegetation dynamics (Krich et al., 2019; Kraft et al., 2019; Pagiannopoulou et al., 2017). Correlations of NDVI– T_{air} and NDVI–Prec were binned into five quantiles and presented in a bivariate color map (Teuling et al., 2011). In addition, we compared differences in the sign of the correlation between

seasonal and longer-term oscillations to detect areas where the correlation was inverted between scales.

2.6 Assessment of land cover change on time series decomposition

We assessed whether land cover change over the 30-year time period influenced our results by extracting pixels with substantial land cover change as determined by Song et al. (2018). While linear trends were removed from the time series before decomposition, changes in amplitude or piecewise linear and nonlinear trends may have an impact on our analyses. First, we aggregated original 0.05° data to match our 0.5° spatial resolution by averaging. We then determined 0.5° pixels with > 25 % gain or loss of trees, short vegetation, or bare ground and assessed whether the observed changes in land cover (Song et al., 2018) were reflected in the NDVI time series to a degree that substantially affected the classification of dominant oscillation regimes.

2.7 Comparison of Fourier transform with empirical mode decomposition

While the FFT approach is the most classical time series decomposition technique, there are more data-adaptive alternatives available (Huang et al., 1998; Ghil, 2002; Paluš and Novotná, 2008). In order to understand whether different methods would lead to different insight, we compared the employed FFT approach with the more data-adaptive empirical mode decomposition (EMD). EMD repeatedly extracts subsignals (intrinsic mode functions, IMFs) from the time series by interpolating a spline between local minima and maxima until the residuals converge to approximately constant values (Huang et al., 1998). We used an ensemble-based modification of the EMD algorithm, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN, Colominas et al., 2014; Torres et al., 2011), and a frequency-binning approach to obtain frequency bands comparable to the ones chosen for FFT. In contrast to the regular EMD, CEEMDAN employs an ensemble approach in which noise is added to the data before decomposition and ensemble averages for each IMF are returned, so that a more robust end result is obtained (Colominas et al., 2014; Torres et al., 2011). Briefly, in CEEMDAN each IMF is computed as the mean of an ensemble of IMFs retrieved from noisy data copies. This IMF is subtracted from the original signal, and the residual signal is used as input for retrieving the next IMF (Colominas et al., 2014; Torres et al., 2011). As such, CEEMDAN is less prone to mode mixing than EMD while still fulfilling the completeness property of EMD (i.e., the sum of all IMFs equals the original signal). As IMFs resulting from EMD do not have a fixed frequency assigned, we then associated each IMF with a timescale by measuring the distance between all local maxima and minima as a proxy for the dominating wavelength of the signal. Distances between

each two maxima or minima were classified as short-term, seasonal, or longer-term depending on their length. The IMF was then categorized by the majority distance category and added into the respective timescale bin. For example, if an IMF contained 25 seasonal cycles and 5 short-term cycles, it was classified as seasonal and added to the seasonal signal bin. IMFs in each bin were combined by summation.

3 Results

3.1 Time series variance across timescales

Assessing the contribution of each timescale subsignal to the signal variance at each grid cell, we find that for NDVI most of the temporal variability is expectedly captured by the seasonal cycle (71 % of the global variance), especially above the Tropic of Cancer (23.5° N) (Fig. 1, Table S2). Short-term oscillations contribute dominantly in parts of tropical America and Southeast Asia, while longer-term components are mainly observed in Australia, South Africa, parts of Argentina, and northern Mexico. Specifically, short-term and longer-term signals together contribute 27 % of the total NDVI variance globally and 38 % in the equatorial region (23.5°N to 23.5°S).

Similarly, T_{air} is strongly dominated by seasonal oscillations in the extratropics above/below 23.5° N/S (94 % and 90 %, respectively, Table S2) as would be expected. Even in the tropics, short-term and longer-term components contribute only 30 % of the variance (and 11 % and 4 % of global variance, respectively, Table S2). In contrast, short-term oscillations dominate global precipitation variance before the seasonal cycle (52 % and 41 % of global variance each, Table S2). An east–west gradient of precipitation over Eurasia stands out, changing from predominantly short-term to predominantly seasonal signal variance. In the tropics, a similar contribution from both oscillations is found (42 % and 41 %, respectively, Table S2). Linear trends removed before FFT decomposition had a minor influence on overall variance (Fig. 1). In summary, short-term and longer-term signals show substantial, regionally focused contributions to signal variance. These regions differ between variables, suggesting complex patterns of temporal interaction.

3.2 Classification of co-oscillations regimes

Given the contrasting, spatially heterogeneous patterns observed in different variables in Fig. 1, we investigated how scale-specific oscillations of biosphere and climate co-occur globally. We combined the dominant scale of variability for each variable in each grid cell (Fig. S4) and found that 84.5 % of the assessed area is dominated by seasonal oscillations of NDVI, 9 % by short-term oscillations in NDVI, and 6.5 % by longer-term oscillations in NDVI (0.03 % captured by the trend). Combining the maps for all three variables into a map of codominant oscillation regimes (Fig. 2, Table S3),

we find that seasonal NDVI regimes co-occur predominantly with seasonal T_{air} , as well as seasonal or short-term Prec regimes (blue regions). Dominant seasonal cycles of NDVI and T_{air} , as well as fast oscillation regimes in Prec, are expected over large parts of the globe, which is reflected by the large extent of the AAS and AAA classes in this analysis. Beyond this expected solar-cycle-induced behavior, a number of differentiated oscillation classes stand out: short-term NDVI oscillations occur mainly in the South American and Asian tropics, in a multitude of combinations with predominantly seasonal or short-term T_{air} and Prec (light-green, red, and light-red regions). Longer-term oscillation regimes of NDVI co-occur with seasonal T_{air} and short-term Prec regimes (dark green regions) around southwestern Africa, southeastern South America, and Australia. Interestingly, the dominant scales in climatic variables are not always associated with similar dominant regimes in NDVI dynamics, suggesting complex or additional driving mechanisms in these heterogeneous regions. In fact, even in areas where temperature or precipitation has a seasonal cycle, NDVI can be dominated by short-term or longer-term oscillations: more than 90 % of the area with short-term NDVI regimes exhibits predominantly seasonal T_{air} , of which 36 % also shows predominantly seasonal Prec (SAA) and 55 % predominantly short-term Prec (SAS, Table S3). All areas where NDVI is predominantly longer term are classified as seasonal T_{air} and short-term Prec regimes (LAS, Table S3).

To account for the influence of clouds and snow cover in the GIMMS NDVI record, especially in the tropics and northern regions, we excluded time points where pixels contained a high proportion of gap-filled values. We found that overall less than 1.5 % of pixels changed their dominant oscillation class when only pixels with more than 0.7 direct observation fraction were considered. Even when the highest-quality threshold was applied (0.95 direct observation fraction), only 2.6 % of pixels changed dominant oscillation class (Fig. S5). Short-term pixels were the most affected by changes in dominant oscillation (12.9 % and 20.8 % for 0.7 and 0.95 direct observation thresholds respectively), while seasonal pixels showed the highest fraction of gap-filling overall (Fig. S5). As a further validation, we found very similar results when repeating the time series decomposition and the dominant oscillation regime classification based on EVI and NDVI from MODIS (Didan et al., 2019; Huete, 1997; Huete et al., 2002) for the years 2001–2015 (Fig. S6).

We investigated to what extent our classification into oscillation regimes shows patterns of temporal vegetation–climate relations that are not represented by conventional static classifications of the land surface. To determine overlap and differences between the classification of temporal vegetation–climate co-oscillations with static classifications of land cover (GLC2000) and Köppen–Geiger climate classes, we assessed their spatial association by the V measure (Nowosad and Stepinski, 2018). The V measures of co-

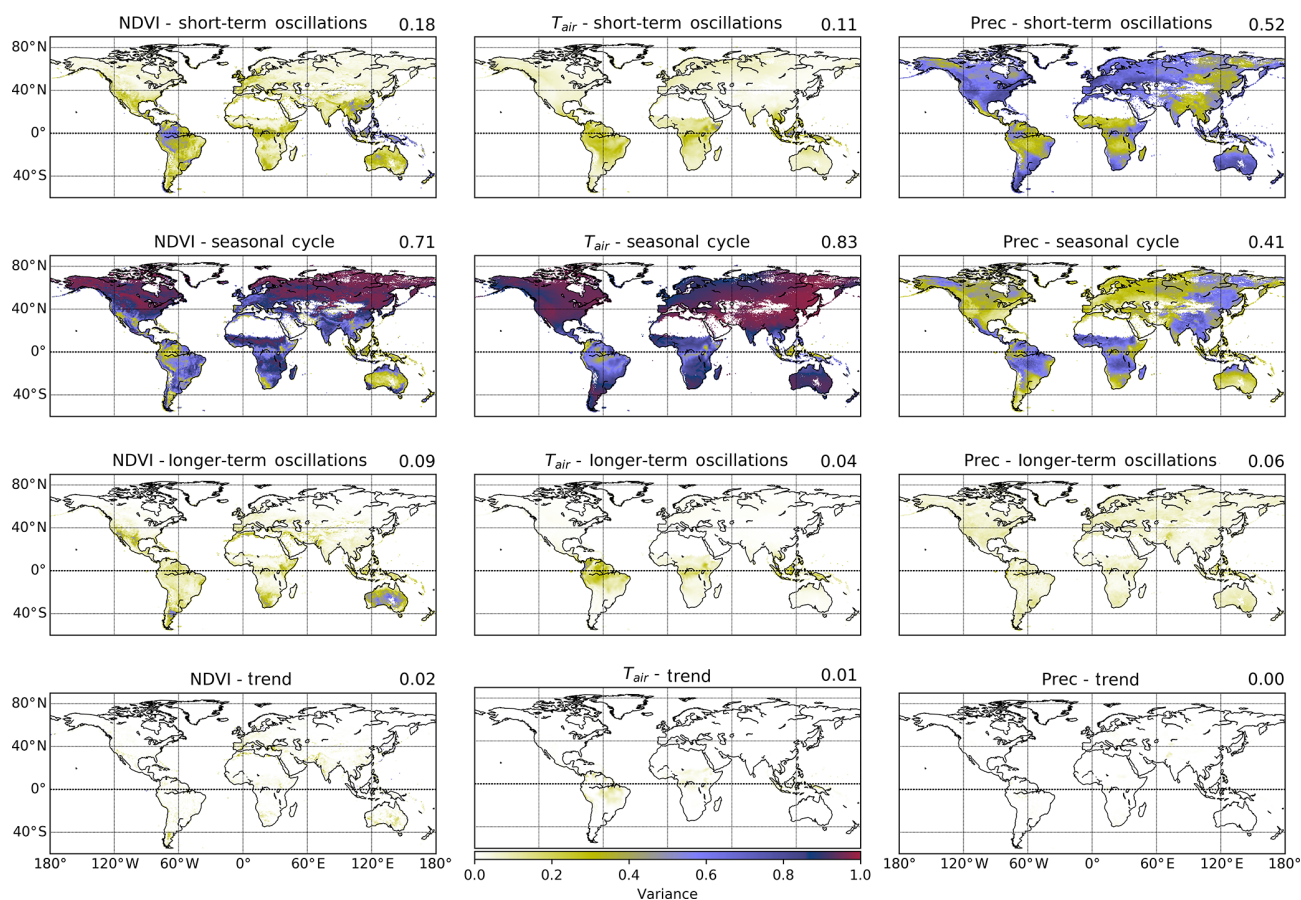


Figure 1. Global distribution of timescale-specific variance (relative spectral powers) of the normalized difference vegetation index (NDVI), air temperature (T_{air}), and precipitation (Prec). Normalized time series of NDVI, T_{air} , and Prec (columns) were decomposed by fast Fourier transformation and reconstructed into short-term (intra-annual), seasonal (annual), and longer-term (interannual) components (rows). The relative contribution of each scale-specific signal to overall variance was determined at each grid cell. Globally, most of the variance of NDVI and T_{air} is contained in the seasonal component (red colors), while Prec shows a high contribution of variance from the short-term component. The semiannual cycle is included in the seasonal band. Upper-right-corner values show the percentage of overall variance explained by each timescale.

oscillation regimes with Köppen–Geiger and GLC2000 were $V = 0.17$ and $V = 0.11$, respectively, indicating weak association with both static classifications. Hence, our classification contains information largely complementary to the compared climate and land cover classifications. Yet we observed a slightly stronger association with Köppen–Geiger than with GLC2000, also when comparing homogeneity and complementarity (Table S4). Comparing the three classifications among each other, we find that dominant temporal patterns in NDVI can be linked to certain land cover types such as shrubs and broadleaf forest: Sankey diagrams (Fig. 2b and c) display which proportion of land surface is commonly classified across different class combinations in the three data layers of the co-oscillation regime, GLC2000, and Köppen–Geiger for evergreen broadleaf forest (EBF, Fig. 2b) and areas dominated by longer-term NDVI (Fig. 2c). We find that EBF is the most diverse among land cover classes in terms of our temporal classification, with 35 % dominated by short-term NDVI

oscillation (Fig. 2b). In contrast, more than 95 % of deciduous and evergreen needleleaf forests (DNF and ENF) and deciduous broadleaf forests (DBF) are dominated by seasonal NDVI regimes (Table S3). We further find a strong association of longer-term NDVI regimes with shrubs (21 % of the area dominated by longer-term NDVI), herbaceous (26 %), and sparse shrubs/herbaceous (49 %) land cover types in arid regions (Fig. 2c, overall 93 % of the LAS area coincides with Köppen–Geiger class B). Thus, differences within and among land cover and climate types exist when assessing temporal co-oscillations of vegetation and climate.

3.3 Assessment of land cover change on time series decomposition

In the above analyses we did not aim to explicitly detect the effect of land cover or land use change (LCLUC), but nevertheless LCLUC could have an influence on our NDVI classi-

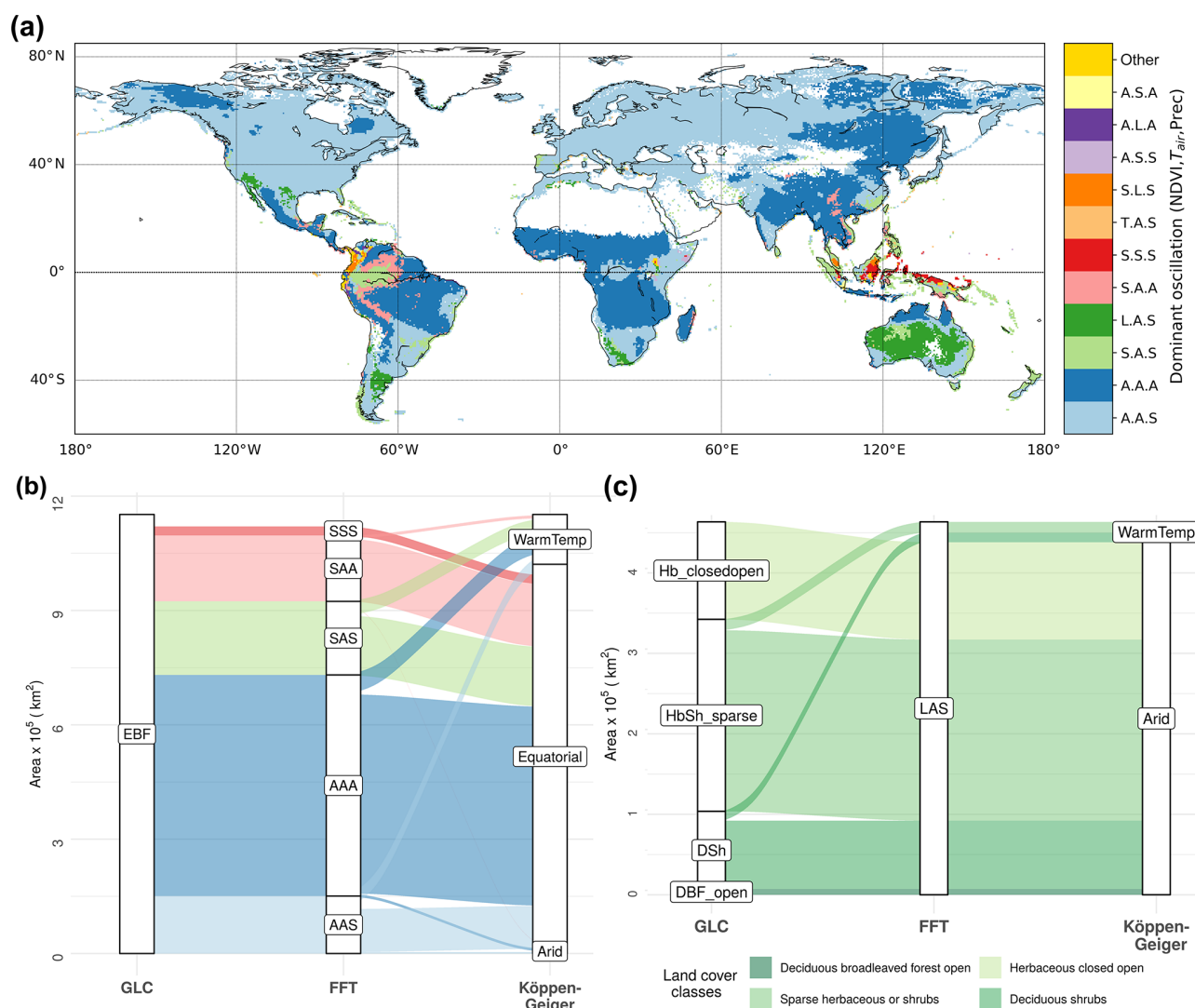


Figure 2. Classification of land surface by dominant scale of variability in NDVI and climate, and its relation to land cover and mean climate. **(a)** Dominant scale of variability was determined for NDVI, T_{air} , and Prec separately for each grid cell and summarized as unique combinations between variables (S – short term, A – seasonal, L – longer term, T – trend, listed in the order of NDVI, T_{air} , and Prec). Only the 11 most common classes are shown. The semiannual cycle is included in the seasonal band. **(b, c)** Sankey diagrams (river plots) showing associations of pixels for **(b)** evergreen broadleaf forest (EBF) and **(c)** regions of dominant long-term oscillations in NDVI (LAS class) to oscillation regime (FFT), land cover class (GLC2000), and Köppen–Geiger (KG) climate class. The width of the ribbons is proportional to the area that is commonly classified into the corresponding GLC2000, KG, or oscillation classes. DBF: deciduous broadleaf forest; Hb_closedopen: herbaceous closed open land cover; DSh: deciduous shrublands; HbSh_sparse: sparse herbaceous and shrub vegetation; Equatorial: KG class A; Arid: KG class B; WarmTemp: KG class C; NDVI: normalized difference vegetation index; T_{air} : air temperature; Prec: precipitation.

fication (Fig. 2). We assessed whether changes in vegetation cover over the 30-year period severely affected our classification by inspecting pixels with $> 25\%$ change in the fraction of trees, short vegetation, or bare ground according to Song et al. (2018). Notably, very few of such pixels showed marked signs of land cover change reflected in NDVI time series at all, which is likely due to the coarse spatial resolution of the data used in this study as compared to previous studies focused on detecting LCLUC (Song et al., 2018;

Fensholt et al., 2015): at 0.5° resolution, most pixels represent mixed signals which obscure most of the details that would allow for detecting land cover changes. In those pixels where we did see a clear progression in NDVI over time, the method did adequately capture this progression, e.g., by correctly reflecting an increasing amplitude of the seasonal cycle and/or shifting baseline (Fig. S7). However, the majority of such pixels with pronounced positive or negative NDVI progression were located in agricultural areas or ar-

eas of urbanization, which had a priori been excluded from downstream analyses. Overall, the change in vegetation over time did not have a widespread influence on the classification of dominant scale and oscillation regimes at the given spatial resolution.

3.4 Correlations of NDVI with climate on multiple scales

To inspect relationships of vegetation with climate at multiple timescales, we correlated NDVI with T_{air} and Prec at each pixel for each timescale (Fig. 3). We found different correlation patterns depending on the timescale: while all possible combinations of correlation between NDVI and T_{air} or Prec exist at the seasonal scale, short-term and longer-term scales show predominantly $T_{\text{air}} + / \text{Prec} -$ or $T_{\text{air}} - / \text{Prec} +$ relationships. On the seasonal scale, NDVI correlates positively with T_{air} and Prec above 40°N , whereas in the other latitudes all possible relations are observed. In particular, South America shows a highly diverse pattern of correlations. Differences exist across the tropics, where South America and Southeast Asia display mainly negative correlation with Prec, whereas African tropics display positive correlation with Prec. Semiarid regions show negative correlations with T_{air} as would be expected. While some of the patterns are known, this correlation of decomposed oscillations reveals a more differentiated picture of ecosystem variability in comparison with the undecomposed data (Fig. S8). Notably, correlations on short- and longer-term scales partially show opposite signs compared to the seasonal scale, e.g., in South America, southern Africa, and Central America. Repeating the analysis with Spearman correlation and partial correlation returned similar results (Figs. S9 and S10). Due to the known saturation effects of NDVI against plant productivity over areas of dense biomass, we repeated the analysis with MODIS EVI. We found overall similar results across timescales, but correlations with T_{air} turned from negative to positive in parts of Central and South America, as well as India (Fig. S11), indicating that NDVI saturation may affect the results obtained from GIMMS long-term records in some areas.

We again compared the observed patterns with vegetation types, to understand how different ecosystems react at different timescales, and found that different land cover classes showed distinct correlation patterns (Fig. 3b and c). Broadleaf evergreen forest shows the most diverse correlations on a seasonal scale (Fig. 3b). For short-term oscillations, the strongest correlations were found in semiarid shrublands and savannas, which spatially coincide with patterns observed in the longer term: for longer-term oscillations, the strong correlation $\text{Prec} +$ and $T_{\text{air}} -$ was again related primarily to shrublands and savannas (Fig. 3a blue areas, Fig. 3c). We also observed a widespread positive longer-term correlation of NDVI with T_{air} in the northern latitudes.

Comparing with static classifications, we found that Köppen–Geiger climate classes had the most prominent dif-

ferentiating effect for correlation patterns, and different climate classes occupy distinct patterns in this correlation space across scales (Fig. S12). Different land cover types generally show similar correlations within one climate zone, but exceptions exist (Fig. S13). Most prominently, EBF shows the most heterogeneous, spatially varying correlations on a seasonal scale. All land cover types show a confined correlation pattern of mainly $T_{\text{air}} + / \text{Prec} -$ or $T_{\text{air}} - / \text{Prec} +$ at the longer-term scale (Fig. 3c), which is further differentiated by climate zone (Fig. S14).

Assessing correlations across the different timescales, we find that the majority of northern temperate regions (Köppen class D) are positively correlated with T_{air} on all timescales, but correlation with Prec varies (zero for short-term and longer-term as well as seasonal scale: generally negative on the coast, positive in the interior continent). The equatorial region, South America, Africa, and Southeast Asia exhibit different correlation patterns with climate despite similar land cover types (tropical forest). In some regions, opposing correlations can be observed across timescales (Fig. 4a). For example, correlation of NDVI with T_{air} in southern Africa varies from negative on the short-term scale to positive on the seasonal scale and back to negative on the longer-term scale. As another example, on the east coast of Australia, NDVI has a low correlation with precipitation on the seasonal scale but high in the longer term. Assessing this globally, correlations between NDVI and T_{air} show inverted signs between seasonal and longer-term scales in 15.4 % of the vegetated land surface area (Fig. 4a and c). The same is true for NDVI and Prec in 27.3 % of the vegetated land surface area (Fig. 4b and c).

In summary, we find that correlations between NDVI and climate variables can change strongly between timescales. Semiarid ecosystems show most prominent short-term and longer-term signatures, while tropical rainforest show the most diverse relationships between variables. These patterns point to complex ecosystem responses to climate at different timescales, indicating that scale-specific ecosystem characterization is necessary to fully understand their temporal dynamics.

3.5 Comparison of fast Fourier transformation with empirical mode decomposition

FFT decomposes a signal in the frequency domain under the assumption that the underlying signals are sinusoidal, time-invariant, and additive (Brockwell and Davis, 2006). Although resulting power spectra and frequency-invariant modes of oscillation are conveniently interpretable, not all ecological processes can be expected to follow regular periodic and additive oscillatory patterns approximated by sine and cosine waves over time. We chose FFT decomposition due to its superior computational speed and stable global applicability, i.e., its ability to return homogeneous spatiotemporal patterns in our analysis. To ensure that the above limita-

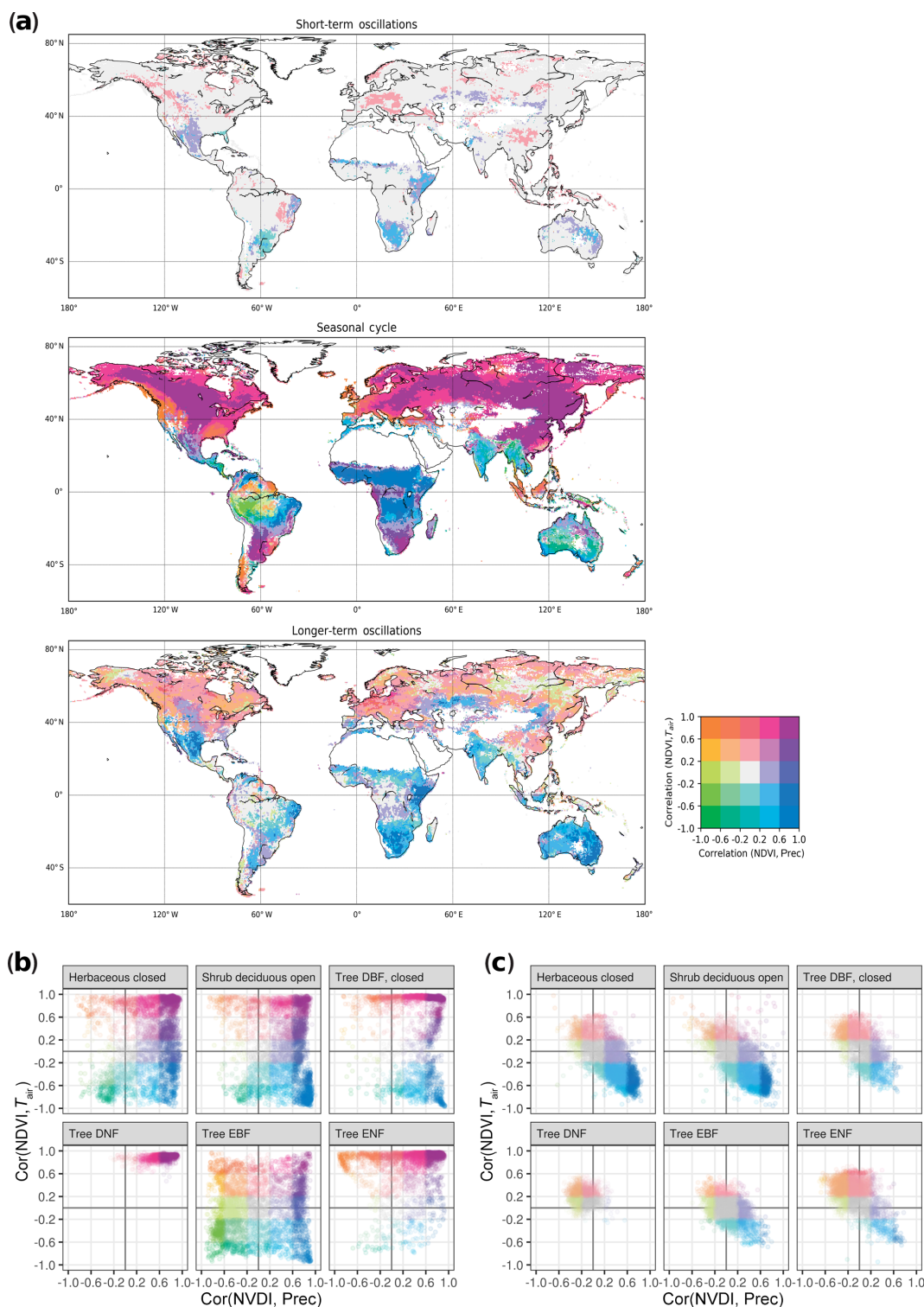


Figure 3. Global distribution of timescale-specific correlation of NDVI with air temperature (T_{air}) and precipitation (Prec). **(a)** Correlations of NDVI with T_{air} and NDVI with Prec were calculated between decomposed signals at each grid cell. NDVI was lagged one time step (15 d) behind precipitation to allow for the response time; T_{air} was correlated instantaneously. Color scale represents both correlations, binned into quantiles (e.g., purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec). Data points with NDVI < 0.2 were excluded to avoid influence of inactive vegetation or nonvegetated time points. **(b, c)** Correlations for different land cover classes (GLC2000) in the seasonal **(b)** and longer-term **(c)** scale.

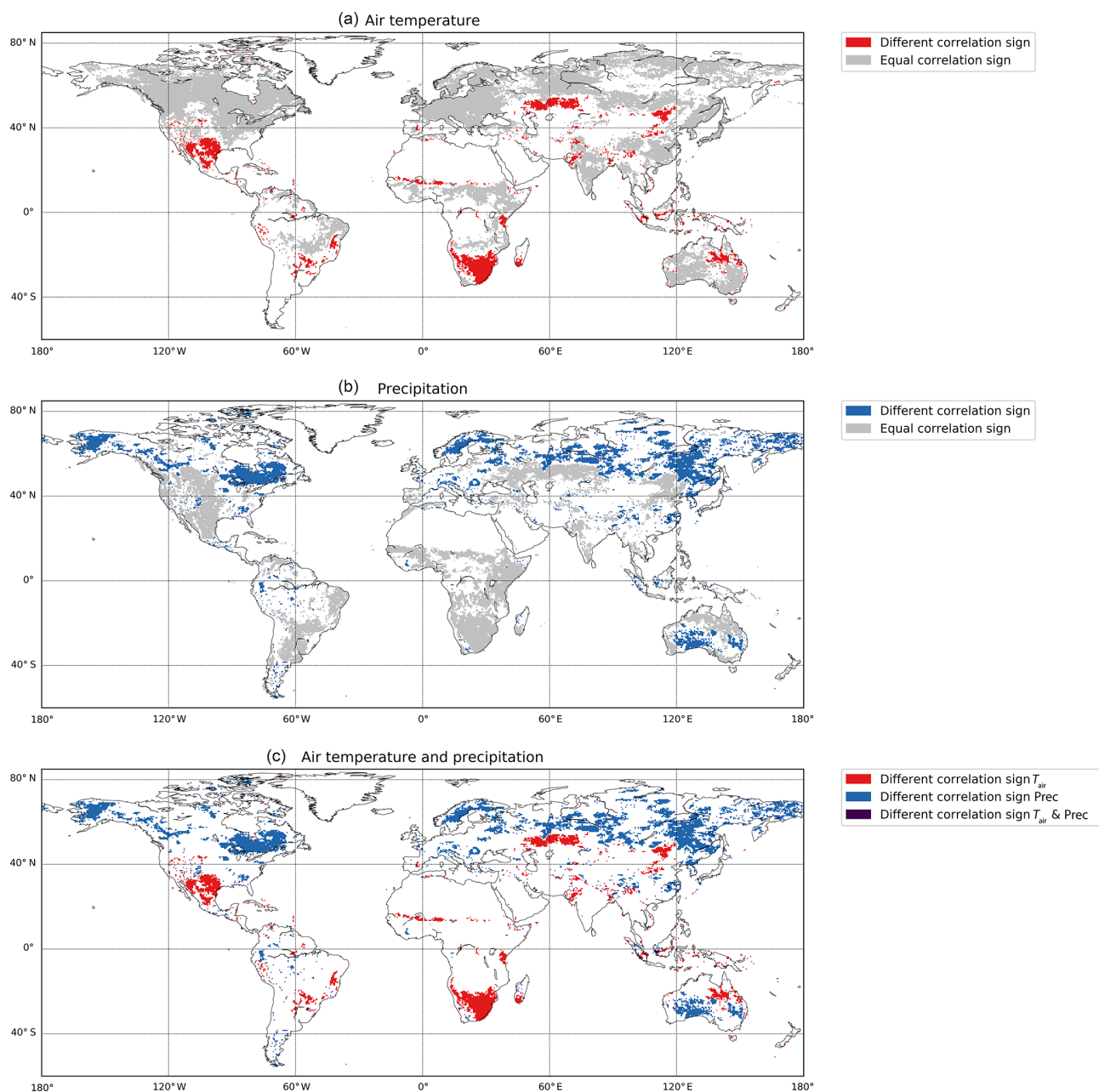


Figure 4. Global comparison of differences in the sign of the correlation between the annual and long-term scale for NDVI and air temperature (a), NDVI and precipitation (b), and summary of both (c). Areas in which the sign of the correlation is inverted between seasonal and longer-term scales are highlighted in color, and areas where the sign of the correlation is identical between scales are highlighted in gray (a, b). Areas with correlations between -0.2 and 0.2 were not considered.

tions did not confound our results, we compared the FFT approach to the data-adaptive empirical mode decomposition, which could be expected to be better suited for exploring nonstationary ecological processes over time. In a test case over Europe, we found that our binning approach resulted in comparable results for the two methods, in terms of both spatial and temporal behavior of the signals (Figs. S15–S18). CEEMDAN generally attributed slightly less signal variance

to the short-term and slightly more to the seasonal cycle for both T_{air} and NDVI and generally showed less modulation in the longer-term signals. Nevertheless, overall results were remarkably comparable. However, because CEEMDAN is a data-adaptive method a higher spatial heterogeneity and spatially varying sensitivity to the noise parameter were observed, which currently constrains a global implementation of the analysis.

4 Discussion

In this study, we present a global characterization of biosphere variability at multiple timescales from weeks to decades where a natural surface classification emerges. We find that a substantial fraction of terrestrial ecosystems is characterized by either short- or longer-term NDVI oscillations (27 % of variance globally). The grid cells dominated by longer-term oscillations in NDVI concentrate mainly in semiarid shrublands, and the short-term-dominated grid cells concentrate mainly in equatorial latitude forests. Patterns in NDVI, air temperature, and precipitation variability are spatially heterogeneous: the classification of codominant oscillations is particularly homogeneous for temperate and boreal regions, while the tropics exhibit complex patterns of codominating timescales in vegetation and climate. This lack of correspondence in dominant temporal oscillations suggests that certain modes of variability in ecosystem–atmosphere interactions can be potentially induced by different exogenous, or even endogenous, dynamics. This picture is further differentiated by the finding that correlations between NDVI and climate variables differ between timescales. This highlights the need to assess vegetation sensitivity to climate specifically on different scales in order to understand complex patterns of atmosphere–biosphere interactions in time, where also confounding factors should be considered.

4.1 Comparison across timescales points to complex temporal signatures

The combination of timescale-specific classification (Fig. 2) and correlation (Fig. 3) allowed us to characterize the major scales of vegetation variability in relation to climate. The classification provides an additional layer of ecosystem characterization beyond common classifications such as land cover classes or the effective Köppen–Geiger climate classifications (Kottek et al., 2006; Koeppen, 1900; Geiger, 1954), which only consider seasonality besides mean climate states, increasing our understanding of dynamic vegetation properties across timescales. The complementarity of this data-driven classification of vegetation dynamics, extracted from the time series and summarized in the co-oscillation classification, is supported by the low spatial association calculated from the V measure. Our findings show that the dominant oscillation of NDVI is often, but not always, related to dominant oscillations of T_{air} and Prec (Fig. 2). For example, most of the land surface is dominated by annual oscillations in NDVI and T_{air} , combined with either seasonal or short-term dominance of Prec (AAA and AAS classes). In many of these regions, air temperature alone or both air temperature and precipitation are limiting factors for plant growth (Nemani, 2003; Seddon et al., 2016) and thus expected to drive vegetation dynamics. In contrast, heterogeneous spatial patterns are observed in equatorial and semiarid regions, where different dominant scales of oscillation are found for NDVI

and climatic variables. Here, the relationship between variables may depend on additional factors, and/or scales may show interactive effects. In the tropics, radiation is proposed to be one of the main drivers of NDVI (Nemani, 2003; Seddon et al., 2016), which could partially explain the lack of temporal coherence between NDVI, T_{air} , and Prec. Dominant short-term oscillations of NDVI (SSS, SAS, SAA) might be explained by climate intraseasonality in the tropics due to the Madden–Julian Oscillation (MJO). The MJO is defined as anomalies in the atmospheric pressure between 10° N and 10° S in the Indian Ocean region that propagate eastward to the eastern Pacific (Madden and Julian, 1971). Depending on the region and phase, its oscillatory period ranges between 20 and 90 d. MJO is considered the dominant component of intraseasonal climate variability in the tropics (Zhang, 2013). We see MJO as one feasible driver of short-term NDVI oscillations through alterations of precipitation and temperature (Zhang, 2013; Hidayat, 2016; Mayta et al., 2019). However, MJO impacts, teleconnections, and predictability are still insufficiently understood (Zhang, 2013; Wang et al., 2019). Short-term oscillations of vegetation in those regions need to be further investigated, including other sources of intraseasonal variation, connections with climatic events, and data constraints. Additionally, regional analysis at higher spatial resolution might reveal details in local climatic variability, as well as other nonclimatic processes such as land use change or crop rotations, among others. Comparing variables across multiple timescales can point to areas with complex temporal signatures that require further attention.

4.2 Nondominant subsignals reveal short- and longer-term ecosystem dynamics

From assessing relationships among variables on multiple timescales, we conclude that (i) nondominant subsignals in climate variables may dominate the biospheric response, (ii) possible links may exist between short-term and longer-term scales, and (iii) correlations of NDVI with climate variables differ between scales.

The dominance of long-term NDVI in semiarid regions coincides with strong correlations of longer-term NDVI with Prec (positive) and T_{air} (negative). This indicates that longer-term variation in precipitation exerts a strong influence on NDVI variability in these regions despite contributing a minor portion of precipitation variation itself. Overall, longer-term correlation of precipitation with NDVI is higher than seasonal correlation (by at least 0.2) in 73 % of the area classified as LAS, where simultaneously longer-term variance of precipitation itself contributes < 20 % of the variance (3 939 362 km²). Due to their highly plastic interannual vegetation dynamics, semiarid ecosystems exert a strong influence on interannual variability of the land CO₂ sink (Ahlstrom et al., 2015; Poulter et al., 2014; Zhang et al., 2016). Longer-term correlations between variables also show broad patterns related to temperature-induced greening in the

northern latitudes (Pan et al., 2018; Keenan and Riley, 2018; Zhu et al., 2016; Park et al., 2016). This is in agreement with previous findings using higher-resolution data (Clinton et al., 2014). Thus, nondominant subsignals in climate variables may dominate the biospheric response, stressing their possible long-term impact on vegetation dynamics.

Vegetation may respond to interannual climate variation on both intra- and interannual scales (Meir et al., 2018). Such interannual climate variation may occur, e.g., in the form of precipitation variation or periodic atmospheric fluctuations like the El Niño–Southern Oscillation (ENSO, Poveda and Salazar, 2004; Kogan and Guo, 2017; Liu et al., 2017), the Pacific Decadal Oscillation (Chen et al., 2017), or Indian Ocean dynamics (Hawinkel et al., 2015). As a prominent example in our study, for semiarid regions both short- and longer-term correlations indicate a strong coupling to variations in water availability for shrublands and herbaceous land cover. These results harmonize with the observed fast response of vegetation to water deficit in arid and semiarid regions (Vicente-Serrano et al., 2013; Wang et al., 2016), as well as the observation of strong water memory effects in these regions (Liu et al., 2018). Some of these patterns match regions where vegetation is stressed during ENSO events due to precipitation decrease (Ahlstrom et al., 2015; Kogan and Guo, 2017), generating a possible link between short-term and longer-term scales. Previous studies suggest that climate forcing on one timescale can be amplified or dampened in corresponding vegetation responses (Stoy et al., 2009), or transferred to another timescale (Katul et al., 2001), preserving the system’s entropy but creating complex interactions across scales. This highlights the need to further investigate interactions between different timescales globally in long-term EO records.

Finally, for some regions the correlation of variables can differ between timescales. In southern Africa, for example, this may be due to a pronounced temperature-dependent annual cycle of vegetation but a longer-term negative effect of warming temperatures on vegetation productivity. Thus, time series decomposition offers important differentiation of atmosphere–biosphere covariation across scales. This may serve as a platform for generating hypotheses in areas where contrasting dominant oscillations and/or correlations across scales are observed.

4.3 Differences between land cover classes highlight the tropics

By characterizing the temporal behavior of NDVI and climate, we observed different vegetation dynamics between land cover types. Differences in power spectra between plant functional types have been shown before on shorter timescales with flux data (Stoy et al., 2009). Assessing this phenomenon globally, we find both homogeneous and heterogeneous behavior within land cover types, showing non-trivial global patterns of the influence of land cover and cli-

mate on vegetation variability across scales. For example, 36 % of evergreen broadleaf forests are dominated by short-term oscillations in NDVI, while other forest types are dominated almost exclusively by seasonal NDVI oscillations. Indeed, the most heterogeneous patterns of codominating oscillations and correlations were found for tropical regions, within and across continents. In African tropics, NDVI is predominantly seasonal and correlation of NDVI with precipitation is always positive, while in most of the remaining tropics, NDVI is dominated by short-term oscillations and shows a predominantly negative correlation with Prec on a seasonal scale in the central Amazon and Southeast Asian tropical forests. This could be explained by different amounts of mean annual precipitation (MAP) falling in these regions, which cause a pronounced wet–dry seasonality in Africa and the central Amazon but not in the northwest or outer regions of the Amazon and SE Asia where MAP is in excess of annual vegetation water demand (Guan et al., 2015). In such areas, correlation with Prec may, e.g., become negative when water is already in excess and clouded/rainy seasons cause limitation in radiation available for plant growth. Similarly, temperature is not usually limiting canopy development in the tropics (rather the contrary, Huang et al., 2019), which may explain negative correlations with T_{air} . As NDVI saturates over regions of dense vegetation, results in the tropics need to be interpreted with caution, and negative correlation with T_{air} could alternatively be explained by underestimation of the seasonal cycle over tropical EBF. In fact, negative correlations with T_{air} were observed less frequently when repeating the analysis with MODIS EVI (Fig. S11), indicating that saturation of NDVI against plant productivity might affect our results in densely vegetated areas such as the tropics. Overall, despite known drawbacks of NDVI as a proxy for plant productivity, the long-term NDVI record generally agrees well with results obtained from the considerably shorter EVI time series, suggesting that it is a good proxy for vegetation activity across timescales over large parts of the global land surface.

It is relevant to emphasize that our results might be affected by noise related to the continuous presence of clouds in the tropics and other atmospheric artifacts. To estimate this effect, we excluded pixels with a low number of direct observations and recalculated the dominant oscillation regime. Short-term oscillations were most affected in this analysis, but roughly 80 % were consistently dominated by short-term oscillation under the strictest scenario, providing higher confidence in our results for the tropics (Fig. S5).

The observed scale-specific patterns highlight the need to assess dynamic vegetation properties in time as differentiating factors beyond land cover type and mean climate.

4.4 Results prove robust against changes in data source and decomposition method

We used long-term GIMMS NDVI records in combination with Fourier transformation in this analysis, well aware of their potential limitations (van Leeuwen et al., 2006; Beck et al., 2011; Fensholt and Proud, 2012; Pinzon and Tucker, 2014). Further consolidating our results against methodological artifacts of the data source and decomposition method, we found that results were not broadly affected or conclusions changed when repeating analyses with MODIS NDVI and EVI or empirical mode decomposition, when excluding gap-filled values as discussed above, or when testing the effect of land use change on decomposed oscillations. Specifically, the analysis of MODIS NDVI and EVI returned a similar classification of dominant timescales in vegetation (Fig. S6). Although short-term oscillations in tropical NDVI may partly reflect noise introduced by cloud cover, heavy aerosol conditions, and biomass burning, our results based on EVI, which is less sensitive to aerosols and haze (Miura et al., 2012), resulted in even more, rather than less, pixels being classified into the short-term oscillation regime (Fig. S6). However, dense clouds are still a limitation when optical remote sensing data are used. EMD decomposition consistently reproduced results of FFT in space and time for all variables (Figs. S15–S18). Excluding gap-filled values originating from snow or cloud inference from the GIMMS NDVI dataset changed the dominant oscillation for only up to 2.3 % of pixels overall, and 20 % of short-term classified pixels, when the strictest threshold was applied (Fig. S5). Land cover and land use change were hardly detectable at the coarse spatial resolution of 0.5° employed and had a minor effect on the distribution of signal variance to the different timescale (Fig. S7). In summary, our results proved robust against data source and decomposition method.

4.5 Limitations and outlook

The current study presents a first global characterization of atmosphere–biosphere variability at multiple timescales from weeks to decades. We chose the longest available satellite-retrieved time series of vegetation, GIMMS NDVI, to be able to assess relations of atmosphere–biosphere co-variability over more than three decades. We find heterogeneous temporal patterns of biosphere–climate responses across timescales. Known limitations of NDVI include saturation effect at high canopy cover, especially relevant in the tropics, as well as influence by soil reflectance in sparsely vegetated areas. These effects could thus influence our results and the emerging patterns should be compared with newer satellite products such as Sun-induced fluorescence (SIF), which are coupled more directly to plant physiology and photosynthesis (Badgley et al., 2017; Koren et al., 2018) but are only available for short time periods. Considering further variables influencing vegetation dynamics, such as radi-

ation, cloud cover, soil moisture, fires, or storms could bring additional insight into the drivers of vegetation dynamics, especially for poorly explained regions in the current analysis, such as the tropics. In future studies, longer-term climate signals could be compared with climate oscillations such as ENSO to gain further understanding of their effect on long-term ecosystem variability.

Analysis of time lag effects between atmospheric forcing and vegetation response may bring additional valuable insight into ecosystem functioning, yet assessing meaningful time lags across timescales is challenging due to a variety of processes involved. Plausible time lags from months to years have been suggested between climate forcing and vegetation response and/or ecosystem carbon exchange through direct and indirect effects (e.g., Braswell et al., 1997, 2005; Vukičević et al., 2001; Krich et al., 2019; Kraft et al., 2019; Papiannopoulou et al., 2017). Assessing lagged vegetation responses across timescales may help to disentangle such co-existing time lags to form a global, timescale-resolved picture of vegetation responses to climate. To account for the confounding effect of autocorrelation and spurious links between variables, methods like causal inference (Runge et al., 2013, 2019; Krich et al., 2019) should be applied in order to retrieve causal time lags between variables.

Our analyses are conducted at 0.5° spatial and 15 d temporal resolution, which may obscure short-term and local vegetation–climate relations, and instead only provide average relationships of variables within each grid cell. Our analyses may thus not be representative in heterogeneous landscapes such as coastlines or mountains. Regions standing out through heterogeneous patterns, such as the Amazon, should be further investigated regionally at higher temporal and spatial resolution whenever consistent data streams permit this to better understand local influence of climate, vegetation and topography on atmosphere–biosphere covariation. Recently, studies in the Amazon based on products such as SIF have detected differences in vegetation anomalies within the basin during El Niño events (Koren et al., 2018). The identified asymmetry in the east–west gradient coincides with observed changes in temperature, soil moisture, and GRACE-derived water storage. Our results pave a way for better understanding the spatial heterogeneity of ecosystem responses to climate variability (van Schaik et al., 2018). Here, assessing temporal patterns beyond correlation (see Wu et al., 2015) will provide additional insight into the temporal evolution of vegetation dynamics and the carbon cycle variability.

5 Conclusions

In conclusion, decomposing vegetation and climate time series into discrete subsignals allows us to disentangle atmosphere–biosphere oscillations from short- to longer-term timescales. A key finding is that short-term and longer-term modes of variability can dominate regional patterns

of ecosystem dynamics: 18 % of land area is effectively characterized by intra-annual variability and 9 % by longer-term modes of NDVI. We derived a global map of dominant patterns of vegetation–climate covariability on multiple timescales. The emerging classification of variability regimes allows us to generate new hypotheses on land–atmosphere interactions. In particular, we can now delineate areas with complex spatiotemporal vegetation signatures. For example, tropical evergreen forests are dominated by short-term oscillations (36 %), while shrublands and herbaceous land cover make up > 90 % of the area dominated by longer-term NDVI, suggesting important roles in intra- and inter-annual biosphere dynamics of these land cover classes. Importantly, changing correlations of NDVI with climate across timescales suggest that climate sensitivities of vegetation can vary with timescale. Globally, 15.4 % of the land area shows opposing correlation of NDVI to T_{air} between annual and long-term modes of variability, while 27.3 % shows opposite correlations of NDVI and Prec. These findings underline the relevance of advancing our understanding of scale-specific climate sensitivities. In southern Africa, for instance, the relation of vegetation to temperature is inverted across scales, as well as in parts of Australia where the same is true for precipitation. Differentiating such responses is essential to fully comprehend long-term biosphere dynamics and project them into the future. Understanding the interaction of climate and vegetation on separate timescales, warranting that independent processes are not obscured by dominant ones, is essential in times where extreme climate conditions of increasing frequency exert a repeated perturbing influence on ecosystem dynamics (Defriez and Reuman, 2017). This needs to be considered for short- and long-term vegetation modeling under changing climate scenarios.

Code and data availability. All remote sensing data are publicly available from the respective supplier. The co-oscillation classification (Fig. 2) is available as a NetCDF file from <https://doi.org/10.5281/zenodo.3611262> (Linscheid et al., 2019). The code underlying all primary figures is made available as a supplementary notebook from <https://doi.org/10.5281/zenodo.3611262> (Linscheid et al., 2019) and in the Supplement.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/bg-17-945-2020-supplement>.

Author contributions. LMES and NL have contributed equally to the paper and performed all analyses. The study was designed by NL, MDM, and LMES with input from all other authors. LMES and NL wrote the paper with substantial input from all other authors. FC and FG implemented the EmpiricalModeDecomposition.jl and libeemd.jl packages in Julia.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This paper has been realized within the Earth System Data Lab project funded by the European Space Agency. The authors acknowledge Lina Fürst for initiation of the preliminary study laying the foundation for this project. The authors acknowledge support from Ulrich Weber for data management and preprocessing. Lina M. Estupinan-Suarez acknowledges the support of the DAAD and its Graduate School Scholarship Programme (57395813). Nora Linscheid acknowledges the support of the TUM Graduate School. Lina M. Estupinan-Suarez and Nora Linscheid acknowledge the continuous support of the International Max Planck Research School for Global Biogeochemical Cycles. Felix Cremer acknowledges the support of the German Research Foundation project HyperSense (grant no. TH 1435/4-1).

Financial support. The article processing charges for this open-access publication were covered by the Max Planck Society.

Review statement. This paper was edited by Akihiko Ito and reviewed by two anonymous referees.

References

- Ahlstrom, A., Raupach, M. R., Schurgers, G., Smith, B., Arneeth, 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., and Zeng, N.: The Dominant Role of Semi-Arid Ecosystems in the Trend and Variability of the Land CO₂ Sink, *Science*, 348, 895–899, <https://doi.org/10.1126/science.aaa1668>, 2015.
- Badgley, G., Field, C. B., and Berry, J. A.: Canopy Near-Infrared Reflectance and Terrestrial Photosynthesis, *Sci. Adv.*, 3, e1602244, <https://doi.org/10.1126/sciadv.1602244>, 2017.
- Bala, G., Caldeira, K., and Nemani, R.: Fast versus Slow Response in Climate Change: Implications for the Global Hydrological Cycle, *Clim. Dynam.*, 35, 423–434, <https://doi.org/10.1007/s00382-009-0583-y>, 2010.
- Bartholomé, E. and Belward, A. S.: GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data, *Int. J. Remote Sens.*, 26, 1959–1977, <https://doi.org/10.1080/01431160412331291297>, 2005.
- Beck, H. E., McVicar, T. R., van Dijk, A. I. J. M., Schellekens, J., de Jeu, R. A. M., and Bruijnzeel, L. A.: Global Evaluation of Four AVHRR–NDVI Data Sets: Intercomparison and Assessment against Landsat Imagery, *Remote Sens. Environ.*, 115, 2547–2563, <https://doi.org/10.1016/j.rse.2011.05.012>, 2011.
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment, *B. Am. Meteorol. Soc.*, 100, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>, 2019.
- Braswell, B. H., Schimel, D. S., Linder, E., and Moore, B.: The Response of Global Terrestrial Ecosystems to In-

- terannual Temperature Variability, *Science*, 278, 870–873, <https://doi.org/10.1126/science.278.5339.870>, 1997.
- Braswell, B. H., Sacks, W. J., Linder, E., and Schimel, D. S.: Estimating Diurnal to Annual Ecosystem Parameters by Synthesis of a Carbon Flux Model with Eddy Covariance Net Ecosystem Exchange Observations, *Glob. Change Biol.*, 11, 335–355, <https://doi.org/10.1111/j.1365-2486.2005.00897.x>, 2005.
- Brockwell, P. J. and Davis, R. A.: *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, New York, NY, 2nd Edn., reprint of the 1991 Edn., oCLC: 819807707, 2006.
- Canisius, F., Turrall, H., and Molden, D.: Fourier Analysis of Historical NOAA Time Series Data to Estimate Bimodal Agriculture, *Int. J. Remote Sens.*, 28, 5503–5522, <https://doi.org/10.1080/01431160601086043>, 2007.
- Cao, L., Bala, G., and Caldeira, K.: Climate Response to Changes in Atmospheric Carbon Dioxide and Solar Irradiance on the Time Scale of Days to Weeks, *Environ. Res. Lett.*, 7, 034015, <https://doi.org/10.1088/1748-9326/7/3/034015>, 2012.
- Chen, M., Parton, W. J., Del Grosso, S. J., Hartman, M. D., Day, K. A., Tucker, C. J., Derner, J. D., Knapp, A. K., Smith, W. K., Ojima, D. S., and Gao, W.: The Signature of Sea Surface Temperature Anomalies on the Dynamics of Semiarid Grassland Productivity, *Ecosphere*, 8, e02069, <https://doi.org/10.1002/ecs2.2069>, 2017.
- Clinton, N., Yu, L., Fu, H., He, C., and Gong, P.: Global-Scale Associations of Vegetation Phenology with Rainfall and Temperature at a High Spatio-Temporal Resolution, *Remote Sens.*, 6, 7320–7338, <https://doi.org/10.3390/rs6087320>, 2014.
- Colominas, M. A., Schlotthauer, G., and Torres, M. E.: Improved Complete Ensemble EMD: A Suitable Tool for Biomedical Signal Processing, *Biomed. Signal Process. Control*, 14, 19–29, <https://doi.org/10.1016/j.bspc.2014.06.009>, 2014.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- Defriez, E. J. and Reuman, D. C.: A Global Geography of Synchrony for Terrestrial Vegetation, *Global Ecol. Biogeogr.*, 26, 878–888, <https://doi.org/10.1111/geb.12595>, 2017.
- De Keersmaecker, W., Lhermitte, S., Tits, L., Honnay, O., Somers, B., and Coppin, P.: A Model Quantifying Global Vegetation Resistance and Resilience to Short-Term Climate Anomalies and Their Relationship with Vegetation Cover: Global Vegetation Resistance and Resilience, *Global Ecol. Biogeogr.*, 24, 539–548, <https://doi.org/10.1111/geb.12279>, 2015.
- Didan, K., Barreto-Munoz, A., Solano, R., and Huete, A.: *MODIS Vegetation Index User's Guide (MOD13 Series). Version 3.00 (Collection 6)*, Vegetation Index and Phenology Lab The University of Arizona, available at: <https://vip.arizona.edu/>, last access: 1 August 2019.
- Fensholt, R. and Proud, S. R.: Evaluation of Earth Observation Based Global Long Term Vegetation Trends – Comparing GIMMS and MODIS Global NDVI Time Series, *Remote Sens. Environ.*, 119, 131–147, <https://doi.org/10.1016/j.rse.2011.12.015>, 2012.
- Fensholt, R., Horion, S., Tagesson, T., Ehammer, A., Ivits, E., and Rasmussen, K.: Global-Scale Mapping of Changes in Ecosystem Functioning from Earth Observation-Based Trends in Total and Recurrent Vegetation: Mapping of Ecosystem Functioning Change from EO Data, *Global Ecol. Biogeogr.*, 24, 1003–1017, <https://doi.org/10.1111/geb.12338>, 2015.
- Fürst, L. M. M.: *Characterizing Global Spatiotemporal Patterns of the Fraction of Absorbed Photosynthetically Active Radiation*, Diploma Thesis, University Bayreuth, Bayreuth, 2009.
- Geiger, R.: *Ch. Klassifikation Der Klimate Nach W. Köppen*, in: *Landolt-Börnstein – Zahlenwerte Und Funktionen Aus Physik, Chemie, Astronomie, Geophysik Und Technik, Band III, alte Serie*, 603–607, Springer, Berlin, 1954.
- Ghil, M.: Advanced Spectral Methods for Climatic Time Series, *Rev. Geophys.*, 40, 3-1–3-41, <https://doi.org/10.1029/2000RG000092>, 2002.
- Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., Caylor, K. K., Sheffield, J., Wood, E. F., Malhi, Y., Liang, M., Kimball, J. S., Saleska, S. R., Berry, J., Joiner, J., and Lyapustin, A. I.: Photosynthetic Seasonality of Global Tropical Forests Constrained by Hydroclimate, *Nat. Geosci.*, 8, 284–289, <https://doi.org/10.1038/ngeo2382>, 2015.
- Guanter, L., Alonso, L., Gómez-Chova, L., Amorós-López, J., Vila, J., and Moreno, J.: Estimation of Solar-Induced Vegetation Fluorescence from Space Measurements, *Geophys. Res. Lett.*, 34, L08401, <https://doi.org/10.1029/2007GL029289>, 2007.
- Hannachi, A., Straus, D. M., Franzke, C. L. E., Corti, S., and Woollings, T.: Low-Frequency Nonlinearity and Regime Behavior in the Northern Hemisphere Extratropical Atmosphere: Nonlinearity and Regime Behaviour, *Rev. Geophys.*, 55, 199–234, <https://doi.org/10.1002/2015RG000509>, 2017.
- Hawinkel, P., Swinnen, E., Lhermitte, S., Verbist, B., Van Orshoven, J., and Muys, B.: A Time Series Processing Tool to Extract Climate-Driven Interannual Vegetation Dynamics Using Ensemble Empirical Mode Decomposition (EEMD), *Remote Sens. Environ.*, 169, 375–389, <https://doi.org/10.1016/j.rse.2015.08.024>, 2015.
- Hidayat, R.: Modulation of Indonesian Rainfall Variability by the Madden–Julian Oscillation, *Proc. Environ. Sci.*, 33, 167–177, <https://doi.org/10.1016/j.proenv.2016.03.067>, 2016.
- Huang, M., Piao, S., Ciais, P., Peñuelas, J., Wang, X., Keenan, T. F., Peng, S., Berry, J. A., Wang, K., Mao, J., Alkama, R., Cescatti, A., Cuntz, M., De Deurwaerder, H., Gao, M., He, Y., Liu, Y., Luo, Y., Myneni, R. B., Niu, S., Shi, X., Yuan, W., Verbeeck, H., Wang, T., Wu, J., and Janssens, I. A.: Air Temperature Optima of Vegetation Productivity across Global Biomes, *Nature Ecol. Evol.*, 3, 772–779, <https://doi.org/10.1038/s41559-019-0838-x>, 2019.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis, *P. Roy. Soc. A-Math. Phys.*, 454, 903–995, <https://doi.org/10.1098/rspa.1998.0193>, 1998.
- Huete, A.: A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS, *Remote Sens. Environ.*, 59, 440–451, [https://doi.org/10.1016/S0034-4257\(96\)00112-5](https://doi.org/10.1016/S0034-4257(96)00112-5), 1997.

- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L.: Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices, *Remote Sens. Environ.*, 83, 195–213, [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2), 2002.
- Katul, G., Lai, C.-T., Schaefer, K., Ellsworth, D., and Oren, R.: Multiscale Analysis of Vegetation Surface Fluxes: From Seconds to Years, *Adv. Water Resour.*, 24, 1119–1132, 2001.
- Keenan, T. F. and Riley, W. J.: Greening of the Land Surface in the World’s Cold Regions Consistent with Recent Warming, *Nature Clim. Change*, 8, 825–828, <https://doi.org/10.1038/s41558-018-0258-y>, 2018.
- Koepfen, W.: Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt, *Geogr. Z.*, 6, 657–679, 1900.
- Kogan, F. and Guo, W.: Strong 2015–2016 El Niño and Implication to Global Ecosystems from Space Data, *Int. J. Remote Sens.*, 38, 161–178, <https://doi.org/10.1080/01431161.2016.1259679>, 2017.
- Koren, G., van Schaik, E., Araújo, A. C., Boersma, K. F., Gärtner, A., Killaars, L., Kooreman, M. L., Kruijt, B., van der Laan-Luijkx, I. T., von Randow, C., Smith, N. E., and Peters, W.: Widespread Reduction in Sun-Induced Fluorescence from the Amazon during the 2015/2016 El Niño, *Philos. T. Roy. Soc. B*, 373, 20170408, <https://doi.org/10.1098/rstb.2017.0408>, 2018.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen–Geiger Climate Classification Updated, *Meteorol. Z.*, 15, 259–263, <https://doi.org/10.1127/0941-2948/2006/0130>, 2006.
- Kraft, B., Jung, M., Körner, M., Requena Mesa, C., Cortés, J., and Reichstein, M.: Identifying Dynamic Memory Effects on Vegetation State Using Recurrent Neural Networks, *Front. Big Data*, 2, 31, <https://doi.org/10.3389/fdata.2019.00031>, 2019.
- Krich, C., Runge, J., Miralles, D. G., Migliavacca, M., Perez-Priego, O., El-Madany, T., Carrara, A., and Mahecha, M. D.: Causal networks of biosphere–atmosphere interactions, *Biogeosciences Discuss.*, <https://doi.org/10.5194/bg-2019-297>, in review, 2019.
- Linscheid, N., Estupinan-Suarez, L. M., Brenning, A., Carvalhais, N., Cremer, F., Gans, F., Rammig, A., Reichstein, M., Sierra, C. A., and Mahecha, M. D.: Supplementary Data and Notebook to “Towards a global understanding of vegetation–climate dynamics at multiple time scales”, *Zenodo*, <https://doi.org/10.5281/zenodo.3611262>, 2020.
- Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., Bloom, A. A., Wunch, D., Frankenberg, C., Sun, Y., O’Dell, C. W., Gurney, K. R., Menemenlis, D., Gierach, M., Crisp, D., and Eldering, A.: Contrasting Carbon Cycle Responses of the Tropical Continents to the 2015–2016 El Niño, *Science*, 358, eaam5690, <https://doi.org/10.1126/science.aam5690>, 2017.
- Liu, L., Zhang, Y., Wu, S., Li, S., and Qin, D.: Water Memory Effects and Their Impacts on Global Vegetation Productivity and Resilience, *Sci. Rep.-UK*, 8, 2962, <https://doi.org/10.1038/s41598-018-21339-4>, 2018.
- Madden, R. A. and Julian, P. R.: Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific, *J. Atmos. Sci.*, 28, 702–708, [https://doi.org/10.1175/1520-0469\(1971\)028<0702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2), 1971.
- Mahecha, M. D., Reichstein, M., Lange, H., Carvalhais, N., Bernhofer, C., Grünwald, T., Papale, D., and Seufert, G.: Characterizing ecosystem–atmosphere interactions from short to interannual time scales, *Biogeosciences*, 4, 743–758, <https://doi.org/10.5194/bg-4-743-2007>, 2007.
- Mahecha, M. D., Fürst, L. M., Gobron, N., and Lange, H.: Identifying Multiple Spatiotemporal Patterns: A Refined View on Terrestrial Photosynthetic Activity, *Pattern Recogn. Lett.*, 31, 2309–2317, <https://doi.org/10.1016/j.patrec.2010.06.021>, 2010a.
- Mahecha, M. D., Reichstein, M., Carvalhais, N., Lasslop, G., Lange, H., Seneviratne, S. I., Vargas, R., Ammann, C., Arain, M. A., Cescatti, A., Janssens, I. A., Migliavacca, M., Montagnani, L., and Richardson, A. D.: Global Convergence in the Temperature Sensitivity of Respiration at Ecosystem Level, *Science*, 329, 838–840, <https://doi.org/10.1126/science.1189587>, 2010b.
- Mahecha, M. D., Reichstein, M., Jung, M., Seneviratne, S. I., Zahle, S., Beer, C., Braakhekke, M. C., Carvalhais, N., Lange, H., Le Maire, G., and Moors, E.: Comparing Observations and Process-Based Simulations of Biosphere–Atmosphere Exchanges on Multiple Timescales, *J. Geophys. Res.-Biogeophys.*, 115, G02003, <https://doi.org/10.1029/2009JG001016>, 2010c.
- Mahecha, M. D., Gans, F., Brandt, G., Christiansen, R., Cornell, S. E., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., Donges, J. F., Dorigo, W., Estupinan-Suarez, L., Gutierrez-Velez, V. H., Gutwin, M., Jung, M., Londoño, M. C., Miralles, D. G., Papastefanou, P., and Reichstein, M.: Earth system data cubes unravel global multivariate dynamics, *Earth Syst. Dynam. Discuss.*, <https://doi.org/10.5194/esd-2019-62>, in review, 2019.
- Martínez, B. and Gilabert, M. A.: Vegetation Dynamics from NDVI Time Series Analysis Using the Wavelet Transform, *Remote Sens. Environ.*, 113, 1823–1842, <https://doi.org/10.1016/j.rse.2009.04.016>, 2009.
- Mayta, V. C., Ambrizzi, T., Espinoza, J. C., and Dias, P. L. S.: The Role of the Madden–Julian Oscillation on the Amazon Basin Intraseasonal Rainfall Variability, *Int. J. Climatol.*, 39, 343–360, <https://doi.org/10.1002/joc.5810>, 2019.
- Meir, P., Mencuccini, M., Binks, O., da Costa, A. L., Ferreira, L., and Rowland, L.: Short-Term Effects of Drought on Tropical Forest Do Not Fully Predict Impacts of Repeated or Long-Term Drought: Gas Exchange versus Growth, *Philos. T. Roy. Soc. B*, 373, 20170311, <https://doi.org/10.1098/rstb.2017.0311>, 2018.
- Miura, T., Huete, A. R., van Leeuwen, W. J. D., and Didan, K.: Vegetation Detection through Smoke-Filled AVIRIS Images: An Assessment Using MODIS Band Passes, *J. Geophys. Res.-Atmos.*, 103, 32001–32011, <https://doi.org/10.1029/98JD00051>, 2012.
- Nemani, R. R.: Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999, *Science*, 300, 1560–1563, <https://doi.org/10.1126/science.1082750>, 2003.
- Nowosad, J. and Stepinski, T. F.: Spatial Association between Regionalizations Using the Information-Theoretical V-Measure, *Int. J. Geogr. Inf. Sci.*, 32, 2386–2401, <https://doi.org/10.1080/13658816.2018.1511794>, 2018.
- Paluš, M. and Novotná, D.: Detecting Oscillations Hidden in Noise: Common Cycles in Atmospheric, Geomagnetic and Solar Data, in: *Nonlinear Time Series Analysis in the Geosciences*, edited by: Donner, R. V. and Barbosa, S. M., Vol. 112, 327–353, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-78938-3_15, 2008.
- Pan, N., Feng, X., Fu, B., Wang, S., Ji, F., and Pan, S.: Increasing Global Vegetation Browning Hidden in Overall Vegetation

- Greening: Insights from Time-Varying Trends, *Remote Sens. Environ.*, 214, 59–72, <https://doi.org/10.1016/j.rse.2018.05.018>, 2018.
- Papagiannopoulou, C., Miralles, D. G., Dorigo, W. A., Verhoest, N. E. C., Depoorter, M., and Waegeman, W.: Vegetation Anomalies Caused by Antecedent Precipitation in Most of the World, *Environ. Res. Lett.*, 12, 074016, <https://doi.org/10.1088/1748-9326/aa7145>, 2017.
- Pappas, C., Mahecha, M. D., Frank, D. C., Babst, F., and Koutsoyiannis, D.: Ecosystem Functioning Is Enveloped by Hydrometeorological Variability, *Nature Ecol. Evol.*, 1, 1263–1270, <https://doi.org/10.1038/s41559-017-0277-5>, 2017.
- Park, T., Ganguly, S., Tømmervik, H., Euskirchen, E. S., Høgda, K.-A., Karlsen, S. R., Brovkin, V., Nemani, R. R., and Myneni, R. B.: Changes in Growing Season Duration and Productivity of Northern Vegetation Inferred from Long-Term Remote Sensing Data, *Environ. Res. Lett.*, 11, 084001, <https://doi.org/10.1088/1748-9326/11/8/084001>, 2016.
- Pinzon, J. and Tucker, C.: A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series, *Remote Sens.*, 6, 6929–6960, <https://doi.org/10.3390/rs6086929>, 2014.
- 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., and van der Werf, G. R.: Contribution of Semi-Arid Ecosystems to Interannual Variability of the Global Carbon Cycle, *Nature*, 509, 600–603, <https://doi.org/10.1038/nature13376>, 2014.
- Poveda, G. and Salazar, L. F.: Annual and Interannual (ENSO) Variability of Spatial Scaling Properties of a Vegetation Index (NDVI) in Amazonia, *Remote Sens. Environ.*, 93, 391–401, <https://doi.org/10.1016/j.rse.2004.08.001>, 2004.
- Runge, J., Petoukhov, V., and Kurths, J.: Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models, *J. Climate*, 27, 720–739, <https://doi.org/10.1175/JCLI-D-13-00159.1>, 2013.
- Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J.: Inferring Causation from Time Series in Earth System Sciences, *Nat. Commun.*, 10, 2553, <https://doi.org/10.1038/s41467-019-10105-3>, 2019.
- Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D., and Willis, K. J.: Sensitivity of Global Terrestrial Ecosystems to Climate Variability, *Nature*, 531, 229–232, <https://doi.org/10.1038/nature16986>, 2016.
- Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., and Townshend, J. R.: Global Land Change from 1982 to 2016, *Nature*, 560, 639–643, <https://doi.org/10.1038/s41586-018-0411-9>, 2018.
- Stoy, P. C., Richardson, A. D., Baldocchi, D. D., Katul, G. G., Stanovick, J., Mahecha, M. D., Reichstein, M., Detto, M., Law, B. E., Wohlfahrt, G., Arriga, N., Campos, J., McCaughey, J. H., Montagnani, L., Paw U, K. T., Sevanto, S., and Williams, M.: Biosphere-atmosphere exchange of CO₂ in relation to climate: a cross-biome analysis across multiple time scales, *Biogeosciences*, 6, 2297–2312, <https://doi.org/10.5194/bg-6-2297-2009>, 2009.
- Teuling, A. J., Stöckli, R., and Seneviratne, S. I.: Bivariate Colour Maps for Visualizing Climate Data, *Int. J. Climatol.*, 31, 1408–1412, <https://doi.org/10.1002/joc.2153>, 2011.
- Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P.: A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4144–4147, IEEE, Prague, Czech Republic, <https://doi.org/10.1109/ICASSP.2011.5947265>, 2011.
- van Leeuwen, W. J., Orr, B. J., Marsh, S. E., and Herrmann, S. M.: Multi-Sensor NDVI Data Continuity: Uncertainties and Implications for Vegetation Monitoring Applications, *Remote Sens. Environ.*, 100, 67–81, <https://doi.org/10.1016/j.rse.2005.10.002>, 2006.
- van Schaik, E., Killaars, L., Smith, N. E., Koren, G., van Beek, L. P. H., Peters, W., and van der Laan-Luijkx, I. T.: Changes in Surface Hydrology, Soil Moisture and Gross Primary Production in the Amazon during the 2015/2016 El Niño, *Philos. T. Roy. Soc. B*, 373, 20180084, <https://doi.org/10.1098/rstb.2018.0084>, 2018.
- Vicente-Serrano, S. M., Gouveia, C., Camarero, J. J., Begueria, S., Trigo, R., Lopez-Moreno, J. I., Azorin-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J., Moran-Tejeda, E., and Sanchez-Lorenzo, A.: Response of Vegetation to Drought Time-Scales across Global Land Biomes, *P. Natl. Acad. Sci. USA*, 110, 52–57, <https://doi.org/10.1073/pnas.1207068110>, 2013.
- Viles, H.: Interannual, Decadal and Multidecadal Scale Climatic Variability and Geomorphology, *Earth-Sci. Rev.*, 61, 105–131, [https://doi.org/10.1016/S0012-8252\(02\)00113-7](https://doi.org/10.1016/S0012-8252(02)00113-7), 2003.
- Vukićević, T., Braswell, B. H., and Schimel, D.: A Diagnostic Study of Temperature Controls on Global Terrestrial Carbon Exchange, *Tellus B*, 53, 150–170, <https://doi.org/10.1034/j.1600-0889.2001.d01-13.x>, 2001.
- Wang, B., Chen, G., and Liu, F.: Diversity of the Madden–Julian Oscillation, *Sci. Adv.*, 5, eaax0220, <https://doi.org/10.1126/sciadv.aax0220>, 2019.
- Wang, J., Zeng, N., and Wang, M.: Interannual variability of the atmospheric CO₂ growth rate: roles of precipitation and temperature, *Biogeosciences*, 13, 2339–2352, <https://doi.org/10.5194/bg-13-2339-2016>, 2016.
- Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., and Zhao, W.: Time-Lag Effects of Global Vegetation Responses to Climate Change, *Glob. Change Biol.*, 21, 3520–3531, <https://doi.org/10.1111/gcb.12945>, 2015.
- Zeng, F.-W., Collatz, G., Pinzon, J., and Ivanoff, A.: Evaluating and Quantifying the Climate-Driven Interannual Variability in Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) at Global Scales, *Remote Sens.*, 5, 3918–3950, <https://doi.org/10.3390/rs5083918>, 2013.
- Zhang, C.: Madden–Julian Oscillation: Bridging Weather and Climate, *B. Am. Meteorol. Soc.*, 94, 1849–1870, <https://doi.org/10.1175/BAMS-D-12-00026.1>, 2013.
- Zhang, Y., Xiao, X., Guanter, L., Zhou, S., Ciais, P., Joiner, J., Sitch, S., Wu, X., Nabel, J., Dong, J., Kato, E., Jain, A. K., Wiltshire, A., and Stocker, B. D.: Precipitation and Carbon-Water Coupling Jointly Control the Interannual Variability of Global Land Gross Primary Production, *Sci. Rep.-UK*, 6, 39748, <https://doi.org/10.1038/srep39748>, 2016.
- Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao,

C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N.: Greening of the Earth and Its Drivers, *Nature Climate Change*, 6, 791–795, <https://doi.org/10.1038/nclimate3004>, 2016.

Chapter 4

A regional Earth system data lab for understanding ecosystem dynamics: An example from tropical South America

Second manuscript

Authors: Lina M. Estupinan-Suarez, Fabian Gans, Alexander Brenning, Victor H. Gutierrez-Velez, Maria C. Londono, Daniel E. Pabon-Moreno, Germán Poveda, Markus Reichstein, Björn Reu, Carlos A. Sierra, Ulrich Weber, and Miguel D. Mahecha.

Status: Published

Journal: Front. Earth Sci. 9:613395, 2021. doi: 10.3389/feart.2021.613395



A Regional Earth System Data Lab for Understanding Ecosystem Dynamics: An Example from Tropical South America

Lina M. Estupinan-Suarez^{1,2*}, Fabian Gans¹, Alexander Brenning^{2,3}, Victor H. Gutierrez-Velez⁴, Maria C. Londono⁵, Daniel E. Pabon-Moreno¹, Germán Poveda⁶, Markus Reichstein^{1,3,7}, Björn Reu⁸, Carlos A. Sierra^{1,9}, Ulrich Weber¹ and Miguel D. Mahecha^{1,7,10,11}

¹Max Planck Institute for Biogeochemistry, Jena, Germany, ²Department of Geography, Friedrich Schiller University Jena, Jena, Germany, ³Michael Stifel Center Jena for Data-Driven and Simulation Science, Jena, Germany, ⁴Department of Geography and Urban Studies, Temple University, Philadelphia, PA, United States, ⁵Alexander Von Humboldt Biological Resources Research Institute, Bogotá, Colombia, ⁶Department of Geosciences and Environment, Universidad Nacional de Colombia, Medellín, Colombia, ⁷German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, ⁸School of Biology, Faculty of Science, Universidad Industrial de Santander, Bucaramanga, Colombia, ⁹Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden, ¹⁰Remote Sensing Centre for Earth System Research, Leipzig University, Leipzig, Germany, ¹¹Helmholtz Centre for Environmental Research-UFZ, Leipzig, Germany

OPEN ACCESS

Edited by:

Alexander Kokhanovsky,
Telespazio Germany GmbH, Germany

Reviewed by:

Gregory Giuliani,
Université de Genève, Switzerland
Rabi Mohtar,
American University of Beirut,
Lebanon

*Correspondence:

Lina M. Estupinan-Suarez
lestup@bgc-jena.mpg.de

Specialty section:

This article was submitted to
Environmental Informatics and Remote
Sensing,
a section of the journal
Frontiers in Earth Science

Received: 02 October 2020

Accepted: 28 June 2021

Published: 20 July 2021

Citation:

Estupinan-Suarez LM, Gans F, Brenning A, Gutierrez-Velez VH, Londono MC, Pabon-Moreno DE, Poveda G, Reichstein M, Reu B, Sierra CA, Weber U and Mahecha MD (2021) A Regional Earth System Data Lab for Understanding Ecosystem Dynamics: An Example from Tropical South America.
Front. Earth Sci. 9:613395.
doi: 10.3389/feart.2021.613395

Tropical ecosystems experience particularly fast transformations largely as a consequence of land use and climate change. Consequences for ecosystem functioning and services are hard to predict and require analyzing multiple data sets simultaneously. Today, we are equipped with a wide range of spatio-temporal observation-based data streams that monitor the rapid transformations of tropical ecosystems in terms of state variables (e.g., biomass, leaf area, soil moisture) but also in terms of ecosystem processes (e.g., gross primary production, evapotranspiration, runoff). However, the underexplored joint potential of such data streams, combined with deficient access to data and processing, constrain our understanding of ecosystem functioning, despite the importance of tropical ecosystems in the regional-to-global carbon and water cycling. Our objectives are: 1. To facilitate access to regional “Analysis Ready Data Cubes” and enable efficient processing 2. To contribute to the understanding of ecosystem functioning and atmosphere-biosphere interactions. 3. To get a dynamic perspective of environmental conditions for biodiversity. To achieve our objectives, we developed a regional variant of an “Earth System Data Lab” (RegESDL) tailored to address the challenges of northern South America. The study region extensively covers natural ecosystems such as rainforest and savannas, and includes strong topographic gradients (0–6,500 masl). Currently, environmental threats such as deforestation and ecosystem degradation continue to increase. In this contribution, we show the value of the approach for characterizing ecosystem functioning through the efficient implementation of time series and dimensionality reduction analysis at pixel level. Specifically, we present an analysis of seasonality as it is manifested in multiple indicators of ecosystem primary production. We demonstrate that the RegESDL has the ability to underscore contrasting patterns of

ecosystem seasonality and therefore has the potential to contribute to the characterization of ecosystem function. These results illustrate the potential of the RegESDL to explore complex land-surface processes and the need for further exploration. The paper concludes with some suggestions for developing future big-data infrastructures and its applications in the tropics.

Keywords: data cubes, data access, time series, dimensionality reduction, tropical ecosystems, ecosystem functioning, seasonality, biodiversity

1 INTRODUCTION

Novel data streams in the Earth system sciences are becoming available at unprecedented rates (Boulton, 2018). Given that many data streams are regularly improved and frequently its spatio-temporal resolution is increased, we overall face a quasi-exponential growth of data volumes (Guo, 2017; Reichstein et al., 2019). The “data rich world” has become a challenge widely acknowledged across disciplines, but also opened a novel window of opportunity. For example, in the Earth system sciences, we expect a deeper understanding of a wide range of processes that remain to be insufficiently understood today (Scholze et al., 2017; Gentine et al., 2018; Reichstein et al., 2019). In particular, these data have large potential to reduce uncertainties in the quantification of global hydrological fluxes (Miralles et al., 2011; Beck et al., 2016; Ciabatta et al., 2018; Shen et al., 2018), atmosphere-biosphere exchange of carbon, water and energy (Dorigo et al., 2011, 2017; Green et al., 2017; Konings and Gentine, 2017; Papagiannopoulou et al., 2017; Ryu et al., 2019; Jung et al., 2020), the estimation of biodiversity patterns (Asner et al., 2015; Ma et al., 2020), and the interactions of all these processes as mediated by key ecosystem functional properties (Reichstein et al., 2014; Musavi et al., 2015; He et al., 2019).

The reason for this optimism comes especially from emerging opportunities in interpreting a wide array of data streams that jointly monitor the same system from different viewpoints. Examples are the monitoring of land ecosystems with multiple sensors at different wavelengths via satellite remote sensing, e.g., from the optical to the radar domain (Joshi et al., 2016; Anaya et al., 2020; Heckel et al., 2020), the joint analysis of field measurements and remotely sensed data (Mahecha et al., 2017; Meyer et al., 2019), and productions of ensembles of multiple data sets that integrate process-based understanding (Musavi et al., 2017). In general, it is the multitude of climate data sets that allow researchers to understand the multivariate and multifaceted nature of land-dynamics in relation to climate variability (Kraemer et al., 2020; Mahecha et al., 2020). Big-data perspectives of this kind in the Earth system context are therefore highly relevant to improve our understanding of ecological processes, e.g., effects of land use and climate change, and other fundamental transformations on the functioning of land ecosystems.

Given that many of the relevant data streams are retrieved from space, they create a unique opportunity to understand dynamics, trends and tipping points in those regions of the Earth that often lack dense *in-situ* observation networks. This is an advantage for low and mid-income countries, which

nevertheless experience the fastest and most severe ecological and social transformations (Hansen et al., 2013; Leblois et al., 2017; McNicol et al., 2018; Song et al., 2018). We specifically point out to tropical ecosystems for two reasons. On the one hand, these regions experience fast ecological transitions e.g., due to rapid socioeconomic development (Dávalos et al., 2011; Bathiany et al., 2018; Armenteras et al., 2019a). For instance, the tropical Andes are the most critical hotspot for biodiversity in the world due to human encroachment, deforestation and land use change (Orme et al., 2005; Etter et al., 2008; Poveda et al., 2011). On the other hand, tropical ecosystems constitute relevant controls on the global carbon and water cycles. The Amazon forest, for instance, plays a significant role in the global carbon balance (Chambers et al., 2001; Pan et al., 2011; Phillips and Lewis, 2014; Hubau et al., 2020), and strongly regulates water and moisture recycling at the continental scale (Poveda et al., 2006; Zemp et al., 2014, 2017). In turn, the Andes and the low-lying Amazon constitute a coupled system whereby the low-lying Amazon exports atmospheric water to the Andes by the winds, while the Andes export surface water, sediments and nutrients to the Amazon, which highlights their mutual interdependence and the deleterious impacts of deforestation on both sides for the integrity of the system’s functioning (Builes-Jaramillo et al., 2018). Additionally, other ecosystems play a crucial role in biogeochemical cycles. In general, savannas have been considered main drivers of interannual variability in the carbon cycle (Ahlström et al., 2015), and wetlands, including flooded savannas and swamp forest, are significant contributors of methane emissions (Bloom et al., 2017).

However, the countries hosting these highly relevant ecosystems often happen to be those where knowledge on Earth system dynamics happens to be at a comparatively early stage. Limited resources for science and technology often hinder dealing with these highly interdisciplinary challenges. One practical reason might be that research e.g., into large-scale biosphere-atmosphere interactions may require big-data infrastructures, data hosting facilities, and numerical preparation that is hardly achievable by local institutions in the long-term. This is why great hopes are today on global data facilities that may become fundamental game changers in this context. Specifically, we refer to such facilities that are not only providing data access but also provide users’ independence for developing and executing analysis. Prominent examples such as Google Earth Engine (GEE) (Gorelick et al., 2017; Tamiminia et al., 2020), or the Climate Data Store give access to a wide set of data streams accompanied by analytics facilities. However, these platforms usually provide the data as is, which means they are a

collection of satellite images that are stored image by image for example as a stack of GeoTIFFS or as NetCDF files chunked as latitude-longitude maps. If the task is understanding the complex dynamics of time series or spatio-temporal patterns, having efficient computational access to the time dimension is key for boosting temporal studies at individual pixels or selected regions. Therefore, a certain amount of data pre-processing is necessary to give analytical tools the possibility for efficient access to data along all axes of the n-dimensional cube (e.g., latitude, longitude, time, variables, ensemble members, etc.). In addition, none of these platforms are particularly developed for the analysis of tropical ecosystems in terms of specific data availability and analytic capacity to trigger the understanding of regional Earth system processes.

In an ongoing joint European-Colombian collaboration since 2016, we identified the lack of open platforms that not only share free and open data, but enable analyzing them for specific problems such as regional conservation issues that should accompany a suite of regional monitoring and synthesis efforts (Sierra et al., 2017), let alone with the accompanying meta-data information. Starting from there, we have worked on the conceptual outline of a data infrastructure to study land ecosystem dynamics in space and time in tropical South America. We developed a regional Earth System Data Lab (RegESDL) for northern South America to facilitate big-data analytics efficiently based on cloud infrastructures. Our approach is twofold: first, it provides analysis ready data cubes (ARDCs) that can be augmented by almost any other conventional spatial data set. Second, and this is key, it provides the opportunity to apply any arbitrary set of user-defined functions and algorithms on the generated data cubes. The idea is that time, space, and variables, are all dimensions that can be equally relevant to the problem under investigation and therefore need to be treated alike (Mahecha et al., 2020).

Many regional data cube efforts have recently emerged around the Open Data Cube (ODC) concept that was originally developed for Australia (Lewis et al., 2017). Mostly, ODCs facilitate access to pre-processed satellite imagery. This idea is now applied to e.g., Armenia (Asmaryan et al., 2019), Colombia (Ariza-Porras et al., 2017; Bravo et al., 2017), Kenya (Killough, 2019), Switzerland (Giuliani et al., 2017), among other countries. Previous work has focused on the implementation and perspectives of these and related initiatives that vary from the software development to its establishment as national projects for supporting decision making (Giuliani et al., 2017; Killough, 2019). However, these initiatives are based on a latitude-longitude grid for data storage. This dramatically limits efficient access to time series which comes with an expensive computational cost when operating on the raw data directly because the spatial dimensions is the main unit of access and processing. Compared to all these initiatives, our approach is committed to support efficient access to any suite of geographical dataset. This allows users to explore time, variables, space, and other dimensions in its equal right (Mahecha et al., 2020). We also regard it as essential to take user-defined functions seriously and prioritize them. Users should be able to use the full power of a programming language to write algorithms, including calling into

third-party libraries and map them over the entire data cube in a way that is equally efficient. In this study, we support moderate spatial resolution for understanding Earth System interactions addressing regional challenges; varying from technical aspects of data quality, acquisition and management, to high complexity due to landscape heterogeneity.

The purpose of this paper is to introduce the RegESDL for northern South America and illustrate its potential to characterize land-surface processes in relation to climatic and land use drivers. Using the example of ecosystems complex seasonality, we illustrate the approach by combining dimensionality reduction and time series analytics tools. The paper is structured as follows: First, we briefly introduce the RegESDL architecture, the implemented facilities and computational approach. Second, we describe the available datasets and the regional context. Third, we showcase how to operate on the RegESDL by characterizing seasonal dynamics in tropical ecosystems. For this, we use a multivariate set of remote sensing derived indicators related to ecosystem productivity. Seasonality in the tropics is well characterized from a climatological point of view, but hardly described from the point of view of ecosystem functioning. Finally, we discuss the findings of our study and provide some guidance on how the RegESDL should help to advance research in the tropics across disciplines such as biodiversity from both a conceptual and technical standpoint.

2 METHODS

2.1 The Regional Earth System Data Lab Architecture

The RegESDL is a twofold approach for big-data analytics of spatio-temporal variables. It is conceptually and technically an extension of the the Earth System Data Lab (ESDL) (Mahecha et al., 2020), and its guiding principle is to treat all data dimensions such as latitude, longitude, time, variables, and new dimensions (i.e., outcomes from processing), all alike. Thus, all data sets are treated as elements of the same “hypercube”. A formal mathematical definition of data cubes and how to operate on them is provided in Mahecha et al. (2020). The first fold is based on ARDCs which are a set of data cubes gridded at the same spatio-temporal resolution (see section 2.2). The second fold is the analytics software that tackles the issue of working with large datasets that are too big to fit into a computer’s RAM. With our approach this is not any longer a critical limitation given that many user functions do not operate on the whole dataset at once, but can operate independently on slices of the data cube along a given dimension.

Traditionally, there are many tools for analyzing data sliced along the space dimensions and well established libraries like the Geospatial Data Abstraction Library (GDAL/OGR Contributors, 2020) used as main dependency in geospatial libraries of different geographical information system software such as QGIS and programming languages such as R and Python. However, in our cubing approach, slicing is not only efficient along the space dimension but also along all dimensions i.e., time, variables, and any other thinkable dimension a cube might

have (e.g., frequency domain after time series decomposition). The idea is that users only have to define and implement their basic functions at the minimum dimensions that the computation would operate individually in the data cube. The system then handles each operation efficiently, i.e., the implemented logic for slicing can be applied to solve spatial, temporal, or multivariate problems all in the same highly efficient way.

The ESDL software uses split-apply-combine methods (Wickham, 2011) to facilitate the repeated application of user-defined functions to sub-cubes. Thus, in the Julia interface, in order to implement a customized function to be applied to the entire cube or its lower-dimensional sub-cubes, the user first simply implements it for the lowest possible dimensionality. As an example, an operation on a single time series only needs to be implemented as a function that takes a vector (i.e., one-dimensional array) as its input. Likewise, a function intended to operate on two-dimensional latitude-longitude slices of the data only needs to be designed in a way that it expects a matrix argument, not a three or four-dimensional cube. The user simply ignores the fact that later the function is applied to a higher-dimensional object, for example along variables or any remaining cube dimension. Then this function is passed to a higher-level processing pipeline, along with the definition of input and output dimensions and a highly optimized computation kernel is generated by the processing package.

The users can, too, allow any ARDCs axis to interact with data stored in other formats such as one or multi-dimensional arrays. For example, indices of climate variability such as El Niño Southern Oscillation (ENSO) could either be defined simply as a vector or transformed into a one-dimensional cube. In both cases, the index interacts with the corresponding ARDCs axis in a similar way than an apply function in the time domain. These split-apply-combine methods are common tools in data-science oriented languages like R, Python and Julia for in-memory datasets. However, for larger-than memory datasets it is not so easy to find a suitable solution at hand. While frameworks like Apache Hadoop and Spark (Vavilapalli et al., 2013; Zaharia et al., 2016) provide solutions for unstructured, table-like data, they would not fit for the challenges provided by structured n-dimensional arrays. A very promising approach is the combination of xarray and dask for efficient and scalable split-apply-combine computations in combination with Zarr as a storage backend.

The implementation of this data cube approach takes advantage of the latest cloud-ready formats for big chunked spatio-temporal data sets. Here we use the Zarr format (<https://zarr.readthedocs.io/en/stable/spec.html>) that focuses in cloud technologie storage and can be efficiently handled in Python (<https://zarr.readthedocs.io/en/stable/>) and Julia (<https://github.com/meggart/Zarr.jl>). In addition to the Python xarray interface we offer another interface implemented in Julia, a novel high-level programming language for scientific computing, in the ESDL.jl package by Fabian Gans (co-author), and the most up-to-date documentation is always available in the respective GitHub repository <https://github.com/esa-esdl/ESDL.jl> (last visit Apr 17, 2020). All the Julia ESDL.jl packages and facilities have been transferred to the RegESDL offering a very flexible and efficient way for processing. The ESDL software is open source and available under the MIT license.

2.2 Analysis Ready Data Cubes

In our study, we defined ARDCs as spatio-temporal datasets, usually provided by different sources, stored in a uniform grid and located at common chunks. **Figure 1** shows a schematics of the ARDCs main features and the stepwise generation. ARDCs support immediate interaction between different data sources. Rechunking is unnecessary for analysis along the spatial or temporal dimension speeding up the processing. ARDCs in the RegESDL extend from latitude 14° N to 14° S and longitude -83° W to -60° W. The spatial resolution is 0.0083° × 0.0083° (approximately 0.9 km × 0.9 km at the equator). We consider this spatial resolution a good compromise for regional studies, preserving sufficient spatial details for general climate patterns, ecosystem gradients and main relief features. Nevertheless, it is limited in very steep areas and heterogeneous landscapes in the Andes. The temporal resolution is 8-daily and the covered period is from 2001 to 2014. When necessary, data has been resampled or interpolated to match the spatio-temporal grid. Temporal and spatial aggregation was done using the package gridtools (<https://github.com/esa-esdl/gridtools>), further details are included in Table S1. The total size of the ARDCs is 3.03 TB. We present the RegESDL ARDCs in three categories; time series, descriptive variables and national layers. A comprehensive list of the data with details of the original resolution, interpolation method, spanning time, source and license is in the **Supplementary Tables S1–S5**.

Time series are mainly data sets from models or satellite products related to vegetation and climate (**Supplementary Table S1**). We compiled data of gross primary productivity (GPP), evapotranspiration, shortwave radiation, photosynthetically active radiation (PAR) and diffuse PAR from the Breathing Earth System Simulator (Ryu et al., 2011; Jiang and Ryu, 2016; Ryu et al., 2018) describing ecosystem functioning. The selected products from the Moderate Resolution Imaging Spectroradiometer (MODIS) are 8-daily composites of leaf area index, fraction of absorbed photosynthetically active radiation (FPAR), and 16-daily composites of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) which are values of standing vegetation and greenness. Day and night land surface temperature was also obtained from MODIS. Albedo data was acquired from the Quality Assurance for Essential Climate Variables project (<http://www.qa4ecv.eu/>). Precipitation data was provided from two sources; the Tropical Rainfall Measuring Mission (TRMM, <https://trmm.gsfc.nasa.gov/>) and The Climate Hazards Group Infrared Precipitation with Stations (Funk et al., 2015). Time series with different temporal resolutions were also included such as the annual land cover maps from the European Space Agency (ESA) (ESA, 2017), annual vegetation cover fraction from MODIS, monthly annual averages of cloud coverage (Wilson and Jetz, 2016) and monthly fire data from MODIS. We also incorporated quality flags from MODIS products. The quality flag criteria was implemented on the original files (i.e., sinusoidal projection) and it is documented in **Supplementary Table S2**. After defining the pixels with acceptable quality, data was

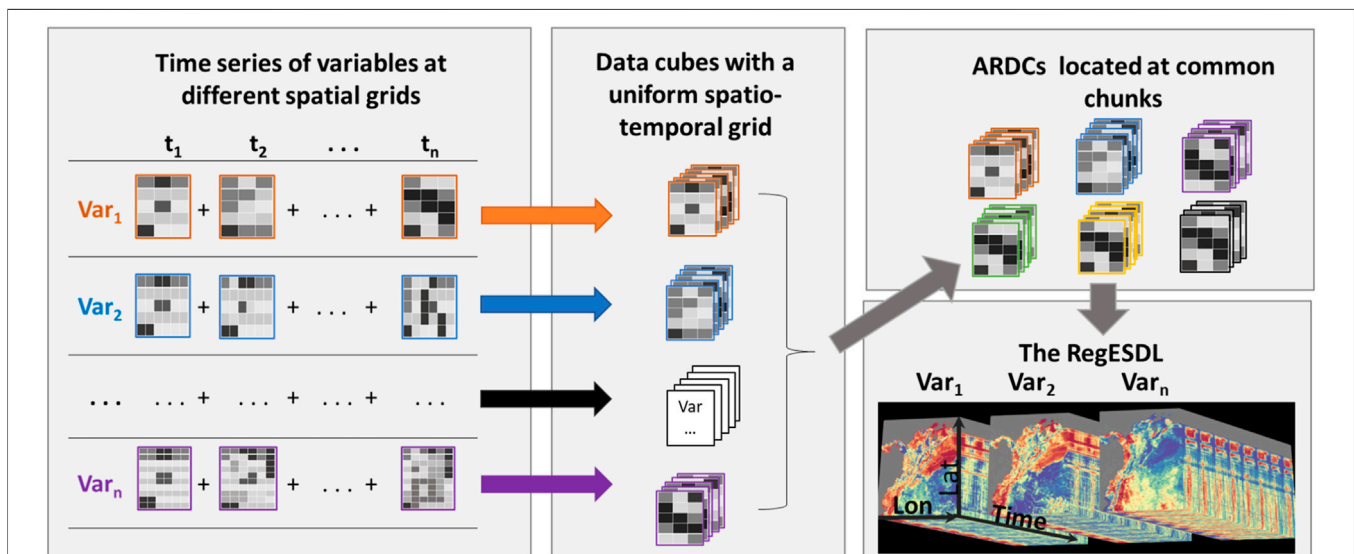


FIGURE 1 | Workflow of the Analysis Ready Data Cubes (ARDCs) generation and dimensions of the Regional Earth System Data Cube (RegESDL). Lat: latitude. Lon: longitude t: time, Var: variables.

reprojected to WGS84. For detailed information of all data products refer to the original documentation, references are included in the supplementary and bibliography.

Another important element relates to descriptive ecosystem variables. This category collects global datasets without a time dimension and describes some ecosystem features (**Supplementary Table S3**). There are 45 layers in total; canopy height (Simard et al., 2011), 11 variables of the Harmonized World Soil Database (FAO and ISRIC, 2012) and 23 of Soil Grids (Hengl et al., 2014) associated with soil composition and chemical properties at different strata. These ecosystem variables reflect a specific ecosystem state which is of importance for characterization, but they lack on offering a dynamic perspective. Alternative products bring the possibility for investigating soil water dynamics (e.g., soil moisture), however the coarse spatial resolution constrains its current use in our study.

Because the RegESDL also aimed to support the Colombian Biodiversity Observation Network (BON) geographical layers of Colombia were included (**Supplementary Table S4**). These data layers have been ingested from governmental web portals or from scientific publications. Layers in vector format were transformed to the target grid. Borders of administrative units (IGAC, 2010) and national natural parks (PNN, 2015) were included. Maps of wetlands (Flórez et al., 2016), agriculture frontier (MADR-UPRA, 2017) and biotic units (Londoño et al., 2017) were also added for further ecological analyses, as well as comparative interpolations of mean annual precipitation (Álvarez-Villa et al., 2011).

The RegESDL can be operated through different ways. First, the RegESDL can be accessed locally using Julia. In this case, the loaded datasets are exclusively the ones required for the analysis. Second, the RegESDL can be completely downloaded to any local machine using Python. Explanatory scripts of how to access the

RegESDL are included in the supplementary and at <http://doi.org/10.5281/zenodo.5068004>.

It is important to highlight that datasets described previously are the first core of the RegESDL. Nevertheless, the RegESDL is prepared for interacting with new data sources and can be easily expanded by users ingesting their own data as NetCDF or CSV files using the ESDL.jl package. The imported data layers must share the same grid extent to warranty Interoperability among multiple datasets. Vector files i.e., shapefiles are also supported, they can be loaded and converted to a grid for further processing in the ESDL environment.

2.3 The Regional Earth System Data Lab Coverage

The RegESDL focuses on tropical ecosystems of northern South America. It covers Colombia, Ecuador, Venezuela, and partially Brasil, Bolivia, Panama and Peru. From now on we refer as regional scale to the area covered by the RegESDL, and local scale to the country level and finer geographical units. We selected this region because it is facing a rapid ecosystem transformation due to land use change and urgently needs tools that help to understand ecosystem dynamics, contributing to fill a knowledge gap in the countries that encompass this region. Having a ready RegESDL to facilitate big-data analytics for ecosystem function is key on understanding one of the most diverse regions of the world.

This region is dominated by extensive natural tropical ecosystems, and multiplicity of climates related to topographic gradients, trade winds, and the meridional migration of the Intertropical Convergence Zone that drives annual variability. The highest altitude is in the peruvian Andes in Cordillera Blanca with more than 6,500 masl. Dominant land cover types are broadleaved evergreen forest

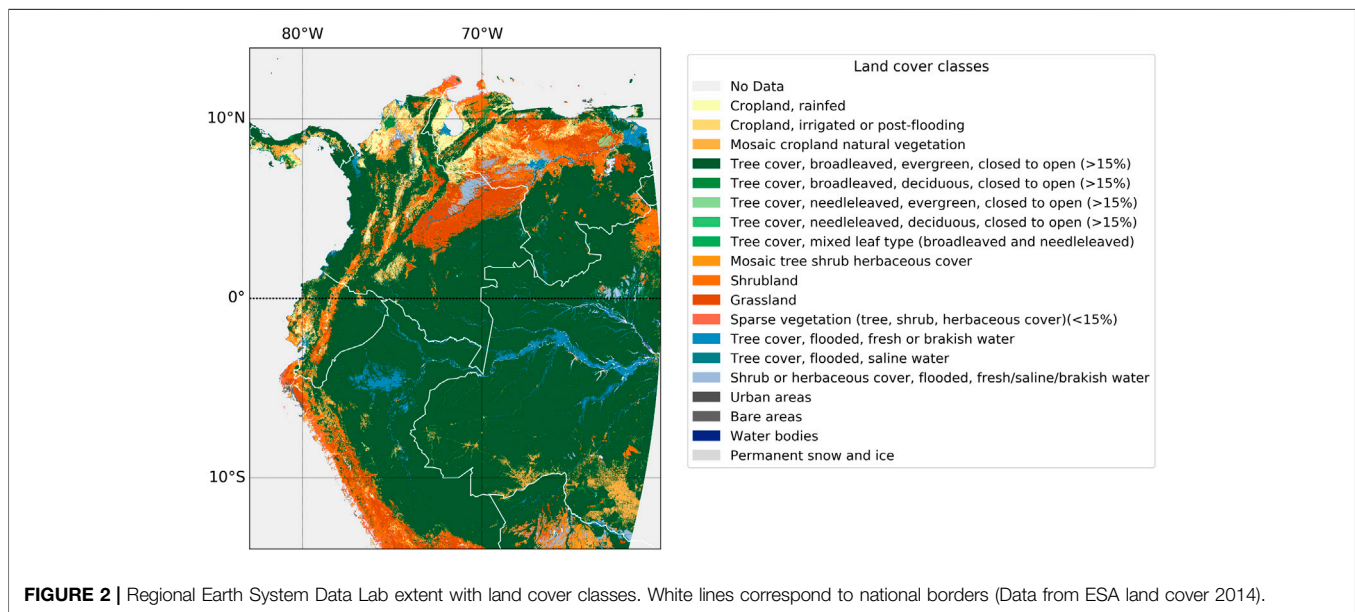


FIGURE 2 | Regional Earth System Data Lab extent with land cover classes. White lines correspond to national borders (Data from ESA land cover 2014).

open or close > 15% (4,740,043 km²), grassland (474,959 km²) and shrubland (266,967 km²) according to the ESA land cover classification (ESA, 2017) (Figure 2). Regional climate interaction with the Andes creates a variety of microclimates that cause two major hydrological regimes; unimodal and bimodal seasonal dynamics in many hydro-meteorological processes. In general, all feedbacks in the water cycle are governed by complex lateral interactions across the Amazon and the Andes, but also locally caused by precipitation recycling due to orographic gradients (Poveda et al., 2006; Bedoya-Soto et al., 2019; Espinoza et al., 2020). The mean annual precipitation records in the RegESDL area range from less than 100 mm in the Peruvian coast, upper limit with Atacama desert, to more than 11,000 mm in the Choco region of Colombia, which is perhaps the rainiest region on Earth (Poveda and Mesa, 2000; Yepes et al., 2019). The diurnal cycle of temperatures is the most salient feature of tropical climatology. Variations in temperature are often larger within a day than throughout the year (Hastenrath, 1991), with strong effects on the diurnal cycle of precipitation in the tropical Andes (Poveda et al., 2005). The region currently faces increasing rates of deforestation and land cover change.

In the following, we emphasize the territory of Colombia which is considered the third most biodiverse country and a hotspot for biological conservation (Myers et al., 2000; Andrade, 2011). Currently, Colombia is facing a massive transformation of natural ecosystems due to various socioeconomic transitions (Baptiste et al., 2017; Sierra et al., 2017; Salazar et al., 2018). Deforestation is now reaching national protected areas (Armenteras et al., 2019b; Clerici et al., 2020), and the agricultural frontier is also expanding to other natural ecosystems (Miles et al., 2006; Etter et al., 2008; Bianchi and Haig, 2013; Patino and Estupinan-Suarez, 2016; Correa Ayram et al., 2020). The RegESDL incorporated national layers to facilitate analysis at this scale with a special

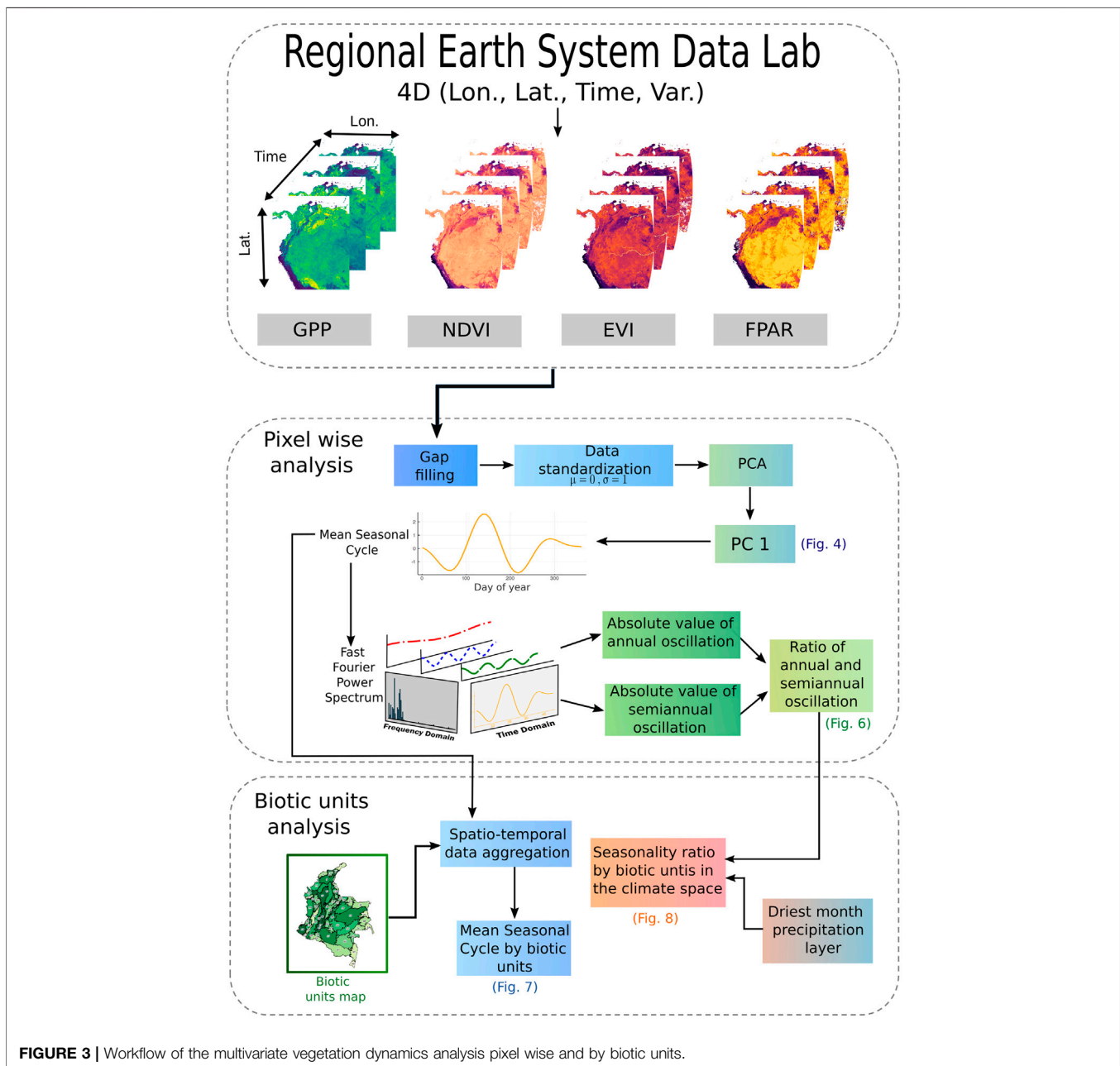
focus on biodiversity and ecosystems research (see section 2.2). Moreover, some variables were selected to provide seasonal and longer-term information to Biomodelos (<http://biomodelos.humboldt.org.co/es>) aiming to get a more dynamic perspective of species distribution models. For this particular reason they both share the same grid extent. The RegESDL also aimed to contribute to the development of Essential Biodiversity Variables (EBVs) in mega-diverse tropical countries. In this case we thought on a top-down approach for biodiversity monitoring. The Colombia BON, one of our partners, has done an extensive development and implementation on the topic.

2.4 Case Study Question

In the following, we showcase the potential of using the RegESDL for assessing seasonality in northern South America. Here, we investigated the joint variability of multiple time series, and contextualized the identified patterns in different spaces (i.e., climate space, geographical space, along biodiversity gradients). This analysis had two purposes. One was to carry out a seasonality analysis pixel wise for the region. The other one emphasized on the Colombian territory and its biotic units. A schematic of the RegESDL and a workflow overview is in Figure 3.

2.4.1 Seasonality in Northern South America

We focused on seasonal phenology, one of the fundamental dynamics of most ecosystems of the world (Schwartz, 1998). In tropical ecosystems, however, seasonal dynamics are hardly understood and not well characterized (Wu et al., 2016). On the one hand, this is due to data quality issues, but on the other hand it reflects the fact that the tropics embrace extremely heterogeneous landscapes. For northern South America, a large level of annual variability in phenological variables has been reported, particularly for savannas, dry forest and wetlands



(Estupinan-Suarez et al., 2015; Hamunyela et al., 2016; Fagua et al., 2019). Subannual data acquired from tree ring cores (Giraldo et al., 2020), and from space via solar induced fluorescence (SIF) (van Schaik et al., 2018) recently also suggested annual seasonality in tropical rainforest. From the functional perspective, Nemani et al. (2003) reported radiation as a limiting factor of plant growth and net primary productivity in the tropics that likewise hints at some seasonality.

What remains unclear from the literature is: i) how strongly are seasonal patterns reflected by terrestrial ecosystem dynamics, ii) whether we can delineate unimodal or bimodal regions, and iii) how do patterns of annual and semi-annual seasonality distribute spatially. If we understand where unimodal and bimodal

seasonalities predominate in land-atmosphere interactions, we could achieve a better predictive understanding of the imprints of extreme climate events such as ENSO, and climate change signals on ecosystems.

Terrestrial seasonal dynamics should be contained in all remote sensing indicators related to green biomass and primary production. Available data sets in the RegESDL are partly direct remote sensing vegetation indices of greenness, i.e., NDVI (Tucker and Sellers, 1986) and EVI (Huete et al., 1997, 2002). But also, we can analyze GPP and FPAR that are closely related to vegetation activity. Conceptually, these variables represent different processes which are related and physiologically connected, but they are not exchangeable. Of course, the closer to the actual process under interest (e.g., GPP), the more

model assumptions are contained in the data (Jiang and Ryu, 2016). The advantage of the original remote sensing signals is that they are closer to the purely observational signal, but not necessarily directly related to the process of interest. Yet, we can assume that all these time series somehow reflect the seasonal cycle of vegetation productivity, yet coupled via more or less direct mechanisms that all reflect seasonal dynamics of green vegetation.

2.4.2 Characterization of the Mean Seasonal Cycle Pixel Wise

In order to capture the seasonal variability of vegetation while accounting for the redundancy of the different vegetation related signals, we conducted a principal components analysis (PCA). PCA is a dimensionality reduction method that seeks new dimensions (components) in the feature space to explain the largest variance, and does it recursively based on orthogonal basis functions. In this sense, it provides common modes of variability at the pixel level that serve to assess seasonality. For our analysis, we selected GPP, NDVI, EVI and FPAR variables spanning from 2001 to 2014 (data set size = 111.24 GB). Variables were gap filled using the Mean Seasonal Cycle (MSC) method. Gaps were clustered in rivers and waterbodies and its amount varied among variables; while GPP excluded all data pixels related to water, MODIS products preserve or exclude them irregularly. The next step was to standardize the time series to mean zero and variance of one ($\mu = 0$ and $\delta = 1$), and then applied PCA at pixel level. Explained variances by the leading components described how representative the leading mode is for the different vegetation variables.

Based on the Fast Fourier power spectrum, we estimated the contribution of annual and semiannual oscillations of the MSC pixel wise using the discrete Fast Fourier transform library (FFTW) included in the Julia programming language. Here, the idea is that the ratio between the spectral power of the annual and semiannual oscillations can be quantified if the annual or semiannual oscillations dominate the seasonality of the signal. Theoretically, values of <1 indicate a dominance of the semiannual oscillations, values close to 1 are related to an equal contribution of seasonal and semi-annual modes, while values of >1 reflect a dominance of the annual oscillation. We classified the FFTW outcomes as follows: The first component is the sinusoidal function offset, the second component corresponds to the annual oscillation, whereas the third and fourth components together (6 and 4 months respectively) correspond to the semiannual oscillation. Subsequently, we computed the contribution of both annual and semiannual oscillation to the entire signal. Only pixels with at least one direct retrieval at each MSC time step were included in the analysis. The quantification of direct retrievals was based on the quality flags from MODIS variables (Supplementary Table S2). The computation is documented in Jupyter notebooks included as supplementary material and at the GitHub repository https://github.com/linamaes/Regional_ESDL.

2.4.3 Seasonality Characterization of Biotic Units in Colombia

For our second study, we used the biotic units map of Colombia produced by Londoño et al. (2017) as a level of aggregation,

aiming to bring a functional understanding of units with biological diversity connotation. The delineation of these biotic units was based on a Beta diversity criteria defined as the taxonomic composition variation (Supplementary Figure S1). Each unit was defined by having a unique set of species that was significantly different to the species set of all other units. Beta diversity was computed from species distribution models of amphibians, birds, mammals, reptiles and plants available in Biomodelos, a collaborative platform that integrates models and expert knowledge (Olaya-Rodríguez et al., 2018; Velásquez-Tibatá et al., 2019). Our approach looks at biodiversity based on the hierarchical concept developed by Noss (1990), and seeks a functional perspective including different taxa similar to Radeloff et al. (2019). It is not related to spectral diversity for ecosystem function assessment.

We calculated the MSC of biotic units using the first principal components (PC) obtained from the pixel wise analysis. First, we computed the average and standard deviation of all pixels within each unit. Then, we calculated the ratio and fraction values of the averaged MSC based on the Fast Fourier power spectrum as we explained in section 2.4.2. We also analyzed the biotic units outcomes in light of climatic variability. For this, we selected variables that describe annual bioclimatic conditions such as precipitation of the driest month, maximum temperature of the warmest month, mean temperature of the driest quarter from WorldClim (Fick and Hijmans, 2017), and median annual cloud frequency (Wilson and Jetz, 2016). These variables were used in the climate space and were related to the MSC ratio by biotic units.

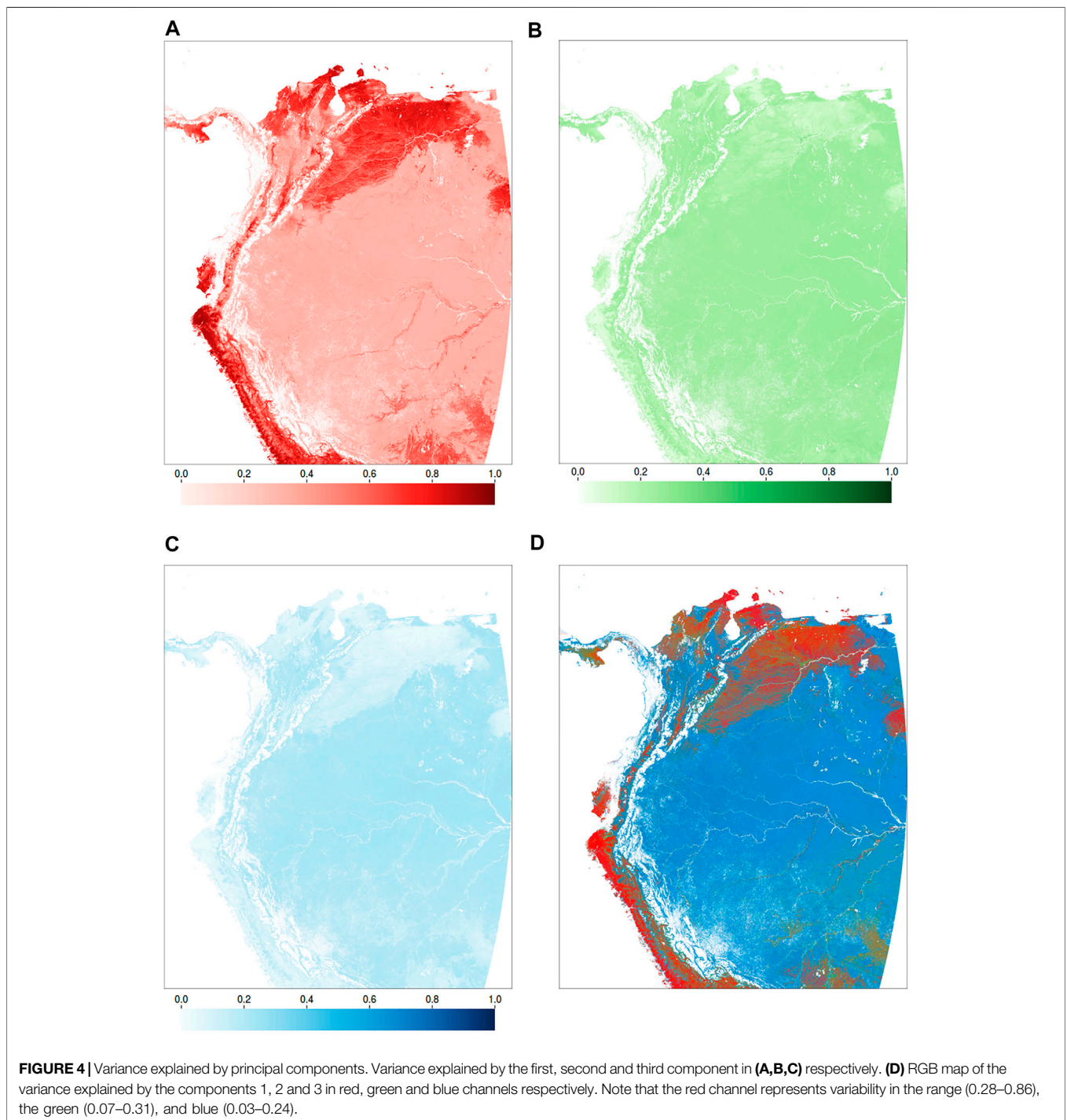
All analysis ran in Julia 1.3 using the ESDL package v0.8.4. Processing was done in an Intel®Xeon®Processor E5-2687W v4 CPU (30M Cache, 3.00 GHz), and used six cores.

3 RESULTS

3.1 Summarizing Multivariate Vegetation Dynamics

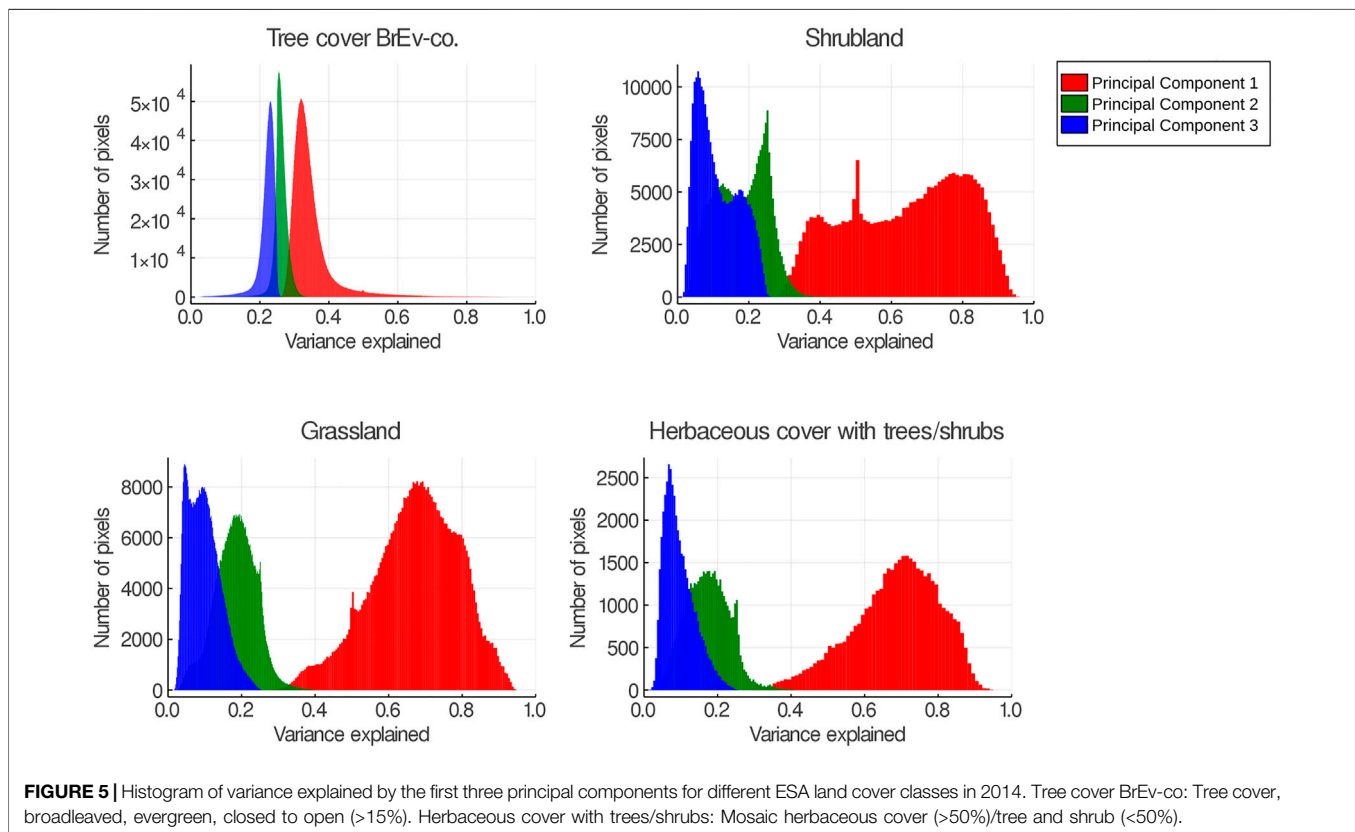
The first PCs from the PCA captured the largest vegetation variability pixelwise (Figure 4). When exploring pixels from different land cover types, we found overall that PC1 captured the main MSC features of each variable (Supplementary Figure S2). In general, for broadleaf evergreen trees and grassland the variables' trajectories and peaks overlapped most of the time, although the signal amplitude differed. The shrubs-herbaceous flooded cover was the one presenting the most contrasting trajectories between variables. The main contributors to PC1 are different along the regions (Supplementary Figure S3), NDVI contributed the most in arid and semiarid regions such as the Caribbean and Orinoquia savannas and at the Pacific coast of Ecuador and Peru, whereas EVI is the larger contributor in the Amazon and Biographic Choco i.e., the wettest regions. Some pixels were excluded from the analysis due to data quality. They are mainly located along the Pacific coast, in the higher Andean mountains, and in the transition between the mountains and lowlands.

We observed different regional patterns of variance explained by the three principal components. PC1 (Figure 4A) explains



the largest variance in ecosystems mainly dominated by savannas in the Orinoco and Caribbean basins of Colombia and Venezuela, and lowlands of the Magdalena-Cauca river basin. Similar pattern occurs in the Pacific coast of Ecuador and Peru. These regions are characterized by very low precipitation during the dry season and are dominated by grassland. PC2 (**Figure 4B**) shows a homogeneous spatial pattern, being slightly lower in arid and semiarid ecosystems. Otherwise, PC3 (**Figure 4C**) dominates the broadleaf evergreen forest. The explained

variance by the PCs range from; 0.28 to 0.86 for PC1, 0.07–0.31 for PC2, and 0.03–0.24 for PC3. These shows that variance PC2 and PC3 can carry similar amount of variance in some regions. Overall, **Figure 5** shows that PC1 explains more than 40% of variance in grassland, shrublands and herbaceous cover with shrubs and trees. But it is certainly limited in broadleaf evergreen forest where captures between 20% and 40%, which is also a region known by large data gaps due to clouds.



3.2 Quantification of Temporal Dynamics

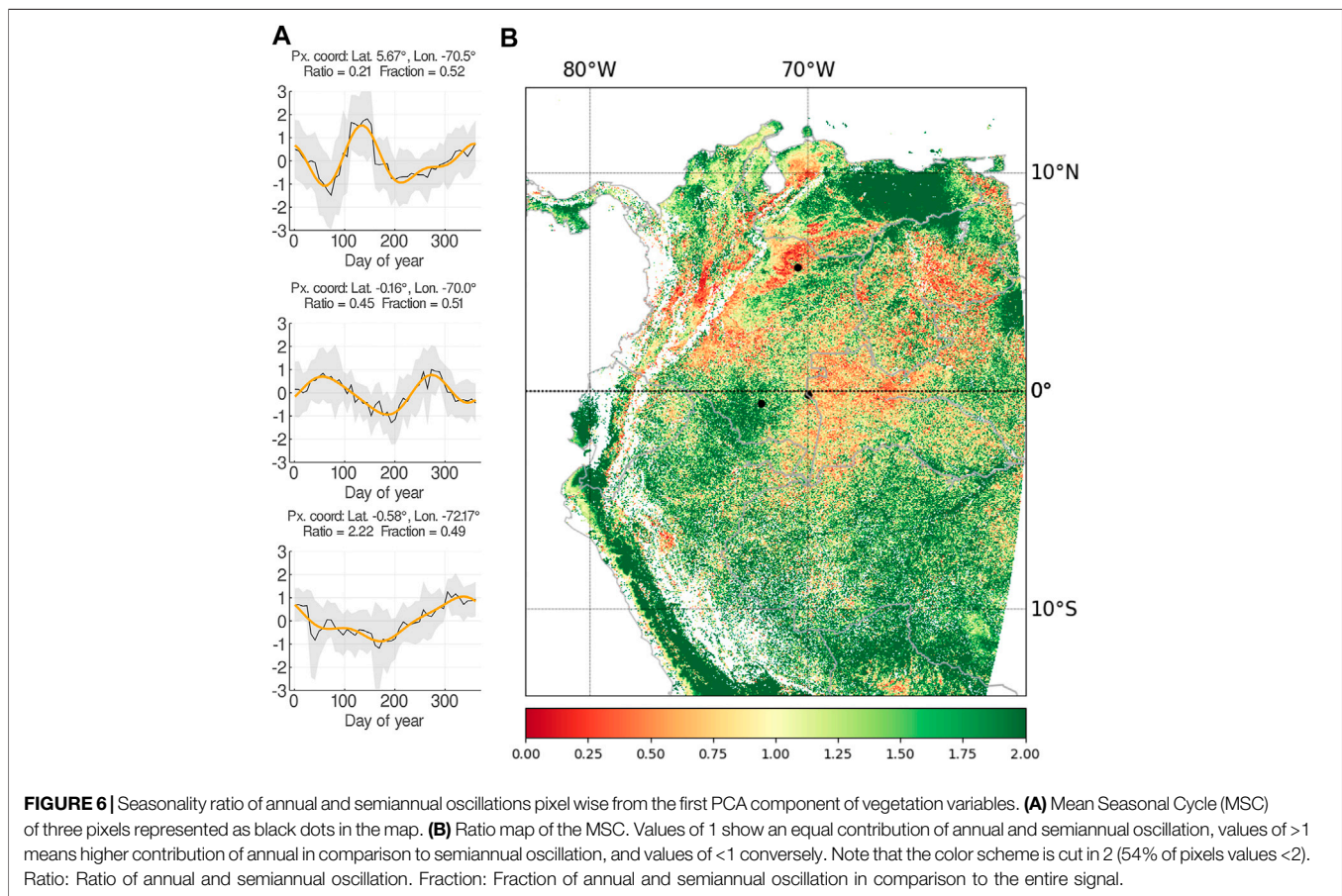
The map of seasonality ratios (**Figure 6**) depicts a spatially heterogeneous dominance of annual variability in the region that can be associated to the major landscapes. Annual cycles dominate the land-surface dynamics along the Ecuadorian and Peruvian coast and northern Venezuela. Large-scale patches of semi-annual oscillations are found in the flooded savannas of Orinoquia, and inter-Andean valleys in Colombia. Similar contributions of annual and semi-annual oscillations are observed partly in the Caribbean coast and foothills. The northwestern Amazon basin does not have a uniform pattern, although it shows a slight trend with dominance of bimodality close to the Equator (the wettest region), and unimodality toward South. Yet, it is hard to determine due to the optical sensors limitations in the rainforest. However, care must be taken in interpreting this figure, taking into the account the variability represented from the total signals (**Figure 4**) and the amplitude of the oscillations. In this regard, we estimated the fraction between both assessed oscillations (i.e., annual and semiannual) and the entire signal based on the power spectrum of Fast Fourier. These values are found in **Figures 6, 7**.

3.3 Seasonal Dynamics of Colombian Biotic Units

In order to understand the spatial variability of vegetation seasonality and its links to biodiversity we used the biotic

units of Colombia. We observe that the units with higher amplitude are also the ones where annual and semiannual variability have a fraction value of >0.7 (**Figure 7**). Overall, we observe that seasonality in the biotic units is extremely different; peaks of vegetation activity are reached at different times across Colombia, unimodal and bimodal seasonality are equally important at the national level, and regions with lower variability explained by these regimes have to be further explored. These could be associated with the dominance of fast oscillations as reported by Linscheid et al. (2020) for some tropical regions.

We used bioclimatic variables as a first proxy to understand seasonal variability. We selected precipitation of the driest month, maximum temperature of the warmest month, mean temperature of the driest quarter (Fick and Hijmans, 2017), and median annual cloud frequency Wilson and Jetz (2016). As an overview, six biotic units with contrasting patterns are presented. We observe that the lowest values of monthly annual precipitation differ significantly within biotic units. In *Arauca* and *Baja Guajira-Cesar* values are less than 30 mm (**Figures 8A,D**) whereas in *Micay* values are above 300 mm on average (**Figure 8F**). Interestingly, in the region *Magdalena-Medio* & *Depresion Momposina*, there are two clear hotspots (**Figure 8E**). Pixels with precipitation in the driest month of >50 mm show a bimodal dynamics in vegetation dynamics, while pixels with drier conditions show a higher importance of annual oscillations. This could be associated with the distribution of



different land cover types, i.e., broadleaf evergreen trees and rainfed croplands, based on the rainfall patterns. When assessing the maximum values of annual temperature, biotic units in the lowlands are very confined in the space gradients. Andean biotic units are the ones showing higher levels of variability such as in the *Patia* region (**Supplementary Figure S4**). It is important to highlight that in Colombia, the largest temperature variability is mostly found on a diurnal basis than along the year (Hastenrath, 1991).

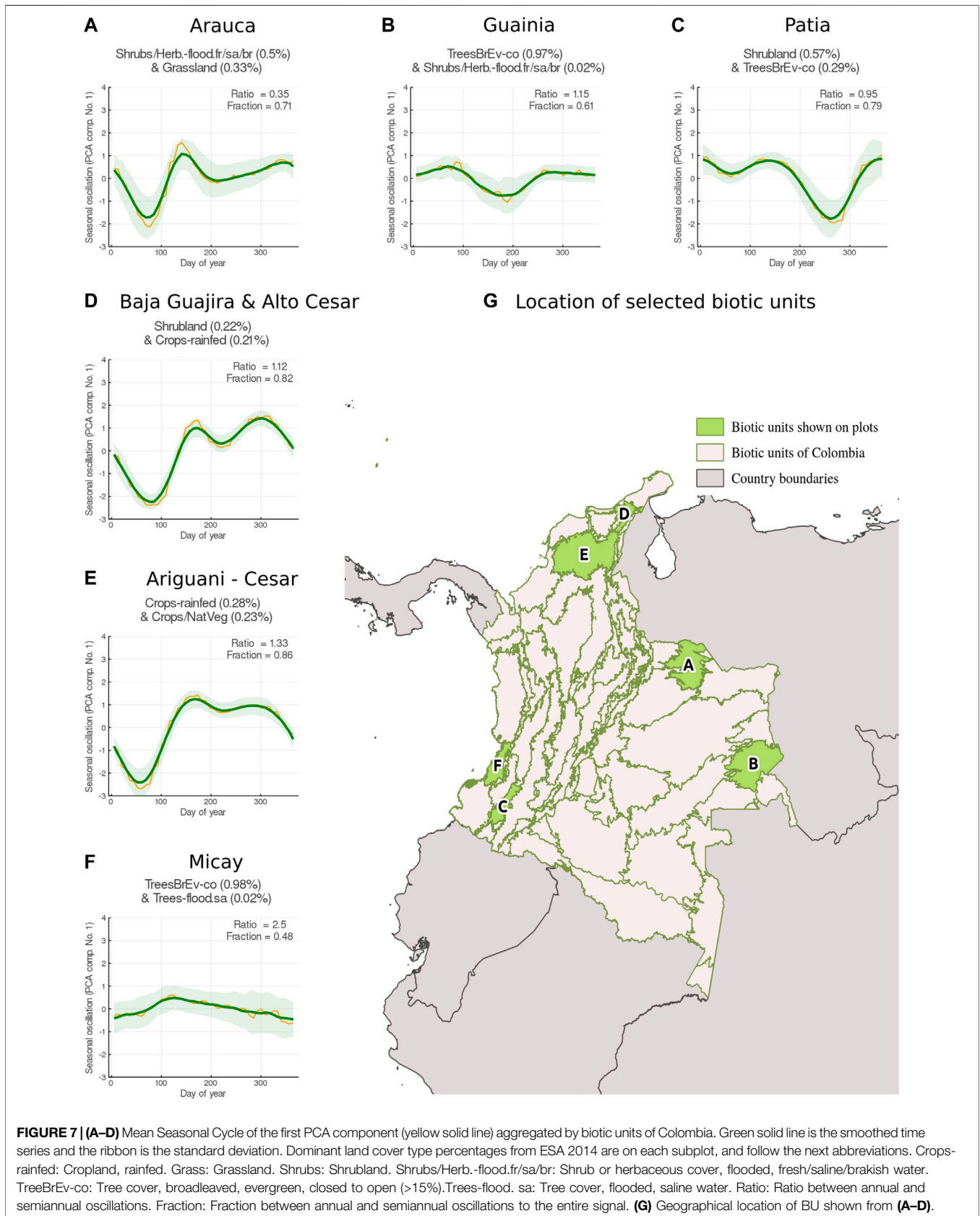
Due to the orographic conditions of our study area, clouds are a major limitation for passive sensors, however recording these conditions is also informative in an ecological context. Biotic units with a strong dry season are the ones with lower values of annual cloud frequency. Nevertheless, they either show a centralized pattern as *Baja Guajira* or a large spectrum as *Magdalena Medio*, similar to the one observed in *Patia* (**Supplementary Figure S5**). *Micay* exposes high cloud cover as it is expected due to its location in the biogeographic Choco region, one of the wettest places on Earth (Poveda and Mesa, 2000; Yepes et al., 2019).

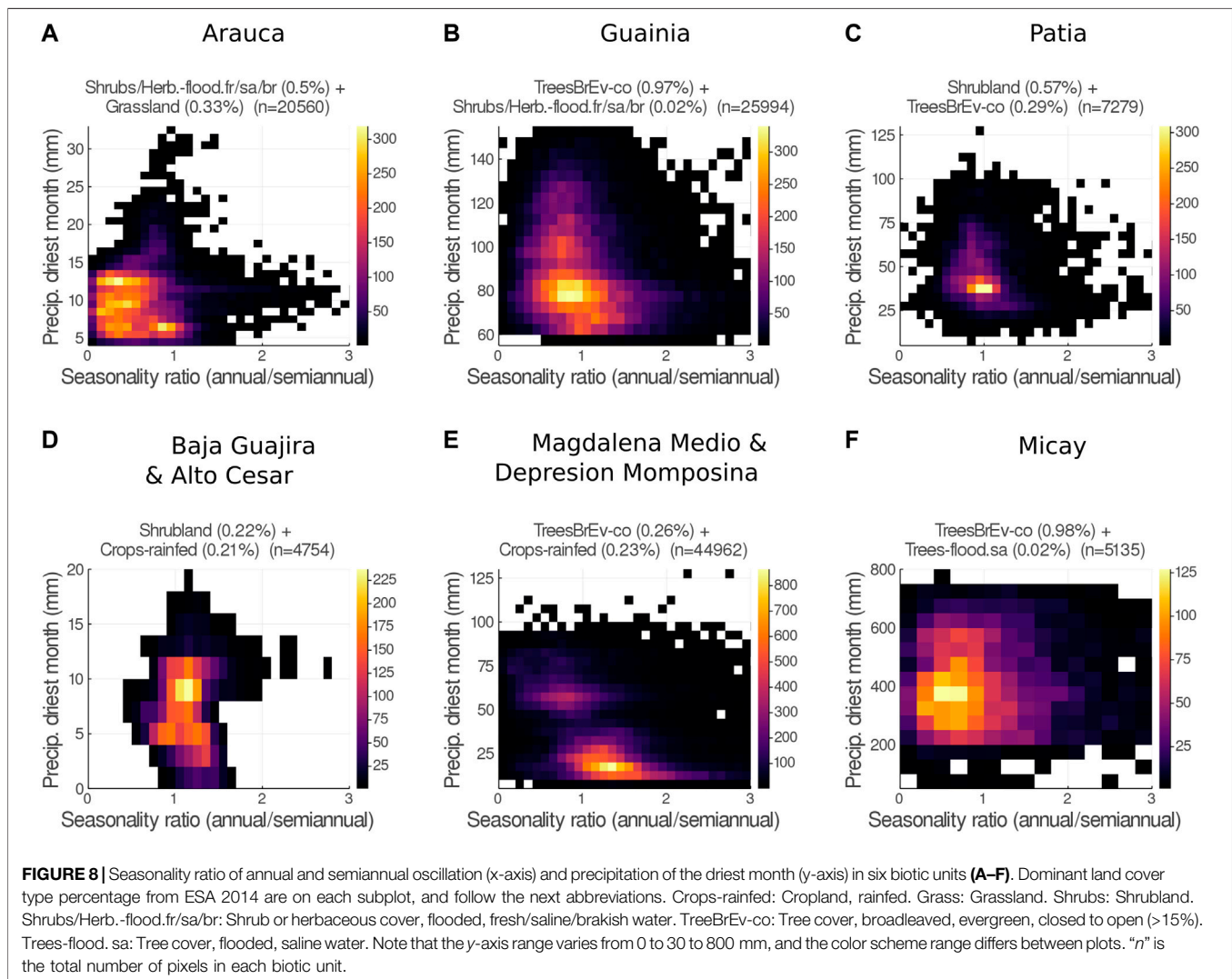
As we have shown in this case study, the seasonal land-surface dynamics in northern South America is surprisingly complex and not trivially explainable by climatic factors alone. Our results are a step forward to reveal interactions between biotic and abiotic components in tropical ecosystems. From a technical perspective, we show that multi-dimensional analysis in any dimension

i.e., along physical variables, time, and space, can be performed very efficiently with a few lines of code. The code (available at <http://doi.org/10.5281/zenodo.5068004>) can be applied to any other data cube, e.g., the ones that are now emerging from the Coupled Model Intercomparison Project Phase 6 (CMIP6) archive and used to understand if patterns as reported here, are similarly identifiable in such global simulations. If they were not, this would suggest that one of the most basic processes of land-surface dynamics in state-of-the-art models is not well represented yet.

3.4 Computational Performance of the Regional Earth System Data Lab

During the multivariate vegetation analysis, we used a set of four variables (i.e., GPP, NDVI, EVI, FPAR) available as ARDCs at the RegESDL with a total size of 111.24 GB. The estimated time processing for the main steps was: 27 min for the time series standardization, 2.5 min for PCA and 2.8 min for computing the Fast Fourier spectrum. All these computations were carried out at the pixel level using the RegESDL for time series analysis. Overall, these estimates are broad because timing for data loading and processing are hardly discernible, furthermore speed is also affected by other features as parallel processing in other cores. Nevertheless, these timing values bring a general picture of the





computational performance. A comparison between the spatial and temporal RegESDL computational performance, for time series analysis, is in the supplementary (**Supplementary Table S6**) and the code is available at <http://doi.org/10.5281/zenodo.5068004>.

4 DISCUSSION

Open data and open source software can be seen as the major triggers for recent advances in environmental and Earth system sciences. A remarkable example are the achievements after the Landsat archive was opened for ecosystem monitoring, assessment of climate impacts, among others (Wulder et al., 2016). Moreover, data streams are continuously improved; spatio-temporal and radiometric resolutions increased with the development of new satellites and sensors (e.g., the Sentinel constellation). This increasing data availability also raises enormous challenges for data management and ways to interact with them. Data cubes have

been a common solution for tackling this problem bridging the code to the data, and therefore simplifying data storage and processing (Lewis et al., 2017; Giuliani et al., 2019; Gomes et al., 2020; Tamimonia et al., 2020; Yao et al., 2020). Nevertheless, most initiatives prioritize spatial grids for data storing (Gorelick et al., 2017; ODC, 2021), and work with specific data sets such as climate or satellite products exclusively. In comparison, our approach considers that all data dimensions are equally important whether they are space, time, variables, or frequency. This has two main advantages: i) Facilitate access to other axes different from latitude and longitude. ii) Implement the cube structure to outputs as is for inputs. With the ESDL package new-axes can be easily added to cubes during processing. Users have full functionality for defining new-axes of a “hypercube”, assign their name and type class. In addition, the RegESDL links data across disciplines offering a multidimensional perspective for environmental sciences. Thus, we are capable of analyzing multiple data sources efficiently at spatial and temporal dimensions which is key for tropical regions facing rapid transformation.

4.1 Perspectives of the Regional Earth System Data Lab

The RegESDL started as an initiative for supporting analysis that integrate Earth science and biodiversity within a tropical geographical range. The main objectives were to get a dynamic perspective of environmental conditions for biodiversity, to contribute to the understanding of ecosystems functioning and to explore atmosphere-biosphere interactions. Until now, studies have focused mostly on evaluating land cover and ecosystems structure for the region. This has been pivotal for the assessment of deforestation and in some extent forest degradation (Armenteras et al., 2016; Pacheco-Pascagaza et al., 2018; Ramírez-Delgado et al., 2018; Meyer et al., 2019; Anaya et al., 2020). As well, different passive and active sensors have been used for ecosystems delineation (Estupinan-Suarez et al., 2015; Flórez et al., 2016), estimation of vegetation biomass and extraction of canopy features from individual trees and forest using Lidar (Asner et al., 2012; Li et al., 2012; Ferraz et al., 2016; Jeronimo et al., 2018; Ferraz et al., 2020). However, fewer studies are carried out for investigating ecosystem function even though global products such as GPP are available from several sources i.e., GPP estimates derived from satellite retrievals (Running et al., 2004), coupled to process-based models (Ryu et al., 2011; Jiang and Ryu, 2016), and data driven methods (Jung et al., 2011, 2020). This can be partially related to the large uncertainties regarding tropical ecosystems due to atmospheric conditions and limited ground data for models calibration and validation. But also to the early state of Earth science development in the region.

The growing availability of new retrievals from upcoming satellite missions and advanced mathematical methods offer new information and alleviate some gaps regarding data accuracy and quality. Promising variables such as SIF open a possibility to improve estimates of GPP and phenological changes (Porcar-Castell et al., 2014; Walther et al., 2016; Sun et al., 2017; Merrick et al., 2019). Recently launched and upcoming satellites missions will deliver hyperspectral and lidar data globally that will facilitate structure and functional biodiversity assessment. Satellite missions such as the DLR Earth Sensing Imaging Spectrometer, the Global Ecosystem Dynamics Investigation (GEDI) or the Surface Biology program will provide key information for evaluating vegetation stress, vegetation traits, and improving carbon and water fluxes estimation. Thus, tools and long-term projects are needed to warrant timely and efficient access to these new data streams. In this sense, the ESDL software offers a suitable framework to address the technical developments required by the unprecedented volume of coming datasets, and most importantly for integrating different suites across research disciplines. Recently, Cremer et al. (2020) implemented the ESDL for higher resolution analysis, specifically for assessing Sentinel-1.

In fact, the RegESDL is an example of a regional effort for offering a common ground to assess Earth system science, ecosystem function, and explore links to biodiversity that could be realized in practice. Specifically during our showcase we presented different modes of seasonality in the region. Studies

of bimodal precipitation regimes have been carried out globally (Knoben et al., 2019) and regionally in East and Central Africa, and northern South America (Poveda et al., 2006; Hawinkel et al., 2015). Nevertheless, such patterns are poorly explored at the ecosystem level. In East and Central Africa double peaks of vegetation greenness were observed by Hawinkel et al. (2015) at annual scales. Recently, Turner et al. (2020) reported bimodality in vegetation activity for savannas in California based on SIF measurements that were previously jeopardized when using vegetation indices. Our outcomes revealed that these double peaks on vegetation activity are also occurring in regions of northern South America. But further analysis are needed to understand the mechanistic process of bimodality in the region. For example, it is expected to observe different timing for leaf flushing in savannas. Hypothetically, this might be driven by water pulses and different vegetation strata. Otherwise, studies using passive sensors are limited in some regions despite of having more than 10 years data. In this context, data integration from upcoming missions with higher spatial and multispectral resolution and active sensors is key. Moreover, the biotic units analysis showed that the dominance and intensity of seasonality in the tropics is very heterogeneous, and it requires deeper analysis on the drivers that are not covered in this paper.

Future studies in the RegESDL aim to exploit time series analysis to evaluate tropical ecosystems dynamics at different time scales. Also, a next step to increase our understanding of how biodiversity is related to ecosystem function is to integrate species distribution data, from platforms such as the Global Biodiversity Facility (GBIF), with data from the RegESDL. This will allow the scientific community to understand how the spatial distribution of specific taxa (not necessarily plants) could respond not only to abiotic variables but also to dynamics of the land surface. Moreover, this can be a benchmark to predict patterns of species migrations by global warming.

4.2 Comparison with Alternative Regional Projects

Colombia has done a major effort to develop strategies for big data generation and management. The National Institute of Environmental Studies and Meteorology has released the latest Climate Forecast System Reanalysis models for the region from hourly to daily temporal resolution, as well as national climate change scenarios and analyses of extreme events thresholds (Ruiz M. et al., 2020) (<http://bart.ideam.gov.co/wrfideam/>). Another example is the development of the Colombian ODC (Ariza-Porras et al., 2017; Bravo et al., 2017) to reduce redundancy in satellite imagery acquisition, pre-processing and storage. Simultaneously, the scientific community is taking advantage of GEE to assess rapid socio ecological challenges such as deforestation using satellite imagery or products derived from passive (Clerici et al., 2020) and active sensors (Anaya et al., 2020). From the biodiversity perspective, Colombia is contributing significantly in an international context; it is the only country with a National BON in the tropics, its biodiversity information system (SIB Colombia, <https://sibcolombia.net/>)

contributes to the GBIF, and since 2017 different projects are supporting the assessment and implementation of EBVs; specifically for improving species distribution model platforms (Velásquez-Tibatá et al., 2019), supporting biodiversity decision making (<http://biotablero.humboldt.org.co/>, <http://www.bosproject.org/en/>), and participates in the calibration of new satellites missions such as GEDI in tropical dry, moist and rain forest (Fagua et al., 2019), TRMM and the Global Precipitation Measurement (Vallejo-Bernal et al., 2020). All of these efforts point to an urgent necessity for developing a common framework that improves data access and management and facilitates ecosystem assessments with a more dynamic and functional perspective, but also for evaluating the rapid natural and anthropogenic changes that are occurring. The RegESDL aimed to bring the initial foundations for these developments, integrating data from biospheric, and atmospheric variables with biodiversity.

4.3 Technical Challenges for the Future

One of the major challenges of all data cube facilities is that data is constantly being updated. Not adopting a “living data paradigm” is one of the major obstacles that reduces the user-update of project-based infrastructures. Also in our case, no matter how convincing the concept and implementation might be, it will not persist as a stand-alone implementation next to a machinery that can update the underlying data archives in near-real time. This is why we hope that the analytic framework as it was developed here can soon be coupled to data cube facilities that solve the underlying data availability issue. In fact, in Europe, for instance, the Data and Information Access Services initiatives are about to realize such an approach such that initiatives like the RegESDL can be placed on top of them. The difficulty today remains that the cube type of data access along all dimensions remains often not well addressed and that the idea is rarely to give the full spectrum of functionalities to the users as ESDL does. This is to enable users to map arbitrary functions. Otherwise, initiatives centered on specific satellites products have opted for automated ingestion based on rapid data acquisition from platforms as GEE (Giuliani et al., 2017). However, this alternative is hard to extrapolate when working with multiple data sources.

We believe that in the future, the ESDL software should not merely be a facility that hosts predefined ARDCs but become a service to generate them tailored to individual user needs. These user-defined cubes should be able to ingest arbitrary gridded and non-gridded data. Particularly, higher resolution datasets are requested when moving from regional to local studies or for *in-situ* data up scaling. Here, accessing high-resolution and latest global data products e.g., SIF from various sources on demand and in tandem with other existing data suites is key. In this sense, ARDCs may be static and pre-processed, transient, automatically updated, or lazy, whatever suits best its intended usage. This means that the ESDL needs to serve data streams from multiple sources, and therefore handle multi-resolution data within the same framework.

Only if we can convince the major infrastructure providers of our concept, we soon will address the most pressing data analytic

bottlenecks that remain open: That is, the need to further generalize spatio-temporal data analytics. The current methods implemented in the ESDL are highly efficient in exploiting high-dimensional time series and maps, but lack one fundamental requirement: Spatio-temporal interactions and spatio-temporal contextualized data analytics are essentially not possible and limit our capacity to study e.g., telecouplings at large scales or lateral transport processes. This is of particular relevance, when aiming to simulate e.g., water transport in space and time via e.g., surface hydrology or atmospheric interactions. Another challenge for the future is certainly brining in latest advances in machine learning i.e., Deep Learning (DL), as one of the most relevant and rapidly developing fields. In principle this is possible already today, but whenever a DL model requires dealing with e.g., structured data (either spatial, temporal or spatio-temporal as in the Earth sciences), the cube-slicing approach is suboptimal and needs to be redesigned to efficiently map DL methods on the cube. This is particularly promising for understanding e.g., biotic dynamics in the tropics that are often not well constrained by our ecohydrological understanding.

Another major challenge is Interoperability. The exponential data generation and advancements in computer and software technology reinforced the urgency of data exchange between research infrastructure systems. This has to be accompanied by common metadata conventions, open algorithms and software documentation (Kissling et al., 2015; Wilkinson et al., 2016; Hardisty et al., 2019). For Earth observation data cubes, Giuliani et al. (2019) defined two tiers to approach this issue: an upstream tier which is the ODC infrastructure, and a downstream tier which is centered on the user's interaction. Here, we focused on data ingestion from several data sources which was crucial for multivariate analysis. With this approach we achieved to integrate different environmental layers from the biosphere and atmosphere. The usage of Zarr format for data storage, that is in the process of becoming the standard for the Open Geospatial Consortium (OGC, 2020), facilitates exchange with other geo initiatives such as the CMIP6 model. As well, the implementation of downstream software as xarray opens the possibility to develop backends with existing data cubes. We followed the Climate and Forecast metadata conventions (<http://cfconventions.org/>) that supports properties of the data. What is still remaining is the possibility of working with multi-resolution data, which is one of the following steps. These characteristics are favorable to Interoperability, but this is indeed a topic on constant development by the community.

4.4 Engaging the User Community

A long-term sustainability of the RegESDL requires an active users community. On one hand, the ESDL software is open, documented and its main features, i.e., the cube generation and analytical tools, can operate independently which brings versatility for forward software development and applications. On the other hand, consolidating such a community requires active dissemination and training to operate the RegESDL efficiently, besides constant technical support. Having a consolidated community is a key step for warranting the establishment and development of our tool. We see as

potential users research groups and universities that are interested in the spatio-temporal understanding of the biosphere-atmosphere interactions, implications of climate variability, land cover change, and biodiversity loss in tropical ecosystems.

Moreover, embracing big data initiatives, such ours, is also beneficial for supporting environmental studies that support environmental policy. For example, the Colombian government has appointed a roadmap for a strong bioeconomy development to the International Mission of Experts in Science, Technology and Innovation (<https://minciencias.gov.co/mision-sabios/documentos>), but there are many unknowns about ecosystems dynamics and functioning nation wise that need to be solved for creating such a plan. Studies bringing new insights for informing the sustainable goals, or the Intergovernmental Panel on Climate Change as well as research for understanding ecosystems functioning and ecosystem resilience under climate change scenarios could be efficiently carried out in the RegESDL. In this sense, our initiative could contribute to existing platforms that provide scientific input for decision making at local level.

5 CONCLUSION

To our knowledge the ESDL is the first data cube implementation with an emphasis on representing interactions across the water cycle, carbon cycle, and climate system (Mahecha et al., 2020). It has been successfully used to understand biosphere-atmosphere interactions at multiple time-scales (Linscheid et al., 2020), analyzing specific variables of ecosystems to climate extremes (Flach et al., 2020), and has enabled studying the multivariate nature of land-surface dynamics globally (Kraemer et al., 2020). The RegESDL has been developed to more specifically explore biodiversity as yet another thematic domain. Our goal was bridging the gap between Earth science and biological diversity that includes ecosystems composition, structure and function (Noss, 1990; Randin et al., 2020).

In this regard, the satellite era has brought us the possibility of gaining ecosystems dynamics understanding from systematic measurements over time and at larger spatial scales. Advances in the EBVs (Pereira et al., 2013), functional traits and functional ecosystem properties (Reichstein et al., 2014; Musavi et al., 2015) are supporting this development. And it is increasingly recognized that one has to consider the functional dimension of biodiversity in its own right, which requires working with high temporal resolutions. In particular, remote sensing observations are of uttermost importance for the analysis of the EBVs (Skidmore et al., 2015; Pettorelli et al., 2016; Giuliani et al., 2020; Randin et al., 2020). Recently, the EBVs for Species Population started to develop a framework for working with space-time-species cubes and proposed it as a suitable model for the oncoming challenges of big data (Jetz et al., 2019) (European BON <http://biodiversity.eubon.eu/essential-biodiversity-variables>). As well, Randin et al. (2020) provided a detailed review of possible synergies between remote sensing products and biophysical process variables that are key for species distribution

models. Furthermore, Hardisty et al. (2019) developed a framework for Interoperability between research infrastructures related to EBVs. Hence, the successful implementation of the RegESDL for studying ecosystem dynamics in northern South America can be seen as a step forward on the development of EBVs related to function and structure in tropical ecosystems.

Specifically on our case study, we show the variety of seasonal trajectories in northern South America. We computed simple metrics to compare semiannual and annual regimes, derived from vegetation variables of productivity, greenness and standing vegetation available at the RegESDL. Nevertheless, our seasonality analysis approach relies on passive sensors which are limited in rainy regions despite of having long time series. This reinforced the necessity of synergies between multiple sensors and data sources. Using the biotic units map of Colombia, we observed heterogeneity on the seasonality modes and the timing of maximum activity among units that are characterized by Beta diversity values. Our findings contribute to elucidate the large seasonal variability in hotspots of biodiversity which is key for differentiated management of tropical ecosystems. Complementary analysis in light of local climate variability needs to be further investigated to identify the main drivers.

DATA AVAILABILITY STATEMENT

The RegESDL can be accessed using Julia or Python. The dataset is available for free download using Python. Complementary, data can also be accessed locally using Julia. In this case, only data used in the analysis is loaded. Explanatory scripts of how to access the RegESDL, and how to reproduce the main analysis and figures are available in the supplementary and at <http://doi.org/10.5281/zenodo.5068004>. **Supplementary Table S7** lists all the scripts and their location. The most up-to-date code can be found at https://github.com/linamaes/Regional_ESDL.

AUTHOR CONTRIBUTIONS

LE-S performed all analyses. The study was designed by LE-S and MM. LE-S and MM wrote the paper with substantial input from all other authors. FG, UW, and LE-S acquired and pre-processed the data. FG and LE-S worked in the RegESDL cube generation.

FUNDING

This research has been supported by the European Space Agency via the Earth System Data Lab Project <https://www.earthsystemdatalab.net/>. LE-S was funded by the German Academic Exchange Service (DAAD) 57395813 grant. DP-M was funded by the European Union's Horizon 2020 research and innovation program via the TRuStEE project under the Marie Skłodowska-Curie grant agreement no. 721995T. The article processing charges for this open-access publication were covered by the Max Planck Society.

ACKNOWLEDGMENTS

The authors thank the European Space Agency for funding the Earth System Data Lab project. All authors thank PEACE 2016 meeting participants where the project initiative started. The authors thank the reviewers for their insightful comments that contributed to the improvement of this manuscript. LE-S acknowledges the support of DAAD and its Graduate School Scholarship Program, as well as the International Max Planck Research School for Global Biogeochemical Cycles. The

Colombian BON and the Alexander von Humboldt Institute for Biological Research were key institutions for starting the project as part of the subvention USAID AID-OAA-A-11-00012.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feart.2021.613395/full#supplementary-material>.

REFERENCES

- Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arnett, A., Jung, M., et al. (2015). The Dominant Role of Semi-arid Ecosystems in the Trend and Variability of the Land CO₂ Sink. *Science* 348, 895–899. doi:10.1126/science.aaa1668
- Álvarez-Villa, O. D., Vélez, J. I., and Poveda, G. (2011). Improved Long-Term Mean Annual Rainfall fields for Colombia. *Int. J. Climatol.* 31, 2194–2212. doi:10.1002/joc.2232
- Anaya, J. A., Gutiérrez-Vélez, V. H., Pacheco-Pascagaza, A. M., Palomino-Ángel, S., Han, N., and Balzter, H. (2020). Drivers of Forest Loss in a Megadiverse Hotspot on the Pacific Coast of Colombia. *Remote Sensing* 12, 1235. doi:10.3390/rs12081235
- Andrade, M. G. (2011). Estado del conocimiento de la biodiversidad en Colombia y sus amenazas. Consideraciones para fortalecer la interacción ciencia-política. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 35, 491–508.
- Ariza-Porras, C., Bravo, G., Villamizar, M., Moreno, A., Castro, H., Galindo, G., Cabera, E., Valbuena, S., and Lozano, P. (2017). “CDCol: A Geoscience Data Cube that Meets Colombian Needs,” in Colombian Conference on Computing, September 19–22, 2017 Cali, Colombia: Springer, 87–99 doi:10.1007/978-3-319-66562-7_7
- Armenteras, D., González, T. M., Retana, J., and Espelta, J. M. (2016). *Degradación de bosques en Latinoamérica: síntesis conceptual, metodologías de evaluación y casos de estudio nacionales*. IBERO-REDD+.
- Armenteras, D., Negret, P., Melgarejo, L. F., Lakes, T. M., Londoño, M. C., García, J., et al. (2019a). Curb Land Grabbing to Save the Amazon. *Nat. Ecol. Evol.* 3, 1497. doi:10.1038/s41559-019-1020-1
- Armenteras, D., Schneider, L., and Dávalos, L. M. (2019b). Fires in Protected Areas Reveal Unforeseen Costs of Colombian Peace. *Nat. Ecol. Evol.* 3, 20–23. doi:10.1038/s41559-018-0727-8
- Asmryan, S., Muradyan, V., Tepanosyan, G., Hovsepyan, A., Saghatelian, A., Atsatryan, H., et al. (2019). Paving the Way towards an Armenian Data Cube. *Data* 4, 117. doi:10.3390/data4030117
- Asner, G. P., Martin, R. E., Anderson, C. B., and Knapp, D. E. (2015). Quantifying forest Canopy Traits: Imaging Spectroscopy versus Field Survey. *Remote Sensing Environ.* 158, 15–27. doi:10.1016/j.rse.2014.11.011
- Asner, G. P., Mascaro, J., Muller-Landau, H. C., Vieilledent, G., Vaudry, R., Rasamoelina, M., et al. (2012). A Universal Airborne LiDAR Approach for Tropical forest Carbon Mapping. *Oecologia* 168, 1147–1160. doi:10.1007/s00442-011-2165-z
- Baptiste, B., Pinedo-Vasquez, M., Gutierrez-Velez, V. H., Andrade, G. I., Vieira, P., Estupinan-Suarez, L. M., et al. (2017). Greening Peace in Colombia. *Nat. Ecol. Evol.* 1, 1–3. doi:10.1038/s41559-017-0102
- Bathiany, S., Dakos, V., Scheffer, M., and Lenton, T. M. (2018). Climate Models Predict Increasing Temperature Variability in Poor Countries. *Sci. Adv.* 4, eaar5809. doi:10.1126/sciadv.aar5809
- Beck, H. E., van Dijk, A. I. J. M., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., et al. (2016). Global-scale Regionalization of Hydrologic Model Parameters. *Water Resour. Res.* 52, 3599–3622. doi:10.1002/2015wr018247
- Bedoya-Soto, J. M., Aristizábal, E., Carmona, A. M., and Poveda, G. (2019). Seasonal Shift of the Diurnal Cycle of Rainfall over Medellín’s Valley, Central Andes of Colombia (1998–2005). *Front. Earth Sci.* 7, 92. doi:10.3389/feart.2019.00092
- Bianchi, C. A., and Haig, S. M. (2013). Deforestation Trends of Tropical Dry Forests in central Brazil. *Biotropica* 45, 395–400. doi:10.1111/btp.12010
- Bloom, A. A., Bowman, K. W., Lee, M., Turner, A. J., Schroeder, R., Worden, J. R., et al. (2017). A Global Wetland Methane Emissions and Uncertainty Dataset for Atmospheric Chemical Transport Models (WetCHARTs Version 1.0). *Geoscientific Model. Dev.* 10, 2141–2156. doi:10.5194/gmd-10-2141-2017
- Boulton, G. (2018). The Challenges of a Big Data Earth. *Big Earth Data* 2, 1–7. doi:10.1080/20964471.2017.1397411
- Bravo, G., Castro, H., Moreno, A., Ariza-Porras, C., Galindo, G., Cabrera, E., Valbuena, S., and Lozano-Rivera, P. (2017). “Architecture for a Colombian Data Cube Using Satellite Imagery for Environmental Applications,” in Colombian Conference on Computing, September 19–22, 2017 Cali, Colombia: Springer, 227–241. doi:10.1007/978-3-319-66562-7_17
- Builes-Jaramillo, A., Marwan, N., Poveda, G., and Kurths, J. (2018). Nonlinear Interactions between the Amazon River basin and the Tropical North Atlantic at Interannual Timescales. *Clim. Dyn.* 50, 2951–2969. doi:10.1007/s00382-017-3785-8
- Chambers, J. Q., Higuchi, N., Tribuzy, E. S., and Trumbore, S. E. (2001). Carbon Sink for a century. *Nature* 410, 429. doi:10.1038/35068624
- Ciabatta, L., Massari, C., Brocca, L., Gruber, A., Reimer, C., Hahn, S., et al. (2018). SM2RAIN-CCI: a New Global Long-Term Rainfall Data Set Derived from ESA CCI Soil Moisture. *Earth Syst. Sci. Data* 10, 267–280. doi:10.5194/essd-10-267-2018
- Clerici, N., Armenteras, D., Kareiva, P., Botero, R., Ramírez-Delgado, J. P., Forero-Medina, G., et al. (2020). Deforestation in Colombian Protected Areas Increased during post-conflict Periods. *Sci. Rep.* 10, 4971–5010. doi:10.1038/s41598-020-61861-y
- Correa Ayram, C. A., Etter, A., Díaz-Timoté, J., Rodríguez Buriticá, S., Ramírez, W., and Corzo, G. (2020). Spatiotemporal Evaluation of the Human Footprint in Colombia: Four Decades of Anthropogenic Impact in Highly Biodiverse Ecosystems. *Ecol. Indicators* 117, 106630. doi:10.1016/j.ecolind.2020.106630
- Cremer, F., Urbazaev, M., Cortés, J., Truckenbrodt, J., Schullius, C., and Thiel, C. (2020). Potential of Recurrence Metrics from Sentinel-1 Time Series for Deforestation Mapping. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 13, 5233–5240. doi:10.1109/jstars.2020.3019333
- Dávalos, L. M., Bejarano, A. C., Hall, M. A., Correa, H. L., Corthals, A., and Espejo, O. J. (2011). Forests and Drugs: Coca-Driven Deforestation in Tropical Biodiversity Hotspots. *Environ. Sci. Tech.* 45, 1219–1227.
- Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011). The International Soil Moisture Network: a Data Hosting Facility for Global *In Situ* Soil Moisture Measurements. *Hydrol. Earth Syst. Sci.*, 15, 1675, 1698. doi:10.5194/hess-15-1675-2011
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI Soil Moisture for Improved Earth System Understanding: State-Of-The Art and Future Directions. *Remote Sensing Environ.* 203, 185–215. doi:10.1016/j.rse.2017.07.001
- ESA (2017). *Land Cover CCI Product User Guide Version 2*. Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf
- Espinoza, J. C., Garreaud, R., Poveda, G., Arias, P. A., Molina-Carpio, J., Masiokas, M., et al. (2020). Hydroclimate of the Andes Part I: Main Climatic Features. *Front. Earth Sci.* 8, 64. doi:10.3389/feart.2020.00064
- Estupinan-Suarez, L. M., Florez-Ayala, C., Quinones, M. J., Pacheco, A. M., and Santos, A. C. (2015). Detection and Characterization of Colombian Wetlands Using Alos Palsar and MODIS Imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XL-7/W3*, 375–382. doi:10.5194/isprsarchives-xl-7-w3-375-2015

- Etter, A., McAlpine, C., and Possingham, H. (2008). Historical Patterns and Drivers of Landscape Change in Colombia since 1500: a Regionalized Spatial Approach. *Ann. Assoc. Am. Geogr.* 98, 2–23. doi:10.1080/00045600701733911
- Fagua, J. C., Jantz, P., Rodriguez-Buritica, S., Duncanson, L., and Goetz, S. J. (2019). Integrating LiDAR, Multispectral and SAR Data to Estimate and Map Canopy Height in Tropical Forests. *Remote Sensing* 11, 2697. doi:10.3390/rs11222697
- FAO, I., and ISRIC, I. (2012). *JRC: Harmonized World Soil Database (Version 1.2)*. Rome, Italy and IIASALaxenburg, Austria: FAO.
- Ferraz, A., Saatchi, S. S., Longo, M., and Clark, D. B. (2020). Tropical Tree Size-Frequency Distributions from Airborne Lidar. *Ecol. Appl.* 30, e02154. doi:10.1002/eap.2154
- Ferraz, A., Saatchi, S., Mallet, C., Jacquemoud, S., Gonçalves, G., Silva, C., et al. (2016). Airborne Lidar Estimation of Aboveground forest Biomass in the Absence of Field Inventory. *Remote Sensing* 8, 653. doi:10.3390/rs08080653
- Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol.* 37, 4302–4315. doi:10.1002/joc.5086
- Flach, M., Brenning, A., Gans, F., Reichstein, M., Sippel, S., and Mahecha, M. D. (2020). Vegetation Modulates the Impact of Climate Extremes on Gross Primary Production. *Biogeosciences Discuss.* 18, 39–53. doi:10.5194/bg-18-39-2021
- Flórez, C., Estupiñán-Suárez, L., Rojas, S., Aponte, C., Quiñones, M., Acevedo, O., et al. (2016). Identificación espacial de los sistemas de humedales continentales de Colombia. *Biota Colombiana* 17, 44–62. doi:10.21068/c2016s01a03
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The Climate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes. *Scientific Data* 2, 1–21. doi:10.1038/sdata.2015.66
- GDAL/OGR Contributors (2020). *Geospatial Data Abstraction Software Library*. Open Source Geospatial Foundation.
- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G. (2018). Could Machine Learning Break the Convection Parameterization Deadlock?. *Geophys. Res. Lett.* 45, 5742–5751. doi:10.1029/2018gl078202
- Giraldo, J. A., del Valle, J. I., Sierra, C. A., and Melo, O. (2020). Dendrochronological Potential of Trees from America's Rainiest Region BT - Latin American Dendroecology: Combining Tree-Ring Sciences and Ecology in a Megadiverse Territory. In *Latin American Dendroecology*, eds. M. Pompa-García and J. J. Camarero (Cham: Springer International Publishing), Chapter 5, 79–119. doi:10.1007/978-3-030-36930-9_5
- Giuliani, G., Camara, G., Killough, B., and Minchin, S. (2019). Earth Observation Open Science: Enhancing Reproducible Science Using Data Cubes. *Data* 4, 147. doi:10.3390/data4040147
- Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J.-P., Allenbach, K., et al. (2017). Building an Earth Observations Data Cube: Lessons Learned from the Swiss Data Cube (SDC) on Generating Analysis Ready Data (ARD). *Big Earth Data* 1, 100–117. doi:10.1080/20964471.2017.1398903
- Giuliani, G., Egger, E., Italiano, J., Poussin, C., Richard, J.-P., and Chatenoux, B. (2020). Essential Variables for Environmental Monitoring: What Are the Possible Contributions of Earth Observation Data Cubes?. *Data* 5, 100. doi:10.3390/data5040100
- Gomes, V., Queiroz, G., and Ferreira, K. (2020). An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sensing* 12, 1253. doi:10.3390/rs12081253
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing Environ.* 202, 18–27. doi:10.1016/j.rse.2017.06.031
- Green, J. K., Konings, A. G., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., et al. (2017). Regionally strong Feedbacks between the Atmosphere and Terrestrial Biosphere. *Nat. Geosci.* 10, 410–414. doi:10.1038/ngeo2957
- Guo, H. (2017). Big Earth Data: A New Frontier in Earth and Information Sciences. *Big Earth Data* 1, 4–20. doi:10.1080/20964471.2017.1403062
- Hamunyela, E., Verbesselt, J., De Bruin, S., and Herold, M. (2016). Monitoring Deforestation at Sub-annual Scales as Extreme Events in Landsat Data Cubes. *Remote Sensing* 8, 651. doi:10.3390/rs08080651
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution Global Maps of 21st-century forest Cover Change. *Science* 342, 850–853. doi:10.1126/science.1244693
- Hardisty, A. R., Michener, W. K., Agosti, D., Alonso García, E., Bastin, L., Belbin, L., et al. (2019). The Bari Manifesto: An Interoperability Framework for Essential Biodiversity Variables. *Ecol. Inform.* 49, 22–31. doi:10.1016/j.ecoinf.2018.11.003
- Hastenrath, S. (1991). *Climate Dynamics of the Tropics*. (Updated ed., *Atmospheric sciences library*; v.8). Dordrecht, Boston. Kluwer Academic.
- Hawinkel, P., Swinnen, E., Lhermitte, S., Verbist, B., Van Orshoven, J., and Muys, B. (2015). A Time Series Processing Tool to Extract Climate-Driven Interannual Vegetation Dynamics Using Ensemble Empirical Mode Decomposition (EEMD). *Remote Sensing Environ.* 169, 375–389. doi:10.1016/j.rse.2015.08.024
- He, N., Liu, C., Piao, S., Sack, L., Xu, L., Luo, Y., et al. (2019). Ecosystem Traits Linking Functional Traits to Macroecology. *Trends Ecol. Evol.* 34, 200–210. doi:10.1016/j.tree.2018.11.004
- Heckel, K., Urban, M., Schratz, P., Mahecha, M., and Schullius, C. (2020). Predicting Forest Cover in Distinct Ecosystems: The Potential of Multi-Source Sentinel-1 and -2 Data Fusion. *Remote Sensing* 12, 302. doi:10.3390/rs12020302
- Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., et al. (2014). SoilGrids1km—global Soil Information Based on Automated Mapping. *PLoS one* 9, e105992. doi:10.1371/journal.pone.0105992
- Hubau, W., Lewis, S. L., Phillips, O. L., Affum-Baffoe, K., Beekman, H., Cuni-Sanchez, A., et al. (2020). Asynchronous Carbon Sink Saturation in African and Amazonian Tropical Forests. *Nature* 579, 80–87. doi:10.1038/s41586-020-2035-0
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sensing Environ.* 83, 195–213. doi:10.1016/s0034-4257(02)00096-2
- Huete, A., Liu, H. Q., Batchily, K. V., and Van Leeuwen, W. (1997). A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sensing Environ.* 59, 440–451. doi:10.1016/s0034-4257(96)00112-5
- IGAC (2010). *Cartografía vectorial a escala 1:100.000 con cobertura total de la República de Colombia. C. Referido al sistema de coordenadas MAGNA-SIRGAS*. Bogotá DC: Instituto Geográfico Agustín Codazzi.
- Jeronimo, S. M. A., Kane, V. R., Churchill, D. J., McGaughey, R. J., and Franklin, J. F. (2018). Applying LiDAR Individual Tree Detection to Management of Structurally Diverse forest Landscapes. *J. For.* 116, 336–346. doi:10.1093/jofore/fyy023
- Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., et al. (2019). Essential Biodiversity Variables for Mapping and Monitoring Species Populations. *Nat. Ecol. Evol.* 3, 539–551. doi:10.1038/s41559-019-0826-1
- Jiang, C., and Ryu, Y. (2016). Multi-scale Evaluation of Global Gross Primary Productivity and Evapotranspiration Products Derived from Breathing Earth System Simulator (BESS). *Remote Sensing Environ.* 186, 528–547. doi:10.1016/j.rse.2016.08.030
- Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., et al. (2016). A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing* 8, 70. doi:10.3390/rs8010070
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global Patterns of Land–Atmosphere Fluxes of Carbon Dioxide, Latent Heat, and Sensible Heat Derived from Eddy Covariance, Satellite, and Meteorological Observations. *J. Geophys. Res. Biogeosciences* 116. doi:10.1029/2010jg001566
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling Carbon Fluxes from Eddy Covariance Sites to globe: Synthesis and Evaluation of the FLUXCOM Approach. *Biogeosciences* 17, 1343–1365. doi:10.5194/bg-17-1343-2020
- Killough, B. (2019). “The Impact of Analysis Ready Data in the Africa Regional Data Cube,” in IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 Aug. 2019 (IEEE), 5646–5649. doi:10.1109/IGARSS.2019.8898321
- Kissling, W. D., Hardisty, A., García, E. A., Santamaria, M., De Leo, F., Pesole, G., et al. (2015). Towards Global Interoperability for Supporting Biodiversity Research on Essential Biodiversity Variables (EBVs). *Biodiversity* 16, 99–107. doi:10.1080/14888386.2015.1068709
- Knoben, W. J. M., Woods, R. A., and Freer, J. E. (2019). Global Bimodal Precipitation Seasonality: A Systematic Overview. *Int. J. Climatol.* 39, 558–567. doi:10.1002/joc.5786

- Konings, A. G., and Gentine, P. (2017). Global Variations in Ecosystem-scale Isohydrycity. *Glob. Change Biol.* 23, 891–905. doi:10.1111/gcb.13389
- Kraemer, G., Camps-Valls, G., Reichstein, M., and Mahecha, M. D. (2020). Summarizing the State of the Terrestrial Biosphere in Few Dimensions. *Biogeosciences* 17, 2397–2424. doi:10.5194/bg-17-2397-2020
- Leblois, A., Damette, O., and Wolfersberger, J. (2017). What Has Driven Deforestation in Developing Countries since the 2000s? Evidence from New Remote-Sensing Data. *World Dev.* 92, 82–102. doi:10.1016/j.worlddev.2016.11.012
- Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., et al. (2017). The Australian Geoscience Data Cube - Foundations and Lessons Learned. *Remote Sensing Environ.* 202, 276–292. doi:10.1016/j.rse.2017.03.015
- Li, W., Guo, Q., Jakubowski, M. K., and Kelly, M. (2012). A New Method for Segmenting Individual Trees from the Lidar point Cloud. *Photogramm. Eng. Remote Sensing* 78, 75–84. doi:10.14358/pers.78.1.75
- Linscheid, N., Estupinan-Suarez, L. M., Brenning, A., Carvalhais, N., Cremer, F., Gans, F., et al. (2020). Towards a Global Understanding of Vegetation-Climate Dynamics at Multiple Timescales. *Biogeosciences* 17, 945–962. doi:10.5194/bg-17-945-2020
- Londoño, M. C., Bello, C., Velásquez, J., Norden, N., Ortiz, C., González, I., et al. (2017). “Documento Técnico: Componente Biótico Mapa de Ecosistemas Continentales, Marinos y Costeros de Colombia.”. *Escala 1:100.000*. Tech. rep. (Bogotá D.C.: Instituto de Investigación de Recursos Biológicos Alexander von Humboldt).
- Ma, X., Migliavacca, M., Wirth, C., Bohn, F. J., Huth, A., Richter, R., et al. (2020). Monitoring Plant Functional Diversity Using the Reflectance and Echo from Space. *Remote Sensing* 12, 1248. doi:10.3390/rs12081248
- MADR-UPRA (2017). “Identificación general de la frontera agrícola en Colombia. Ministerio de Agricultura y Desarrollo Rural Agropecuario - Unidad de Planificación Rural Agropecuaria.”. Tech. rep. (Bogotá D.C.: Ministerio de Agricultura y Desarrollo Rural Agropecuario - Unidad de Planificación Rural).
- Mahecha, M. D., Gans, F., Brandt, G., Christiansen, R., Cornell, S. E., Fomferra, N., et al. (2020). Earth System Data Cubes Unravel Global Multivariate Dynamics. *Earth Syst. Dynam.* 11, 201–234. doi:10.5194/esd-11-201-2020
- Mahecha, M. D., Gans, F., Sippel, S., Donges, J. F., Kaminski, T., Metzger, S., et al. (2017). Detecting Impacts of Extreme Events with Ecological *In Situ* Monitoring Networks. *Biogeosciences* 14, 4255–4277. doi:10.5194/bg-14-4255-2017
- McNicol, I. M., Ryan, C. M., and Mitchard, E. T. A. (2018). Carbon Losses from Deforestation and Widespread Degradation Offset by Extensive Growth in African Woodlands. *Nat. Commun.* 9, 3045. doi:10.1038/s41467-018-05386-z
- Merrick, T., Pau, S., Jorge, M. L. S. P., Bennartz, T. S. F., and Silva, R. (2019). Spatiotemporal Patterns and Phenology of Tropical Vegetation Solar-Induced Chlorophyll Fluorescence across Brazilian Biomes Using Satellite Observations. *Remote Sensing* 11, 1746. doi:10.3390/rs11151746
- Meyer, V., Saatchi, S., Ferraz, A., Xu, L., Duque, A., García, M., et al. (2019). Forest Degradation and Biomass Loss along the Chocó Region of Colombia. *Carbon Balance Manage* 14, 2. doi:10.1186/s13021-019-0117-9
- Miles, L., Newton, A. C., DeFries, R. S., Ravilious, C., May, I., Blyth, S., et al. (2006). A Global Overview of the Conservation Status of Tropical Dry Forests. *J. Biogeogr.* 33, 491–505. doi:10.1111/j.1365-2699.2005.01424.x
- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J. (2011). Global Land-Surface Evaporation Estimated from Satellite-Based Observations. *Hydrol. Earth Syst. Sci.* 15, 453–469. doi:10.5194/hess-15-453-2011
- Musavi, T., Mahecha, M. D., Migliavacca, M., Reichstein, M., van de Weg, M. J., van Bodegom, P. M., et al. (2015). The Imprint of Plants on Ecosystem Functioning: A Data-Driven Approach. *Int. J. Appl. Earth Observation Geoinformation* 43, 119–131. doi:10.1016/j.jag.2015.05.009
- Musavi, T., Migliavacca, M., Reichstein, M., Kattge, J., Wirth, C., Black, T. A., et al. (2017). Stand Age and Species Richness Dampen Interannual Variation of Ecosystem-Level Photosynthetic Capacity. *Nat. Ecol. Evol.* 1, 48. doi:10.1038/s41559-016-0048
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity Hotspots for Conservation Priorities. *Nature* 403, 853–858. doi:10.1038/35002501
- Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., et al. (2003). Climate-driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999. *Science* 300, 1560–1563. doi:10.1126/science.1082750
- Noss, R. F. (1990). Indicators for Monitoring Biodiversity: a Hierarchical Approach. *Conservation Biol.* 4, 355–364. doi:10.1111/j.1523-1739.1990.tb00309.x
- ODC (2021). *Open Data Cube. Documentation on Data Loading*. Limitations and Problems.
- Olaya-Rodríguez, M. H., Velásquez-Tibatá, J., and Estupiñán-Suárez, L. M. (2018). Integrando la información de sensores remotos con modelos de distribución de especies para el monitoreo de la biodiversidad. Caso de estudio para las especies *Zamia amazonum* y *Zamia chigua*. *Biodiversidad en la Práctica* 3, 74–98.
- Orme, C. D. L., Davies, R. G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V. A., et al. (2005). Global Hotspots of Species Richness Are Not Congruent with Endemism or Threat. *Nature* 436, 1016–1019. doi:10.1038/nature03850
- OGC (2020). Considering Zarr as Community Standard; seeks Public Comment on New Work Item. The Open Geospatial Consortium. <https://www.ogc.org/presroom/pressreleases/3275>. (Accessed August 21, 2020).
- Pacheco-Pascagaza, A. M., García, M., Rodríguez-Veiga, P., and Baltzer, H. (2018). “The Use of Multifrequency SAR Data for Assessing Levels of Forest Disturbance in Bajo Calima Colombia,” in IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018 (IEEE), 7015–7018. doi:10.1109/IGARSS.2018.8518871
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., et al. (2011). A Large and Persistent Carbon Sink in the World’s Forests. *Science* 333, 988–993. doi:10.1126/science.1201609
- Papagiannopoulou, C., Miralles, D. G., Decubber, S., Demuzere, M., Verhoest, N. E. C., Dorigo, W. A., et al. (2017). A Non-linear Granger-causality Framework to Investigate Climate-Vegetation Dynamics. *Geosci. Model. Dev.* 10, 1945–1960. doi:10.5194/gmd-10-1945-2017
- Patino, J. E., and Estupinan-Suarez, L. M. (2016). Hotspots of Wetland Area Loss in Colombia. *Wetlands* 36, 935–943. doi:10.1007/s13157-016-0806-z
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., et al. (2013). Essential Biodiversity Variables. *Science* 339, 277–278. doi:10.1126/science.1229931
- Pettorelli, N., Wegmann, M., Skidmore, A., Múcher, S., Dawson, T. P., Fernandez, M., et al. (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, O. L., and Lewis, S. L. (2014). Evaluating the Tropical forest Carbon Sink. *Glob. Change Biol.* 20, 2039–2041. doi:10.1111/gcb.12423
- PNN (2015). *Áreas protegidas registradas en el registro único de áreas protegidas. Registro Único Nacional de Áreas Protegidas RUNAP. Decreto único 1076 del 2015. Shapefile*. Bogotá: Parques Nacionales Naturales de Colombia. Available at <http://mapas.parquesnacionales.gov.co/services>.
- Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van der Tol, C., Flexas, J., Pfündel, E. E., et al. (2014). Linking Chlorophyll a Fluorescence to Photosynthesis for Remote Sensing Applications: Mechanisms and Challenges. *J. Exp. Bot.* 65, 4065–4095. doi:10.1093/jxb/eru191
- Poveda, G., Álvarez, D. M., and Rueda, Ó. A. (2011). Hydro-climatic Variability over the Andes of Colombia Associated with ENSO: a Review of Climatic Processes and Their Impact on One of the Earth’s Most Important Biodiversity Hotspots. *Clim. Dyn.* 36, 2233–2249. doi:10.1007/s00382-010-0931-y
- Poveda, G., and Mesa, O. J. (2000). On the Existence of Lloró (The Rainiest Locality on Earth): Enhanced Ocean-Land-Atmosphere Interaction by a Low-Level Jet. *Geophys. Res. Lett.* 27, 1675–1678. doi:10.1029/1999gl006091
- Poveda, G., Mesa, O. J., Salazar, L. F., Arias, P. A., Moreno, H. A., Vieira, S. C., et al. (2005). The Diurnal Cycle of Precipitation in the Tropical Andes of Colombia. *Monthly Weather Rev.* 133, 228–240. doi:10.1175/mwr-2853.1
- Poveda, G., Waylen, P. R., and Pulwarty, R. S. (2006). Annual and Inter-annual Variability of the Present Climate in Northern South America and Southern Mesoamerica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 234, 3–27. doi:10.1016/j.palaeo.2005.10.031
- Radeloff, V. C., Dubinin, M., Coops, N. C., Allen, A. M., Brooks, T. M., Clayton, M. K., et al. (2019). The Dynamic Habitat Indices (Dhis) from Modis and Global Biodiversity. *Remote Sensing Environ.* 222, 204–214. doi:10.1016/j.rse.2018.12.009

- Ramírez-Delgado, J., Galindo, G., Yepes, A., and Cabrera, E. (2018). *Estimación de la degradación de bosques de Colombia a través de un análisis de fragmentación*. Bogotá D.C.: Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM, Ministerio de Ambiente y Desarrollo Sostenible – MADS, Programa ONU-REDD Colombia.
- Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., et al. (2020). Monitoring Biodiversity in the Anthropocene Using Remote Sensing in Species Distribution Models. *Remote Sensing Environ.* 239, 111626. doi:10.1016/j.rse.2019.111626
- 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, 13697–13702. doi:10.1073/pnas.1216065111
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al. (2019). Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* 566, 195–204. doi:10.1038/s41586-019-0912-1
- Ruiz, M., J. F., and Melo, J. Y. (2020). *Modelación Numérica de Tiempo y Clima IDEAM*.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H. (2004). A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *Bioscience* 54, 547–560. doi:10.1641/0006-3568(2004)054[0547:acsmog]2.0.co;2
- Ryu, Y., Baldocchi, D. D., Kobayashi, H., Van Ingen, C., Li, J., Black, T. A., et al. (2011). Integration of MODIS Land and Atmosphere Products with a Coupled-Process Model to Estimate Gross Primary Productivity and Evapotranspiration from 1 Km to Global Scales. *Glob. Biogeochem. Cycles* 25, a–n. doi:10.1029/2011GB004053
- Ryu, Y., Berry, J. A., and Baldocchi, D. D. (2019). What Is Global Photosynthesis? History, Uncertainties and Opportunities. *Remote Sensing Environ.* 223, 95–114. doi:10.1016/j.rse.2019.01.016
- Ryu, Y., Jiang, C., Kobayashi, H., and Detto, M. (2018). MODIS-derived Global Land Products of Shortwave Radiation and Diffuse and Total Photosynthetically Active Radiation at 5 Km Resolution from 2000. *Remote Sensing Environ.* 204, 812–825. doi:10.1016/j.rse.2017.09.021
- Salazar, A., Sanchez, A., Villegas, J. C., Salazar, J. F., Ruiz Carrascal, D., Sitch, S., et al. (2018). The Ecology of Peace: Preparing Colombia for New Political and Planetary Climates. *Front. Ecol. Environ.* 16, 525–531. doi:10.1002/fee.1950
- Scholze, M., Buchwitz, M., Dorigo, W., Guanter, L., and Quegan, S. (2017). Reviews and Syntheses: Systematic Earth Observations for Use in Terrestrial Carbon Cycle Data Assimilation Systems. *Biogeosciences* 14, 3401–3429. doi:10.5194/bg-14-3401-2017
- Schwartz, M. D. (1998). Green-wave Phenology. *Nature* 394, 839–840. doi:10.1038/29670
- Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., et al. (2018). HESS Opinions: Incubating Deep-Learning-Powered Hydrologic Science Advances as a Community. *Hydrol. Earth Syst. Sci. (Online)* 22, 5639–5656. doi:10.5194/hess-22-5639-2018
- Sierra, C. A., Mahecha, M., Poveda, G., Álvarez-Dávila, E., Gutierrez-Velez, V. H., Reu, B., et al. (2017). Monitoring Ecological Change during Rapid Socio-Economic and Political Transitions: Colombian Ecosystems in the post-conflict Era. *Environ. Sci. Pol.* 76, 40–49. doi:10.1016/j.envsci.2017.06.011
- Simard, M., Pinto, N., Fisher, J. B., and Baccini, A. (2011). Mapping forest Canopy Height Globally with Spaceborne Lidar. *J. Geophys. Res. Biogeosciences* 116. doi:10.1029/2011jg001708
- Skidmore, A. K., Pettorelli, N., Coops, N. C., Geller, G. N., Hansen, M., Lucas, R., et al. (2015). Environmental Science: Agree on Biodiversity Metrics to Track from Space. *Nature* 523, 403–405. doi:10.1038/523403a
- Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., et al. (2018). Global Land Change from 1982 to 2016. *Nature* 560, 639–643. doi:10.1038/s41586-018-0411-9
- Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., et al. (2017). OCO-2 Advances Photosynthesis Observation from Space via Solar-Induced Chlorophyll Fluorescence. *Science* 358, eaam5747. doi:10.1126/science.aam5747
- Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., and Brisco, B. (2020). Google Earth Engine for Geo-Big Data Applications: A Meta-Analysis and Systematic Review. *ISPRS J. Photogrammetry Remote Sensing* 164, 152–170. doi:10.1016/j.isprsjprs.2020.04.001
- Tucker, C. J., and Sellers, P. J. (1986). Satellite Remote Sensing of Primary Production. *Int. J. Remote Sensing* 7, 1395–1416. doi:10.1080/01431168608948944
- Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C. (2020). A Double Peak in the Seasonality of California's Photosynthesis as Observed from Space. *Biogeosciences* 17, 405–422. doi:10.5194/bg-17-405-2020
- Vallejo-Bernal, S. M., Urrea, V., Bedoya-Soto, J. M., Posada, D., Olarte, A., Cárdenas-Posso, Y., et al. (2020). Ground Validation of TRMM 3B43 V7 Precipitation Estimates over Colombia. Part I: Monthly and Seasonal Timescales. *Int. J. Climatology* 41, 1–24.
- van Schaik, E., Killaars, L., Smith, N. E., Koren, G., van Beek, L. P. H., Peters, W., et al. (2018). Changes in Surface Hydrology, Soil Moisture and Gross Primary Production in the Amazon during the 2015/2016 El Niño. *Phil. Trans. R. Soc. B* 373, 20180084. doi:10.1098/rstb.2018.0084
- Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., et al. (2013). "Apache Hadoop Yarn: Yet Another Resource Negotiator," in Proceedings of the 4th annual Symposium on Cloud Computing, Santa Clara, California, New York, NY: Association for Computing Machinery, 1–16.
- Velásquez-Tibatá, J., Olaya-Rodríguez, M. H., López-Lozano, D., Gutiérrez, C., González, I., and Londoño-Murcia, M. C. (2019). BioModelos: A Collaborative Online System to Map Species Distributions. *PLoS One* 14, e0214522. doi:10.1371/journal.pone.0214522
- Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Köhler, P., et al. (2016). Satellite Chlorophyll Fluorescence Measurements Reveal Large-Scale Decoupling of Photosynthesis and Greenness Dynamics in Boreal evergreen Forests. *Glob. Change Biol.* 22, 2979–2996. doi:10.1111/gcb.13200
- Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* 40, 1–29. doi:10.18637/jss.v040.i01
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific data* 3, 1–9. doi:10.1038/sdata.2016.18
- Wilson, A. M., and Jetz, W. (2016). Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *Plos Biol.* 14, e1002415. doi:10.1371/journal.pbio.1002415
- Wu, J., Albert, L. P., Lopes, A. P., Restrepo-Coupe, N., Hayek, M., Wiedemann, K. T., et al. (2016). Leaf Development and Demography Explain Photosynthetic Seasonality in Amazon evergreen Forests. *Science* 351, 972–976. doi:10.1126/science.aad5068
- Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., et al. (2016). The Global Landsat Archive: Status, Consolidation, and Direction. *Remote Sensing Environ.* 185, 271–283. doi:10.1016/j.rse.2015.11.032
- Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., et al. (2020). Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges. *Remote Sensing* 12, 62.
- Yepes, J., Poveda, G., Mejía, J. F., Moreno, L., and Rueda, C. (2019). CHOCO-JEX: A Research Experiment Focused on the Chocó Low-Level Jet over the Far Eastern Pacific and Western Colombia. *Bull. Am. Meteorol. Soc.* 100, 779–796. doi:10.1175/bams-d-18-0045.1
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., et al. (2016). Apache Spark. *Commun. ACM* 59, 56–65. doi:10.1145/2934664
- Zemp, D. C., Schlessner, C.-F., Barbosa, H. M. J., Hirota, M., Montade, V., Sampaio, G., et al. (2017). Self-amplified Amazon forest Loss Due to Vegetation-Atmosphere Feedbacks. *Nat. Commun.* 8, 1–10. doi:10.1038/ncomms14681
- Zemp, D. C., Schlessner, C.-F., Barbosa, H. M. J., Van der Ent, R. J., Donges, J. F., Heinke, J., et al. (2014). On the Importance of Cascading Moisture Recycling in South America. *Atmos. Chem. Phys.* 14, 13337–13359. doi:10.5194/acp-14-13337-2014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Estupinan-Suarez, Gans, Brenning, Gutierrez-Velez, Londono, Pabon-Moreno, Poveda, Reichstein, Reu, Sierra, Weber and Mahecha. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Chapter 5

Spatial patterns of vegetation activity related to ENSO in northern South America

Third manuscript

Authors: Lina M. Estupinan-Suarez, Miguel D. Mahecha, Alexander Brenning, Guido Kraemer, Germán Poveda, Markus Reichstein, and Carlos A. Sierra.

Status: Submitted

Spatial patterns of vegetation activity related to ENSO in northern South America

Abstract

Interannual variability of vegetation activity (i.e. photosynthesis) is regionally correlated with El Niño Southern Oscillation (ENSO). Globally, a reduction in carbon uptake by terrestrial ecosystems has been observed during El Niño (warm phase) and the opposite during La Niña (cold phase). However, this global perspective obscures the heterogeneous impacts of ENSO on vegetation activity at regional scales. Particularly, ENSO has contrasting impacts on climate in northern South America (NSA) depending on the ENSO phase and geographical location. Furthermore, changes in vegetation activity related to multiple ENSO events in NSA are not well understood yet. In this study, we investigated time series of vegetation variables (i.e. GPP, NDVI, EVI, FPAR) from 2001 to 2014 at moderate spatial resolution (0.0083°). Data were aggregated through dimensionality reduction analysis, specifically we used the Global Principal Component Analysis. The leading principal component served as proxy of vegetation activity (VAC). VAC was correlated to the multivariate ENSO index separately for each ENSO phase. Our aim was to understand differences between El Niño and La Niña in NSA, a region dominated by rainforest and savannas, and to identify hotspots of ENSO impacts. Our results show that El Niño phase has a stronger impact on vegetation activity both in intensity and duration than La Niña phase. Moreover, seasonally dry ecoregions were more susceptible to El Niño impacts on vegetation activity. Understanding these differences is key for regional adaptation and differentiated management of ecosystems.

Keywords: ENSO, tropical ecosystems, vegetation activity, dimensionality reduction analysis, lagged effects, hotspots, northern South America

5.1 Introduction

The response of vegetation activity – photosynthesis in particular – to changes in climate is strongly controlled by both anthropogenically induced climate change and natural drivers induced by internal variability of the climate system (Arias et al., 2021b). In northern South America (NSA), the El Niño Southern Oscillation (ENSO) is considered the main driver of interannual variability of cloud cover, rainfall and temperature (Cai et al., 2020; Cess et al., 2001; Eleftheratos et al., 2011; Hilker et al., 2014; Philander, 2018; Poveda, Waylen, and Pulwarty, 2006; Tedeschi, Grimm, and Cavalcanti, 2016),

with important impacts on the productivity of vegetation within this region (Bastos et al., 2018; Köhler et al., 2018; Liu et al., 2017; Luo et al., 2018; Salas et al., 2020; Schaik et al., 2018). ENSO is a coupled oceanic-atmospheric event in the Pacific ocean and it is expected to increase in frequency and intensity in the wake of climate change (Cai et al., 2014; Fasullo, Otto-Bliesner, and Stevenson, 2018; Yun et al., 2021), which can lead to excesses or deficits of rainfall for different parts of NSA depending on the season. Therefore, a regional understanding of climate variability, in particular of ENSO and its impacts on vegetation, is of paramount importance to predict the impacts of global change on natural ecosystems, and, in consequence, its further implications to human populations that inhabit this region.

Hydro-climatological studies have shown important changes in precipitation and temperature associated with ENSO phases (i.e., El Niño, La Niña and Neutral) both at regional and subregional scales (Bolaños et al., 2021; Cai et al., 2020; Poveda, Waylen, and Pulwarty, 2006; Poveda et al., 2001; Waylen and Poveda, 2002). These studies have shown that each ENSO phase does not manifest homogeneously across the entire NSA domain, but rather at different intensities and with contrasting patterns of precipitation. The complex topography imposed by the Andean mountains as well as the diversity of ecosystem types characteristic of the tropical rain and dry forests, savannas and wetlands of NSA pose big challenges to understanding the impacts of different ENSO phases on climate and vegetation activity. In particular, there is limited knowledge on how variability among ENSO phases affects vegetation activity taking into account the diverse topography and ecosystems in NSA.

In general, the impacts of ENSO on vegetation are studied for individual events (e.g., El Niño 2015-2016) from global to local scales. Local studies have used ground data and permanent plots to assess plant functional traits, and changes in floristic biodiversity in regions affected by ENSO (González-M et al., 2021; Muenchow et al., 2013, 2020). Regional and global analyses have assessed changes on vegetation productivity and photosynthesis (Bastos et al., 2018; Liu et al., 2017; Luo et al., 2018; Patra et al., 2017). Specifically, for tropical South America, research has focused extensively on productivity in the Amazon biome. For example, Schaik et al. (2018) compared GPP anomalies among the Amazon sub-basins for El Niño 2015–2016 and found that the northeastern sub-basin hosted the largest GPP loss (56% from October to March). In contrast to analyses based on El Niño events, analyses related to La Niña are scarce and usually carried out at global scales. For instance, Bastos et al. (2013) found that La Niña of 2011 explained more than 40% of the variance in global net primary productivity. For the same La Niña event, Pandey et al. (2017) reported an increase in methane emissions by wetlands (5%). Analyses of individual events contribute to explain yearly changes in global vegetation activity but they are limited in: (i) detecting

systematic/concurrent hotspots of ENSO across time; and, (ii) identifying differences in ecosystem responses due to local conditions.

An alternative to evaluating individual ENSO events is to implement time series analyses of climatological and ecosystem-productivity variables that cover multiple ENSO events. This approach can be very powerful in assessing and comparing changes over time that could be associated with lagged effects during transitions between ENSO phases. Currently, multiple proxies of vegetation activity are available (e.g., greenness, productivity). Some of these variables are either calculated from direct satellite retrievals such as vegetation indices (Huete et al., 1997; Huete et al., 2002; Tucker and Sellers, 1986) or data products obtained from model-data fusion methods or data-driven models (Jiang and Ryu, 2016; Running, Mu, and Zhao, 2015). Nevertheless, tropical regions such as NSA usually face challenges related to data availability and quality due to cloud cover and lack of ground-truthing (Estupinan-Suarez, Leon, and Sarmiento Pinzon, 2017; Hilker et al., 2012).

Another important challenge in analyzing vegetation activity metrics is that multiple data products are available that may provide contrasting results concerning the relationships between climatic and ecosystem variables. It is often difficult to determine whether one single vegetation variable is preferable over others, but data analysis methods such as dimensionality reduction contributes to capturing a clearer vegetation signal and to reducing noise (Estupinan-Suarez et al., 2021; Kraemer et al., 2020). In particular, these methods can combine the information from multiple data streams into a small set of indicator variables retaining most of the shared information in the data, which offers an opportunity to aggregate different metrics of vegetation activity without a major loss of information.

In this study, we investigated the spatial and temporal differences in vegetation activity associated with ENSO phases (i.e., El Niño, La Niña) in NSA from 2001 to 2014. In particular, we address three main research questions; How well is vegetation activity captured by a single indicator variable obtained from a dimensionality reduction analysis? Are there differences in time lags of vegetation activity between ENSO phases? And, Where are the hotspots in which vegetation activity responds more strongly to inter ENSO variability in NSA?

5.2 Methods

5.2.1 Study area

This study was carried out in NSA, in a domain that includes the countries of Panama, Colombia, Venezuela, Ecuador, and partially Peru and Brazil. This region hosts one

TABLE 5.1: Climatic impacts of El Niño and La Niña in northern South America. *: Anomalies not reported. Source: NOAA (2021a).

Region	El Niño		La Niña	
	Austral summer	Austral winter	Austral summer	Austral winter
Pacific coast	Warm and wet	Warm	Cool and dry	Cool
Caribbean coast	*	Warm and dry	*	Cool and wet
Amazon	Dry	*	Wet	*

of the largest biodiversity levels on Earth with a high degree of species endemism (Andrade, 2011; Myers et al., 2000). In addition, the region is home to more than 100 million inhabitants (United Nations, 2019) that rely on a number of ecosystem services, but are also a major driver of deforestation and land use change. The most extensive natural land covers are broadleaved evergreen forests (71.8%), grasslands (7.1%), and shrublands (4.0%) (ESA, 2017).

In this region, the impacts of ENSO on weather are contrasting depending on the season (i.e., austral summer and winter), ENSO phase, and geographical location (NOAA, 2021a). The main climatic impacts are summarized on Table 5.1.

5.2.2 Vegetation variables

As a proxy of vegetation activity, we selected two vegetation indices and two data sets derived from models. These variables are the normalized difference vegetation index (NDVI) (Tucker and Sellers, 1986), the enhanced vegetation index (EVI) (Huete et al., 1997; Huete et al., 2002) and the fraction of absorbed photosynthetically active radiation (FPAR) (Knyazikhin et al., 1998; Sellers et al., 1997) from the Moderate Resolution Imaging Spectroradiometer (MODIS) Terra. Gross primary productivity (GPP) was obtained from the Breathing Earth System Simulator (Ryu et al., 2011). Data sets were acquired from the Regional Earth System Data Lab (RegESDL) where they are available in a harmonized grid at 0.0083° spatial resolution and 8-daily temporal resolution (Estupinan-Suarez et al., 2021). Their geographical extend is from latitude 14°N to 14°S and longitude 83°W to 60°W. Quality flag filters were applied previously to the MODIS products by the RegESDL. Data spans from 2001 to 2014, a period when all four variables overlapped.

5.2.3 Global Principal Component Analysis

Each variable was normalized pixel-wise to a global mean of zero and a variance of one. To extract the main features of vegetation activity, we used dimensionality reduction analysis. Specifically, we used an online version of the principal component analysis (PCA) developed to deal with very large data sets and spherical distortions (Kraemer et al., 2020), and refer to it as Global PCA. Within this framework, the entire data set was projected into a unified principal component space regardless of the time and space features. The Global PCA steps are: (i) to create covariance matrices for each time step; (ii) to combine the covariance matrices of each time step into a global covariance matrix; (iii) to calculate the loadings of the PCA from the global covariance matrix; and (iv) to project data pixel by pixel to the new principal component (PC) space.

In addition, the first PC (PC_1) trajectory was cross-checked with the one from the input variables. This is because the PCA orthogonal rotation is arbitrary and not sensitive to the variables directions. In our case, the PC_1 trajectory was opposite to the one from the input variables, then PC_1 was flipped with a simple multiplication by -1 . Moreover, the Global PCA handles no data values, and also it assigns a weighted factor to each covariance matrix per time step. Therefore data was not gap-filled. To account for pixel size variations and accurately estimate the final covariance matrix we used the `WeightedOnlineStats.jl` package (doi: 10.5281/zenodo.6494412) that efficiently handles large data sets.

Furthermore, we conducted a separate Global PCA for each of the dominant natural land cover classes, and referred to it as the land cover PCA for simplicity (see Appendix D). With this approach, we compared the fraction of variance captured for PCs when doing a separate analysis for broadleaved evergreen forests, shrublands and grasslands.

Subsequent analyses were performed with the PC_1 that captures the largest fraction of variation, and holds the main information of vegetation activity (see section 5.3.1). Henceforth, we refer to PC_1 as the leading vegetation activity component (VAC).

5.2.4 Time-lagged correlations between ENSO phases and the vegetation activity component

We calculated the Spearman correlation (ρ) between VAC and the bi-monthly Multivariate ENSO index v2.0 (MEI) on a pixel-by-pixel level. MEI is calculated based on five atmospheric and oceanic variables (i.e., sea level pressure, sea surface, temperature, surface zonal winds, surface meridional winds, and outgoing longwave radiation) between 30° S – 30° N and 100° E – 70° W over the tropical Pacific, and it takes into account seasonal and sub-seasonal variability (NOAA, 2021b). Time series of monthly

MEI are available from the National Oceanic and Atmospheric Administration (NOAA) (<https://psl.noaa.gov/enso/mei/>, last access Sep. 1st, 2021). To match the variable's temporal resolution with the one of MEI, the index was interpolated to 8-daily time steps using B-splines. Based on MEI thresholds by NOAA, the index values were classified as follows: ≤ -0.5 correspond to La Niña whereas values ≥ 0.5 to El Niño, values from -0.5 to 0.5 are Neutral (NOAA, 2021b).

The Spearman correlation was computed separately for El Niño, La Niña and the Neutral phases. Furthermore, we calculated cross-correlations between MEI and VAC to assess time-lagged correlations from zero to six months. Due to seasonal variability, we only considered the months that have data from all three ENSO phases (January to March and August to December). Finally, we used the Neutral phase as a baseline to compare VAC during El Niño and La Niña. Thus, we subtracted the correlation coefficient difference between El Niño (La Niña) and the Neutral phase pixel-wise. Henceforth, we refer to this difference as the correlation anomalies, and interpreted as a measure of the effect of the ENSO phases on variation in vegetation activity.

5.2.5 Hotspots of correlations between ENSO and vegetation activity by ecoregions

To investigate regions with the strongest association between ENSO phases and vegetation activity, we evaluated the study area by terrestrial ecoregions (Figure 5.1). The ecoregion map integrates biogeographical principles with spatial distribution models of species and communities (Olson et al., 2001). Within these ecoregions, we computed the median Spearman correlation between ENSO phases and the vegetation component.

All data analyses were conducted in Julia 1.3 and used the ESDL.jl package (<https://esa-esdl.github.io/ESDL.jl/latest/>) that allows efficient processing over time series and multivariate statistics (Estupinan-Suarez et al., 2021; Mahecha et al., 2020).

5.3 Results

5.3.1 Global PCA

We calculated a unified principal component space for the entire study area. VAC, the leading vegetation activity component, captured a fraction of 0.45 of the variance. The second and third PC explained a fraction of 0.21 and 0.19 of total variance, respectively. Vegetation indices such as EVI and NDVI were the main variables contributing to the leading component (Table 5.2). In general, VAC had a good agreement with the

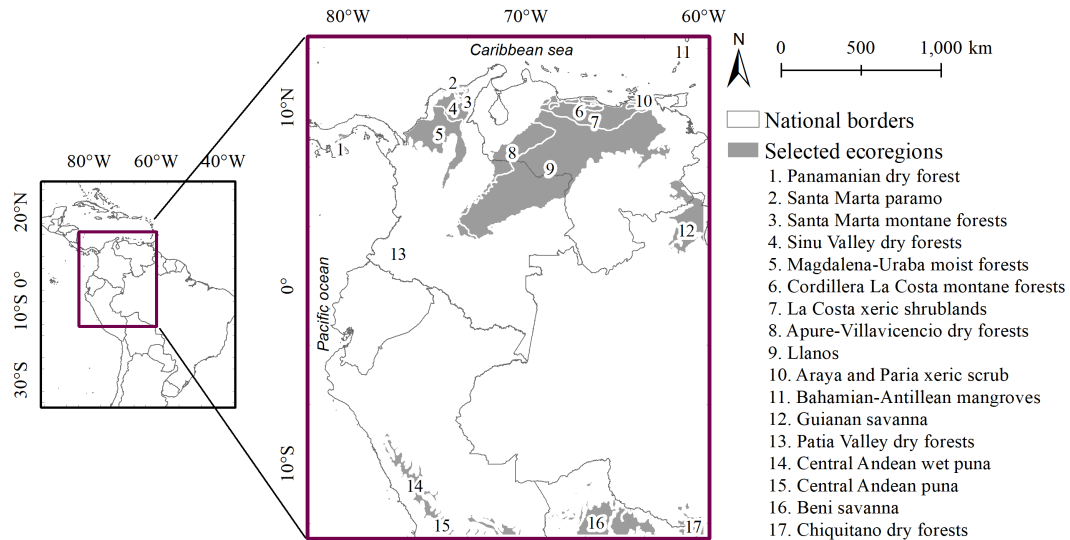


FIGURE 5.1: Map of northern South America. Grey polygons are selected ecoregions from Olson et al. (2001). The numbers indicate the ecoregion name. The black lines are the country borders.

input variables in grasslands and shrublands, with correlation coefficients $\rho > 0.7$ (Figure D.1). Nevertheless, the agreement between VAC and GPP decreases when $GPP \geq 6$ gCm^2/day , highlighting that VAC is not sensitive when productivity is high. Regarding broadleaf evergreen forests, the correlation between VAC and the input variables dropped dramatically $\rho \leq 0.5$, except for EVI. In fact, EVI has a nearly linear relationship with VAC within each of the assessed land cover classes, which contrasts with the observed pattern for NDVI. In this sense, we found that the correlation with NDVI highly varies among land cover classes with a $\rho \sim 0.3$ in broadleaved evergreen forests in comparison with $\rho \sim 0.7$ in savannas.

When conducting the Global PCA separately by land cover class (i.e., land cover PCA), we found that VAC accounted for 0.68 and 0.71 of the variance in shrublands and savannas, respectively, and decreased to 0.33 in broadleaved evergreen forests (Table D.1). Additionally, EVI, NDVI and GPP had similar loadings for savannas and shrublands (from -0.51 to -0.53), whereas EVI was the main variable in broadleaved evergreen forest (-0.67 vs. ≥ -0.42).

5.3.2 Vegetation activity during different ENSO phases along lags

We compared the variability of VAC-MEI correlations among ENSO phases. The correlation anomalies map (Figure 5.2) shows the differences of the Spearman correlation coefficients between the El Niño (La Niña) and the Neutral phase. The differences of correlation coefficients at each pixel and for each time lag illustrate how the strength

TABLE 5.2: Loadings from the Global PCA by (normalized) variable. EVI: Enhanced vegetation index. FPAR: Fraction of absorbed photosynthetically active radiation. GPP: Gross primary productivity. NDVI: Normalized difference vegetation index. VAC: Vegetation activity component. PC: Principal component.

Variable	VAC (PC ₁)	PC ₂	PC ₃	PC ₄
EVI	-0.58	-0.22	0.09	-0.78
FPAR	-0.41	0.85	-0.34	0.02
GPP	-0.49	-0.49	-0.59	0.43
NDVI	-0.51	0.04	0.73	0.46

of these correlations change spatially and temporally. Specifically, areas near the Pacific coast of Ecuador and Peru, the Caribbean coast, and northern inland savannas (i.e., *Llanos* and surroundings) are the areas with the highest variability in vegetation activity associated to El Niño and La Niña.

In general, we observed contrary patterns between areas north and south from the equator during all ENSO phases. In the northern region, correlation anomalies were negative during El Niño while positive during La Niña. Additionally, the correlation anomalies were stronger during El Niño than La Niña either in intensity (ρ) and duration (along lags). For example, in the Caribbean coast of Venezuela, the correlation anomalies were stronger for all lags during El Niño and only from 4 to 6 month during La Niña. In the *Sinu valley*, the correlation anomalies was stronger during El Niño in the first lags (0 to 2 month), whereas for La Niña occurred from 3 to 6 month lags.

The opposite was reported for the southern regions near the Ecuadorian and northern Peruvian coast, it was an increase on the correlation anomalies during El Niño that also became stronger along time (2 to 6 month lag). During La Niña, the correlation anomalies turned negative and stronger from 3 to 6 month lag. The Amazon biome included in the study area, i.e., the upper western Amazon basin, showed a more homogeneous pattern during La Niña than during El Niño. Mostly, during La Niña the correlation diminished from 2 to 6-month lag. The strongest period was at 3-month lag, then it slightly increased in the following months. Contrary, there was not a unified response of vegetation activity during El Niño. Around 10°S, the correlations were negative whereas in the northern part of the Amazon they were positive (Figure 5.2a). This heterogeneity is more discernible between 2 to 6 month lag, and highlights the variety of responses within the Amazon biome.

Despite these strong spatial differences, these results consistently show that time lags are more pronounced during El Niño than during La Niña phase. In brief, variability of the vegetation activity is higher during El Niño in the mountain ranges near the

southern Pacific coast, Caribbean coast and the *Llanos* along all lags. On the contrary, changes in vegetation activity related to La Niña are narrowed to specific lags.

5.3.3 Hotspots of ENSO impacts on vegetation activity

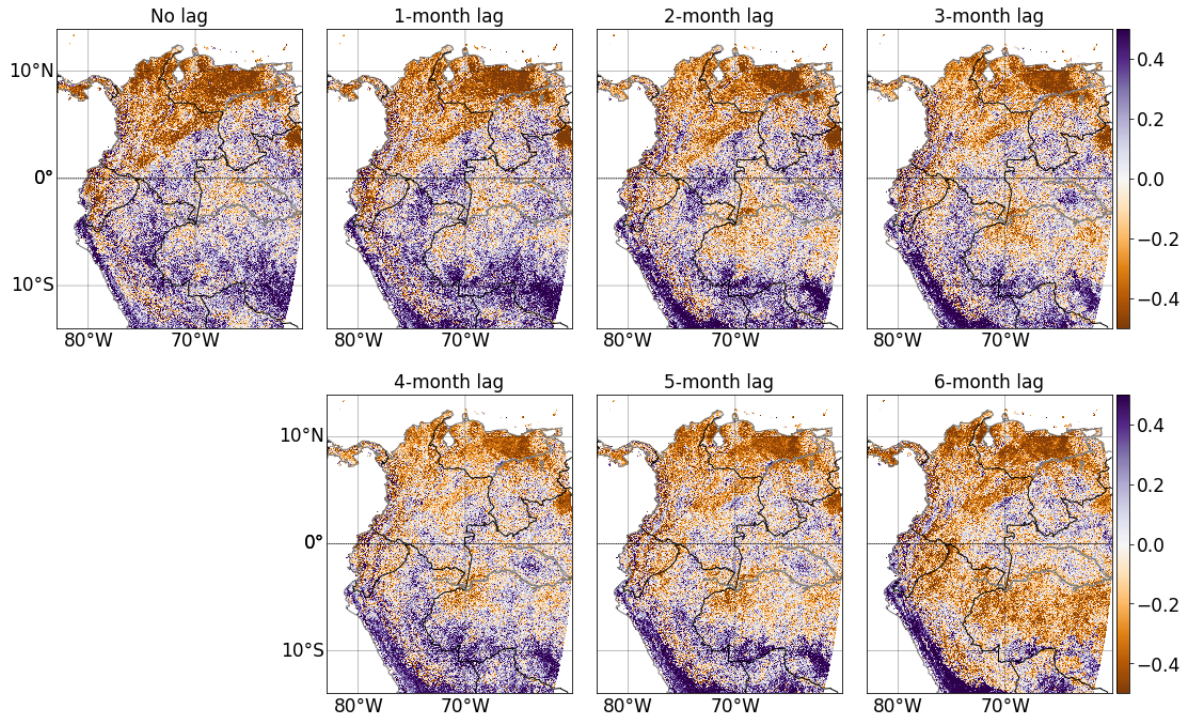
We also calculated the lagged Spearman correlation between VAC and MEI by ecoregions for each ENSO phase. We found that El Niño had the strongest correlation with VAC, and followed the same trajectory that the Neutral phase which had weaker values (Figure 5.3). On the contrary, the trajectory observed during La Niña is opposite to the ones from other ENSO phases, and also had weak correlations.

Contrasting correlation patterns emerged from the ecoregions along the latitude gradient. In fact, we clearly observed that in the northern region, the VAC–MEI correlations were stronger and negative during El Niño while they were weakly positive during La Niña. Conversely, we found a positive (negative) correlation with El Niño (La Niña) in the southern region. Interestingly, El Niño showed the strongest correlation in ecoregions located north and south from the equator, despite the opposite signs of the correlation i.e., a positive correlation in the southern region and a negative one in the north.

In particular for El Niño, the strongest correlation, and smallest standard deviation, was found in ecoregions closer to the coasts. Specifically, ecoregions closer to the Pacific ocean (i.e., *Central Andean Puna* and *Central Andean wet Puna*) and the Caribbean coast (i.e., *La costa xeric shrublands*) have the strongest correlation $|\rho| > 0.6$. Continental ecoregions such as *Guianan savanna*, *Llanos* and *Beni savanna* had $|\rho| \sim 0.5$ and a larger standard deviation. Overall, we observed a higher variability of VAC associated to ENSO in dry ecosystems; from the 16 ecoregions with the strongest correlation ($|\rho| > 0.4$), 69% are either savannas, dry forests, or xeric shrubs/scrubs. Figure S2 illustrates the gradient between ρ and precipitation of the driest quarter, this suggests that the correlation becomes weaker in wetter ecoregions.

To understand the climatic drivers of these ecoregional changes on vegetation, we explored time series of precipitation and air temperature. In general, plots of MEI against the climate variables showed similar trajectories that the ones observed with VAC (Figure D.3). As expected rainfall increases during La Niña in the northern ecoregions and decreases during El Niño in the southern ecoregions highlighting the regional heterogeneity. Otherwise the trajectories regarding the MEI-temperature correlation were not uniform along lags and often overlapped (Figure D.4).

(a) El Niño – Neutral



(b) La Niña – Neutral

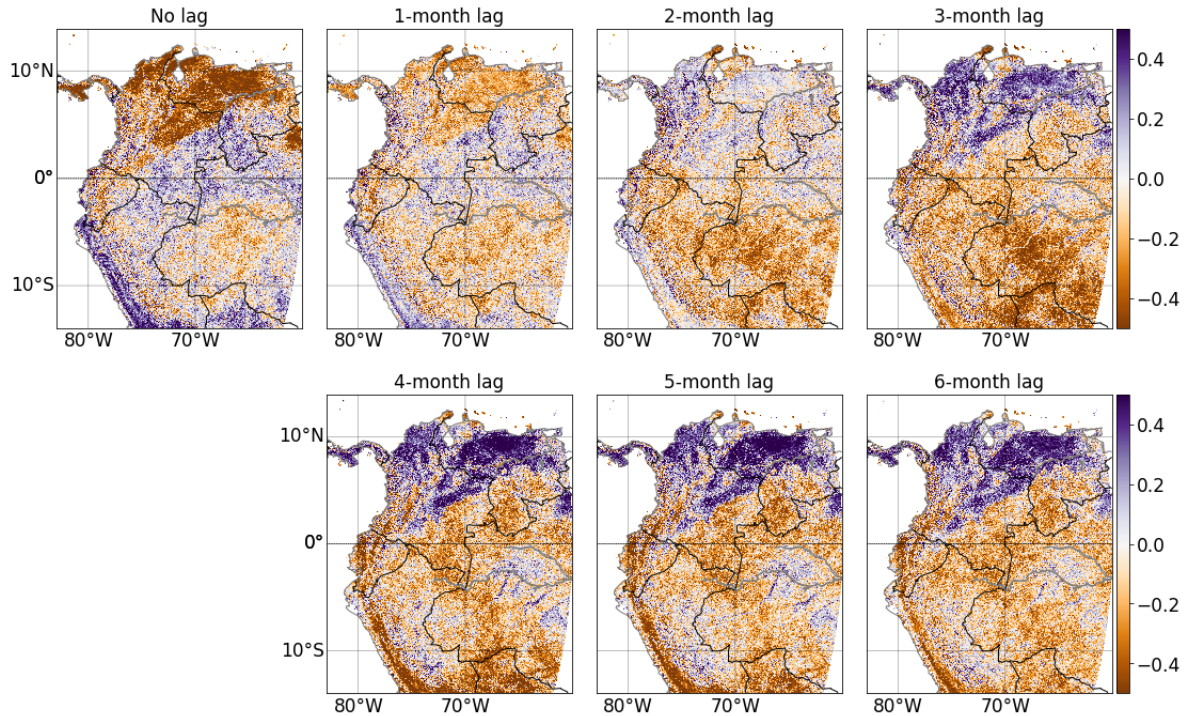


FIGURE 5.2: Map of correlation anomalies. The color map represents the difference in lagged Spearman correlations between MEI and the leading vegetation activity component (VAC) for (a) the El Niño and Neutral phase, and (b) the La Niña and Neutral phase. The black lines are country borders.

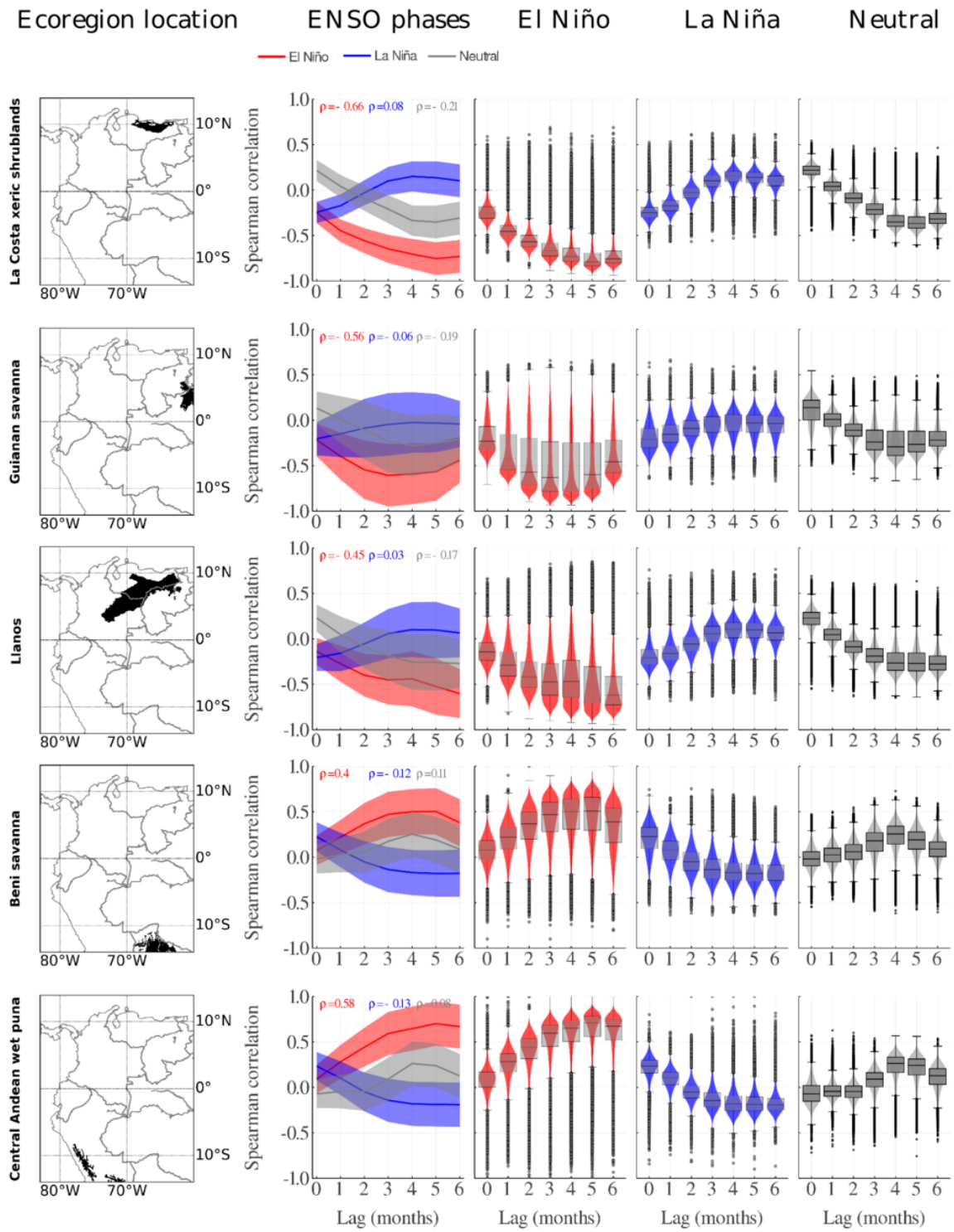


FIGURE 5.3: Median values of the lagged Spearman correlation between VAC and MEI by ecoregions. Ecoregions are in rows. From left to right: first column shows the ecoregion location. Second column represents the median correlations within the ecoregion by ENSO phases. The ribbon correspond to the standard deviation. Data distribution from El Niño, La Niña and Neutral phases by lags is showed in columns three to five respectively. Showed ecoregions have the highest correlation and an area larger than 50000 km²

5.4 Discussion

We analyzed time series of four vegetation variables to assess the effects of inter ENSO variability across time and space in NSA. To do this, we implemented the Global PCA approach (Kraemer et al., 2020) with the aim to capture a clearer vegetation activity signal that incorporated information from different sources. The derived leading component (i.e., VAC) was correlated to MEI within the different ENSO phases. Our findings show: (i) VAC captures the largest amount of variation from vegetation variables but its performance varies among land cover classes. (ii) There are opposite patterns of vegetation activity along time lags during El Niño and La Niña. In addition, (iii) vegetation in drier ecoregions show higher variability in El Niño regardless of the correlation sign with MEI. And, (iv) the trajectories of the lagged correlations are opposite in La Niña in comparison to the Neutral phase. In the following we discuss these findings in detail.

5.4.1 The vegetation activity component from the Global PCA

Our understanding of vegetation activity in the tropics is still limited. Recent studies are clarifying the climatic drivers at seasonal scales in different ecosystems (Chen et al., 2021; Hashimoto et al., 2021; Li et al., 2021; Uribe, Sierra, and Dukes, 2021), but its variability at longer temporal scales is still in a very early phase. One of the major challenges is that clear-sky observations are limited due to high cloud cover, and occasionally aerosols from fire (Hilker et al., 2012). In addition, ground data for models calibration is limited in the tropics reducing their regional predictability (Jung et al., 2020). Using the Global PCA approach, we captured the main characteristics of vegetation variables related to greenness and productivity. The derived vegetation activity proxy defined as VAC captured the largest amount of variation in seasonally dry climate ecosystems such as savannas and grasslands, but it was limited in the rainforest regions (Figure D.1). This was expected due mostly to two main reasons. First, areas such as the western Amazon or the Biogeographic Chocó, where annual precipitation could reach levels above 2500 mm and 11000 mm respectively (Poveda and Mesa, 2000), present large seasonal data gaps (Hilker et al., 2012) offering limited yearly data. Second, there are known drawbacks of optical satellites to detect changes in vegetation activity in ecosystem with very dense canopies such as the rainforest (Asner and Alencar, 2010; Huete et al., 1997; Köhler et al., 2018).

In general, our results show that VAC is dominated by the EVI signal. As a consequence, the information content in VAC is closer to vegetation greenness than to productivity. When looking at the dominant land cover classes, we found certain characteristics. For example, all variables show similar linear trends and correlations in

grasslands with the exception of a nonlinear relationship between VAC and GPP (Figure D.1). This clearly delineates the limitations of VAC regarding its ability to represent spatio-temporal variation of photosynthetic activity. With respect to broadleaved evergreen forests, EVI is the variable with the highest Spearman correlation ($\rho = 0.67$), and NDVI has the weakest correlation ($\rho = 0.3$). This was unexpected considering that both are measuring greenness, but reinforces the idea that EVI is more suitable for broadleaved evergreen vegetation because its lower sensitivity to atmospheric effects as well as less canopy saturation (Huete et al., 1997; Huete et al., 2002).

5.4.2 Lagged effects of ENSO in vegetation activity in NSA

Regions near the coast of southern Ecuador and northern Peru, including *La Sierra*, experienced a progressive increase on the correlations anomalies from lag zero to six during El Niño (Figure 5.2). In this area, the dominant land cover is grasslands and shrublands that might benefit from an increase in rainfall associated to El Niño. A similar response was observed in the Caribbean coast and neighbor inland areas such as the *Sinu Valley* and Orinoquia savannas (*Llanos*), but for La Niña phase and from lag three to six. Conversely, these ecoregions showed a negative correlation between VAC and El Niño across all lags, which can be related to a stronger water deficit during the dry season.

In comparison to studies that pointed to a reduction of vegetation activity during El Niño episodes in the Amazon basin (Hilker et al., 2014; Liu et al., 2017; Luo et al., 2018; Patra et al., 2017), we observed a more heterogeneous pattern. During the first three months, the western Amazon basin, covered by our study area, had a boost on the correlation anomalies. This could be explained by the fact that solar radiation is a limiting factor in tropical rainforests (Graham et al., 2003; Nemani et al., 2003), which benefit from less clouds during El Niño (Moura et al., 2019). Furthermore, this area is part of the wettest sub-basin in the Amazon (Schaik et al., 2018), therefore plants are not exposed to water stress in the early stages. However, at a six month lag, the correlation of vegetation activity with MEI became predominantly negative. This reveals contrasting vegetation responses related to a lagged effect, and it could be indicative of a link between prolonged solar radiation and the crossing of other environmental thresholds in the Amazon (Brando et al., 2010).

In general, we found negative correlations in the Amazon region during both ENSO phases at lag six, although spread of these correlations cover a larger area during La Niña phase. Taking into account that higher precipitation and less solar radiation is expected during La Niña, the vegetation activity should drop (Moura et al., 2019). This is consistent with observations by Graham et al. (2003) at the site level in the Panama

rainforest, who reported light as the main limiting factor for forest carbon uptake. Nevertheless, this result has to be carefully interpreted when working with optical satellite data due to quality issues in tropical forests and saturation of vegetation indices.

Overall, we summarize the interplay between ENSO phases and VAC as follows. There is a clear and contrasting response between the Ecuadorian and Peruvian regions near the coast on the one hand, and the Caribbean coast and Orinoquia savannas on the other hand, during both the El Niño and La Niña phases. This highlights the importance of spatial heterogeneity within the region, with a clear separation of the northern and southern regions of the study area. This variability is associated to the complex climatology of the region. For instance, Salas et al. (2020) showed that anomalies in precipitation and streamflow during ENSO phases are interlinked with moisture advection of regional jets affecting the anomalies in magnitude and spatial distribution. Thus, the Caribbean Jet, Choco Jet, and Orinoco Jet caused different anomaly patterns in the Caribbean, northern Pacific coast (*Choco*) and the Amazon, respectively.

Finally, the VAC–MEI correlation anomalies are stronger during El Niño than La Niña both in terms of intensity and delayed effects in the southern Pacific and Caribbean coasts as well as in the Orinoquia savannas. Elucidating these differences in ENSO phases is key for assessing changes, not only for the natural ecosystems but also in terms of further impacts for the population depending on them.

5.4.3 ENSO hotspots by ecoregions

Of significant importance is our finding of a high correlation between vegetation activity and MEI during El Niño in dry ecoregions. About 67% of ecoregions with $\rho > |\pm 0.4|$ are considered arid or semi-arid ecosystems. Currently, there is limited understanding on the physiological and climatological mechanisms that may lead to this response in such ecosystems, but some authors have highlighted some implications. For example, González-M et al. (2021) reported a negative net biomass balance in tropical dry forests during El Niño of 2015 for NSA. For similarly dry ecosystems such as the Cerrado in Brazil, Zanella De Arruda et al. (2016) found that this grassland dominated ecosystem is a source of CO₂ during the drier years. In general, research on tropical ecosystems with seasonal dry climate is still limited (Pennington, Lehmann, and Rowland, 2018), and even scarcer regarding their response to ENSO. Taking into account that semi-arid ecosystems are considered the main driver of interannual variability of GPP (Ahlström et al., 2015), and that ENSO is the main climatic driver at the same scale, it is crucial to increase our understanding in this context. Moreover, research focused on tropical savannas and dry forests is extremely relevant in light of decreased rainfall scenarios in places such as the Amazon due to the amplified effects of deforestation, drought, and

climate change (Hilker et al., 2014; Parsons et al., 2018; Shiogama et al., 2011; Vilanova et al., 2021; Zemp et al., 2017b).

In addition, our results also show that ecoregions with the highest correlation anomalies ($\rho > |\pm 0.6|$) were in relative proximity to the coast, i.e., *La Costa xeric shrublands* and *Central Andean puna*, followed by continental ecoregions such as *Guianan savanna*, *Llanos* and *Benni savanna*. This suggests that ecoregions near the coast are slightly more sensitive to MEI than the continental ones independent of their latitudinal location. Another geographical remark is the clear inverse response of VAC in ecoregion's located north and south from the equator during the same ENSO phase. In summary, while the sign correlation between VAC and MEI is positive in the North it becomes negative in the South and conversely. This pattern can be explained by the opposite effects that ENSO phases have in the different geographical locations in NSA, and it is aligned with the diverse changes on temperature and rainfall previously described for the region (NOAA, 2021a; Salas et al., 2020). Overall, our results suggests that due to the warm ENSO phase, the identified ecoregions have a higher sensitivity to changes in climate. Therefore, we considered them as the hotspots of ENSO effects on vegetation activity.

Lastly, we highlight the opposite trajectory of the lagged correlations during La Niña in comparison to the Neutral phase (Figure 5.3). Although the correlations between VAC and MEI are weaker during La Niña (from -0.23 to 0.09), the trend consistently has the opposite direction than the Neutral phase along lags. We repeated the same lagged correlation analysis but replacing VAC by precipitation (Figure D.3). Our results showed the same trajectories along lags for each ENSO phase and are consistent with the trajectories observed with VAC. These results were unexpected because La Niña is usually considered and enhanced Neutral phase from the atmospheric point of view (Bureau of Meteorology, 2012; Philander, 2018). These findings require further investigation to assess the local climate conditions during La Niña in the identified ecoregions. Currently, the assessment of land vegetation variability during La Niña and the Neutral phase is limited. But studies such as Pandey et al. (2017) reporting a global methane increase ($6-9 \text{ Tg CH}_4 \text{ yr}^{-1}$) during La Niña 2011 reveal the large impact and relevance of the ENSO cold phase. Regionally, important questions emerged after the extensive flooding by La Niña 2011. For example; How did the ecosystems recover after nutrients were recharged by the river sediments? For how long were soils water-logged? And, What are the consequences in the carbon cycle? Unfortunately, assessments of such impacts by La Niña are still pending.

5.4.4 Limitations and opportunities

The Global PCA approach aggregates the largest vegetation variability from the entire region into a unified PCA space. This has the advantage that all pixels are projected using the same loadings, and prevents opposite loading signs among neighboring pixels in a pixel-wise analysis. Nevertheless, PCA relies on orthogonal linear transformations and is limited for analysing nonlinear processes. There is an opportunity to use nonlinear dimensionality reduction methods such as Isomap (Tenenbaum, Silva, and Langford, 2000) or locally linear embeddings (Roweis and Saul, 2000) in future studies. Until now, these methods have not been computationally implemented for capturing simultaneously the variability in multiple dimensions (e.g., variables, space, time).

This study was carried out in a limited temporal and spatial domain that was defined by the overlap in the time spans of the assessed variables, together with consistency requirements for their spatial resolution. Nevertheless, there are opportunities to expand this analysis to cover larger areas such as the entire tropical domain and include longer time windows as more information becomes available. Furthermore, our method can be used to assess different ENSO indices in addition to the MEI used here.

5.5 Conclusion

Our results show that in NSA, a region with high heterogeneity in terms of climate and biophysical characteristics, vegetation activity presents opposite patterns within the same ENSO phase. Thus, when the correlation between vegetation activity and ENSO is positive during El Niño south of the equator, it is negative towards the north, and vice versa during La Niña. Understanding these differences is key for regional adaptation to more frequent and prolonged droughts or floods. Moreover, this could also contribute to the development of differentiated management plans for different ecosystem types that respond differently to the extreme phases of ENSO.

When assessing the lagged effects of ENSO, we observed that variability of vegetation activity is stronger in intensity and more prolonged during El Niño than during La Niña. Interestingly, this is independent of the correlation sign. In addition, we found that seasonally dry ecosystems are the most sensitive to El Niño. This is key in a region where research efforts are focused on tropical rainforest, and highlights the urgency of research in other ecosystem types.

Furthermore, we found a shift in the correlation sign in the Amazon around five to six month lag. This suggests the role of different environmental thresholds over time that have to be further investigated. Analyses of prolonged periods of solar radiation as well as water availability could contribute to elucidate ecosystems thresholds over consecutive months for specific ENSO events.

Overall, our approach captured the integrated response of vegetation as assessed by the combined information content of different metrics of vegetation activity. Although we observed a large heterogeneity of responses in space and time, our analysis unequivocally shows the contrasting variability of vegetation activity during ENSO phases in NSA. By integrating different data sources, and by combining spatial and temporal analyses, our results provide relevant information to disentangle the different responses related to large-scale climate events such as ENSO considering those specific characteristics that are endemic to each ecoregion in NSA.

5.6 Acknowledgments

LMES acknowledges the support of DAAD and its Graduate School Scholarship Program (57395813), as well as the International Max Planck Research School for Global Biogeochemical Cycles.

Chapter 6

Discussion

The overarching goal of my thesis was *to understand dynamics of vegetation activity occurring across time scales globally and in a regional context*. To achieve this, I analyzed variables related to vegetation activity acquired by satellites or generated by models through two approaches: (i) time series decomposition, and (ii) dimensionality reduction analysis. The first approach disentangles information contained in the seasonal cycle, short-term (< 1 year), and longer-term (> 1 year) time scales. I applied it to vegetation greenness and climate variables at the global scale. The second approach sought to extract the largest amount of information from a combination of vegetation variables through dimensionality reduction analysis in NSA. NSA is a region with a large diversity of climate and ecosystems, but also with scarcity of data. This chapter summarizes the main results and explains the implications of my work from a more integrated perspective. I also emphasize on the relevance and implications for NSA, by putting the obtained results into a regional context.

6.1 Dominant modes of variability in NDVI and climate variables across time scales

Investigating vegetation activity at different time scales is of notable importance, as well as it is recognizing processes that occur beyond the often dominant seasonal cycle, and among different ecosystems. I assessed more than 30 years of NDVI records in tandem with climate variables. The research question addressed was RQ1: *What are the modes of variability in vegetation greenness and what is its co-variability with climate across time scales globally?* In summary, my results show that; (i) 27 % of NDVI variability is dominated either by short- or longer-term oscillations. Short-term oscillations are mainly found in broadleaved evergreen forest while longer-term oscillations are predominantly found in shrublands; (ii) the dominant co-variability map between

NDVI and climate highlights homogeneous regions in the northern latitudes and heterogeneous regions in the tropics; and, (iii) the sign of co-variability between NDVI, temperature and precipitation differs across time scales.

In general, my results provide a comprehensive understanding of vegetation greenness considering the temporal component. Of significant importance is the assessment of climate-vegetation interactions. As I reported, correlations between NDVI and climate variables can have opposite signs when assessing the seasonal cycle and longer-term oscillations in different regions (Chapter 3, Figure 4). Considering the expected higher climate variability, it is important to understand how strongly vegetation greenness varies at these time scales, and which are the main co-varying climate drivers. In addition, it is known that estimations of terrestrial vegetation productivity are limited at capturing interannual variability (Jung et al., 2020). Providing differentiated spatio-temporal frameworks of NDVI will improve model predictions of vegetation productivity as well as forecasting ecosystem dynamics in climate change scenarios.

Another key result was the regional differences of the dominant co-variability between NDVI and climate variables (Chapter 3, Figure 2). I identified three main patterns along latitudes. At northern latitudes (above 23° N), the seasonal cycle of NDVI and air temperature dominate in combination with either short-term oscillations or the seasonal cycle of precipitation. At southern latitudes (below 23° S), longer-term oscillation of NDVI become significant, whereas the strong annual seasonality of air temperature prevails. Long-term oscillations of NDVI were dominant in deciduous forest and shrubs, and herbaceous covers. This is aligned with studies highlighting the role of semi-arid ecosystems as main drivers of interannual variability of the terrestrial carbon sink (Ahlström et al., 2015). Furthermore, the tropics present multiple combinations of dominant co-variability, especially in South America and South East Asia, where short-term oscillation of NDVI and precipitation are remarkable (see Section 6.2).

A remaining task that could be addressed in future studies, is a more detailed assessment of short-term oscillations of vegetation activity, which until now has been constrained by data availability. Such analyses are generally carried out using eddy covariance data (Braswell et al., 2005; Mahecha et al., 2010), and therefore they are site specific. In our global study with GIMMS NDVI, we hypothesized that the dominant variability of short-term oscillations in the tropics might be explained by the Madden Julian Oscillation, i.e. atmospheric oscillations at frequencies between 20 and 90 days (Madden and Julian, 1971). Assessing this interseasonal variability using weekly composites, based on daily observations such as GIMMS, is insufficient because of limited data sampling and quality. New avenues are opened with geostationary satellite missions such as the Geostationary Operational Environmental Satellite 16 (GOES-16)

that carries the Advanced Baseline Imager (ABI) instrument. ABI includes the visible and infrared spectral bands needed to compute NDVI (Schmit et al., 2018). In comparison to AVHRR and MODIS, ABI observations are acquired every 10 to 15 minutes preserving a moderate spatial resolution (1 km), and at the same time increasing significantly the number of clear-sky observations (Hashimoto et al., 2021). In fact, geostationary satellites open an opportunity to explore vegetation activity in shorter time periods and beyond the seasonal cycle such as the rapid growing season of annual plants, plants' response to water-pulses in water limited ecosystems, among others (Küçük et al., 2022).

6.2 Analysis of vegetation activity in northern South America

In NSA, the patterns of dominant co-variability between vegetation and climate are complex as shown in Chapter 3. The ecosystems in NSA are key for the water and carbon cycles at annual and interannual time scales. Currently, there is limited research assessing land-atmosphere interactions, and therefore limited understanding of ecosystem dynamics in the region. At the same time, and despite their continental and global importance, ecosystems in NSA are highly threatened due to the expansion of the agricultural frontier, which imposes large threats on these ecosystems that may be amplified by climate change (Anaya et al., 2020; Armenteras et al., 2019; Barona et al., 2010; Clerici et al., 2020; Seymour and Harris, 2019; Song et al., 2018).

Spatio-temporal analyses of vegetation in NSA are often constrained because of limited data access and processing power, low data availability due to clouds, and large uncertainty in global products. To address the first constraints, I implemented the RegESDL and addressed RQ2: *How to carry out computationally efficient time series analysis in NSA with moderate spatial resolution?* Overall, the RegESDL facilitates big data analytics, and contributes to integrating geographical layers from different fields such as climate, vegetation, biodiversity, among others (see section 6.3).

6.2.1 Diverse modes of seasonality of vegetation activity in NSA

As a first application of the RegESDL, I explored the seasonal cycle of vegetation activity in the region. To understand seasonality in NSA, it is important to consider two key points. First, climate seasonality is dominated by precipitation regimes; unimodal and bimodal. Second, seasonality of vegetation in tropical ecosystems has been related either to precipitation or solar radiation (Hashimoto et al., 2021; Uribe, Sierra, and

Dukes, 2021). In fact, tropical ecosystems with strong seasonality are mostly water-limited such as tropical dry-forests and savannas. By contrast, the presence of seasonal cycles of vegetation activity remains a controversial topic for tropical rainforest (i.e. broadleaved evergreen forest). However, different fields of research have reported structural ecosystem changes such as leaf flushing and leaf litter falling throughout the year (Chave et al., 2010; Chen et al., 2021; Li et al., 2021; Lopes et al., 2016). The research question addressed in this study was RQ3: *What are the modes of seasonality of vegetation activity in NSA?*. To do this, I combined a set of variables acquired by satellites and models through dimensionality reduction analysis. The resulting leading component was used as a proxy of vegetation activity and as input for the following computations.

I proposed a seasonality ratio (SR) metric (Chapter 4) based on the contribution of annual and semiannual oscillations. The results depict contrasting patterns of SR in ecosystems with a strong climate seasonality (e.g. water-limited ecosystems). Regions with unimodal precipitation regimes such as tropical savannas surprisingly exhibited one cluster with a dominant bimodality of vegetation activity (Chapter 4, Figure 8). This was the case of a biotic unit (i.e., *Arauca*) covered by a mosaic of flooded shrubs/grass and grasslands. From here, the emerging hypothesis was that different vegetation types peak at different times. Thus, the observed double-peak signal of vegetation activity could be explained by a mosaic of different vegetation layers in savannas and their functional diversity. In contrast, regions with bimodal precipitation regimes such as lowland inter-Andean valleys (i.e., *Magdalena Medio & depression Momposina*) showed two different SR clusters. The first cluster highlighted unimodal seasonality in drier regions. The second cluster was related to wetter regions and it was dominated by bimodal values. Besides water availability, the emergence of these two SR clusters might be explained by the dominant vegetation types (i.e., broadleaved evergreen forest and crop-rainfed). A detailed assessment of vegetation types, with higher spatial resolution imagery and multiple sensors such as Sentinel-1 and 2, would contribute to elucidate new hypotheses regarding ecosystem function together with climate.

In addition, the seasonality ratio map (Chapter 4, Figure 6) revealed a pattern of high complexity in western Amazon. In this part of the basin, the SR covered the entire range; starting from bimodal values ($SR < 1$), going through pixels where both oscillations had equivalent contribution ($SR \sim 1$), as well as pixels with unimodal seasonality ($SR > 1$). A remark is that the vegetation activity proxy used to calculate the SR only explained ~ 0.3 of variability in the majority of broadleaved evergreen forest pixels. This is aligned with results from Chapter 3, where we reported that the dominant variability of NDVI was enclosed in the short-term oscillations instead of the

seasonal cycle. Nevertheless, these findings showed that even though the seasonal cycle is not dominant in broadleaved evergreen forests, and was partially captured by the vegetation activity component, heterogeneous patterns emerge when isolated annual and semiannual oscillations are assessed. As introduced in section 6.1, first analyses from geostationary satellites are providing new evidence of interseasonal changes in vegetation activity at regional scales. In this sense, results from ABI have revealed significant seasonality in 4 from 7 sites in the Amazon despite the limitations of working with NDVI in dense canopies (Hashimoto et al., 2021).

Revealing seasonality of vegetation activity, and the corresponding climate drivers is still an ongoing process in tropical ecosystems. These results are a first step for understanding different seasonalities of vegetation activity in very diverse ecosystems. They highlight that not only climatic factors explain vegetation seasonality in NSA but the composition of the ecosystems as well. An important step forward is to integrate ground measurements and remote sensing observations from active and passive sensors. A good example is the integration of phenocams and vegetation indices from satellites by Lopes et al. (2016), who evaluated leaf canopy flushing in the Amazon. Similar studies have to be extended, together with the development of interdisciplinary frameworks and collaborations between big data analyses and ground measurements. Outcomes from such analyses will go beyond assessing seasonality of vegetation activity and its implications on carbon cycling, they will be also relevant to understand other ecological processes such as nutrient cycling, animal migrations, among others.

6.2.2 Heterogeneous spatio-temporal activity of vegetation during ENSO phases in NSA

In my last study, I investigated vegetation variables related to greenness and photosynthesis during the ENSO phases (i.e., El Niño, La Niña, Neutral)(Chapter 5). The main research question addressed was RQ4: *Where are the hotspots of vegetation activity observed during the ENSO phases in NSA?* To achieve this, I integrated the time series of GPP, FPAR, NDVI and EVI using the Global PCA approach over the entire domain of NSA from 2001 to 2014. The method facilitated obtaining a clearer signal by retaining most of the variability in the data. I used the resulting leading component (i.e., PC1) as a proxy of vegetation activity, and correlated it with MEI for each ENSO phase. Additionally, I analyzed the results by ecoregions.

My findings depicted a large heterogeneity of vegetation activity in NSA during El Niño and La Niña phases. The results highlight the contrasting variability in vegetation activity on regions near the southern Pacific coast compared to the northern savannas and Caribbean coast during the same ENSO phase. In general, changes in

vegetation activity, in magnitude and duration, were stronger during El Niño than La Niña (Chapter 5, Figure 2). Moreover, when aggregating the results by ecoregions, the strongest correlations with MEI was found in the drier ecoregions during El Niño phase suggesting their high vulnerability to the ENSO cold phase (Chapter 5, Figure 3).

From the analysis at the ecoregion level, an unexpected outcome was the similarity in trajectories between the Neutral phase and El Niño along lags. Interestingly, the Neutral phase followed the same trajectory as El Niño but with lower correlation values, and at the same time they both are opposite to La Niña trajectory (Chapter 5, Figure 3). This is contrary to the meteorological definition of the ENSO phases in the Pacific ocean. From the atmospheric perspective, La Niña and the Neutral phase are more similar than El Niño. In fact, La Niña is considered as an enhanced Neutral phase with an intensification of the Walker circulation and convection. By contrast, El Niño presents a weaker Walker circulation, with often a displacement of convection cells due to reversed winds (Bureau of Meteorology, 2012; Philander, 2018). I carried out an equivalent analysis with precipitation and it confirmed the similar trajectories between El Niño and the Neutral phase (Figure D.3). This underlines the limited knowledge about vegetation activity during the Neutral phase as well as the unknown local forcing under regular climatic conditions which has to be further investigated.

Another remark is that the Global PCA is a linear dimensionality reduction method. Linear methods demand less computing resources in comparison to nonlinear methods, as well they have a broader implementation by different software. Nevertheless, they have limited ability to reveal non linearities in the system. I selected the Global PCA for its capability to capture data variability along the spatial and variable dimensions, which was aligned with my objective of assessing the regional variability of vegetation activity in NSA. Otherwise, nonlinear methods such as Isomap (Tenenbaum, Silva, and Langford, 2000) do not have the computational implementation for reducing both dimensions simultaneously. I carried out a complementary analysis to compare PCA and Isomap by reducing only one data dimension (univariate analysis). The idea was to learn about the possible differences between both methods. The results showed that when reducing the variables pixel-wise both methods have a similar performance. In fact, the resulting correlation maps between the leading component and MEI showed similar patterns (Appendix A, Figure A.1). Furthermore, when reducing the spatial dimension by watershed using a single variable (i.e., GPP), the correlation between Isomap components and MEI was slightly higher in comparison to PCA (Appendix A, Figure A.2).

Finally, results from both regional analyses have to be seen in the light of data availability. NSA is exposed to the vertical migration of the ITCZ that in combination with the recycling moisture from the Amazon create seasonal dense cloud cover. This

constrains the number of reliable observations with optical sensors. In my research, I applied data quality flags to the input variables (see Appendix C, Table S2). The implemented criterion diminishes the number of final observations and therefore the results are particularly limited in the wet seasons. In consequence, some regions were totally excluded (i.e., northern Pacific coast *Choco Biogeografico* - Chapter 4, Figures 4 and 6) due to the reduced or null values in the *SR* analysis.

6.3 Opportunities from big data infrastructures

A common solution to facilitate access to big data and processing is to bridge the gap between user code and data repositories. This is the principle of the RegESDL approach implemented in this research. Similar initiatives are the well-known Google Earth Engine or the national data cube projects led by CSIRO. Google Earth Engine offers extensive data sets operated with closed source software (Lewis et al., 2017; Tamiminia et al., 2020), while the CSIRO data cubes are open source software but have focused exclusively on satellite imagery (Bravo et al., 2017; Giuliani et al., 2017, 2019; Lewis et al., 2017). By contrast, the RegESDL is open source software that provides different data sets to link Earth science, biodiversity and ecosystem function.

In this dissertation, I show the RegESDL deployment, its successful implementation to efficiently carry out time series analysis and multivariate statistics. I used time series decomposition to calculate annual and semiannual oscillations pixel-wise taking advantage of the data storage which is optimized for temporal analysis (Chapter 4). As well, the performance was not degraded when working over the spatial domain to calculate the Global PCA which required to load all land pixels from all variables per time step (Chapter 5). Another application of the RegESDL was by Burbano Girón et al. (2020) who assessed ecosystem functional types in tropical dry forest. Using break-point detection to assess and characterize seasonality, they found that the most relevant indicators for vegetation monitoring are mean time series of productivity, photosynthetic activity, and water use.

The RegESDL offers unique properties to infer a more dynamic understanding of vegetation activity in NSA across time scales. My results reveal the diverse modes of seasonality of vegetation activity, and the contrasting variability of vegetation during the ENSO phases. Furthermore, they were also contextualized within a biodiversity framework, and the results can be seen to benchmark the assessment of essential biodiversity variables. Currently, the RegESDL is ready to assess land-atmospheric impacts of deforestation or land use change, explore ecosystem functional traits, and could potentially integrate species distribution data from biodiversity platforms (e.g.

GBIF, Biomodelos). The latter is key for understanding the response of species to land surface dynamics.

Yet, widespread use of the RegESDL is an ongoing process. One of the bottlenecks is to have up-to-date data which is a challenge for any data infrastructure. To tackle this, the ESDL package recently implemented an option to ingest new data sets independently. With this new development, users could easily extend the core data set and leverage the software for efficient data processing over any data dimension (e.g., time, space). Moreover, the latest ESDL package supports different spatial resolutions (Cremer et al., 2020) compared to the preset grid from the RegESDL. These recent developments demonstrate the rapid advances in technology and the importance of a continuous maintenance of data infrastructures. Finally, a major challenge is to consolidate a community that often requires technological support and up-to-date documentation. This task implies a team and long-term funding to ensure a learning phase, consolidation of analysis and timely technological solutions.

6.4 New avenues to study vegetation activity across time scales

In my research, the assessment of vegetation activity across time scales required long records of vegetation variables, mostly to capture long-term oscillations and assess interannual variability. Hence, the temporal span of the variables was one of the main criteria for selecting the input variables. For the global analysis, I used the longest available records of NDVI resampled at 0.5° . For the regional analysis, I accepted a compromise between the temporal span and the higher spatial resolution i.e., 0.0083° . In general, long records are suitable for assessing vegetation activity beyond the seasonal cycle, which is important for studying modes of interannual variability (e.g., ENSO), or vegetation trends such as browning and greening (Chen et al., 2019; Cortés et al., 2021; Piao et al., 2020). In addition, a longer data span obviously increases the number of good quality observations in regions with frequent cloud-cover such as NSA. In brief, the higher the number of observations the more comprehensive understanding of processes throughout the seasons could be obtained.

New mathematical methods are promising to improve estimates of vegetation indices. Examples of recent vegetation indices are the near-infrared reflectance of terrestrial vegetation (NIR_V) (Badgley, Field, and Berry, 2017), and the kernel NDVI (Camps-Valls et al., 2021). They both can be calculated from any retrieval that has been previously used to compute NDVI, which is advantageous for working with long operational satellites such as AVHRR or MODIS. NIR_V has achieved a good agreement with

GPP (Badgley, Field, and Berry, 2017), and it has been recently used as a substitute of NDVI on the latest SIF downscaled product (Duveiller et al., 2020). Complementary, the newest Kernel NDVI showed less saturation than traditional vegetation indices, and a stable performance at different spatial and temporal resolutions (Camps-Valls et al., 2021). Further analysis of these indices in regions such as the tropics, where data is scarce and model performance is low, could bring more robust results when estimating vegetation activity at different time scales. A follow-up analysis would be to assess their performance in tropical ecosystems using the *SR* metric proposed in this thesis (Chapter 4). This will contribute to gaining a better understanding of seasonality in the region.

In addition, new satellite products are improving our understanding of vegetation activity over short periods of time. These satellite observations are suitable for assessing seasonality or intraseasonal variability of vegetation. On the one hand, satellite missions to measure SIF are promising for improving estimates of vegetation carbon uptake due to its clear relationship with GPP, and its lower sensitivity to cloud-cover than traditional vegetation indices (Duveiller et al., 2020; Guanter et al., 2012; Yang et al., 2015). In fact, SIF has revealed different seasonal cycles in the Amazon, but further investigation is needed to get a consensus between SIF seasonality and environmental drivers (Köhler et al., 2018; Koren et al., 2018; Lee et al., 2013). On the other hand, new geostationary satellites are providing high-frequency data sets (e.g., GOES-16, Meteosat, see section 6.1). This latest generation of satellites have included sensors to assess vegetation without compromising spatial resolution (~ 1 to 5 km), and are optimal for regional analyses (Hashimoto et al., 2021; Küçük et al., 2022).

Overall, these new satellite products and methodologies open new avenues to increase our understanding of vegetation activity across time scales. In fact, they offer a step forward to continue analyses based on the results from this dissertation. Specifically, I see a clear opportunity to use high-frequency data to investigate regional vegetation-climate variability at intraseasonal time scales such as the connection between dominant short-term oscillations in tropical South America and the Madden Julian oscillation hypothesized in Chapter 3. In addition, analysis with high-frequency data will require efficient access over the time dimension, which can be efficiently carried out with the RegESDL.

Chapter 7

Concluding remarks

Overall, this thesis has contributed to advance in our understanding of vegetation activity at different temporal scales globally, and in the domain of NSA. It started with a global assessment of vegetation greenness and climate variables at the short-term, seasonal cycle and longer-term time scales. Later, I emphasized on the seasonal and interannual variability of vegetation activity in NSA, a region where vegetation-climate interactions are complex, and often reduced to analysis of evergreen broadleaved forest. Through this research, I took advantage of open data sets, different mathematical approaches for time series decomposition and dimensionality reduction, and the state-of-the-art computing technology to carry out effectively time series analyses over different geographical domains.

In my first study, the analysis of global modes of variability of vegetation greenness and climate (Chapter 3), I identified regions where the dominant NDVI variability occurs at short- and long-term scales. The emerging dominant variability map of NDVI, precipitation and temperature delineated regions with complex land-atmosphere interactions such as NSA and South East Asia. This emphasizes the regional dominance of spatio-temporal processes occurring beyond the seasonal cycle. Another relevant result was the opposite correlation sign observed between NDVI and climate variables for the seasonal and longer-term scales. In face of higher expected climate variability, differentiating the co-variability between vegetation and climate across time scales by regions will improve the spatial and temporal accuracy of climate-vegetation models and projections.

In my second study, I deployed the RegESDL and investigated the modes of seasonality on vegetation activity in NSA (Chapter 4). I showed the heterogeneous seasonal patterns based on the integration of vegetation variables and the computation of the *SR* metric. A remarkable result was that peaks of seasonality observed with remote sensing do not always correspond to the climate seasonality driven by precipitation (i.e., unimodal and bimodal regimes). This suggests that other environmental conditions besides climate are driving the seasonal cycle of vegetation. Analyses with higher spatial resolution from optical satellites in tandem with other sensors will elucidate

other features of seasonality related to ecosystem structure and functional diversity.

In my last study, I compared the vegetation activity in NSA during the ENSO phases (Chapter 5). My results showed that during the same ENSO phase there is an inverse variability of vegetation activity in areas near the southern Pacific coast in comparison with the northern inland savannas (*Orinoquia*) and the Caribbean coast. Also, I found that the strongest correlations occur in drier ecoregions during El Niño, which suggest a high vulnerability of these regions to the ENSO cold phase. This finding underlines the role of arid and semiarid ecosystems in a region where research is mostly focused on tropical rainforest. In addition, due to the remarkable heterogeneous patterns of vegetation activity during ENSO phases, governments must consider differentiated responses and adaptation strategies especially when these events become more frequent and stronger.

In addition, the RegESDL deployment facilitated the implementation of multivariate approaches to assess seasonality and inter-annual variability in NSA. Currently, the RegESDL is prepared to support efficient analyses considering the increasing open data sets from remote sensing and models. Moreover, it is an open source infrastructure to handle big data, which is crucial for warranting independence in data access and processing. And, it is an alternative approach to tackle the standard time series processing in geoscience.

In conclusion, disentangling these processes and assessing them separately allowed me to formulate new hypotheses about mechanisms of ecosystem-climate interactions, reveal hidden spatial and temporal patterns in these interactions, and produce relevant information for ecosystems conservation and management. This is critical for the tropics where there is a limited understanding of dynamics of vegetation activity and land-atmosphere interactions. In addition, there is a potential for incorporating this knowledge on vegetation and climate models to improve predictions of future ecosystem responses to climate change.

Finally, I hope this dissertation motivates follow-up studies that emphasize research in three topics; assessment of intraseasonal variability associated with short-term oscillations, further investigation of tropical semiarid ecosystems regarding their seasonality and vulnerability to ENSO, and a deeper understanding of the ENSO Neutral phase. As last remark, the methods and analyses implemented in this dissertation can be applied to the increasing amount of satellite products and new vegetation indices, and can be extended to other regions. This would help to develop new hypotheses and improve our understanding of vegetation-climate interactions across multiple time scales.

Bibliography

- Ahlström, Anders et al. (2015). “The dominant role of semi-arid ecosystems in the trend and variability of the land CO₂ sink”. In: *Science* 348.6237, pp. 895–899. ISSN: 0036-8075.
- Anaya, Jesús A et al. (2020). “Drivers of Forest Loss in a Megadiverse Hotspot on the Pacific Coast of Colombia”. In: *Remote Sensing* 12.8, p. 1235.
- Andrade, M Gonzalo (2011). “Estado del conocimiento de la biodiversidad en Colombia y sus amenazas. Consideraciones para fortalecer la interacción ciencia-política”. In: *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 35.137, pp. 491–508. ISSN: 0370-3908.
- Arias, P. A. et al. (2021a). “Technical Summary”. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by V. Masson-Delmotte et al. Cambridge University Press. In Press.
- Arias, Paola A et al. (2021b). “Technical summary”. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by V. Masson-Delmotte et al. Cambridge University Press.
- Armenteras, Dolores et al. (2019). “Curb land grabbing to save the Amazon”. In: *Nature Ecology & Evolution* 3.11, p. 1497. ISSN: 2397-334X.
- Asner, Gregory P and Ane Alencar (2010). “Drought impacts on the Amazon forest: the remote sensing perspective”. In: *New Phytologist* 187.3, pp. 569–578.
- Baccini, A et al. (2017). “Tropical forests are a net carbon source based on aboveground measurements of gain and loss”. In: *Science* 358.October, pp. 230–234.
- Badgley, Grayson, Christopher B Field, and Joseph A Berry (2017). “Canopy near-infrared reflectance and terrestrial photosynthesis”. In: *Science advances* 3.3, e1602244. ISSN: 2375-2548.
- Barona, Elizabeth et al. (2010). “The role of pasture and soybean in deforestation of the Brazilian Amazon”. In: *Environmental Research Letters* 5.2, p. 24002. ISSN: 1748-9326.
- Bastos, Ana et al. (2013). “The global NPP dependence on ENSO: La Niña and the extraordinary year of 2011”. In: *Journal of Geophysical Research: Biogeosciences* 118.3, pp. 1247–1255. ISSN: 2169-8953.

- Bastos, Ana et al. (2018). “Impact of the 2015/2016 El Niño on the terrestrial carbon cycle constrained by bottom-up and top-down approaches Climate Processes Section, Environment and Climate Change”. In: *Phil. Trans. R. Soc. B* 373: 20170.
- Bloom, A Anthony et al. (2017). “A global wetland methane emissions and uncertainty dataset for atmospheric chemical transport models (WetCHARTs version 1.0).” In: *Geoscientific Model Development* 10.6. ISSN: 1991-959X.
- Bolaños, Silvana et al. (2021). “GRACE reveals depletion of water storage in north-western South America between ENSO extremes”. In: *Journal of Hydrology* 596, p. 125687. ISSN: 0022-1694.
- Bonan, Gordon (2015). *Ecological climatology: concepts and applications*. Third edit. Cambridge University Press. ISBN: 1316425193. DOI: 10.1017/CB09781107339200.
- Brando, Paulo M et al. (2010). “Seasonal and interannual variability of climate and vegetation indices across the Amazon”. In: *Proceedings of the National Academy of Sciences* 107.33, 14685 LP –14690.
- Brando, Paulo M et al. (2019). “Droughts, wildfires, and forest carbon cycling: A pantropical synthesis”. In: *Annual Review of Earth and Planetary Sciences* 47, pp. 555–581. ISSN: 0084-6597.
- Braswell, Bobby H et al. (2005). “Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations”. In: *Global Change Biology* 11.2, pp. 335–355. ISSN: 1354-1013.
- Bravo, Germán et al. (2017). “Architecture for a Colombian data cube using satellite imagery for environmental applications”. In: *Colombian Conference on Computing*. Springer, pp. 227–241.
- Brockwell, Peter J. and Richard A. Davis (2002). *Introduction to time series and forecasting*. Second. Springer, p. 434.
- Burbano Girón, Jaime et al. (2020). “Identification of ecosystem functional types in tropical drylands: a functional outlook for monitoring EBVs”. In: *GEO BON open science conference and all hands meeting*.
- Bureau of Meteorology (2012). *Record-breaking La Niña events*. Bureau of Meteorology, Commonwealth of Australia, p. 24.
- Cai, Wenju et al. (2014). “Increasing frequency of extreme El Niño events due to greenhouse warming”. In: *Nature Climate Change* 4.2, pp. 111–116.
- Cai, Wenju et al. (2020). “Climate impacts of the El Niño–southern oscillation on South America”. In: *Nature Reviews Earth & Environment* 1.4, pp. 215–231. ISSN: 2662-138X.
- Camps-Valls, Gustau et al. (2021). “A unified vegetation index for quantifying the terrestrial biosphere”. In: *Science Advances* 7.9, eabc7447. ISSN: 2375-2548.

- Cess, Robert D et al. (2001). “The influence of the 1998 El Niño upon cloud-radiative forcing over the Pacific warm pool”. In: *Journal of climate* 14.9, pp. 2129–2137. ISSN: 1520-0442.
- Charney, Jule G et al. (1979). *Carbon dioxide and climate: a scientific assessment*. National Academy of Sciences, Washington, DC.
- Chave, Jerome et al. (2010). “Regional and seasonal patterns of litterfall in tropical South America”. In: *Biogeosciences* 7.1, pp. 43–55. ISSN: 1726-4170.
- Chen, Jing M et al. (2019). “Vegetation structural change since 1981 significantly enhanced the terrestrial carbon sink”. In: *Nature communications* 10.1, pp. 1–7. ISSN: 2041-1723.
- Chen, Xiuzhi et al. (2021). “Vapor pressure deficit and sunlight explain seasonality of leaf phenology and photosynthesis across Amazonian evergreen broadleaved forest”. In: *Global Biogeochemical Cycles* 35.6, e2020GB006893. ISSN: 0886-6236.
- Clerici, N et al. (2020). “Deforestation in Colombian protected areas increased during post-conflict periods”. In: *Scientific Reports* 10.1, pp. 1–10. ISSN: 2045-2322.
- Cortés, José et al. (2021). “Where are global vegetation greening and browning trends significant?” In: *Geophysical Research Letters* 48.6, e2020GL091496. ISSN: 0094-8276.
- Cremer, Felix et al. (2020). “Potential of Recurrence Metrics from Sentinel-1 Time Series for Deforestation Mapping”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, pp. 5233–5240. ISSN: 1939-1404.
- Defriez, Emma J et al. (2016). “Climate change-related regime shifts have altered spatial synchrony of plankton dynamics in the North Sea”. In: *Global Change Biology* 22.6, pp. 2069–2080. ISSN: 1354-1013.
- Duveiller, Gregory et al. (2020). “A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity”. In: *Earth System Science Data* 12.2, pp. 1101–1116. ISSN: 1866-3508.
- Eleftheratos, K et al. (2011). “Interannual variability of cirrus clouds in the tropics in El Niño Southern Oscillation (ENSO) regions based on International Satellite Cloud Climatology Project (ISCCP) satellite data”. In: *International Journal of Remote Sensing* 32.21, pp. 6395–6405.
- Eltahir, Elfatih A B and Rafael L Bras (1994). “Precipitation recycling in the Amazon basin”. In: *Quarterly Journal of the Royal Meteorological Society* 120.518, pp. 861–880. ISSN: 0035-9009.
- Erfanian, Amir, Guiling Wang, and Lori Fomenko (2017). “Unprecedented drought over tropical South America in 2016 : significantly under-predicted by tropical SST”. In: *Scientific Reports* October 2016, pp. 1–11. DOI: 10.1038/s41598-017-05373-2.
- ESA (2017). *Land Cover CCI Product User Guide Version 2*. URL: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2{_}2.0.pdf.

- Estupinan-Suarez, L. M., Olga Adiana Leon, and Carlos E. Sarmiento Pinzon (2017). “Disturbance Regime, section 2.6.” In: *A Sourcebook of Methods and Procedures for Monitoring Essential Biodiversity Variables in Tropical Forests with Remote Sensing*. Ed. by M Gill et al. GOF-C-GOLD Land Cover Project Office, Wageningen University, The Netherlands.
- Estupinan-Suarez, Lina M et al. (2021). “A Regional Earth System Data Lab for Understanding Ecosystem Dynamics: An Example from Tropical South America”. In: *Frontiers in Earth Science* 9, p. 574.
- Fasullo, J T, B L Otto-Bliesner, and S Stevenson (2018). “ENSO’s Changing Influence on Temperature, Precipitation, and Wildfire in a Warming Climate”. In: *Geophysical Research Letters* 45.17, pp. 9216–9225.
- Giraldo, Jorge A et al. (2020). “Dendrochronological Potential of Trees from America’s Rainiest Region BT - Latin American Dendroecology: Combining Tree-Ring Sciences and Ecology in a Megadiverse Territory”. In: *Latin American Dendroecology*. Ed. by Marín Pompa-García and J Julio Camarero. Cham: Springer International Publishing. Chap. Chapter 5, pp. 79–119.
- Giuliani, Gregory et al. (2017). “Building an earth observations data cube: Lessons learned from the Swiss data cube (SDC) on generating analysis ready data (ARD)”. In: *Big Earth Data* 1.1-2, pp. 100–117. ISSN: 2096-4471.
- Giuliani, Gregory et al. (2019). “Earth observation open science: Enhancing reproducible science using data cubes”. In: *Data* 4.4, p. 147.
- González-M, Roy et al. (2021). “Diverging functional strategies but high sensitivity to an extreme drought in tropical dry forests”. In: *Ecology Letters* 24.3, pp. 451–463. ISSN: 1461-023X.
- Gorelick, Noel et al. (2017). “Google Earth Engine: Planetary-scale geospatial analysis for everyone”. In: *Remote Sensing of Environment* 202, pp. 18–27. ISSN: 0034-4257.
- Graham, Eric A et al. (2003). “Cloud cover limits net CO₂ uptake and growth of a rainforest tree during tropical rainy seasons”. In: *Proceedings of the National Academy of Sciences* 100.2, 572 LP –576.
- Guanter, Luis et al. (2012). “Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements”. In: *Remote Sensing of Environment* 121, pp. 236–251. ISSN: 0034-4257.
- Hashimoto, Hirofumi et al. (2021). “New generation geostationary satellite observations support seasonality in greenness of the Amazon evergreen forests”. In: *Nature communications* 12.1, pp. 1–11. ISSN: 2041-1723.
- Hilker, Thomas et al. (2012). “Remote sensing of tropical ecosystems: Atmospheric correction and cloud masking matter”. In: *Remote Sensing of Environment* 127, pp. 370–384. ISSN: 0034-4257.

- Hilker, Thomas et al. (2014). "Vegetation dynamics and rainfall sensitivity of the Amazon". In: *Proceedings of the National Academy of Sciences* 111.45, pp. 16041–16046.
- Holmgren, Milena et al. (2013). "Effects of interannual climate variability on tropical tree cover". In: *Nature Climate Change* 3.8, pp. 755–758. ISSN: 1758-678X. DOI: 10.1038/nclimate1906. URL: <http://dx.doi.org/10.1038/nclimate1906>.
- Huang, Norden E and Zhaohua Wu (2008). "A review on Hilbert-Huang transform: Method and its applications to geophysical studies". In: *Reviews of geophysics* 46.2. ISSN: 8755-1209.
- Huang, Norden E et al. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". In: *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454.1971, pp. 903–995. ISSN: 1364-5021.
- Huete, A R et al. (1997). "A comparison of vegetation indices over a global set of TM images fo EOS MODIS". In: *Remote Sensing of Environment* 59.440-451.
- Huete, Alfredo et al. (2002). "Overview of the radiometric and biophysical performance of the MODIS vegetation indices". In: *Remote Sensing of Environment* 83.1-2, pp. 195–213. ISSN: 0034-4257.
- Jarvis, P G (1976). "The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field". In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 273.927, pp. 593–610. ISSN: 0080-4622.
- Jiang, Chongya and Youngryel Ryu (2016). "Remote Sensing of Environment Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS)". In: *Remote Sensing of Environment* 186, pp. 528–547. ISSN: 0034-4257.
- Jolliffe, Ian T and Jorge Cadima (2016). "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065, p. 20150202. ISSN: 1364-503X.
- Jung, Martin et al. (2011). "Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations". In: *Journal of Geophysical Research: Biogeosciences* 116.G3. ISSN: 0148-0227.
- Jung, Martin et al. (2020). "Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach". In: *Biogeosciences* 17.5, pp. 1343–1365. ISSN: 1726-4170.
- Junk, Wolfgang J, Peter B Bayley, and Richard E Sparks (1989). "The flood pulse concept in river-floodplain systems". In: *Canadian special publication of fisheries and aquatic sciences* 106.1, pp. 110–127.

- Junk, Wolfgang J et al. (2014). “Brazilian wetlands: their definition, delineation, and classification for research, sustainable management, and protection”. In: *Aquatic Conservation: marine and freshwater ecosystems* 24.1, pp. 5–22. ISSN: 1052-7613.
- Keeling, Charles D et al. (1976). “Atmospheric carbon dioxide variations at Mauna Loa observatory, Hawaii”. In: *Tellus* 28.6, pp. 538–551. ISSN: 0040-2826.
- Knoben, Wouter J.M., Ross A. Woods, and Jim E. Freer (2018). “Global bimodal precipitation seasonality : A systematic overview”. In: *International Journal of Climatology* June, pp. 1–10.
- Knyazikhin, Y (1999). “MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD 15) algorithm theoretical basis document”. In: <http://eosps0.gsfc.nasa.gov/atbd/modistabls.html>.
- Knyazikhin, Yuri et al. (1998). “Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data”. In: *Journal of Geophysical Research: Atmospheres* 103.D24, pp. 32257–32275. ISSN: 0148-0227.
- Kogan, Felix and Wei Guo (2017). “Strong 2015 – 2016 El Niño and implication to global ecosystems from space data”. In: *International Journal of Remote Sensing* 38.1, pp. 161–178. ISSN: 13665901.
- Köhler, Philipp et al. (2018). “Assessing the potential of sun-induced fluorescence and the canopy scattering coefficient to track large-scale vegetation dynamics in Amazon forests”. In: *Remote Sensing of Environment* 204, pp. 769–785. ISSN: 0034-4257.
- Koren, Gerbrand et al. (2018). “Widespread reduction in sun-induced fluorescence from the Amazon during the 2015/2016 El Niño”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1760, p. 20170408. ISSN: 0962-8436.
- Kraemer, Guido, Markus Reichstein, and Miguel D Mahecha (2018). “dimRed and coRanking—unifying dimensionality reduction in R”. In: *R Journal* 10.1, pp. 342–358. ISSN: 2073-4859.
- Kraemer, Guido et al. (2020). “Summarizing the state of the terrestrial biosphere in few dimensions”. In: *Biogeosciences* 17.9, pp. 2397–2424. ISSN: 1726-4170.
- Küçük, Çağlar et al. (2022). “Characterizing the Response of Vegetation Cover to Water Limitation in Africa Using Geostationary Satellites”. In: *Journal of Advances in Modeling Earth Systems* 14.3, e2021MS002730.
- Lee, Jung-Eun et al. (2013). “Forest productivity and water stress in Amazonia: Observations from GOSAT chlorophyll fluorescence”. In: *Proceedings of the Royal Society B: Biological Sciences* 280.1761, p. 20130171. ISSN: 0962-8452.
- Lewis, Adam et al. (2017). “The Australian geoscience data cube foundations and lessons learned”. In: *Remote Sensing of Environment* 202, pp. 276–292. ISSN: 0034-4257.

- Li, Qian et al. (2021). "Remote sensing of seasonal climatic constraints on leaf phenology across pantropical evergreen forest biome". In: *Earth's Future* 9.9, e2021EF002160. ISSN: 2328-4277.
- Liu, Junjie et al. (2017). "Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño". In: *Science* 358.6360, eaam5690.
- Lopes, Aline Pontes et al. (2016). "Leaf flush drives dry season green-up of the Central Amazon". In: *Remote Sensing of Environment* 182, pp. 90–98. ISSN: 00344257. arXiv: arXiv:1011.1669v3.
- Luo, Xiangzhong et al. (2018). "The impact of the 2015 / 2016 El Niño on global photosynthesis using satellite remote sensing". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1760.
- Madden, Roland A and Paul R Julian (1971). "Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific". In: *Journal of the atmospheric sciences* 28.5, pp. 702–708. ISSN: 1520-0469.
- Mahecha, Miguel D et al. (2010). "Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales". In: *Journal of Geophysical Research: Biogeosciences* 115.G2. ISSN: 0148-0227.
- Mahecha, Miguel D et al. (2020). "Earth system data cubes unravel global multivariate dynamics". In: 1, pp. 201–234.
- Manabe, Syukuro and Wetherald (1967). "Thermal equilibrium of the atmosphere with a given distribution of relative humidity". In: *Journal of Atmospheric Sciences* 24.3, pp. 242–259.
- Moura, Marks Melo et al. (2019). "Relation of El Niño and La Niña phenomena to precipitation, evapotranspiration and temperature in the Amazon basin". In: *Science of The Total Environment* 651, pp. 1639–1651. ISSN: 0048-9697.
- Muenchow, Jannes et al. (2013). "Coupling ordination techniques and GAM to spatially predict vegetation assemblages along a climatic gradient in an ENSO-affected region of extremely high climate variability". In: *Journal of vegetation science* 24.6, pp. 1154–1166. ISSN: 1100-9233.
- Muenchow, Jannes et al. (2020). "Monitoring and predictive mapping of floristic biodiversity along a climatic gradient in ENSO's terrestrial core region, NW Peru". In: *Ecography* 43.12, pp. 1878–1890. ISSN: 0906-7590.
- Myers, Norman et al. (2000). "Biodiversity hotspots for conservation priorities". In: *Nature* 403.February, pp. 853–858.
- Myneni, Ranga B et al. (1997). "Estimation of global leaf area index and absorbed PAR using radiative transfer models". In: *IEEE Transactions on Geoscience and remote sensing* 35.6, pp. 1380–1393. ISSN: 0196-2892.

- Nemani, Ramakrishna R et al. (2003). “Climate-driven increases in global terrestrial net primary production from 1982 to 1999”. In: *Science* 300.5625, pp. 1560–1563. ISSN: 0036-8075.
- NOAA (2021a). *Global impacts of El Niño and La Niña*. Tech. rep. <https://www.climate.gov/>. URL: <https://www.climate.gov/>.
- (2021b). *Multivariate ENSO Index Version 2 (MEI.v2)*. Physical Sciences Laboratory. Tech. rep. <https://psl.noaa.gov/enso/mei/>. URL: <https://psl.noaa.gov/enso/mei/>.
- ODC (2021). *Open Data Cube. Documentation on Data Loading. Limitations and Problems*. URL: https://datacube-core.readthedocs.io/en/latest/architecture/data_{_}loading.html?highlight=dataloading{#\}limitations-and-problems (visited on 03/23/2021).
- Olson, David M et al. (2001). “Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity”. In: *BioScience* 51.11, pp. 933–938. ISSN: 0006-3568.
- Pan, Yude et al. (2011). “A large and persistent carbon sink in the world’s forests”. In: *Science* 333.6045, pp. 988–993. ISSN: 0036-8075.
- Pandey, Sudhanshu et al. (2017). “Enhanced methane emissions from tropical wetlands during the 2011 La Niña”. In: *Nature Scientific Reports* 7.45759, pp. 1–8.
- Parsons, L A et al. (2018). “The threat of multi-year drought in western Amazonia”. In: *Water resources research* 54.9, pp. 5890–5904. ISSN: 0043-1397.
- Patra, Prabir K et al. (2017). “The Orbiting Carbon Observatory (OCO-2) tracks 2–3 peta-gram increase in carbon release to the atmosphere during the 2014–2016 El Niño”. In: *Scientific reports* 7.1, pp. 1–12. ISSN: 2045-2322.
- Pearson, Karl (1901). “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11, pp. 559–572. ISSN: 1941-5982.
- Pennington, R Toby, Caroline E R Lehmann, and Lucy M Rowland (2018). “Tropical savannas and dry forests.” eng. In: *Current biology : CB* 28.9, R541–R545. ISSN: 1879-0445 (Electronic). DOI: 10.1016/j.cub.2018.03.014.
- Philander, S George (2018). “9. El Niño, La Niña, and the Southern Oscillation”. In: *Is the Temperature Rising?: The Uncertain Science of Global Warming*. Princeton University Press, pp. 143–158.
- Piao, Shilong et al. (2020). “Characteristics, drivers and feedbacks of global greening”. In: *Nature Reviews Earth & Environment* 1.1, pp. 14–27. ISSN: 2662-138X.
- Pinzon, Jorge E. and Compton J Tucker (2014). “A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series”. In: *Remote Sensing* 6, pp. 6929–6960. DOI: 10.3390/rs6086929.

- Poveda, G, Peter R Waylen, and Roger S Pulwarty (2006). “Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica”. In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 234, pp. 3–27.
- Poveda, Germán and Oscar J Mesa (2000). “On the existence of Lloró (the rainiest locality on Earth): Enhanced ocean-land-atmosphere interaction by a low-level jet”. In: *Geophysical Research Letters* 27.11, pp. 1675–1678. ISSN: 0094-8276.
- Poveda, Germán et al. (2001). “Seasonally in ENSO-related precipitation, river discharges, soil moisture, and vegetation index in Colombia”. In: *Water resources research* 37.8, pp. 2169–2178. ISSN: 0043-1397.
- Roweis, Sam T and Lawrence K Saul (2000). “Nonlinear dimensionality reduction by locally linear embedding”. In: *science* 290.5500, pp. 2323–2326. ISSN: 0036-8075.
- Running, S., Q. Mu, and M. Zhao (2015). *MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006 [Data set]*. URL: <https://doi.org/10.5067/MODIS/MOD17A2H.006> (visited on 03/09/2022).
- Ryu, Youngryel, Joseph A Berry, and Dennis D Baldocchi (2019). “What is global photosynthesis? History, uncertainties and opportunities”. In: *Remote Sensing of Environment* 223, pp. 95–114. ISSN: 0034-4257.
- Ryu, Youngryel et al. (2011). “Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales”. In: *Global Biogeochemical Cycles* 25.4, pp. 1–24. ISSN: 08866236.
- Salas, Hernán D et al. (2020). “Generalized synchronization between ENSO and hydrological variables in Colombia: A recurrence quantification approach”. In: *Frontiers in Applied Mathematics and Statistics* 6, p. 3. ISSN: 2297-4687.
- Schaik, Erik van et al. (2018). “Changes in surface hydrology, soil moisture and gross primary production in the Amazon during the 2015/2016 El Niño”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1760, p. 20180084. ISSN: 0962-8436.
- Schmit, Timothy J et al. (2018). “Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI)”. In: *J. Operational Meteor.* 6.4, pp. 33–46.
- Sellers, P J et al. (1997). “Modeling the exchanges of energy, water, and carbon between continents and the atmosphere”. In: *Science* 275.5299, pp. 502–509. ISSN: 0036-8075.
- Seymour, Frances and Nancy L Harris (2019). “Reducing tropical deforestation”. In: *Science* 365.6455, pp. 756–757. ISSN: 0036-8075.
- Shiogama, Hideo et al. (2011). “Observational constraints indicate risk of drying in the Amazon basin”. In: *Nature Communications* 2.1, pp. 1–7. ISSN: 2041-1723.

- Siyum, Zenebe Girmay (2020). "Tropical dry forest dynamics in the context of climate change: syntheses of drivers, gaps, and management perspectives". In: *Ecological Processes* 9.1, pp. 1–16. ISSN: 2192-1709.
- Song, Xiao-Peng et al. (2018). "Global land change from 1982 to 2016". In: *Nature*. ISSN: 0028-0836.
- Stan, Kayla and Arturo Sanchez-Azofeifa (2019). "Tropical dry forest diversity, climatic response, and resilience in a changing climate". In: *Forests* 10.5, p. 443.
- Stoy, Paul C et al. (2009). "Biosphere-atmosphere exchange of CO₂ in relation to climate: a cross-biome analysis across multiple time scales". In: *Biogeosciences* 6.10, pp. 2297–2312. ISSN: 1726-4170.
- Tamiminia, Haifa et al. (2020). "Google Earth Engine for geo-big data applications: A meta-analysis and systematic review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 164, pp. 152–170. ISSN: 0924-2716.
- Tedeschi, Renata G, Alice M Grimm, and Iracema F A Cavalcanti (2016). "Influence of Central and East ENSO on precipitation and its extreme events in South America during austral autumn and winter". In: *International Journal of Climatology* 36.15, pp. 4797–4814. ISSN: 0899-8418.
- Tenenbaum, Joshua B, Vin De Silva, and John C Langford (2000). "A Global Geometric Framework for Nonlinear Dimensionality Reduction". In: *Science* 290.December, pp. 2319–2324.
- Tramontana, Gianluca et al. (2016). "Supplement of Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms". In: *Biogeosciences* 13, pp. 4291–4313.
- Tucker, C J and P J Sellers (1986). "Satellite remote sensing of primary production". In: *International Journal of Remote Sensing* 7.11, pp. 1395–1416. ISSN: 0143-1161.
- United Nations (2019). "World Population Prospects 2019". In: *Department of Economic and Social Affairs, Population Division* Online Edi.
- Uribe, M R, C A Sierra, and J S Dukes (2021). "Seasonality of Tropical Photosynthesis: A Pantropical Map of Correlations With Precipitation and Radiation and Comparison to Model Outputs". In: *Journal of Geophysical Research: Biogeosciences* 126.11, e2020JG006123. ISSN: 2169-8953.
- Vilanova, Regiane Souza et al. (2021). "Vegetation degradation in ENSO events: Drought assessment, soil use and vegetation evapotranspiration in the Western Brazilian Amazon". In: *Remote Sensing Applications: Society and Environment* 23, p. 100531.
- Wang, Jun, Ning Zeng, and Meirong Wang (2016). "Interannual variability of the atmospheric CO₂ growth rate: Roles of precipitation and temperature". In: *Biogeosciences* 13.8, pp. 2339–2352. ISSN: 17264189. DOI: 10.5194/bg-13-2339-2016.

- Waylen, Peter and Germán Poveda (2002). “El Niño–Southern Oscillation and aspects of western South American hydro-climatology”. In: *Hydrological Processes* 16.6, pp. 1247–1260. ISSN: 0885-6087.
- Yang, Xi et al. (2015). “Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest”. In: *Geophysical Research Letters* 42.8, pp. 2977–2987. ISSN: 0094-8276.
- Yepes, Johanna et al. (2019). “Choco-jex: A research experiment focused on the Chocó low-level jet over the far eastern Pacific and western Colombia”. In: *Bulletin of the American Meteorological Society* 100.5, pp. 779–796. ISSN: 0003-0007.
- Yun, Kyung-Sook et al. (2021). “Increasing ENSO–rainfall variability due to changes in future tropical temperature–rainfall relationship”. In: *Communications Earth & Environment* 2.1, pp. 1–7. ISSN: 2662-4435.
- Zanella De Arruda, Paulo Henrique et al. (2016). “Large net CO₂ loss from a grass-dominated tropical savanna in south-central Brazil in response to seasonal and interannual drought”. In: *Journal of Geophysical Research: Biogeosciences* 121.8, pp. 2110–2124.
- Zemp, D C et al. (2014). “On the importance of cascading moisture recycling in South America”. In: *Atmospheric Chemistry and Physics* 14.23, pp. 13337–13359.
- Zemp, Delphine Clara et al. (2017a). “Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks”. In: *Nature Communications* 8, pp. 1–10.
- Zemp, Delphine Clara et al. (2017b). “Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks”. In: *Nature communications* 8.1, pp. 1–10.
- Zeng, N., A. Mariotti, and Patrick Wetzel (2005). “Terrestrial mechanisms of interannual CO₂ variability”. In: *Global Biogeochemical Cycles* 19.1, pp. 1–15.

Declaration on authorship and copyright in a cumulative doctoral thesis

Reprint permissions have been obtained for all manuscripts used as a part of the doctoral thesis.

The co-authors of the manuscripts used in the present cumulative doctoral thesis have been informed about the use of the manuscripts and about the declared individual contributions; they have given their consent.

The declared individual contributions of the doctoral candidate and the other doctoral candidates participate as co-authors in the publications are listed in the attachment.

Lina M. Estupinan Suarez

Name of doctoral candidate	Place	Date	Signature
----------------------------	-------	------	-----------

I give my consent to the submission of a cumulative doctoral thesis and confirm the correctness of the information provided above.

Prof. Dr. Alexander Brenning

Name of supervisor	Place	Date	Signature
--------------------	-------	------	-----------

Publication # 1: Towards a global understanding of vegetation–climate dynamics at multiple timescales.

Status: Published in *Biogeosciences*, 17, 945–962, 2020. doi: 10.5194/bg-17-945-2020.

Author	Conceptual research design	Planning of research activities	Data collection	Data analyses and interpretation	Manuscript writing	Suggested publication equivalence value
Lina M. Estupinan-Suarez¹	*	*	*	*	*	1
Nora Linscheid ¹	*	*	*	*	*	
Alexander Brenning					*	
Nuno Carvalhais				*		
Felix Cremer				*		
Fabian Gans				*		
Anja Rammig					*	
Markus Reichstein	*					
Carlos A. Sierra					*	
Miguel D. Mahecha	*	*			*	

¹ This publication has two first authors. Both authors contributed equally. Each of them more than 33%.

Publication # 2: A regional Earth system data lab for understanding ecosystem dynamics: An example from tropical South America.

Status: Published in *Front. Earth Sci.* 9:613395, 2021. doi: 10.3389/feart.2021.613395.

Author	Conceptual research design	Planning of research activities	Data collection	Data analyses and interpretation	Manuscript writing	Suggested publication equivalence value
Lina M. Estupinan-Suarez	*	*	*	*	*	1
Fabian Gans	*		*		*	
Alexander Brenning					*	
Victor H. Gutierrez-Velez					*	
Maria C. Londono					*	
Daniel E. Pabon-Moreno					*	
Germán Poveda					*	
Markus Reichstein					*	
Björn Reu					*	
Carlos A. Sierra	*				*	
Ulrich Weber			*		*	
Miguel D. Mahecha	*			*	*	

Publication # 3:

Status: Submitted.

Author	Conceptual research design	Planning of research activities	Data collection	Data analyses and interpretation	Manuscript writing	Suggested publication equivalence value
Lina M. Estupinan-Suarez	*	*	*	*	*	1
Miguel D. Mahecha	*			*		
Alexander Brenning				*		
Guido Kraemer				*		
German Poveda				*		
Markus Reichstein				*		
Carlos A. Sierra.		*		*	*	

Acknowledgments

I want to express my deepest gratitude to my supervisor Carlos Sierra for his continuous support since I started my PhD, and especially in the last year when his contributions were critical. Thank you for your always constructive comments, the motivation that you gave me and for teaching me a different way of thinking and making science. *¡Muchas gracias Carlos!*

As well, my gratitude to my supervisor Miguel Mahecha for his continuous input of ideas and methods, his scientific support and for helping me to settle in Jena, without these steps my dissertation project would not start. Special thanks to my supervisor Alexander Brenning for his scientific insights, his critical thinking, but also for opening the door of the former Grietgasse to me which was key during the pandemic. I am very grateful to Markus Reichstein for the scientific discussions and support during my dissertation development, his overarching perspective and the opportunity to join the exciting BGI department.

Additionally, many thanks to the Theoretical Ecosystem Ecology group, the GI-Science group and the former Empirical Inference of the Earth System group. The time with all of you has been a great exchange and learning experience.

I am deeply grateful to the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC); Steffi Rothhardt, John Kuhla and Stefanie Burkert. Thank you for supporting me not only to achieve my professional goals, but most of all for caring about the well-being of PhDs, our mental health and for seeking a safe working environment. You made this process safer and supportive. Also, I am quite privileged to finalize the last details of my dissertation from my research stay, after two years of pandemic because of my IMPRS grant and the supportive staff. I am very grateful to the German Academic Exchange Service (DAAD) for funding my PhD and the institutional support.

Special thanks to Fabian Gans for taking the time when needed to help me learning Julia, navigate between software configurations, and conversations about scientific and other cross-cutting topics. Also, for his support in translating the abstract of this dissertation into German. Thanks to Uli Weber for his impeccable, efficient, and always kind assistance in data management.

My great thanks to Tea Thum and Sophia Walter who read and commented on my previous scientific reports, and in particular thanks to Sophia for proofreading this

dissertation. I want to thank Nora Linscheid for such a harmonious team effort, it was a pleasure to collaborate with you. In addition, I like to thank all my co-authors for their insights and comments. As well, I am grateful to my previous mentors and project coordinators, you have guided and contributed to my scientific and personal skills.

I am thankful to Linda Mack, Alice Ratajczak and Sarah Nuschke who have continuously helped me to navigate the daily bureaucracy and have significantly facilitated my life in Jena.

Besides my academic sphere, I like to thank my cycling-mates; Sungmin, Goekben, Naixin and Xuanlong for all the fun during our trips and gatherings. Many thanks to Tea, Sophia, Nora and Fabian who listened to me in both the darkest and happiest moments. My thanks to Daniel, Andrea, Rossana and Giudo for caring about me, and bringing Latin vibes to Jena that made me feel less homesick. A special thanks to my friends of decades; Liz, Dani and Xime who are always by my side even from miles away. I am grateful to all for your friendship and making my life in Jena more enjoyable, happier and smoother, especially during covid times.

Finally, I am very grateful to my family; my mother Nidya, my siblings Milena and Juan Pablo, my aunts Flor and Isaura, for their unconditional and infinite love. Your love and support has been always shown besides distance and pandemic conditions. So many thanks for your support throughout my scientific life even though this often means to be away from you. A special thanks to the Suarez Camacho family for their unlimited warm support. Last, I am very grateful to Martin Schoder for bringing so much joy to my life. Your patience and support always helped me to be stronger and calmer at the same time, especially during turning moments. It is a pleasure to grow up by your side.

Appendix A

Supplementary figures

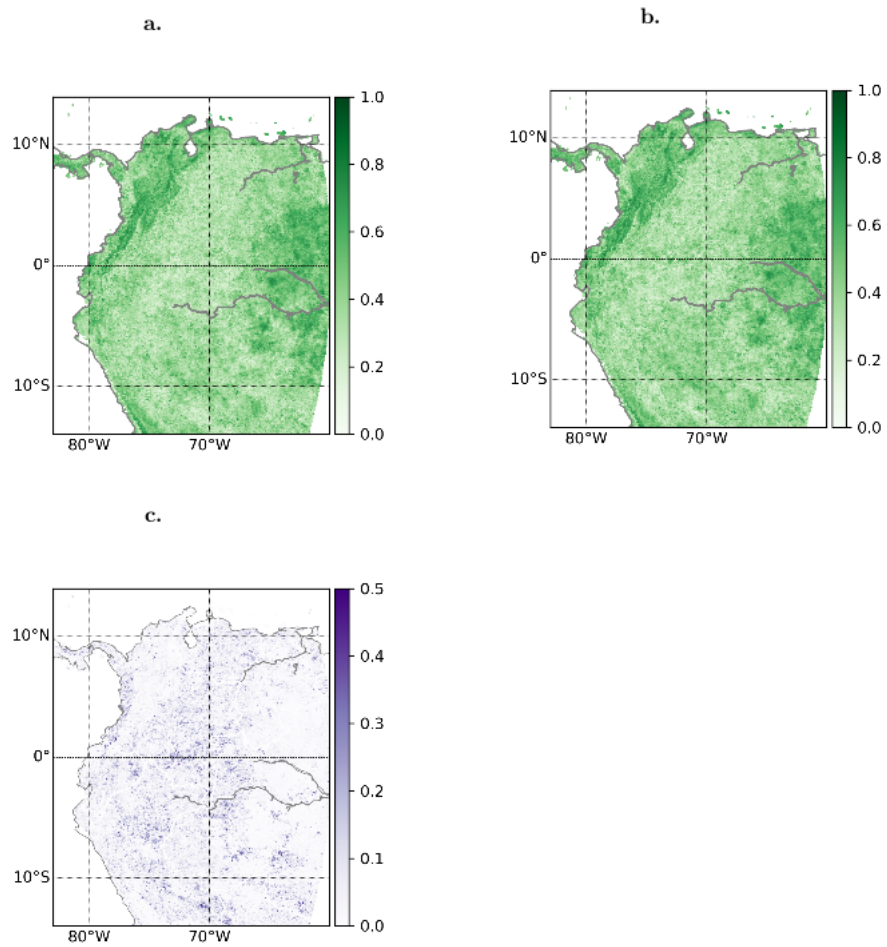


FIGURE A.1: Lagged Pearson correlation between the Multivariate ENSO Index and Isomap (PCA) components pixelwise. The showed value correspond to the highest lagged correlation between MEI and respective components from (a) Isomap and (b) PCA. (c) Difference between Isomap and PCA lagged correlations.

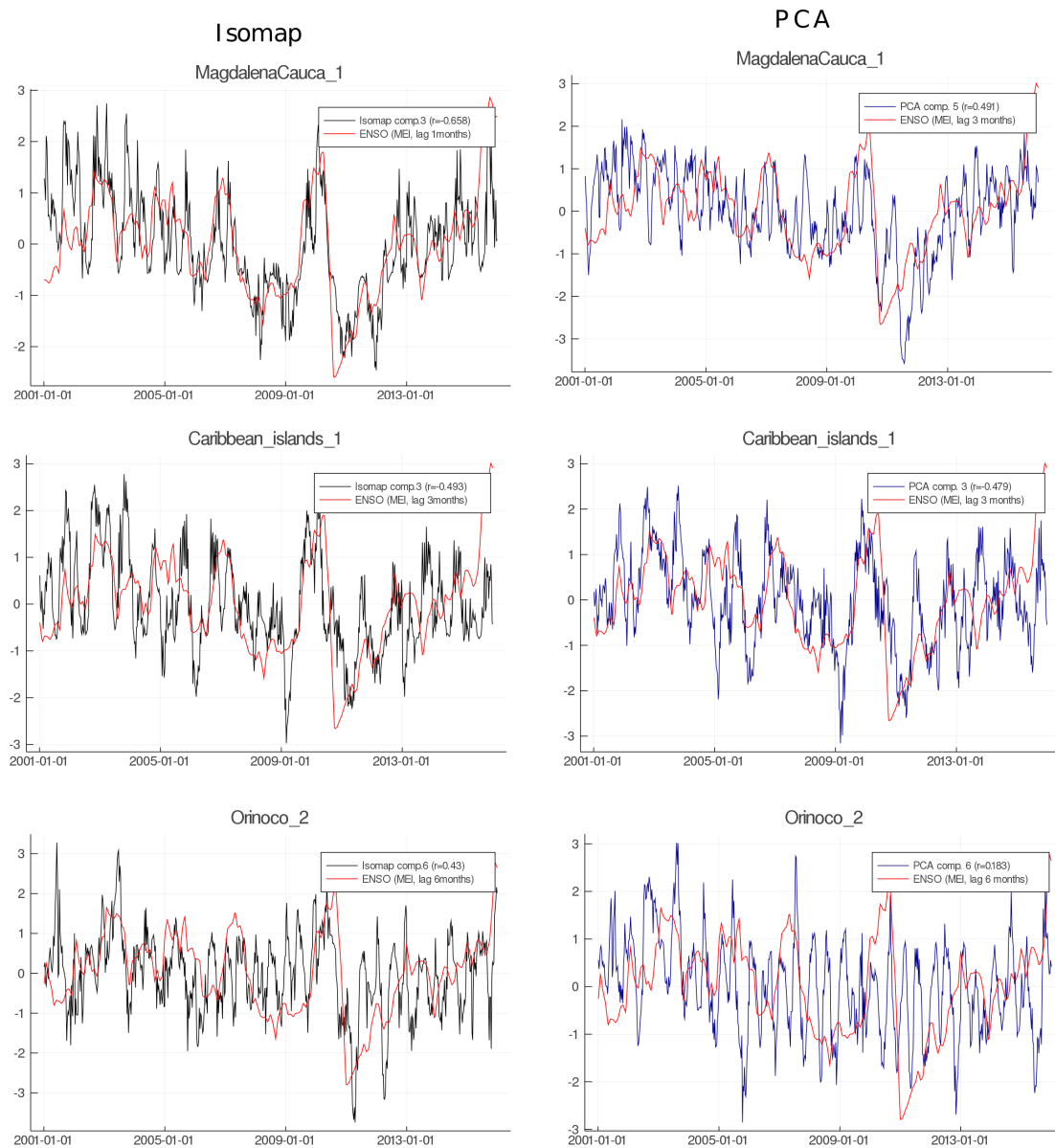


FIGURE A.2: Time series of the Multivariate ENSO Index (MEI) and a selected output component from multidimensionality regression analysis by watersheds. The assessed variable was global primary productivity from BESS. MEI has been standardized to $\mu = 0$ and $\sigma = 1$. The showed component was selected because it has the highest absolute correlation with MEI. Watersheds are from HydroBASINS (level 4), and are in the rows. The dimensionality reduction methods are in the columns. Time is in the x-axis. Each subplot has information about the selected components, correlation value and respective lag in the legend.

Appendix B

Supplement manuscript 1

Supplement of Biogeosciences, 17, 945–962, 2020
<https://doi.org/10.5194/bg-17-945-2020-supplement>
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.



Supplement of

Towards a global understanding of vegetation–climate dynamics at multiple timescales

Nora Linscheid et al.

Correspondence to: Nora Linscheid (nlinsch@bgc-jena.mpg.de) and Lina M. Estupinan-Suarez (lestup@bgc-jena.mpg.de)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

Supplementary Figures

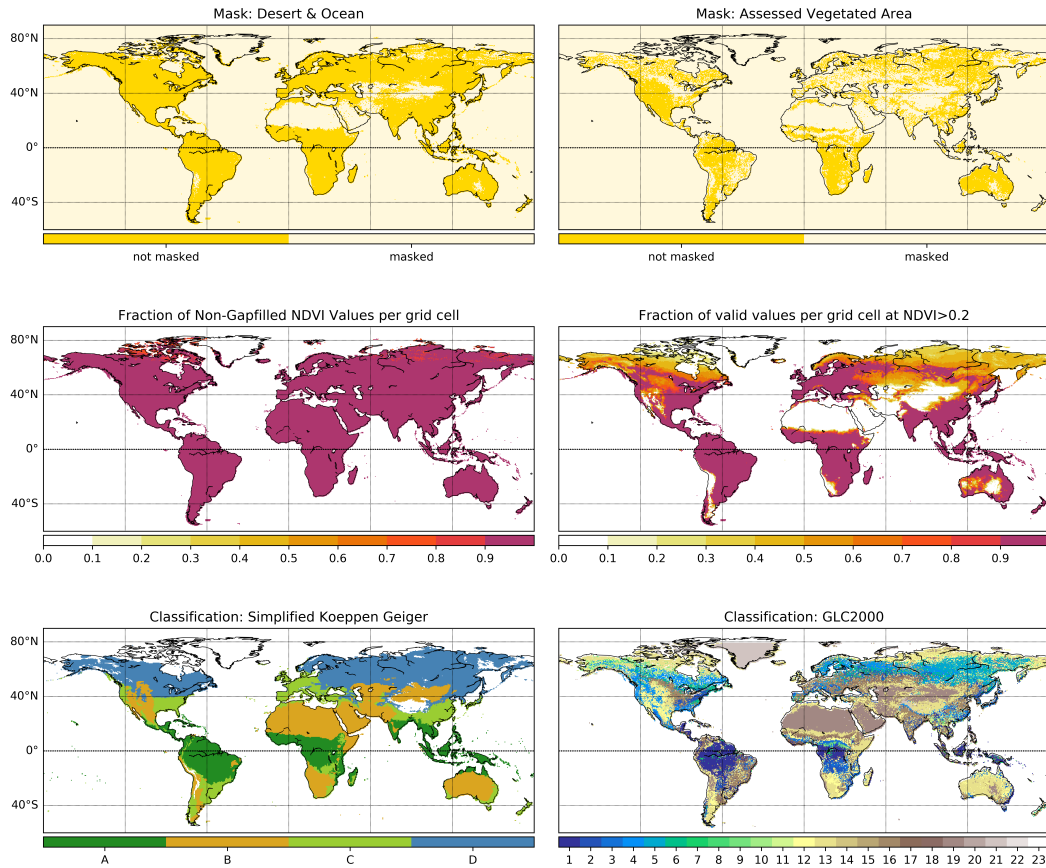


Figure S1: **Masks and Classification schemes used in the analyses.** **a.** Mask for deserts and oceans, **b.** Mask for natural vegetated area based on GLC2000, **c.** Fraction of non-gapfilled NDVI values per grid cell based on GIMMS NDVI, **d.** Fraction of valid values per grid cell after filtering for $NDVI > 0.2$, **e.** Simplified Köppen–Geiger Classification, A: equatorial, B: arid, C: warm temperate, D: snow, **f.** Classification of land cover classes after Global Land Cover 2000 (GLC2000). Numbers from 1–23 represent: 1 – Tree Cover broadleaved evergreen, 2 – Tree Cover broadleaved deciduous closed, 3 – Tree Cover broadleaved deciduous open, 4 – Tree Cover needle leaved evergreen, 5 – Tree Cover needle leaved deciduous, 6 – Tree Cover mixed leaf type, 7 – Tree Cover regularly flooded fresh water, 8 – Tree Cover regularly flooded saline water, 9 – Mosaic: Tree Cover and other natural vegetation, 10 – Tree Cover burnt, 11 – Shrub Cover closed open evergreen, 12 – Shrub Cover closed open deciduous, 13 – Herbaceous Cover closed open, 14 – Sparse herbaceous or sparse shrub cover, 15 – Regularly flooded shrub and or herbaceous cover, 16 – Cultivated and managed areas, 17 – Mosaic: Cropland Tree Cover Other natural vegetation, 18 – Mosaic: Cropland Shrub and or grass cover, 19 – Bare Areas, 20 – Water Bodies, 21 – Snow and Ice, 22 – Artificial surfaces and associated areas, 23 – no data.

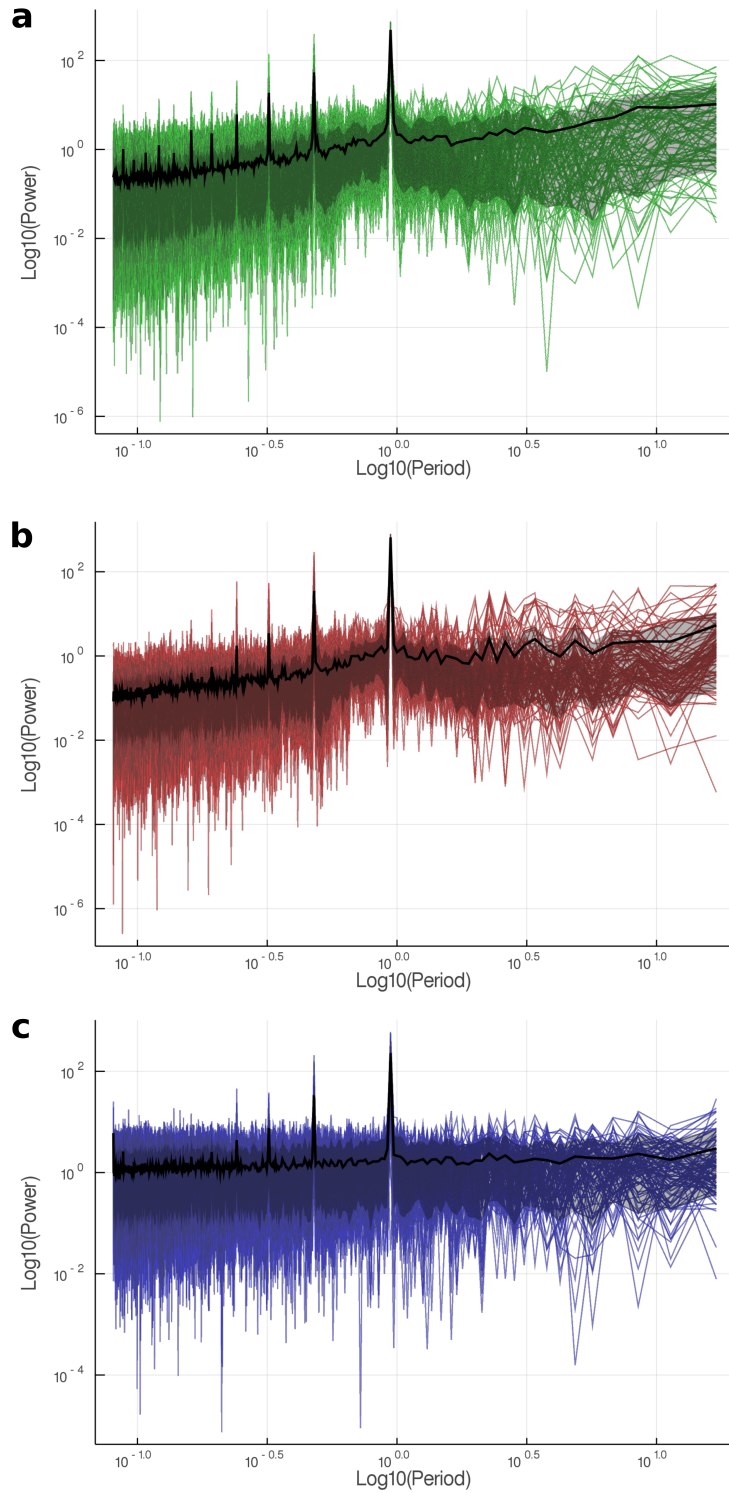


Figure S2: **Representative power spectra of Fourier decomposed NDVI (a), air temperature (b) and precipitation (c) time series.** Mean and 10th–90th percentile of power spectra are plotted as black line (mean) and band (percentiles), overlaid by 10000 sample spectra. Shortest signal periods (fastest frequencies) are plotted on the left side of the x-axis, longest periods on the right side of the x-axis. The annual period is located at 10^0 .

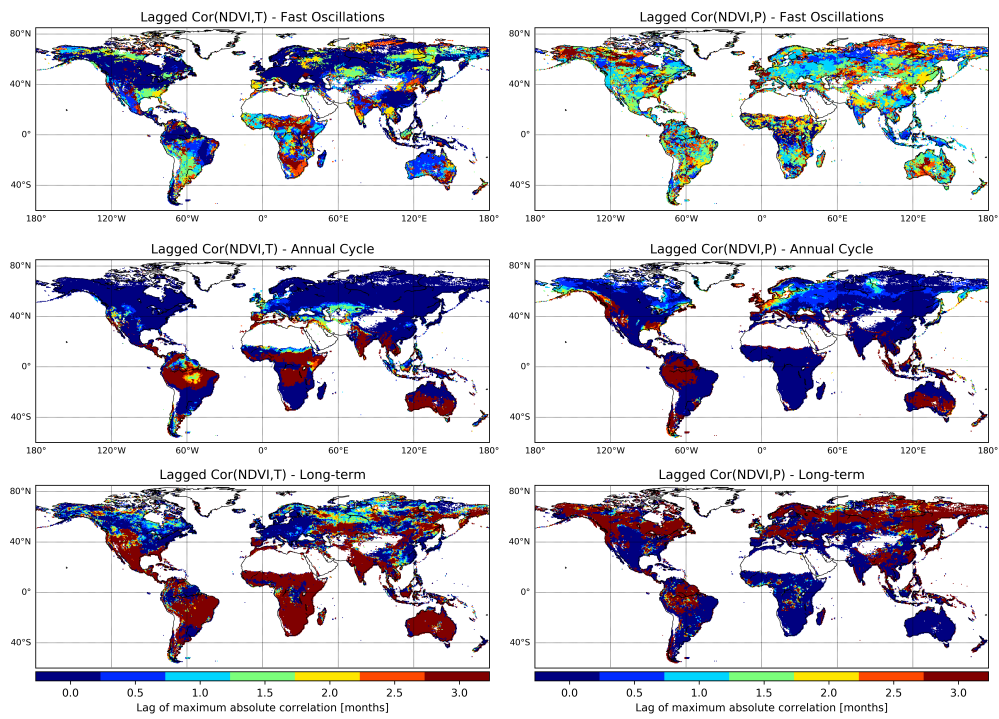


Figure S3: **Lag of maximum absolute correlation** between NDVI and air temperature (T, left panel) and NDVI and precipitation (P, right panel) at each grid cell. The time step used is 15 days, which is equivalent to a 0.5-month lag in the color key.

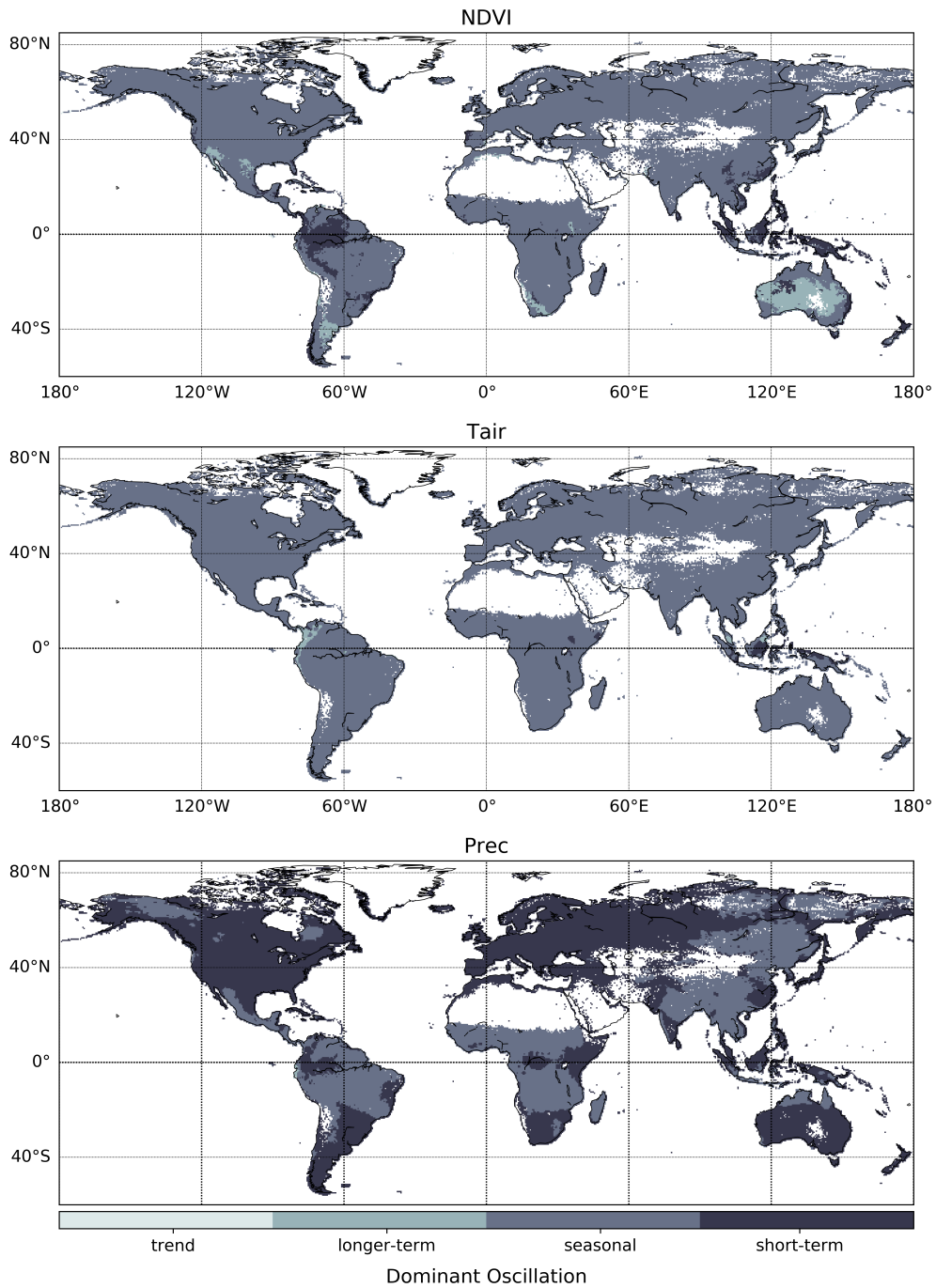


Figure S4: **Dominant Oscillation of NDVI, air temperature (T_{air}) and precipitation (Prec) per grid cell.** Dominant scale of variability was determined from normalized, detrended and decomposed time series as the time scale containing highest relative variance (cf. Fig. 1).

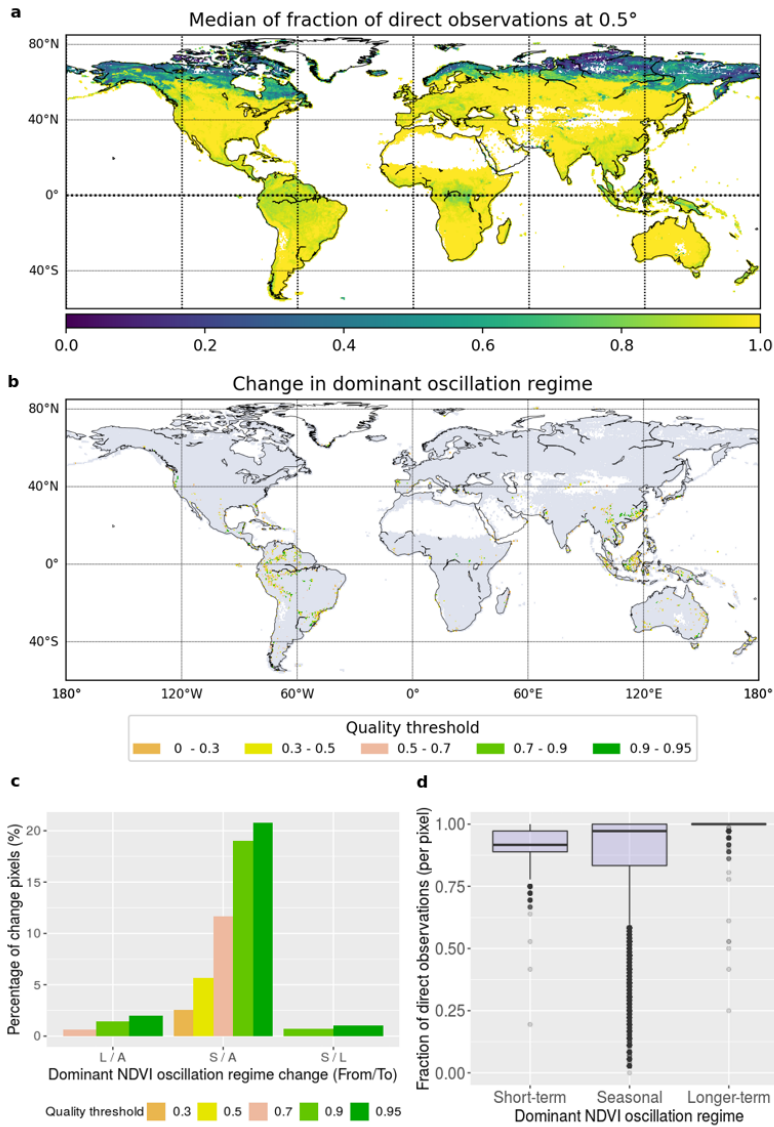


Figure S5: **Assessment of NDVI GIMMS quality flags; direct observations and effect of retrieval values.** **a.** Median of fraction of direct observations at 0.5° per grid cell calculated overall time period (1982–2015). Fraction of direct observations ranges from 0 to 1, and corresponds to the number of pixels with direct observation after data aggregation (from 0.083° to 0.5°). Quality flag 1 is obtained when all aggregated pixels are direct observations, 0 if none are direct observations, **b.** Pixels that change NDVI dominant oscillation class when 0.3, 0.5, 0.7, 0.9, and 0.95 quality threshold is applied (quality is defined as the fraction of pixels originating from direct observations after aggregation), **c.** Percentage of pixels with change per dominant oscillation class. S: Short-term, A: Seasonal, L: Longer-term, T: Trend, in order from / to. Categories with change <0.05% are omitted. **d.** Median fraction originating from direct observation per pixel shown as box plot per oscillation regime. Lowest percentage of direct observation is found in seasonal NDVI regimes.

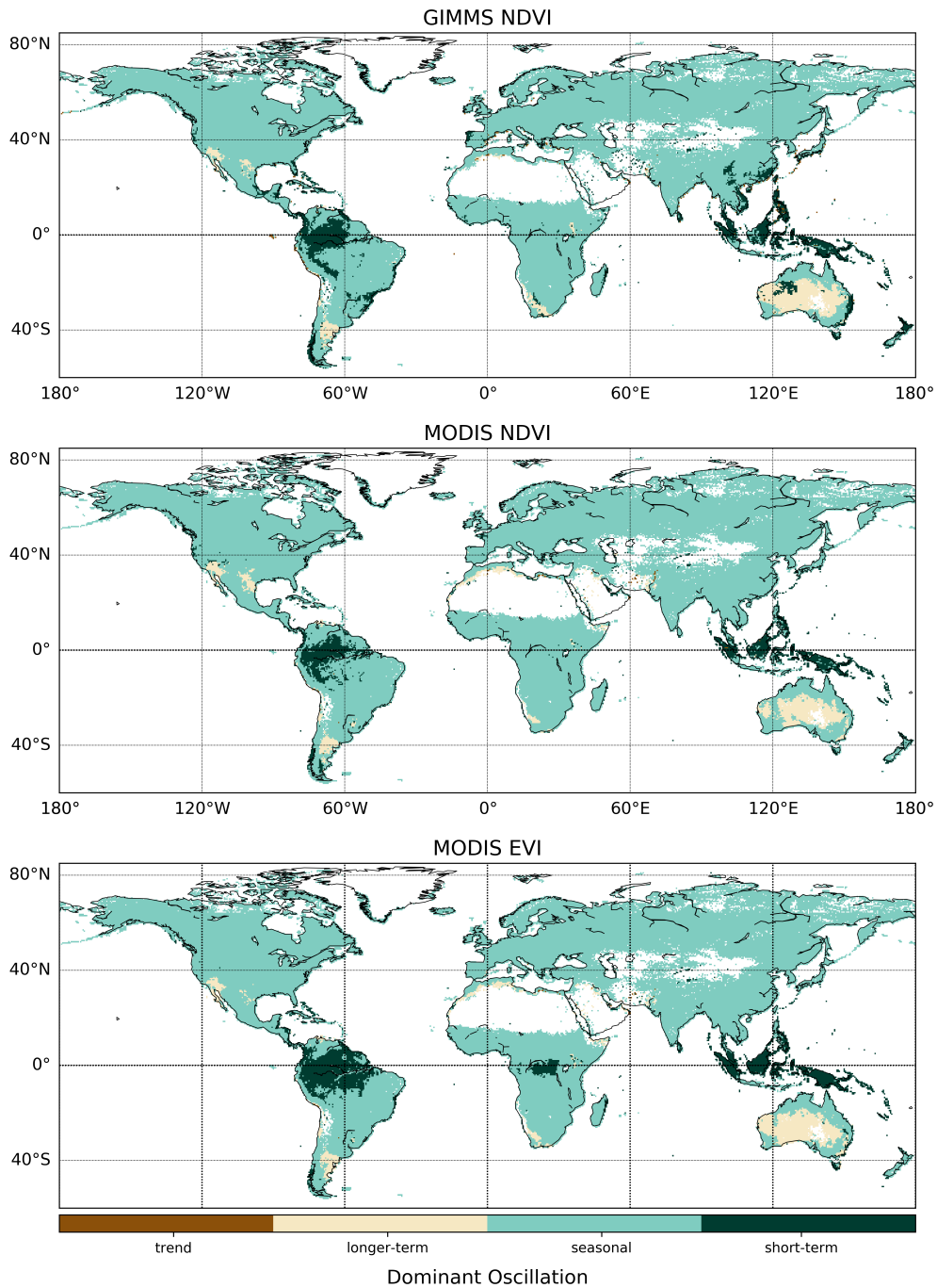


Figure S6: **Comparison of dominant oscillation classification between vegetation indices.** Dominant scale of variability for GIMMS NDVI from 1982 to 2015 (top), MODIS NDVI from 2001 to 2015 (center), and EVI MODIS from 2001 to 2015 (bottom). Dominant scale of variability was determined per pixel from normalized, detrended and Fourier-decomposed time series as the time scale containing highest relative variance.

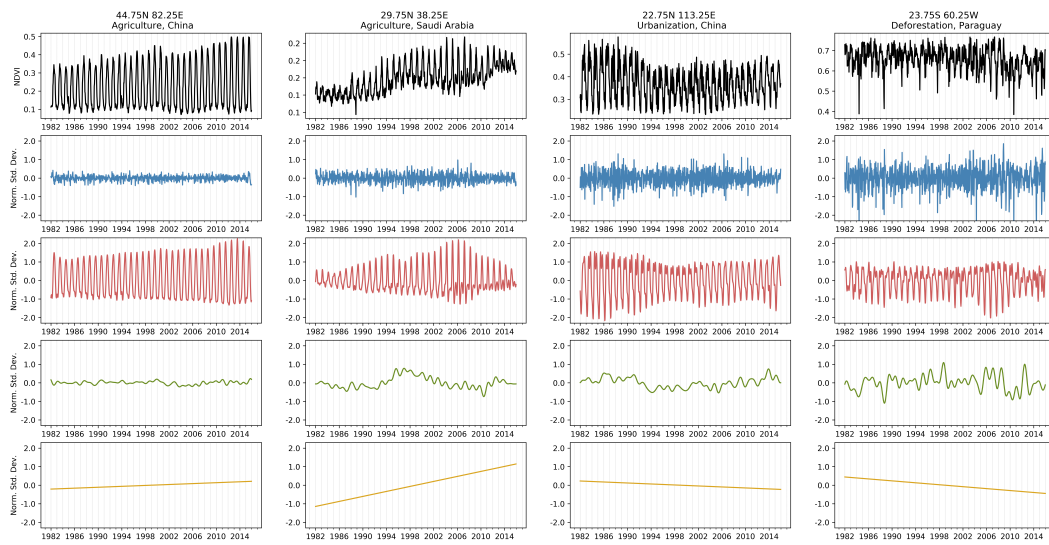


Figure S7: Assessment of the effect of land cover change over time on decomposition results of NDVI time series Four pixels with $>25\%$ change in vegetation type according to Song et al. (2018) are displayed (columns), representing from left to right: (i) short vegetation gain, (ii) bare ground loss, (iii) bare ground gain, and (iv) tree loss. Rows from top to bottom: integrated NDVI signal (black), short-term oscillation (blue), seasonal oscillation (red), longer-term oscillation (green), and trend (yellow). Time series were normalized and detrended before Fourier decomposition.

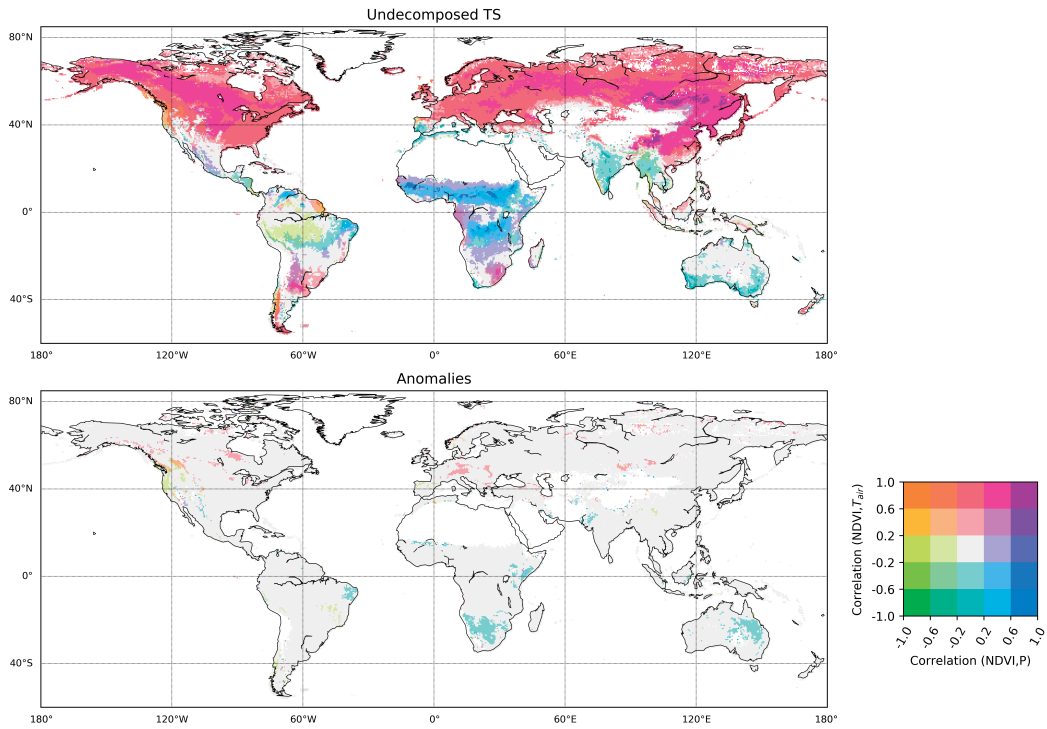


Figure S8: **Bicolor map of undecomposed time series (top) and detrended, deseasonalized anomalies (bottom)**. Pearson correlation of NDVI with precipitation (Prec, legend x axis) and air temperature (T_{air} , legend y axis) is shown at each grid cell. NDVI was lagged one time step (15 days) behind precipitation to allow response time, T_{air} was correlated instantaneously. Color scale represents both correlations, binned into quantiles (e. g. purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec). Data points where NDVI < 0.2 were excluded to avoid influence of inactive vegetation or non-vegetated time points.

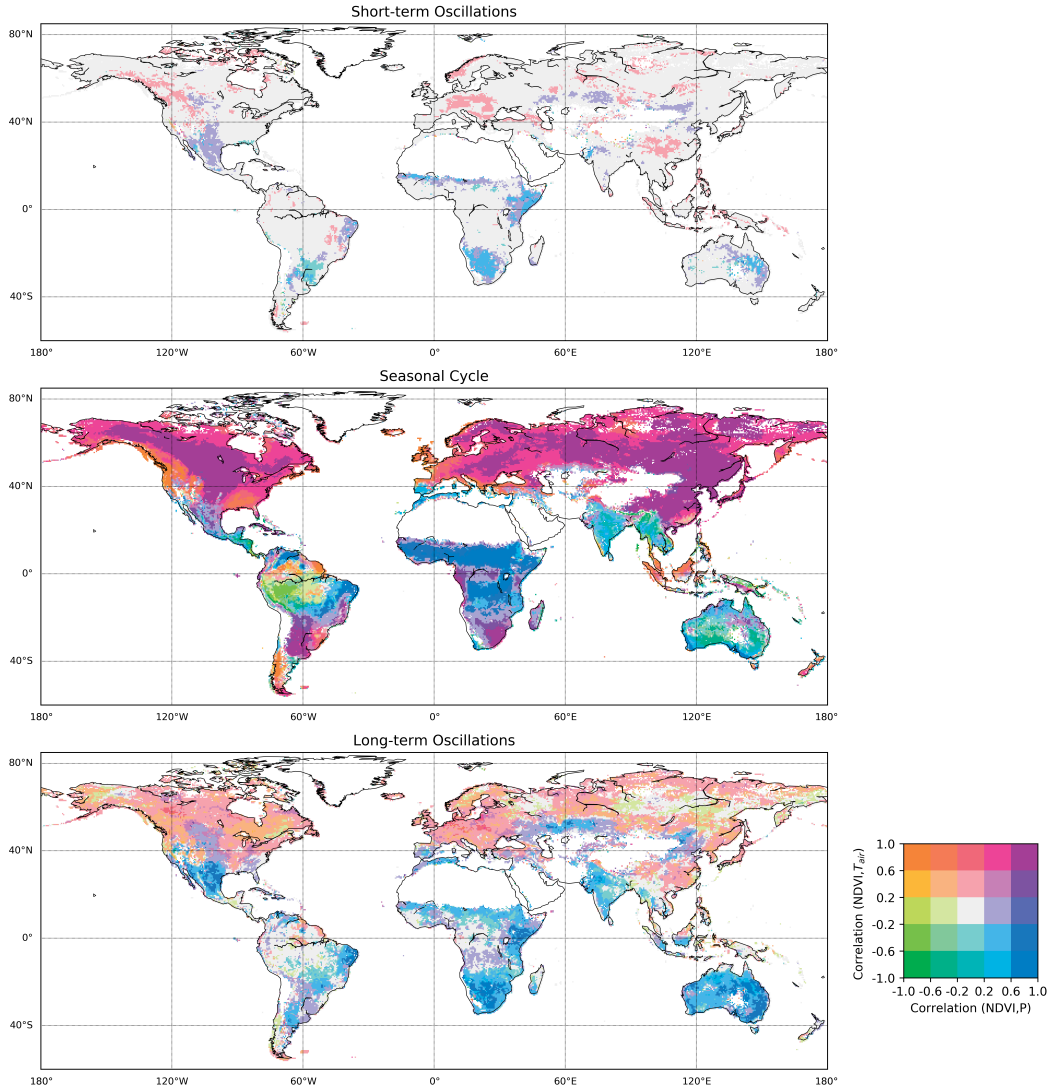


Figure S9: **Bicolor map of Spearman correlations between NDVI, air temperature (T_{air}) and precipitation (Prec).** Correlation of NDVI with T_{air} (legend y axis) and NDVI with Prec (legend x axis) were calculated between decomposed signals at each grid cell for each time scale (rows). NDVI was lagged one time step (15 days) behind precipitation to allow response time, T_{air} was correlated instantaneously. Color scale represents both correlations, binned into quantiles (e. g. purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec). Data points where NDVI < 0.2 were excluded to avoid influence of inactive vegetation or non-vegetated time points. The semi-annual cycle is included in the seasonal band.

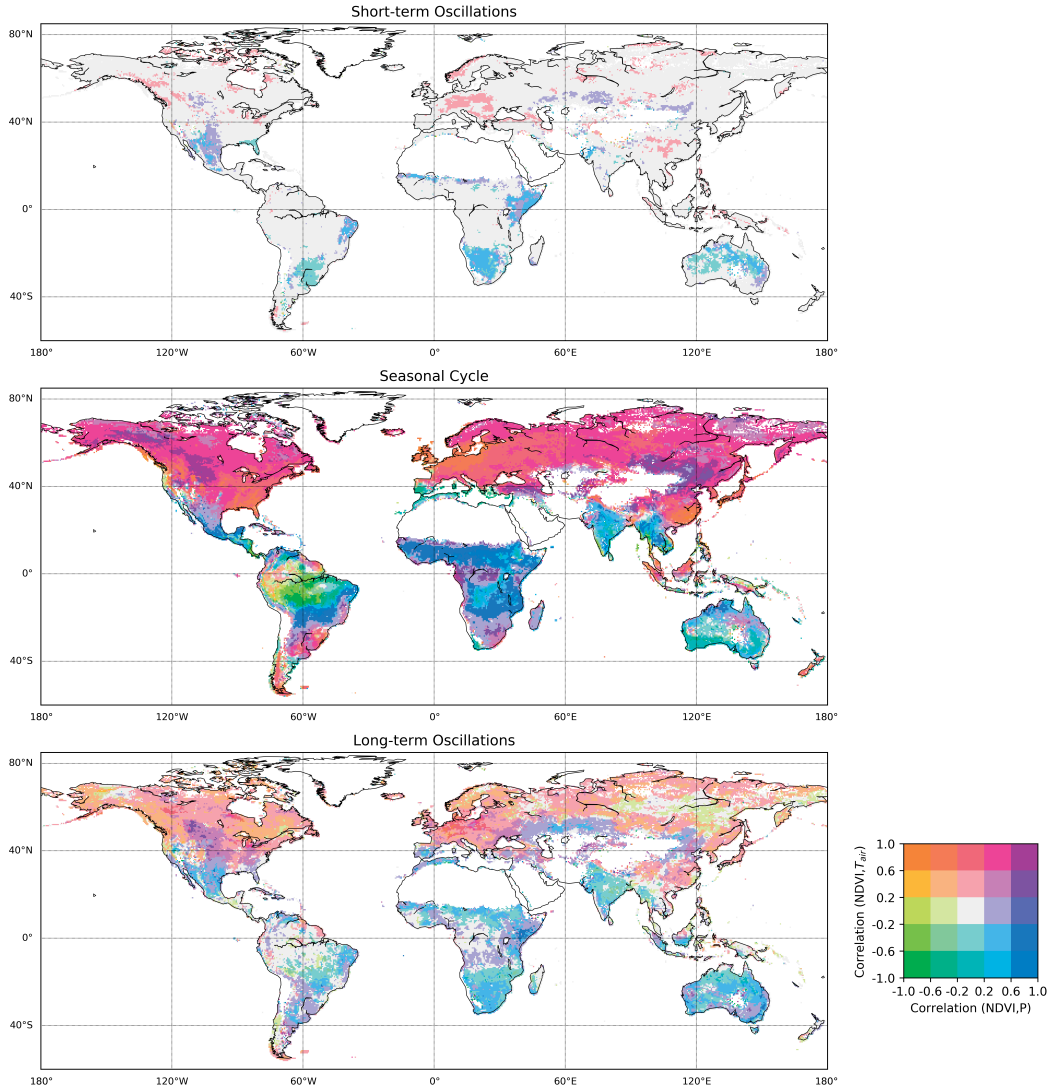


Figure S10: Bicolor map of Partial correlations between NDVI, air temperature (T_{air}) and precipitation (Prec). Correlation of NDVI with T_{air} (legend y axis) and NDVI with Prec (legend x axis) were calculated between decomposed signals at each grid cell for each time scale (rows). NDVI was lagged one time step (15 days) behind precipitation to allow response time, T_{air} was correlated instantaneously. Color scale represents both correlations binned into quantiles (e. g. purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec). Data points where $NDVI < 0.2$ were excluded to avoid influence of inactive vegetation or non-vegetated time points. The semi-annual cycle is included in the seasonal band.

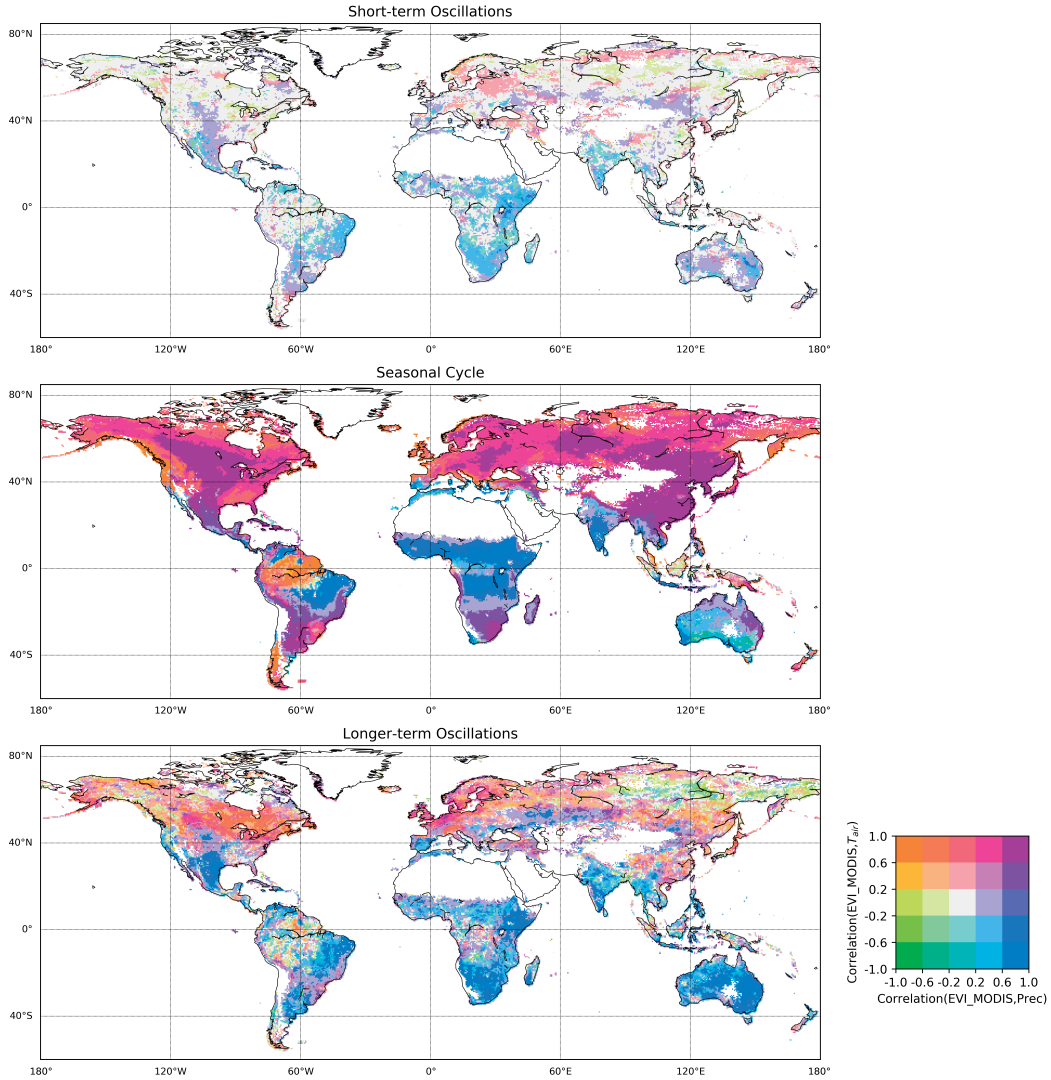


Figure S11: **Bicolor map of Pearson correlations between MODIS EVI, air temperature (T_{air}) and precipitation (Prec).** Correlation of EVI with T_{air} (legend y axis) and MODIS EVI with Prec (legend x axis) were calculated between decomposed signals at each grid cell for each time scale (rows) for the years 2007–2015. EVI was lagged one time step (15 days) behind precipitation to allow response time, T_{air} was correlated instantaneously. Color scale represents both correlations binned into quantiles (e. g. purple – high positive correlation of EVI with both T_{air} and Prec, green – high negative correlation of with both T_{air} and Prec). The semi-annual cycle is included in the seasonal band.

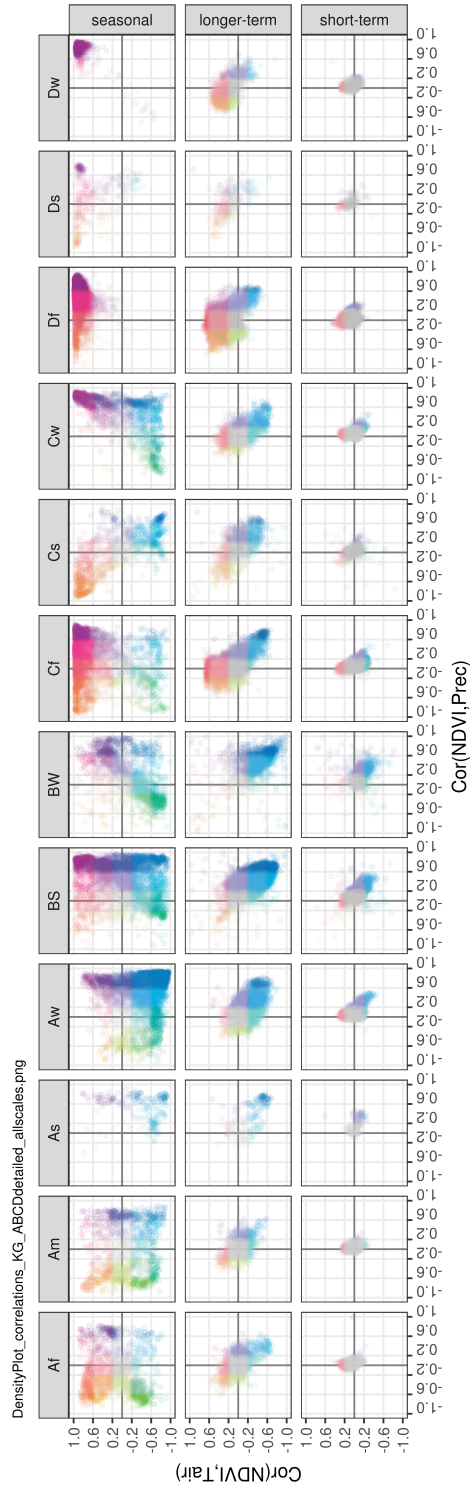


Figure S12: Köppen-Geiger classes in “correlation space” across time scales. Correlations of NDVI with precipitation (x axis) and air temperature (y axis) as determined for Fig. 3 were plotted for each time scale (rows) and Köppen-Geiger class (columns). Each point represents one 0.5° grid cell from the global map. Transparency was scaled with the area represented by the grid cell, colors represent both correlations in accordance with Fig. 3 binned into quantiles (e. g. purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec).

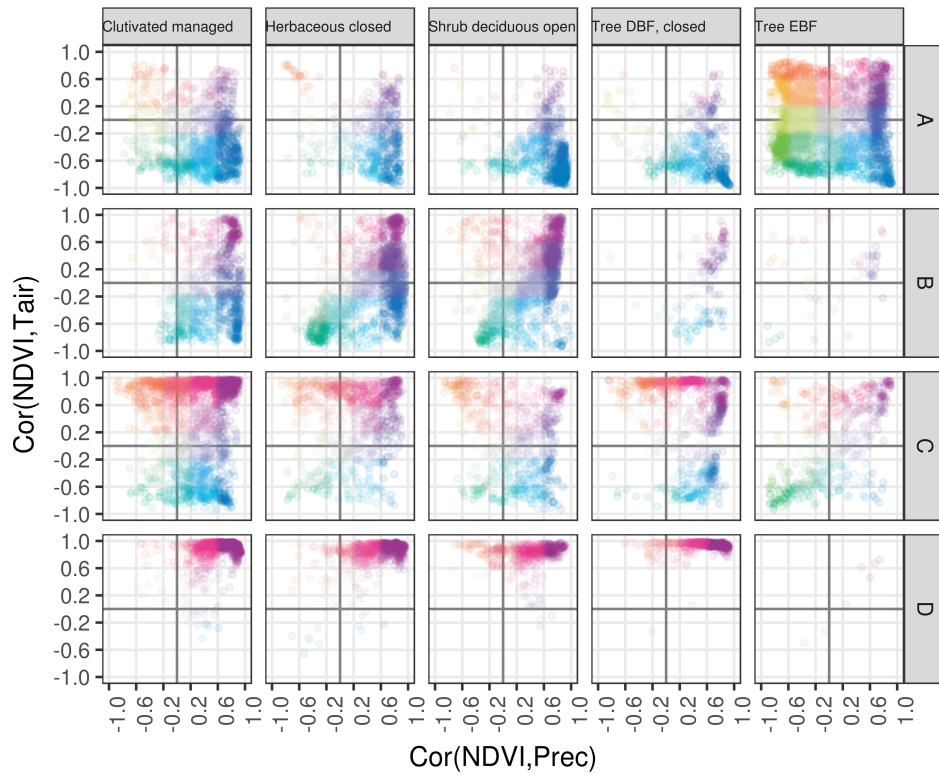


Figure S13: Land cover classes in “correlation space” across Köppen–Geiger classes for seasonal scale. Correlations of NDVI with precipitation (x axis) and air temperature (y axis) as determined for Fig. 3 were plotted for major land cover classes (GLC2000, columns) across Köppen–Geiger classes (rows). Each point represents one 0.5° grid cell from the global map. Transparency was scaled with the area represented by the grid cell, colors represent both correlations in accordance with Fig. 3 binned into quantiles (e. g. purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec). A – equatorial, B – arid, C – warm temperate, D – snow)

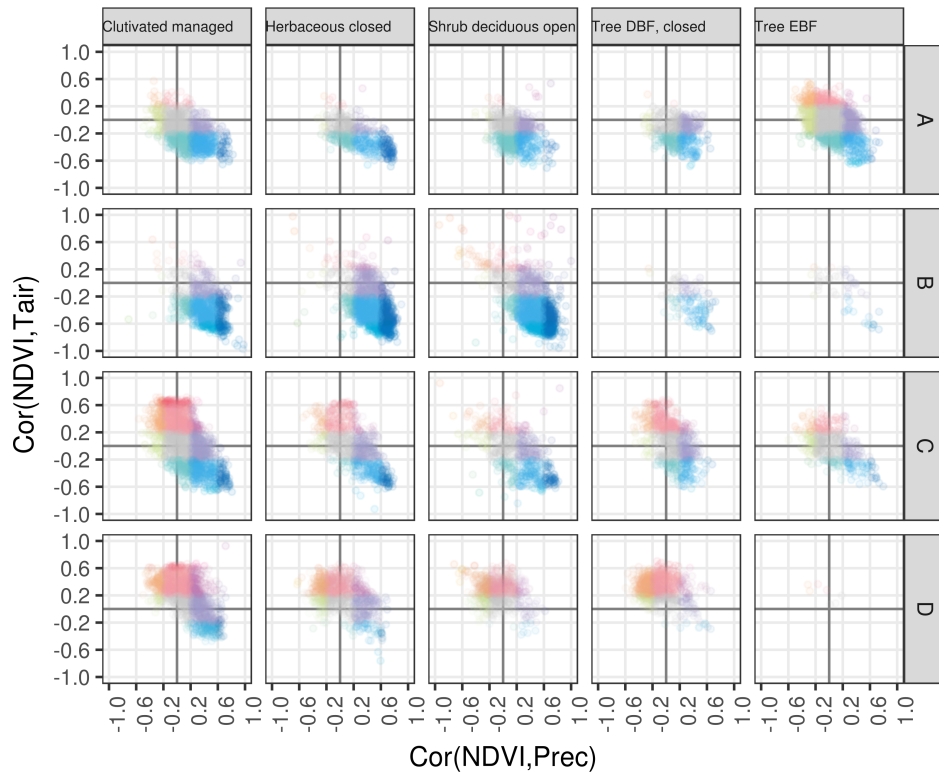


Figure S14: Land cover classes in “correlation space” across Köppen–Geiger classes for longer–term scale. Correlations of NDVI with precipitation (x axis) and air temperature (y axis) as determined for Fig. 3 were plotted for major land cover classes (GLC2000, columns) across Köppen–Geiger classes (rows). Each point represents one 0.5° grid cell from the global map. Transparency was scaled with the area represented by the grid cell, colors represent both correlations in accordance with Fig. 3 binned into quantiles (e. g. purple – high positive correlation of NDVI with both T_{air} and Prec, green – high negative correlation of NDVI with both T_{air} and Prec). A – equatorial, B – arid, C – warm temperate, D – snow)

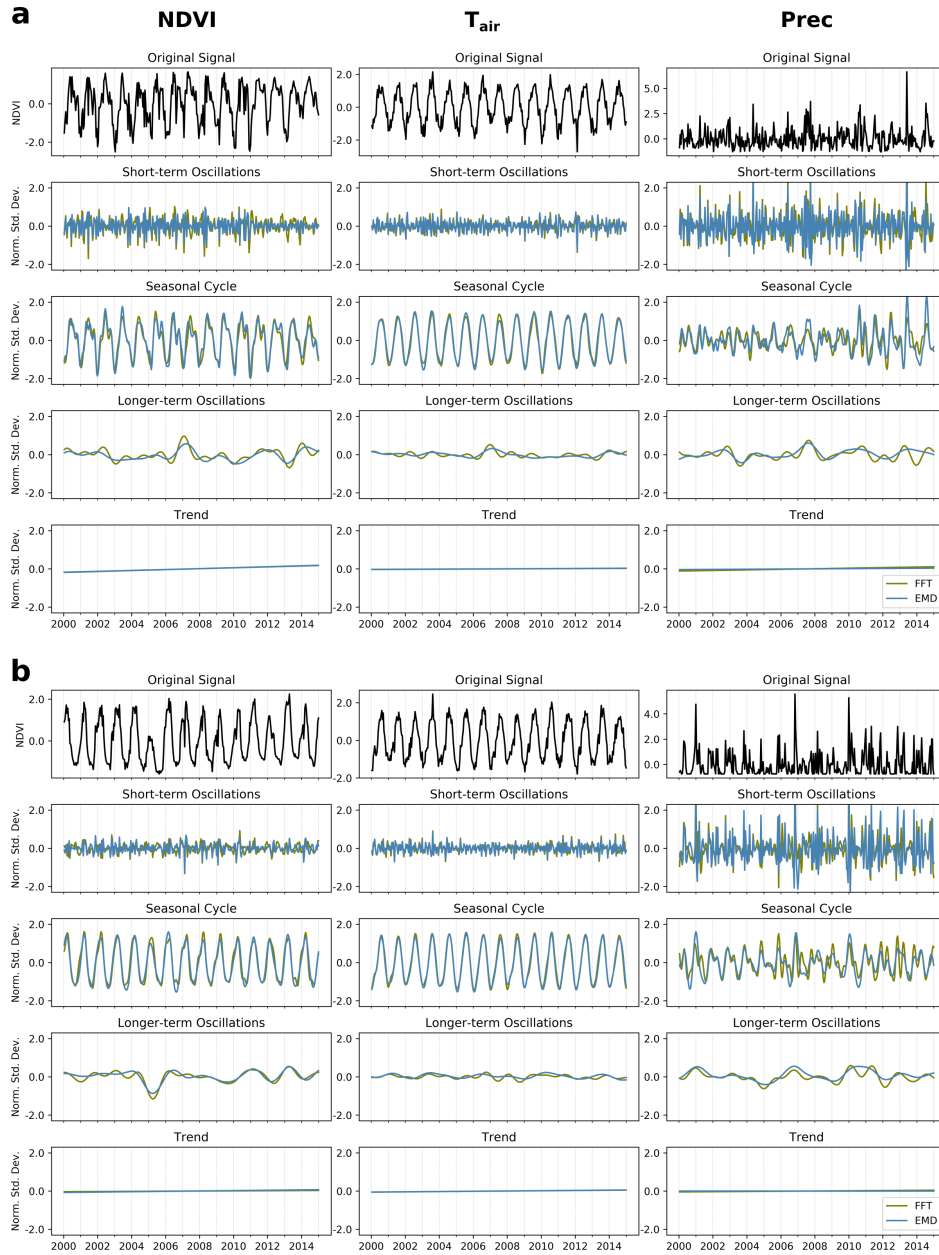


Figure S15: **Temporal comparison of Fourier transformation (FFT) and Empirical Mode Decomposition (EMD)**. Time series examples for **a.** Germany (lon. 11°, lat. 51°), and **b.** southern Portugal (lon. -8°, lat. 38°) of decomposed time series of NDVI, air temperature (T_{air}) and precipitation (Prec) from 2000–2014. FFT signals are colored green, EMD signals are colored blue.

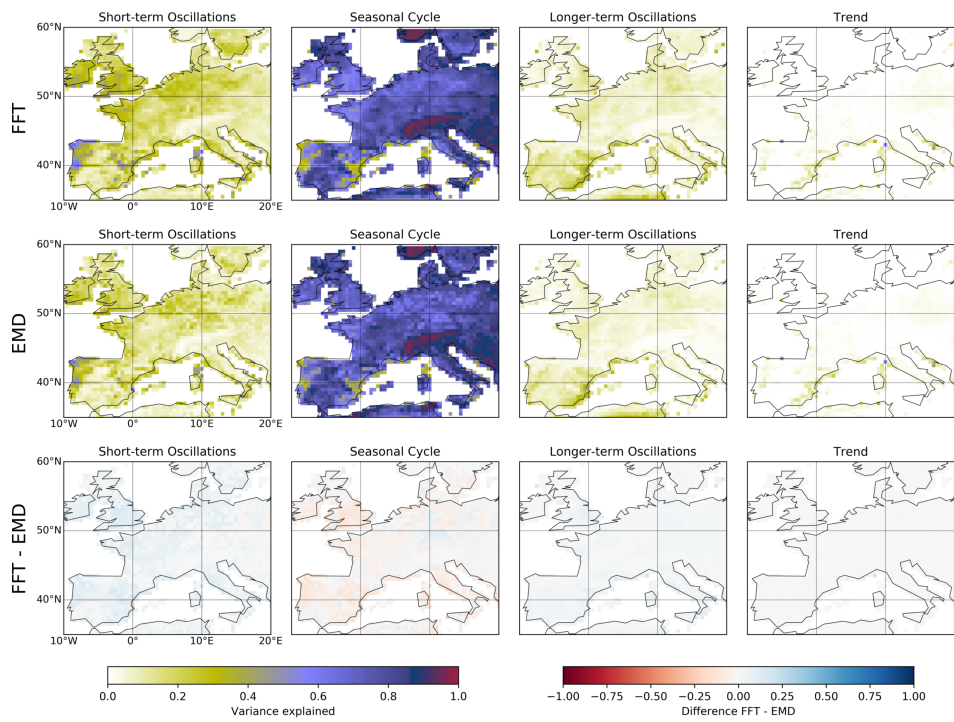


Figure S16: **Spatial comparison of Fourier transformation (FFT) and Empirical Mode Decomposition (EMD)**. Comparison of variance explained per pixel as determined for NDVI time series (2000–2014) by Fourier transformation (FFT, upper row) and empirical mode decomposition (EMD), as well as their difference (lower row) over Europe.

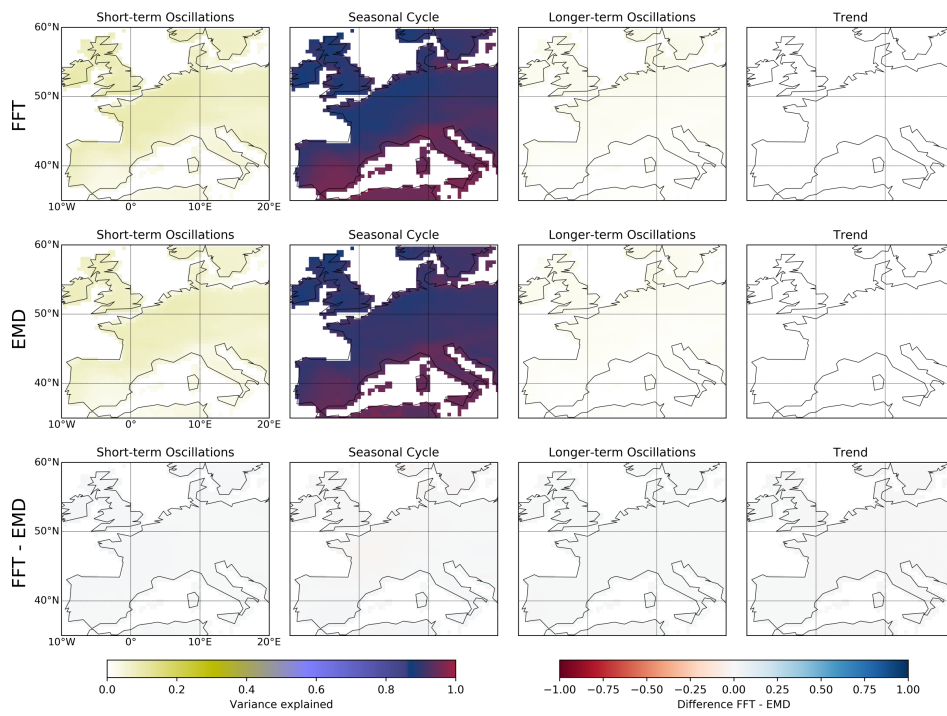


Figure S17: **Spatial comparison of Fourier transformation (FFT) and Empirical Mode Decomposition (EMD)**. Comparison of variance explained per pixel as determined for air temperature time series (2000–2014) by Fourier transformation (FFT, upper row) and empirical mode decomposition (EMD), as well as their difference (lower row) over Europe.

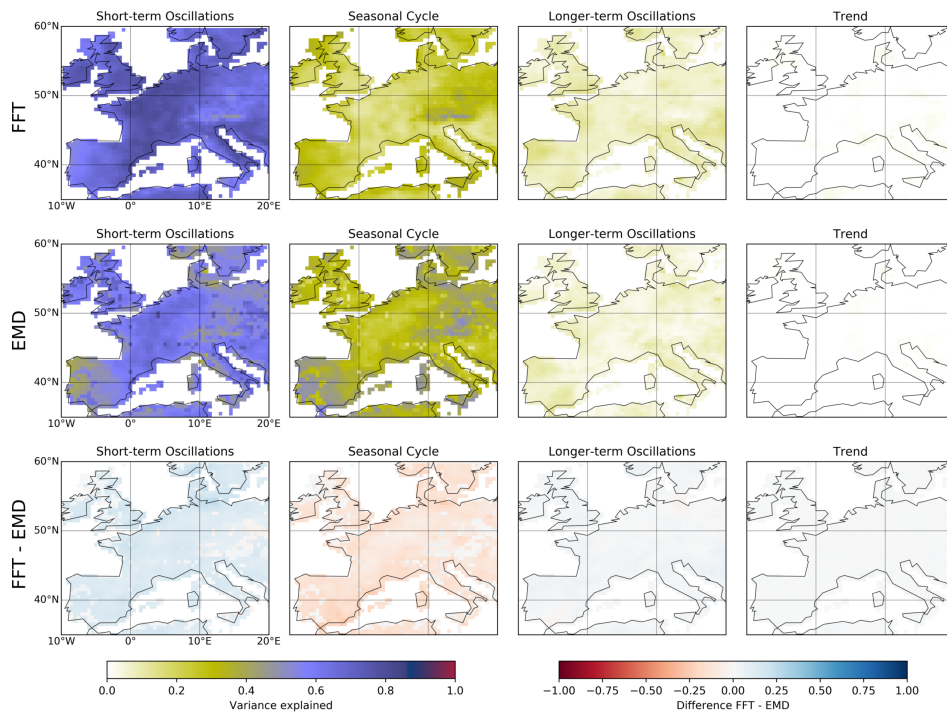


Figure S18: **Spatial comparison of Fourier transformation (FFT) and Empirical Mode Decomposition (EMD)**. Comparison of variance explained per pixel as determined for precipitation time series (2000–2014) by Fourier transformation (FFT, upper row) and empirical mode decomposition (EMD), as well as their difference (lower row) over Europe.

Supplementary Tables

Table S1: Selected and excluded classes from GLC 2000.

Selected land cover classes (natural vegetation)	
Herbaceous cover closed open	Tree cover broadleaved deciduous open
Mosaic: Tree cover other natural vegetation	Tree cover broadleaved evergreen
Regularly flooded shrub and or herbaceous cover	Tree cover mixed leaf type
Shrub cover closed open deciduous	Tree cover needle leaved deciduous
Shrub cover closed open evergreen	Tree cover needle leaved evergreen
Sparse herbaceous or sparse shrub cover	Tree cover regularly flooded fresh water
Tree cover broadleaved deciduous closed	Tree cover regularly flooded saline water
Excluded land cover classes	
Artificial surfaces and associated areas	Mosaic: Cropland / Tree cover / Other natural vegetation
Bare areas	Snow and ice (natural & artificial)
Cultivated and managed areas	Tree cover burnt
Mosaic: Cropland / Shrub or grass cover	Water bodies (natural & artificial)

Table S2: Global weighted mean of decomposed oscillations and three latitudinal bands (i) extratropics northern hemisphere (above 23.5° N), (ii) tropics (23.5° N to 23.5° S) and (iii) extratropics southern hemisphere (below 23.5° S). The mean weights are based on pixel area.

Variable	Region	Short-term	Seasonal	Longer-term	Trend
NDVI	Global	0.18	0.71	0.09	0.02
NDVI	Above 23.5° N	0.1	0.84	0.05	0.01
NDVI	Tropics	0.27	0.59	0.11	0.02
NDVI	Below 23.5° S	0.25	0.46	0.25	0.03
Tair	Global	0.11	0.83	0.04	0.01
Tair	Above 23.5° N	0.05	0.94	0.01	0
Tair	Tropics	0.21	0.68	0.09	0.02
Tair	Below 23.5° S	0.08	0.9	0.02	0
Prec	Global	0.52	0.41	0.06	0
Prec	Above 23.5° N	0.57	0.36	0.06	0
Prec	Tropics	0.42	0.51	0.06	0
Prec	Below 23.5° S	0.68	0.24	0.09	0

Table S3: Summary statistics of total area assessed and percentage of dominant NDVI oscillations by Köppen–Geiger, vegetated land cover classes and dominant oscillations of climatic variables. A: Annual, L: Longer-term, S: Short-term, T: Trend. Values of T are solely presented for area calculations.

Köppen-Geiger classes	Area assessed (km ²)												DOMINANT OSCILLATION OF NDVI																			
	S				A				L				T				Overall percentage (%)				Percentage by class (% rows)				Percentage by dominant NDVI (% cols)							
	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T				
ColdTemp	20666.8	23707941.1	0.0	10321.0	0.0	31.2	0.0	0.1	99.9	0.0	0.0	0.0	0.0	31.2	0.0	0.1	99.9	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	37.1	0.0	0.0	0.0
WarmTemp	1531328.9	11281304.1	336226.0	78194.2	2.0	14.9	0.4	11.6	85.3	2.5	0.4	0.4	2.0	14.9	0.4	11.6	85.3	2.5	0.4	0.4	22.4	17.6	0.0	0.0	22.4	17.6	0.0	0.0	17.6	6.9	0.0	0.0
Arid	775689.8	13231895.0	4534532.5	79587.2	1.0	17.4	6.0	4.2	71.1	24.4	6.0	4.2	1.0	17.4	6.0	4.2	71.1	24.4	6.0	4.2	11.3	20.7	92.5	0.0	11.3	20.7	92.5	0.0	20.7	92.5	0.0	0.0
Equatorial	4520718.3	15702426.7	30631.2	30043.6	6.0	20.7	0.0	22.3	77.4	0.2	0.0	0.0	6.0	20.7	0.0	22.3	77.4	0.2	0.0	0.0	66.0	24.6	0.6	0.0	66.0	24.6	0.6	0.0	24.6	0.6	0.0	0.0
Vegetated land cover classes	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T
Herbaceous cover closed open	514350.8	9193800.1	1255570.0	19281.0	0.7	12.1	1.7	4.7	83.7	11.4	0.0	0.0	0.7	12.1	1.7	4.7	83.7	11.4	0.0	0.0	7.5	14.4	0.0	0.0	7.5	14.4	0.0	0.0	14.4	0.0	0.0	0.0
Mosaic: Tree cover other natural vegetation	138263.4	2050350.8	5415.6	5717.1	0.2	2.7	0.0	6.3	93.2	0.2	0.0	0.0	0.2	2.7	0.0	6.3	93.2	0.2	0.0	0.0	2.0	3.2	0.1	0.0	2.0	3.2	0.1	0.0	3.2	0.1	0.0	0.0
Regularly flooded shrub and or herbaceous cover	94256.8	1425427.4	51408.3	2849.5	0.1	1.9	0.1	6.0	90.6	3.3	0.1	0.1	0.1	1.9	0.1	6.0	90.6	3.3	0.1	0.1	1.4	2.2	1.0	0.0	1.4	2.2	1.0	0.0	2.2	1.0	0.0	0.0
Shrub cover closed open deciduous	208024.5	9421620.4	966549.2	33200.1	0.3	12.4	1.3	2.0	88.6	9.1	0.0	0.0	0.3	12.4	1.3	2.0	88.6	9.1	0.0	0.0	3.0	14.7	19.7	0.0	3.0	14.7	19.7	0.0	14.7	19.7	0.0	0.0
Shrub cover closed open evergreen	210446.9	1688229.8	41985.2	7422.2	0.3	2.2	0.1	10.8	86.7	2.2	0.1	0.1	0.3	2.2	0.1	10.8	86.7	2.2	0.1	0.1	3.1	2.6	0.9	0.0	3.1	2.6	0.9	0.0	2.6	0.9	0.0	0.0
Sparse herbaceous or sparse shrub cover	637336.7	7667753.4	2404181.6	61753.6	0.8	10.1	3.2	5.9	71.2	22.3	0.0	0.0	0.8	10.1	3.2	5.9	71.2	22.3	0.0	0.0	9.3	12.0	49.1	0.0	9.3	12.0	49.1	0.0	12.0	49.1	0.0	0.0
Tree cover broadleaved deciduous closed	140333.1	6241346.3	17724.3	19055.9	0.2	8.2	0.0	2.2	97.2	0.3	0.0	0.0	0.2	8.2	0.0	2.2	97.2	0.3	0.0	0.0	2.0	9.8	0.4	0.0	2.0	9.8	0.4	0.0	9.8	0.4	0.0	0.0
Tree cover broadleaved deciduous open	97232.9	3473408.2	81498.9	0.0	0.1	4.6	0.1	2.7	95.1	2.2	0.1	0.1	0.1	4.6	0.1	2.7	95.1	2.2	0.1	0.1	1.4	5.4	1.7	0.0	1.4	5.4	1.7	0.0	5.4	1.7	0.0	0.0
Tree cover broadleaved evergreen	4107891.8	7421853.2	17685.7	26460.1	5.4	9.8	0.0	35.5	64.1	0.2	0.0	0.0	5.4	9.8	0.0	35.5	64.1	0.2	0.0	0.0	60.0	11.6	0.4	0.0	60.0	11.6	0.4	0.0	11.6	0.4	0.0	0.0
Tree cover mixed leaf type	37583.7	3049469.0	2419.7	9124.3	0.0	4.0	0.0	1.2	98.4	0.1	0.0	0.0	0.0	4.0	0.0	1.2	98.4	0.1	0.0	0.0	0.5	4.8	0.0	0.0	0.5	4.8	0.0	0.0	4.8	0.0	0.0	0.0
Tree cover needle leaved deciduous	0.0	3427749.0	0.0	0.0	0.0	4.5	0.0	0.0	100.0	0.0	0.0	0.0	0.0	4.5	0.0	0.0	100.0	0.0	0.0	0.0	0.0	5.4	0.0	0.0	0.0	5.4	0.0	0.0	5.4	0.0	0.0	0.0
Tree cover needle leaved evergreen	389722.8	8498764.0	56951.3	13282.3	0.5	11.2	0.1	4.4	94.9	0.6	0.0	0.0	0.5	11.2	0.1	4.4	94.9	0.6	0.0	0.0	5.7	13.3	1.2	0.0	5.7	13.3	1.2	0.0	13.3	1.2	0.0	0.0
Tree cover regularly flooded fresh water	257565.2	303364.8	0.0	0.0	0.3	0.4	0.0	0.0	45.9	54.1	0.0	0.0	0.3	0.4	0.0	0.0	45.9	54.1	0.0	0.0	3.8	0.5	0.0	0.0	3.8	0.5	0.0	0.0	0.5	0.0	0.0	0.0
Tree cover regularly flooded saline water	15375.3	60430.6	0.0	0.0	0.0	0.1	0.0	0.0	20.3	79.7	0.0	0.0	0.0	0.1	0.0	0.0	20.3	79.7	0.0	0.0	0.2	0.1	0.0	0.0	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.0
Dominant oscillations of Temperature and precipitation*	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T	S	A	L	T
A.A.	2505656.1	26763039.4	0.0	0.0	3.3	35.3	0.0	8.6	91.4	0.0	0.0	0.0	3.3	35.3	0.0	8.6	91.4	0.0	0.0	0.0	36.6	41.9	0.0	0.0	36.6	41.9	0.0	0.0	41.9	0.0	0.0	0.0
A.S.	3767937.1	36854191.1	4901389.7	198146.0	5.0	48.6	6.5	8.2	80.6	10.7	0.0	0.0	5.0	48.6	6.5	8.2	80.6	10.7	0.0	0.0	55.0	57.7	100.0	0.0	55.0	57.7	100.0	0.0	57.7	100.0	0.0	0.0
S.A.	0.0	33532.8	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0
S.S.	338242.5	196420.7	0.0	0.0	0.4	0.3	0.0	63.3	36.7	0.0	0.0	0.0	0.4	0.3	0.0	63.3	36.7	0.0	0.0	0.0	4.9	0.3	0.0	0.0	4.9	0.3	0.0	0.0	0.3	0.0	0.0	0.0
L.S.	236346.1	0.0	0.0	0.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	3.5	0.0	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
L.A.	0.0	76383.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0
Percentage area assessed (%)	9.03	84.25	6.46	0.26																												
Total area assessed (km²)	75871486.5																															

*A: Annual, S: Short-term oscillation, L: Longer-term oscillation, T: Trend

Table S4: Spatial association between co-oscillation regimes and Köppen–Geiger or Global Land Cover (GLC2000), respectively. c = complementarity, h = homogeneity, m = number of classes, V = V-measure.

	Co-oscillations regime (11 classes)			
Static maps	m	h	c	V
Köppen–Geiger	4	0.19	0.16	0.17
Global land cover	9	0.16	0.09	0.11

Appendix C

Supplement manuscript 2

Supplementary Material

1 SUPPLEMENTARY TABLES

Table S1. Time series available at the RegESDL. BESS: Breathing Earth System Simulator. CHIRPS: Climate Hazards group Infrared Precipitation with Stations. MODIS: Moderate Resolution Imaging Spectroradiometer. TRMM: Tropical Rainfall Measuring Mission. QA4ECV: Quality Assurance for Essential Climate Variables project. Ori.res: Original resolution.

Variable	Abbreviation	Data product or project	Timestep*	Firstyear	Lastyear	Orig.res.
Bihemispherical reflectance near infrared	BHR_NIR	QA4ECV	daily	2001	2018	0.5°
Bihemispherical reflectance shortwave	BHR_SW	QA4ECV	daily	2001	2018	0.5°
Bihemispherical reflectance visible	BHR_VIS	QA4ECV	daily	2001	2018	0.5°
Directional hemispherical reflectance near infrared	DHR_NIR	QA4ECV	daily	2001	2018	0.5°
Directional hemispherical reflectance shortwave	DHR_SW	QA4ECV	daily	2001	2018	0.5°
Directional hemispherical reflectance visible	DHR_VIS	QA4ECV	daily	2011	2018	0.5°
Diffuse Photosynthetically Active Radiation	BESS_PARdiff_Daily	BESS	daily	2001	2016	0.5°
BESS_RSDN_Daily	BESS_RSDN_Daily	BESS	daily	2001	2016	0.5°
Gross Primary Productivity	GPP	BESS	8day	2000	2015	0.0083°
Photosynthetically active radiation	BESS_PAR	BESS	daily	2001	2016	0.5°
Evapotranspiration	ET	BESS	8day	2001	2015	0.0083°
Enhanced Vegetation Index Aqua	EVI_aqua	MYD13A2.EVI	16day	2003	2017	0.0083°
Enhanced Vegetation Index Terra	EVI_terra	MOD13A2.EVI	16day	2001	2017	0.0083°
Fire dates	Fire_date	MCD64A1	monthly	2001	2017	0.0042°
Fire quality	Fire_quality	MCD64A1	monthly	2001	2017	0.0042°
Fire uncertainty	Fire_date_uncertainty	MCD64A1	monthly	2001	2017	0.0042°
Fraction of Absorbed Photosynthetically Active Radiation	FPAR	MOD15A2	8day	2001	2017	0.0042°
Land Surface Temperature day	LST_day	MOD11A2	8day	2001	2017	0.0083°
Land Surface Temperature night	LST_night	MOD11A2	8day	2001	2017	0.0083°
Leaf Area Index	LAI	MOD15A2H	8day	2001	2017	0.0042°
Normalized Difference Vegetation Index Aqua	NDVI_aqua	MYD13A2.NDVI	16day	2003	2017	0.0083°
Normalized Difference Vegetation Index Terra	NDVI_terra	MOD13A2.NDVI	16day	2001	2017	0.0083°
Vegetation Cover Fraction Trees	VCT_tree	MOD44B	annual	2001	2016	0.0083°
Vegetation Cover Fraction no vegetation	VCF_nonVeg	MOD44B	annual	2001	2016	0.0083°
Vegetation Cover Fraction nonTrees	VCF_nonTreeVeg	MOD44B	annual	2001	2016	0.0083°
ESA_land_cover	ESA_LC	ESA Land Cover CCI Product	annual	2001	2015	300 m
Precipitation	Precipitation	TRMM	monthly	1998	2012	0.25°
Precipitation	Precipitation	CHIRPS	daily	2001	2018	0.05°

Table S2. Quality flags of MODIS products, and selected flag settings. QBP: Quality band position in HDF file.

Variable	Data product	Dataset name	QBP	Data type	Quality criteria settings
LST day	MOD11A2	QC.Day	2	Binary	Bit 0 to 1 == 00 & any combination Bit 0 to 1 == 01 & Bit 2 to 3 == 00 or 01 or 11
LST night	MOD11A2	QC.Night	6	Binary	Bit 0 to 1 == 00 & any combination Bit 0 to 1 == 01 & Bit 2 to 3 == 00 or 01 or 11
NDVI/EVI terra	MOD13A2	1 km 16 days pixel reliability	12	8bit	Values of 0 or 1
NDVI/EVI aqua	MYD13A2	1 km 16 days pixel reliability	12	8bit	Values of 0 or 1
FPAR/LAI	MOD15A2	FparLai.QC	3	Binary	Bit 0 == 0 & any combination Bit 0 == 1 & Bit 5 to 7 == 000 or 001 or 010
Vegetation Cover Fraction Trees/ No trees/No vegetation	MOD44B	Quality	4	Binary	All values <1111111 (This is an annual layer. Flags are fixed for dates range)
Fire quality	MCD64A1	Quality	3	Binary	Bit 1 == 1 & any combination Bit 1 == 0 & Bit 5 to 7 == from 001 to 101

Table S3. Descriptive variables available at the RegESDL. EarthEnv: Global 1-km Cloud Cover. GCH-lidar: Mapping forest canopy height globally with spaceborne lidar. HWSD: Harmonized world soil database. SoilGrids: Global Gridded Soil Information.

Variable	Source	Resolution	Units	Citation
canopy_height	GCH-lidar	1 km	m	Simard, M., Pinto N., Fisher J.B. and Baccini A. 2011. Mapping forest canopy height globally with spaceborne lidar. <i>Journal of Geophysical Research</i> , vol. 116.
clouds_monthlymean_01	EarthEnv	1 km	%	Wilson A.M. and Jetz W. 2016. Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. <i>PLoS Biol</i> 14(3): e1002415. doi:10.1371/journal.pbio.1002415 https://www.earthenv.org/cloud
clouds_monthlymean_02	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_03	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_04	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_05	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_06	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_07	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_08	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_09	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_10	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_11	EarthEnv	1 km	%	Wilson and Jetz (2016)
clouds_monthlymean_12	EarthEnv	1 km	%	Wilson and Jetz (2016)
available_water_storage_capacity	HWSD	1 km	categorical_mm/m	FAO/IIASA/ISRIC/ISSCAS/JRC, 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria. http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html?sb=1
clay_percentage_in_subsoil	HWSD	1 km	%	FAO et al. (2012)
gravel_volume_percentage_in_subsoil	HWSD	1 km	%	FAO et al. (2012)
organic_carbon_weight_percentage_in_subsoil	HWSD	1 km	%	FAO et al. (2012)
sand_percentage_in_subsoil	HWSD	1 km	%	FAO et al. (2012)
silt_percentage_in_subsoil	HWSD	1 km	%	FAO et al. (2012)
clay_percentage_in_topsoil	HWSD	1 km	%	FAO et al. (2012)
gravel_volume_percentage_in_topsoil	HWSD	1 km	%	FAO et al. (2012)
organic_carbon_weight_percentage_in_topsoil	HWSD	1 km	%	FAO et al. (2012)
sand_percentage_in_topsoil	HWSD	1 km	%	FAO et al. (2012)
silt_percentage_in_topsoil	HWSD	1 km	%	FAO et al. (2012)
depth_to_bedrock_R_horizon_up_to_200_cm	SoilGrids	250 m	cm	Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, et al. 2014. SoilGrids1km — Global Soil Information Based on Automated Mapping. <i>PLoS ONE</i> 9(8): e105992. doi:10.1371/journal.pone.0105992 https://www.isric.org/explore/soilgrids
predicted_probability_of_occurrence_of_R_horizon	SoilGrids	250 m	%	Hengl et al. (2014)
absolute_depth_to_bedrock	SoilGrids	250 m	cm	Hengl et al. (2014)
Soil_organic_carbon_stock	SoilGrids	250 m	tonnes/ha	Hengl et al. (2014)
Soil_organic_carbon_stock	SoilGrids	250 m	tonnes/ha	Hengl et al. (2014)
Bulk_density_fine_earth	SoilGrids	250 m	kg/m ³	Hengl et al. (2014)
Bulk_density_fine_earth	SoilGrids	250 m	kg/m ³	Hengl et al. (2014)
Clay_content_0.2_micro_meter_mass_fraction	SoilGrids	250 m	%	Hengl et al. (2014)
Clay_content_0.2_micro_meter_mass_fraction	SoilGrids	250 m	%	Hengl et al. (2014)
Coarse_fragments_volumetric	SoilGrids	250 m	%	Hengl et al. (2014)
Coarse_fragments_volumetric	SoilGrids	250 m	%	Hengl et al. (2014)
Silt_content_250_micro_meter_mass_fraction	SoilGrids	250 m	%	Hengl et al. (2014)
Silt_content_250_micro_meter_mass_fraction	SoilGrids	250 m	%	Hengl et al. (2014)
Sand_content_50–2000_micro_meter_mass_fraction	SoilGrids	250 m	%	Hengl et al. (2014)
Sand_content_50–2000_micro_meter_mass_fraction	SoilGrids	250 m	%	Hengl et al. (2014)
Cation_exchange_capacity_of_soil	SoilGrids	250 m	cmolc/kg	Hengl et al. (2014)
Cation_exchange_capacity_of_soil	SoilGrids	250 m	cmolc/kg	Hengl et al. (2014)
Soil_organic_carbon_content_fine_earth_fraction	SoilGrids	250 m	g/kg	Hengl et al. (2014)
Soil_organic_carbon_content_fine_earth_fraction	SoilGrids	250 m	g/kg	Hengl et al. (2014)
Soil_pH_x_10_in_H2O	SoilGrids	250 m	pH	Hengl et al. (2014)
Soil_pH_x_10_in_H2O	SoilGrids	250 m	pH	Hengl et al. (2014)
Soil_pH_x_10_in_KCl	SoilGrids	250 m	pH	Hengl et al. (2014)
Soil_pH_x_10_in_KCl	SoilGrids	250 m	pH	Hengl et al. (2014)

Table S4. Layers of Colombia available at the RegESDL.

Variable	Data product or project	Format-Res.	Units	Citation
annual_mean_precipitation_interpolation_standardized_cokriging*	Improved long-term mean annual rainfall fields for Colombia	Raster 0.017°	mm	Alvarez-Villa, O. D., Velez, J. I., & Poveda, G. 2011. Improved long-term mean annual rainfall fields for Colombia. International Journal of Climatology, 31(14), 2194–2212. https://doi.org/10.1002/joc.222
annual_mean_precipitation_markov_regionalization*	Improved long-term mean annual rainfall fields for Colombia	Raster 0.017°	mm	Alvarez et al. (2011)
precipitation_annual_mean_interpolation_colocated_cokriging*	Improved long-term mean annual rainfall fields for Colombia	Raster 0.017°	mm	Alvarez et al. (2011)
precipitation_annual_mean_interpolation_kriging_with_external_drift*	Improved long-term mean annual rainfall fields for Colombia	Raster 0.017°	mm	Alvarez et al. (2011)
biotic_units**	Ajuste de la capa del componente biótico, incorporada dentro del Mapa Nacional de Ecosistemas Terrestres, Marinos y Costeros de Colombia, elaborado en 2016 a escala 1:100.000	Shapefile	categorical	Londono, M. C., Bello, C., Velasquez, J., Norden, N., Ortiz, C., Gonzalez, I., Lopez, D., Gutierrez, C., Olaya, H., and Saavedra, K.: Documento Tecnico: Componente Biotico Mapa de Ecosistemas Continentales, Marinos y Costeros de Colombia, Escala 1 : 100 000, Tech. rep., Instituto de Investigacion de Recursos Biologicos Alexander von Humboldt, Bogota, D.C., 2017 http://geonetwork.humboldt.org.co/geonetwork/srv/spa/catalog.search#/metadata/a1afc35c-db98-4110-8093-98e599d1571e
municipalities_administrative_units***	Cartografia Basica Digital Integrada. Republica de Colombia.. Escala 1:100.000	Shapefile	categorical	IGAC. 2014. Cartografia Basica Digital Integrada. Republica de Colombia.. Escala 1:100.000. http://metadatos.igac.gov.co/geonetwork/srv/spa/catalog.search#/metadata/c1b4bfe5-f7c7-44a4-8849-c5f7c4257937
wetlands**	Delimitacion de ecosistemas estrategicos	Raster 90 m	categorical	IAVH 2016. Identificacion de humedales, escala 1:100.000. Proyecto: Insumos para la delimitacion de ecosistemas estrategicos: Paramos y Humedales. Convenio No. 13-014 (FA. 005 de 2013) suscrito entre el Fondo Adaptacion y el Instituto Humboldt. Bogota D.C., Colombia. http://geonetwork.humboldt.org.co/geonetwork/srv/spa/catalog.search#/metadata/d68f4329-0385-47a2-8319-8b56c772b4c0
agriculture_frontier***	Identificacion General de la Frontera Agricola en Colombia a escala 1:100.000 Ministerio de Agricultura y Desarrollo Rural Agropecuario	Shapefile	categorical	Unidad de Planificacion Rural Agropecuaria (UPRA). 2017. Identificacion general de la frontera agricola en Colombia. Escala 1:100.000. MADR y UPRA. Bogota (Colombia). https://www.minagricultura.gov.co/Normalidad/Projects_Documents/IDENTIFICACION%20GENERAL%20DE%20LA%20FRONTERA%20AGROPECUARIA%20EN%20COLOMBIA%20A%20ESCALA%201%3A%20100.000%20MINISTERIO%20DE%20AGRICULTURA%20Y%20DESARROLLO%20RURAL%20AGROPECUARIO%202017%20.pdf shp: https://sipra.upra.gov.co/
national_parks***	Parques Nacionales Naturales de Colombia	Shapefile	categorical	Limites de los Parques Nacionales Naturales de Colombia, en Sistema de Referencia Magna - Sirgas y multiescala (1:100.000 y 1:25.000). http://geonetwork.parquesnacionales.gov.co/geonetwork/srv/spa/metadata.show?id=10048&currTab=simple

*Raster resample to RegESDL using the mean value *Categorical raster resample to RegESDL using the mode value **Categorical shapefiles rasterized in R using the funcion 'last'.

Table S5. Data license information. BESS: Breathing Earth System Simulator. CHIRPS: Climate Hazards group Infrared Precipitation with Stations. EarthEnv: Global 1-km Cloud Cover. HWSD: Harmonized world soil database. QA4ECV: Quality Assurance for Essential Climate Variables project. MODIS: Moderate Resolution Imaging Spectroradiometer. SoilGrids: Global Gridded Soil Information. TRMM: Tropical Rainfall Measuring Mission.

Category	Data product or project	License information
Time series	QA4ECV	License details in Global Attributes. Check info from ESDL
	BESS	License issued directly by data owner. No restriction for use. Citation to original data is mandatory.
	MODIS	No restrictions on subsequent use or redistribution https://modaps.modaps.eosdis.nasa.gov/services/faq/LAADS_Data-Use_Citation_Policies.pdf (visited on 29.07.2020)
	TRMM	Data are freely available https://gpm.nasa.gov/data/policy
	CHIRPS	Creative Commons Attribution 4.0 International (CC BY 4.0)
Static layers	Canopy	This dataset is openly shared, without restriction, in accordance with the NASA Data and Information Policy.
	Harmonized World Soil Data Base	Creative Commons Attribution-Noncommercial 3.0 (CC BY-NC 3.0)
	SoilGrids	Creative Commons Attribution 4.0 International (CC BY 4.0)
	EarthEnv	Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
Layers for Colombia	Improved long-term mean annual rainfall fields for Colombia	From authors
	biotic_units	License issued directly by data owner. No restriction for use. Citation to original data is mandatory.
	municipalities_administrative_units	Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
	wetlands	TBD
	agriculture_frontier	Governmental information for public use without any use restrictions. Data must be properly cited.
	national_parks	Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) License provided by Colombian National Parks Rad. 20172400004243 (http://www.parquesnacionales.gov.co/) available at: url tbf

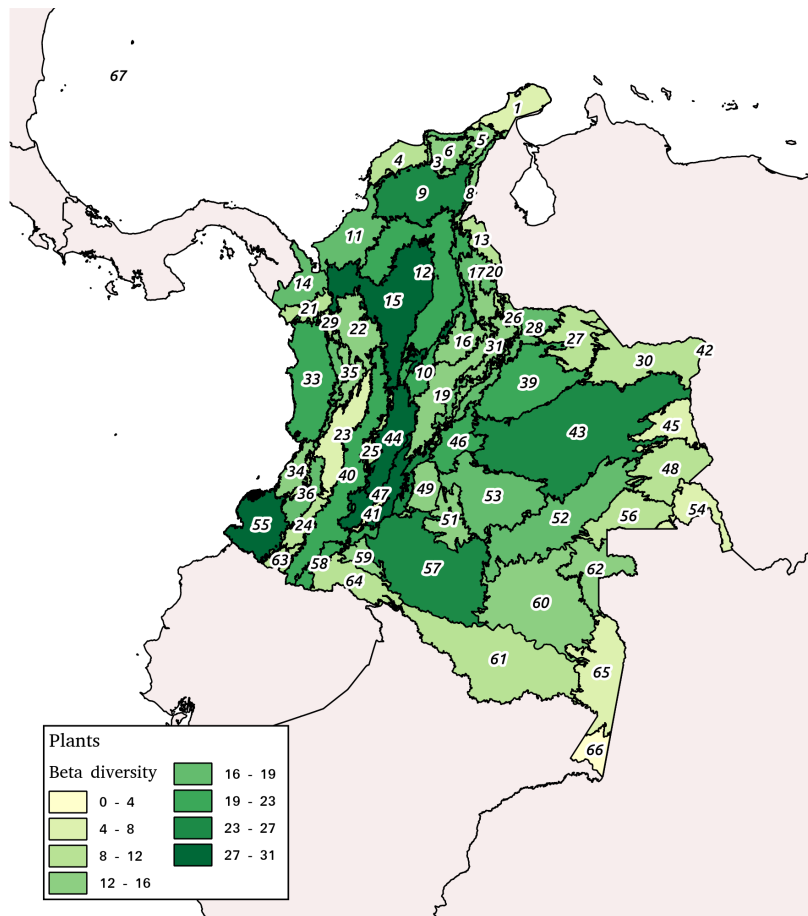
Table S6. Timing of computer processing when using different RegESDL version.

Examples	Reg ESDL for Temporal analysis Time (minutes:seconds)	Reg ESDL for Spatial analysis Time (minutes:seconds)
Example 1: Time series decomposition using Fast Fourier	02 min : 55 sec	36 min : 33 sec
Example 2: Find the maximum value along time series	00 min : 41 sec	12 min : 38 sec

Table S7. Scripts available in the supplementary, and at the Zenodo (<http://doi.org/10.5281/zenodo.5068004>) and GitHub (https://github.com/linamaes/Regional_ESDL) repositories. RegESDL: Regional Earth System Data Lab

Scripts	Supplementary	Zenodo	GitHub
Access to the RegESDL			
01_script_for_downloading_regional_ESDL_using_python.py	X	X	X
02_loading_regional_ESDL.jl	-	X	X
Computing performance			
03_comparison_TS_decomposition_using_different_DC_storage.jl	X	X	X
Main figures			
01_figure_2_land_cover_map_ESA_2014.jl	-	X	X
02_figure_4_principal_components_processing_&_plotting.jl	X	X	X
03_figure_5_histograms_PCs_variance_explained_by_land_cover.jl	X	X	X
04_figure_6_seasonality_ratio_annual_semiannual_oscillations_PC1.jl	X	X	X
05_figure_7_processing_biotic_units_mean_seasonal_cycle_PC1.jl	-	X	X
06_figure_7_plotting_biotic_units_mean_seasonal_cycle_PC1.jl	X	X	X
07_figure_8_&_figure_S4_S5_heatmaps_seasonality_ratio_&_climate.jl	X	X	X
Supplementary figures			
08_figure_S2_plotting_px_MSC_4vars&pca.jl	-	X	X
09_figure_S3_loadings_principal_components.jl	-	X	X

2 SUPPLEMENTARY FIGURES



Id and Name

1. Alta Guajira
2. Estribación norte Sierra Nevada de Santa Marta
3. Estribación sur Sierra Nevada de Santa Marta
4. Cartagena y delta del Magdalena
5. Baja Guajira y alto Cesar
6. Sierra nevada de Santa Marta
7. Perijá y montes de Oca
8. Perijá
9. Ariguaní-Cesar
10. Cordillera oriental Magdalena medio
11. Sinú
12. Magdalena medio y depresión momposina
13. Zulia
14. Darién –Taccarcuna
15. Nechí-San Lucas
16. Guane-Yariguies
17. Catatumbo
18. Tamá
19. Altoandino cordillera oriental
20. Cúcuta
21. Truandó
22. Cauca alto

Biotic units

23. Cauca medio
24. Patía
25. Chaparral
26. Vertiente llanera cordillera oriental
27. Arauca
28. Piedemonte Orinoquia
29. Alto Murri
30. Bitá
31. Uwa
32. Altoandino influencia llanera
33. San Juan
34. Micay
35. Estribaciones Pacífico norte
36. Estribaciones Pacífico sur
37. Vertiente Pacífico-Chocó
38. Vertiente Pacífico-Cauca
39. Casanare
40. Cordillera central
41. Caquetá influencia cordillera central
42. Maipures
43. Altilanura
44. Tolima grande
45. Matavén
46. Villavicencio
47. Huila-Caquetá
48. Guainía
49. Macarena
50. Picachos
51. Alto Guayabero
52. Inírida-Papunaua
53. Guaviare – Guayabero
54. Serranía del Naquén
55. Pacífico nariñense-Tumaco
56. Puinawai
57. Yari-Chiribiquete
58. Piedemonte Amazonas
59. Alto Caquetá
60. Apaporis
61. Huitoto-Cahuinari
62. Bajo Vaupés
63. Nudo de los pastos
64. Alto Putumayo
65. Bajo Caquetá - Puré
66. Ticuna – Amacayacu
67. San Andrés y Providencia

Figure S1. Biotic units of Colombia. Values correspond to Beta diversity values of plants (Magnoliopsida).

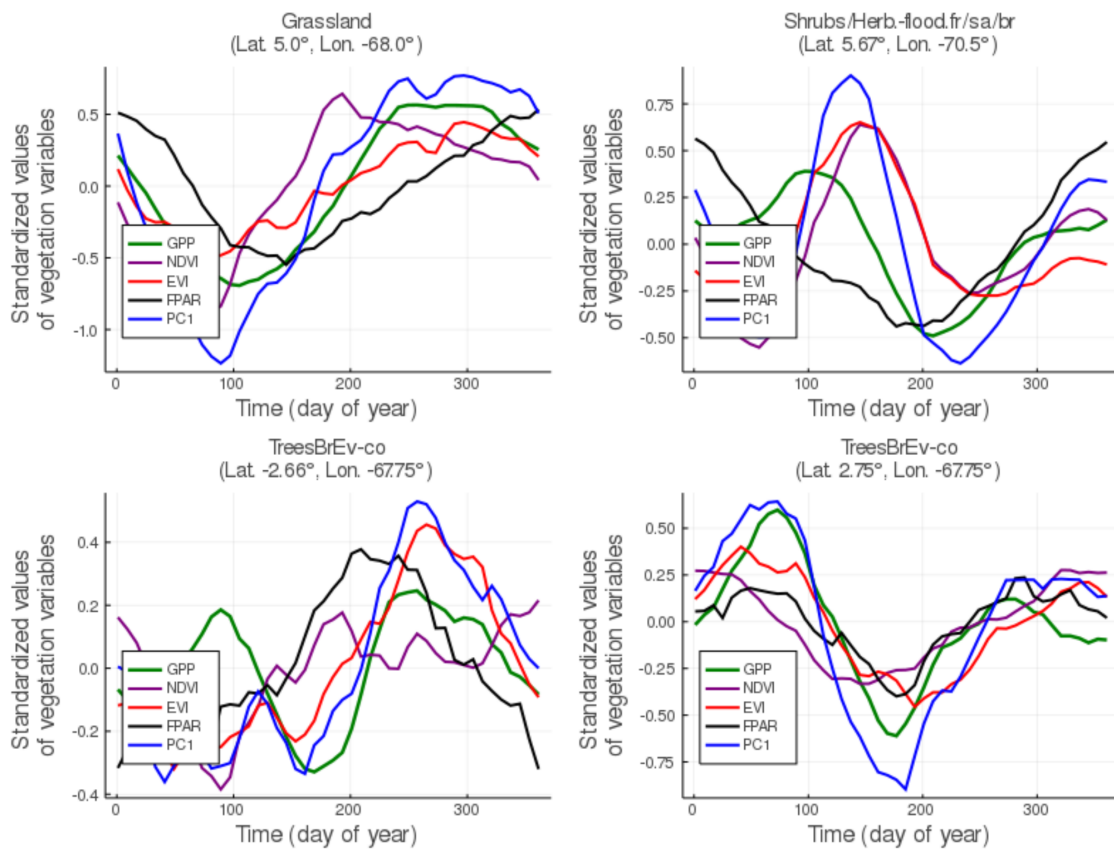


Figure S2. Mean seasonal cycle of the vegetation variables set calculated from the first principal component. Variables have been standardized to $\mu = 0$ and $\delta = 1$. Shrubs/Herb.-flood.fr/sa/br: Shrub or herbaceous cover, flooded, fresh/saline/brakish water. TreeBrEv-co: Tree cover, broadleaved, evergreen, closed to open ($> 15\%$).

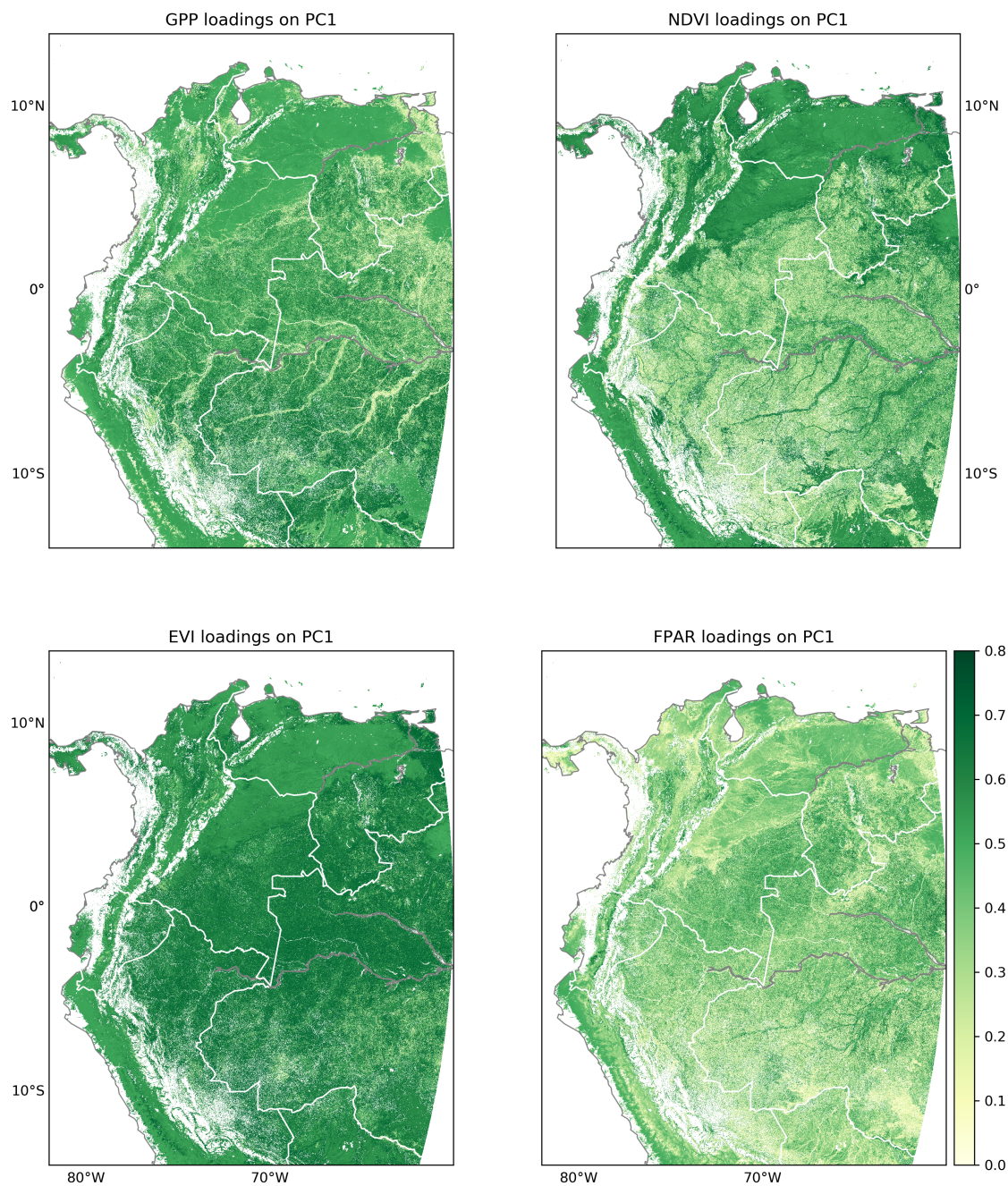


Figure S3. Loading values of vegetation variables for the first component of PCA. GPP: Gross Primary Productivity, NDVI: Normalized Difference Vegetation Index. EVI: Enhanced Vegetation Index. FPAR: Photosynthetically Active Radiation. National borders are delineated in white. Color map shows absolute values.

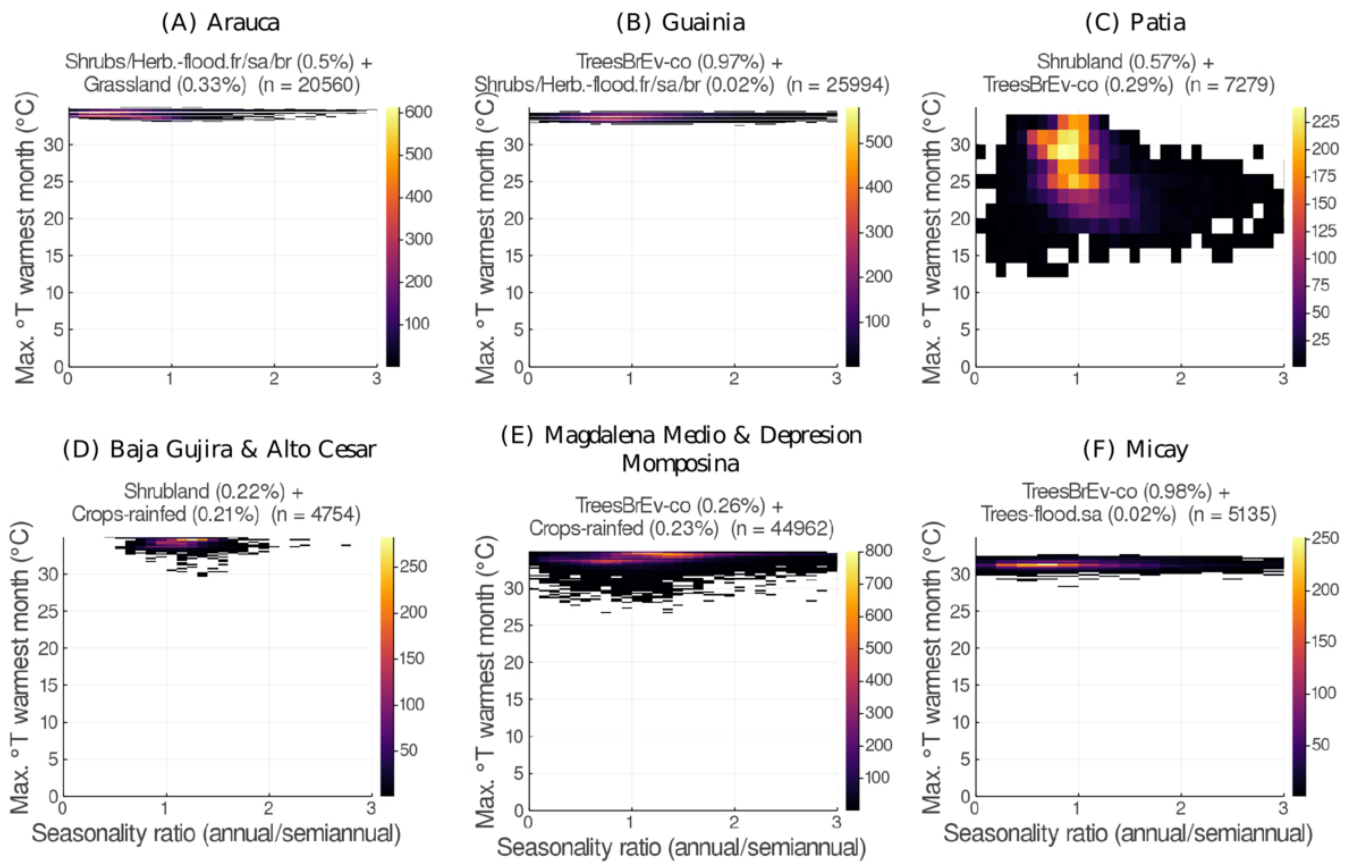


Figure S4. Seasonality ratio of annual and semiannual oscillation (x-axis) and maximum temperature of the warmest month (y-axis). Note that the color scheme range varies between figures. n is the total number of pixels in each biotic unit

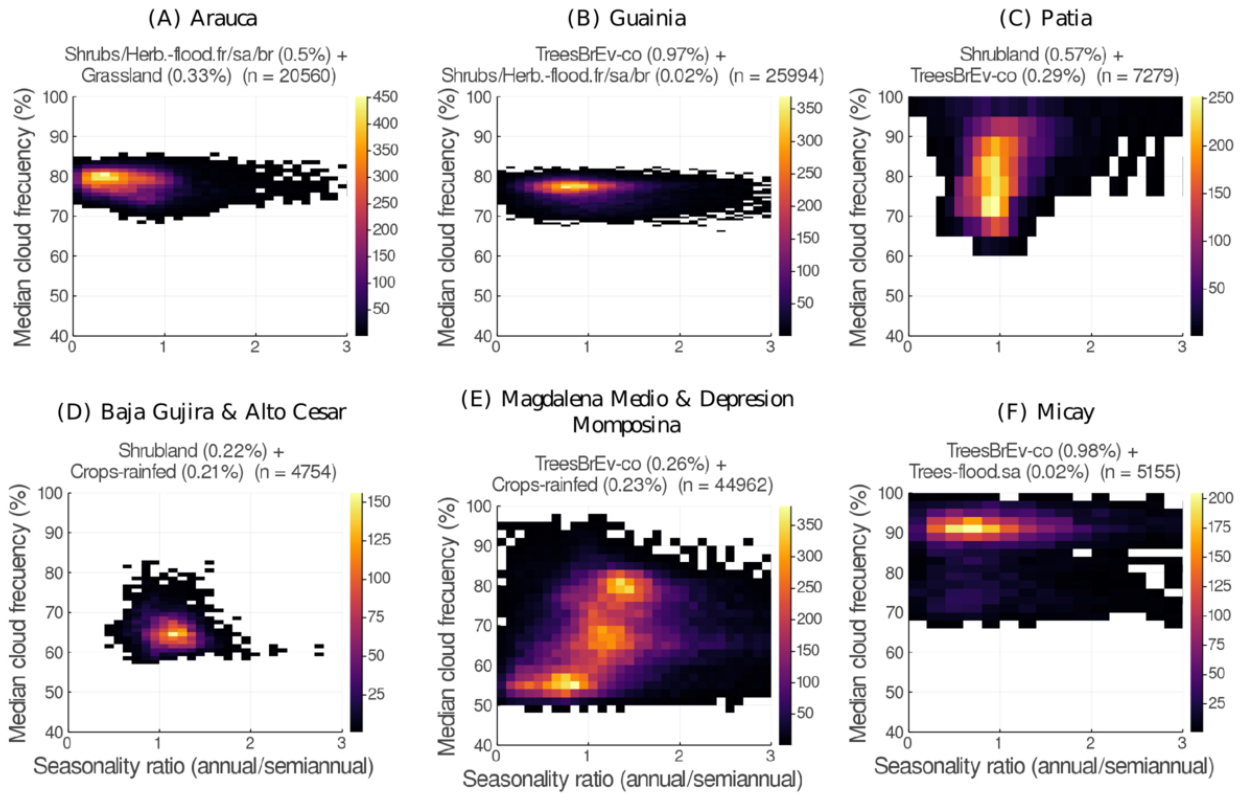


Figure S5. Seasonality ratio of annual and semiannual oscillation (x-axis) and annual median of clouds frequency (y-axis). Note that the color scheme range differs between plots. “n” is the total number of pixels in each biotic unit.

3 SCRIPTS

The following scripts were exported from Jupyter notebooks and correspond to the scripts specified in Table S7.

Appendix D

Supplement manuscript 3

Supplement of:

Spatial patterns of vegetation activity related to ENSO in northern South America

Estupinan-Suarez et al.

- The separated Global PCA analysis by land cover classes was carried out using the ESA land cover map for 2014 (ESA, 2017), maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf) resampled to 0.0083° and available at the RegionalESDL (Estupinan-Suarez et al., 2021).
- Data of precipitation of the driest month, used on Figure D.2 correspond to the bioclimatic variables of WorldClim.

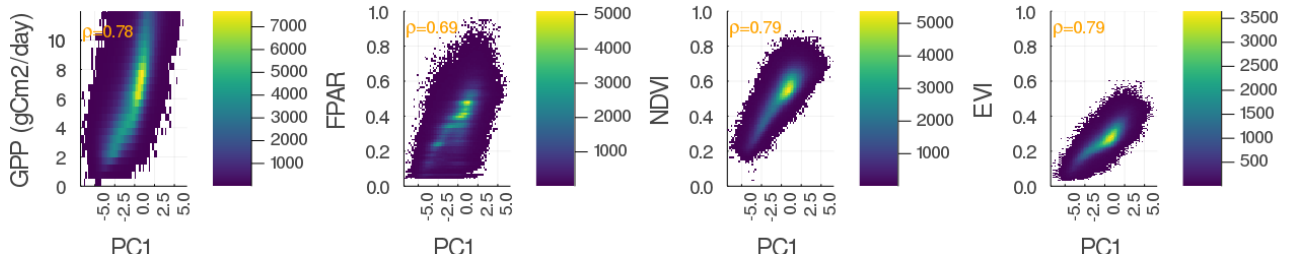
TABLE D.1: Explained fraction of variance by principal components (PC) from the Global PCA and the land cover PCA.

Principal component (PC)	Global PCA	Land cover PCA		
		Broadleaved Evergreen	Shrublands	Grasslands
PC ₁	0.45	0.33	0.68	0.71
PC ₂	0.21	0.26	0.15	0.15
PC ₃	0.19	0.23	0.10	0.09
PC ₄	0.15	0.18	0.07	0.05

TABLE D.2: First principal component loadings from the Global PCA and the land cover PCA.

Variable	Global PCA	Land cover PCA		
		Broadleaved evergreen forest	Shrublands	Grasslands
	PC ₁	PC ₁	PC ₁	PC ₁
EVI	-0.58	-0.67	-0.53	-0.53
FPAR	-0.41	-0.38	-0.44	-0.42
GPP	-0.49	-0.48	-0.51	-0.51
NDVI	-0.51	-0.42	-0.52	-0.53

(a) Grasslands



(b) Broadleaved Evergreen

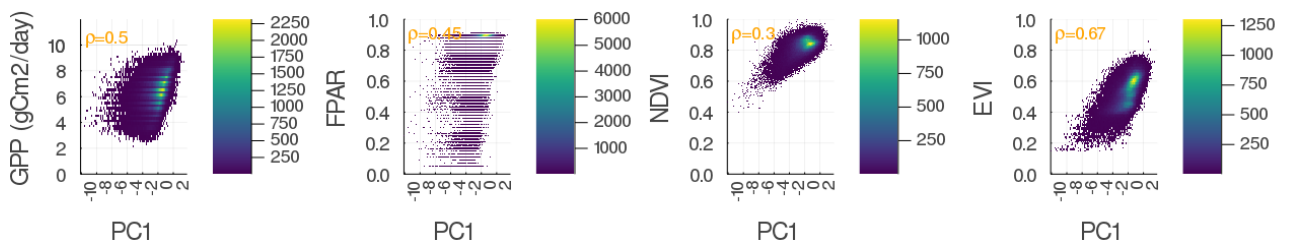


FIGURE D.1: Density plots of vegetation variables (y-axis) against PC₁ (x-axis) for homogeneous land covers.

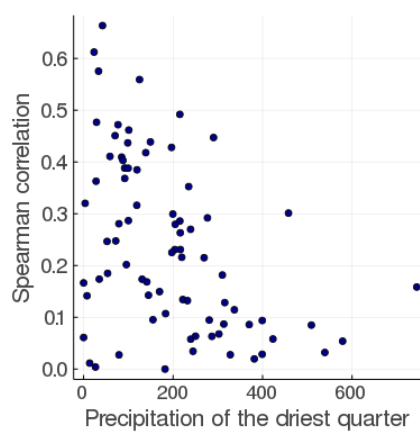


FIGURE D.2: Absolute values of Spearman correlation (y-axis) versus precipitation of the driest month (x-axis) for all ecoregions in the study area.

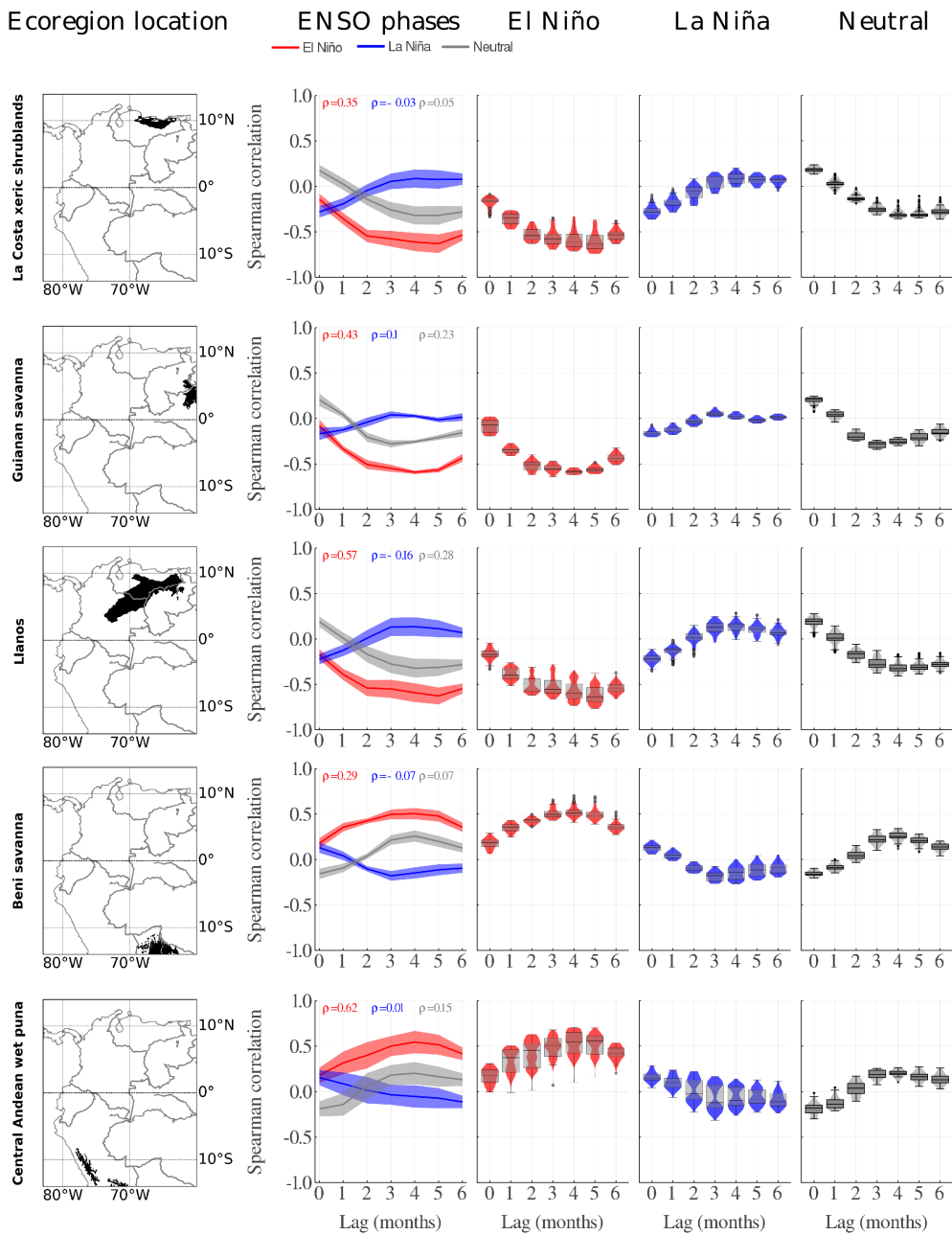


FIGURE D.3: Median values of the lagged Spearman correlation between MEI and precipitation (source: Global Precipitation Climatology Project) by ecoregions. Ecoregions are in rows. From left to right: first column shows the ecoregion location. Second column is the median correlation values within the ecoregion by ENSO phases. The ribbon correspond to the standard deviation. Data distribution from El Niño, La Niña and Neutral phases by lags is shown in columns three to five respectively. Shown ecoregions have the highest correlation and an area larger than 50000 km².

