Genetic Algorithm Applications of Gaussian Process Kernels toward Mortality Surface Inference

RCLR Modelling and Societal Impact of Longevity and Ageing

Dr. Jimmy Risk
Cal Poly Pomona

5/25/23

Preliminaries

- Model Log Mortality Rate

$$\mathbf{x} = (x_{age}, x_{year}, x_{cohort})^1 \qquad y_{\mathbf{x}} = \log\left(\frac{D_{\mathbf{x}}}{L_{\mathbf{x}}}\right) = f(\mathbf{x}) + \epsilon_{\mathbf{x}}$$

- y, D, L are observed
- $f(\cdot)$ drives the underlying mortality dynamics (unknown)
- ϵ_x is unaccountable (white) noise

[1] $x_{cohort} = x_{year} - x_{age}$ is year of birth

LRZ Mortality Models

- **Ludkovski-Risk-Zail** (2018) $\boxed{f = \mathcal{GP}(m, k)}$ is a **Gaussian Process**

$$\mathbb{E}[f(\mathbf{x}_i)] = m(\mathbf{x}_i), \qquad \text{(prior mean function)}$$
$$\operatorname{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = k(\mathbf{x}_i, \mathbf{x}_j), \qquad \text{(covariance kernel)}$$

Definition

$\{f(\mathbf{x})\}_{x \in \mathbb{R}^d}$ is a **Gaussian process** if for all $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$, the vector $[f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)]^\top$ has a **multivariate normal distribution**.

Gaussian Process Regression

- Gaussian process regression assumes $\epsilon(\mathbf{x}) \overset{iid}{\sim} N(0, \sigma^2)$

$$y(\mathbf{x}) = f(\mathbf{x}) + \epsilon(\mathbf{x}) \qquad \Rightarrow^{2} \qquad \{y(\mathbf{x}_i)\}_{i=1}^{\infty} \sim \mathcal{GP}(m, k + \sigma^2 \delta_{i=j})$$

- Denote $\mathbf{y} = [y(\mathbf{x}_1), \cdots, y(\mathbf{x}_n)]^\top$
- Properties of MVN[3] imply that conditional on observed $\mathbf{y}$,

$$\boxed{f_* := f_* | y \sim \mathcal{GP}(m_*, k_*)} \qquad \text{(posterior GP)}$$

where

$$m_*(\mathbf{x}) = K(\mathbf{x}, \mathbf{X}) \left[K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_n\right]^{-1} \mathbf{y}, \tag{1}$$

$$k_*(\mathbf{x}, \mathbf{x}') = K([\mathbf{x}, \mathbf{x}']^\top, \mathbf{X}) \left[K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_n\right]^{-1} K(\mathbf{X}, [\mathbf{x}, \mathbf{x}']^\top), \tag{2}$$

[2] since f is a GP

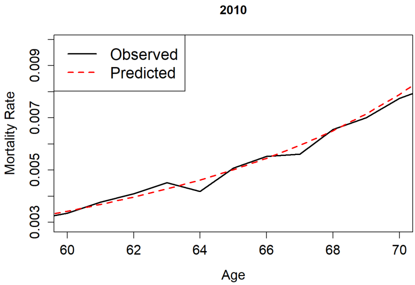
[2] since f is a GP

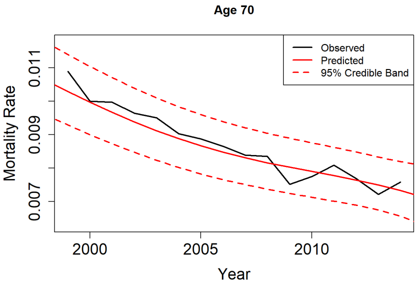
[3] Or, using Bayes' theorem

LRZ 2018 Smoothing

$$m(\mathbf{x}) = \beta_0 + \beta_{ag} x_{ag} \tag{3}$$

$$k(\mathbf{x}, \mathbf{x}') = \eta^2 \exp\left(-\frac{(x_{ag} - x'_{ag})^2}{2\ell_{ag}^2} - \frac{(x_{yr} - x'_{yr})^2}{2\ell_{yr}^2}\right) \tag{4}$$





Kernels

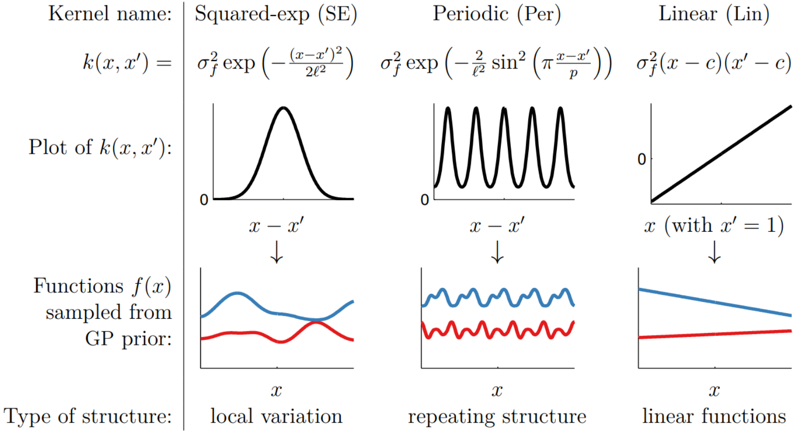
- Covariance functions are valid Kernel functions[4] and vice versa
- Govern the underlying properties of the prior and posterior process f
- This can be seen e.g. through the posterior mean:

$$
\begin{aligned}
m_*(\cdot) &= K(\cdot, \mathbf{X}) \underbrace{\left[K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_n\right]^{-1} \mathbf{y}}_{:=\mathbf{c}}, \\
&= \sum_{i=1}^{n} c_i k(\cdot, \mathbf{x}_i)
\end{aligned}
$$

- Governs smoothness, periodicity, etc., of process itself

[4] e.g. used in support vector machines, kernel ridge regression

A Few Basic Kernels



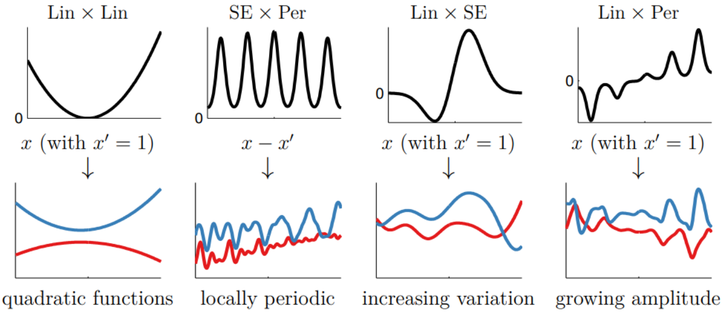
[5]

[5]Credit to Automatic Model Construction with Gaussian Processes by Duvenaud https://www.cs.toronto.edu/~duvenaud/thesis.pdf

Kernel Algebra

- The space of positive definite functions enjoy a rich collection of algebraic properties

$$k_1, k_2 \text{ p.d. kernels} \quad \Rightarrow \quad \begin{cases} k_1 + k_2 \text{ is a p.d. kernel} & \text{(closed under addition)} \\ k_1 \cdot k_2 \text{ is a p.d. kernel} & \text{(closed under mult.)} \end{cases}$$



[6]

[6] Credit to Automatic Model Construction with Gaussian Processes by Duvenaud
https://www.cs.toronto.edu/~duvenaud/thesis.pdf

Kernels in the Mortality Setting

- Kernel use already exists historically *(possibly unknowingly)*
- E.g. time series covariance functions define kernels!

Ex. Cairns-Blake-Dowd (CBD) (2006)

$$\boxed{f(\mathbf{x}) = \kappa^{(1)}_{x_{yr}} + (x_{ag} - \overline{x_{ag}})\kappa^{(2)}_{x_{yr}}}$$

If $(\kappa^{(1)}_{x_{yr}}, \kappa^{(2)}_{x_{yr}})$ is multivariate random walk, then

$$\begin{aligned} k_{CBD}(\mathbf{x}, \mathbf{x}') = \operatorname{cov}(f(\mathbf{x}), f(\mathbf{x}')) = \\ = \big[\sigma_1^2 + \rho\sigma_1\sigma_2(x_{ag} + x'_{ag} - 2\overline{\mathbf{x}_{ag}}) \\ +(x_{ag} - \overline{\mathbf{x}_{ag}})(x'_{ag} - \overline{\mathbf{x}_{ag}})\sigma_2^2\big] \cdot \min\left(x_{yr}, x'_{yr}\right). \end{aligned}$$

Kernels in the Mortality Setting (Part II)

Definition

A *separable* kernel over $\mathbb{R}^d$ is one that can be written as a product: $k(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^{d} k_j(x^{(j)}, x'^{(j)})$, where $x^{(j)}$ is the jth coordinate of $\mathbf{x}$.

- The global kernel is separated as a product over its dimensions, each having its own kernel.
- Provides a rich interpretation

Earlier example:

$$k(\mathbf{x}, \mathbf{x}') = \eta^2 \exp\left(-\frac{(x_{ag} - x'_{ag})^2}{2\ell_{ag}^2} - \frac{(x_{yr} - x'_{yr})^2}{2\ell_{yr}^2}\right)$$

$$= k_{RBF}(x_{ag}, x'_{ag}; \ell ag) \cdot k_{RBF}(x_{yr}, x'_{yr}; \ell_{yr})$$

- Two smooth components in each coordinate;
- Correlations decay rapidly as age distance increases;
- Similar for year.

Kernels in the Mortality Setting (Part II)

Definition

A *separable* kernel over $\mathbb{R}^d$ is one that can be written as a product: $k(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^{d} k_j(x^{(j)}, x'^{(j)})$, where $x^{(j)}$ is the jth coordinate of $\mathbf{x}$.

- The global kernel is separated as a product over its dimensions, each having its own kernel.
- Provides a rich interpretation

Earlier example:

$$k(\mathbf{x}, \mathbf{x}') = \eta^2 \exp\left(-\frac{(x_{ag} - x'_{ag})^2}{2\ell_{ag}^2} - \frac{(x_{yr} - x'_{yr})^2}{2\ell_{yr}^2}\right)$$
$$= k_{RBF}(x_{ag}, x'_{ag}; \ell ag) \cdot k_{RBF}(x_{yr}, x'_{yr}; \ell_{yr})$$

- Two smooth components in each coordinate;
- Correlations decay rapidly as age distance increases;
- Similar for year.

What Kernel to Use?

- For many tasks, precise kernel is of secondary importance
- Default of multiplicative (separable) stationary kernels are reasonable starting point; sufficient for many tasks
- For some tasks, the kernel plays a more significant role:
  - Longer-range extrapolation
  - Short-term mortality improvement (kernel smoothness dramatically affects results)
- Interested in mortality dependence structure from fitting GP models:
  - Smoothness of mortality experience across Age and Year
  - Models for Cohort effects
  - Additive structures
  - Comparing populations (how does discovered structure vary among country/gender)

Too Many Kernels

Kernel Name	Abbv.	Formula $k(x, x'; \theta)$	Properties
Matérn-1/2	M12	$\exp\left(-\frac{\lvert x-x'\rvert}{\ell_{\text{len}}}\right), \quad \ell_{\text{len}} > 0$	C^0
Matérn-3/2	M32	$\left(1 + \frac{\sqrt{3}}{\ell_{\text{len}}}\lvert x - x'\rvert\right) \exp\left(-\frac{\sqrt{3}}{\ell_{\text{len}}}\lvert x - x'\rvert\right), \quad \ell_{\text{len}} > 0$	C^1
Matérn-5/2	M52	$\left(1 + \frac{\sqrt{5}}{\ell_{\text{len}}}\lvert x - x'\rvert + \frac{5}{3\ell_{\text{len}}^2}\lvert x - x'\rvert^2\right) \exp\left(-\frac{\sqrt{5}}{\ell_{\text{len}}}\lvert x - x'\rvert\right)$	C^2
Cauchy	Chy	$\frac{1}{1+\lvert x-x'\rvert^2/\ell_{\text{len}}^2}, \quad \ell_{\text{len}} > 0$	C^∞
Radial Basis	RBF	$\exp\left(-\frac{(x-x')^2}{2\ell_{\text{len}}^2}\right), \quad \ell_{\text{len}} > 0$	C^∞
AR2	AR2	$\exp(-\alpha\lvert x - x'\rvert)\left\{\cos(\omega\lvert x - x'\rvert) + \frac{\alpha}{\omega}\sin(\omega\lvert x - x'\rvert)\right\}$	Periodic, C^1
Linear	Lin	$\sigma_0^2 + x \cdot x', \quad \sigma_0 > 0$	Non-stationary
Minimum	Min	$t_0^2 + \min(x, x'), \quad t_0 > 0$	Non-stat, C^0
Mehler	Meh	$\exp\left(-\frac{\rho^2(x^2+x'^2)-2\rho x x'}{2(1-\rho^2)}\right), \quad -1 \le \rho \le 1$	Non-stationary

- Can use x_{ag}, x_{yr}, **or** x_{co} as input
- Each offers a different mortality interpretation
- Can combine kernels with addition and multiplication

Too many kernels!

Genetic Algorithm for Kernels

Employ genetic algorithm for kernel search

1. **(Generation 0)** Randomly generate n_g kernels uniformly[7]
2. **(Generation** g**,** $g > 0$**)** For n_g times:
   - Randomly sample T **parents** from the $i-1$th generation
   - Mutate or crossover the "**fittest**"[8] parents
   - **Offspring** inserted into generation g
3. Repeat for all $g = 1, 2, \ldots, G$

Parameter	Value	Notes
Population Size	$n_g = 200$	Number of individuals per generation
Generations	$G = 20$	Number of generations
Tournament Size	$T = 7$	Run double tournament; select smaller winner w/ prob. $D/2$
	$D = 1.2$	Smaller is winner with probability 0.60

[7] Choosing 2–7 base kernels and coordinates, and combining with $+$ or $\cdot$ randomly

[8] Using BIC as the criterion

Genetic Algorithm for Kernels

Employ genetic algorithm for kernel search

1. **(Generation 0)** Randomly generate n_g kernels uniformly[7]
2. **(Generation g, $g > 0$)** For n_g times:
   - Randomly sample T **parents** from the $i-1$th generation
   - Mutate or crossover the "**fittest**"[8] parents
   - **Offspring** inserted into generation g
3. Repeat for all $g = 1, 2, \ldots, G$

Parameter	Value	Notes
Population Size	$n_g = 200$	Number of individuals per generation
Generations	$G = 20$	Number of generations
Tournament Size	$T = 7$	Run double tournament; select smaller winner w/ prob. $D/2$
	$D = 1.2$	Smaller is winner with probability 0.60

[7] Choosing 2–7 base kernels and coordinates, and combining with $+$ or $\cdot$ randomly
[8] Using BIC as the criterion

Genetic Algorithm for Kernels

Employ genetic algorithm for kernel search

1. **(Generation 0)** Randomly generate n_g kernels uniformly[7]
2. **(Generation g, $g > 0$)** For n_g times:
   - Randomly sample T **parents** from the $i-1$th generation
   - Mutate or crossover the "**fittest**"[8] parents
   - **Offspring** inserted into generation g
3. Repeat for all $g = 1, 2, \ldots, G$

Parameter	Value	Notes
Population Size	$n_g = 200$	Number of individuals per generation
Generations	$G = 20$	Number of generations
Tournament Size	$T = 7$	Run double tournament; select smaller winner w/ prob. $D/2$
	$D = 1.2$	Smaller is winner with probability 0.60

[7] Choosing 2–7 base kernels and coordinates, and combining with $+$ or $\cdot$ randomly

[8] Using BIC as the criterion

GA Operations

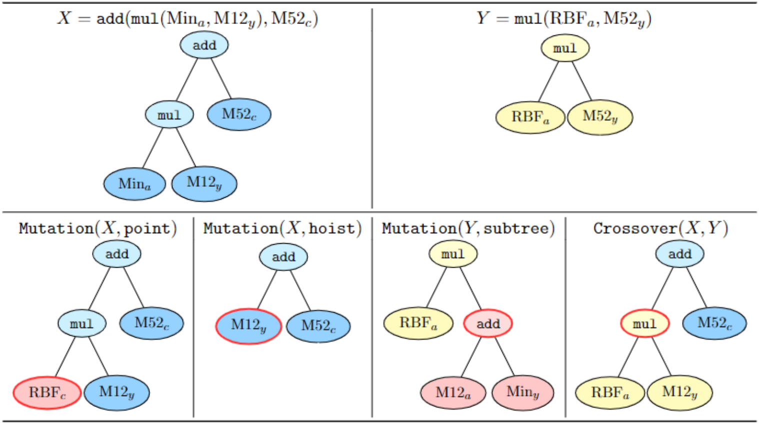


Figure 1: Representative compositional kernels and GA operations. Bolded red ellipses indicate the node of X (or Y) that was chosen for mutation or crossover.

GA Specific Hyperparameters

Probability	GA Operation	Notes
$p_c = 0.45$	Crossover	
$p_s = 0.2$	Subtree Mutation	
$p_h = 0.1$	Hoist Mutation	
$p_p = 0.05$	Point Mutation	Each node is mutated with another node of same arity with prob. q_p
$p_r = 0.15$	Respectful Point Mutation	Each node is mutated with another node of same arity and same $(\text{age}, \text{year}, \text{cohort})$ with prob. q_r
$p_o = 0.05$	Copy	
$q_p = 0.25$	Point Replace	
$q_r = 0.35$	Respectful Replace	
$q_a = 0.5$		Probability that add/mul is included when initializing trees

Table: Operator specific GA hyperparameters with description. Note that $p_c + p_s + p_h + p_p + p_r + p_0 = 1$.

Synthetic Experiments

- Generate synthetic mortality data with specified k_0
- Run GA to see if k_0 can be recovered

Exprmnt	Ground Truth Kernel (k_0)	$\sigma^2(\mathbf{x})$	β_0	β_{ag}
SYA	$0.04 \cdot \text{RBF}_a(13.6) \cdot \text{RBF}_y(9.0)$	0.001	-11.7	0.1
SYB	$0.08 \cdot \text{RBF}_a(19.93) \cdot M12_y(400) + 0.02 \cdot M52_c(5)$	0.0004	-11.4276	0.0875
SYC	$0.0134 \cdot M52_a(38.49) \cdot \text{Min}_y(26.33) \cdot M12_c(6062.76) \cdot \cdot \text{Meh}_c(0.8483)$	$1.0783/D_{\mathbf{x}}$	-12.58	0.0994

Table: Description of synthetic data sets. Data is generated as multivariate normal realizations with parametric mean function $m(\mathbf{x}) = \beta_0 + \beta_{ag} x_{ag}$. SYA and SYB are homoskedastic. In generating SYC's heteroskedastic noise, $D_{\mathbf{x}}$ comes from the JPN Female data.

GA Convergence

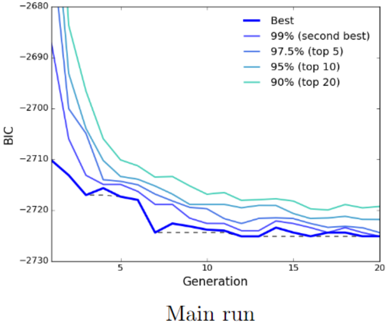


Figure: GA convergence for JPN Female run (later). Convergence for synthetic experiments was quicker than these plots show.

SYA Results

SYA-1			SYA-2		
BIC	$\widehat{\text{BF}}(k, k_0)$	Kernel	BIC	$\widehat{\text{BF}}(k, k_0)$	Kernel
-2034.23	**1.0000*** **	$\mathbf{RBF_a RBF_y}$	-2066.93	1.1907***	$\text{M52}_a\,\text{RBF}_y$
-2034.04	0.8264***	$\text{M52}_a\,\text{RBF}_y$	**-2066.76**	**1.0000*** **	$\mathbf{RBF_y RBF_a}$
-2031.82	0.0902*	$\text{M52}_a\text{M52}_y$	-2064.63	0.1216**	$\text{M52}_a\text{M52}_a\,\text{RBF}_y$
-2031.29	0.0526*	$\text{M52}_a\,\text{RBF}_a\,\text{RBF}_y$	-2064.24	0.0801*	$\text{M52}_a\,\text{RBF}_a\,\text{RBF}_y$
-2031.09	0.0433*	$\text{M52}_a\text{M52}_a\,\text{RBF}_y$	-2063.88	0.0561*	$\text{M52}_a\text{M52}_a\,\text{RBF}_y$

Table: Top five fittest non-duplicate kernels for the first synthetic case study SYA. Bolded is $k_0 = \text{RBF}_y\,\text{RBF}_a$, the true kernel used in data generation. SYA-1 and -2 denote the realization trained on.

- **SYA:** $k_0 = \text{RBF}_y\,\text{RBF}_a$
- Goals:
  - Generate two surfaces to check consistency
  - Investigate mortality "smoothness"

Synthetic Results (cont'd)

- SYB results: correctly identify age · year + cohort
- SYC results: correctly identify # terms, multiplicative structure, nonstationarity
- Closely recover ground truth GP hyperparameters
- Observe substitution effect (BIC-wise plausible alternatives)

Applications

All data is from HMD[9] using years 1990–2018 and ages 50–84

1. Japan Female $\mathcal{K}_f$
   - Analyze smoothness, nonstationary, additive structure, etc.
   - Robustness: Rerun
   - Compare with $\mathcal{K}_r$ (*is remainder* $\mathcal{K}_f \backslash \mathcal{K}_r$ *needed?*)
   - Robustness: Expand to 1988-2018 and 48–86
   - Compare with Japan Males
2. US Males
3. Sweden Females
4. Compare across all country+gender pairs analyzed
5. Necessity of cohort

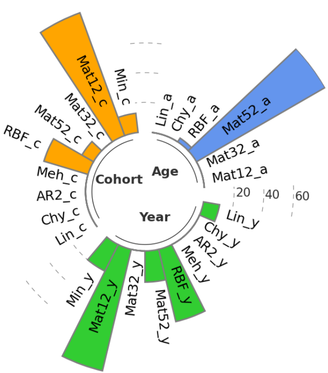
[9] Human mortality database `https://www.mortality.org/`

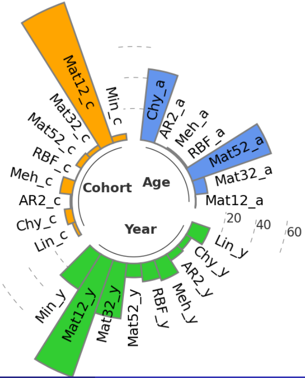
Japanese Female Preliminary Results

($G = 20$ generations, $n_g = 200$ individuals per generation, analyze over $20 \cdot 200 = 4000$ results)

Japan Female HMD Dataset for 1990-2018 and Ages 50-84

$\mathcal{K}_r$			$\mathcal{K}_f$		
BIC	$\widehat{\text{BF}}$	Kernel	BIC	$\widehat{\text{BF}}$	Kernel
-2725.288	0.995	$M52_a(RBF_y\ M12_y)M12_c$	-2725.293	1	$M52_a(Chy_y\ M12_y)M12_c$
-2725.270	0.977	$M52_a(M52_y M12_y)M12_c$	-2725.270	0.977	$M52_a(M52_y M12_y)M12_c$
-2725.233	0.941	$M52_a(RBF_y\ Min_y)M12_c$	-2725.221	0.931	$M52_a(M52_y\ Min_y)M12_c$
-2725.221	0.931	$M52_a(M52_y\ Min_y)M12_c$	-2724.623	0.512	$M52_a(M52_y M12_y)\ Min_c$
-2724.640	0.520	$M52_a(M52_y M12_y)\ Min_c$	-2724.510	0.457	$M52_a(M32_y M12_c)M12_c$





Japanese Female Top Kernel Posterior Forecasts

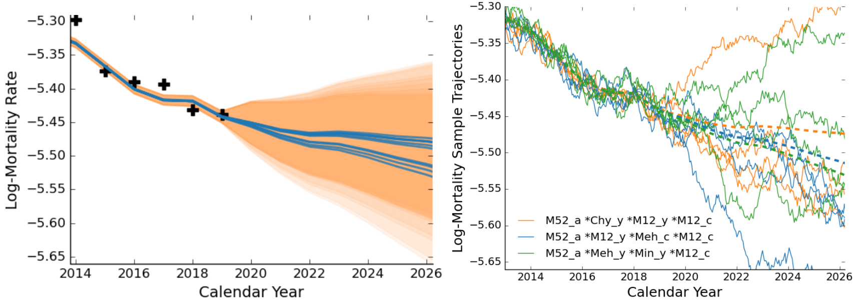


Figure: Predictions from the top 10 kernels in $\mathcal{K}_f$ for JPN Females Age 65. *Left:* predictive mean and 90% posterior interval from the top-10 kernels. For comparison we also display (black plusses) the 5 observed log-mortality rates during 2014–2019. *Right:* 4 sample paths from 3 representative kernels.

Japanese Female Best Kernel Heatmaps

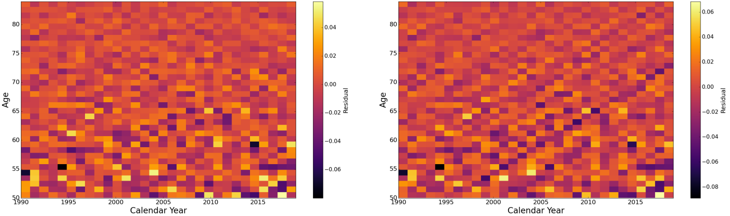


Figure: Left: residuals from the best kernel in $\mathcal{K}_f$. Right: residuals from a run that removed all cohort kernels.

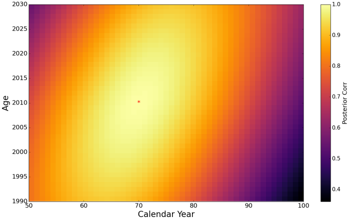


Figure: Implied prior correlation of the best kernel.

Japanese Female Comparative Results

Range	BIC max	BIC min	length	addtv. comp's	non-stat.	all age	all year	all coh	rough age	rough year	rough cohort
JPN Female											
1-10	-2723.68	-2725.29	4.00	1.00	0%	1.00	1.80	1.20	0%	100%	100%
1-50	-2720.64	-2725.29	4.34	1.08	10%	1.12	1.90	1.32	0%	100%	100%
51-100	-2718.24	-2720.62	4.60	1.20	18%	1.12	2.20	1.28	0%	100%	100%
101-150	-2717.03	-2718.17	5.02	1.14	4%	1.30	2.18	1.54	6%	98%	100%
151-200	-2715.77	-2717.01	5.10	1.48	12%	1.28	2.36	1.46	6%	100%	100%
JPN Female Rerun											
1-10	-2723.05	-2725.29	4.00	1.00	0%	1.00	1.60	1.40	0%	100%	100%
1-50	-2719.82	-2725.29	4.14	1.08	10%	1.08	1.58	1.48	0%	98%	100%
51-100	-2718.30	-2719.82	4.46	1.26	8%	1.14	1.62	1.70	6%	100%	100%
JPN Female Search in $\mathcal{K}_r$											
1-10	-2724.11	-2725.27	4.00	1.00	0%	1.00	1.70	1.30	0%	100%	100%
1-50	-2721.19	-2725.27	4.48	1.10	8%	1.14	1.96	1.38	0%	100%	100%
51-100	-2718.06	-2721.19	4.72	1.50	18%	1.16	1.96	1.60	0%	100%	100%
JPN Female trained on $\mathcal{D}_{rob}$											
1-10	-2724.11	-2725.29	4.00	1.40	40%	1.00	1.50	1.50	0%	100%	100%
1-50	-2716.84	-2725.29	4.42	1.12	18%	1.14	1.64	1.64	0%	100%	100%
51-100	-2714.96	-2716.58	4.70	1.16	12%	1.18	1.68	1.84	0%	100%	100%

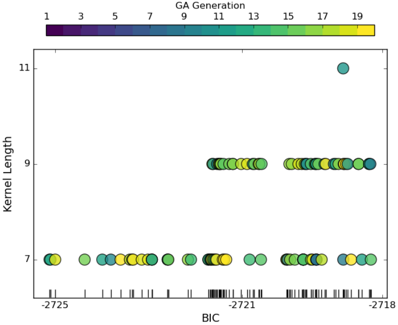
Table: Results for various settings over entire GA using $G = 20$ generations and $n_g = 200$ kernels per generation. Range indicates the top results over the $20 \cdot 200 = 4000$ results obtained.

Full Comparison

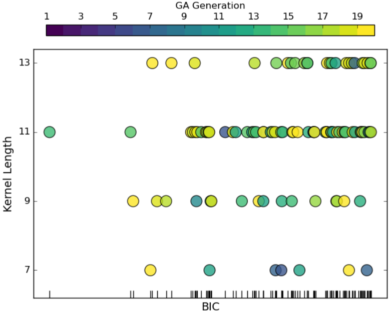
Range	BIC max	BIC min	length	addtv. comp's	non-stat.	all age	all year	all coh	rough age	rough year	rough cohort
					JPN Female						
1-10	-2723.68	-2725.29	4.00	1.00	0%	1.00	1.80	1.20	0%	100%	100%
1-50	-2720.64	-2725.29	4.34	1.08	10%	1.12	1.90	1.32	0%	100%	100%
51-100	-2718.24	-2720.62	4.60	1.20	18%	1.12	2.20	1.28	0%	100%	100%
					JPN Male						
1-10	-2978.43	-2980.53	4.10	1.00	0%	1.00	1.60	1.50	0%	100%	100%
1-50	-2975.36	-2980.53	4.26	1.10	0%	1.06	1.70	1.50	18%	100%	100%
51-100	-2974.25	-2975.32	4.60	1.00	0%	1.04	2.14	1.42	64%	100%	100%
					US Male						
1-10	-3163.54	-3170.29	5.70	2.30	0%	1.50	1.50	2.70	100%	100%	100%
1-50	-3160.32	-3170.29	5.78	2.24	0%	1.40	1.54	2.84	100%	100%	100%
51-100	-3157.93	-3160.24	6.14	2.38	2%	1.46	1.72	2.96	100%	100%	98%
					SWE Female						
1-10	-1624.34	-1625.57	3.00	1.00	0%	1.00	1.00	1.00	0%	100%	0%
1-50	-1622.74	-1625.57	3.02	1.00	6%	1.00	1.24	0.78	0%	100%	14%
51-100	-1622.04	-1622.74	3.42	1.04	16%	1.10	1.38	0.94	0%	100%	6%

Table: Results from GA runs on JPN Male, US Male and SWE Female. Throughout we search within the full set $\mathcal{K}_f$.

JPN Female vs USA Male Kernel Length Properties



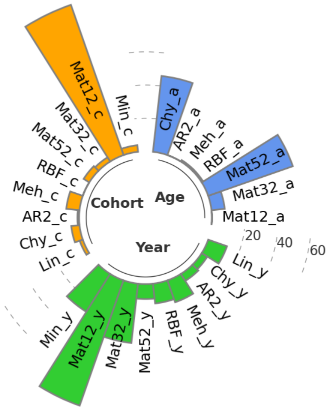
JPN Female



USA Male

Figure: Properties of the top 100 kernels found by GA.

Comparison Across Gender (Frequency of Appearance)



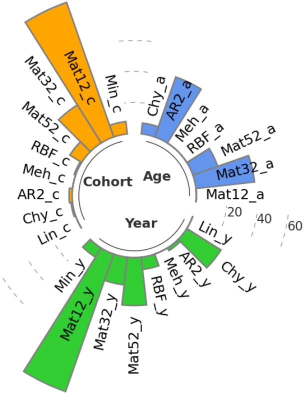


Figure: Frequency of appearance of different kernels from $\mathcal{K}_f$

JPN Female | US Male | SWE Female

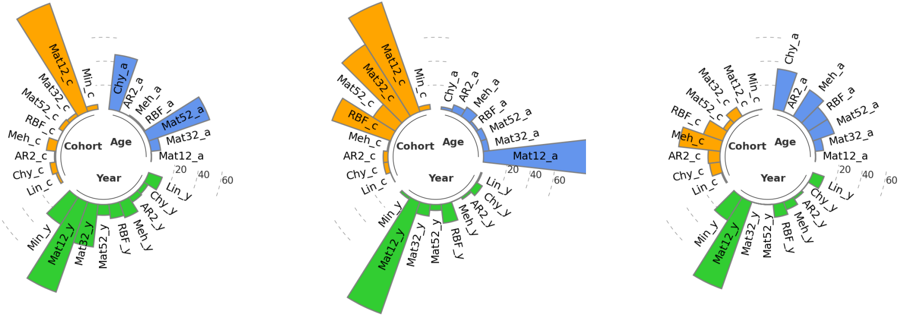


Figure: Frequency of appearance of different kernels from $\mathcal{K}_f$

Takeaways

- GP-GA works to discover structures in mortality surfaces.
- Real world data contains: smooth age, rough year, rough cohort
- "Residual kernel" often present
- GA is confident in overall structure (e.g. purely multiplicative, size)
    - US Males contain exceptions
- Kernel replacement effect exists. E.g.
    - Min vs M12
    - Chy vs RBF vs M52

Future Work

- Jointly model countries using **multi-output Gaussian processes**
- Kernel choices
  - Refining $\mathcal{K}_f$
  - Including more kernels *(express more structure)*

$$\exp(-[x_{ag}, x_{yr}]^\top A[x_{ag}, x_{yr}]) \qquad \text{(non-separable example)}$$

$$\sigma(x)k_1(x,x')\sigma(x') + \overline{\sigma}(x)k_2(x,x')\overline{\sigma}(x') \qquad \text{(changepoint kernel)}$$

- Analyze limitations of GA and hyperparameters

Thank You!

Our work:

- Risk, Jimmy and Ludkovski, Mike. "Expressive Mortality Models through Gaussian Process Kernels". *Working version available on request.*
- Ludkovski, Mike, Jimmy Risk, and Howard Zail. "Gaussian process models for mortality rates and improvement factors." *ASTIN Bulletin: The Journal of the IAA 48.3 (2018): 1307-1347.*

Top Kernels Across Countries

Pop'n/Search Set	N_{pl}	Top Kernel
JPN Female $\mathcal{K}_r$	90	$0.464 \cdot \text{M52}_a(1.1) \cdot \text{RBF}_y(1.33)\text{M12}_y(62.51) \cdot \text{M12}_c(118.06)$
JPN Female $\mathcal{K}_f$		$0.4638 \cdot \text{M52}_a(1.11) \cdot \text{Chy}_y(1.95)\text{M12}_y(62.42) \cdot \text{M12}_c(117.11)$
JPN Male $\mathcal{K}_r$	89	$0.1491 \cdot \text{M52}_a(0.95) \cdot \text{RBF}_y(1.15)\text{M12}_y(26.24) \cdot \text{M12}_c(24.90)$
JPN Male $\mathcal{K}_f$	112	$0.2130 \cdot \text{M52}_a(1.09) \cdot \text{M12}_y(39.09) \cdot M32_c(0.86)\text{M12}_c(40.73)$
US Male $\mathcal{K}_r$	57	$0.017 \cdot \text{M12}_a(5.04) \cdot \text{M52}_y(0.50)M12_y(10.33) \cdot \text{M52}_c(0.36)\text{M12}_c(5.00)$
US Male $\mathcal{K}_f$	35	$0.01 \cdot \text{AR2}_a(1.12, 1.88) \cdot \text{M12}_y(24.18) \cdot \text{M32}_c(0.72) \cdot [4.6211 \cdot \text{M12}_c(13.49) + 0.01 \cdot \text{M32}_a(0.02) \cdot \text{M52}_c(0.1)]$
SWE Female $\mathcal{K}_r$	200+	$0.2527 \cdot \text{RBF}_a(0.52) \cdot \text{M12}_y(73.74) \cdot \text{RBF}_c(0.62)$
SWE Female $\mathcal{K}_f$	200+	$0.2094 \cdot \text{Chy}_a(1.05) \cdot \text{M12}_y(67.27) \cdot \text{Meh}_c(0.60)$

Table: Best performing kernel in $\mathcal{K}_r$ and $\mathcal{K}_f$ for each of the 4 populations considered. N_{pl} is the number of alternate kernels that have a BIC within 6.802 of the top kernel and hence are judged "plausible" based on the BF criterion.