Population projections with Bayesian hierarchical models

Violeta Calian

violeta.calian@hagstofa.is

https://github.com/violetacln

Statistics Iceland

Nordic Demographic Symposium, June 10-12, 2025

Outline

Results!

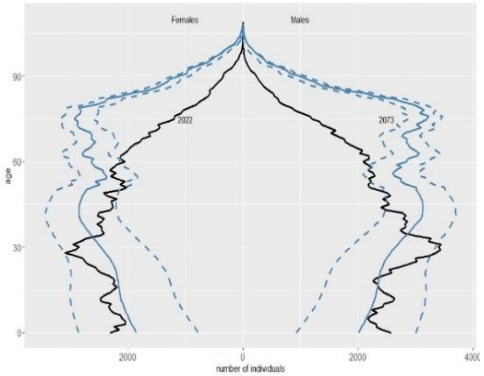
Formulation of the problem(s)

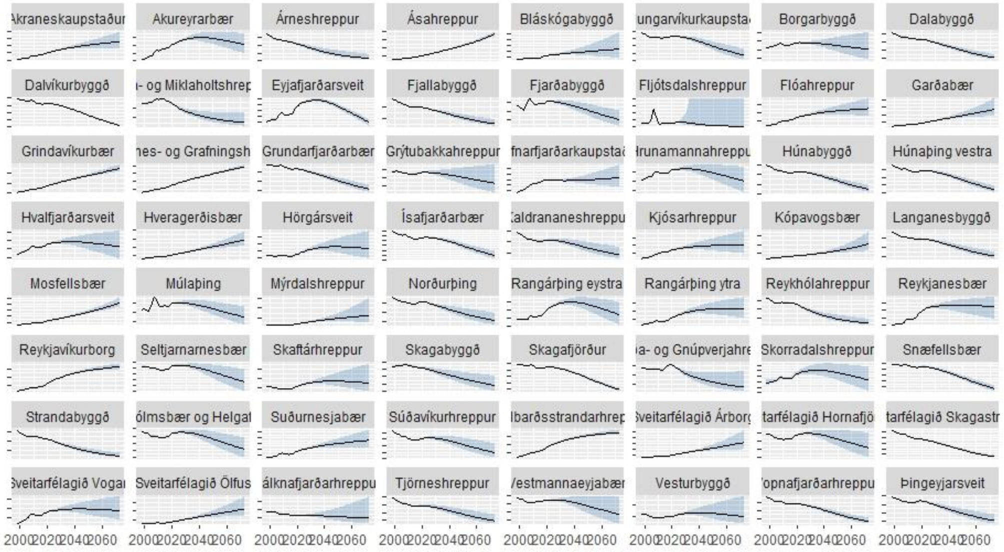
Bayesian solution

Applications

Future work & new projects!

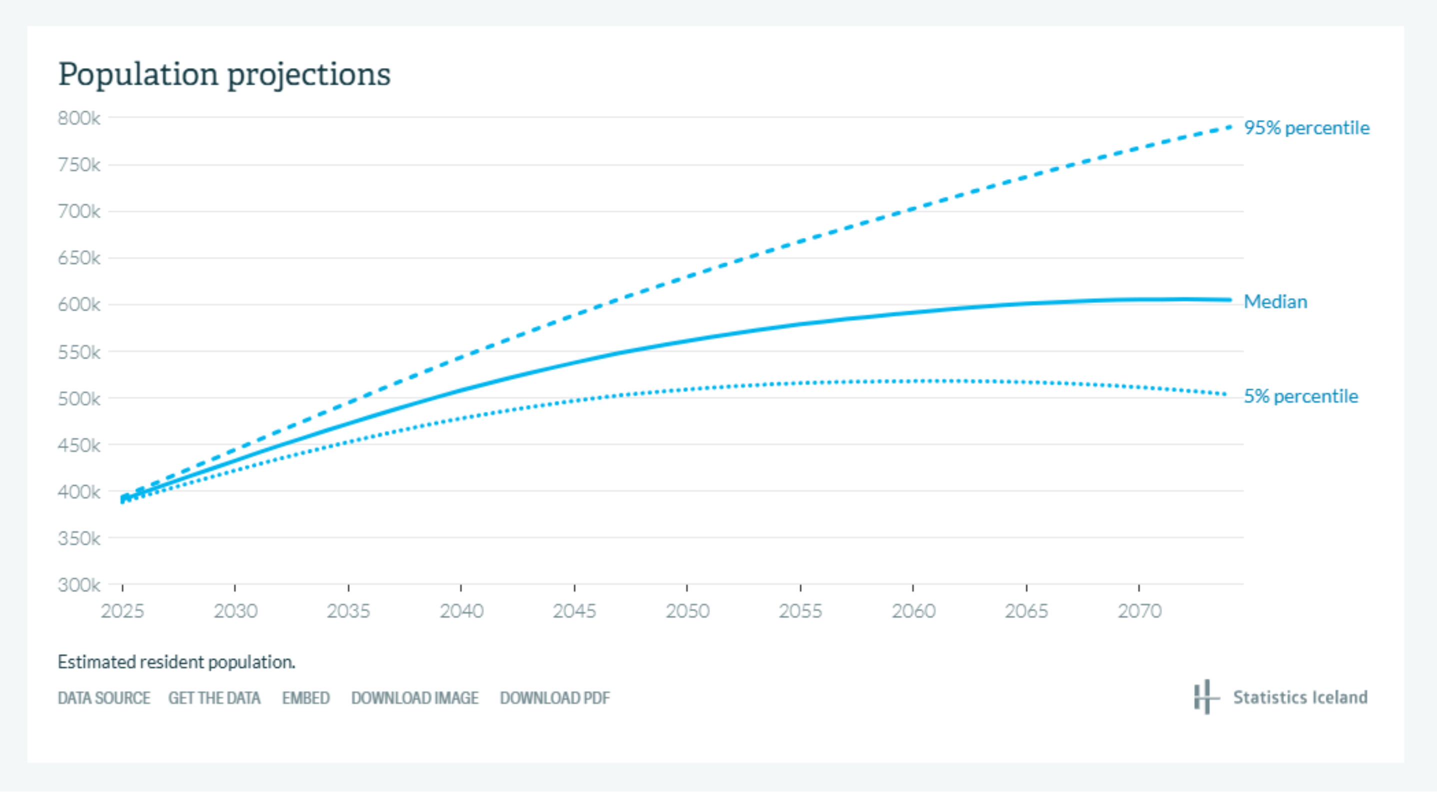
Conclusions





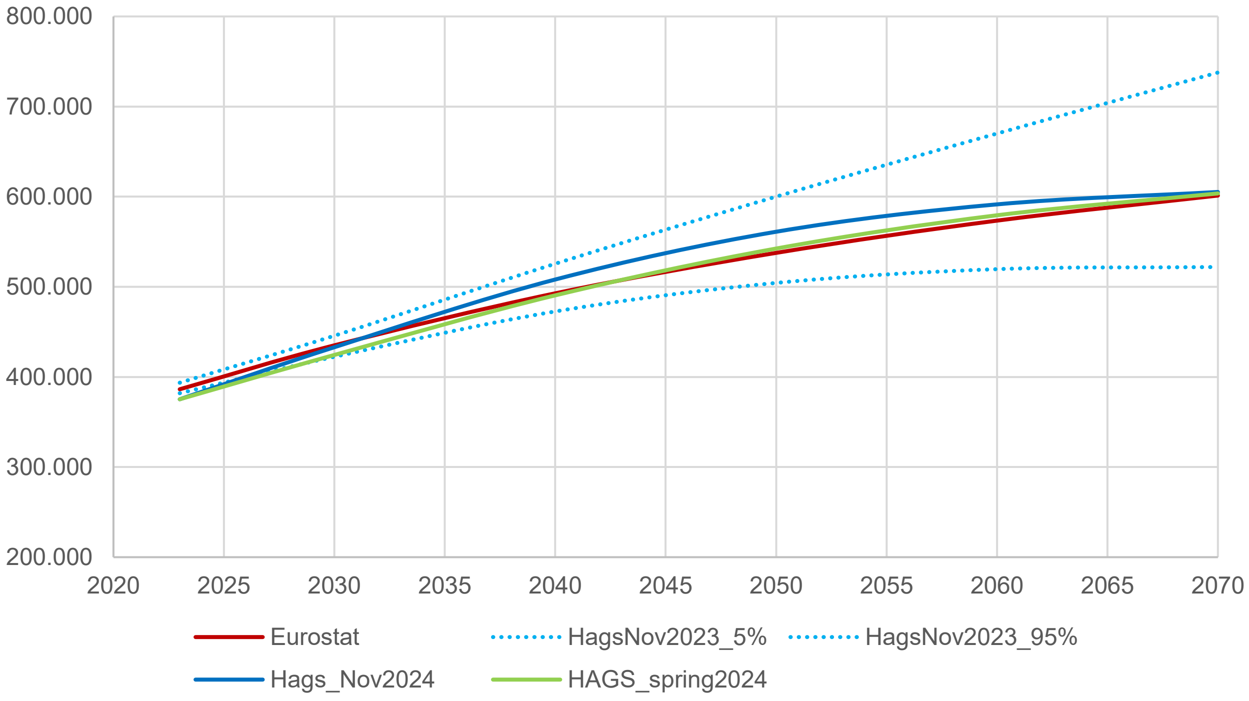
Experimental statistics: Population by municipality 1998-2073

Most recent population projection



Simple comparisons and causes of changes

Eurostat prediction include: constrained migration flows



Formulation of the general problem

All **demographic components** (*migration, fertility, mortality)*

as functions of: age, time and/or gender, POB, education, location, … →

To model and predict while taking into consideration:

- small area/population issues and/or rare events/ shocks
- expert assumptions
- uncertainty of model parameters
- errors in *input* data* (e.g. *deregistration*)
- (missing data)

and *reporting uncertainty* of estimates/predictions,

in order to be combined *via a **stochastic** version of the **cohort component** pop.proj.meth.*

Complex issues influencing modeling and forecast:

- *measurement errors (input data)* & estimates of past counts

- prior information & interpretability (hyperparam)

- correlations between (lagged age/time):

  im/e-migration

- LE of migrants, evolution of 'older' return rates

- (user driven!) **conditional forecasting** & multivariate-dynamics effects for components/local/total population projections

Simple view on Bayesian modeling and forecasting

Posterior_distribution ~

Prior_distribution * *Likelihood_data*

| |

| *greater role in model fitting*

greater role in forecasting

Hierarchical models for all demographic components)

Response:

count data of demographic events

Predictors:

$x \in$ (*year, age, (location), gender, place of birth, (municipality attributes), education, …*)

N_0 - *„exposed" population*

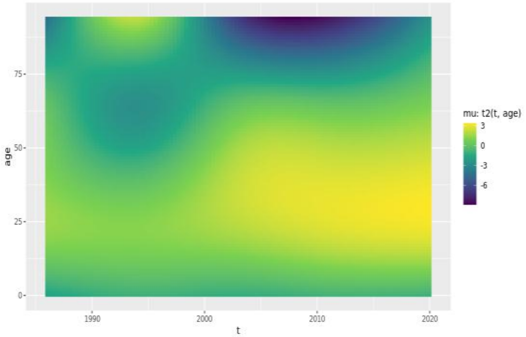
***Note**:*

simplest local projections as experimental statistics

```
## using brms ---
m_svf_brms <- brms::brm(nr ~  gp(ar, by=svf) + (1|svf)
                             + offset(log_nrtot)
                             , family=poisson(), data=pop_loc)
```

(*https://github.com/violetacln/SIPP/blob/main/models_Bayesian_chosen.R*)

HGAM& R-tools



$$response \sim \boldsymbol{P}(f(..(f(\eta)))$$

$$\eta = \beta \boldsymbol{x} + \gamma \boldsymbol{z} + \sum s_k(a_k, by = c) + \ \ldots$$

		smooth dependencies (including *GP*)
	group level (cross-/ clustering) attributes and effects
population level attributes and effects

or **unknown form/shape**

GP : (as prior probability distribution over functions) :: connection between Machine Learning and Statistical Modeling :: it learns from data the *strength* of the relation between observations (not the shape)

Implementation: ***R-tools*** **(*brms, mvgam* and *mgcv* packages)**

can include: *me*(…), *mi*(…), mi & me, mo(…), cs(…), autocor(…)

posterior distribution checks, LOO-validation, …

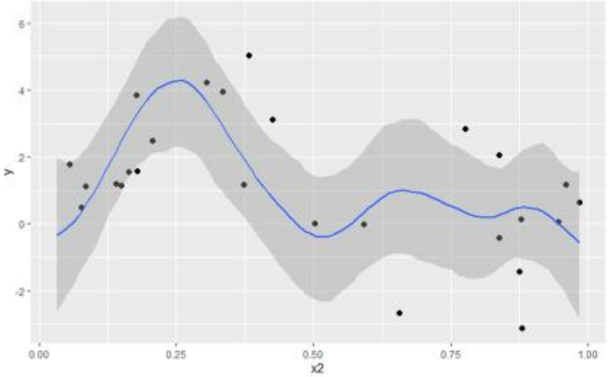
Experimenting with ***H***GAMs

→ not/ sharing

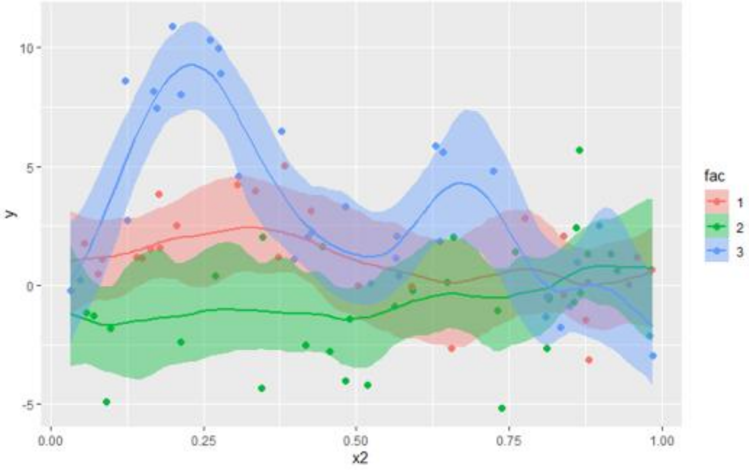
smoothness/trend s(…)

structure of gp(…)

across groups




```
fit5 <- brm(y ~ gp(x2) +fac, dat2, chains = 2)
summary(fit5)
plot(conditional_effects(fit5), points = TRUE)
```




```
# fit separate gaussian processes for different levels of 'fac'
fit4 <- brm(y ~ gp(x2, by = fac), dat2, chains = 2)
summary(fit4)
plot(conditional_effects(fit4), points = TRUE)
```

https://paul-buerkner.github.io/brms/reference/gp.html

Special component, GP

Advantages of modelling with Gaussian Processes (GP):

- rich structure of the covariance function capturing
quasi-periodicity, non-stationarity, complex interactions
& trend, periodicity, seasonality, autoregressive character
that affect the model predictions both in interpolation and extrapolation

- *good forecasting* performance

- *related/equivalent* to*:* state space models, kriging, Bayes-Neural Networks, more

The *stochastic cohort component method*

- demographic rates are sampled from their posterior distributions at each time step $(t+1)$ by using: the models *and* the exposed population numbers at time step t

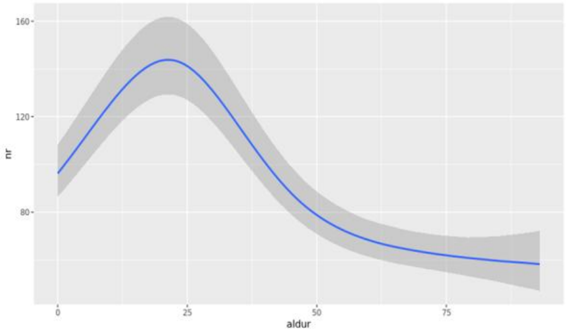
- these samples are combined according to the standard balance equation

$P_{t+1} = P_t + M_{t,t+1} - D_{t,t+1} + B_{t,t+1}$, written for all extra-dimensions* (e.g. age, gender, POB)

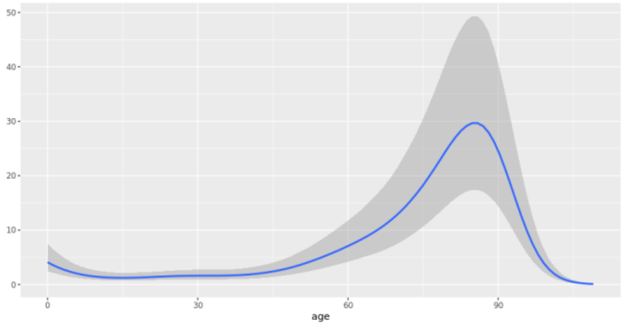
- the process is repeated a large number of times (e.g. K=10,000) in order to generate the posterior distributions of $N(t, x, \ldots)$ of $t > t_0$ where t_0 is the starting time of projections.

Demographic components

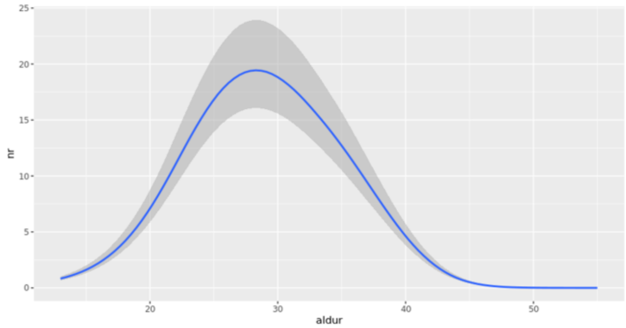
Migration
(conditional effect, brms)



Mortality
(conditional effect, brms)



Fertility
(conditional effect, brms)



Discussion / To do!

- Testing and validation for the *population* projections (not only of demographic *components*)
- Performing sensitivity analyses for testing the effect of changes in prior distributions
- Length of time series / length of forecasting horizon
- Improve the m. learning of the correlation im/e-migration

- Incorporating new significant attributes in the production version, such as education
- Building a unique model for local and total population projections
- Testing the use of Markov Random Fields for modelling local effects more accurately?
- Adding more municipality-level attributes into the model and **prior information** (such as building plans)

- Improve and share the code https://github.com/violetacln/SIPP/

References

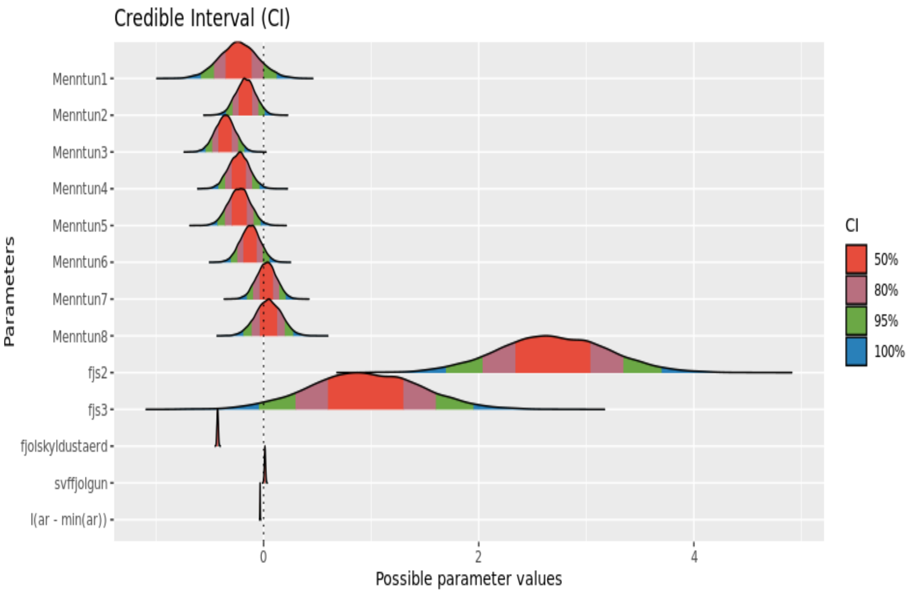
- Gelman, A. et al. Bayesian data analysis. (2021) https://stat.columbia.edu/~gelman/book/BDA3.pdf
- Wood, S.N. (2017). Generalized Additive Models: An Introduction with R, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279
- https://cran.r-project.org/web/packages/mgcv/index.html
- Rasmussen, C. E., Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. United Kingdom: MIT Press.
- 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
- Stan Development Team (2022). "RStan: the R interface to Stan." R package version 2.21.7, https://mc-stan.org/.
- Jakub Bijak & John Bryant (2016) Bayesian demography 250 years after Bayes, Population Studies, 70:1, 1-19, DOI: 10.1080/00324728.2015.1122826Wood S (2017). *Generalized Additive Models: An Introduction with R*, 2 edition. Chapman and Hall/CRC.
- 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
- 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
- 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
- Simpson, G.L. (2018) Modelling palaeoecological time series using generalized additive models. *Frontiers in Ecology & Evolution* 6;149.
- Pedersen EJ, Miller DL, Simpson GL, Ross N. 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* 7:e6876 https://doi.org/10.7717/peerj.6876
- Piiwronen A., Piironen J, Laaksonen T. (2022) Predicting spatio-temporal distributions of migratory populations using Gaussian process modelling. Journal of Applied Ecology, https://doi.org/10.1111/1365-2664.14127

Appendices

Appendix: micro and aggregated data

Effects of fertility characteristics

(events; *lme4, arm, bayestestR*)

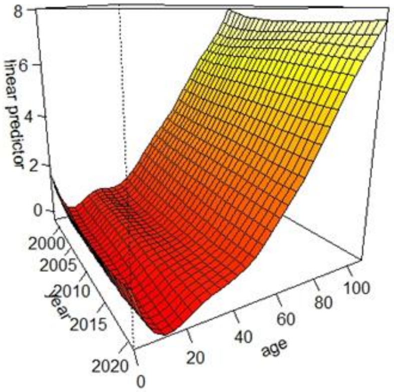


Number of births, by time, age, POB

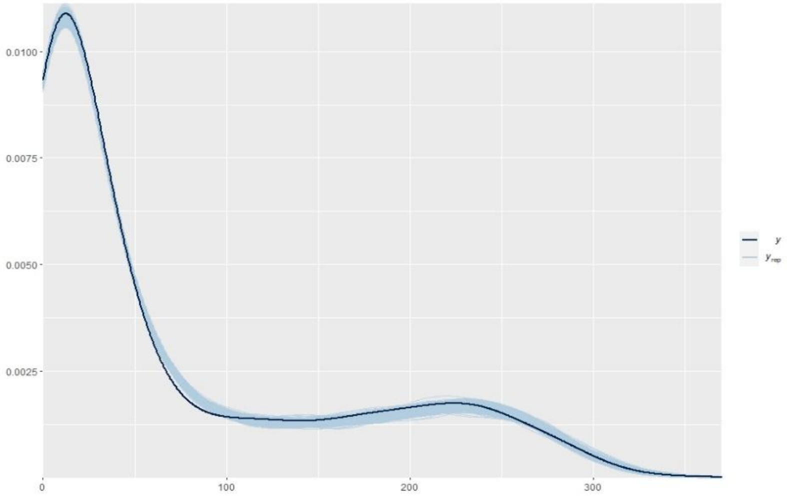
(*mgcv*)

Appendix: different implementations

Mortality rates (*mgcv*)

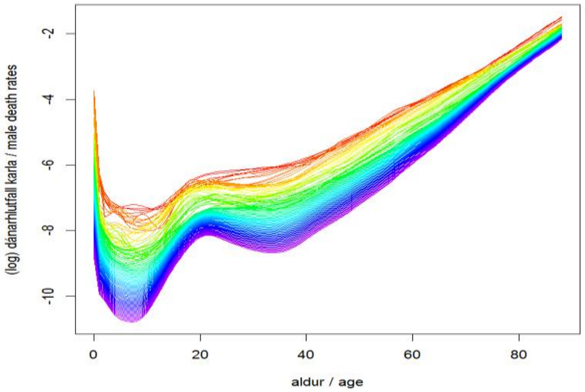


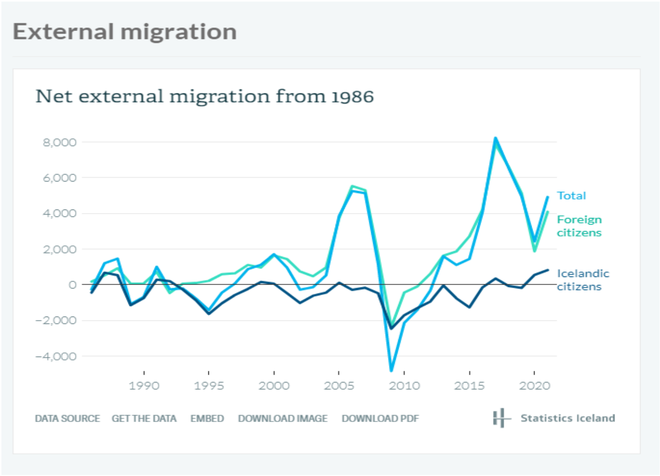
Posterior checks (*brms*)– example

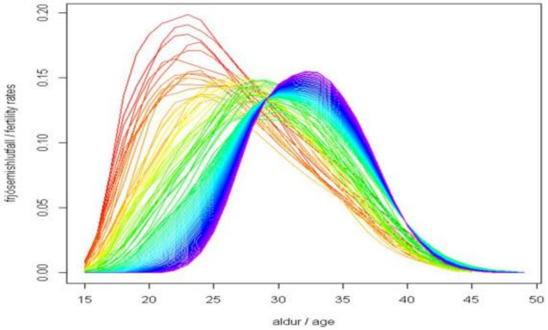


Appendix: past methodology

Functional models (fertility, mortality) [Hyndman]
Plus ARDL models (short term migration) +
Long term migration (model+assumptions)







Thank you!