Ecological Informatics 80 (2024) 102502

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

# GEE_xtract: High-quality remote sensing data preparation and extraction for multiple spatio-temporal ecological scaling

Francesco Valerio [a,c,d,*], Sérgio Godinho [d,e], Ana T. Marques [a,b,c], Tiago Crispim-Mendes [f], Ricardo Pita [f], João Paulo Silva [a,c]

[a] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal

[b] CIBIO – Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa 1349-017, Portugal

[c] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão 4485-661, Portugal

[d] EaRSLab—Earth Remote Sensing Laboratory, University of Évora, Évora 7000-671, Portugal

[e] Institute of Earth Sciences, Universidade de Évora, Rua Romão Ramalho, 59, Évora 7002-554, Portugal

[f] MED – Mediterranean Institute for Agriculture, Environment, and Development & CHANGE – Global Change and Sustainability Institute, Institute for Advanced Studies and Research & Unidade de Biologia da Conservação, Departamento de Biologia, Universidade de Évora, Pólo da Mitra, Ap. 94, Évora 7006-554, Portugal.

A R T I C L E I N F O

*Keywords:*
Multiple spatiotemporal scales
Time series
Sentinel
Landsat
MODIS
Google earth engine

A B S T R A C T

Environmental sensing via Earth Observation Satellites (EOS) is critically important for understanding Earth' biosphere. The last decade witnessed a "Klondike Gold Rush" era for ecological research given a growing multidisciplinary interest in EOS. Presently, the combination of repositories of remotely sensed big data, with cloud infrastructures granting exceptional analytical power, may now mark the emergence of a new paradigm in understanding spatio-temporal dynamics of ecological systems, by allowing appropriate scaling of environmental data to ecological phenomena at an unprecedented level.

However, while some efforts have been made to combine remotely sensed data with (near) ground ecological observations, virtually no study has focused on multiple spatial and temporal scales over long time series, and on integrating different EOS sensors. Furthermore, there is still a lack of applications offering flexible approaches to deal with the scaling limits of multiple sensors, while ensuring high-quality data extraction at high resolution.

We present GEE_xtract, an original EOS-based (Sentinel-2, Landsat, and MODIS) code operational within Google Earth Engine (GEE) to allow for straightforward preparation and extraction of remote sensing data matching the multiple spatio-temporal scales at which ecological processes occur. The GEE_xtract code consists of three main customisable operations: (1) time series imageries filtering and calibration; (2) calculation of comparable metrics across EOS sensors; (3) scaling of spatio-temporal remote sensing time series data from ground-based data.

We illustrate the value of GEE_xtract with a complex case concerning the seasonal distribution of a threatened elusive bird and highlight its broad application to a myriad of ecological phenomena. Being user-friendly designed and implemented in a widely used cloud platform (GEE), we believe our approach provides a major contribution to effectively extracting high-quality data that can be quickly computed for metrics time series, converted at any scale, and extracted from ground information. Additionally, the framework was prepared to facilitate comparative research initiatives and data-fusion approaches in ecological research.

* Corresponding author at: CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal.

*E-mail address:* fvalerio@cibio.up.pt (F. Valerio).

https://doi.org/10.1016/j.ecoinf.2024.102502
Received 12 May 2023; Received in revised form 8 November 2023; Accepted 26 January 2024
Available online 28 January 2024

## 1. Background

Ecology science has undergone several technological revolutions, with environmental sensing standing out as a new paradigm for data collection (Chave, 2013; Kwok, 2018). The use of remotely sensed data historically relied on public sensors aboard Earth Observation Satellites (EOS), given their several benefits, which include inexpensive data for users, the long operational duration (lifespan) of sensors,extensive area coverage (footprint/swath), and repeated locations revisits (Lillesand et al., 2015). Importantly, multi-temporal EOS data consisting of long time series and derived metrics are increasingly required for ecological applications (Perrone et al., 2023; Valerio et al., 2023), jointly with the combination of different sensors to overcome specific data limits in terms of spatial, temporal, and radiometric resolution (Lillesand et al., 2015; Pettorelli et al., 2018; Schulte to Bühne, H., and Pettorelli, N, 2018). Each mission and sensor has particularly helped to understand Earth-system processes (Tatem et al., 2008), and to track spatial and temporal dynamic factors (i.e., abiotic and biotic) linked to ecological events, at functional scales (Chave, 2013; Pettorelli et al., 2005; Roughgarden et al., 1991). Increased knowledge diffusion has also boosted ecological research, facilitating scientific collaboration and enhancing multidisciplinary thinking for solving global issues (Chave, 2013). The huge growth and continuity of EOS imagery have thus promoted cross-disciplinary research interests towards remote sensing more than ever before (Pettorelli et al., 2014a; Ustin and Middleton, 2021). Undoubtedly, public remote sensing data are succeeding in turning the page on ecological applications, with the prospective to represent a step forward in addressing questions and gaps on spatio-temporal scaling of environmental data to ecological processes (Chave, 2013; Pettorelli et al., 2014).

The scale at which the problems are addressed remains a central issue in ecology (Fritsch et al., 2020), and it intertwines with the modelling procedure whenever the causal interaction between the observed pattern and the ecological process under scrutiny is sought to be identified (Levin, 1992; Wiens, 1989). In detail, the pattern-process relationship is scale-dependent from a spatio-temporal viewpoint, and a long-standing concern remains on the inconsistencies between measured environmental data and ecological phenomena, which may lead to equivocal understanding of whole ecosystems functioning (Wiens, 1989). To paraphrase Levin (1992), "the description of pattern is the description of variation, and the quantification of variation requires the determination of scales". This is, however, particularly difficult when investigating dynamic ecological systems because scale identification is usually context-dependent, meaning that some ecological patterns or processes are examined at coarse spatio-temporal scales, while others require a finer approach (Fu et al., 2011; Wiens, 1989).

Flexible data scaling is challenging not only because data may substantially differ across sensors in terms of spatial and temporal characteristics, but also because EOS data acquisition often includes hundreds or even thousands of scenes to process. Despite the promising possibility to collect multi-sensor multi-temporal time series spanning decades, sequential process steps of such EOS big data are only possible through advanced frameworks and enhanced computational power (Ma et al., 2015). In this direction, online data repositories and cloud computing platforms (e.g. Google Earth Engine; GEE; Gorelick et al., 2017) are progressively providing a compelling solution, given their ever-increasing capacity to store data from multiple satellites, combined with big data fast computation (Gorelick et al., 2017; Ma et al., 2015). In particular, the GEE platform has witnessed a significant increase in popularity among users as a geospatial processing service, and has made substantial contributions to research advancements and originality across several disciplines, mostly given its advantages for fast processing multi-petabyte of EOS data (Amani et al., 2020).

Despite providing tremendous opportunities for ecological research and multidisciplinary studies, only in recent years the use of GEE platform for remote sensing applications has become more widespread (Amani et al., 2020). In this regard, while several recent frameworks related to GEE represent a valid step forward for many applications, several important limitations still remain. For instance, efforts have recently been made to develop operational frameworks capable of matching ground truth data and remote sensing time series (multi-temporal calibration) using pre-existing available products (Crego et al., 2021; Crego et al., 2022; Dobson et al., 2023; Remelgado et al., 2019). However, the vast majority of calculable metrics are not represented in online repositories (e.g., GEE), which inevitably leads to limited research capacity and flexibility of previous frameworks in producing original metrics and variants. Likewise, data scalability is also limited because spatial and temporal data scaling parameters are largely overlooked, meaning that a single scale in multitemporal calibration is selected as the best functional one, which may not be a reasonable assumption in most cases (Levin, 1992; Wiens, 1989). Metrics calculation has become easier given the recent rise of online EOS catalogues including formulae across sensors and metrics description, potentially helping a very large user pool targeting information from various satellites (Montero et al., 2023). At present, such relevant but overlooked issues can be addressed within GEE given the high computation capacity of this web-based cloud computing platform (Gorelick et al., 2017). The increased need to reproduce, scale, and compare EOS metrics also calls for frameworks interfacing with different satellites, though addressing multiple sensors and how to soften associated shortcomings (e.g. memory limits) is routinely neglected. Therefore, it remains challenging how these issues can be bridged in a replicable framework, especially when considering that working on native imageries may lead to other problematics (e.g. clouds), potentially hampering data extraction. Motivated by this, our overall objective was to develop and describe flexible and ready-to-use codes (JavaScript and Python language) that can be tested within the GEE interface for different satellites (Sentinel-2, Landsat, and MODIS). This framework was designed for a plethora of applications requiring the collection of high-quality EOS data, the calculation of novel metrics, and the extraction of data at multiple spatio-temporal scales matching (near) ground ecological observations. The GEE_xtract codes also provide considerable speed increases in the calculation and scalability of metrics over long time series given the GEE large parallel processing system (Gorelick et al., 2017), thus useful for coherent customisation and standardisation across sensors. By allowing a quick processing of EOS data and advanced customisation of environmental metrics, we present a unique framework to facilitate the identification of the functional spatial and temporal scales that best explain a given ecological process. Therefore, this approach may help users to soften scale mismatches between environmental variables and ecological data, which remains a technical constraint for many ecological modelling approaches (Essl et al., 2023).

## 2. GEE_xtract code features and replicability

To retrieve high-quality information, the GEE_xtract code was prepared for Sentinel-2, Landsat, and MODIS, containing user-friendly parameters and organised into three main sections, namely (1) Data filtering, (2) Metrics calculation, and (3) Data scaling (Fig. 1).

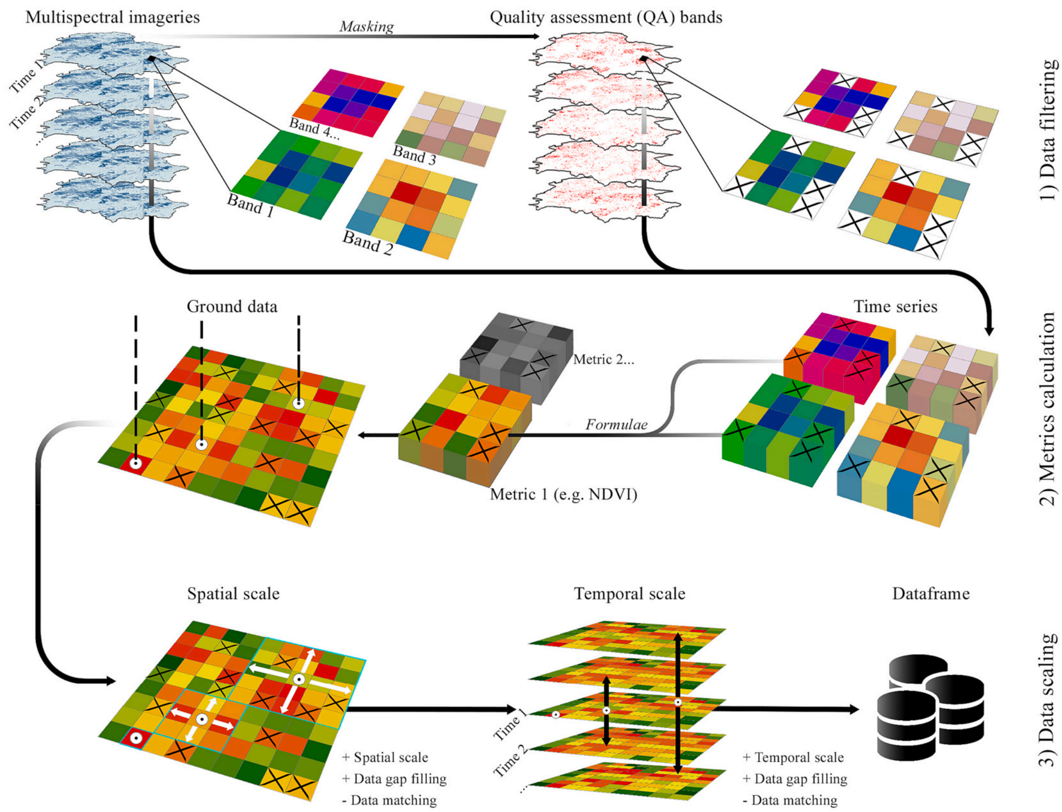


**Fig. 1.** GEE_xtract workflow according to the three main steps: (1) Data filtering, (2) Metrics calculation, and (3) Data scaling. The code utility may further depend on the complexity of involved operations around ground data (Fig. S1), which may regard data scaling to the studied ecological process, while also data gap filling and data matching features, explained further below.

### 2.1. *Data filtering*

High-quality environmental data are key for accurate ecological modelling. Accordingly, this pre-processing step consists of optimising multispectral imagery acquisition by masking out inadequate or unwanted pixels for users' goals, and subsequently retaining calibrated products. This is firstly ensured through quality assessment bands (QA bands or Bitmasks), each reflecting the EOS imagery times of observation (Fig. 1). The information within QA bands is stored in bits for each pixel, where values represent unique properties for image-masking customisation. The uptake of particular bits can ensure the exclusion of "noisy" pixels associated with atmospheric anomalies at a particular time of observation. The confidence interval in object detection can be defined according to different degrees of inclusion. For advanced parameters, several bit scores can be selected to filter out unwanted elements of terrestrial surface (e.g. water). Besides, datasets can be selected with atmospheric correction, and radiometric calibrations can be applied, such as scaling factors, and coefficients to solve inhomogeneities in bandwidth across sensors (Roy et al., 2016). Such operations should take place within a study area and period defined a priori for each study case, where EOS sequential scenes are identified on overlapping swaths, then clipped, masked, and organised into time series reflecting satellite temporal resolution (Gorelick et al., 2017).

### 2.2. *Metrics calculation*

Remotely sensed metrics may be defined and routinely analysed along time series for each multispectral imagery (Fig. 1). The aim is to provide users with a collection of formulas that are useful in applied ecology to infer biophysical characteristics of the environment (see Supplementary information S1.1 for a list of candidate metrics, as well as Montero et al. (2023) for other implementable metrics).

Multispectral bands are organised to adequately meet the wavelength interval of selected metrics for each sensor. This protocol was conceived to retrieve standardised metrics, facilitating research initiatives comparing performances across sensors with sharp technological differences.

### 2.3. *Data scaling*

The final step refers to data scaling, and intrinsically to gap filling and data matching levels required by the user. Different scaling domains have been suggested by Fritsch et al. (2020). As the interest relies on aggregating data across scales prior to modelling, our approach conforms to a "pre-modelling scaling" protocol (Fritsch et al., 2020). Sequential processes were implemented within GEE for scaling up environmental time series from one spatial and/or temporal scale to another during data extraction.

The spatial scale originally refers to area extents, wherein sub-units are referred to as grain (e.g. pixels) (Fritsch et al., 2020). The grain reflects the native spatial resolution of EOS sensors, and can range from high resolution (fine scales; e.g. 10-30 m) to moderate or low resolution (coarse scales; e.g. 250-1000 m). Here, the scaling dimension relates to the characterisation of a "spatial moving window" (Gorelick et al., 2017), to aggregate information by spatially identifying a number of neighbours pixels (sub-units; Fritsch et al., 2020), for each imagery (per-band basis) of the time series. The first advantage of this scaling process is to aggregate information while maintaining the sensor native grain size, while the second advantage relies on a vast selection of aggregation formulae, unlike traditional resampling procedures (see Park and Schowengerdt, 1983). More specifically, this scaling process was designed by parameterising the shape and size of the spatial moving window applied to each suitable pixel, and by a formula defining the spatial aggregation. The window may be configured to vary in shape, namely round or square, and size by determining the number of neighbouring sub-units included (e.g. 3 × 3; Fig. 1). The aggregating formula ranges from mean, median or standard deviation, to more advanced, like the textural metrics mentioned in Supplementary information S1.1.

The temporal scale indicates the study period, where the sub-units (sensu Fritsch et al., 2020) constitute the observation times of the multispectral imageries. The temporal scaling here relies on the characterisation of a "temporal moving window" that aggregates information per-pixel coincident with ground observations, from a composite of different imageries (Fig. 1). Here, the number of imageries is determined by a time period (e.g. number of days, months, etc.) set by the user, which reflects the "size" of the temporal moving window around observations. In detail, the temporal references of logging events (timestamps; e.g. year/month/day) associated with ground observations are independently used to aggregate environmental information forward and backward in time, depending on the size of the temporal moving window (Fig. 1). For instance, while a fine temporal scaling may be selected to perfectly match environmental information with ground observations (e.g. wildlife movements), other ecological processes may require coarser temporal scales, hence more diluted temporal aggregation (e.g. ecosystem resilience). As for spatial scaling, a formula defines how data is temporally aggregated, namely by a mean, median, or standard deviation.

Multiple spatio-temporal moving windows can be parameterised across sensors to standardise metrics from different products, facilitating the identification of functional ecological scales linked to the phenomena under study. Also, data scaling may be particularly useful for gap-filling operations, offering opportunities to avoid missing information within extracted data. Briefly, by increasing the spatio-temporal scale around ground observations, users may ensure higher chances of avoiding missing data, through the aggregation of spatio-temporal information within moving windows (Fig. 1). However, this implies also a decrease in data-matching accuracy (Fig. 1). Key decisions and trade-offs must be therefore made when extracting spatio-temporal multiscale data, possibly involving particular strategic decisions to overcome eventual implementation constraints (see the Supplementary information S1.2 section).

### 2.4. Worfklow replicability

To illustrate the replicability and flexibility of geospatial analysis within the GEE computing platform (https://earthengine.google.com/), we present a workflow highlighting the mandatory aspects in ground-based data preparation, as well as the main customisable parameters and optional functions, respectively depicted in blue and yellow type in the demonstrative code lines provided below. These codes are entirely editable and processable within a web interactive development environment (IDE), the Earth Engine Code Editor (EECE; Gorelick et al., 2017; https://code.earthengine.google.com/).

### 2.5. Preparing the ground-based data

While ground observations in the presented case study are based on point patterns, the EECE accepts multiple geometry data types, namely points, polygons, and lines, provided that information is vectorised, and two meta-data properties are respected in the vector attribute table: i) an 'ID' field (to join the extracted data table with the user shapefile), identifying each ground-data observation with a unique number, and ii) a 'timestamp' field (e.g. 'timestamp2'), where it is mandatory for the observations timestamps to present a GEE-compatible time format (e.g. "YYYY/MM/dd"; https://developers.google.com/earth-engine/apidocs/ee-date-parse) (Supplementary information S2.1). The ground-data vector, as well as the study area polygon, should be imported within the EECE, as a private (or public) asset through the Asset Manager (https://developers.google.com/earth-engine/guides/asset_manager) (Supplementary information S2.2). This is essential so that each vector is retrievable by the user in its own folder path in the EECE, so to be operational as a code object (shown in red below), via the *ee.FeatureCollection()* constructor.

```
Area = ee.FeatureCollection("users/valeriofrank/GEE_xtract/Study_Area")
user object          user path in the Asset Manager          user shapefile
pts = ee.FeatureCollection("users/valeriofrank/GEE_xtract/Ground_Observations_2017")
```

### 2.6. Coding GEE_xtract within GEE

The GEE_xtract code integrates several functions and constructors, parameterisable as command-lines by the user according to the three main steps.

The first step (data filtering) involves data acquisition and masks application (Fig. 1). From the EECE, users can retrieve the EOS data catalogue from the GEE online repository (https://developers.google.com/earth-engine/datasets/), using the *ee.ImageCollection()* constructor. Constructor arguments must also be included, such as *filterDate*, which limits the collection to a selected period, and the inner *clip* function that cuts it within a specific study area extent. The identified public image series (time series) may be preprocessed, but also undergo the previously mentioned QA bands-based masking processes for each satellite, through optionally implementable masking functions (in yellow as depicted below).

```
start = "2017-01-01"   — starting period

end = "2017-12-31"   — ending period

S2 = ee.ImageCollection("COPERNICUS/S2_SR")   — collection

                  temporal filter          collection intersection to a geometry

filtered = S2.filterDate(start, end).filterBounds(Area)
           .map(function(img){return img.clip(Area)})

                                           collection clip to a geometry

HighQuality_Area_Land_FreeClouds = filtered.map(maskClouds).map(maskCirrus)

                                     collection mask (examples) to QA bands
```

The second step (metrics calculation) concerns time series analysis involving the calculation of metrics along ingested imageries, interesting single (per-band basis) and/or multiple bands (band-to-band basis; Gorelick et al., 2017) (Fig. 1). Users can process the *compute_metrics* function (see below), which includes band math equations for calculating the metrics present in the GEE_xtract codes, as well to customise metric variants and/or introduce entirely novel metrics.

The last step (data scaling) concerns spatial and temporal scaling. Spatial scaling regards the scaling process applied to each metric of the time series, and is exemplified below by a parametrizable function, namely *addTexture_3 × 3_Mean,* which includes neighbourhood operations by applying spatial moving windows (Fig. 1), here with a size (bandwidth) of 3 pixels, and the aggregation formula given by the mean (but see also *addTexture_3 × 3_stdDev* and *NDVI_GLCM16bit* functions in GEE_xtract codes).

```
compute_metrics = function(image) {    ← function to compute metrics concomitantly

                    example of formula     selected spectral bands     output metric
ndvi = image.normalizedDifference(['NIR1', 'RED']).rename('NDVI');

                    example of variant formula
ndvi_nir2 = image.normalizedDifference(['NIR2', 'RED']).rename('NDVI_NIR2');

                    example of personalised formula
usermetric = image.expression(
'(BandX - BandY)/(BandX + BandY)', {
'BandX': image.select('BandX'),
'BandX': image.select('BandY')
}).rename(UserMetric);

                    add metrics as new bands
return image.addBands(ndvi).addBands(ndvi_nir2).addBands(usermetric)
}
                    applies the function to the original collection and add the new metrics
Spectral_Metrics = HighQuality_Area_Land_FreeClouds.map(compute_metrics)
```

```
addTexture_3X3_Mean = function(image) {← function to cumpute spatial moving windows

                    selected metric          moving window inner function
texture = image.select(['NDVI').reduceNeighborhood({

      aggregation formula (mean, median, SD, min, max, etc.)
reducer: ee.Reducer.mean().unweighted(),← no weight is given to pixels within windows

shape (square, circle, etc.)|size (3, 9, 12, etc.) of moving windows
kernel: ee.Kernel.circle(3),}).rename(['NDVI_3S_Mn']);

                    add scaled metric(s) as new band(s)
return image.addBands(texture)
}
                    applies the function to the collection and add the new metric(s)
Spectral_Metrics = Spectral_Metrics.map(addTexture_3X3_Mean)
```

Diversely, temporal scaling relies on information extraction of metrics time series from ground-based observations (the ‘pts’ object previously defined) (Fig. 1). Within the *ExtractedData* function depicted below, users may first configure the size of the temporal window through a data interval score (e.g. 3) and a time unit (e.g. days, months, etc.), and then define an aggregation formula (e.g. mean). Specifically, the timestamps of the observations are recorded and a date range is associated with each observation. These date ranges are then used repeatedly to filter and aggregate portions of the time series, allowing the ground-data vector to extract pixel statistics, by using the constructor *reduceRegion()*.

```
ExtractedData = pts.map(function(feature){          — function to cumpute
                                                      temporal moving windows
  numbers the timestamps associated to each observation
date = ee.Number(feature.get('timestamp2'));

               parses the string representation of the timestamps
date = ee.Date.parse('YYYY/MM/dd', date);

                                       size (3, 9, 12, etc.) of temporal moving windows
startTempWind = date.advance(-3,'days');backward and forward from timesteps
endTempWind = date.advance(3,'days');

                             aggregating formula (e.g., mean, min, max, SD, etc.)
FiltTempWind = Spectral_Metrics.filterDate(startTempWind, endTempWind).mean();

                 extract a metric value from each observation
data = FiltTempWind.reduceRegion(ee.Reducer.first(), feature.geometry(), 10);

        returns a data table with properties              sensor resolution (e.g., Sentinel-2)
return ee.Feature(feature.setMulti(data));
})
```

Finally, the data is formatted into a table that may be converted into a CSV file and/or attributable to the original user shapefile, then exportable into a specific Google Drive folder for external operations.

```
Export.table.toDrive({                — function to export data to Google Drive

                selected dataset to export
collection: ExtractedData,

                   dataset name
description:'Landsat_PA_Multispectral_2017',

            data format
fileFormat: 'CSV',

            Google Drive folder
'folder': 'Landsat_PA_Multispectral'
})
```

### 2.7. Worfklow applicability

To highlight a potential ecological application, we extensively tested our codes for different satellites (Landsat, Sentinel, and MODIS) using a highly complex case of spatio-temporal matching between multi-sensor environmental data and GPS tracking data from an elusive bird of conservation concern, the little bustard (*Tetrax tetrax*). We aimed to identify, for each satellite alone and across satellites, meaningful ecological variables explaining the species’ winter habitat selection (Fig. 2). For this steppe bird species, the selected area coincided with Extremadura (Spain) and Alentejo (Portugal) regions, where remotely sensed imageries were pre-processed (section 2.1) to remove noisy pixels, then variables were calculated (section 2.2) and finally scaled (3×3 window) before ground data extraction (section 2.3). Metrics were produced to represent habitat attributes in semi-arid environments: the Normalized

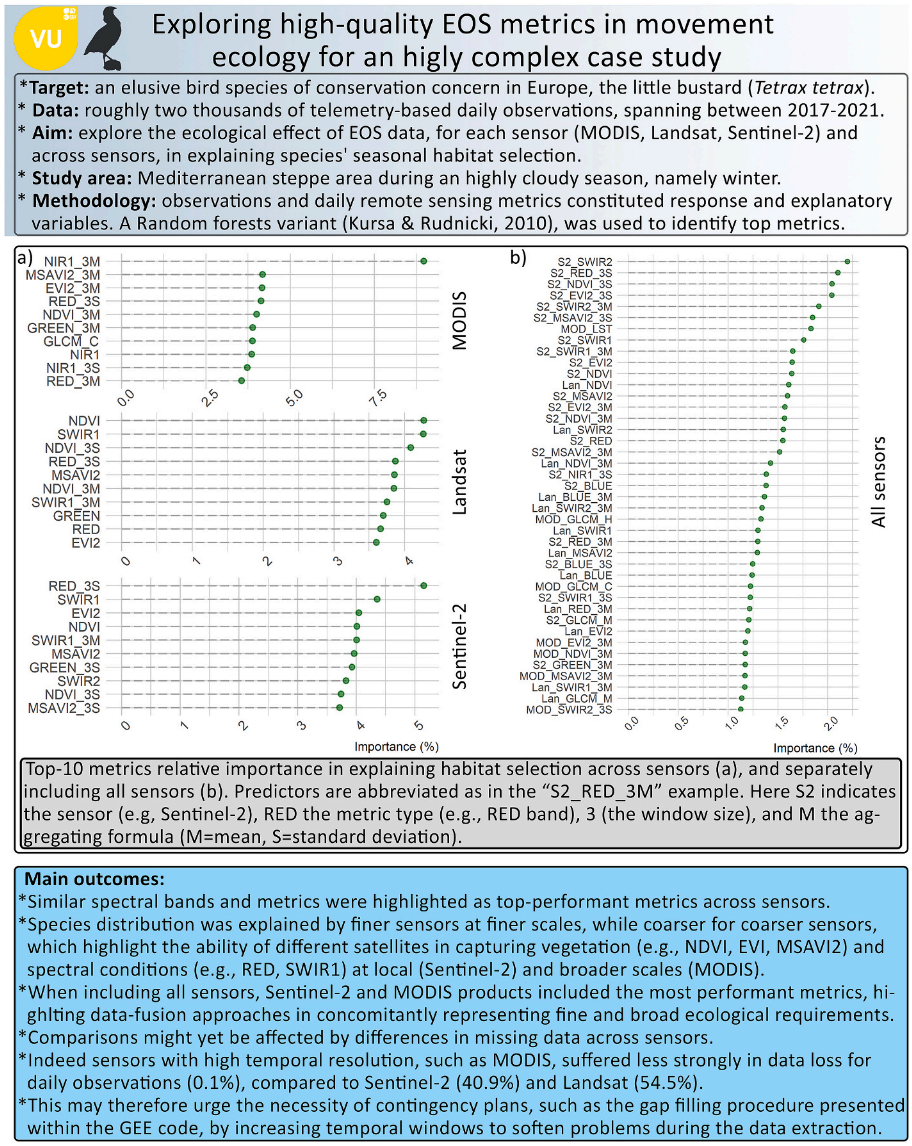


**Fig. 2.** Case study description and results from EOS-based multiscale analyses. Interpretation of metrics is presented in the grey box (see also Supplementary information S1.1), while the main outcomes are in the cyan box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Difference Vegetation Index (NDVI), the Enhanced vegetation index (EVI2), the Modified Soil Adjusted Vegetation Index 2 (MSAVI2), then satellite-derived spectral bands (BLUE, GREEN, RED, Near-Infrared–NIR1, Short-wave Infrared - SWIR1&2) and NDVI-derived textural metrics (GLCM; further detail in Supplementary information S1.1). We used a Random Forests variant (*Boruta*; Kursa and Rudnicki, 2010) to discriminate significant variables in the R environment (v.4.2.0; R Core Team, 2021), being the present GEE tool focused primarily on data preparation and scaling. This represents an example of how to explore relevant variables and associated scales, as through *Boruta* permutations are performed to understand whether the importance of each variable significantly differs from the respective randomised copy (Kursa and Rudnicki, 2010).

## 3. Applicability in ecological research

The GEE system possesses great potential to address ecological scaling issues (Gorelick et al., 2017), and the present GEE_xtract codes were designed to suit numerous ecological applications. Fine-scale approaches are often required to overcome drawbacks of coarse approaches, which traditionally involve products with relatively limited spatio-temporal resolution (e.g. Corine Land cover maps), being therefore considered as static 'snapshots' of the environment (Zeller et al., 2020). Focus has been devoted in this direction, including as regard to broad-scale ecological processes related for instance to land-use change, ecosystem status and landscape connectivity (Gustafson, 1998; Zeller et al., 2020), for which gradient-oriented EOS data at finer scales may provide valuable further insights into the factors driving such processes (Wiens, 1989; Zeller et al., 2020). Similarly, accurate spatio-temporal approaches have proven particularly useful for monitoring human-related activities, namely biophysical and socio-economic drivers in agriculture and forestry (Atzberger, 2013), as well livestock biomass consumption (Ali et al., 2016), soil and vegetation condition, moisture content and water bodies, among others (Ali et al., 2016; Henderson and Lewis, 2008; Silva et al., 2008) (Anthropogenic Domain; Fig. 3).

Fine-scale approaches have also facilitated advancements in the

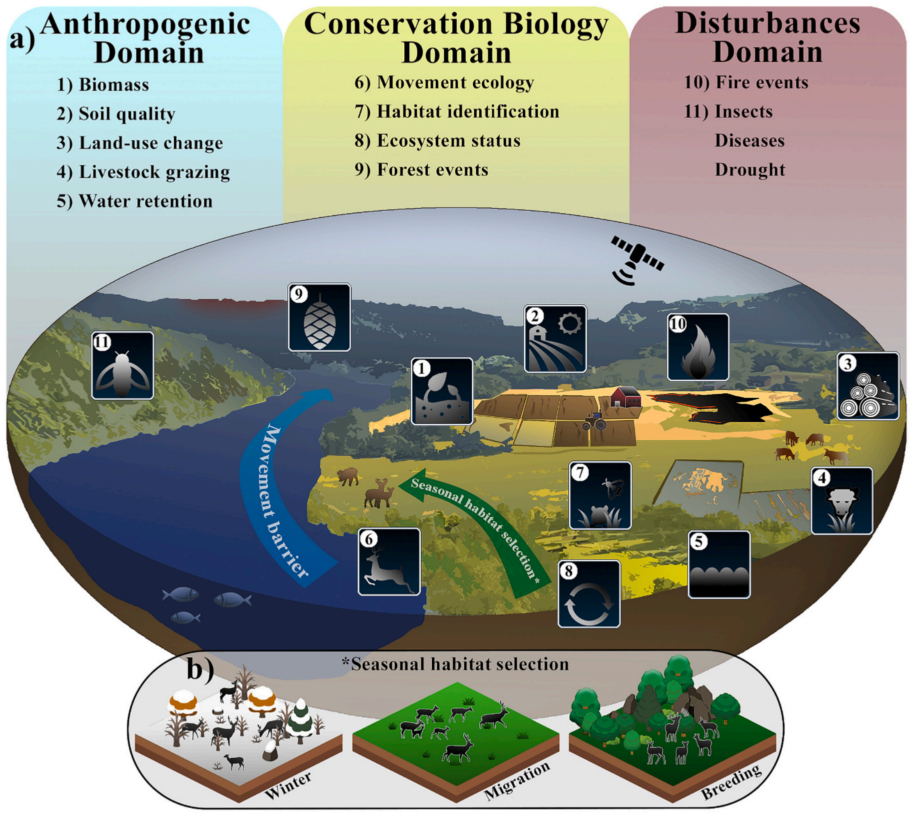


**Fig. 3.** Multiple research areas and ecological processes (represented by the different symbols) within three mutually-related domains of study (a), potentially benefitting from the application of multi-scale spatio-temporal remote sensing frameworks, enabling straightforward matching procedures between environmental data and (near-) ground-based ecological data. Linked to our case study (Section 3), the lower part (b) illustrates a possible application within the field of wildlife ecology, where habitat selection research may be disentangled along different periods (seasons) in highly dynamic landscapes.

fields of conservation research: macroecology to explain fine-scale ecological processes (Beck et al., 2012) and to map species distribution (He et al., 2015), but also movement ecology to gain mechanistic insights into wildlife behaviour and seasonal habitat selection, now possible due to significant advancements in bio-logging technologies (Wilmers et al., 2015) (Domain of Conservation Biology; Fig. 3a). Moreover, opportunities are now emerging to improve precision in the identification of irregular and dispersed ecological agents, such as fires, diseases and pests, useful for natural hazard studies (van Lierop et al., 2015) (Disturbances Domain; Fig. 3). There is therefore a plethora of ecological research areas that can benefit from the present application, especially in terms of data compatibility, transparency, and extraction speed along time series.

## 4. Conclusions

Scaling is an inevitable and complex task in ecological modelling. While some ecological processes are better explained at coarse scales, others may only be traceable at finer, or by combining multiple scales. The GEE_xtract codes presented here were designed to handle multiple spatial and temporal scales from calculated metrics over long time series from various EOS sensors and associated satellite generations, while allowing quick data extraction from broad to fine scales given the variegated nature of ground-based data. Spatial and temporal scales can be defined separately by the user and applied over spectral bands/metrics to address a broad range of ecological scaling questions and gaps. Furthermore, issues related to possible difficulties in handling large datasets or dealing with different sources of noise may be minimized following a set of proposed routines for increasing implementation effectiveness. The efficiency and flexibility of the application have thus been enhanced, not only to be multifunctional towards diverse issues (data quality, metrics, scales), but also in data comparability across sensors in terms of spectral information and deriving metrics. Therefore, we believe that the GEE_xtract codes will contribute to boosting ecological research across multiple areas, given its substantial improvements in both spatial and temporal data preparation and standardization beyond existing GEE tutorials.

## Author contributions

**Francesco Valerio**: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Visualization, Data curation, Methodology, Software, Formal Analysis. **Sérgio Godinho**: Conceptualization, Writing - Review & Editing, Visualization, Methodology, Investigation, Supervision. **Ana Teresa Marques**: Methodology, Writing - Review & Editing, Visualization. **Tiago Crispim-Mendes**: Data curation, Writing - Review & Editing. **Ricardo Pita**: Writing - Review & Editing. **João Paulo Silva**: Conceptualization, Methodology, Writing - Review & Editing, Data curation, Project administration, Funding acquisition.

## Funding

This work, including the scholarships of FV and ATM, was co-funded by the project NORTE-01-0246-FEDER-000063, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). FCT funded TM doctoral grant SFRH/BD/145156/2019, RP contract 2022.02878.CEECIND, and JPS contract DL57/2019/CP 1440/CT 0021.

## Declaration of competing interest

The author declares no conflict of interest.

## Data availability

The GEE_xtract codes are disposable on Zenodo (https://zenodo.org/records/10477901) and the GitHub repository (https://github.com/FrankVal/GEE_xtract). Data samples and exemplificative analyses in R are disposable as vectors in the FigShare repository: https://figshare.com/articles/dataset/Annual_Ground_Observations_and_Study_Area/21641564 (DOI: 10.6084/m9.figshare.21641564). Alternatively, for users authenticated to GEE, the codes may be disposable and promptly executable in the EECE, following different EOS-based paths: the path https://code.earthengine.google.com/?scriptPath=users%2Fvaleriofrank%2FGEE_xtract%3AExtract_Points_MODIS which interests MODIS collections and the path https://code.earthengine.google.com/?scriptPath=users%2Fvaleriofrank%2FGEE_xtract%3AExtract_Points_MODIS_For_Data_Fusion interesting MODIS data-fusion approaches; then the path https://code.earthengine.google.com/?scriptPath=users%2Fvaleriofrank%2FGEE_xtract%3AExtract_Points_Sentinel interesting Sentinel-2 collection; finally the path https://code.earthengine.google.com/?scriptPath=users%2Fvaleriofrank%2FGEE_xtract%3AExtract_Points_Landsat that interests Landsat (4, 5, 8) collections.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102502.

## References


Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. J. Plant Ecol. 9 (6), 649–671.

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., et al., 2020. Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. IEEE J. Select. Top.Appl. Earth Observat. Remote Sens. 13, 5326–5350.

Atzberger, C., 2013. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens. 5 (2), 949–981.

Beck, J., Ballesteros-Mejia, L., Buchmann, C.M., Dengler, J., Fritz, S.A., Gruber, B., et al., 2012. What's on the horizon for macroecology? Ecography 35 (8), 673–683.

Chave, J., 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? Ecol. Lett. 16, 4–16.

Crego, R.D., Masolele, M.M., Connette, G., Stabach, J.A., 2021. Enhancing animal movement analyses: spatiotemporal matching of animal positions with remotely sensed data using google earth engine and R. Remote Sens. 13 (20), 4154.

Crego, R.D., Stabach, J.A., Connette, G., 2022. Implementation of species distribution models in Google earth engine. Divers. Distrib. 28 (5), 904–916.

Dobson, R., Challinor, A.J., Cheke, R.A., Jennings, S., Willis, S.G., Dallimer, M., 2023. dynamicSDM: an R package for species geographical distribution and abundance modelling at high spatiotemporal resolution. Methods Ecol 14, 1190–1199.

Essl, F., García-Rodríguez, A., Lenzner, B., Alexander, J.M., Capinha, C., Gaüzère, P., et al., 2023. Potential sources of time lags in calibrating species distribution models. J 51, 89–102.

Fritsch, M., Lischke, H., Meyer, K.M., 2020. Scaling methods in ecological modelling. Methods Ecol. Evol. 11 (11), 1368–1378.

Fu, B., Liang, D., Lu, N., 2011. Landscape ecology: coupling of pattern, process, and scale. Chin. Geogr. Sci. 21 (4), 385–391.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Gustafson, E.J., 1998. Quantifying landscape spatial pattern: what is the state of the art? Ecosystems 1 (2), 143–156.

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.N., Schmidtlein, S., et al., 2015. Will remote sensing shape the next generation of species distribution models? Remote Sens. Ecol. Conservat. 1 (1), 4–18.

Henderson, F.M., Lewis, A.J., 2008. Radar detection of wetland ecosystems: a review. Int. J. Remote Sens. 29 (20), 5809–5835.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13.

Kwok, R., 2018. Ecology's remote-sensing revolution. Nature 556 (7699), 137–138.

Levin, S.A., 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology 73 (6), 1943–1967.

Lillesand, T., Kiefer, R.W., Chipman, J., 2015. Remote Sensing and Image Interpretation. John Wiley & Sons.

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., Jie, W., 2015. Remote sensing big data computing: challenges and opportunities. Futur. Gener. Comput. Syst. 51, 47–60.

Montero, D., Aybar, C., Mahecha, M.D., Martinuzzi, F., Söchting, M., Wieneke, S., 2023. A standardized catalogue of spectral indices to advance the use of remote sensing in earth system research. Sci. Data 10 (1), 197.

Park, S.K., Schowengerdt, R.A., 1983. Image reconstruction by parametric cubic convolution. Computer vision, graphics, and image processing 23 (3), 258–272.

Perrone, M., Di Febbraro, M., Conti, L., Divíšek, J., Chytrý, M., Keil, P., et al., 2023. The relationship between spectral and plant diversity: disentangling the influence of metrics and habitat types at the landscape scale. Remote Sens. Environ. 293, 113591.

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 Ecol. Evol. 20 (9), 503–510.

Pettorelli, N., Safi, K., Turner, W., 2014. Satellite remote sensing, biodiversity research and conservation of the future. Philos. Trans. R. Soc. B 369 (1643), 20130190.

Pettorelli, N., Schulte to Bühne, H, Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A. M., et al., 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. Remote Sens. Ecol. Conservat. 4 (2), 71–93.

Remelgado, R., Wegmann, M., Safi, K., 2019. Rsmove—an r package to bridge remote sensing and movement ecology. Methods Ecol. Evol. 10 (8), 1212–1221.

Roughgarden, J., Running, S.W., Matson, P.A., 1991. What does remote sensing do for ecology? Ecology 72 (6), 1918–1922.

Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., et al., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens. Environ. 185, 57–70.

Schulte to Bühne, H., & Pettorelli, N, 2018. Better together: integrating and fusing multispectral and radar satellite imagery to inform biodiversity monitoring, ecological research and conservation science. Methods Ecol. Evol. 9 (4), 849–865.

Silva, T.S., Costa, M.P., Melack, J.M., Novo, E.M., 2008. Remote sensing of aquatic vegetation: theory and applications. Environ. Monit. Assess. 140 (1), 131–145.

Tatem, A.J., Goetz, S.J., Hay, S.I., 2008. Fifty years of earth observation satellites: views from above have lead to countless advances on the ground in both scientific knowledge and daily life. Am. Sci. 96 (5), 390.

Team, R. C, 2021. R: A Language and Environment for Statistical Computing. Published online 2020.

Ustin, S.L., Middleton, E.M., 2021. Current and near-term advances in earth observation for ecological applications. Ecol. Process. 10 (1), 1–57.

Valerio, F., Godinho, S., Salgueiro, P., Medinas, D., Manghi, G., Mira, A., et al., 2023. Integrating remote sensing data on habitat suitability and functional connectivity to inform multitaxa roadkill mitigation plans. Landsc. Ecol. 1–18.

van Lierop, P., Lindquist, E., Sathyapala, S., Franceschini, G., 2015. Global forest area disturbance from fire, insect pests, diseases and severe weather events. For. Ecol. Manag. 352, 78–88.

Wiens, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3 (4), 385–397.

Wilmers, C.C., Nickel, B., Bryce, C.M., Smith, J.A., Wheat, R.E., Yovovich, V., 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. Ecology 96 (7), 1741–1753.

Zeller, K.A., Lewsion, R., Fletcher, R.J., Tulbure, M.G., Jennings, M.K., 2020. Understanding the importance of dynamic landscape connectivity. Land 9 (9), 303.