Local population projections with Bayesian hierarchical models

https://github.com/violetacln/SIPP

Violeta Calian

violeta.calian@hagstofa.is

Statistics Iceland

Why (local/total) population projections?

- One of the key products of Official Statistics← huge demand, from:
  - National and local administration
  - Policy makers
  - Planning infrastructure: constructions, heating, water and electricity supply
  - Planning future developments in social sectors: health, pensions, education
  - Researchers in multiple domains
  - Individual users

For all demographic components (*migration, fertility, mortality)* as functions of: age, time and/or gender, citizenship, education, locality, …,
in order to be combined *stochastically, via ccpm, into:* **population projections** (*total and local*)

- ***Formulation of the statistical problem****:*
- *estimating and predicting* in the presence of:
  - small area/population issues and/or rare events/ shocks
  - expert assumptions
  - uncertainty of model parameters
  - errors in input data (e.g. *deregistration*)
- while reporting *uncertainty* of estimates/predictions

- ***Solution must be based on***:
- Modeling and combining (prior) information

New Models and Implementation in R

$P(response|parameters\_0 \vee fcts) = F1(attributes\_0)$ *where*
$P(parameters\_0 \vee fcts|parameters\_1) = F2(attributes\_1)$ *where…*

- Type of models: **Hierarchical / Multilevel**, Bayesian / frequentist

  - data used: *individual or aggregated* response (Y) versions

  - what to model: ***processes,*** since *time series* data $P(Y_i, \ldots Y_n|a)$ plus other grouping effects -->

    need to account for complex ***(auto-) correlations***

- Why:

  - {components/prior information/characteristics/local&total} & **uncertainty** report

- ***R***-Tools:

  several *Stan*-running R packages like ***brms***, arm, bayestestR, …

  and the faster: *lme4, **mgcv**, …* for frequentist / as initializations / *not only* (see Laplace approx of mgcv)

Structure and novelty

x=(year, age, gender, citizenship, municipality, education, …)

Response = rates/events of local/total demographic components

I. Counts/rates

- $N(x, \dots) \sim Poisson(e^{f(x,\dots)}, N_0(x, \dots))$ (or: neg. binom, see over-disperssion issues, etc)
- $f = \sum f_m(x, \dots)$, $f_m \sim GP(\, m(x, \dots), C(\{x, \dots\}, \{x', \dots\})\,)$ + …

II. Microdata (events)

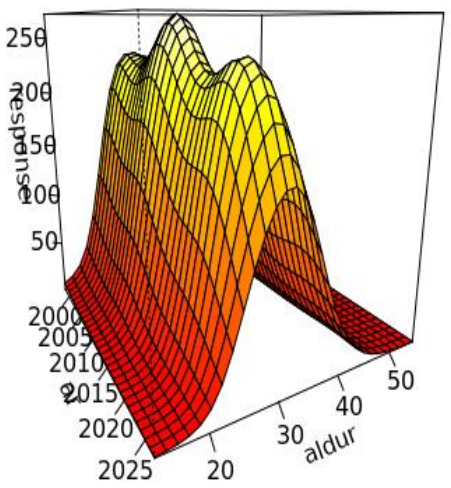
- $N(x, \dots) \sim B\,(L(f(x, \dots), N_0(x, \dots))$
- $f = \sum f_m(x, \dots)$, $f_m \sim GP(\, m(x, \dots), C(\{x, \dots\}, \{x', \dots\})\,)$ + …

- f_m – smooths/re/t2 and/or by {grouping factors} OR unknown functions with Gaussian Processes as priors (with m -> long term behaviour, C -> correl between any two response values)

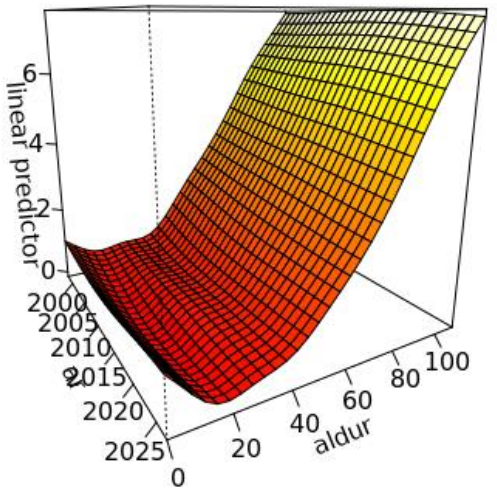
Count data input

*https://cran.r-project.org/web/packages/**mgcv**/index.html*

Births

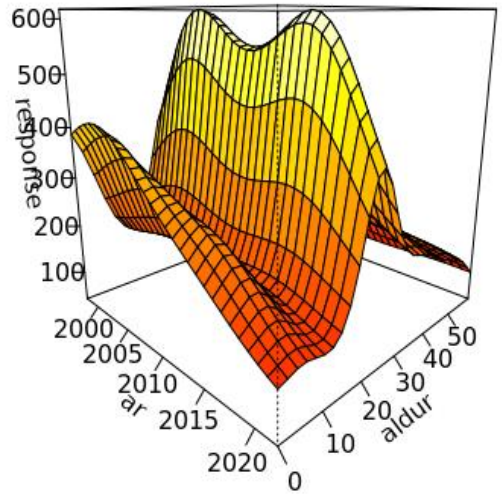


Death rates (log)

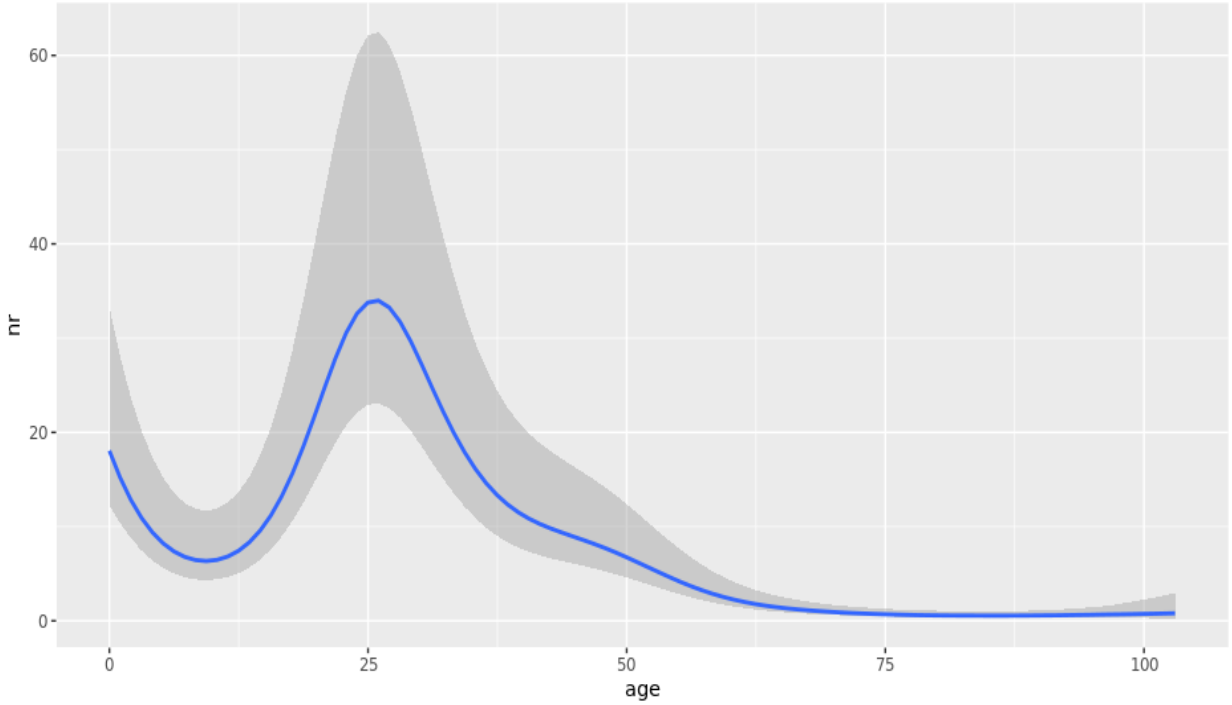


Count data input: migration

*https://cran.r-project.org/web/packages/**mgcv**/index.html*

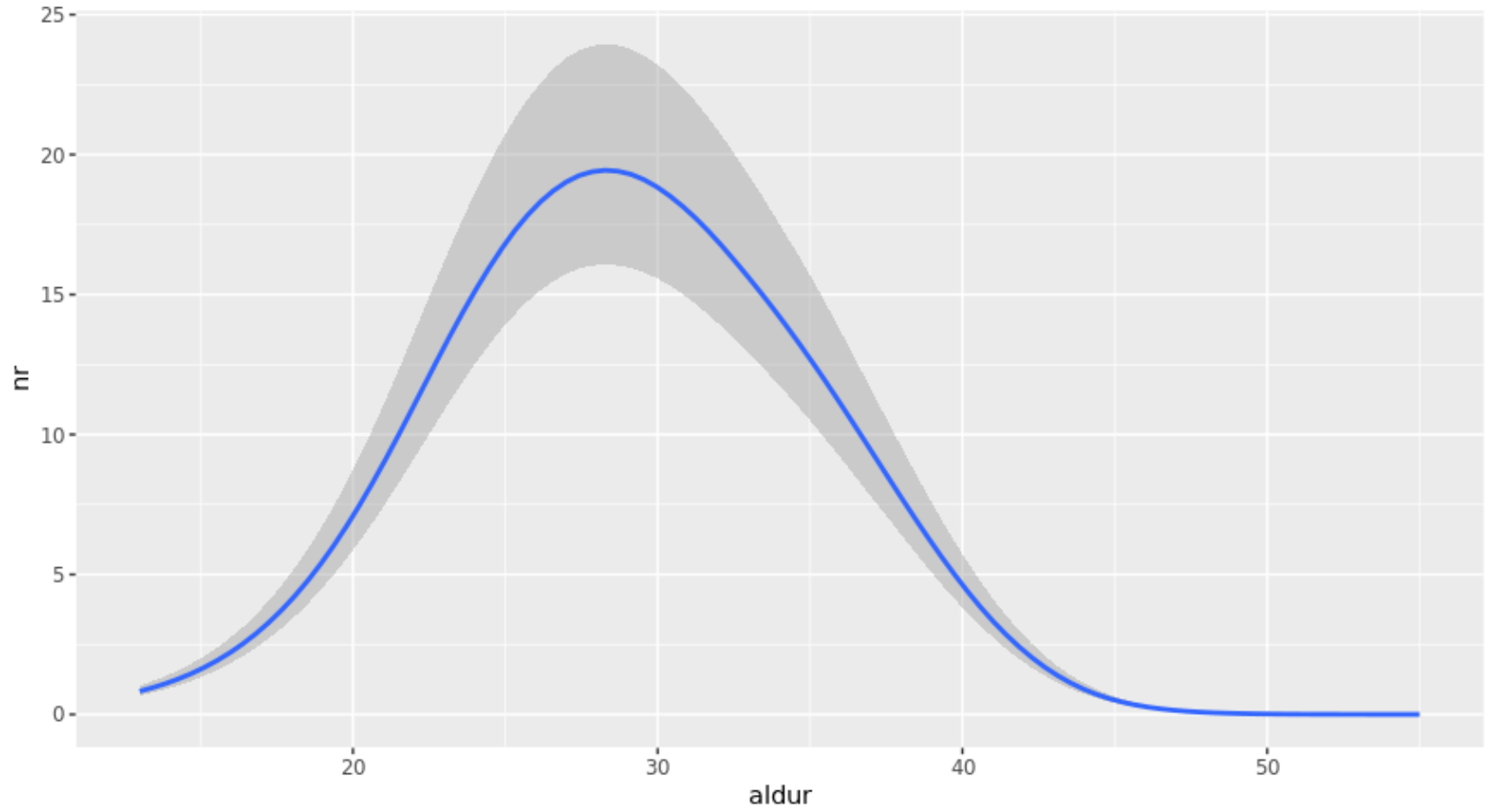


*https://cran.r-project.org/web/packages/**brms**/index.html*



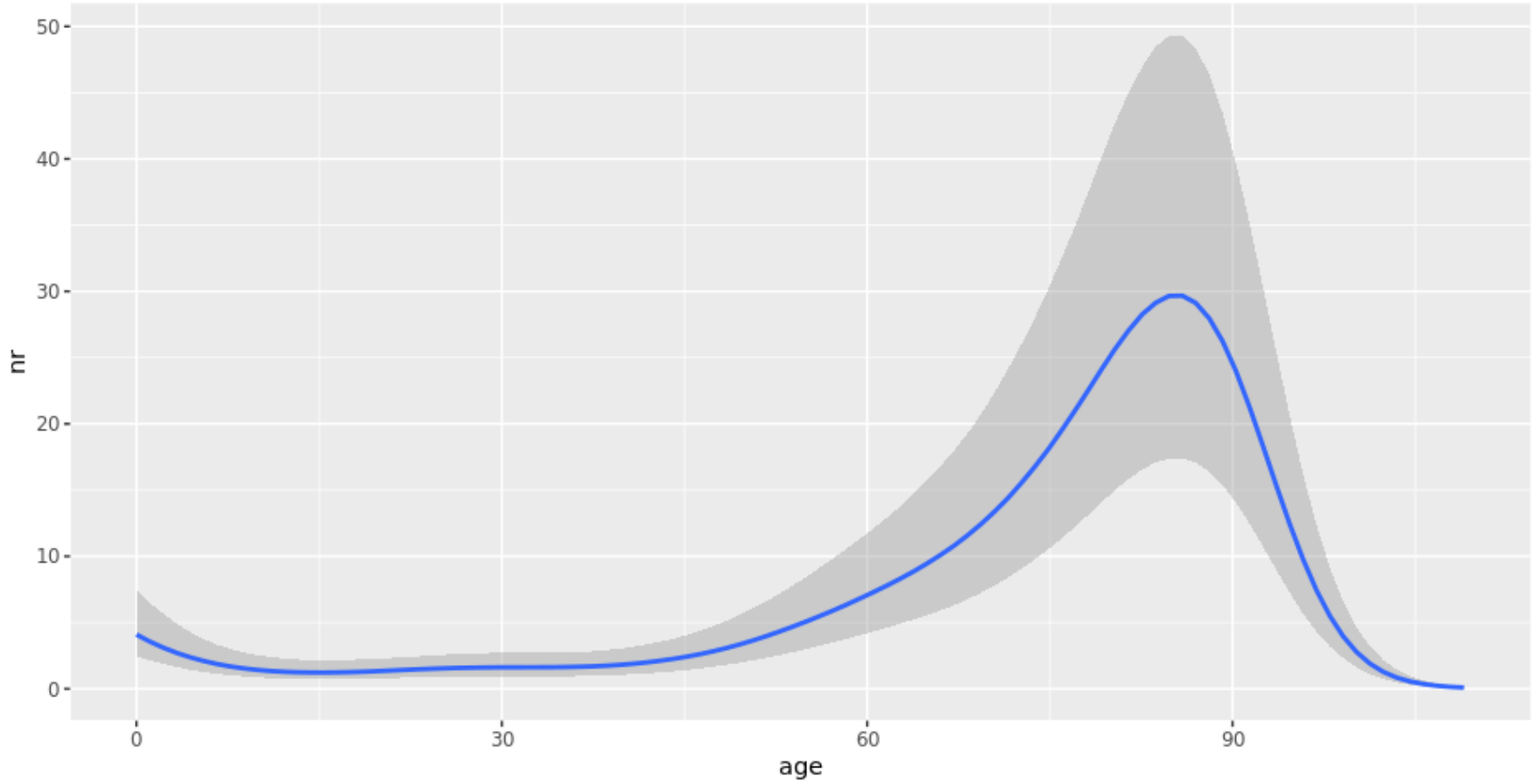
Counts/rates: fertility models

*https://cran.r-project.org/web/packages/**brms**/index.html*



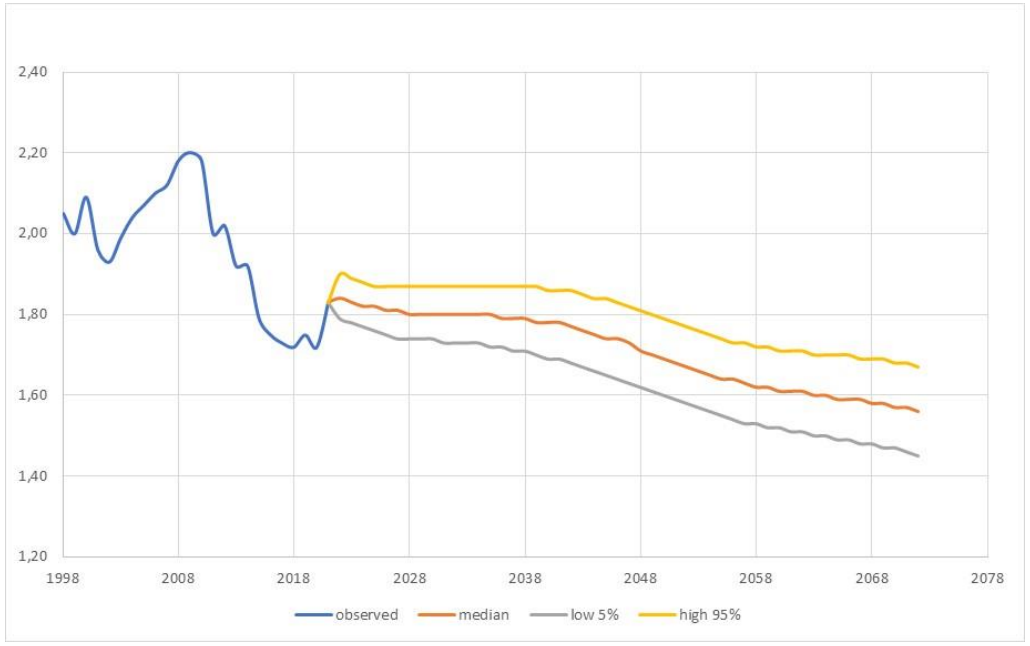
Counts/rates: mortality models (addit, s)

*https://cran.r-project.org/web/packages/**brms**/index.html*

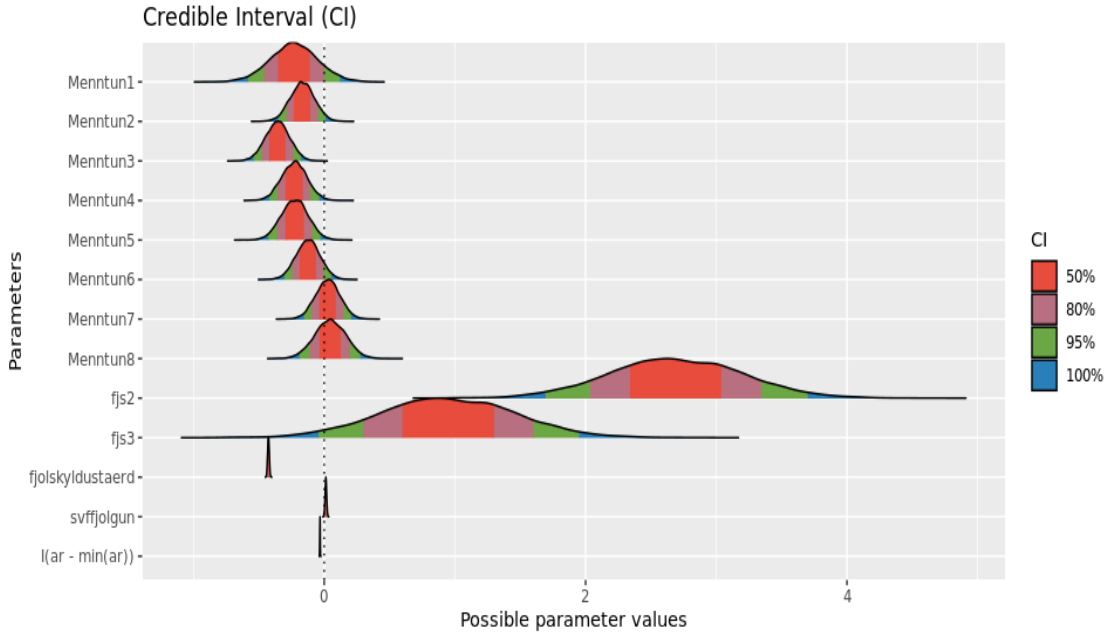


Examples: demographic *rates*‘ projections and significance of *attributes*‘ effects (using microdata)

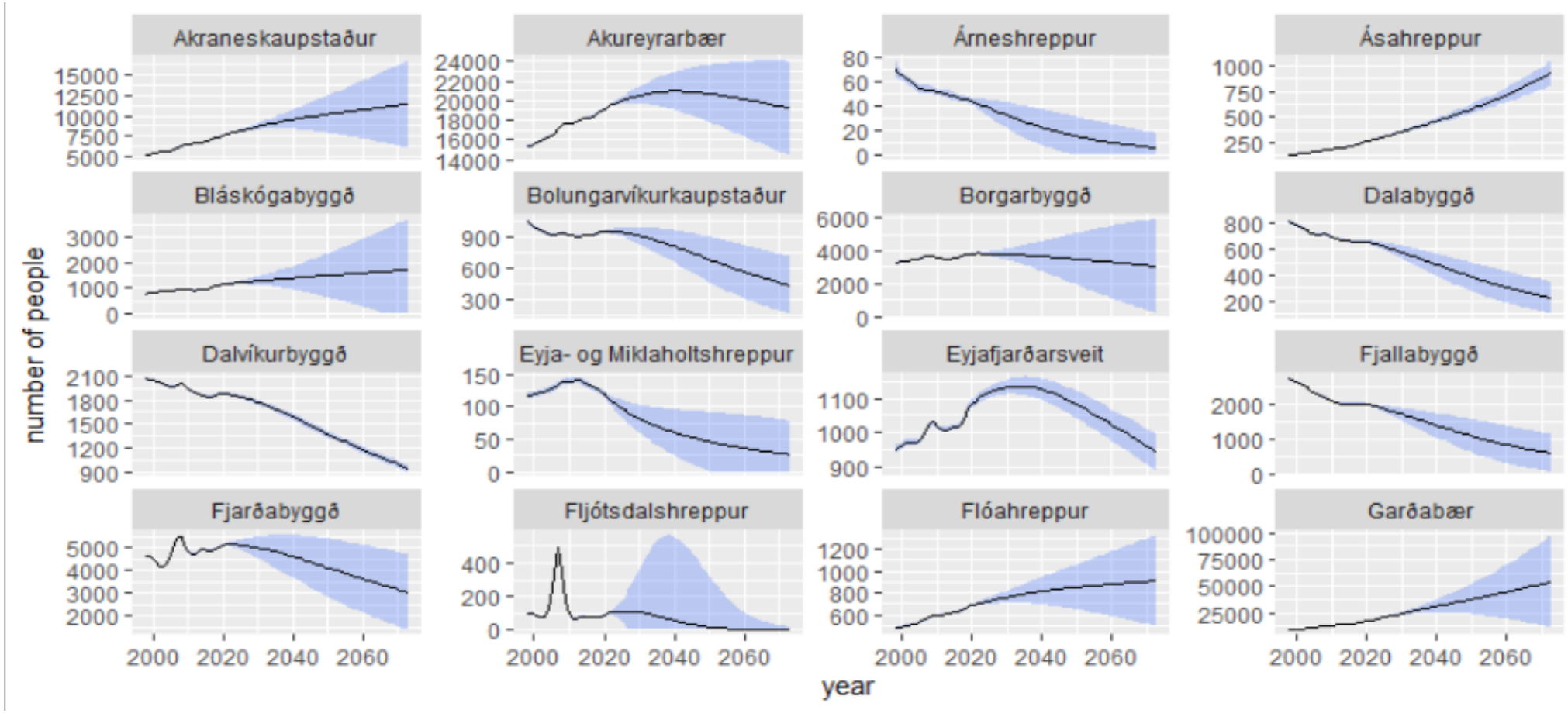
Total fertility rates: median and credible intervals



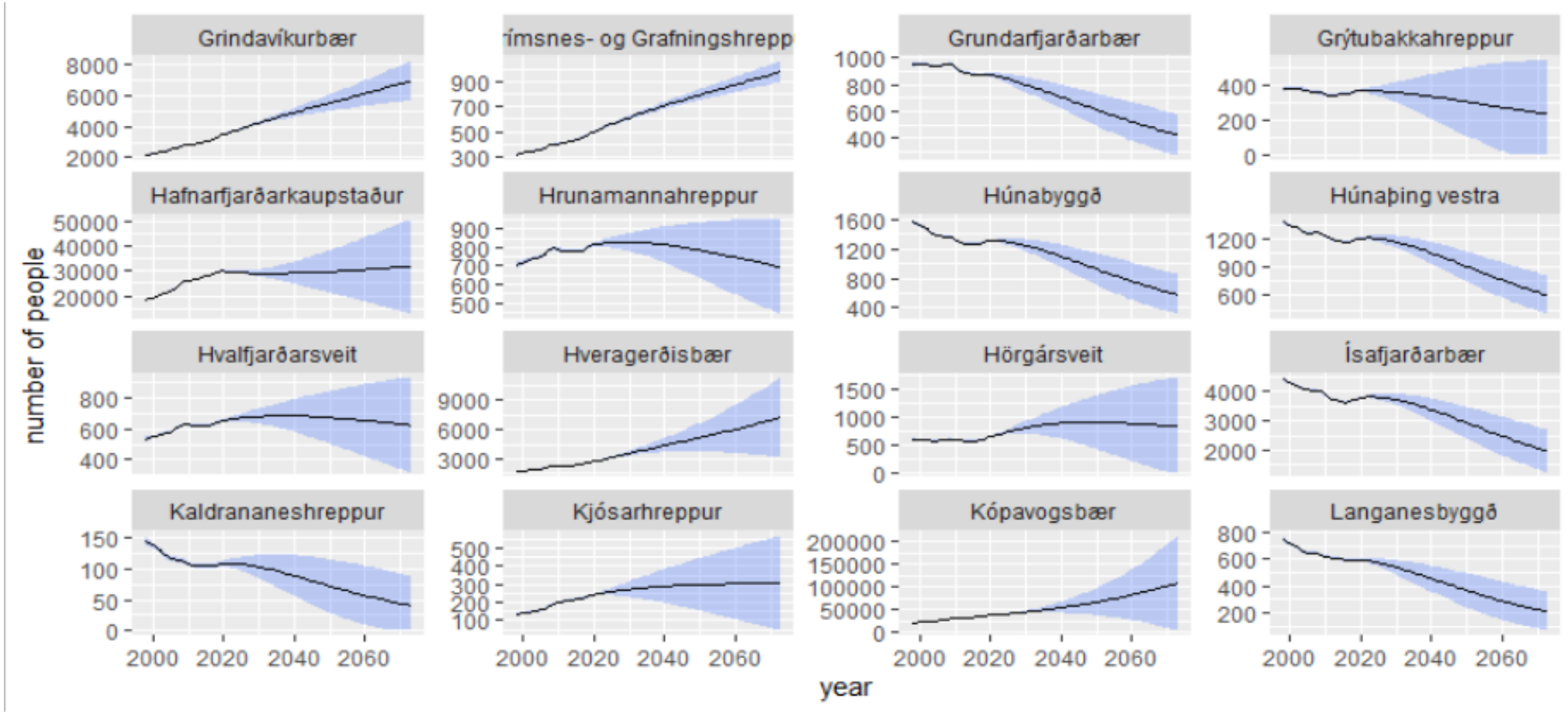
Effects of (individual, family, municipality) attributes on fertility (events)



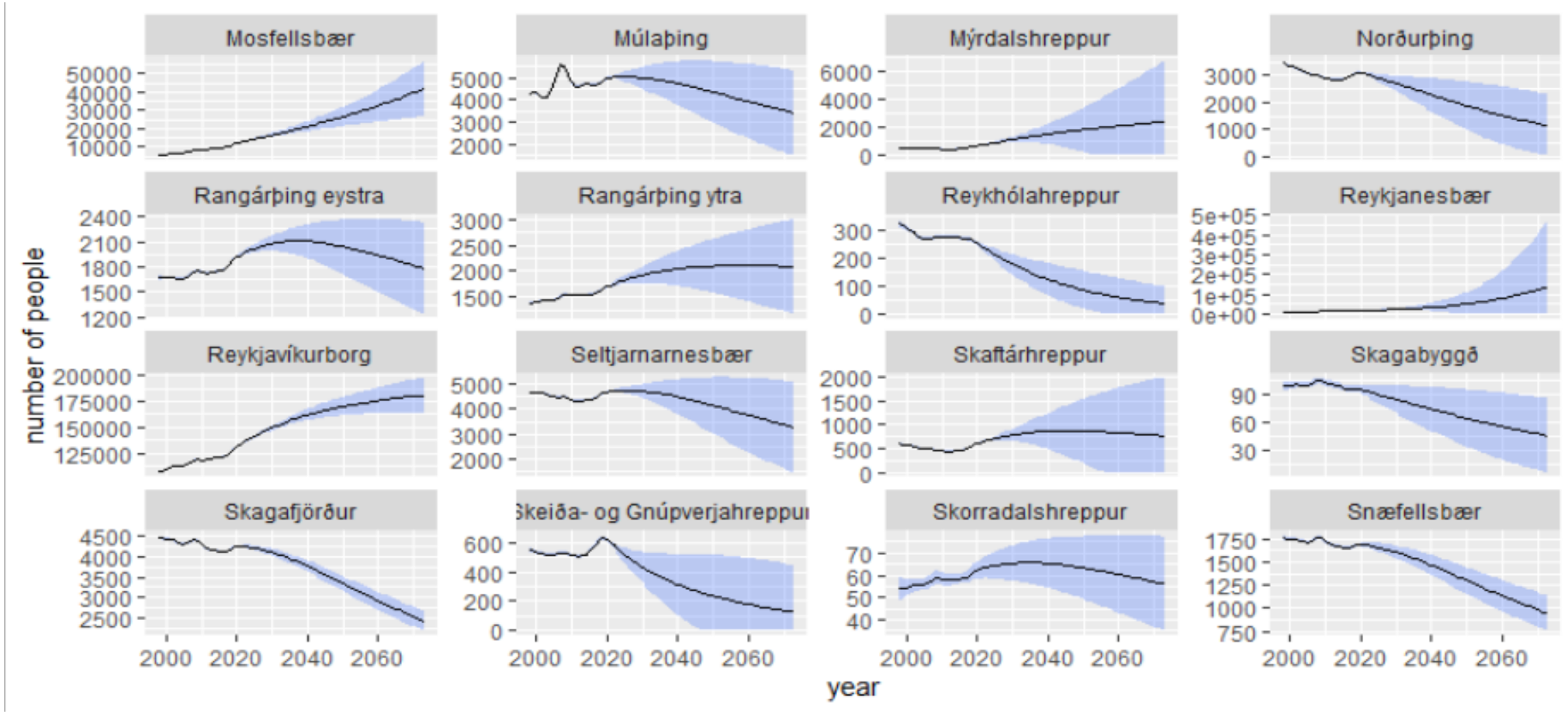
Local population projections



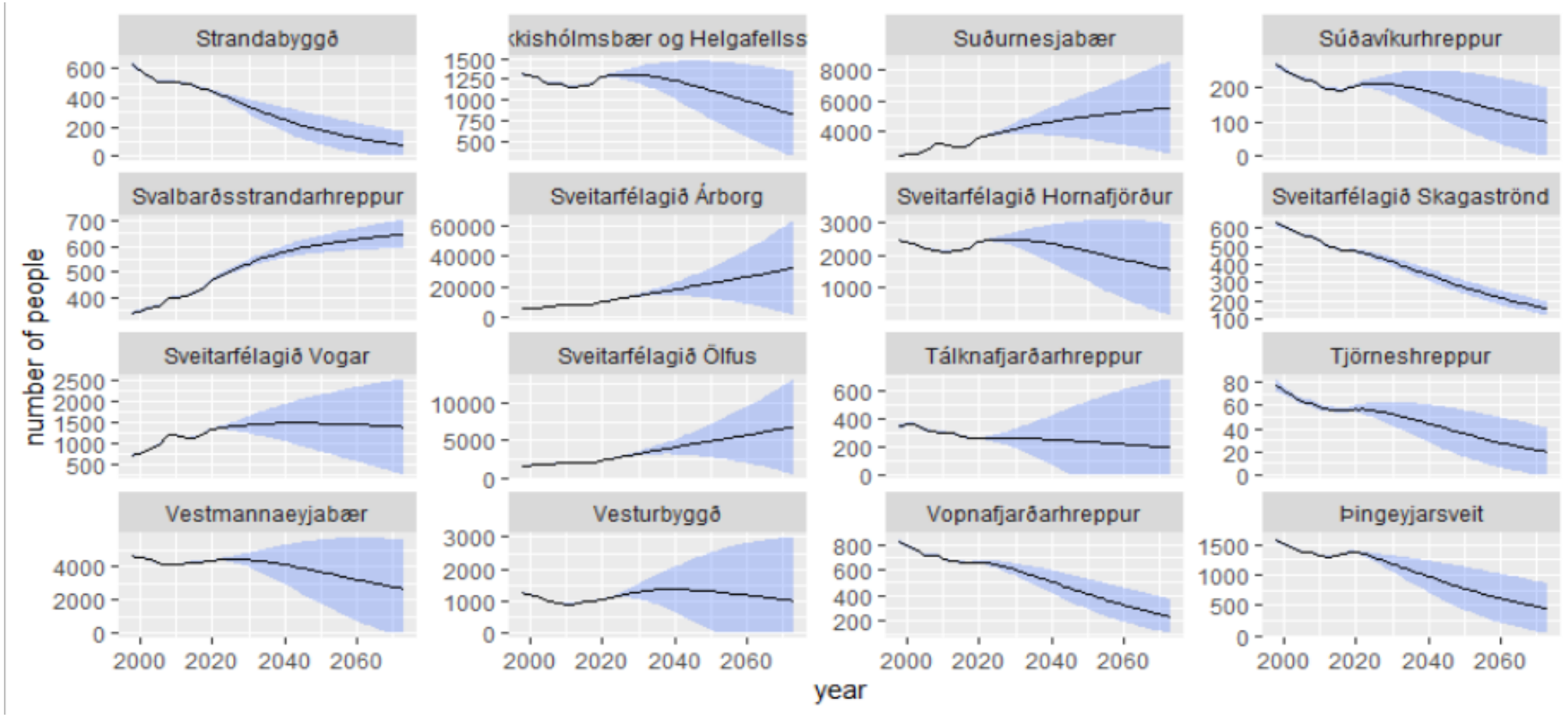
Local population projections



Local population projections



Local population projections



Future work and work in progress

- compare local population projections based on non- and informative *priors* (with users/experts‘ feedback)
- finalise the evaluation of the *model performance*
- forecasting with *multiple-model averaging*

References

- Jakub Bijak & John Bryant (2016) Bayesian demography 250 years after Bayes, Population Studies, 70:1, 1-19, DOI: 10.1080/00324728.2015.1122826
- https://cran.r-project.org/web/packages/bsts/index.html
- Wood S (2017). *Generalized Additive Models: An Introduction with R*, 2 edition. Chapman and Hall/CRC.
- https://cran.r-project.org/web/packages/mgcv/index.html
- Bürkner P (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software*, **80**(1), 1–28. doi: 10.18637/jss.v080.i01
- Bürkner P (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms." *The R Journal*, **10**(1), 395–411. doi: 10.32614/RJ-2018-017
- Bryant, J., & Zhang, J. L. (2016). Bayesian forecasting of demographic rates for small areas: emigration rates by age, sex and region in New Zealand, 2014-2038. *Statistica Sinica*, *26*(4), 1337–1363. http://www.jstor.org/stable/44114337
- Lam, K.K.; Wang, B. (2021). Robust Non-Parametric Mortality and Fertility Modelling and Forecasting: Gaussian Process Regression Approaches. *Forecasting* 2021, *3*, 207-227. https://doi.org/10.3390/forecast3010013
- Ludkovski, M., Risk, J., Zail, H. Gaussian process models for mortality rates and improvement factors. ASTIN Bulletin: The Journal of the IAA , Volume 48 , Issue 3 , September 2018 , pp. 1307 – 1347, https://doi.org/10.1017/asb.2018.24
- Stan Development Team (2022). "RStan: the R interface to Stan." R package version 2.21.7, https://mc-stan.org/.