Supplementary Information - ECOGRAPHY

P. CHAMBAULT et al.

09/11/2020

Contents

1 Context

The purpose of this document was to test the influence of uneven tracking durations and number of locations per individual on model outputs.

2 Methods

A similar sample size was tested for each individual to reduce the spatial autocorrelation and give the same weight to each turtle.
The dataset was first reduced to 60 days and subsampled to 2 locations per day per turtle to give the same weight to each individual. Fourteen individuals were left (21 from the full dataset). The 60 days tracking duration was chosen based on a trade-off between the number of individuals and the total number of occurrences (tracking duration of the original dataset: from 17 to 266 days, mean=103 days). If the selected tracking duration is too high, the number of remaining individuals will be drastically reduced and vice versa.

1) *Pseudo-absence data generation*: generate prediction maps on the reduced dataset.

2) *K-fold*: the dataset was split into the training (2/3 of the data) and the validation dataset (1/3 of the data). The model of each pseudo-asbence simulation was run on the training dataset and the performance metrics calculated on the validation dataset.

3) *GLM and GLMM*: run models on the reduced dataset and compare the performance metrics to the model containing the full dataset.

4) *Spatial autocorrelation* was tested by generating a variogram based on each model residuals.

3 Results

```
# Summary of the tracking data for each turtle.
df
```

```
      ptt  StartDate    EndDate Nloc NlocPerDay Duration
1  121819 2012-11-28 2013-06-21  357          4      205
2  32888c 2012-11-28 2013-02-05 1406         20       69
3  121814 2012-12-28 2013-02-03  233          6       37
4  121815 2012-12-12 2013-09-05 4909         19      267
5  121816 2012-12-05 2013-03-06  198          3       91
6  121817 2012-11-29 2013-07-03 3264         15      216
7  121818 2012-11-29 2013-04-20 1027          7      142
8  121821 2012-12-21 2013-07-02  788          6      193
9  32888d 2013-03-08 2013-03-26  106          6       18
10 32897c 2012-11-28 2013-03-31 1225         11      123
11 32899c 2012-11-28 2013-04-11 1662         12      134
12 32900b 2010-10-25 2011-02-06  471          5      104
13 167852 2019-02-07 2019-04-27 1250         16       79
14 169513 2018-05-14 2018-07-01  168          4       48
15 169514 2018-05-14 2018-06-28   75          2       45
16 169515 2018-05-14 2018-07-29   98          3       76
17 169516 2018-05-01 2018-09-01 2391         22      123
18 169518 2019-01-24 2019-04-01  446          7       67
19 169519 2019-01-24 2019-04-08 1407         19       74
20 169520 2019-01-25 2019-02-21  123          4       27
21 169521 2019-02-07 2019-03-19  391         10       40
```

3.1 Spatial autocorrelation

3.1.1 Global models

```
# Variogram of the Global model on the full dataset
v_g_f$var_model
```

```
  model      psill    range kappa
1   Nug 0.00000000    0.000   0.0
2   Ste 0.03416412 1126.339   0.2
```

```
# Variogram of the Global model on the reduced dataset
v_g_r$var_model
```

```
  model      psill   range kappa
1   Nug 0.00000000   0.000   0.0
2   Ste 0.02826748 464.179   0.3
```

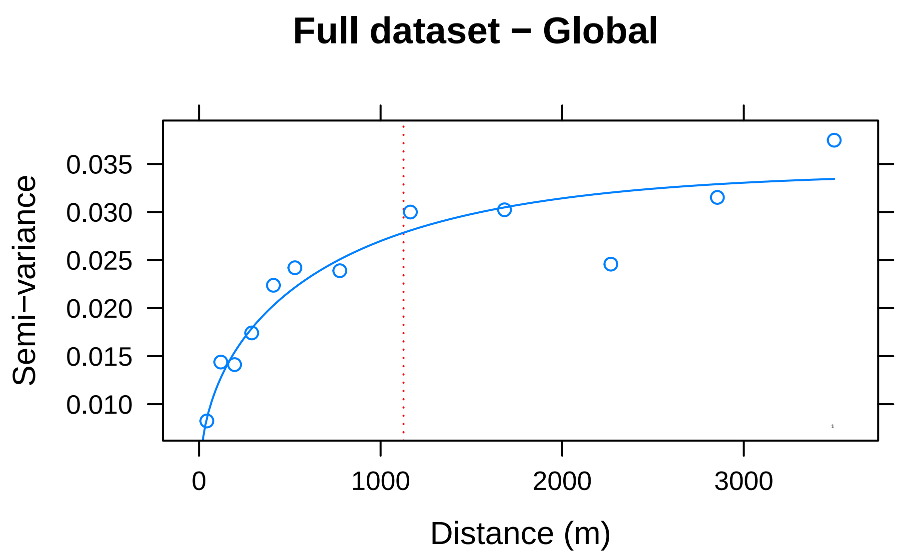

Figure 1: Variogram on the residuals of the Global model from the full dataset. Semi-variance (Y axis) according to the distance bewteen pairs of locations (X axis).

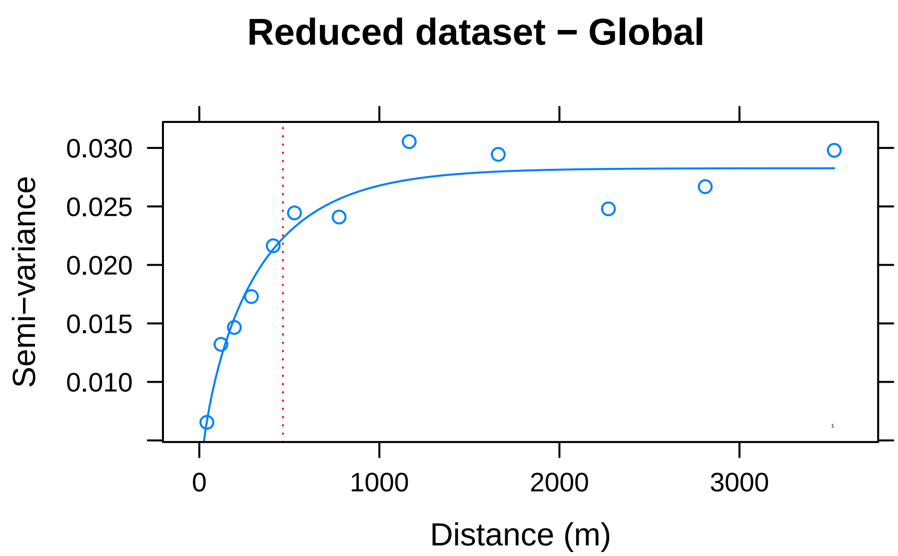

Figure 2: Variogram on the residuals of the Global model from the reduced dataset. Semi-variance (Y axis) according to the distance bewteen pairs of locations (X axis).

3.1.2 Individual models

```
# Variogram of the Individual model on the full dataset
v_i_f$var_model
```

```
  model        psill    range
1   Nug 0.0066474389    0.000
2   Sph 0.0005737639 1592.276
```

```r
# Variogram of the Individual model on the reduced dataset
v_i_r$var_model
```

```
  model       psill    range
1   Nug 0.004871553   0.0000
2   Sph 0.003680714 119.7026
```

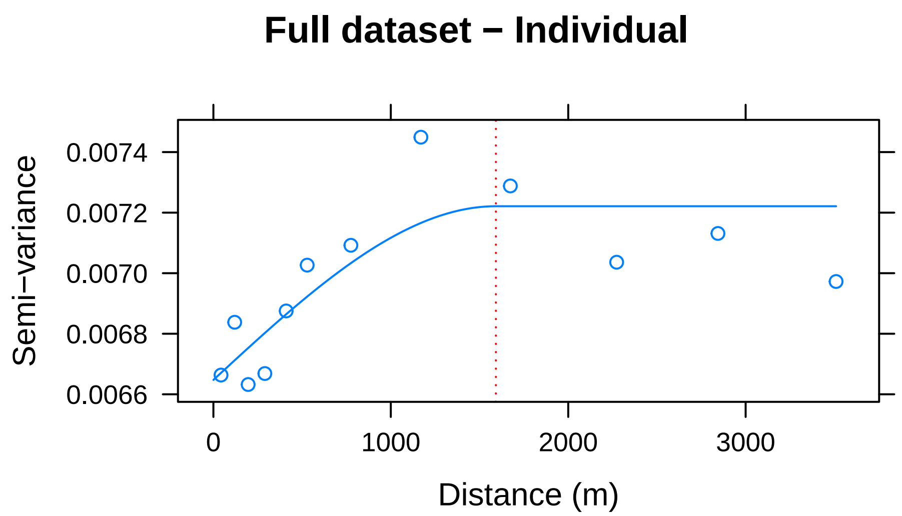

Figure 3: Variogram on the residuals of the Individual model from the full datasets.

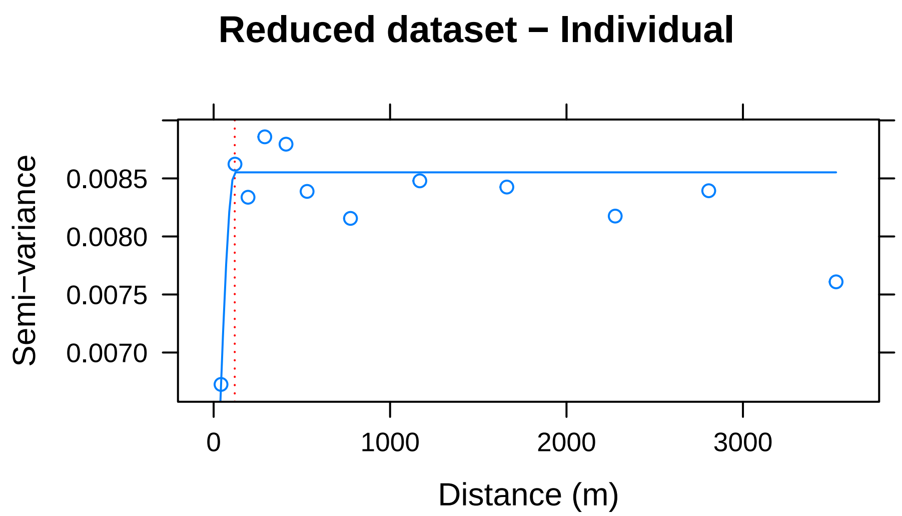

Figure 4: Variogram on the residuals of the Individual model from the reduced dataset.

3.2 Performance metrics and goodness-of-fit

3.2.1 Global models

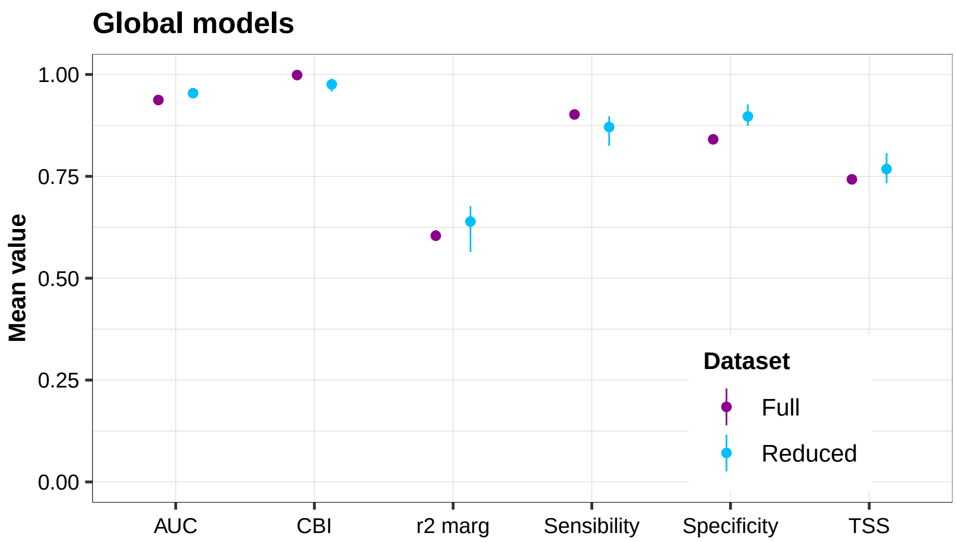

Figure 5: Performance metrics and goodness-of-fit for the Global models using the full dataset and the reduced dataset. The predictive power and the goodness-of-fit are similar for both datasets.

3.2.2 Individual models

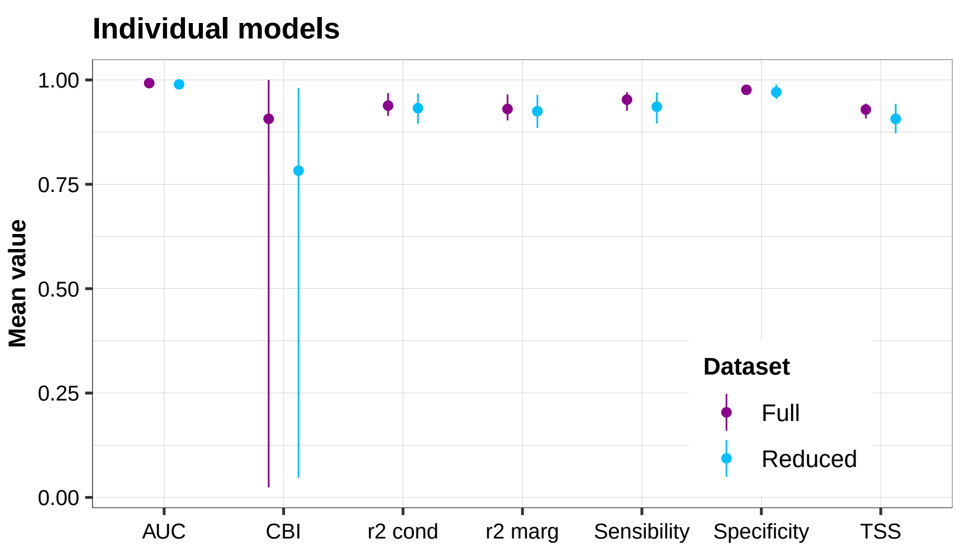

Figure 6: Performance metrics and goodness-of-fit for the Individual models using the full dataset and the reduced dataset. The predictive power and the goodness-of-fit are similar for both datasets.

3.3 Response curves

3.3.1 Global models

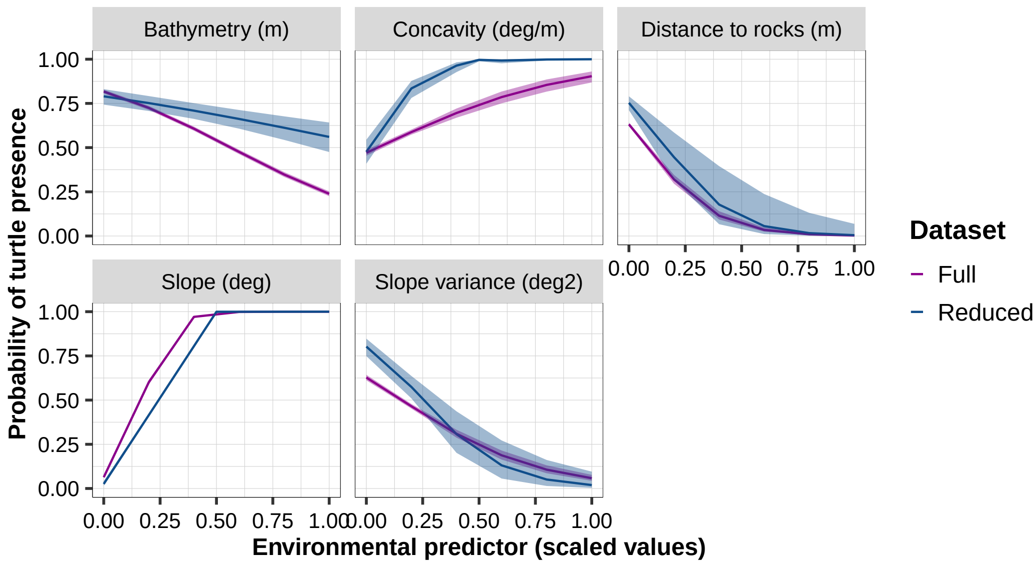

Figure 7: Response curves for the Global models using the full and reduced datasets. Similar curves to the model on the full dataset. All variables were significant. Similar curves were obtained from both datasets.

3.3.2 Individual models

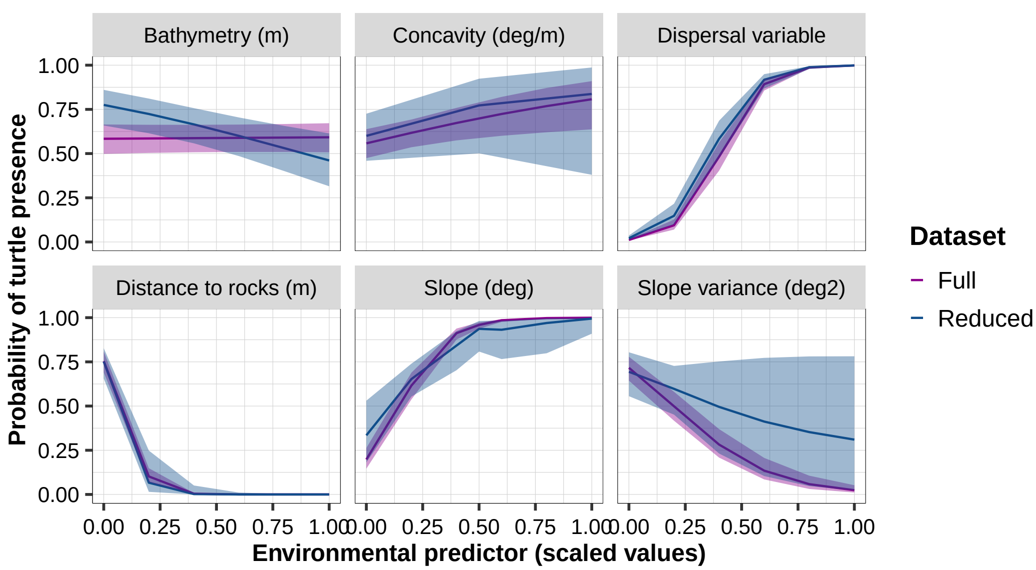

Figure 8: Response curves for the Individual models using the reduced dataset (120 locs/turtle).

Similar curves were obtained from both datasets, with larger confidence intervals with the reduced dataset. Only variables significance varied among datasets with *concavity* not significant for the reduced dataset whereas *litto* was not significant for the full dataset.

3.4 Prediction maps

3.4.1 Global models

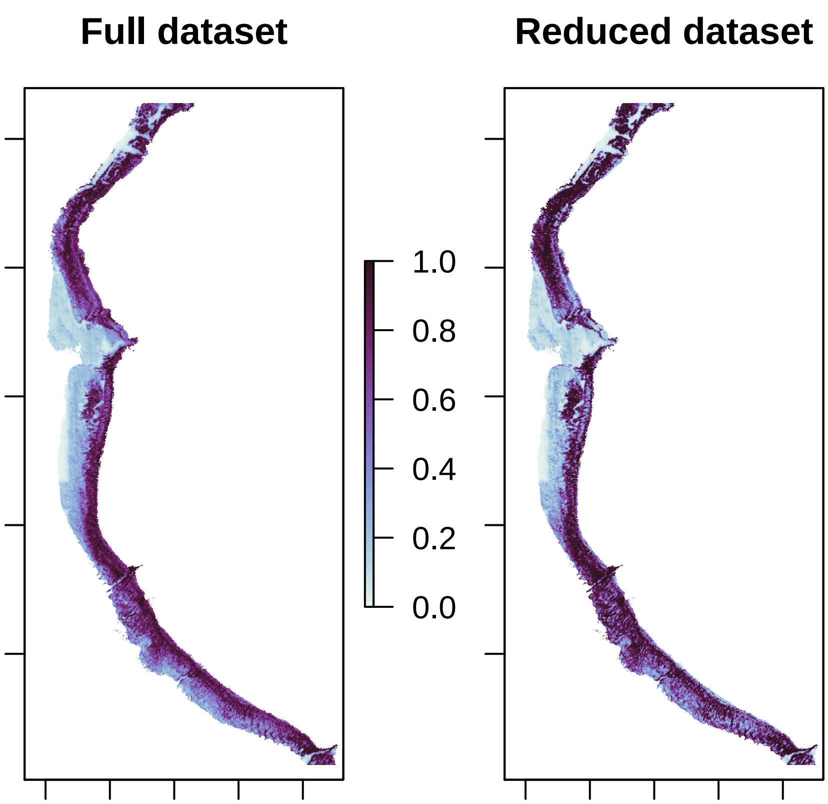

Figure 9: Averaged prediction maps for the Global models using the full dataset (right) and the reduced dataset (left).

The predictions maps for the Global models were similar for both the full and reduced datasets.

3.4.2 Individual models

The predictions maps for the Individual models were very similar to the ones of the full model, e.g. patches of highly suitable habitats localized around the three tagging locations.

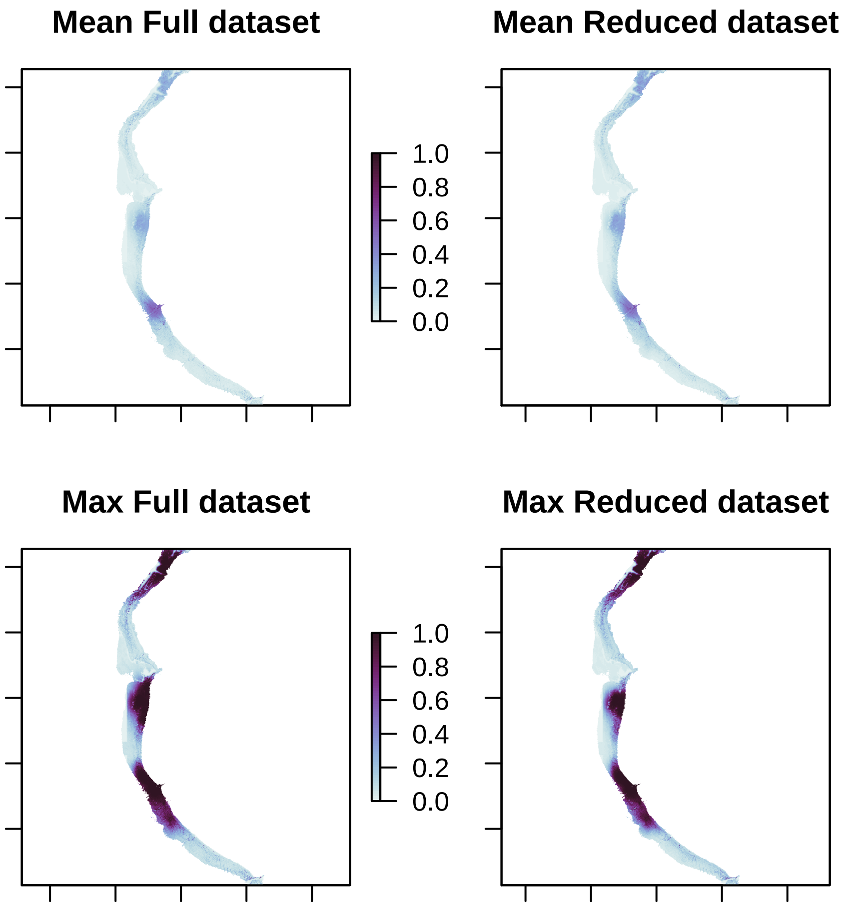

Figure 10: Averaged prediction maps for the Individual models using the full dataset (right) and the reduced dataset (left).

4 Conclusion

Reducing the dataset slightly reduced the spatial autocorrelation but gave similar results regarding performance metrics, response curves and prediction maps. Even when using the reduced dataset, there is still some spatial autocorrelation so we won't be able to get rid of it due to the very fine-scale movements of the turtles (<1 km width). We therefore decided to neglect the remaining spatial autocorrelation due to the average travel speed of the turtles and our low sampling interval, e.g. we definitely missed many occurrences that were not transmitted by the satellites so we could easily assume that the locations are spatially independent.