

Phenological Diversity Trends With Remote Sensing Datacubes

03 April 2024

*Corresponding author: marcantonio.matteo@gmail.com

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

Introduction:

The B-Cubed Hackathon brought together computer scientists and ecologists 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 biodiversity indices from numerical matrices in the R programming environment.

Biodiversity indices included in **rasterdiv** currently focus on the spatial component. Here we outline how our extension to the pre-existing implementation of Rao's Q diversity indices (Rocchini, Marcantonio, and Ricotta 2017) can explicitly account for the temporal dimension of data, alongside the relevant biological context to our extension.

The Importance of Biodiversity Indices:

Ecosystems with heterogeneous biota have been shown both experimentally and theoretically to provide greater utility to all the agents which comprise that ecosystem (Zhang et al. 2022). This is through the provision of more resources and a greater variety of niches available for species. This subsequently increases the value of ecosystem services provided to the people in communities surrounding an ecosystem. Heterogeneous ecosystems are typically also more resilient to disturbances they experience, likely because they have functional redundancy (Mace, Norris, and Fitter 2012). 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 patterns and climate change.

Novel satellite remote sensing tools generate large amounts of data about the Earth's surface at increasing rate, due to their almost constant temporal coverage and increasing spatial resolution (Frazier and Hemingway 2021). The “big data” generated by these systems need to be interpreted effectively to accurately detect and describe ecosystem trends, such as a change in ecosystem diversity. Consequently, information theory has been used to build the analytical tools which are now regularly used to assess remote sensing datasets. For example, Shannon's H index, and other related indices only based on mathematical transformation of frequency data, have been widely used as a proxy for biodiversity. But these can be inadequate when applied to time series data generated by remote sensing platforms. In fact, these indices do not consider the distance between each sampled point (whether they are species incidences, pixels, or any other quantitative

abstractions of an observation). This approach therefore treats all objects within a dataset as equidistant from one another, so measures of Shannon's H value are prone to saturation, especially when only marginal differences are observed between the objects in a remote sensing dataset.

Rao's Quadratic Diversity index (Rao's Q) adds "space" (but not necessarily geographical space) as a trait to its abstraction of biodiversity by accounting for the distance between traits of the observations within a study site. As a trait 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 et al. 2021), for which pixels are the discrete observation unit. However, Rao's Q remains limited by its inability to assess trait change over time. Current implementations of the index (for example in the **rasterdiv** R package (Rocchini et al. 2021)) 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 minimising 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 et al. 2019). Many other ecosystem processes, such as seasonal migrations and pollination, need to coincide with phenological events, so phenology timing represents an important differentiating factor for time series representing ecosystems with plants (Fitchett, Grab, and Thompson 2015).

The TWDTW approach rectifies this by imposing a "cost" when aligning observations (which in our study are pixels) with greater temporal separation. Therefore, the TWDTW function is less likely to match a time series to others which exhibit substantially different phenologies. This has been successfully demonstrated by Maus *et al* (V. Maus et al. 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.

In addition to the standard cost matrix of the DTW function, they also propose a method to implement a cost due to temporal separation of temporal trends (Equation 1). In Equation 1, α is the steepness of the logistic function which is used to impose a distance penalty on temporal separation, and β is the midpoint of the curve, the threshold below which separation in time does not substantially impact the other components of the equation. Lastly, $g(t_i, t_j)$ represents the time elapsed between the dates evaluated in the match (t_i and t_j times of the i th and j th observations).

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

In this manuscript, we used Copernicus SENTINEL-2 mission's optical data of a small, grazed grassland site in Calabria, Italy to demonstrate and evaluate our R-based **rasterdiv** implementation of phenology into Rao's Q index. We also evaluated its efficacy in comparison to the Shannon's H and unmodified Rao's Q indices.

Case Study & Results:

Implementation within rasterdiv:

We implemented this method within the existing `paRao()` function of the `rasterdiv` R package (Marcantonio, Thouverau, and Rocchini 2024). We used the `twtwd` function from the `twdtw` R package (Victor Maus 2023) which is a wrapper for C++ implementations of TWDTW functions. In particular, the `twtwd` method was implemented through the internal hidden function `.mtwtdtw`:

```
.mtwtdtw <- function(x, time_vector = 0, stepness = -0.5, midpoint = 35,
cycle_length = "year", time_scale = "day") {
  twdtw::twdtw(
    x = data.frame(time = time_vector, v = x[[1]]),
    y = data.frame(time = time_vector, v = x[[2]]),
    time_weight = c(stepness = stepness, midpoint = midpoint),
    cycle_length = cycle_length,
    time_scale = time_scale)
}
```

stored in the `rasterdiv` file `mdist.R`. The `paRao()` function uses the "`distm_m`" argument to internally call this function, applying it to pairwise pixel time series (`Z`). Within each moving window, Rao's Q index is calculated using the computed distances and the specified `alpha` value. By leveraging the TWDTW distance over a time-series of matrices or images, Rao's Q index can be calculated using the following R function:

```
paRao(x = time.series,
       time_vector = time,
       window = 3,
       alpha = 2,
       na.tolerance = 0,
       simplify = 3,
       method = "multidimension",
       dist_m = "twdtw",
       midpoint = 35,
       stepness = -0.5
     )
```

The arguments and our input parameters of which are:

`time.series` is 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 (with daily `Z` temporal resolution) of our study site.

`time_vector` A vector of dates corresponding to every `Z` layer 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, ...)`). If any `x, y` pixel `i` along the `Z` temporal vector has any invalid numeric data (e.g., `NA`, `NULL`, `Inf`, etc.), its Rao's Q index value will be `NA`, since DTW does not allow for no data values.

`steepness` A continuous numeric value corresponding to the α variable from the time-weighting function in Maus *et al* (V. Maus et al. 2016). Different values of α can 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 *et al* (V. Maus et al. 2016). The input data must be of the unit specified by the `time_scale` argument (e.g. in our example, it is expressed in days).

`cycle_length` This argument indicates the length of a single cycle. This argument can accept either a string or numeric value. Valid string input values are “year”, “month”, “day”, “hour”, “minute”, and “second”. String inputs are also passed to the `time_scale` argument. Numeric input values must be the same unit specified by the `time_scale` argument.

`time_scale` This argument sets the units of the TWDTW function’s cycle length. Valid string input values are “year”, “month”, “day”, “hour”, “minute”, and “second”. If the input given to `cycle_length` is a string value, then this argument will change the units given in the output. Alternatively, if a numeric value is given to `cycle_length`, then this argument will compute the elapsed time in seconds.

Other arguments remain unchanged from (Victor Maus 2023).

Study Site Description & Plant Biodiversity Data:

The study site was a small (6 hectare) area within the protected area of “Macchia Sacra” (Site of Community Importance (SCI): SICIT9310073). It was selected as it was suitable for thorough imaging by drone, which formed the basis of our ground-truthed biodiversity observations. After drone image acquisition, we defined eight types of plant communities within the study site, with the expertise in classification imparted by an expert botanist (Figure 1). The study site is characterized by the presence of a road on the north-east part of the site. From the level of the road the elevation declines to a lower part which features a sharp canyon running south to west, the result of a previous small stream which had dried up by the time of our drone survey. This part of the study site is characterized by hydrophilic vegetation. Between these two extremes is a small hill which culminates in a plateau. The plateau is the resting area of a herd of cows which graze in the area. This area is much dryer and subject to strong pasture pressure and mechanical disruption, but is more nutrient rich, due to the presence of cow manure.

Efficacy Evaluation of our Results:

We used a time series of 144 Plant Phenological Index (PPI) images derived from SENTINEL-2 imagery and available through the Copernicus Service High Resolution Vegetation Phenology and Productivity (HRVPP) (data acquired via: <https://scihub.copernicus.eu/>) encompassing all data which were available in 2023 (Figure 2). The dimensions of each image were 20 pixels vertically by 27 pixels horizontally, and each pixel was 10m². The PPI index was chosen as it is minimally influenced by soil signal and the presence of shadows (Karkauskaite, Tagesson, and Fensholt 2017).

Using these data, we applied three analytical approaches to measure biodiversity: i) The Shannon’s Biodiversity index applied to the mean yearly PPI trajectory, ii) the classical Rao’s Q index, applied to the same dataset and with three digits numeric resolution, and iii) the Rao’s Q index with our implementation of the TWDTW function across the full time series of 144 images. A gross visual inspection of Figure 3 illustrates the unsuitability of Shannon’s H index, as when using a 3 pixel wide moving window, all pixels inside each window exhibited different values and the index was always equal to its maximum (i.e., $\ln(S)$). It was therefore impossible to classify the ecosystem into different groups. The standard Rao’s Q index identified the main biodiversity hotspot as where the road intersects with the study site, and a secondary hotspot as the plateau atop the hill. Finally, our new implementation of Rao’s Q index with the DTW distance metrics resulted in two meaningful differences from the standard Rao’s Q index: the road is no longer a “diversity” hotspot, and the main biodiversity hotspot moved to the borders between two of the communities identified by our expert.

Discussion:

In this Hackathon, we developed a streamlined method for implementing the TWDTW algorithm to calculate Rao’s Quadratic Diversity index within the `rasterdiv` R package. This addition introduces a solid

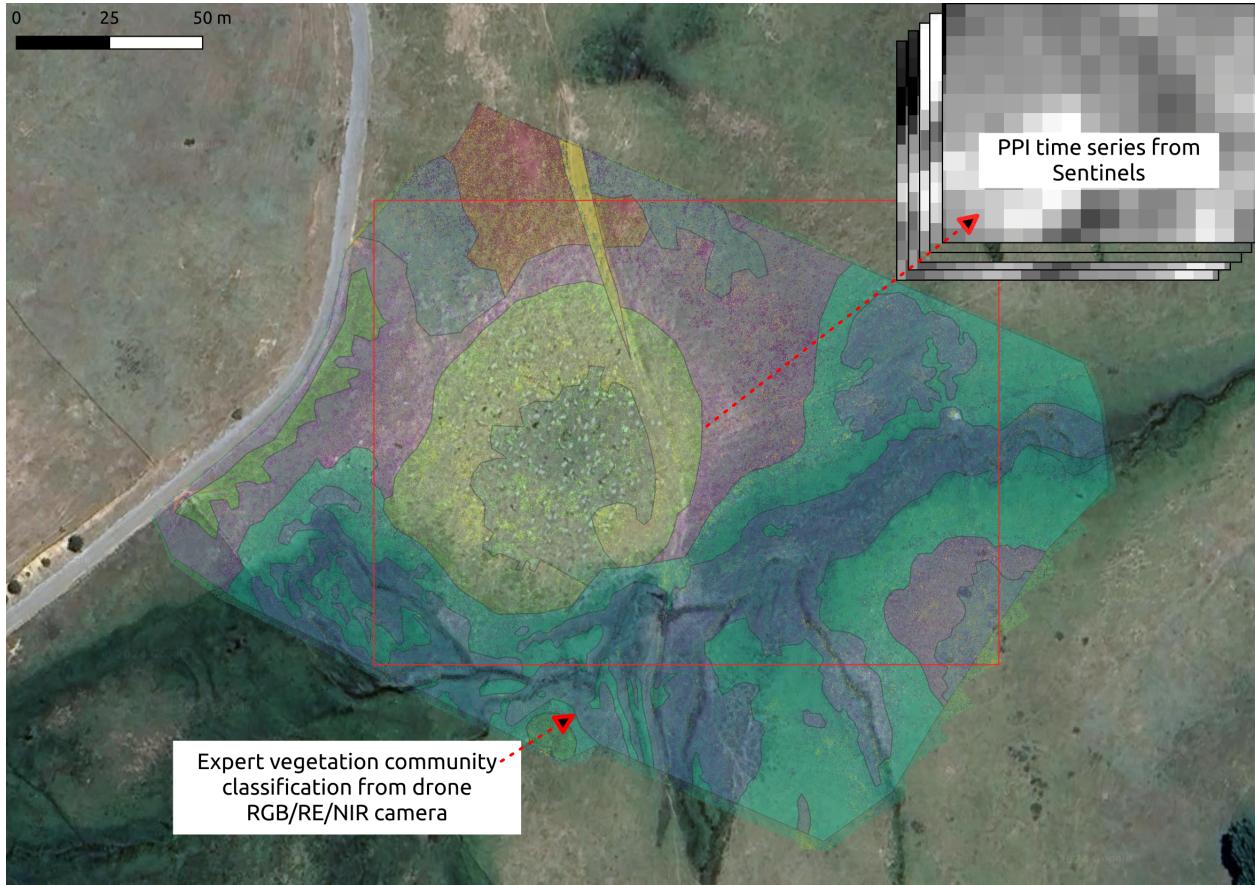


Figure 1: A true colour RGB satellite image (Map data: Google, Maxar Technologies 2023) giving an overview of the study site and its surroundings within Calabria. The eight different plant community types are overlaid as differently coloured masks in transparency upon the image. The subset of the study site which was extracted from SENTINEL-2 data is indicated within the red rectangle.

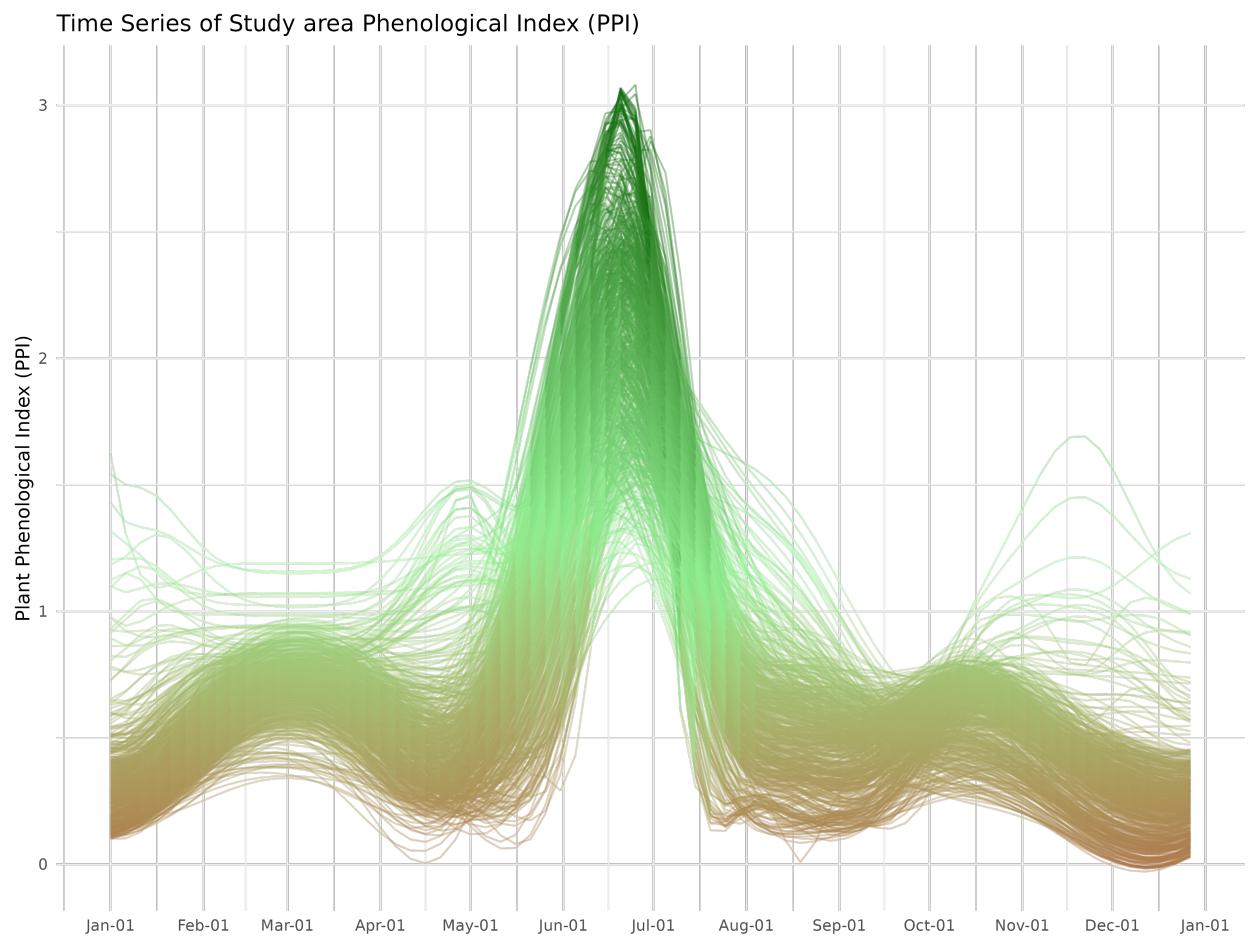


Figure 2: A time series illustrating changes in the value of the Plant Phenological Index (PPI) for every pixel within the Macchia Sacra Special Protection Area study site in Calabria.

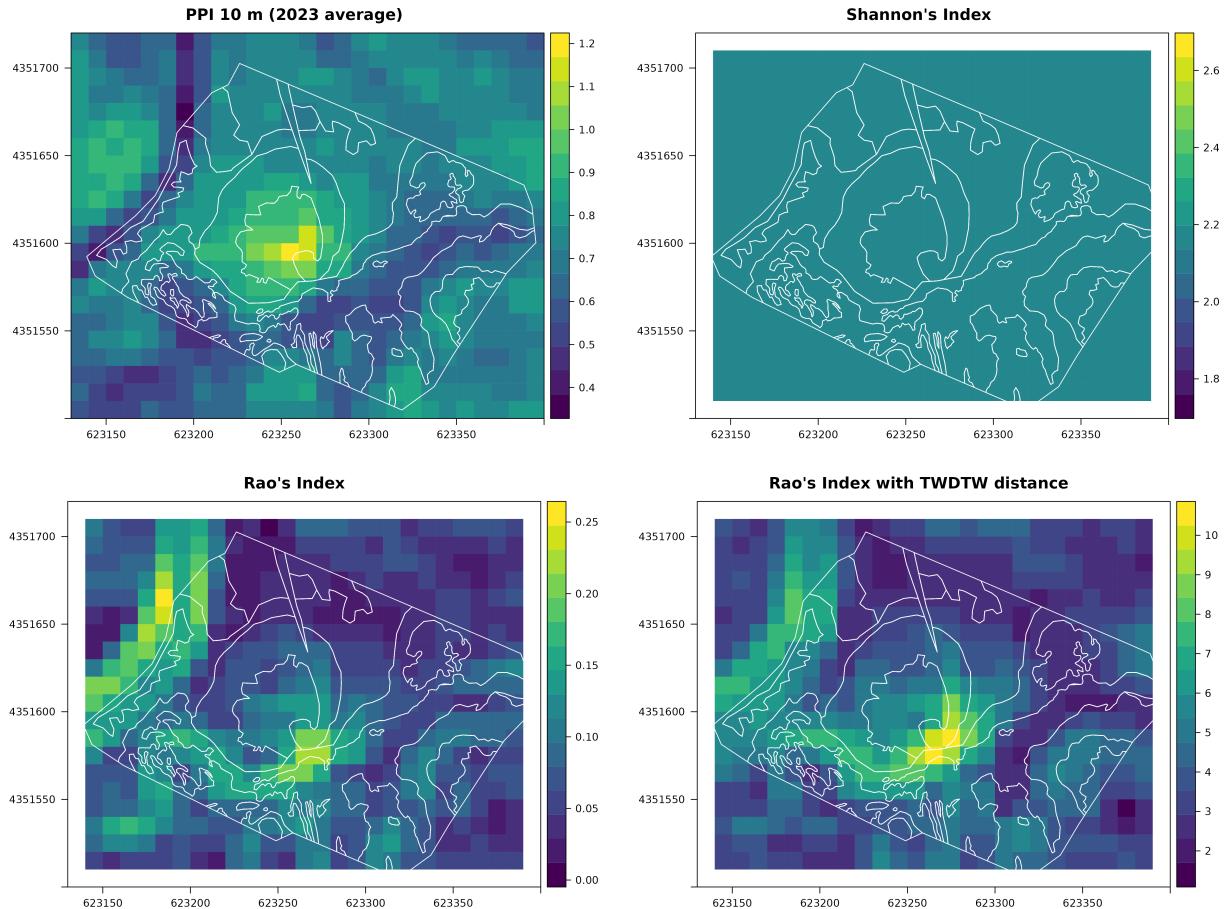


Figure 3: A four panel plot comparing the efficacy of different diversity indices at measuring biodiversity within our grassland ecosystem in Calabria. The white contour lines represent different plant communities classified by an expert botanist from the drone high-resolution image. The scale bar to the right of each plot indicates the assessed biodiversity value in the index being tested (e.g. Shannon's H index or Rao's Quadratic Diversity index).

method to account for the temporal dimension to the traditional spatial analysis of landscape diversity. Recognising the dynamic nature of plant communities and ecosystems over time, our method integrates phenological variation into diversity assessments derived from satellite imagery. Notably, our case study found that when this technique was applied to multiband remotely sensed data from disturbed grasslands, accounting for phenological cycles can improve diversity indices by filtering out artefacts. For instance, it can help to distinguish between semi-natural habitats and artificial land cover types, like roads, which lack temporal phenological shifts. These artificial features tend to form clusters of minimal DTW distances when considering DTW as an inter-voxel distance, leading to lower values of Rao's Q which more accurately reflect their level of biodiversity.

In healthy, biodiverse ecosystems, a variety of flora are also likely to be present to fully utilise the various niches present in the landscape. As Figure 2 from our case study demonstrated, phenology trends can differ substantially within an ecosystem to reflect its biodiversity. PPI rose sharply across the whole site before reaching a peak ranging from approximately 1 to 3 in June, after which point, the site wide PPI drops sharply, mostly remaining between 0.2 and 1 for the rest of the year. As this sharp rise indicates, any analysis based on a single temporal snapshot is likely to be unrepresentative of how the ecosystem in question functions throughout the year. This primary spike in PPI also presents the widest observed range in PPI values, which enables biodiversity indices to better classify the different types of floral communities present. Furthermore, Figure 2 highlights two interesting secondary PPI peaks: one in spring prior to the primary spike in PPI, and the other beginning in October and continuing throughout the winter. Though neither peak is as consistent site wide nor presents the same magnitude of increase in PPI, they are nevertheless ecologically significant. These secondary peaks likely represent floral communities which are exploiting different temporal niches, and like an oasis in the desert, likely also create unique niches to support a more biodiverse ecosystem. Again, an analytical approach which does not consider phenology would misrepresent the functioning of the ecosystem. Additionally, using a traditional DTW approach to consider phenology would erase ecological detail, as the smaller secondary peaks would have been warped to match the rest of the site. Thus, our approach using TWDTW can provide a valuable new level of insight into ecosystem functioning and biodiversity.

Remote sensing via Earth observation remains a rapidly developing area of scientific research with wide applicability to conservation practice (Pettorelli, Safi, and Turner 2014). As the technologies continue to mature, the possibilities remote sensing approaches like ours offer continues to grow. For instance, Schulte to Bühne *et al* (Schulte to Bühne et al. 2022) recently demonstrated the power of Earth observation satellites to assess rewilding efforts at the landscape scale. Rewilding programmes are notoriously challenging technically, and are often expensive, too. Thus, the ability to assess the progress of rewilding inexpensively at the landscape scale via satellite is eminently valuable. However, as Maus *et al* (V. Maus et al. 2016) observed, markedly different land use types, such as soy bean plantations and primary rainforest, can be erroneously classified as the same via an NDVI based classification protocol which does not consider phenology. Multiple studies (V. Maus et al. 2016) (Lopes et al. 2020) also found that accounting for the effects of phenology significantly increased the accuracy of land use classifications. As rewilding and reforestation programmes continue, our implementation of TWDTW reduces the barrier for conservationists trying to assess the temporal dimension of diversity in addition to the spatial dimension. Our case study further demonstrated that incorporation of phenological data enhances the inferential power of analyses, as this dimension more accurately identified biodiversity hotspots than Rao's Q index alone (Figure 3). Since increasing biodiversity is typically a core goal of rewilding programmes, of which land cover diversity is a core component (Skidmore et al. 2015), more accurate classification of landscape diversity can highlight where efforts are succeeding or where further efforts are needed.

Conclusion

By incorporating temporal dynamics into the `paRao()` function of the `rasterdiv` R package, we broaden the scope for analysing remotely sensed time series. This advancement enriches the suite of diversity indices obtainable from remote sensing data, enhancing our understanding of landscape heterogeneity, and in so doing, enhancing our ability to conserve it.

GitHub & Data Repositories:

This manuscript, previous revisions, open source data, and scripts can all be found on the open source GitHub repository “Samuel-Green/B-3-Hackathon-Project-6” via the URL: <https://github.com/Samuel-Green/B-3-Hackathon-Project-6>. The implementation of the Rao’s Q index described in this manuscript can be found via the URL: https://github.com/mattmar/rasterdiv/tree/bcubed_hackathon

Acknowledgements:

We thank:

- Dr. Quentin Groom of Meise Botanic Garden for co-ordinating the project.
- Laura Abraham of Meise Botanic Garden for organising the event.
- Prof. Nicodemo Passalacqua (University of Calabria) who kindly shared the ground-truthing data within pilot project 2.3.2 of the Tech4you PNRR project.
- The European Union’s Horizon Europe Research and Innovation Programme (ID No 101059592) for funding the B-Cubed programme, and thus, this event.

References:

- Fitchett, Jennifer M, Stefan W Grab, and Dave I Thompson. 2015. “Plant phenology and climate change: Progress in methodological approaches and application.” *Progress in Physical Geography: Earth and Environment* 39 (4): 460–82. <https://doi.org/10.1177/0309133315578940>.
- Frazier, Amy E., and Benjamin L. Hemingway. 2021. “A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery.” *Remote Sensing* 13 (19). <https://doi.org/10.3390/rs13193930>.
- Fu, Yongshuo H, Shilong Piao, Xuancheng Zhou, Xiaojun Geng, Fanghua Hao, Yann Vitasse, and Ivan A Janssens. 2019. “Short photoperiod reduces the temperature sensitivity of leaf - out in saplings of *Fagus sylvatica* but not in horse chestnut,” no. February: 1696–1703. <https://doi.org/10.1111/gcb.14599>.
- Karkauskaite, Paulina, Torben Tagesson, and Rasmus Fensholt. 2017. “Evaluation of the Plant Phenology Index (PPI), NDVI and EVI for Start-of-Season Trend Analysis of the Northern Hemisphere Boreal Zone.” *Remote Sensing* 9 (5): 485.
- Lopes, Mailys, Pierre-Louis Frison, Sarah M. Durant, Henrike Schulte to Bühne, Audrey Ipavec, Vincent Lapeyre, and Nathalie Pettorelli. 2020. “Combining Optical and Radar Satellite Image Time Series to Map Natural Vegetation: Savannas as an Example.” *Remote Sensing in Ecology and Conservation* 6 (3): 316–26. <https://doi.org/10.1002/rse2.139>.
- Mace, Georgina M, Ken Norris, and Alastair H Fitter. 2012. “Biodiversity and ecosystem services: a multilayered relationship.” *Trends in Ecology & Evolution* 27 (1): 19–26. <https://doi.org/10.1016/j.tree.2011.08.006>.
- Marcantonio, Matteo, Elisa Thouverau, and Duccio Rocchini. 2024. “The Rasterdiv Package for Measuring Diversity from Space,” February. <https://doi.org/10.1101/2024.02.07.579266>.
- Maus, V, G Câmara, R Cartaxo, A Sanchez, F M Ramos, and G R de Queiroz. 2016. “A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (8): 3729–39. <https://doi.org/10.1109/JSTARS.2016.2517118>.
- Maus, Victor. 2023. *Twdtw: Time-Weighted Dynamic Time Warping*. <https://CRAN.R-project.org/package=twdtw>.
- Pettorelli, Nathalie, Kamran Safi, and Woody Turner. 2014. “Satellite remote sensing, biodiversity research and conservation of the future.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1643): 20130190. <https://doi.org/10.1098/rstb.2013.0190>.

- Rocchini, Duccio, Matteo Marcantonio, and Carlo Ricotta. 2017. "Measuring Rao's Q diversity index from remote sensing: An open source solution." *Ecological Indicators* 72: 234–38. <https://doi.org/10.1016/j.ecolind.2016.07.039>.
- Rocchini, Duccio, Elisa Thouverai, Matteo Marcantonio, Martina Iannacito, Daniele Da Re, Michele Tornesani, Giovanni Bacaro, et al. 2021. "rasterdiv—An Information Theory tailored R package for measuring ecosystem heterogeneity from space: To the origin and back." *Methods in Ecology and Evolution* 12 (6): 1093–1102. <https://doi.org/10.1111/2041-210X.13583>.
- Schulte to Bühne, Henrike, Bethany Ross, Christopher J Sandom, and Nathalie Pettorelli. 2022. "Monitoring rewilding from space: The Knepp estate as a case study." *Journal of Environmental Management* 312: 114867. <https://doi.org/10.1016/j.jenvman.2022.114867>.
- Skidmore, Andrew K, Nathalie Pettorelli, Nicholas C Coops, Gary N Geller, Matthew Hansen, Richard Lucas, Caspar A Mücher, et al. 2015. "Environmental science: Agree on biodiversity metrics to track from space." *Nature* 523 (7561): 403–5. <https://doi.org/10.1038/523403a>.
- Zhang, Yixin, Zhenhong Wang, Yonglong Lu, and Li Zuo. 2022. "Editorial: Biodiversity, Ecosystem Functions and Services: Interrelationship with Environmental and Human Health." *Frontiers in Ecology and Evolution* 10. <https://doi.org/10.3389/fevo.2022.1086408>.