---

title: 'Project 6: Phenological Diversity Trends By Remote Sensing Related Datacubes'

tags:

- Rao Q Index

- Time-Weighted Dynamic Time Warping

- Landscape Heterogeneity

- Remote Sensing

- Time Series

authors:

- name: First Last

orcid: 0000-0000-0000-0000

affiliation: 1

- name: Second Last

orcid: 0000-0000-0000-0000

affiliation: 2

affiliations:

- name: Institution 1, address, city, country

index: 1

- name: Institution 1, address, city, country

index: 2

date: 03 April 2024

bibliography: paper.bib

authors\_short: Last et al. (2021) BioHackrXiv template

group: B-Cubed

event: B-Cubed Hackathon 2024 - Hacking biodiversity data cubes for policy

editor\_options:

markdown:

wrap: 72

---

# Introduction:

The R package "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 an extention to our

implementation of Rao's diversity indices to account for the temporal

dimension of data, alongside the relevant biological context.

## 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, probably due to functional redundancy. 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 [Explain what a pixel is and how it can be any sort of

entity] 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 [@Rocchini:2021], 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 [@Fu:2019]. 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

[@Maus:2016] 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:

![TWDTW Equation from Maus

2016](.TWDTW%20Equation%20from%20Maus%202016.png)

Reproduced from Maus [@Maus:2016]. In addition to the standard cost matrix of the DTW function, they also apply the above equation to implement a temporal cost. In the above equation $α$ is the steepness of the logistic function used for penalisation of time distance, and $β$ is the midpoint, and lastly, $g(ti,tj)$ represents the time elapsed between the dates evaluated in the match ($ti$and $tj$ times of the ith and jth observations ).

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 [@Maus:2023]. 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 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.

`steepness` An continuous corresponding to the $α$ variable from the time-weighting function in Maus [@Maus:2016]. Lower or higher values of $α$ ...increase or decrease?... penalisation for deviations from the pattern time.

`midpoint` An integer corresponding to the $β$ variable expressed in day from the time-weighting function in Maus [@Maus:2016]. The input data must be of the scale specified by the `time\_scale`

cycle\_length= "year", time\_scale="day"

## case study:

The small patch of 5 hectares within Macchia Sacra, Special Protection Area, was selected thanks to the availability of a detailed drone flight that could be used to define a ground truth for biodiversity. Thanks to expert knowledge, guided by classification, we defined 8 community type areas. The area is characterized on the northeast part of road. From the level of the road the elevation declines to a lower part that cut from south to west were a small stream pass. This area is characterized by hydrophile vegetation. Between these two extremes sit a shoulder with on a top a small flat patch. This patch is the resting area of the local cow herd. This area is much dryer and subject to strong pasture pressure and mechanical disruption, but greater amount of nutrients.

## validation phase

We used 144 sentinel2 images from HRVPP of Phenological Plant Index (PPI) covering all 2023 collecting an image 20 by 27 pixels. The PPI is optimized to be least influenced by soil signal and problem of shading (ref). Using this starting data we applied 3 biodiversity analysis approaches: a Shannon index applied on the mean yearly value with 3 significant digits of the PPI trajectory; a Rao Q index with different alpha, applied on the same dataset and the Rao Q with the addition of the TWDTW function on the full 144 time-depth data set.

Looking a figure x is possible to see that Shannon was not working given that using 3 digits all pixels had different mean values and is not possible to identify grouping. Simple Rao Q correctly identified the main diversity spot on the top of the shoulder and as secondary spot the side of the road. Notice that Rao Q gives does not change, changing alpha given that all pixels are different. Our new distance has two main differences with classical Rao Q, the road is not anymore, a secondary hotspot, and the main spot position moved on the border between two community identified by our expert opinion. # Discussion:

# GitHub and Data Repositories:

# Acknowledgements:

# References: