Appendix III

Ecological memory patterns of virtual taxa: analyses and results

Blas M. Benito

Contents

Summary

This appendix presents the code required to run the analyses and to generate the figures presented in the paper. The aim of this appendix is to ensure the reproducibility of our work. To reach a full understanding of what the code in this appendix is doing requires to at least read the paper, and go through the contents of **Appendix II**. Please bear in mind that the code shown in this appendix can take hours to execute.

IMPORTANT: An Rmarkdown version of this document can be found at: https://github.com/BlasBenito/EcologicalMemory.

1 Cuantifying ecological memory patterns for the simulated taxa and dataset types

We evaluated ecological memory on 16 virtual taxa and 5 dataset types ("Annual", "1cm", "2cm", "6cm", and "10cm") by using separatedly as exogenous component the values of the *driver* and the *suitability* values returned by the niche functions of the taxa. To compute ecological memory patterns from the simulated data, these steps are required:

- Run Random Forest models, as explained in **Appendix II** on each combination of taxa and dataset type by using the *runExperiment* function.
- Plot the outcome of *runExperiment* to examinate results visually.
- Organize the results into a single table with *experimentToTable*.

After these steps, results are ready for further analyses.

1.1 Running Random Forest models

General parameters for the simulation.

```
#dataset types and taxa to be analyzed
selected.dataset.types = 1:5
selected.taxa = 1:16

#number of Random Forest models to fit
repetitions = 300

#names of these columns
sampling.names = c("Annual",
                   "1cm",
                   "2cm",
                   "6cm",
                   "10cm")

#response variable
response.column = "Pollen"

#time column
time.column = "Time"
```

```
#lags
lags = seq(20, 240, by=20)

#species traits to include in the output
traits <- c("maximum.age",
            "fecundity",
            "niche.A.mean",
            "niche.A.sd")
```

Using **suitability** as exogenous component.

```
E1.suitability <- runExperiment(
  simulations.file = simulation,
  selected.rows = selected.taxa,
  selected.columns = selected.dataset.types,
  parameters.file = parameters,
  parameters.names = traits,
  sampling.names = sampling.names,
  driver.column = "Suitability",
  response.column = response.column,
  subset.response = "none",
  time.column = time.column,
  time.zoom = NULL,
  lags = lags,
  repetitions = repetitions)
```

Using **driver** as exogenous component.

```
E1.driver <- runExperiment(
  simulations.file = simulation,
  selected.rows = selected.taxa,
  selected.columns = selected.dataset.types,
  parameters.file = parameters,
  parameters.names = traits,
  sampling.names = sampling.names,
  driver.column = "Driver.A", #only difference
  response.column = response.column,
  subset.response = "none",
  time.column = time.column,
  time.zoom = NULL,
  lags = lags,
```

```
  repetitions = repetitions)
```

1.2 Plotting ecological memory patterns

Plotting the **suitability** results.

```
plotExperiment(
  experiment.output=E1.suitability,
  experiment.title="Ecological memory patterns with
  suitability as exogenous component.",
  legend.position="bottom",
  R2=TRUE,
  sampling.names=sampling.names,
  strip.text.size=7,
  axis.x.text.size=7,
  axis.y.text.size=12,
  axis.x.title.size=14,
  axis.y.title.size=14,
  title.size=18,
  filename = "E1_suitability")
```

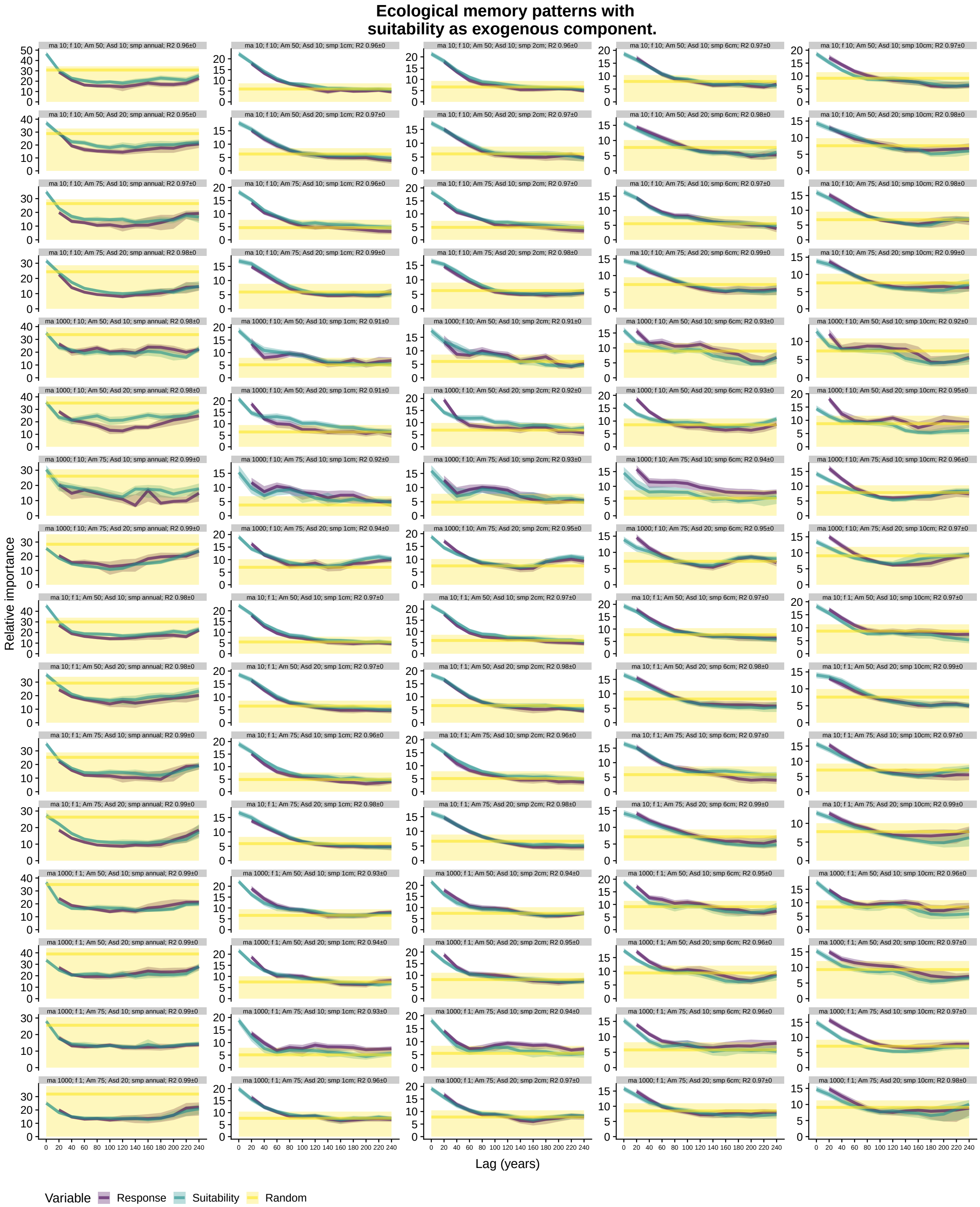

Figure 1: Ecological memory patterns of all virtual species (rows) and dataset types (columns) when using suitability (blue) as exogenous component (blue curves).

Plotting the **driver** results.

```
plotExperiment(
  experiment.output=E1.driver,
  experiment.title="Ecological memory patterns with
  driver as exogenous component.",
  legend.position="bottom",
  R2=TRUE,
  sampling.names=sampling.names,
  strip.text.size=7,
  axis.x.text.size=7,
  axis.y.text.size=12,
  axis.x.title.size=14,
  axis.y.title.size=14,
  title.size=18,
  filename = "E1_driver.pdf")
```

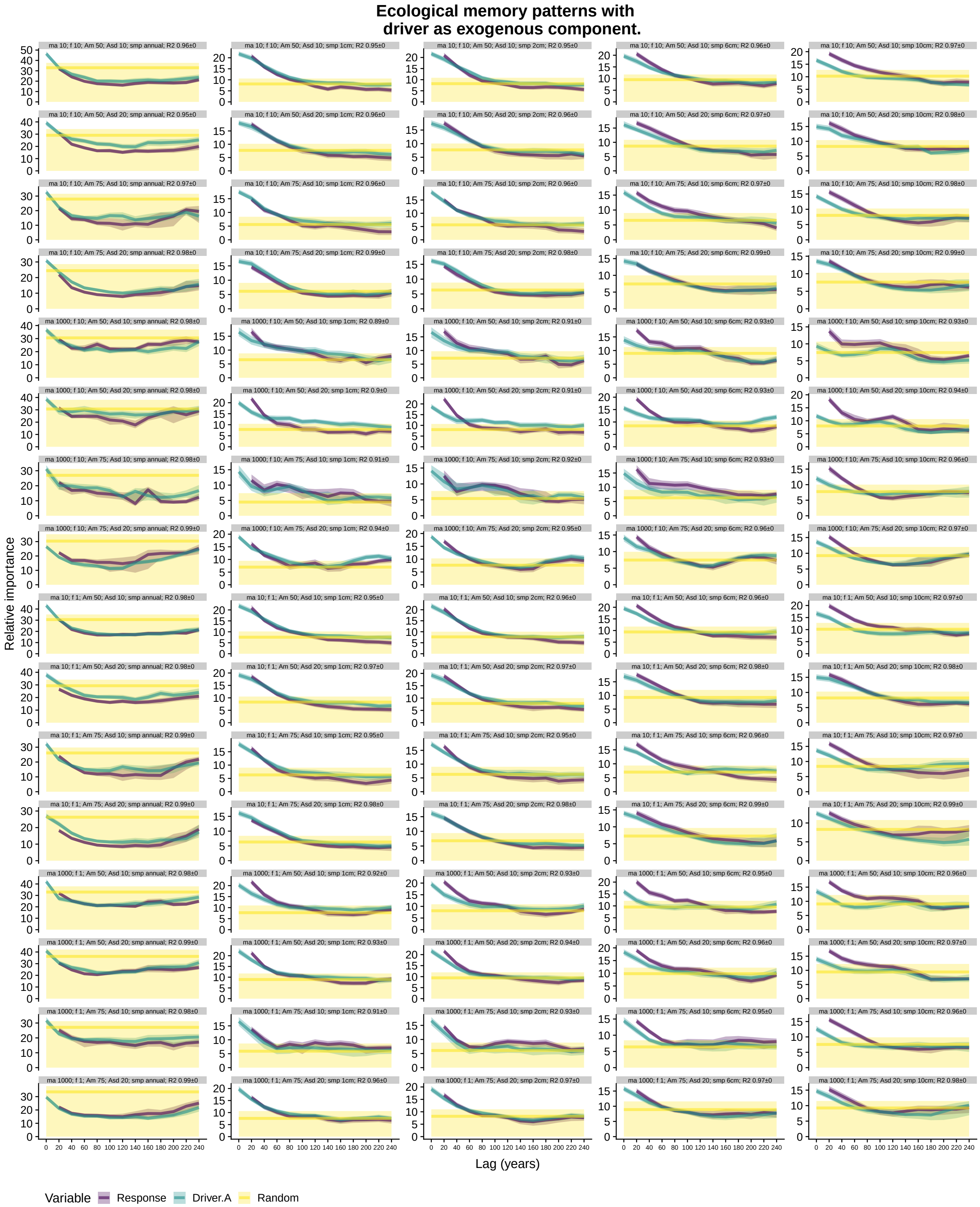

Figure 2: Ecological memory patterns of all virtual species (rows) and dataset types (columns) when using driver (blue) as exogenous component (blue curves).

1.3 Organizing results into dataframes to facilitate further analyses

When applying *experimentToTable* to **E1.suitability** or **E1.driver** the resulting columns are (see **Table 1**):

- **median**: median importance of the given **Variable** at a given **Lag**.
- **sd**: standard deviation of the importance.
- **min**: 0.05 percentile of the variable importance across repetitions.
- **max**: 0.95 percentile of the variable importance.
- **Variable**: name of the given variable (response, suitability, driver, random).
- **Lag**: lag ID.
- **R2mean**: mean pseudo R-squared of Random Forest models across repetitions.
- **R2sd**: standard deviation of pseudo R-squared.
- **VIFmean**: average variance inflation factor (VIF) of the predictors for the given dataset.
- **VIFsd**: standard deviation of VIF values.
- **label**: name of the virtual taxa as shown in the parameters dataframe.
- **maximum.age**: maximum life-span of the virtual taxa.
- **reproductive.age**: reproductive age.
- **fecundity**: maximum fecundity under ideal conditions.
- **growth.rate**: growth rate.
- **maximum.biomass**: maximum biomass of the individuals (100).
- **carrying.capacity**: carrying capacity of the landscape (10000).
- **niche.A.mean**: mean of the niche function. Also, *niche position*.
- **niche.A.sd**: standard deviation of the niche function. Also, *niche breadth*.
- **sampling**: one of “Annual”, “1cm”, “2cm”, “6cm”, “10cm”

The code below generates the tables for the **suitability** and **driver** data. A sample output of *experimentToTable* is shown below.

```
E1.suitability.df <- experimentToTable(
  experiment.output=E1.suitability,
  parameters.file=parameters,
  sampling.names=sampling.names, R2=TRUE)

E1.driver.df <- experimentToTable(
  experiment.output=E1.driver,
  parameters.file=parameters,
  sampling.names=sampling.names, R2=TRUE)
```

Table 1: First rows of the table E1.suitability.df

median	sd	min	max	Variable	Lag	R2mean	R2sd	VIFmean	VIFsd	label	maximum.age	reproductive.age	fecundity	growth.rate	maximum.biomass	carrying.capacity	niche.A.mean	niche.A.sd	sampling
28.80	0.93	27.12	30.17	Response	20	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
20.92	1.05	19.31	22.71	Response	40	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
16.29	0.91	14.84	17.72	Response	60	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
15.43	0.96	13.86	16.92	Response	80	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
15.34	1.18	13.17	17.13	Response	100	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
14.56	2.04	10.31	17.46	Response	120	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
16.06	1.49	13.38	18.21	Response	140	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
18.16	1.58	15.44	20.56	Response	160	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
16.95	1.65	14.17	19.44	Response	180	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
16.75	1.57	14.25	19.25	Response	200	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
18.03	1.51	15.39	20.28	Response	220	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
22.65	1.60	19.95	25.10	Response	240	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
46.50	1.35	44.44	48.83	Suitability	0	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.30	1.09	28.73	32.33	Suitability	20	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
22.77	1.01	21.15	24.43	Suitability	40	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
20.38	0.97	18.81	22.19	Suitability	60	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
18.74	1.23	16.75	20.68	Suitability	80	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
19.40	0.93	17.74	20.72	Suitability	100	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
18.16	1.31	15.95	20.36	Suitability	120	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
19.89	1.49	17.29	22.38	Suitability	140	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
21.08	1.43	18.99	23.42	Suitability	160	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
22.96	1.63	20.33	25.53	Suitability	180	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
22.06	1.46	19.84	24.50	Suitability	200	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
20.96	1.46	19.04	23.80	Suitability	220	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
25.20	1.76	22.22	27.88	Suitability	240	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	20	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	40	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	60	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	80	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	100	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	120	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	140	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	160	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	180	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
30.77	3.22	0.00	34.34	Random	200	0.96	0.00025	10	3.22863	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual

2 Paper figures and tables

2.1 Figure 1

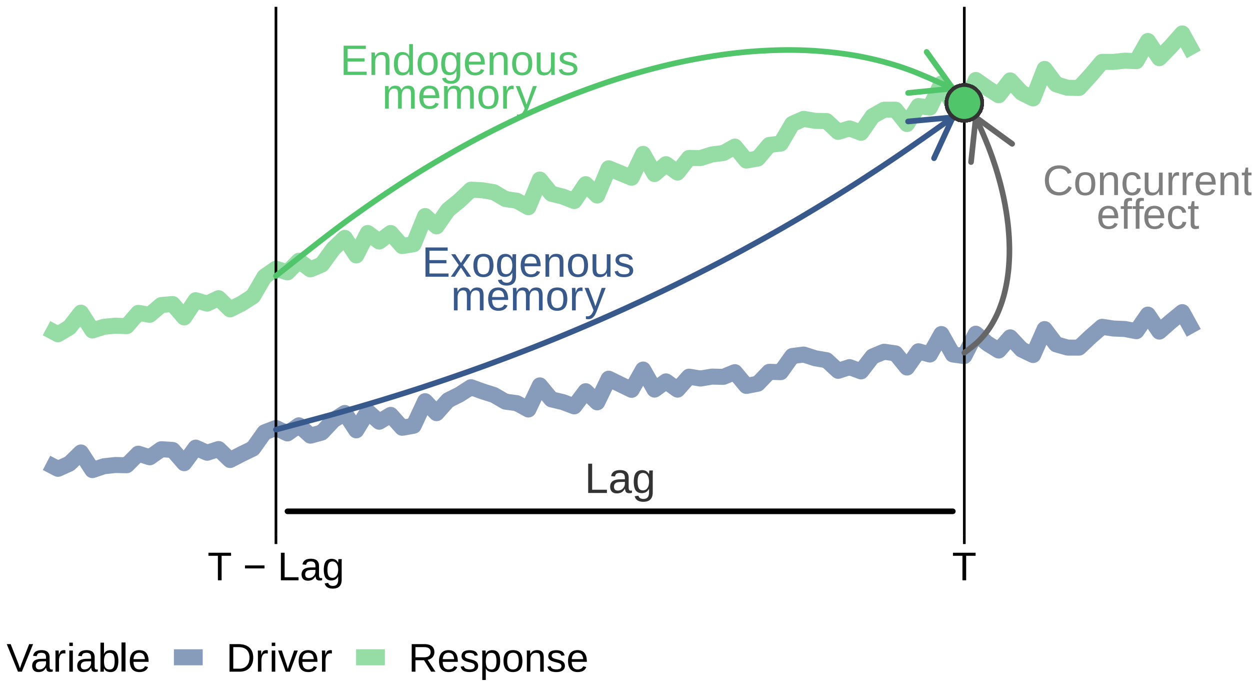

Figure 1: Components of ecological memory. Antecedent values of the driver and the response for a given lag length are located at T – Lag. Arrows represent the relative contribution of endogenous memory, exogenous memory, and the concurrent effect on the value of interest of the response at time T, represented by the green dot. This conceptual structure can be scaled up to any number of drivers and lags.

2.2 Figure 2

This figure was generated with the package *yEd* (URL: www.yworks.com/products/yed).

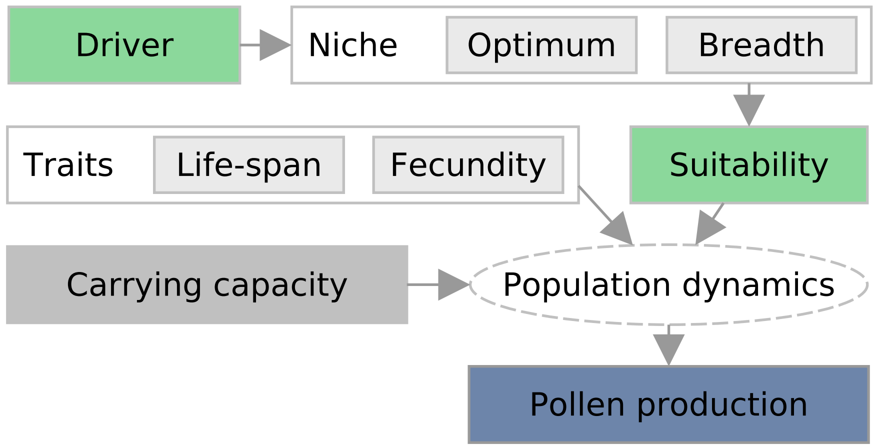

Figure 2: Model structure. Green and white boxes represent respectively the environment and elements relative to the virtual taxa. Light-gray boxes are the experimental variables, changing across model executions. Population dynamics (dashed ellipse) is the emergent process from which pollen production (blue box), the response variable, results.

2.3 Figure 3

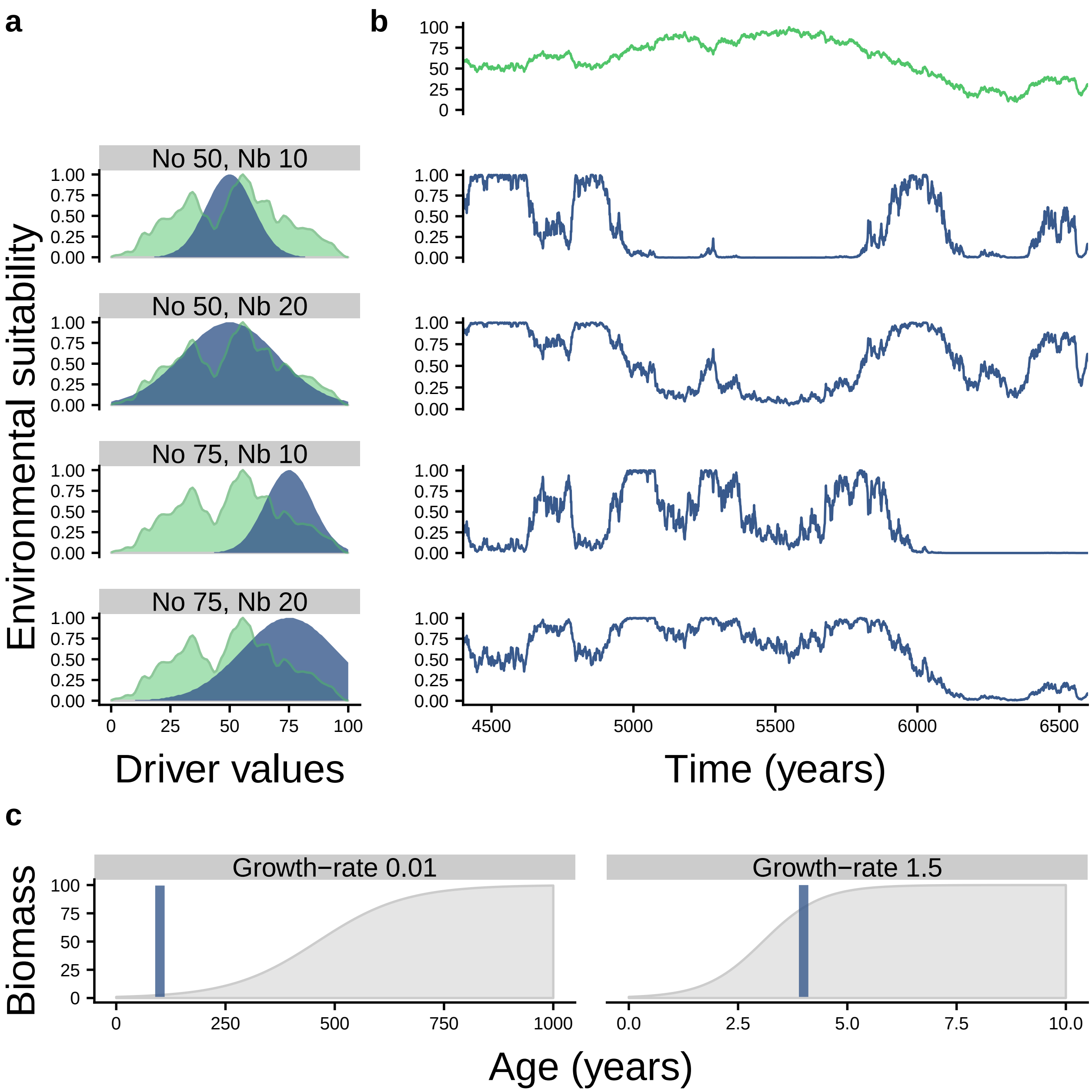

Figure 3: Panel a represents the different ecological niches used in the simulations. Light green colour represents the driver and marine blue represents the niche of the virtual taxa. Numbers in grey strips represent the mean (m) and standard deviation (sd) of the Gaussian functions used to define the niche. Panel b shows the dynamics of the driver during 2000 years, and the suitability values returned by the different niche functions in panel a. Panel c shows the two combinations of life-span used in the simulations, with their respective growth rates, and sexual maturity indicated by a blue vertical line.

2.4 Table 1

Table 1: Parameters of simulated taxa. The unites of Life-span and Reproductive age are years. Fecundity is the maximum number of viable seeds produced by an individual under ideal conditions. Growth rate is the slope of the logistic curve defining biomass growth. Niche optimum and niche breadth are expressed in driver units.

ID	Life-span	Reproductive age	Fecundity	Growth rate	Niche optimum	Niche breadth
1	10	4	1	1.50	50	10
2	10	4	1	1.50	50	20
3	10	4	1	1.50	75	10
4	10	4	1	1.50	75	20
5	10	4	10	1.50	50	10
6	10	4	10	1.50	50	20
7	10	4	10	1.50	75	10
8	10	4	10	1.50	75	20
9	1000	100	1	0.01	50	10
10	1000	100	1	0.01	50	20
11	1000	100	1	0.01	75	10
12	1000	100	1	0.01	75	20
13	1000	100	10	0.01	50	10
14	1000	100	10	0.01	50	20
15	1000	100	10	0.01	75	10
16	1000	100	10	0.01	75	20

2.5 Figure 4

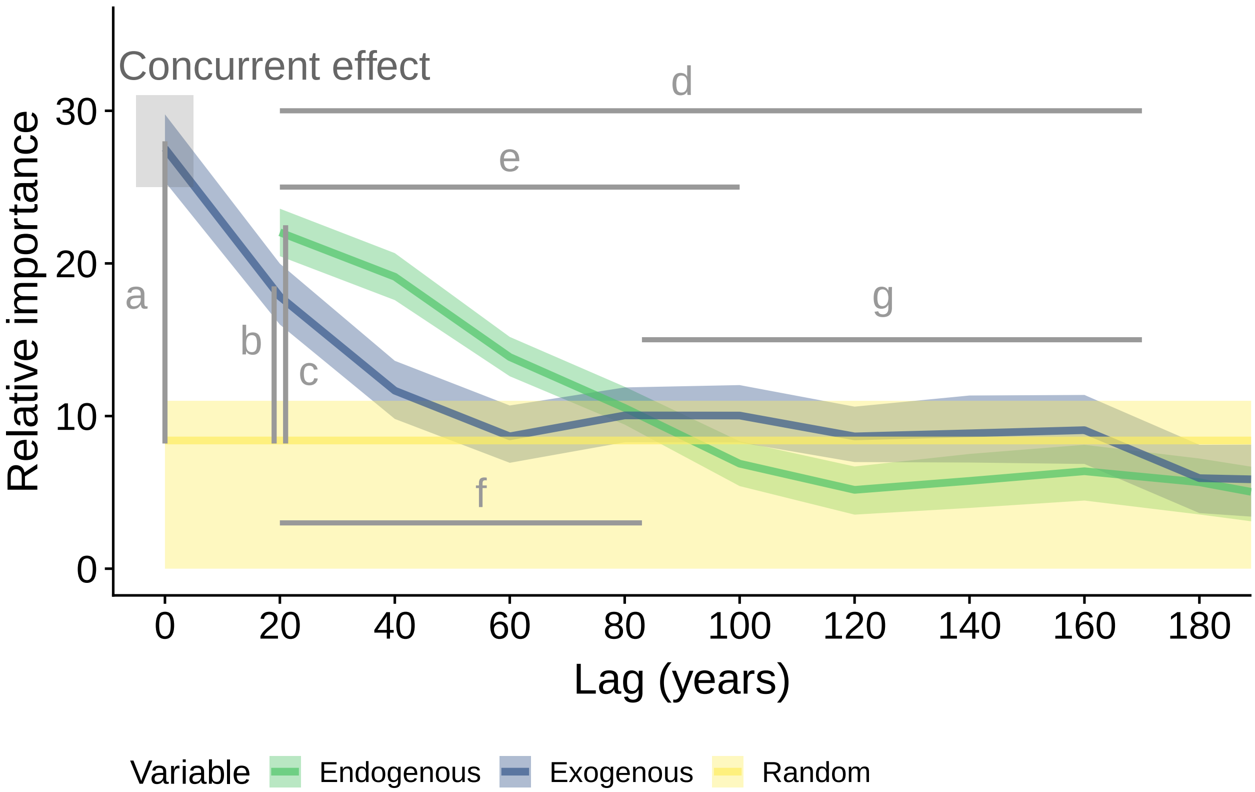

Figure 4: Example of ecological memory pattern showing the relative importance of each component at each time lag as computed by Random Forest. The ecological memory features measured are: a) strength of the concurrent effect (highlighted by a gray box); b) strength of the exogenous memory; c) strength of the endogenous component; d) length of the exogenous component; e) length of the endogenous component; f) dominance of the endogenous component; g) dominance of the exogenous component. Note that only data above the median of the Random variable (yellow line) are considered for the computations.

2.6 Figure 5

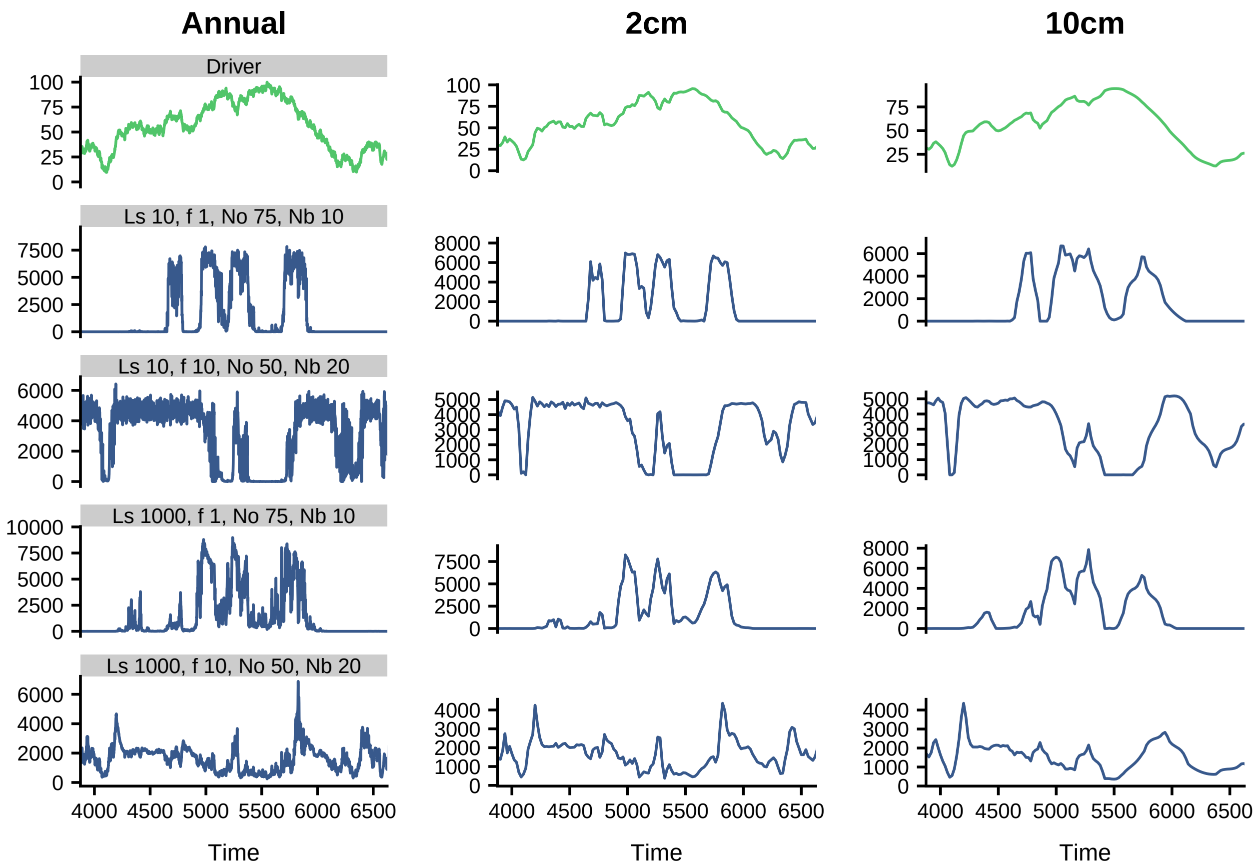

Figure 5: Driver (light green) and a subset of simulations (marine-blue) at the original Annual resolution (left panel), and resampled at 2 and 10cm intervals interpolated at 20 years resolution (centre and right panels).

2.7 Table 2

Table 2: Pearson correlation index between simulated pollen curves and the driver at different sampling resolutions (columns Annual, 1cm, 2cm, 6cm, and 10cm), compared with niche and life traits used for the simulations.

Life-span	Fecundity	Niche optimum	Niche breadth	Annual	1cm	2cm	6cm	10cm
10	1	50	10	0.04	0.05	0.06	0.04	0.08
10	1	50	20	0.02	0.02	0.01	0.00	0.01
10	10	50	10	0.03	0.03	0.04	0.03	0.12
10	10	50	20	0.05	0.06	0.05	0.01	0.03
1000	1	50	10	0.00	0.01	0.02	0.04	0.03
1000	1	50	20	0.05	0.05	0.03	0.01	0.04
1000	10	50	10	0.02	0.02	0.01	0.07	0.04
1000	10	50	20	0.02	0.03	0.01	0.09	0.09
10	1	75	10	0.63	0.64	0.64	0.65	0.73
10	1	75	20	0.88	0.89	0.90	0.89	0.91
10	10	75	10	0.70	0.73	0.73	0.73	0.78
10	10	75	20	0.86	0.89	0.90	0.90	0.91
1000	1	75	10	0.59	0.61	0.63	0.65	0.69
1000	1	75	20	0.88	0.89	0.90	0.90	0.90
1000	10	75	10	0.64	0.66	0.67	0.68	0.71
1000	10	75	20	0.76	0.77	0.79	0.81	0.81

2.8 Figure 6

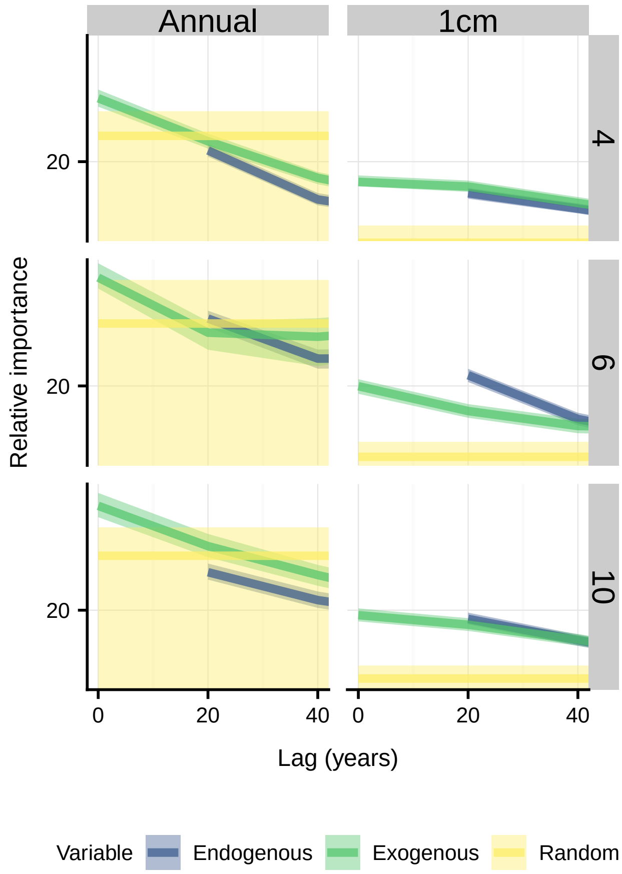

Figure 6: Effect of sediment accumulation rate on ecological memory patterns. Each row represents a virtual taxa (ID codes in vertical stripes correspond with the ones in Table 1) for which ecological memory showed contrasting patterns between the Annual and the 1cm datasets.).

2.9 Figure 7

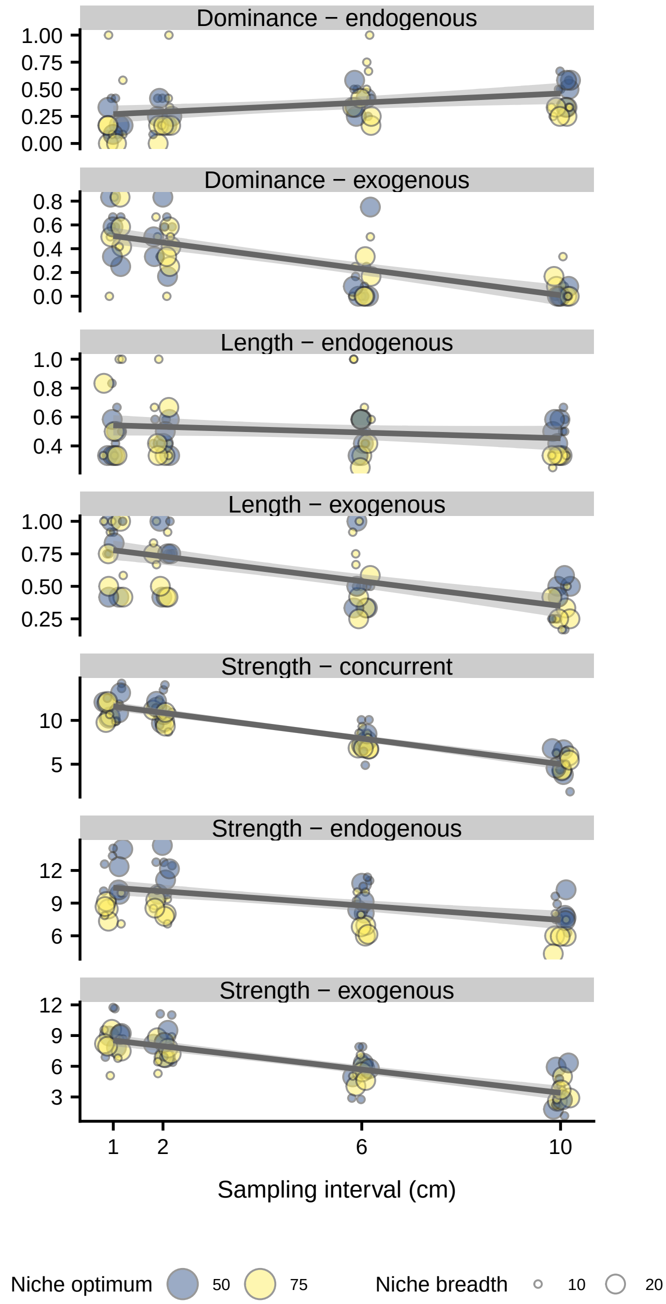

Figure 7: Ecological memory features across decreasing sampling densities. Dots represent virtual taxa (horizontal jitter was applied to the points for ease of examination). Dot colour and size represent niche traits. Linear model was fitted with the lm function of the R package.

2.10 Figure 8

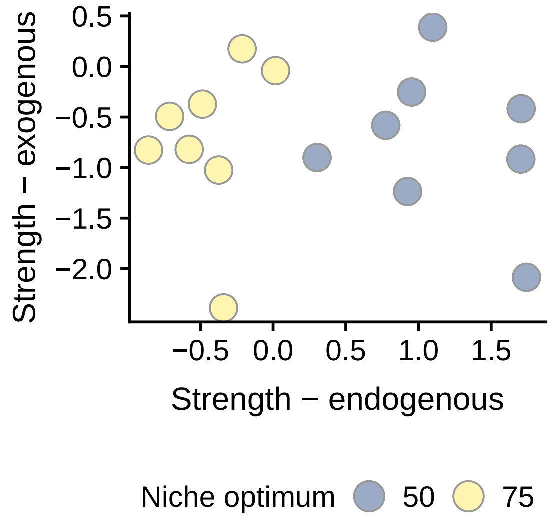

Figure 8: Change in strength of the endogenous and exogenous memory components when using driver instead of suitability to quantify ecological memory. Values represent changes in endogenous (horizontal axis) and exogenous (vertical axis) memory strength when using the driver to analyse ecological memory, instead of the suitability index produced from the driver by the niche function of the virtual taxa. Colours represent values of the niche optimum of the virtual taxa.

2.11 Figure 9

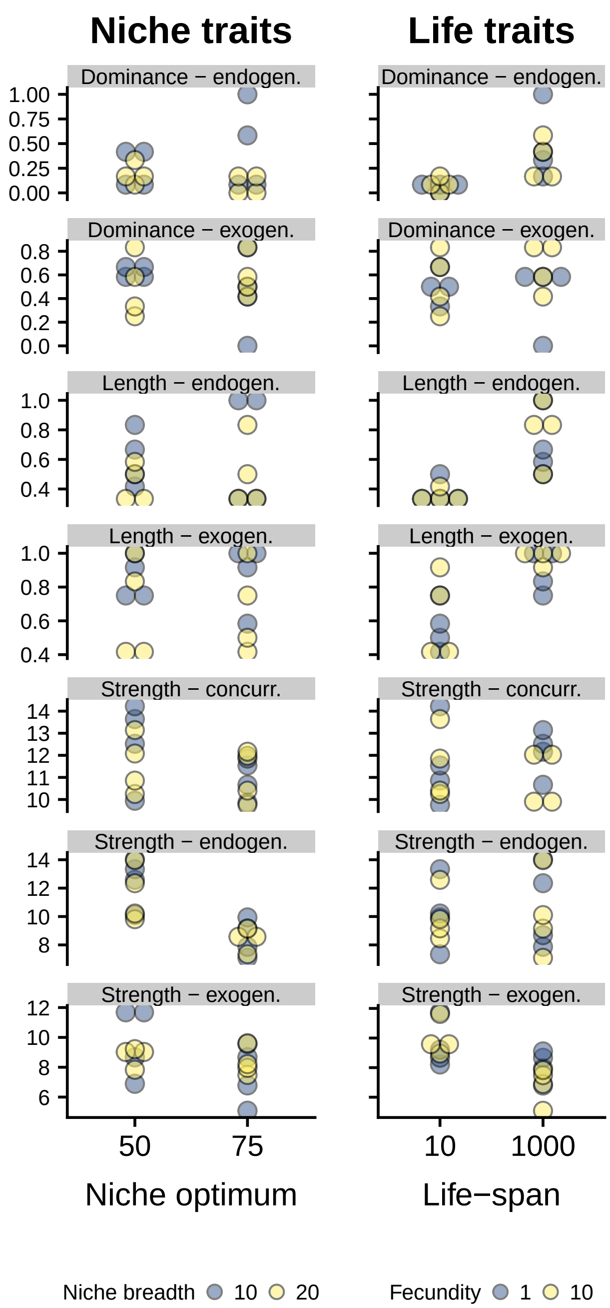

Figure 9: Ecological memory features for combinations of niche optimum and niche breadth (left panel) and life-span and fecundity (right panel). Each point represents a virtual taxa. Horizontal jitter was applied to overlapping points.