AI in longevity risk management: improved long-term projections by machine learning

Ronald Richman

Peter Vekas

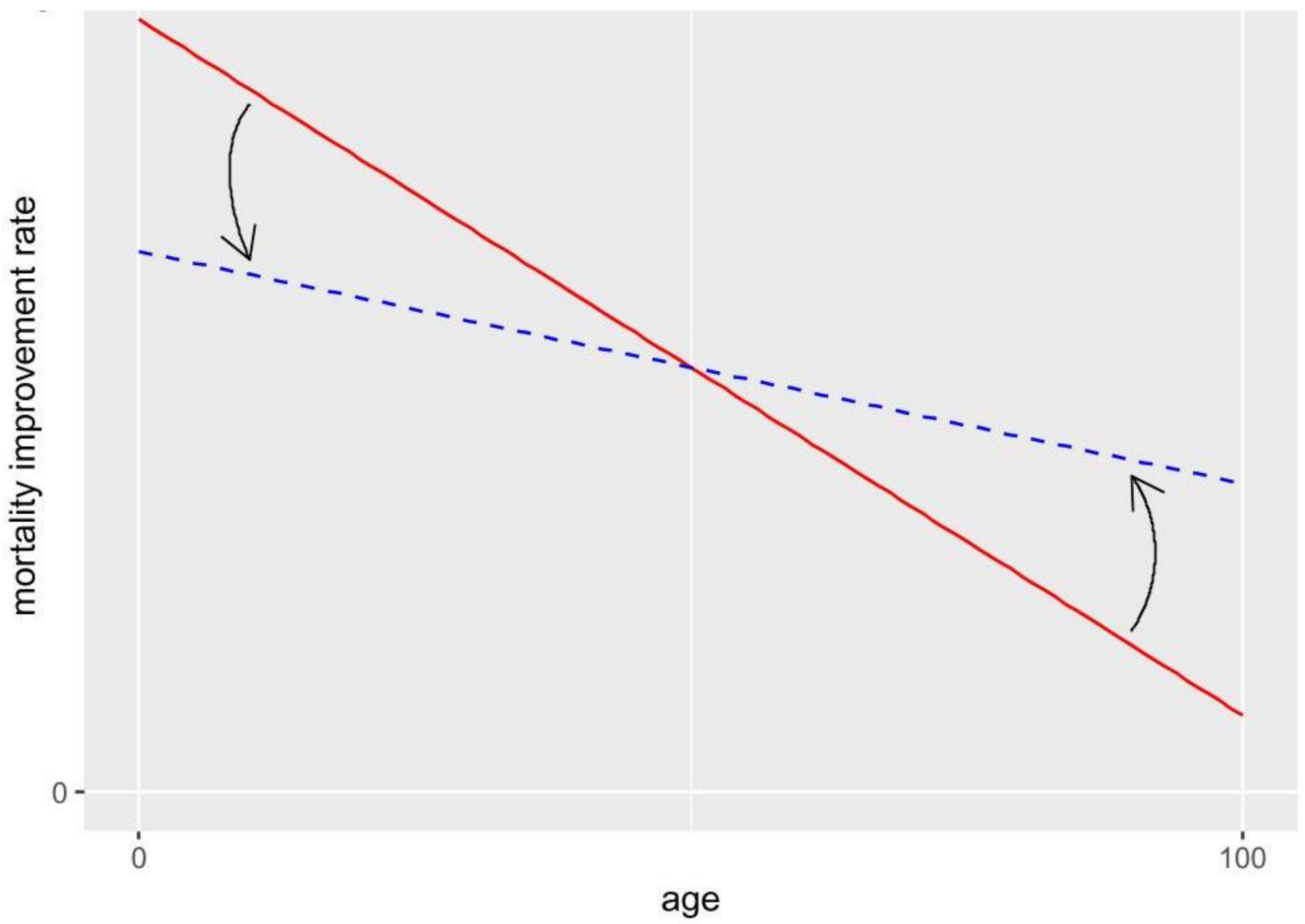
Laszlo Kovacs

DAV/DGVFM Jahrestagung 2022

Agenda

Introduction

- **Several authors have noticed that the speed of mortality decline tends to slow down in younger ages and speed up in older ages in the long run.**
- **Li, Lee and Gerland (2013) call this the 'rotation' of the age pattern of mortality decline.**
- **Possible reasons:**
    - less room left for spectacular advances in preventing child mortality (e.g. due to vaccination),
    - improved costly medical technology to extend life at older ages.



Literature overview

- Kannisto et al. (1994) find accelerating mortality improvements in ages 80 to 99 years between 1950 and 1989 in 27 countries.

- Horiuchi and Wilmoth (1995) argue for a shift in mortality decline from younger towards older ages in Sweden.

- Lee and Miller (2001) examine rotation by comparing the average rates of mortality decline by age in the first and second halves of the 20th century.

- Carter and Prskawetz (2001) estimate Lee–Carter models on Austrian data using sliding time windows to demonstrate the rotation.

- Rau et al. (2008) and Christensen et al. (2009) note that mortality decline in ages 80 years or older has accelerated since 1950 in some countries out of 30.

- Vékás (2020) proposes a data-driven measure of rotation and finds evidence for rotation since 1950 in several countries of the European Union.

Practical significance

1. **Differences between rotated and unrotated forecasts may be minor in the short run but highly significant in the long run!**

2. **Ignoring rotation leads to the underestimation of the old-age population and overestimation of the young-age population.**

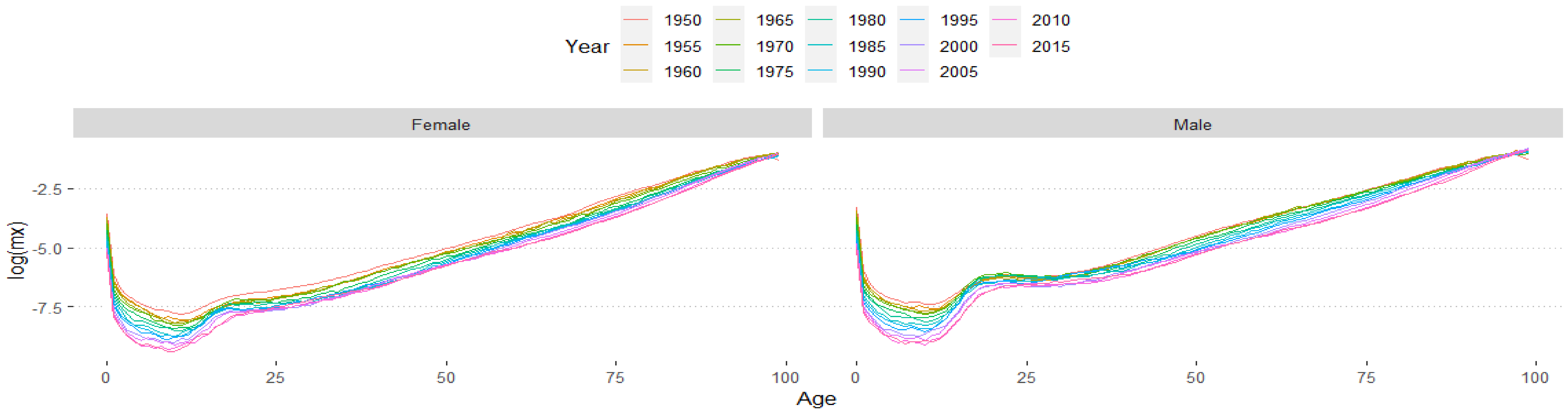
3. **This exacerbates longevity risk in life insurance and pensions valuations, as well as the risk to social security systems.**

4. **In long-term forecasts, it is crucial to assess rotation and to model it appropriately.**

Agenda

Demographic data

- **Human Mortality Database (HMD) provides mortality data for 38 countries**
- **Data downloaded as of February 2020**
  - Used for main work in this talk
  - Subsequent data used for assessing impact of COVID-19
- **Used mx from HMD in the period 1950-2018 (data shown for USA)**
- **Partitioned into training, validation and test sets**
  - Train: 1950 – 1990
  - Validation: 1990 – 1999
  - Test: 2000 - 2018
- **Models fit in two rounds:**
  - Fit on Train and then tested on Validation – hyperparameter tuning
  - Optimal models then fit on Train + Validation and tested on Test set – out of sample performance



Lee−Carter model variant incl. rotation

- **Original ("vanilla") Lee−Carter (1992) model:**

$$\ln m_{xt} = a_x + b_x k_t + \varepsilon_{xt}$$

- **As k_t declines over time, the coefficients b_x determine the rates of improvement by age.**

- **Age-specific improvement rates are assumed to be independent of time!**

- **Rotated variant of Li, Lee and Gerland (2013):**

$$\ln m_{xt} = a_x + B(x,t)k_t + \varepsilon_{xt}$$

- **Improvement rates are weighted means of initial and ultimate values if life expectancy is at least 80 years:**

$$B(x,t) = (1 - w_s(t))b_0(x) + w_s(t)b_u(x)$$

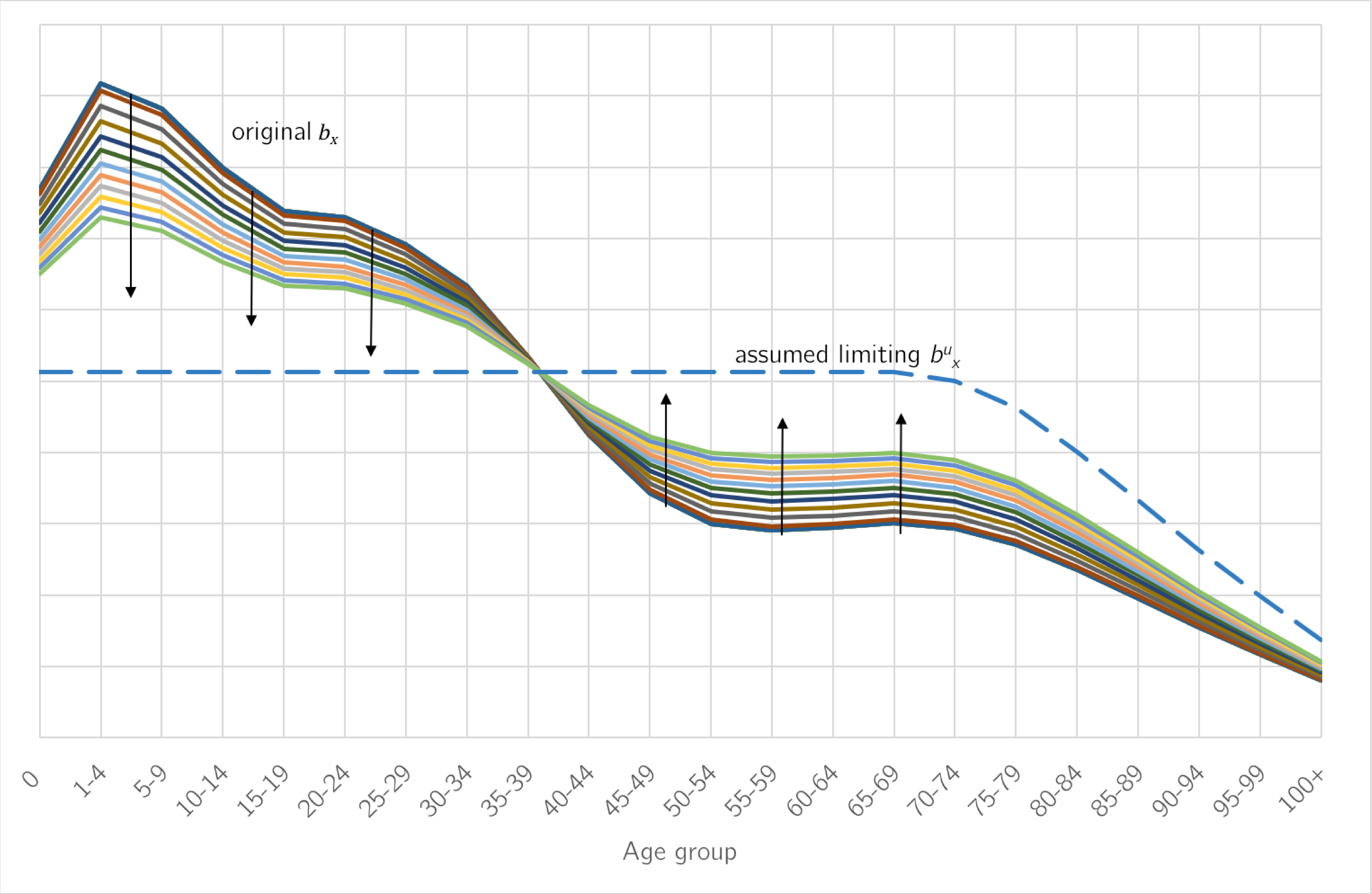
- **"Smooth" weights are computed from "raw" weights:**

$$w_s(t) = \left\{0.5\left[1 + \sin\left[\tfrac{\pi}{2}(2w(t) - 1)\right]\right]\right\}^p$$

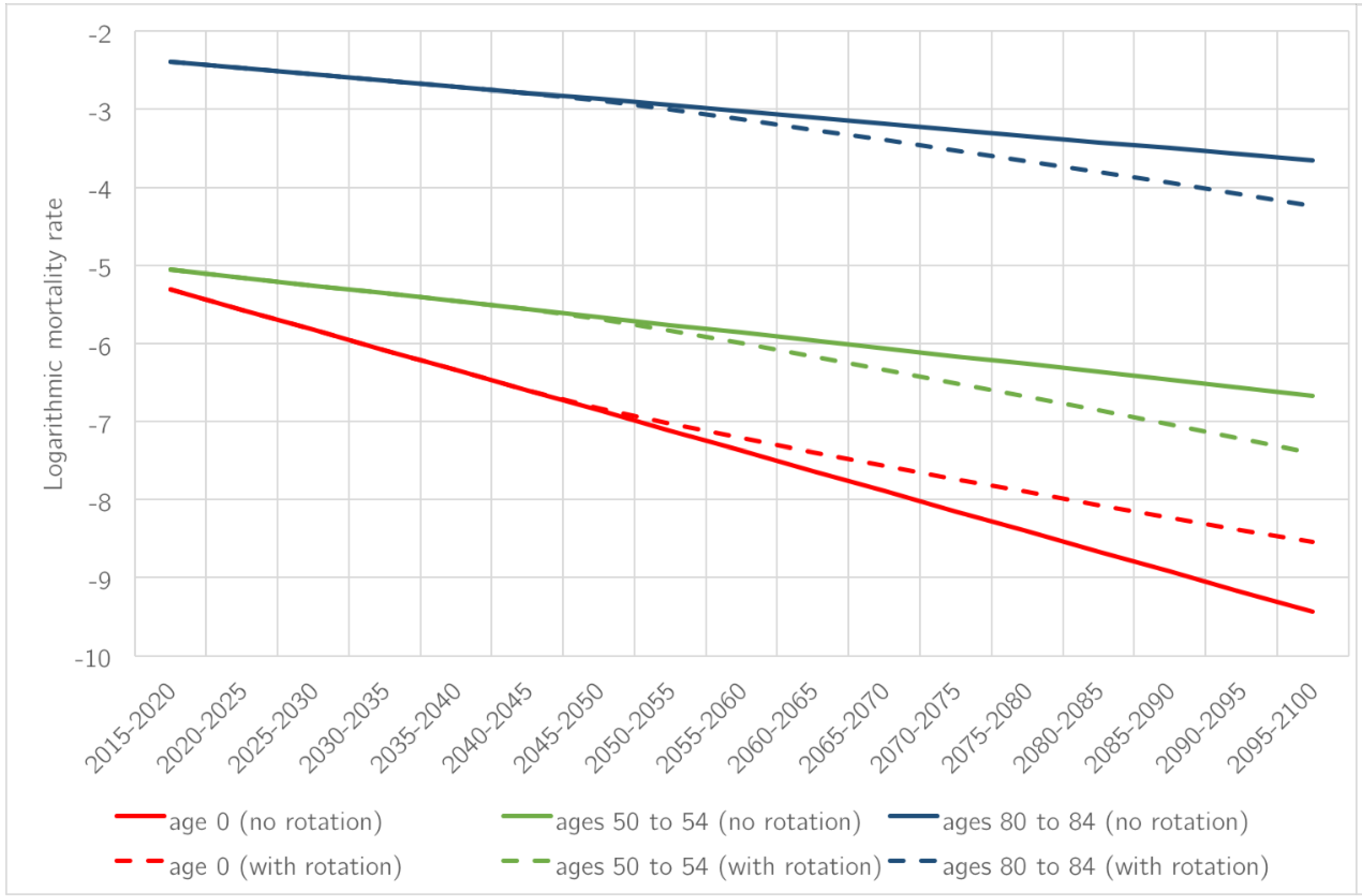
$$w(t) = \frac{e_0(t) - 80}{e_0^u - 80}$$

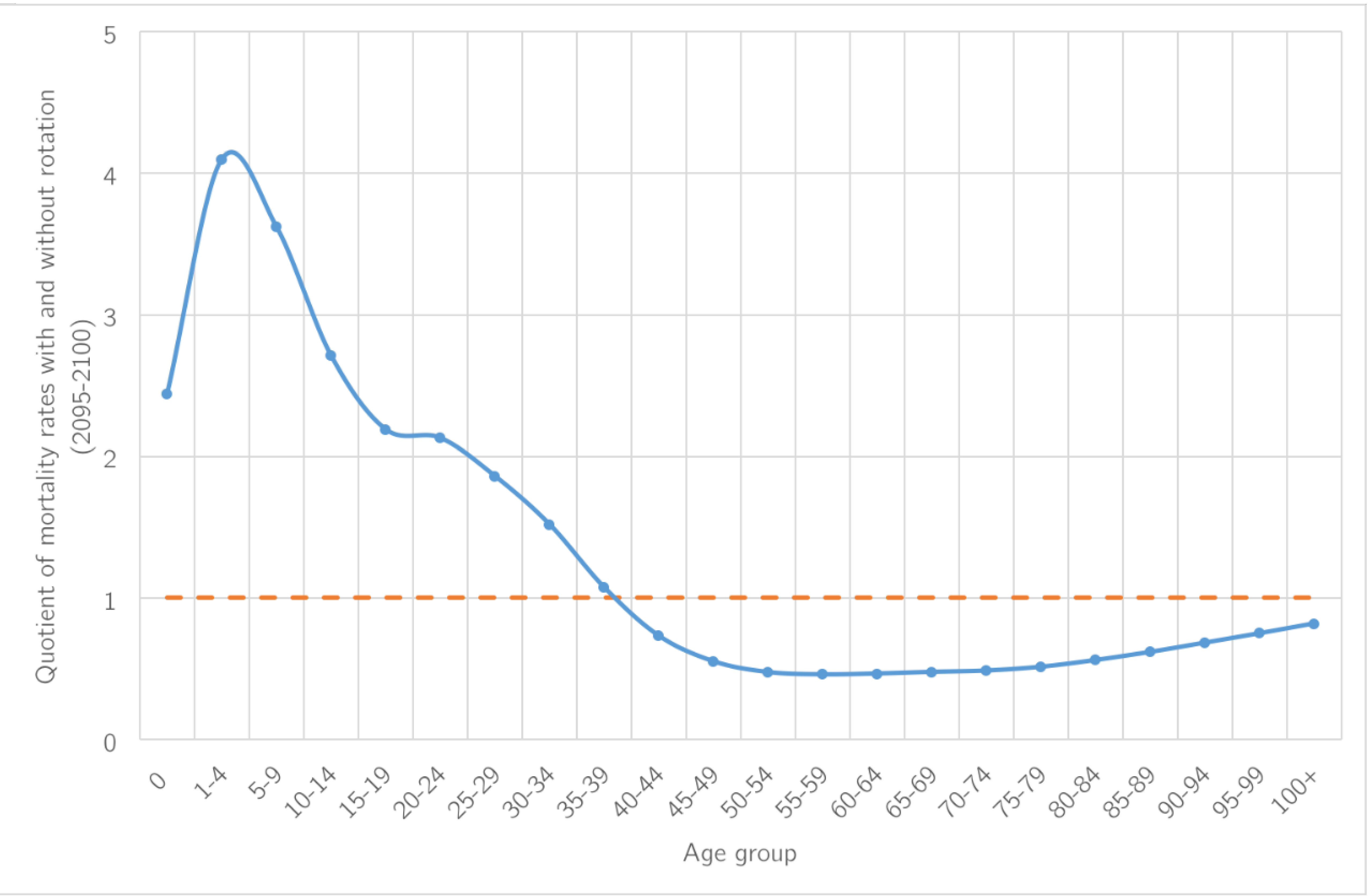
- **The authors do not optimize the hyperparameters *p* (the speed of rotation) and the life expectancy at birth where the rotation starts but propose 0.5 and 80 years, respectively.**

Rotation of B(x,t) in the rotated variant



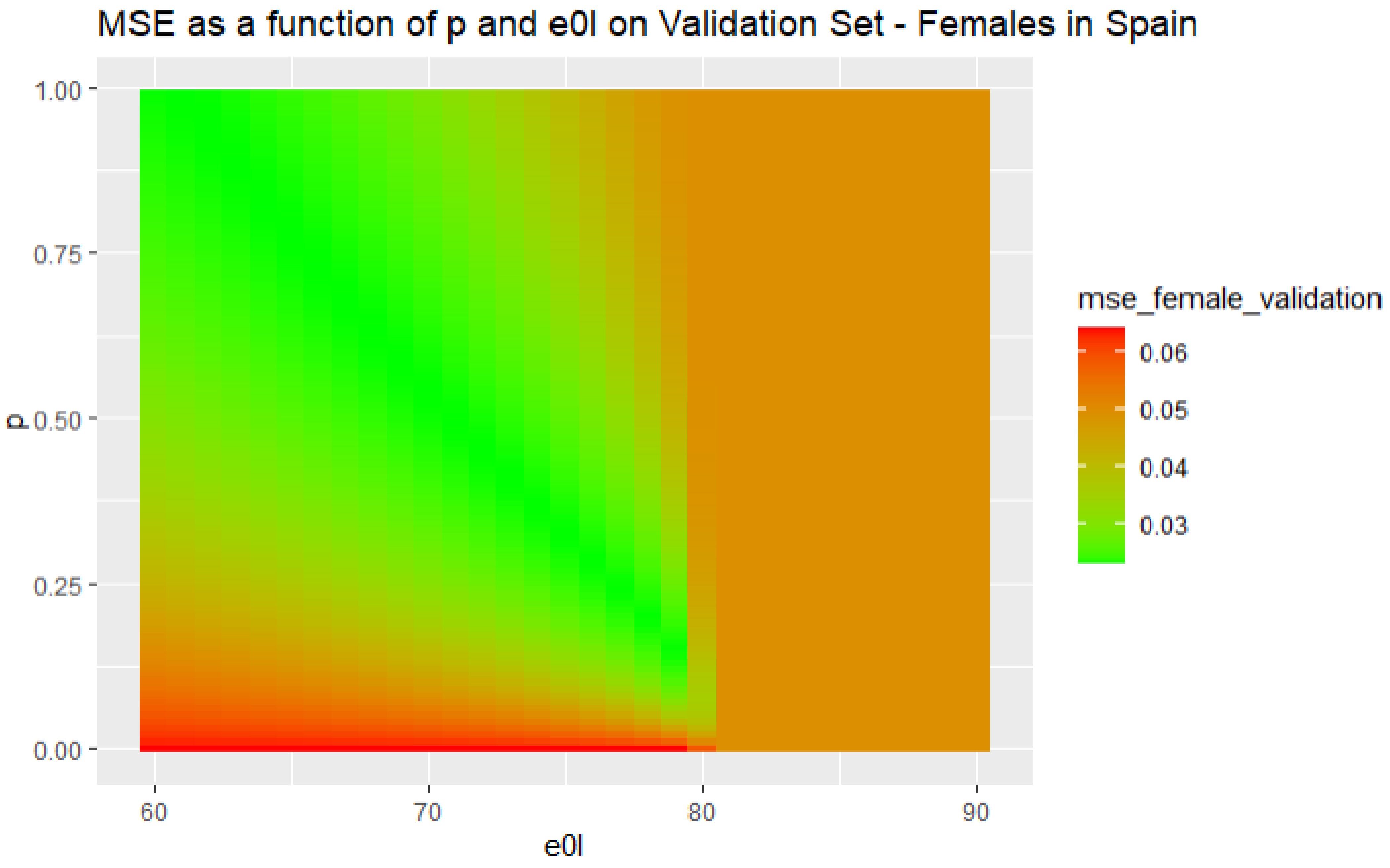
Projection with and without rotation





Hyperparameter Tuning for Rotated L-C

- **Grid Search on two hyperparameters**
  - p: „speed of rotation” $\rightarrow [0,1]$ Default is 0.5
  - e_0^l: life expectancy at birth at which the rotation starts $\rightarrow$ Default is 80
  - $e_0(t)$ is projected by vanilla Lee-Carter to check against e_0^l
- **Parameter space on the Grid Shearch is**
  - $[0,1]$ increased by 0.01 for p
  - integers in $[60,90]$ for e_0^l
- **Technical setup of Grid Search**
  - **Training set** for parameter estimation is 1950-1990
  - **Validation set** for checking hyperparameter performance in MSE of $\ln m_{x,t}$ is 1991-1999
    - Selecting $\{p, e_0^l\}$ that minimize MSE of $\ln m_{x,t}$ on validation set
  - **Retraining** the model on training + validation set (1950-1999)
  - **Projecting** $\ln m_{x,t}$ from the retained model to **test set**: 2000-
    - Get test performance as MSE of $\ln m_{x,t}$
  - *Multiple optimal hyperparameter settings is possible!*
- **Separately for *male* and *female* populations**



Optimal hyperparameters: $\{p = 0.99, e_0^l = 61\}$

GAM on the Lee-Carter residuals

- **„Vanilla” Lee-Carter formulation**

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}$$

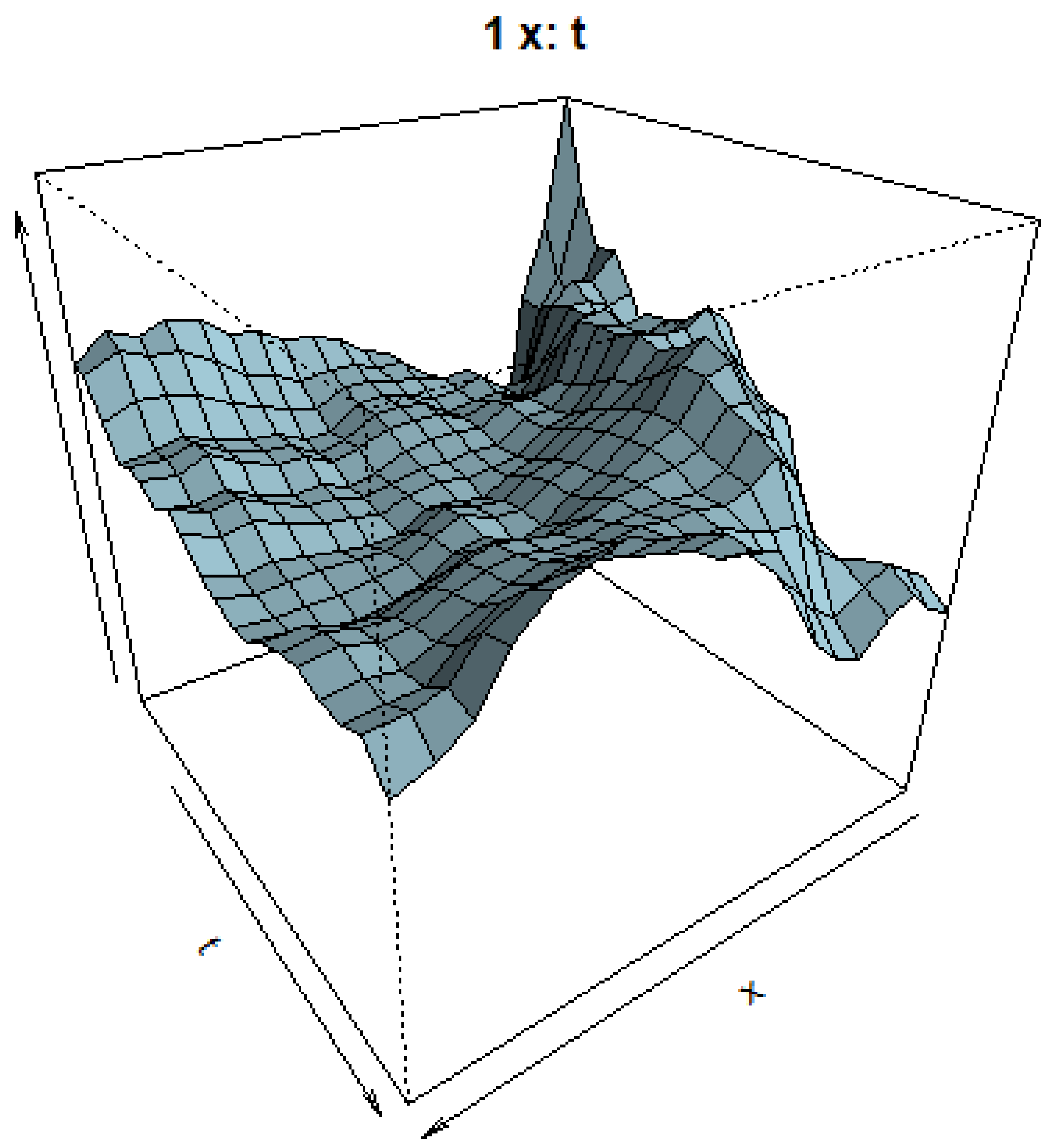
$\rightarrow \varepsilon_{x,t} \sim N(0, \sigma^2)$ and i.i.d.

- **When rotation is present $\varepsilon_{x,t} \sim N(0, \sigma^2)$ and i.i.d. assumption does not hold**
  - $\varepsilon_{x,t}$ is still dependent on x and t
  - $\varepsilon_{x,t} = \mathrm{f}(x,t) + u_{x,t}$ where $u_{x,t}$ has homogeneous σ^2 and i.i.d.
  - $\mathrm{f}(x,t)$ is a function that represents the rotation patterns
- **$\mathrm{f}(x,t)$ can be estimated in a GAM framework**
  - $\mathrm{f}(x,t)$ can be decomposed additively as

$$\mathrm{f}(x,t) = g_1(x) + g_2(t) + h(x,t)$$

  - $g_1(x), g_2(t)$ and $h(x,t)$ can be represented as **spline** functions in GAMs
- **The approach of the *mgcv* R package by Simon N. Wood is considered**

$h(x,t)$ fitted on
Spanish female population



Repesenting the f(x, t)s

- **See for 1D spline functions $f(x)$s first – basis expansion**

$$f(x) = \sum_{k=1}^{K} \gamma_k b_k(x)$$

→ $b_k(x)$s are usually 3rd degree polinomials (cubic basis function)

→ K is an integer hyperparameter called *knots*

→ K should not be too large to avoid overfitting

- **Curve fitting – Objective function**

$$\sum_i (y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx \rightarrow \min$$

λ can be optimized through cross-validation

→ maximize the marginal likelihood (REML)

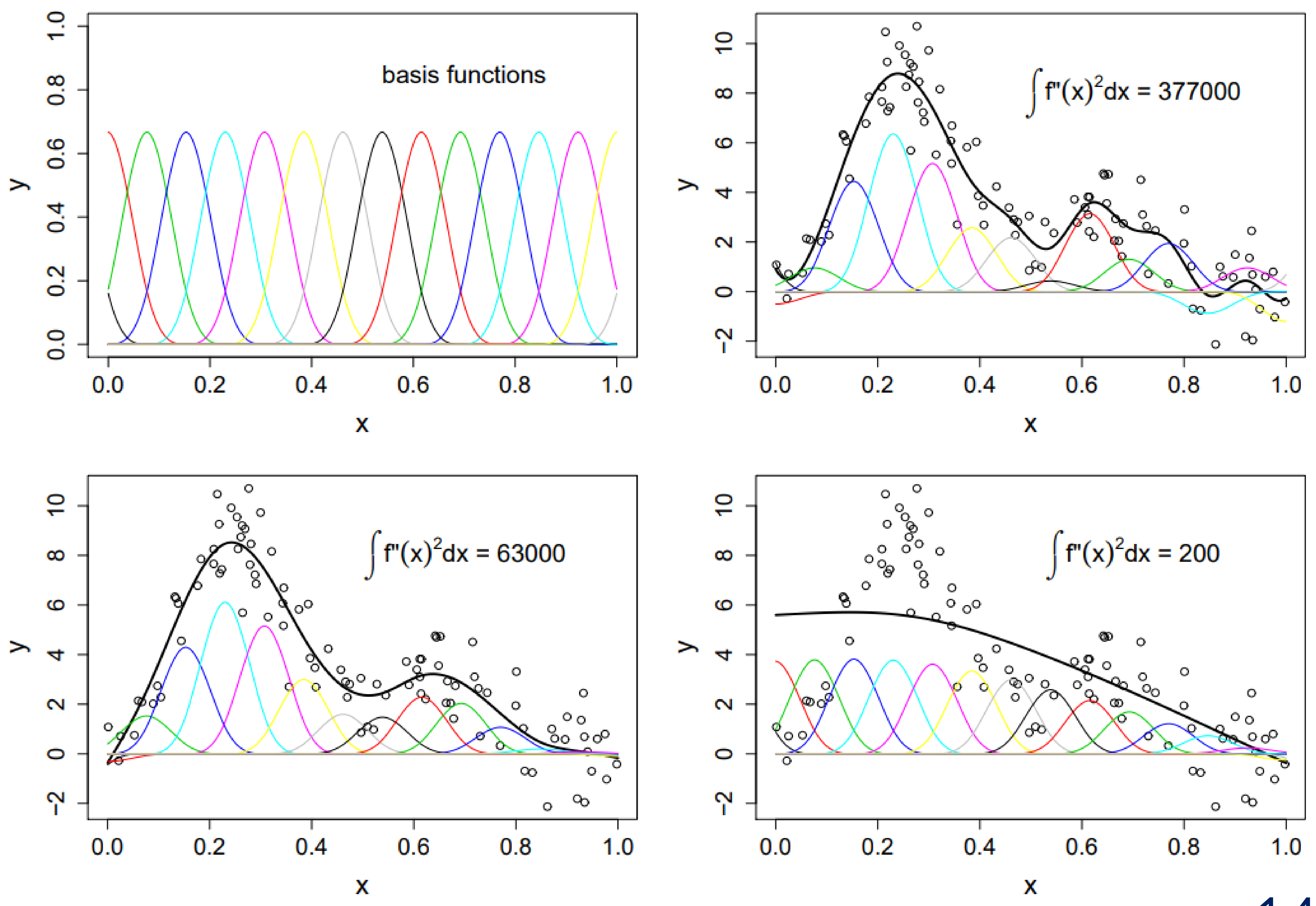
$\int f''(x)^2 dx$ is the ‚wiggliness' penalty

$\int f''(x)^2 dx$ can be expressed in matrix form: $\gamma^T \boldsymbol{S} \gamma$ ($\boldsymbol{S}$ known)

- **Objective function can be extended for 2D**

$$\sum_i (y_i - f(x_i, t_i))^2 + \lambda \int f_{xx}^2 + f_{xt}^2 + f_{tt}^2 dxdt \rightarrow \min$$

Basis & penalty example
Source: Wood (2017)



Spline fitting methods in *mgcv* for 2D

- **Cubic regression splines ’*cr*’ and ’*cs*’**

  $b_k(x)$s are 3rd degree polinomials

  knots spread evenly through the covariate values

- **Thin plate regression splines ’*tp*’**

  starts from a full spline

  → where number of knots equals the number of observations

  takes a reduced rank eigen approximation of this full spline

  gets an optimal low rank basis

  avoids the knot placement problem this way

- **The ’*cs*’ and ’*ts*’**

  special cases for ’cr’ and ’tp’ respectively

  extra penalty on the null space in the objective function

  get the eigen decomposition of $\boldsymbol{S}$ fom $\gamma^T \boldsymbol{S} \gamma$

  define $\boldsymbol{S}^* = \boldsymbol{U}^* {\boldsymbol{U}^*}^T$ where the matrix of eigenvectors corresponding to the zero eigenvalues of $\boldsymbol{S}$

  apply the double penalty term $\lambda \gamma^T \boldsymbol{S} \gamma + \lambda^* \gamma^T \boldsymbol{S}^* \gamma$ in the objective function

  the whole term can therefore be shrunk to zero

Hyperparameter Tuning for GAMs

- **Grid Search on two hyperparameters**

  bs: „type of spline" → $\{cr, tp, cs, ts\}$

  K: number of „knots" → $\{2, 3, 4, 5, 6, 7\}$

  some pairs might not be fitted to certain data in reasonable time

  → these points in the grid are omitted

  forecasting the L-C residuals is done by extrapolating the basis functions after the last „knot" along t

- **Technical setup of Grid Search**

  Same as for Rotated Lee-Carter

  Measure of fit is MSE of $\ln m_{x,t}$

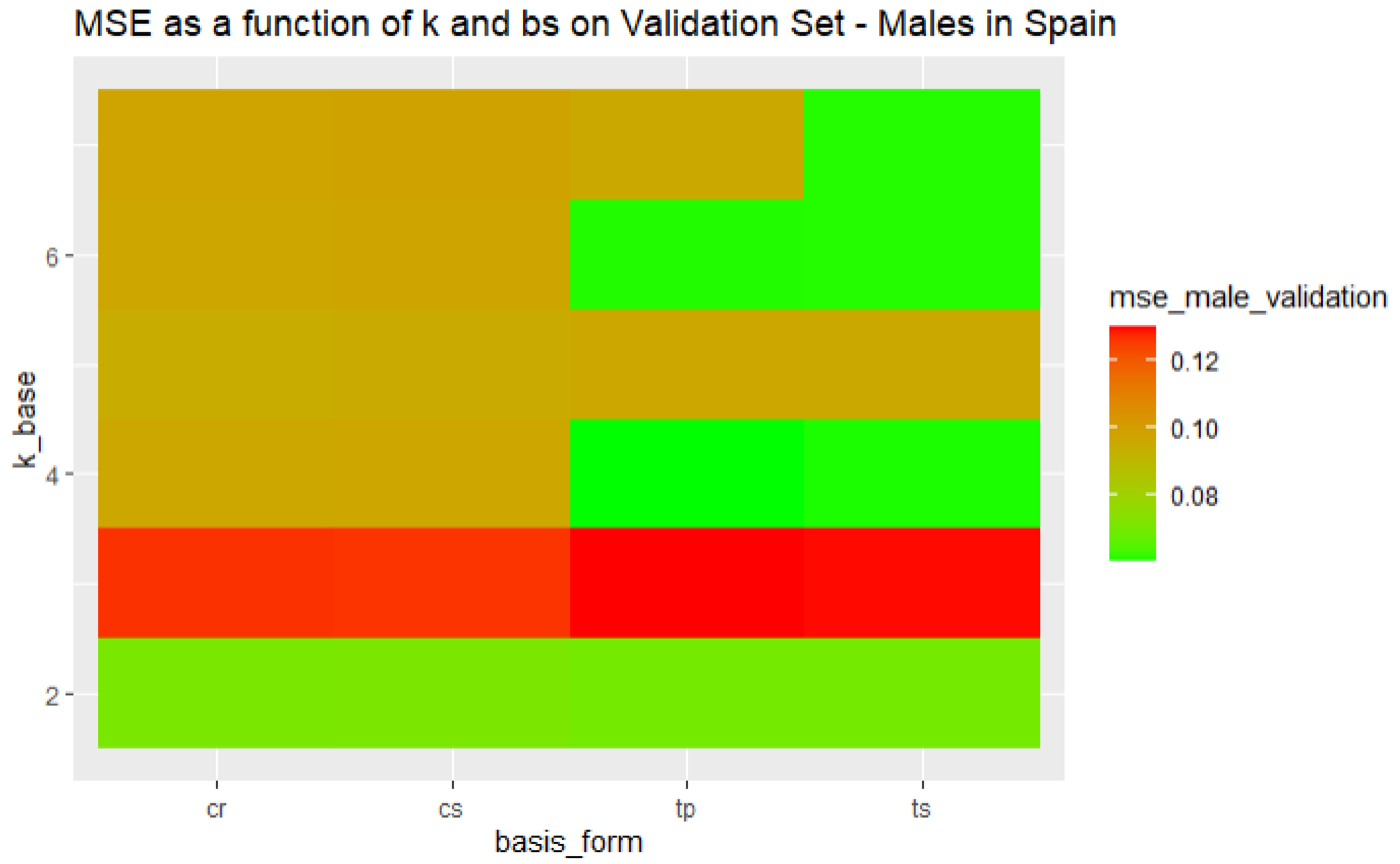
  Training set is 1950-1990

  Validation set is 1991-1999

  Retraining on 1950-1999

  Test set: 2000-

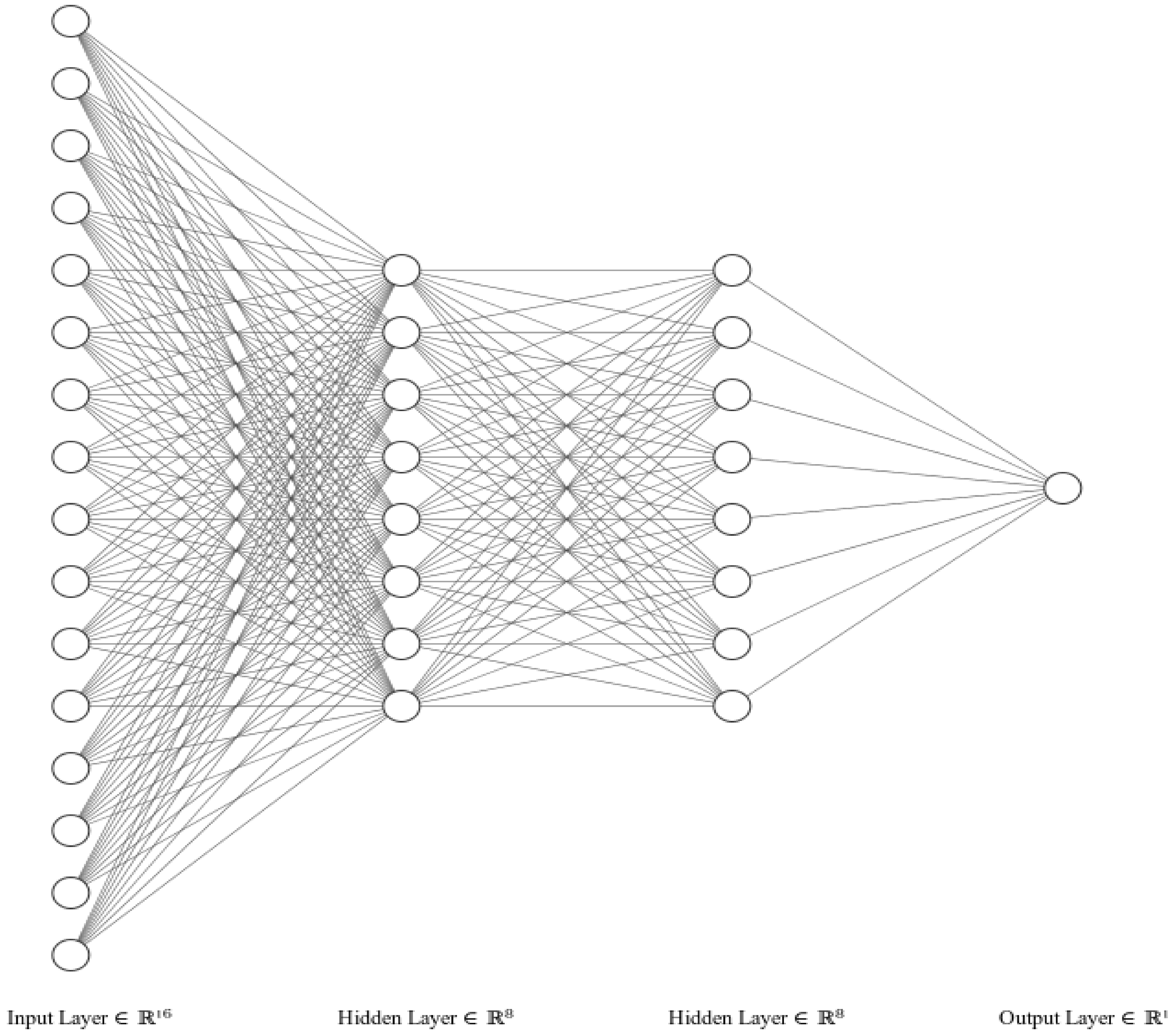
- **Separately for *male* and *female* populations**



Optimal hyperparameters: $\{K = 4, bs = 'tp'\}$

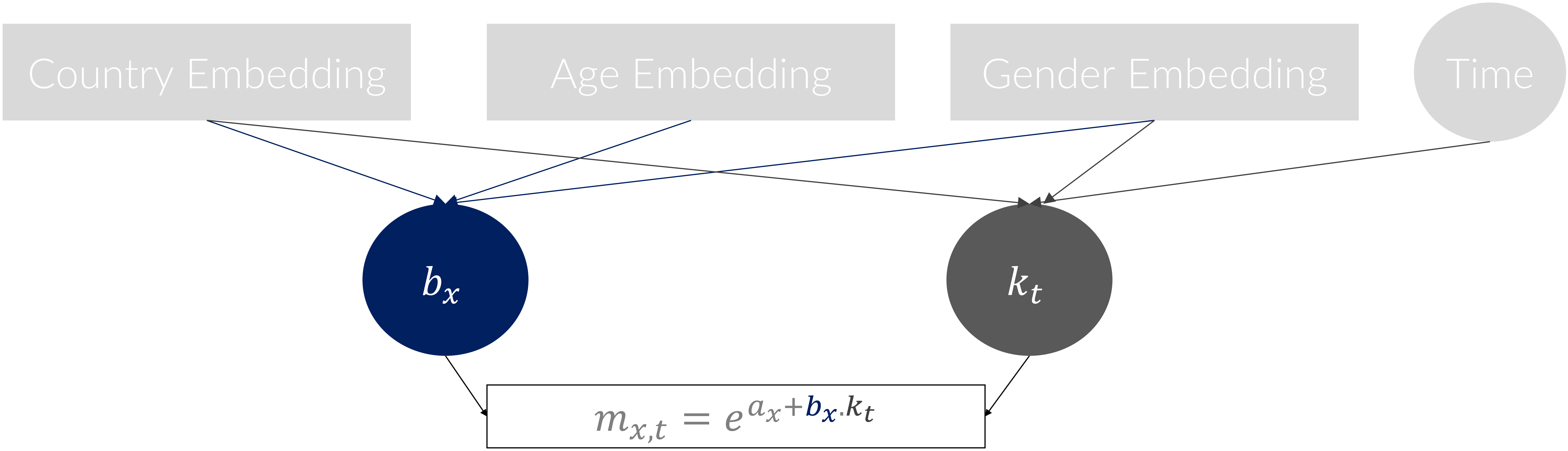
Deep Feedforward Net

- **Deep = multiple layers**
- **Feedforward = data travels from left to right**
- **Fully connected network (FCN) = all neurons in layer connected to all neurons in previous layer**
- **More complicated representations of input data learned in hidden layers - subsequent layers represent regressions on the variables in hidden layers**



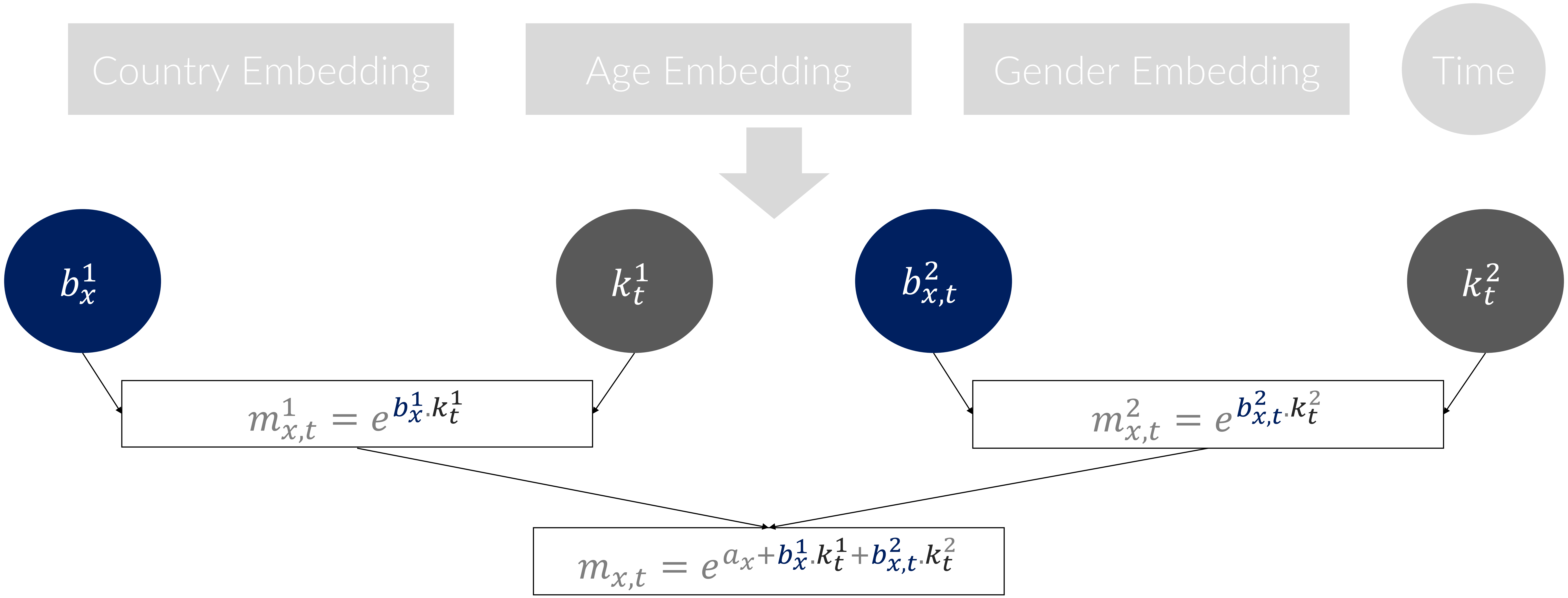
Application to Forecasting (1)

- **Deep neural networks used to predict parameters of Lee-Carter model**
  - Concept referred to as a hypernetwork by Ha et al. (2016)
  - FFNs + country embeddings used here
- **Pooling across countries and genders -> better performance than vanilla LC model in some cases, even with simple networks**
- **Applied using convolutional networks in Perla et al. (2021) and Scognamiglio (2021)**
- **4 key inputs to network**
  1. Country
  2. Gender
  3. Age
  4. Year
- **1-3 treated as categorical using embeddings and 4 treated as continuous input to network**
- **Basic Lee-Carter network:**



Application to Forecasting (2)

- **Previous model allows only for mortality improvement rates (bx) that vary with Country, Age and Gender, but not with time.**
- **Adapt neural nets to allow for time-varying mortality improvement rates (bx,t)**
- **Boosting = improving previous model by adding new model component, see e.g. Hothorn et al. (2010)**
- **Fit in two stages:**
  1. Fit basic Lee-Carter Hypernetwork, then keep parameters
  2. Boost the network using a second network that includes $b_{x,t}$



Stacking

- **Well known result in time series literature that averaging over models boosts performance see e.g. Jose and Winkler (2008)**

- **Performance of simple average usually close to optimal**

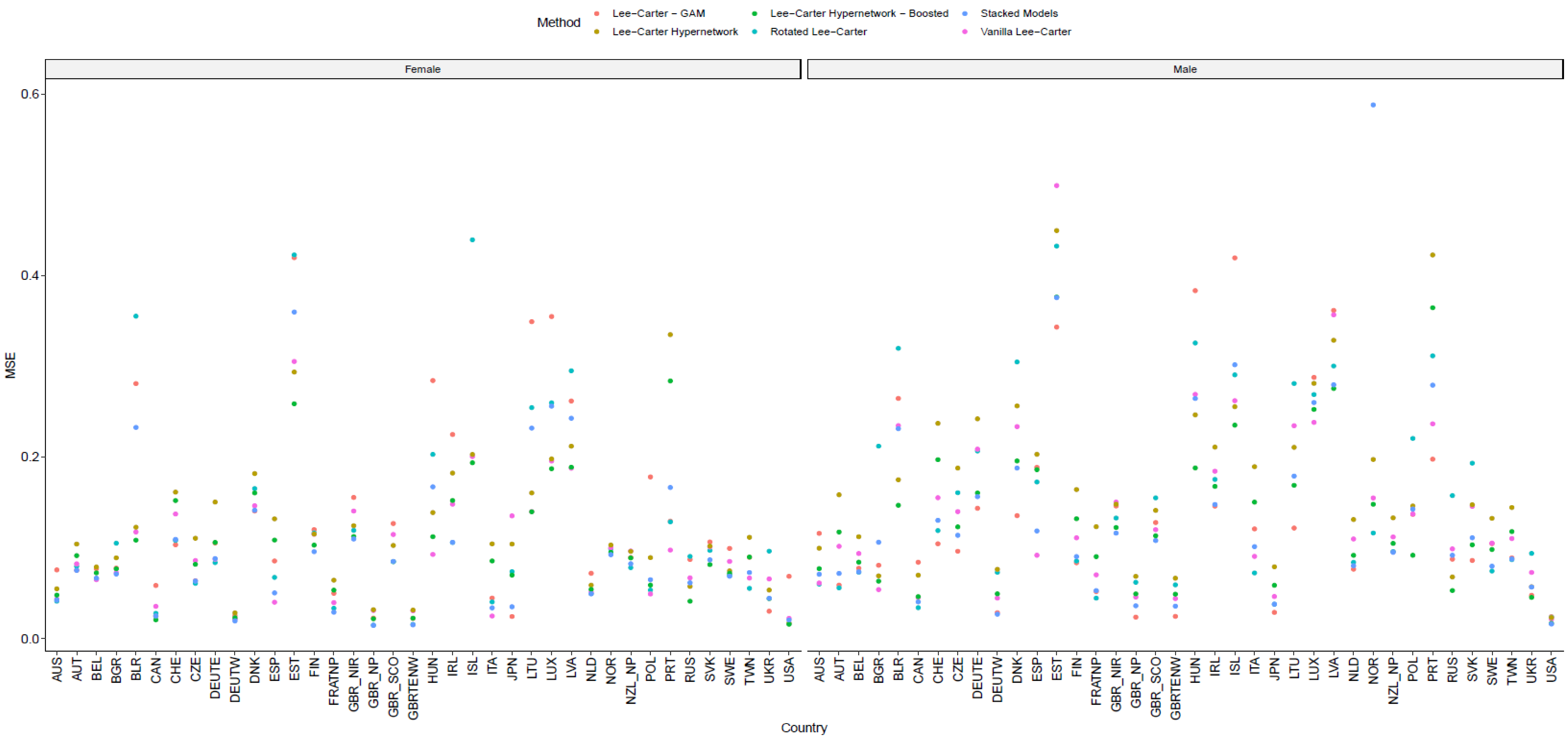
- **For stacked model, used simple average of 3 models best performing models:**

    1. Lee-Carter Hypernetwork - Boosted
    2. Rotated Lee-Carter
    3. Lee-Carter - GAM

Agenda

1. Introduction
2. Data and methods
3. **Results**
4. Conclusions

Performance evaluation (1)



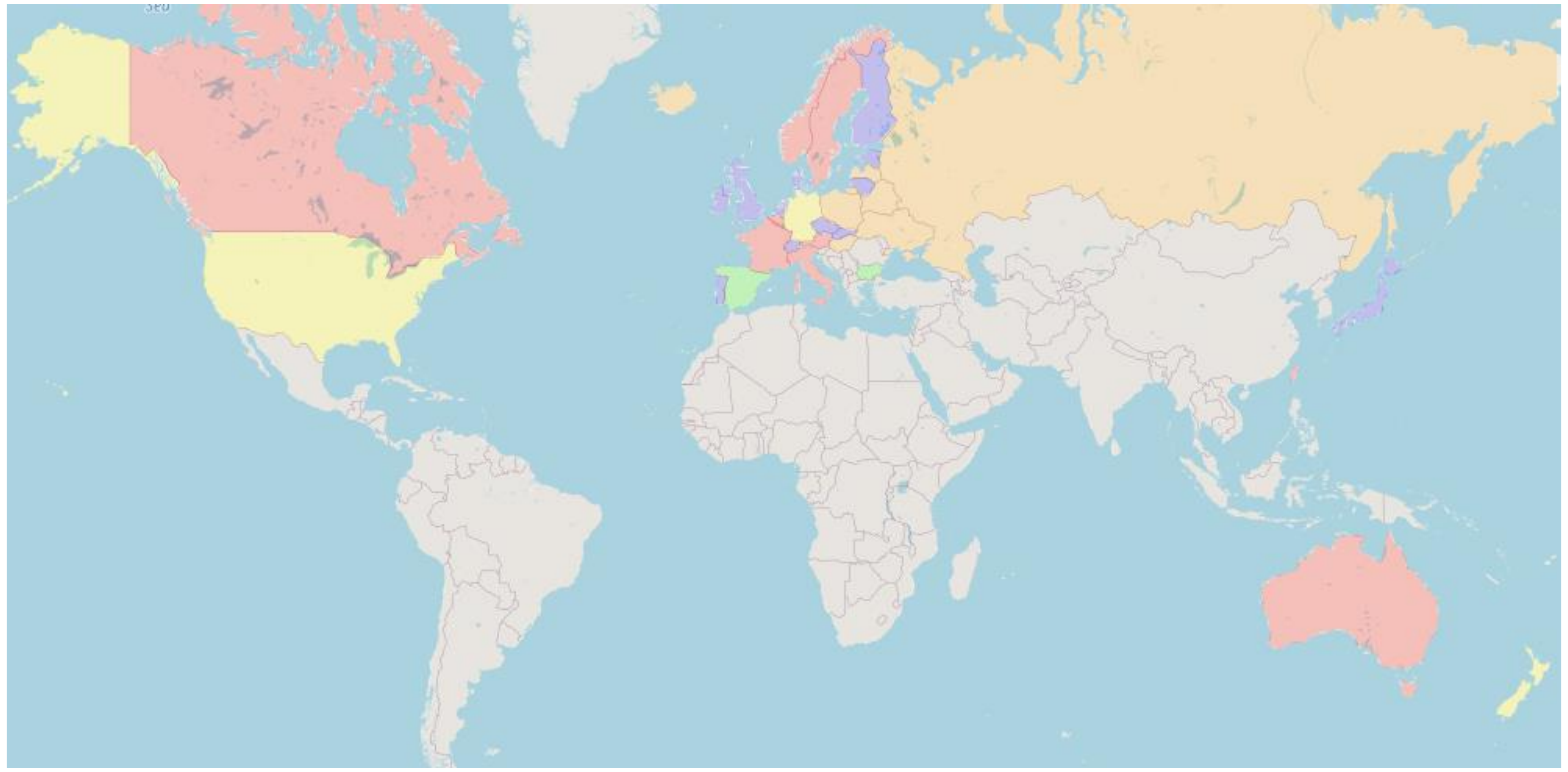
Performance evaluation (2)

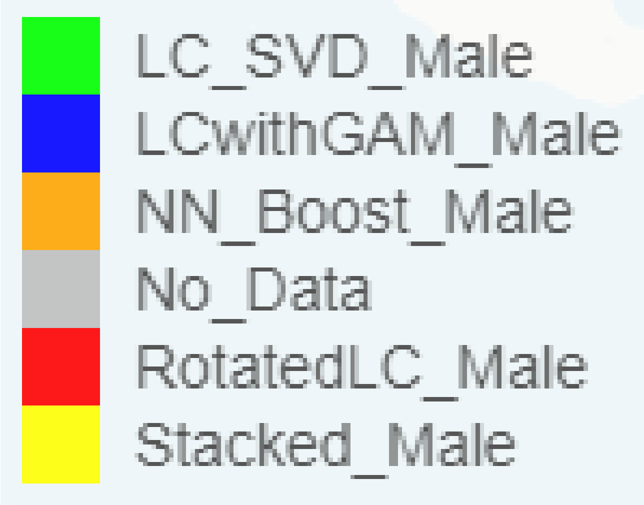
- **Table shows number of times minimum MSE is attained by each method…**
- **… when considering each country and gender separately**
- **Best method overall is the Lee-Carter – GAM method**
- **Results change when considering for each gender**
- **Strongest single models are the LC boosted with neural networks for females and LC boosted with GAMs for males**
- **Stacked models perform optimally for females**

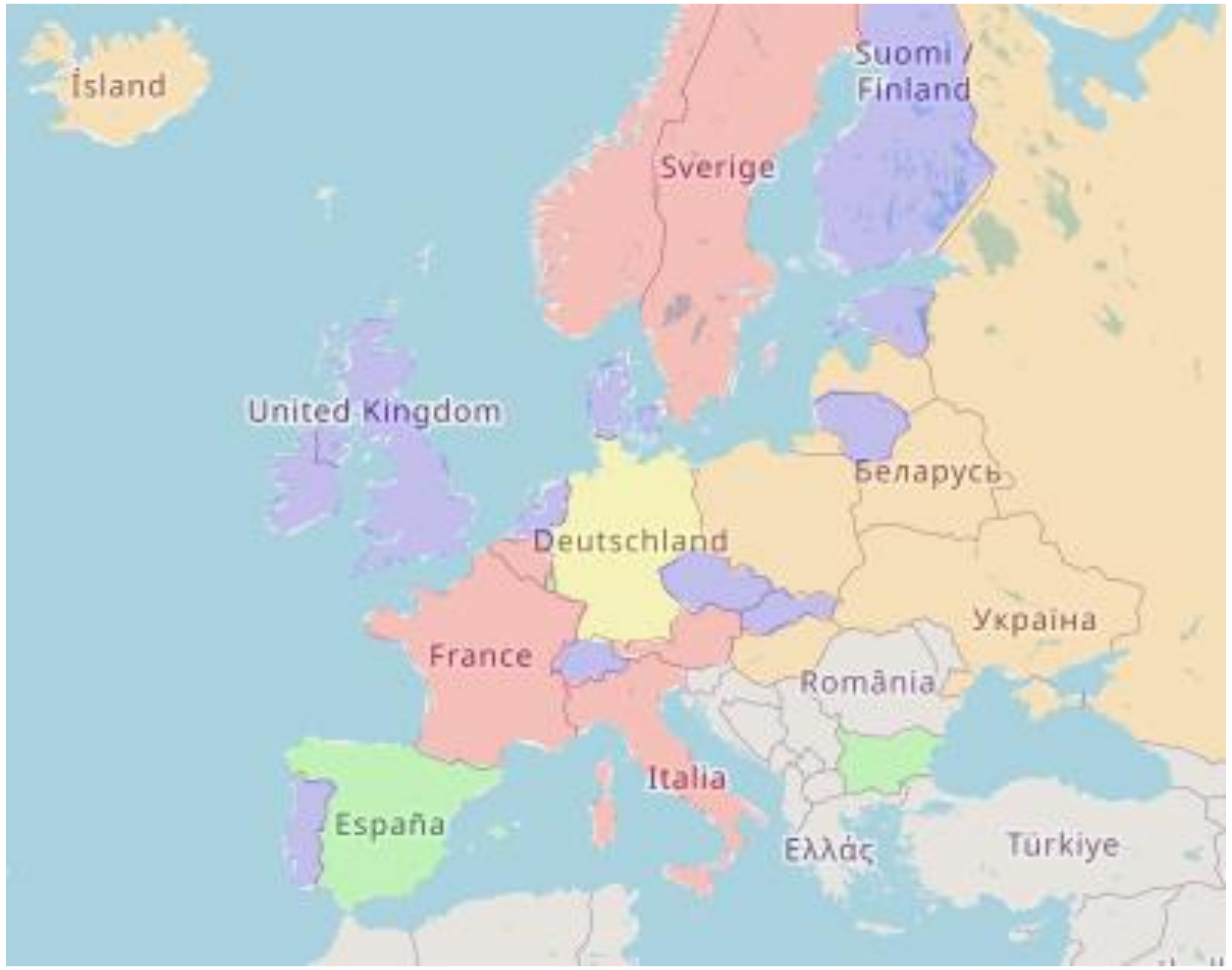
Method	Number of Wins	Percentage of Wins
Lee-Carter - GAM	18	24%
Lee-Carter Hypernetwork - Boosted	16	21%
Rotated Lee-Carter	16	21%
Stacked Models	16	21%
Vanilla Lee-Carter	10	13%
Total	76	100%

Method	Female	Male
Lee-Carter - GAM	4	14
Lee-Carter Hypernetwork - Boosted	9	7
Rotated Lee-Carter	7	9
Stacked Models	11	5
Vanilla Lee-Carter	7	3
Total	38	38

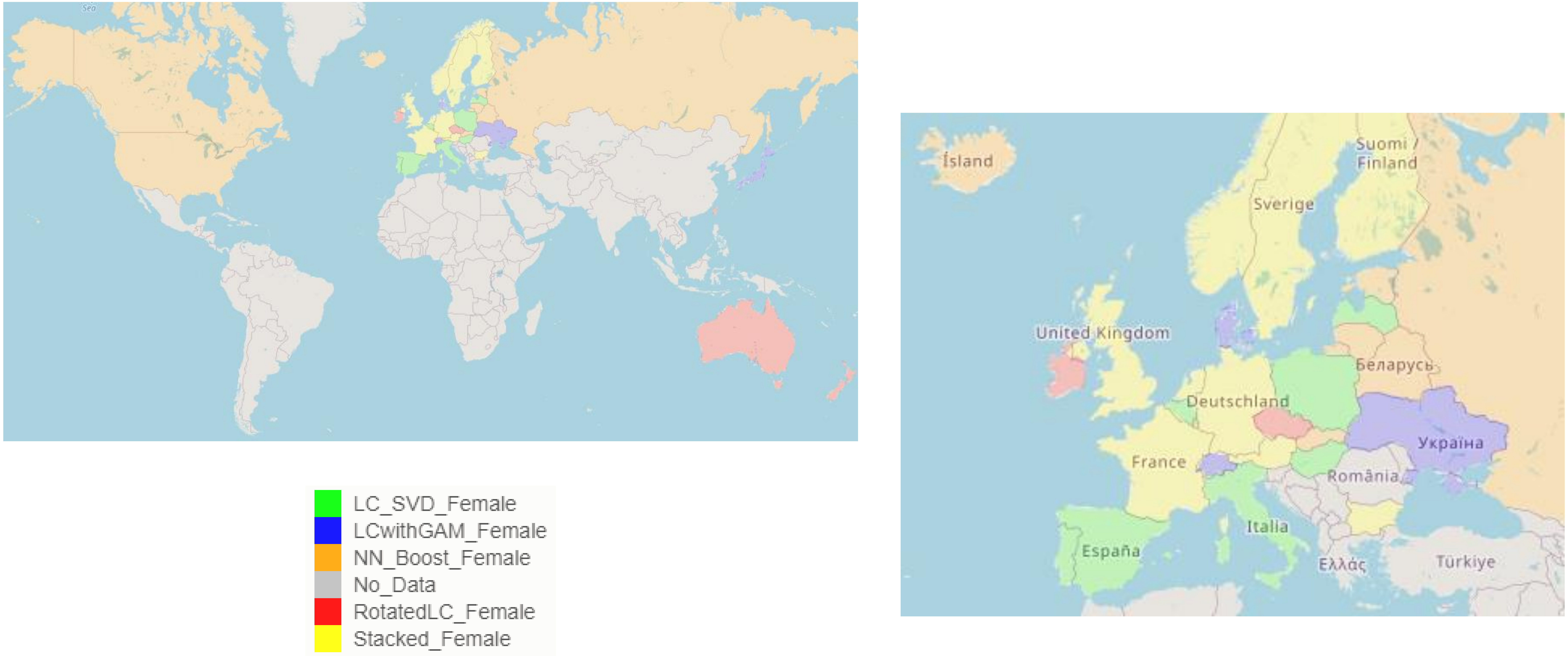
Spatial Analysis – Best Model for Males



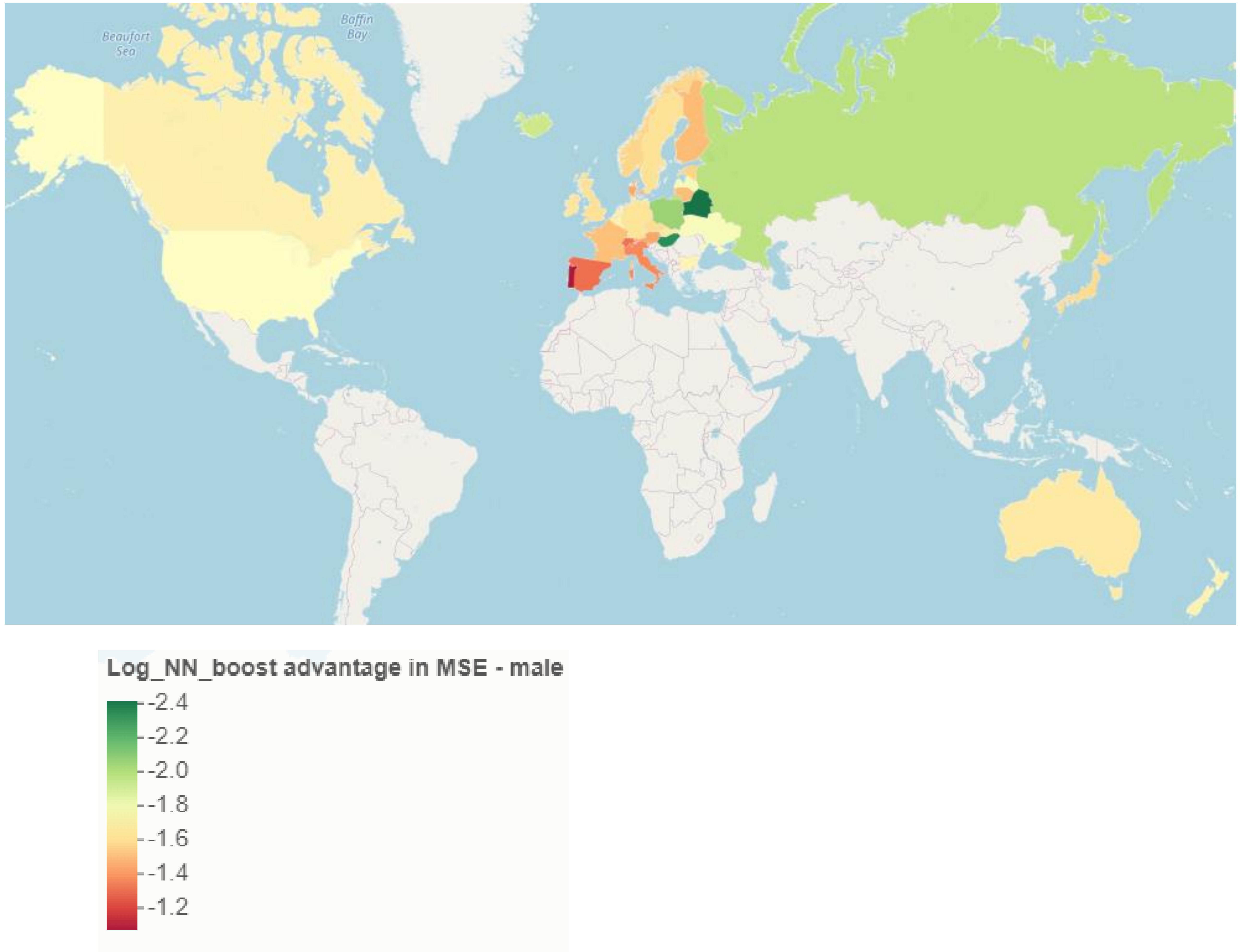




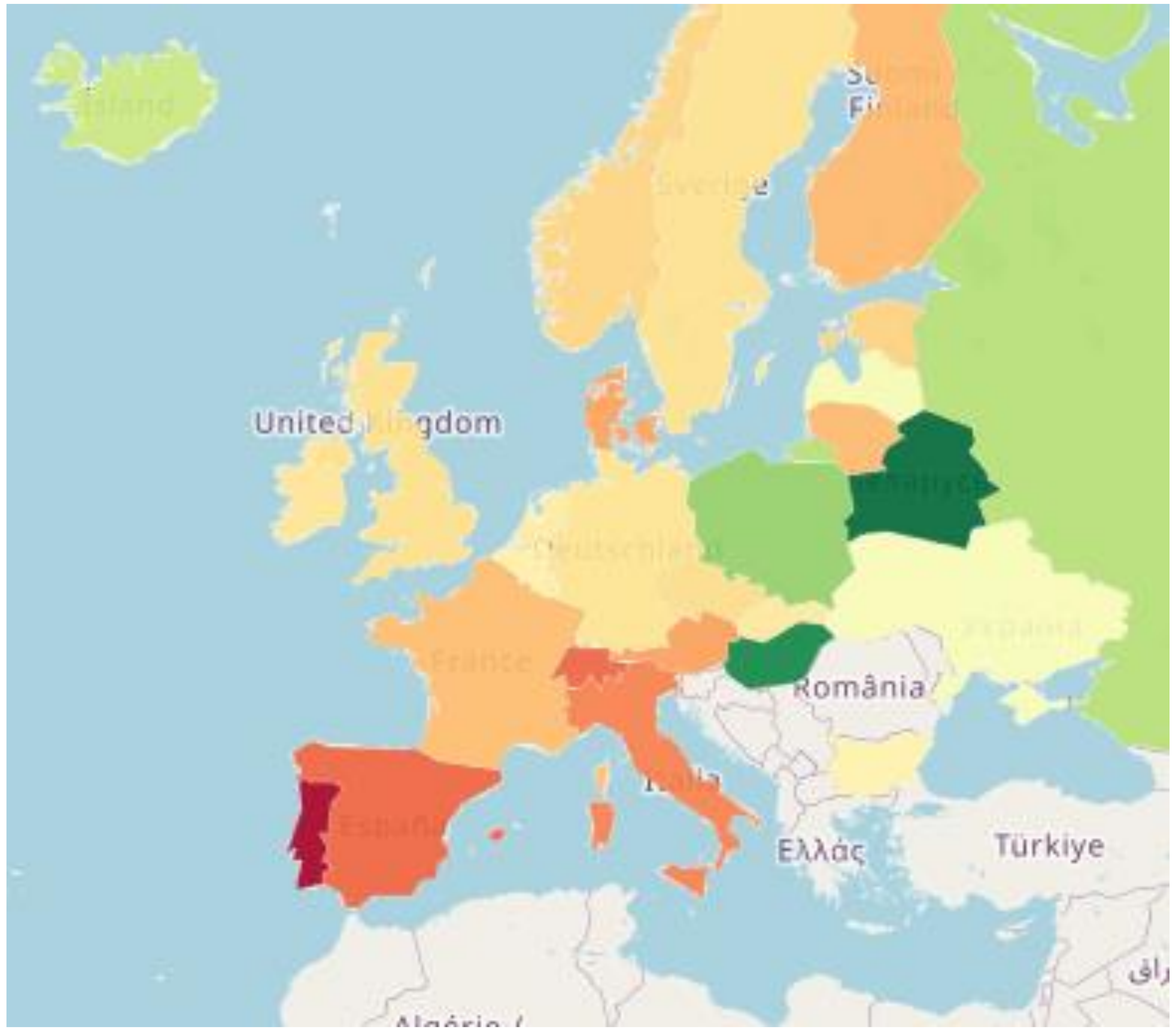
Spatial Analysis – Best Model for Females



Spatial – NN Boost Advantage for Males

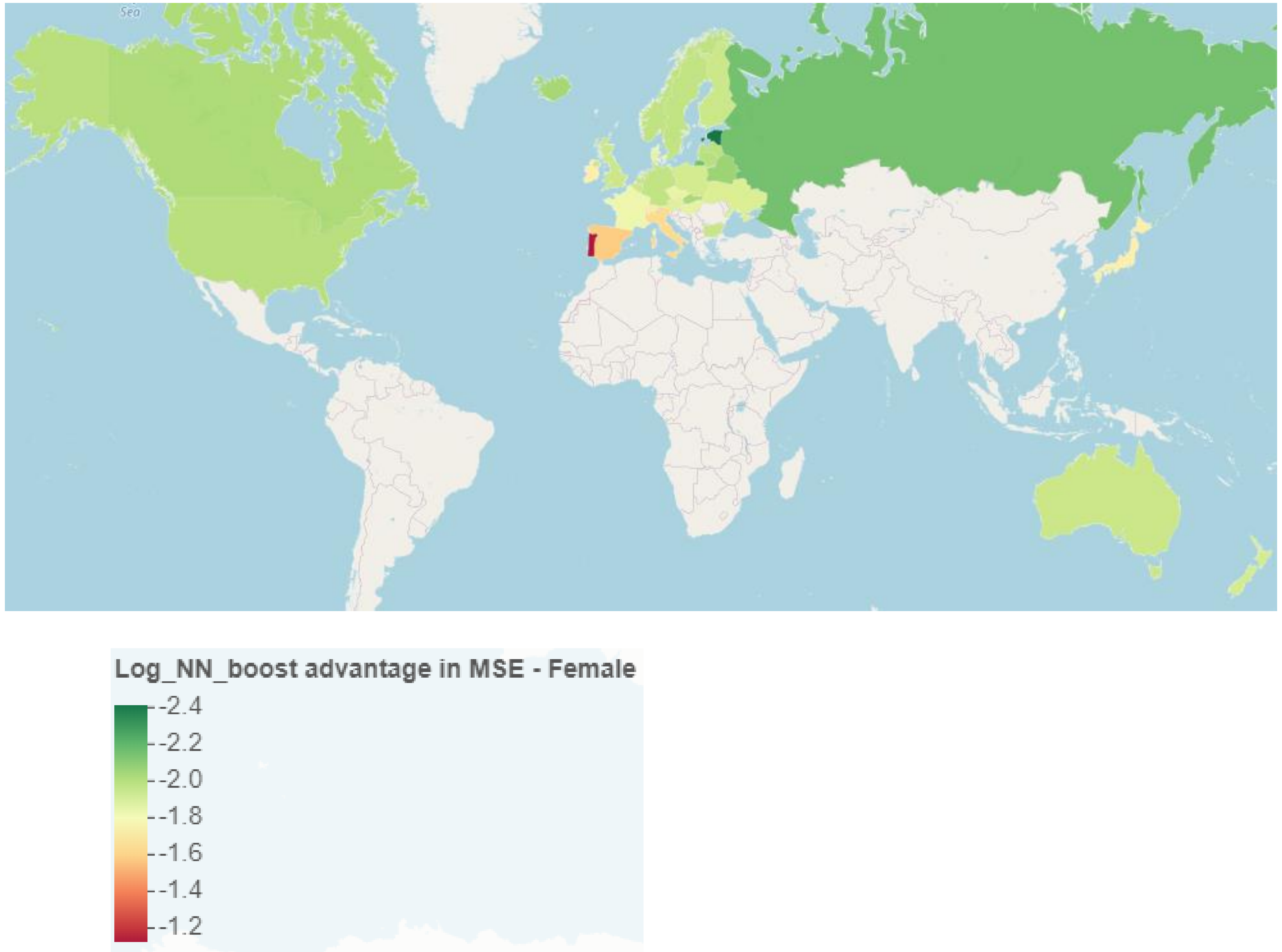


Moran’s I = 0.52*

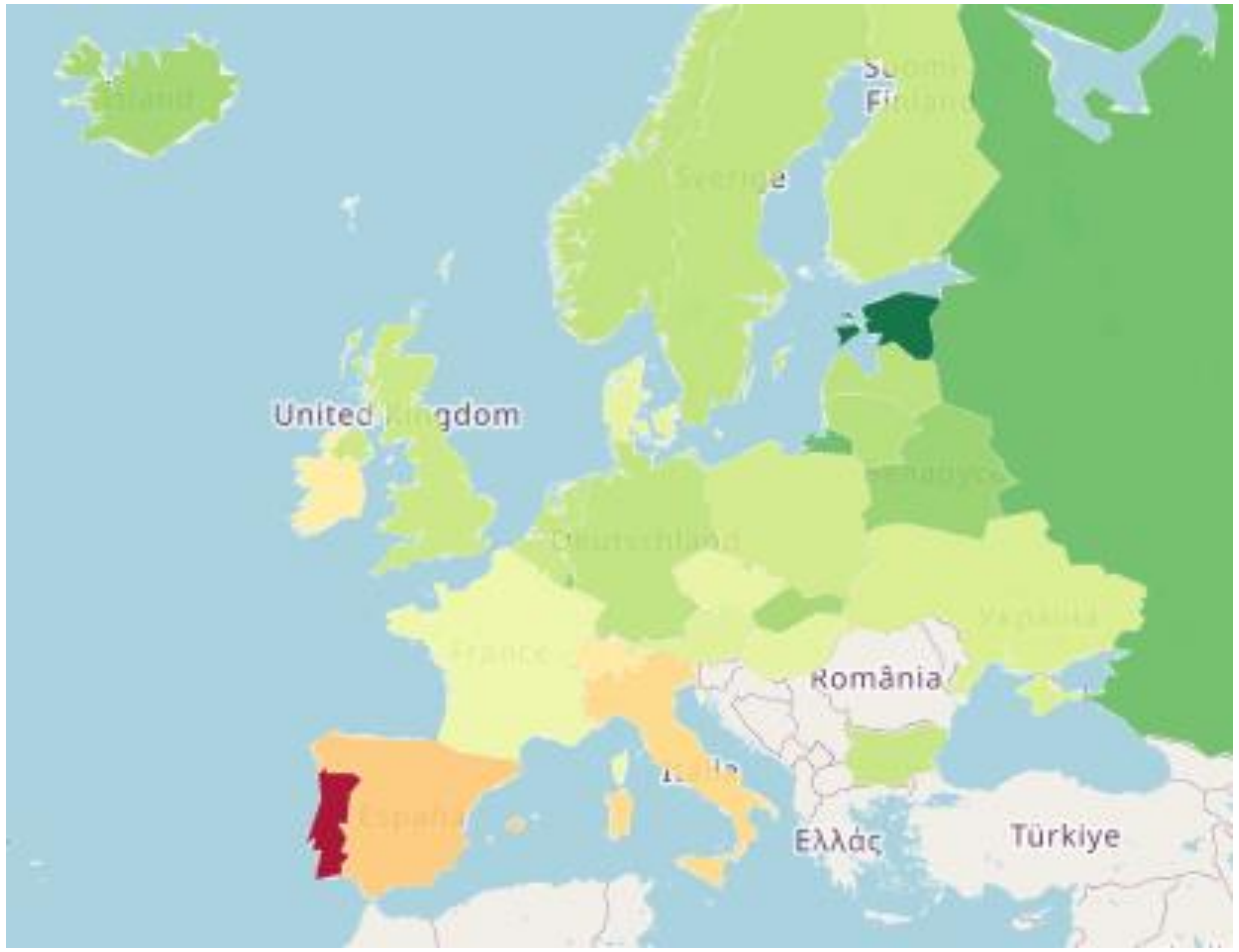


Compared to the best not NN solution – Log Scale

Spatial – NN Boost Advantage for Females



Moran's I = 0.51*



Compared to the best not NN solution – Log Scale

Agenda

1. Introduction
2. Data and methods
3. Results
4. **Conclusions**

Conclusions

- **Significant gains in forecasting ability by applying more advanced methods than original Lee-Carter model**
- **Machine learning approach to tuning hyperparameters of Rotated Lee-Carter model proposed here**
- **New methods demonstrated – boosting a Lee-Carter model using a GAM and a deep neural network**
- **Benefits of boosting LC with neural nets mainly observed for Eastern European countries**
- **Next steps:**

  Demonstrate financial impacts on annuity valuations
  Can we evaluate COVID impacts in the same framework?

References

- **Carter, L.R. and Prskawetz, A. (2001). Examining structural shifts in mortality using the Lee -Carter method (working paper). Max Planck Institute for Demographic Research. https://www.demogr.mpg.de/Papers/Working/wp-2001-007.pdf**
- **Christensen, K., Doblhammer, G., Rau, R. and Vaupel, J.W. (2009). Ageing populations: the challenges ahead. Lancet 374(9696):1196–1208. https://doi.org/10.1016/S0140-6736(09)61460-4**
- **Horiuchi, S. and Wilmoth, J.R. (1995). The aging of mortality decline. In: Annual meeting of the population Association of America, San Francisco, CA**
- **Jose, V.R.R. and Winkler, R.L., 2008. Simple robust averages of forecasts: Some empirical results. International journal of forecasting, 24(1), pp.163-169.**
- **Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting US mortality. J Am Stat Assoc 87:659–671. https://doi.org/10.2307/2290201**
- **Kannisto, V., Lauritsen, J., Thatcher, A.R. and Vaupel, J.W. (1994). Reductions in mortality at advanced ages: several decades of evidence from 27 countries. Popul Dev Rev 20(4):793–810. https://doi.org/10.2307/2137662**
- **Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. Demography 38(4):537–549. https://doi.org/10.1353/dem.2001.0036**
- **Li, N., Lee, R. and Gerland, P. (2013). Extending the Lee–Carter method to model the rotation of age patterns of mortality-decline for long-term projection. Demography 50(6):2037–2051. https://doi.org/10.1007/s13524-013-0232-2**
- **Rau, R., Soroko, E., Jasilionis, D. and Vaupel, J.W. (2008). Continued reductions in mortality at advanced ages. Popul Dev Rev 34:747–68. https://doi.org/10.1111/j.1728-4457.2008.00249.x**
- **Vékás, P. (2020). Rotation of the age pattern of mortality improvements in the European Union. *Central European Journal of Operations Research*, 28: pp. 1031–1048. https://doi.org/10.1007/s10100-019-00617-0**
- **Wood, S. N. (2017) Generalized Additive Models: An Introduction with R (2nd edition).Chapman and Hall/CRC.**