

# Automatic Dynamic Relevance Determination for Gaussian process regressions with functional inputs

**Luis Damiano**<sup>†1</sup>, Joaquim Teixeira<sup>2</sup>, Margaret Johnson<sup>2</sup>, Jarad Niemi<sup>1</sup>

<sup>1</sup>Department of Statistics, Iowa State University

<sup>2</sup>NASA Jet Propulsion Laboratory

SIAM Conference on Uncertainty Quantification  
April 13th, 2022

Funded, in part, by

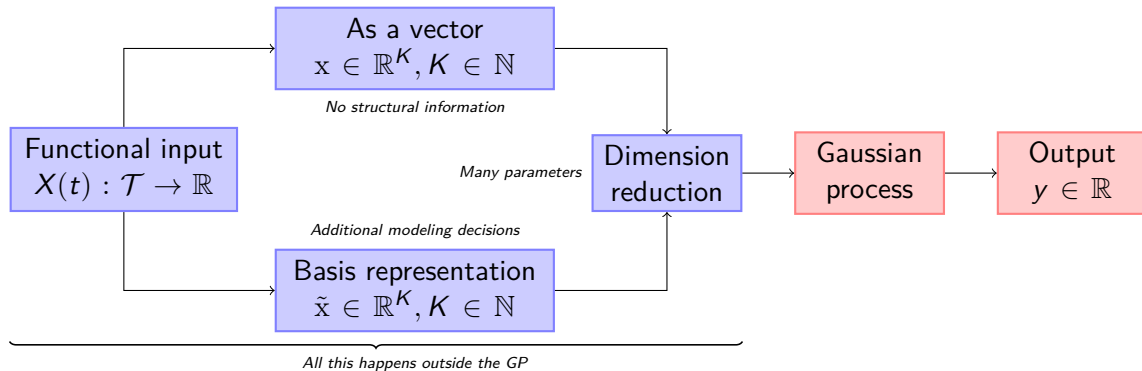
- ISU Presidential Interdisciplinary Research Initiative on C-CHANGE: Science for a Changing Agriculture
- Foundation for Food and Agriculture Research Grant ID: CA18-SS-0000000278

---

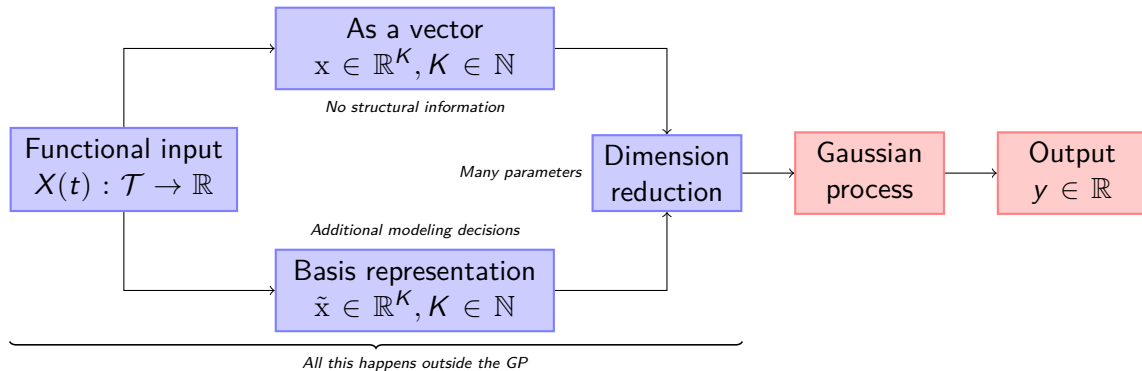
<sup>†</sup>[ldamiano@iastate.edu](mailto:ldamiano@iastate.edu)

# Overview & Motivation

# Gaussian process with functional input

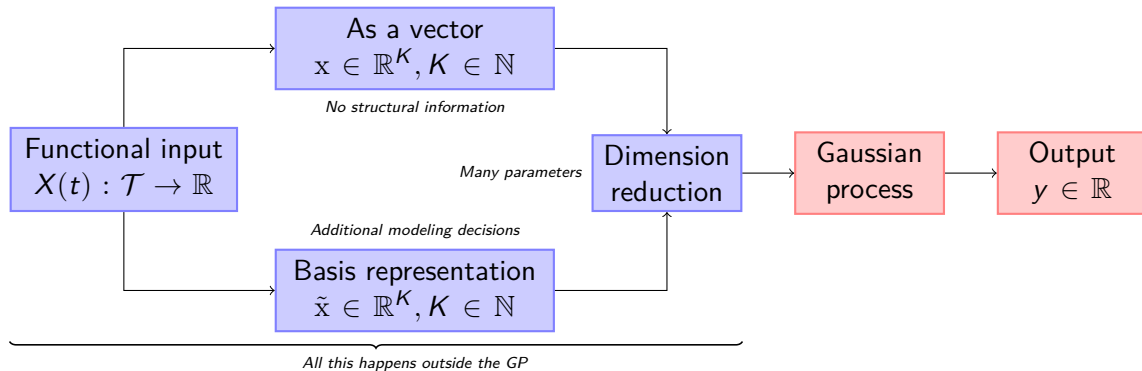


# Gaussian process with functional input



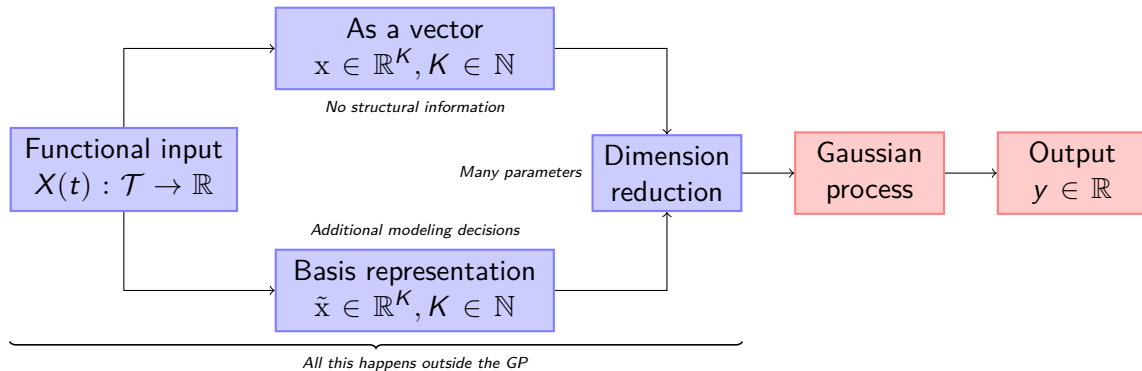
- Can we connect the functional input structure to a physical mechanism?

# 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 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	Index $t \in \mathcal{T}$	Index subspaces $t \in \mathcal{T}_u$	Mechanism

# Input structure information

Output	Input $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	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 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	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 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	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 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	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 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	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 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$	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?*

# Overview & Motivation

# 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  $x_2$



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  $x_2$



Some index subspaces  
matter more than others  
 $X(t_1)$  vs  $X(t_2)$

Screening  
(*exploration*  
& *diagnostics*)

Permutation  
Feature  
Importance



Permutation  
Feature  
*Dynamic Importance*<sup>[1]</sup>

Model tuning  
(*learning*)

Automatic  
Relevance  
Determination

---

<sup>[1]</sup> Forthcoming paper

# Extending relevance

Some inputs  
matter more than others  
 $x_1$  vs  $x_2$



Some index subspaces  
matter more than others  
 $X(t_1)$  vs  $X(t_2)$

Screening  
(*exploration*  
& *diagnostics*)

Permutation  
Feature  
Importance



Permutation  
Feature  
*Dynamic Importance*<sup>[1]</sup>

Model tuning  
(*learning*)

Automatic  
Relevance  
Determination



Automatic  
*Dynamic Relevance*  
Determination

---

<sup>[1]</sup> Forthcoming paper

# Modeling *dynamic* relevance

Model tuning  
(*learning*)

Automatic  
Relevance  
Determination



Automatic  
*Dynamic Relevance*  
Determination

Distance  
 $d(X_i, X_j)$

$$\sum_{k=1}^K \frac{(x_{i,k} - x_{j,k})^2}{\ell_k^2}$$

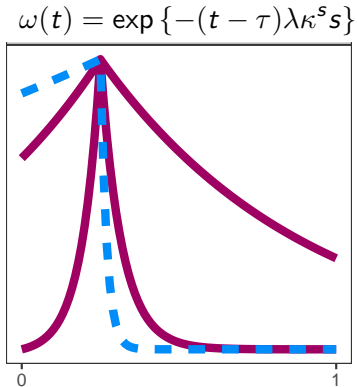
$$\int_{\mathcal{T}} \omega(t) (X_i(t) - X_j(t))^2 dt$$

Weights  
(*relevance*)

$$\ell_1^{-2}, \dots, \ell_K^{-2} > 0$$

$$\omega(t) : \mathcal{T} \rightarrow \mathbb{R}^+$$

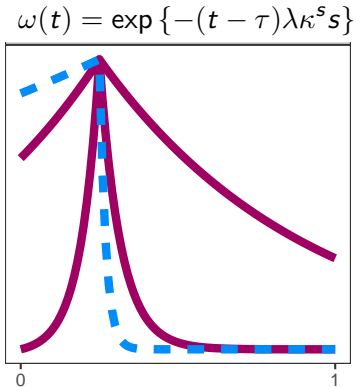
# A flexible weight function



---

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

# A flexible weight function

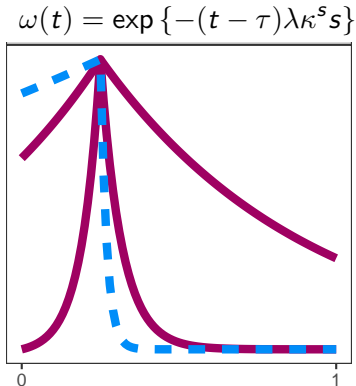


- The index is most relevant at  $\tau$  (relevance peak)

---

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

# A flexible weight function

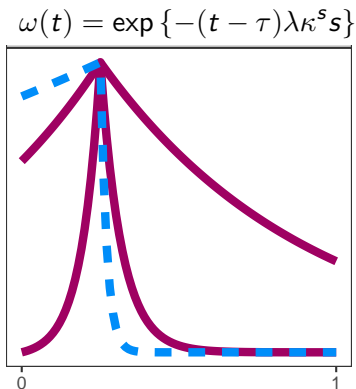


- The index is most relevant at  $\tau$  (relevance peak)
- Relevance increases at a  $\lambda_1 = \lambda\kappa^{-1}$  rate from  $t = 0$  to the peak

---

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

# A flexible weight function

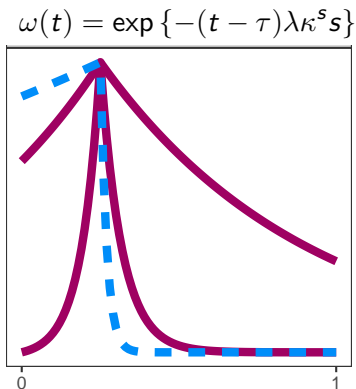


- The index is most relevant at  $\tau$  (relevance peak)
- Relevance increases at a  $\lambda_1 = \lambda\kappa^{-1}$  rate from  $t = 0$  to the peak
- Relevance decreases at a  $\lambda_2 = \lambda\kappa$  rate from the peak to  $t = 1$

---

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

# A flexible weight function



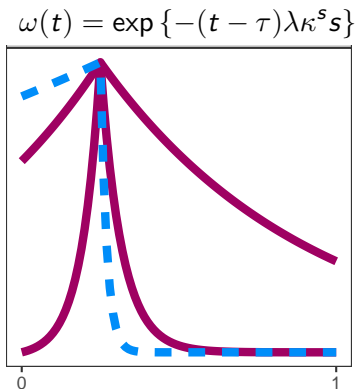
- The index is most relevant at  $\tau$  (relevance peak)
- Relevance increases at a  $\lambda_1 = \lambda\kappa^{-1}$  rate from  $t = 0$  to the peak
- Relevance decreases at a  $\lambda_2 = \lambda\kappa$  rate from the peak to  $t = 1$
- To predict the output, look for input profiles that are similar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

---

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



# A flexible weight function



- The index is most relevant at  $\tau$  (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  $\tau$   
circumvent input dimension reduction

---

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

# Functional Input Gaussian Process (fiGP)

$$y \sim \mathcal{N}(0, \sigma_f^2 \mathbf{R}_f + \sigma_\varepsilon^2 \mathbf{I}) \quad (1)$$

$$(\mathbf{R}_f)_{ij} = \exp \left\{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \right\} \quad (2)$$

$$(3)$$

$$(4)$$

---

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

# Functional Input Gaussian Process (fiGP)

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

$$(\mathbf{R}_f)_{ij} = \exp \left\{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \right\} \quad (2)$$

$$d_f(\mathbf{X}_i, \mathbf{X}_j) = \int_{\mathcal{T}} \omega(t) (\mathbf{X}_i(t) - \mathbf{X}_j(t))^2 dt \quad (3)$$

$$(4)$$

---

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

# Functional Input Gaussian Process (fiGP)

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

$$(\mathbf{R}_f)_{ij} = \exp \left\{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \right\} \quad (2)$$

$$d_f(\mathbf{X}_i, \mathbf{X}_j) = \int_{\mathcal{T}} \omega(t) (\mathbf{X}_i(t) - \mathbf{X}_j(t))^2 dt \quad (3)$$

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

---

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

# Functional Input Gaussian Process (fiGP)

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

$$(\mathbf{R}_f)_{ij} = \exp \left\{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \right\} \quad (2)$$

$$d_f(\mathbf{X}_i, \mathbf{X}_j) = \int_{\mathcal{T}} \omega(t) (\mathbf{X}_i(t) - \mathbf{X}_j(t))^2 dt \quad (3)$$

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

**Weak priors**  $\phi \sim \text{INV GAMMA}(\cdot, \cdot)$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim \mathbf{N}^+(\cdot, \cdot)$ ,  $\log(\kappa) \sim \mathbf{N}(\cdot, \cdot)$

---

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

# Functional Input Gaussian Process (fiGP)

$$y \sim \mathcal{N}(0, \sigma_f^2 \mathbf{R}_f + \sigma_\varepsilon^2 \mathbf{I}) \quad (1)$$

$$(\mathbf{R}_f)_{ij} = \exp \{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \} \quad (2)$$

$$d_f(\mathbf{X}_i, \mathbf{X}_j) = \int_{\mathcal{T}} \omega(t) (\mathbf{X}_i(t) - \mathbf{X}_j(t))^2 dt \quad (3)$$

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

**Weak priors**  $\phi \sim \text{INV GAMMA}(\cdot, \cdot)$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim \mathbf{N}^+(\cdot, \cdot)$ ,  $\log(\kappa) \sim \mathbf{N}(\cdot, \cdot)$

**Multiple inputs** e.g., correlation product

---

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

# Functional Input Gaussian Process (fiGP)

$$y \sim \mathcal{N}(0, \sigma_f^2 \mathbf{R}_f + \sigma_\varepsilon^2 \mathbf{I}) \quad (1)$$

$$(\mathbf{R}_f)_{ij} = \exp \left\{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \right\} \quad (2)$$

$$d_f(\mathbf{X}_i, \mathbf{X}_j) = \int_{\mathcal{T}} \omega(t) (\mathbf{X}_i(t) - \mathbf{X}_j(t))^2 dt \quad (3)$$

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

**Weak priors**  $\phi \sim \text{INV GAMMA}(\cdot, \cdot)$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim \mathbf{N}^+(\cdot, \cdot)$ ,  $\log(\kappa) \sim \mathbf{N}(\cdot, \cdot)$

**Multiple inputs** e.g., correlation product

**Complex index spaces** e.g., spatio-temporal structures

---

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

# Functional Input Gaussian Process (fiGP)

$$y \sim \mathcal{N}(0, \sigma_f^2 \mathbf{R}_f + \sigma_\varepsilon^2 \mathbf{I}) \quad (1)$$

$$(\mathbf{R}_f)_{ij} = \exp \left\{ -0.5 \phi^{-2} d_f(\mathbf{X}_i, \mathbf{X}_j) \right\} \quad (2)$$

$$d_f(\mathbf{X}_i, \mathbf{X}_j) = \int_{\mathcal{T}} \omega(t) (\mathbf{X}_i(t) - \mathbf{X}_j(t))^2 dt \quad (3)$$

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

**Weak priors**  $\phi \sim \text{INV GAMMA}(\cdot, \cdot)$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim \mathbf{N}^+(\cdot, \cdot)$ ,  $\log(\kappa) \sim \mathbf{N}(\cdot, \cdot)$

**Multiple inputs** e.g., correlation product

**Complex index spaces** e.g., spatio-temporal structures

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

---

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



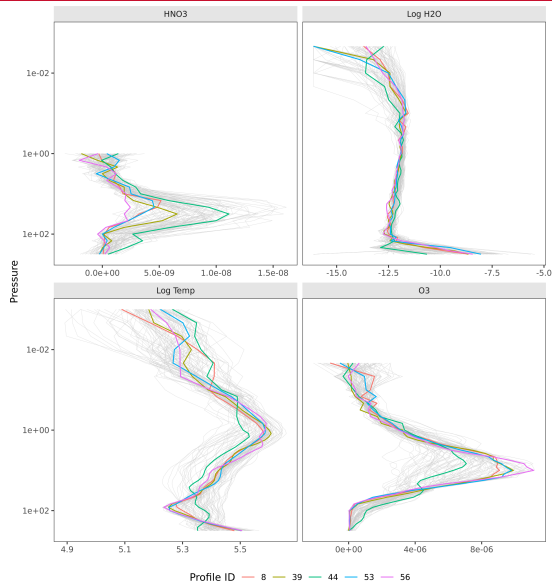
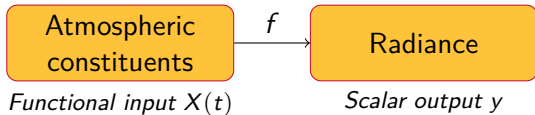
# NASA's Microwave Limb Sounder

*a case study*

# Data structure



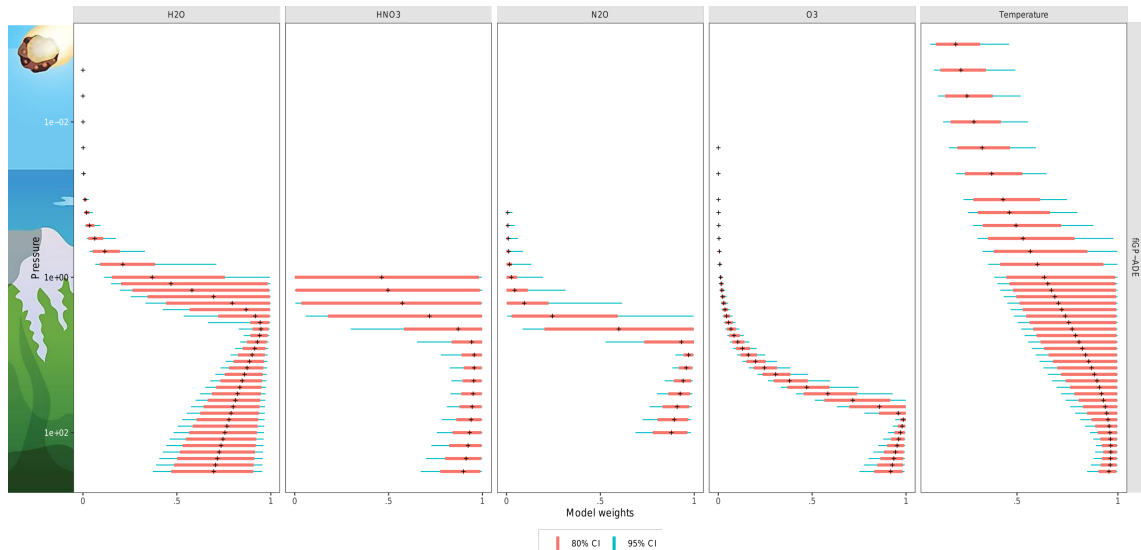
*Credit: NASA Aura*



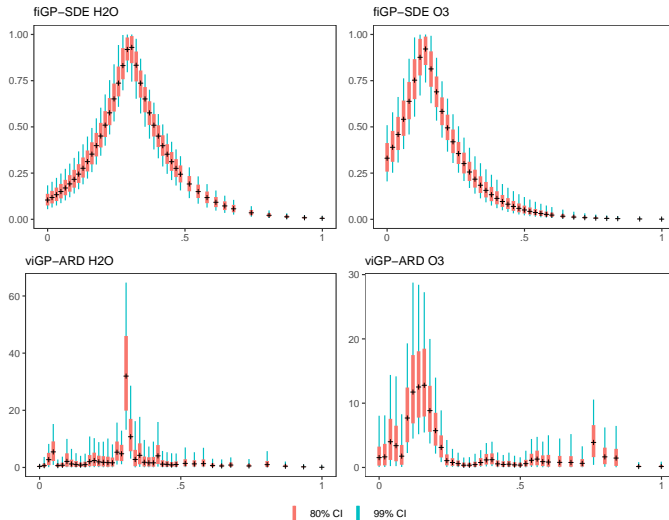
# Implementation

- 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  $3 \ll K$
  - ▶ Scales up for applications with higher input resolution  $\uparrow K$

---

<sup>[1]</sup>Forthcoming paper

<sup>[2]</sup>Future research

# Why a fiGP?

- + High dimensional inputs with no dimension reduction
  - ▶ Reduce unknowns  $3 \ll K$
  - ▶ 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

---

<sup>[1]</sup>Forthcoming paper

<sup>[2]</sup>Future research

# Why a fiGP?

- + High dimensional inputs with no dimension reduction
  - ▶ Reduce unknowns  $3 \ll K$
  - ▶ 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<sup>[1]</sup>

---

<sup>[1]</sup>Forthcoming paper

<sup>[2]</sup>Future research



# Why a fiGP?

- + High dimensional inputs with no dimension reduction
  - ▶ Reduce unknowns  $3 \ll K$
  - ▶ 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<sup>[1]</sup>
- ++ Extensible to **complex index spaces**, e.g., spatio-temporal spectral inputs<sup>[2]</sup>

---

<sup>[1]</sup>Forthcoming paper

<sup>[2]</sup>Future research

# Acknowledgments

The MLS team at JPL, California Institute of Technology

Thank you!

► mail [ldamiano@iastate.edu](mailto: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

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

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

---

See [1] for a B-spline approach

# Out-of-sample prediction

	H2O	HNO3	N2O	O3	Temp	Mean		H2O	HNO3	N2O	O3	Temp	Mean
SE	.34	<b>.48</b>	<b>.44</b>	.32	<b>.25</b>	.37	SE	273	<b>614</b>	<b>585</b>	138	<b>-7</b>	323
ARD	<b>.31</b>	<b>.47</b>	<b>.43</b>	<b>.30</b>	<b>.25</b>	.35	ARD	<b>196</b>	<b>619</b>	<b>581</b>	<b>92</b>	<b>-13</b>	295
FPCA	.67	.91	.99	.46	.54	.71	FPCA	1024	1320	1406	637	802	1038
FFPCA	.46	.54	<b>.46</b>	.38	.33	.44	FFPCA	535	<b>646</b>	<b>630</b>	295	268	475
EDN	.33	<b>.47</b>	<b>.44</b>	<b>.29</b>	<b>.25</b>	.36	EDN	261	<b>623</b>	<b>585</b>	<b>90</b>	<b>4</b>	312
SDE	<b>.31</b>	<b>.47</b>	<b>.44</b>	<b>.29</b>	<b>.25</b>	.35	SDE	<b>202</b>	<b>623</b>	<b>585</b>	<b>85</b>	<b>4</b>	300
ADE	<b>.31</b>	<b>.47</b>	<b>.43</b>	<b>.29</b>	<b>.25</b>	.35	ADE	<b>202</b>	<b>610</b>	<b>581</b>	<b>87</b>	<b>2</b>	297
Mean	.39	.55	.52	.33	.31	.42	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  $\tau, \kappa, \lambda$  all free; ARD as many free parameters as measurements per vertical profile.