# Multi-population longevity modeling with Gaussian Processes

2021 Insurance Data Science Conference

June 16, 2021

Mike Ludkovski

Dept of Statistics & Applied Probability UC Santa Barbara

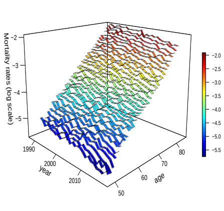
joint w/Nhan Huynh

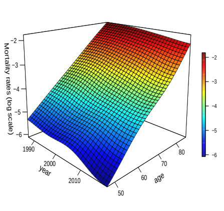
STATISTICS & APPLIED PROBABILITY CENTER FOR FINANCIAL MATHEMATICS & ACTUARIAL RESEARCH

UCSB

# Mortality modeling

- Mortality rates indexed by Age and Year: $x = (x^n_{ag}, x^n_{yr})$
- Raw deceased count $D^n$; Exposures $E^n$; Mid-Year Lives $L^n$
- Observed raw log-rates $Y(x^n) = \log \frac{D^n}{L^n}$
- Model $\boxed{Y(x) = f(x) + \varepsilon}$ where $Var(\varepsilon(x)) = \sigma^2(x)$ (additive noise)
- $f(\cdot)$ is the latent log-mortality surface




Consider several popn's $Y^{(\ell)}$ simultaneously:

- Improve fitting
- Information fusion
- Joint forecasts

Desired model features:

- Good predictive power
- Uncertainty quantification
- Interpretable covariance modeling
- Coherent forecasts
- Ability to handle non-rectangular inputs
- Scalability

# Statistical Framework for a Single Population

- Training Dataset $\mathcal{D} = (x^{1:n,(\ell)}, y^{1:n,(\ell)})$
- Specify prior distribution and then compute conditional distribution given the data $p(f|\mathcal{D}) \propto p(\boldsymbol{y}|f, \boldsymbol{x})p(f) = \{likelihood\} \cdot \{prior\}$
- Covariance structure: knowing response at $x$ will greatly influence response at "neighboring" $x$'s
- Gaussian prior + Gaussian likelihood $\Rightarrow$ Gaussian posterior
- Gaussian random field w/prior $f \sim GP(\boldsymbol{m}(\mathrm{x}), \boldsymbol{C}(\mathrm{x}, \mathrm{x}))$
- Observation likelihood $p(\boldsymbol{y}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{f}, \Sigma)$ (Gaussian conjugate!)
- The posterior is Gaussian $f(x)|\mathcal{D} \sim \mathcal{N}(m_*(x), s_*^2(x))$
- Point forecast $m_*(x)$ and credible band $m_*(x) \pm z_\alpha s_*(x)$

# Correlating Populations

- Treat population as a factor covariate with hot encoding: $x^n = (x^n_{ag}, x^n_{yr}, x^n_{ctr_2}, \ldots, x^n_{ctr_L})$
- Full-rank, squared-exponential kernel:

$$C(x^i, x^j) = \underbrace{\eta^2 \exp\left[ -\frac{(x^i_{ag} - x^j_{ag})^2}{2\theta^2_{ag}} - \frac{(x^i_{yr} - x^j_{yr})^2}{2\theta^2_{yr}} \right]}_{\text{Covariance over Age \& Year}} \overbrace{\prod_{\{l_1, l_2\}} \exp\left[ -\theta_{l_1,l_2} \delta^{ij}_{l_1,l_2} \right]}^{\text{Cross-population covariance}}$$

- Cross-population correlation is an exponential function of $\theta_{l_1,l_2}$: $r_{l_1,l_2} = \exp(-\theta_{l_1,l_2})$.
- Large value of $\theta_{l_1,l_2} \rightarrow$ low correlation between two populations.
- Separability between cross-population covariance and covariance over the Age-Year inputs.
- Observations across $L$ populations share same spatial covariance kernel: commonality via $\theta_{ag}$ and $\theta_{yr}$.
- Estimating the cross-covariance kernel requires $L(L-1)/2$ parameters $\theta_{l_1,l_2}$.

# Interlude

Multi-Output GP Framework: each population has its own surface $f_l$, inferred jointly

Everything looks like a nail:

- Cause-of-death: $x = (Year, AgeGroup, Cause)$
- Demographics: $x = (Year, Age, Gender, Demographic)$
- Birth Cohort effect: $x = (Year, Age, BirthYear)$
- Covid-19 Excess Deaths: $x = (Week, Year, AgeGroup)$
- Single-population vs joint modeling:
  - Credibility
  - Commonality
  - Tractability

# Dimension Reduction

# Coregionalization

- Curse of dimensionality is a major challenge
- Intrinsic coregionalization model (Alvarez et al 2011)
- The $\ell$-th surface $f_\ell$ is a linear comb. of $Q$ indep. latent GPs: $f_\ell(\mathrm{x}) = \sum_{q=1}^{Q} a_{\ell,q} u_q(\mathrm{x})$.
  - $a_{\ell,q}$'s: factor loadings & $u_1(\mathrm{x}), \ldots, u_Q(\mathrm{x})$: independent latent functions from a GP prior with covariance $C^{(u)}$.
- The ICM-MOGP covariance:

$$\mathrm{Cov}(\mathrm{f}(\mathrm{x}), \mathrm{f}(\mathrm{x}')) = \left(\sum_{q=1}^{Q} \mathrm{a}_q \mathrm{a}_q^T\right) \otimes C^{(u)}(\mathrm{x}, \mathrm{x}') = B \otimes C^{(u)}(\mathrm{x}, \mathrm{x}')$$

  - $B \in \mathbb{R}^{L \times L}$: cross-population covariance or *coregionalization matrix* with rank $Q$.
  - $\mathrm{C}^{(u)} \in \mathbb{R}^{N \times N}$: covariance over Age-Year inputs.
- Number of parameters in cross-population covariance is $Q \times L$.
  - $Q < L/2 \rightarrow$ reduce hyper-parameter space + alleviate computational budget.
  - $Q$ is chosen based on Bayesian Information Criterion (BIC).

# Semiparametric Latent Factor Model

- Teh et al 2004: allow for spatial heterogeneity

$$f_\ell(\mathrm{x}) = \sum_{q=1}^{Q} a_{\ell,q} u_q(\mathrm{x}).$$

where $u_q(\cdot)$ has covariance kernel $C_q^{(u)}(\mathrm{x}, \mathrm{x}')$.

- The covariance for $\mathrm{f}(\mathrm{x})$ is:

$$\mathrm{Cov}(\mathrm{f}(\mathrm{x}), \mathrm{f}(\mathrm{x}')) = \sum_{q=1}^{Q} A_q A_q^T C_q^{(u)}(\mathrm{x}, \mathrm{x}') = \sum_{q=1}^{Q} B_q C_q^{(u)}(\mathrm{x}, \mathrm{x}')$$

where $A_q = \mathrm{a}_q = (a_{1,q}, a_{2,q}, \ldots, a_{L,q})^T$ and each $B_q$ has rank one.

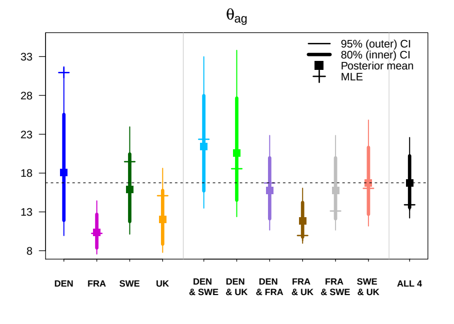
# Hierarchies

- Can further decompose if have several types of populations (e.g cause + country)
- Take $f_l(\mathrm{x}) = \sum_{i=1}^{PQ} a_{l,i} u_i(\mathrm{x})$ and $\mathrm{Cov}(\mathrm{f}(\mathrm{x}), \mathrm{f}(\mathrm{x}')) = B \otimes C^{(u)}(\mathrm{x}, \mathrm{x}')$
- Kronecker product: $B = C \otimes D$: $C \in \mathbb{R}^{L_c \times L_c}$ is the cross-country covariance and $D \in \mathbb{R}^{L_d \times L_d}$ is the cross-cause covariance: effective dimension $P + Q$
- Means $a_{l,i} = (A_c \otimes A_d)_{li}$ where $A_c$ are the factor loadings on countries and $A_d$ on cause
- $A_c = (\mathrm{e}_1, \ldots, \mathrm{e}_Q)$ w/vectors $\mathrm{e}_q = (e_{1,q}, \ldots, e_{L_c,q})^T$
- Hierarchical SLFM also possible
- Can handle up to 25 populations

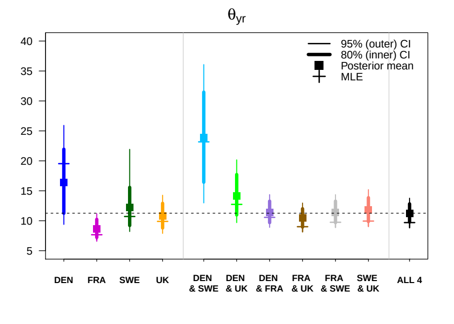
# Features of MOGP

# Global Mortality Structure

- Estimation of GP hyperparameters is challenging; MLE might be unstable
- Common dependence structure mitigates calibration errors
- A joint model fit on multiple populations has tighter hyper-parameter posteriors



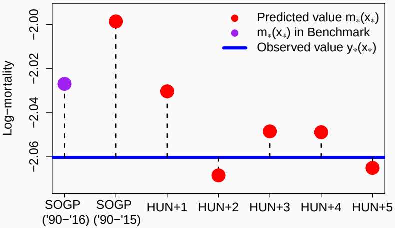

Length-scale in Age dimension




Length-scale in Year dimension

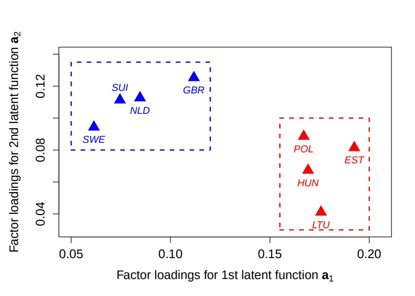
# Information Fusion/Clustering Populations

- Data from different countries arrives non-synchronously
- Borrow latest information from others
  - 2016 Hungarian data is missing; incorporate 1990-2016 neighboring data
  - Borrowing latest information from highly-corr. popns is better than having latest domestic data




Prediction errors for 84-year-old Hungarian Males

- Factor loadings provide insight on dependence across populations:
  - Two well-separated clusters among 8 countries in this example.
  - Countries within the same cluster are more correlated.

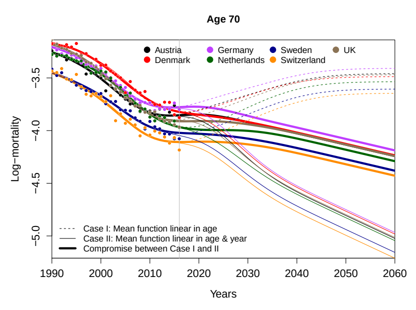


# Coherent forecasts

| SMAPE | 2013 (1-yr out) | | 2016 (4-yr out) | |
|---|---|---|---|---|
| | Single-pop | Multi-pop | Single-pop | Multi-pop |
| Denmark | 1.58 | **1.52** | 1.26 | **1.22** |
| Sweden | 1.05 | **0.82** | 2.53 | **0.83** |

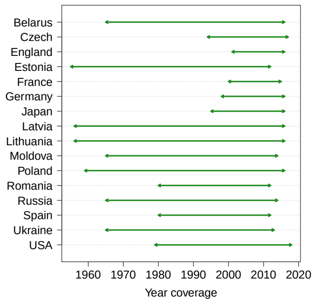
- Point forecasts assessed via mean absolute percentage error
- Probabilistic forecasts assessed via CRPS
- Multi-population models tend to reduce posterior variance

- Mortality across populations moves in unison over time
- Forecasts via MOGP maintain historical characteristics
- Extrapolation reverts to the prior $m_*(x) \to m(x)$
- Transition controlled by $\theta_{yr}$



# Cause of Death

- causeofdeath.org
- Causes based on ICD codes
- By-cause patterns are heterogenous
- Data much more noisy: fusion is important to see the forest for the trees



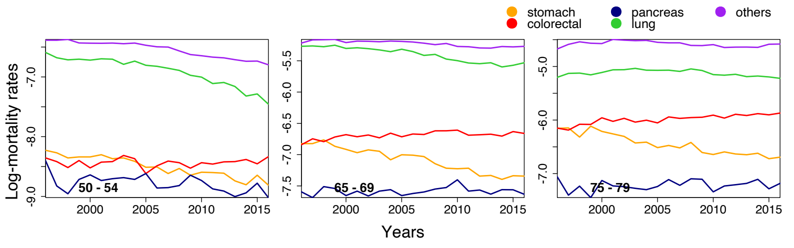
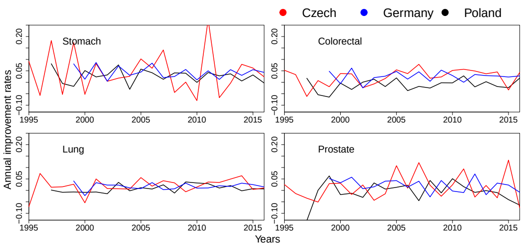



**Figure 1:** Raw Log-mortality rates in common variations of cancer in Poland. Lung and Colorectal cancers are the leading cause of cancer deaths while the prevalence of Pancreas and Stomach cancers are relatively lower.

Raw YoY improvement rates:



# Smoothed Improvement Rates

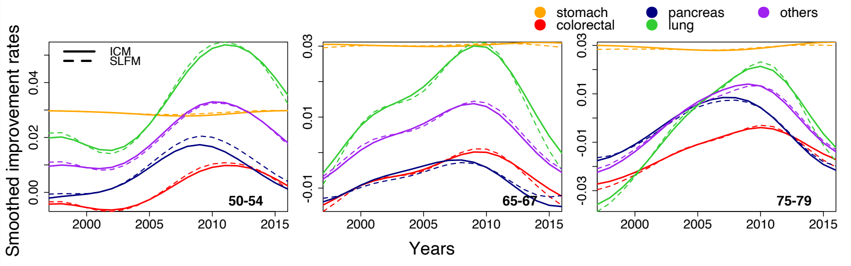
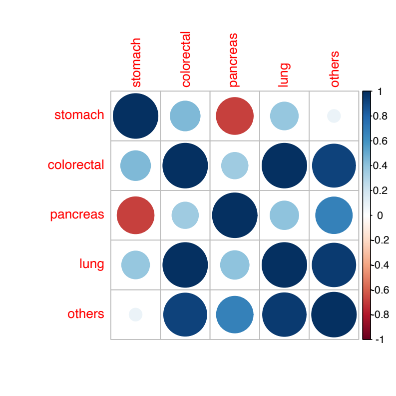
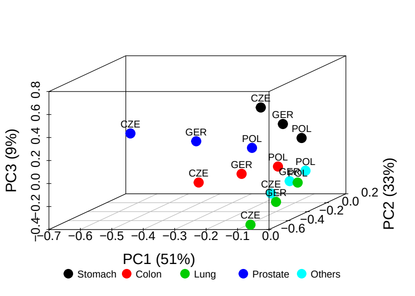



**Figure 2:** Smoothed YoY improvement rates for Poland via MOGP models by countries and age groups. Stomach cancer has the largest improvement rate.

# Correlation across Causes

- Negative correlation is empirically possible
- Not necessarily expect coherence/long-term convergence
- Causes vs countries dependence



# Summary

- Hierarchical GP approach to handle tensor mortality data
- Natural framework for enforcing/uncovering common dependence structure
- Investigate different dimension reduction techniques
- Wide scope for further enhancements:
  - Kernel selection
  - Trend modeling
  - Noise modeling
  - Nonstationary covariance
  - Approximation for large datasets (eg Kronecker structure)
  - Multiple software implementations (R, Python)

# Summary

- Hierarchical GP approach to handle tensor mortality data
- Natural framework for enforcing/uncovering common dependence structure
- Investigate different dimension reduction techniques
- Wide scope for further enhancements:
  - Kernel selection
  - Trend modeling
  - Noise modeling
  - Nonstationary covariance
  - Approximation for large datasets (eg Kronecker structure)
  - Multiple software implementations (R, Python)

**Thank You!**

# References


**Williams, C. K. and Rasmussen, C. E. 2006.**
*Gaussian processes for machine learning*, the MIT Press.

**M. Ludkovski, J. Risk, H. Zail**
Gaussian Process Models for Mortality Rates and Improvement Factors
ASTIN Bulletin, 48(3), pp. 1307–1347, 2018
Reproducible R notebook:

**N. Huynh, M. Ludkovski**
Multi-Output Gaussian Processes for Multi-Population Longevity Modeling
Annals of Actuarial Science, to Appear, 2021 arXiv:2003.02443

**N. Huynh, M. Ludkovski, H. Zail**
Multipopulation Longevity Analysis: a Spatial Random Field Approach
SOA 2020 Living to 100 Symposium


RShiny apps and Tutorials:

- https://nhanhuynh46.github.io/MOGPTutorials/
- https://rosalia1010.shinyapps.io/COVID19_ExcessDeaths/
- https://rosalia1010.shinyapps.io/Longevity_Forecasting_Tool/
- github.com/jimmyrisk/GPmortalityNotebook