Project 6: Phenological Diversity Trends By Remote Sensing Related Datacubes

Authors: E. Shayle (1), M. Marcantonio (2), C. Richiardi (3), R. Labadessa (4), & S. Vicario (4)

1: Environmental Informatics, University of Marburg, Deutschhausstr. 12 - 35032, Marburg, Germany

2: Joint European Research Center (Freelancer scientist), Ispra, Italy

3: ENEA, Strada per Crescentino, 13, Turin, Italy

4: CNR - IIA, Via Orabona, 4, Bari, Italy

Keywords: Rao's Q Index, Time-Weighted Dynamic Time Warping, Landscape Heterogeneity, Remote Sensing, Time Series, & Phenology

Introduction:

The B3 Hackathon brought together informaticians from a variety of institutions to rapidly create novel informatics solutions to the biodiversity challenges facing the planet. We identified that the addition of time-weighting to the R package "rasterdiv" would be a worthwhile contribution to the environmental informatics community. Rasterdiv was created to calculate diversity indices with data of the class "raster layer". Biodiversity indexes commonly focus on the spatial component. Here we outline how our extention to the pre-existing implementation of Rao's diversity indices [@Rocchini2017] can account for the temporal dimension of data, alongside the relevant biological context to our extension.

The Importance of Biodiversity Indices:

Heterogeneous ecosystems have been shown both experimentally and theoretically to provide greater utility to all the agents which comprise that ecosystem. This is through the provision of more and more varied niches for flora and fauna to propagate. This subsequently increases the value of ecosystem services provided to the communities surrounding an ecosystem. Heterogeneous ecosystems are typically also more resilient to disturbances they experience. Due to the centrality of biodiversity to healthy ecosystem functioning, quantitative measures of biodiversity are required to understand how ecosystems are responding to ongoing environmental changes, such as shifting land use.

Shannon's H value has been widely used as a proxy for biodiversity, but can be inadequate when applied to the new kinds of data generated by remote sensing platforms (e.g. images from Earth observation satellites). To create quantified data from ecosystems, most analytical approaches assess discrete points within the ecosystem, such as those from a quadrat, or pixels in the case of aerial remote sensing datasets. One limitation is that Shannon's H value is that it does not consider the distance between each sampled point (whether they are species, pixel, or any other quantitative abstractions of an observation). This approach treats all objects within a dataset as equally distant from one another.

Rao's Quadratic Diversity Index (Rao's Q) adds space as a trait to its abstraction of biodiversity by accounting for the distance between observations within a study site. As a spatially informed alternative to Shannon's H, Rao's Q has been demonstrated experimentally to offer greater efficacy when representing biodiversity in aerial remote sensing datasets [@Rocchini2021], for which pixels are the discrete observation units. However, Rao's Q remains limited by its inability to assess trait change over time. Current implementations of the index only assess one snapshot of the data at a time. We set out to overcome this limitation by incorporating Time-Weighted Dynamic Time Warping (TWDTW) to include time as a component of the distance variable within Rao's Q.

The Purpose of (Time-Weighted) Dynamic Time Warping & its Ecological Utility:

Dynamic Time Warping (DTW) is a mathematical approach used to compare data series when the timing of observations differs. It has been used in a variety of disciplines. DTW works by finding the smallest distance between two time series.

However, by flattening the differences in timing, biologically significant differences can also be obscured, such as when comparing plant phenology. For instance, many tree species require a minimum number of Growing Degree Hours (GDH) to commence their springtime budburst [@Fu2019]. Other ecosystem processes typically need to coincide with phenological events, so phenology timing represents an important differentiating factor for time series representing ecosystems with plants.

The TWDTW approach rectifies this by including a cost to aligning pixels with greater temporal separation. Therefore, the TWDTW function is less likely to match the time series to others which exhibit substantially different phenologies. This has been successfully demonstrated by [@Maus2016] to classify changing land use patterns in the Brazilian Amazon, and was a more effective tool than standard DTW when applied to heterogeneous biological environments like these.

Equation:

$$\omega_{i,j} = \frac{1}{1 + e^{-\alpha(g(t_i,t_j)-\beta)}}$$

Reproduced from Maus [@Maus2016]. In addition to the standard cost matrix of the DTW function, they also propose the equation above to implement a temporal cost. In the equation α is the steepness of the logistic function used for penalisation of time distance, and β is the midpoint of the curve. Lastly, $g(ti,tj)$ represents the time elapsed between the dates (ti in the original pattern, and tj in the time series).

In this manuscript, we used optical aerial remote sensing data derived from a small, grazed grassland site in Calabria, Italy to demonstrate and evaluate our R-based implementation of phenology into Rao's Q index. We also evaluate its efficacy in comparison to Shannon's H and unmodified Rao's Q indices.

Results:

implementation in rasterdiv

We implemented this method within the existing `paRao()` function of the rasterdiv R package. We used the `twtwd` function from the `twdtw` R package [@Maus2019]. This package uses C++ to compute the TWDTW.

The resulting implementation of our code is as follows: `paRao(x=time.series, time_vector=time, window=11, alpha=1, na.tolerance=0, method="multidimension", dist_m="twdtw", simplify=4, np=8)`

The arguments and our input parameters of which are:

`x` An `(X,Y,Z)` raster stack (or cube) of spectral data, where the X and Y axes represent discrete pixel values, and each layer of the Z axis is a a different temporal snapshot of the raster layer. In our study, this is the Sentinel derived time series of our study site in Calabria.

`time_vector` A vector of dates corresponding to every point in the raster time series, which must be the same as the `Z` axis from the `x` variable. All pixels in the input time series must share the same temporal spacing as the temporal pattern to which it is being compared (i.e. if the time series has observations on days `c(1, 3, 7, ...)`, then the pattern it is being compared to must also have observations on days `c(1, 3, 7, ...)`.

`steepness` A numeric value corresponding to the α variable from the time-weighting function in Maus [@Maus2016]. Lower or higher values of α ...increase or decrease?... penalisation for deviations from the pattern time.

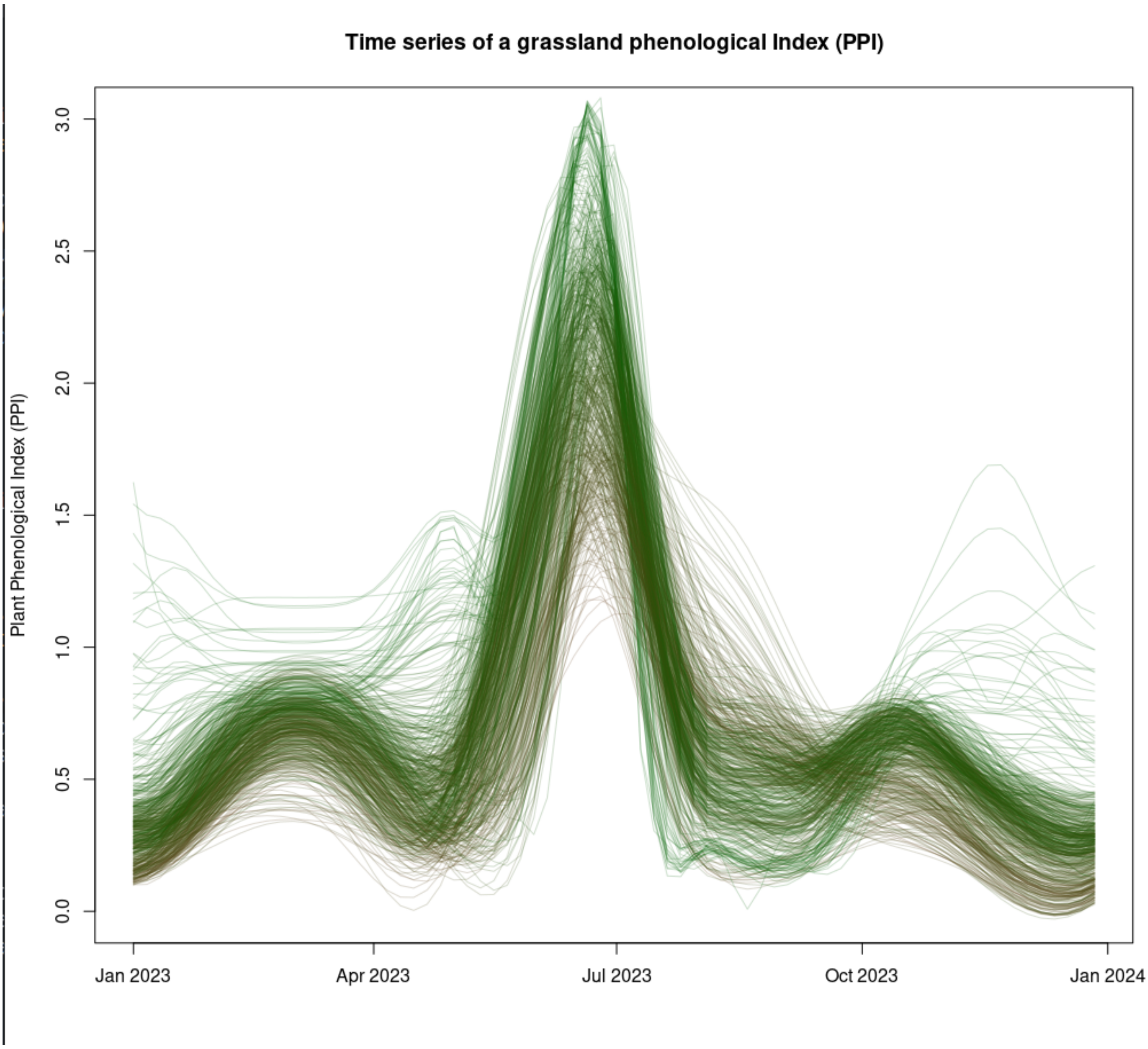
`midpoint` A numeric value corresponding to the β variable from the time-weighting function in Maus [@Maus2016]. The input data must be of the unit specified by the `time_scale` argument.

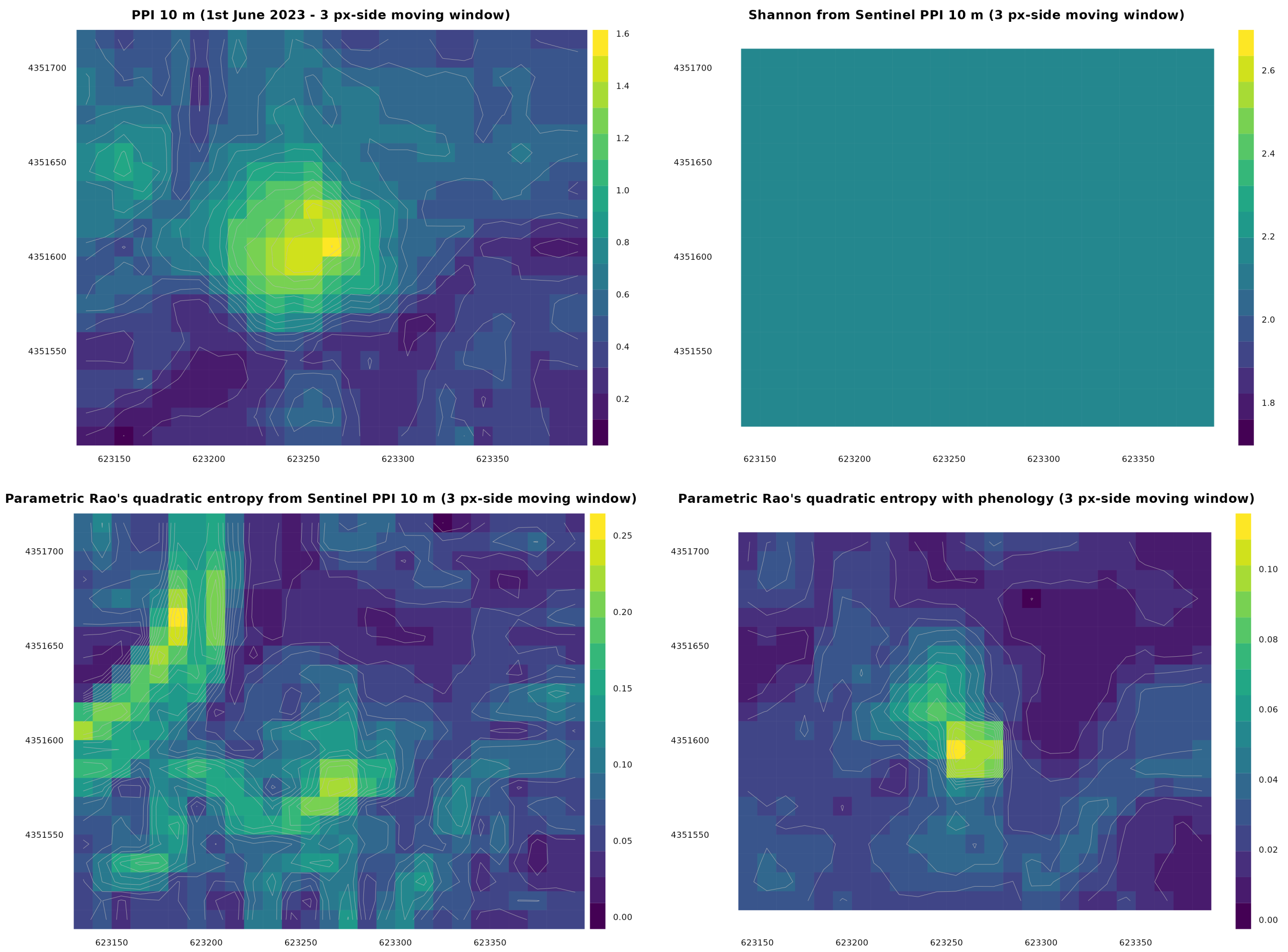
`cycle_length` A string value. Valid input arguments are

`time_scale`

Other arguments remain unchanged.

Case Study:


Time series of a grassland phenological Index (PPI)
Plant Phenological Index (PPI)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan 2023
Apr 2023
Jul 2023
Oct 2023
Jan 2024



Discussion:

GitHub and Data Repositories:

Acknowledgements:

References: