ENVIRONMENTAL RESEARCH
COMMUNICATIONS

PAPER • OPEN ACCESS

Quantifying vegetation response to environmental changes on the Galapagos Islands, Ecuador using the Normalized Difference Vegetation Index (NDVI)

To cite this article: E Herrera Estrella *et al* 2021 *Environ. Res. Commun.* **3** 065003

Environmental Research Communications

PAPER

Quantifying vegetation response to environmental changes on the Galapagos Islands, Ecuador using the Normalized Difference Vegetation Index (NDVI)

E Herrera Estrella[1,2,3], A Stoeth[2,4], N Y Krakauer[3,5] and N Devineni[3,5]

[1] The City College of New York, City University of New York, New York, 10031, United States of America
[2] The Graduate Center, City University of New York, New York, 10016, United States of America
[3] NOAA—Cooperative Science Center for Earth System Sciences and Remote Sensing Technologies, The City College of New York, City University of New York, New York, 10031, United States of America
[4] Queens College, School of Earth and Environmental Sciences, City University of New York, Flushing, NY 11367, United States of America
[5] Department of Civil Engineering, City University of New York (City College), New York, 10031, United States of America

E-mail: eherrera@gradcenter.cuny.edu

Keywords: The Galapagos Islands, NDVI, climate change, seasonal fluctuations

OPEN ACCESS

RECEIVED
24 September 2020

REVISED
25 May 2021

ACCEPTED FOR PUBLICATION
16 June 2021

PUBLISHED
25 June 2021

Abstract

The vegetation of the Galapagos Islands (Ecuador) is strongly influenced by climate. El Niño events, seasonality, isolation, volcanism, and increasing human activity define the ecosystems of the archipelago. Given their socio-cultural and economic importance, it is critical to monitor the response of Galapagos vegetation to changes in climate and assess its vulnerability. This study explores the potential to use Normalized Difference Vegetation Index (NDVI) as a proxy to describe trends in primary productivity in the Galapagos (2000–2019) and models the relationship between NDVI and climate variables including evaporation and atmospheric carbon dioxide concentration.

From numerous possible co-variates compiled from reanalysis and satellites, we identify the independent variables that most strongly influence NDVI using the least absolute shrinkage and selection operator (LASSO) method. Significant variables, including carbon dioxide concentration, evaporation, and autocorrelation (1-month and 12-months lagged NDVI) are then used to model NDVI in a generalized linear model (GLM) framework. The model predicts NDVI more effectively where values for NDVI are high (high elevation, lush vegetation), and clearly reflects seasonality. Validation of the model across pixels produces R^2 values ranging from 0.05 to 0.94, and the mean R^2 is 0.57 (0.65 for elevation >20 m). This methodology has the potential to continuously and non-intrusively monitor vegetation changes in sensitive ecological regions, such as the Galapagos.

1. Introduction

The Galapagos Islands, located ~1000 km off the west coast of Ecuador in the equatorial Pacific, are world famous. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), these islands and the surrounding marine reserve are a unique 'living museum and showcase of evolution' [1]. Island ecosystems often boast uniquely diverse biota, due to their isolation and micro-climates, which can drive species endemism [2]. Even so, the Galapagos have an exceptional scientific legacy. Charles Darwin conducted field research in the Galapagos in 1835 that helped inform his theory of evolution by natural selection—a theory that has fundamentally changed scientific understanding of biological diversity [3]. The organisms he studied and chronicled in *On the Origin of Species* [4], including tortoises and finches, are legendary; they are textbook examples of adaptive radiation and provide living evidence of evolution in progress [1]. They also help draw approximately 170,000 annual visitors to the islands, bestowing considerable economic importance to the Galapagos [1].

Galapagos vegetation, which supports the famous animal life of the archipelago, is influenced by a unique climate system shaped by ocean-atmosphere dynamics, the El Niño Southern Oscillation (ENSO), active volcanic island formation, and—more recently—human activities. The topography of the islands defines their ecoregions, with the volcanic island peaks receiving the most precipitation and therefore containing the lushest vegetation. Nevertheless, arid ecosystems at or near sea level dominate the Galapagos, accounting for an estimated 83% of land area, with humid and transitional zones occurring only at elevations >200 meters above sea level (MASL) and littoral zones fringing the islands [5]. New volcanism and island formation provide continuing opportunities for primary succession, speciation, and the formation of novel colonizer communities, both through the creation of new land and through the destruction of existing habitat [6].

Despite their dynamism, the Galapagos ecosystems have been resistant to past climate changes, perhaps due to their continual exposure to ENSO-driven natural climate variability. ENSO creates extreme inter-annual variability in temperature, humidity, and precipitation, which have acclimatized the ecological communities of the Galapagos to climate disturbance. A fossil pollen study by Restrepo *et al* 2012 [7] suggests that plant regimes in the Galapagos have been remarkably stable over the last 2690 years; researchers found few changes to ecological community composition as a result of major climate events such as the Little Ice Age (1550–1880 CE) and the Medieval Warm Period (800–1250 CE). On the other hand, these ecosystems are vulnerable to anthropogenic disturbance. Humans have known about and regularly used the Galapagos Islands since the 16th century for piracy, whaling, and sealing, creating permanent settlements in the 1800s [8]. Since the 1960s and 1970s, human activities have intensified; population has increased, fishing has become more lucrative (and thus, more intense), and land use has changed to accommodate agriculture, urbanization, and tourism [9]. Climate change and introduced species are additional stressors to the archipelago's ecosystems.

As a result, vegetation regimes in the archipelago are now changing in unprecedented ways [2, 10]. Non-native herbivores, such as goats, have negatively affected net primary productivity (NPP) across the islands through prolific and indiscriminate grazing. Invasive plant species now dominate the humid highlands, where favorable growing conditions have enabled them to outcompete native species. Most endemic Galapagos species are now confined to the xeric lowlands, where conditions are harsher. Rivas-Torres *et al* 2018 [2] estimate that 40% of vascular plants in the Galapagos are found nowhere else on Earth; of these, 62% are considered to be rare and/or have vulnerable populations. El Niño events, which mimic rainy seasons in the Galapagos, are hypothesized to benefit invasive plant species at the expense of natives [7] by easing water scarcity, and recent research suggests that extreme El Niño events are becoming more frequent [11]. Moreover, anthropogenic climate change is expected to cause higher ocean temperatures and increased precipitation in the eastern equatorial Pacific, analogous to sustained El Niño conditions [12]. These climate trends point to the possibility of major plant community shifts in the coming decades. Vegetation is the base of the Galapagos food chain, meaning changes in species composition, biomass, or productivity have the potential to affect consumers at higher trophic levels, and ultimately, to change the structure and stability of the overall ecosystem. For these reasons, it is immensely important to assess, monitor, and understand the Galapagos plant community.

Remote sensing can non-intrusively gather enormous amounts of data and has been employed for monitoring purposes in many research fields, including ecology, oceanography, and geography [13]. The remote sensing product Normalized Difference Vegetation Index (NDVI) is a common proxy for vegetation condition/phenological stage [14] and above-ground primary productivity (AGPP).

Despite widespread scientific interest in both climate change and the unique ecology of the Galapagos Islands, few studies have used NDVI to explore the effect of climate variability on the archipelago's plant life. Previous studies using remote sensing to assess the environmental health of the Galapagos Islands have focused on the impact of herbivory on native vegetation [10, 15] or on specific floral species [2, 16] and land cover changes [9]. None of these studies used NDVI as their primary dependent variable, and none analyzed the NDVI of the entire Galapagos archipelago, focusing instead on individual islands or groups of islands relevant to their given research question. We seek to examine whether NDVI can be used in an island biodiversity hotspot like the Galapagos to monitor vegetation and measure its response to climate dynamics in an era of anthropogenic global change, making this study the first to assess the association between NDVI and climate change in the Galapagos holistically.

We use NDVI calculated from satellite remote sensing to quantify the response of Galapagos vegetation to climate variability, looking at both decadal trends (that could show signals of anthropogenic climate change) and cyclical patterns associated with natural variability (e.g., those driven by seasonality and ENSO cycles). We analyze the NDVI of the Galapagos archipelago using data collected from February 2000 to February 2019 by MODIS satellites (Terra and Aqua) and compiled on a monthly basis. This study has two principal objectives: 1) describe any observed trends in NDVI over the 19-year period of the MODIS data, and 2) develop a robust multiple regression model to explain NDVI variability in the Galapagos, using independent variables related to climate, geology, and human activities. The list of considered independent variables is presented in section 2.2.3.

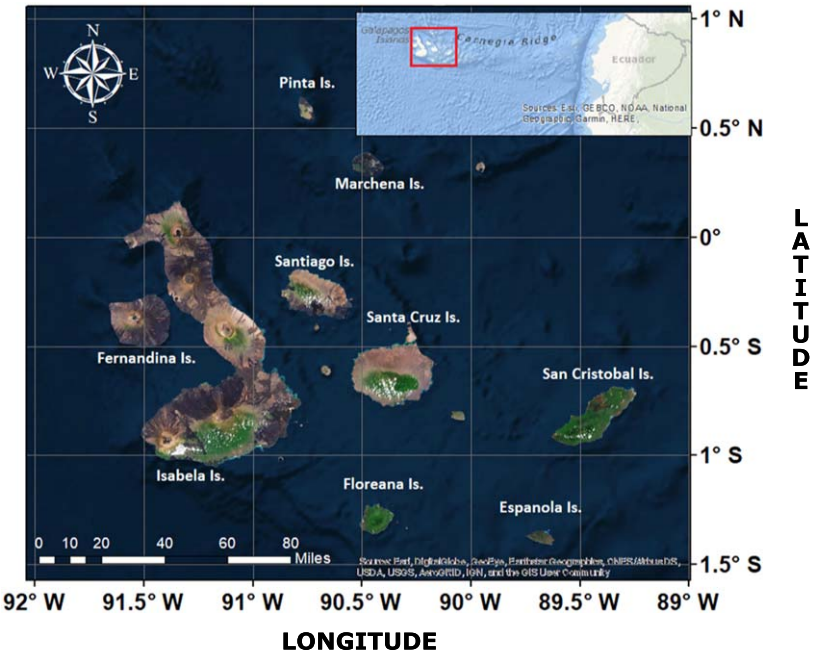

Figure 1. Map of the Galapagos Islands. (Esri, 'World imagery'. Sources: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community [20]; and 'World ocean base'. Sources: Esri, Garmin, GEBCO, NOAA NGDC, and other contributors [21].)

2. Study region, data, and methods

2.1. Study region

The Galapagos archipelago (figure 1) sits on the Equator and extends for approximately 260 km E–W (90°01′ W to 89°16′ W) and approximately 200 km N–S (1°40′ S to 1°36′ N) [17], with a total area near 52,000 km^2. It is composed of 128 named islands [6]; four of the islands, including the largest, Isabela, are inhabited, with a combined population of approximately 30,000 [1, 18]. The archipelago is volcanic in origin and geologically young; The oldest islands formed 3–6 million years ago (mya) and the youngest formed 0.05–0.5 mya [19]. Lavas represent ∼44% of land area.

Monthly average daytime/nighttime surface temperatures range from 24–42 °C/14–23 °C, respectively, with average diurnal temperature swings of 14 °C. Average precipitation varies from 88–263 mm/year and is largely seasonal, driven by the interaction of nearby air and sea currents. From January-May, the Inter-Tropical Convergence Zone (ITCZ) is to the south of the Islands and the Panama Current brings tropical heat and rain to the Galapagos. [5]. From June-December, the ITCZ moves north of the Galapagos and the Humboldt Current keeps the archipelago unusually cool and dry for its latitude [5, 6]. Garúa (misty/drizzly air blown inland and upslope from the ocean) is characteristic of this period, and rainfall is uncommon [22]. El Niño events, which occur every 2–8 years [5, 23], resemble sustained rainy seasons in the Galapagos. Along with volcanism, ENSO is primarily responsible for interannual climate variability in the archipelago. Of vegetated land in the Galapagos, ∼61% is dry forest, which dominates the lowlands. ∼21% is evergreen forest and scrubland, which occurs at higher elevations (>200MASL). A study of the Galapagos National Park found that ∼54% had vegetation cover characteristic of native ecosystems, while ∼2% was dominated by invasive species [2]. The dominant species associated with dry and evergreen forest ecosystems are summarized in table A1–2 in the appendix.

2.2. Data

2.2.1. Normalized difference vegetation index

Many Earth-observing satellites are equipped with sensors designed to measure near-infrared (NIR) and red reflectances (ρ). These are converted into NDVI [24] as:

$$\mathrm{NDVI} = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}. \quad (1)$$

NDVI values range from −1.0 and 1.0 [25]. High values indicate a higher density of green vegetation, low values indicate scarce, moisture-stressed, or dryland vegetation [26, 27], and values close to 0 are likely to

correspond to non-vegetated surfaces (such as bare soil, urbanized areas, and exposed rock/lava flows). Negative values correspond primarily to clouds, water, or snow.

Average NDVI in the islands varies from −0.04 to 0.78. Lower values of NDVI are found on Isabela, Fernandina, Santiago, and Marchena Islands. These values correspond to active volcanic areas near the shore, which also have lower elevations and predominantly arid climates. Higher NDVI values are mainly associated with elevations 200–800 MASL. Rivas-Torres *et al* (2018) [2] classified the mean values of NDVI across the diverse ecosystems of the Galapagos Islands, and showed that values 0–0.2 corresponded to rocky outcrops and/or sparse, deciduous vegetation; values from 0.2–0.59 represented deciduous vegetation, and values higher than 0.6 indicated dense and evergreen vegetation. Zones of mixed vegetation (such as transitional zones) had intermediate mean NDVI values [2]. Section 3.1 gives additional details on the distribution of NDVI across the archipelago.

2.2.2. Independent variables

The set of independent variables used in this study was divided in 5 categories: air composition, atmospheric state, soil, ENSO, and topographical.

Variations in the concentration of air molecules and particles, which can potentially affect vegetation, are both natural and man-made. The concentration of carbon dioxide, a major greenhouse gas, varies due to anthropogenic emissions, biological activity, and air-sea fluxes. Carbon monoxide is associated with fires, as well as transport of polluted air from industrial areas. Dust is carried by wind, and can be generated by lofting from volcanoes or deserts. It can negatively affect plant photosynthesis, or in some cases supply valuable nutrients. Ozone is another species whose concentration is affected by anthropogenic activity and atmospheric transport pathways and that in high concentration can negatively affect plant development by oxidizing tissues exposed to it. These variables are listed in table 2 as variables 1–4.

The atmospheric variables include day and night air and surface temperatures. Excessively high temperatures stress plants and increase respiration rates. Wind can affect plant water loss rate and physical integrity, depending on the direction, velocity, and duration. The water cycle is very important in vegetation growth. The amount of rain (precipitation) is a key component of ecosystem water balance, along with the amount of water that evaporates from earth or plants. The variables thus considered are listed in table 2 as variables 5–12.

Water from precipitation can be stored by the soil at different rates depending on the type of soil. This is known as soil moisture. Along with the temperature of soil at different depths, this influences the amount of water and nutrients plants can absorb. Soil variables are listed in table 2 as variables 15–18.

The El Niño-Southern Oscillation (ENSO) includes the warm (El Niño) and the cool (La Niña) phases of a recurrent climate pattern across the tropical Pacific. Affecting the air currents and the ocean temperature, it can impact the growth of plants. Indices of ENSO are listed in table 2, variables 19–20.

The topographical variable consists of an elevation data set, which allows identifying the elevation ranges where vegetation is generally more dense. In addition, from the gridded elevation we can calculate aspect (slope face direction) and slope (change in elevation per unit horizontal distance) over the archipelago, which could potentially affect light and water fluxes and hence local temperature and moisture status. This is listed in table 2 as variable 21.

2.2.3. Datasets

Data were obtained from the following sources:

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a sensor on board the Terra and Aqua satellites, which were launched in 1999 and 2002 respectively. As polar-orbiting satellites, they move around the Earth in a north-south orientation, with Terra crossing the equator in the morning and Aqua crossing in the afternoon. Terra and Aqua observe the entire Earth surface every 1–2 days, acquiring data in 36 spectral bands (ranges of wavelengths) [28]. The MODIS Collection 6 products provide vegetation index (VI) values on a pixel basis using blue, red, and near-infrared reflectances with a spatial resolution of 0.05° latitude/longitude (5,600m at the equator). These MODIS products are a monthly composite of cloud-free spatial Level 3 products. The NDVI products, used independently in this study, are MOD13C2 (Terra) and MYD13C2 (Aqua). The new Enhanced Vegetation Index (EVI) is another product from MYD13C2, and uses the blue band to remove atmospheric contamination, minimizes canopy variations, and maintains sensitivity over dense vegetation. NDVI is much more commonly used as a vegetation index in the Galapagos and other tropical research, compared to EVI [2, 9, 10]. MODIS products also include land surface temperature and emissivity (MOD11C3), both of which average the corresponding daytime and nighttime observations over each month.

The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) incorporates and synthesizes data gathered since 1980 within a global climate model framework. MERRA-2 incorporates space-based observations of aerosols to represent their interactions with other physical processes in the climate

Table 1. List of abbreviations and acronyms used on this manuscript.

Abbreviations and Acronyms	Explanation
AGPP	Above-ground primary productivity
AR	Autoregression spectrum
ASTER	The Advanced Spaceborne Thermal Emission and Reflection Radiometer
CO	Carbon monoxide
CO_2	Carbon dioxide
DEM	Digital elevation model
E	Evaporation rate over land
ENSO	El Niño Southern Oscillation
GDEM V2	Global Digital Elevation Model Version 2
GLM	Generalized linear model
GNP	Galapagos National Park
IRLS	Iteratively reweighted least squares
ITCZ	Inter-Tropical Convergence Zone
LASSO	The least absolute shrinkage and selection operator
MASL	Meters above sea level
MATLAB	Matrix laboratory
MEI	Multivariate ENSO Index
MERRA-2	The Modern-Era Retrospective analysis for Research and Applications, Version 2
MODIS	The Moderate Resolution Imaging Spectroradiometer
NASA	The National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index
NIR	Near infrared
NPP	Net primary productivity
O_3	Ozone
ONI	Oceanic Niño Index
R^2	Coefficient of determination
RMSE	Root Mean Square Error
TA	Air Temperature
TRMM	The Tropical Rainfall Measuring Mission
TS	Surface Temperature
UNESCO	The United Nations Educational, Scientific and Cultural Organization
VEI	Volcanic explosivity index

system, and has a spatial resolution of 50 km in the latitudinal direction [29]. From the different products that MERRA-2 produces, we selected 15 products, including meteorological, atmospheric, and geologic variables that have the potential to affect vegetative phenology (as summarized above, and listed in table 2).

The Tropical Rainfall Measuring Mission (TRMM) (1997–2015) was designed to improve the understanding of the distribution and variability of precipitation within the tropics [30]. The TRMM-3B43-7 monthly precipitation product is used, which is based on data from TRMM 3B42 product (3-hours merged microwave and infrared based precipitation (mm/hr), calibrated to station precipitation gauges).

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2), released on October 2011, has a coverage from 83° N-83° S latitude, spanning 99% of Earth's landmass, with a resolution of 30 meters [31].

Monthly CO_2 dry air mixing ratio from Mauna Loa, Hawaii (20° N) is used to quantify the potential impact of global carbon dioxide (CO_2) concentration on the vegetation of the Galapagos Archipelago.

MEI (Multivariate ENSO Index) and ONI (Oceanic Niño Index) are used to quantify the effect of ENSO over the islands. ONI, an index of the National Oceanic and Atmospheric Administration (NOAA), uses the sea surface anomaly for the Niño 3.4 region in the Equatorial Pacific, with El Niño defined as when the anomaly exceeds +0.5 °C for three months. MEI is a weighted average of the anomaly of six meteorological variables as associated with ENSO: sea surface temperature, sea level pressure, surface air temperature, surface wind (meridional and zonal components), and cloud fraction [32].

Table 2. List of explanatory independent variables. Each independent variable is numbered, named, and presented with its unit, source, spatial resolution, and product name. The temporal resolution of NDVI and potential independent variables is monthly, expcet for topography which is constant in time. The table is divided in sections. The first row is the dependent variable, followed by the numbered rows of independent variables categorized under air molecules and particles, atmosphere, soil, ENSO, and topography.

#	Dataset name	Unit	Source	Resolution	Product version
—	NDVI	—	MODIS	0.05°	MOD13C2 (Terra) and MYD13C2 (Aqua)
	Air molecules and particles				
1	Carbon dioxide (CO_2)	ppm	NOAA	global	Mauna Loa CO2 monthly mean data
2	Carbon monoxide (CO) emission	kg m^{-2} s^{-1}	MERRA-2	0.5 × 0.625°	M2TMNXCHM_5_12_4_COEM
3	Dust	kg m^{-3}	MERRA-2	0.5 × 0.625°	M2TMNXAER_5_12_4_DUSMASS
4	Ozone (O_3) mix ratio	kg/kg	MERRA-2	0.5 × 0.625°	M2IMNPASM_5_12_4_O3-1000hPa
	Atmosphere				
5	Evaporation from land	kg m^{-2} s^{-1}	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_EVLAND
6	Precipitation	mm hr^{-1}	TRMM	0.25°	TRMM_3B43_7_precipitation
7	Surface air Temperature	K	MERRA-2	0.5 × 0.625°	M2TMNXFLX_5_12_4_TLML
8	TA (Air Temperature)	K	MODIS	0.5°	M2IMNPANA_5_12_4_T-1000hPa
9	TS day (Surface Temperature)	K	MERRA-2	0.5 × 0.625°	MOD11C3_006_LST_Day_CMG
10	TS night (Surface Temperature)	K	MERRA-2	0.5 × 0.625°	MOD11C3_006_LST_Night_CMG
11	Wind Speed	m s^{-1}	MERRA-2	0.5 × 0.625°	M2TMNXFLX_5_12_4_SPEED
12	Wind Speed max	m s^{-1}	MERRA-2	0.5 × 0.625°	M2TMNXFLX_5_12_4_SPEEDMAX
	Soil				
13	Soil Moisture	m^{-3} m^{-3}	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_GWETTOP
14	Temperature (T) Soil 1	K	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_TSOIL1
15	T Soil 2	K	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_TSOIL2
16	T Soil 3	K	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_TSOIL3
17	T Soil 4	K	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_TSOIL4
18	T Soil 5	K	MERRA-2	0.5 × 0.625°	M2TMNXLND_5_12_4_TSOIL5
	ENSO				
19	MEI (Multivariate ENSO Index)	—	NOAA	global	Version 2 (MEI.v2)
20	ONI (Oceanic Niño Index)	—	NOAA	global	Version 5 (ONI_v5)
	Topography				
21	Digital elevation model (DEM)	m	ASTER	30 m	ASTER TIF

CO_2 , MEI, and ONI are globally representative time series and were obtained from NOAA sources [33–35]. All other datasets presented in table 2 (below) and NDVI datasets are spatially resolved and were downloaded from NASA's Giovanni web interface [36]. All data represent monthly values, except for the digital elevation model (DEM) which is temporally constant. Data were processed in MATLAB v9.4.

2.3. Methods

The methods used in this paper to model the spatiotemporal patterns in NDVI are the Least Absolute Shrinkage and Selection Operator (LASSO) regression for selecting predictor variables, followed by the generalized linear model (GLM) for fitting the selected variables. First, we preprocessed each dataset by filling in missing values using nearest-neighbor or linear interpolation, as long as the percentage of the missing data did not exceed 5% of the study period (≈11.5 months). Filled data represent 13.6% of the total study area. Then the remaining selected pixels are used as a mask to eliminate independent variables that have more than 5% of missing data. Each dataset was spatially regridded from its original resolution (shown in table 2) to match the NDVI resolution ($0.05° \approx 5.6$ km). The regridded method used is the nearest neighbor interpolation. Then, each series was standardized using the mean and standard deviation for that pixel. In addition to the original set of 20 variables (table 2, variables 1–20), lags of 1 month, 3 months, 6 months, 9 months, and 12 months was applied to the NDVI, soil moisture, and precipitation data to consider possible relationships between long-term moisture accumulation, stress, and vegetation. Finally, all data sets were divided into two parts. The first 183 months of the data set (80% of the data) was used for calibration of the model, and the most recent 49 months (20% of the data) was used in validation of the model. The model was additionally tested using two other subsets, 70%–30% and 60%–40% for calibration-validation respectively (table 4).

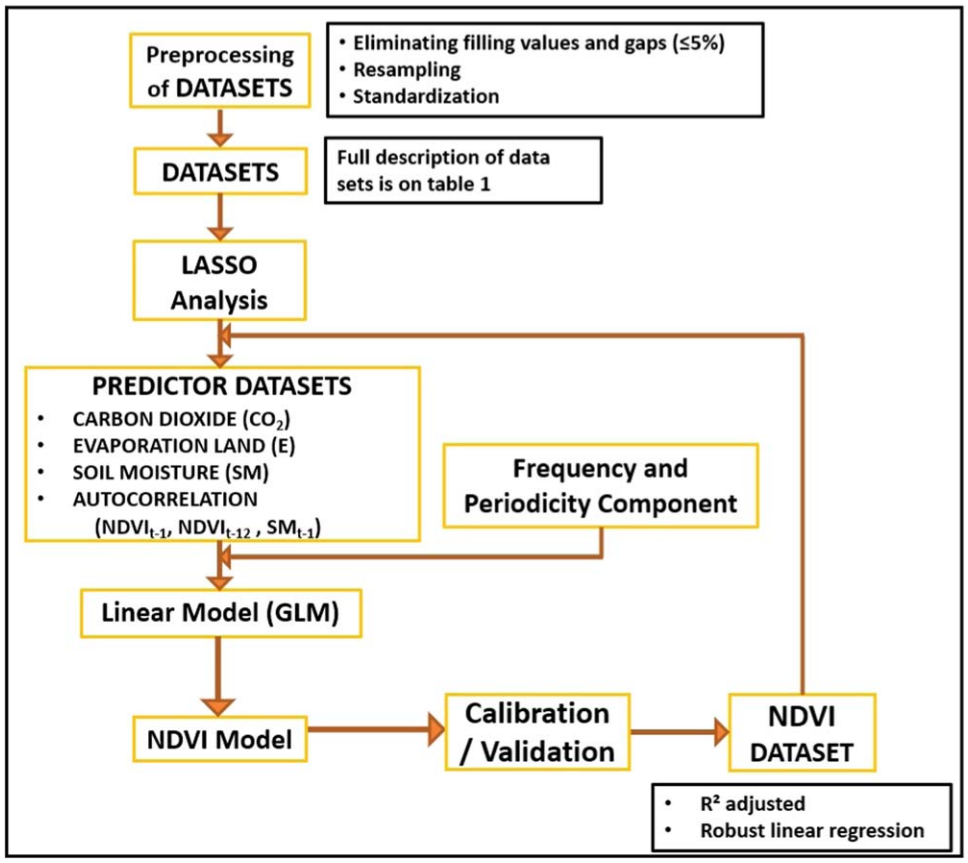

Figure 2. Diagram of the overall analysis structure and workflow.

For calibration, we used a regression-based technique known as LASSO [37, 38], to identify the best set of independent variables that can explain NDVI variability while dealing with multicollinearity between the possible independent variables for each pixel. The LASSO method assumes a linear relationship between dependent and independent variables, with Gaussian noise, but constrains the L1 norm of the regression coefficients in the least-squares optimization. The inclusion of the L1 constraint ($\alpha = 1$) results in the shrinkage of certain coefficients to zero, hence providing a way to obtain the best subset of independent variables. The variables associated with the coefficients that were shrunk to zero were dropped, giving a final set of 10 independent variables.

Then, we developed a generalized linear regression model, which used maximum likelihood [39] to model NDVI with independent variables selected from the LASSO analysis along with functions of time to account for seasonality and interannual variability. The GLM used a Gaussian link function, which makes our model equivalent to a linear model. Autoregressive (AR) spectrum analysis was used to detect frequencies in the NDVI and select the peak NDVI cycle amplitudes. Thus, the model took the following form:

$$g(\boldsymbol{Y_i}) = \beta_0 + \beta_1 \boldsymbol{x_{1i}} + \ldots + \beta_n \boldsymbol{x_{ni}} + \alpha_1 \sin(\omega_1 \boldsymbol{t}) + \ldots + \alpha_n \sin(\omega_n \boldsymbol{t}) + \boldsymbol{\epsilon_i} \quad (2)$$

where i indicate that the values are different for each grid point, β_0 is the intercept, β_i (α_i) is the independent variable (frequency) coefficient, x_{ji} is an independent variable, ω_i is a frequency component, and ϵ_i is an error term, assumed to be a Gaussian distributed random variable. Depending on the number of frequencies present in the data, the model could use up to three sine components.

Different combinations of the ten possible independent variables, including lags, were individually tested, and p-values were calculated. A series of GLMs with a minimum of one independent variable and a maximum of ten (including autocorrelation) were constructed and run; the models with higher correlation with the NDVI trend, using the least number of independent variables and best p-value were selected. Models were compared using adjusted R^2 (hereafter referred to as R^2) and a robust linear regression model. The robust regression model used iteratively reweighted least squares (IRLS) to assign a weight to each data point. This weight was assigned equally to each data point in the first iteration and model coefficients were estimated using ordinary least squares. At following iterations, points further from the model predictions in the previous iteration are given lower weight, then model coefficients are recomputed using weighted least squares. This process continued until the values of the coefficient estimates converged within a specified tolerance. This weighting ensured that the

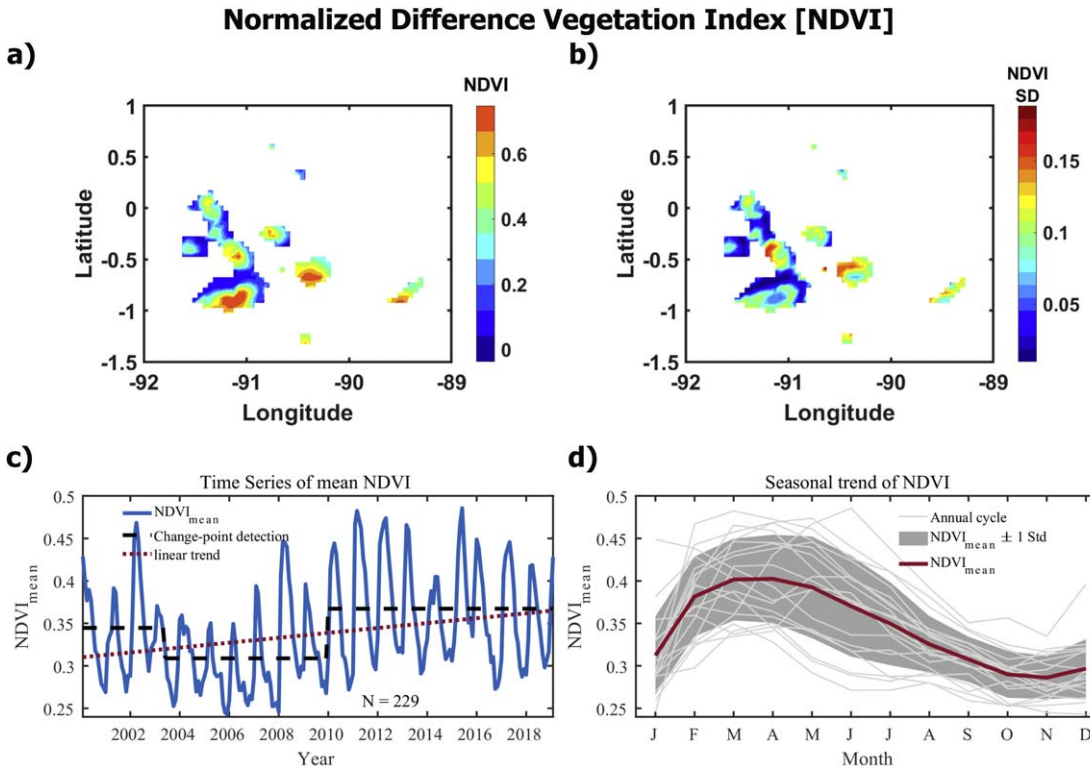

Figure 3. (a) Spatial distribution of the average normalized difference vegetation index (NDVI) on the Galapagos Archipelago, and (b) standard deviation of each NDVI pixel. (c) Time series of mean normalized difference vegetation index (NDVI) for 229 months with two step changes showing three phases and a linear trend for the whole period, (d) The mean seasonal cycle of NDVI ±1 standard deviation represented with a red line and gray shading on the Galapagos Archipelago.

final model was not much affected by outliers [40] and allowed us to identify a smaller set of independent variables that robustly improve NDVI prediction.

Finally we compared the model-predicted NDVI with the observed data at each pixel, and we validate the model using the second part of the data. Figure 2 shows the overall workflow.

3. Results and analysis

3.1. Description of NDVI data

Figure 3(a) shows the spatial distribution of the average NDVI value, and figure 3(b) shows the standard deviation (SD). The SD values reflect the temporal variability of NDVI at each pixel. The maximum standard deviation is 0.18 (in NDVI units). Regions with low NDVI values also have low standard deviations, as shown in figure 3(b). However, higher standard deviations are not necessary indication of higher NDVI values.

To get a general overview of the NDVI trend for the years of this study, NDVI values were averaged spatially, giving a mean NDVI value for the whole archipelago per month. Figure 3(c) shows a linear trend and a two step-change model for mean NDVI, along with a clear seasonal cycle. Using a linear regression on the whole period, mean NDVI shows a 1% annual increase ($R^2 = 0.065$, p-value = 5.73E-05, RMSE = 64.1). Two significant change points were found using Pettitt's test. In significance order, one occurred at January 2010 (month = 120, p-value = 2.36E-08) and the other occurred at June 2003 (month = 41, P-value = 3.50E-03). During the first period, the NDVI average value was 0.3447 for 3.33 years (40 months). Then, during the second step, NDVI decreased, with an average NDVI value of 0.3089 for the next 6.58 years (79 months). Then, for the last step, NDVI increased, with an average NDVI value of 0.3673 for the next 9.17 years (110 months) figure 3(d) shows the seasonal cycle of NDVI. NDVI values are higher from January to June (max: 0.33–0.48 in May [interannual range]), which is a warm and wet season, and lower from July to December (min: 0.26–0.36 in October), which is a cool and dry season. The dry season has lower interannual variability in NDVI, while the wet season has higher interannual variability.

In figure 4, topographical features such as aspect, slope, and elevation are compared with average NDVI to observe relationships between vegetation and the island topography. Figure 4(a) (left) shows the spatial distribution of the aspect angle over the whole archipelago, and (right) the radial distribution in degrees of each

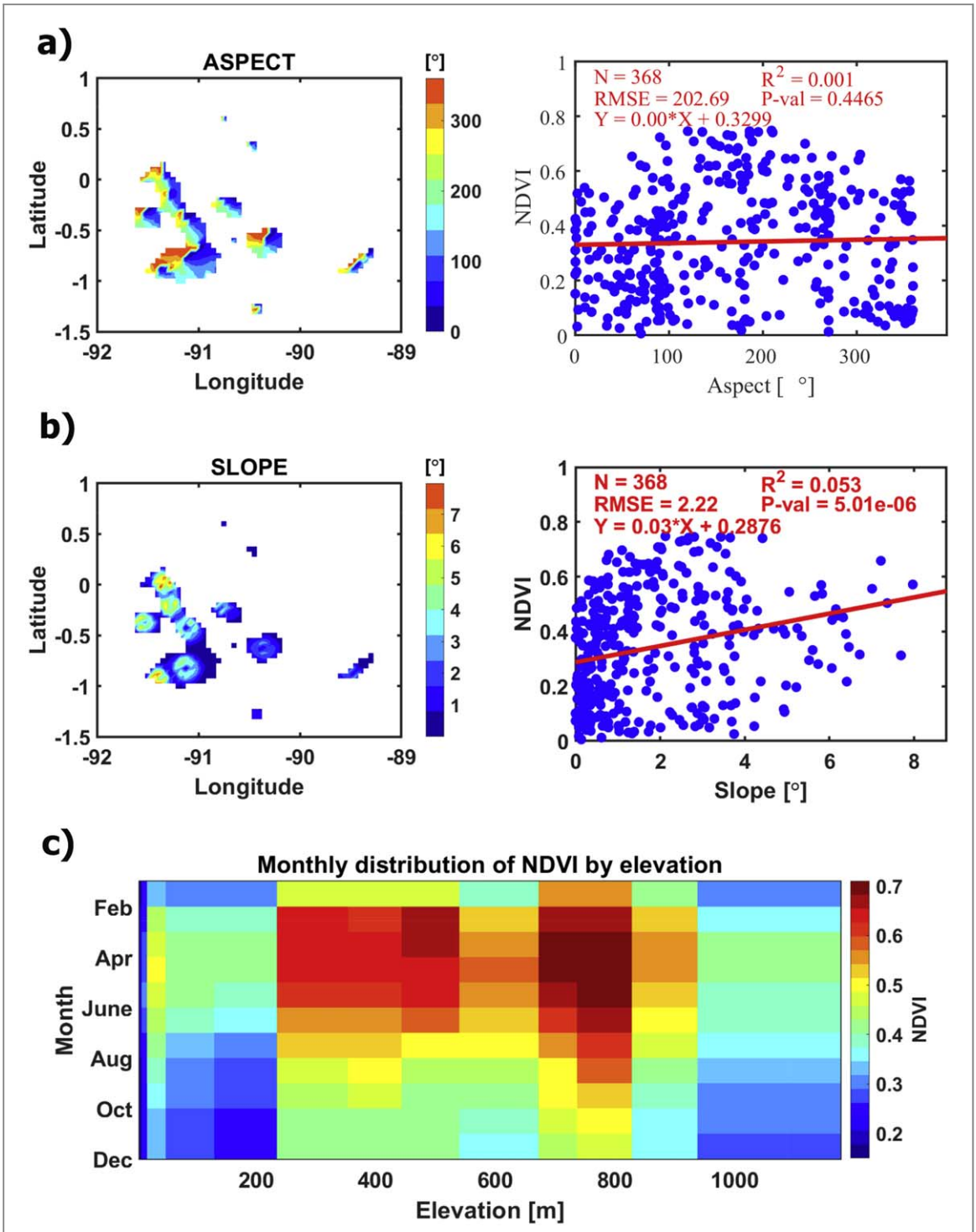

Figure 4. (a) Spatial distribution of aspect across the Galapagos Archipelago and Aspect versus NDVI. (b) Spatial distribution of slope across the archipelago and Slope versus NDVI. (c) Monthly distribution of NDVI by elevation.

pixel. There is not a clear relationship between aspect and NDVI values. $R^2 = 0.1\%$ (p-value = 4.46E-1) figure 4(b) (left) shows the spatial distribution of the angle of the slope over the whole archipelago, and (right) the scatter plot of the dependence of NDVI to the slope angle. $R^2 = 5.29\%$ (p-value = 5.01E-6). Elevation has the clearest relationship with NDVI. There are two marked elevation zones where NDVI values are very high (over 0.6). One is between 2–16 MASL and the other is between 250–800 MASL (figure 4(c)). Higher variability of the seasonal trend in NDVI is observed at elevations lower than 20 MASL. Because of this, later analysis is divided in two elevation ranges. One elevation range includes the pixels with an elevation lower or equal to 20 meters above sea level (MASL) (35% of the area of study) and the other includes pixels with an elevation higher than 20 MASL (65% of the area of study). Elevation is unevenly distributed, resulting in gaps in NDVI data for certain elevation ranges; thus, figure 4(c) shows monthly mean NDVI based on a 100 m elevation interval.

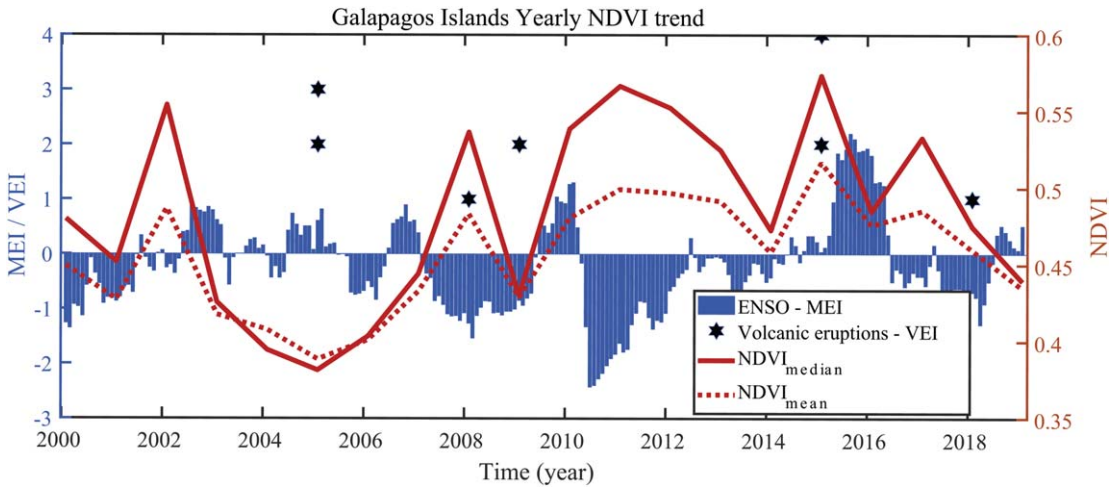

Figure 5. Yearly maximum NDVI mean (red dashed line) and median (red line) values for the whole archipelago, monthly MEI values (blue bars), and registered volcanic eruption using VEI scale (black asterisks). Negative MEI values indicate La Niña events and positive MEI values indicate El Niño events.

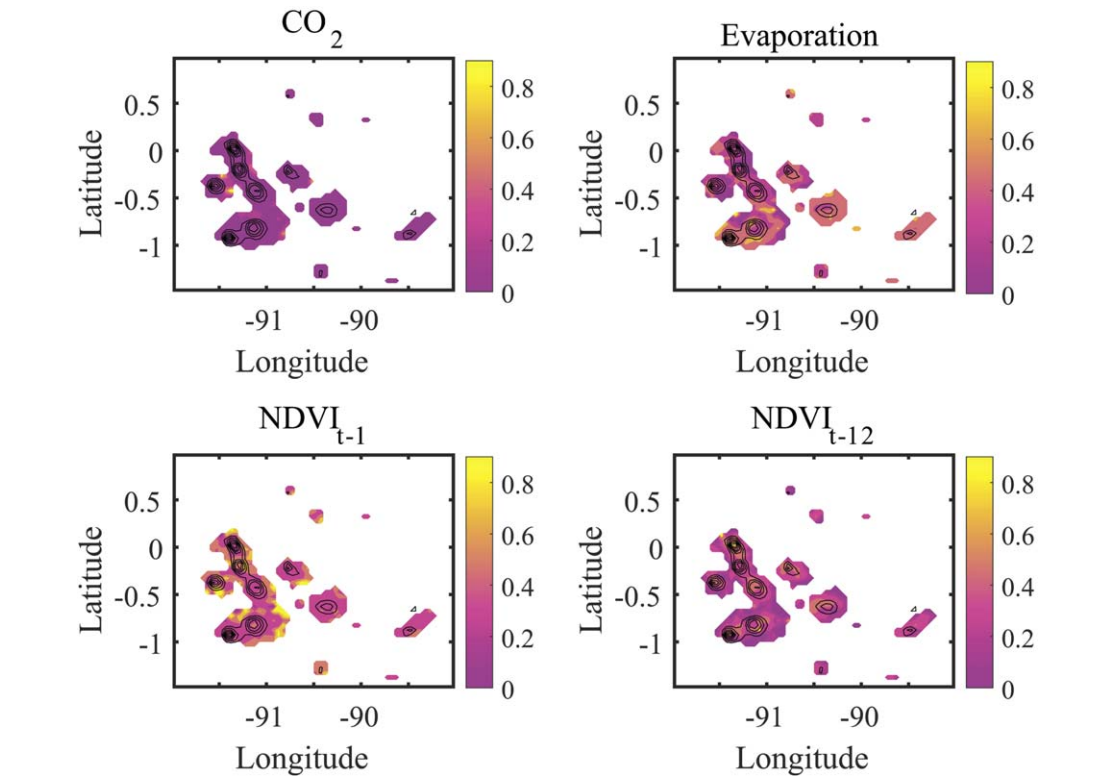

Figure 6. Spatial representation of the variance decomposition of each independent variable used in the model with elevation in black contour lines.

Therefore, this figure does not show well trends within the lower elevation range. Even though pixels representing 2–16mof elevation comprise approximately 10% of the data, they are lost when averaged with the rest of the pixels in this range because of the decreasing number of pixels with respect to increasing elevation.

Figure 5 shows maximum annual NDVI (mean and median, using dashed and solid red lines, respectively). Monthly MEI values are presented in blue bars, and volcanic eruptions that registered on the VEI scale are presented with black asterisks. Negative MEI values indicate La Niña events, and positive MEI values indicate El Niño events. Figure 5 shows that larger magnitude ENSO events tend to correlate with higher average maximum NDVI values, while minor events (such as those indicated between 2003 and 2008) tend to correlate with lower

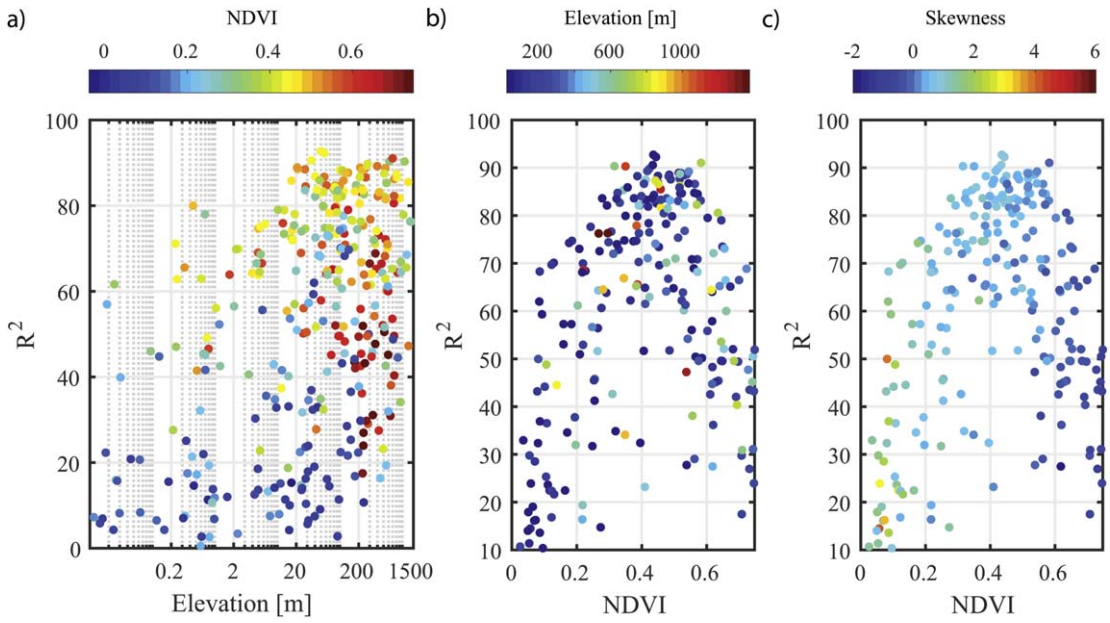

Figure 7. (a) Distribution of R^2 based on average NDVI and elevation in log scale, (b) based on average NDVI and elevation, and (c) based on average NDVI and NDVI skewness. (b) and (c) show those pixels that have an elevation higher than 20 MASL.

Table 3. List of independent variables used to build the model and the number of pixels that they influence (expressed as percentage of the study area), RMSE, adjusted R^2, and P-value for carbon dioxide, evaporation, autocorrelation of one-month and 12-months lag in NDVI for the study period, 2000–2019.

Independent variable Name	N (%)	RMSE	Adj. R^2	p-value
Carbon dioxide (CO_2)	189 (51%)	0.079	0.1597	2.39E-10
Evaporation land (E)	156 (42%)	0.662	0.4000	8.81E-27
Autocorrelation of NDVI ($NDVI_{t-1}$)	251 (68%)	0.306	0.7991	8.83E-77
Autocorrelation of NDVI ($NDVI_{t-12}$)	168 (46%)	0.480	0.4772	3.49E-32

overall NDVI values. Higher NDVI values are often found with negative MEI values, suggesting that during these events vegetation may benefit from extended effects of El Nino, even after it passes.

3.2. Selection of independent variables

After gap filling and standardization, 368 (374) pixels from Terra (Aqua) were available to be used in the analyses. The ten independent variables selected after lasso regression are CO_2, evaporation, precipitation, surface temperature day, and lags of 1 month for NDVI and soil moisture, 3 months for soil moisture, and 12 months for NDVI, soil moisture and precipitation. The average R^2 for NDVI was calculated using the independent variables named previously for each elevation region. For pixels with elevation less than or equal to 20 MASL, $R^2 = 46.88\%$, and for elevation higher than 20 MASL, $R^2 = 69.03\%$. NDVI autocorrelation ($NDVI_{t-1}$), CO_2, evaporation (E), and NDVI autocorrelation ($NDVI_{t-12}$) were selected as independent variables to build the final model (table 3); each was the dominant independent variable for 68%, 51%, 42%, and 46% of the area of study, respectively. Using these 4 independent variables (section 3.3), the R^2 values changed to 43.28% ($\Delta R^2 = -3.6\%$ from using all the lasso-selected independent variables); for pixels with elevations $\leqslant$20 MASL, and R^2 changed to 66.41% ($\Delta R^2 = -2.62\%$ from using all the lasso-selected independent variables) for elevations higher than 20 MASL.

3.3. Building of the model and calibration

The following results are based in the calibration set. R^2 values presented on the graphs are based on the mean NDVI, and modeled mean NDVI for the whole archipelago. Figure 8 shows the spatial distribution of R^2 for the two elevation ranges previously mentioned. In the top left panel, shoreline pixels with elevations less than or equal to 20 MASL, are highly variable in R^2 values. The number of pixels in this range is 130, representing 35% of the study area. Pixels with elevations higher than 20 MASL (top right panel) generally have higher R^2 values. For each range, each pixel shows the

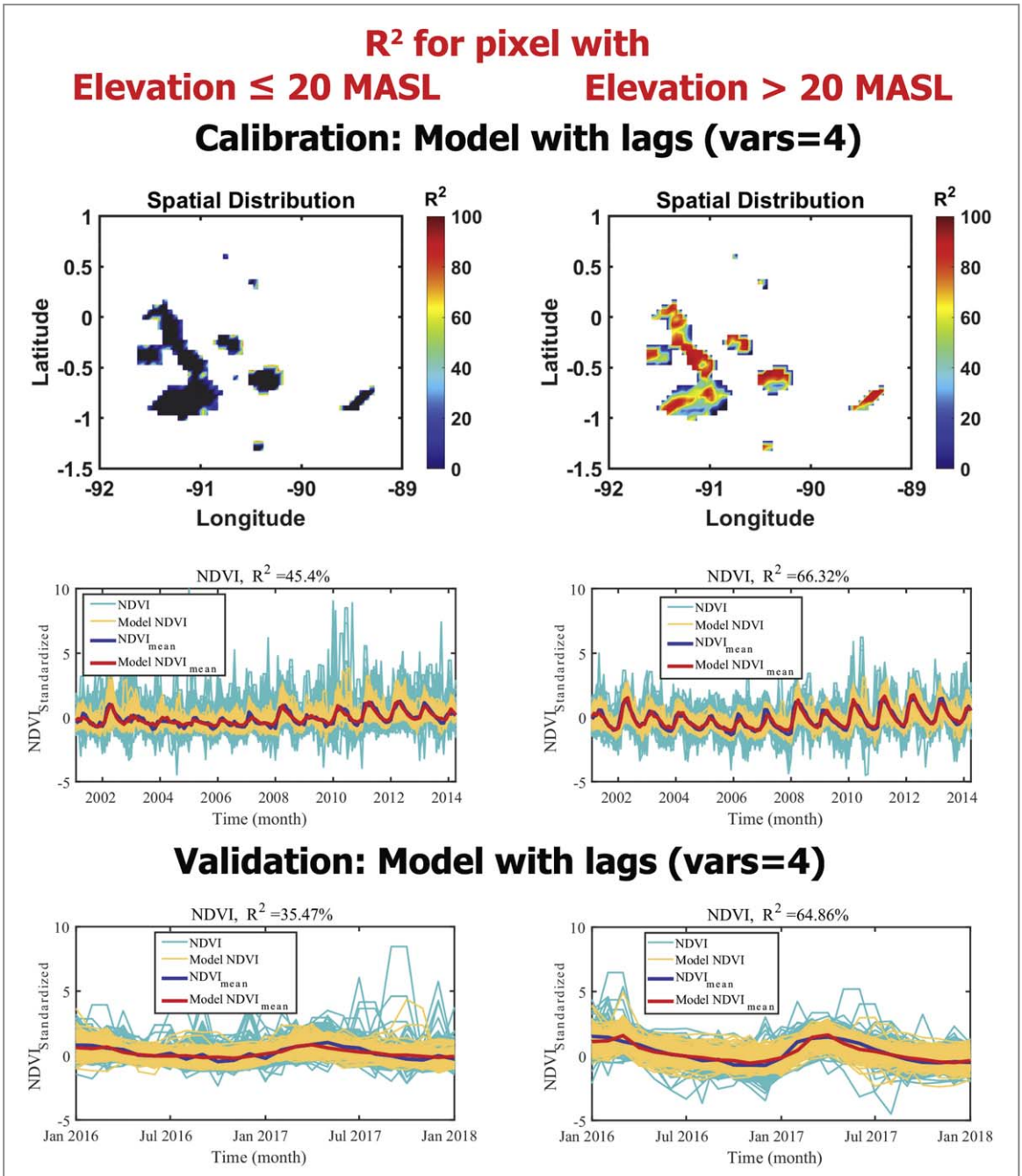

Figure 8. On the left side R^2 values for pixel with an elevation lower or equal than 20 MASL, and on the right side R^2 values for pixel with an elevation higher than 20 MASL. The first row shows the spatial correlation of the normalized difference vegetation index (NDVI) R^2. Pixels in black represent pixels outside of the area of analysis. The second row shows the time series of the normalized difference vegetation index (NDVI)(light blue) and its average (blue) and the generalized linear regression model by pixel (mustard) and its average (red). These results are for the calibration section of the model, and on the third row are the results for the validation section of the model.

time series of the modeled NDVI, the actual NDVI trend, and their average for the whole archipelago (figure 8, second row). The peaks of the modeled NDVI have lower amplitude than the real observations, but the periodicity and trends match precisely.

Figure 9 top panel shows a bar plot of all usable pixels in the archipelago from which R^2 was calculated. The lowest R^2 is 4.5% and the highest R^2 is 93.8%. The values of R^2 range from 4.5%–89% and 16–93.8% in lower and higher elevation zones, respectively. The overall mean R^2 is 58.5%; for pixels with an elevation higher than 20 MASL mean R^2 is 66.32% and for pixels with an elevation lower than 20 MASL mean R^2 is 45.4% .

Variance decomposition was used on the independent variables retained in the final model (figure 6). NDVI autocorrelation of one month lag explained the largest proportion of the modeled NDVI variance (42%), followed by evaporation (33%), NDVI autocorrelation at twelve months lag (21%), and CO_2 (5%). After subtracting the effect of the NDVI autocorrelation, evaporation explained 83% of the NDVI variance, followed

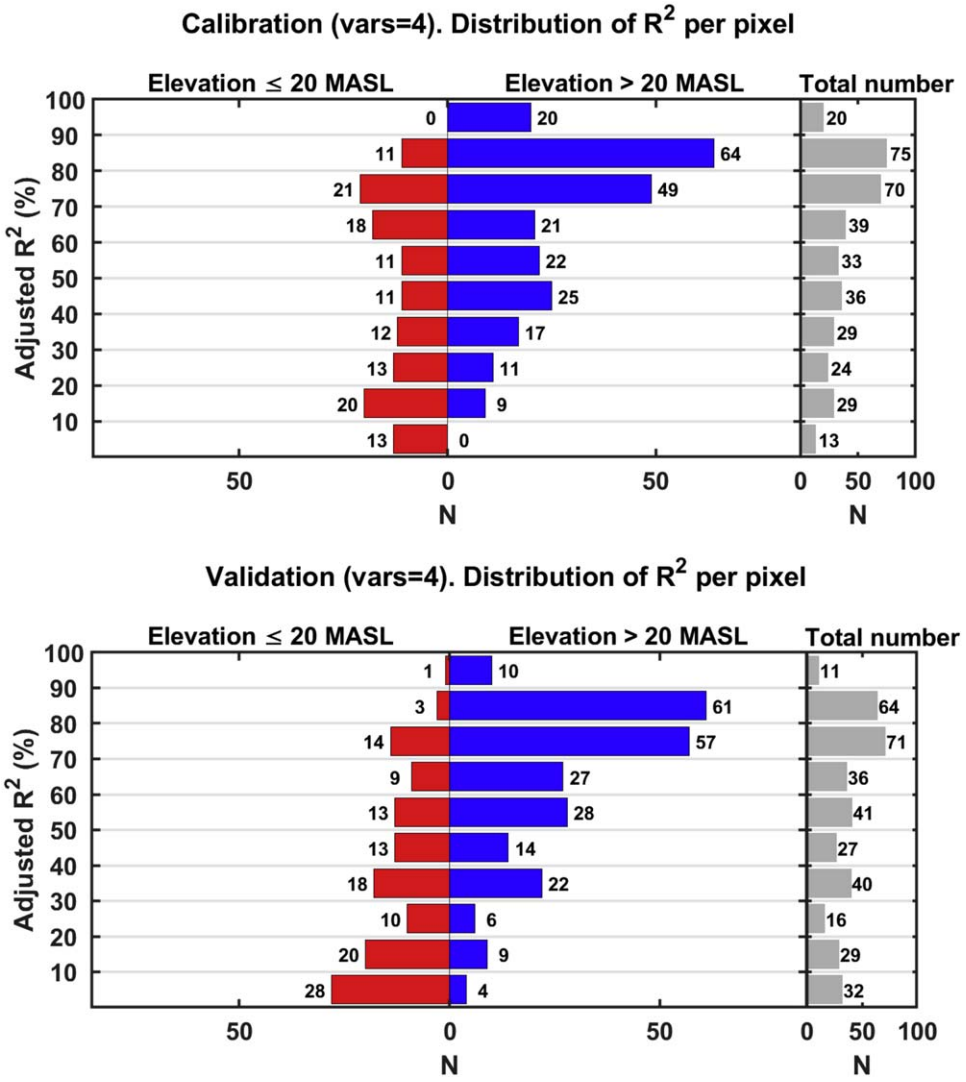

Figure 9. Bar plot of R^2 of the calibration section (top) and validation section (bottom) for the whole archipelago, segmented by two elevation ranges. One segment shows pixels with an elevation lower or equal to 20 MASL in red, and the other segment shows pixels with an elevation higher than 20 MASL in blue. On the right side of the plot, the total number of pixels corresponding to each R^2% is presented in grey.

by CO_2, which explained 17%. CO_2 was more important as an independent variable in areas with very low NDVI values, as shown in figure 3 of Rivas Torres *et al* (2018) [2], corresponding to sites with relatively recent lavas and scarce or absent vegetation. Fernandina Island and Isabela Island, which have the most active volcanoes of the archipelago [41] present good examples of low NDVI values associated with recent and active lava flows [42]. Evaporation increases in importance with the presence of vegetation, but its predictive power is also influenced by the type of vegetation present, which varies across the archipelago. NDVI on Santa Cruz Island, which is one of the main agricultural zones of the archipelago, is highly influenced by evaporation. NDVI autocorrelation (lag-1 month) shows higher predictive power close to shore on Isabela, Fernandina, Santiago, San Cristobal, and Floreana islands. The area with dense vegetation such as between active volcanoes on Isabela and Fernandina islands, an unusual ecoregion, also corresponds to higher values of NDVI autocorrelation (lag-12 months).

Different combinations of independent variables and periodic trends were used in an attempt to improve the predictive power of the model for the entire archipelago. The AR spectrum analysis showed that the peak NDVI cycle amplitudes (in decreasing order) have periods of 17, 34, and 13 months. After a series of iterations, the best fit to the NDVI data included periodic trends with 17 and 34-month periods. These timescales are influenced by ENSO [34]. However, adding these periodic terms to the model only resulted in a slight increase (0.002%) in the average values of R^2 across pixels. Therefore, they were not included in the final model.

Removing the NDVI autocorrelation resulted in a dramatic decrease in R^2 to a maximum of 27.82% for pixels with an elevation $\leqslant$20 MASL and 39.04% for pixels with an elevation $>$20 MASL (appendix, figure A2–2).

Table 4. List of R^2 from calibration (higher% data) and validation (lower% data) using different data sets length. Validation values are compared to NDVI from Aqua data set (last two columns).

Elevation	% NDVI data set	RMSE	R^2	RMSE	R^2
Lower or equal to 20 MASL	60	0.1621	0.3919	—	—
Greater to 20 MASL	60	0.2033	0.6257	—	—
Lower or equal to 20 MASL	40	0.1558	0.3477	0.1261	0.3417
Greater to 20 MASL	40	0.1839	0.5977	0.2051	0.5231
Lower or equal to 20 MASL	70	0.1761	0.4426	—	—
Greater to 20 MASL	70	0.2125	0.6537	—	—
Lower or equal to 20 MASL	30	0.1654	0.3725	0.1228	0.3465
Greater to 20 MASL	30	0.2043	0.6068	0.2099	0.5216
Lower or equal to 20 MASL	80	0.1730	0.4540	—	—
Greater to 20 MASL	80	0.2707	0.6632	—	—
Lower or equal to 20 MASL	20	0.1514	0.3547	0.1134	0.3612
Greater to 20 MASL	20	0.2318	0.6486	0.2407	0.5496

In both elevation regions, the model without autocorrelation does not fit with the observed NDVI trend during the period of lower NDVI in 2003–2008. The peaks of modeled NDVI have lower amplitude than the observations, but the periodicity and trends match exactly for pixels with elevations higher than 20 MASL. This match is weaker for pixels with elevations $\leqslant$20 MASL. Figure A2–3 (appendix) shows a bar plot of all usable pixels in the archipelago from which R^2 was calculated. For elevation less than or equal to 20 MASL, R^2 is between 0%–71%. The number of pixels with an R^2 greater than 50% is 26. For elevation higher than 20 MASL, R^2 range is between 0%–75%. The number of pixels with an R^2 lower than 20% is 41. The different ranges of R^2, 0%–70%, have an average number of pixels of 50.

Because NDVI autocorrelation (one month-lag, and 12-months lag) improved the fit of the model, especially for the period of lower NDVI (2003–2008), NDVI autocorrelation was retained in the model. The final model was reduced to

$$\boldsymbol{NDVI} = \beta_0 + \beta_1 \boldsymbol{CO_2} + \beta_2 \boldsymbol{E} + \beta_3 \boldsymbol{NDVI_{t-1}} + \beta_4 \boldsymbol{NDVI_{t-12}} + \boldsymbol{\epsilon} \quad (3)$$

Residuals were temporally independent, indicating that the method produced an adequate fit to the NDVI dataset. Kolmogorov-Smirnov (K-S) test was applied to check the fit of the residuals of each pixel to a normal distribution. Residuals did not follow a standard normal distribution (p-value $= 6 \times 10^{-6}$). The mean value of residuals for the whole archipelago is 0.0036; the model very slightly underestimates observed NDVI, which is most apparent in peak values, but otherwise matches the observed trend well. Ten other normality tests [43] ($\alpha = 0.05$) were applied on the residuals. These results are presented in table A1–1.

Figure 7 shows the comparison of R^2 with elevation, average NDVI, and NDVI skewness. Figure 7(a) presents the distribution of R^2 based on pixel elevation and average NDVI value. Pixels with R^2 values $<$50% are generally those with elevation $<$20 MASL and NDVI values $<$0.3. These pixels could be representative of disturbed or arid zones. Those pixels with higher R^2 values and elevations generally have NDVI values $>$0.3, indicating denser vegetation. Figure 7(b) shows R^2 based on the average NDVI for those pixels that have an elevation higher than 20 MASL Pixels show an inverted u-pattern where pixels with lower ($<$0.15) and higher ($>$0.7) values of average NDVI tend to have a lower R^2 value. Figure 7(c) shows the distribution of the NDVI skewness for the comparison of R^2 with average NDVI. Pixels that represent extreme average NDVI values likewise have extreme skewness values (< -1 or >1). Based on these results, NDVI under 20 MASL is more variable and not as influenced by the independent variables used in the current model.

3.4. Validation

The model was validated using data from January 2016 to January 2018 (the validation subset described in 2.3). The period of analysis is shorter than the full 4 years because of the need to incorporate NDVI autocorrelation (12 month lag) in the model. The bottom left panel of figure 8 shows the average R^2 value for those pixels with elevation $\leqslant$20 MASL is 35.47%, while the average R^2 values for higher elevations is 64.86% (bottom right panel). At higher elevation, the values of R^2 are similar to those values obtained from calibration processes with differences (Δ) $<$5%. On the other hand, $\leqslant$20 MASL, Δ may be up to 10%. RMSE values are low for all elevations, and lower for elevations $\leqslant$20 MASL at any percentage of the data set. Table 4 shows RMSE and R^2 for three sets of calibration and validation data. Additionally, the model was run with NDVI from the Aqua data set for the same validation period tested using data from Terra (table 4, last two columns). Results for lower

elevation are similar in both data sets. For higher elevation, R^2 associated with the Aqua data set are lower (up to 10%). Values are stable with different lengths of data for validation.

In the bottom panel of figure 9, a large number of pixels with elevation lower or equal to 20 MASL have an R^2 value that range between 0 to 20% while at higher elevation, many pixels have R^2 values between 70 to 90%. Overall, the maximum R^2 is 75%.

4. Discussion

4.1. Interpretation of results

Climate change is expected to trigger changes to temperature and precipitation patterns across the globe. In the Galapagos, climate change is expected to create warmer and wetter conditions with stronger El Niño events; all of these are conditions that enhance plant growth and greenness, which can be measured through NDVI. Our results show a statistically significant increase in NDVI of 1% per year over the 19 years of study, which could be a signal of anthropogenic climate change (e.g. changes in temperature, precipitation, or CO_2 concentration), although a change-point detection (step change) analysis shows a decrease in NDVI values from 2003 to 2010, which could be related to ENSO events and volcanic eruptions at that time. Furthermore, small peaks in average/median annual maximum NDVI (2002, 2008, 2011, 2015, and 2017) could reflect the multi-year periodicity associated with ENSO [34]. Importantly, not all the above years correspond to El Niño events, e.g. 2008, but this year is noted as having been unusually warm and rainy in the Galapagos [5]).

Our results show that, out of over 20 potential predictors considered, temporal autocorrelation at 1-month and 12-months lag, evaporation, and carbon dioxide levels are the most important independent variables for modeling NDVI in the Galapagos Islands, yielding values of R^2 for the GLM that range from 5%–93%. Prior research supports the importance of considering temporal autocorrelation in explanatory NDVI models [44]; given the uniquely stable character of Galapagos ecosystems, it is unsurprising that intrinsic variables and/or system memory are important for modeling vegetative dynamics.

We offer conjectural interpretations of some of the spatial patterns seen in the coefficients of our model for NDVI variability. These interpretations are subject to refinement as longer and finer-resolution data sets become available.

Evaporation appears to be an important proxy of NDVI in the ecosystems that exhibit strong seasonality, as shown in figure 6. This is particularly true of the deciduous forests and shrublands across the archipelago (see Rivas-Torres *et al* (2018) [2], figure 3). In Santiago, Santa Cruz, San Cristobal, Genovesa, and Pinzon, the dominant ecosystem is dry forest, where primary productivity is limited by moisture availability and dependent on seasonal rainfall [45]; importantly, precipitation is extremely variable in the Galapagos (spatially, inter-annually, and intra-annually) [45] and its relationship to NDVI is not direct—especially in the xeric ecosystems that dominate the archipelago [2, 45]. On the other hand, increased evaporation—even at a coarse scale—can indicate overall conditions that are favorable for plant growth, including increased moisture availability, higher temperatures, and increased solar irradiation [5, 45, 46]. Evaporation and leafiness are positively associated in seasonal forests, because plants only produce larger leaves with larger surface areas when they can afford to lose water (figure A2–1). The inverse is also true; leaf shedding is a phenological response of deciduous organisms to water stress. Indeed, other studies of vegetative response to atmospheric and climate variables have shown evaporation to be more strongly correlated with above ground primary productivity (AGPP) than precipitation [46–48]. Chen *et al* (2019) [46] offer a compelling explanation for this phenomenon, arguing that evaporation provides a more accurate estimation of NDVI by considering the interaction of moisture availability with available solar energy—both of which are limiting factors for plant growth [46]. A CO_2 fertilization effect has been observed in the tropics as well as, more famously, in the northern hemisphere; Krakauer *et al* [49] showed that for Nepal, global CO_2 concentration was better at modeling NDVI intrannually than climate variables, (which were more important interannually). Moreover, we suggest that in volcanic zones of the Galapagos, pioneer species make global CO_2 concentration a better indicator of NDVI than local climatic variables. Fast growing plants (such as pioneers) are more sensitive to elevated CO_2 than slow growers [50–52]. On primary lava, temperature and moisture are highly variable due to the lack of surface vegetation and soils, yet early successional species are adapted to full sun conditions and photosynthesize more rapidly than later successional species. Idealized curves for photosynthesis rates from Bazzaz (1979) [53] show that in full sun, early successional species achieve photosynthesis rates of $\approx$25mg $CO_2 dm^{-2} h^{-1}$, compared to $\approx$15 mg $CO_2 dm^{-2} h^{-1}$ for mid successional species, and $\approx$2–5 mg $CO_2 dm^{-2} h^{-1}$ for late successional species. Pioneer species in the Galapagos include numerous representatives of the fast growing 'weeds' of the Poaceae family (including *Aristida repens* and *Eragrostis mexicana* (lovegrass)) and Asteraceae family (including *Sonchus oleraceus L.*, a dandelion relative, and *Baccharis gnidifolia*). Just outside of the ashy barren zones, pioneer communities also include members of the Solanacae family (e.g. *Solanum erianthum*, the potato tree,*Lycopersicon* (tomato),

Exedeconus miersii (shore petunia), and numerous vines. Solanaceae are C3 photosynthesizers, which have been shown to respond to elevated CO_2 levels with higher levels of photosynthesis, compared to C4 species [54]. The predictive power of CO_2 for NDVI in areas that correspond to recent lava flows suggests a quickly colonizing, resilient pioneer community that is adapted to local climate variation and geologic disturbance. At the same time, sensitivity to CO_2 levels suggests that global CO_2 increase may have the capacity to drive successional dynamics and pioneer community composition in the future [55].

In the same way as volcanic zones, shoreline areas have a very dynamic and challenging environment that promotes pioneer communities. In these areas, the importance of previous-month NDVI as a predictor of current-month NDVI speaks to the sustainability of these communities within seasons in these harsh environments. Finally, previous-year NDVI as a predictor represents the adaptability of vegetation to slower changes in climate factors on the region. This is especially visible for dense vegetation where NDVI values are higher.

Since our model works best to explain NDVI at higher elevations (>20 MASL), and works especially well for sites with elevation >200 MASL, low elevation areas and littoral zones may represent unique plant communities (endemic Galapagos dry forest, shrub land, and/or Mangrove forests), whose overall productivity responds most strongly to variables not considered in this study. The model works best to explain NDVI in the dry season (June-December), with higher variability during rainy periods, including El Niño events, a trend also observed by Kalisa *et al* (2019) [56]. Seasonally anomalous NDVI values could be representative of a nearly universal greening response to water availability during these periods.

4.2. Potential implications for management

Trueman *et al* (2011) [5] present a detailed description of how a changing climate might affect the phenology of the islands' varied vegetation. Increased El Niño, temperatures, and precipitation are generally expected to enhance greenness across the islands, by stimulating growth and producing larger flowers, leaves, and bushes, as well as through longer growing seasons and leaf retention periods. Although climate change is broadly expected to benefit Galapagos flora, it is the non-native species that are expected to benefit most [5]. Among endemic species, sustained wet conditions can be detrimental. *Opunita echios* (prickly-pear cactus) *Jasminocerus thouarsii* (candelabra cactus), and *Bursera graveolens* (palo santo tree) can become waterlogged, lose large proportions of seedlings, and even break or collapse under the weight of rapidly growing vines and herbaceous species, which proliferate under wet conditions[2, 57].

In the highlands, *Scalesia pedunculata* (daisy tree) is also speculated to be vulnerable to climate change because historically, it has been shown to die back in response to El Niño events [5, 57]. In addition to direct mortality, dramatic die offs of Scalesia could fundamentally alter the ecosystem by ceding valuable growing territory in the humid highlands to opportunistic non-native species that have been shown to respond to El Niño conditions with increased growth [5].

Over the last several decades, there has been a concerted effort in the Galapagos to remedy direct anthropogenic disturbances to the ecosystem, including creation of the Galapagos National Park (GNP) Service and delineation of the official Park boundaries, the establishment of the Galapagos Biosphere Reserve and Whale Sanctuary, and concentrated eradication campaigns directed at large non-native herbivores such as goats and pigs [10, 58]. Watson *et al* (2010) [58] attempted to quantify the proportion of the archipelago that had been severely degraded by direct human activities (defined as active and abandoned agricultural areas and urban centers). The team arrived at an overall estimate of ~5%, which varied across the five inhabited islands from a maximum of 17% and 14% degraded land area on San Cristobal and Santa Cruz Islands respectively, to 8%, 5%, and 0.1% of total land area on Floreana, Isabela, Santiago, respectively. The team highlighted that the humid highlands are significantly more degraded than the lowlands, because growing conditions are more favorable for agriculture and invasive colonizers [58]. The Galapagos National Park currently occupies over 96% of the land area of the Galapagos Islands, and is protected from direct anthropogenic stressors such as land conversion and agricultural development [2]. Accordingly, anthropogenic impacts on the ecosystems of the archipelago are primary indirect for the period of our study, related to the dynamics of invasive species and—we argue—the early signals of a changing global climate.

The variability of our model's predictability across the Galapagos highlights the diversity of the flora that NDVI is designed to assess. Although climate is fundamentally important to vegetative activity and phenology, the variables that most strongly influence NDVI have been shown to vary, both by ecosystem and by plant type [56, 59]. The Galapagos has a unique climate and a stable biological community that has existed for thousands–perhaps tens of thousands–of years [22], sorted into unique ecosystems across the archipelago. Many of the ecosystems that flourish in the Galapagos are poorly quantified by NDVI because they are adapted to aridity and characterized by low levels of greenness; indeed, the evolutionary advantage of many endemic Galapagos plants comes from their toughness and ability to withstand extreme environmental conditions. Wetter summers and

more frequent El Niño events, which are expected over the next century, could benefit non-native plants whose life strategy is to grow fast and die young, at the expense of natives that evolved over millennia to survive extremes and poor growing conditions, and rely on seed banks that are similarly adapted to the severity of the environmental context.

Due to the lack of public local information from the different islands, it was not possible to corroborate the different remote sensing datasets used in the manuscript. Resolution of the different sets plays a significant role in exploring and understanding the vegetation on the Galapagos Islands. Since the finest resolution on this study is ~5.6 km from the NDVI dataset, a large number of small islands were not taken into account in the analysis. It is also worth mentioning that sparse vegetation, such as that common in dryland systems, is underestimated by the spatial resolution of MODIS (~5.6 km). This effect would be dominant at elevations lower than 20 MASL where vegetation is primarily deciduous grass, shrubland, and dry forest. Moreover, the 50 km resolution of the MERRA-2 reanalysis does not capture topographic variability across the archipelago well, which could be hindering our effort to model the factors driving variability in NDVI. In addition, global data such as atmospheric CO_2 concentrations can affect the credibility of the results since they are based on observations from higher latitude than the Galapagos Islands. We consider that those measurements give us a general idea of how atmospheric CO_2 affects vegetation in the archipelago, although they may not fully represent site-specific interactions between atmospheric CO_2 and flora at the different islands.

5. Conclusions

Our study demonstrates that NDVI can quantify the effect of changing climate variables on vegetation (NDVI $\geqslant$ 0.3) in the Galapagos Archipelago. This is significant, given the cultural, economic, educational, and scientific importance of these islands. Their capacity to serve as a museum and evolutionary laboratory for future generations will depend on the resiliency of the ecosystem and of its anchor vegetation in the face of global change, as well as on its responsible management. Remote sensing products and models such as ours should help scientists and conservationists develop tools to monitor the ecosystem continually and non-intrusively, especially in the highlands where our model works well to explain NDVI variability, and where species invasion is active and change in community composition is likely.

Studies of vegetation regime shifts on volcanic islands and in the tropics using remote sensing are still sparse, so there is much research to be done. It remains to be seen whether remote sensing products such as NDVI might develop the resolution to detect fine scale community dynamics or species invasions, or whether additional data and non-linear approaches [60] can clearly distinguish NDVI responses to various components of climate change, such as natural versus anthropogenic. Such studies will be useful to the global community as we monitor the responses of our natural systems to a changing climate.

Acknowledgments

This study is supported and monitored by The National Oceanic and Atmospheric Administration—Cooperative Science Center for Earth System Sciences and Remote Sensing Technologies (NOAA-CESSRST) under the Cooperative Agreement Grant #: NA16SEC4810008. The authors would like to thank The City College of New York, and the NOAA Educational Partnership Program with Minority Serving Institutions for fellowship support for Eder Herrera Estrella and NOAA Center for Earth System Sciences and Remote Sensing Technologies. The statements contained within the research article are not the opinions of the funding agency or the U.S. government, but reflect the author's opinions. We acknowledge the mission scientists and Principal Investigators who provided the data used in this research effort. We would also like to extend our sincere thanks to two anonymous peer reviewers who substantially improved the content and readability of our manuscript.

Data availability statement

No new data were created or analyzed in this study.

Appendix

A.1. Appendix A1

Climate Parameters of the Galapagos Islands

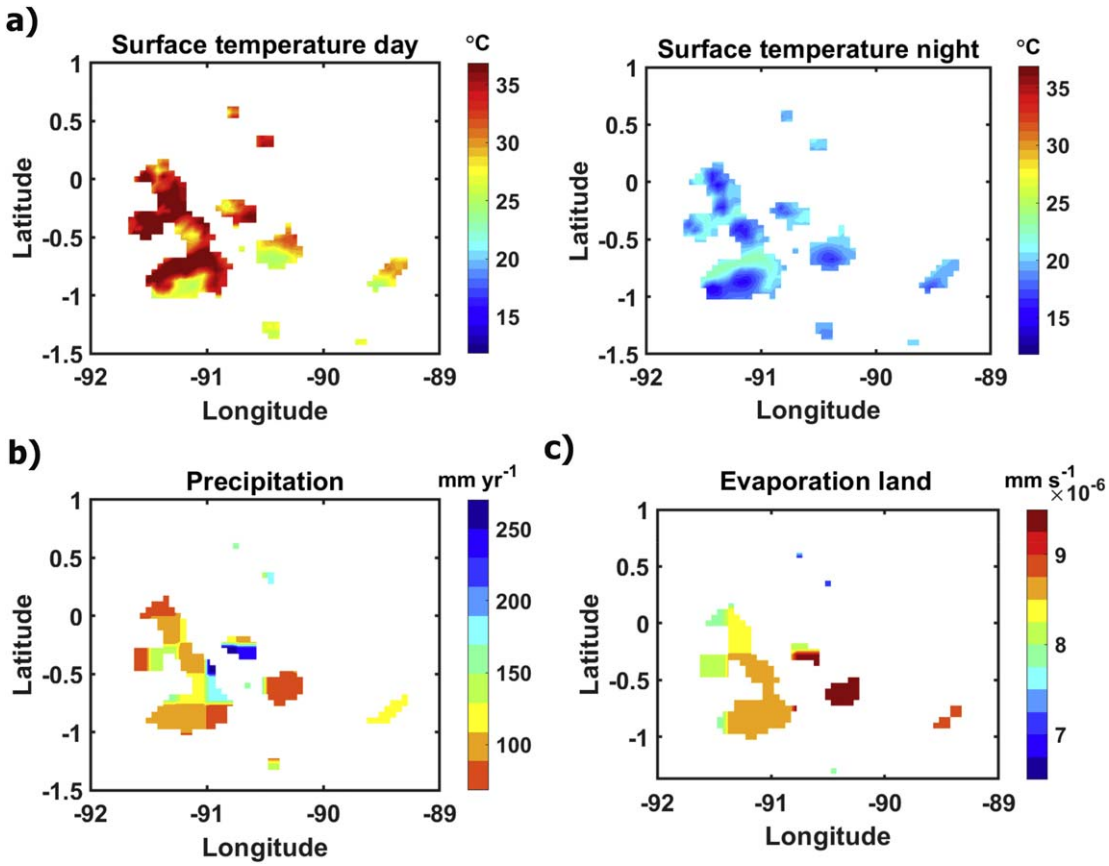

Figure A1-1. (a) Spatial distribution of the average surface temperature during the day (left) and at night (right) in Celsius (C). (b) Georeferenced Precipitation in mm yr^{-1} over the Galapagos Archipelago. (c) Georeferenced evaporation from land in mm s^{-1} over the Galapagos Archipelago.

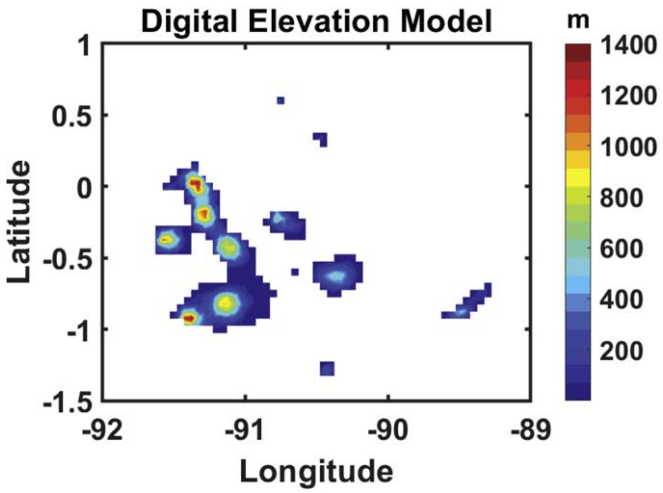

Figure A1-2. Digital elevation Model of the Galapagos Island in meters and a resolution of 0.05° from ASTER v2.

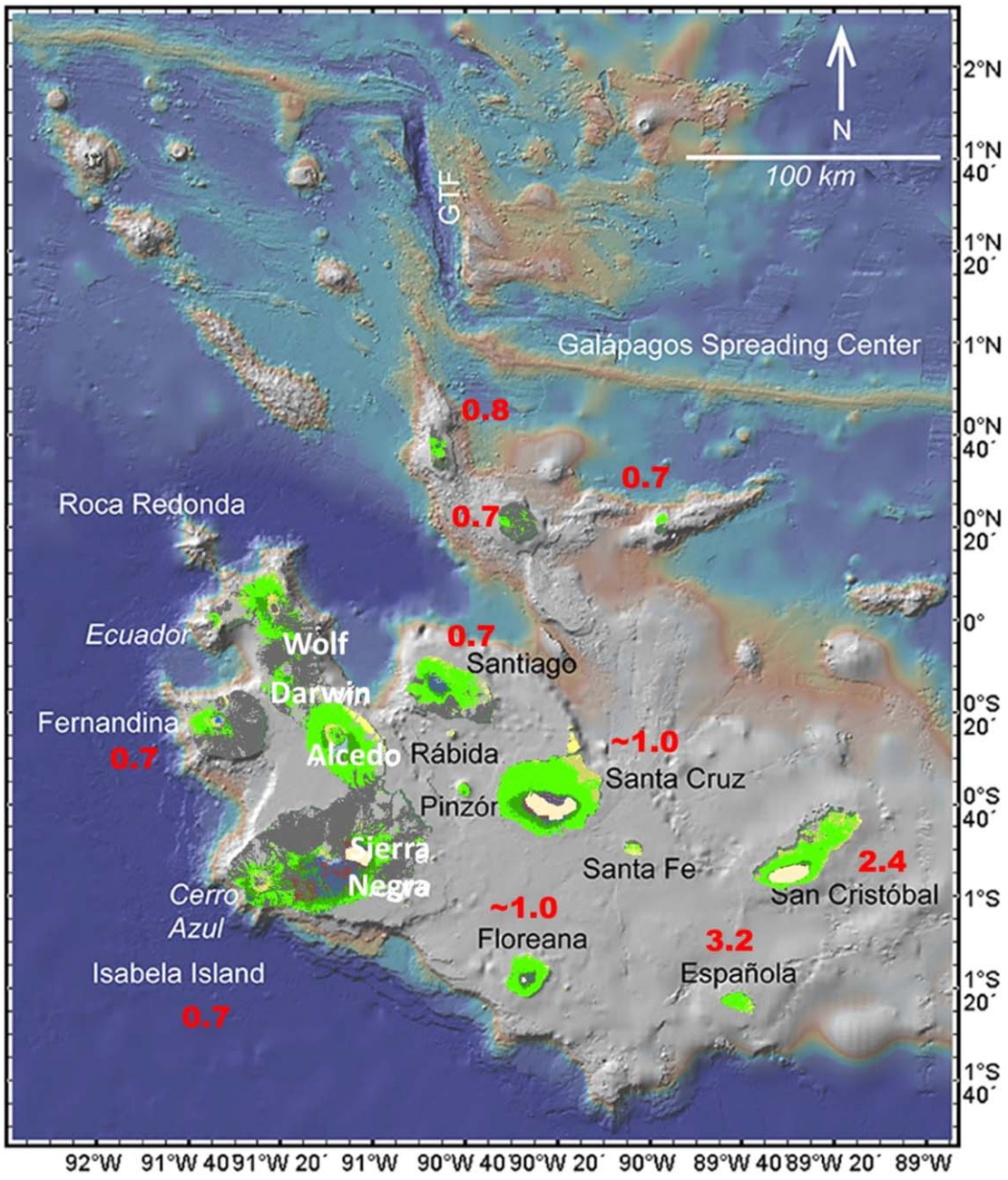

Figure A1-3. A combined map of the Galapagos Islands geology and land cover. The background map shows the Galapagos and surrounding bathymetry (figure 1 from Harpp *et al* (2018) [61]) which let us observe the Galapagos transform fault (GTF) north of the archipelago; Names and geological Ages (in million years) [19] of the islands are shown in black and red font, respectively, and volcanoes are shown in white. Land cover layer is presented as in Rivas-Torres *et al* (2018), figure 3 [2] with its respective legend.

Table A1-1. List of volcanic eruptions [42] for the archipelago.

Volcano name	Year of eruption	VEI
Fernandina	2018	1
Sierra Negra	2015	2
Fernandina	2009	2
Cerro Azul	2008	1
Fernandina	2005	2
Sierra Negra	2005	3
Wolf	2005	4

Table A1-2. List of dominant species associated with dry and evergreen forest ecosystem.

Vegetation zone	Common species
Dry Forest (lowlands, ~61% of vegetated area)	Species of cacti (e.g. *Brachycereus nesioticus* (lava cactus), *Opunita echios* (prickly-pear cactus), and *Jasminocerus thouarsii* (candelabra cactus)), as well as trees (*Scalesia heller* (daisy) and *Hippomane mancinella* (poison apple)) and shrubs (*Cordia lutea* (glue bush), *Gossypium darwinii* (Galapagos cotton), and *Lantana peduncularis* (Galapagos lantern)) [8].
Evergreen Forest and Shrubland (transitional zones and highlands, ~21% of vegetated area)	Natives: woody species such as *Scalesia pedunculata* (daisy tree) and *Miconia robinsoniana* (cacaotillo), as well as numerous ferns, epiphytes, mosses, and liverworts [5]. Invasives: *Psidium guajava* (guava), *Syzigium jambos* (rose apple), *Pennisetum purpureum* (Napier grass), *Lantana camara* (verbena), *Ricinus communis* (castor oil), *Cinchona pubescens* (red cinchona), *Rubus niveus* (raspberry sp.) and *Caesalpinia bonduc* (yellow nicker) [5].

A.2. Appendix A2
Calibration of the model

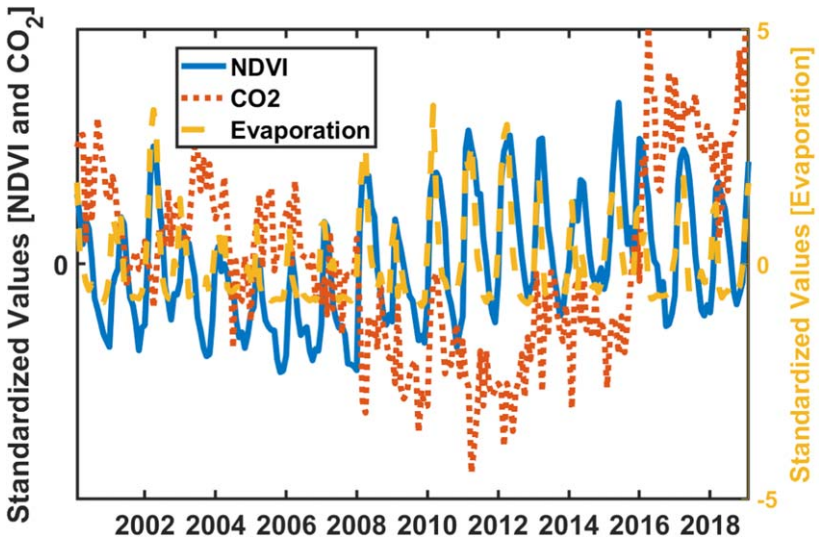

Figure A2-1. Standardized mean NDVI trend of the Galapagos Islands, Carbon dioxide, and Evaporation for period 2000–2019. Because the primary purpose of this analysis was to identify cyclical trends, all variables were standardized and linear trends (e.g. increasing CO_2 were removed prior to graphing).

Table A2-1. List of Normality tests [43] applied on residuals of the calibration set. The table showed the number of pixels (and it percentage of area) that follows normality ($\alpha = 0.05$) in the archipelago, the mean p-value and median p-value for the whole archipelago.

Test Name	Number of pixels that follow Normality (%)	p-value (mean)	p-value (median)
KS Limiting Form	216(58.70)	0.2023	0.0926
KS Marsaglia Method	65(17.66)	0.0333	0.0100
KS Lilliefors Modification	212(57.61)	0.1942	0.0865
KS Stephens Modification	65(17.66)	0.0332	0.0007
Cramer- Von Mises Test	37(10.05)	0.0305	0.0002
Anderson-Darling Test	49(13.32)	0.0394	0.0001
D'Agostino & Pearson Test	47(12.77)	0.0282	0.0002
Jarque-Bera Test	49(13.32)	0.0313	0.0004
Shapiro-Wilk Test	81(22.01)	0.0575	0.0002
Shapiro-Francia Test	69(18.75)	0.0482	0.0002

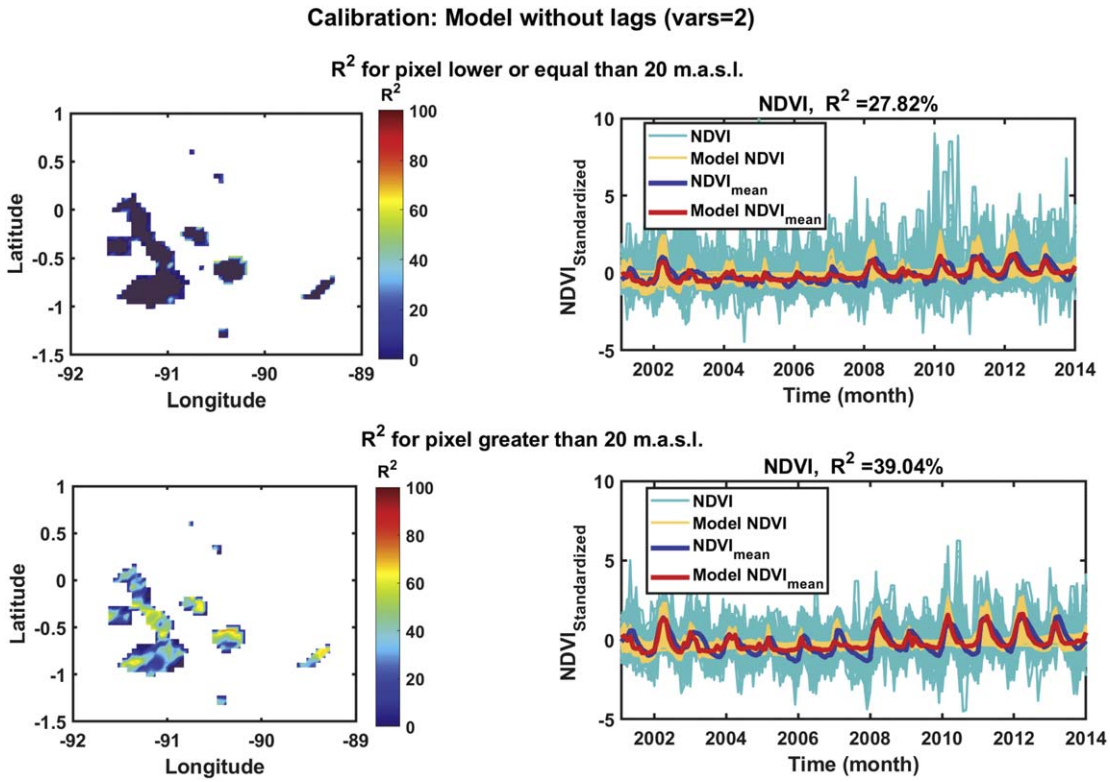

Figure A2-2. Calibration of the model without lags. On the left side, spatial correlation of the normalized difference vegetation index (NDVI) R^2 with respect to the generalized linear regression model for pixel with an elevation lower or equal than 20 MASL (top), and higher than 20 MASL (bottom). Pixels in black represent pixels outside of the area of analysis. On the right side, time series of the normalized difference vegetation index (NDVI)(light blue) and its average (blue) and the generalized linear regression model by pixel (mustard) and its average (red) with an elevation lower or equal than 20 MASL (top), and higher than 20 MASL (bottom).

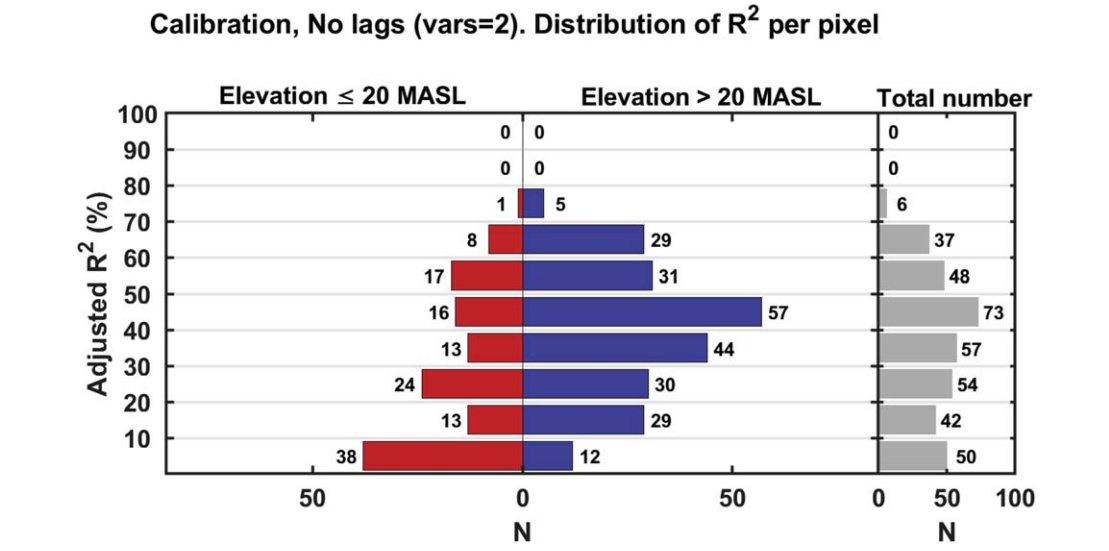

Figure A2-3. Calibration of the model without lags. Bar plot of R^2 for the whole archipelago, segmented by two elevation ranges. One segment shows pixels with an elevation lower or equal to 20 MASL in red, and the other segment shows pixels with an elevation higher than 20 MASL in blue. On the right side of the plot, the total number of pixels corresponding to each R^2% is presented in grey.

ORCID iDs

E Herrera Estrella https://orcid.org/0000-0003-0965-8661
A Stoeth https://orcid.org/0000-0002-5222-1759
N Y Krakauer https://orcid.org/0000-0002-4926-5427
N Devineni https://orcid.org/0000-0003-0311-8752

References

[1] UNESCO World Heritage Centre n.d.Last accessed 10 November 2018 Galàpagos Islands (https://whc.unesco.org/en/list/1)
[2] Rivas-Torres G F, Benitez F L, Rueda D, Sevilla C and Mena C F 2018 A methodology for mapping native and invasive vegetation coverage in archipelagos: an example from the Galapagos Islands *Prog. Phys. Geog.* **42** 83–111
[3] UNESCO TheWonders of the Iguana: Galápagos Islands 2018 Last accessed 10 December 2018 (http://www.unesco.org/archives/multimedia/?pg=33&s=films_details&id=871)
[4] Darwin C 1859 *On The Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life* (London: J. Murray)
[5] Trueman M, Hannah L and d'Ozouville N 2011 *Climate Change Vulnerability Assessment of the Galàpagos Islands* (*Chapter 2: Terrestrial Ecosystems in Galàpagos: Potential Responses to Climate Change*) ed I Larrea and G Di Carlo (USA: Conservation International and WWF)
[6] WorldWildlife Fund. Galàpagos Islands, off the coast of Ecuador 2019 Last accessed 20 May 2019 (https://www.worldwildlife.org/ecoregions/nt1307)
[7] Restrepo A, Colinvaux P, Bush M, Correa-Metrio A, Conroy J, Gardener M R, Jaramillo P, Steinitz-Kannan M and Overpeck J 2012 Impacts of climate variability and human colonization on the vegetation of the Galàpagos Islands ecology *Ecology* **93** 1853–66
[8] Hamann O 2004 Vegetation changes over three decades on Santa Fe Island, Galapagos, Ecuador *Nordic Journal of Botany* **23** 143–52
[9] Benitez F L, Mena C F and Zurita-Arthos L 2017 Urban land cover change in ecologically fragile environments: the case of the Galapagos Islands *Land* **7** 21
[10] Bastille-Rousseau G, Gibbs J P, Campbell K, Yackulic C B and Blake S 2017 Ecosystem implications of conserving endemic versus eradicating introduced large herbivores in the Galapagos Archipelago *Biological Conservation* **209** 1–10
[11] Wang B, Luo X, Yang Y-M, Sun W, Cane M A, Cai W, Yeh S-W and Liu J 2019 Historical change of El Niño properties sheds light on future changes of extreme El Niño *Proc. Natl Acad. Sci.* **116** 22512–7
[12] Di Carlo G, d'Ozouville N, Henderson S, de Koning F, Larrea I, Ortiz F, Pidgeon E, Spurrier L and Suàrez L 2011 *Climate Change Vulnerability Assessment of the Galàpagos Islands (ch. Introduction)* (USA: Conservation International and WWF)
[13] Roughgarden J, Running S W and Matson P A 1991 What does remote sensing do for ecology? *Ecology, Ecological society of America* **72** 1918–22
[14] Lawley V, Lewis M, Clarke K and Ostendorf B 2016 Site-based and remote sensing methods for monitoring indicators of vegetation condition: an Australian review *Ecol. Indic.* **60** 1273–83
[15] Blake S, Yackulic C B, Cabrera F, Tapia W, Gibbs J P, Kümmeth F and Wikelski M 2013 Vegetation dynamics drive segregation by body size in Galapagos tortoises migrating across altitudinal gradients *Journal of Animal Ecology* **82** 310–21
[16] Moity N, Delgado B and de Leon P S 2019 Mangroves in the Galapagos Islands: distribution and dynamics *PLoS One* **14** e0209313
[17] Nagasaki University About the Galapagos Islands [Galapagos Islands image database] 2007 Last accessed 12 January 2019
[18] Acharya K 2000 Life & Nature Paradise in peril. October 27-31
[19] White B and Burdick B Volcanic Galapagos: Formation of an Oceanic Archipelago University of Oregon Last accessed 20 October 2020 (https://darkwing.uoregon.edu/~drt/Research/Volcanic%20Galapagos/presentation.view@_id=9889959127044&_page=1&_part=3&.html)
[20] Esri 2019 The Galapagos Islands 1 °n–1.5 °s, 92 °w–89 °w [basemap: The world imagery. Scale not given] Last accessed 20 December 2019
[21] Esri 2019 The Ecuadorian territorial sea and land 2 °n–6 °s, 97 °w–75 °w [basemap: The world ocean basemap. Scale not given] Last accessed 20 December 2019
[22] Hamann O 1979 On climatic conditions, vegetation types, and leaf size in the Galapagos Islands *Biotropica* **11** 101–22
[23] Dos Santos Coelho C A and Ambrizzi T 1999 Climatological studies of the influences of El Niño Southern Oscillation events in the precipitation pattern over South America during austral summer *Sixth International Conference on Southern Hemisphere Meteorology and Oceanography (Santiago, Chile)*
[24] Tucker C J 1979 Red and photographic infrared linear combinations for monitoring vegetation *Remote Sens. Environ.* **8** 127–50
[25] The Earth Observatory Measuring vegetation (NDVI & EVI) 2000 Last accessed 12 March 2019 (https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php)
[26] Wardlow B D, Egbert S L and Kastensc J H 2007 Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. central great plains *Remote Sens. Environ.* **108** 290–310
[27] Javadnia E, Mobasheri M R and Kamali G A 2009 MODIS NDVI quality enhancement using ASTER images *Journal of Agricultural Science and Technology (JAST)* **11** 549–58
[28] Muskett R R and 2014 Arctic Diurnal Land-Surface Temperature Range Changes Derived by NASA MODIS-Terra and -Aqua 2000 through 2012 *Atmospheric and Climate Sciences* **4** 231–40
[29] Gelaro R *et al* 2017 The Modern-Era Retrospective Analysis for research and applications, version 2 (MERRA-2) *J. Clim.* **30** 5419–54
[30] Global Precipitation Measurement The Tropical Rainfall Measuring Mission (TRMM) 2019 Last accessed 8 April 2019 (https://gpm.nasa.gov/missions/trmm)
[31] JetPropulsion Laboratory ASTER Global Digital Elevation Map 2019 Last accessed 8 April 2019 (https://asterweb.jpl.nasa.gov/gdem.asp)
[32] NationalWeather Service ENSO Indices 2020 Last accessed 20 August 2020 (https://www.weather.gov/fwd/indices)
[33] GlobalMonitoring Laboratory Carbon Cycle Greenhouse Gases 2020 Last accessed 4 August 2020 (https://gml.noaa.gov/ccgg/trends/)
[34] NOAA 2019 *State of the Climate: Global Climate Report for June 2019* NOAA National Centers for Environmental Information published online July 2019, Last accessed 4 August 2020 (https://www.ncdc.noaa.gov/sotc/global/201906/supplemental/page-3)

[35] National Weather Service Climate Prediction Center Cold & Warm Episodes by Season Last accessed 16 September 2018 (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)
[36] Acker J G, Leptoukh G and 2007 Online Analysis Enhances Use of NASA Earth Science Data *EOS* **88** 14-17 https://giovanni.gsfc.nasa.gov/giovanni/
[37] Tibshirani R 1996 Regression shrinkage and selection via the lasso *Journal of the Royal Statistical Society Series* B **58** 267–88
[38] Hastie T, Tibshirani R and Friedman J 2001 *The Elements of Statistical Learning Data Mining, Inference, and Prediction* (*Series in Statistics*) 2nd (New York, NY: Springer) (https://doi.org/10.1007/978-0-387-84858-7)
[39] McCullagh P and Nelder J A 1989 *Generalized Linear Models* 2nd (London: Chapman and Hall)
[40] Holland P W and Welsch R E 1977 Robust regression using iteratively reweighted least-squares *Commun. Stat. - Theory Methods* **6** 813–27
[41] Amelung F, Jónsson S, Zebker H and Segall P 2000 Widespread uplift and 'trapdoor' faulting on Galàpagos volcanoes observed with radar interferometry *Nature* **407** 993–6
[42] Global Volcanism Program 2013 Volcanoes of the world, v. 4.9.0 (4 June 2020). Venzke, E (ed.). Smithsonian Institution. Last accessed 04 July 2019 (https://doi.org/10.5479/si.GVP.VOTW4-2013)
[43] Ipek 2020 Normality test package 2020 MATLAB Central File Exchange. Last accessed 10 November 2020 (https://www.mathworks.com/matlabcentral/fileexchange/60147-normality-test-package)
[44] Wu D, Zhao X, Liang S, Zhou T, Huang K, Tang B and Zhao W 2015 Time-lag effects of global vegetation responses to climate change *Global Change Biol.* **21** 3520–31
[45] Trueman M and d'Ozouville N 2010 Characterizing the Galapagos terrestrial climate in the face of global climate change *Galapagos Research* **67** 26–37
[46] Chen M *et al* 2019 Assessing precipitation, evapotranspiration, and NDVI as controls of U.S. great plains plant production *Ecosphere* **10** e02889
[47] Mo X, Liu S, Lin Z, Wang S and Hu S 2015 Trends in land surface evapotranspiration across China with remotely sensed NDVI and climatological data for 1981–2010 *Hydrol. Sci. J.* **60** 2163–77
[48] Szilagyi J, Rundquist D, Gosselin D and Parlange M 1998 NDVI relationship to monthly evaporation *GeophysicalResearch Letters* **25** 1753–6
[49] Krakauer N Y, Lakhankar T and Anadon J D 2017 Mapping and attributing normalized difference vegetation index trends for Nepal *Remote Sensing* **9** 986
[50] Sun J and Qin X 2016 Precipitation and temperature regulate the seasonal changes of NDVI across the tibetan plateau *Environmental Earth Science* **75** 291
[51] Taub D 2010 Effects of Rising Atmospheric Concentrations of Carbon Dioxide on Plants *Nature Education Knowledge* **3** 21
[52] Poorter H and Navas M 2013 Plant growth and competition at elevated CO_2: on winners, losers and functional groups *New Phytologist.* **157** 175–98
[53] Bazzaz F A 1979 The physiological ecology of plant succession *Annual Review of Ecology and Systematics* **10** 351–171
[54] Ainsworth E and Rogers A 2007 The response of photosynthesis and stomatal conductance to rising [co2]: mechanisms and environmental interactions *Plant, Cell & Environment* **30** 258–70
[55] Lovelock C E, Winter K, Mersits R and Popp M 1998 Responses of communities of tropical tree species to elevated CO_2 in a forest clearing *Oecologia* **116** 207-218
[56] Kalisa W, Igbawua T, Henchiri M, Ali S, Zhang S, Bai Y and Zhang J 2019 Assessment of climate impact on vegetation dynamics over East Africa from 1982 to 2015 *Sci. Rep.* **9**
[57] Aldàz I and Tye A 1999 Effects of the 1997-98 El Niño event on the vegetation of Alcedo volcano, Isabela Island *Noticias de Galapagos* **60** 25–8
[58] Watson J, Trueman M, Tufet M, Henderson S and Atkinson R 2010 Mapping terrestrial anthropogenic degradation on the inhabited islands of the Galapagos Archipelago *Oryz* **44** 79–82
[59] Zhang X, Tarpley D and Sullivan J T 2007 Diverse responses of vegetation phenology to a warming climate *Geophysical Research Letter* **34** L19405
[60] Winter K and Lovelock C E 1999 Growth responses of seedlings of early and late successional tropical forest trees to elevated atmospheric CO_2 *Flora* **194** 221–7
[61] Harpp K S and Geist D J 2018 The Evolution of Galápagos Volcanoes: An Alternative Perspective *Frontiers in Earth Science* **6** 1–50