PEOPLE

Ecosystem Restoration

D7: ALGORITHM THEORETICAL BASELINE DOCUMENT

December 2023

esa · Hatfield · THE UNIVERSITY OF BRITISH COLUMBIA · VTT · ICAS

Prepared for:

ESA/ESRIN

Frascati, Italy

Hatfield Consultants LLP

#200 - 850 Harbourside Drive
North Vancouver, British Columbia, Canada V7P 0A3
Tel: 1.604.926.3261 • Fax: 1.604.926.5389
www.hatfieldgroup.com

PIONEER EARTH OBSERVATION APPLICATIONS FOR THE ENVIRONMENT – ECOSYSTEM RESTORATION (PEOPLE-ER)

D7: ALGORITHM THEORETICAL BASELINE DOCUMENT

Prepared for:

ESA/ESRIN
LARGO GALILEO GALILEI 1
00044 FRASCATI (ROMA)
ITALY

Prepared by:

HATFIELD CONSULTANTS LLP
#200 - 850 HARBOURSIDE DRIVE
NORTH VANCOUVER, BC
CANADA V7P 0A3
TEL: 1.604.926.3261 • WWW.HATFIELDGROUP.COM

DECEMBER 2023

ESA11339
VERSION 2.0

TABLE OF CONTENTS

LIST OF ACRONYMS ii
LIST OF FIGURES iii
DISTRIBUTION LIST iv
AMENDMENT RECORD iv

1.0 INTRODUCTION 1
- 1.1 PURPOSE 1
- 1.2 TARGET AUDIENCE 1

2.0 VEGETATION SPECTRAL RECOVERY 2
- 2.1 OBJECTIVE 2
- 2.2 SCIENTIFIC BASIS 2
- 2.3 USE CASES 5
- 2.4 IMPLEMENTATION 6
- 2.5 LIMITATIONS 10
- 2.6 TESTING AND VALIDATION 10
- 2.7 FUTURE WORK 11

3.0 K-NN ESTIMATION OF FOREST STRUCTURAL VARIABLES 12
- 3.1 OBJECTIVE 12
- 3.2 SCIENTIFIC BASIS 12
- 3.3 USE CASES 13
- 3.4 IMPLEMENTATION 14
- 3.5 LIMITATIONS 17
- 3.6 TESTING AND VALIDATION 18

4.0 WETLAND FUNCTION RECOVERY 19
- 4.1 OBJECTIVE 19
- 4.2 SCIENTIFIC BASIS OR OTHER ALGORITHMS 19
- 4.3 USE CASES 21
- 4.4 IMPLEMENTATION 22
- 4.5 TESTING AND VALIDATION 26
- 4.6 FUTURE WORK 26

5.0 REFERENCES 27

LIST OF ACRONYMS

ALS	Airborne Laser Scanning
ARD	Analysis Ready Data
ATBD	Algorithm Theoretical Baseline Document (ATBD)
BAP	Best Available Pixel
CA	Cluster Analysis
CSV	Comma-Separated Value
DTW	Dynamic Time Warping
EBV	Essential Biodiversity Variable
EO	Earth Observation
ER	Ecosystem Restoration
ESA	European Space Agency
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization of the United Nations
GIS	Geographic Information Systems
IW	Interferometric Wide
JNB	Jenks' Natural Breaks
JRC	Joint Research Centre
k-NN	K-Nearest Neighbors
LiDAR	Light Detection and Ranging
NBR	Normalized Burn Ratio
NIR	Near Infrared
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
PEOPLE-ER	Pioneer Earth Observation apPlications for the Environment Ecosystem Restoration
R80P	Ratio of 80P
R&D	Research & Development
RGB	Red, Green, Blue
RI	Recovery Indicator
RRI	Relative Recovery Indicator
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SWIR	Short-Wave Infrared
TCB	Tasseled Cap Brightness
TCG	Tasseled Cap Greenness
TCW	Tasseled Cap Wetness
TEP	Thematic Exploitation Platform
t-SNE	t-distributed Stochastic Neighbor Embedding
UN	United Nations
VHR	Very-High Resolution
VI	Vegetation Index
WS	Watershed Segmentation
Y2R	Years to Recovery
YrYr	Year on Year Average

LIST OF FIGURES

Figure 1 Vegetation recovery – spectral trajectory tool workflow....8

Figure 2 Example output from the Spectral Recovery Tool – raster format....9

Figure 3 Example output from the Spectral Recovery Tool – chart format....9

Figure 4 Example of a k-NN tool input feature bank csv file....15

Figure 5 k-NN tool workflow....16

Figure 6 Example output from the k-NN tool. Input Sentinel-2 image (real colour) below and volume estimation map (0-160 m^3/ha) on top. 4 x 10 km area in Northern Finland....17

Figure 7 Wetland and wetness trends tool workflow....23

Figure 8 Example output from Wetland Function tool – tSNE map with HDBSCAN clustering....24

Figure 9 Example output from Wetland *Function tool – time series of tSNE clusters.*....25

Figure 10 Example output from Wetland Function tool – classification of floodplain connectivity....25

DISTRIBUTION LIST

The following individuals/firms have received this document:

Name	Firm	Email	SharePoint	Project Website
Frank Martin Seifert	ESA	✓	-	-
Early Adopters representative	IUCN Vietnam, African Parks, Luke, INCDS, SERNbc	✓	✓	-
	Public		-	✓

AMENDMENT RECORD

This report has been issued and amended as follows:

Issue	Description	Date	Approved by	
1	Draft version of D7 – Algorithm Theoretical Baseline Document	20230630	Andy Dean Project Manager	Marcos Kavlin Assistant Project Manager
2	Final version of D7 – Algorithm Theoretical Baseline Document	20231231	Andy Dean Project Manager	Marcos Kavlin Assistant Project Manager

1.0 INTRODUCTION

1.1 PURPOSE

The Algorithm Theoretical Baseline Document (ATBD) provides a detailed description of the algorithms that are developed within the Pioneer Earth Observation apPlications for the Environment Ecosystem Restoration (PEOPLE-ER) project financed by the European Space Agency (ESA). PEOPLE-ER aims to develop innovative high-quality Earth observation (EO) based application products, indicators, and methods, targeting ER research and development (R&D) priorities.

The purpose of this document is to describe the algorithm theoretical baseline for each of the PEOPLE-ER solutions:

1. **Vegetation Spectral Recovery** – a method for analysis of time series of high-resolution multi-spectral imagery to assess spectral vegetation index recovery metrics that can be linked to structural or functional recovery.

2. **KNN estimation of forest structural variables** – a method for deriving forest structural variable (e.g., height, diameter, basal area and volume) estimation by combining field reference data and EO datasets.

3. **Wetland Function Recovery** – a method for analysis of time series of high-resolution radar imagery to assess surface water dynamics in natural to heavily modified wetland ecosystems that can be linked to wetland functions or floodplain connectivity.

1.2 TARGET AUDIENCE

This document targets:

1. Remote sensing experts interested in the analysis of EO time series for ecosystem restoration assessment.

2. ER practitioners and users of EO-derived products who want to obtain a more in-depth understanding of the algorithms.

2.0 VEGETATION SPECTRAL RECOVERY

2.1 OBJECTIVE

The objective of the PEOPLE-ER Vegetation Spectral Recovery solution is a method to assess a time series of multi-spectral satellite data using spectral vegetation indexes (VIs) to generate metrics related to forest vegetation recovery. The algorithm is able to use a variety of VIs, spatial or temporal reference conditions, and generate several recovery metrics. The solution should support applications in a wide variety of contexts, with flexible reference conditions enabling integration with current ER initiatives and guidelines. The output raster products will be suitable for further analysis.

2.2 SCIENTIFIC BASIS

Remote sensing of ER offers a solution in continuous monitoring of large spatial and temporal extents. Free and open satellite observation programs such as Landsat and Sentinel-2 have increased ER monitoring potential (Wulder et al. 2012), with further development of VIs and remotely sensed essential biodiversity variables (RS-EBVs). VIs and RS-EBVs enable the extraction of vegetation health and recovery metrics (Skidmore et al. 2021), subsequently allowing estimation of ecosystem vegetation structure, diversity, and functioning (Cabello et al. 2012; Cordell et al. 2017).

Following the opening of the Landsat archive and increased availability of multi-spectral ARD, time series composites have been widely used for assessing land cover, land cover change, and ecosystem and vegetation health (Woodcock et al. 2008; Senf 2022). This has resulted in rapid development of methodologies and algorithms which support the use of remote sensing for ER monitoring (Banskota et al. 2014; Cordell et al. 2017). These improvements have aimed to increase the effectiveness and reduce error of time series analyses and include composite methodologies, noise reduction techniques, change or trend detection algorithms, indices, and metrics. Subsequently this has resulted in transitions in the use of remotely sensed time series: detection of disturbances to regrowth, characterizing of abrupt changes to gradual changes, and the observation of transitional change to conditional change (Woodcock et al. 2020). Together, these improvements enable the use of time series analysis to monitor ecosystem conditions.

Vegetation Indices

The use of spectral variables as proxies for vegetation conditions is well established and includes spectral bands, ratio indices (e.g., NDVI), tasseled cap indices (e.g., Tasseled Cap Wetness), and spectral mixture analysis indices (Banskota et al. 2014). VIs have been specifically developed to maximize reflectance signals in a way that maximizes vegetation spectral signatures, providing information about the status of vegetation (Zeng et al. 2022). (Cabello et al. 2012). For example, NDVI can be used as a proxy for net primary production which is an essential process for ecosystem functioning (Cabello et al. 2012). (Skidmore et al. 2021).

Different VIs provide distinct information depending on the spectral bands or wavelengths involved, as well as the transformation that occurs in the calculation of the indices. For instance, SWIR-based (shortwave infrared) VIs are often used for estimating vegetation moisture content as some SWIR wavelengths are strongly absorbed by water (Zeng et al. 2022). Commonly used indices for monitoring vegetation conditions include NBR (Normalized Burn Ratio), EVI (Enhanced Vegetation Index), TCB (Tasseled Cap Brightness), TCW (Tasseled Cap Wetness), TCG (Tasseled Cap Greenness), NDMI (Normalized Difference Moisture Index), with NDVI (Normalized Difference Vegetation Index) being the most popular index (Cohen et al. 2018; Huang et al. 2021; Zeng et al. 2022). While most vegetation monitoring studies have focused on only one or a small number of indices, research suggests that using a multitude of indices is most appropriate for monitoring recovery as it provides a more complete picture of ground conditions (Cohen et al. 2018).

In addition to providing distinct information concerning the structural or functional conditions of vegetation, (Kennedy et al. 2010) note that different VIs may be more effective for subtle or gradual change detection, as they found that NBR and TCW have greater efficacy for long-term trends than NDVI. (Pickell et al. 2016) and (Schroeder et al. 2011) echoed the finding, both finding that NDVI is adequate for short-term recovery detection but does not provide recovery information after five years post-disturbance. This is due to rapid saturation effects of spectral indices, such as NDVI (Schroeder et al. 2011; Pickell et al. 2016). Thus, the incorporation of a suite of VIs into the spectral recovery tool provides the user with a comprehensive look at a multitude of vegetation and ecosystem characteristics, while also enabling the assessment of both short-term and long-term recovery trends. This fits current ER guidelines, which try to incorporate measures of diversity or abundance, ecosystem structure, and ecosystem functioning into estimates of ecosystem recovery (Ruiz-Jaen and Mitchell Aide 2005; Gann et al. 2019).

The value of using VIs for analysis is demonstrated by the development of a spectral indices package which offers a standardized catalogue of 234 indices readily implemented into a project environment (Montero et al. 2023). The Python package, called Spyndex, provides this project with a number of standardized VI equations, which were integrated into the spectral recovery tool. Further information about the formulas used in the calculations of these indices can be found in Montero et al.'s (2023) paper.

Recovery Metrics

Ecosystem recovery can be measured using recovery metrics, which utilize trends or trajectories, segments, and breakpoints of time series spectral data to characterize changes in conditions through time (Wulder et al. 2019). Recovery metrics used in prior time series analyses include RI (Recovery Indicator), RRI (Relative Recovery Indicator), Y2R (Years to Recovery), R80P (Ratio of 80P), $\Delta NBR_{regrowth}$, and YrYr (Year on Year Average) (Frazier et al. 2018; De Keersmaecker et al. 2022). 2018). As with VIs, recovery metrics present distinct and complementary information about ecosystem recovery.

Recovery metrics can be modified to better suit study objectives (Frazier et al. 2018). For example, R80P was derived from Y2R, whilst (Frazier et al. 2018) modified the popular Recovery Indicator (RI) metric to create the Relative Recovery Indicator (RRI) metric, more appropriate to their analysis of unfitted spectral data. Consequently, previous analyses justify this project's modifications of recovery metrics. Modifications include the application of $\Delta NBR_{regrowth}$ to multiple indices (e.g., the absolute change in any index from the start of restoration to five years later), as well as flexibility in defining the years to use for spectral values within the recovery metric calculations. The inclusion of multiple established and modified recovery metrics in the spectral recovery tool gives extensive information about current and potential recovery progress.

There are numerous successful time series analyses which use multi-spectral information to determine recovery progress. For example, (Pickell et al. 2016) calculated Y2R to characterize the recovery trends of multiple indices, including NBR, NDVI, and TCG, for disturbed North American boreal forests between 1985 and 2010. (Frazier et al. 2018) used a temporal segmentation and Thiel-Sen regression approach to characterize annual unfitted time series and calculate RRI, R80P, and YrYr. (White et al. 2022) used the $\Delta NBR_{regrowth}$ and Y2R metrics, validated with ALS (airborne laser scanning) data, to estimate baseline recovery rates for Canadian ecozones. Similarly, (White et al. 2017) performed a short-term and long-term recovery analysis, extracting $\Delta NBR_{regrowth}$, RI, and Y2R using segmentation and fitted trajectory curves. Validation for these analyses often use ground measurements or ALS, however Woodcock et al. (2020) argue that field data and ALS are often unattainable due to large study areas and a lack of resources, and in their absence, manual visual interpretation of high spatial resolution imagery has proven adequate for validation datasets. Previous research has thus established multi-spectral time series analysis as a worthwhile approach to determine a wealth of information about vegetation conditions over large spatial and temporal extents.

Overall, the utilization of time series analysis for the extraction of recovery metrics measuring changes in spectral indices can consequently be used to monitor recovery of vegetation and ecosystem conditions.

Recovery metrics equation variables

Variable	Description
D_s	Index value: Disturbance start
D_e	Index value: Disturbance end
D_{pre}	Index value: Pre-disturbance (average of 2 years prior to D_s)
$D_{average}$	Index value: average value from D_s to D_e
R_0	Index value: Restoration start (usually 1 year after D_e)
R_i	Index value: years after R_0, subscripts indicate the # of years
$R_{current}$	Index value: current time step or last year of time series
R_{target}	Index value: recovery target
R_{slope}	Recovery slope/rate
P	Percent, user-defined, default is 80
t	Years/time step

RRI (Relative Recovery Indicator)

The RRI provides unique information in that the recovery magnitude is divided by the magnitude of the disturbance, meaning the recovery is scaled to the disturbance severity (Kennedy et al. 2012; Frazier et al. 2018).

$$\frac{max\ (R_5, R_4) - R_0}{D_S - D_e}$$

Y2R (Years to Recovery)

Y2R has typically represented the number of years required for the restoration site to reach 80% of its pre-disturbance condition, providing an estimate of how long the recovery process takes (Pickell et al. 2016; White et al. 2017; White et al. 2018).

$$i \; where \; R_i = R_{target} * P$$
Default:
$$i \; where \; R_i = R_{target} * 0.8$$

R80P (Ratio of 80P)

R80P compares the current condition of an ecosystem to target conditions, as the extent to which a pixel has reached 80% of pre-disturbance values (Frazier et al. 2018; De Keersmaecker et al. 2022).

$$\frac{R_{current}}{R_{target} * P}$$
OR
$$\frac{\max(R_5, R_4)}{R_{target} * P}$$

ΔNBR$_{regrowth}$ (change in NBR from the start of the restoration process to five years after)

ΔNBR$_{regrowth}$ is the absolute change in NBR from the start of the restoration process to five years after (White et al. 2017; White et al. 2022). ΔNBR$_{regrowth}$ provides a recovery rate distinct from the reference conditions and thus provides a more direct measure of regrowth.

$$R_i - R_0$$

Default:

$$R_5 - R_0$$

YrYr (Year on Year Average)

YrYr uses the rate of recovery, or spectral trajectory, to determine the average annual rate of spectral change (Frazier et al. 2018; De Keersmaecker et al. 2022). YrYr provides a recovery rate distinct from the reference conditions and thus provides a more direct measure of regrowth.

$$\frac{R_i - R_0}{\Delta t(R_i - R_0)}$$

Default:

$$\frac{R_5 - R_0}{5}$$

2.3 USE CASES

The primary use case for the PEOPLE-ER Spectral Recovery tool is a user who has administered or is monitoring an ER intervention and wants to assess the efficacy of the intervention using remote sensing-based recovery metrics.

The user may wish to:

- Determine recovery from a historic perspective – assess a restoration area's recovery with metrics that use a historic period (prior to ER) as the target conditions.
- Determine recovery from a reference site(s) perspective – assess a restoration area's recovery with metrics based on reference site(s) and a reference period as the target conditions.
- Determine relevant metrics – generate multiple ER metrics to determine which most effectively assess various aspects of recovery (e.g., structure, landscape) over an area of interest (AOI).

In Scope / Out of Scope

In scope features of the PEOPLE-ER Spectral Recovery tool are:

- Multi-spectral recovery metrics for restoration site(s).
- Generation of a recovery baseline from reference period (prior to ER) or from reference site(s).
- Computation of spectral indices relevant to ER as input to recovery metric methods.
- Visualization of recovery metrics for restoration site(s).
- Analysis of multi-spectral recovery metrics – e.g., spatial or time series clustering.

Out-of-scope features of the PEOPLE-ER Spectral Recovery tool are:

- Multi-spectral cloud-free compositing to provide the input EO data
- Spatial or temporal delineation of restoration or reference site(s)

Preconditions

The tool requires that a user have pre-existing domain knowledge of the restoration sites they wish to assess. Specifically, users must have:

- Knowledge and spatial delineation of restoration sites.
- Knowledge of the year of restoration intervention.
- Knowledge of reference period to determine target conditions.
- (Optional) Knowledge and spatial delineation of reference site(s).

A user must also be able to provide multi-spectral EO data to the tool, related to the following preconditions:

- Access to, or ability to produce cloud-free analysis ready data (ARD) Landsat or Sentinel 2 derived annual time series composites over the restoration and reference sites for multiple spectral bands.
- (Optional) Access to, or ability to produce a data mask for masking out undesirable areas (e.g. water bodies, naturally unvegetated features, settlements) in the provided time series composites.

2.4 IMPLEMENTATION

Platform Environment

The PEOPLE-ER Spectral Recovery tool is developed using the Python 3 programming language because it provides ease-of-use, a breadth of high-quality packages, and is now one of the standard languages in the world of open-source scientific computing.

To facilitate scalable analysis, the following third-party Python packages will be used to implement data models and spectral recovery algorithms:

- Dask: open-source package for parallel computing, capable of scaling Python code on multi-core systems or to distributed clusters in the cloud.
- XArray: open-source package providing label-based data models and operations on top of NumPy-like arrays, making N-dimensional data processing easier for users and developers.

These packages, and any additional packages, are selected as to be platform-independent and interoperable with the scientific Python ecosystem and open-source scientific computing initiatives like the Pangeo project. Thus, any processing platforms which support similar ecosystems will be able to support the PEOPLE-ER Spectral Recovery tool. Specific platforms that should be compatible are Microsoft's Planetary Computer Hub, VTT's Forestry TEP platform, and the new Copernicus Data Access Service (based on information available).

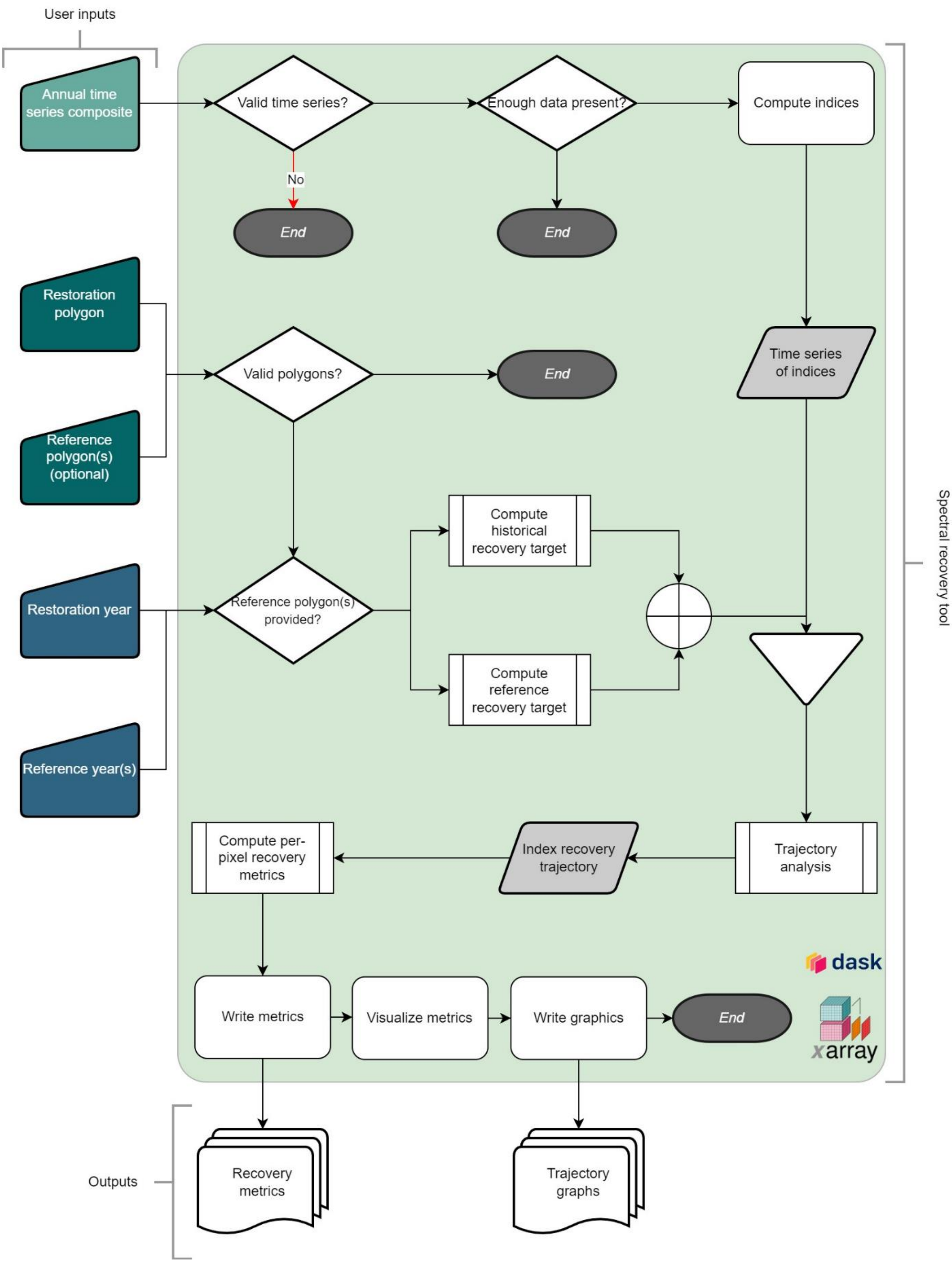
A User Guide is available at: https://people-er.github.io/Spectral-Recovery/about/

Inputs

- A cloud-free, ARD Landsat or Sentinel-2 derived multi-spectral annual time series composite.
- Spatial and temporal delineation of a restoration site.
- An optional delineation of a reference target condition, or a historic reference temporal range to provide a baseline condition.
- Selection of indices.
- Selection of metrics.

Workflow

Figure 1 Vegetation recovery – spectral trajectory tool workflow.



Outputs

- Computed recovery metrics for selected indices: RRI, Y2R, R80P, $\Delta NBR_{regrowth}$, YrYr
- Recovery trajectory graphs for restoration sites.
- Raster of restoration site recovery relative to baseline or target conditions for each index/metric combination.

Figure 2 Example output from the Spectral Recovery Tool – raster format.

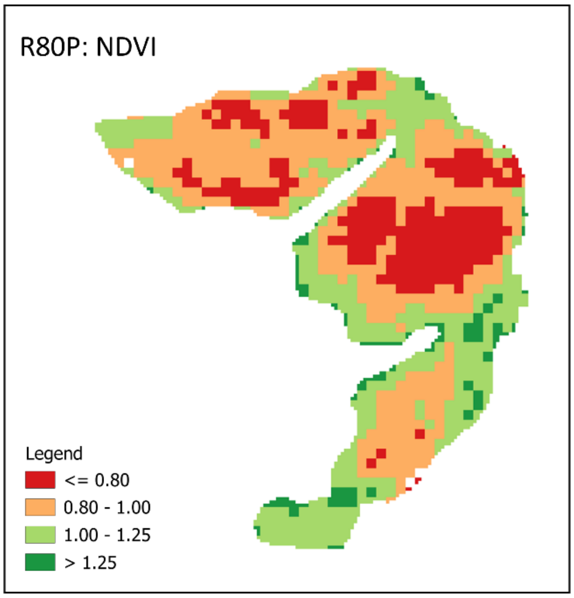


Figure 3 Example output from the Spectral Recovery Tool – chart format.

2.5 LIMITATIONS

- Sensor support – The tool supports Landsat and Sentinel derived data. Support for other sensors will be beyond the scope of the toolset, as the formulae for index calculations are sensor dependent.

- Spatial resolution – Sentinel and Landsat data are limited to 10 m and 30 m spatial resolutions respectively. This may not provide the level of detail necessary for small restoration sites or fine-scale analyses, or to reveal the true recovery variation in a landscape. The tool will be best applied to landscape approaches, and it is recommended to follow with sampled field-based data if used for local restoration efforts to ensure recovery is adequately assessed.

- Spectral information as sole input – Resulting interpretation of ecosystem structural and compositional characteristics can only be estimated by using spectral indices as proxies.

- User-provided time series composites – Noise may be introduced into the analysis if composites are not high quality. This can be mitigated by users following established pre-processing composite approaches (e.g., best-available-pixel (BAP), and using cloud-masking algorithms such as FMask (Zhu and Woodcock 2012), SEN2COR (Tarrio et al. 2020), LaSRC (Vermote et al. 2016; Skakun et al. 2019), MAJA (Hagolle et al. 2010), TMask (Zhu and Woodcock 2014), or s2cloudless (Zupanc 2017).

2.6 TESTING AND VALIDATION

Evaluation of the vegetation spectral recovery metrics is required to ensure understanding of the linkage between spectral measures of forest recovery and manifestations of forest structure and composition.

Data that can be used for validation and/or testing include:

- **Airborne Lidar** – Airborne lidar provides the key validation dataset and can be used to determine structural benchmarks of recovery. Availability will depend on the validation/test site and time-period.

- **Vegetation Inventory Data** – Availability of the dataset may depend on the testing/validation site and time-period.

- **Disturbance Polygons** – Availability of the dataset may depend on the testing/validation site and time-period.

- **Forestry spatial data** – Availability may vary depending on the testing/validation site and time-period.

White et al. (2022) completed validation of the spectral recovery method using Landsat time series, NBR as the spectral index, and a two-year pre-disturbance temporal baseline period. Airborne lidar data were used generate benchmark thresholds of canopy cover (e.g., >10%) and height (>5 m) to assess forest recovery that are related to the minimum values required to satisfy the FAO's definition of forest (FAO 2018; White et al. 2018; White et al. 2022).

PEOPLE-ER Validation is reported in Deliverable D6 (Results and Validation) available from https://www.people-er.info/

2.7 FUTURE WORK

Future options to improve the initial solution include:

- Expansion of sensor support beyond Landsat and Sentinel-2.
- Inclusion of a time series composite creation tool.
- Integration of other data, such as climate and soil variables.

3.0 K-NN ESTIMATION OF FOREST STRUCTURAL VARIABLES

3.1 OBJECTIVE

The objective of the PEOPLE-ER k-NN tool is to provide a generic tool to conduct k Nearest Neighbour (k-NN) estimation of target variables of interest. In the context of ER monitoring, the tool allows wall-to-wall propagation of the variables of interest using field reference data and selected EO datasets.

Together with the spectral recovery tool, it provides the PEOPLE ER user with the possibility to approximate the ecosystems' status (e.g., in the form of forest structural variable development) at any given time during the recovery process. It allows users to use any available combination of EO and auxiliary layers, together with their own or external field reference data. The field reference data can be in the form of field plot measurements or forest compartment (stand) level data.

3.2 SCIENTIFIC BASIS

There is a long tradition of forest structural variable (e.g., height, diameter, basal area and volume) estimation combining field reference data and EO datasets. Satellite based methods for estimation of forest structural variables in 10-30 m resolution have been developed since the 1990's (Tomppo and Katila 1991; Tokola et al. 1996; Hame et al. 2013) are still in operational use together with new data sources like LiDAR (Kangas et al. 2018). The k-Nearest Neighbor method (Alt 2001) has been widely used in forest monitoring (Chirici et al. 2016) and is operationally used for example in the Finnish National Forest Inventory (Mäkisara et al. 2022). One of the key benefits of the k-NN method is the capability for simultaneous estimation of all variables of interest. This helps to retain natural ecological relationships between the variables (e.g., height and volume) in any given point, better than methods that estimate one variable at a time.

The provision of a generic k-NN tool allows users to derive additional information on the status of the ecosystem with a variety of datasets. The tool is not limited to any specific type of reference or remotely sensed data but can be used with the datasets available for the area of interest. The PEOPLE-ER k-NN tool should be seen as a supporting tool for the main ecosystem recovery monitoring tools developed in the project. This tool can be used to provide further information within and around the restoration areas, benefiting from the field reference data the users have collected or that is available from public sources.

The tool is an implementation of a generic non-parametric and distribution-free k-NN method (Alt, 2001). The estimates for the target variable values are obtained as linear combinations of target variable values in a set of observations selected from a reference feature bank. The reference feature bank should contain combinations of reference target variable values with the corresponding EO and auxiliary data features. The observations are selected by Euclidean distance on the EO and auxiliary variable space. The reference observations with the smallest distances to the target pixel in the EO and auxiliary space are selected. Simultaneous estimation of all variables of interest can be conducted. The k-NN is a non-parametric estimator since estimations can be made without any parameters, as well as distribution-free estimation method because estimations can be made without any prior distributional assumptions. The reference observations can be field plots or stands.

The k-NN predicted target variable value Y_p for pixel p is calculated as:

$$\hat{y}_p = \sum_{i \in I} w_{i,p}\, y_i,$$

where y_i is the target variable value of the i^{th} contributing observation in the reference set, I is the subset of nearest observations used for the estimation, and $w_{i,p}$ is the weight of i^{th} contributing observation. In the implementation described here, the observations are weighed inversely to the Euclidean distance.

The Euclidean distance is calculated as:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i\right)^2}$$

where x_i is the target pixel feature value for feature i and y is the corresponding value for an observation in the training feature bank. The number of nearest observations used in the estimation typically varies from three to over 15, depending on the number and characteristics of the available field reference data.

In addition to estimation of the target variable calculated as the weighted mean of the chosen observations as described above, the standard deviation of the observations used to derive the pixel level estimates is calculated. This provides information on the variability of the observations and thereby the uncertainty of the estimation. The standard deviation is calculated as:

$$\text{Standard deviation} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

where x_i is the target variable value for observation i and the $\bar{x}$ is the mean of the observations used in the estimation.

In addition to the output map layers (i.e. the mean estimate and standard deviation layers), general uncertainty metrics are calculated from the feature banks. The user has an option to divide the feature bank to training and testing feature banks. The training feature bank is used to predict values for the observations in the testing feature bank. The estimations are then compared to the observations to calculate Root Mean Square Error (RMSE) and bias of the estimations as:

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}} \qquad BIAS = \frac{\sum_i (y_i - \hat{y}_i)}{n}$$

where y represents the observed reference values, $\hat{y}$ represents the predicted values and n is the number of samples. Both RMSE and bias are provided in absolute terms and as % of the mean. All of the uncertainty statistics are calculated for each target variable.

3.3 USE CASES

The primary use case for the PEOPLE-ER k-NN tool is a user who is monitoring an ecosystem restoration intervention and wants to know the past or current status of the ecosystem in the area of interest.

The user may wish to:

- Know the pre-restoration characteristics of the ecosystem.

- Monitor the development of ecosystem characteristics during the restoration process.
- Compare ecosystem characteristics in the restoration area and in ecosystem outside the restoration area.

In Scope / Out of Scope

In scope features of the PEOPLE-ER k-NN tool are:

- Generation of pixel-wise estimates of target variable values with the k-NN algorithm using EO (and optionally auxiliary) data features together with field reference data.

Out-of-scope features of the PEOPLE-ER Spectral Recovery tool are:

- Collection, screening, and pre-processing of the field reference data
- Preprocessing of the potential auxiliary datasets
- Creation of the feature bank, which combines target variable values and EO feature values (and potential auxiliary feature values)

Preconditions

When the k-NN tool is used to support ecosystem restoration the user needs to have pre-existing domain knowledge of the restoration sites and other areas they wish to create the target variable map over. Specifically, users must have:

- Knowledge and spatial delineation of restoration sites
- Knowledge of the year of restoration intervention
- Representative field reference data collected on a year that also has EO data available. The representativeness of the field data is essential. It should cover the entire range of target variable values and include typically at least 100 observations.

In addition, it is a crucial pre-condition that the variable of interest correlates with the EO and/or auxiliary data features. Otherwise, the tool does not provide reliable estimates of the target variable (e.g., forest height) based on the EO and auxiliary data features. Please see more considerations on the limitations of the tool in Section 3.5.

3.4 IMPLEMENTATION

Platform Environment

The implementation follows a generic non-parametric and distribution-free k-Nearest Neighbour method (Alt 2001) which has been widely used in forest monitoring. The PEOPLE-ER k-NN tool has been implemented on the Forestry TEP online platform (https://f-tep.com/) to allow easy exploitation of Earth observation (EO) datasets. It is also provided as a python script package (with example datasets) in the PEOPLE-ER github repository (https://github.com/PEOPLE-ER) for flexible implementation by more advanced users. A User Guide is available at: https://people-er.github.io/k-NN/

Inputs

The required inputs can be divided into two categories:

- Input required for the feature bank creation:
  - EO (and any auxiliary feature as desired) layers from the time of the reference data collection
  - Representative reference data from the area of interest
- Input required for the k-NN estimation:
  - Selected EO and auxiliary feature layers for the desired date
  - Training feature bank .csv file including the target and EO/auxiliary variables (Figure 4)
  - (Optional) Testing feature bank .csv file
  - Selection of the number of neighbours (k) to be used

Figure 4 Example of a k-NN tool input feature bank csv file.

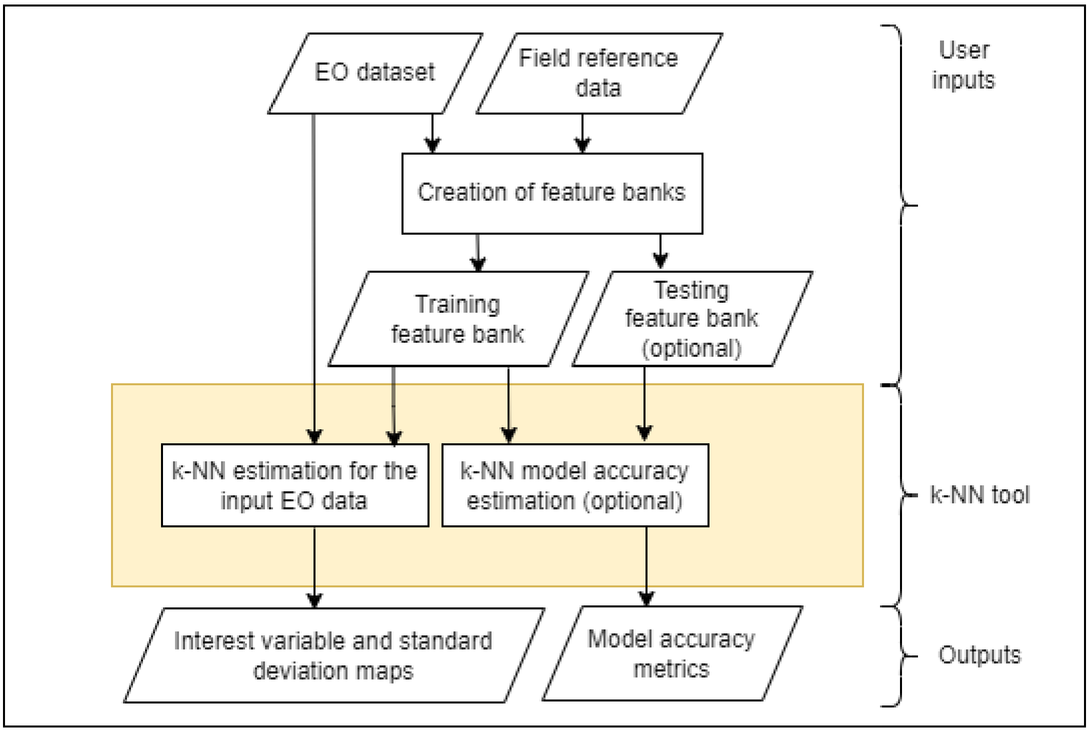
```
PLOTID,G,V,D,H,PINE,SPRUCE,BL,B2,B3,B4,B8,B5,B11,B12
536719,0,0,0,0.3,0,0,0,97.21,254.99,241.08,1578.89,628.91,1456.94,741.88
535639,0,0,0,1.1,0,0,0,139.88,428.34,276.41,3132.56,912.91,1954.96,930.72
535424,0.2,0,0.3,0.9,100,0,0,193.86,511.42,375.75,2917.17,990.62,1690.4,825.55
536661,0,0,0,0.5,0,0,0,253.45,550.84,549.33,2715.1,1230.06,1860.79,983.99
543307,0,0.5,0.4,3,0,0,0,145.46,431.06,322.49,2782.42,931.85,1508.38,752.85
536569,0.3,1,0.9,1.6,100,0,0,169.8,432.05,310.13,2455.36,789.69,1132.67,558.71
```

Workflow

The workflow of the process is illustrated in Figure 5. The tool can be run with an EO image and a training feature bank. The feature bank can be provided to the user, or the user can create the feature bank, e.g., in QGIS by extracting EO feature values to their field data locations. In an optimal case, part (e.g., 1/3) of the feature bank is extracted to a testing feature bank, while the remainder (e.g., 2/3) of the reference observations are used as training feature bank for the target variable estimation. When the feature bank is available and uploaded to Forestry TEP, can be run for any given EO image directly on the platform.

The EO and auxiliary dataset needs to be in .tif format and it needs to contain the same layers in the same order as provided in the Feature bank .csv files. The .csv files can be created in any GIS software by extracting image band and auxiliary data values for the corresponding target variable values, e.g., in pixel or polygon level, depending on the field data area unit. In addition to the comma separated target variable and EO feature values, the .csv file needs to include a header row providing the column names. EO features are used to search for the closest neighbours to be used in the prediction of the target feature values. The EO features and target variable features are listed in the knnsettings.csv file.

Figure 5 k-NN tool workflow.



Outputs

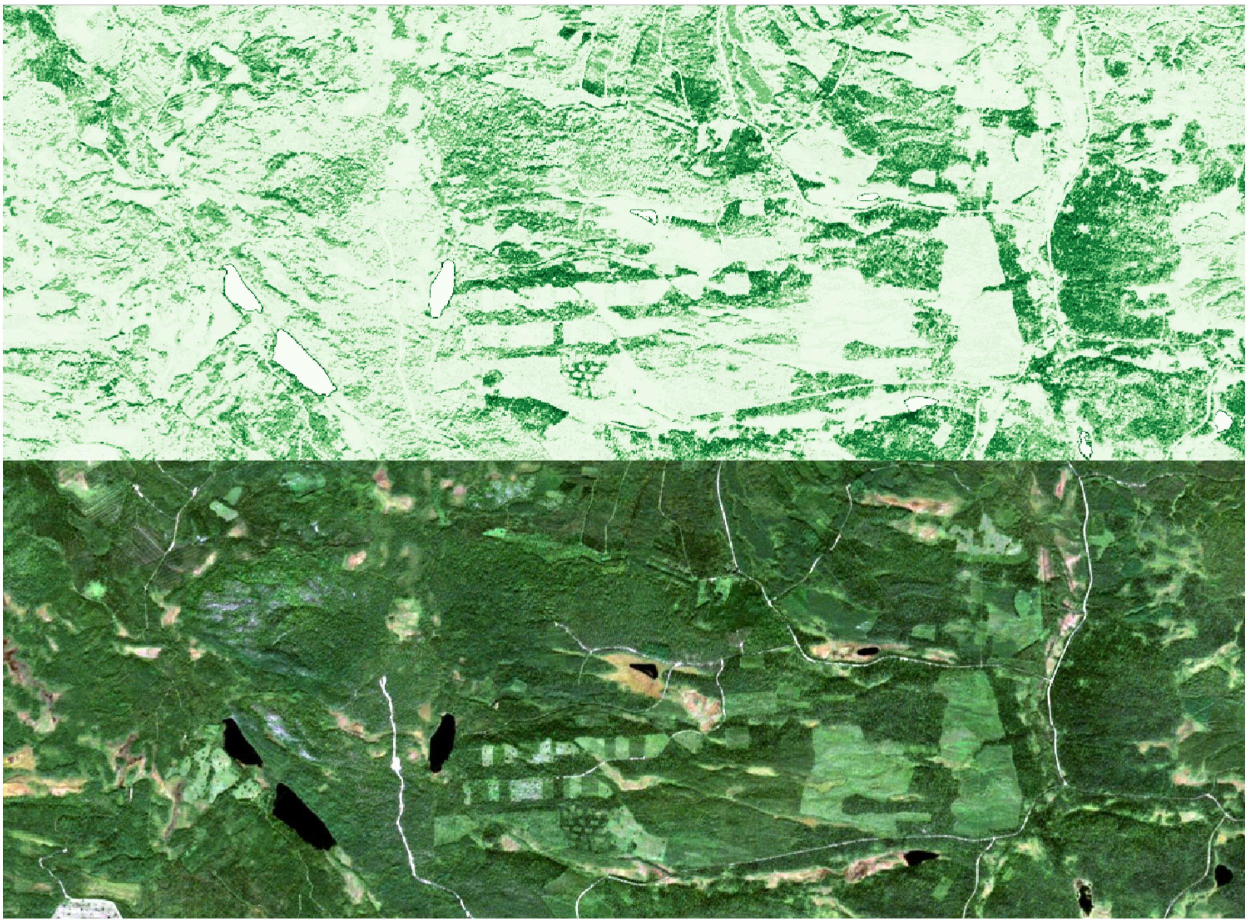
The k-NN tool creates two different types of outputs:

1. Target variable estimate and standard deviation maps for the area of interest
2. Accuracy metrics (in case optional testing feature bank provided)

The output maps include a layer of estimates for each of the forest variables of interest. These maps are output in .tif format in the same spatial resolution as the input EO data (Figure 6). In addition to the estimate maps, each variable is also accompanied with a layer of standard deviation of the estimates. Note that the standard deviation maps are only available in the Forestry TEP version of the tool, not in the python code distributed in the GitHub.

The uncertainty metrics (see Section 3.2) are provided in .csv files. They have been calculated from the testing feature bank by using the training feature bank to predict values for the testing feature bank items and comparing the predictions to the reference values. Note that in the case of very small testing feature bank, the metrics may not be reliable.

Figure 6 Example output from the k-NN tool. Input Sentinel-2 image (real colour) below and volume estimation map (0-160 m^3/ha) on top. 4 x 10 km area in Northern Finland.



3.5 LIMITATIONS

The tool requires that a user have pre-existing domain knowledge of the restoration sites and other areas they wish to create the target variable map over. Specifically, users must have:

- Knowledge and spatial delineation of restoration sites
- Knowledge of the year of restoration intervention
- Representative field reference data collected on a year that also has available EO data

The representativeness of the field data is essential. It should cover the entire range of target variable values and include typically at least 100 observations. Without a representative field reference data, the tool will not produce reliable results. Typically, the representative reference data needs to be provided by the user. In some cases, suitable open field reference datasets may be available from the area of interest.

The tool does not provide pre-processing of the reference, EO or auxiliary datasets. The user needs to create the feature bank .csv file and a corresponding stack of the EO and auxiliary layers.

In addition, it is a crucial pre-condition that the variable of interest correlates with the EO and/or auxiliary data features. Otherwise, the tool does not provide reliable predictions of the target variable (e.g., forest height) based on the EO and auxiliary data features. Similarly, the tool only provides reliable predictions for ecosystems which are included in the feature bank. For example, the example feature bank provided here with the python code at GitHub only applies for forest ecosystem. Results created with the example feature bank are invalid in any other land cover types.

Furthermore, it is important to note that the levels of reflectance in the EO images used for the tool should be similar to levels of the image used to create the feature banks. The user should beware of variation in the levels of reflectance, even in Analysis Ready Datasets (ARD), like the Sentinel-2 L2A surface reflectance product. Atmospheric and seasonal changes may cause variation in the levels of reflectance between images. This type of variation affects the predictions of the k-NN tool.

Finally, application of a feature bank outside its geographic extent must be conducted with caution. Based on an empirical method, the k-NN tool is sensitive to the geographic variation in ecosystem characteristics. It is not recommended to apply a feature bank for target variable prediction outside its geographical extent.

3.6 TESTING AND VALIDATION

The testing and validation of the tool was conducted in two countries: Finland and Romania. The validation results are reported in Deliverable D6 (Results and Validation) available from https://www.people-er.info/

4.0 WETLAND FUNCTION RECOVERY

4.1 OBJECTIVE

The objective of the PEOPLE-ER Wetland Function Recovery solution is to provide a method for high-resolution satellite EO data time series analysis to enable the monitoring and comparison of surface water dynamics in natural to heavily modified wetland ecosystems. With recent innovation in cloud computing and the availability of long-term Synthetic Aperture Radar (SAR) EO datasets at high temporal and spatial resolution, the technical objective is to develop analysis tools in such a way that it is not tied to a singular EO exploitation platform, but instead can be distributed to a variety of platforms.

4.2 SCIENTIFIC BASIS OR OTHER ALGORITHMS

For wetland ecosystems, the location and persistence of surface water (inland and coastal) is a key driver of biological diversity and ecosystems services. Restoring of natural wetland inundation function is often an important feature of wetland restoration because the hydrological regime drives the nutrient fluxes, water quality, and habitat suitability for plant and animal species and other biodiversity. For example, enabling reconnection of wetlands within a floodplain and restoring wetland inundation functions can be a key indicator of wetland restoration.

The *biological effects of irregular inundation* is recognized as a high-priority remote sensing biodiversity product, related to the "ecosystem disturbance and habitat structure" remote sensing enabled essential biodiversity variables (RS-EBVs) (Skidmore et al. 2021).

EO time series have a proven capability in the detection of surface water location and vegetation inundation seasonality. Several initiatives aimed to use multi-spectral and radar time series (e.g., Joint Research Centre (JRC) global surface water permanence dataset based on Landsat time series (Pekel et al. 2016) and the ESA financed WorldWater project[1] provide valuable information but are limited to the detection of surface water. Ecosystem restoration practitioners require tools to assess complex wetland ecosystems. However, challenges include working with large volumes of EO data, handling EO data time series, the complexity of wetland structure, and methods to assess wetland restoration, i.e., using reference sites or reference time periods. Another factor is the inter-annual variability of climate and short to long-term responses of wetlands to restoration processes and climate variability.

SAR sensors are capable of distinguishing open water from other land cover classes relatively well due to shallow scattering signature and thus strong contrast with other land cover classes. C-band SAR image time series are particularly suitable for mapping water bodies with high accuracy (Lamarche et al. 2017). The extent of permanent water bodies and dynamics can be monitored. Recent land cover classification experiments with deep learning semantic segmentation models suggest very high accuracies can be obtained with dual-pol IW mode Sentinel-1 data. Further improvements can be expected from combining SAR and optical datasets for delineating inland water bodies, for example with Copernicus datasets (Gao et al. 2018).

Initial results of a round robin exercise organized within the WorldWater project confirm that high accuracies in surface water mapping can be achieved with optical and SAR data. Particularly, optical

[1] https://worldwater.earth/

images are better at capturing spatial detail while, SAR data provide a better seasonal characterization when looking at the classification performance across several study sites as well as through time.

By combining both optical and SAR data the PEOPLE-ER wetland function recovery tool aims to provide meaningful analyses of wetland patterns and function to restoration practitioners.

Landscape Structure

The wetland function recovery tool is object-based rather than pixel-based. The rationale is that natural and modified wetland landscapes are often managed in terms of hydrological units (Department of Environment and Science, Queensland 2021) and assessing objects can reduce the noise that may be associated with pixel based analysis.

There are numerous image segmentation methods available. An example open algorithm is introduced as an optional pre-processing component of the wetland function recovery tool. This is based on cloud-free Sentinel-2 composite images and a combination of Scharr edge detection and Watershed Segmentation (WS), referred to as the CEWS workflow (Watkins and Van Niekerk 2019). Firstly, edge layers for each band of each Sentinel-2 composite will be generated using the Scharr algorithm. Afterward, the multi-temporal edge layers will be combined into one composite edge layer using equal-weight summation, to enhance the magnitudes of the landscape boundary. The output of the Scharr edge detection algorithm will be reclassified using Jenks' Natural Breaks (JNB) algorithm. The result of this step will be used as the input to the WS algorithm. The objective of the WS algorithm is to recognize the high gradient magnitudes (boundaries) that divide low gradient regions (homogenous landscapes). However, the WS algorithm is prone to overly segment images and consequently produces many small objects. Therefore, the use of JNB will effectively avoid the influence of noise in the edge layer and alleviate the over-segmentation issue (Xu et al. 2023). The landscape units extracted by the WS algorithm will be converted into vector format for the subsequent analysis.

See https://people-er.github.io/Wetland-Function-Assessment/02_landscape_segmentation/

t-distributed Stochastic Neighbor Embedding (t-SNE) and cluster analysis

t-distributed Stochastic Neighbor Embedding (t-SNE) is a state-of-the-art embedding technique to visualize the similarity of temporal patterns (Maaten and Hinton 2008), and can be used to project time series onto a 2-Dimensional map. t-SNE outperforms conventional dimensionality reduction methods, as it produces low-dimensional representations that preserve the local and global structure of high dimensional data, and effectively improves the separability of datapoints (Maaten and Hinton 2008; Tang and Carey 2022). Time series sharing similar temporal patterns tend to stay clustered on the t-SNE map (Tang and Carey 2022).

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN; McInnes et al, 2017) is a density-based clustering algorithm widely used for identifying clusters or groups within data, especially when dealing with datasets of varying densities and shapes. Unlike traditional clustering methods, HDBSCAN integrates density-based clustering principles, allowing it to identify clusters of varying sizes while also classifying noise points as outliers. It operates by constructing a hierarchical cluster tree based on density-reachability, enabling the algorithm to effectively identify clusters and label points as core, border, or noise, providing a comprehensive analysis of complex data structures. HDBSCAN's versatility and robustness make it particularly advantageous for uncovering meaningful clusters within datasets that may contain irregular shapes or differing densities. In time-series analyses, the HDBSCAN algorithm can be employed to identify major temporal patterns within the dataset.

The algorithm is used in the Wetland Function recovery tool to help identify the time series trends present on the landscape. It is assumed that the time series of each cluster (returned by the HDBSCAN algorithm) will exhibit a distinctive temporal pattern that can then be attributed with a specific wetland function. An average time series can then be generated from the individual time series' associated to each cluster, as the average time series usually shows a more stable temporal trend than any individual observation. These average time series can then be used as reference for the wetland functions of interest.

See: https://people-er.github.io/Wetland-Function-Assessment/04_cluster_analysis/

Dynamic Time Warping

The Wetland Function Recovery tool classifies landscape objects into specific wetland functions based on the degree of similarity between their time series and the reference time series of each wetland function. Each landscape object is assigned to the wetland function with the most similar trend. The similarity of time series is measured by Dynamic Time Warping (DTW) (Salvador and Chan 2007). The calculation is implemented using the fastdw Python library (version 0.3.4). The DTW algorithm returns a similarity metric between each object and each wetland function. A similarity metric threshold can then be defined by the user, and is used to determine which wetland function each landscape unit displays. If a landscape unit has a similarity metric lower than the threshold with all reference time series, it will be binned into an "unknown" wetland function. Each landscape unit's wetland function should be determined annually and the change of a landscape unit between wetland functions indicates a potential change in wetland function.

$$\mathrm{DTW}(s,t) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(d(s_i,t_j)\right)^2}$$

where s_i and t_j denote the elements in the time sequences of $[s_1, s_2, \ldots, s_n]$ and $[t_1, t_2, \ldots, t_m]$, and $d(s_i, t_j)$ represents the edulicadian distance between elements s_i and t_j.

See: https://people-er.github.io/Wetland-Function-Assessment/05_calculate_dtw/

4.3 USE CASES

The primary use case for the PEOPLE-ER Wetland and Wetness Trends toolset is a user that has completed a wetland restoration project targeting the restoration of natural wetland functions.

The user may wish to:

- Conduct baseline characterization – assess a wetland's hydrological regime prior to an intervention.
- Conduct reference site or target regime characterization – assess a wetland's hydrological regime to use as a reference site.
- Compare restored wetland areas to the pre-restoration reference period or reference sites. They will be able to visualize and quantify the similarity of the restored hydrological regime to the reference site.

The length of the time series must capture the periodicity and variability of wetland inundation (key target function) and provide the user the ability to identify changes for individual wetland features as well as for the entire landscape. Therefore, associated to this use case is the segmentation of the wetland landscape into meaningful features or objects for analysis.

In Scope / Out of Scope

In scope features of the PEOPLE-ER Wetland and Wetness Trends toolset are:

- Landscape segmentation: demonstrate how to break up the landscape into meaningful units of analysis.
- Sentinel-1 time series clustering: demonstrate how to cluster their time series to explore their data and better understand existing trends within the landscape of interest.
- Sentinel-1 time series classification: show users how to classify each landscape unit based on its time series' similarity to a reference wetland function.

Out of scope features of the PEOPLE-ER Wetland and Wetness Trends toolset are:

- Multi-spectral image cloud-free compositing.
- SAR image time series pre-processing and stacking.
- Spatial or temporal delineation of a reference(s) to use as a baseline.

Preconditions

- User knowledge about and ability to delineate a restoration area
- User knowledge of the date of restoration interventions to determine a reference period (if preferred assessment option)
- User knowledge of the date of restoration intervention to determine an adequate reference polygon (if preferred assessment option)
- Landscape objects to complete the analysis (can be created, if needed)

4.4 IMPLEMENTATION

Platform Environment

The PEOPLE-ER Wetland and Wetness Trends is developed using the Python 3 programming language because it provides ease-of-use, a breadth of high-quality packages, and is now one of the standard languages in the world of open-source scientific computing. To facilitate scalable analysis, the following third-party Python packages are used to implement data models and Wetland and Wetness trend evaluation algorithms:

- Dask: enables work with large datasets by enabling a distributed workflow and taking advantage of scalable cloud-computing
- XArray: enables the addition of labels in the form of dimensions, coordinates, and attributes on top of NumPy-like arrays making it easier for users to work with multi-band raster datasets
- SciPy – provides a multitude of algorithms that the tool will use, KMeans, t-SNE, and Scharr edge-detection

These packages, and any additional packages, are selected so as to be platform-independent and interoperable with the scientific Python ecosystem and open-source scientific computing initiatives like the Pangeo project. Thus, any processing platforms which support similar ecosystems will be able to support the PEOPLE-ER Spectral Recovery tool. Specific platforms that are compatible are Microsoft's Planetary Computer Hub, VTT's Forestry TEP platform, and the new Copernicus Data Access Service.

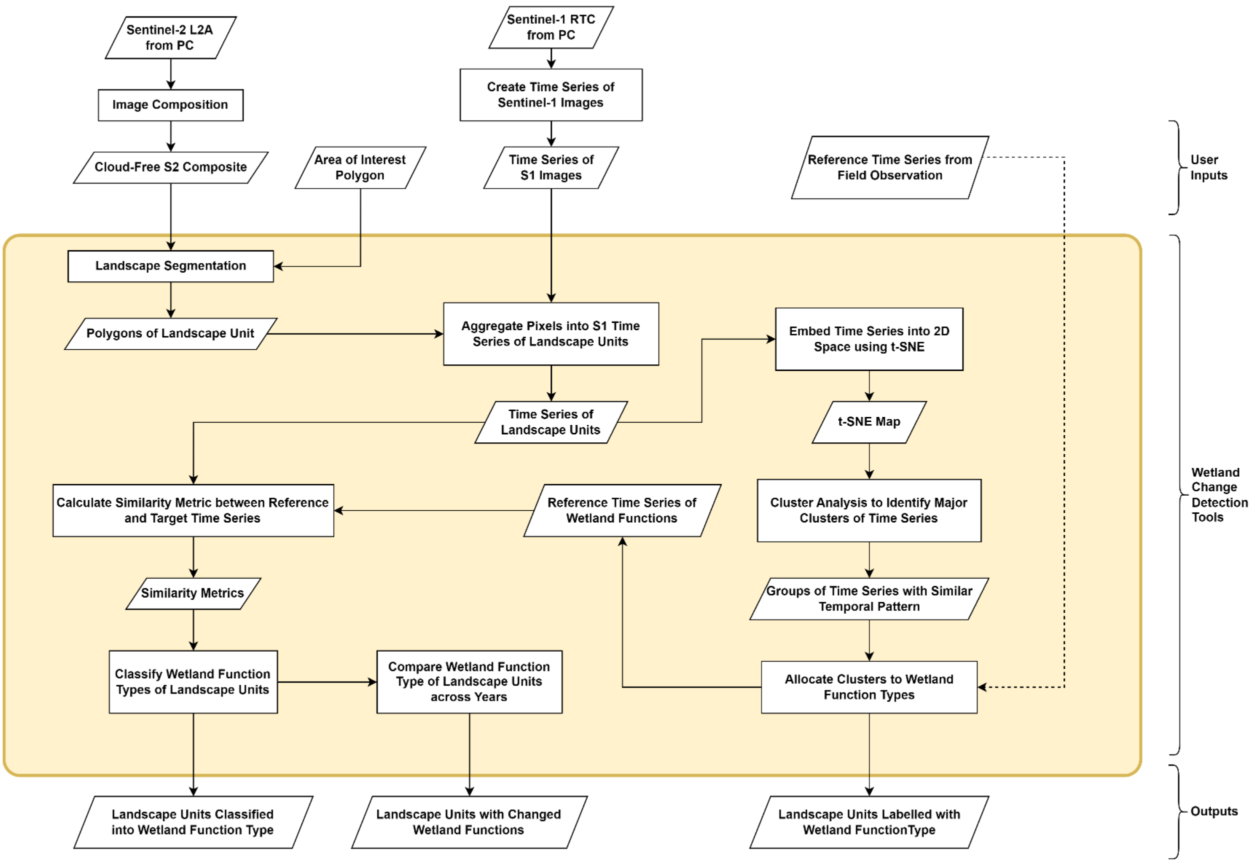
Inputs

- Cloud-Free Sentinel-2 images, or Cloud-Free composites, only using the 10m high-resolution bands (RGB and NIR)
- Time series of Sentinel-1 images
- An optional reference time series

Workflow

The workflow of the process is illustrated in Figure 7.

Figure 7 Wetland and wetness trends tool workflow.



Outputs

- tSNE map displaying the clusters returned by the HDBSCAN clustering method (Figure 8).
- Graphs displaying clustered time series profiles to support interpretation of wetland function types (Figure 9).
- Landscape units classified by wetland function type – result of the DTW classification that matches each landscape unit's time series to a reference wetland function time series (Figure 10).
- Landscape units whose wetland function have changed over time – a visual layer depicting landscape units whose wetland function has changed over the course of the studied time period.

Figure 8 Example output from Wetland Function tool – tSNE map with HDBSCAN clustering.

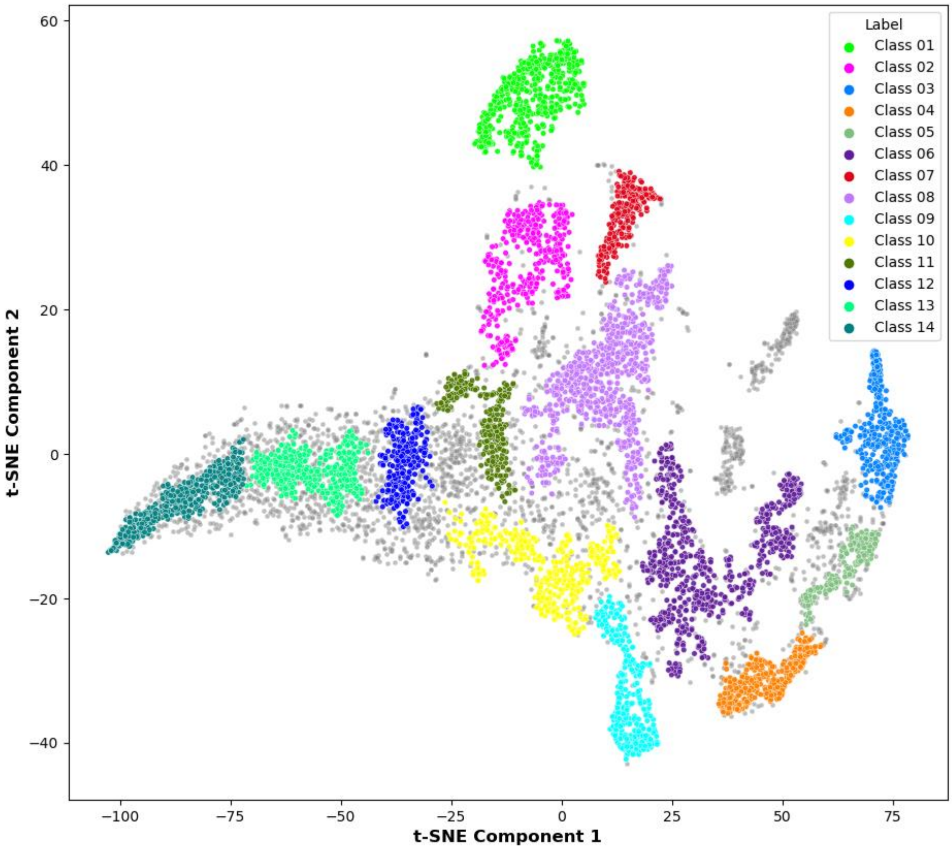


Figure 9 Example output from Wetland *Function tool – time series of tSNE clusters.*

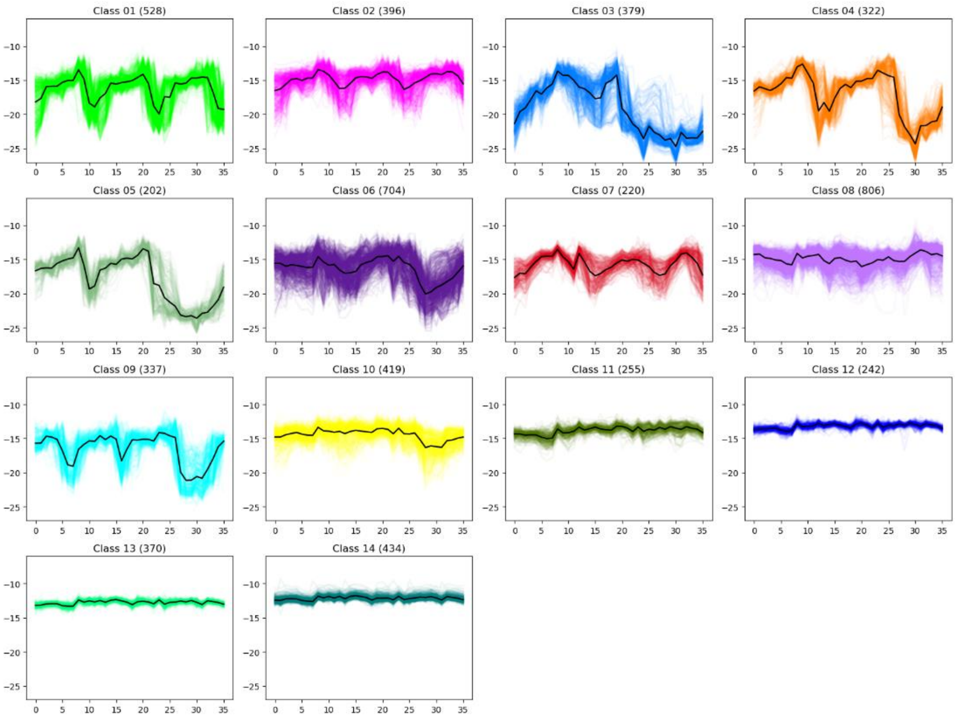
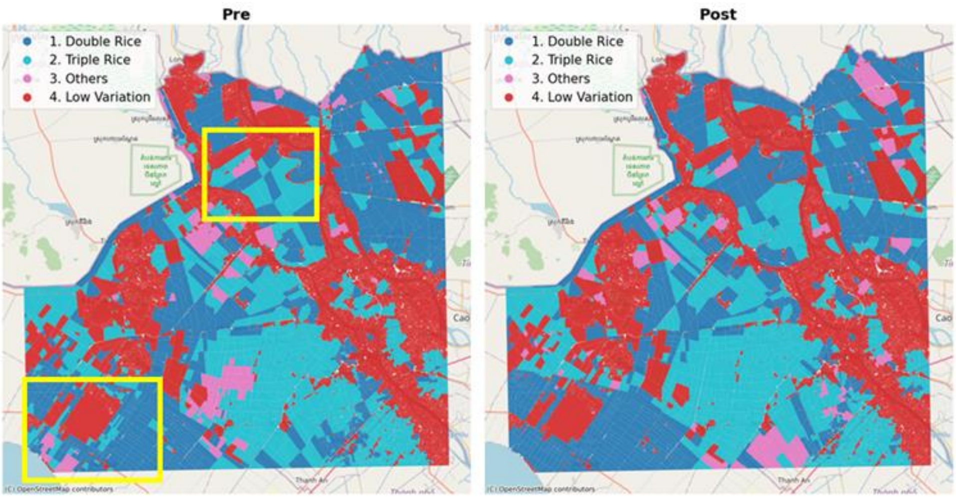


Figure 10 Example output from Wetland Function tool – classification of floodplain connectivity.



Limitations

Limitations of the PEOPLE-ER Wetland and Wetness Trends toolset are:

- Minimum mapping unit – Sentinel-1 and Sentinel-2 provide high spatial resolution 10 m data. This is typically sufficient to characterize complex wetland landscapes, but the analysis is best implemented for landscape features/objects with a minimum mapping unit of at least 1 ha.
- Vegetation density affects the ability of Sentinel-1 to detect important surface hydrological processes, meaning that certain wetland functions may not be captured in densely forested areas. The analytical processes should be applicable to other SAR sensors, i.e., L-band with longer wavelengths and greater vegetation penetration

- Inter-annual variability may influence the results of the tool, as the natural functions of each individual unit can change based on climate, and weather patterns.

4.5 TESTING AND VALIDATION

The testing and validation of the tool was conducted in the Mekong Delta area of Vietnam. The validation results are reported in Deliverable D6 (Results and Validation) available from https://www.people-er.info/

4.6 FUTURE WORK

Future options to improve the initial solution:

- Improved Sentinel-2 segmentation methods and defined methods to validate the results of segmentation (e.g., compared to other datasets in a validation area).
- Address inter-annual variability using annual hydrographs (where available). For example, consistency with annual hydrographs (particularly during the flood season) could provide supplementary information when determining wetland types and functions.

5.0 REFERENCES

Alt H. 2001. The Nearest Neighbor. In: Alt H, editor. Computational Discrete Mathematics. Vol. 2122. Berlin, Heidelberg: Springer Berlin Heidelberg. (Goos G, Hartmanis J, Van Leeuwen J, editors. Lecture Notes in Computer Science). p. 13–24. [accessed 2023 Jun 28]. http://link.springer.com/10.1007/3-540-45506-X_2.

Banskota A, Kayastha N, Falkowski MJ, Wulder MA, Froese RE, White JC. 2014. Forest Monitoring Using Landsat Time Series Data: A Review. Canadian Journal of Remote Sensing. 40(5):362–384. doi:10.1080/07038992.2014.987376.

Cabello J, Fernández N, Alcaraz-segura D, Oyonarte C, Piñeiro G, Altesor A, Delibes M, Paruelo JM. 2012. The ecosystem functioning dimension in conservation: insights from remote sensing. Biodiversity & Conservation. 21(13):3287–3305. doi:10.1007/s10531-012-0370-7.

Chirici G, Mura M, McInerney D, Py N, Tomppo EO, Waser LT, Travaglini D, McRoberts RE. 2016. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. Remote Sensing of Environment. 176:282–294. doi:10.1016/j.rse.2016.02.001.

Cohen WB, Yang Z, Healey SP, Kennedy RE, Gorelick N. 2018. A LandTrendr multispectral ensemble for forest disturbance detection. Remote Sensing of Environment. 205:131–140. doi:10.1016/j.rse.2017.11.015.

Cordell S, Questad EJ, Asner GP, Kinney KM, Thaxton JM, Uowolo A, Brooks S, Chynoweth MW. 2017. Remote sensing for restoration planning: how the big picture can inform stakeholders. Restoration Ecology. 25(S2):S147–S154. doi:10.1111/rec.12448.

De Keersmaecker W, Rodríguez-Sánchez P, Milencović M, Herold M, Reiche J, Verbesselt J. 2022. Evaluating recovery metrics derived from optical time series over tropical forest ecosystems. Remote Sensing of Environment. 274:112991. doi:10.1016/j.rse.2022.112991.

Department of Environment and Science, Queensland. 2021. Hydrology – Land unit and subunit. WetlandInfo website. https://wetlandinfo.des.qld.gov.au/wetlands/ecology/processes-systems/water/hydrology/habitat.html.

FAO. 2018. FRA 2020 Terms and Definitions. https://www.fao.org/3/I8661EN/i8661en.pdf.

Frazier RJ, Coops NC, Wulder MA, Hermosilla T, White JC. 2018. Analyzing spatial and temporal variability in short-term rates of post-fire vegetation return from Landsat time series. Remote Sensing of Environment. 205:32–45. doi:10.1016/j.rse.2017.11.007.

Gann GD, McDonald T, Walder B, Aronson J, Nelson CR, Jonson J, Hallett JG, Eisenberg C, Guariguata MR, Liu J, et al. 2019. International principles and standards for the practice of ecological restoration. Second edition. Restoration Ecology. 27(S1):S1–S46. doi:10.1111/rec.13035.

Gao F, Ma F, Wang J, Sun J, Yang E, Zhou H. 2018. Visual Saliency Modeling for River Detection in High-Resolution SAR Imagery. IEEE Access. 6:1000–1014. doi:10.1109/ACCESS.2017.2777444.

Hagolle O, Huc M, Pascual DV, Dedieu G. 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images. Remote Sensing of Environment. 114(8):1747–1755. doi:10.1016/j.rse.2010.03.002.

Hame T, Kilpi J, Ahola HA, Rauste Y, Antropov O, Rautiainen M, Sirro L, Bounpone S. 2013. Improved Mapping of Tropical Forests With Optical and SAR Imagery, Part I: Forest Cover and Accuracy Assessment Using Multi-Resolution Data. IEEE J Sel Top Appl Earth Observations Remote Sensing. 6(1):74–91. doi:10.1109/JSTARS.2013.2241019.

Huang S, Tang L, Hupy JP, Wang Y, Shao G. 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J For Res. 32(1):1–6. doi:10.1007/s11676-020-01155-1.

Kangas A, Astrup R, Breidenbach J, Fridman J, Gobakken T, Korhonen KT, Maltamo M, Nilsson M, Nord-Larsen T, Næsset E, et al. 2018. Remote sensing and forest inventories in Nordic countries – roadmap for the future. Scandinavian Journal of Forest Research. 33(4):397–412. doi:10.1080/02827581.2017.1416666.

Kennedy RE, Yang Z, Cohen WB. 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sensing of Environment. 114(12):2897–2910. doi:10.1016/j.rse.2010.07.008.

Kennedy RE, Yang Z, Cohen WB, Pfaff E, Braaten J, Nelson P. 2012. Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. Remote Sensing of Environment. 122:117–133. doi:10.1016/j.rse.2011.09.024.

Lamarche C, Santoro M, Bontemps S, d'Andrimont R, Radoux J, Giustarini L, Brockmann C, Wevers J, Defourny P, Arino O. 2017. Compilation and Validation of SAR and Optical Data Products for a Complete and Global Map of Inland/Ocean Water Tailored to the Climate Modeling Community. Remote Sensing. 9(1):36. doi:10.3390/rs9010036.

Maaten L van der, Hinton G. 2008. Visualizing Data using t-SNE. Journal of Machine Learning Research. 9(86):2579–2605.

Mäkisara K, Katila M, Peräsaari J. 2022. The Multi-Source National Forest Inventory of Finland - methods and results 2017 and 2019. Natural Resources Institute Finland (Luke) Report No.: 90/2022.

Montero D, Aybar C, Mahecha MD, 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. doi:10.1038/s41597-023-02096-0.

Pekel J-F, Cottam A, Gorelick N, Belward AS. 2016. High-resolution mapping of global surface water and its long-term changes. Nature. 540(7633):418–422. doi:10.1038/nature20584.

Pickell PD, Hermosilla T, Frazier RJ, Coops NC, Wulder MA. 2016. Forest recovery trends derived from Landsat time series for North American boreal forests. International Journal of Remote Sensing. 37(1):138–149. doi:10.1080/2150704X.2015.1126375.

Ruiz-Jaen MC, Mitchell Aide T. 2005. Restoration Success: How Is It Being Measured? Restoration Ecology. 13(3):569–577. doi:10.1111/j.1526-100X.2005.00072.x.

Salvador S, Chan P. 2007. Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis. 11(5):561–580. doi:10.3233/IDA-2007-11508.

Schroeder TA, Wulder MA, Healey SP, Moisen GG. 2011. Mapping wildfire and clearcut harvest disturbances in boreal forests with Landsat time series data. Remote Sensing of Environment. 115(6):1421–1433. doi:10.1016/j.rse.2011.01.022.

Senf C. 2022. Seeing the System from Above: The Use and Potential of Remote Sensing for Studying Ecosystem Dynamics. Ecosystems. 25(8):1719–1737. doi:10.1007/s10021-022-00777-2.

Skakun S, Vermote EF, Roger J-C, Justice CO, Masek JG. 2019. Validation of the LaSRC Cloud Detection Algorithm for Landsat 8 Images. IEEE J Sel Top Appl Earth Observations Remote Sensing. 12(7):2439–2446. doi:10.1109/JSTARS.2019.2894553.

Skidmore AK, Coops NC, Neinavaz E, Ali A, Schaepman ME, Paganini M, Kissling WD, Vihervaara P, Darvishzadeh R, Feilhauer H, et al. 2021. Priority list of biodiversity metrics to observe from space. Nat Ecol Evol. 5(7):896–906. doi:10.1038/s41559-021-01451-x.

Tang W, Carey SK. 2022. Classifying annual daily hydrographs in Western North America using T-DISTRIBUTED stochastic neighbour embedding. Hydrological Processes. 36(1):e14473. doi:10.1002/hyp.14473.

Tarrio K, Tang X, Masek JG, Claverie M, Ju J, Qiu S, Zhu Z, Woodcock CE. 2020. Comparison of cloud detection algorithms for Sentinel-2 imagery. Science of Remote Sensing. 2:100010. doi:10.1016/j.srs.2020.100010.

Tokola T, Pitkänen J, Partinen S, Muinonen E. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing. 17(12):2333–2351. doi:10.1080/01431169608948776.

Tomppo E, Katila M. 1991. Satellite image-based national forest inventory of finland for publication in the igarss’91 digest. In: [Proceedings] IGARSS’91 Remote Sensing: Global Monitoring for Earth Management. Vol. 3. Espoo, Finland: IEEE. p. 1141–1144. [accessed 2023 Jun 28]. http://ieeexplore.ieee.org/document/579272/.

Vermote E, Justice C, Claverie M, Franch B. 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment. 185:46–56. doi:10.1016/j.rse.2016.04.008.

Watkins B, Van Niekerk A. 2019. A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal Sentinel-2 imagery. Computers and Electronics in Agriculture. 158:294–302. doi:10.1016/j.compag.2019.02.009.

White JC, Hermosilla T, Wulder MA, Coops NC. 2022. Mapping, validating, and interpreting spatio-temporal trends in post-disturbance forest recovery. Remote Sensing of Environment. 271:112904. doi:10.1016/j.rse.2022.112904.

White JC, Saarinen N, Kankare V, Wulder MA, Hermosilla T, Coops NC, Pickell PD, Holopainen M, Hyyppä J, Vastaranta M. 2018. Confirmation of post-harvest spectral recovery from Landsat time series using measures of forest cover and height derived from airborne laser scanning data. Remote Sensing of Environment. 216:262–275. doi:10.1016/j.rse.2018.07.004.

White JC, Wulder MA, Hermosilla T, Coops NC, Hobart GW. 2017. A nationwide annual characterization of 25years of forest disturbance and recovery for Canada using Landsat time series. Remote Sensing of Environment. 194:303–321. doi:10.1016/j.rse.2017.03.035.

Woodcock CE, Allen R, Anderson M, Belward A, Bindschadler R, Cohen W, Gao F, Goward SN, Helder D, Helmer E. 2008. Free access to Landsat imagery. Science. 320(5879):1011–1011.

Woodcock CE, Loveland TR, Herold M, Bauer ME. 2020. Transitioning from change detection to monitoring with remote sensing: A paradigm shift. Remote Sensing of Environment. 238:111558. doi:10.1016/j.rse.2019.111558.

Wulder M, Masek J, Cohen W, Loveland T, Woodcock C. 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment. 122:2–10. doi:10.1016/j.rse.2012.01.010.

Wulder MA, Loveland TR, Roy DP, Crawford CJ, Masek JG, Woodcock CE, Allen RG, Anderson MC, Belward AS, Cohen WB, et al. 2019. Current status of Landsat program, science, and applications. Remote Sensing of Environment. 225:127–147. doi:10.1016/j.rse.2019.02.015.

Xu S, Zhu Xiaolin, Chen J, Zhu Xuelin, Duan M, Qiu B, Wan L, Tan X, Xu YN, Cao R. 2023. A robust index to extract paddy fields in cloudy regions from SAR time series. Remote Sensing of Environment. 285:113374. doi:10.1016/j.rse.2022.113374.

Zeng Y, Hao D, Huete A, Dechant B, Berry J, Chen JM, Joiner J, Frankenberg C, Bond-Lamberty B, Ryu Y, et al. 2022. Optical vegetation indices for monitoring terrestrial ecosystems globally. Nat Rev Earth Environ. 3(7):477–493. doi:10.1038/s43017-022-00298-5.

Zhu Z, Woodcock CE. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote sensing of environment. 118:83–94.

Zhu Z, Woodcock CE. 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment. 144:152–171. doi:10.1016/j.rse.2014.01.011.

Zupanc A. 2017. Improving Cloud Detection with Machine Learning. Sentinel Hub Blog. [accessed 2023 Jan 4]. https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13.