EUPollMap: The european atlas of contemporary pollen distribution maps

Fabio Oriani[1,2], Manuel Chevalier[3], and Gregoire Mariethoz[1]

[1]Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne, Switzerland
[2]Agroscope, Remote Sensing Team, Division Agroecology and Environment, CH-8046 Zürich, Switzerland
[3]Institute for Geosciences, University of Bonn, 53115 Bonn, Germany

Correspondence: Fabio Oriani (fabio.oriani@protonmail.com)

Abstract. Contemporary pollen presence data is widely used in environmental research as climate, vegetation, and environmental indicator. To make regional characterizations, a reliable and continuous spatial representation of the pollen presence is needed, but these data are mainly available as point datasets. We present here a comprehensive collection of european pollen presence maps including 143 pollen taxa. These raster maps are realized by interpolation of point data sourced from the EMPD pollen database encompassing Eurasia. An automatic interpolation workflow based on Kriging has been developed to choose an optimal geostatistical model to describe the spatial variability. The output consists in a series of multivariate predictive maps of Europe at 25-km scale, showing the occurrence probability of pollen, the predicted presence, and its uncertainty for each taxon respectively. The visual inspection of the maps and a systematic cross-validation test removing 50% of the data, show that the ensemble of the predictions is reliable for missing-data zones, with a relatively low uncertainty and robust to complex and non-stationary pollen distributions. The maps, freely distributed as GeoTIFF files, are proposed as a ready-to-use tool for spatial geoenvironmental characterization and as a contemporary analog for paleo-reconstructions.

1 Introduction

Pollen data has a primary role in last-generation geoenvironmental studies for paleoclimate (Cao et al., 2021; Herzschuh et al., 2022), paleoenvironmental characterization (Binney et al., 2017; Marcolla et al., 2021; Lu et al., 2022), contemporary vegetation distribution assessment (Leru et al., 2019; Zhao et al., 2022), and atmospheric science (Galveias et al., 2022; Verstraeten et al., 2022; Adams-Groom et al., 2022; Frisk et al., 2022).

Different pollen datasets encompass the European Continent (Giesecke and various contributors, 2007; Davis et al., 2020; Li et al., 2022) and many studies using this kind of data, e.g. for spatial biome reconstruction (Prentice et al., 2000; Binney et al., 2017; Sun et al., 2022), result in distributed point characterizations. This approach is reliable since based on direct measurements, but limited to discontinuous point-cloud areal representations. Conversely, continuous pollen maps would allow a better quantification of the surface extension, the geometry, and the connectivity of the characterized processes. These are important factors in the quantitative analysis of ecosystem and climate evolution, also to relate the obtained spatial analysis with already existing cartographic products used as reference or auxiliary data.

The goal of the present study is to realize, for the first time to our knowledge, a collection of raster maps representing the contemporary long-term distribution of pollen over Europe. The atlas, called EUPollMap, encompasses 143 main pollen taxa, based on the point dataset EMPD v.2 (Davis et al., 2020). The dataset contains modern pollen data from the latest 200 years (see section 2) over Eurasia for a total of 8134 samples, here interpolated in space to generate continuous raster maps.

Realizing raster maps of pollen presence for Europe is not a trivial task for the large domain considered, the data quantity and variability, and the complex spatial distribution of the different taxa. A commonly used polynomial interpolation of the point data would be a simplistic solution, imposing an arbitrary spatial model (e.g. linear, quadratic, or cubic) and not taking into account the uneven distribution of the samples in space. To make a more reliable estimation of the pollen presence in space, the interpolation was based on Kriging, a robust geostatistical technique using a spatial model, representing the observed spatial variability of the data and minimizing the local bias and error variance. An automatic framework has been developed for the data preprocessing, the choice of the best model, and the generation of the maps. For every taxon considered, the output data consists of three raster maps showing respectively: 1) the pollen occurrence probability, 2) the discrete occurrence based on probability thresholds, and 3) the uncertainty of the predictions.

The reliability of the obtained predictive maps is assessed by mean of a cross-validation test removing 50% of the available point data and comparing them with the estimated occurrence probability at the same locations. The obtained pollen maps can be employed in studies on current climate and environment as well as contemporary analogs in paleo reconstructions.

The paper is organized as follows: section 2 describes the dataset used, section 3 the developed methodology, section 4 presents the obtained cartographic product with visual examples and the reliability assessment results, and section 6 is dedicated to the conclusions.

2   Source data

The pollen-presence point data used in this study come from the Eurasian Modern Pollen Database (EMPD) v.2 (Davis et al., 2020), an open-source and participated database including 8134 pollen samples over Eurasia. As specified by the authors, the accepted samples are trap samples with pollen counts averaged over a period of at least 10 years, to represent long-term variability comparable with fossil pollen assemblages from sediments cores.

2.1   Preprocessing

The source data contain raw pollen counts from different source surveys, which cannot assure full consistency due to lack of a standard protocol and different long-term survey conditions. Also, the number of taxa and the geographical frame is too large for the European domain represented in the present atlas. For this reasons, the data were preprocessed improving the consistency of the dataset and pertinence to the study zone. The applied preprocessing workflow consists in the following operations:

- The fields considered for every dataset record are: the sample location coordinates, the pollen count, the taxon name and group ID.

– The data coordinates are transformed from the coordinate reference system (CRS) EPSG:4326 "WGS84" to EPSG:3034 "ETRS89-extended / LCC Europe" to reduce the local deformation for the studied domain.

– Records with missing or invalid coordinates are discarded and sample locations with the same coordinates are merged.

– The dataset is limited to data located inside and near the spatial interpolation grid chosen to define the maps (see metadata in Table 1). The distance limits for data outside the grid is defined as 5% of the total easting length of the map (approximately 242 Km). This assured ending grid points to be contoured by nearby available data, avoiding extrapolation as much as possible.

– Only taxa belonging to a selected series of groups present in the study domain are considered. Following the original dataset grouping, those are (with the corresponding group ID): Dwarf shrubs (DWAR), Herbs (HERB), Liana (LIAN), Palms (PALM), Succulents (SUCC), Trees and Shrubs (TRSH), and Uplands herbs (UPHE). After all the preprocessing steps, not all of them are observed in the final point dataset (see the taxa list in the supplemental material 1).

– The taxa are further grouped into a consolidated pollen taxa series given in the source dataset. This includes pollen types which are not distinguishable in the count, e.g. for different Cedrus species, whose pollen is only distinguished for their sample location. This allows a more consistent point-presence distribution.

– The pollen count is transformed to a binary variable indicating the presence (1) or absence (0) at every location. This datum constitutes a less accurate but more reliable information, less dependent on the long-term uncertainty of the surveys. Based on the hypothesis that all pollen in the trap samples has been detected and counted, any sample location where a taxon has not been detected is considered as an absence datum for that taxon. This results in a constant point dataset for all taxa.

The obtained dataset consists in a total of 143 consolidated taxa (see supplemental material 1), each of them including 5632 point data indicating the presence/absence of pollen (see section 4).

2.2 Reference plant dataset

The visual examples of the obtained pollen spatial interpolations shown in this paper are compared with contemporary plant distributions. For this purpose, plant-presence point data are sourced from the EU-Forest database (Mauri et al., 2017), covering Europe and largely based on national forest-monitoring surveys over 1-km regular grids.

3 Methods

In the following we describe the core interpolation technique as well as the whole workflow developed to generate the pollen presence maps.

3

3.1 The Kriging method

Formalized in the early '60s in the field of georesources (Matheron, 1963), Kriging is still nowadays the standard geostatistical interpolation technique which estimates the target variable as a weighted mean of nearby data values. The weights used in the estimation are computed with the Kriging system, based on the variogram function γ, quantifying the observed variability as a function of the distance between point couples. Given the spatial variable $Z(x)$, defined at n-dimensional coordinates x, with the hypothesis of stationarity (i.e. the statistical properties of Z are uniform in space), the variogram for Z can be estimated on the available data as $\hat{\gamma}(h) = E[(Z(x) - Z(x+h))^2]/2$, where $Z(x)$ and $Z(x+h)$ are any known values of Z at distance h and $E[\cdot]$ is the average operator among all couples of points with a similar h, grouped in discrete h intervals (lags). A parametric variogram model γ is then fitted to the experimental variogram $\hat{\gamma}$, to model the variability of Z in the Kriging system. Examples of experimental and model variograms are visible in figures 1, 2, and 3 panel b. Since long lags are not enough represented by data couples, it is common practice to limit the model fitting to a fixed maximum lag, here set as 3.1e6 m, including 80% of the closest data couples.

We considered here the most common variogram models (γ): linear, power, exponential, spherical, and gaussian (for more details see e.g. Cressie, 1985). Every model is fitted using a least-square approach and the one minimizing the error $|\gamma - \hat{\gamma}|$ is used in the Kriging system. In this study, the system is solved for every raster map grid, using the python package PyKrige (https://pypi.org/project/PyKrige/). Two types of Kriging were considered: Ordinary Kriging (OK), which implies Z to be stationary with no regional trend, and Universal Kriging with external drift (UK), implying the existence of a regional trend, given or estimated. From preliminary tests comparing OK with UK using the elevation as external drift for different taxa (see supplemental material 3), no sensible difference in the resulting interpolations was observed. OK was then adopted as Kriging type, since it allowed including all point data with a contained computational burden and an easy update of the future map releases. As shown in several comparative studies (see e.g. Oriani et al., 2020), OK is overall a robust approach for regional smooth interpolations, minimizing the error variance and bias for missing-data locations.

In this study, the given data for Z are binary, indicating the presence ($Z = 1$) or absence ($Z = 0$) of pollen over an established regular grid (Indicator Kriging). The first Kriging output for Z is the expected value map, varying continuously in space between 0 and 1. The rare values barely outside this interval are approximated to 0 or 1. This map is interpretable as the occurrence probability of pollen. The second output map shows the Kriging variance, which indicates the uncertainty (higher variance) in space of the estimated value. This quantity depends on the data density, their variability, and the variogram model chosen.

3.2 Interpolation workflow

Starting from a preprocessed pollen-presence point dataset for one taxon (section 2.1), the relative map is generated with the following steps implemented in Python:

1. Define the interpolation domain as a rectangular grid in the dataset CRS covering the whole continental European territory (see the metadata, Table 1).

2. If the dataset present all-0 (total absence) or all-1 (total presence) data, generate on the defined grid a 0/1 field as output Kriging mean and a 0 field as output Kriging variance, then go to step 6.

3. Compute the experimental variogram and find the best variogram model for the given point dataset (see section 3.1).

4. Solve the Kriging system with the found model parameters for all points in the interpolation grid. A mask based on the coastal perimeter is used to exclude water bodies from the interpolation. The output is the pollen presence Kriging mean and variance maps.

5. By applying a series of fixed thresholds (0.2, 0.4, 0.5, 0.6, 0.8) on the mean map, generate a discrete pollen occurrence map.

6. Using the open-source python packages gdal (https://gdal.org) and shapefile (http://geospatialpython.com) export a ESRI GeoTIFF file of the georeferenced output maps (see the metadata, Table 1): Kriging mean (band 1), occurrence map (band 2), and uncertainty map (band 3). Also export the attached preprocessed point dataset as ESRI Shapefile.

3.3 Validation

The dataset is first visually examined with map examples presenting different types of spatial distribution and the visual corre-spondence with the corresponding plant distribution is assessed by superposing plant-presence point data from an independent dataset (section 2.2).

The output map composing the present atlas is based on a probabilistic forecast of the pollen presence. To assess its reliability, we used reliability plots (Murphy and Winkler, 1977), a simple quantitative and graphical representation already applied in geo and atmospheric science (Bröcker and Smith, 2007; Allard et al., 2012). The graph is generated by removing part of the data and comparing the predicted probability of occurrence with the occurrence probability observed in the data. For example, for pixels with predicted probability of the pollen occurrence around 0.2, the sample probability computed on the corresponding removed data should be close to 0.2 for the prediction to be reliable. The predicted probability range [0-1] is divided in discrete bins to group the validation locations and co-located occurrence data. Then the sample occurrence probability values are plotted against the associated predicted probabilities. If the plot points lie along the bisector, the predictions can be considered reliable since they correspond to the observed occurrence probability. In some cases the predictions do not cover uniformly the [0-1] range so the reliability graph may in some cases not be representative (Jolliffe and Stephenson, 2012). For this reason, the validation data were increased to 50% of all the available data. The reliability values are displayed in both form of ensemble graph (Figure 4 b) and table containing the same reliability indicator for the different taxa separately (supplemental 1). This allows identifying taxa which do not show reliable predictions for any probability classes.

To analyze the uncertainty of the predictions the average Kriging variance map is used, for which a higher variance indicate the zones in the map where the prediction is less certain. For every separate taxa, this map is part of the output map in the dataset (Table 1 band 3).

4   Output dataset description

The dataset is composed by a multivariate cartographic product covering 143 pollen taxa organized in different folders. The
geographic domain covers the whole continental Europe and the main islands at 25-km resolution. This resolution is related
to the data density and the goal which is a regional long-term representation of the contemporary pollen distribution. For each
taxon, the output data is a multivariate raster map in the GeoTIFF file format including three bands (see the metadata in Table
1 a and the map examples in section 4.1): 1) the pollen occurrence probability map, corresponding to the Kriging interpolation
mean, 2) a discrete pollen occurrence map, obtained by imposing fixed thresholds on 1), and 3) the occurrence uncertainty
map (Kriging variance). Each map file is associated to a georeferenced shapefile with the preprocessed source dataset (Table 1
b) and both can be imported in any GIS software. The shapefile includes 5362 point data indicating the pollen presence with
the attribute POLLEN_PRE. Any taxa folder also contains a preview pdf file of the the output maps, point data and variogram
model plot.

4.1   Map examples

In the following three examples of maps are shown and discussed. Those are representative of the visual quality of the dataset
and serve as examples of map reading. A larger selection of visual examples can be found in the supplemental material 2 and
a quantitative analysis of the map ensemble is present in section 5.

Figure 1 shows the output maps for the taxon Abies, whose pollen is mainly observed in central and eastern Europe. In the
occurrence probability map (Figure 1 a), the superposed source dataset indicates the observed pollen presence with yellow
points and the observed absence in blue, with close neighbor areas respectively in yellow (probability = 1) and blue colors
(probability = 0). Conversely, interpolated areas from presence to absence data, or where these are densely mixed, are occupied
by blue/green shades indicating a probability of pollen occurrence between 0 and 1. Figure 1 b shows the variogram model
used in the Kriging interpolation to estimate the occurrence probability map. As indicated in the legend, the gaussian model
is the one chosen in the automatic setup. The point at about 1e6 m lag called variogram range, where the model reaches the
horizontal asymptote, represents the average correlation length of the data in space.

By imposing thresholds to the probability map, a discrete occurrence map is obtained (Figure 1 c), delimiting zones related
to discrete probability intervals. This version of the probability map is proposed as a ready-to-use tool for practitioners who
want to quantify discrete areas of pollen presence. Frequent Abies pollen presence is represented by yellow and green patches
covering large parts of central and eastern Europe, and northern Russia. This distribution fairly matches the plant distribution
from the considered external dataset (red dots in Figure 1 c). Note that the plant distribution dataset (section 2.2), being
independent and different for data type and sampling, is not used here as a reference for a quantitative validation but rather as
a general term of comparison.

Panel d in Figure 2 shows the Kriging variance map which gives information on the uncertainty of the interpolation. This
indicates the possible variability of the interpolation determined by the distance from the available data, their variability, and

Table 1. Main metadata of the pollen presence maps, contained in each GeoTIFF and Shapefile output files.

a) Raster maps

Name	<taxon name>
CRS	EPSG:3034 - ETRS89-extended / LCC Europe
Extent	1993992.0, 449652.0 : 6843992.0, 5224652.0
Unit	meters
Width	194
Height	191
Pixel size	25000, -25000
Data type	Float32
GDAL Driver Description	GTiff
Data type	GeoTIFF
Band count	3
Band 1	Occurrence probability (Kriging mean)
Band 2	Occurrence map (<= probability thresholds)
Band 3	Occurrence uncertainty (Kriging variance)

b) Point data

Name	<taxon name>
Data type	ESRI Shapefile
Geometry	Point (Point)
CRS	EPSG:3034 - ETRS89-extended / LCC Europe
Unit	meters
Feature count	5362
Attribute Count	1
POLLEN_PRE	String (T=True, F=False)

180   the chosen variogram model. In the bottom part of the map, the lack of neighbor data sensibly increases the uncertainty of the pollen presence estimation.

The second example is Betula (Figure 2), widely present in central Europe, Scandinavia, and northern Russia. The variogram model selected and parametrized by the automatic workflow is a power function tending to linear. As shown in the plot (Figure 2 b), the model fits well the experimental variogram (red stars) in its lower-lag portion, representative of the data variability and

185   used for calibration (see section 3.1). With respect to the previous example (Figure 1), here the model allows longer correlation structures as observed in the large north-south pollen presence zone covering Europe (Figure 2 a, c), matching well the plant distribution data (red dots, Figure 2 c). The eastern part of the map is mainly classified as a low-probability (0-0.4) zone
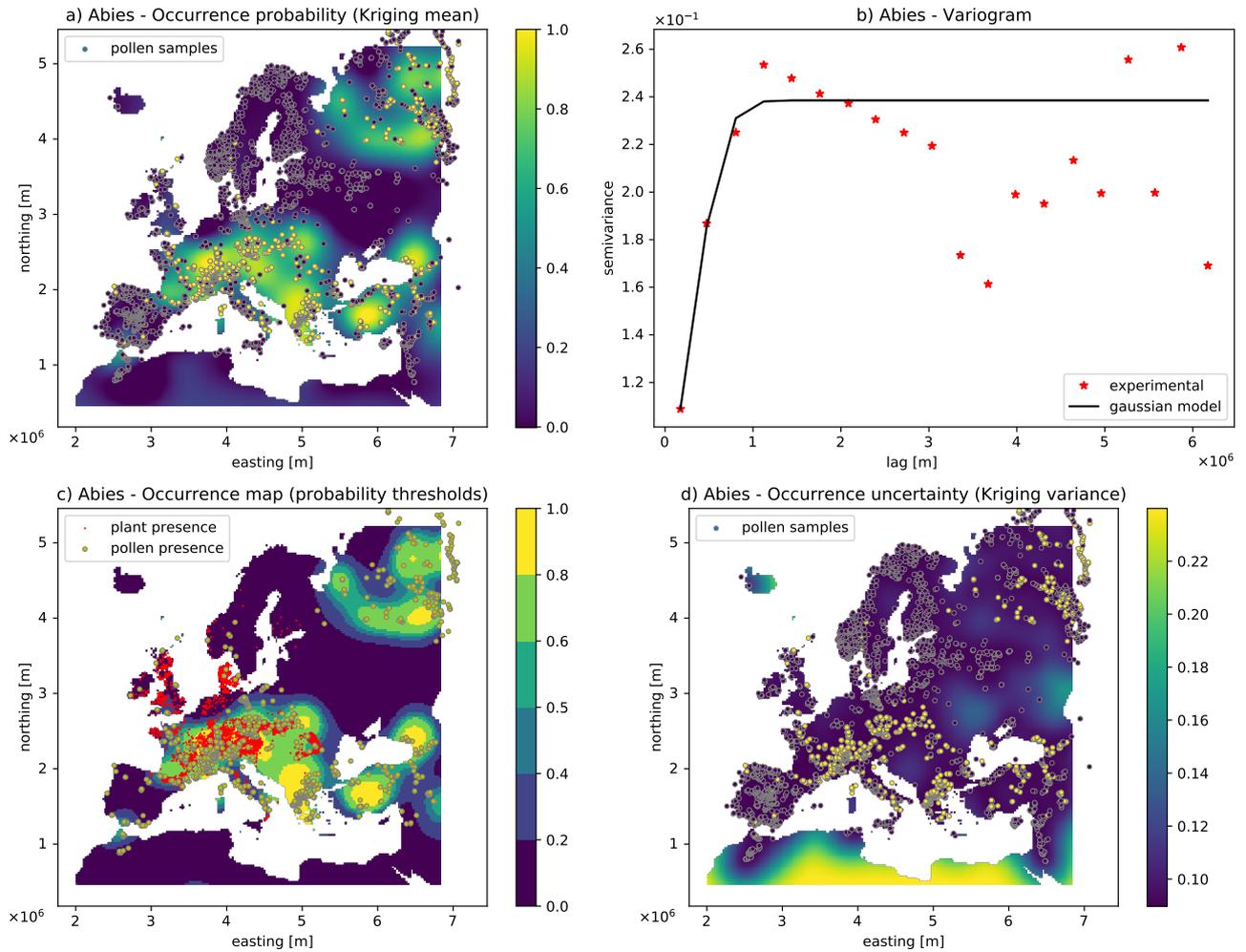
Figure 1. Output maps for Abies: a) Pollen occurrence probability map, b) Variogram model, c) Occurrence map based on probability thresholds, d) Uncertainty map based on the Kriging variance.

for the sporadic pollen detection surrounded by points of non-detection. This local variability moderately increases the model uncertainty (Kriging variance around 0.16, Figure 2 d) with respect to the Abies example (Kriging variance around 0.12, Figure 1 d).

The third example is the Olea pollen distribution (Figure 3), which spans over the whole Mediterranean region and fades uniformly towards center Europe. The gaussian model, here with a larger variogram range until 3e6 m, allows long correlated structures covering the whole bottom of the map from west to east, where the pollen presence is highly probable (Figure 3 a). Towards central Europe, the density of detected pollen points decreases progressively until total absence. Since it mainly
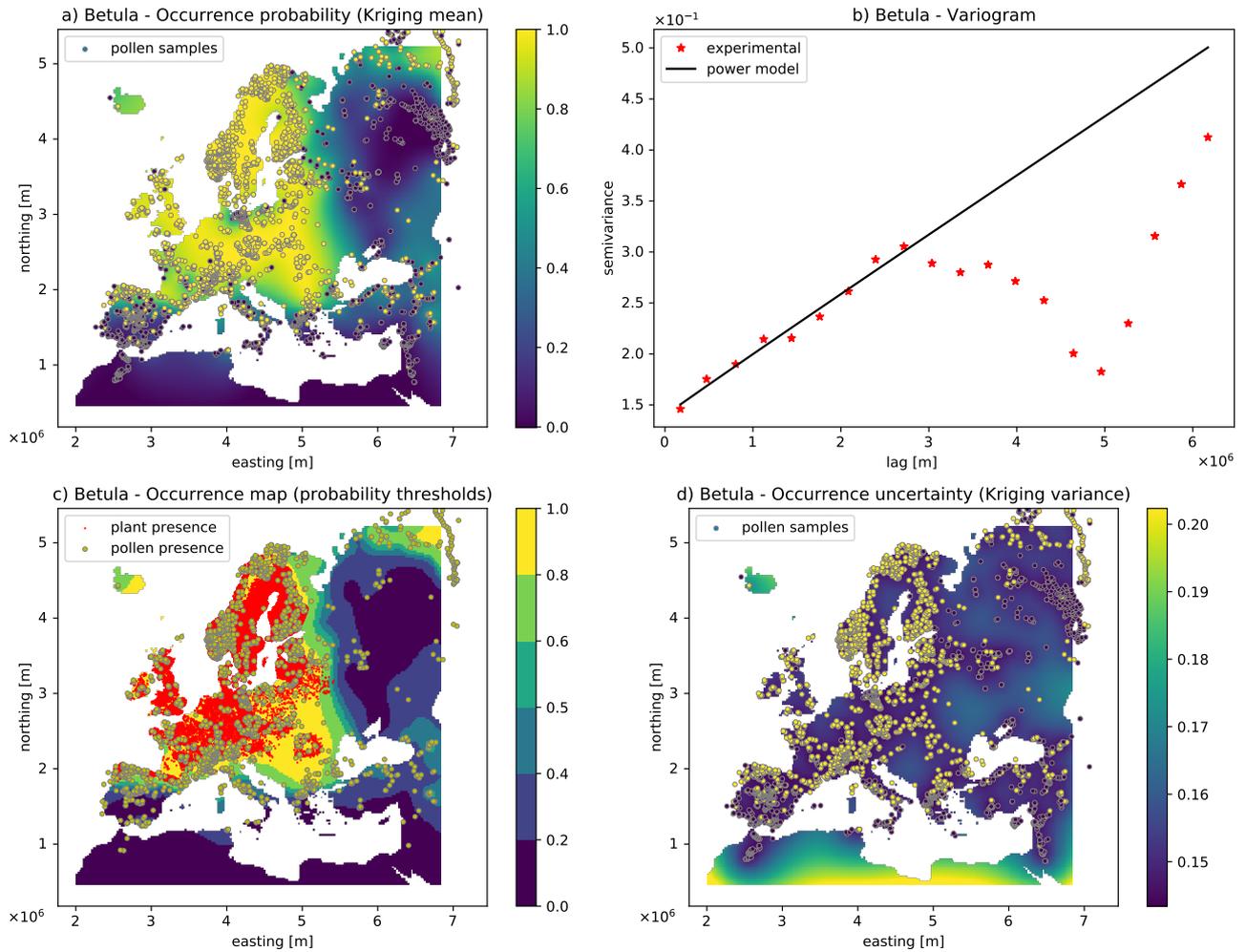
Figure 2. Output maps for Betula: a) Pollen occurrence probability map, b) Variogram model, c) Occurrence map based on probability thresholds, d) Uncertainty map based on the Kriging variance.

follows a regional gradient with no complex structures, the uncertainty on this map (Figure 3 d) is low and uniform in space, excepted for the lower limit of the map, where the lack of data increases the uncertainty.

5 Ensemble reliability assessment

To asses the overall uncertainty of the predictive maps ensemble, the mean Kriging variance map is considered, which is the mean of all taxa variance maps (band 3 in Table 1). With the variance ranging in this case between 0 and 1, the map presents low values in the order of 10e-2 (Figure 4 a), with no zones of high uncertainty in the European continent and its variability
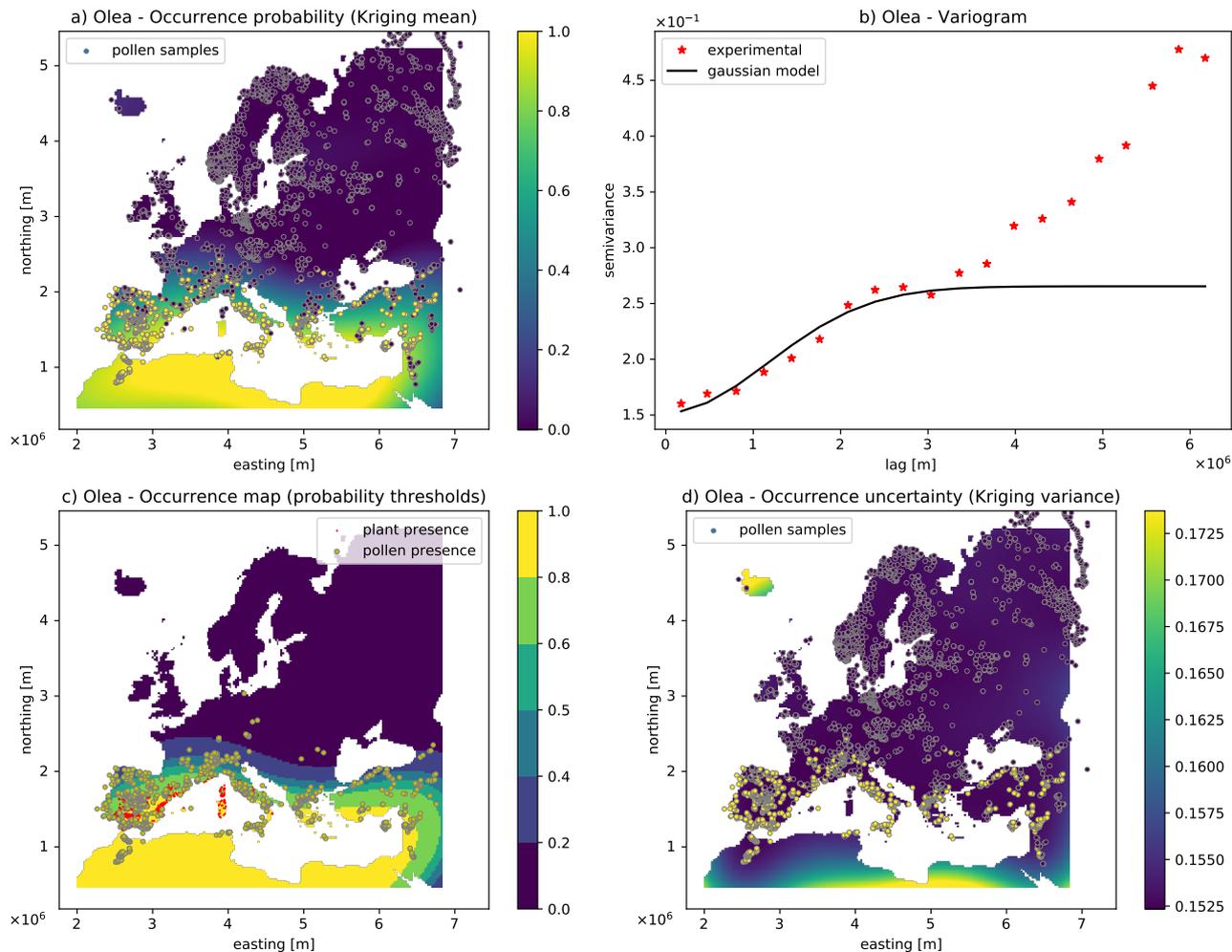
Figure 3. Output maps for Olea: a) Pollen occurrence probability map, b) Variogram model, c) Occurrence map based on probability thresholds, d) Uncertainty map based on the Kriging variance.

controlled by the distance from the data. This suggests the available data provide a statistically accurate information on the pollen distribution over the study zone.

To assess not only the accuracy, but also the reliability of the probabilistic predictions, reliability plots are examined. From section 3.3, recall those are realized by removing 50% of the data and then plotting the predicted probability of pollen occurrence for these locations with the corresponding observed occurrence probability. This rather high amount of removed data may penalize the prediction performance, but it is necessary to observe the pollen occurrence for all predicted probability classes. In the ensemble reliability plot (Figure 4 b), the majority of the taxa (.25-.75 quantile interval) lie along the bisector, meaning the predicted occurrence probability matches well the observed one. Nevertheless, some taxa show a tendency to overestimate (be-
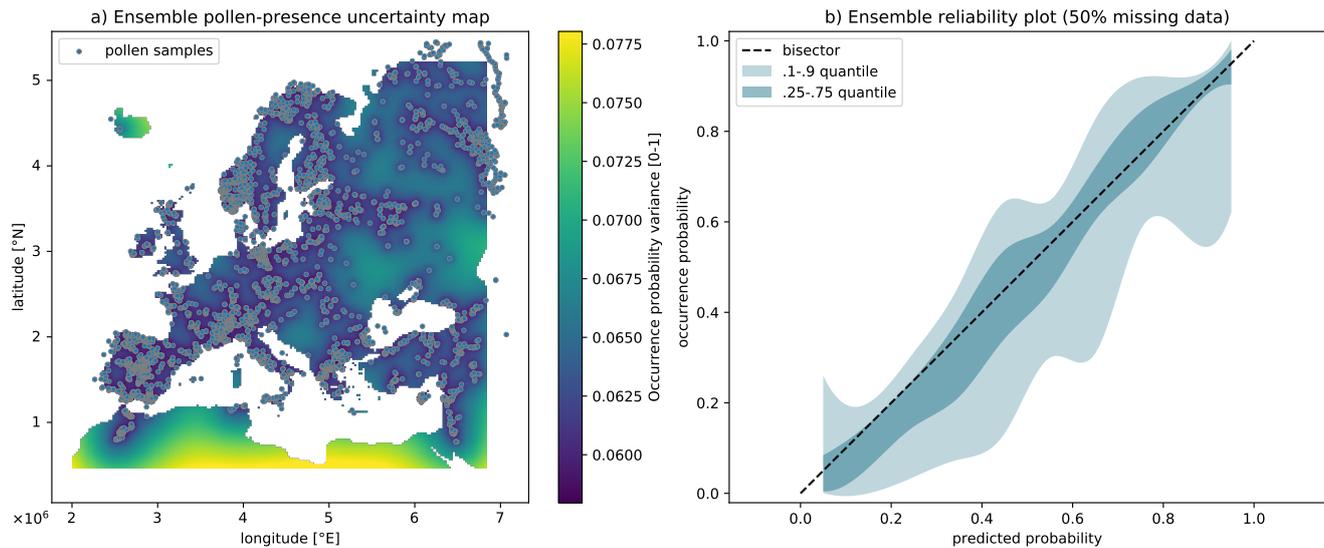
Figure 4. Ensemble statistical indicators for the generated maps: a) Average Kriging Variance map and b) Reliability plot obtained from the cross-validation test removing 50% of the data. The latter included the observed occurrence probability (y axis) for all taxa shows as a distribution for every predicted probability class (x axis).

low the bisector) or underestimate (above the bisector) the pollen occurrence. These cases can be identified in the table present in the supplemental material 1, where the observed occurrence probability is shown for each taxon separately, with the biased values marked in bold. By examining the pollen maps, it resulted that the biased values largely correspond to taxa with poorly represented local pollen variability, where isolated pollen presence data are surrounded by non-detection points (e.g. Myrica, supplemental material 2), or vice versa single non-detection points surrounded by presence points (e.g. Quercus, supplemental material 2). In these cases, important local probability biases can be observed in the cross-validation exercise. This limitation is usually overcome by using all data points, which allow more reliable local estimations. The mentioned maps of Myrica and Quercus seldom show such local bias in their definitive version, realized with all available data (see supplemental material 2).

6 Conclusions

In this paper we presented the European atlas of contemporary pollen-distribution maps (EUPollMap) covering 143 pollen taxa spread over the European continent. Realized starting from the point dataset EMPD v.2, including long-term data samples collected in the latest 200 years, the atlas consists of predictive raster maps for long-term pollen distribution at the 25-km scale. The maps are realized by interpolation of the point dataset using Kriging, a classical geostatistical approach which minimizes the error variance and it is based on a spatial model fitting the observed data variability. The workflow developed here allowed the automatic choice and optimization of the spatial model for every pollen taxon.

11

The accuracy and reliability of the maps are assessed by a cross-validation and a statistical analysis comparing the predicted pollen occurrence probability with the one observed in the data. The results show the map ensemble is overall reliable and robust to non-stationary structures such as regional gradients in the pollen presence or heterogeneous distributions. Nevertheless, the predicted pollen presence can present local bias in case of isolated and poorly represented pollen detection points. The map reliability scores for each taxon are available from the attached supplemental material 1 and its uncertainty from the output map variance (band 2). This allows the user to check the reliability of the raster-map data specifically for the studied taxa and region.

The output maps, available as georeferenced tiff maps, are multivariate, including for each taxa a probability of occurrence map, a discrete pollen presence map, and an uncertainty map. Therefore, this dataset constitutes a ready-to-use cartographic tool for spatial studies on biodiversity and environment which can benefit of both a discrete or probabilistic datum on pollen presence. This includes also paleoclimate and paleoenvironmental reconstructions, where these maps can be used as contemporary analog.

7 Data availability

Dataset name: EUPollMap Version: 1.0

Release date: ...

Developer: Fabio Oriani

Format: ESRI GeoTIFF, ERSI Shapefile

Repository: https://zenodo...

Licence: GNU v.3

References

Adams-Groom, B., Selby, K., Derrett, S., Frisk, C. A., Pashley, C. H., Satchwell, J., King, D., McKenzie, G., and Neilson, R.: Pollen season trends as markers of climate change impact: Betula, Quercus and Poaceae, Science of The Total Environment, 831, 154 882, https://doi.org/10.1016/j.scitotenv.2022.154882, 2022.

Allard, D., Comunian, A., and Renard, P.: Probability Aggregation Methods in Geoscience, Mathematical Geosciences, 44, 545–581, https://doi.org/10.1007/s11004-012-9396-3, 2012.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V.: Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns, Quaternary Science Reviews, 157, 80–97, https://doi.org/10.1016/j.quascirev.2016.11.022, 2017.

Bröcker, J. and Smith, L. A.: Increasing the Reliability of Reliability Diagrams, Weather and Forecasting, 22, 651–661, https://doi.org/10.1175/WAF993.1, 2007.

Cao, X., Tian, F., Li, K., Ni, J., Yu, X., Liu, L., and Wang, N.: Lake surface sediment pollen dataset for the alpine meadow vegetation type from the eastern Tibetan Plateau and its potential in past climate reconstructions, Earth System Science Data, 13, 3525–3537, https://doi.org/10.5194/essd-13-3525-2021, 2021.

Cressie, N.: Fitting variogram models by weighted least squares, Journal of the International Association for Mathematical Geology, 17, 563–586, https://doi.org/10.1007/BF01032109, 1985.

Davis, B. A. S., Chevalier, M., Sommer, P., Carter, V. A., Finsinger, W., Mauri, A., Phelps, L. N., Zanon, M., Abegglen, R., Åkesson, C. M., Alba-Sánchez, F., Anderson, R. S., Antipina, T. G., Atanassova, J. R., Beer, R., Belyanina, N. I., Blyakharchuk, T. A., Borisova, O. K., Bozilova, E., Bukreeva, G., Bunting, M. J., Clò, E., Colombaroli, D., Combourieu-Nebout, N., Desprat, S., Di Rita, F., Djamali, M., Edwards, K. J., Fall, P. L., Feurdean, A., Fletcher, W., Florenzano, A., Furlanetto, G., Gaceur, E., Galimov, A. T., Gałka, M., García-Moreiras, I., Giesecke, T., Grindean, R., Guido, M. A., Gvozdeva, I. G., Herzschuh, U., Hjelle, K. L., Ivanov, S., Jahns, S., Jankovska, V., Jiménez-Moreno, G., Karpińska-Kołaczek, M., Kitaba, I., Kołaczek, P., Lapteva, E. G., Latałowa, M., Lebreton, V., Leroy, S., Leydet, M., Lopatina, D. A., López-Sáez, J. A., Lotter, A. F., Magri, D., Marinova, E., Matthias, I., Mavridou, A., Mercuri, A. M., Mesa-Fernández, J. M., Mikishin, Y. A., Milecka, K., Montanari, C., Morales-Molino, C., Mrotzek, A., Muñoz Sobrino, C., Naidina, O. D., Nakagawa, T., Nielsen, A. B., Novenko, E. Y., Panajiotidis, S., Panova, N. K., Papadopoulou, M., Pardoe, H. S., Pędziszewska, A., Petrenko, T. I., Ramos-Román, M. J., Ravazzi, C., Rösch, M., Ryabogina, N., Sabariego Ruiz, S., Salonen, J. S., Sapelko, T. V., Schofield, J. E., Seppä, H., Shumilovskikh, L., Stivrins, N., Stojakowits, P., Svobodova Svitavska, H., Święta Musznicka, J., Tantau, I., Tinner, W., Tobolski, K., Tonkov, S., Tsakiridou, M., Valsecchi, V., Zanina, O. G., and Zimny, M.: The Eurasian Modern Pollen Database (EMPD), version 2, Earth System Science Data, 12, 2423–2445, https://doi.org/10.5194/essd-12-2423-2020, 2020.

Frisk, C. A., Apangu, G. P., Petch, G. M., Adams-Groom, B., and Skjøth, C. A.: Atmospheric transport reveals grass pollen dispersion distances, Science of The Total Environment, 814, 152 806, https://doi.org/10.1016/j.scitotenv.2021.152806, 2022.

Galveias, A., Ribeiro, H., Guimarães, F., Costa, M. J., Rodrigues, P., Costa, A. R., Abreu, I., and Antunes, C. M.: Differential Quercus spp. pollen-particulate matter interaction is dependent on geographical areas, Science of The Total Environment, 832, 154 892, https://doi.org/10.1016/j.scitotenv.2022.154892, 2022.

Giesecke, T. and various contributors: The European Pollen Database, http://www.europeanpollendatabase.net/, 2007.

285   Herzschuh, U., Böhmer, T., Li, C., Chevalier, M., Dallmeyer, A., Cao, X., Bigelow, N. H., Nazarova, L., Novenko, E. Y., Park, J., Peyron, O., Rudaya, N. A., Schlütz, F., Shumilovskikh, L. S., Tarasov, P. E., Wang, Y., Wen, R., Xu, Q., and Zheng, Z.: LegacyClimate 1.0: A dataset of pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the late Quaternary, Earth System Science Data Discussions, pp. 1–29, https://doi.org/10.5194/essd-2022-38, 2022.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: A Practitioner's Guide in Atmospheric Science, John Wiley & Sons, google-
290   Books-ID: DCxsKQeaBH8C, 2012.

Leru, P. M., Eftimie, A. M., Anton, V. F., and Thibaudon, M.: Assessment of the risks associated with the invasive weed Ambrosia artemisi-iifolia in urban environments in Romania, Ecocycles, 5, 56–61, https://doi.org/10.19040/ecocycles.v5i1.142, 2019.

Li, C., Postl, A. K., Böhmer, T., Cao, X., Dolman, A. M., and Herzschuh, U.: Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0), Earth System Science Data, 14, 1331–1343, https://doi.org/10.5194/essd-14-1331-2022, 2022.

295   Lu, L.-L., Jiao, B.-H., Qin, F., Xie, G., Lu, K.-Q., Li, J.-F., Sun, B., Li, M., Ferguson, D. K., Gao, T.-G., Yao, Y.-F., and Wang, Y.-F.: Artemisia pollen dataset for exploring the potential ecological indicators in deep time, Earth System Science Data Discussions, pp. 1–48, https://doi.org/10.5194/essd-2022-23, 2022.

Marcolla, A., Miola, A., Mozzi, P., Monegato, G., Asioli, A., Pini, R., and Stefani, C.: Middle Pleistocene to Holocene palaeoenviron-mental evolution of the south-eastern Alpine foreland basin from multi-proxy analysis, Quaternary Science Reviews, 259, 106 908,
300   https://doi.org/10.1016/j.quascirev.2021.106908, 2021.

Matheron, G.: Principles of geostatistics, Economic Geology, 58, 1246–1266, https://doi.org/10.2113/gsecongeo.58.8.1246, 1963.

Mauri, A., Strona, G., and San-Miguel-Ayanz, J.: EU-Forest, a high-resolution tree occurrence dataset for Europe, Scientific Data, 4, 160 123, https://doi.org/10.1038/sdata.2016.123, 2017.

Murphy, A. H. and Winkler, R. L.: Reliability of Subjective Probability Forecasts of Precipitation and Temperature, Journal of the Royal
305   Statistical Society. Series C (Applied Statistics), 26, 41–47, https://doi.org/10.2307/2346866, 1977.

Oriani, F., Stisen, S., Demirel, M. C., and Mariethoz, G.: Missing Data Imputation for Multisite Rainfall Networks: A Comparison between Geostatistical Interpolation and Pattern-Based Estimation on Different Terrain Types, Journal of Hydrometeorology, 21, 2325–2341, https://doi.org/10.1175/JHM-D-19-0220.1, 2020.

Prentice, I. C., Jolly, D., and Participants, B. .: Mid-Holocene and glacial-maximum vegetation geography of the northern continents and
310   Africa, Journal of Biogeography, 27, 507–519, https://doi.org/10.1046/j.1365-2699.2000.00425.x, 2000.

Sun, Y., Xu, Q., Gaillard, M.-J., Zhang, S., Li, D., Li, M., Li, Y., Li, X., and Xiao, J.: Pollen-based reconstruction of total land-cover change over the Holocene in the temperate steppe region of China: An attempt to quantify the cover of vegetation and bare ground in the past using a novel approach, CATENA, 214, 106 307, https://doi.org/10.1016/j.catena.2022.106307, 2022.

Verstraeten, W. W., Kouznetsov, R., Hoebeke, L., Bruffaerts, N., Sofiev, M., and Delcloo, A. W.: Reconstructing multi-decadal air-
315   borne birch pollen levels based on NDVI data and a pollen transport model, Agricultural and Forest Meteorology, 320, 108 942, https://doi.org/10.1016/j.agrformet.2022.108942, 2022.

Zhao, Y., Li, Y., Zhang, Z., Fan, B., Zhu, Y., and Zhao, H.: Relationship between modern pollen assemblages and vegetation in the Bashang typical steppe region of North China, Ecological Indicators, 135, 108 581, https://doi.org/10.1016/j.ecolind.2022.108581, 2022.