European Vegetation Archive

Data Request Form

To obtain data from the European Vegetation Archive (EVA), including the ReSurveyEurope Database, please first enquire the EVA database administrator Ilona Knollová (ikuzel@sci.muni.cz) whether the data that meet your needs are available. If they are, please fill in the form below and submit it to Ilona or another member of the EVA Coordinating Board (or ReSurveyEurope Board if you ask for data from the ReSurveyEurope Database).

- Applicant's name:
  Zeke Marshall
- Applicant's institutional address:
  UK Centre for Ecology & Hydrology, Lancaster, United Kingdom
- Applicant's e-mail:
  zekmar@ceh.ac.uk
- Project title:
  Ecological Niche Models for the Plants of Europe
- Are you asking for core EVA data (non-repeated vegetation surveys) or for ReSurveyEurope data (repeated vegetation surveys)?
  EVA
- Brief description of the aims and methods of the study:
  The aim of the project is to produce two sets of realised ecological niche models; classification-based occurrence models and regression-based cover models.

  The construct these models, the Hutchinsonian conceptualisation of the ecological niche as an N-dimensional space will be adopted. The environmental variables comprising this space will consist of community-weighted light, moisture, reaction, nutrients, salinity, grazing pressure, and soil disturbance values. In addition to these plot-derived metrics, three climatic variables retrieved from Copernicus Climate Change Service dataset "Downscaled bioclimatic indicators for selected regions from 1950 to 2100 derived from climate projections", namely minimum January temperature, maximum July temperature, and annual precipation will be used.

  The models will be created using Gaussian processes, allowing for the creation of classification models with probabilistic predictions, and the creation of regression models with probabilistic predictions and associated uncertainties. The use of Gaussian processes for ecological niche modelling has been limited by their computational demands, however the use of undersampling methods which preserve the geometries of the presence and absence hypervolumes and use of a Laplacian kernel provide a path to access the benefits of Gaussian processes for large vegetation plot datasets.

  Model performance will be assessed through two sets of prevalence/imbalance-insensitive performance metrics, with a subset of models assessed qualititatively for

their ecological realism using interpretable machine learning outputs such as accumulated local effect plots.

The models will be made available through atleast two means: 1) integrated into the RMAVIS platform, and 2) as an R package.

It is expected that a significant proportion of extant European vascular plants, and hopefully bryophytes, can be modelled; this will hopefully provide a means to predict the occurrence and habitat suitability of thousands of taxa considering the affects of multiple interacting drivers.

- Will someone else be involved in data editing or analysis in addition to the applicant?

I will be the sole person performing the formal analysis and having access to the data, though there will be other paper co-authors.

- Estimated time of delivery of results (e.g., manuscript submission):

April 2025.

- Geographic area needed (e.g., countries or range of geographic coordinates):

All countries in the EVA target area and excluding: Algeria, Egypt, Greenland, Isreal, Kazakhstan (European part), Lebanon, Libya, Mauritania, Morocco, Russia (European part), Svalbard and Jan Mayen, and Tunisia.

- Do you need plots to be georeferenced? If so, what is the minimum accuracy of plot location (in metres or kilometres) needed for your project?

Yes and 10km is the minimum accuracy I would need, ideally 1km.

- Vegetation types needed (syntaxa):

All.

- Other data selection criteria:

-

- Envisaged publications:

A Journal of Vegetation Science article entitled *Ecological Niche Models for the Plants of Europe*

- Data deposition. Some journals require data used for the analysis to be stored in a public repository to ensure the repeatability of the analysis. According to EVA Rules, you are not allowed to store the original vegetation-plot data obtained from EVA. However, if you plan to publish in such a journal, you may deposit a reduced EVA-derived dataset that (1) would make it possible to repeat the analysis published in the paper and (2) does not contain any information not used in the analysis. For example, such a dataset can contain only a subset of species (e.g., only angiosperms or only neophytes), or replace species names with codes, or replace species cover values with presences/absences, or remove all the header data, or replace the exact plot coordinates by coarse grid-cell coordinates etc. If you plan to deposit reduced information from vegetation plots, please describe here what might be deposited. If the project developed so that you

needed to deposit more information than specified here, you would need to ask specific permission from the Custodians of the EVA databases used in your analysis before the dataset is deposited.

I may include a subset of the training data to allow for partial reproduction of the results, but this would be data that I hold for Britain.

- Plant trait data from the TRY consortium. If you plan to combine your analysis of vegetation-plot data with plant trait data, you can also request a dataset of 18 gap-filled traits for a large number of plant taxa prepared by the TRY consortium. These traits include Leaf area, Specific leaf area, Leaf fresh mass, Leaf dry matter content, Leaf C, Leaf N, Leaf P, Leaf N per area, Leaf N:P ratio, Leaf delta15N, Seed mass, Seed length, Seed number per reproductive unit, Dispersal unit length, Plant height, Stem specific density, Stem conduit density, and Conduit element length. This dataset can be provided to you by the EVA manager together with the vegetation-plot data. If you use this dataset, you must inform about your project the TRY data contributors who might be potentially interested and invite them as potential co-authors, assuming they will make an intellectual contribution to your paper. The list of the TRY data contributors will be sent to you together with the gap-filled trait dataset.

Yes

- Specification of the co-authorship arrangements in publications based on the requested data. Note that the EVA Rules recommend that co-authorship is offered to a representative of each database providing data that are particularly important for the project (e.g., a relatively large proportion of the final dataset used in the analyses or data from unique vegetation types or under-represented geographic areas). This database representative should be an expert in the topic of the project (not necessarily the custodian or deputy custodian), and this person should contribute to the project more than just by providing the existing data, e.g. by intellectual contribution to the concept of the paper, preparation of new data, or helping with data analysis, interpretation of the results or writing parts of the paper (see the IAVS Code of Professional Ethics: https://www.iavs.org/page/governance_code-of-proffesional-ethics). The project leader should enable active participation by regularly informing potential co-authors about the progress of the project from its early stage. The project leader should also make final co-authorship arrangements based on the real input of the individual contributors.

Happy to adhere to the EVA rules detailed above.

- Eligibility of the applicant to receive EVA or ReSurveyEurope data. Specify to which EVA or ReSurveyEurope database the applicant has contributed; if the applicant is not the custodian or deputy custodian of an EVA or ReSurveyEurope database, give a name of a custodian or deputy custodian who supports this data request.

I have not yet contributed to the EVA, but I am in the process of collating ~250,000 vegetation plots from Britain for the establishment of a National Vegetation Plot database for Britain for submission to the EVA in Q1 2025. This request is supported by Milan Chytrý.

- I agree with the terms of EVA Data Property and Governance Rules as approved on 26 May 2012 (http://euroveg.org/download/eva-rules.pdf).
- If I ask for ReSurveyEurope data, I agree with the terms of ReSurveyEurope Data Property and Governance Rules as approved on 6 April 2022 (http://euroveg.org/download/resurveyeurope-rules.pdf).
- In any result obtained based on EVA core data (non-repeated vegetation surveys), I will cite the EVA report article (Chytrý et al. 2016; https://doi.org/10.1111/avsc.12191). In any result obtained based on the ReSurveyEurope data (repeated vegetation surveys), I

will cite the ReSurveyEurope report article as soon as it is published. In addition, I will cite individual source databases used in my project (if possible, in the list of References; if not possible, at least as a list of databases in the electronic supplementary material).

- If I ask for the plant trait data from TRY, I agree to invite to my project the TRY data contributors following the list received from the EVA database manager.

Lancaster, United Kingdom.

Zeke Marshall, 05/12/2024.