

Title: Finite mixtures in capture-recapture surveys for modelling residency patterns in marine wildlife populations

Author: Gianmarco Caruso, Pierfrancesco Alaimo Di Loro, Marco Mingione, Luca Tardella, Daniela Silvia Pace, Giovanna Jona Lasinio

Corresponding Author for Code:

Gianmarco Caruso (Email: gianmarco.caruso@uniroma1.it)

Configurations:

R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: windows 10 x64 (build 22000)

Matrix products: default

locale:

```
[1] LC_COLLATE=Italian_Italy.utf8 LC_CTYPE=Italian_Italy.utf8
[3] LC_MONETARY=Italian_Italy.utf8 LC_NUMERIC=C
[5] LC_TIME=Italian_Italy.utf8
```

attached base packages:

```
[1] parallel stats graphics grDevices utils datasets methods base
```

other attached packages:

```
[1] doRNG_1.8.6          rngtools_1.5.2      doParallel_1.0.17   iterators_1.0.14
[5] foreach_1.5.2        R2jags_0.7-1        rjags_4-13          coda_0.19-4
[9] cowplot_1.1.1        tricolore_1.2.2     ggtern_3.4.2        magrittr_2.0.3
[13] lubridate_1.9.2      forcats_1.0.0       stringr_1.5.0       purrr_1.0.1
[17] readr_2.1.4          tidyr_1.3.0         tidyverse_2.0.0     dplyr_1.1.1
[21] ggplot2_3.4.2        tibble_3.2.1        HDInterval_0.2.4
```

loaded via a namespace (and not attached):

```
[1] splines_4.2.3      assertthat_0.2.1    tensorA_0.36.2      robustbase_0.99-1
[5] remotes_2.4.2      R2WinBUGS_2.1-21    pillar_1.9.0        lattice_0.20-45
[9] quantreg_5.95      glue_1.6.2          digest_0.6.31       colorspace_2.1-0
[13] Matrix_1.5-3       plyr_1.8.8          pkgconfig_2.0.3     Rfast_2.0.1.4
[17] SparseM_1.81       scales_1.2.1        processx_3.8.2      RANN_2.6.1
[21] tzdb_0.3.0         MatrixModels_0.5-1 timechange_0.2.0    generics_0.1.3
[25] withr_2.5.0        hexbin_1.28.3       cli_3.6.1           proto_1.0.0
[29] survival_3.5-3     crayon_1.5.2        mcmc_0.9-7          ps_1.7.5
[33] fansi_1.0.4        MASS_7.3-58.2       compositions_2.0-6  pkgbuild_1.4.0
[37] RcppZigurat_0.1.6  tools_4.2.3         loo_2.6.0           prettyunits_1.1.1
[41] hms_1.1.3          lifecycle_1.0.3     matrixStats_0.63.0  MCMCpack_1.6-3
[45] munsell_0.5.0      callr_3.7.3         Rfast_2.0.7         compiler_4.2.3
[49] rlang_1.1.0        grid_4.2.3          rstudioapi_0.14     boot_1.3-28.1
[53] codetools_0.2-19  gtable_0.3.3        abind_1.4-5         curl_5.0.0
[57] R6_2.5.1           bayesm_3.1-6        gridExtra_2.3       utf8_1.2.3
[61] rprojroot_2.0.3    latex2exp_0.9.6     desc_1.4.2          stringi_1.7.12
[65] Rcpp_1.0.10        vctrs_0.6.1         DEoptimR_1.1-3      tidyselect_1.2.0
```

Content of the folder “codeForReproducibility”:

An R project has been created in the folder “codeForReproducibility” to make it easier for users to navigate through the content of the folder.

The file “JSsimfit_fun.R” contains the function to simulate a dataset from the RPT model and the functions to fit the RPT model and the ten JS-type alternative models. The BUGS scripts to fit the previous models is contained in the folder “BUGSmodel”.

The file “margPriors.R” contains the code for producing Figure 5 (a,b) in the Appendix.

The folder “results” contains every resulting table (.csv format) and figure (.jpg).

Real data analysis:

- “dolphins_uneven_3years.R” contains the code to load the data for the real data analysis. The dataset is stored in CSV format in the file “bottlenoseDolph.csv”;
- “fitRPTDolph.R” and “fitJStyleDolph.R” contain the code for fitting the RPT model and the ten JS-type alternative models. Since the model fitting via MCMC is computationally intensive (indeed, we suggest exploiting a High Performance Computing), we make the corresponding RData files available at the following link (folder: “outputDolph”):
<https://drive.google.com/drive/folders/1T-HDNR7xrmhj10Ti5rIhwz-GQRkrVdRP?usp=sharing>
- “output_real.R” contains all codes to reproduce the results of the real data application in the paper, once the models have been estimated. Note that, due to the limited RAM availability of the PCs used for the analysis, the outputs are produced analysing one RData at a time. The RData files which are analysed in this script are available in the Drive folder “outputDolph”, as already mentioned above;
- “realData_modelchoice.R” contains the code for reproducing Table 3. The resulting output is contained in “realDataResults.RData”.

Simulation experiment:

- “simstudy_setup.R” contains the details of the simulation experiment;
- “simstudy_RPT.R” and “simstudy_JStyle.R” are used to fit the candidate models on the simulated datasets. Since the model fitting via MCMC is computationally intensive (indeed, we suggest exploiting a High-Performance Computing and parallelization), we make the corresponding RData files available at the following link (folder: “outputSim”):
<https://drive.google.com/drive/folders/1FBX0FPFgYqVTaNGU20k9Bp8EwVVat0h6?usp=sharing>
- “output_sim.R” contains all codes to reproduce the results of the simulation study in the paper (Table 1,2 and Figure 2), once all models have been estimated. The RData files which are analysed in this script are available in the Drive folder “outputSim”, as already mentioned above. **In case you cannot download the RData files for the simulation study (due to large file size – it may take a lot of time indeed), please load the R workspace “simResults.RData”;**
- “simResults.RData” is the workspace which contains the results of “output_sim.R”, after the very large objects have been deleted; you can find the result of Table 1, Table 2 and Figure 2.