Preliminary results paleoarea project

Pedro Santos Neves

19 of the 141 island age runs have been completed and all have yielded a best model. Here, I present an overview of the best models selected for each age. I submitted jobs with intervals of roughly 5ky in order to cover a wider temporal parameter space, the exception being other time slices of interest such as the "representative" age of 16ky and the oldest available age of 140ky.

```
library(islandpaleoarea)
results <- read_results(file.path("results", "archipelagos41_paleo"))
```

Basic sanity checks

```
nrow(results)
```

```
## [1] 7143
```

```
length(unique(results$age))
```

```
## [1] 19
```

7143 runs completed. Submitted jobs consisted of 15 randomly sampled starting parameters for optimization for each of the 28 models in each island age. Thus, out of a total of 7980 submissions, 89.5% was successful.

The following table lists the age and model combinations for which the optimisation procedure was unable to estimate any paremeters. A higher number in n_na how many of the 15 randomly sampled initial parameter combinations failed at obtaining results. Note that this does not account for HPCC jobs that crashed or timed out.

```
saved_results_tally <- results |>
  dplyr::group_by(age, model) |>
  dplyr::summarise(n_na = sum(is.na(lambda_c0))) |>
  dplyr::arrange(dplyr::desc(n_na))
```

```
## `summarise()` has grouped output by 'age'. You can override using the `.groups`
## argument.
```

```
knitr::kable(head(saved_results_tally))
```

age	model	n_na
30	17	12
35	17	12
45	18	12
65	17	12
15	18	10
20	18	10

For every age/model of the total to 532, 504 age/model combinations finished. All of the 504 sets have at least 3 successful seeds.

Results

Model selection

The following table presents the best model for each completed island age run.

```
ordered_results <- results |>
  dplyr::mutate(bic = calc_bic(loglik, df, 1000)) |>
  dplyr::group_by(age) |>
  dplyr::arrange(dplyr::desc(bic), .by_group = TRUE) |> dplyr::slice_min(bic)
knitr::kable(ordered_results, digits = 2)
```

age	model	seed	lambda_0	c_0	mu_0	x	K_0	z	gamma_0	alpha	lambda_beta	beta	loglik	df	conv	bic
1	19	279	0.04	0	2.35	0.17	Inf	0	73.38	0.30	0.06	0.39	-3650.90	8	0	7371.77
5	9	122	0.20	0	3.41	0.21	Inf	0	55.96	0.26	0.04	0.44	-3662.37	7	0	7385.95
10	9	123	0.20	0	3.62	0.21	Inf	0	59.83	0.27	0.04	0.45	-3662.26	7	0	7385.74
15	17	255	0.12	0	3.47	0.20	Inf	0	78.10	0.31	0.05	0.40	-3650.02	9	0	7378.75
16	17	247	0.11	0	3.44	0.20	Inf	0	80.50	0.32	0.05	0.39	-3650.32	9	0	7379.35
17	17	248	0.11	0	3.47	0.20	Inf	0	79.75	0.31	0.05	0.39	-3650.77	9	0	7380.25
18	17	241	0.12	0	3.51	0.20	Inf	0	80.32	0.32	0.05	0.40	-3651.08	9	0	7380.86
20	17	250	0.11	0	3.51	0.20	Inf	0	78.57	0.31	0.05	0.40	-3651.41	9	0	7381.52
25	17	252	0.11	0	3.49	0.20	Inf	0	80.37	0.32	0.05	0.39	-3650.89	9	0	7380.50
30	17	245	0.11	0	3.43	0.20	Inf	0	81.53	0.32	0.05	0.39	-3650.02	9	0	7378.75
35	17	255	0.11	0	3.36	0.20	Inf	0	81.20	0.32	0.05	0.39	-3650.44	9	0	7379.59
40	17	245	0.12	0	3.37	0.20	Inf	0	81.76	0.32	0.05	0.40	-3650.37	9	0	7379.45

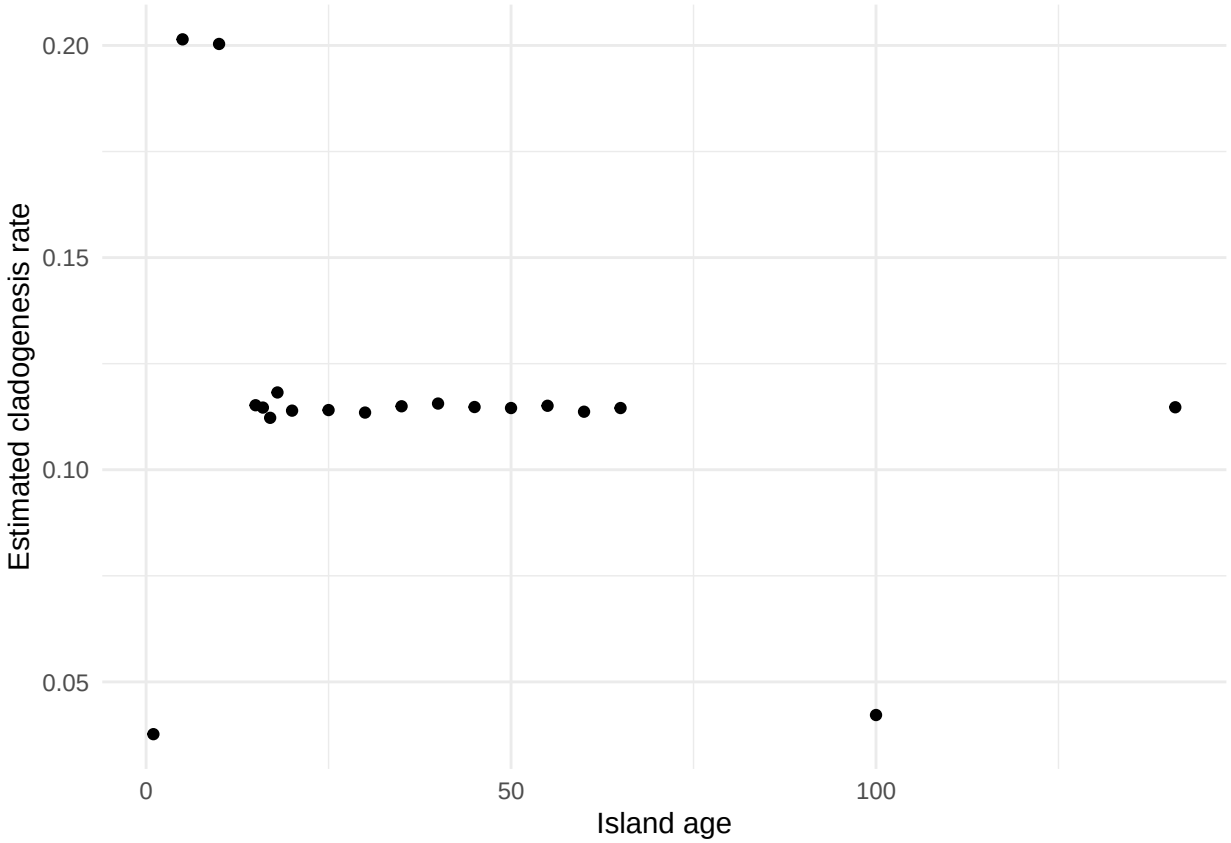
age	model	seed	lambda_c_0	mu_0	x	K_0	z	gamma_0	alpha	lambda_a_0	beta	loglik	df	conv	bic
45	17	246	0.11	0	3.45	0.20	Inf	0	79.45	0.31	0.05	0.40	-3650.02	9	0	7378.75
50	17	243	0.11	0	3.45	0.20	Inf	0	81.50	0.32	0.05	0.40	-3650.01	9	0	7378.73
55	17	249	0.12	0	3.46	0.20	Inf	0	81.46	0.32	0.05	0.40	-3650.01	9	0	7378.73
60	17	242	0.11	0	3.39	0.20	Inf	0	81.11	0.32	0.05	0.39	-3650.16	9	0	7379.03
65	18	258	0.11	0	3.39	0.20	Inf	0	81.56	0.32	0.05	0.40	-3650.53	9	0	7379.77
100	19	271	0.04	0	2.56	0.17	Inf	0	76.46	0.31	0.06	0.39	-3651.79	8	0	7373.54
141	17	242	0.11	0	3.25	0.20	Inf	0	82.65	0.32	0.05	0.39	-3649.01	9	0	7376.72

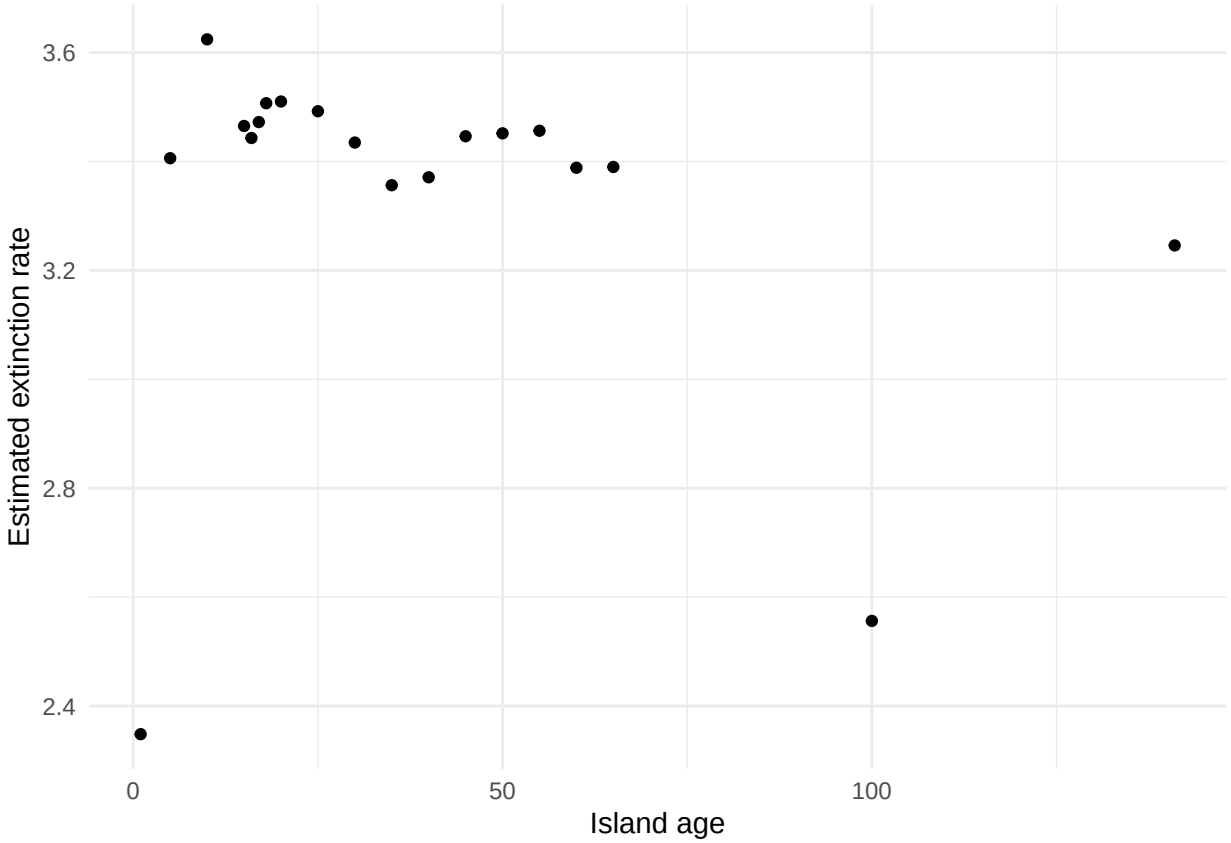
There is a clear preference for models 17 and 19. Models 14 and 19 were chosen as the best models in the Nature paper. Models 17 and 19 are both *post hoc* power models and vary only in the exponent of the power law in the cladogenesis rate.

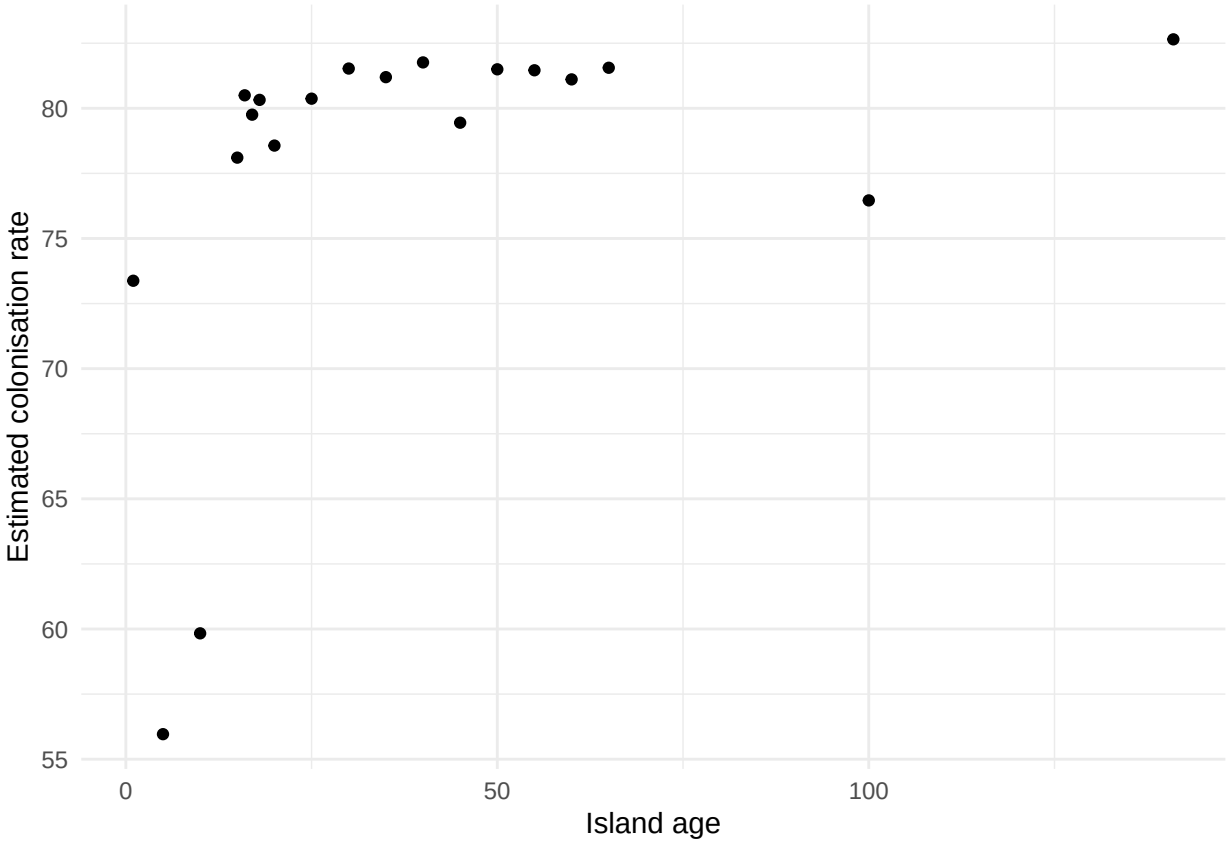
- M19 $\lambda^c = \lambda_0^c A^{d_0 log(D)}$
- M17 $\lambda^c = \lambda_0^c A^{y + \frac{D}{d_0}}$

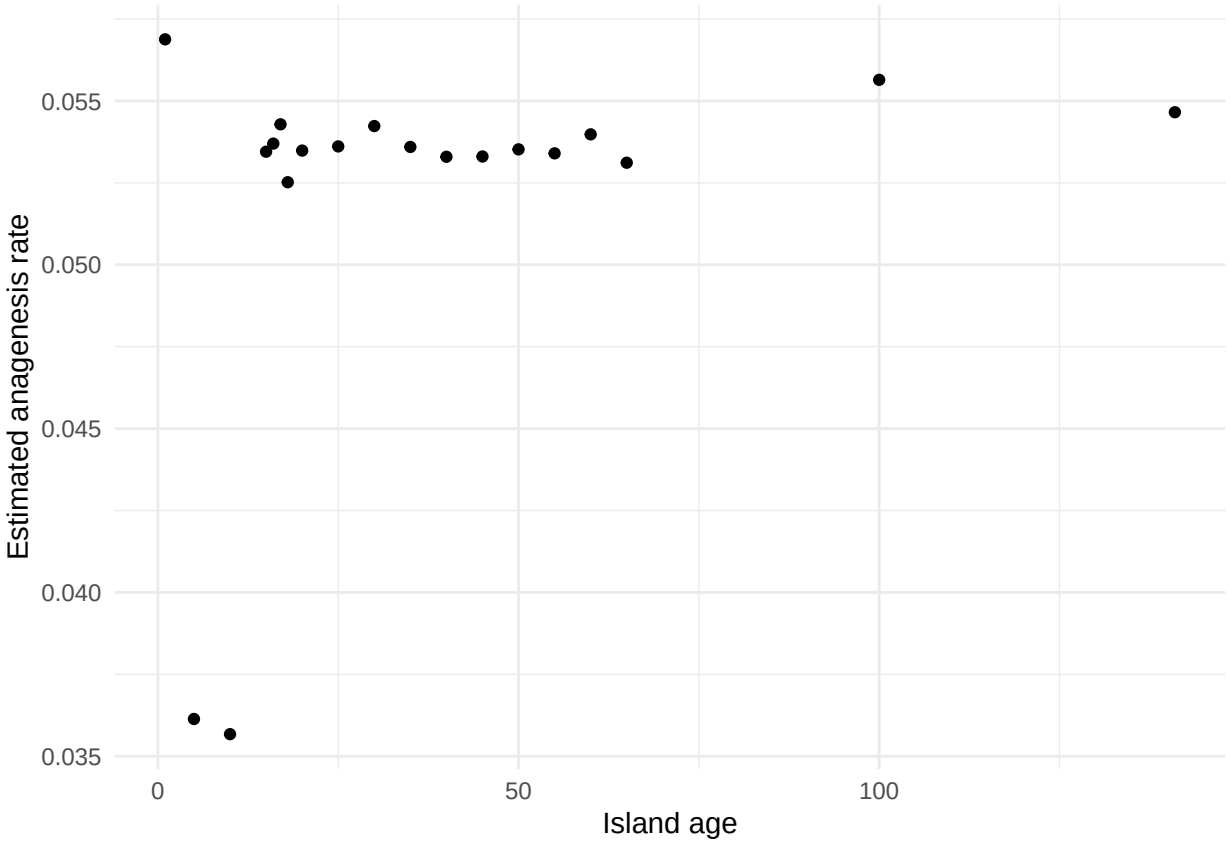
Parameter estimates

For a very course grained visualisation of the effect of area on estimates, below I present plots of estimates over time. Note that I make no distinction by the selected model, so these plots should be taken as a mere overview and not a detailed analysis.


Estimated cladogenesis rate
0.20
0.15
0.10
0.05
0
50
100
Island age


3.6
3.2
2.8
2.4
Estimated extinction rate
0
50
100
Island age


Estimated colonisation rate
80
75
70
65
60
55
0
50
100
Island age



The most striking observation relates to the large discrepancy of all CES estimates from present (age = 0) to any of the older island age estimates, being substantially lower at present and consistently higher up to 60kya. Whether this linear pattern will hold to the oldest available area estimates remains to be seen as more results become available.