

# Modelling the relationship between local biodiversity and remotely-sensed vegetation indices: the effect of spatio-temporal scale

Sean Tuck

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science of Imperial College London and Diploma of Imperial College London

September 2012

## Abstract

---

Biodiversity is being lost at an alarming rate and the intensity of environmental pressures behind the decline show no sign of abating. Many limitations of biodiversity indicators are shared: models relate the response of biodiversity components to mechanisms of environmental change, but are often data deficient and assign qualitative classifications to intensities of change. By utilising remotely-sensed MODIS Enhanced Vegetation Index (EVI), quantitative measures of habitat destruction intensity can be globally applied at local scales and related to local biodiversity. The spatio-temporal scale of EVI data must be large enough to capture environmental dynamics acting upon local biodiversity, but not so large as to include irrelevant dynamics: the optimal scale at which EVI can be related to local abundance is unclear. This study aims to: devise a method for MODIS remotely-sensed data to be retrieved and applied to statistical modelling, and; identify an optimal spatio-temporal scale of EVI for globally modelling local responses of abundance to environmental change. An R package – MODISTools – was developed to improve access and implementation of MODIS data to statistical modelling, and is available for download. MODISTools was used to retrieve EVI, which was related to local abundance using Generalized Linear Mixed Models (GLMMs) and Generalized Additive Mixed Models (GAMMs). Abundance data were a pooled collection from publications in the literature, producing abundances for many species in many locations: space was a surrogate for time, looking at the local relationship within each site, among many sites, building a global response. Random effects account for differences between studies, locations, and species. EVI data were spatially-weighted using a Gaussian kernel function. Effects of spatio-temporal scale were explored by fitting the model to data of varying scales and using model comparison via AIC and parameter estimates. EVI was considered within increasingly larger spatial extents to compare with smaller scales, and repeated for increasingly larger temporal extents, to find a scale that maximises model fit and provides meaningful parameter estimates for inference. Results show that the effects of scale are complex; AIC marginally decreases for increased spatial extents with little difference between temporal scales containing sufficient data. But spatial dependency in EVI parameters produce scales that are the best fit but poorly represent local dynamics. At low temporal scales AIC is unreliable but parameter estimates provide an increasingly detailed description of local dynamics with increasing spatial extent. Larger temporal scales facilitate more quality EVI data, allowing better discrimination between sites, but include dynamics irrelevant to the local response. Whereas, smaller temporal scales are data deficient in some locations, but where data is reliable, a more detailed response is elucidated and increasing spatial extent improves parameter estimates. In conclusion, there is a trade-off between model fit and a detailed description of local dynamics that must be met by balancing spatial and temporal scales. All models show high mean EVI has a positive effect on local abundance, but a nonlinear threshold effect is identified with locally scaled EVI.

---

# Contents

Introduction .....	4
MODISTools R Package .....	9
Method .....	11
Data Collection .....	11
Modelling Framework .....	15
Results .....	17
Discussion.....	25
Synthesis.....	25
Limitations & Further Work.....	27
Concluding Remarks .....	30
Acknowledgements.....	30
References.....	31
Appendices.....	40

## List of tables and figures:

Figure 1	8
Figure 2	8
Figure 3	12
Figure 4	15
Figure 5	17
Figure 6	18
Figure 7	19
Figure 8	19
Figure 9	20
Figure 10	22
Figure 11	23
Figure 12	24
Table 1	24

## Introduction

Biodiversity is being lost at a steady or accelerating rate and the intensity of environmental pressures behind the decline show no sign of abating (Butchart *et al.*, 2010; Mace & Baillie, 2007; Mace *et al.*, 2010). In the species-rich tropics there have been unparalleled rates of habitat destruction, degradation and fragmentation (Achard *et al.*, 2002). This translates into estimated extinction rates drastically greater than the background rate (Pimm *et al.*, 1995; CBD, 2010). To compound this trend, extinction debts lead to a time-lag in the realisation of species-loss as a result of current or past environmental pressures; the impacts of current pressures will continue to emerge in the future (Wearn *et al.*, 2012; Rangel *et al.*, 2012). Long-term studies are showing that the accumulation of diversity-dependent feedbacks cause high-diversity species combinations to become less functionally redundant over time (Reich *et al.*, 2012); this means short-term studies may have underestimated the negative impacts of species loss.

There is scope for intervention to reduce biodiversity loss through global policy (Pereira *et al.*, 2010). Targets were set by the 192 states and the EU that are party to the Convention on Biological Diversity (CBD, 2010) to reduce biodiversity loss by 2010. There has been interest from policy-makers in methods of quantifying biodiversity change to assess if targets have been met: current global biodiversity indicators suggest they have not (Alkemade *et al.*, 2009; Certain *et al.*, 2011; Collen *et al.*, 2009; Gibson *et al.*, 2011; Nichols *et al.*, 2007; Scholes & Biggs, 2005). The new 2011-2020 Strategic Plan adopts the Aichi Biodiversity Targets set out in Nagoya 2010 (CBD, 2011); with the upcoming Global Biodiversity Outlook 4 – the mid-plan assessment – and the emergence of GEO-BON (Scholes *et al.*, 2008), there is a continual stream of opportunity to better inform policy.

In response to CBD and the World Summit on Sustainable Development (WSSD) GLOBIO3 was developed, combining some Global Environment Outlook 3 and Millennium Ecosystem Assessment (MA) approaches (MA, 2005; UNEP, 2001, 2002). GLOBIO3 (Alkemade *et al.*, 2009) is a biodiversity model designed to assess the intensity of impacts on biodiversity as a result of human activity, and make projections on future impacts based on given activity scenarios. GLOBIO3 is a meta-analysis of publications that establishes cause-effect relationships between environmental drivers and biodiversity impacts – measured in Mean Species Abundance (MSA) and defined as change in MSA relative to an undisturbed

equivalent ecosystem. The MSA metric used is similar to the Biodiversity Integrity Index (Majer & Beeston, 1996) and Biodiversity Intactness Index (Scholes & Biggs, 2005), the main difference being spatial weighting, where MSA gives equal weight to all spatial units and Biodiversity Intactness Index (BII) gives more weight to species rich areas. Other methods, such as the Living Planet Index (Collen *et al.*, 2009; Loh *et al.*, 2005) also use a reference to measure relative biodiversity loss, but unlike GLOBIO3 the Living Planet Index (LPI) uses a baseline of conditions at 1970 (Collen *et al.*, 2009), making fewer assumptions of reference points being pristine habitats.

Biodiversity can be measured in different ways due to its complex multidimensionality, encompassing many components, such as composition, structure and function all at many spatial scales. As a result, there can be no all-inclusive single metric for biodiversity (Magurran, 2004; Purvis & Hector, 2000). Species richness – the most commonly used measure of biodiversity – does not alone provide enough information for diversity estimates (Lyashevskaya & Farnsworth, 2012; Wisley *et al.*, 2005); functional, structural and other taxonomic levels uncorrelated with species richness provide independent information (Gallardo *et al.*, 2011; Cadotte *et al.*, 2010). Many of the limitations of biodiversity indicator approaches are shared. Due to data requirements of methodologies, the measures are often restricted in their scope of studies. By averaging among species and giving no method of weighting, individual responses of species, or ecologically similar taxa, are lost. However, it is well established that taxa do not respond equally to threat, in terms of timescales and intensity of response (Mace *et al.*, 2003; Isaac & Cowlshaw, 2004). GLOBIO3, unlike other approaches, makes no mention of calculating variances around MSA estimates, meaning less is known about the uncertainty inherent in estimates; this has not been included in implementations of GLOBIO3 at a later date (Alkemade *et al.*, 2011; Alkemade *et al.*, 2012).

With some exceptions, biodiversity indicators use Generalized Linear Mixed Models (GLMMs) to model the response of some diversity metric or effect size to the chosen explanatory variables. GLMMs have emerged in recent years as a powerful statistical tool, providing greater flexibility than classical procedures when dealing with non-normal data with complex structures, as is often the case in ecology (Bolker *et al.*, 2009; Zuur *et al.*, 2009). This framework better deals with count data, such as abundances, and allows the indicators to account for idiosyncratic variations between publications, locations and

species, focusing on the response to changes in the environmental variables of interest. Generalized Additive Mixed Models (GAMMs) relax the assumption of linearity by fitting non-parametric smoothing parameters to fixed effects, allowing any non-linear responses such as threshold effects to emerge from the data, reducing to a linear form if the data permit (Wood, 2006).

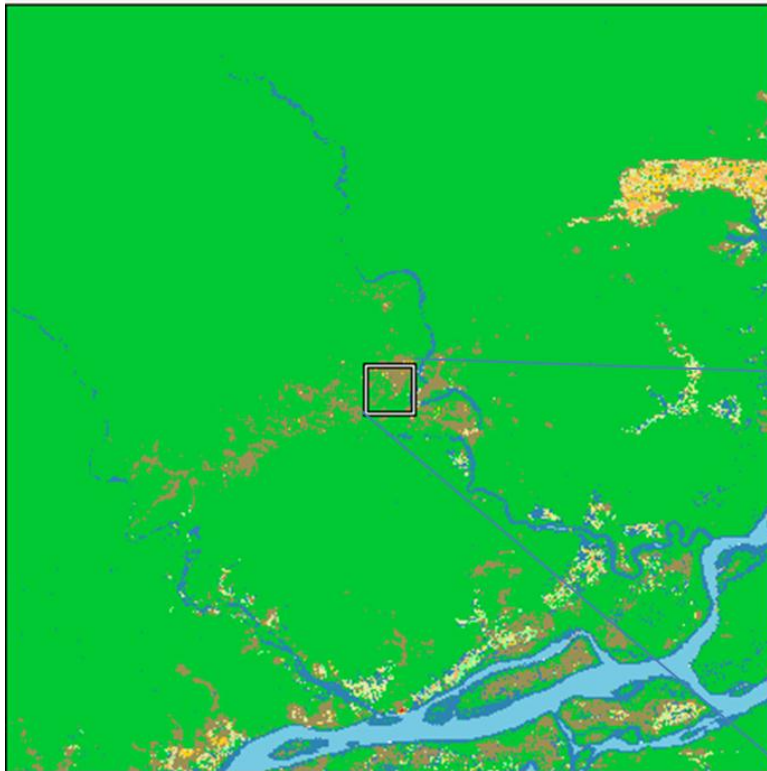
In order to effectively estimate biodiversity loss, there are things we need to be true of our models and data: to incorporate as much data as possible through few restrictions and data requirements; to incorporate the given data in a way that is representative of biodiversity, reducing effects of sampling bias taxonomically and geographically; to model the effects of sources of environmental change in simple yet not overly-simplistic ways; to make estimates applicable to many spatial scales so that it can be disaggregated to estimate for different components or resolutions; to quantify degrees of uncertainty in estimates; to incorporate as many ecological measures as possible to capture the different aspects of biodiversity. Some of these requirements mirror identified areas for progression in macroecology (Beck *et al.*, 2012). Meeting data requirements are key to macroecological development: not only do datasets tailored for macroecological questions need to be developed, but also the inclusion of past data so that analyses over larger temporal scales can be implemented. Further consideration of local processes is important to understand variability at different scales. The application of remotely-sensed data can help meet these aims for development.

Remote-sensing provides large amounts of regularly collected quantitative data that can be used to investigate a wide variety of questions in versatile ways. The data are often collected globally, and at increasingly finer resolutions, facilitating questions that need to be asked at different spatial scales. Application of remote-sensing in ecology, conservation biology and global change research is being increasingly recognised (Bai *et al.*, 2008; Chawla *et al.*, 2012; Donoghue, 2002; Justice *et al.*, 1998; Kerr & Ostrovsky, 2003; Naeem *et al.*, 2012). Many datasets are in their infancy, but represent an opportunity for long-term data collection on time-scales otherwise difficult to achieve in ecology. One such resource providing useful data for ecological analyses is the freely-available Land Processes Distributed Active Archive Center (LP DAAC). LP DAAC, one of the discipline-specific data centres within the NASA Earth Observing System Data and Information System (EOSDIS), contains land observations collected by NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) instrument

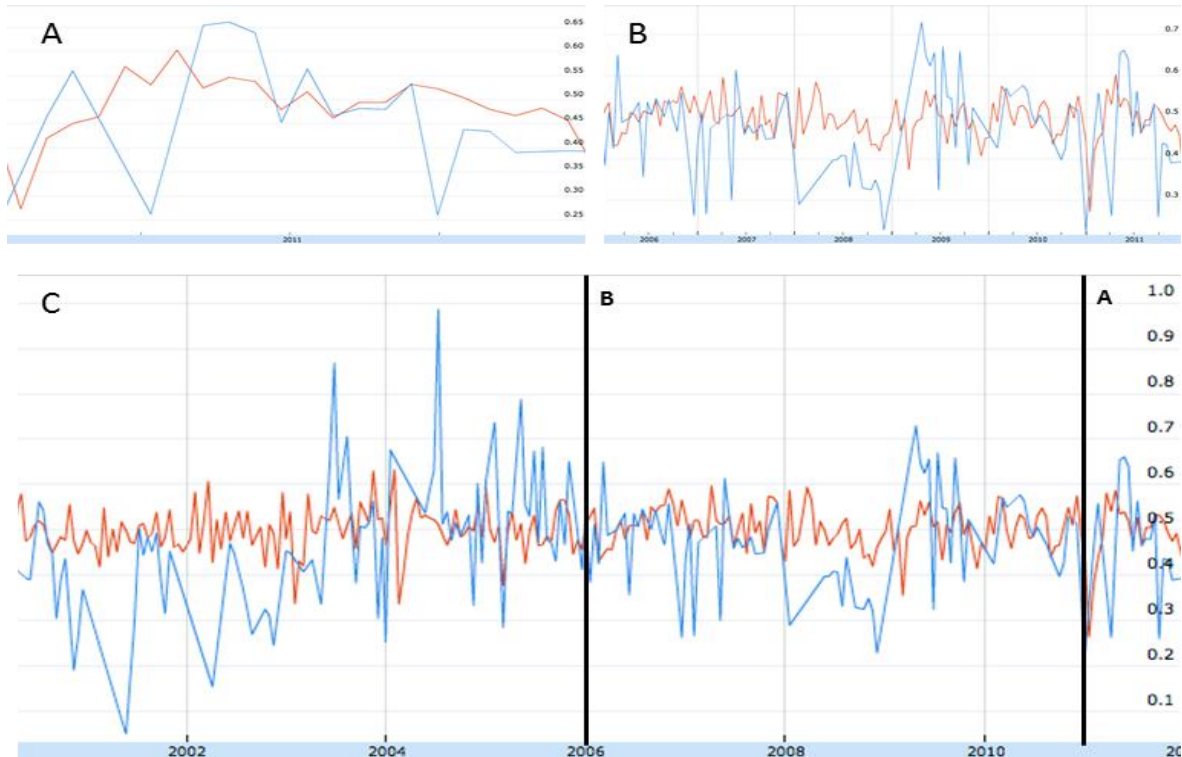
(NASA, 2012). The MODIS archive encompasses a wide range of data for many applications, including vegetation indices that provide consistent, regular, and quantifiable comparisons on vegetation conditions.

The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) monitor global vegetation conditions and can be used in modelling climatic, biogeochemical and hydrologic processes, or land surface biophysical processes, such as primary production or land cover conversion (Justice *et al.*, 1998, 2002; Running *et al.*, 2004; Sellers *et al.*, 1994; Zhang *et al.*, 2003). The use of NDVI in ecological studies, particularly on environmental change in an ecosystem context, is now well documented (Gao *et al.*, 2012; Gould, 2000; Pettorelli *et al.*, 2005). EVI presents some modifications of the older NDVI, providing the same use but with more reliable data. These vegetation indices can be used to represent environmental change in a whole range of contexts, including habitat destruction, degradation and fragmentation (Holm *et al.*, 2003; Thiam, 2003), and have been used in correlative studies to identify declines in population abundances (Carey *et al.*, 2001) or species richness (Bar-Massada *et al.*, 2012). Remotely-sensed data, such as the MODIS archive, are becoming increasingly available and user-friendly. But, the potential for widespread application in fields such as macroecology or conservation biology may be realised further by providing simpler 'canned' methods for data collection, handling, and implementation in modelling.

With data on both diversity and environmental change, the responses of diversity to environmental change can be elucidated. The question remains, however, at what spatio-temporal scale should environmental data be collected in order to effectively capture the dynamics of the environmental change acting upon diversity? Diversity data – such as abundance – are often collected at very small-scaled locations, via traps, quadrats or similar methods. But these methods infer upon abundance in a wider spatial context and clearly the sampling level is not a sensible scale to collect environmental data. So what is? The spatial extent of environmental data must be large enough to sensibly capture information on the changes which are of interest. But, not so large as to include dynamics that are unrelated to the locations with respect to diversity data, blurring out the environmental picture that is important for abundance at the location of interest (Figure 1). It is not obvious which spatial extent has the optimum explanatory power for modelling local biodiversity responses to



**Figure 1.** The environment picked up depends on the spatial scale at which MODIS data is extracted. Colours represent different land cover classes. By taking the centre pixel only nearby land cover clustered around the location, which may be relevant, are not considered. But extending the spatial scale too far may include land cover, and dynamics, that are unrelated to the location of interest.



**Figure 2.** A time-series of EVI fluctuations in a location (centre pixel in blue, an average of all pixels within 5km<sup>2</sup> in red). A shows one year of EVI data (2011-2012), B shows 6 years (2006-2012), and C shows 12 years (2000-2012). Seasonality and stochastic differences between years may mean A does not contain enough information to optimally describe the environmental conditions in an area. However, extending beyond B seems to include much larger fluctuations, possibly indicative of previous dynamics that are unrelated to the abundance found at that location in 2012. The averaged pixel time-series is much more stable than the centre pixel, suggesting space and time have complementary effects on providing a clear picture of EVI dynamics in a given location.

Images courtesy of Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC, 2012).



environmental change, or how this varies for different habitat types or biomes, or when related to different taxa.

Similarly, how can the temporal characteristics of the environmental change be most optimally represented? A time-series of environmental data can be collected for a given length prior to the date of abundance collection to better represent the dynamics that have led to the abundances recorded in a given location. There, a balance must be found between accounting for seasonal and inter-annual variations, and including time-series data that are not relevant in determining the abundances recorded at the time of collection (Figure 2).

In light of this, the following questions are presented: Is just the focal pixel overlaying the known coordinate location for diversity data enough? Or, is a wider spatial extent around the location needed to optimally explain how abundance responds to change in EVI? What temporal scale should a mean EVI value be taken over to optimally represent pertinent EVI dynamics? The study aims are: to devise a method for MODIS remotely-sensed data to be easily retrieved and applied to statistical modelling of scientific questions – ecological and beyond, and; identify an optimal spatial and temporal scale at which this data can be used in a global model of local responses of abundance to mechanisms of environmental change.

## **MODISTools R Package**

MODIS environmental data can be retrieved from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), available online. Current available tools online allow for a subset to be requested for a given location, specifying the subset description in terms of product type (e.g. Vegetation Indices – MOD13Q1), size of surrounding spatial extent (centred on the location), and time-series length each time. By writing a tool to automate this process for a batch of locations researcher effort, and most importantly opportunity for human error in the data collection process, is greatly reduced. This tool would also be applicable to the extraction of any MODIS subsets: given the broad interest that MODIS data attracts, this tool would be useful to a wide-ranging scope of research, not only to this study.

As such, a software package for R Statistical Computing Language and Environment – MODISTools – was created for use in this study and beyond, housing a set of tools useful for MODIS archive extraction and application that can be distributed to MODIS users

(MODISTools pdf manual and R code for some key functions can be found in the appendices). MODISTools uses the GNU General Public License (GPL-2) and is available for download from a github repository (Tuck, 2012). The R package has been checked and installed on linux, mac and windows.

MODISTools uses an SOAP web service – processing the web services description language and setting up an SOAP-client interface in R – to interrogate the MODIS database remotely. Various functions within the package return the requested subsets of different forms into R and downloaded to a local server. For organisation and analysis practicality, a file is written for each location specified, containing all the requested information – all data band values for a given product at the pixel level for all requested dates. R, being a software primarily for statistical data analysis, is a reasonable environment to apply this SOAP client to, because many researchers who may wish to use MODIS data are not programmers and whose programming experience may be restricted to R. MODISTools imports the packages SSOAP and XMLSchema (Temple Lang, 2012a, 2012b) for web service functionality.

One of the most important functions is MODISSubsets, which allows the user to request subsets from a given MODIS product for a specified time-series of a given area surrounding many locations at a time, greatly reducing time, effort, and human error. These locations are described by a lat-long coordinate in WGS-1984 coordinate system. Alternatively MODISTransects expands upon MODISSubsets by extracting MODIS data along a transect and a specified surrounding neighbourhood. Downloaded subsets are saved in ascii files along with a subset download log that lists all the unique time-series downloaded and their download status; any error and warning messages encountered during downloading are caught and can be traced to the problematic time-series.

The subsets can then be accessed by MODISSummaries for computing summary statistics and data processing. The processed data are also organised back with the original input data into one csv file that can be easily used for modelling; this provides efficient storage of data and a transparent process from data collection to processing into a form that is ready for final use. MODISTools also includes other minor functions that facilitate its main purpose, including: a lat-long coordinate conversion tool (ConvertToDD) for converting degrees, minutes seconds, or degrees, decimal minutes coordinates into the decimal degrees format,

and; functions for listing product types and data bands available through the MODIS Web Service.

## Method

### Data Collection

The explanatory variables used to model how biodiversity responds to human impacts were retrieved from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Terra (EOS AM) and Aqua (EOS PM) satellites. Vegetation indices were used as continuous proxies for habitat destruction and degradation, allowing modelling of biodiversity responses to the intensity of habitat destruction. EVI is preferred here to NDVI as it improves upon detection of changes in canopy structure, such as canopy type and leaf area index (Gao *et al.*, 2000) and is less saturating in high-biomass regions such as the tropics (Huete *et al.*, 2002). It also reduces noise from atmospheric effects like aerosol conditions (Xiao *et al.*, 2003). These vegetation indices are direct transformations of calculated surface bidirectional reflectance values ( $\rho$ ) from several spectral bands, making no assumptions of land cover type, soil type or climate. EVI takes the form:

$$EVI = G \frac{\rho_{NIR} - \rho_{red}}{(\rho_{NIR} + C_1 \cdot \rho_{red}) - (C_2 \cdot \rho_{blue} + L)}$$

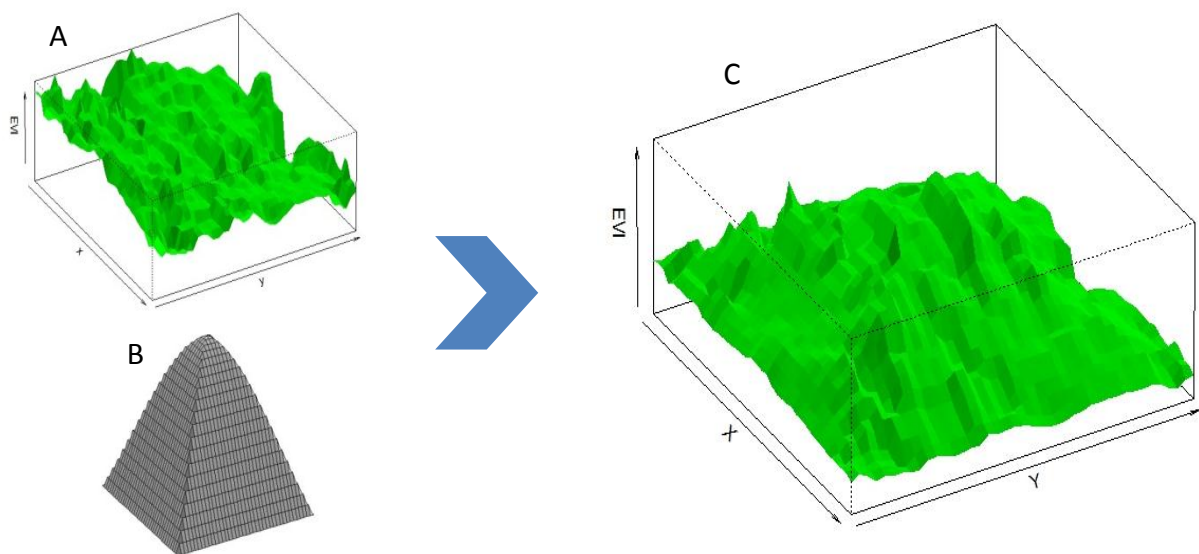
where G is a gain factor that limits EVI values to a fixed range,  $\rho$  are atmospherically corrected reflectance values for near-infrared (NIR), red and blue spectral bands respectively, L is an adjustment for NIR and red radiant transfer, for canopy-related improvements on NDVI, and  $C_1$  and  $C_2$  are coefficients dealing with atmospheric noise, using the blue band to account for aerosol effects on the red band. The MODIS algorithm adopts the following coefficient values:  $G=2.5$ ;  $L=1$ ;  $C_1=6$ ;  $C_2=7.5$ . The valid range of values for EVI data is between -2000 and 10000, which are then multiplied by a scale factor of 0.0001.

These data can be globally downloaded from MODIS at a pixel resolution of 250x250m, for 16 day intervals, from 2000 to present. The MODISSubsets function was used to retrieve EVI and a measure of data reliability at each pixel from the MOD13Q1 product for 197 locations where diversity data have been attained. The data were downloaded for 10km<sup>2</sup> spatial extents, centred on the site location, with several time-series lengths prior to the date the specific abundance data were recorded.

Any missing values or with a reliability score  $> 0$  (which, non-intuitively, indicate problems with the reliability of EVI scores), were removed from each time-series and replaced by linearly interpolating between reliable data. A mean EVI value across the whole time-series was then calculated for each pixel.

Mean EVI values at each pixel were aggregated into distance bands, averaging all values that were the same distance from the centre pixel as each other. This was done for modelling simplicity and assumes that the underlying habitat characteristics and corresponding spatial relationships are uniformly distributed in accordance with complete spatial randomness. Therefore, there is an assumption of no directional bias in the distance-decay of relatedness to the centre pixel, and a spatially stationary response to EVI. Mean EVI values for each distance band were then spatially weighted, using a distance-decay function to approximate how the importance of an EVI value attenuates with distance from the centre pixel.

**Figure 3.** A surface of EVI values for a tile of pixels at a given location centred on a coordinate of abundance data (A) is transformed by a Gaussian weighting kernel (B), which weights each pixel EVI value by its distance from the centre of the tile, changing the topography of EVI (C) so that values close to the location of the abundance data have a greater influence than those further away.



A Gaussian weighting kernel function of the following form was used:

$$\omega_n = e^{\left(-\frac{n^2}{b^2}\right)}$$

Where  $\omega_n$  is the weighting term for an observation at distance band  $n$  from the centre pixel, and  $b$  is the basal width of the kernel function (the outermost distance band).  $\omega$  decreases according to a Gaussian curve as distance from the centre pixel increases. The centre pixel is weighted at 1. All weighted EVI distance band values were then summed, to give a cumulative tile EVI value:

$$EVI = \sum_{n=0}^b EVI_n \cdot e^{\left(-\frac{n^2}{b^2}\right)}$$

This means that the equivalent EVI change in two different distance bands have differing impacts on the total tile EVI value, as defined by this Gaussian weighting kernel.

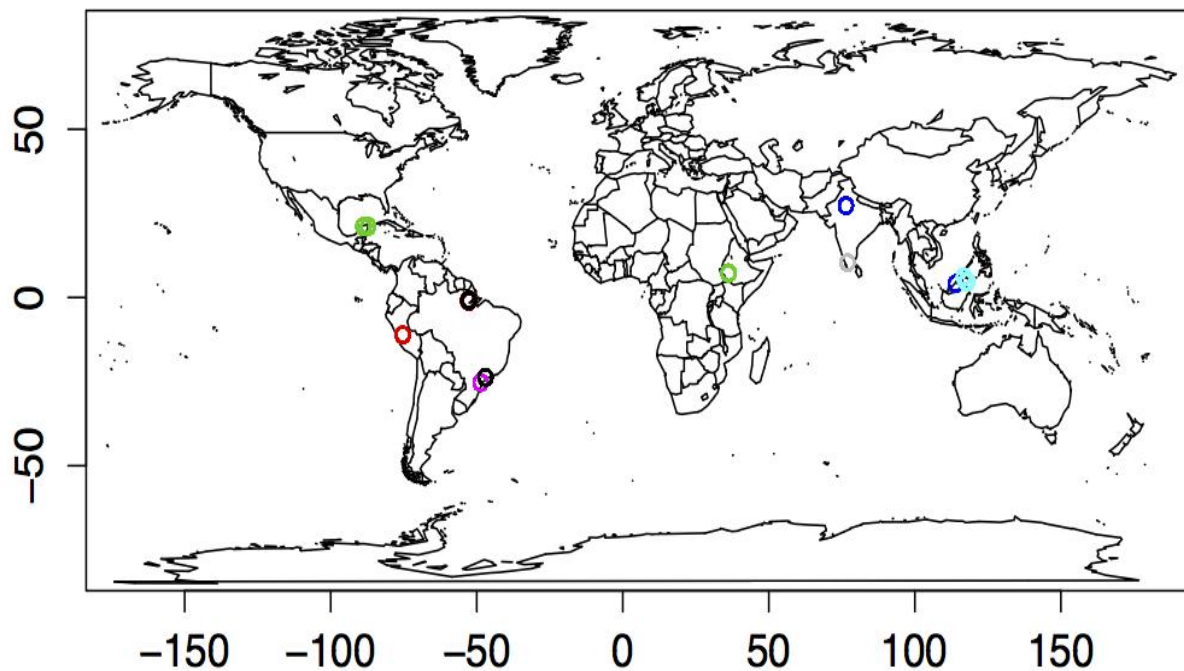
The diversity data, provided by Helen Phillips, was collected from relevant publications between 2000 and present; the time period for which MODIS data is recorded. The decision tree for selecting relevant papers (appendices) screens papers for data collected in relevant times that consider the human impact on named taxa at multiple sites, giving useful biodiversity measures as a result. Species abundances were used as a measure for biodiversity response, as opposed to observed species-level changes such as species richness. As population decline occurs on shorter timescales to species-level extinction, and provides more information at different spatial scales, it may be more responsive to changes in threat intensity and be a more sensitive indicator of biodiversity loss on the short time-scales dealt with (Ceballos & Ehrlich, 2002; Balmford *et al.*, 2003).

The data to look at how biodiversity responds to environmental change over time simply does not exist. Instead, space is used as a surrogate for time; looking at the local relationship between abundance and environmental conditions within each site, among many sites, to ascertain a global response of abundance to environmental change. The data were collected from as many tropical locations as possible, incorporating as many taxa as possible, including

plants, invertebrates, mammals, and birds. Within each major taxon many species were recorded, spanning different groups that perform distinct ecosystem functions. There are few requirements for data based on habitat type: as four different habitats and land-covers are included, with no requirement for 'pristine' conditions based on an arbitrary baseline. By modelling the response locally within sites – allowing the variation between studies and habitat types to be treated as a random effect – a wider scope of locations and their inherent habitats/land-covers can be incorporated.

A data structure within publications was established: data collected using distinctly different methodologies or at different times are considered different studies, and within studies, data collected is divided into sites. Sites are defined as subsets of data that are either spatially explicit – being at least 1km apart – or drawn from methodologically distinct collections. Site is the level within which the relationship is modelled. In total, data was collected from 14 publications, divided into 20 studies. From these, a total of 197 sites are identified. Within each site are abundance values for each species recorded, giving 21,611 records of abundance for all species within the 197 locations. This dataset, which was provided for this study, contains a fantastically large amount of abundance data for a wide taxonomic variety, including under-represented taxa (e.g. invertebrates and plants) and for the tropical biome – an under-represented region – making this a high-quality dataset for such analysis that can facilitate parameter intensive modelling.

Location data received from the publication authors were in various formats, projections and coordinate systems. These had to be converted into WGS-1984 (Transverse Mercator projection) decimal degrees format, both for consistency and so coordinates were compatible with MODISTools functions: only WGS-1984 decimal degrees could be processed by the MODIS web service. To do this, degrees coordinates of different formats were converted to decimal degrees using ConvertToDD included in MODISTools, and locations in other coordinate/projection systems were re-projected in WGS-1984 using ArcGIS.



**Figure 4.** Map showing the distribution of abundance data collected. Data points are colour coded at the publication level (14). Publications are relatively evenly distributed across tropical regions, well representing the whole tropical biome. The map highlights Africa as the most data deficient continent, particularly the West African tropical forests.

### Modelling Framework

The relationship between biodiversity and intensity of mechanisms for environmental change was tested by fitting several models of increasing complexity to the data.

Generalized Linear Mixed Models (GLMMs) were fitted to the data, with the environmental data as fixed-effects. The error structure specification was explored by looking at residual diagnostics from model outputs.

The structure of the data was accounted for by fitting random effects: site nested within study crossed with species as grouping structures. There are unexplainable variations in the responses of different species that must be accounted for, and variations as a result of location and differences in study methodologies. This random effects specification disregards the variation among taxa and among studies, looking at the local responses within the context of each site and taxon. The random effects specification was then optimised by model comparison using Likelihood Ratio Tests (LRT). If overdispersion is identified and accounted for, the inclusion of fixed-effects can be justified using AIC

comparison (Richards, 2008) and testing of parameter estimates using Wald  $\chi^2$  (Bolker *et al.*, 2009).

Additional variation may be explained by including random slopes with respect to EVI; a random intercept and slope at the species-level, and an uncorrelated random intercept at the site-level and slope at the study-level were all tested. This specification not only allows species and locations to have different intercepts, but also for abundances of different species or in different sites to respond differently to mean EVI. The uncorrelated slope and intercept for locations assumes that all sites within a study – as they are in close proximity with each other – show the same response to EVI, but that the intercept may vary: slopes are allowed to vary at the study level.

Generalized Additive Mixed Models (GAMMs) were then fitted to the data to test for significant nonlinearity. The inclusion of a non-parametric smoothing parameter relaxes the assumption of linearity in the relationship and allows curvature to emerge from the data. This may be important in identifying nonlinear behaviour in the relationship, such as threshold effects in the impact of threat intensity on biodiversity.

Overdispersion and zero-inflation were explored in the response data and attempts were made to account for them. Some zero-inflated poisson distribution is likely to be the best fit, but is often difficult to achieve with additive mixed-models that have complex random-effects structures, exemplified by the low level of support in R. A logarithmic transformation may deal with overdispersion, but there is little reason to expect the error structure be multiplicative. Also, with zeros in the data a log-transformation can only be applied to adjusted data – most commonly  $\log(\text{data}+1)$ . The level of adjustment made can bias the model fit (O'Hara & Kotze, 2010).

Overdispersion was tested for by computing:  $D = (\text{Observed } \sigma^2) / (\text{Theoretical } \sigma^2) \times (n - 1)$ , where the theoretical variance is of data drawn from a poisson process, and  $n$  is the number of observations – overdispersion can be tested for. If  $D$  is approximately  $\chi^2$  distributed with  $n-1$  degrees of freedom, then a poisson distribution would be a good fit to the data. It can be accounted for using a negative binomial (poisson-gamma) or a poisson-lognormal model (Breslow, 1990). Including a normally-distributed, observation-level random-effect in a model with a poisson error structure and log link effectively specifies a poisson-lognormal

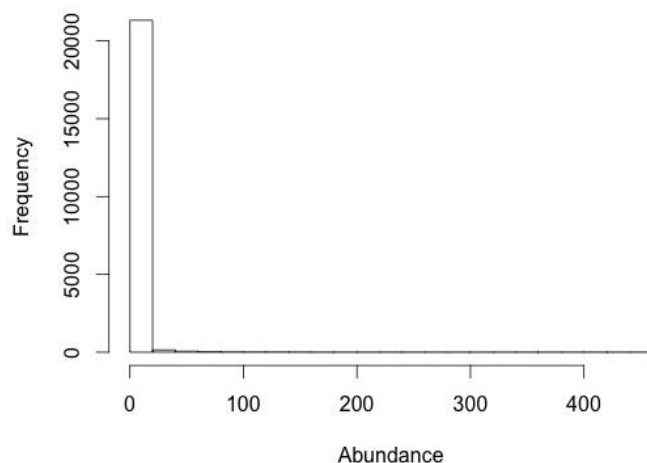


mixture distribution (Elston *et al.*, 2001; Breslow, 1990; Lawless, 1987; Hinde, 1982). This poisson-lognormal model is computationally intensive but can be estimated using Gauss-Hermite Quadrature Approximation (Hinde, 1982).

The question of spatio-temporal extents was approached by fitting the minimal adequate model to the data of varying spatial and temporal extents and using model comparison via AIC and parameter estimates. To begin, mean EVI was considered in terms of only the centre pixel overlaying the abundance coordinate location derived from time-series of three years for each pixel; the spatial extent of environmental data was then increased to compare with results that use smaller extents. By successively increasing the spatial extent of EVI by one distance band of pixels and assessing model fit each time, an optimum spatial extent can be identified. This process was then carried out for mean EVI calculated from time-series of increasingly longer lengths, finding the optimal environmental scale that maximises model likelihood estimates and provides the best confidence intervals around the parameter estimates.

## Results

Figure 5 shows the distribution of abundance is highly skewed. The overriding characteristic of the data is zero-inflation. The data appear to conform to a zero-inflated poisson, with high overdispersion. Looking at mean=1.4, and variance=99.2, we can see there is high overdispersion ( $\chi^2$ , d.f. = 21610:  $p < 0.001$ ).



**Figure 5.** Histogram of abundance. The high frequency of zeros shows the data conform to a zero-inflated poisson distribution. High overdispersion is also an important characteristic in the data.

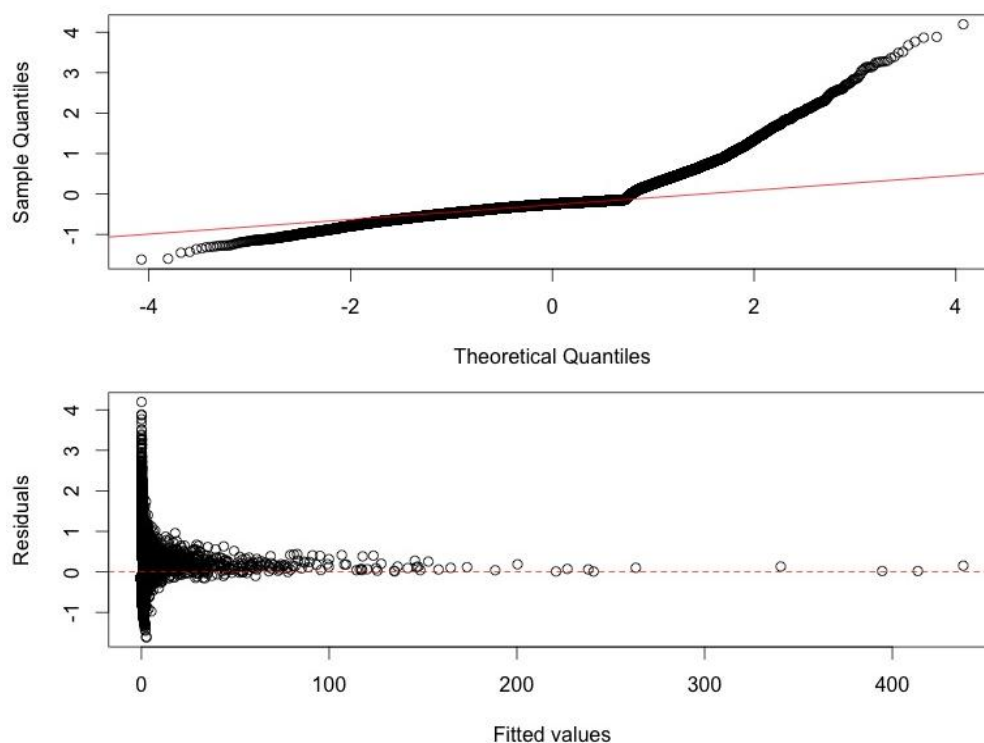
Species grouping structure is most the important random-effect – when specified as random intercepts – explaining 58% of the variation in a null model. The random slopes significantly improve model fit (LRT, d.f. = 2:  $p < 1.379e^{-15}$ ), so they are kept in the model. Again, the

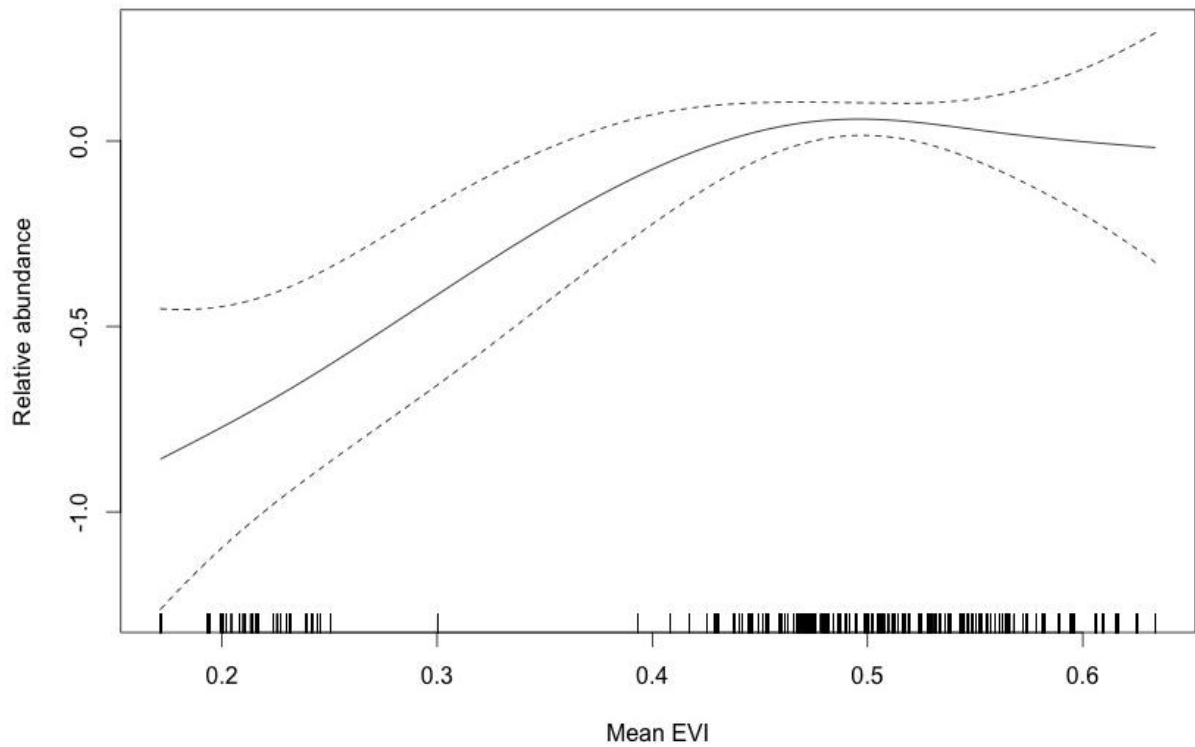
additional complexity of an observation-level random-effect is justified and a poisson-lognormal model is a much better fit to the data than a poisson (LRT, d.f. = 1:  $p < 2.2e^{-16}$ ). Two categorical fixed-effects – one that defines habitat type and one that defines the intensity of land-use in the area – are introduced individually and then together, to show the  $\Delta AIC$  that each variable causes, and then combined. The tests show that inclusion of habitat type significantly improves the models explanatory power, but that land-use intensity does not.

The minimal adequate model therefore considers the response of abundance to intensity of environmental pressures by including mean EVI and habitat type as fixed effects, with site nested within study, crossed with species grouping structures of random slope and intercept models, and a poisson-lognormal error structure. Residual diagnostics (figure 6) of the minimal adequate model show some non-normality in errors and heteroscedasticity around low fitted values: these issues are most likely caused by the zero-inflated data that is not easily accounted for.

---

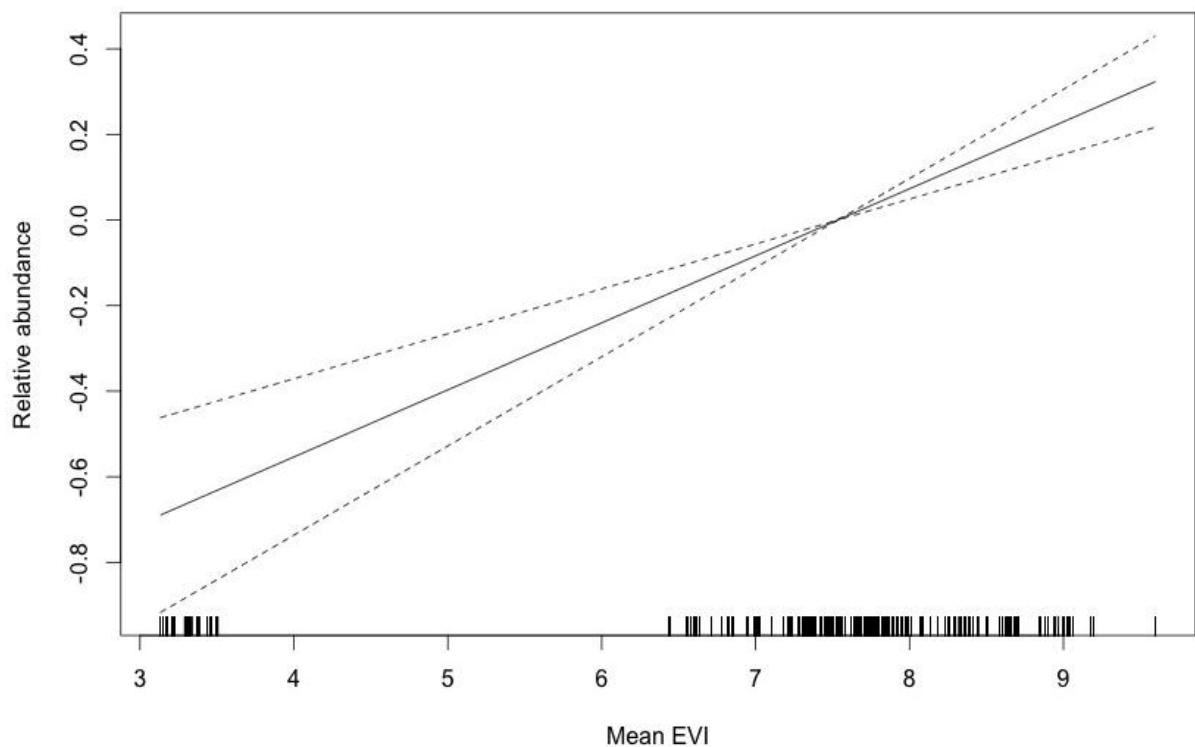
**Figure 6.** Model residuals diagnostics. Top: the Q-Q Plot shows some deviation from normality, suggesting the assumption of normally-distributed errors is being violated. The non-normal behaviour is exclusive to the high end of the range. This could be due to the zero-inflation: possibly this change in behaviour indicates the additional zeros arise from a separate process, indicating a two-step hurdle model could be appropriate. Bottom: a plot of residuals vs fitted values clearly shows the impact of the zero-inflated data. There is heterogeneity in the data, arising almost entirely from high variance in the zeros. Zero-inflation is clearly an important component in the data, but is difficult to specify in models complex as those used in this study.

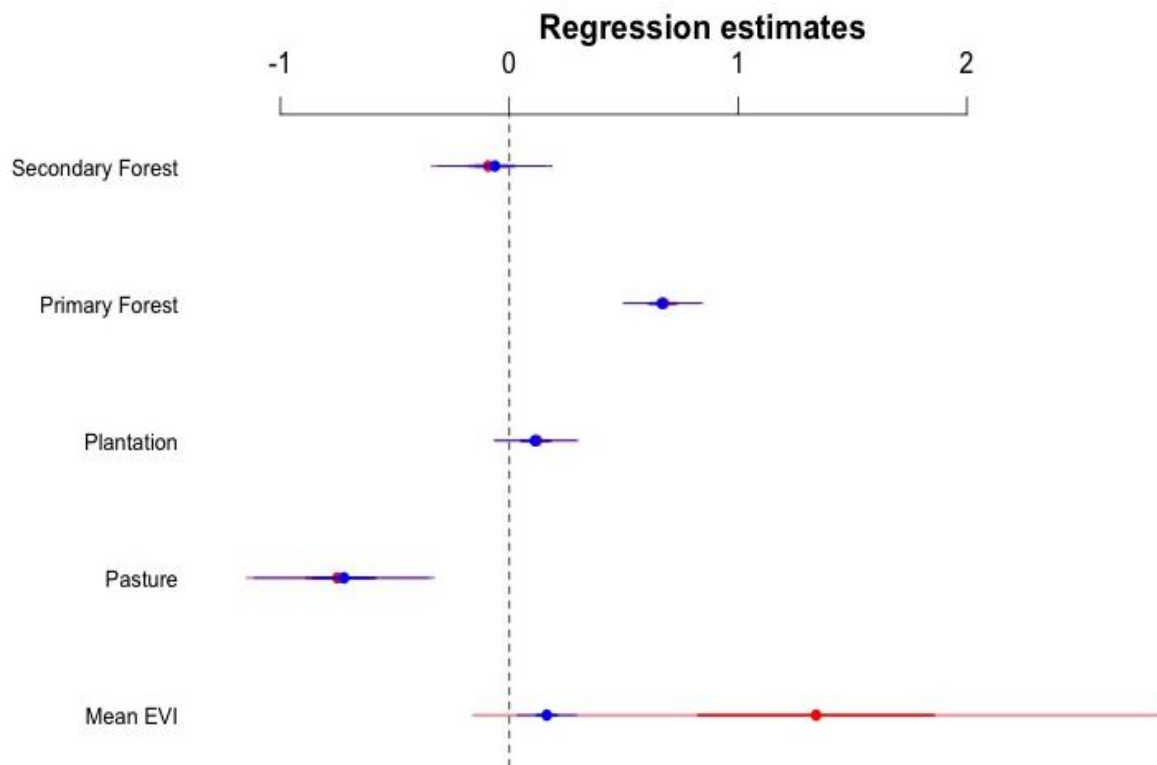




**Figure 7.** Plot of a Generalised Additive Mixed Model showing how relative abundance changes in response to 3-year mean EVI at a spatial extent of 250m<sup>2</sup> (centre pixel only). A saturation of EVI occurs where any additional increase does not determine any significant change in relative abundance. However, this nonlinearity may not be significantly improves the models explanatory power.

**Figure 8.** Plot of a Generalised Additive Mixed Model showing how relative abundance changes in response to 3-year mean EVI at a spatial extent of 10km<sup>2</sup>. The additional spatial extent has estimated a response with narrower confidence intervals, producing a slope for EVI significantly different from 0, which has reduced to a linear form.





**Figure 9.** Plot showing the parameter estimates for the minimal adequate model with 3-year mean EVI taken from the centre pixel only (250m<sup>2</sup> extent) in red, and from a spatial extent of 10km<sup>2</sup> in blue. The habitat type coefficients have no appreciable difference when including the greater EVI extent, but our confidence on the slope value for EVI dramatically increases, with much narrower intervals on the parameter estimate, becoming significantly different from 0. The EVI estimate also dramatically decreases, indicating some unexpected processes occurring (see figure 10).

Figure 7 shows a minimal adequate GAMM for 3-year mean EVI using only the central pixel of data. There is a nonlinear response of abundance to increase in EVI, showing a steady increase at low EVI followed by saturation at mean EVI of approximately 0.5; any further increase beyond this point produces no significant change in relative abundance. Figure 7, however, shows large confidence intervals around the response and figure 9 shows the parameter estimate for EVI is not significantly different from 0. By increasing the spatial extent for 3-year mean EVI, the relationship becomes linear (figure 8): the meaning of this can be interpreted by exploring  $\Delta AIC$  and behaviour of the EVI parameter estimate with increasing spatial extent.

Upon increasing the spatial extent of 3-year mean EVI to the maximum analysed (21 pixels: 10km<sup>2</sup> tiles), the AIC decreased ( $\Delta AIC = -8.2$ ; no incremental distance band increases yielded a significant drop in AIC – figure 10), confidence intervals narrowed, and the slope for EVI became significantly different from 0 ( $p < 0.05$ ), shown in figure 9. This pattern in the changes in AIC with spatial extent are matched at larger temporal scales, from 3 years and

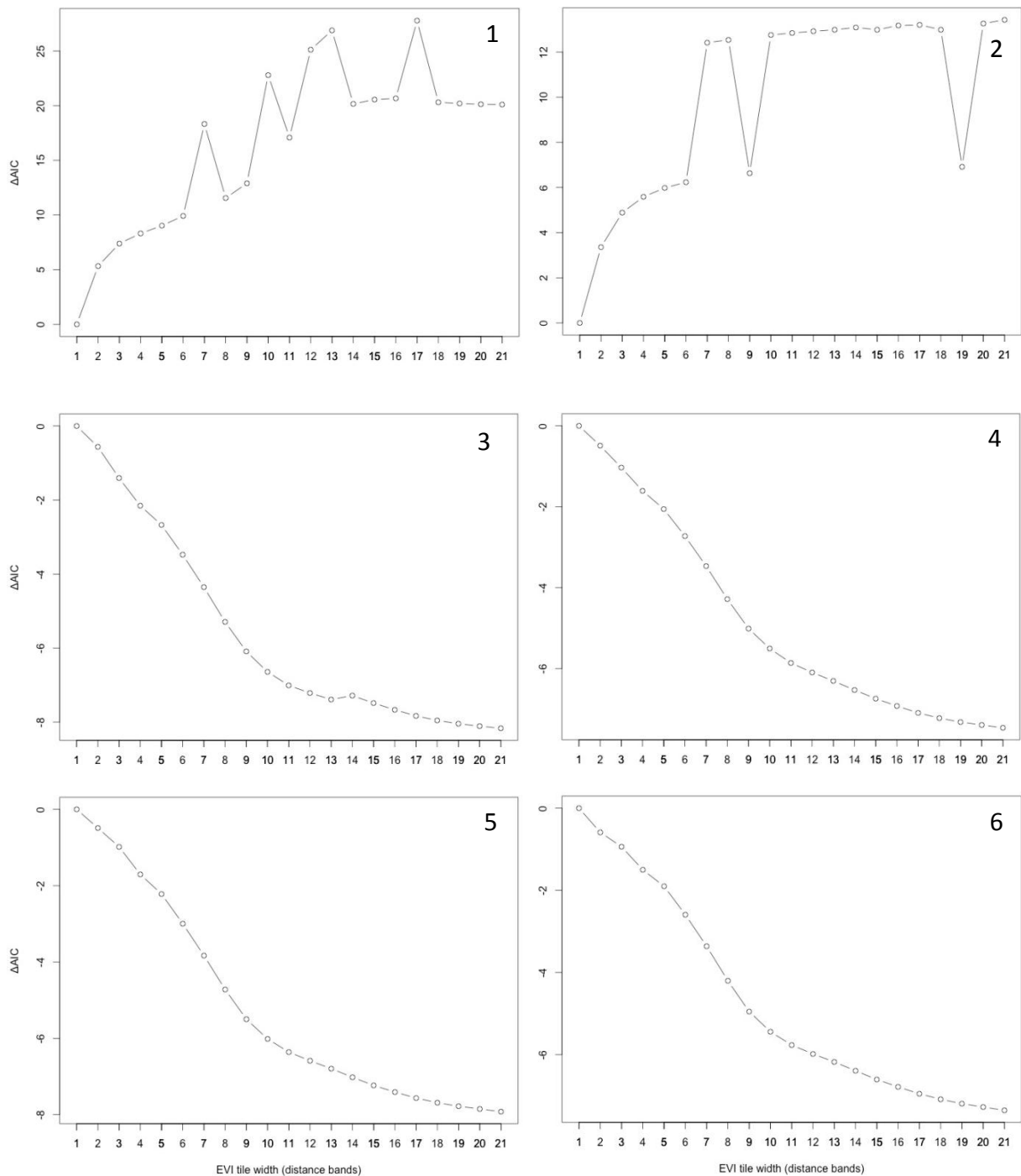
above. The changes in EVI parameter estimates at these temporal scales also follow the same pattern: a rapid decrease in the mean and variance of the parameter estimate.

The decrease in AIC indicates that increasing the spatial extent of environmental data has improved our confidence on estimates and enabled a better description of the relationship between abundance and mean EVI. However, as the AIC value and confidence on the parameter estimate improve, the coefficient value decreases and is no longer representative of the dynamics estimated at local scales (figure 11). When modelled using local EVI data, at a 3-year mean EVI value of 0.5 the estimated relative abundance yields a 0.6706 increase on the intercept value, whereas a 0.06799 increase is estimated from the equivalent EVI modelled using a 10km<sup>2</sup> extent: the local response estimated using local EVI is roughly 10 times stronger than that from 10km<sup>2</sup>, at the 3-year temporal scale. Inferences made on EVI at greater spatial scales should therefore be questioned (see discussion).

Increasing the time-series length beyond 3 years seems to have little impact on AIC, as the pattern of decrease remains similar and no additional significant decreases in AIC, relative to the 3-year mean, are evident (table 1). The parameter estimates are much more responsive to changes in spatial or temporal extent and are providing a more informative description of how our inference from the models change with varying the spatio-temporal scale of EVI, and how confident we can be of these inferences. The lowest AIC produced is with a spatial extent of 10km<sup>2</sup> using 3-year mean (AIC=22356.9), but not sufficiently different to claim it is the optimal extent; this spatio-temporal scale does not produce the most informative parameter estimates, therefore selection of this scale as optimal based on AIC is rejected.

The parameter estimates at 1- and 2-year temporal scales show interesting differences compared with larger temporal scales (figure 11). Initially, local parameter estimates are similar for all temporal scales, showing some marginal increase in significance with increasing temporal scale. As spatial scale is increased, however, parameter estimates increase and then stabilise, with narrowed confidence intervals; parameters remain representative of dynamics that were estimated at the local scale. Therefore, increasing spatial extent at short temporal scales is beneficial, but doing so at long temporal scales hinders drawing meaningful model inferences.

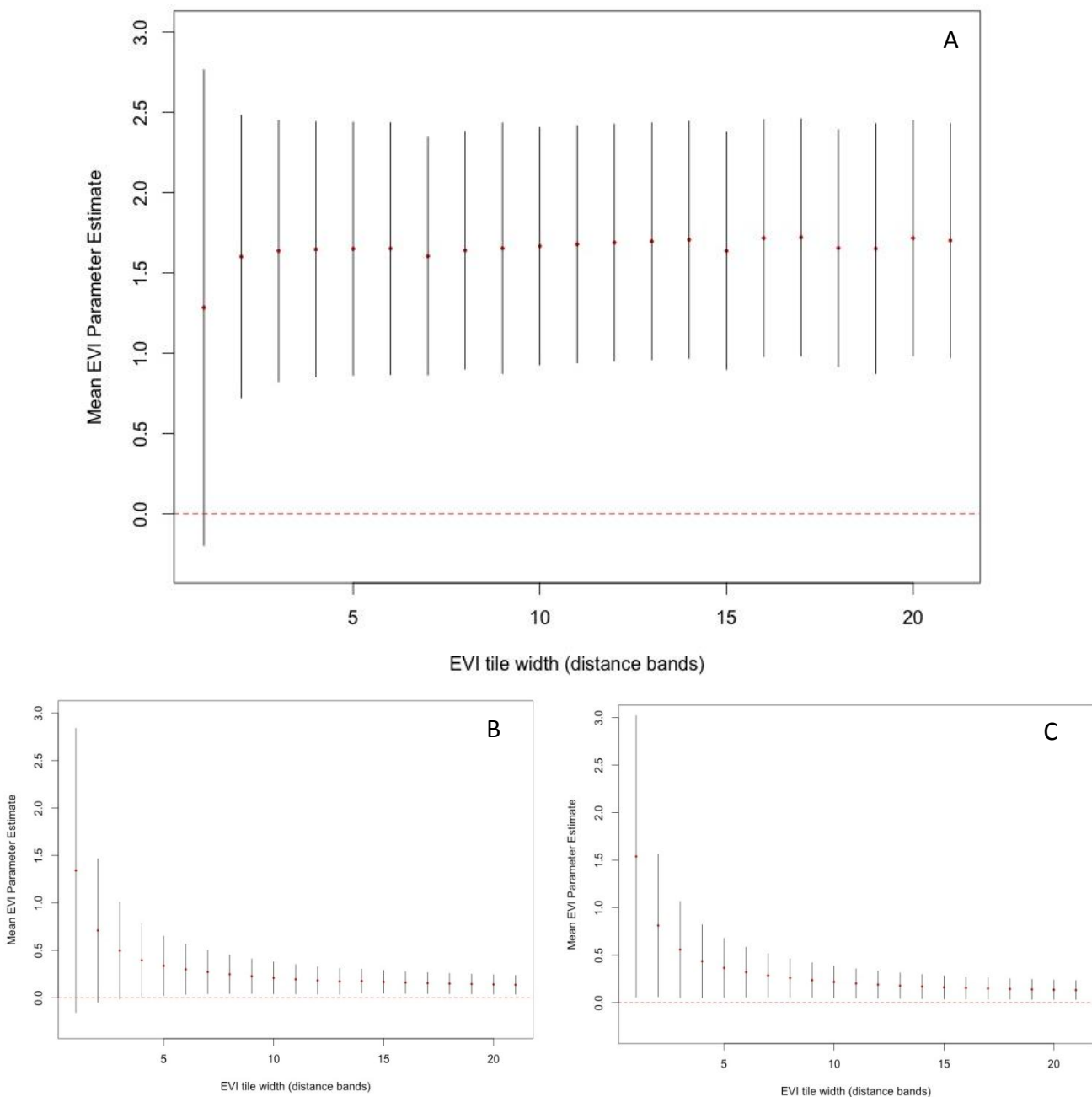
**Figure 10.** AIC values for minimal adequate models change as spatial extents are increased for all temporal scales (1-6 years).  $\Delta$ AIC is defined relative to AIC at 1 (considering only the centre pixel). Mean EVI calculated from 1 and 2 years are unreliable due to few high quality data in some places, as the prominence of short-term interference increases (e.g. cloud-cover). Increasing the time-series length provides more quality data, producing expected patterns for  $\Delta$ AIC with increasing spatial extent. AIC may increase for short time-series because more pixels allow for more inadequate data. There is a decrease in AIC for mean EVI calculated from time-series of 3, 4, 5 and 6 years: all show similar patterns in the decrease of AIC, and minimum values do not vary drastically. This suggests increasing the time-series beyond 3-years does not increase model fit.



**Figure 11.**

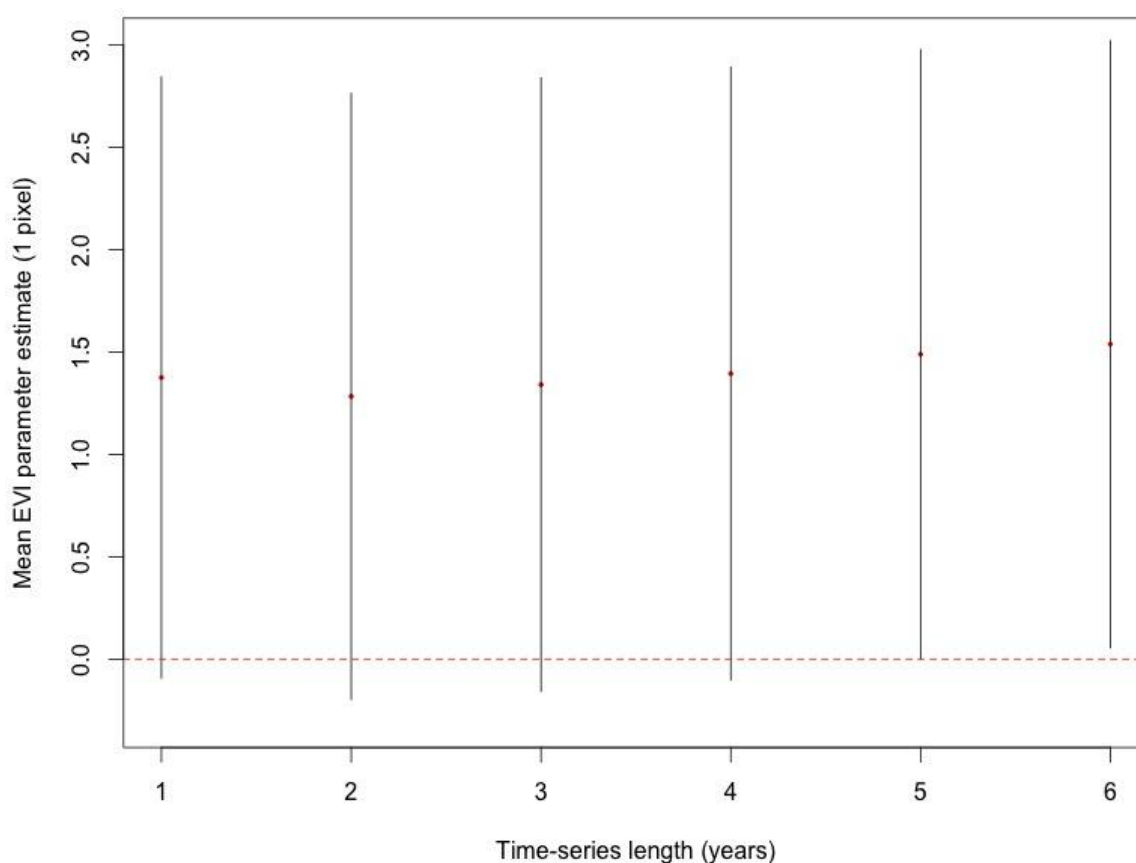
A: Changes in parameter estimate with increasing spatial extent for 2-year mean EVI. The AIC (figure 10) for 1- and 2-year mean EVI are erratic and identified poor quality data. The parameter estimates, however, show an interesting result: increasing spatial extent causes the parameter estimate to slightly increase and rapidly settle on a value that seems informative and maintains details of local dynamics. Confidence intervals around the parameter estimates also improve. This finding, coupled with the AIC behaviour suggests that, in some places, there is insufficient data at small temporal scales, but where there are sufficient data, the model provides a much richer description of local dynamics.

B & C: The same plot for time-series lengths of 3-years (B) and 6-years (C). Parameter estimates at the local scale are similar to A, but a different pattern with increased spatial extent is shown. The parameter estimate mean and variance decreases as the estimate declines until the estimate becomes statistically significant. However, AIC suggests an improved fit. By increasing EVI spatial extent, the model better discriminates between low and high EVI sites, but includes increasingly irrelevant EVI information, blurring the local abundance response to environmental pressures at that site. These findings suggest some trade-off between model fit (and reliable data) and meaningful parameter estimates.



Time-series length (years)	AIC	d.f.	$\Delta$ AIC
1	22365.6	12	0.8
2	22369.2	12	4.4
3	22365.0	12	0.3
4	22365.3	12	0.5
5	22365.4	12	0.7
6	22364.8	12	0.0

**Table 1.** The AIC values from minimal adequate models, using an EVI extent of the central pixel only, for differing temporal scales. There is no significant decrease in AIC as temporal scale is increased. This suggests that by taking the mean of EVI over a longer time period model fit is not improved. Changes in parameter estimates should also be considered before drawing inferences, but this suggests including additional years is not worthwhile. Time-series lengths of 1 and 2 years have already been identified as unreliable, so the most parsimonious time-series length would be 3 years.



**Figure 12.** The local scaled mean EVI parameter estimate for all temporal scales. As time-series length is increased the significance of the estimate improves, but only marginally. Estimates remain as variable. This indicates that 6-year mean EVI at the most local of scales is the most informative, although there is little room for discrimination between time-series lengths. 1- and 2-year means follow a similar pattern at this local scale, but as shown above (figure 11), may prove to be the most informative temporal scales when spatial extent is increased.



## Discussion

### Synthesis

MODISTools has provided a simple and efficient method for accessing and making use of remotely-sensed data, and models have shown that the retrieved EVI data from MODIS has been informative as a proxy for habitat destruction and degradation, in modelling how local abundance responds to the intensity of mechanisms of environmental change.

Whilst the data have been informative, it is not clear what spatio-temporal extent of EVI provides the most explanatory power to answer this biological question. There appears to be a trade-off between model fit according to AIC and capturing a detailed relationship between local abundance and the mean EVI in that local area. Prior to undertaking this study, it was assumed that local processes could be modelled using only the local environmental conditions, but by increasing the spatial extent to encompass an environmental neighbourhood a more reliable estimate of environmental pressures in that area could be derived. What we see at increased spatial extent, however, is a decline on the coefficient of mean EVI and its variance shrinking; as the coefficient value declines the estimated slope for EVI becomes weaker. The increase in significance corresponds with AIC decrease at large temporal scales, suggesting an improved model fit. By increasing EVI spatial extent the model better discriminates between low and high EVI sites, but increasingly irrelevant EVI information is included, blurring the local response of abundance to environmental pressures at that site. This causes an underestimation of coefficient values and a less detailed picture of how abundance is responding to local environmental pressures.

This finding is corroborated when the GAMM plots in figures 7 and 8 are compared: when unaware of the spatial dependency of the parameter estimate (it should be noted, only the mean EVI parameter is spatially dependent – see figure 9) the improved AIC suggests that the reduction to a linear form at larger extents is the product of a better description of the relationship between abundance and EVI. With the identified spatial dependency in mind, however, it is more likely that the nonlinear response found in figure 7 is a more detailed local response, and that the increased spatial extent is producing a less clear picture, giving a

linear form as a result. Therefore, in order to find an optimal extent, a trade-off between AIC and local detail must be met.

Figure 11 shows how the coefficient of mean EVI is affected by increasing the spatial and temporal extent. At temporal scales of 3 years and above, the estimate rapidly decreases as more pixels are included. For these models and data, the decrease in AIC with spatial extent (-8.2 when increasing extent from 1 250x250m pixel, to 10km<sup>2</sup> tile of pixels, for 3-year mean EVI) is not sufficient enough to justify losing any detail in the local response. Therefore, for a detailed local picture to be maintained the spatial extent should be limited to only the central pixel overlaying the abundance location.

The EVI parameter estimate marginally increases as temporal extent is increased. In fact, the increase from 3 to 6 years is enough for the estimate to become significantly different from 0. AIC has no appreciable change, however, the slightly improved confidence on the parameter estimate whilst maintaining a detailed local response warrants including a greater temporal extent when calculating mean EVI. As the parameter estimate actually increases slightly, it could be argued that by including a greater time span of the local environmental conditions prior to the abundance recording an even clearer picture of the local dynamics are being elucidated.

As shown in figure 11, at temporal scales lower than 3 years, a very different inference can be made from the models. The AIC values at varying spatial extents for 1- and 2-year mean EVI are unreliable and erratic, as a result of insufficient data in the time-series at some locations. The parameter estimates, however, show a much more meaningful trend as spatial extent is increased. A value similar to the most local estimate is converged upon and confidence intervals have narrowed, maintaining the detailed local response that is identified when considering only the most local dynamics. The behaviour of AIC values and parameter estimates with increasing spatial scale show contrasting results, in that improved (but marginal) model fit (figure 10) is met with less meaningful parameter estimates (figure 11).

Larger temporal scales facilitate a greater proportion of high quality EVI data, which allows better discrimination between high and low EVI sites, but includes EVI dynamics that are increasingly irrelevant to the local response. Whereas, smaller temporal scales are data

deficient in some locations, but at locations where EVI data is high quality, a more detailed response of abundance can be elucidated and increasing spatial extent can help improve estimates on the relationship between abundance and mean EVI. This trade-off between model fit and meaningful parameter estimates manifests itself as a trade-off between temporal and spatial extent.

For these models and data, increasing spatial and temporal scale of environmental data does not improve our inference of local abundance response to EVI. Increasing either temporal or spatial scale separately, however, can be beneficial. The trade-off between scales witnessed highlights the importance of using environmental data that is pertinent to the biological question being asked. Here, a local response of abundance to EVI was modelled. By taking either the local dynamics for a long period of time, or a wider surrounding extent of data around the local dynamics, descriptive model parameters can be estimated and meaningful inferences drawn. But, including wider spatial extents for long time-periods prior to abundance observation can blur out the environmental picture that is pertinent to the local dynamics of interest, producing less meaningful results. Changes in abundance, as opposed to species-level biodiversity measures, occur on relatively short timescales (Balmford *et al.*, 2003); this means increasing the temporal extent may easily surpass the scale of abundance dynamics, including irrelevant past conditions as a result. Judgement on the boundary between relevant and irrelevant inference is open to interpretation.

### **Limitations & Further Work**

The form of distance-decay of EVI from the location of interest is assumed to follow a Gaussian curve. This was chosen because it has previously been used when testing spatial dependency and nonstationarity in vegetation indices (Gao *et al.*, 2012). It is possible that other distance-decay functions better describe how importance of EVI decreases with distance from the location of interest. A different exponential form, for example, or a power law decay on a constant that could be parameterised using maximum likelihood estimation. A power law decay function was attempted, but parameterisation of the function to converge on a value for the constant, for models as complex as those used in this study, proved very difficult and computationally intensive. Other spatial models such as Geographically Weighted Regression (GWR) or Spatial Autoregressive Models (SAR) may have been useful, but specification when fitting to a dataset with as complex a structure as

that used in this study would be problematic, and would require more detailed coordinate data down to the pixel-level.

The Modifiable Areal Unit Problem always exists in spatial analysis: our inferences for spatial extent may be valid using these models and data, but if EVI becomes available at a finer resolution it is possible our inference would change. Further spatial extents may give more insights into the impact of environmental scale. However, there must be a reasonable, finite limit to the maximum scale implemented. The results indicate that local dynamics are already being lost at spatial scales of  $10\text{km}^2$ , so it could be argued that increasing spatial extent further is unlikely to be informative.

An assumption of the models used is that the explanatory variables are measured without error. EVI measurements do vary as a result of environmental conditions, although it does perform better than NDVI in removing effects of aerosol and background soil interference. Additional steps are taken to maximise the quality of EVI by removing values with a low reliability score (as part of the MOD13Q1 product downloaded) and linearly interpolating between high quality values to take a time-series mean.

The model residuals are shown to be non-normal and there is heteroscedasticity (figure 6): these are most likely the result of unaccounted for zero-inflation. Zero-inflation was not dealt with because there is little support within R for parameterising complex GAMMs with zero-inflation incorporated. It may have been possible to implement a two-step hurdle model (Welsh *et al.*, 1996; Heilbron, 1994), where the additional zeros not expected from a poisson process are dealt with using a binomial model, after which, the remaining data is treated with a truncated poisson model (or similar). However, assumptions arise from this method too: the additional zeros are assumed to be produced by a separate process to the rest of the data (e.g. observation error – an additional complication of dealing with this is that observation error will vary between studies).

Methods for model evaluation and selection with GLMMs and GAMMs are deceptively difficult (Zuur *et al.*, 2009), as maximum likelihood estimates can only be found by integrating over all possible values for the random-effects (Bolker *et al.*, 2009). Analysing mixed models can be problematic because tests used to make inferences on results assume that the null values of the parameters to be estimated are not at the edge of their possible

values (Bolker *et al.*, 2009; Zuur *et al.*, 2009): this assumption is violated when standard deviations for a random effect are estimated because its value has to be  $\geq 0$  (null value is 0). There is also ambiguity in how degrees of freedom are calculated: here the minimum degrees of freedom contributed by random effects that affect the term being tested is used.

The habitat type fixed-effect was determined based on the descriptions given in each publication and given a land-cover class that adheres to the classification system used in the major biodiversity indicator, GLOBIO3 (Alkemade *et al.*, 2009). This gives consistency and opportunity for comparison, but may not be the best match for EVI. MODIS provide a surface of land-cover data calculated using EVI, recorded at the same resolution and time intervals. In so far as it uses EVI to differentiate between habitat types, using MODIS land-cover as a categorical fixed-effect may provide a better description of how the local response of abundance varies in different habitats. However, any unreliability inherent in EVI would therefore propagate through to land-cover data.

The impact of spatio-temporal extent was explored on a global level in the sense that all data were included, irrespective of habitat type or geographic region (e.g. continent). This analysis could be disaggregated to different levels to see if conclusions on environmental data scales differ between habitats or regions. The data used are exclusive to the tropics. Further work could also look at the same questions in other biomes. This may facilitate a more mechanistic approach to describing an optimal spatio-temporal environmental scale.

The response of local abundance to intensity of environmental pressures may be better modelled by including additional explanatory variables that provide information on other mechanisms for environmental change. Layers of population density or distance from roads, for example, are commonly used as proxies for intensity of anthropogenic impact (Luck, 2007). Remotely-sensed data on fire damage may explain effects of increased frequency and intensity of disturbance events (Cochrane & Laurance, 2002; Kerr & Ostrovsky, 2003), but may also be collinear with EVI. Physical characteristics of habitat quality – biogeochemical flux, for instance – may also be beneficial (Lee *et al.*, 2010; Foster *et al.*, 2011).

Models need biological traits too; incorporating functional traits of specific taxa (Díaz & Cabido, 2001; Dawson *et al.*, 2011) or community-level processes (Mokany *et al.*, 2012, Magurran *et al.*, 2010) to provide a more mechanistic explanation of biodiversity responses

environmental change. In terms of inferring upon wider biodiversity, abundance goes some way but an assessment of biodiversity would need to model the response of other measures, such as phylogenetic diversity, to capture its many components (Magurran, 2004; Lyashevskaya & Farnsworth, 2012).

### **Concluding Remarks**

Conclusions made suggest that only the most local and current of EVI dynamics are important in describing the local response of abundance to EVI, but that reliability of these dynamics can be improved by including a small surrounding neighbourhood of dynamics or a short length of time previous to abundance observation. For a parsimonious but meaningful model of local abundance response to environmental pressures intensity, an EVI spatio-temporal scale of 2-year mean for 2 pixel-bands (area  $\approx 1\text{km}^2$ ) could be suggested (figure 11). The most detailed description of local dynamics presented in the study is most likely the nonlinear response identified in figure 7. This model suggests that abundance saturates at high EVI values. From a biodiversity conservation perspective, this presents a threshold effect: habitat destruction/degradation up to a certain level may produce no observed abundance change over timescales concerned, but beyond a given threshold limit further EVI decline causes a significant decline in local abundance.

The optimal spatio-temporal scale is debatable and its definition requires some judgement on behalf of the researcher in balancing model fit and meaningful model inference. Asking questions of scale in studies on biodiversity and wider ecological contexts is important, but the effects of scale are complex (Bar-Massada *et al.*, 2012) and statistically challenging to infer upon (Whittaker *et al.*, 2001; Foody, 2004).

### **Acknowledgements**

Thanks to Andy Purvis for encouraging and diligent supervision throughout, to Helen Phillips for the provision of abundance data, to Andy Purvis' lab group for being R package guinea pigs, to Dan Reuman and all MSc Quantitative Biology teachers for fantastic training, and my family for always being interested and supportive.

## References

- Achard, F., Eva, H. D., Stibig, H.-J., Mayaux, P., Gallego, J., Richards, T., Malingreau, J.-P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*. Vol.297(5583), p.999-1002.
- Alkemade, R., Bakkenes, M., Eickhout, B. (2011). Towards a general relationship between climate change and biodiversity: an example for plant species in Europe. *Reg. Environ. Change*. Vol.11, p.S143-S150.
- Alkemade, R., Reid, R. S., van den Berg, M., de Leeuw, J., Jeuken, M. (2012). Assessing the impacts of livestock production on biodiversity in rangeland ecosystems. *PNAS*. doi:10.1073/pnas.1011013108.
- Alkemade, R., van Oorschot, M., Miles, L., Nellemann, C., Bakkenes, M., ten Brink, B. (2009). GLOBIO3: A framework to investigate options for reducing global terrestrial biodiversity loss. *Ecosystems*. Vol.12, p.374-390.
- Bai, Z. G., Dent, D. L., Olsson, L., Schaepman, M. E. (2008). Proxy global assessment of land degradation. *Soil Use and Management*. Vol.24, p.223-234.
- Balmford, A., Green, R. E., Jenkins, M. (2003). Measuring the changing state of nature. *TRENDS in Ecology and Evolution*. Vol.18(7), p.326-330.
- Bar-Massada, A., Wood, E. M., Pidgeon, A. M., Radeloff, V. C. (2012). Complex effects of scale on the relationships of landscape pattern versus avian species richness and community structure in a woodland savannah mosaic. *Ecography*. Vol.35, p.393-411.
- Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M., Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*. Vol.35, p.673-683.
- Bolker, B., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, H. H. M., White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *TRENDS in Ecology and Evolution*. Vol.24(3), p.127-135.

- Breslow, N. (1990). Tests of hypotheses in overdispersed poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association*. Vol.85(410), p.565-571.
- Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E. M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K. E., Carr, G. M., Chanson, J., Chenery, A. M., Csirke, J., Davidson, N. C., Dentener, F., Foster, M., Galli, A., Galloway, J. N., Genovesi, P., Gregory, R. D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M. A., McRae, L., Minasyan, A., Morcillo, M. H., Oldfield, T. E. E., Pauly, D., Quader, S., Revenga, C., Sauer, J. R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S. N., Symes, A., Tierney, M., Tyrrell, T. D., Vié, J.-C., Watson, R. (2010). Global biodiversity: indicators of recent declines. *Science*. Vol.328, p.1164-1168.
- Cadotte, M. W., Davies, J., Regetz, J., Kembel, S. W., Cleland, E., Oakley, T. H. (2010). Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecology Letters*. Vol.13, p.96-105.
- Carey, C., Heyer, W. R., Wilkinson, J., Alford, R. A., Arntzen, J. W., Halliday, T., Hungerford, L., Lips, K. R., Middleton, E. M., Orchard, S. A., Rand, S. A. (2001). Amphibian declines and environmental change: use of remote-sensing data to identify environmental correlates. *Conservation Biology*. Vol.15(4), p.903-913.
- CBD. (2010). *Global Biodiversity Outlook 3*. Montreal: Secretariat of the Convention on Biological Diversity.
- CBD. (2011). Report of the tenth meeting of the conference of the parties to the Convention on Biological Diversity. *COP10*. Nagoya, Japan: 18-29 October 2010. UNEP.
- Ceballos, G., Ehrlich, P. R. (2002). Mammal population losses and the extinction crisis. *Science*. Vol.296, p.904-907.
- Certain, G., Skarpaas, O., Bjerke, J.-W., Framstad, E., Lindholm, M., Nilsen, J.-E., Norderhaug, A., Oug, E., Pedersen, H.-C., Schartau, A.-K., van der Meeren, G. I., Aslaksen, I., Engen, S., Garnåsjordet, P.-A., Kvaløy, P., Lillegård, M., Yoccoz, N. G., Nybø, S. (2011). The



Nature Index: A general framework for synthesizing knowledge on the state of biodiversity. *PLoS ONE*. Vol.6(4), p.e18930.

Chawla, A., Yadav, P. K., Uniyal, S. K., Kumar, A., Vats, S. K., Kumar, S., Ahuja, P. S. (2012). Long-term ecological and biodiversity monitoring in the western Himalaya using satellite remote sensing. *Forest Remote Sensing, Biodiversity and Climate*. Vol.102(8), p.1143-1156.

Cochrane, M. A., Laurance, W. F. (2002). Fire as a large-scale edge effect in Amazonian forests. *Journal of Tropical Ecology*. Vol.18, p.311-325.

Collen, B., Loh, J., Whitmee, S., McRae, L., Amin, R., Baillie, J. E. M. (2009). Monitoring change in vertebrate abundance: The Living Planet Index. *Conservation Biology*. Vol.23(2), p.317-327.

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, C. I., Mace, G. M. (2011). Beyond predictions: biodiversity conservation in a changing climate. *Science*. Vol.332, p.53-58.

Díaz, S., Cabido, M. (2001). Vive la différence: plant functional diversity matters to ecosystem processes. *TRENDS in Ecology and Evolution*. Vol.16(11), p.646-655.

Donoghue, D. N. M. (2002). Remote sensing: environmental change. *Progress in Physical Geography*. Vol.26(1), p.144-151.

Elston, D. A., Moss, R., Boulinier, T., Arrowsmith, C., Lambin, X. (2001). Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks. *Parasitology*. Vol.122(5), p.563-569.

Foody, G. M. (2004). Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*. Vol.13, p.315-320.

Foster, W. A., Snaddon, J. L., Turner, E. C., Fayle, T. M., Cockerill, T. D., Ellwood, F. M. D., Broad, G. R., Chung, A. Y. C., Eggleton, P., Khen, C. V., Yusah, K. M. (2011). Establishing

the evidence base for maintaining biodiversity and ecosystem function in the oil palm landscapes of South East Asia. *Phil. Trans. R. Soc. B.* Vol.366, p.3277-3291.

Gallardo, B., Gascón, S., Quintana, X., Comín, F. A. (2011). How to choose a biodiversity indicator - redundancy and complementarity of biodiversity metrics in a freshwater ecosystem. *Ecological Indicators.* Vol.11(5), p.1177-1184.

Gao, X., Huete, A. R., Ni, W., Miura, T. (2000). Optical-biophysical relationships of vegetation spectra without background contamination. *Remote Sensing of Environment.* Vol.74, p.609-620.

Gao, Y., Huang, J., Li, S., Li, S. (2012). Spatial pattern of non-stationarity and scale-dependent relationships between NDVI and climatic factors - A case study in Qinghai-Tibet Plateau, China. *Ecological Indicators.* Vol.20, p.170-176.

Gibson, L., Lee, T. M., Koh, L. P., Brook, B. W., Gardner, T. A., Barlow, J., Peres, C. A., Bradshaw, C. J. A., Laurance, W. F., Lovejoy, T. E., Sodhi, N. S. (2011). Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature.* Vol.478, p.378-381.

Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications.* Vol.10(6), p.1861-1870.

Heilbron, D. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal.* Vol.36, p.531-547.

Hinde, J. (1982). Compound poisson regression models. In: *GLIM 82: Proceedings of the International Conference on Generalized Linear Models.* ed. Gilchrist, R. Springer: Berlin. pp.109-121.

Holm, A. M., Cridland, S. W., Roderick, M. L. (2003). The use of time-integrated NOAA NDVI data and rainfall to assess landscape degradation in the arid shrubland of Western Australia. *Remote Sensing of Environment.* Vol.85, p.145-158.

Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment.* Vol.83, p.195-213.

- Isaac, N. J. B., Cowlshaw, G. (2004). How species respond to multiple extinction threats. *Proc. R. Soc. B.* Vol.271, p.1135-1141.
- Justice, C. O., Townshend, J. R., Vermote, E., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., Morisette, J. T. (2002). An overview of MODIS land data processing and product status. *Remote Sensing of Environment.* Vol.83, p.3-15.
- Justice, C. O., Vermote, E., Townshend, J. R., Defries, R., Roy, D. P., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A., Lucht, W., Myneni, R. P., Knyazikhin, Y., Running, S. W., Nemani, R. R., Wan, Z., Huete, A. R., van Leeuwen, W., Wolfe, R. E., Giglio, L., Muller, J.-P., Lewis, P., Barnsley, M. J. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *NASA Publications.* <http://digitalcommons.unl.edu/nasapub/28>
- Kerr, J. T., Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *TRENDS in Ecology and Evolution.* Vol.18(6), p.299-305.
- Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics.* Vol.15(3), p.209-225.
- Lee, M., Manning, P., Rist, J., Power, S. A., Marsh, C. (2010). A global comparison of grassland biomass responses to CO<sub>2</sub> and nitrogen enrichment. *Phil. Trans. R. Soc. B.* Vol.365, p.2047-2056.
- Loh, J., Green, R. E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., Randers, J. (2005). The Living Planet Index: using species population time series to track trends in biodiversity. *Phil. Trans. R. Soc. B.* Vol.360(289-295).
- Luck, G. W. (2007). A review of the relationships between human population density and biodiversity. *Biological Reviews.* Vol.82(4), p.607-645.
- Lyashevskaya, O., Farnsworth, K. D. (2012). How many dimensions of biodiversity do we need? *Ecological Indicators.* Vol.18, p.485-492.
- MA. (2005). *Millennium ecosystem assessment. Ecosystems and human well-being: scenarios2.* Washington D.C: Island Press.

- Mace, G. M., Baillie, J. E. M. (2007). The 2010 biodiversity indicators: challenges for science and policy. *Conservation Biology*. Vol.21(6), p.1406-1413.
- Mace, G. M., Cramer, W., Díaz, S., Faith, D. P., Larigauderie, A., Le Prestre, P., Palmer, M., Perrings, C., Scholes, R. J., Walpole, M., Walther, B. A., Watson, J. E. M., Mooney, H. A. (2010). Biodiversity targets after 2010. *Current Opinion in Environmental Sustainability*. Vol.2, p.1-6. doi:10.1016/j.cosust.2010.03.003.
- Mace, G. M., Gittleman, J. L., Purvis, A. (2003). Preserving the tree of life. *Science*. Vol.300, p.1707-1709.
- Magurran, A. E. (2004). *Measuring Biological Diversity*. Oxford: Blackwell Science Ltd.
- Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. McP., Elston, D. A., Scott, M. A., Smith, R. I., Somerfield, P. J., Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *TRENDS in Ecology and Evolution*. Vol.25(10), p.574-582.
- Majer, J. D., Beeston, G. (1996). The biodiversity integrity index: an illustration using ants in Western Australia. *Conservation Biology*. Vol.10, p.65-73.
- Mokany, K., Harwood, T. D., Williams, K. J., Ferrier, S. (2012). Dynamic macroecology and the future for biodiversity. *Global Change Biology*. doi: 10.1111/j.1365-2486.2012.02760.x
- Naeem, S., Duffy, E. J., Zavaleta, E. (2012). The functions of biological diversity in an age of extinction. *Science*. Vol.336(6087), p.1401-1406.
- NASA Land Processes Distributed Active Archive Center (LP DAAC). (2012). *MODIS Overview*. ORNL DAAC. URL: [https://lpdaac.usgs.gov/products/modis\\_overview](https://lpdaac.usgs.gov/products/modis_overview).
- Nichols, E., Larsen, T., Spector, S., Davis, A. L., Favila, M., Vulinec, K., Network., T. S. R. (2007). Global dung beetle response to tropical forest modification and fragmentation: a quantitative literature review and meta-analysis. *Biol. Conserv.* Vol.137, p.1-19.
- Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-Manjarrés, J. F., Araújo, M. B., Balvanera, P., Biggs, R., Cheung, W. W. L., Chini, L.,

- Cooper, D. H., Gilman, E. L., Guénette, S., Hurtt, G. C., Huntington, H. P., Mace, G. M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R. J., Sumaila, U. R., Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science*. Vol.330, p.1496-1501.
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *TRENDS in Ecology and Evolution*. Vol.20(9), p.503-510.
- Pimm, S. L., Russell, G. J., Gittleman, J. L., Brooks, T. M. (1995). The future of biodiversity. *Science*. Vol.269, p.347-350.
- Purvis, A., Hector, A. (2000). Getting the measure of biodiversity. *Nature*. Vol.405, p.212-219.
- R Core Team. (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. URL: <http://www.R-project.org/>.
- Rangel, T. F. (2012). Amazonian Extinction Debts. *Science*. Vol.337(6091), p.162-163.
- Reich, P., Tilman, D., Isbell, F., Mueller, K., Hobbie, S. E., Flynn, D. F. B., Eisenhauer, N. (2012). Impacts of biodiversity loss escalate through time as redundancy fades. *Science*. Vol.336, p.589-592.
- Richards, S. A. (2008). Dealing with overdispersed count data in applied ecology. *Journal of Applied Ecology*. Vol.45, p.218-227.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience*. Vol.54(6), p.547-560.
- Scholes, R. J., Biggs, R. (2005). A biodiversity intactness index. *Nature*. Vol.434, p.45-49.
- Scholes, R. J., Mace, G. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., Muchoney, D., Walther, B. A., Mooney, H. A. (2008). Toward a global biodiversity observing system. *Science*. Vol.321, p.1044-1045.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S., Justice, C. O., Dazlich, D. A., Randall, D. A. (1994). A global 10\*10 NDVI data set for climate studies. Part 2 - The adjustment of the NDVI and generation of global fields of terrestrial biophysical parameters. *International Journal of Remote Sensing*. Vol.15, p.3519-3545.

Temple Lang, D. (2012a). *SSOAP: Client-side SOAP access for S*. R package version 0.91-0. URL: <http://www.omegahat.org/SSOAP>.

Temple Lang, D. (2012b). *XMLSchema: R facilities to read XML schema*. R package version 0.8-0. URL: <http://www.omegahat.org/XMLSchema>.

Thiam, A. K. (2003). The causes and spatial pattern of land degradation risk in Southern Mauritania using multitemporal AVHRR-NDVI imagery and field data. *Land Degradation & Development*. Vol.14, p.133-142.

Tuck, S. (2012). *MODISTools: MODIS Subsetting Tools*. R package version 0.9. URL: <https://github.com/seantuck12/MODISTools/downloads>.

UNEP. (2001). Global methodology for mapping human impacts on the biosphere. *Report UNEP/DEWA/TR 25*.

UNEP. (2002). *Global environment outlook 3*. London: Earth-scan Publications Ltd.

Wearn, O. R., Reuman, D. C., Ewers, R. M. (2012). Extinction debt and windows of conservation opportunity in the Brazilian Amazon. *Science*. Vol.337, p.228-232.

Welsh, A. H., Cunningham, R. B., Donnelly, C. F., Lindenmayer, D. B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*. Vol.88, p.297-308.

Whittaker, R. J., Willis, K. J., Field, R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*. Vol.28, p.453-470.

Wisley, B., Chalcraft, D. R., Bowles, C. M., Willig, M. R. (2005). Relationships among indices suggest that richness is an incomplete surrogate for grassland biodiversity. *Ecology*. Vol.86(5), p.1178-1184.

Wood, S. N. (2006). *Generalized Additive Models: an introduction with R*. Chapman & Hall.

Xiao, X., Braswell, B., Zhang, Q., Boles, S., Frohking, S., Moore III, B. (2003). Sensitivity of vegetation indices to atmospheric aerosols: continental-scale observations in Northern Asia. *Remote Sensing of Environment*. Vol.84, p.385-392.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A., Hodges, J. C. F., Gao, F., Reed, B. C., Huete, A. R. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*. Vol.84, p.471-475.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

# Appendices

## MODISSubsets function R code:

```
MODISSubsets <- function(LoadDat, LoadMethod='object' | 'ext.file', FileSep=NULL, Product,
Bands, Size=c(), SaveDir=NULL, StartDate=FALSE, TimeSeriesLength=2, DateFormat='year' |
'posixt', Transect=FALSE)
{
  if(LoadMethod == 'object') { dat<- data.frame(LoadDat) } # Load external file, or R object.

  if(LoadMethod == 'ext.file') { dat<- read.delim(LoadDat, sep=FileSep) }

  if(any(is.na(dat$lat) != is.na(dat$long)) == TRUE) { stop('Not equal amount of lats and
longs: there must be locations with incomplete coordinate information.') } # Test for missing
lat/long data

  Start<- rep(StartDate, length(dat$lat[!is.na(dat$lat)]))

  ifelse(Start == TRUE, lat.long<-
unique(cbind(lat=dat$lat[!is.na(dat$lat)],long=dat$long[!is.na(dat$lat)],end.date=dat$end.date
[!is.na(dat$lat)],start.date=dat$start.date[!is.na(dat$lat)]), lat.long<-
unique(cbind(lat=dat$lat[!is.na(dat$lat)],long=dat$long[!is.na(dat$lat)],end.date=dat$end.date
[!is.na(dat$lat)])) # Finds all unique time-series wanted, for each unique location.

  print(paste('Found ',nrow(lat.long),' unique time-series to download.',sep=''))

  Start<- rep(StartDate, nrow(lat.long))

  if(nrow(lat.long) != length(unique(dat$ID))) {

    ifelse(Start == TRUE, ID<-
paste(lat.long[,1],lat.long[,2],lat.long[,3],lat.long[,4],sep=''), ID<-
paste(lat.long[,1],lat.long[,2],lat.long[,3],sep=''))

    lat.long<- data.frame(SubsetID=ID,lat.long,Status=rep(NA,nrow(lat.long)))

    print('IDs do not contain unique time-series: using subset IDs instead.')
  } else {

    lat.long<- data.frame(SubsetID=unique(dat$ID),lat.long,Status=rep(NA,nrow(lat.long)))
  }

  if(DateFormat == 'year') {

    ifelse(Start == FALSE,start.date<-strptime(paste(lat.long[,4]-TimeSeriesLength,'-01-
01',sep=''),'%Y-%m-%d'),start.date<-strptime(paste(lat.long[,5],'-01-01',sep=''),'%Y-%m-%d'))

    end.date<- strptime(paste(lat.long[,4],'-12-31',sep=''),'%Y-%m-%d')# Dates in POSIXlt.

    start.day<- start.date$yday

    start.day[nchar(start.day) == 2]<- paste(0, start.day[nchar(start.day) == 2], sep='')

    start.day[nchar(start.day) == 1]<- paste(0, 0, start.day[nchar(start.day) == 1], sep='')

    end.day<- end.date$yday

    end.day[nchar(end.day) == 2]<- paste(0, end.day[nchar(end.day) == 2], sep='')

    end.day[nchar(end.day) == 1]<- paste(0, 0, end.day[nchar(end.day) == 1], sep='')

    MODIS.start<- paste('A', substr(start.date, 1, 4), start.day, sep='')

    MODIS.end<- paste('A', substr(end.date, 1, 4), end.day, sep='')

  }

  if(DateFormat == 'posixt') {

    ifelse(Start == FALSE, start.date<- strptime(paste(lat.long[,4]-TimeSeriesLength,'-01-
01',sep=''),'%Y-%m-%d'), start.date<- strptime(lat.long[,5],'%Y-%m-%d'))
  }
}
```



```

end.date<- strptime(lat.long[,4],'%Y-%m-%d')

start.day<- start.date$yday

start.day[nchar(start.day) == 2]<- paste(0, start.day[nchar(start.day) == 2], sep='')

start.day[nchar(start.day) == 1]<- paste(0, 0, start.day[nchar(start.day) == 1], sep='')

end.day<- end.date$yday

end.day[nchar(end.day) == 2]<- paste(0, end.day[nchar(end.day) == 2], sep='')

end.day[nchar(end.day) == 1]<- paste(0, 0, end.day[nchar(end.day) == 1], sep='')

MODIS.start<- paste('A', substr(start.date, 1, 4), start.day, sep='')

MODIS.end<- paste('A', substr(end.date, 1, 4), end.day, sep='')

}

#####
# Get the MODIS Web Service Description Language and set up SOAP-Client interface.

ornlMODIS = processWSDL('http://daac.ornl.gov/cgi-
bin/MODIS/GLBVIZ_1_Glb_subset/MODIS_webservice.wsdl')

ornlMODISFuncs = genSOAPClientInterface(operations=ornlMODIS@operations[[1]], def=ornlMODIS)

##### Retrieve the list of date codes to be requested and organise them in batches of
time series' of length 10.

dates = ornlMODISFuncs@functions$getdates(lat.long[1,2], lat.long[1,3], Product)

#####
# Use the getsubset function as described (http://daac.ornl.gov/MODIS/MODIS-
menu/modis_webservice.html) to retrieve data subsets for each location of a set of product
bands, at a defined surrounding area, saving the data for each location into separate ascii
files in /pixels dir in the working directory.

for(i in 1:nrow(lat.long)) { # Loop set up to make request and
write a subset file for each location.

start.dates<- which(dates >= MODIS.start[i]) # Finds the start date and end date
specific for each subset.

end.dates<- which(dates >= MODIS.end[i])

date.res<- start.dates[which(start.dates %in% end.dates == FALSE)]

options(warn=-1); date.list<- matrix(dates[date.res], nrow=10); options(warn=0)

result<- list(NA)

subsets<- c()

print(paste('Getting subset for location ',i,' of ',nrow(lat.long),'...', sep=''))

for(n in 1:length(Bands)) { # Loop for each band specified, to be requested
individually and then dropped into subsets.

if(ncol(date.list) > 1) { # Stops (ncol(date.list)-1) = 0 in the for loop.

for(x in 1:(ncol(date.list)-1)) { # getsubset function return object of ModisData
class, with a subset slot that only allows 10 elements (i.e. 10 dates), looped until all
requested dates have been retrieved.

result[[n]] = try(ornlMODISFuncs@functions$getsubset(lat.long[i,2], lat.long[i,3],
Product, Bands[n], date.list[1,x], date.list[10,x], Size[1], Size[2]))

if(class(result[[n]]) == 'try-error'){

timer<- 1

```

```

while(timer <= 50){

    print(paste('Connection to the MODIS Web Service failed: trying again in
30secs...attempt ',timer,sep=''))

    Sys.sleep(30)

    result[[n]] = try(ornlMODISFuncs@functions$getsubset(lat.long[i,2],
lat.long[i,3], Product, Bands[n], date.list[1,x], date.list[10,x], Size[1], Size[2]))

    timer<- timer+1

    ifelse(class(result[[n]]) == 'try-error', next, break)

}

ifelse(class(result[[n]]) == 'try-error', print('Connection to the MODIS Web
Service failed: Subset requested timed out after 10 failed attempts...stopping subset
download.'), break)

stop(result[[n]])

}

subsets<- as.vector(c(subsets, result[[n]]@subset)) # Stores retrieved data.

}

}

result[[n]] = try(ornlMODISFuncs@functions$getsubset(lat.long[i,2], lat.long[i,3],
Product, Bands[n], date.list[1,ncol(date.list)], # Final batch of dates, finishes at
end.date

date.list[which(date.list[,ncol(date.list)] >= dates[max(date.res)]),
ncol(date.list)], Size[1], Size[2]))

if(class(result[[n]]) == 'try-error'){

timer<- 1

while(timer <= 50){

print(paste('Connection to the MODIS Web Service failed: trying again in
30secs...attempt ',timer,sep=''))

Sys.sleep(30)

result[[n]] = try(ornlMODISFuncs@functions$getsubset(lat.long[i,2],
lat.long[i,3], Product, Bands[n], date.list[1,ncol(date.list)], # Final batch of dates,
finishes at end.date

date.list[which(date.list[,ncol(date.list)] >= dates[max(date.res)]),
ncol(date.list)], Size[1], Size[2]))

timer<- timer+1

ifelse(class(result[[n]]) == 'try-error', next, break)

}

ifelse(class(result[[n]]) == 'try-error', print('Connection to the MODIS Web Service
failed: Subset requested timed out after 10 failed attempts...stopping subset download.'),
break)

stop(result[[n]])

}

if(regexpr('Server is busy handling other requests in queue. Please try your subset
order later.We apologize for the inconvenience', result[[n]]@subset[1]) != -1) {

stop('Server is busy handling other requests in queue. Please try your subset order
later.')

```

```

    }

    subsets<- as.vector(c(subsets, result[[n]]@subset)) # Stores all useful retrieved data.
}

if(length(subsets) != (length(date.res)*length(Bands))){

  ifelse(StartDate == TRUE, lat.long[i,6]<- 'Missing data in subset: try downloading
again', lat.long[i,5]<- 'Missing data in subset: try downloading again')

  print(paste('There is missing information in the subset downloaded for time-series
',lat.long$ID[i],'. See subset download file.',sep=''))

  } else {

    ifelse(StartDate == TRUE, lat.long[i,6]<- 'Successful download', lat.long[i,5]<-
'Successful download')

  }

  if(Transect == FALSE){ write(subsets, file=paste(SaveDir,lat.long[i,1],'_',Product,'.asc',
sep=''), sep='') } # Writes an ascii file for time-series into the wd.

  if(Transect == TRUE){

    if(i == 1){ write(subsets, file=paste(SaveDir,lat.long[i,1],'_',Product,'.asc', sep=''),
sep='') }

    if(i != 1){ write(subsets, file=paste(SaveDir,lat.long[i,1],'_',Product,'.asc', sep=''),
sep='', append=TRUE) }

  }

  if(i == nrow(lat.long)) { print('Full subset download complete. Writing the subset
download file...') }

}

if(Transect == FALSE){ write.table(lat.long, file=paste(SaveDir,'Subset Download
',Sys.Date(),'.csv',sep=''), col.names=TRUE, row.names=FALSE, sep=',') } # Writes
an ascii file with all dates for each band at a given location into the working directory.

if(Transect == TRUE){

  ifelse(SaveDir == NULL, DirList<- list.files(), DirList<- list.files(path=SaveDir))

  if(any(DirList == paste(SaveDir,'Subset Download ',Sys.Date(),'.csv',sep='')) == FALSE){
write.table(lat.long, file=paste(SaveDir,'Subset Download ',Sys.Date(),'.csv',sep=''),
col.names=TRUE, row.names=FALSE, sep=',') }

  if(any(DirList == paste(SaveDir,'Subset Download ',Sys.Date(),'.csv',sep=''))){
write.table(lat.long, file=paste(SaveDir,'Subset Download ',Sys.Date(),'.csv',sep=''),
col.names=FALSE, row.names=FALSE, sep=',',append=TRUE) }

}

if(Transect == FALSE){ print('Done! Check the subset download file for correct subset
information and any download messages.') }

}

```