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 Quantifying 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 Fitting Random Forest models

General parameters for the simulation.

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

The experiments ran in the previous section can be easily plotted with the *plotExperiment()* function as follows:

```
plotExperiment(
  experiment.output=E1.suitability,
  experiment.title="plot title",
  legend.position="bottom",
  R2=TRUE,
  sampling.names=sampling.names,
  strip.text.size=7,
  axis.x.text.size=9,
  axis.y.text.size=12,
  axis.x.title.size=14,
  axis.y.title.size=14,
  title.size=18,
  filename = "filename",
  caption = "figure caption")
```

The result of this function applied on the experiments is shown below.

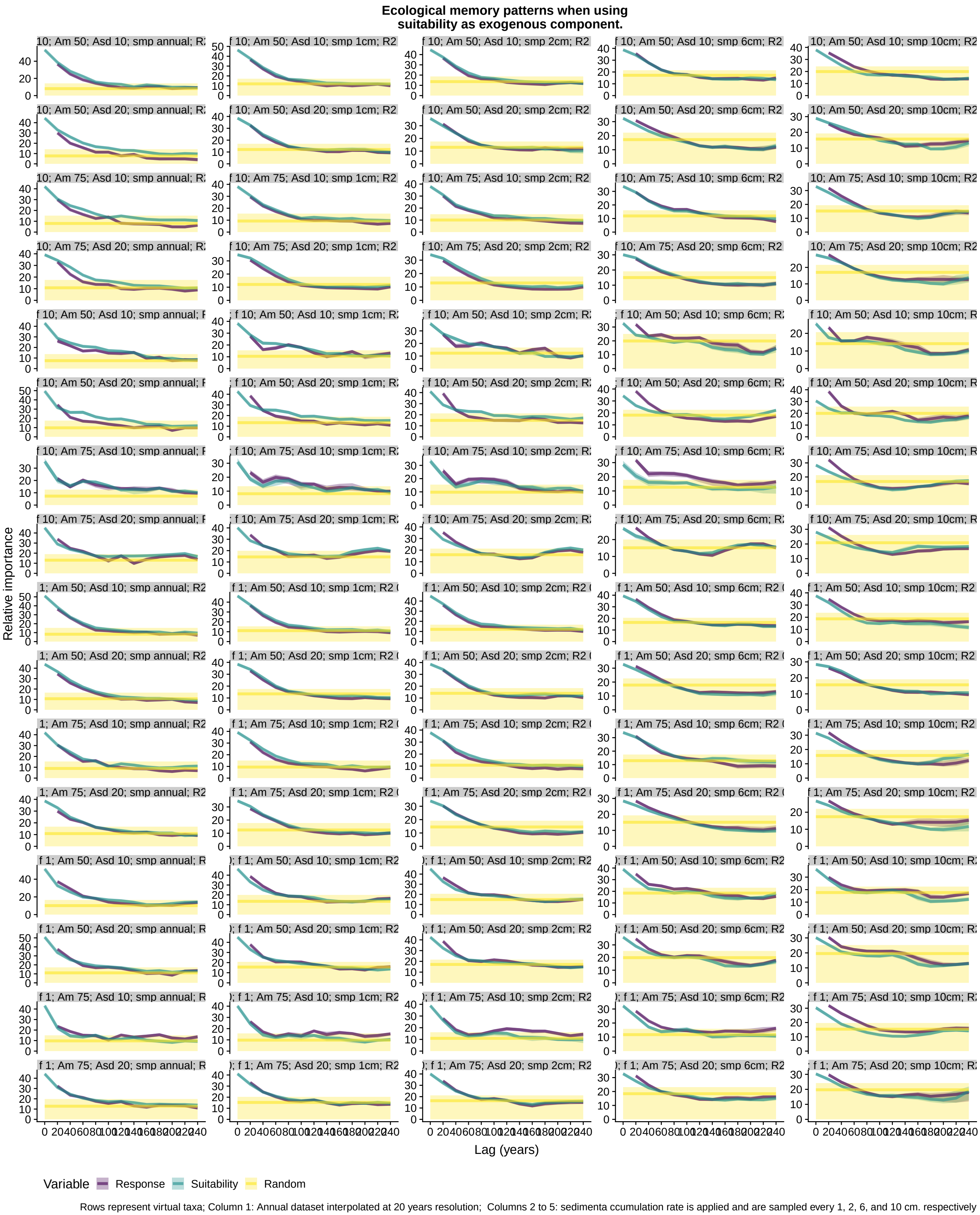

Figure 1: Ecological memory patterns of all virtual species (rows) and dataset types (columns) when using suitability (blue) as exogenous component (blue curves).

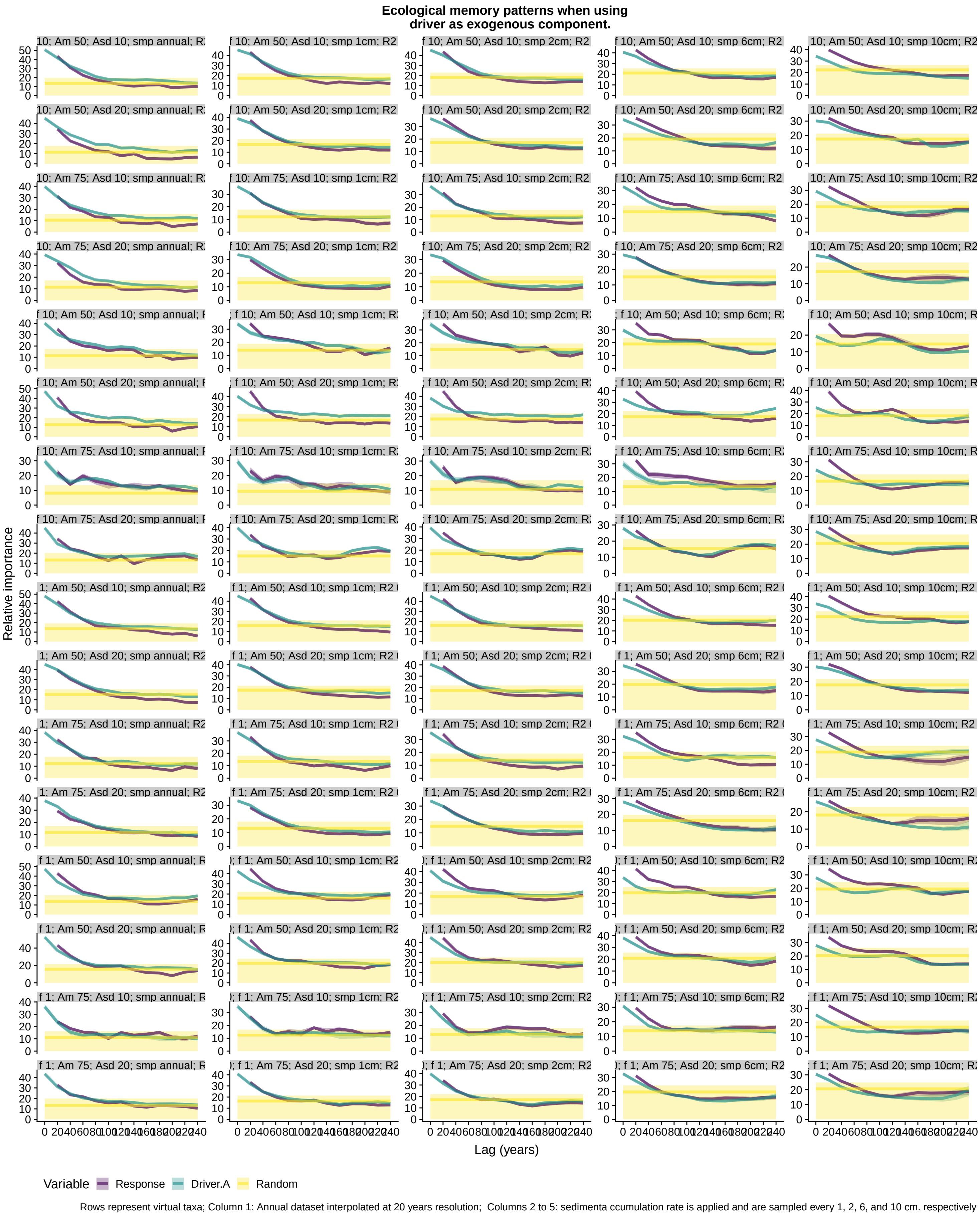

Figure 2: Ecological memory patterns of all virtual species (rows) and dataset types (columns) when using driver (blue) as exogenous component (blue curves).

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 optimum* or *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
36.64	0.71	35.40	37.69	Response	20	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
25.02	0.65	23.95	26.11	Response	40	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
18.32	0.63	17.29	19.41	Response	60	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
14.33	0.57	13.39	15.26	Response	80	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
11.38	0.63	10.40	12.44	Response	100	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
9.22	0.65	8.23	10.45	Response	120	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.77	0.76	7.40	10.00	Response	140	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
10.28	0.75	9.03	11.46	Response	160	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
10.42	0.65	9.29	11.43	Response	180	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.20	0.68	7.10	9.34	Response	200	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.94	0.70	7.87	10.06	Response	220	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
9.24	0.75	7.99	10.42	Response	240	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
53.19	0.81	51.98	54.54	Suitability	0	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
38.65	0.71	37.58	39.96	Suitability	20	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
28.34	0.64	27.29	29.34	Suitability	40	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
22.33	0.60	21.42	23.38	Suitability	60	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
15.83	0.60	14.82	16.72	Suitability	80	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
13.96	0.65	12.82	14.99	Suitability	100	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
13.12	0.69	11.93	14.19	Suitability	120	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
10.21	0.69	9.24	11.44	Suitability	140	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
12.48	0.76	11.21	13.64	Suitability	160	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
11.15	0.72	9.98	12.26	Suitability	180	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
9.72	0.71	8.62	10.88	Suitability	200	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
9.84	0.71	8.70	10.95	Suitability	220	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
9.18	0.79	7.88	10.37	Suitability	240	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	20	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	40	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	60	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	80	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	100	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	120	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	140	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	160	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	180	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual
8.09	3.66	0.00	14.25	Random	200	0.96	0.00045	10.85	3.48436	S10A50-5_f10	10	4	10	1.5	100	10000	50	10	Annual

2 Paper figures and tables

2.1 Figure 1

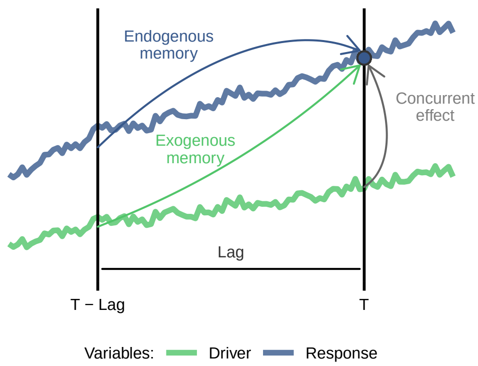

Figure 1: Components of ecological memory. Antecedent values of the driver and the response for a given lag length are located at T – Lag. Arrows represent the relative contribution of endogenous memory, exogenous memory, and the concurrent effect on the value of interest of the response at time T, represented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

2.2 Figure (removed from paper)

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

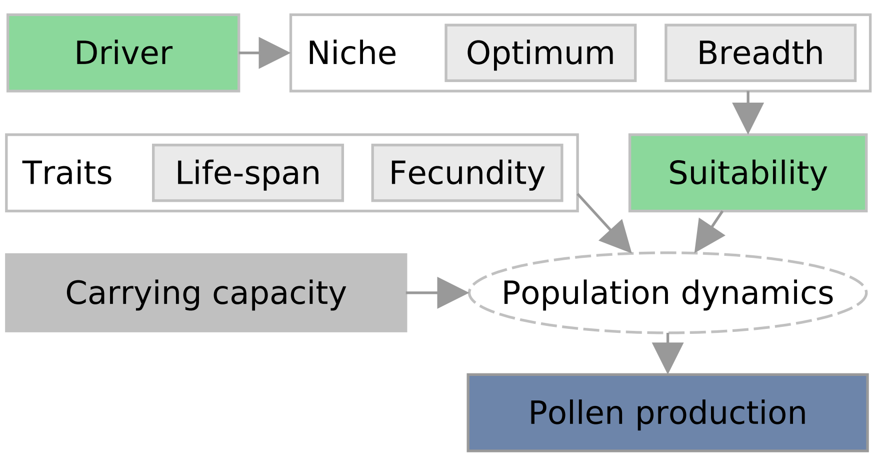

Figure 2: Model structure. Green and white boxes represent respectively the environment and elements relative to the virtual taxa. Light-gray boxes are the experimental variables, changing across model executions. Population dynamics (dashed ellipse) is the emergent process from which pollen production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

2.3 Figure (removed from paper)

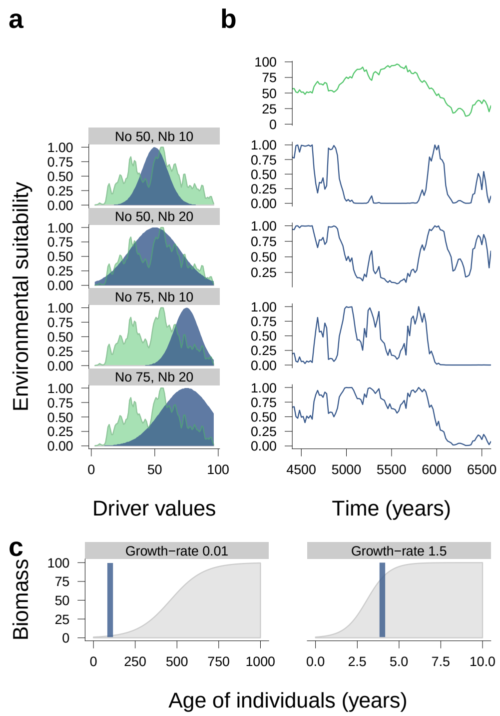

Figure 3: Panel a represents the different ecological niches used in the simulations. Light green colour represents the driver and marine blue represents the niche of the virtual taxa. Numbers in grey strips represent the mean (m) and standard deviation (sd) of the Gaussian functions used to define the 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 Figure 2

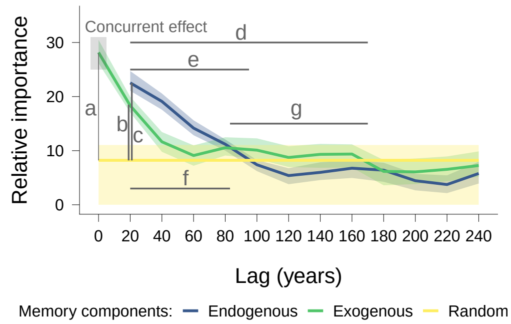

Figure 4: Example of ecological memory pattern showing the relative importance of each component at each time lag as computed by Random Forest. The ecological memory features measured are: a) strength of the concurrent effect (highlighted by a gray box); b) strength of the exogenous memory; c) strength 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.

2.5 Figure (removed from paper)

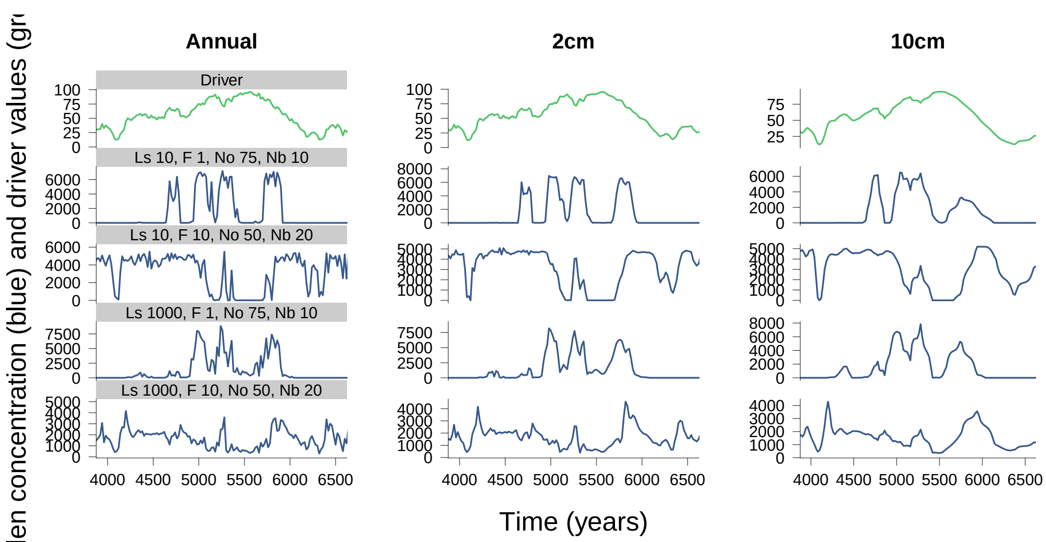

Figure 5: Driver (light green) and a subset of simulations (marine-blue) derived from the original Annual resolution (left panel), and resampled at 2 and 10cm (centre and right panels).

2.6 Table 1 (removed from paper)

Table 1: 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	0.05	0.09
10	1	50	20	0.02	0.03	0.02	0.01	0.01
10	10	50	10	0.02	0.03	0.04	0.03	0.13
10	10	50	20	0.05	0.06	0.04	0.02	0.00
1000	1	50	10	0.00	0.01	0.02	0.04	0.03
1000	1	50	20	0.06	0.05	0.03	0.01	0.04
1000	10	50	10	0.03	0.02	0.01	0.07	0.04
1000	10	50	20	0.03	0.02	0.01	0.10	0.08
10	1	75	10	0.63	0.64	0.64	0.64	0.70
10	1	75	20	0.88	0.90	0.90	0.89	0.91
10	10	75	10	0.69	0.72	0.74	0.73	0.77
10	10	75	20	0.86	0.89	0.90	0.90	0.92
1000	1	75	10	0.58	0.61	0.63	0.65	0.69
1000	1	75	20	0.87	0.89	0.90	0.90	0.90
1000	10	75	10	0.63	0.66	0.67	0.68	0.72
1000	10	75	20	0.76	0.78	0.80	0.81	0.82

2.7 Figure 3

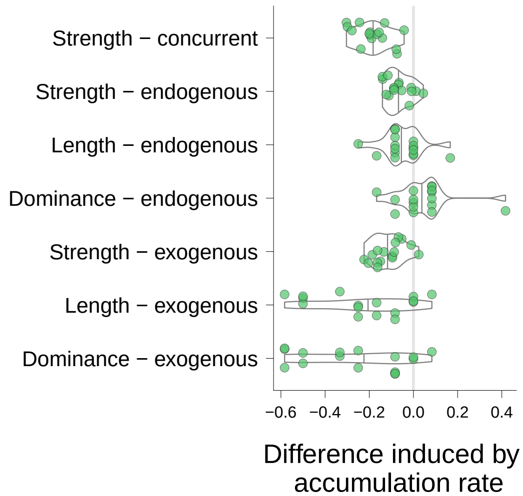

Figure 6: Effect of variable sediment accumulation rate on ecological memory features. Dots represent virtual taxa. Values of the x axis show the difference in ecological memory features when substracting those of the 1cm dataset (variable accumulation rate) to those of the Annual dataset (constant accumulation rate). Violin-plots, in blue, show the underlying distribution of the data. Vertical blue lines represent the median.

2.8 Figure 4

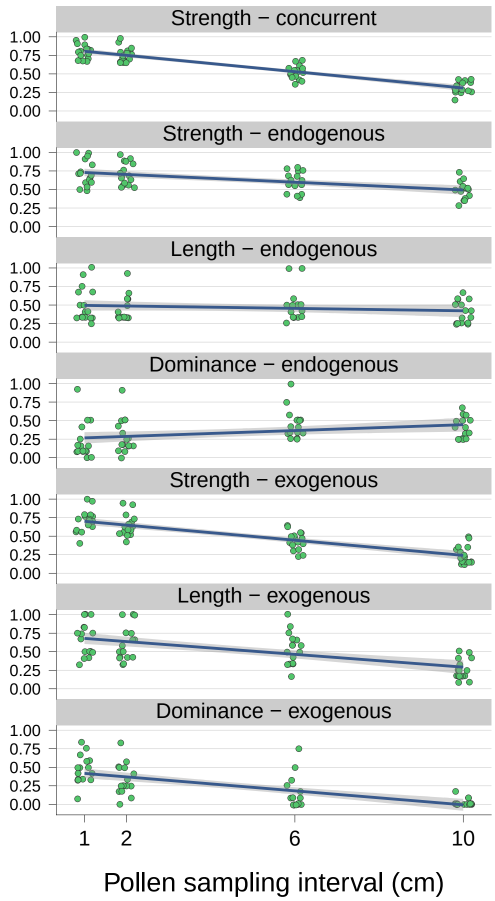

Figure 7: Ecological memory features across decreasing sampling densities. Dots represent virtual taxa (horizontal jitter was applied to the points for ease of examination).

2.9 Figure 5

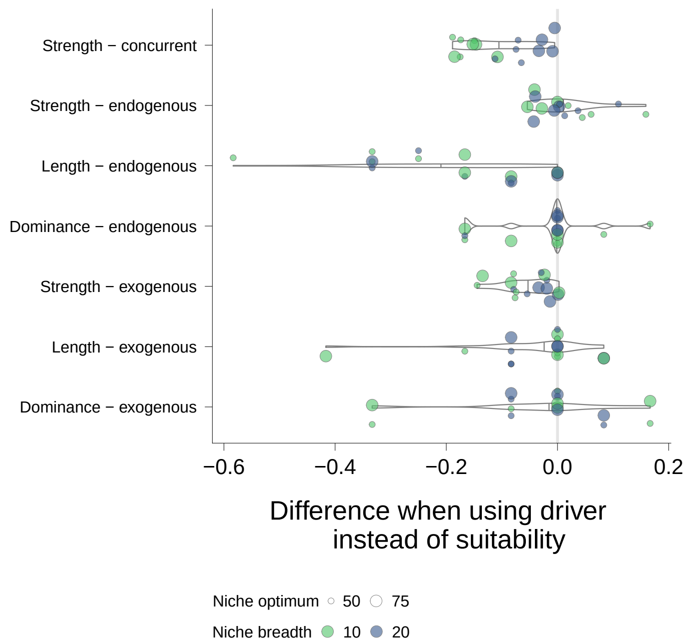

Figure 8: Change in values of the different ecological memory features when using driver as exogenous memory component instead of suitability. Colors represent niche breadth (higher if blue), and size represent niche position (larger for a marginal niche position).

2.10 Figure 6

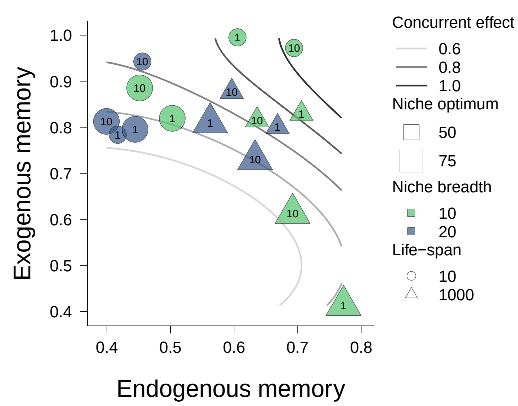

Figure 9: Relationship between life traits, niche features, and ecological memory components, when using environmental suitability as exogenous component. Colors represent niche breadth, size represents niche optimum, and shape represents life-span. Numbers within figures represent fecundity. X and Y axes are computed as the average of the length, dominance, and strength of the endogenous and exogenous components of ecological memory. Contour lines represent the relative strength of the concurrent effect (synchronous influence of the driver over the suitability response).