# Methods

## Species Distribution Models

### Preparing the data

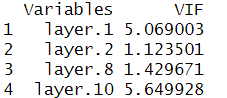
#### Environmental data

Variables included:

* Bio1: annual temp
* Bio2: diurnal range (temp)
* Bio8: temp of wettest quarter
* Bio10: temp of warmest quarter
* Elevation
* Percentage freshwater cover
* Land cover type
* Distance from coast
* Used the *gadm* function in the “geodata” R package (Hijmans et al, 2024) to get shapefiles of Sweden, Norway and Finland. When combined, these were used to define the extents of all future environmental layers
* Downloaded elevation, land cover and freshwater cover for all three countries from the DIVA-GIS version 7.5 database (<https://diva-gis.org/data.html>) at 30-second resolution
  + Calculated each cell’s percentage of freshwater cover using the *rasterize* function in the “terra” package in R (Hijmans, 2024)
  + Combined classes of landcover type for clarity and simplicity

Original landcover types taken from [this pdf](https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fforobs.jrc.ec.europa.eu%2Fstatic%2Fglc2000%2Flegend%2FGLC2000_legend_summary.doc&wdOrigin=BROWSELINK)

|  |  |  |  |
| --- | --- | --- | --- |
| Original Code | Original Habitat Type | New Habitat Type | New Code |
| 1 | Tree Cover, broadleaved, evergreen | Tree Cover | 1 |
| 2 | Tree Cover, broadleaved, deciduous, closed | Tree Cover | 1 |
| 3 | Tree Cover, broadleaved, deciduous, open | Tree Cover | 1 |
| 4 | Tree Cover, needle-leaved, evergreen | Tree Cover | 1 |
| 5 | Tree Cover, needle-leaved, deciduous | Tree Cover | 1 |
| 6 | Tree Cover, mixed leaf type | Tree Cover | 1 |
| 7 | Tree Cover, regularly flooded, fresh water | Tree Cover Flooded | 2 |
| 8 | Tree Cover, regularly flooded, saline water, | Tree Cover Flooded | 2 |
| 9 | Mosaic: Tree cover / Other natural vegetation | Mosaic | 3 |
| 10 | Tree Cover, burnt | Tree Cover | 1 |
| 11 | Shrub Cover, closed-open, evergreen | Shrubland | 4 |
| 12 | Shrub Cover, closed-open, deciduous | Shrubland | 4 |
| 13 | Herbaceous Cover, closed-open | Shrubland | 4 |
| 14 | Sparse Herbaceous or sparse Shrub Cover | Shrubland | 4 |
| 15 | Regularly flooded Shrub and/or Herbaceous Cover | Shrubland Flooded | 5 |
| 16 | Cultivated and managed areas | Cultivated and managed areas | 6 |
| 17 | Mosaic: Cropland / Tree Cover / Other natural vegetation | Mosaic | 3 |
| 18 | Mosaic: Cropland / Shrub or Grass Cover | Mosaic | 3 |
| 19 | Bare Areas | Bare Areas | 7 |
| 20 | Water Bodies | Water Bodies | 8 |
| 21 | Snow and Ice | Snow and Ice | 9 |
| 22 | Artificial surfaces and associated areas | Artificial surfaces and associated areas | 10 |

* Calculated each cell’s distance to the coast using terra package’s *distance* function (Hijmans, 2024)
* CHELSA data (Karger et al, 2017; Karger et al, 2018)
  + Resolution: 30 arc sec, ~1km
  + NB: all temperatures are recorded in K/10
  + Calculated the required “bioclim equivalent” parameters using the *biovars* function in the “dismo” R package (Hijmans et al, 2023)
  + Variables
    - Bio1: annual temp
    - Bio2: diurnal range (temp)
    - Bio8: temp of wettest quarter
    - Bio10: temp of warmest quarter
  + Testing for collinearity using the *vif* and *vifstep* functions from the “usdm” package in R (Naimi et al 2014)
    - Taking all 19 variables into account, bioclim variables 1 (annual temp), 2 (diurnal range of temperature), 4, 5, 11, 12, 13, 16 would have been excluded
    - Choosing those variables thought to be most relevant to damselflies, (1,2,8,10) and testing that subset for collinearity, there were no issues
    - 
    - A threshold value of 10 would imply collinearity problems
  + This covers the years 2000-2018!
* ERA-5 dataset for 2019-2023 (C3S, 2017; Hersbach et al 2020)
  + Downloaded the data from [ERA5-Land monthly averaged data from 1950 to present](https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land-monthly-means?tab=download) using the following parameters
    - 2m Temperature
      * Monthly averaged reanalysis by hour of day
      * Years: 2019-2023
      * Months: January-December
      * 0:00-23:00
      * Sub-region extraction: x (4,32), y(54, 71)
      * NB: all temperatures are recorded in K!
    - Precipitation
      * Monthly averaged reanalysis
      * Years: 2019-2023
      * Months: January-December
      * Hours: 0:00
      * Sub-region extraction: x (4,32), y(54, 71)
  + To calculate bioclim variables using the dismo package, I extracted tmin and tmax from the monthly averaged reanalysis by hour of day of the temperature dataset
  + Using the “biovars” function in the dismo package, I calculated all 19 bioclim variables and then extracted Bio1, Bio2, Bio8, and Bio 10
  + I disaggregated this dataset using the *resample* function in the “terra” package to match the resolution of the CHELSA data and reprojected the Era5 data to match projections
* Created a raster stack of each year’s environmental variables, 2000-2023
  + Bio1: annual temp
  + Bio2: diurnal range (temp)
  + Bio8: temp of wettest quarter
  + Bio10: temp of warmest quarter
  + Elevation
  + Percentage freshwater cover
  + Land cover type
  + Distance from coast

#### Presence data

* Pulled records from local repositories
  + Norway
    - <https://artskart.artsdatabanken.no/>
    - <https://www.artsobservasjoner.no/>
    - Species map.artsdatabanken.no 9.9.2022. Find data from: Bergen museum, UiB, Natural History Museum UiO, Tromsø museum UiT, NTNU Science Museum. Download from Artskart.
  + Sweden
    - <https://www.artportalen.se/>
    - <https://artportalen.se/ViewSighting/SharedSearch?storedSearchId=12449&identifier=88EB82C6>
    - 10/31/2022
    - SLU Swedish Species Information Centre (2022). Artportalen: Detailed find information. https://artportalen.se/ViewSighting/SharedSearch?storedSearchId=12449&identifier=88EB82C6 [2022-10-31]
  + Finland:
    - <https://laji.fi/en>
    - Finnish Biodiversity Information Facility/FinBIF. http://tun.fi/HBF.66820 (accessed 9.9.2022).
* Cleaned them
  + Data were filtered for complete records and coordinate accuracy < 1000m as well as obvious geographical outliers outside of the known range and specific repository’s country boundaries.

#### Matching presence points with the environmental information and creating background points

* Using the *extract* function of the “terra” package (Hijmans, 2024), created a dataset matching each presence point with the environmental values for the year in which it was observed
* Using the *background* function in the “terra” package, randomly sampled 100 background points for each year to create a pseudo-absence dataset
  + This resulted in 2400 background points
* Dealing with NAs
  + Deleting all coordinates with some missing environmental variables would have discarded 2/3 of the dataset. Additionally, most of these coordinates were located along the coast where coordinates might be in estuarine or shallow environments and environmental data might be missing
  + Therefore, I took the coordinates with missing data and imputed the information using the *preProcess* function from the “caret” package (Kuhn, 2008)
  + This preserved ~8000 observations
  + NB: in the case of Land cover, where values are categorical, this has taken means and created new values. We could round up or down, but that doesn't mean much so I have just left it. There are only 66 of them so won't skew the model too much.
  + I still removed any rows with NaNs or NAs remaining

#### Addressing spatial autocorrelation

* The first step is to remove the duplicates from the presence dataset. The resolution is ~1km x 1km, so only one presence point per cell is retained
  + This resulted in 3914 presence points across 2000-2022
* Tested for spatial autocorrelation using the *spatiotemp\_autocorr* function in the “spThin” package. There were no spatial correlation issues.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Observed | Expected | sd | P value |
| Annual Temperature | -0.1594322 | -0.0002555584 | 0.000170611 | 0 |
| Diurnal range | -0.0293214 | -0.0002555584 | 0.0001706231 | 0 |
| Temp of wettest quarter | -0.01009252 | -0.0002555584 | 0.000170629 | 0 |
| Temp of warmest quarter | -0.02989776 | -0.0002555584 | 0.0001706002 | 0 |
| Elevation | -0.01180911 | -0.0002555584 | 0.0001705965 | 0 |
| Landcover | -0.003217194 | -0.0002555584 | 0.0001706406 | 1.777229e-67 |
| Percentage water cover | -0.004101675 | -0.0002555584 | 0.0001706123 | 1.572157e-112 |
| Distance to coast | -0.04831741 | -0.0002555584 | 0.0001705428 | 0 |

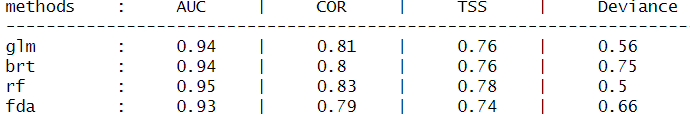
#### Separating for training and testing purposes

* Combine the presence and pseudo-absence points into one dataset
* To ensure a clean test dataset, I separated this dataset into a 70:30 split, retaining 30% of all points for testing purposes

## Running the SDM

Using the “sdm” package (Naimi et al 2016), I built an ensemble SDM, combining the following methods: generalised linear model, boosted regression tree, random forest and Flexible Discriminant Analysis.

This model was trained using 70% of the data, extracted with environmental variables across all years of presence (2000-2022)

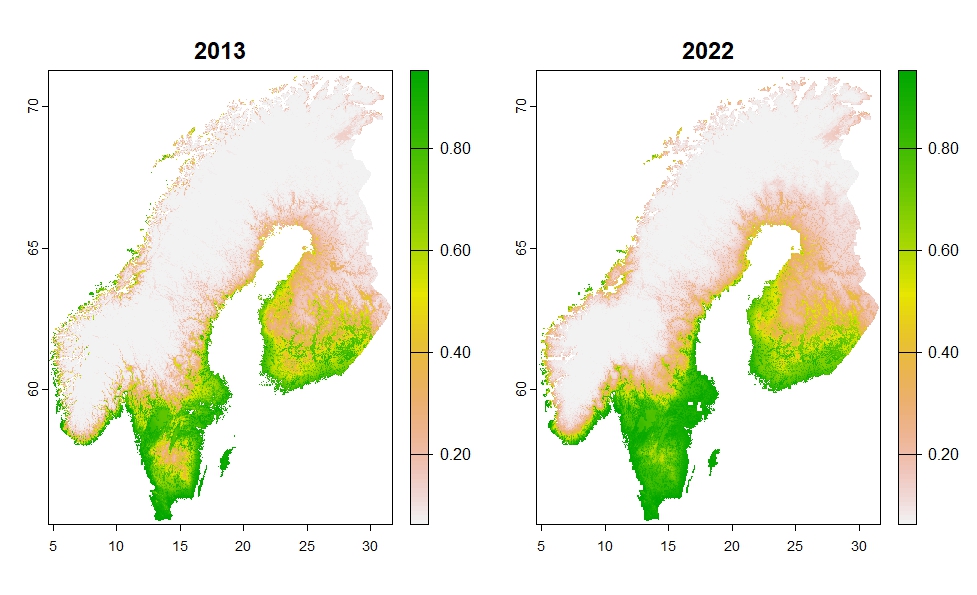


Training the model on the entire presence record and environmental range ensures we capture the full niche of *Ischnura,* allowing it to follow the environmental shift northwards over time. If we had restricted our training to, for example, only 2013 environment, we run the risk of marking certain conditions as unfavourable just because it wasn’t observed in that year, when in reality *Ischnura* could happily inhabit that environmental range.

### Creating regional scale predictions

This model can then be used to predict probability of presence across the regional scale (Sweden, Finland and Norway) for each year, using the *predict* function in the “sdm” package (Naimi et al 2016).

Comparing 2013 vs 2022

Figure 1: Predicted habitat suitability for the entire region in 2013 vs 2022

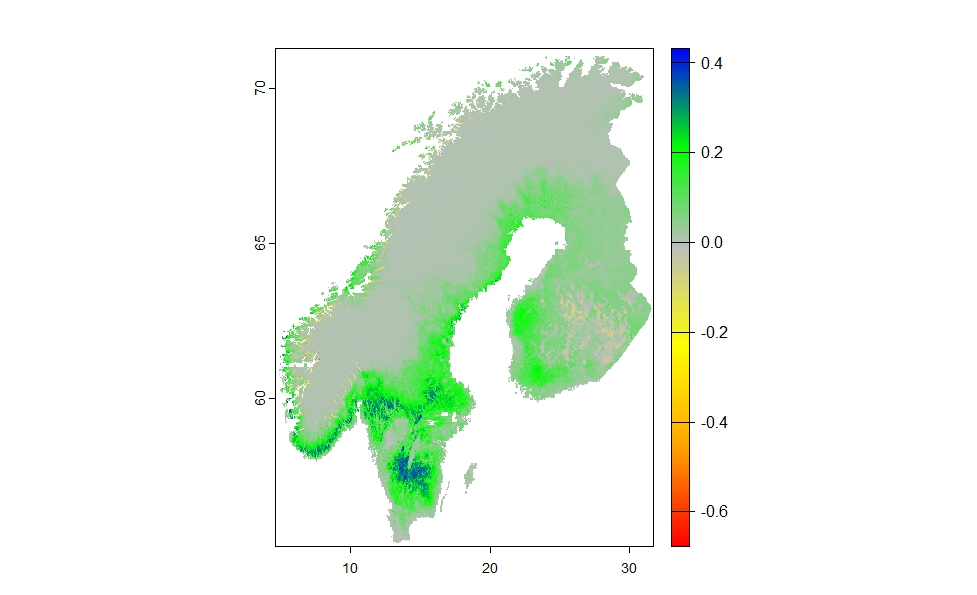


Figure 2: The difference in habitat suitability between 2022 and 2013. Green to blue values show positive habitat suitability change, while yellow to red values show negative change. Most of the region has experienced positive habitat suitability change for Ischnura

### Predicting across Sweden + Finland

* By subsetting the regional-scale environmental and land layers, we can apply the model across various regions including Sweden + Finland, and just Sweden on its own

#### Country-specific water information

* We have access to highly detailed data on Sweden’s freshwater bodies via Dr. Keith W. Larson, Director of the Arctic Centre at Umeå University, and Maren Wellenreuther.
* These consisted of 263 .gml files, each containing detailed data of a small “tile” of Sweden and the freshwater bodies it contains
* Using the *st\_read* function from the “sf” package (Pebesma, 2018) and the *rasterize* function from the “terra” package, I combined these 263 files into one raster and calculated the percentage water cover each cell contained

#### Applying the regional model

* While the regional model was in standard WGS84 projection, Sweden’s freshwater data is in EPSG:3006 (also referred to as SWEREF99 TM), the projected coordinate system for Sweden. For the initial sdm prediction, this new water layer needs to be projected to WGS84, using terra package’s *project* function
* I then extracted all layers to Sweden’s outline to create the environmental layers for predicting the regional sdm model onto Sweden
* I also extracted all layers to Sweden + Finland’s outline to do the same

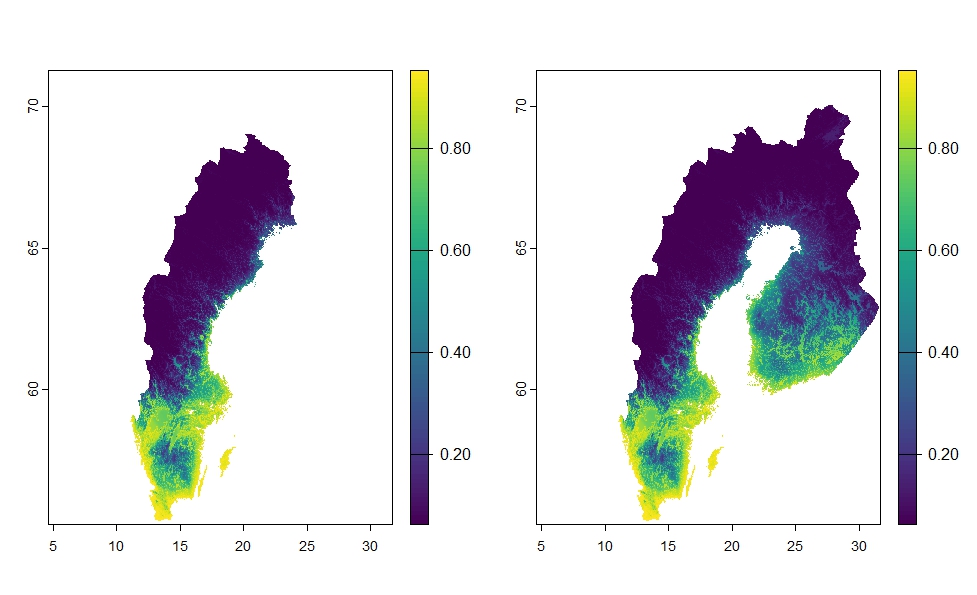


Figure 3: Predicted habitat suitability for Sweden and Sweden+Finland in 2013

* The regional model can now be used to predict habitat suitability for each year from 2013-2023

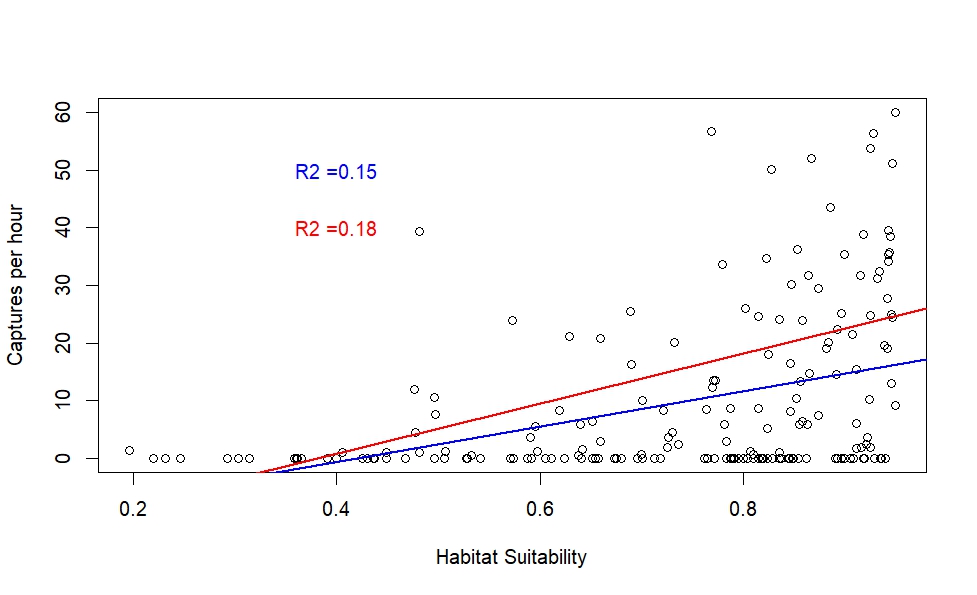
## Preparing RangeShifter input files

### Projection and resolution

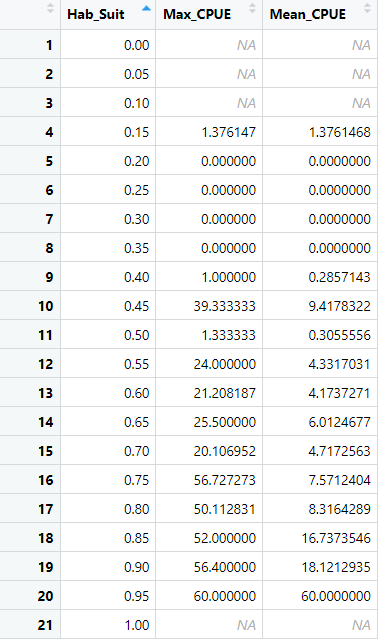
* Since these input layers are intended for use with RangeShifter, which requires the resolution to be in metres, all layers have to be projected back into the EPSG:3006 projection.
* The resolution can now be adjusted to be exactly 1km x 1km using the *resample* function in the “terra” package.

### Converting habitat suitability to “habitat classes”

* Empirical data collected on transect surveys in 2013, 2022 and 2023 provide catch per unit effort (CPUE) estimates of densities. In this case, the number of damselflies captured per hour of survey effort.
* I then extracted the predicted habitat suitability of the lat/long location of the survey transect to test the relationship between CPUE and predicted habitat suitability. This relationship was significant, with p<0.01 both when surveys with 0 captures were included and excluded. The Adjusted R2 ranged from 0.15 to 0.18, respectively, so does not explain most of the variability of this relationship.

Figure 4: linear relationships between habitat suitability and CPUE (captures per hour), when surveys with 0 captures are included (Blue) and excluded (Red). Adjusted R2 values are presented for both models.

* I then extracted the CPUE values associated with “bins” of habitat suitability and calculated the maximum and mean CPUE within those bins. The resulting data looked like this:



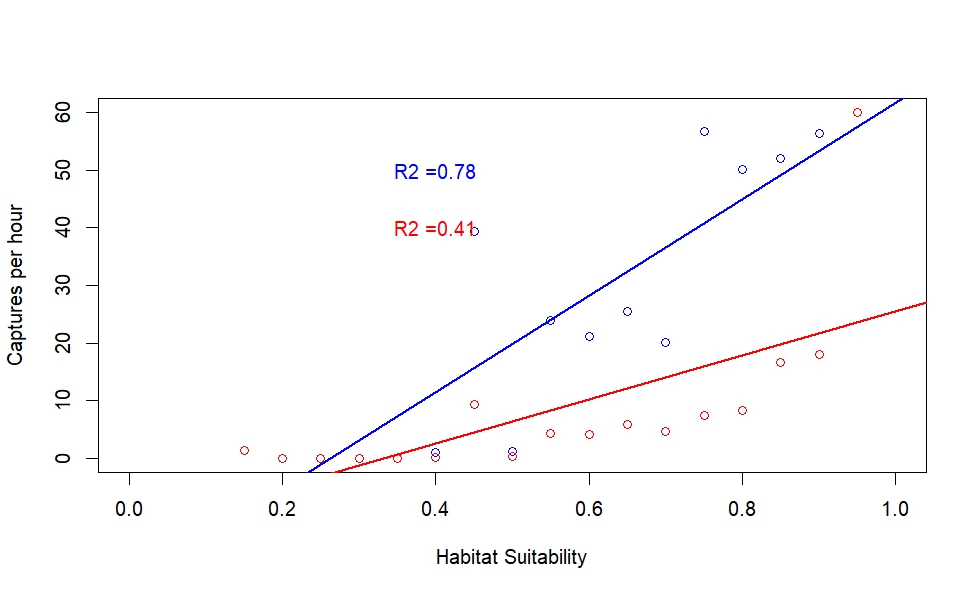
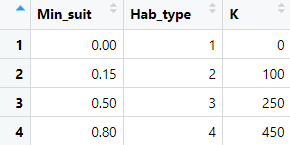
* I then performed the linear models again on these data which resulted in these relationships: 

Figure 5: linear relationships between habitat suitability and CPUE (captures per hour), for maximum (Blue) and mean (Red) values of CPUE. Adjusted R2 values are presented for both models.

* These relationships were significant with p<0.005 and adjusted R2 values of 0.78 and 0.41 for maximum and mean CPUE, respectively.
* The information that is missing here is how CPUE scales with absolute density, so we don’t know what it means in absolute terms to have a CPUE=60, for example. What we can glean from these data is:
  + The lowest habitat suitability that was non-zero is 0.15, so we can assume that any cells with values <0.15 will be completely unsuitable
  + Very low CPUEs below habitat suitability 0.5, so while they are possible, cells with habitat suitability 0.15-0.5 have a low carrying capacity
  + The next jump is between 0.7 and 0.8 habitat suitability, which could simply be due to missing data at that range. But we can assume that habitat suitability >0.8 is high quality and has a high carrying capacity
  + Between 0.5-0.8, there is little change in maximum or mean CPUE, so we can assume there is no difference in carrying capacity here
  + Therefore, going forward, the model will have 4 habitat types: unsuitable (<0.15), low carrying capacity (0.15-0.5), medium carrying capacity (0.5-0.8) and high carrying capacity (>0.8)
  + **NB: this is potentially an area to play with!** Breaking these into finer groups if dispersal is not facilitated enough. However, at this stage, I didn’t want to introduce more assumptions than necessary, given the lack of further data
* Further assumptions were made for carrying capacity values:
* 
* **NB: these can also be played with easily!**
* Given these specifications, the habitat layers for Sweden and Sweden\_Finland for 2013 and 2023look like this:

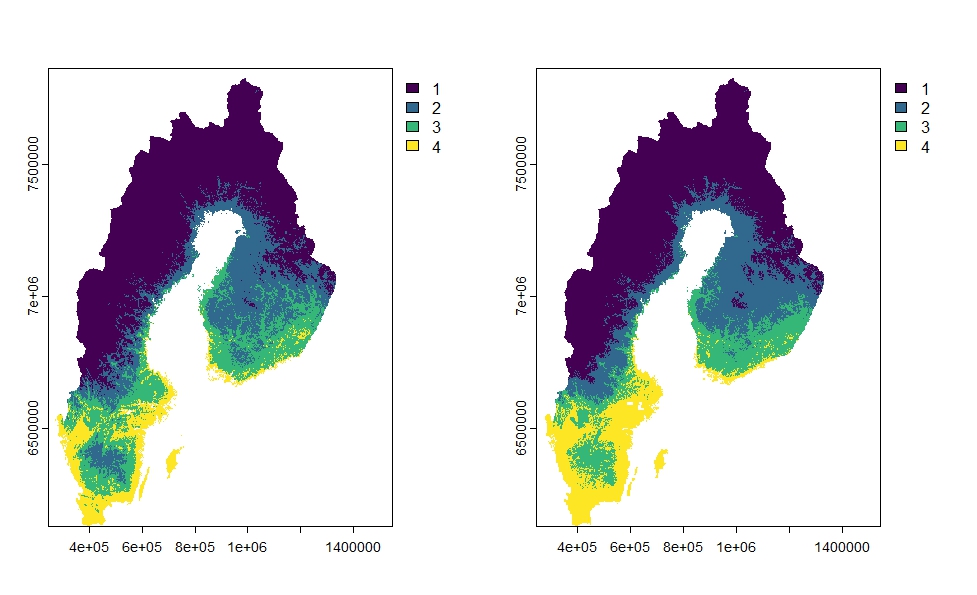


Figure 6: habitat types for 2013 (left) and 2023 (right)

### Creating initial species distribution

* Using presence records up to 2013 to get starting distribution, I projected the latitude/longitude coordinates into the Sweden-specific projection EPSG:3006 so coordinates are now in meters
* Using the *extract* function in the “terra” package, I extracted the cell numbers of the presence points and created a presence/absence raster of 0s and 1s

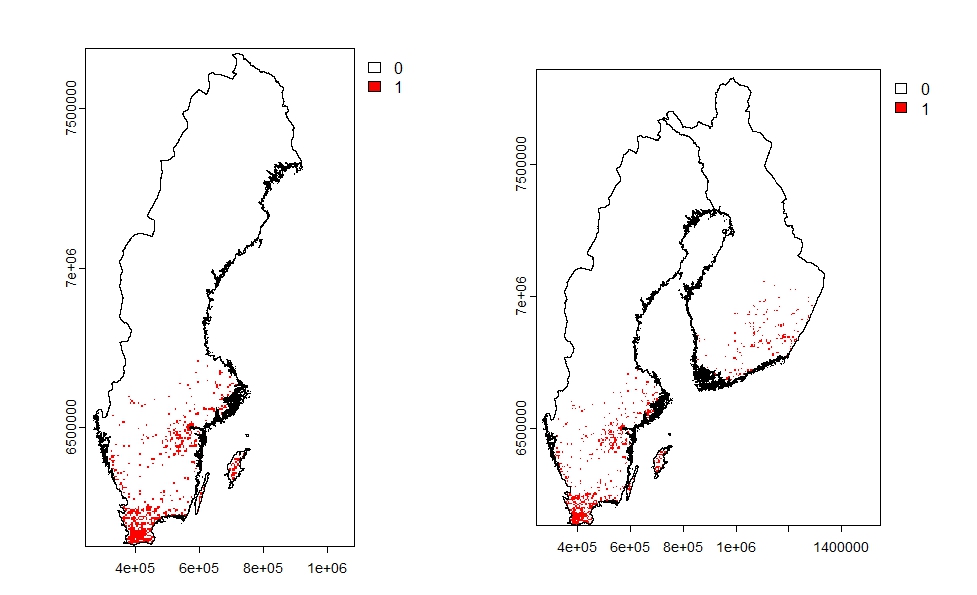


Figure 7: initial distribution of Ischnura elegans for baseline in 2013

## RangeShifter parameters

### Cropping to the range front

* For computational reasons, I chopped off the southern extent of the range, since this foundational run is focusing on the range front and not interrogating genetics yet

### Setting demographic and dispersal parameters

* Demography
  + Non-overlapping generations because larvae stay in the pond for 1 year, adults live only a few weeks, disperse, mate, lay eggs. So, everything happens within a year, therefore unstructured model
  + Rmax: 2.2
  + Proportion of males: 0.5
* Dispersal
  + Emigration probability: 0.8 with a double negative exponential dispersal kernel with 90% of individuals dispersing 1500m and 10% dispersing up to 10km

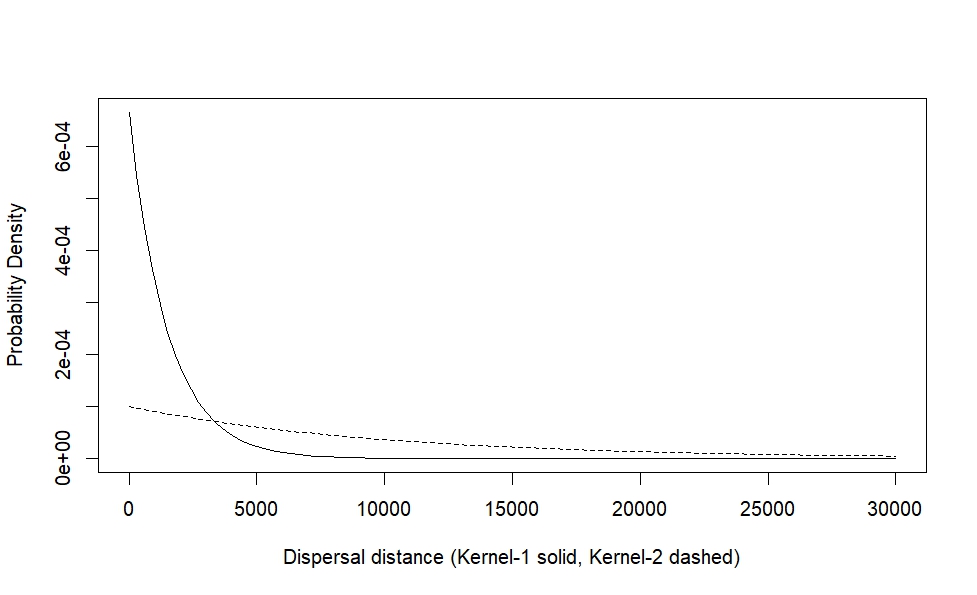


Figure 8: double negative exponential dispersal kernel, 90% 1500m, 10% 10km

# Results

## 2013 static landscape

* To test range shifting under static environmental conditions, I ran a RangeShiftR (Malchow et al 2021) simulation using 2013 habitat type maps and kept them constant throughout the model run

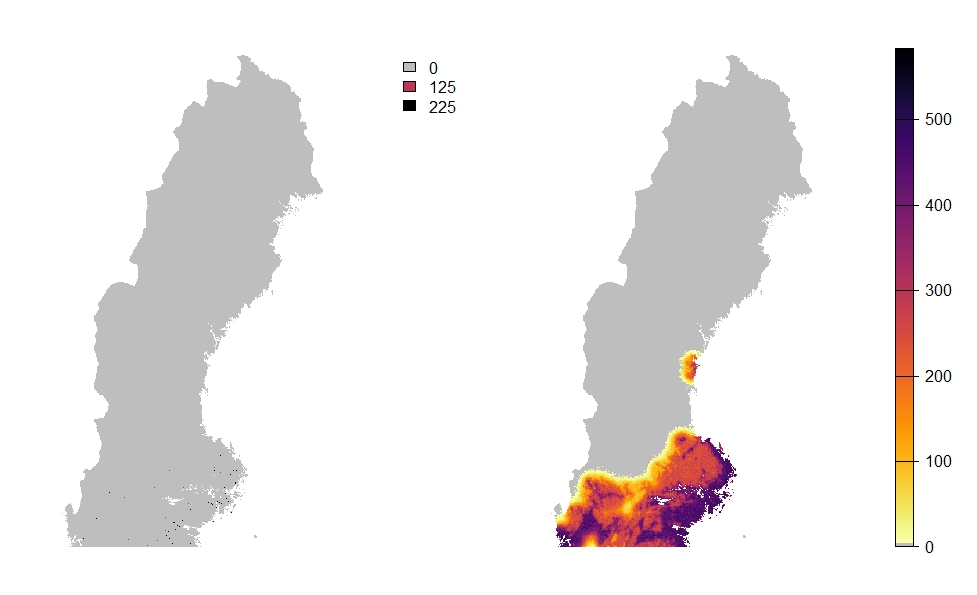
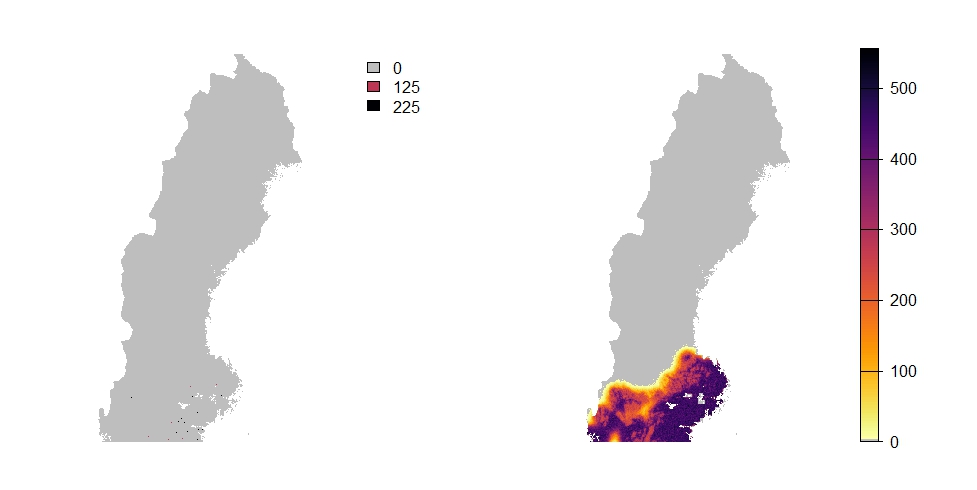


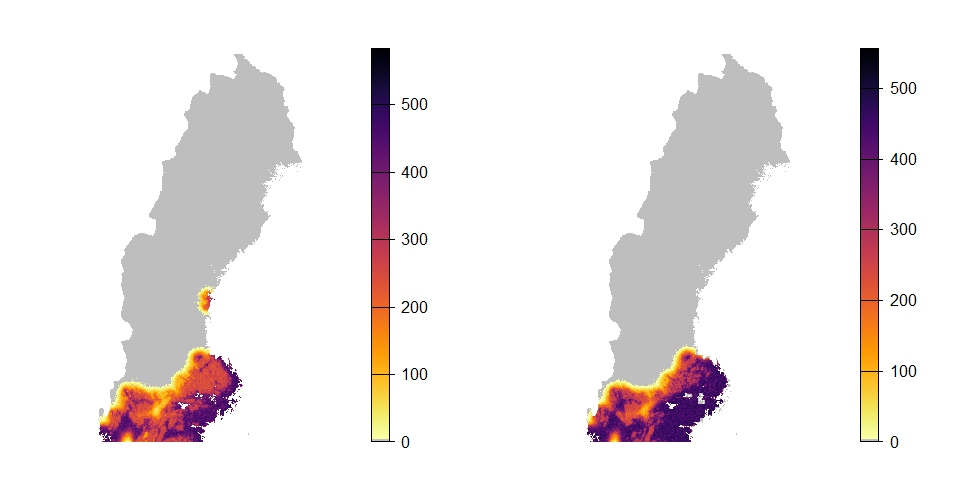
Figure 9: Population growth and range expansion over a 20 year simulation, starting from 2013, using static environmental conditions from 2013. Cooler colours indicate higher population densities

## Dynamic landscapes

* It’s important to make the habitat maps match in terms of NAs, otherwise RangeShiftR can’t work with them. Because we have combined environmental data from two different sources (CHELSA and era5), our layers differ in their NAs. I have used the *mask* function in the “terra” package to “fill in” some of the NAs so that at least they are in the same locations. This makes very little visible difference to the habitat maps but means the simulations run smoothly

## Compare static and dynamic

* After 20 years:



# Notes going forward

* Need to interrogate where that northern cluster has gone
  + This might be due to how I got rid of some NAs in the 2019-2023 data (see above) so perhaps that got rid of cells with species presence?
* We are not seeing the northwards spread that we expect, so a few things can be played with:
  + Some kind of burn in period using data from before 2013?
  + We simply don’t have the numbers, perhaps, so running the sims on the cluster with higher numbers might see more of the long distance dispersal events making a difference
* The pattern just looks like they’re spreading in absolutely every cell, which I'm not sure is realistic
  + Could look into editing the habitat maps slightly to make any cell that doesn’t have water in it unsuitable even if the SDM predicts some kind of suitability there. That might give a bit less of a “carpeted” feel to the expansion. However, this might overlook some cells with small ponds etc that are not in our datasets
* Future edits to be made to parameters
  + Emig prob: 0.1. the simulations right now have Emigration rate 0.8 but I think we talked about reducing it to 0.1
  + Make dispersal sex-dependent, make males more likely to disperse
  + Ref: rachel’s paper based on genetics
  + Calibrate with chance of leaving 1kmx1km cell
  + Future sims: Emig prob as evolving trait? To capture behaviour at range edge
* A note on superindividuals: I’m not sure how to represent them when looking at genetic structure, this might require some tweaking!
* Obviously, the plan was to also run this for the map covering Sweden and Finland to see if individuals are coming up and over from Finland (with the genetic structure from empirical data in mind!) so those simulations would need to be set up as well, most likely on the cluster to accommodate higher memory requirements

# References

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS), **date of access june 2 2025**. <https://cds.climate.copernicus.eu/cdsapp#!/home>

Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; 146: 1999–2049. <https://doi.org/10.1002/qj.3803>

Hijmans RJ, Phillips S, Leathwick J, Elith J (2023). \_dismo: Species Distribution Modeling\_. R package version 1.3-14, <https://CRAN.R-project.org/package=dismo>.

Hijmans R (2024). \_terra: Spatial Data Analysis\_. R package version 1.7-78, <https://CRAN.R-project.org/package=terra>.

Hijmans RJ, Barbosa M, Ghosh A, Mandel A (2024). \_geodata: Download Geographic Data\_. R package version 0.6-2, <https://CRAN.R-project.org/package=geodata>.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, P., Kessler, M. (2017): Climatologies at high resolution for the Earth land surface areas. Scientific Data. 4 170122. <https://doi.org/10.1038/sdata.2017.122>

Karger D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E, Linder, H.P., Kessler, M. (2018): Data from: Climatologies at high resolution for the earth’s land surface areas. EnviDat. <https://doi.org/10.16904/envidat.228.v2.1>

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>

Malchow A, Bocedi G, Palmer SCF, Travis JM, Zurell D (2021). “RangeShiftR: an R package for individual-based simulation of spatial eco-evolutionary dynamics and species' responses to environmental change.” Ecography. doi:10.1111/ecog.05689 <https://doi.org/10.1111/ecog.05689>.

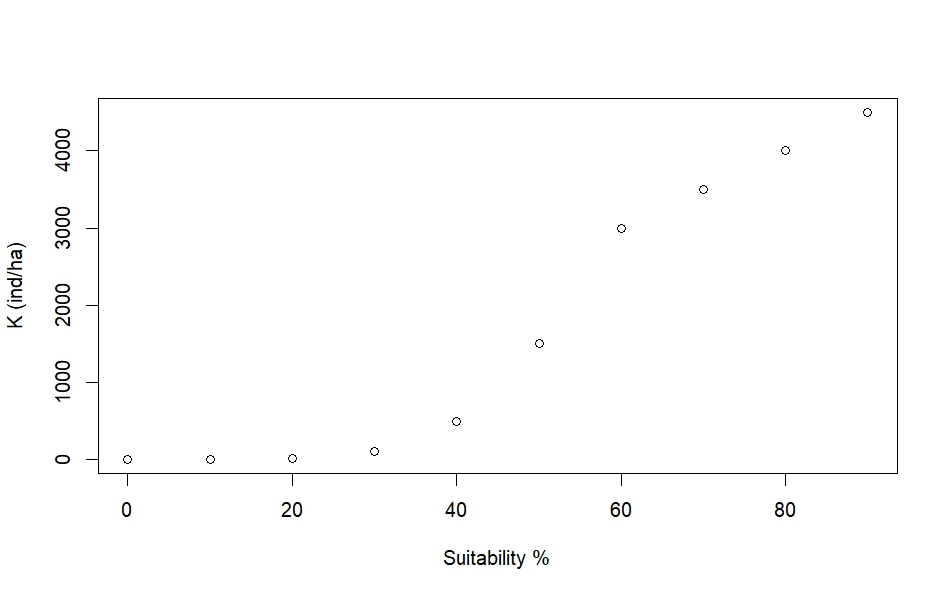
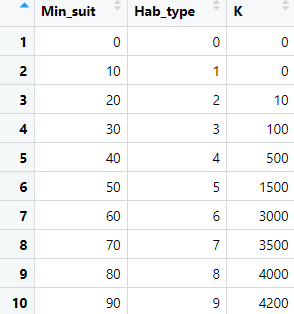
Naimi B, Hamm Na, Groen TA, Skidmore AK, Toxopeus AG (2014). “Where is positional uncertainty a problem for species distribution modelling.” Ecography, \*37\*, 191-203.  
 doi:10.1111/j.1600-0587.2013.00205.x <https://doi.org/10.1111/j.1600-0587.2013.00205.x>.

Naimi B, Araujo MB (2016). “sdm: a reproducible and extensible R platform for species distribution modelling.” \_Ecography\_, \*39\*, 368-375. doi:10.1111/ecog.01881  
 <https://doi.org/10.1111/ecog.01881>.

Pebesma, E., & Bivand, R. (2023). Spatial Data Science: With Applications in R. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>   
  
 Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>

# Supplementary

## Alternative method for habitat type conversion

Figure 4: Relationship between habitat suitability and carrying capacity

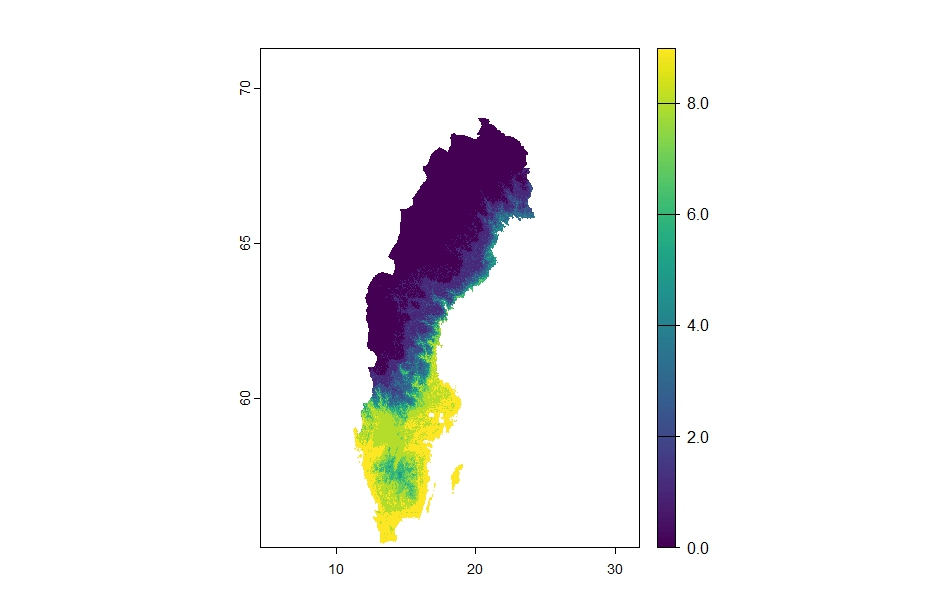


Figure 5: Discrete habitat types associated with carrying capacity