

Received:
2 October 2017
Revised:
10 April 2018
Accepted:
31 July 2018

Cite as: G. Klarenberg,
R. Muñoz-Carpena,
M. A. Campo-Bescós,
S. G. Perz. Highway paving in
the southwestern Amazon
alters long-term trends and
drivers of regional vegetation
dynamics.
Heliyon 4 (2018) e00721.
doi: [10.1016/j.heliyon.2018.e00721](https://doi.org/10.1016/j.heliyon.2018.e00721)



Highway paving in the southwestern Amazon alters long-term trends and drivers of regional vegetation dynamics

G. Klarenberg^a, R. Muñoz-Carpena^{a,*}, M. A. Campo-Bescós^b, S. G. Perz^a

^a *University of Florida, Gainesville, FL, USA*

^b *Public University of Navarre, Pamplona, Spain*

* Corresponding author.

E-mail address: carpena@ufl.edu (R. Muñoz-Carpena).

Abstract

Infrastructure development, specifically road paving, contributes socio-economic benefits to society worldwide. However, detrimental environmental effects of road paving have been documented, most notably increased deforestation. Beyond deforestation, we hypothesize that road paving introduces “unseen” regional scale effects on forests, due to changes to vegetation dynamics. To test this hypothesis, we focus on the tri-national frontier in the southwestern Amazon that has been subject to construction of the Inter-Oceanic Highway (IOH) between 1987 and 2010. We use a long-term remotely sensed vegetation index as a proxy for vegetation dynamics and combine these with field-based socio-ecological data and biophysical data from global datasets. We find 4 areas of shared vegetation dynamics associated with increasing extent of road paving. Applying Dynamic Factor Analysis, an exploratory dimension-reduction time series analysis technique, we identify common trends and covariates in each area. Common trends, indicating underlying unexplained effects, become relatively less important as paving increases, and covariates increase in importance. The common trends are dominated by lower frequency signals possibly embodying long-term climate variability. Human-related covariates become more important in explaining vegetation dynamics as road paving extent increases, particularly family density

and travel time to market. Natural covariates such as minimum temperature and soil moisture become less important. The change in vegetation dynamics identified in this study indicates a possible change in ecosystem services along the disturbance gradient. While this study does not include all potential factors controlling dynamics and disturbance of vegetation in the region, it offers important insights for management and mitigation of effects of road paving projects. Infrastructure planning initiatives should make provisions for more detailed vegetation monitoring after road completion, with a broader focus than just deforestation. The study highlights the need to mitigate population-driven pressures on vegetation like family density and access to new markets.

Keywords: Environmental science, Geography

1. Introduction

The Amazon region in South America holds the largest areas of forest in the world [1, 2], which are of great importance to the global and regional climate system, biodiversity and carbon sequestration [3, 4]. There is great concern about changes in forest cover in the Amazon, and the global and local impacts of such changes [4, 5, 6]. Anthropogenic disturbances have direct impacts on forest cover by deforestation, and on forest structure and composition that vary with the extent and frequency of the disturbances. For example, logging and burning result in structural and phenological changes like reduction in basal area, vegetation composition [7], understory composition and vine cover [8]. Degradation of healthy forest reduces local, regional and global ecosystem services [4].

Road construction and infrastructure upgrading have been found to be a main driver of deforestation [9, 10, 11, 12]. Large infrastructure projects also cause land degradation, effects on abiotic processes (such as hydrology), disruption of movement of organisms and increased mortality, alteration of natural disturbance regimes (e.g. fire), and pollution. While infrastructure has been demonstrated to bring socio-economic benefits, it can also lead to rural-urban migration, and violent conflict over natural resources [11, 13, 14, 15, 16]. Increasing interest in the relationships between infrastructure development and the environment has given rise to the field of ‘road ecology’ [17], mostly focused on effects in the vicinity of roads. Most studies have considered straightforward direct impacts such as deforestation [10, 12], and localized impacts such as edge effects on vegetation [13, 18, 19]. For example, environmental impact assessments conducted for Amazonian highway construction in Brazil only considered very limited areas of impact along the roads, neglecting the potential regional impacts of road construction [11], such as forest degradation and changing vegetation dynamics in larger areas. Vegetation dynamics refer to temporal fluctuations and spatial variability in vegetation associated with

disturbance regimes resulting from natural and anthropogenic drivers. Vegetation can experience changes in for instance density, composition, growth rate, establishment and mortality in time and space. Changes in vegetation dynamics are indicative of shifting phenology and structure (the morphology and architecture of a plant community [20]), and the corresponding ecosystem services, which have social as well as ecological impacts [4]. It is therefore important to study changes beyond land and forest cover changes, and to do an analysis that addresses both natural and human processes, thus considering the system from a socio-ecological perspective.

An area where many of these issues have come to the fore is a tri-national frontier in the southwestern Amazon, the so-called MAP area after the provinces of Madre de Dios (Peru), Acre (Brazil) and Pando (Bolivia) [21, 22] (Fig. 1). It has been characterized as a “biodiversity hotspot” [1, 3, 15] where rural livelihoods are partly dependent on management of natural resources. The area is a useful example of a complex socio-ecological or coupled natural-human system [15, 16, 21]. This means that disturbances to the system can result from natural and anthropogenic drivers and their interactions [6, 23, 24, 25, 26, 27]. One important anthropogenic disturbance in MAP is infrastructure upgrading, which has been a key part of regional economic integration and development in recent decades [15, 28, 29]. In particular, the Inter-Oceanic Highway (IOH), connecting the Atlantic and Pacific Oceans and traversing the MAP region, is one of the key infrastructure projects of the Peru-Brazil axis of integration [30]. This infrastructure upgrade, road paving, is a major part of the Initiative for the Integration of Regional Infrastructure in South America (IIRSA) that targeted infrastructure investments for “strategic growth corridors for regional commerce” [16].

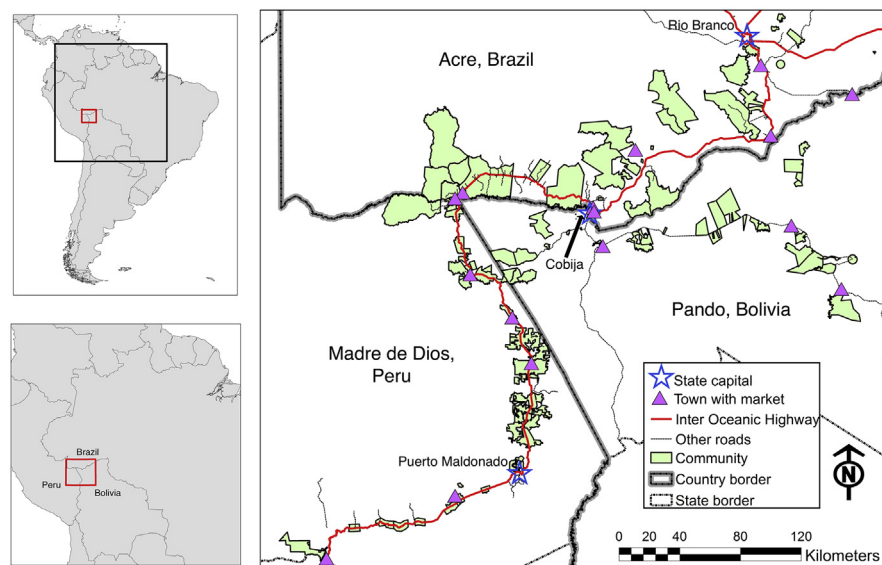


Fig. 1. Location and map of the study area and communities. Maps were created using Esri ArcGIS Arc-Map 10.5 (<http://www.arcgis.com>).

Previous research on natural and anthropogenic disturbances and vegetation dynamics found that even limited logging disturbances, without significant forest cover loss, had a permanent local effect on forest structure in Madagascar [31]. Differences in vegetation dynamics between natural and anthropogenic treefall gaps 1–4 years after logging were also identified in a Bolivian forest, despite almost identical forest cover percentages (mean 88% for the anthropogenic gaps, and 91% for the natural gaps), with lower mean numbers of flowering and fruiting plants in anthropogenic gaps, as well as more regeneration of non-commercial pioneer species in these gaps [8]. Vegetation recovery after fire differs when there is anthropogenic activity [27], and a review study concluded that in Neotropical secondary forests, the recovery trajectory of vegetation in terms of various characteristics is uncertain in anthropogenic settings, and dependent on site-specific factors and land use [32]. Changes in vegetation dynamics will in turn impact ecosystem service provision: changes in tropical forest structure have been linked to modifications in wildlife populations in Panama [33], and ecosystem productivity was found to be driven by canopy phenology in the Amazon [34]. A forest inventory study in the MAP region [35] found differences in forest value (based on biodiversity, carbon stocks and timber and non-timber forest products) across the frontier, and highlighted that deforestation and degradation do not necessarily respond similarly to road paving. Unfortunately, *in situ* vegetation structure assessments, while extremely valuable for purposes of biodiversity valuation [35], are time-intensive, limited to discrete locations and usually have a limited time dimension (synoptic), making it difficult to assess changes in vegetation dynamics over longer time periods and larger areas. This study comprehensively assessed spatially distributed changes in vegetation dynamics due to road paving to gain a better understanding of the long-term system-wide effects of infrastructure development, and thus of the resilience of the system [22, 36, 37].

Based on evidence from these previous studies, we formulated a hypothesis that even if forest cover is maintained, the vegetation dynamics will be regionally impacted over time in the presence of advancing road paving. Our hypothesis asserts that there are different forest dynamics across a gradient of road paving, from dirt road to fully paved. Further we suggest that forest change responds to different drivers as one moves along the paving gradient. Specifically, anthropogenic covariates are expected to become more important under conditions of advanced paving, which in turn is associated with increasing regional disturbances that are integral to driving and changing vegetation dynamics [38, 39].

To test this hypothesis, we required a long-term time series of vegetation dynamics that coincided with periods before and after road paving. The long-term monthly Enhanced Vegetation Index (EVI), a remote sensing product, has been found to indicate vegetation dynamics [40, 41]. It is preferred over the Normalized Difference Vegetation Index (NDVI), as the latter is more sensitive to background reflectance

and tends to saturate in high leaf areas such as tropical forests [40, 42]. Comparisons between EVI and biophysical measures such as Leaf Area Index and leaf flushing [40, 43] and other local radiometric measurements [40, 42, 43] have shown the usefulness of EVI. While EVI cannot provide information on the exact vegetation change at the local level (without site-specific validation), it is able to provide a regional overview of vegetation dynamics and variability. It has been used in previous studies to evaluate phenology, structure and ecosystem functioning [44, 45, 46, 47, 48]. We were able to match the vegetation index data with long-term field-based socioeconomic and biophysical data, which served as indicators of drivers of vegetation change. Socio-ecological survey data was available for the research area, as well as biophysical covariates from global data sets (see Methods). Paving of the IOH has occurred at different times in different countries across the MAP frontier starting in 1984, offering a valuable long-term regional study site with a road paving gradient (before, during and after).

The objective of this study was to test our road-paving regional disturbance hypotheses by applying advanced statistical time series analyses. We employed an innovative analysis framework based on hierarchical clustering and dynamic factor analysis to explore the structural and phenological changes of forests at a regional scale and determine how they are associated with the progression of road paving and natural and human drivers. We analyzed a full set of 99 unique, southwestern Amazonian communities based long-term time series of vegetation dynamics and human and environmental covariates, without having to resort to simplification techniques. Our approach provides a new way for systematic and continuous assessment of forest degradation, represented in our study area by the shift in the importance of common trends and human and natural covariates under increased highway paving.

2. Materials and methods

2.1. Study area

Included in this study are 99 areas defined as communities in the MAP region [15, 16]. Of these 99 communities, 38 are located in Madre de Dios, Peru, 25 in Acre, Brazil, and 36 in Pando, Bolivia (Fig. 1, Supplementary Table S1). The area lies between 9°48' S and 13°1' S latitude and 67°10' W and 70°31' W longitude, at the foot of the Andes Mountains and in the headwaters of tributaries of the Amazon River. Elevations are low across all communities with an average of 239.7 meters above sea level ($\sigma = 47.6$) [21]. The climate is classified as ranging from equatorial winter dry to monsoonal to fully humid (Aw, Am and Af in the Köppen-Geiger climate classification [49, 50]). The types of forest in the area are dense tropical forest, open tropical forest with palm trees, and open forests dominated by bamboo – with many

locations containing a mix of these forest types [51, 52, 53]. Detailed information on soil types in the area was unavailable as classification schemes differed between countries [21]. A previous Amazon-wide study classified soils in the MAP frontier as Acrisols, Cambisols, Ferralsols and some Lixisols. Fluvisols and Gleysols are found along rivers [54].

The Inter-Oceanic Highway running through the area was paved between 1992 (Brazil) and 2010 (Peru). Communities in Bolivia still had unpaved roads during the time period studied since road paving had not progressed much from Cobija by 2010 (Fig. 1). The communities are resource-dependent rural communities that were part of earlier studies [15, 16, 21]. They are defined as being “distinct land tenure units and/or population centers” [15]. Population densities are low, with an average family density of 0.07 families/ha and a maximum of 3.17 families/ha for the study period. Land use is described as complex and shifting, and includes urban areas, subsistence agriculture, cattle pasture, logging, gold mining, conservation areas, secondary forest and old-growth forest in which non-timber forest products (such as brazil nuts and rubber) are harvested [55].

2.2. Data

We sourced data on dynamic variables from global data sets and from previous studies in the area, for the period of road paving. Based on the time period with most available data, the period selected for analysis was 1987 (January 1987) to the start of 2010 (December 2009). We sought out the highest spatial resolutions available. Variables such as elevation and soils were assumed to be stationary throughout the study period and were not included in the time series analyses.

2.2.1. Vegetation dynamics (EVI2)

Data for the enhanced vegetation index (EVI2) used in this analysis were obtained from the University of Arizona’s Vegetation Index and Phenology (VIP) lab (Table 1). EVI2 is a relatively new product [56], consisting of EVI extended back in time from two-band Advanced Very High Resolution Radiometer (AVHRR) data (1982–1999) and three-band EVI from the Moderate Resolution Imaging Spectroradiometer (MODIS, 2000 and after). An algorithm has been applied by the VIP lab to translate two-band data from the AVHRR into MODIS EVI [56] to create the long-term EVI2 time series. The data used in this analysis (1987–2010) were obtained from the VIP lab in October 2013 in monthly time steps and at a 0.05° resolution. Area-weighted time series were extracted for each community polygon; each monthly value was an average of the values of the grid cells intersecting the community polygon, weighted by the area they covered.

Table 1. List of covariates used in the analysis, their unit of measure, and source.

	Covariate		Units	Source
Response variable	EVI2	Enhanced Vegetation Index	0 to 1	University of Arizona Vegetation Index and Phenology lab, https://vip.arizona.edu/viplab_data_explorer.php [56]
Human covariates	ENF	Enforcement of tenure rules	0 to 1: with 0 = least, 1 = most	University of Florida, Department of Sociology [15, 16, 21]
	FAM	Number of families in the community (polygon)	Count	University of Florida, Department of Sociology [15, 16, 21]
	FAMD	Family density	Families/ha	University of Florida, Department of Sociology [15, 16, 21]
	PAV	Paving	0 to 1, with 0 = no paving, 1 = fully paved	University of Florida, Department of Sociology [15, 16, 21]
	PNC	Population at state capital	Count	University of Florida, Department of Sociology [15, 16, 21]
	PNM	Population at nearest market	Count	University of Florida, Department of Sociology [15, 16, 21]
	TEN	Tenure rules: fraction of deforestation allowed of community area	0 to 1	University of Florida, Department of Sociology [15, 16, 21]
	TTC	Travel time to capital	Minutes	University of Florida, Department of Sociology [15, 16, 21]
	TTM	Travel time to nearest market	Minutes	University of Florida, Department of Sociology [15, 16, 21]
Natural covariates	AVET	Mean temperature	°C	University of East Anglia, Climate Research Unit, https://crudata.uea.ac.uk/cru/data/hrg/ [54]
	FOR	Forest area	as fraction community area	University of Florida, Department of Geography [12, 21, 59]
	MAXT	Maximum temperature	°C	University of East Anglia, Climate Research Unit, https://crudata.uea.ac.uk/cru/data/hrg/ [57]
	MINT	Minimum temperature	°C	University of East Anglia, Climate Research Unit, https://crudata.uea.ac.uk/cru/data/hrg/ [57]
	P	Precipitation	mm	University of East Anglia, Climate Research Unit, https://crudata.uea.ac.uk/cru/data/hrg/ [57]
	PET	Potential evapotranspiration	mm	University of East Anglia, Climate Research Unit, https://crudata.uea.ac.uk/cru/data/hrg/ [57]
	SM	Soil moisture	mm	NOAA Climate Prediction Center (PCP), https://www.esrl.noaa.gov/psd/data/gridded/data.cpcsoil.html [58]
	SR	Species richness	Alpha diversity	University of Florida, Department of Agricultural and Biological Engineering [60, 61]

Data correction was performed to address outliers and discrepancies between AVHRR-derived EVI (1982–1999) and MODIS EVI (2000–2010), because AVHRR-derived EVI exhibited consistently lower values (an overall average of 0.387 vs. 0.513). This was attributed to the lower quality of AVHRR data in areas with high cloud density (pers. comm. Dr. K. Didan). The final data set generally had the lowest EVI2 occurring in the driest months (June, July and August). Minima across the 99 communities ranged from 0.05 to 0.38. EVI2 peaks during the wet season, in November, December and January. Maxima ranged between 0.56 and 0.76. Variations in monthly medians across communities were larger during the wet season than the dry season (Supplementary Fig. S1).

2.2.2. Candidate covariates related to human activity

Annual data sets for a variety of covariates associated with human activity in the study area were available from previous studies [15, 21]. These data were calculated from surveys conducted in communities in 2007 and 2008 by faculty and students from the University of Florida, the National Amazonian University of Madre de Dios, the Amazonian University of Pando and the Federal University of Acre, supplemented with key informant interviews on paving status to certain cities by a given year, up to 2011. Variables used in this study were: number of families in a community (FAM), family density (FAMD, number of families/ha), population in the capital (PNC), population in the nearest market (PNM), paving (PAV), deforestation allowed under tenure rules (TEN), enforcement of tenure rules (ENF), travel time to the capital (TTC) and travel time to the nearest market (TTM), see [Table 1](#). Annual data sets were linearly interpolated to create monthly time series.

Paving was a value between 0 and 1 representing proportion of a road segment paved and was derived from field work. It is referred to as ‘paving extent’. The year when paving of the road segment along which a community sat was finalized was taken as the starting point to estimate increments in the proportion paved. For example, if a community was along a highway segment between two towns that was fully paved by 1999, that community got a value of “1” for 1999 (and subsequent years). Proportions paved in the years prior were derived from field notes of the timing of the onset and conclusion of paving of that segment, with a linear interpolation of paving proportions during the intervening time period. The value “0” for a period indicated that the road or highway segment was not paved or upgraded (yet).

Travel times, representing connectivity, were measured in minutes and based on the distance to the capital or nearest market, and average travel speeds taking into account paved and non-paved segments. For some communities, the nearest market was also the capital, rendering the same time series for these covariates ([Fig. 1](#)).

Deforestation allowed under tenure rules and enforcement of tenure rules were similarly based on fieldwork, which included workshops with stakeholders, and official rules for resource use given by governments in policy and planning documents. Deforestation allowed under tenure rules was simply the percentage of forest a community was allowed to cut down. Values vary between 0 and 1, representing 0–100% deforestation allowed. Enforcement of tenure rules were perceptions by experts as to the extent to which those use rules are actually enforced by government agencies responsible for oversight, which roughly corresponds to the probability of infractions being detected and punished. Also taken into account was knowledge of known police actions in communities. Here again, the values ran from 0 to 1, with higher values indicating more likely enforcement.

2.2.3. Candidate covariates associated with biophysical processes

Monthly data sets were sourced in June 2013 from the Climatic Research Unit (CRU) at the University of East Anglia (Table 1). Covariates included the mean, minimum and maximum temperatures, precipitation and potential evapotranspiration (AVET, MINT, MAXT, P, PET). The mean, minimum and maximum temperatures (in °C) and precipitation (in mm) were obtained at a resolution of $0.5 \times 0.5^\circ$ [57], and were assigned to each community polygon in an area-weighted manner. Potential evapotranspiration (in mm) was also included in the CRU data set, and was calculated from a variant of the Penman-Monteith formula, using mean, minimum and maximum temperature, vapor pressure and cloud cover [57].

Soil moisture (SM) came from the NOAA Climate Prediction Center (CPC) model at a resolution of $0.5 \times 0.5^\circ$, which uses CPC precipitation data and temperature data from the NCEP/NCAR Reanalysis (Table 1) [58]. The data were provided as average soil moisture in terms of water height equivalents (mm). As with the other data sets, the soil moisture data were calculated as an area-weighted time series for each community polygon.

Forest area as a percentage of community polygon area (FOR) was sourced from deforestation and land cover change studies conducted in the area (Table 1) [12, 21, 59]. Forest and non-forest percentages for each community were available for the years 1986, 1991, 1996, 2000, 2005 and 2010, and were interpolated to obtain monthly values.

Inferred species richness (SR, alpha diversity), was computed from Landsat imagery by applying a texture-based method from earlier studies; this method is based on the Shannon entropy of pixel reflectance for the green band (Table 1) [60, 61]. At the microscale different species have different reflectance levels, and at the macroscale the gradient of Shannon entropy is linearly proportional to the gradient of species richness. This method regards higher ranges of reflectivity as higher entropy, thus

assessing species diversity as a macro-ecological variable rather than a precise count of all species in a geographical area. The method from the previous studies [60, 61] was applied to 30 m resolution Landsat images at the annual scale [62], with linear interpolation to obtain monthly time steps.

2.3. Statistical analysis

We applied three main steps aimed at reducing the dimensionality of multivariate time series and identifying drivers of vegetation dynamics (Fig. 2a–c). Dimension reduction is an important step since the number of parameters in a time series model increases quickly with the number of time series covariates involved [63, 64, 65]. After compiling and quality checking the extensive time series data set (Fig. 2a), we identified areas of similar spatiotemporal vegetation dynamics along a road paving gradient by means of cluster analysis (Fig. 2b). For each of the resulting clusters with common vegetation dynamics drivers, we separated and attributed the relative importance of latent effects (unexplained shared variance, also termed common trends) and explanatory natural and human covariates (explained variance or direct

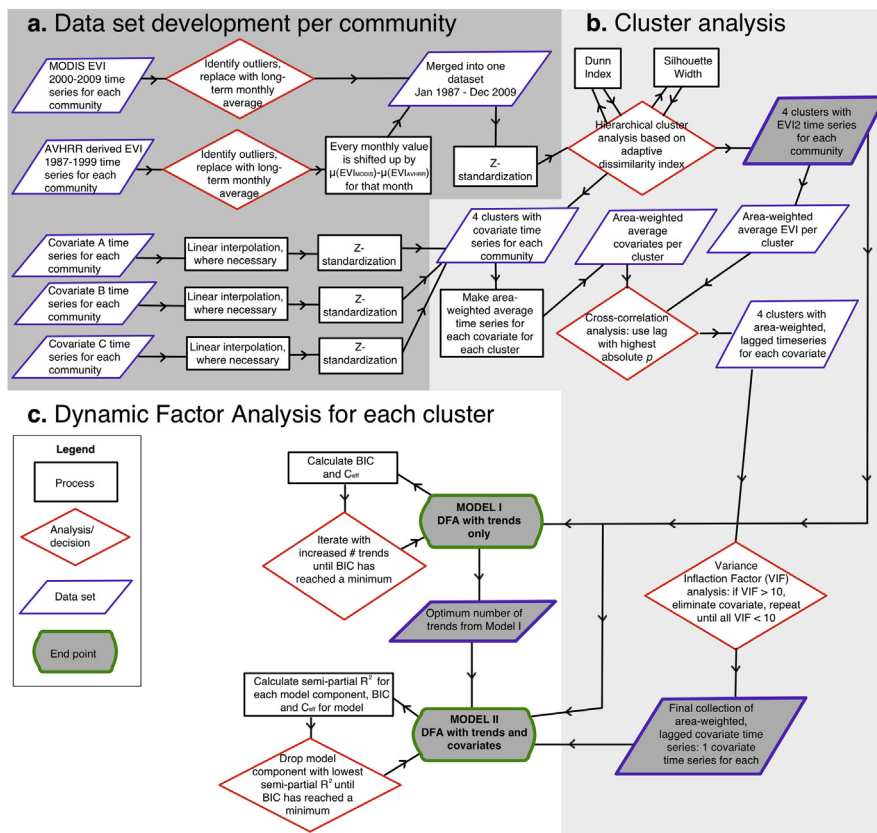


Fig. 2. Flow chart of methods.

effects) by means of Dynamic Factor Analysis (DFA, Fig. 2c), a specialized dimension-reduction time-series analysis technique [66, 67, 68, 69].

2.3.1. Clustering, lagging and reduction of covariate data set

To find the “optimum” number of clusters representing shared vegetation dynamics, 3 steps were implemented (Fig. 2b): 1) calculation of a dissimilarity matrix, 2) application of Ward’s hierarchical clustering, and 3) calculation of the Dunn index and Silhouette width as measures of compactness and separation of the clusters.

Each EVI2 time series was normalized by subtracting the mean and dividing by the standard deviation, and a time series-based dissimilarity matrix D was established [70]. D contained aspects of Euclidian distance between time series, as well as a temporal correlation measure. First, the temporal correlation coefficient was

$$\text{cort}(S_1, S_2) = \frac{\sum_{i=1}^{N-1} (u_{i+1} - u_i)(v_{i+1} - v_i)}{\sqrt{\sum_{i=1}^{N-1} (u_{i+1} - u_i)^2} \sqrt{\sum_{i=1}^{N-1} (v_{i+1} - v_i)^2}} \quad (1)$$

in which S_1 and S_2 represent time series of length N , u and v their respective points at time i . Second, the conventional Euclidian distance measure was

$$\delta_{\text{conv}}(S_1, S_2) = \sqrt{\left(\sum_{i=1}^N (u_i - v_i)^2 \right)} \quad (2)$$

These measures were used in the dissimilarity matrix D calculation as

$$D(S_1, S_2) = f(\text{cort}(S_1, S_2)) \cdot \delta_{\text{conv}}(S_1, S_2) \quad (3)$$

with

$$f(x) = \frac{2}{1 + e^{k \cdot x}}, \quad x \geq 0 \quad (4)$$

with parameter $k \geq 0$. Therefore, the proximity measure was related to similarity of values and time series behavior. Behavior is defined as the increase or decrease of values between points in time, as well as the rate of this change – incorporated in the temporal correlation coefficient. Parameter k in the automatic adaptive tuning function (Eq. (4)) determined how much behavior and value proximity contribute to D . At $k = 0$, D is only based on values, and at $k > 5$ behavior contributes 100%.

Ward’s hierarchical clustering was then applied to D to group communities together [71, 72]. Ward’s method calculates the distance between clusters as the increase in the sum of squares if they are merged. All series start out with the value zero, as they are all regarded as their own cluster, and series or clusters are merged for those that have the smallest distance increase. The Dunn Index [73] and the Silhouette Width

[74] were used as metrics to determine the suitability of the clusters. These are widely used metrics that focus on compactness and separation of clusters [75, 76, 77, 78, 79]. They are known as internal validation measures: evaluation of the clusters is based on the data and clustering only and no additional ('outside') information is required. The number of clusters with the highest Dunn index indicated the optimum number of clusters; i.e. the highest separation between clusters, and the least spread of data within clusters. The index was the ratio of the minimum distance between observations not in the same cluster, and the maximum distance between observations within a cluster. The Silhouette width also needed to be optimized. It ranged from -1 to 1, with higher values meaning that the clusters were cohesive and well separated. This measure took the lowest average distance of a series with other clusters and subtracted the average distance of the series with its own cluster. The "silhouette" is then the ratio of this number to whichever average distance is the highest: if the average distance with the neighboring cluster is larger than the average distance within-cluster, this yields a positive number ≤ 1 . The average silhouette is the final metric. The Dunn index and Silhouette width were calculated for 2 to 10 clusters with varying values for k to determine the appropriate k for Eq. (4) and number of clusters. All candidate covariate time series were clustered according to the regions that resulted from the EVI2 clustering analysis, and subsequent analyses were applied to each cluster independently.

To evaluate paving in relation to the clusters, two paving measures were calculated from the paving extent (PAV) dataset. The first was a stationary measure, the average paving for each community for the study period. The second entailed a time series, the area-weighted average paving extent per cluster over the study period.

To account for delayed responses of vegetation dynamics to covariates, lags were applied to candidate covariates (Fig. 2b). All candidate covariate time series (1987–2010) were reduced to a single time series for each cluster as an area-weighted average. Based on the highest statistically significant cross-correlation with area-weighted EVI2 of its clustered spatial region, each regional candidate covariate was lagged anywhere from 0 to 19 time steps (months, τ). Variance Inflation Factor (VIF) analysis was applied to candidate covariates in each cluster to detect collinearity ($VIF \geq 10$, Fig. 2c) [80, 81], with covariates with the highest VIF excluded from the data set in an iterative manner until a final covariate data set remained with all VIFs < 10 .

2.3.2. Dynamic factor analysis

Dynamic Factor Analysis (DFA) is a dimension-reducing statistical analysis that was applied to explore relationships between response covariates and common (shared) covariates in dynamic systems over time [63, 81, 82, 83]. It aims to explain shared

variation of observed time series using a set of common trends, with the number of common trends being substantially smaller than the number of observed time series. These common trends across time series model temporal variation across the response covariates as linear combinations and represent driving factors across time series. The coefficient (factor loading) associated with a common trend in a model for a time series gives an indication of the importance of that common trend for that time series. The trends are estimated using an Expectation Maximization (EM) algorithm, which performs maximum likelihood estimation in situations where there are latent random covariates in a model [81, 84, 85]. Covariates can be added to the model in a linear fashion. The Dynamic Factor Model (DFM) being estimated was

$$S_n(t) = \sum_{m=1}^M \gamma_{m,n} \alpha_m(t) + \mu_n(t) + \sum_{k=0}^K \beta_{k,n} \nu_k(t) + \varepsilon_n(t) \quad (5)$$

$$\alpha_m(t) = \alpha_m(t-1) + \eta_m(t) \quad (6)$$

$S_n(t)$ is the value of the n -th response variable (i.e. EVI2 time series for communities within each cluster) at time t ; $\alpha_m(t)$ is the m -th unknown common trend at time t ; $\gamma_{m,n}$ represents the unknown factor loadings; $\mu_n(t)$ is the n -th constant level parameter for displacing each linear combination of common trends up and down; $\beta_{k,n}$ represents the unknown regression parameters for the K covariate time series (i.e. one time series for each covariate type per cluster) $\nu_k(t)$; $\varepsilon_n(t)$ and $\eta_m(t)$ are the error components. $\varepsilon_n(t)$ can be interpreted as the process errors of the hidden trends, and $\eta_m(t)$ as the observation errors, which are independent Gaussian noise with zero mean and a variance-covariance matrix that can take different forms. For this analysis, the variance-covariance matrices were implemented as symmetric (diagonal) matrices. The constant level parameter ($\mu_n(t)$) was set to zero since normalized data was used.

The Dynamic Factor Models (DFMs) were assessed based on their goodness-of-fit (Nash-Sutcliffe coefficient of efficiency, C_{eff}) and parsimony (Bayesian Information Criterion, BIC). The Nash-Sutcliffe coefficient of efficiency [82, 86] was calculated as

$$C_{eff} = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (7)$$

where O and P represent the observed values and model estimates of length N , and \bar{O} the mean of the observed values. The ratio of the mean squared error of observed and predicted values and the variance of the observed values (the subtrahend in Eq. (7)) will be very small at good model fits, hence C_{eff} takes the values $-\infty \leq C_{eff} \leq 1$. The BIC [87] evaluates the goodness-of-fit and penalizes for the number of parameters in order to obtain a parsimonious model:

$$BIC = -2 \ln \mathcal{L} + k \ln(N) \quad (8)$$

where \mathcal{L} is the maximized value of the likelihood function for the model, k is the number of parameters estimated in the model and N is time series length. The aim throughout model development was to add enough covariates to explain most of the shared variance across all response covariate time series, so that reliance on the trends was minimized, with minimal lowering of the goodness-of-fit or worsening model parsimony [66, 67, 68, 82, 83, 88].

DFMs I contained only common trends, DFMs II contained both trends and covariates. The best DFMs II (with covariates) for each cluster were obtained by stepwise backward elimination of covariates or trends until the lowest BIC was reached. We started with initial ‘full’ models with all covariates and the number of trends from the best DFM I (lowest BIC). Elimination was based on the average importance of each model component per cluster. The importance of trends and covariates in the DFMs was examined by variance partitioning [81, 89, 90]. This was implemented as the average semi-partial R^2 over all possible orders of model components, as well as their ‘relative importance’ [91, 92]. The latter measure is relative to the total coefficient of determination (R^2), is non-negative and the values for all model components add up to 1, or 100% of R^2 (also known as the Lindeman, Merenda and Gold (LMG) method [92]). Since DFA estimated a unique DFM for each community, based on common trends and covariates, we averaged the relative importances of each trend and covariate across the cluster, and then eliminate the one with the lowest average importance. This approach of taking the average semi-partial R^2 over all possible orders of model components was driven by the fact that the order in which non-orthogonal regressors appear in the regression determines the amount of variance they explain – which could lead to biased results if not all orders are taken into account.

Finally, to gain insight into the periodic components of the common trends, spectral analysis was performed [93]. If time series of physical processes are (partially) a result of the sum of various frequencies [94], spectral analysis can help disentangle them. Frequencies with the highest power spectral density contribute most to explaining variance in relative terms (this method does not quantify absolute contribution). Dominant periods or frequencies were identified by spectral density estimation: higher values indicate relatively more importance of a particular frequency in explaining oscillations in the trend. For each trend, the spectral density was estimated using a fast Fourier transform with a modified Daniell kernel with dimension 2 as a smoother, and the 3 frequencies with the highest densities were selected for display in the results.

2.4. Software used and data availability

All data cleaning was done in R 3.2.4 [95] and Python 2.7 [96]. Maps were rendered in ArcGIS ArcMap 10.5. R (3.2.4–3.3.3) was used for all subsequent analyses and

figures. The following R packages were used: TSclust (adaptive dissimilarity index [70, 97]), cIValid (cluster validation [98], adjusted to be able to work with the adaptive dissimilarity index as input), fmsb (VIF [99]), MARSS (DFA [85]), relaimpo (relative importance in linear regressions [92]), hydroGOF (Nash-Sutcliffe coefficient [100]), ggplot2 [101], reshape2 [102], zoo [103], gridExtra [104], cowplot [105] and gtable [106] (figures).

The datasets generated and/or analyzed for the current study, and code, are available from a figshare repository, <https://doi.org/10.6084/m9.figshare.c.3858064>.

3. Results

3.1. Identification of clusters of vegetation dynamics and their association with road paving extent

Hierarchical cluster analysis of normalized monthly EVI2 time series (1987–2010) showed that for $k = 2$ (Eq. (4)) and 4 clusters, the Dunn index was highest (0.085) as well as the Silhouette width (0.36, Supplementary Figs. S2–3). These 4 clusters will be referred to as Vegetation Dynamics Clusters (VDCs) in this study (Fig. 3a).

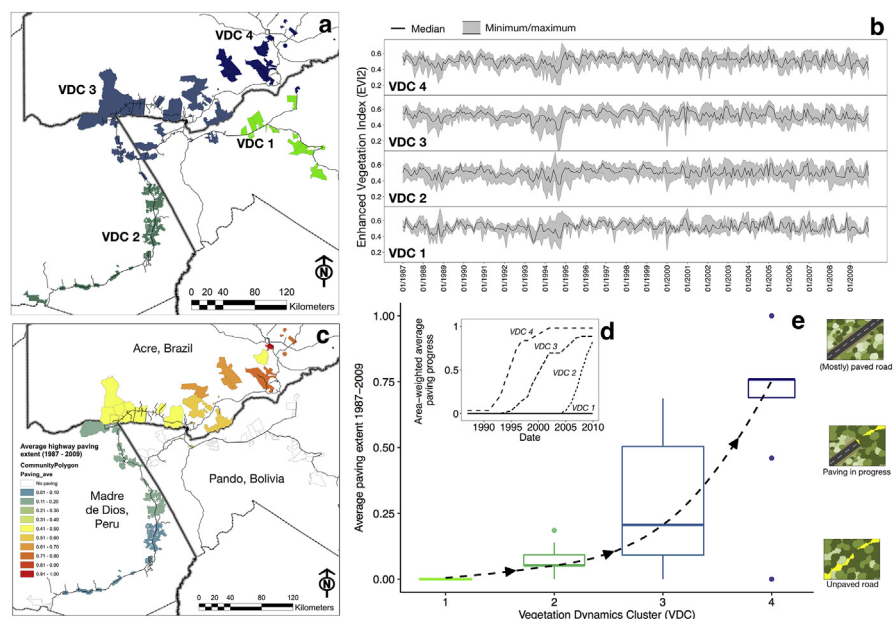


Fig. 3. Characteristics of the study area and clustering analysis results. a) Map of the study area, with 4 Vegetation Dynamics Clusters (VDCs). VDCs are based on the adaptive dissimilarity index of the Enhanced Vegetation Index (EVI2). Maps were created using Esri ArcGIS ArcMap 10.5 (<http://www.arcgis.com>). b) Minimum, median, maximum monthly EVI2 time series per VDC. c) The study area with average paving extent for the period 1987–2009 for 99 communities. Values range from 0 to 1, indicating that road sections associated with communities are unpaved, to fully paved. d) Area-weighted average paving extent over time per VDC. e) Average paving extent of the communities in each VDC for the study period, with an upward non-linear tendency from VDC 1 to 4. The tendency is a loess curve through the medians.

Characteristics of the EVI2 time series of each VDC are given in Fig. 3b and Supplementary Table S2, and of the candidate covariates in Supplementary Table S3. The normalized values that were used in this study retained the focus on the time series dynamics and facilitated interpretation of model results.

Average paving for each community for the period January 1987 to December 2009 (Fig. 3c) and the area-weighted average paving extent per VDC over the study period (Fig. 3d) were calculated in relation to the 4 VDCs. Similar patterns of increased paving extent were identified between VDCs and their average paving over time (Fig. 3d), and between VDCs and average paving extent of communities (Fig. 3e). Thus, VDC 1 represented the unpaved system state, VDC 2 and VDC 3 transition states, and VDC 4 the mostly paved system state. Subsequent analyses and models were conducted and developed separately for each VDC. The candidate covariates were grouped per VDC (Supplementary Table S3), and area-weighted averages were calculated. The appropriate lags (τ) were determined based on the highest cross-correlation and applied (Supplementary Table S4). VIF analyses yielded the final covariate sets for each VDC (Supplementary Table S5).

3.2. Common trends and shared variability in each VDC (DFM I)

Dynamic Factor Models (DFMs) for VDC 1 through 4 simulated EVI2 with common trends only (DFMs I) and identified 4, 7, 6 and 3 distinct trends, respectively (Table 2), based on the lowest Bayesian Information Criterion (BIC). The median goodness-of-fit (Nash-Sutcliffe coefficient, C_{eff}) ranged between 0.67 and 0.76, indicating that the models captured the shared variability within regions well. The majority of DFMs I for VDC 1 through 4 had an acceptable $C_{\text{eff}} > 0.60$, for 83%, 75%, 81% and 71% of the communities, respectively [69] (Supplementary Table S6 for details). Overall, 98% of C_{eff} were higher than 0.50 (97 out of 99), which is deemed a good overall fit [67, 107] considering the possible residual noisiness of AVHRR-transformed data [108]. The analysis thus identified regional, shared variance – common trends – underlying the vegetation dynamics for each VDC.

3.3. Importance of covariates across VDCs (DFM II)

Based on the identification of common trends, we developed DFMs with both trends and the area-weighted averaged covariates (DFM II) to attempt to explain the shared variance within each VDC. The covariates included in model building differed across VDCs, based on the VIF analysis results (Supplementary Table S5), and as a result the order of importance of covariates was not the same for the models (Table 2, see Supplementary Table S7 for more detailed results). The results showed that final DFM IIs (bold in Table 2) were based on the same number of trends as DFM Is, but with 6–8 covariates each. Each DFM II contained both natural and human covariates. The models' goodness-of-fit C_{eff} stayed the same or increased (C_{eff}

Table 2. Results of Dynamic Factor Analyses of Enhanced Vegetation Index (EVI2) for 4 Vegetation Dynamics Clusters (VDCs). Dynamic Factor Model I (DFM I): only trends are fitted, no covariates. Dynamic Factor Model II (DFM II): both trends and covariates are fitted. Covariates are listed according to their relative importance in each model, covariates in italics are human (see Table 1 for the abbreviations). BIC is the Bayesian Information Criterion, C_{eff} is the Nash-Sutcliffe coefficient of efficiency. Model results in bold are the selected models for further discussion.

VDC	DFM	Number of trends	Covariates	BIC	Median C_{eff} (95% confidence interval)
1, unpaved ($n = 18$)	I	1		10205	0.54 (0.27–0.81)
	I	2		10011	0.62 (0.30–0.85)
	I	3		9876	0.71 (0.32–0.87)
	I	4		9789	0.74 (0.49–0.87)
	I	5		9796	0.75 (0.50–0.89)
	II	4	<i>ENF AVET MINT PET FOR SR SM FAMD</i> MAXT P	9994	0.75 (0.52–0.87)
	II	4	<i>ENF AVET MINT PET SR SM FOR MAXT</i> <i>FAMD</i>	9958	0.75 (0.52–0.88)
	II	4	FOR <i>ENF AVET MINT PET SR SM MAXT</i>	9923	0.75 (0.51–0.87)
	II	4	FOR AVET MINT <i>ENF</i> PET SM SR	9896	0.75 (0.50–0.87)
2, transition ($n = 24$)	II	3	FOR AVET MINT SR PET <i>ENF</i> SM	9939	0.73 (0.40–0.86)
	I	1		14379	0.41 (0.14–0.84)
	I	2		10915	0.49 (0.18–1.00)
	I	3		10305	0.60 (0.24–1.00)
	I	4		10144	0.72 (0.29–1.00)
	I	5		10055	0.77 (0.39–1.00)
	I	6		10043	0.80 (0.48–1.00)
	I	7		10007	0.80 (0.51–1.00)
	I	8		10034	0.82 (0.53–1.00)
	II	7	<i>TTM PAV PET SM MAXT P MINT AVET</i> SR <i>TEN</i>	10390	0.83 (0.54–1.00)

(continued on next page)

Table 2. (Continued)

VDC	DFM	Number of trends	Covariates	BIC	Median C_{eff} (95% confidence interval)
3, transition ($n = 43$)	II	7	PAV TTM PET SR MAXT SM P MINT AVET	10323	0.83 (0.54–1.00)
	II	7	PAV TTM SR PET SM MAXT P MINT	10266	0.83 (0.54–1.00)
	II	7	PAV TTM PET SM MAXT P SR	10244	0.82 (0.53–1.00)
	II	7	PAV TTM PET SM MAXT SR	10262	0.83 (0.46–1.00)
	I	1		23129	0.63 (0.40–0.76)
	I	2		22261	0.66 (0.41–0.82)
	I	3		21573	0.71 (0.48–0.85)
	I	4		21223	0.74 (0.55–0.90)
	I	5		18230	0.74 (0.55–1.00)
	I	6		18082	0.77 (0.56–1.00)
	I	7		18134	0.78 (0.56–1.00)
	II	6	PET SM MINT FAMD AVET MAXT P TEN SR ENF	19010	0.79 (0.59–1.00)
	II	6	PET MINT SM MAXT AVET FAMD P TEN SR	18900	0.79 (0.59–1.00)
	II	6	PET MINT SM MAXT AVET FAMD P TEN	18817	0.79 (0.59–1.00)
	4, paved ($n = 14$)	II	6	FAMD PET MINT SM MAXT AVET P	18695
II		6	FAMD PET SM MINT AVET MAXT	18590	0.78 (0.58–1.00)
II		5	FAMD PET SM MINT AVET MAXT	18664	0.77 (0.58–1.00)
I		1		8224	0.59 (0.35–0.77)
I		2		7992	0.66 (0.35–0.89)
I		3		7986	0.68 (0.36–0.89)
I		4		8007	0.70 (0.39–0.91)
II		3	FAMD TTM PET MINT AVET MAXT P ENF SM TEN	8100	0.69 (0.38–0.90)
II		3	FAMD TTM PET MINT AVET MAXT ENF P SM	8076	0.68 (0.38–0.90)
II		3	FAMD TTM PET MINT AVET MAXT ENF P	8054	0.69 (0.38–0.90)
II	3	FAMD PET MINT TTM AVET MAXT ENF	8056	0.69 (0.38–0.90)	

≥ 0.50 for 98% of DFM IIs, see Supplementary Fig. S4 for an example of model fits), indicating that the importance of the trends became smaller. In DFM II a portion of the unexplained variance of DFM I was shifted to the covariates (explained variance) while maintaining model fit. The importance of all trends and covariates was calculated and compared to average paving extent in communities. While total explained variance (R^2) stayed fairly constant across all average paving extent (Fig. 4a), once average paving extent passed 0.50 (50%), the importance of trends and covariates shifted (Fig. 4b). Specifically, covariates explained approximately 25% of variance under unpaved conditions, and 50% under paved conditions, while trends decreased their explanatory power. This implies that for communities where paving started longer ago, much of the vegetation dynamics can be explained with covariates directly. Supplementary Tables S8–S11 contain details on covariate time series, model fits, relative importance of model components, and weighting and loading coefficients for each community.

3.4. Importance of human covariates across VDCs (DFM II)

The explained variance by covariates in the DFMs II was subdivided into that associated with natural and human covariates, indicating the importance of each group in the models (Table 1, Fig. 4c). Communities with more recent road paving (low average road paving extent, 0.10) showed an increase in variance explained by human covariates compared to communities with no road paving at all (Fig. 4c), but this decreased again for communities with higher average road paving (0.25). Eventually though, the importance of human covariates increased as paving extent increased, and there was an inversion of importance of natural and human covariates that occurred after paving extent reached past the halfway point (Fig. 4c).

Travel time to market, family density and enforcement of tenure rules were the most important human covariates in the DFM IIs (highest average semi-partial R^2 , Fig. 5). The covariate time series used in the DFM IIs for each VDC are summarized in Fig. 6. The standardized regression coefficients (β) of the DFM IIs gave additional information beyond the explained variance as their signs indicate an inverse (negative) or direct (positive) effect of the covariates on vegetation dynamics (Fig. 7). Human covariates were applied with lags ranging from 5 to 13 months at higher paving extents.

Family density showed the largest increase in importance with increased paving extent. The effect is negative (negative β): higher family density implies lower EVI2. Similarly, the importance of travel time to market is related to the average paving extent of communities in each VDC. For later paved communities (low average paving extent, VDC 2), the decrease in travel time is associated with decreased EVI2 (positive β). However, for communities that had paved roads earlier in the study period (higher average paving extent, VDC 4), decreased travel time is

associated with increased EVI2 (negative β). Enforcement of tenure rules on deforestation also had a positive effect on EVI2 in communities that had unpaved roads during the study period (VDC 1), as expected. While enforcement also increased over time for the communities that had paving longer, in VDC 4 (Fig. 6) its effects appeared negative (i.e. lower EVI2).

3.5. Importance of natural covariates across VDCs (DFM II)

Temperature time series played a role in all VDCs (Fig. 5). Normalized minimum, maximum and average temperature had different temporal patterns (Fig. 6) and combined appeared relevant in simulating vegetation dynamics (Fig. 5): on average, across all communities, the variance explained by the three temperature time series (average semi-partial R^2) taken together is 32% of all variance explained by natural covariates. In terms of hydrological controls, precipitation was less important than potential evapotranspiration and soil moisture (Fig. 5). For the latter two, their importance decreased in areas with higher average road paving extent, soil moisture more so than potential evapotranspiration. The type of effect of soil moisture (the β coefficient) was mixed in cases with less road paving (Fig. 7). As road paving increased to 0.5, it became more distinctly negative ($-0.40 < \beta < 0.0$), meaning

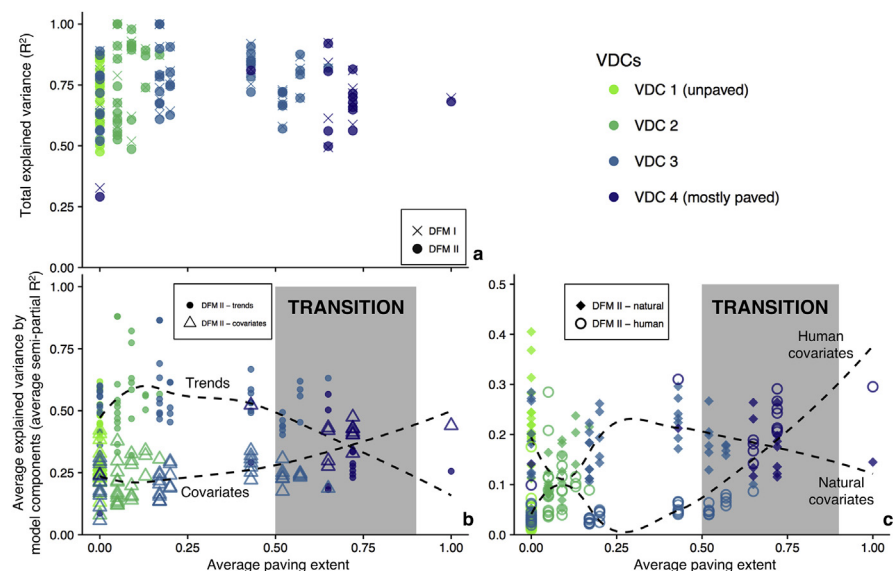


Fig. 4. Contributions of model components of the final VDC DFMs II for each community to explaining variance in vegetation dynamics. Average paving extent of each community is plotted along the x-axis, colors indicate the VDC. a). Proportion of explained variance, R^2 , of the DFMs I and II. b) Average proportion of explained variance over all possible orders of the model components, average semi-partial R^2 , for trends and covariates. Loess curves indicate respectively a downward and upward tendency with increased paving extent, with a transition identified between a paving extent of 0.50 and 0.90. c) Average proportion of explained variance over all possible orders of the model components, average semi-partial R^2 , for natural and human covariates. Loess curves indicate a downward and upward trend respectively with increased paving extent, with a transition identified between a paving extent of 0.50 and 0.90.

lower soil moisture lead to a higher EVI2. Overall, soil moisture dynamics appear to be an important control in vegetation dynamics, but this effect disappears when average road paving extent increases.

While the importance of potential evapotranspiration was fairly constant across communities (Fig. 5), the effect was negative for communities in a transition paving state (Fig. 7) until the most paved state, where potential evapotranspiration had a positive effect on EVI2. The early negative effect indicated that higher evapotranspiration results in lower EVI2. Species richness appeared important when average paving extent was below 0.25 (Fig. 5), with mostly negative coefficients in the models (Fig. 7). The forest cover percentage of a community predictably had a positive effect on EVI2 (positive β coefficient, Fig. 7), but only played a role in VDC 1 with no paving (Figs. 5 and 7).

3.6. Remaining trends and their frequency components (DFM II)

The fact that the goodness-of-fit from DFM I to DFM II remained the same was a sign that the effects of certain covariates on vegetation dynamics were present in the

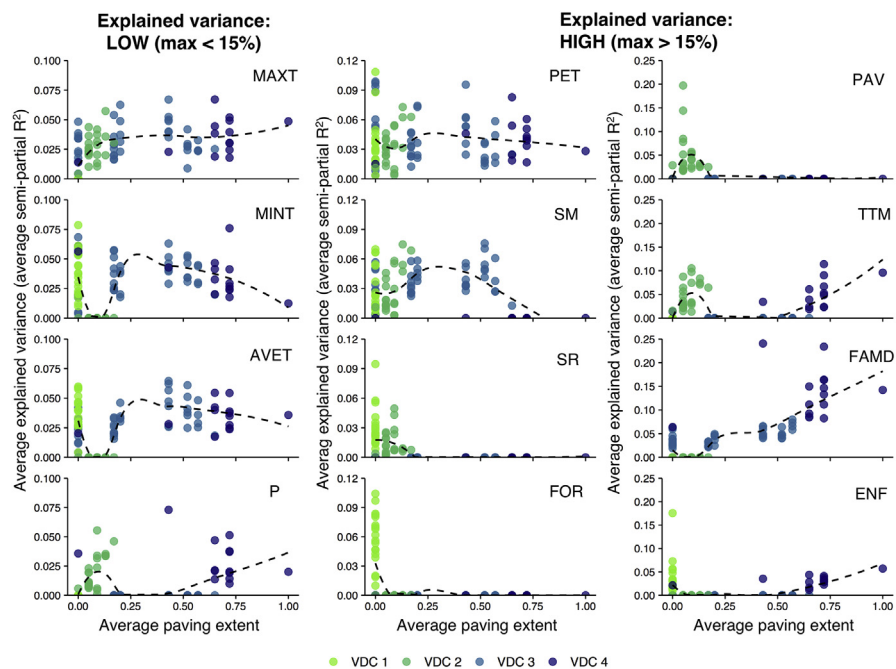


Fig. 5. Proportion of variance explained by each covariate for each community. Average paving extent of each community is plotted along the x-axis, colors indicate the VDC. Covariate plots are grouped in columns according to the maximum variance explained: low (left column) and moderate to high (middle and right column). MAXT, MINT, AVET and P are maximum, minimum and average temperature and precipitation. PET, SM, SR and FOR are potential evapotranspiration, soil moisture, species richness and forest area. PAV, TTM, FAMD and ENF are paving, travel time to market, family density and enforcement of tenure rules on deforestation.

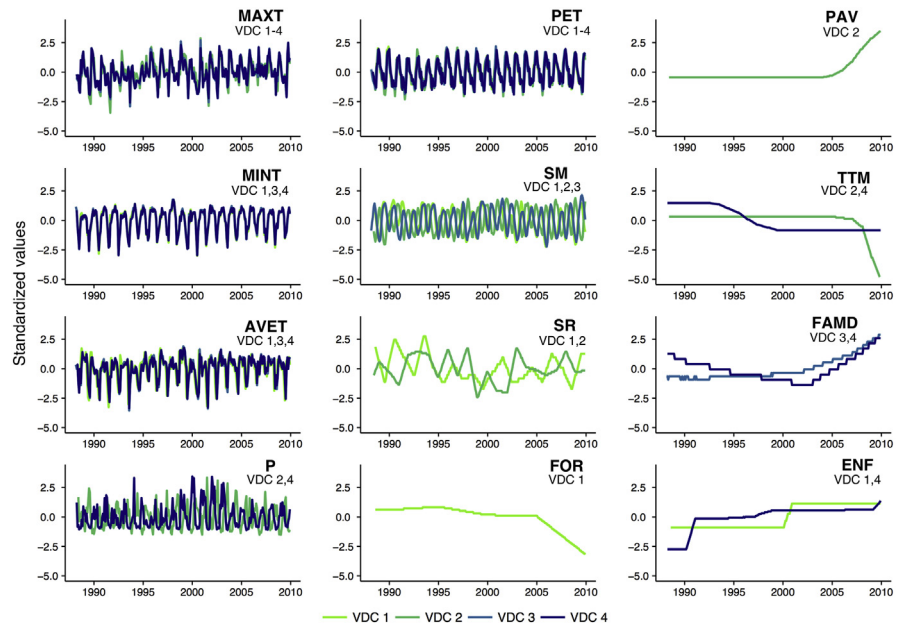


Fig. 6. Lagged time series of the covariates used in the final DFMs II for each VDC. The applied lags are specified in Supplementary Table S4. Colors are associated with VDCs. Not every covariate is used for each VDC, selection is based on Variation Inflation Factor (VIF) analysis. MAXT, MINT, AVET and P are maximum, minimum and average temperature and precipitation. PET, SM, SR and FOR are potential evapotranspiration, soil moisture, species richness and forest area. PAV, TTM, FAMD and ENF are paving, travel time to market, family density and enforcement of tenure rules on deforestation.

common trends in DFM I and were removed from the trends in DFM II. The variance explained by each of the common trends for each community (expressed as average semi-partial R^2) differed per community (Fig. 8a). The common trends over time are summarized in Fig. 8b. It is of interest to note that even for communities with a similar paving history (x-axis for Fig. 8a), there was a range in the variance explained by each common trend across communities.

The spectral power density of signals contained in the common trends was determined, and for each common trend the three signals with the highest density were visualized in Fig. 8c. Low-frequency signals (22.5, 11.25 and 5.625 years) appear across all VDCs. A number of common trends in DFMs II contain 4.5 and 7.5 year frequencies. There was a decrease in higher frequency signals (<18 months) across the VDCs with increased paving extent (Fig. 8c). Especially for VDC 4, more of the variance of the common trends was captured in the low frequency signals.

3.7. Unpaved and paved states as alternative system states of vegetation dynamics

A transition state between unpaved and paved conditions was identified across VDCs (Fig. 4b–c). There was a shift in the importance of trends and covariates

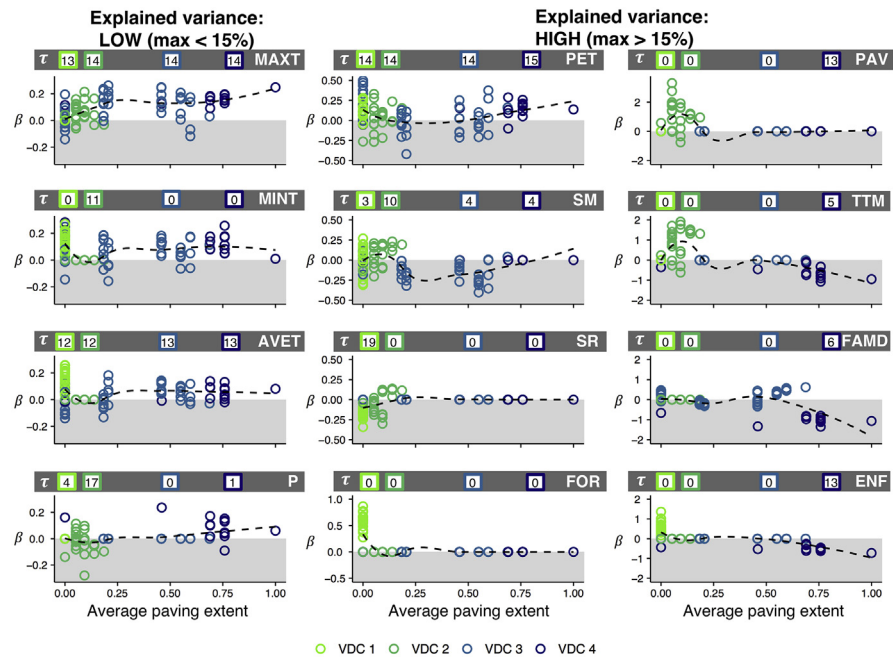


Fig. 7. β coefficients of the covariates used in the final Dynamic Factor Models (DFMs II) for each VDC. The applied lags (τ) are specified in the grey bar above the plots, and in Supplementary Table S4. Negative covariates (in the grey area in the plot) imply that the covariate has an inverse (negative) effect on EVI2. Colors are associated with VDCs. Grouping of subplots in columns is according to the extent of explained variance of the covariates from Fig. 5: note the differences in y-axis scaling for the columns. MAXT, MINT, AVET and P are maximum, minimum and average temperature and precipitation. PET, SM, SR and FOR are potential evapotranspiration, soil moisture, species richness and forest area. PAV, TTM, FAMD and ENF are paving, travel time to market, family density and enforcement of tenure rules on deforestation.

explaining EVI2 dynamics, as well as a shift in importance between natural and human covariates, as was noted above. Both of these shifts happened around an average paving extent of 0.75.

4. Discussion

4.1. Main findings

We have found distinct areas with common temporal vegetation dynamics, associated with road paving progression in the southwestern Amazon, supporting our hypothesis that vegetation dynamics are altered with increased road paving. While a previous study put the area in the same phenoregion [109], this research showed there are differences at the regional scale, associated with road paving. With a time series dimension reduction technique, DFA, we uncovered common trends and a number of (lagged) socio-economic and biophysical covariates that explain shared variance of EVI2 for each VDC. We found differences in the DFMs between VDCs in terms of covariates and common trends included, their importance in

explaining variance, and whether effects were positive or negative. At the extremes of the paving extent, the unpaved and paved states both have a small number of common trends but differ in terms of the relative importance of human and natural factors driving vegetation dynamics.

Human covariates are relatively more important in the paved state (family density and travel time to markets); conversely, the natural covariates are important in the unpaved state. In the unpaved state and the transition states, there is greater influence from natural variables like potential evapotranspiration, soil moisture, and minimum and average temperatures. The importance of temperature is probably due to its effect on photosynthesis, which affects phenological signatures. With increased average paving, maximum temperature becomes relatively more important than minimum and average temperature (Fig. 5). Minimum temperature is used unlagged in the models, indicating an immediate effect of this covariate in the simulation; average and maximum temperatures are lagged 12–13 and 13–14 months, respectively (Fig. 7). Maximum and average temperature potentially have a stronger effect on germination, flowering and other reproductive processes [110], which would explain the longer lag in their effect. Previous work (on deciduous broad leaf and evergreen needleleaf forest phenology) found similar importance of minimum and maximum temperature [111] on vegetation dynamics. Particularly the increased importance of maximum temperature with higher paving extent points to potentially more change in the future. Climate change is projected to affect temperatures significantly in the future, including its anomalies [112],

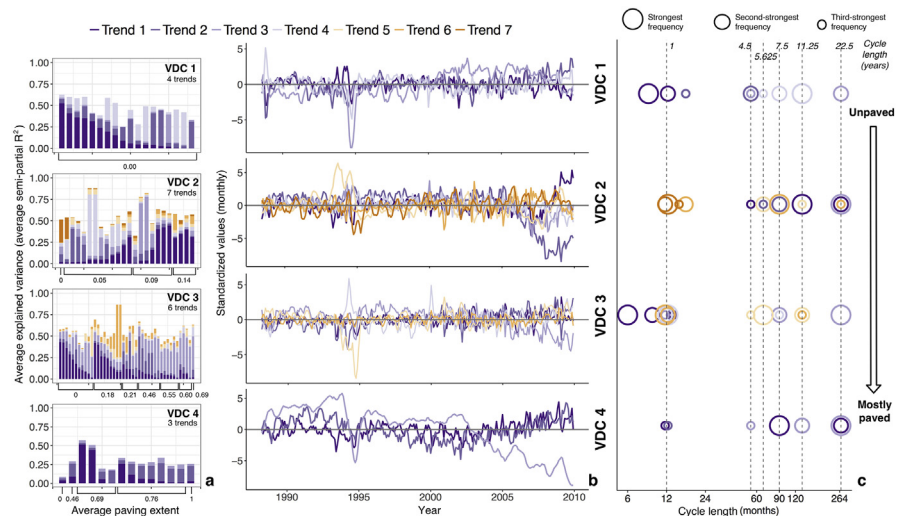


Fig. 8. Characteristics of the trends (unknown explained variance) in the selected DFMs II. a) Average semi-partial R^2 of trends for each community indicates trends contribute differently to explaining variance per community and across paving extent (x-axis). b) Monthly values of trends over time. c) The three strongest frequencies for each trend in each VDC, identified with spectral density estimation, are depicted by large, medium and small circles. Where trends have signals of the same frequency in common, not all n trends \times 3 signals are visible due to overlap.

indicating that areas under the influence of road paving disturbances will be more sensitive to these changes.

The findings on soil moisture also correspond to results from previous work. The importance of soil moisture for VDC 1, VDC 2 and VDC 3 is supported by previous research [113] which pointed to the relevance of the hydrologic cycle in combination with soil properties in these areas with mostly natural disturbances and limited anthropogenic disturbances. A key component in interpreting the negative effect of soil moisture in most communities is the lag of 3–4 months applied to this covariate. While soil moisture usually peaks at the beginning of the year during the later portion of the wet season (March–April), EVI2 generally reaches its lowest points around the middle of the year during the dry season (June–September). By implementing the soil moisture with a lag, EVI2 and soil moisture essentially displayed opposite dynamics, and thus the timing of soil moisture peaks and valleys became a factor in simulating EVI2 dynamics for most communities. For the communities where soil moisture had mostly positive effects (VDC 2) the longer lag of 10 months that was applied ensured that soil moisture and vegetation dynamics lined up again: March–April soil moisture peaks were now aligned with EVI2 peaks in December–February.

While the relative importance of potential evapotranspiration was expected [111], the negative effect on vegetation dynamics in some communities was initially counterintuitive. We expected that higher evapotranspiration would relate to more photosynthetic activity and thus higher EVI2 values. However, potential evapotranspiration data were only driven by climatic variables and did not take into account available water [68]. Since our analyses focused on normalized values and not absolute values this result did not necessarily implied a water deficit; but it could indicate that in the communities where potential evapotranspiration and vegetation dynamics behaved in an opposite manner, available water might play a role. Considering the importance of the soil moisture covariate, this was plausible: especially in the light of results for VDC 4, where soil moisture did not feature as important (Fig. 5), and all potential evapotranspiration effects were positive (Fig. 7). The opposite effect (negative coefficients) of species richness (Fig. 7) for some communities were similar to earlier findings [114], which established that there are trade-offs between plant diversity and forest structure covariates in the region, i.e. increased forest structure was associated with less diversity.

In terms of human covariates, the initial rise and then decline of their importance before the 0.25 point of paving extent (Fig. 3c) could be due to disturbance and then adaptation of the ecosystem to the disturbance regime. The detailed results show that this pattern was mostly driven by the importance of paving and travel time to markets (Fig. 5). The initial effect of travel time to markets was positive (meaning lower travel times correspond to less EVI2 and vice versa), potentially due to increased pressure and exploitation associated with better access to markets

for locals and better access to forest products by outsiders. For VDC 4, at higher paving extents, the effect is reversed. The change in travel time for this VDC is less pronounced and occurs over a longer time period (Fig. 6), so it is possible that over time communities shift their focus to more urban sources of income such as wages rather than forest-based activities, as observed in another study [115], prompting higher EVI2 over time.

A possible explanation for the seemingly counterintuitive result of a negative effect of enforcement of tenure rules on vegetation dynamics for VDC 4 are the specific circumstances in this area in Acre, Brazil. Extraction of forest resources and conversion to agriculture had been going on since the 1970s in this area, and in the 1980s several groups (ranchers, rubber tappers, colonists) clashed over land use and deforestation [116]. Most of the enforcement increase occurred in the early years (1990–1991, Fig. 6), after which enforcement stayed high and fairly constant (Supplementary Table S3 indicates the non-standardized values). While enforcement even increased slightly in subsequent years, vegetation may have been too degraded already to experience positive effects. Another possibility is that people engaged in covert or illegal forest exploitation. This is an option if the payoff from these activities is worthwhile despite enforcement.

The importance of travel time to markets was found in a previous study on land cover change in the area [21], but family density did not appear significant in that case, as it did in this study. The importance and negative effect of family density are intuitive: increasing family density (Fig. 6) leading to decreasing EVI2 could be linked to increased pressure on vegetation due to rising forest resources exploitation.

The number of trends in the unpaved and paved state was similar and low (VDC 1 and VDC 4 in Table 2 and Fig. 8a). The presence of only a few common trends explaining most of the variance across a number of communities pointed to the spatial uniformity of vegetation dynamics in the communities in these VDCs. This was possibly because the dynamics reached a similar state in all communities considering the longer duration that they were unpaved or mostly paved. This assumed that the levels of disturbance had stabilized. The transition states, with 6 and 7 trends, had more heterogeneity in paving and impacts across the communities in their regions and a variety of vegetation responses. This highlights that while paving was ongoing, changes and system states at a regional level were more heterogeneous and more difficult to predict. Once the system stabilized again (VDC 4), the system state was different than before the disturbance (VDC 1): this is visible, in addition to the DFM specification of covariates, in the frequencies contained in the trends. But the shift in importance of common trends and covariates at a paving extent of 0.75 indicated that vegetation dynamics are potentially resilient to the effects of early paving. An average paving extent of 0.75 means that paving has to be quite advanced before a shift takes place.

When paving was present sufficiently advanced though, the results show that there was a change in common trends from the unpaved to the paved system state, with trends losing both importance (Fig. 4b) and overall number of unique signals, notably high-frequency signals (Fig. 8c). We deduced that these phenomena were related: less high-frequency signals were probably due to covariates explaining more variance in the paved state directly. Within VDCs, common trends varied in their contribution to explained variance. This suggested that there were missing covariates or non-linear interactions between variables that had spatial variation in importance. Low-frequency signals that showed up consistently in each VDC (5.625, 11.25 and 22.5 years) were attributed to solar and climatic influences. Solar activity and its associated radiation have an influence on the earth's biosphere [117, 118] and its variation is acknowledged as a natural climate forcing [112]. The solar cycle has an 11- to 22-year frequency, and a 5.6-year periodicity [119]. The Pacific Decadal Oscillation (PDO), a pattern of ocean-atmosphere variability, has been found to play a role in the Amazon in previous studies [120, 121, 122]. The PDO contains a 4.8- and 8-year frequency (Supplementary Table S12), close to the 4.5- and 7.5-year frequencies also found in a number of the DFM trends.

This study adds an important dimension to our understanding of the effects of roads on vegetation. It complements a previous finding that forest degradation and deforestation do not always respond in a similar way to infrastructure change [35]: in that study road paving did not always increase degradation as it did deforestation. By adding a temporal component and a number of human and natural covariates to our analysis, we were able to distinguish the positive and negative effects of these covariates with increased road paving; contributing to the expanding body of knowledge on road impacts.

Potential phenological changes due to global change have been identified in previous studies [110, 123], and the combination of these changes with anthropogenic impacts such as road paving increases the uncertainty of future trajectories. As this could alter ecosystem functions, ecosystem services would also be affected [124]. While this study does not quantify these possible effects, it contributes to the growing body of research on road paving impacts in the Amazon [11, 21, 59]. It also highlights the effects beyond deforestation alone [12, 21] and the changes in drivers of vegetation dynamics with increased road paving.

4.2. Limitations and further research

While remote sensing products are increasingly available and useful for doing large-scale regional analyses, inherent noise in the data should be carefully evaluated. This is especially true in humid tropical regions, which exhibit extended periods of cloud cover contributing to observation error. While the models exhibit an acceptable goodness-of-fit for most communities, their performance is low in some areas partly due to observation error. Further research on signals and frequencies in EVI2 data

would be beneficial to quantify differing vegetation dynamics in more detail in order to assess errors, as well as field studies that measure biomass and vegetation structure combined with higher resolution remote sensing studies.

DFA is efficient at capturing variance in ecological time series because of its autoregressive approach to modeling trends, but if the variance is not accounted for among the covariates in expanded models and instead remains captured in the trends, it stays unexplained. This requires more detailed investigations into issues such as logging, the role of fire, and species composition in terms of functional groups. It is however challenging to obtain reliable data on these concepts for sufficiently long periods of time, for a large area.

DFM IIs add covariates linearly, and while possible lagged responses have been accounted for, interactions are not explicitly accounted for. Future research should focus on untangling interactions between both human and natural covariates, in particular with mechanistic models. This would also assist in identifying causal relationships, since DFA does not account for this. The mechanisms behind the impacts of changes in travel time and family density on vegetation dynamics should also be investigated in more detail, as these are factors that could be managed (as opposed to e.g. temperature as a driver).

5. Conclusions

This study has shown the power of Dynamic Factor Analysis when applied to a complex data set where time series are subject to change over an extended period of time and space. DFA offers a continuous overview of the dynamics in the importance of the effects of different covariates, as opposed to snapshots in time. DFA thus provides a systematic framework to study how forest degradation evolves over time as a process like road paving unfolds. Other areas in which this method could prove useful are experimental interventions in biological and health fields to assess how complex systems respond, or how application of governmental regulations affects constituency behaviors with regard to intended outcomes.

This research offers support for the hypothesis that there are areas that have distinct vegetation dynamics along a road paving gradient and provides support for the second part of the hypothesis, that different covariates and mechanisms drive vegetation dynamics across the road paving gradient. The results of the study can benefit local and regional planners, as well as conservation initiatives at a practical level. The findings show which covariates impact regional vegetation dynamics positively or negatively, and show how these differ with road paving progress. The results thus point to covariates that could be managed to minimize the impact of road paving on regional vegetation dynamics, such as enforcement of tenure rules and population density. The analysis also provides insights into the relationships of natural covariates with

vegetation dynamics; and some of these covariates could be affected by climate change. With temperature anomalies projected to change in the future [112], the effect on areas in a paved state will be stronger. Most importantly, the study indicates that vegetation dynamics - and thus potentially ecosystem services - will differ with road paving progress. Future research on changes in ecosystem services due to altered vegetation dynamics is relevant for both conservation as well as future economic opportunities in the study area. Previous research has considered degradation and ecosystem services in general [4], but more research on roads impacts specifically is needed. Both roads and ecosystem services are important to livelihoods and economic development, so infrastructure planning initiatives should account for their potential effects on regional vegetation dynamics and ecosystem services. Provisions should be made in the planning stages of a road project to monitor vegetation dynamics after project completion, beyond only deforestation. More detailed and localized technologies, such as Light Detection And Ranging (LiDAR) or products from the Fluorescence Explorer (FLEX), will provide useful in extending this analysis to more local scales [37, 125, 126]. Data from these technologies is not yet available for the longer term, and thus could not be used in this current study. Governments and research institutes should invest in these technologies for ongoing monitoring.

Lastly, this study adds a large scale regional analysis to the field of road ecology. As national governments and banks continue to pursue development and trade via infrastructure investments, road paving is likely to extend into many forested areas such as the Amazon, Central Africa and Southeast Asia, and thus pose threats of disturbing ecosystem services provision [127].

Declarations

Author contribution statement

Geraldine Klarenberg: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Rafael Muñoz-Carpena: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Miguel A. Campo-Bescós: Conceived and designed the experiments; Contributed analysis tools; Edited the paper.

Stephen G. Perz: Contributed materials and data; Edited the paper.

Funding statement

This work was supported by funding from NSF CNH Award number 1114924, USA.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2018.e00721>.

Acknowledgements

Y. Xia, L. Zhu, E. Bunting, M. Convertino, M. Marsik contributed remote sensing data and consulted on some of the methods used in this paper. R.M.C and G.K. acknowledge the University of Florida Research Computing (High-Performance Computing, <http://rc.ufl.edu>) for providing computational resources and support that have contributed to the results reported in this paper.

References

- [1] T.J. Killeen, L.A. Solorzano, Conservation strategies to mitigate impacts from climate change in Amazonia, *Phil. Trans. Biol. Sci.* 363 (2008) 1881–1888.
- [2] M. Keller, *Amazonia and Global Change*, American Geophysical Union, Washington, DC, 2009.
- [3] N. Myers, R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities, *Nature* 403 (2000) 853–858.
- [4] J.A. Foley, G.P. Asner, M.H. Costa, M.T. Coe, R.S. DeFries, H.K. Gibbs, et al., Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin, *Front. Ecol. Environ.* 5 (2007) 25–32.
- [5] C.A. Nobre, L.D.S. Borma, “Tipping points” for the Amazon forest, *Curr. Opin. Environ. Sustain.* 1 (2009) 28–36.
- [6] E.A. Davidson, A.C. de Araújo, P. Artaxo, J.K. Balch, I.F. Brown, M.M. C Bustamante, et al., The Amazon basin in transition, *Nature* 481 (2012) 321–328.
- [7] J. Zhu, Z. Mao, L. Hu, J. Zhang, Plant diversity of secondary forests in response to anthropogenic disturbance levels in montane regions of north-eastern China, *J. Res.* 12 (2007) 403–416.

- [8] A. Felton, A.M. Felton, J. Wood, D.B. Lindenmayer, Vegetation structure, phenology, and regeneration in the natural and anthropogenic tree-fall gaps of a reduced-impact logged subtropical Bolivian forest, *For. Ecol. Manage.* 235 (2006) 186–193.
- [9] D. Nepstad, G. Carvalho, A. Cristina Barros, A. Alencar, J. Paulo Capobianco, J. Bishop, et al., Road paving, fire regime feedbacks, and the future of Amazon forests, *For. Ecol. Manage.* 154 (2001) 395–407.
- [10] W.F. Laurance, A.K.M. Albernaz, C.D. Costa, Is deforestation accelerating in the Brazilian Amazon? *Environ. Conserv.* 28 (2002) 1–7.
- [11] W.F. Laurance, M. Goosem, S.G.W. Laurance, Impacts of roads and linear clearings on tropical forests, *Trends Ecol. Evol.* 24 (2009) 659–669.
- [12] M. Marsik, J. Southworth, F.R. Stevens, Amazon deforestation: rates and patterns of land cover change and fragmentation in Pando, northern Bolivia, 1986 to 2005, *Prog. Phys. Geogr.* 35 (2011) 353–374.
- [13] A.W. Coffin, From roadkill to road ecology: a review of the ecological effects of roads, *J. Transport Geogr.* 15 (2007) 396–406.
- [14] W.F. Laurance, T.E. Lovejoy, H.L. Vasconcelos, E.M. Bruna, R.K. Didham, P.C. Stouffer, et al., Ecosystem decay of Amazonian forest fragments: a 22-year investigation, *Conserv. Biol.* 16 (2002) 605–618.
- [15] S.G. Perz, L. Cabrera, L.A. Carvalho, J. Castillo, R. Chacacanta, R.E. Cossio, et al., Regional integration and local change: road paving, community connectivity, and social–ecological resilience in a tri-national frontier, southwestern Amazonia, *Reg. Environ. Change* 12 (2011) 35–53.
- [16] S.G. Perz, A. Shenkin, G. Barnes, L. Cabrera, L.A. Carvalho, J. Castillo, Connectivity and resilience: a multidimensional analysis of infrastructure impacts in the southwestern Amazon, *Soc. Indic. Res.* 106 (2011) 259–285.
- [17] R.T.T. Forman, *Road Ecology: Science and Solutions*, Island Press, 2003.
- [18] A.E. Lugo, H. Gucinski, Function, effects, and management of forest roads, *For. Ecol. Manage.* 133 (2000) 249–262.
- [19] R.C.G. Mesquita, P. Delamônica, W.F. Laurance, Effect of surrounding vegetation on edge-related tree mortality in Amazonian forest fragments, *Biol. Conserv.* 91 (1999) 129–134.
- [20] United States Department of Agriculture (USDA), National Agricultural Library (NAL) Glossary, 2015. <https://agclass.nal.usda.gov/glossary.shtml>.

- [21] S.G. Perz, Y. Qiu, Y. Xia, J. Southworth, J. Sun, M. Marsik, et al., Trans-boundary infrastructure and land cover change: highway paving and community-level deforestation in a tri-national frontier in the Amazon, *Land Use Pol.* 34 (2013) 27–41.
- [22] G.S. Cumming, G. Barnes, S. Perz, M. Schmink, K.E. Sieving, J. Southworth, et al., An exploratory framework for the empirical measurement of resilience, *Ecosystems* 8 (2005) 975–987.
- [23] R.L. Chazdon, Tropical forest recovery: legacies of human impact and natural disturbances, *Perspectives in Plant Ecology, Evol. Syst.* 6 (2003) 51–71.
- [24] O.L. Phillips, S.L. Lewis, T.R. Baker, K.J. Chao, N. Higuchi, The changing Amazon forest, *Phil. Trans. Biol. Sci.* 363 (2008) 1819–1827.
- [25] G.S. Cumming, J. Southworth, X.J. Rondon, M. Marsik, Spatial complexity in fragmenting Amazonian rainforests: do feedbacks from edge effects push forests towards an ecological threshold? *Ecol. Complex.* 11 (2012) 67–74.
- [26] M. Rodrigues, A. Jiménez, J. de la Riva, Analysis of recent spatial–temporal evolution of human driving factors of wildfires in Spain, *Nat. Hazards* 84 (2016) 2049–2070.
- [27] L. Vlassova, F. Pérez-Cabello, Effects of post-fire wood management strategies on vegetation recovery and land surface temperature (LST) estimated from Landsat images, *Int. J. Appl. Earth Obs. Geoinf.* 44 (2016) 171–183.
- [28] S.G. Perz, C. Overdevest, M.M. Caldas, R.T. Walker, E.Y. Arima, Unofficial road building in the Brazilian Amazon: dilemmas and models for road governance, *Environ. Conserv.* 34 (2007) 112–121.
- [29] S. Perz, S. Brillhante, F. Brown, M. Caldas, S. Ikeda, E. Mendoza, et al., Road building, land use and climate change: prospects for environmental governance in the Amazon, *Phil. Trans. Biol. Sci.* 363 (2008) 1889–1895.
- [30] K. Rapp, The Brazil-Peru Trans Oceanic Highway. Project Summary, Bank Information Center, 2005.
- [31] K.A. Brown, J. Gurevitch, Long-term impacts of logging on forest diversity in Madagascar, *Proc. Natl. Acad. Sci. Unit. States Am.* 101 (2004) 6045–6049.
- [32] M.R. Guariguata, R. Ostertag, Neotropical secondary forest succession: changes in structural and functional characteristics, *For. Ecol. Manage.* 148 (2001) 185–206.

- [33] S.J. DeWalt, S.K. Maliakal, J.S. Denslow, Changes in vegetation structure and composition along a tropical forest chronosequence: implications for wildlife, *For. Ecol. Manage.* 182 (2003) 139–151.
- [34] N. Restrepo-Coupe, H.R. da Rocha, L.R. Hutyrá, A.C. da Araujo, L.S. Borma, B. Christoffersen, et al., What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network, *Agric. For. Meteorol.* 182–183 (2013) 128–144.
- [35] C. Baraloto, P. Alverga, S.B. Quispe, G. Barnes, N.B. Chura, I.B. da Silva, et al., Effects of road infrastructure on forest value across a tri-national Amazonian frontier, *Biol. Conserv.* 191 (2015) 674–681.
- [36] J. Redwood, *The Environmental and Social Impacts of Major IDB-financed Road Improvement Projects: the Interoceanica IIRSA Sur and IIRSA Norte Highways in Peru*, Inter-American Development Bank, 2012.
- [37] O.J. Abelleira Martínez, A.K. Fremier, S. Günter, Z. Ramos Bendaña, L. Vierling, S.M. Galbraith, et al., Scaling up functional traits for ecosystem services with remote sensing: concepts and methods, *Ecol. Evol.* 6 (2016) 4359–4371.
- [38] W.P. Sousa, The role of disturbance in natural communities, *Annu. Rev. Ecol. Systemat.* 15 (1984) 353–391.
- [39] K. Thonicke, S. Venevsky, S. Sitch, W. Kramer, The role of fire disturbance for global vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model, *Global Ecol. Biogeogr.* 10 (2001) 661–677.
- [40] A. Huete, K. Didan, T. Miura, E.P. Rodriguez, X. Gao, L.G. Ferreira, Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sens. Environ.* 83 (2002) 195–213.
- [41] G. Asner, A.R. Townsend, B.H. Braswell, Satellite observation of El Niño effects on Amazon forest phenology and productivity, *Geophys. Res. Lett.* (2000) 981–984.
- [42] A.V. Rocha, G.R. Shaver, Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes, *Agric. For. Meteorol.* 149 (2009) 1560–1563.
- [43] P.M. Brando, S.J. Goetz, A. Baccini, D.C. Nepstad, P.S.A. Beck, M.C. Christman, Seasonal and interannual variability of climate and vegetation indices across the Amazon, *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010) 14685–14690.

- [44] B.C. Reed, J.F. Brown, D. VanderZee, T. Loveland, J.W. Merchant, D.O. Ohlen, Measuring phenological variability from satellite imagery, *J. Veg. Sci.* 5 (1994) 703–714.
- [45] B.A. Bradley, E. Fleishman, Can remote sensing of land cover improve species distribution modelling? *J. Biogeogr.* 35 (2008) 1158–1159.
- [46] J. Cabello, N. Fernández, D. Alcaraz-Segura, C. Oyonarte, G. Piñeiro, A. Altesor, et al., The ecosystem functioning dimension in conservation: insights from remote sensing, *Biodivers. Conserv.* 21 (2012) 3287–3305.
- [47] J.N. Volante, D. Alcaraz-Segura, M.J. Mosciaro, E.F. Viglizzo, J.M. Paruelo, Ecosystem functional changes associated with land clearing in NW Argentina, *Agric. Ecosyst. Environ.* 154 (2012) 12–22.
- [48] D. Peng, C. Wu, C. Li, X. Zhang, Z. Liu, H. Ye, et al., Spring green-up phenology products derived from MODIS NDVI and EVI: intercomparison, interpretation and validation using National Phenology Network and Ameri-Flux observations, *Ecol. Indicat.* 77 (2017) 323–336.
- [49] M. Kottek, J. Grieser, C. Beck, B. Rudolf, F. Rubel, World Map of the Köppen-Geiger climate classification updated, *Metz* 15 (2006) 259–263.
- [50] F. Rubel, K. Brugger, K. Haslinger, I. Auer, The climate of the European Alps: shift of very high resolution Köppen-Geiger climate zones 1800–2100, *Metz* 26 (2017) 115–125.
- [51] C.I. Salimon, F.E. Putz, L. Menezes-Filho, A. Anderson, M. Silveira, I.F. Brown, et al., Estimating state-wide biomass carbon stocks for a REDD plan in Acre, Brazil, *For. Ecol. Manag.* 262 (2011) 555–560.
- [52] A.L.d. Carvalho, B.W. Nelson, M.C. Bianchini, D. Plagnol, T.M. Kuplich, D.C. Daly, Bamboo-dominated forests of the Southwest Amazon: detection, spatial extent, life cycle length and flowering waves, *PLoS One* 8 (2013), e54852.
- [53] C.A. Rockwell, K.A. Kainer, M.V.N. d'Oliveira, C.L. Staudhammer, C. Baraloto, Logging in bamboo-dominated forests in southwestern Amazonia: caveats and opportunities for smallholder forest management, *For. Ecol. Manage.* 315 (2014) 202–210.
- [54] C.A. Quesada, J. Lloyd, L.O. Anderson, N.M. Fyllas, M. Schwarz, C.I. Czimczik, Soils of Amazonia with particular reference to the RAINFOR sites, *Biogeosciences* 8 (2011) 1415–1440.

- [55] O.L. Phillips, S. Rose, A.M. Mendoza, P.N. Vargas, Resilience of southwestern Amazon forests to anthropogenic edge effects, *Conserv. Biol.* 20 (2006) 1698–1710.
- [56] Z. Jiang, A. Huete, K. Didan, T. Miura, Development of a two-band enhanced vegetation index without a blue band, *Remote Sens. Environ.* 112 (2008) 3833–3845.
- [57] I. Harris, P.D. Jones, T.J. Osborn, D.H. Lister, Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, *Int. J. Climatol.* 34 (2013) 623–642.
- [58] Y. Fan, Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, *J. Geophys. Res.* 109 (2004).
- [59] J. Southworth, M. Marsik, Y. Qiu, S. Perz, G. Cumming, F. Stevens, et al., Roads as drivers of change: trajectories across the tri-national frontier in MAP, the southwestern Amazon, *Rem. Sens.* 3 (2011) 1047–1066.
- [60] M. Convertino, R.S. Mangoubi, I. Linkov, N.C. Lowry, M. Desai, Inferring species richness and turnover by statistical multiresolution texture analysis of satellite imagery, *PLoS One* 7 (2012).
- [61] M. Convertino, R. Muñoz-Carpena, G.A. Kiker, S.G. Perz, Design of optimal ecosystem monitoring networks: hotspot detection and biodiversity patterns, *Stoch. Environ. Res. Risk Assess.* (2015) 1–18.
- [62] USGS, EarthExplorer. Landsat Data Access, (n.d.). <https://earthexplorer.usgs.gov/>.
- [63] A.F. Zuur, I.D. Tuck, N. Bailey, Dynamic factor analysis to estimate common trends in fisheries time series, *Can. J. Fish. Aquat. Sci.* 60 (2003) 542–552.
- [64] D. Pena, Dimension reduction in time series and the dynamic factor model, *Biometrika* 96 (2009) 494–496.
- [65] C. Lam, Q. Yao, N. Bathia, Estimation of latent factors for high-dimensional time series, *Biometrika* 98 (2011) 901–918.
- [66] D. Kaplan, R. Muñoz-Carpena, A. Ritter, Untangling complex shallow groundwater dynamics in the floodplain wetlands of a southeastern U.S. coastal river, *Water Resour. Res.* 46 (2010).
- [67] Y.-M. Kuo, H.-J. Lin, Dynamic factor analysis of long-term growth trends of the intertidal seagrass *Thalassia hemprichii* in southern Taiwan, *Estuarine, Coast. Shelf Sci.* 86 (2010) 225–236.

- [68] M.A. Campo-Bescós, R. Muñoz-Carpena, D.A. Kaplan, J. Southworth, L. Zhu, P.R. Waylen, Beyond precipitation: physiographic gradients dictate the relative importance of environmental drivers on savanna vegetation, *PLoS One* 8 (2013), e72348.
- [69] A. Ritter, R. Muñoz-Carpena, Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments, *J. Hydrol.* 480 (2013) 33–45.
- [70] A.D. Chouakria, P.N. Nagabhushan, Adaptive dissimilarity index for measuring time series proximity, *Adv. Data Anal. Classif.* 1 (2007) 5–21.
- [71] J.H. Ward Jr., Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.* 58 (1963) 236–244.
- [72] F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* 31 (2014) 274–295.
- [73] J.C. Dunn, Well-separated clusters and optimal fuzzy partitions, *J. Cybern.* 4 (2008) 1–13.
- [74] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.* 20 (1987) 53–65.
- [75] I. Biurrun, J.A. Campos, I. García-Mijangos, M. Herrera, J. Loidi, Floodplain forests of the Iberian Peninsula: vegetation classification and climatic features, *Appl. Veg. Sci.* 19 (2016) 336–354.
- [76] J.R. Vicente, J. Gonçalves, J.P. Honrado, C.F. Randin, J. Pottier, O. Broennimann, et al., A framework for assessing the scale of influence of environmental factors on ecological patterns, *Ecol. Complex.* 20 (2014) 151–156.
- [77] A. Abdolhay, B. Saghafian, M.A.M. Soom, A.H.B. Ghazali, Identification of homogenous regions in Gorganrood basin (Iran) for the purpose of regionalization, *Nat. Hazards* 61 (2012) 1427–1442.
- [78] N. Bolshakova, F. Azuaje, Cluster validation techniques for genome expression data, *Signal Process.* 83 (2003) 825–833.
- [79] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Pérez, I. Perona, An extensive comparative study of cluster validity indices, *Pattern Recogn.* 46 (2013) 243–256.
- [80] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, *Applied Linear Regression Models*, McGraw-Hill, Boston, Massachusetts, USA, 1996.

- [81] A.F. Zuur, E.N. Ieno, G.M. Smith, *Analysing Ecological Data*, Springer, New York, 2007.
- [82] R. Muñoz-Carpena, A. Ritter, Y.C. Li, Dynamic factor analysis of ground-water quality trends in an agricultural area adjacent to Everglades National Park, *J. Contam. Hydrol.* 80 (2005) 49–70.
- [83] D. Kaplan, R. Muñoz-Carpena, Complementary effects of surface water and groundwater on soil moisture dynamics in a degraded coastal floodplain forest, *J. Hydrol.* 398 (2011) 221–234.
- [84] A.F. Zuur, R.J. Fryer, I.T. Jolliffe, R. Dekker, J.J. Beukema, Estimating common trends in multivariate time series using dynamic factor analysis, *Environmetrics* 14 (2003) 665–685.
- [85] E.E. Holmes, E.J. Ward, M.D. Scheuerell, *Analysis of Multivariate Time Series Using the MARSS Package*, NOAA Fisheries, 2014.
- [86] J.E. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models. Part I - a discussion of principles, *J. Hydrol.* 10 (1970) 282–290.
- [87] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* 6 (1978) 461–464.
- [88] A. Ritter, C.M. Regalado, R. Muñoz-Carpena, Temporal common trends of topsoil water dynamics in a humid subtropical forest watershed, *Vadose Zone J.* 8 (2009) 437–449.
- [89] A. Chevan, M. Sutherland, Hierarchical partitioning, *Am. Statistician* 45 (1991) 90–96.
- [90] J.W. Johnson, J.M. Lebreton, History and use of relative importance indices in organizational research, *Organ. Res. Meth.* 7 (2004) 238–257.
- [91] R.H. Lindeman, P.F. Merenda, R.Z. Gold, *Introduction to Bivariate and Multivariate Analysis*, Glenview, Scott, Foresman and Company, Ill, 1980.
- [92] U. Grömping, Relative importance for linear regression in R: the package relaimpo, *J. Stat. Software* 17 (2006) 1–27.
- [93] P. Bloomfield, *Fourier Analysis of Time Series*, John Wiley & Sons, 2004.
- [94] L.H. Koopmans, *The Spectral Analysis of Time Series*, Elsevier, 1995.
- [95] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. <http://www.R-project.org/>.

- [96] Python Software Foundation, Python language Reference, version 2.7, 2012. <http://www.python.org>.
- [97] P. Montero, J.A. Vilar, TSclust: an R package for time series clustering, *J. Stat. Software* 62 (2014) 1–43.
- [98] G. Brock, S. Datta, V. Pihur, S. Datt, clValid, An R package for cluster validation, *J. Stat. Software* 25 (2008) 1–22.
- [99] M. Nakazawa, Package “fmsb”: Functions for Medical Statistics Book with Some Demographic Data, CRAN, 2017.
- [100] M. Zambrano-Bigiarini, Package “hydroGOF”: Goodness-of-fit Functions for Comparison of Simulated and Observed Hydrological Time Series, CRAN, 2017.
- [101] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, second ed., Springer New York, New York, NY, 2016.
- [102] H. Wickham, Package “reshape2”: Flexibly Reshape Data: a Reboot of the Reshape Package, CRAN, 2017.
- [103] A. Zeileis, G. Grothendieck, Zoo: an S3 Class and Methods for Indexed Totally Ordered Observations, arXiv, 2005.
- [104] B. Auguie, Package ‘gridExtra’: Miscellaneous Functions for ‘grid’ Graphics, CRAN, 2017.
- [105] C.O. Wilke, Package ‘cowplot’: Streamlined Plot Theme and Plot Annotations for ‘ggplot2’, CRAN, 2017.
- [106] H. Wickham, Package ‘gtable’: Arrange ‘grobs’ in Tables, CRAN, 2016.
- [107] D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, T.L. Veith, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE* 50 (2007) 885–900.
- [108] X. Zhang, M.A. Friedl, C.B. Schaaf, A.H. Strahler, J.C.F. Hodges, F. Gao, et al., Monitoring vegetation phenology using MODIS, *Remote Sens. Environ.* 84 (2003) 471–475.
- [109] F.B. Silva, Y.E. Shimabukuro, L.E.O.C. Aragão, L.O. Anderson, G. Pereira, F. Cardozo, et al., Large-scale heterogeneity of Amazonian phenology revealed from 26-year long AVHRR/NDVI time-series, *Environ. Res. Lett.* 8 (2013), 024011.

- [110] J. Forrest, A.J. Miller-Rushing, Toward a synthetic understanding of the role of phenology in ecology and evolution, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365 (2010) 3101–3112.
- [111] N. Kathuroju, M.A. White, J. Symanzik, M.D. Schwartz, J.A. Powell, R.R. Nemani, On the use of the advanced very high resolution radiometer for development of prognostic land surface phenology models, *Ecol. Model.* 201 (2007) 144–156.
- [112] R.K. Pachauri, L. Meyer, G.K. Plattner, T. Stocker, *Climate Change 2014: Synthesis Report, Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, 2015.
- [113] C.A. Quesada, O.L. Phillips, M. Schwarz, C.I. Czimczik, T.R. Baker, S. Patiño, et al., Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate, *Biogeosciences* 9 (2012).
- [114] C. Baraloto, P. Alverga, S. Baéz Quispe, G. Barnes, N. Bejar Chura, I. Brasil da Silva, et al., Trade-offs among forest value components in community forests of southwestern Amazonia, *E&S* 19 (2014) art56–11.
- [115] S.G. Perz, M. Rosero, F.L. Leite, L. Araujo Carvalho, J. Castillo, C. Vaca Mejia, Regional integration and household resilience: infrastructure connectivity and livelihood diversity in the southwestern Amazon, *Hum. Ecol.* 41 (2013) 497–511.
- [116] J. Hoelle, Convergence on cattle: political ecology, social group perceptions, and socioeconomic relationships in Acre, Brazil, *Cult. Agric. Food Environ.* 33 (2011) 95–106.
- [117] R. Brajša, H. Wöhl, A. Hanslmeier, G. Verbanac, D. Ruždjak, E. Cliver, et al., On solar cycle predictions and reconstructions, *A&A* 496 (2009) 855–861.
- [118] D. Offermann, P. Hoffmann, P. Knieling, R. Koppmann, J. Oberheide, W. Steinbrecht, Long-term trends and solar cycle variations of mesospheric temperature and dynamics, *J. Geophys. Res.* 115 (2010).
- [119] K. Ramanuja Rao, Short periodicities in solar activity, *Sol. Phys.* 29 (1973) 47–53.
- [120] G.A.M.D. Silva, A. Drumond, T. Ambrizzi, The impact of El Niño on South American summer climate during different phases of the Pacific decadal oscillation, *Theor. Appl. Climatol.* 106 (2011) 307–319.

- [121] P.A. Arias, R. Fu, C.D. Hoyos, W. Li, L. Zhou, Changes in cloudiness over the Amazon rainforests during the last two decades: diagnostic and potential causes, *Clim. Dynam.* 37 (2010) 1151–1164.
- [122] S.K. Herzog, R. Martinez, P.M. Jorgenson, H. Tiessen (Eds.), *Climate Change and Biodiversity in the Tropical Andes*, Inter-American Institute for Global Change (IAI) and Scientific Committee on Problems of the Environment (SCOPE), 2011.
- [123] E. Cleland, I. Chuine, A. Menzel, H. Mooney, M. Schwartz, Shifting plant phenology in response to global change, *Trends Ecol. Evol.* 22 (2007) 357–365.
- [124] B.A. Bradley, J.F. Mustard, Comparison of phenology trends by land cover class: a case study in the Great Basin, USA, *Global Change Biol.* 14 (2007) 334–346.
- [125] J. Verrelst, G. Camps-Valls, J. Muñoz-Marí, J.P. Rivera, F. Veroustraete, J.G.P.W. Clevers, et al., Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - a review, *ISPRS J. Photogrammetry Remote Sens.* 108 (2015) 273–290.
- [126] P. Coppo, A. Taiti, L. Pettinato, M. Francois, M. Taccola, M. Drusch, Fluorescence imaging spectrometer (FLORIS) for ESA FLEX mission, *Rem. Sens.* 9 (2017) 649.
- [127] W.F. Laurance, G.R. Clements, S. Sloan, C.S. O’Connell, N.D. Mueller, M. Goosem, et al., A global strategy for road building, *Nature* 513 (2014) 229–232.