Automatic Dynamic Relevance Determination
for Gaussian process regressions with functional inputs

Luis Damiano[†1], Joaquim Teixeira[2], Margaret Johnson[2], Jarad Niemi[1]

[1]Department of Statistics, Iowa State University
[2]NASA Jet Propulsion Laboratory

SIAM Conference on Uncertainty Quantification
April 13th, 2022

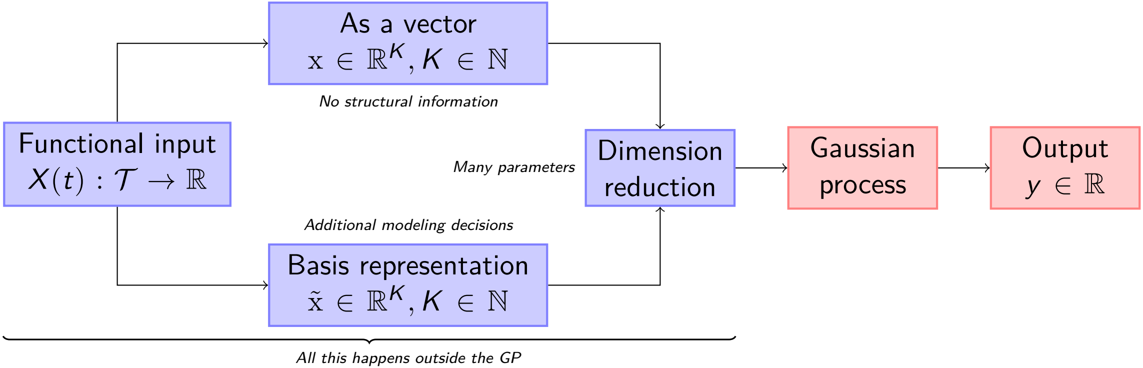
Funded, in part, by

- ISU Presidential Interdisciplinary Research Initiative on C-CHANGE: Science for a Changing Agriculture
- Foundation for Food and Agriculture Research Grant ID: CA18-SS-0000000278

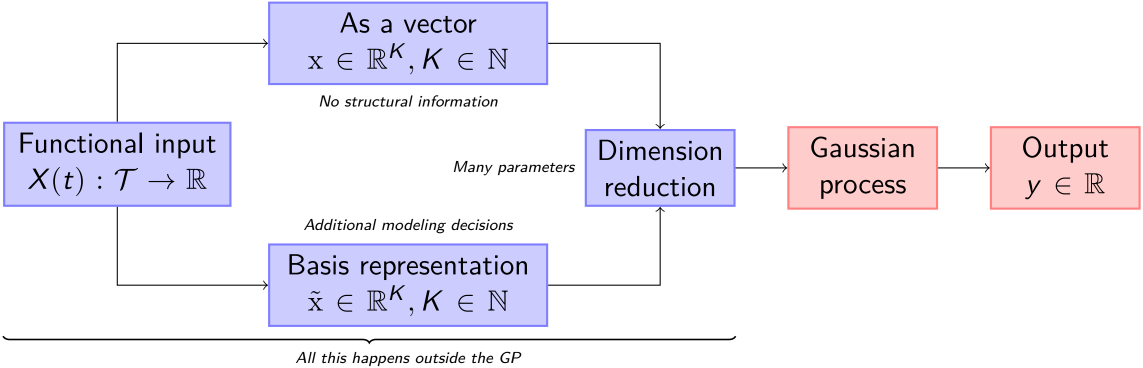
[†] ldamiano@iastate.edu

Overview & motivation

Gaussian process with functional input

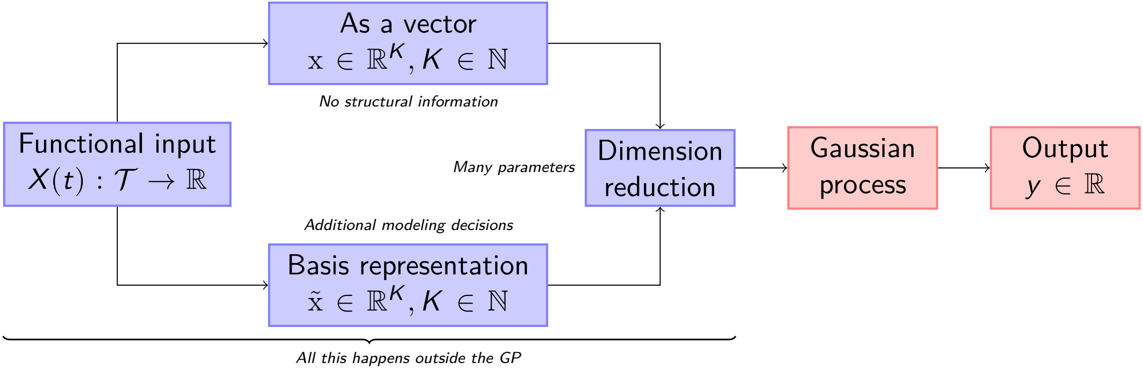


Gaussian process with functional input



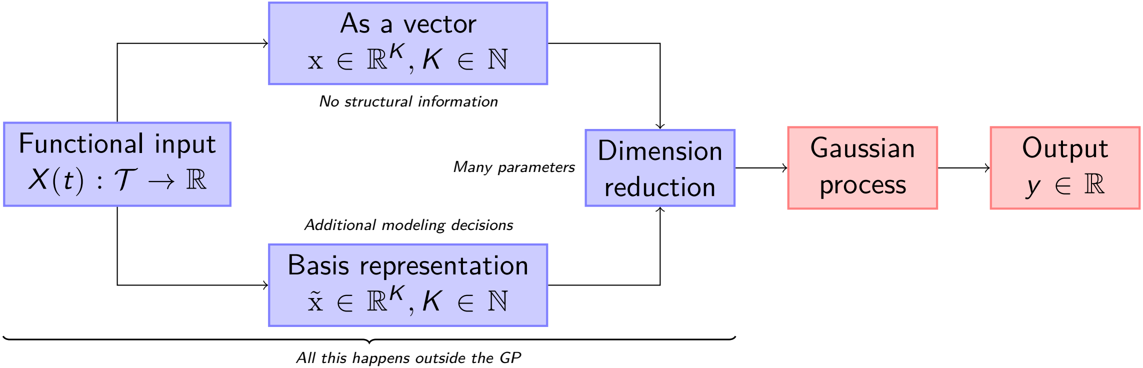
- Can we connect the functional input structure to a physical mechanism?

Gaussian process with functional input



- Can we connect the functional input structure to a physical mechanism?
- Can we integrate the functional input structure into the GP?

Gaussian process with functional input



- Can we connect the functional input structure to a physical mechanism?
- Can we integrate the functional input structure into the GP?
- Can we circumvent input dimension reduction?

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism
Plant growth	Phosphorus	Depth	Soil layers	Root biomass

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism
Plant growth	Phosphorus	Depth	Soil layers	Root biomass
	Precipitation	Time	Cycles, seasons	Germination photosynthesis nutrient absorption

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism
Plant growth	Phosphorus	Depth	Soil layers	Root biomass
	Precipitation	Time	Cycles, seasons	Germination photosynthesis nutrient absorption
Soil erosion	Elevation	Distance	Up/down slope	Water erosion

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism
Plant growth	Phosphorus	Depth	Soil layers	Root biomass
	Precipitation	Time	Cycles, seasons	Germination photosynthesis nutrient absorption
Soil erosion	Elevation	Distance	Up/down slope	Water erosion
Radiance	Chemical species	Elevation	Atmospheric layers	Reflectivity

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism
Plant growth	Phosphorus	Depth	Soil layers	Root biomass
	Precipitation	Time	Cycles, seasons	Germination photosynthesis nutrient absorption
Soil erosion	Elevation	Distance	Up/down slope	Water erosion
Radiance	Chemical species	Elevation	Atmospheric layers	Reflectivity

Index subspaces can provide a meaningful representation of the physical process

Input structure information

Output	Input $X(t) : \mathcal{T} \to \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism
Plant growth	Phosphorus	Depth	Soil layers	Root biomass
	Precipitation	Time	Cycles, seasons	Germination photosynthesis nutrient absorption
Soil erosion	Elevation	Distance	Up/down slope	Water erosion
Radiance	Chemical species	Elevation	Atmospheric layers	Reflectivity

Index subspaces can provide a meaningful representation of the physical process

Can we establish an explicit link $X(t) \xrightarrow{f} y$ *for UQ?*

From relevance to *dynamic* relevance

Extending relevance

Some inputs
matter more than others
x_1 *vs* x_2

Screening
(exploration
& diagnostics)

Permutation
Feature
Importance

Model tuning
(learning)

Automatic
Relevance
Determination

[1] Forthcoming paper

Extending relevance

	Some inputs matter more than others x_1 *vs* x_2	$\longrightarrow$	Some index subspaces matter more than others $X(t_1)$ *vs* $X(t_2)$
Screening *(exploration & diagnostics)*	Permutation Feature Importance		
Model tuning *(learning)*	Automatic Relevance Determination		

[1] Forthcoming paper

Extending relevance

	Some inputs matter more than others x_1 *vs* x_2	$\longrightarrow$	Some index subspaces matter more than others $X(t_1)$ *vs* $X(t_2)$
Screening *(exploration & diagnostics)*	Permutation Feature Importance	$\longrightarrow$	Permutation Feature *Dynamic Importance*[1]
Model tuning *(learning)*	Automatic Relevance Determination		

[1] Forthcoming paper

Extending relevance

	Some inputs matter more than others x_1 *vs* x_2	$\longrightarrow$	Some index subspaces matter more than others $X(t_1)$ *vs* $X(t_2)$
Screening *(exploration & diagnostics)*	Permutation Feature Importance	$\longrightarrow$	Permutation Feature *Dynamic Importance*[1]
Model tuning *(learning)*	Automatic Relevance Determination	$\longrightarrow$	Automatic *Dynamic Relevance* Determination

[1] Forthcoming paper

Modeling *dynamic* relevance

Model tuning *(learning)*	Automatic Relevance Determination
Distance $d(X_i, X_j)$	$\sum_{k=1}^{K} \frac{(x_{i,k}-x_{j,k})^2}{\ell_k^2}$
Weights *(relevance)*	$\ell_1^{-2}, \cdots, \ell_K^{-2} > 0$

Modeling *dynamic* relevance

Model tuning *(learning)*	Automatic Relevance Determination	$\longrightarrow$	Automatic *Dynamic Relevance* Determination
Distance $d(X_i, X_j)$	$\sum_{k=1}^{K} \frac{(x_{i,k}-x_{j,k})^2}{\ell_k^2}$		
Weights *(relevance)*	$\ell_1^{-2}, \cdots, \ell_K^{-2} > 0$		

Modeling *dynamic* relevance

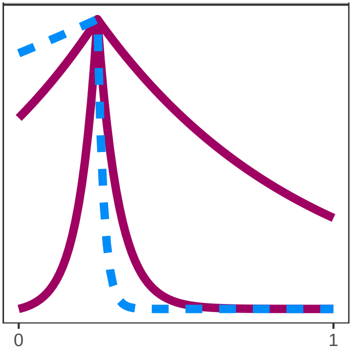
Model tuning *(learning)*	Automatic Relevance Determination	$\longrightarrow$	$\boxed{\text{Automatic } \textit{Dynamic Relevance} \text{ Determination}}$
Distance $d(X_i, X_j)$	$\sum_{k=1}^{K} \frac{(x_{i,k} - x_{j,k})^2}{\ell_k^2}$		$\int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 \mathrm{d}t$
Weights *(relevance)*	$\ell_1^{-2}, \cdots, \ell_K^{-2} > 0$		

Modeling *dynamic* relevance

Model tuning *(learning)*	Automatic Relevance Determination	$\longrightarrow$	Automatic *Dynamic Relevance* Determination
Distance $d(X_i, X_j)$	$\sum_{k=1}^{K} \frac{(x_{i,k} - x_{j,k})^2}{\ell_k^2}$		$\int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 \mathrm{d}t$
Weights *(relevance)*	$\ell_1^{-2}, \cdots, \ell_K^{-2} > 0$		$\omega(t) : \mathcal{T} \to \mathbb{R}^+$

A flexible weight function

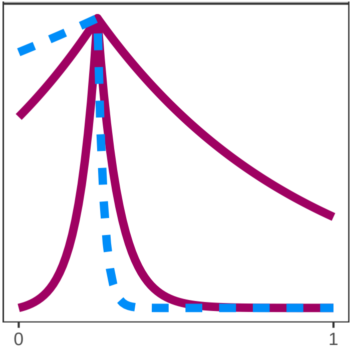
$$\omega(t) = \exp\{-(t-\tau)\lambda\kappa^s s\}$$



$$s = \text{sign}(t-\tau),\ \tau > 0,\ \lambda > 0,\ \kappa > 0,\ \mathcal{T} = [0,1]$$

A flexible weight function

$$\omega(t) = \exp\{-(t-\tau)\lambda\kappa^s s\}$$

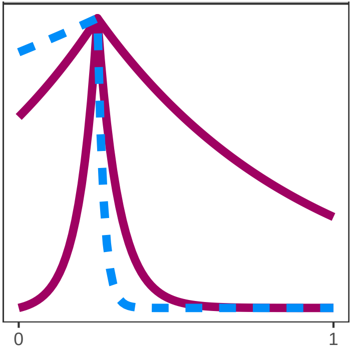


- The input is most relevant at τ (relevance peak)

$s = \mathrm{sign}(t-\tau)$, $\tau > 0$, $\lambda > 0$, $\kappa > 0$, $\mathcal{T} = [0,1]$

A flexible weight function

$$\omega(t) = \exp\{-(t-\tau)\lambda\kappa^s s\}$$

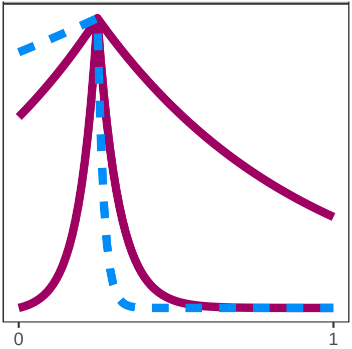


- The input is most relevant at τ (relevance peak)
- Relevance increases at a $\lambda_1 = \lambda\kappa^{-1}$ rate from $t = 0$ to the peak

$s = \text{sign}(t-\tau)$, $\tau > 0$, $\lambda > 0$, $\kappa > 0$, $\mathcal{T} = [0,1]$

A flexible weight function

$$\omega(t) = \exp\{-(t-\tau)\lambda\kappa^s s\}$$

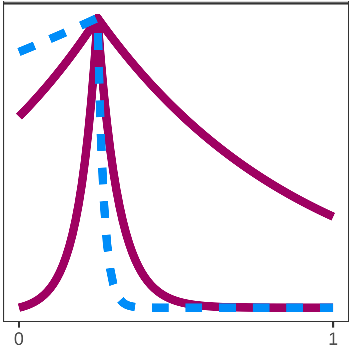


- The input is most relevant at τ (relevance peak)
- Relevance increases at a $\lambda_1 = \lambda\kappa^{-1}$ rate from $t = 0$ to the peak
- Relevance decreases at a $\lambda_2 = \lambda\kappa$ rate from the peak to $t = 1$

$s = \mathrm{sign}(t-\tau)$, $\tau > 0$, $\lambda > 0$, $\kappa > 0$, $\mathcal{T} = [0, 1]$

A flexible weight function

$$\omega(t) = \exp\{-(t-\tau)\lambda\kappa^s s\}$$

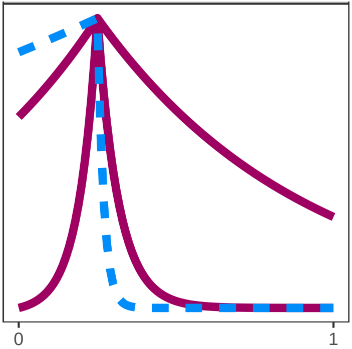


- The input is most relevant at τ (relevance peak)
- Relevance increases at a $\lambda_1 = \lambda\kappa^{-1}$ rate from $t = 0$ to the peak
- Relevance decreases at a $\lambda_2 = \lambda\kappa$ rate from the peak to $t = 1$
- To predict the output, look for input profiles that are similar everywhere *but especially* near τ

$s = \text{sign}(t-\tau)$, $\tau > 0$, $\lambda > 0$, $\kappa > 0$, $\mathcal{T} = [0, 1]$

A flexible weight function

$$\omega(t) = \exp\{-(t-\tau)\lambda\kappa^s s\}$$



- The input is most relevant at τ (relevance peak)
- Relevance increases at a $\lambda_1 = \lambda\kappa^{-1}$ rate from $t = 0$ to the peak
- Relevance decreases at a $\lambda_2 = \lambda\kappa$ rate from the peak to $t = 1$
- To predict the output, look for input profiles that are similar everywhere *but especially* near τ

circumvent input dimension reduction

$s = \text{sign}(t-\tau)$, $\tau > 0$, $\lambda > 0$, $\kappa > 0$, $\mathcal{T} = [0,1]$

Functional Input Gaussian Process (fiGP)

$$y \sim \mathcal{N}\left(0, \sigma_f^2 \, R_f + \sigma_\varepsilon^2 I\right) \tag{1}$$

$$(R_f)_{ij} = \exp\left\{-0.5\phi^{-2} \, d_f(X_i, X_j)\right\} \tag{2}$$

(3)

(4)

$\sigma_\varepsilon^2 > 0, \; \sigma_f^2 > 0, \; \phi > 0, \; i, j = 1, \ldots, N$

Functional Input Gaussian Process (fiGP)

$$\mathrm{y} \sim \mathcal{N}\left(0, \sigma_f^2\ \mathrm{R}_f + \sigma_\varepsilon^2 \mathrm{I}\right) \tag{1}$$

$$(\mathrm{R}_f)_{ij} = \exp\left\{-0.5\phi^{-2}\ d_f(X_i, X_j)\right\} \tag{2}$$

$$d_f(X_i, X_j) = \int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \tag{3}$$

$$\tag{4}$$

$\sigma_\varepsilon^2 > 0,\ \sigma_f^2 > 0,\ \phi > 0,\ i, j = 1, \dots, N$
$\omega(t) : \mathcal{T} \to (0, \infty)$

Functional Input Gaussian Process (fiGP)

$$\mathrm{y} \sim \mathcal{N}\left(0, \sigma_f^2\ \mathrm{R}_f + \sigma_\varepsilon^2 \mathrm{I}\right) \tag{1}$$

$$(\mathrm{R}_f)_{ij} = \exp\left\{-0.5\phi^{-2}\ d_f(X_i, X_j)\right\} \tag{2}$$

$$d_f(X_i, X_j) = \int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \tag{3}$$

$$\omega(t) = \exp\left\{-(t-\tau)\lambda\kappa^s s\right\} \tag{4}$$

$\sigma_\varepsilon^2 > 0,\ \sigma_f^2 > 0,\ \phi > 0,\ i, j = 1, \ldots, N$
$\omega(t) : \mathcal{T} \to (0, \infty)$

Functional Input Gaussian Process (fiGP)

$$\mathrm{y} \sim \mathcal{N}\left(0, \sigma_f^2\, \mathrm{R}_f + \sigma_\varepsilon^2 \mathrm{I}\right) \tag{1}$$

$$(\mathrm{R}_f)_{ij} = \exp\left\{-0.5\phi^{-2}\; d_f(X_i, X_j)\right\} \tag{2}$$

$$d_f(X_i, X_j) = \int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \tag{3}$$

$$\omega(t) = \exp\left\{-(t-\tau)\lambda\kappa^s s\right\} \tag{4}$$

Weak priors $\phi \sim \textsc{InvGamma}(\cdot,\cdot),\ \tau \sim \textsc{Beta}(\cdot,\cdot),\ \lambda \sim \mathrm{N}^+(\cdot,\cdot),\ \log(\kappa) \sim \mathrm{N}(\cdot,\cdot)$

$\sigma_\varepsilon^2 > 0,\ \sigma_f^2 > 0,\ \phi > 0,\ i,j = 1,\ldots,N$
$\omega(t) : \mathcal{T} \to (0,\infty)$

Functional Input Gaussian Process (fiGP)

$$\mathrm{y} \sim \mathcal{N}\left(0, \sigma_f^2\ \mathrm{R}_f + \sigma_\varepsilon^2 \mathrm{I}\right) \tag{1}$$

$$(\mathrm{R}_f)_{ij} = \exp\left\{-0.5\phi^{-2}\ d_f(X_i, X_j)\right\} \tag{2}$$

$$d_f(X_i, X_j) = \int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \tag{3}$$

$$\omega(t) = \exp\left\{-(t-\tau)\lambda\kappa^s s\right\} \tag{4}$$

Weak priors $\phi \sim \textsc{InvGamma}(\cdot,\cdot)$, $\tau \sim \textsc{Beta}(\cdot,\cdot)$, $\lambda \sim \mathrm{N}^+(\cdot,\cdot)$, $\log(\kappa) \sim \mathrm{N}(\cdot,\cdot)$

Multiple inputs e.g., correlation product

$\sigma_\varepsilon^2 > 0$, $\sigma_f^2 > 0$, $\phi > 0$, $i, j = 1, \ldots, N$
$\omega(t) : \mathcal{T} \to (0, \infty)$

Functional Input Gaussian Process (fiGP)

$$\mathrm{y} \sim \mathcal{N}\left(0, \sigma_f^2\ \mathrm{R}_f + \sigma_\varepsilon^2 \mathrm{I}\right) \tag{1}$$

$$(\mathrm{R}_f)_{ij} = \exp\left\{-0.5\phi^{-2}\ d_f(X_i, X_j)\right\} \tag{2}$$

$$d_f(X_i, X_j) = \int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \tag{3}$$

$$\omega(t) = \exp\left\{-(t-\tau)\lambda\kappa^s s\right\} \tag{4}$$

Weak priors $\phi \sim \textsc{InvGamma}(\cdot,\cdot)$, $\tau \sim \textsc{Beta}(\cdot,\cdot)$, $\lambda \sim \mathrm{N}^+(\cdot,\cdot)$, $\log(\kappa) \sim \mathrm{N}(\cdot,\cdot)$

Multiple inputs e.g., correlation product

Complex index spaces e.g., spatio-temporal spectral structures AKA tesseract

---

$\sigma_\varepsilon^2 > 0$, $\sigma_f^2 > 0$, $\phi > 0$, $i, j = 1, \ldots, N$
$\omega(t) : \mathcal{T} \to (0, \infty)$

Functional Input Gaussian Process (fiGP)

$$\mathrm{y} \sim \mathcal{N}\left(0, \sigma_f^2 \, \mathrm{R}_f + \sigma_\varepsilon^2 \mathrm{I}\right) \tag{1}$$

$$(\mathrm{R}_f)_{ij} = \exp\left\{-0.5\phi^{-2} \, d_f(X_i, X_j)\right\} \tag{2}$$

$$d_f(X_i, X_j) = \int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \tag{3}$$

$$\omega(t) = \exp\left\{-(t-\tau)\lambda\kappa^s s\right\} \tag{4}$$

Weak priors $\phi \sim \textsc{InvGamma}(\cdot,\cdot)$, $\tau \sim \textsc{Beta}(\cdot,\cdot)$, $\lambda \sim \mathrm{N}^+(\cdot,\cdot)$, $\log(\kappa) \sim \mathrm{N}(\cdot,\cdot)$

Multiple inputs e.g., correlation product

Complex index spaces e.g., spatio-temporal spectral structures AKA tesseract

Flexibility no need to match input-output structure nor index space

---

$\sigma_\varepsilon^2 > 0$, $\sigma_f^2 > 0$, $\phi > 0$, $i, j = 1, \ldots, N$
$\omega(t) : \mathcal{T} \rightarrow (0, \infty)$

NASA's Microwave Limb Sounder

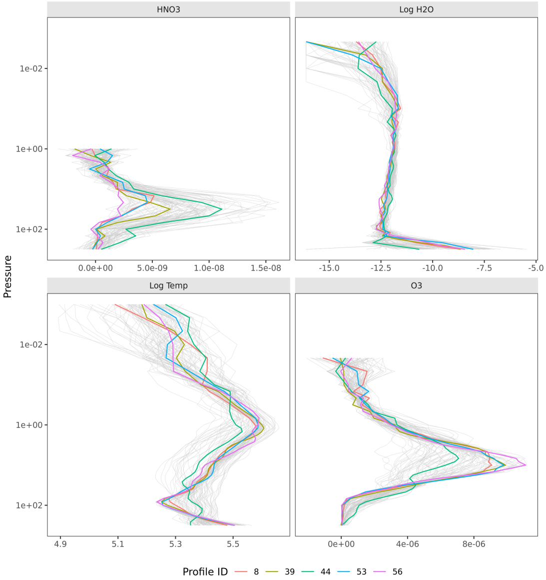
a case study

Data structure

Credit: NASA Aura

Atmospheric constituents $\xrightarrow{f}$ Radiance

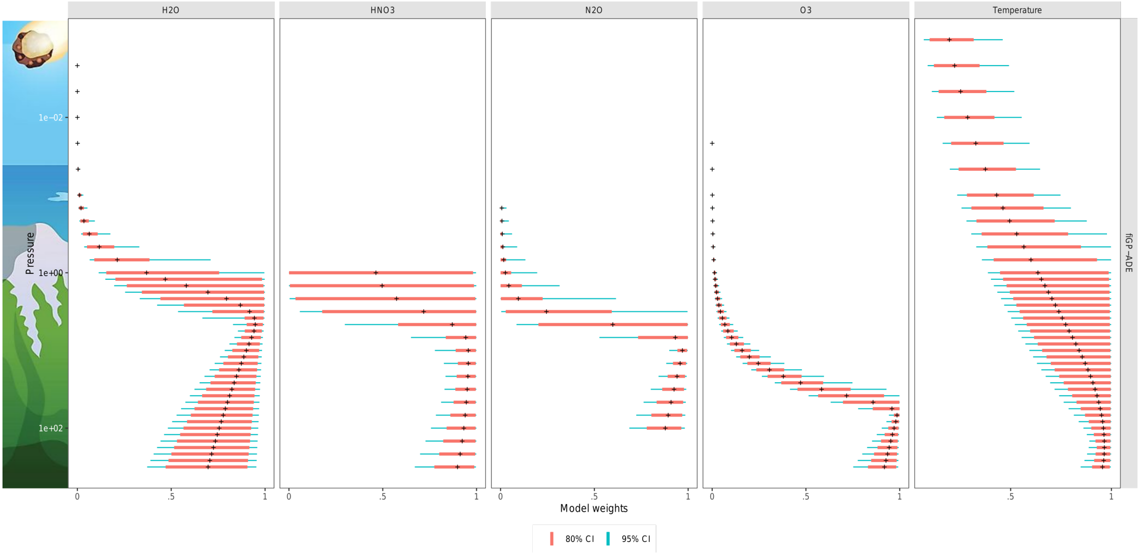
Functional input $X(t)$ — *Scalar output y*



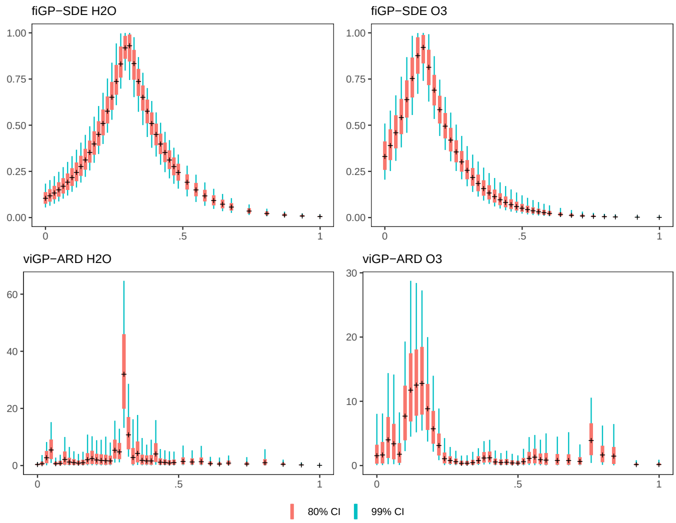
Implementation

- 8 training, 8 test complementary sets
- 1,000 soundings each
- One model fit separately per input-output pair
- Fully Bayesian inference
- Hamiltonian Monte Carlo using Stan
- 1 long chain
- Extensive search for an initial value
- 500 post-warmup iterations
- 1,500 posterior samples

Weight function posterior samples



fiGP vs a vector-input GP



Why a fiGP?

+ High dimensional inputs with no dimension reduction
  - Reduce unknowns $3 << K$
  - Scales up for applications with higher input resolution $\uparrow K$

[1] Forthcoming paper

[2] Future research

Why a fiGP?

+ High dimensional inputs with no dimension reduction
  - Reduce unknowns $3 << K$
  - Scales up for applications with higher input resolution $\uparrow K$
+ Explicit link between output correlation and input functional structure
  - Can incorporate domain-specific knowledge
  - Tangible for prior elicitation
  - Interpretation $\rightarrow$ insight?
  - Smooths out erratic relevance patterns

[1] Forthcoming paper

[2] Future research

Why a fiGP?

- High dimensional inputs with no dimension reduction
  - Reduce unknowns $3 << K$
  - Scales up for applications with higher input resolution $\uparrow K$
- Explicit link between output correlation and input functional structure
  - Can incorporate domain-specific knowledge
  - Tangible for prior elicitation
  - Interpretation $\rightarrow$ insight?
  - Smooths out erratic relevance patterns
- Similar predictive power as vector-input GP[1]

[1] Forthcoming paper
[2] Future research

Why a fiGP?

+ High dimensional inputs with no dimension reduction
  - Reduce unknowns $3 << K$
  - Scales up for applications with higher input resolution $\uparrow K$
+ Explicit link between output correlation and input functional structure
  - Can incorporate domain-specific knowledge
  - Tangible for prior elicitation
  - Interpretation $\rightarrow$ insight?
  - Smooths out erratic relevance patterns
+ Similar predictive power as vector-input GP[1]

++ Extensible to **complex index spaces**, e.g., spatio-temporal spectral inputs[2]

[1] Forthcoming paper

[2] Future research

Acknowledgments

The MLS team at JPL, California Institute of Technology

Thank you!

mail ldamiano@iastate.edu

repo https://github.com/luisdamiano/SIAMUQ22

Appendix

References

Thomas Muehlenstaedt, Jana Fruth, and Olivier Roustant.
Computer experiments with functional inputs and scalar outputs by a norm-based approach.
Statistics and Computing, 27(4):1083–1097, July 2017.

Trapezoidal approximation

$$\int_{\mathcal{T}} \omega(t)(X_i(t) - X_j(t))^2 dt \approx \sum_{k=2}^{K} (t_k - t_{k-1}) \frac{\Delta_{i,j,k} + \Delta_{i,j,k-1}}{2} \tag{5}$$

$$\Delta_{i,j,k} = \omega(t_{k-1})(x_{i,k} - x_{j,k})^2 \tag{6}$$

See [1] for a B-spline approach

Out-of-sample prediction

	H2O	HNO3	N2O	O3	Temp	Mean
SE	.34	**.48**	**.44**	.32	**.25**	.37
ARD	**.31**	**.47**	**.43**	**.30**	**.25**	.35
FPCA	.67	.91	.99	.46	.54	.71
FFPCA	.46	.54	**.46**	.38	.33	.44
EDN	.33	**.47**	**.44**	**.29**	**.25**	.36
SDE	**.31**	**.47**	**.44**	**.29**	**.25**	.35
ADE	**.31**	**.47**	**.43**	**.29**	**.25**	.35
Mean	.39	.55	.52	.33	.31	.42

	H2O	HNO3	N2O	O3	Temp	Mean
SE	273	**614**	**585**	138	**-7**	323
ARD	**196**	**619**	**581**	**92**	**-13**	295
FPCA	1024	1320	1406	637	802	1038
FFPCA	535	**646**	**630**	295	268	475
EDN	261	**623**	**585**	**90**	**4**	312
SDE	**202**	**623**	**585**	**85**	**4**	300
ADE	**202**	**610**	**581**	**87**	**2**	297
Mean	385	722	708	204	152	434

Mean validation statistics: RMSE (left) and negative posterior predictive log-density (right). Smaller values are better. Bold: best in column. EDN $\tau = 0, \kappa = 1$; SDE $\tau = 0$; ADE τ, κ, λ all free; ARD as many free parameters as measurements per vertical profile.