

UNIGIS

Master Thesis

submitted within the UNIGIS Master`s program
“Geographical Information Science & Systems – (UNIGIS MSc)”
at Department of Geoinformatics - Z_GIS,
Paris Lodron University of Salzburg

Leveraging least-cost modelling to uncover multiple introductions: a case study of an invasive wild bee species

submitted by

BSc Christa Rohrbach

u106871

Supervisor:

Dr. Gudrun Wallentin

A thesis submitted in partial fulfilment of the requirements of
the degree of
“Master of Science”, abbreviated “MSc”

Oberhofen am Thunersee, June 2024

Contents

- Acknowledgements..... 2
- Framework Text..... 3
- 1 Manuscript..... 4
 - 1.1 Introduction 6
 - 1.2 Methods..... 8
 - 1.3 Results..... 15
 - 1.4 Discussion 18
- 2 Report 20
 - 2.1 Introduction 20
 - 2.2 Exploring the data as a time series map 23
 - 2.3 Preparing the presence data (Script module *thinning*) 26
 - 2.4 Least-cost modelling..... 28
 - 2.4.1 Limitations and solutions 28
 - 2.4.2 The complexity of cost surface creation..... 29
 - 2.4.3 Script module *leastcostpaths* (open-source)..... 30
 - 2.4.4 Script modules *arcgis_distacc* and *arcgis_optpaths* (ArcGIS Pro) 32
 - 2.4.5 Esri Bug Report for *Optimal Path As Line* tool..... 33
 - 2.5 Delineating populations (Script module *populations*) 36
 - 2.6 Calculating expansion rate (Script module *expansionrate*)..... 42
- References..... 44

Acknowledgements

First and foremost, I'd like to thank my supervisor Dr. Gudrun Wallentin. Her engagement, helpfulness, and sense of humor made this journey both productive and enjoyable. Julia Lanner, Ph.D., contributed significantly by providing data about my study species and numerous useful references. Her proofreading and insightful advice on scientific writing greatly improved the manuscript. My Python program code was reviewed by Yusuf Khasbulatov, who provided insights to enhance the shareability of my code. I haven't managed to incorporate all the feedback yet, but I'll keep at it, I promise! The environmental suitability map, which I used as input for least-cost modelling, was kindly exported for me by Nicolas Dubos, Ph.D., under challenging circumstances. I would also like to thank Stefano Spadaccia, Ph.D., for occasionally lending me his mathematical expertise. He managed to explain math concepts which ChatGPT had not been able to before. And last but definitely not least, my partner Patrick Doll, who bore with me when I was very absent-minded at times and who was always there for moral support. Thank you all for making this thesis possible.

Framework Text

This master thesis is composed in a manuscript-based format, consisting of two complementary parts: a concise paper manuscript intended for submission to a scientific journal and a detailed report which provides thorough documentation of materials and methods. This dual structure aims to ensure both a clear presentation of the research findings and a detailed documentation of my individual contribution. The manuscript and report are inherently linked, as I tried to avoid duplication of content, and therefore, a comprehensive list of references is included at the end of the document. The paper manuscript is prepared for submission to *Landscape Ecology* with a Journal Impact Factor of 5.2. Its interdisciplinary scope focusing on the intersection of geography and ecology fits the current study perfectly, one key topic being “Flows and redistributions of organisms, materials, and energy in landscape mosaics” (Springer Nature Switzerland AG, n.d.). The journal’s article type *short communication* with a maximum of 3’500 words (+/- 10%) appears to be a good fit for the present manuscript. The manuscript largely complies with the submission guidelines with the following open points, which will be dealt with after submission of the master thesis:

- Abstract: is required to be structured
- Citation style: switch to “springer-basic-author-date-no-et-al-with-issue.csl”, which is not suitable for the present thesis because websites are not referenced correctly
- Figures: customize name (“**Fig. 1**” instead of the standard bookdown “Figure 1:”)
- Data availability: clarify with data owner(s) and create data repository

For the composition of this document, the R Markdown format and bookdown R package was chosen, which supports multiple output formats, including HTML, PDF, and DOCX. This choice aligns with the principles of reproducible research, as content can be dynamically generated from underlying data. Two Github repositories were used:

- Public repository for the methodology (Python code):
<https://github.com/thunwal/bioinvasionanalysis>
- Private repository for the present thesis document (R Markdown code):
<https://github.com/thunwal/manuscript>

All outputs from the Python scripts were written to a data directory in the manuscript repository and used to render figures, tables and inline data.

1 Manuscript

Leveraging least-cost modelling to uncover multiple introductions: a case study of an invasive wild bee species

Christa Rohrbach^a ([E-Mail](#)) ([Phone](#)) ([ORCID](#)), Julia Lanner^b ([ORCID](#)), Gudrun Wallentin^a ([ORCID](#))

^a Paris-Lodron-University of Salzburg, Interfaculty Department of Geoinformatics – Z_GIS, Schillerstr. 30, 5020 Salzburg, Austria

^b University of Natural Resources and Life Sciences, Vienna; Department of Integrative Biology and Biodiversity Research; Institute of Zoology, Gregor Mendel Str. 33, A-1180 Vienna, Austria

Abstract Invasive species significantly impact biodiversity, necessitating accurate methods to estimate and predict their spread. Traditional expansion rate estimation often assumes a single introduction point, despite common multiple introductions. This study introduces a novel approach using sequential least-cost modelling to delineate populations before estimating expansion rates, without requiring prior knowledge of population structure. We focused on *Megachile sculpturalis*, the first non-native bee species in Europe, using presence data from 2008 to 2023 covering large parts of Europe. In annual time steps, new observations were connected to previously known observations via least-cost paths, delineating populations by excluding high-cost paths. This resulted in six populations with a sufficiently large sample size. The populations created align with genetic analyses published for the territory of France, Switzerland and Austria, which underlines the great potential of our spatially explicit approach and suggests at least one more independent introduction in Italy. Expansion rates for the six populations, calculated using distance regression, ranged from 9.9 km/year for a population spatially concentrated on Vienna, Austria, to 82.4 km/year for a population originating in Belgrade, Serbia, averaging 45.5 km/year. No advancement of invasion fronts was observed in 2022 and 2023, suggesting limits of suitable habitat or reduced data collection. By incorporating landscape heterogeneity, our approach uncovers multiple introductions, and supports targeted monitoring, management and genetic sampling efforts. Future research could make the step from retrospective to prediction by deriving accumulated costs per year and extrapolate future spread using cost accumulation.

Keywords multiple introductions, invasive species, dispersal, least-cost modelling, cost-distance, *Megachile sculpturalis*

Statements and Declarations

Funding (potential) The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests (potential) The authors have no relevant financial or non-financial interests to disclose.

Author Contribution (potential) Christa Rohrbach conceived and conducted the study and wrote the first draft of the manuscript. Julia Lanner acquired the data and advised on the study species. Gudrun Wallentin supervised the study conception and conduction. All authors contributed to the final manuscript and read and approved it.

Data Availability (potential) The datasets analyzed during the study are available from the corresponding author on reasonable request. The datasets generated during the study are freely available under [persistent link to data repository]. The Python code used to generate the datasets is freely available under <https://github.com/thunwal/bioinvasionanalysis/>.

1.1 Introduction

Invasive species are a major driver of biodiversity loss (Bellard et al., 2016). Accurately estimating and predicting their spread is crucial for assessing potential impacts and optimizing resource allocation for management efforts (Tobin et al., 2007). Methods for estimating expansion rates - such as distance regression, square root area regression, and boundary displacement - are typically applied to all presence data points in a study area, assuming one population with a single point of introduction. However, multiple introductions are common (Bossdorf et al., 2005; Ellstrand & Schierenbeck, 2006; Lawson Handley et al., 2011), as evidenced by genetic studies across various taxa including reptiles (Kolbe et al., 2004), gastropods (Facon et al., 2008), mammals (Zalewski et al., 2010), insects (Lanner et al., 2021; Miller et al., 2005; Ortego et al., 2021) and crustaceans (Robinson et al., 2018). Some researchers address the presence of distinct populations by manually dividing presence data along spatial barriers and calculating expansion rates for each population separately (Fraser et al., 2015; Tobin et al., 2007), but this requires prior knowledge of population structure, often obtained through genetic methods.

This paper presents a novel approach using sequential least-cost modelling to delineate populations before estimating expansion rates, without requiring prior knowledge of population structure. Least-cost modelling is a spatial analysis technique originally developed in the field of transport geography to solve corridor location problems (Etherington, 2016). It allows to incorporate landscape heterogeneity in terms of spatially varying “cost”, “resistance” or “friction” into a distance measure (Adriaensen et al., 2003) when modelling movement or spread on a continuous surface. In ecology, least-cost modelling is widely used to assess habitat connectivity (Etherington, 2016; Zeller et al., 2012), but is not applied in the emerging field of invasion biology to analyze spread dynamics, to our knowledge. Sequential least-cost modelling refers to an iterative application of least-cost modelling in time steps. In archaeology, sequential least-cost modelling has been applied to model ancient seafaring by allowing a virtual navigator to reevaluate and adapt to changing wind conditions in time steps (Perttola, 2022).

To estimate expansion rates, we applied the distance regression method. The cumulative Euclidean distance of the advancing invasion front to the first observation is regressed against time (Hui & Richardson, 2017, p. 22; Liebhold et al., 1992). This method proved least sensitive to small samples and sample point distribution in a performance comparison of three methods against a simulated expansion with known expansion rate (Gilbert & Liebhold, 2010).

We proceeded in three steps: (1) least-cost modelling: connecting each observation with the nearest earlier observation via a least-cost path, (2) population delineation: excluding paths

with the highest accumulated cost to isolate populations, and (3) calculating the expansion rates for all populations using the distance regression method. Our study organism is the Sculptured Resin Bee, *Megachile sculpturalis*, the first non-native bee species accidentally introduced to Europe. Initially observed in 2008 in Allauch, France (Vereecken & Barbier, 2009), this wild bee has since expanded its range dramatically, now ranging from the Pyrenees to the Black Sea. Genetic evidence indicates at least two independent introductions in the area of France, Switzerland and Austria (Lanner et al., 2021), making *M. sculpturalis* an ideal subject for this study.

1.2 Methods

Study species

The Sculptured Resin Bee, *Megachile sculpturalis* Smith, 1853 is a non-managed solitary bee species, commonly referred to as “wild bee”. Its native range is East Asia. Females are considerably larger than males and reach a body length of 22 to 27 mm (Parys et al., 2015). The species builds nests in existing cavities in deadwood and readily accepts artificial nesting aids (Figure 1). *M. sculpturalis* is univoltine in summer months, i.e. only one generation is produced per year. The development of the species in wood suggests that it may have reached Europe via timber trade (Quaranta et al., 2014).



Figure 1: A female Megachile sculpturalis is closing her nest with tree resin after having provisioned her brood with pollen and nectar. The nest is built into a drill hole of an artificial nesting aid

Study area

The study area encompasses the spatial extent of the presence data, covering large parts of Europe (Figure 2). It stretches from the westernmost observation in the Atlantic Pyrenees, Spain, to the easternmost on the Crimean Peninsula, Ukraine, and from the southernmost observation in Calabria, Italy, to the northernmost in Lower Franconia, Germany. To leave enough room for the least-cost modelling algorithm, at least 200 km were added to each axis.

Data sources and preparation

Presence data for *M. sculpturalis* in Europe (Figure 2) was provided by BeeRadar.info (Lanner et al., 2020) and contains 1849 records for the period from 2008 to 2023. Most records ($n = 675$) were retrieved from public naturalist platforms and Social Media (beewatching.it, cardobs.mnhn.fr, entomologiitaliani.net, facebook.com, faune-france.org, flickr.com, inaturalist.org, inpn.mnhn.fr, insecte.org, instagram.com, izeltlabuak.hu, naturamediterraneo.com, naturbeobachtung.at, naturgucker.de, observation.org, silene.eu, spipoll.org), followed by scientific literature ($n = 612$) (Aguado et al., 2018; Amiet, 2012; Bila Dubaić, Plećaš, et al., 2022; Bila Dubaić, Lanner, et al., 2022; Fortel et al., 2016; Gühr & Westrich, 2013; Gogala et al., 2018; Gradinarov et al., 2023; Grossi et al., 2018; Guariento et al., 2019; Ivanov & Fateryga, 2019; Kovács, 2015; Lanner et al., 2020; Le Féon et al., 2018; Le Féon & Geslin, 2018; Ortiz-Sánchez et al., 2018; Poggi et al., 2020; Polidori & Sánchez-Fernández, 2020; Quaranta et al., 2014; Rickenbach & Sprecher, 2018; Ruzzier et al., 2020; Vereecken & Barbier, 2009; Westrich et al., 2015; Westrich, 2020; Zandigiacoimo & Grion, 2017). The remaining records were obtained from citizen scientist projects ($n = 534$) (BeeRadar.info, srbee.bio.bg.ac.rs/english) and from personal communication with scientific professionals ($n = 28$) (Giovanni, Celia; Guariento, Elia; Diaz Calafat, Joan; Bartolotti, Laura; Wanzenböck, Silvia; Dötterl, Stefan; Kovács, Tibor; Westrich, Paul). The data was projected to Mollweide (ESRI:54009) to match the projection of the cost surface. For least-cost modelling, the presence data was reduced to the cost surface resolution, retaining one presence data point per cell with the earliest observation year. This is required to avoid duplication of paths and zero-length paths, as the least-cost algorithm maps input points to cell centers and finds paths from cell center to cell center.

The cost surface used for least-cost modelling was derived from the environmental suitability consensus map published by Lanner et al. (2021), which shows the mean prediction of various species distribution models. The map was provided as a GeoTIFF file in the equal-area Mollweide projection (ESRI:54009) and a spatial resolution of 10 km, with the ocean surface encoded as NoData. After clipping the file to our study area, the value range of [33.04,938.572] was inverted, under the assumption that suitable habitat facilitates spread, and scaled to [1,100] (Figure 2). These steps were executed in QGIS 3.34.5-Prizren.

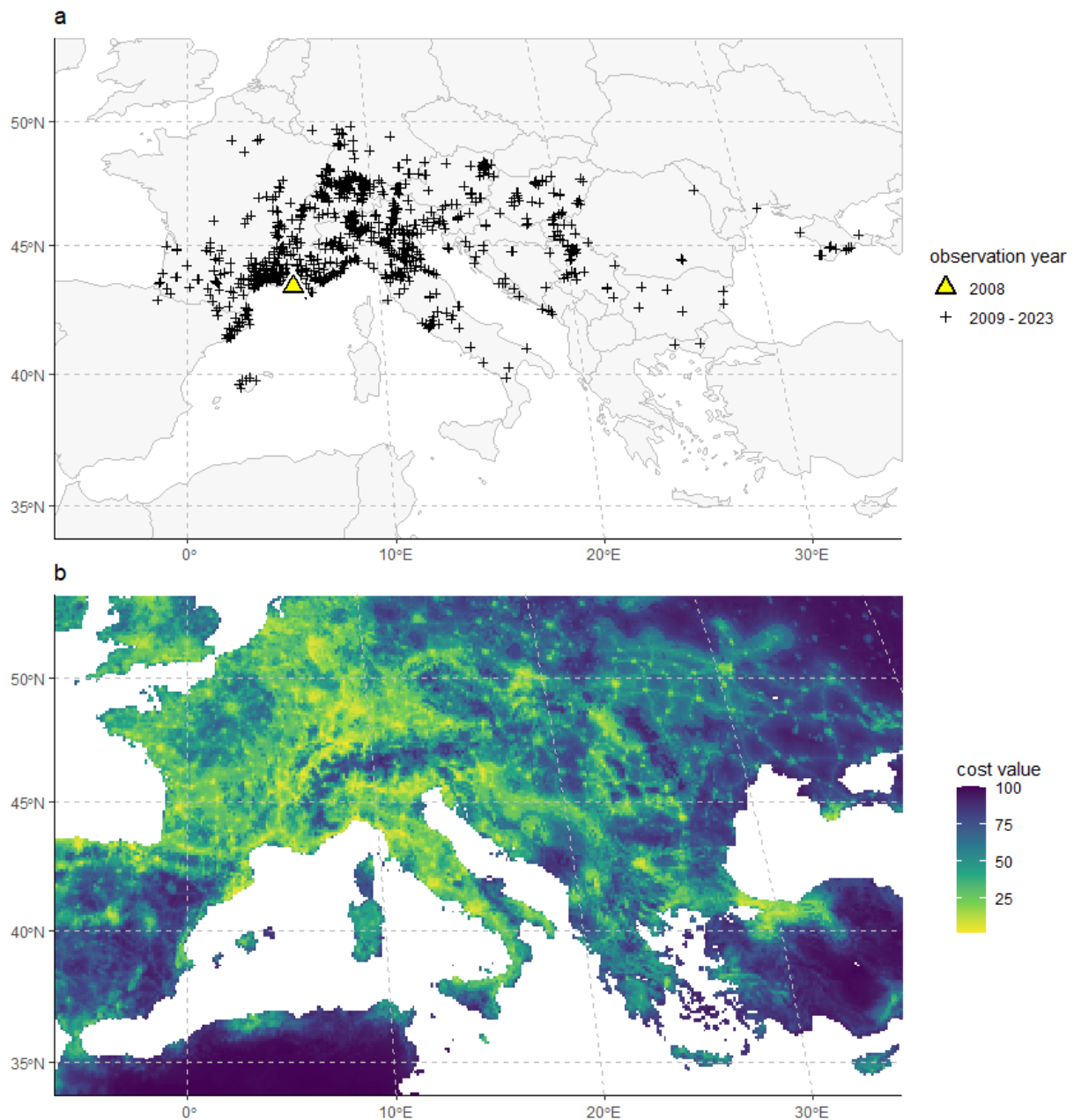


Figure 2: (a) Spatial distribution of the presence data for *M. sculpturalis* in Europe. The data covers the period from the first observation in 2008 until 2023. (b) The cost surface used for least-cost modelling, derived from an environmental suitability consensus map (Lanner et al., 2021). The original raster file was clipped to our study area, its value range inverted, and scaled to [1,100]

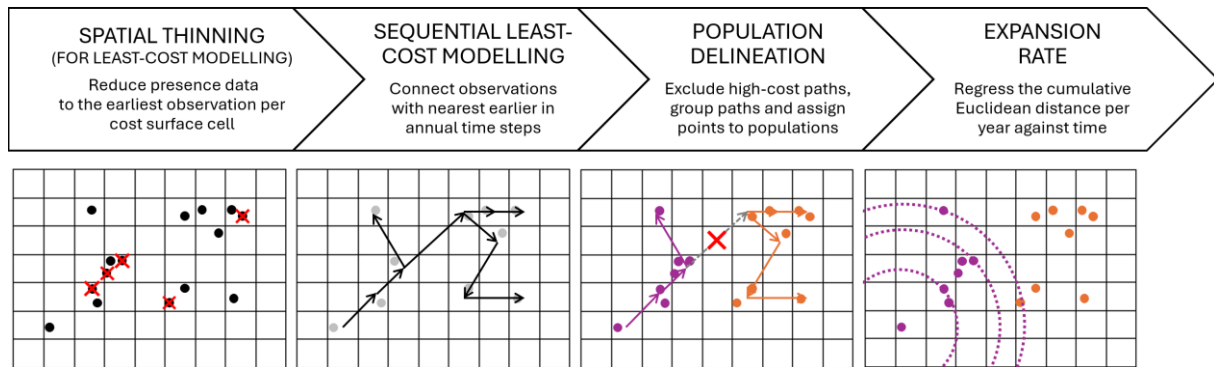


Figure 3: We proceeded in the following steps: (1a) spatial thinning of presence data in preparation for least-cost modelling, (1b) least-cost modelling, (2) population delineation, and (3) calculation of expansion rates

Least-cost modelling

One of the first computer programs to support the least-cost modelling process was named *Generalized Computer-Aided Route Selection (GCARS)* and published by A. K. Turner & Miles (1971). The methodology implemented in GCARS is still predominant today, due to its intuitive appeal and broad adoption in software solutions (Driezen et al., 2007). Three concepts are combined: (1) Cost surface: A landscape is modeled in terms of the “cost”, “resistance” or “friction” of moving through different areas; (2) Landscape graph: The cost surface is converted to a lattice graph based on a cell neighborhood definition. The most common neighborhood definition allows for an 8-way, i.e. orthogonal and diagonal movement from cell to cell; (3) Shortest-path algorithm: A path-finding algorithm is applied to the landscape graph to accumulate costs along possible paths (Adriaensen et al., 2003), and determine the path of least cost between two specified nodes. The most applied algorithm is Dijkstra’s (Dijkstra, 1959).

For the present study, we calculated least-cost paths using the *Graph* module of the *scikit-image* package for Python (Walt van der et al., 2014), which supports landscape graph-based least-cost modelling and works with an 8-way neighborhood definition. The *MCP_Geometric* class was used to accumulate distance-weighted costs, i.e. to account for the fact that orthogonal and diagonal moves are of different length. The class is instantiated with an n-dimensional array, so we converted the cost surface to a two-dimensional array at runtime. In this process, we overwrote NoData values with infinite values, which are ignored by the algorithm. This effectively turns the ocean surface into an impassable barrier.

In our sequential approach to least-cost modelling, we accumulate costs for each annual time step and trace back the least-cost path from each new presence data point in a given year to the nearest presence data point of any previous year (Figure 3). In this process, least-cost

paths are created sequentially until least-cost paths for all presence data points have been traced back, except of isolated presence data points surrounded by a barrier. The result can be thought of as one large dispersal network, which, however, contains the unfiltered entirety of least-cost paths, including paths with extremely high accumulated cost that are assumed to bridge populations.

Delineation of populations

To isolate groups of paths - and ultimately populations -, we applied an accumulated cost threshold and removed least-cost paths with higher accumulated cost from the result set. We then assigned unique group identifiers to the remaining paths based on their connectivity, which was tested by comparing path endpoint coordinates. To identify a suitable threshold, we ran a sensitivity analysis for the effect of the accumulated cost threshold on the number of groups formed (Figure 4). We chose to apply the 95th quantile as a threshold so that the resulting populations match the known genetic groups in France, Switzerland and Austria (Lanner et al., 2021) as closely as possible and that the threshold value falls within a value range of low sensitivity. Lastly, the presence data points were assigned to the population of the nearest path. The nearest path was identified by searching for a path endpoint within a search radius of half the diagonal of a cell, calculated as:

$$\frac{a\sqrt{2}}{2}$$

with a representing the cell side length. The search radius ensures that points isolated either by a barrier or by the removal of the costliest paths are not assigned to a population.

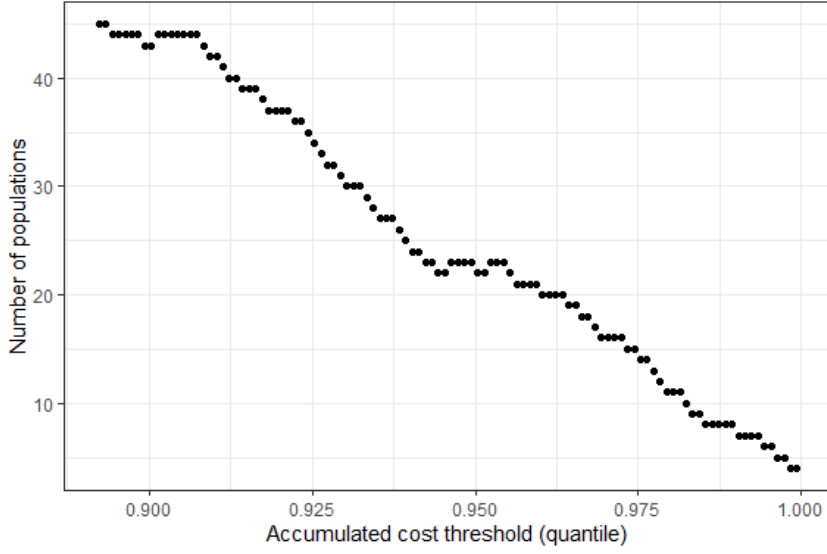


Figure 4: A sensitivity test revealed the inverse relationship of the accumulated cost threshold and the number of delineated populations. The higher the threshold, the fewer least-cost paths are removed from the result set and the fewer populations isolated

Expansion rate

To determine the expansion rate for each population, we adopted the common approach of regressing the cumulative distance of the advancing front to the first observation as a function of time (Hui & Richardson, 2017, p. 22; Liebhold et al., 1992). For each population, we identified the first observation and regressed the cumulative Euclidean distance to other observations against time. The calculation was carried out in the following steps:

1. Distance calculation: For the given population, denote the first observation as p_0 . For each point p_i in the population, calculate the Euclidean distance to p_0 as:

$$d(p_i, p_0) = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

where (x_i, y_i) and (x_0, y_0) are the coordinates of points p_i and p_0 , respectively.

2. Cumulative maximum distance: For the given population, calculate the cumulative maximum distance for each year y as:

$$D_y = \max(\{d(p_i, p_0) | y_i \leq y\})$$

3. Linear regression: For the given population, perform a linear regression of D_y against year y . The regression model is:

$$D_y = \beta_0 + \beta_1 y + \epsilon$$

The estimated slope β_1 of this regression line represents the expansion rate. It is divided by 1000 to convert from the native unit m/a to km/a .

Sample size index

The data basis of the delimited populations should be comparable to be able to assess the reliability of the calculated expansion rates, which we assume to increase with the number and temporal span of observations. To quantify the sample size in both respects, we normalize both measures by scaling them to their respective maximum values and calculate a combined measure. For each population i , the *Sample size index* s_i was computed as follows:

$$s_i = \frac{o_i}{o_{max}} \times \frac{y_i}{y_{max}}$$

where o_i is the number of observations for population i and y_i is the year span for population i . o_{max} and y_{max} are the maximum values for the number of presence data point and year span, respectively, across all populations.

Code repository

We followed a reproducible approach to facilitate further research. All methodological steps - except the cost surface creation - were implemented as standalone Python scripts which are freely available under <https://github.com/thunwal/bioinvasionanalysis>.

1.3 Results

We delineated 22 populations (Table 1 and Figure 5), with 25 presence data points remaining unassigned due to isolation by ocean barriers or the removal of the high-cost paths (Figure 5). Six populations reached a sample size index ($\times 100$) of at least 1 and were deemed supported by sufficient data and suitable for further analysis. Populations with a lower sample size index were excluded from further analysis.

Expansion rates range from 9.9 km/year to 82.4 km/year, with an average rate of 45.5 km/year. The highest expansion rate of 82.4 km/year is observed in population no. 7 originating in Belgrade, Serbia, while the lowest is 9.9 km/year in population no. 5 originating in Vienna, Austria. This population also shows the smallest spatial extent with a concentration of presence data points in Vienna. The two populations with the highest sample size indices, population no. 1 ($n = 843$, France) and no. 2 ($n = 466$, Italy), exhibit a logistic increase in cumulative distance from the first point of observation, with a lag phase until 2011 (Figure 5). Populations with a lower sample size index do not show a clear pattern. There was no increase in cumulative distance in any population during 2022 and 2023.

Table 1: Resulting populations along with expansion rates and R2 value of the linear regression model. A sample size index was calculated to quantify both the number of observations and time span covered (100 = best, 0 = least)

| ID | First observation country | First observation year | Latest observation year | Year span | n | Sample size index x 100 | R2 | Rate [km/year] |
|----|---------------------------|------------------------|-------------------------|-----------|-----|-------------------------|------|----------------|
| 1 | France | 2008 | 2023 | 15 | 843 | 100.00 | 0.88 | 61.6 |
| 2 | Italy | 2009 | 2023 | 14 | 466 | 51.59 | 0.88 | 79.4 |
| 5 | Austria | 2017 | 2023 | 6 | 170 | 8.07 | 0.55 | 9.9 |
| 7 | Serbia | 2017 | 2021 | 4 | 179 | 5.66 | 0.85 | 82.4 |
| 3 | France | 2015 | 2023 | 8 | 25 | 1.58 | 0.71 | 20.9 |
| 4 | Hungary | 2015 | 2023 | 8 | 23 | 1.46 | 0.71 | 18.8 |
| 10 | Hungary | 2018 | 2023 | 5 | 18 | 0.71 | 0.64 | 12.8 |
| 0 | Italy | 2009 | 2017 | 8 | 10 | 0.63 | 0.38 | 2.9 |
| 17 | Ukraine | 2018 | 2023 | 5 | 16 | 0.63 | 0.68 | 19.7 |
| 8 | Hungary | 2019 | 2023 | 4 | 18 | 0.57 | 0.36 | 27.4 |
| 9 | Serbia | 2018 | 2022 | 4 | 12 | 0.38 | 0.68 | 9.2 |
| 6 | Croatia | 2018 | 2023 | 5 | 9 | 0.36 | 0.87 | 31.8 |
| 11 | Hungary | 2019 | 2021 | 2 | 8 | 0.13 | 0.93 | 9.7 |
| 14 | France | 2016 | 2021 | 5 | 2 | 0.08 | 1.00 | 15.6 |
| 15 | Hungary | 2018 | 2021 | 3 | 3 | 0.07 | 1.00 | 10.0 |
| 12 | Hungary | 2018 | 2020 | 2 | 4 | 0.06 | 0.36 | 4.4 |
| 16 | Serbia | 2020 | 2021 | 1 | 7 | 0.06 | 0.87 | 23.6 |
| 21 | Romania | 2019 | 2023 | 4 | 2 | 0.06 | 1.00 | 1.5 |
| 19 | Spain | 2020 | 2023 | 3 | 2 | 0.05 | 1.00 | 14.0 |
| 13 | Spain | 2020 | 2021 | 1 | 2 | 0.02 | 1.00 | 15.2 |
| 18 | Bulgaria | 2021 | 2022 | 1 | 3 | 0.02 | 1.00 | 112.2 |
| 20 | Bulgaria | 2022 | 2023 | 1 | 2 | 0.02 | 1.00 | 87.3 |

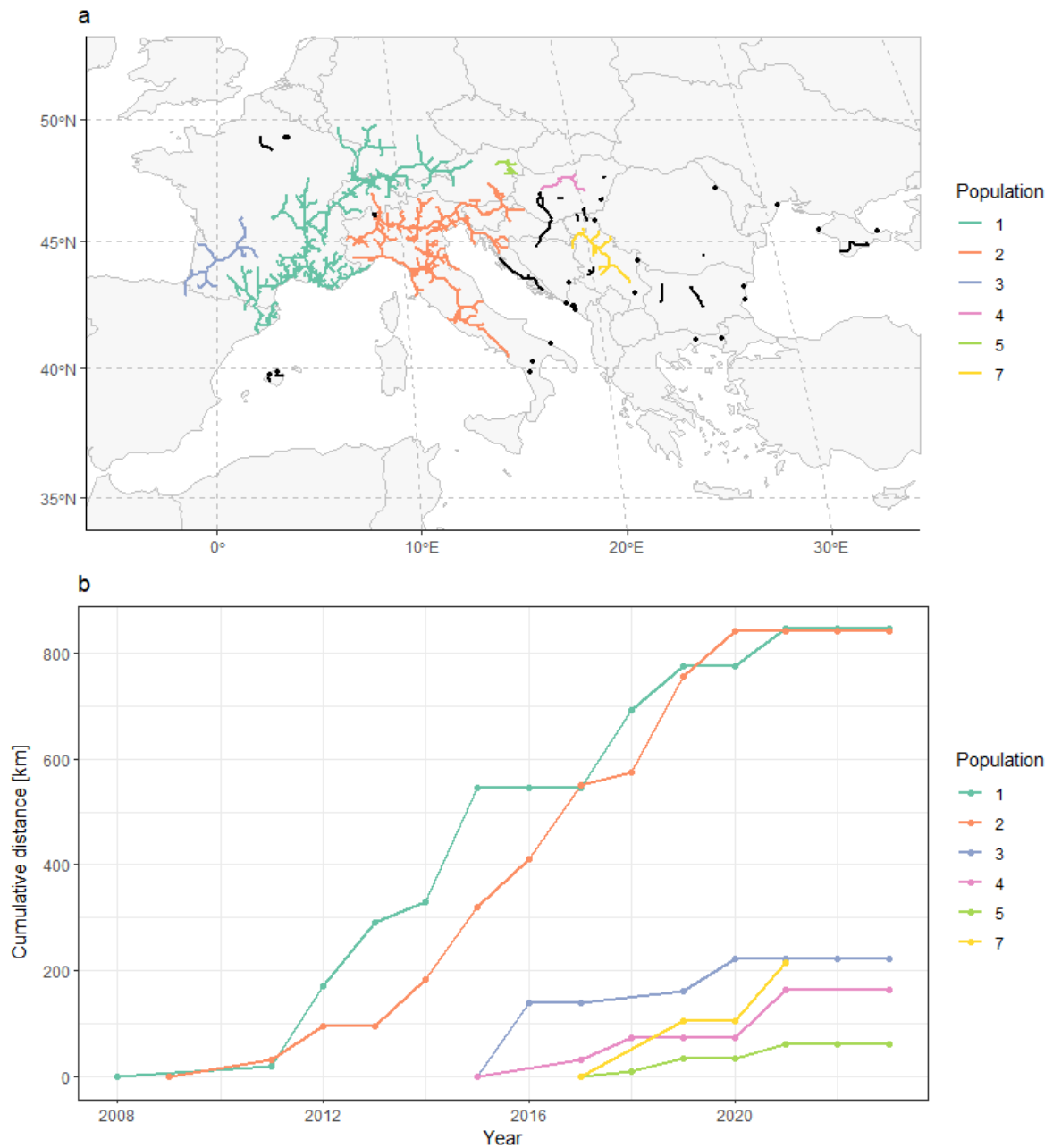


Figure 5: (a) Population assignment of the least-cost paths which connect the presence data points. Populations supported by sufficient data (sample size index $\times 100 \geq 1$) are considered suitable for further analysis and symbolized in colour. Remaining populations and unassigned isolated presence data points are symbolized in black. (b) Diagram showing the increase in cumulative distance of observations to the first observation over time for populations supported by sufficient data

1.4 Discussion

We delineated 22 populations by applying least-cost modelling to European presence data of the multiple-introduced solitary bee species *M. sculpturalis* (n = 1849) and calculated expansion rates for all populations. Expansion rates of populations supported by sufficient data (n = 6) ranged from 9.9 km/year to 82.4 km/year, averaging at 45.5 km/year. These rates were derived by regressing the cumulative Euclidean distance of observations to the first observation against time.

Population delineation

The populations as identified by our model align with genetic analysis results by Lanner et al. (2021). The authors analyzed samples from the area of France, Switzerland and Austria and found strong genetic evidence for at least two genetic groups, one consisting of samples from Vienna, Austria, and another consisting of samples from France and Switzerland. In the present study, we were able to reconstruct the same pattern, with one population being spatially concentrated on the greater area of Vienna, Austria (ID 5), and the other originating in France and spreading eastward along the north side of the Alps (ID 1). With regard to the solid data basis in Italy, we hypothesize that the large population in Italy, as we constructed it (ID 2), is also the result of a separate introduction event. The same is possible for the populations in southwestern France (ID 3), Hungary (ID 4) and Serbia (ID 7), but these populations are supported by less data than the before-mentioned populations.

Our results also clearly show that a low density of presence data reduces the significance of the results. We were only able to evaluate few populations in the periphery of the invaded area and in Southeast Europe. While a lower presence data density is to be expected on the periphery, we anticipate large observation gaps in Southeast Europe. This impedes a meaningful delineation of populations, with the result that an unrealistically large number of truncated populations was created. But on the other hand, this is precisely what makes the assumed observation gaps visible, which is also useful information.

Expansion rate

The expansion rate of *M. sculpturalis* outside its native range has not yet been quantified. We therefore compare our results with findings for *Vespa velutina*, another large winged insect species currently invading Europe. For this species, an expansion rate of 18.3 ± 3.3 km/year was found in northwestern Italy (Bertolino et al., 2016), 37.4 ± 13.2 km/year in Portugal (Carvalho et al., 2020) and 100 km/year in France (Rome et al., 2013). The discrepancies between the estimates can largely be attributed to methodological differences. In the two first-

mentioned studies, the authors aimed to estimate the expansion rate through active dispersal, excluding human-mediated jump dispersal. In the latter study, the authors included all observations, also such resulting from assumed jump dispersal. Bertolino et al. (2016) also mention unfavorable conditions which might have slowed spread. We did not attempt to differentiate between active and human-mediated dispersal; but we might well have delineated some subpopulations resulting from jump dispersal, as - without genetic analyses - we cannot distinguish jump dispersal events and introduction events. Three of the six expansion rates we calculated for *M. sculpturalis* are remarkably high (ID 1, France, 61.6 km/year; ID 2, Italy, 79.4 km/year; ID 7, Serbia, 82.4 km/year), indicating that jump dispersal events might be included. Two rates are comparable to the expansion rate identified for *V. velutina* in northwestern Italy (ID 3, France, 20.9 km/year; ID 4, Hungary, 18.8 km/year), and one rate is remarkably low (ID 5, Austria / Vienna, 9.9 km/year). Both the lowest (ID 5, Austria / Vienna) and highest rate we identified (Serbia, 82.4 km/year) are situated in the eastern periphery of the invaded area, where observation gaps might cause more random effects.

Temporal dynamics

In the last two years of the study period (2022 and 2023), the invasion front was not advancing in any of the six analyzed populations. This stagnation is particularly surprising in the northwestern periphery of the invaded area (France, Germany), where the environmental suitability is considered high (Lanner et al., 2022). In the southwestern periphery (Spain), a reduced environmental suitability might slow down spread. As Citizen Science data contributes substantially to our data basis, we should also note that the populations grew beyond the boundaries of the focus area of the BeeRadar citizen science project (DACH region) what might reduce the detection rate. In Southeast Europe, new observations continue being made each year far off known observations, so there might be a dynamic population growth, which we cannot evaluate due to constrained data availability.

Conclusion

By applying a spatially explicit method, our approach enables multiple introductions to be detected and possible populations to be delineated without prior knowledge of the population structure. The successful validation against genetic data in part of the study area underlines the potential to serve as a basis for targeted monitoring and genetic sampling efforts. Future genetic studies on *M. sculpturalis* should be geographically extended to allow a more robust validation of our approach. Furthermore, a follow-up study could make the step from retrospective to predictive, regressing accumulated costs against time instead of Euclidean distance and extrapolating into the future. However, the appropriate handling of jump dispersal remains a challenge even with least-cost modelling.

2 Report

2.1 Introduction

Apart from the one-time preparation of the cost interface, I implemented all methodological steps as Python scripts so that all steps are easily reproducible, and the methodology can be versioned with git. With the goal of publication in mind, I also attached great importance to:

- Separation of configuration and program logic: Parameters are stored in a dedicated module *params.py*. This makes it easier to use the scripts and keep the code clean.
- Modularity: Thematically related functions are grouped into different scripts and these are loaded as modules from a main script. This simplifies the overview.
- Organization of results: Each script run can be named (parameter `run`) and produces correspondingly named outputs to facilitate the organization of results.

All Python code used for this thesis is published in the Github repository <https://github.com/thunwal/bioinvasionanalysis>. Figure 6 provides an overview of the script modules, functions, inputs and outputs. The log messages below, from a script run including all modules, illustrate the sequence of individual methodological steps, with parameters replaced by a placeholder (`<parameter>`). The subsequent chapters follow the order of these methodological steps and explain methods and code in depth.

```
[16:58:28] Loading cost raster from '<workdir_path>\<cost_name>'...
[16:58:28] Cost raster has CRS ESRI:54009 and cell size 10000.0 x 10000.0.
[16:58:28] Loading presence data from '<workdir_path>\<presence_name>'...
[16:58:28] Presence data has CRS EPSG:4326 and 1849 rows, of which 1849 rows with non-null year and geometry.
[16:58:28] Projected presence data to ESRI:54009.
[16:58:29] Imported presence data saved to '<workdir_path>\<run>.gpkg', layer '<run>_points'.
[16:58:29] Creating temporary fishnet with raster properties...
[16:58:31] Joining observations and fishnet polygons...
[16:58:31] Selecting earliest observation per fishnet polygon...
[16:58:31] Thinned imported presence data saved to '<workdir_path>\<run>.gpkg', layer '<run>_points_thinned'.
[16:58:31] Calculating least-cost paths for year 2009...
[16:58:32] Calculating least-cost paths for year 2010...
[16:58:32] Calculating least-cost paths for year 2011...
[16:58:32] Calculating least-cost paths for year 2012...
[16:58:32] Calculating least-cost paths for year 2013...
[16:58:32] Calculating least-cost paths for year 2014...
[16:58:32] Calculating least-cost paths for year 2015...
[16:58:32] Calculating least-cost paths for year 2016...
[16:58:32] Calculating least-cost paths for year 2017...
```

```
[16:58:32] Calculating least-cost paths for year 2018...
[16:58:32] Calculating least-cost paths for year 2019...
[16:58:32] Calculating least-cost paths for year 2020...
[16:58:32] INFO: No path found for a point.
[16:58:32] INFO: No path found for a point.
[16:58:32] INFO: No path found for a point.
[16:58:32] INFO: No path found for a point.
[16:58:32] Calculating least-cost paths for year 2021...
[16:58:32] Calculating least-cost paths for year 2022...
[16:58:32] Calculating least-cost paths for year 2023...
[16:58:32] Least-cost paths saved to '<workdir_path>\<run>.gpkg', layer '<run>_paths'.
[16:58:33] Upper outlier fence (Q3 + 1.5 x IQR) for accumulated cost is 269.424 (Q0.892).
[16:58:33] Testing accumulated cost thresholds from Q0.892 to Q0.999...
[16:59:16] Test results saved to '<workdir_path>\<run>_sensitivity_test.csv'.
[16:59:16] Least-cost paths with accumulated cost < 473.695 (Q<threshold>) loaded.
[16:59:16] Grouping least-cost paths by their connectivity...
[16:59:16] Grouped least-cost paths saved to '<workdir_path>\<run>.gpkg', layer '<run>_paths_grouped'.
[16:59:17] Assigning observations to the groups formed by least-cost paths...
[16:59:17] Grouped observations saved to '<workdir_path>\<run>.gpkg', layer '<run>_points_grouped'.
[16:59:17] Calculating expansion rates for groups (populations)...
[16:59:17] Raw data saved to '<workdir_path>\<run>_cumulative_distances.csv'.
[16:59:17] Expansion rates saved to '<workdir_path>\<run>_expansion_rates.csv'.
```

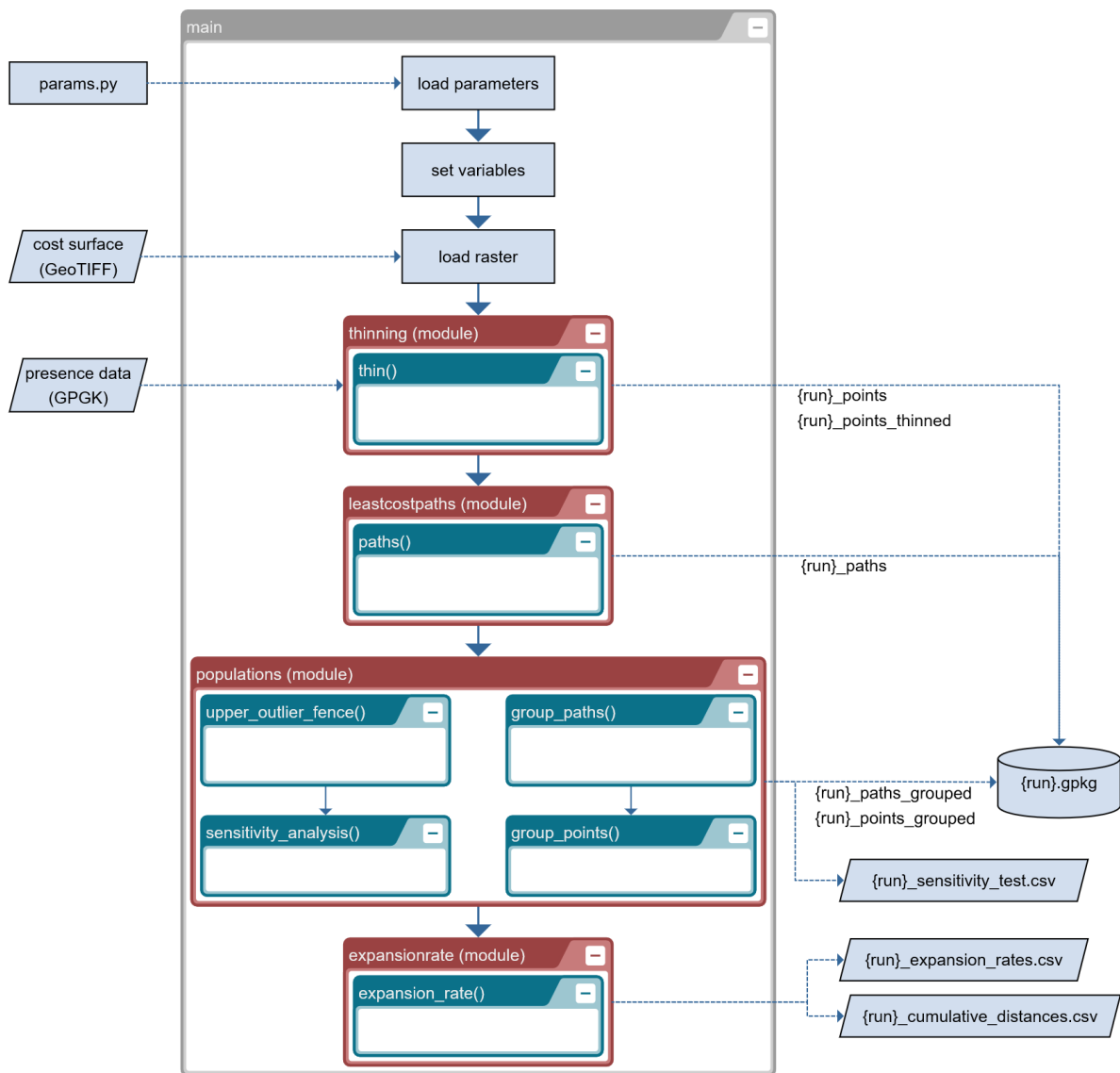


Figure 6: The Python code used for the methodology of this thesis is organized into modules which are loaded by the main script. The script reads configuration parameters from a dedicated file (*params.py*)

2.2 Exploring the data as a time series map

A common saying goes: “A picture is worth a thousand words”. A map is worth at least as much. But only a time series map optimally supports the recognition of dynamic processes. For this reason, I familiarized myself with the data in this way and used the same visualization to evaluate results. The Figures 7, 8, and 9 show a full time series (2008 - 2023) for presence data and final least-cost paths.

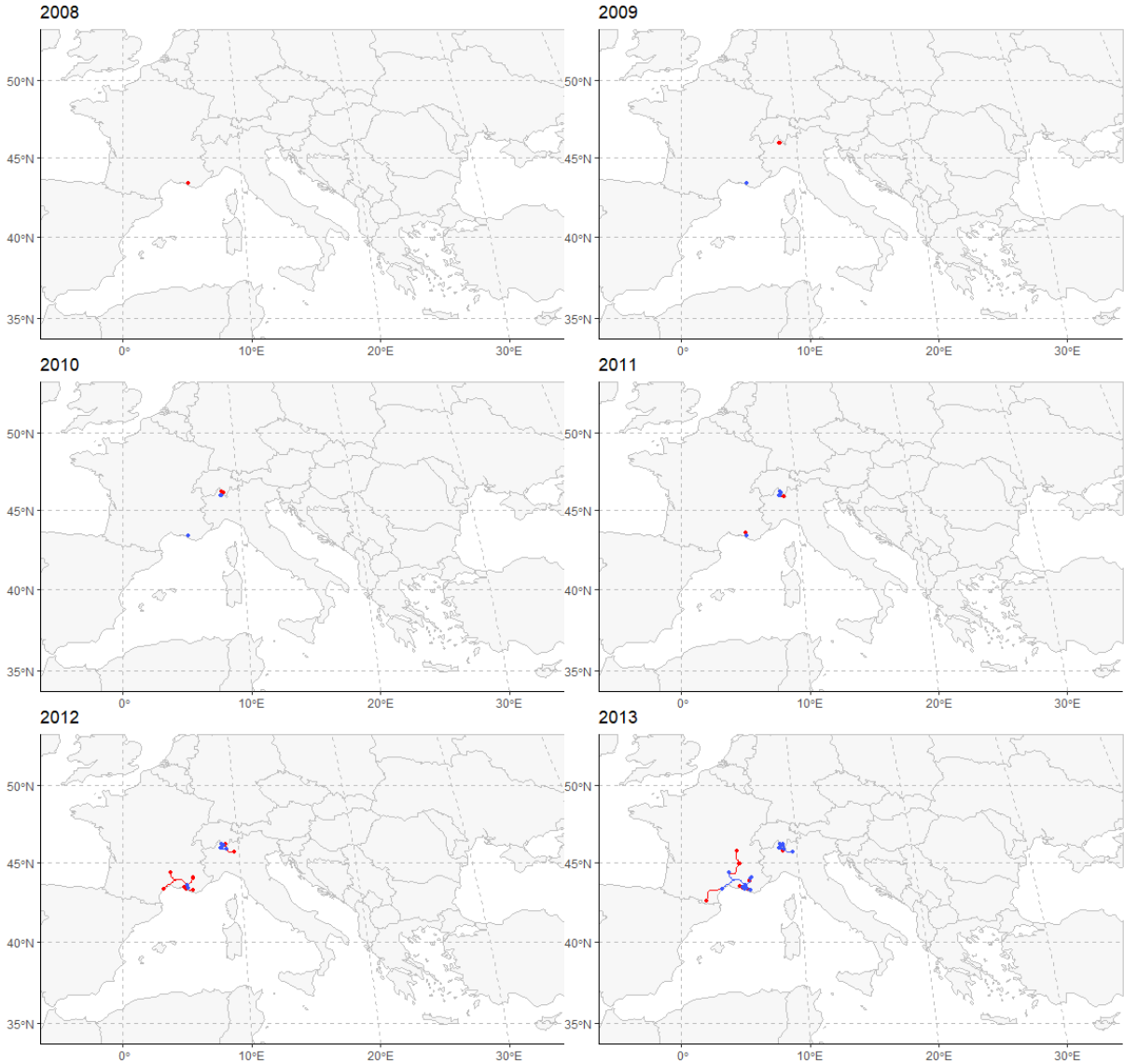


Figure 7: Time series from 2008 to 2013 with remaining least-cost paths after application of the accumulated cost threshold and isolated presence data points. Blue points and lines represent known observations and least-cost paths, red points and lines new ones

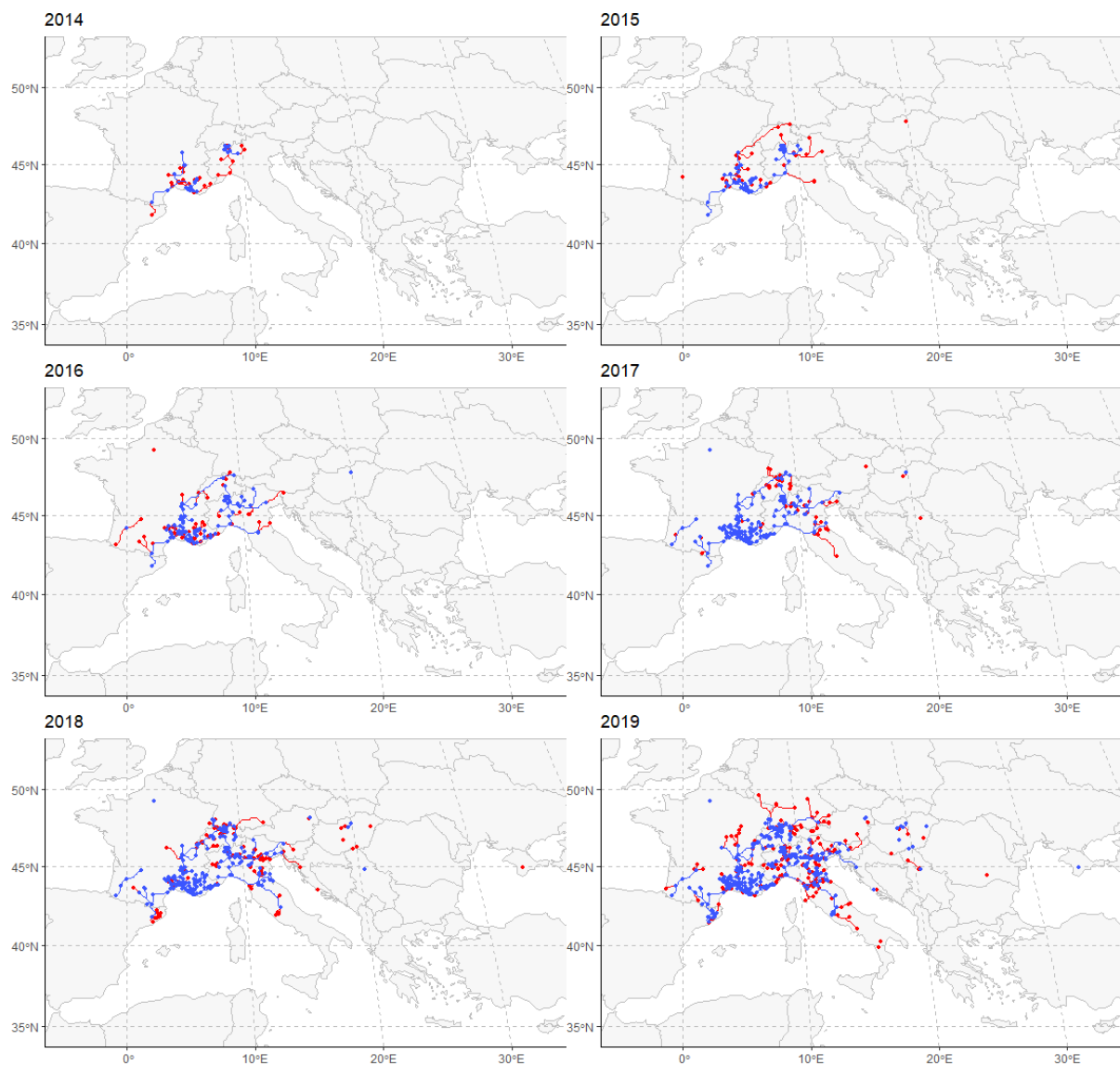


Figure 8: Time series from 2014 to 2019 with remaining least-cost paths after application of the accumulated cost threshold and isolated presence data points. Blue points and lines represent known observations and least-cost paths, red points and lines new ones

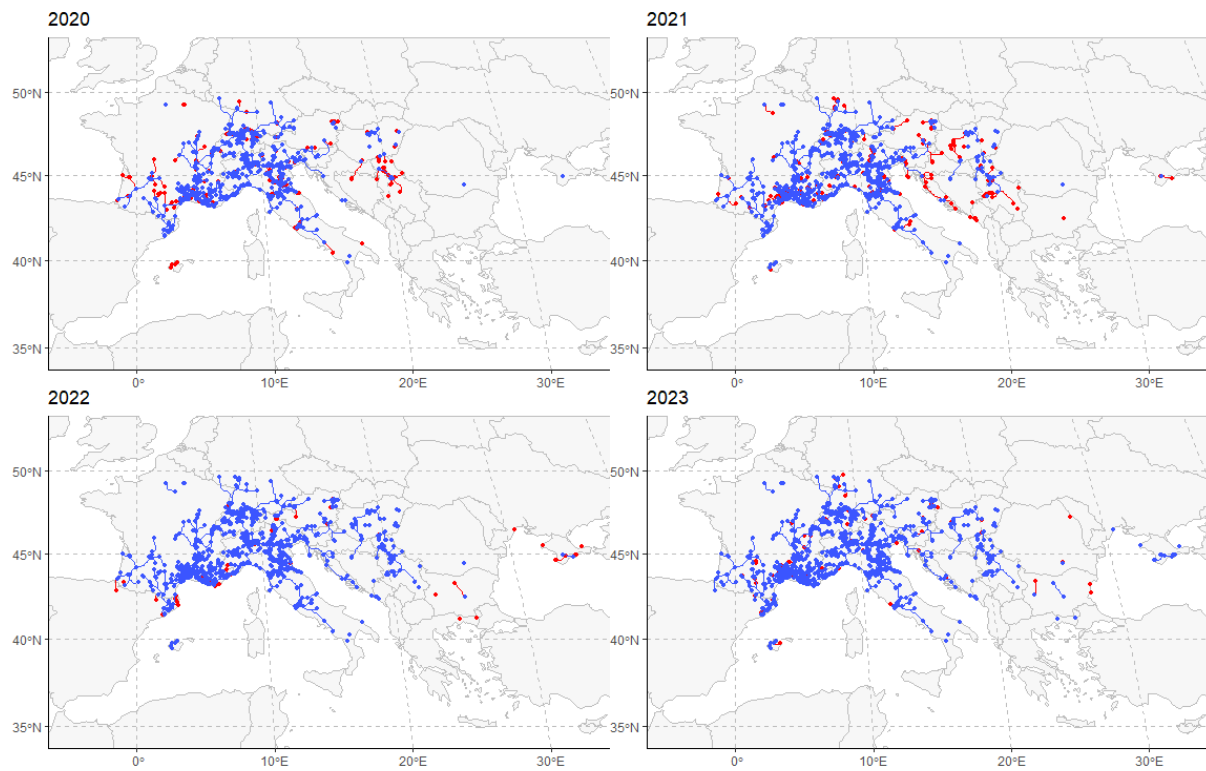


Figure 9: Time series from 2020 to 2023 with remaining least-cost paths after application of the accumulated cost threshold and isolated presence data points. Blue points and lines represent known observations and least-cost paths, red points and lines new ones

2.3 Preparing the presence data (Script module *thinning*)

In first experiments, I found that the presence data needs to be reduced to a maximum of one point per cost surface cell to avoid generating zero-length and duplicated least-cost paths. This spatial thinning process needs to be time-aware, retaining the earliest observation per cell, so that the spread over time can be reconstructed. I solved this by creating a fishnet based on cost surface properties, joining the presence data points with the fishnet polygons and selecting the earliest observation per polygon. For enhanced robustness - as the code is to be shared -, I handle the following two cases: (a) For a spatial join, the involved datasets need to have the same spatial reference system (SRS). Instead of bothering the user with an error message, I project the presence data to the cost surface SRS before the join operation. This is the better option than vice versa, as resampling a raster file would result in a loss of information. (b) As the fishnet is created from square polygons, I implemented a check if the raster file has square cells ($x = y$). If not, an exception is raised and the code execution aborted, as the thinning process would not produce correct results. Figure 10 provides an overview of the program logic applied to prepare the presence data for least-cost modelling.

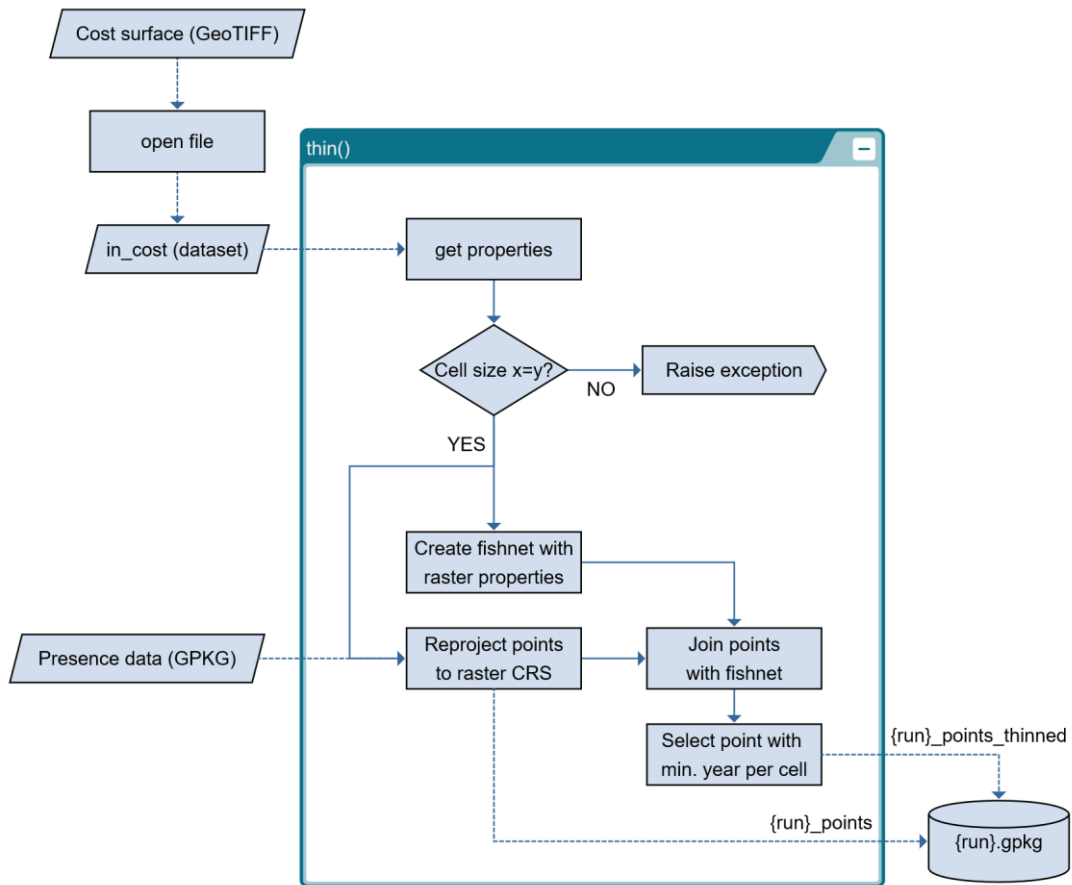


Figure 10: Program logic of the script module 'thinning', which prepares the presence data for subsequent steps by projecting the data to the cost surface SRS and applying spatial thinning for least-cost modelling

2.4 Least-cost modelling

2.4.1 Limitations and solutions

Deterministic path-finding

The application of traditional least-cost modelling in ecology bases on the assumption that the individual is familiar with the entire study area and chooses the optimal path (Etherington, 2016), which in many cases might not represent a realistic assumption for the behavior of the species on the spatial scale under investigation. For my thesis, however, this limitation is not relevant, as my goal is to understand broader patterns of spread, not to identify actual dispersal pathways. Moreover, Etherington (2016) hypothesizes that least-cost modelling might be well suited for studies where an optimistic view of connectivity is required, such as for invasive species management. In addition, as a side effect of applying least-cost modelling sequentially in annual time steps, I limit the all-knowingness of the individual to one time step. So even if not relevant here, the problem is reduced. Before concluding the problem is not relevant, I anyhow researched known solutions, which I present here.

Various authors extended the shortest-path algorithm commonly used in least-cost modelling (Dijkstra, 1959) with stochastic elements to overcome its deterministic nature. Pinto & Keitt (2009) developed a tool named *Multiple Shortest Paths (MSP)* which randomly deletes edges in the graph in proportion to edge weight. Saerens et al. (2009) presented procedures to generate *Randomized Shortest-Paths (RSP)*, as the authors named it. This algorithm is implemented in the R package *gdistance* (Etten, 2017) and allows the user to define the degree of randomness with a parameter θ . Lewis (2021) integrated randomness not into the shortest-path algorithm, but into the cost surface by the use of Monte Carlo simulation to propagate uncertainty (different realizations of vertical RMSE) in the input data (DEM) on the resulting least-cost paths. All these procedures (Lewis, 2021; Pinto & Keitt, 2009; Saerens et al., 2009) produce a set of competitive least-cost paths when run repeatedly.

Other proposed methods to generate alternative paths do not make use of randomization methods, to mention *k-shortest path algorithm* (Eppstein, 1998; Guerriero et al., 2001; Katoh et al., 1982), *near shortest path algorithm* (Carlyle & Wood, 2005), *Iterative Penalty Method (IPM)* (Ayad, 1967; A. K. F. Turner, 1968), *Gateway Shortest Path (GSP) model* (Lombard & Church, 1993) and its extension *Multi-Gateway Shortest Path (MGSP) model* (Scaparra et al., 2014). The latter two models maximize the spatial difference between competitive alternative least-cost paths, an important requirement for infrastructure projects, where alternatives must be offered in the approval process.

While the proposed solutions mentioned above are all modifications of the traditional least-cost modelling approach, other approaches build upon a completely different foundation. Approaches based on Circuit Theory (McRae et al., 2008), when applied in an ecological context, assume a random walker, whilst approaches based on Network Flow Theory (Ahuja et al., 1993; Phillips et al., 2008) inherently account for multiple paths. Agent-based models, i.e. models based on the simulation of individual behavior, can include stochastic elements and limit the perception and knowledge of an individual (Tang & Bennett, 2010).

Directional and length distortion

Goodchild (1977) found that the application of path-finding algorithms to a lattice graph created from raster data is affected by directional and length distortion. The author extended the common 8-way neighborhood definition with the Knight's move to allow for a 16-way movement from cell to cell. This reduces the distortion problem but does not solve it. An obvious manifestation of the problem is zigzag patterns in areas of uniform cost, which is why the problem was also named *zigzag problem* (Tomlin, 2010).

Alternative methods overcome the distortion issues by not converting the landscape to a lattice graph, but constructing a continuous surface from the accumulated cost and tracking the least-cost path along the slope line (Douglas, 1994; Esri Inc., n.d.-a) or modelling the accumulation of cost as the propagation of waves (Tomlin, 2010).

2.4.2 The complexity of cost surface creation

The development of an ecologically meaningful cost surface is the most challenging aspect of least-cost modelling in ecology (Etherington, 2016). Effects resulting from the choice of spatial and thematic scale have to be expected (Etherington, 2016; Zeller et al., 2012), the handling of linear elements during rasterization must be defined (Adriaensen et al., 2003), and the researcher must be aware of the difference between habitat suitability and permeability and implications for the study species (Adriaensen et al., 2003; Zeller et al., 2012). Technical limitations like the lattice graph induced distortions might be far less impactful than the ecological assumptions underlying the cost surface (Adriaensen et al., 2003). Gonzales & Gergel (2007) confirmed that contradictory ecological assumptions produce different results, but also found that absolute and relative cost values for different landscape categories have a greater influence on the results than expected. Due to these uncertainties associated with cost surfaces, I chose to use existing material as a starting point and focus on the delineation of populations and calculation of expansion rates.

2.4.3 Script module *leastcostpaths* (open-source)

The key software I used for this script module is the *Graph* module of the [scikit-image](#) package for Python (Walt van der et al., 2014) which supports landscape graph-based least-cost modelling and works with an 8-way neighborhood definition. The *MCP_Geometric* class is used to accumulate distance-weighted costs. The class is instantiated with an n-dimensional array, so the raster input needs to be converted first. Impassable barriers can be encoded with infinite (`numpy.inf`) or negative values. I used infinite values. The essential elements of this script module are two nested FOR loops as follows (simplified):

FOR each time step (year):

1. Instantiate an object of the *MCP_Geometric* class with the cost array:

```
mcp = MCP_Geometric(cost_array)
```

2. Accumulate costs with (known) start points and (new) end points:

```
acc_cost_array, traceback_array = mcp.find_costs(starts=known_coords,  
ends=new_coords, find_all_ends=True)
```

3. FOR each new point:

- 3.1. Trace back least-cost path:

```
path = mcp.traceback(new_coords[i, :])
```

- 3.2. Get the associated cost:

```
acc_cost = acc_cost_array[row_index][col_index]
```

The traced back coordinates, their associated accumulated cost and the observation year are then used to create a feature with a *MultiLineString* geometry. The complete code can be found in the code repository (<https://github.com/thunwal/bioinvasianalysis>). Figure 11 provides an overview of the program logic.

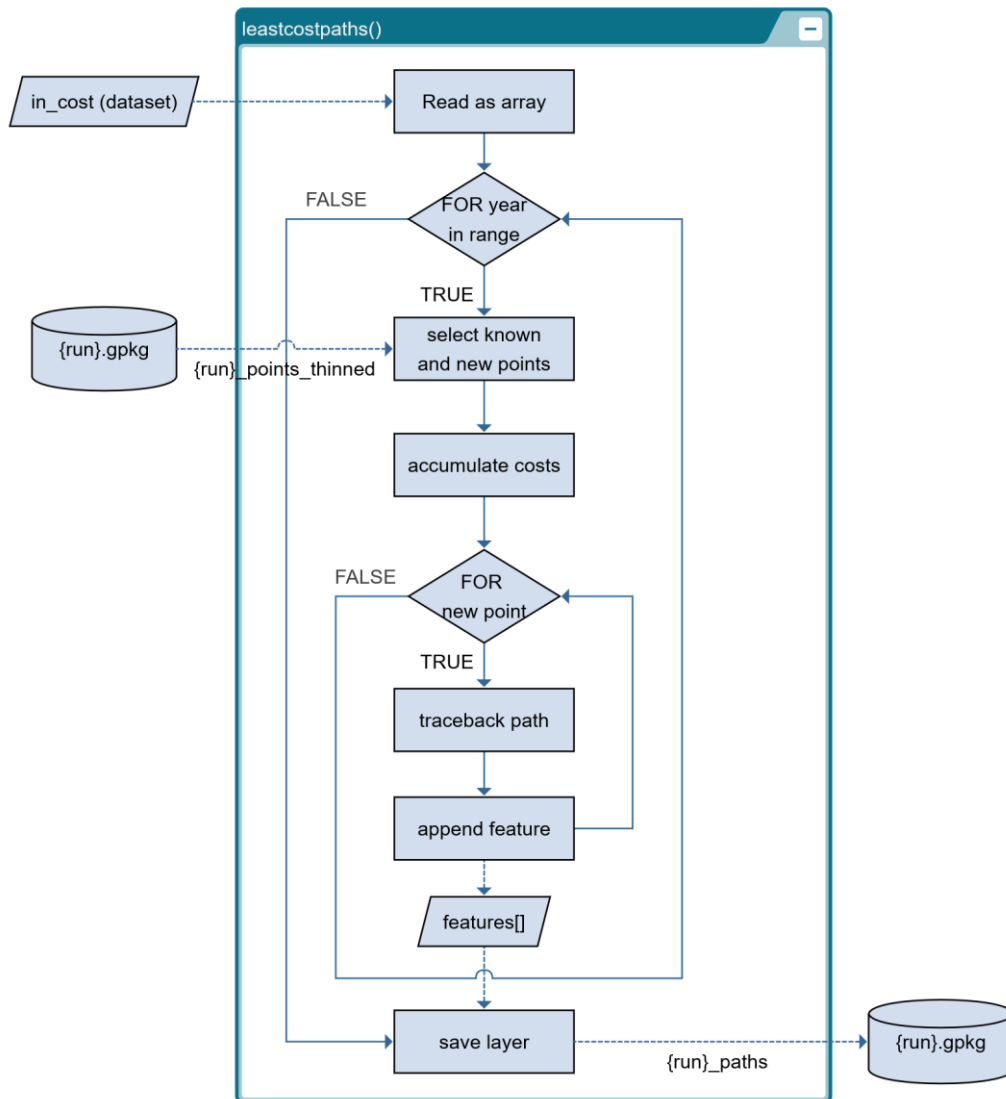


Figure 11: Program logic of the script module 'leastcostpaths', which is based on the open-source Python package scikit-image and serves to connect each observation with the nearest earlier observation according to least cost

2.4.4 Script modules *arcgis_distacc* and *arcgis_optpaths* (ArcGIS Pro)

The two script modules *arcgis_distacc* and *arcgis_optpaths* represent my initial code for least-cost modelling, which worked successfully until I ran into a blocking ArcGIS Pro bug associated with certain input data (see next chapter). While waiting for a bugfix, I wrote a new script module implementing open-source packages (*leastcostpaths*). When the bugfix was released, I confirmed the resolution, but found no time to seamlessly re-integrate the affected script modules. They are therefore currently commented out in the main script.

These ArcGIS Pro-based script modules implement the *Distance Accumulation* and *Optimal Path As Line* tools from ArcGIS Pro's *Distance toolset*. With version 2.5 of ArcGIS Pro, a new algorithm was introduced to the *Distance toolset* which does not work with a network representation of the cost surface, but accumulates cost along the slope of a continuous surface (Esri Inc., n.d.-b) (Figure 12). This algorithm calculates true distance and costs in all directions and is not affected of directional and length distortion.

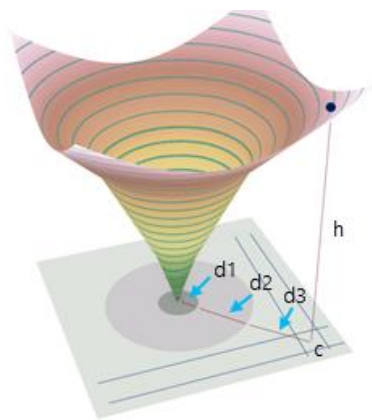


Figure 12: Illustration of the cost accumulation algorithm available since ArcGIS Pro 2.5. The height h of the accumulated cost surface at cell c is the sum of the slopes from the cost raster multiplied by the distances over which those slopes are active ($h = 3 * d1 + 2 * d2 + 1 * d3$). © Esri Inc.

The *Distance Accumulation* tool is used to accumulate costs and create back direction rasters in addition to the accumulated cost surfaces. Both outputs are required inputs for the *Optimal Path As Line* tool to determine least-cost paths. Thus, the essential elements of the two script modules are FOR loops iteratively calling the two tools as follows (simplified):

Script module *arcgis_distacc*

FOR each time step (year), create accumulated cost rasters and back direction rasters:

```
dist_acc_raster = arcpy.sa.DistanceAccumulation("known_points", "", "", path_dist_acc_raster, "", "", "", "", "", path_back_dir_raster, "", "", "", "", "", "", "GEODESIC")
```

Script module *arcgis_optpaths*

FOR each time step (year), traceback least-cost paths and save them as polyline features:

```
arcpy.sa.OptimalPathAsLine("new_points", path_dist_acc_raster, path_back_dir_raster, path_optpaths, year_field, "EACH_CELL", False)
```

2.4.5 Esri Bug Report for *Optimal Path As Line* tool

After I scaled the cost surface to a new value range and re-ran the ArcGIS Pro based script modules, which call *Distance Accumulation* followed by *Optimal Path As Line* (see previous chapter), the latter tool got stuck in an infinite loop at specific time steps, i.e. with specific combinations of accumulated cost rasters and presence data points. I reported the issue in the Esri Community Forum on December 12, 2023. On January 5, 2024, Esri opened a [Bug record](#), and on February 8, 2024, I was informed about the bugfix.

All related communication is listed below:

- 03.12.2023 - Issue reported in the Esri Community Forum: <https://community.esri.com/t5/arcgis-spatial-analyst-questions/optimal-path-as-line-issues-with-memory-leak-and/m-p/1356136/highlight/true#M12055>
- 03.12.2023 - Elizabeth, a product engineer of the Spatial Analyst team asks for reproducible data: <https://community.esri.com/t5/arcgis-spatial-analyst-questions/optimal-path-as-line-issues-with-memory-leak-and/m-p/1356166/highlight/true#M12061>

- 08.12.2023 - I send her an e-mail (as requested by her) with data, steps to reproduce and screenshots (Figure 13). One screenshot included is shown in Figure 14.
- In late December, Elizabeth informs me via e-mail that “a dev is working on it”. This E-Mail is no longer available.
- 05.01.2014 - Date of the bug record’s “Submitted” date.
- Somewhen in January, I ask Elizabeth for an update via e-mail, but do not hear back. This E-Mail is no longer available.
- 08.02.2024 - I am informed via forum that the bug is fixed:
<https://community.esri.com/t5/arcgis-spatial-analyst-questions/optimal-path-as-line-issues-with-memory-leak-and/m-p/1379996/highlight/true#M12116>
- 18.02.2024 - After reviving the affected script modules, I confirm the issue is no longer reproducible: <https://community.esri.com/t5/arcgis-spatial-analyst-questions/optimal-path-as-line-issues-with-memory-leak-and/m-p/1383892/highlight/true#M12126>

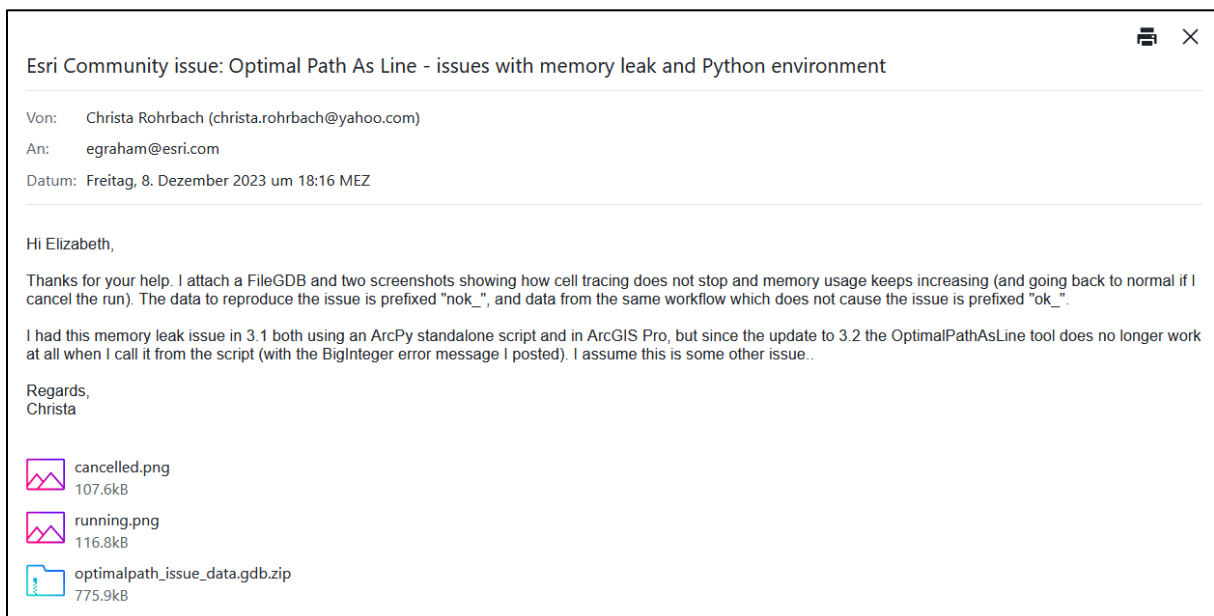


Figure 13: Screenshot of the first e-mail sent to Elizabeth Graham, a product engineer at Esri’s Spatial Analyst team, with data and steps needed to reproduce the issue

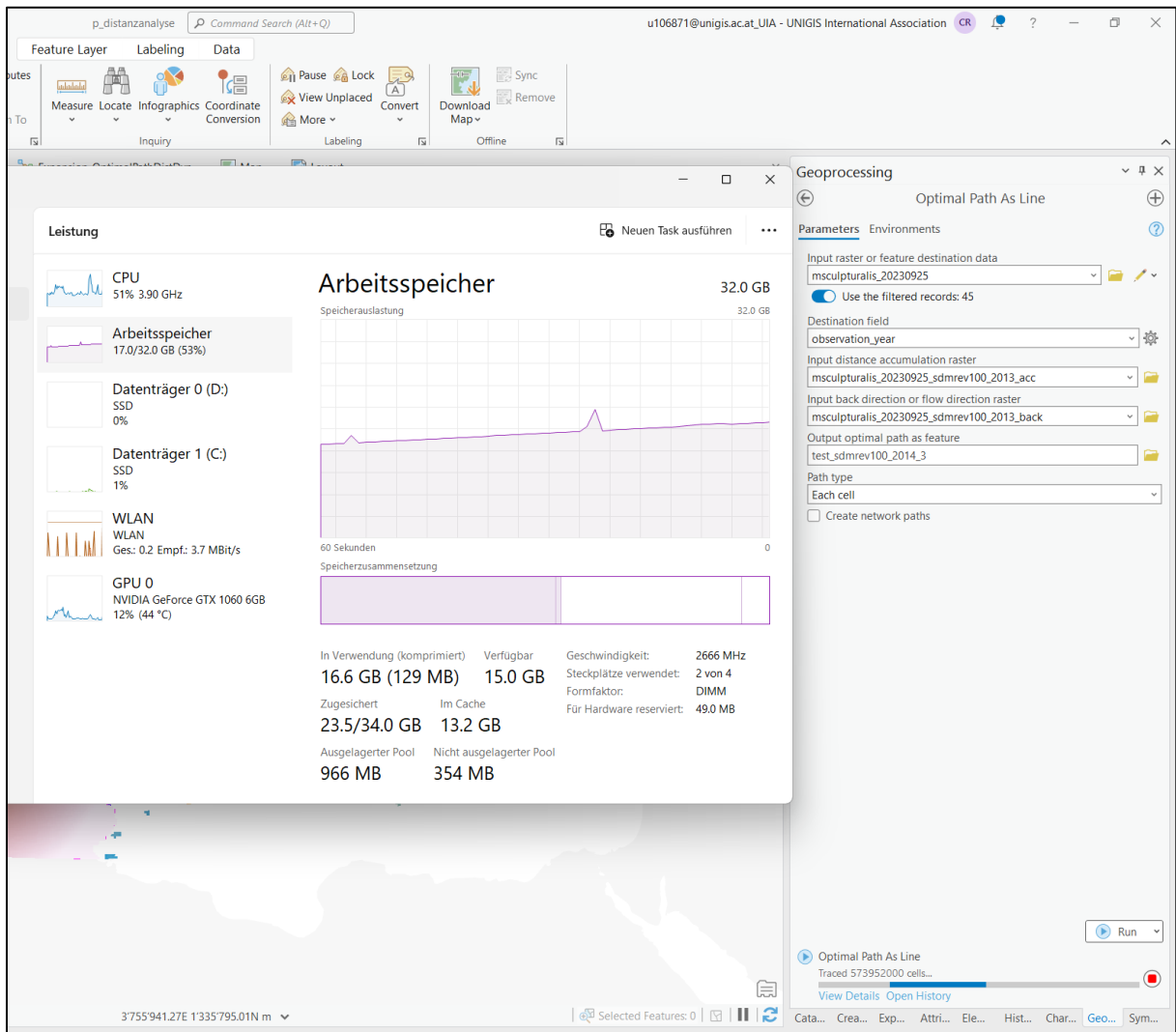


Figure 14: Screenshot of the 'Optimal Path As Line' tool being stuck in an infinite loop. The number of traced cells and memory usage would increase endlessly

2.5 Delineating populations

(Script module *populations*)

The result of the least-cost modelling step is one large network of dispersal paths (technically speaking, it is just a collection of MultiLineStrings). To isolate populations, an accumulated cost threshold in the form of a quantile value (e.g. 0.95) is applied to remove least-cost paths with higher accumulated cost from the result set. The remaining paths are tested for shared endpoints and each group of connected paths is assigned a unique population ID (function *group_paths*, Figure 15). The presence points are subsequently assigned the population ID of the nearest path (function *group_points*, Figure 16). The nearest path is identified by searching for a path endpoint within a search radius of half the diagonal of a cell, calculated as:

$$\frac{a\sqrt{2}}{2}$$

with a representing the cell side length. The search radius ensures that points isolated either by a barrier or by the removal of high-cost paths are not assigned to a population.

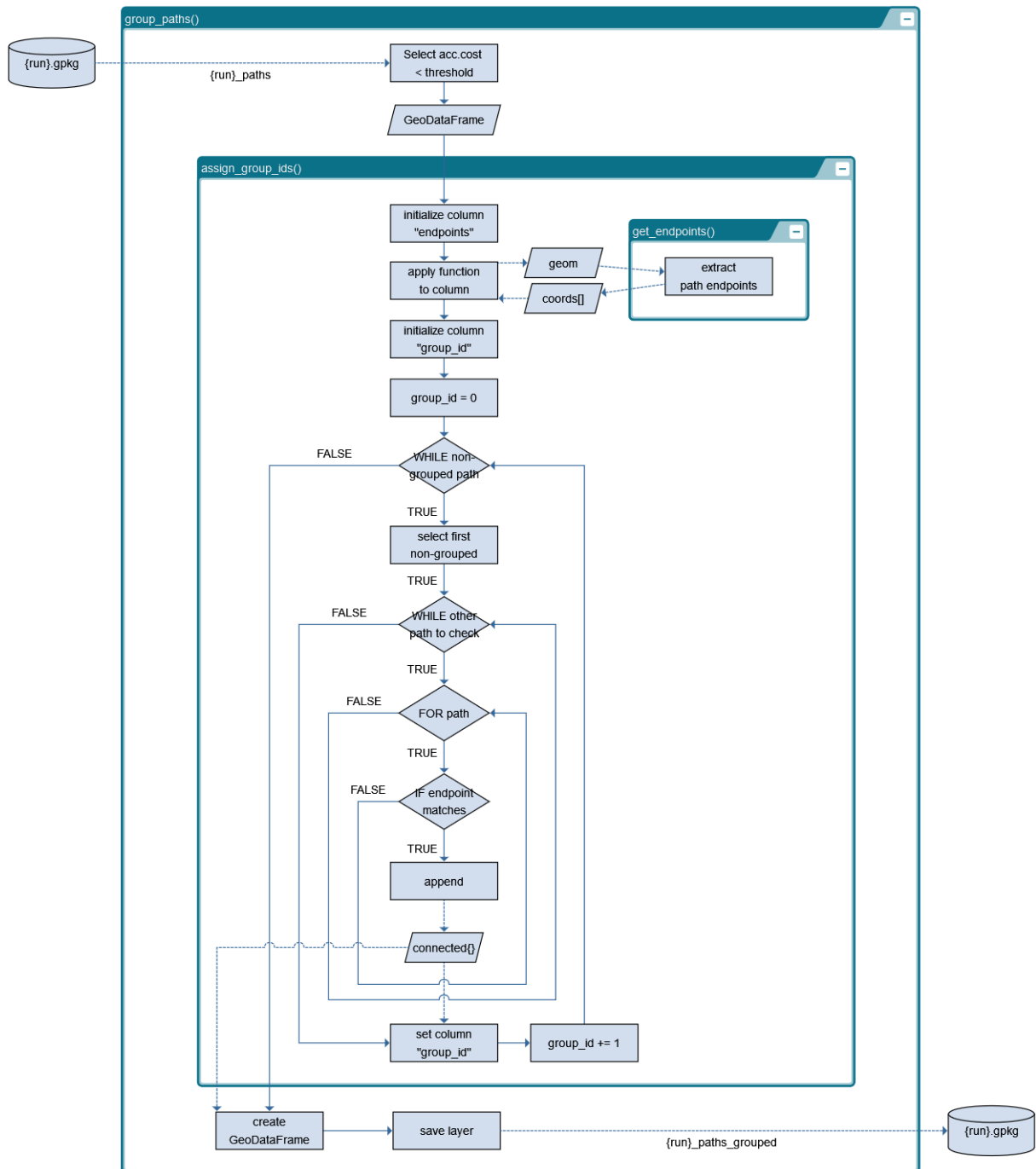


Figure 15: Program logic of the function 'group_paths'. Least-cost paths exceeding an accumulated cost threshold are removed from the result set, the remaining least-cost paths checked for connectivity by comparing endpoint coordinates, and population identifiers assigned to each set of connected paths

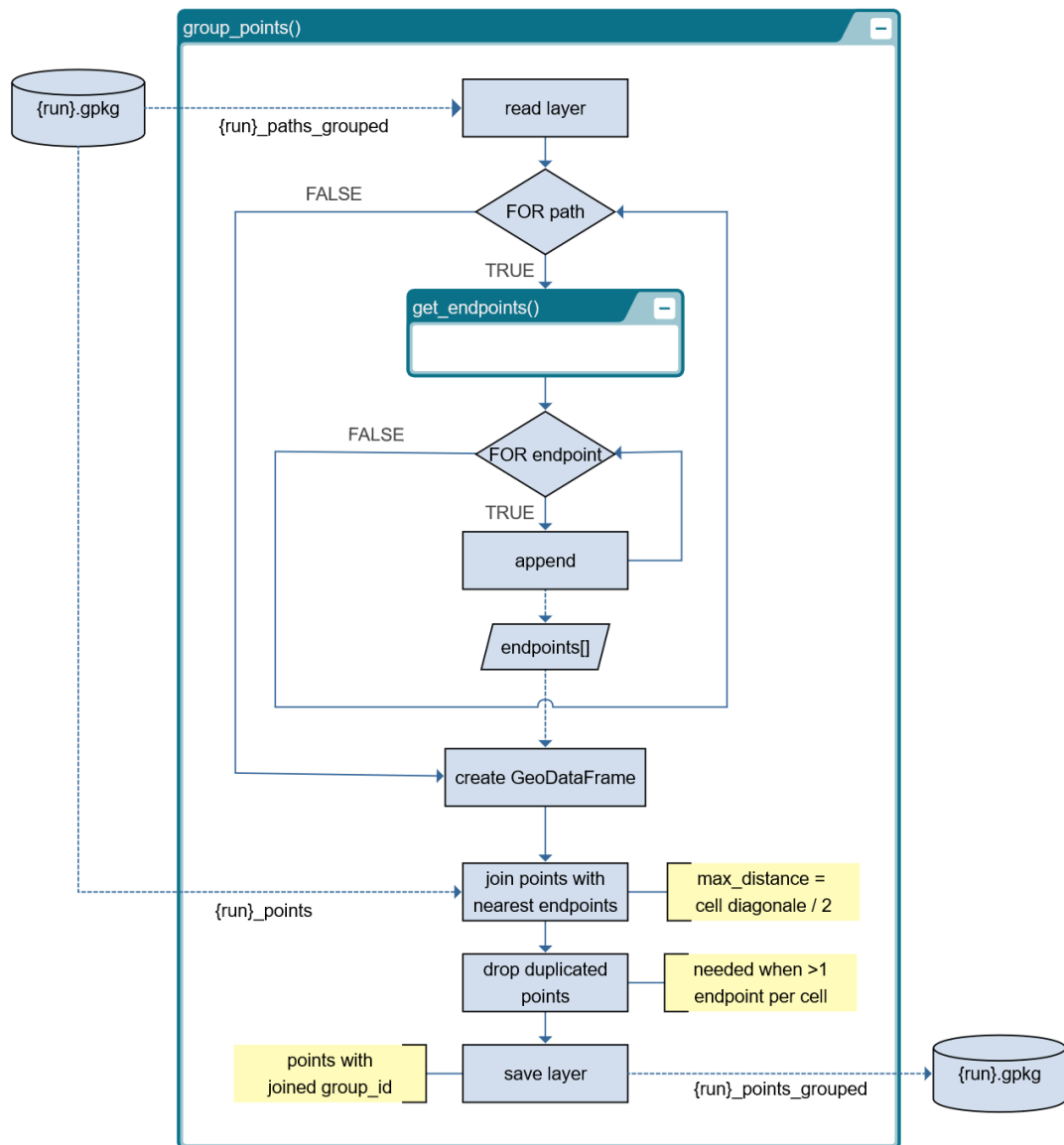


Figure 16: Program logic of the function 'group_points'. Presence data points are assigned to populations by a spatial join with the nearest endpoint of a least-cost path. The search radius is restricted to a half cell diagonal to avoid assigning isolated points to a population

To identify a suitable accumulated cost threshold, the function *sensitivity_analysis* runs a sensitivity analysis for the effect of the accumulated cost threshold on the number of delineated populations (Figure 17). The function runs the delineation process for a quantile range from an upper outlier fence to 1.00 in 0.001 increments and logs the results to a CSV file. The upper outlier fence is calculated by the function *upper_outlier_fence* as $Q3 + 1.5 \times IQR$, followed by conversion to a quantile. The results clearly show the inverse relationship of the threshold and the number of resulting populations (Figure 4). I chose to apply the 95th quantile as a threshold so that the resulting populations match the known genetic groups (Lanner et al., 2021) as closely as possible and that the threshold value falls within a range of low sensitivity. Figure 18 shows the resulting populations after application of the 90th quantile, respectively, the 95th quantile as a threshold.

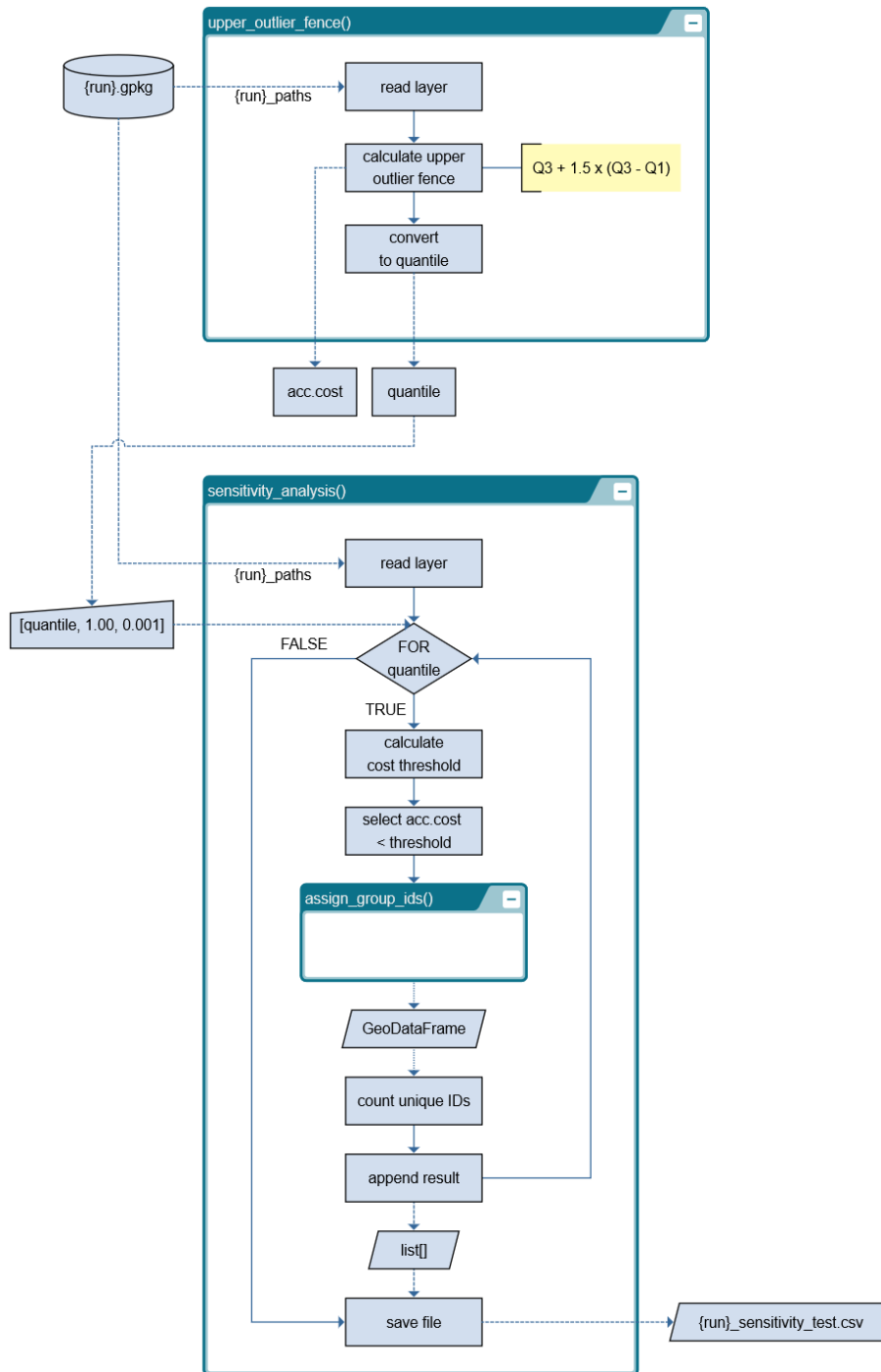


Figure 17: Program logic of the functions 'upper_outlier_fence' and 'sensitivity_analysis'. These functions assist the identification of a suitable accumulated cost threshold to remove high-cost paths and delineate populations

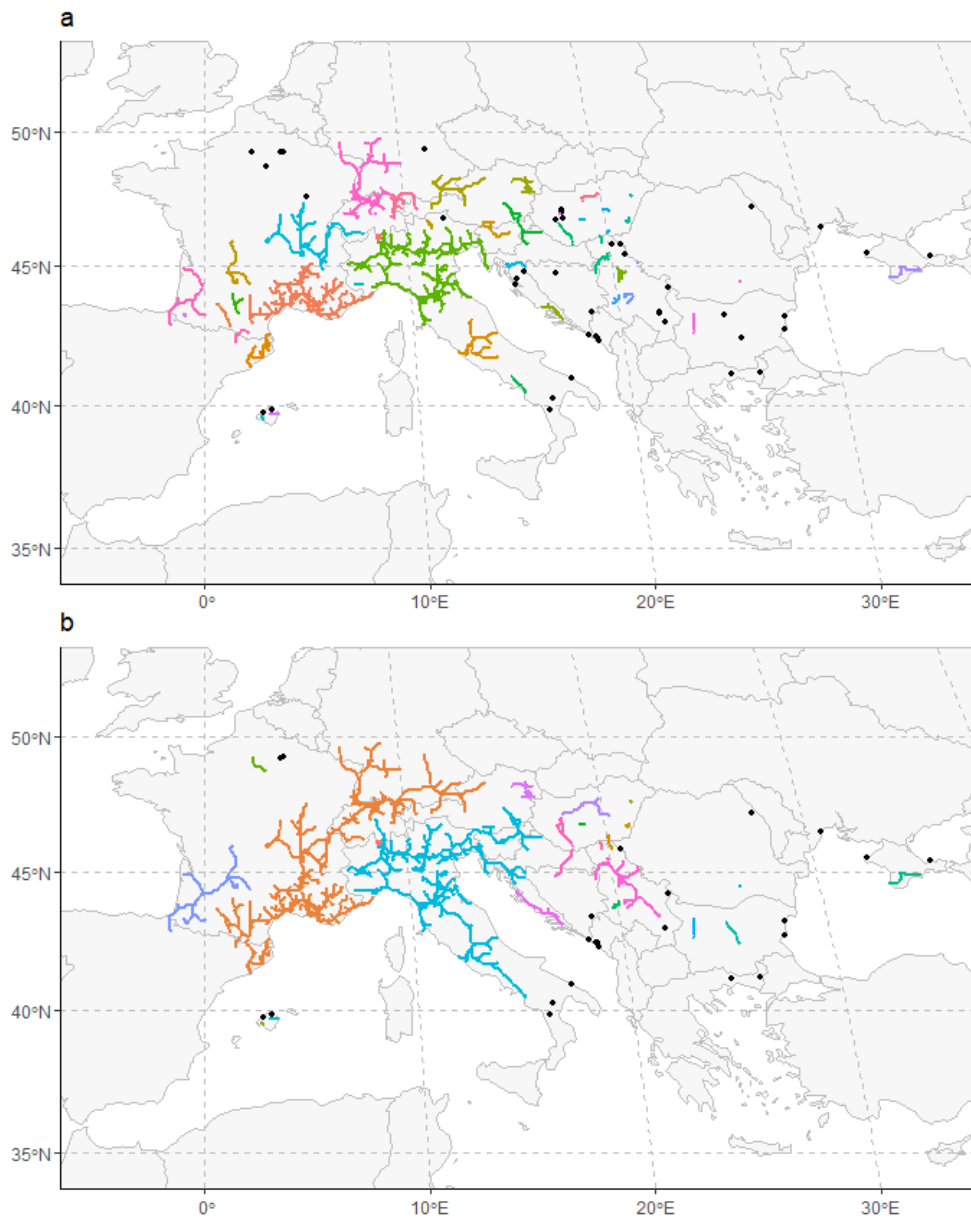


Figure 18: Resulting populations after application of an accumulated cost threshold of Q0.90 (a) and Q0.95 (b). The higher the threshold value, the fewer paths are removed and the fewer populations result. Black dots indicate presence data points which are isolated by barriers or by the removal of high-cost paths

2.6 Calculating expansion rate

(Script module *expansionrate*)

The expansion rate of a non-native species is usually estimated by a regression approach, after reconstruction of the temporal dynamics of range expansion (Hui & Richardson, 2017, p. 22). Most methods are based on the construction of the annual invaded area, which is constructed using a variety of different methods:

- cumulative area of occupied grid cells (Sandvik, 2020; Weber, 1998)
- cumulative area of occupied administrative units (Roques et al., 2016)
- alpha convex hull of observations (Lanner et al., 2020)
- aggregated buffered observations (Verdasca et al., 2021)
- probabilistic range, with the boundary delineated at the 95% contour line (Bertolino et al., 2016; Carvalho et al., 2020; Fraser et al., 2015)
- interpolation of trap data, with the boundary delineated at certain trap catch thresholds (Sharov et al., 1995; Tobin et al., 2007)

Next, the invaded area is regressed as a function of time, resulting in the measure km²/year (Fraser et al., 2015; Lanner et al., 2020; Roques et al., 2016; Verdasca et al., 2021; Weber, 1998). Alternatively, one of the following derivatives of area, measured in km/year, is used for regression:

- The side length $s = \sqrt{A}$ of a square with the area A of the occupied area (Gilbert & Liebhold, 2010)
- The radius $r = \sqrt{A \times \pi^{-1}}$ of a circle with the area A of the occupied area (Sandvik, 2020)
- The annual displacement of the invaded area boundary, either averaged (Sharov et al., 1997; Tobin et al., 2007) or measured along specific axes (Fraser et al., 2015)

I found only one method not depending on the reconstruction of an invaded area: Distance regression, introduced by Liebhold et al. (1992). The cumulative Euclidean distance of observations to the point of introduction is regressed as a function of time. In a performance comparison of three methods against a simulated expansion with known expansion rate, this

method proved least sensitive to small samples and sample point distribution (Gilbert & Liebhold, 2010). With regard to the small sample sizes after the delineation of populations and due to its simplicity, the distance regression method appears well suited for my study. Figure 19 shows how I implemented the method. Distance regression as well as the boundary displacement method require knowledge of the introduction point, though, and the spatial allocation of the introduction point influences the regression result (Gilbert & Liebhold, 2010). It should also be noted that some authors regress the length of least-cost paths instead of Euclidean distance (Kaláb et al., 2021; Mineur et al., 2010). For better comparability with findings from literature, I chose to use Euclidean distance.

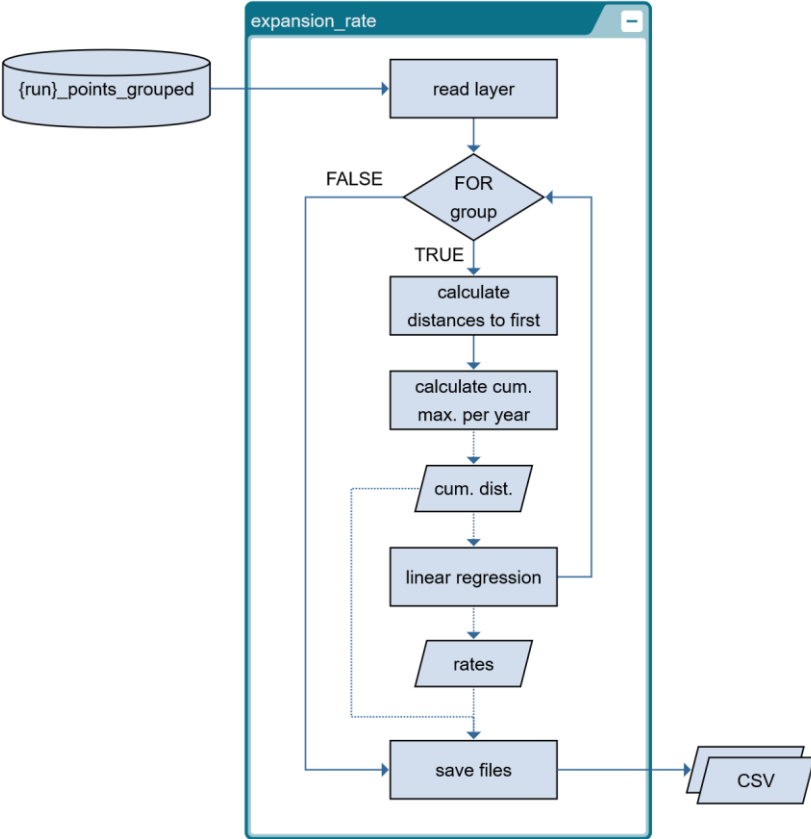


Figure 19: Program logic of the function 'expansion_rate'. For each population, the cumulative Euclidean distance of observations to the point of introduction is regressed against time

References

- Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of “least-cost” modelling as a functional landscape model. *Landscape and Urban Planning*, 64(4), 233–247. [https://doi.org/10.1016/S0169-2046\(02\)00242-6](https://doi.org/10.1016/S0169-2046(02)00242-6)
- Aguado, O., Hernández-Castellano, C., Isamat, E. B., Cassina, M. M., Navarro, D., Stefanescu, C., & Vicens, N. (2018). Megachile (Callomegachile) sculpturalis Smith, 1853 (Apoidea: Megachilidae): a new exotic species in the Iberian Peninsula, and some notes about its biology. *Butlletí de la Institució Catalana d’Història Natural*, 157–162.
- Ahuja, R., Magnanti, T., & Orlin, J. (1993). Network Flows: Theory, Algorithms and Applications. In *Journal of The Operational Research Society - J OPER RES SOC* (Vol. 45). <https://doi.org/10.2307/2583863>
- Amiet, F. (2012). Die Blattschneiderbiene *Megachile sculpturalis* Smith, 1853 (Hymenoptera, Apidae) nun auch in der Schweiz. *Entomo Helvetica*, 5, 157–159.
- Ayad, H. (1967). *System Evaluation by the Simplified Proportional Assignment Technique* [PhD thesis]. Purdue University.
- Bellard, C., Cassey, P., & Blackburn, T. M. (2016). Alien species as a driver of recent extinctions. *Biology Letters*, 12(2), 20150623. <https://doi.org/10.1098/rsbl.2015.0623>
- Bertolino, S., Lioy, S., Laurino, D., Manino, A., & Porporato, M. (2016). Spread of the invasive yellow-legged hornet *Vespa velutina* (Hymenoptera: Vespidae) in Italy. *Applied Entomology and Zoology*, 51(4), 589–597. <https://doi.org/10.1007/s13355-016-0435-2>
- Bila Dubaić, J., Lanner, J., Rohrbach, C., Meimberg, H., Wyatt, F., Čačija, M., Galešić, M., Ješovnik, A., Samurović, K., Plećaš, M., Raičević, J., & Četković, A. (2022). Towards a real-time tracking of an expanding alien bee species in Southeast Europe through citizen science and floral host monitoring. *Environmental Research Communications*, 4(8), 085001. <https://doi.org/10.1088/2515-7620/ac8398>
- Bila Dubaić, J., Plećaš, M., Raičević, J., Lanner, J., & Četković, A. (2022). Early-phase colonization by introduced sculptured resin bee (Hymenoptera, Megachilidae, *Megachile sculpturalis*) revealed by local floral resource variability. *NeoBiota*, 73, 57–85. <https://doi.org/10.3897/arphapreprints.e80975>

- Bossdorf, O., Auge, H., Lafuma, L., Rogers, W. E., Siemann, E., & Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, *144*(1), 1–11. <https://doi.org/10.1007/s00442-005-0070-z>
- Carlyle, M. W., & Wood, K. R. (2005). Near-shortest and K-shortest simple paths. *Networks*, *46*(2), 98–109. <https://doi.org/10.1002/net.20077>
- Carvalho, J., Hipólito, D., Santarém, F., Martins, R., Gomes, A., Carmo, P., Rodrigues, R., Grosso-Silva, J., & Fonseca, C. (2020). Patterns of *Vespa velutina* invasion in Portugal using crowdsourced data. *Insect Conservation and Diversity*, *13*(5), 501–507. <https://doi.org/10.1111/icad.12418>
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271. <https://doi.org/10.1007/BF01386390>
- Douglas, D. H. (1994). Least-cost Path in GIS Using an Accumulated Cost Surface and Slope-lines. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *31*(3), 37–51. <https://doi.org/10.3138/D327-0323-2JUT-016M>
- Driezen, K., Adriaensen, F., Rondinini, C., Doncaster, C. P., & Matthysen, E. (2007). Evaluating least-cost model predictions with empirical dispersal data: A case-study using radiotracking data of hedgehogs (*Erinaceus europaeus*). *Ecological Modelling*, *209*(2), 314–322. <https://doi.org/10.1016/j.ecolmodel.2007.07.002>
- Ellstrand, N. C., & Schierenbeck, K. A. (2006). Hybridization as a stimulus for the evolution of invasiveness in plants? *Euphytica*, *148*(1), 35–46. <https://doi.org/10.1007/s10681-006-5939-3>
- Eppstein, D. (1998). Finding the k Shortest Paths. *SIAM Journal on Computing*, *28*(2), 652–673. <https://doi.org/10.1137/S0097539795290477>
- Esri Inc. (n.d.-a). Distance accumulation algorithm. In *ArcGIS Pro Tool Reference*. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/distance-accumulation-algorithm.htm>.
- Esri Inc. (n.d.-b). Migrating from legacy distance tools to distortion free distance tools. In *ArcGIS Pro Documentation*. <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-analyst/migrate-from-legacy-distance-tools-to-distortion-free-distance-tools.htm>.
- Etherington, T. R. (2016). Least-Cost Modelling and Landscape Ecology: Concepts, Applications, and Opportunities. *Current Landscape Ecology Reports*, *1*(1), 40–53. <https://doi.org/10.1007/s40823-016-0006-9>
- Etten, J. van. (2017). R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, *76*, 1–21. <https://doi.org/10.18637/jss.v076.i13>

- Facon, B., Pointier, J.-P., Jarne, P., Sarda, V., & David, P. (2008). High Genetic Variance in Life-History Strategies within Invasive Populations by Way of Multiple Introductions. *Current Biology*, 18(5), 363–367. <https://doi.org/10.1016/j.cub.2008.01.063>
- Fortel, L., Henry, M., Guilbaud, L., Mouret, H., & Vaissière, B. E. (2016). Use of human-made nesting structures by wild bees in an urban environment. *Journal of Insect Conservation*, 20(2), 239–253. <https://doi.org/10.1007/s10841-016-9857-y>
- Fraser, E. J., Lambin, X., Travis, J. M. J., Harrington, L. A., Palmer, S. C. F., Bocedi, G., & Macdonald, D. W. (2015). Range expansion of an invasive species through a heterogeneous landscape – the case of American mink in Scotland. *Diversity and Distributions*, 21(8), 888–900. <https://doi.org/10.1111/ddi.12303>
- Gühr, C., & Westrich, P. (2013). Ein Brutnachweis der adventiven Riesen-Harzbiene (*Megachile sculpturalis* Smith 1853) in Südfrankreich (Hymenoptera, Apidae). *Eucera*, 7, 1–9.
- Gilbert, M., & Liebhold, A. (2010). Comparing methods for measuring the rate of spread of invading populations. *Ecography*, 33(5), 809–817. <https://doi.org/10.1111/j.1600-0587.2009.06018.x>
- Gogala, A., Zadavec, B., et al. (2018). First record of *Megachile sculpturalis* Smith in Slovenia (Hymenoptera: Megachilidae). *Acta Entomologica Slovenica*, 26(1), 79–82.
- Gonzales, E. K., & Gergel, S. E. (2007). Testing assumptions of cost surface analysis – a tool for invasive species management. *Landscape Ecology*, 22(8), 1155–1168. <https://doi.org/10.1007/s10980-007-9106-6>
- Goodchild, M. F. (1977). An Evaluation of Lattice Solutions to the Problem of Corridor Location. *Environment and Planning A: Economy and Space*, 9(7), 727–738. <https://doi.org/10.1068/a090727>
- Gradinarov, D., Petrova, Y., & Ljubomirov, T. (2023). Nesting observation of the Sculptured Resin bee *Megachile sculpturalis* F. Smith, 1853 (Hymenoptera: Megachilidae) in Bulgaria. *ZooNotes*.
- Grossi, A., Greggio, M., & Corazza, C. (2018). Prima segnalazione di *Ceratina dallatorreana* Friese 1896 nella laguna di Venezia e prime osservazioni di *Megachile sculpturalis* Smith 1853 (Hymenoptera Apoidea) nella provincia di Ferrara. *Quaderni Del Museo Civico Di Storia Naturale Di Ferrara*, 6, 87–88.
- Guariento, E., Lanner, J., Staggl, M. A., & Kranebitter, P. (2019). *Megachile sculpturalis* (Smith, 1853) (Hymenoptera: Megachilidae), the giant resin bee new for South Tyrol with a newly described plant species interaction. *Gredleriana*, 19.

- Guerriero, F., Musmanno, R., Lacagnina, V., & Pecorella, A. (2001). A Class of Label-Correcting Methods for the K Shortest Paths Problem. *Operations Research*, 49(3), 423–429. <https://doi.org/10.1287/opre.49.3.423.11217>
- Hui, C., & Richardson, D. M. (2017). *Invasion Dynamics* (C. Hui & D. M. Richardson, Eds.). Oxford University Press.
- Ivanov, S. P., & Fateryga, A. V. (2019). First record of the invasive giant resin bee *Megachile* (*Callomegachile*) *sculpturalis* Smith, 1853 (Hymenoptera: Megachilidae) in the Crimea. *Far Eastern Entomologist*, 395, 7–13. <https://doi.org/10.25221/fee.395.2>
- Kaláb, O., Pyszko, P., & Kočárek, P. (2021). Estimation of the Recent Expansion Rate of *Ruspolia nitidula* (Orthoptera) on a Regional and Landscape Scale. *Insects*, 12(7), 639. <https://doi.org/10.3390/insects12070639>
- Katoh, N., Ibaraki, T., & Mine, H. (1982). An efficient algorithm for K shortest simple paths. *Networks*, 12(4), 411–427. <https://doi.org/10.1002/net.3230120406>
- Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, 431(7005), 177–181. <https://doi.org/10.1038/nature02807>
- Kovács, T. (2015). *Megachile sculpturalis* Smith, 1853 in Hungary (Hymenoptera, Megachilidae). *Folio Historico-Naturalia Musei Matraensis*, 39, 73–76.
- Lanner, J., Dubos, N., Geslin, B., Leroy, B., Hernández-Castellano, C., Dubaić, J. B., Bortolotti, L., Calafat, J. D., Četković, A., Flaminio, S., Le Féon, V., Margalef-Marrase, J., Orr, M., Pachinger, B., Ruzzier, E., Smagghe, G., Tuerlings, T., Vereecken, N. J., & Meimberg, H. (2022). On the road: Anthropogenic factors drive the invasion risk of a wild solitary bee species. *Science of The Total Environment*, 827, 154246. <https://doi.org/10.1016/j.scitotenv.2022.154246>
- Lanner, J., Gstötenmayer, F., Curto, M., Geslin, B., Huchler, K., Orr, M. C., Pachinger, B., Sedivy, C., & Meimberg, H. (2021). Evidence for multiple introductions of an invasive wild bee species currently under rapid range expansion in Europe. *BMC Ecology and Evolution*, 21, 17. <https://doi.org/10.1186/s12862-020-01729-x>
- Lanner, J., Huchler, K., Pachinger, B., Sedivy, C., & Meimberg, H. (2020). Dispersal patterns of an introduced wild bee, *Megachile sculpturalis* Smith, 1853 (Hymenoptera: Megachilidae) in European alpine countries. *PLOS ONE*, 15(7), e0236042. <https://doi.org/10.1371/journal.pone.0236042>
- Lawson Handley, L.-J., Estoup, A., Evans, D. M., Thomas, C. E., Lombaert, E., Facon, B., Aebi, A., & Roy, H. E. (2011). Ecological genetics of invasive alien species. *BioControl*, 56(4), 409–428. <https://doi.org/10.1007/s10526-011-9386-2>

- Le Féon, V., Aubert, M., Genoud, D., Andrieu-Ponel, V., Westrich, P., & Geslin, B. (2018). Range expansion of the Asian native giant resin bee *Megachile sculpturalis* (Hymenoptera, Apoidea, Megachilidae) in France. *Ecology and Evolution*, *00*, 1–9. <https://doi.org/10.1002/ece3.3758>
- Le Féon, V., & Geslin, B. (2018). Écologie et distribution de l'abeille originaire d'Asie *Megachile sculpturalis* SMITH 1853 (Apoidea – Megachilidae – Megachilini) : Un État des connaissances dix ans après sa première observation en Europe. *Osmia*, *7*, 31–39.
- Lewis, J. (2021). Probabilistic Modelling for Incorporating Uncertainty in Least Cost Path Results: A Postdictive Roman Road Case Study. *Journal of Archaeological Method and Theory*, *28*(3), 911–924. <https://doi.org/10.1007/s10816-021-09522-w>
- Liebhold, A. M., Halverson, J. A., & Elmes, G. A. (1992). Gypsy Moth Invasion in North America: A Quantitative Analysis. *Journal of Biogeography*, *19*(5), 513–520. <https://doi.org/10.2307/2845770>
- Lombard, K., & Church, R. L. (1993). The gateway shortest path problem: Generating alternative routes for a corridor location problem. *Geographical Systems*, *1*(1), 25–45.
- McRae, B. H., Dickson, B. G., Keitt, T. H., & Shah, V. B. (2008). Using Circuit Theory to Model Connectivity in Ecology, Evolution, and Conservation. *Ecology*, *89*(10), 2712–2724. <https://doi.org/10.1890/07-1861.1>
- Miller, N., Estoup, A., Toepfer, S., Bourguet, D., Lapchin, L., Derridj, S., Kim, K. S., Reynaud, P., Furlan, L., & Guillemaud, T. (2005). Multiple Transatlantic Introductions of the Western Corn Rootworm. *Science*, *310*(5750), 992–992. <https://doi.org/10.1126/science.1115871>
- Mineur, F., Davies, A. J., Maggs, C. A., Verlaque, M., & Johnson, M. P. (2010). Fronts, jumps and secondary introductions suggested as different invasion patterns in marine species, with an increase in spread rates over time. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1694), 2693–2701. <https://doi.org/10.1098/rspb.2010.0494>
- Ortego, J., Céspedes, V., Millán, A., & Green, A. J. (2021). Genomic data support multiple introductions and explosive demographic expansions in a highly invasive aquatic insect. *Molecular Ecology*, *30*(17), 4189–4203. <https://doi.org/10.1111/mec.16050>
- Ortiz-Sánchez, F. J., Navarro, J. F., & Taeger, U. (2018). *Megachile* (*Callomegachile*) *sculpturalis* Smith, 1853, nueva especie para la fauna Ibérica (Hymenoptera, Megachilidae). *Boletín de la Sociedad Entomológica Aragonesa*, *63*, 259–261.
- Parys, K. A., Tripodi, A. D., & Sampson, B. J. (2015). The Giant Resin Bee, *Megachile sculpturalis* Smith: New Distributional Records for the Mid- and Gulf-south USA. *Biodiversity Data Journal*, *3*, e6733. <https://doi.org/10.3897/BDJ.3.e6733>

- Perttola, W. (2022). Digital Navigator on the Seas of the Selden Map of China: Sequential Least-Cost Path Analysis Using Dynamic Wind Data. *Journal of Archaeological Method and Theory*, 29(2), 688–721. <https://doi.org/10.1007/s10816-021-09534-6>
- Phillips, S. J., Williams, P., Midgley, G., & Archer, A. (2008). Optimizing Dispersal Corridors for the Cape Proteaceae Using Network Flow. *Ecological Applications*, 18(5), 1200–1211. <https://doi.org/10.1890/07-0507.1>
- Pinto, N., & Keitt, T. H. (2009). Beyond the least-cost path: Evaluating corridor redundancy using a graph-theoretic approach. *Landscape Ecology*, 24(2), 253–266. <https://doi.org/10.1007/s10980-008-9303-y>
- Poggi, R., Tavano, M. L., & Bonifacino, M. (2020). Reperti liguri di Megachile (Callomegachile) sculpturalis Smith, 1853 (Hymenoptera, Megachilidae). *Annali del museo civico di storia naturale "G. Doria" Genova*, 9(410), 1–6.
- Polidori, C., & Sánchez-Fernández. (2020). Environmental niche and global potential distribution of the giant resin bee Megachile sculpturalis, a rapidly spreading invasive pollinator. *Global Ecology and Conservation*. <https://doi.org/10.1016/j.gecco.2020.e01365>
- Quaranta, M., Sommaruga, A., Balzarini, P., & Felicioli, A. (2014). A new species for the bee fauna of Italy: Megachile sculpturalis continues its colonization of Europe. *Bulletin of Insectology*, 67(2), 287–293.
- Rickenbach, F., & Sprecher, E. (2018). Neue Bestäuberin von Bienenbäumen: Die Asiatische Mörtelbiene. *Schweizerische Bienen-Zeitung*, 9, 15–17.
- Robinson, C. V., Garcia de Leaniz, C., James, J., Cable, J., Orozco-terWengel, P., & Consuegra, S. (2018). Genetic diversity and parasite facilitated establishment of the invasive signal crayfish (*Pacifastacus leniusculus*) in Great Britain. *Ecology and Evolution*, 8(18), 9181–9191. <https://doi.org/10.1002/ece3.4235>
- Rome, Q., Dambrine, L., Onate, C., Muller, F., Villemant, C., García-Pérez, A., Maia, M., Carvalho-Esteves, P., & Bruneau, E. (2013). Spread of the invasive hornet *Vespa velutina* Lepeletier, 1836, in Europe in 2012 (Hym., Vespidae). *Bulletin de La Société Entomologique de France*, 118(1), 21–22. <https://doi.org/10.3406/bsef.2013.2580>
- Roques, A., Auger-Rozenberg, M.-A., Blackburn, T. M., Garnas, J., Pyšek, P., Rabitsch, W., Richardson, D. M., Wingfield, M. J., Liebhold, A. M., & Duncan, R. P. (2016). Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years. *Biological Invasions*, 18(4), 907–920. <https://doi.org/10.1007/s10530-016-1080-y>
- Ruzzier, E., Menchetti, M., Bortolotti, L., Selis, M., Monterastelli, E., & Forbicioni, L. (2020). Updated distribution of the invasive Megachile sculpturalis (Hymenoptera:

- Megachilidae) in Italy and its first record on a Mediterranean island. *Biodiversity Data Journal*, 8, e57783.
- Saerens, M., Achbany, Y., Fouss, F., & Yen, L. (2009). Randomized Shortest-Path Problems: Two Related Models. *Neural Computation*, 21(8), 2363–2404.
<https://doi.org/10.1162/neco.2009.11-07-643>
- Sandvik, H. (2020). Expansion Speed as a Generic Measure of Spread for Alien Species. *Acta Biotheoretica*, 68(2), 227–252. <https://doi.org/10.1007/s10441-019-09366-8>
- Scaparra, M. P., Church, R. L., & Medrano, F. A. (2014). Corridor location: The multi-gateway shortest path model. *Journal of Geographical Systems*, 16(3), 287–309.
<https://doi.org/10.1007/s10109-014-0197-8>
- Sharov, A. A., Liebhold, A. M., & Roberts, A. E. (1997). Methods for monitoring the spread of gypsy moth (Lepidoptera: Lymantriidae) populations in the Appalachian Mountains. *Journal of Economic Entomology*, 90(5), 1259–1266.
- Sharov, A. A., Roberts, E. A., Liebhold, A. M., & Ravlin, F. W. (1995). Gypsy moth (Lepidoptera: Lymantriidae) spread in the central Appalachians: Three methods for species boundary estimation. *Environmental Entomology*, 24(6), 1529–1538.
<https://doi.org/10.1093/ee/24.6.1529>
- Springer Nature Switzerland AG. (n.d.). Landscape Ecology - Aims and scope. In *SpringerLink*. <https://link.springer.com/journal/10980/aims-and-scope>.
- Tang, W., & Bennett, D. A. (2010). Agent-based Modeling of Animal Movement: A Review. *Geography Compass*, 4(7), 682–700. <https://doi.org/10.1111/j.1749-8198.2010.00337.x>
- Tobin, P. C., Liebhold, A. M., & Anderson Roberts, E. (2007). Comparison of methods for estimating the spread of a non-indigenous species. *Journal of Biogeography*, 34(2), 305–312. <https://doi.org/10.1111/j.1365-2699.2006.01600.x>
- Tomlin, D. (2010). Propagating radial waves of travel cost in a grid. *International Journal of Geographical Information Science*, 24(9), 1391–1413.
<https://doi.org/10.1080/13658811003779152>
- Turner, A. K. F. (1968). *Computer-assisted procedures to generate and evaluate regional highway alternatives* (1-6-1). Purdue University.
- Turner, A. K., & Miles, R. D. (1971). THE GCARS SYSTEM: A COMPUTER-ASSISTED METHOD OF REGIONAL ROUTE LOCATION. *Highway Research Record*, 348, 1–15.
- Verdasca, M. J., Rebelo, H., Carvalheiro, L. G., & Rebelo, R. (2021). Invasive hornets on the road: Motorway-driven dispersal must be considered in management plans of *Vespa velutina*. *NeoBiota*, 69, 177–198. <https://doi.org/10.3897/neobiota.69.71352>

- Vereecken, N. J., & Barbier, E. (2009). Premières données sur la présence de l'abeille asiatique *Megachile* (*Callomegachile*) *sculpturalis* SMITH (Hymenoptera, Megachilidae) en Europe. *Osmia*, 3, 4–6. <https://doi.org/10.47446/OSMIA3.3>
- Walt van der, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in Python. *PeerJ*, 2, e453. <https://doi.org/10.7717/peerj.453>
- Weber, E. (1998). The dynamics of plant invasions: A case study of three exotic goldenrod species (*Solidago* L.) in Europe. *Journal of Biogeography*, 25(1), 147–154. <https://doi.org/10.1046/j.1365-2699.1998.251119.x>
- Westrich, P. (2020). Neues zur Ausbreitung der Mörtelbiene *Megachile sculpturalis* Smith 1853 (Hymenoptera: Anthophila) in Deutschland – Stand Oktober 2019. *Eucera*, 14, 12–15.
- Westrich, P., Knapp, A., & Berney, I. (2015). *Megachile sculpturalis* Smith 1853 (Hymenoptera, Apidae), a new species for the bee fauna of Germany, now north of the Alps. *Eucera*, 9, 3–10.
- Zalewski, A., Michalska-Parda, A., Bartoszewicz, M., Kozakiewicz, M., & Brzeziński, M. (2010). Multiple introductions determine the genetic structure of an invasive species population: American mink *Neovison vison* in Poland. *Biological Conservation*, 143(6), 1355–1363. <https://doi.org/10.1016/j.biocon.2010.03.009>
- Zandigiaco, P., & Grion, M. (2017). First finding of *Megachile sculpturalis* Smith (Hymenoptera, Megachilidae) in Friuli Venezia Giulia (North-Eastern Italy). *Gortania. Botanica, Zoologica*, 39, 37–40.
- Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, 27(6), 777–797. <https://doi.org/10.1007/s10980-012-9737-0>