

# 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)

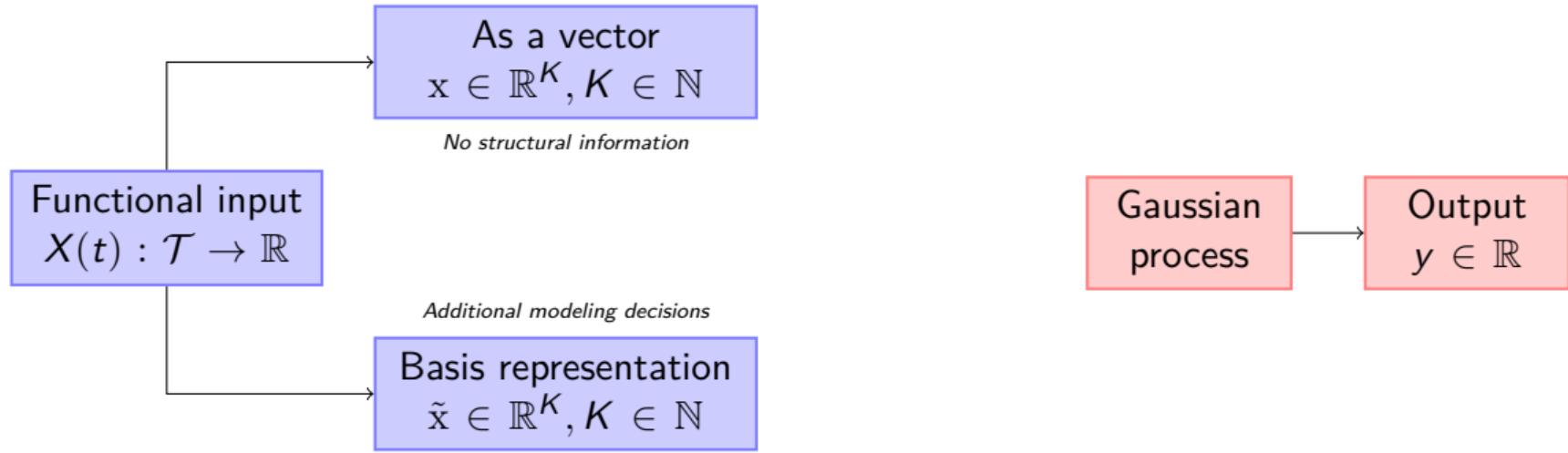
# Overview & motivation

# Gaussian process with functional input

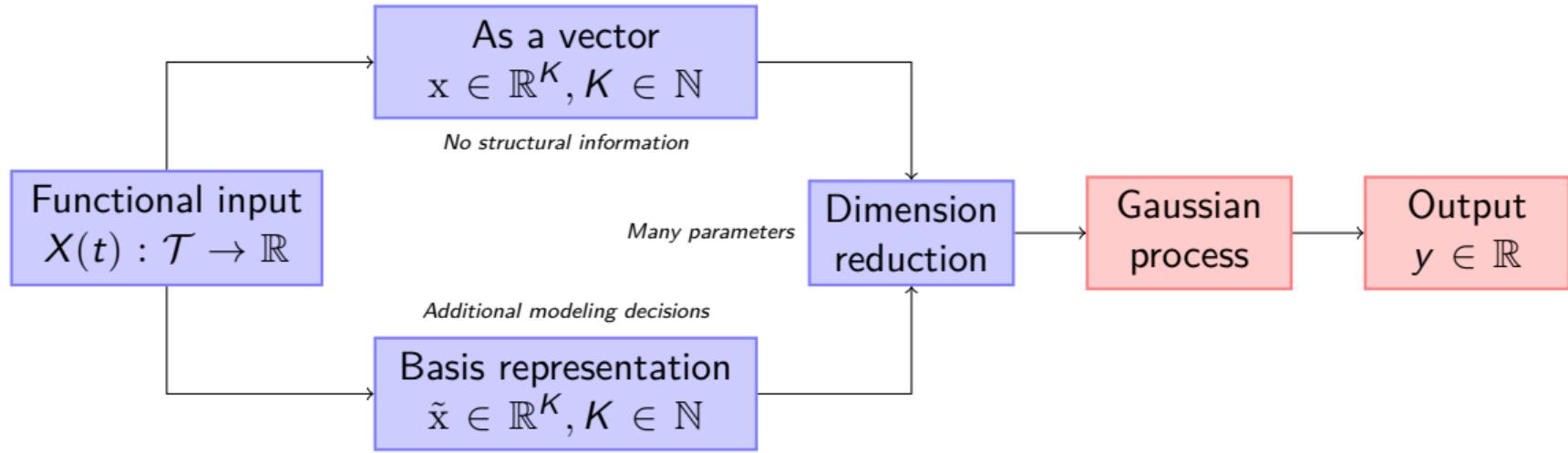
Functional input  
 $X(t) : \mathcal{T} \rightarrow \mathbb{R}$

Gaussian process → Output  
 $y \in \mathbb{R}$

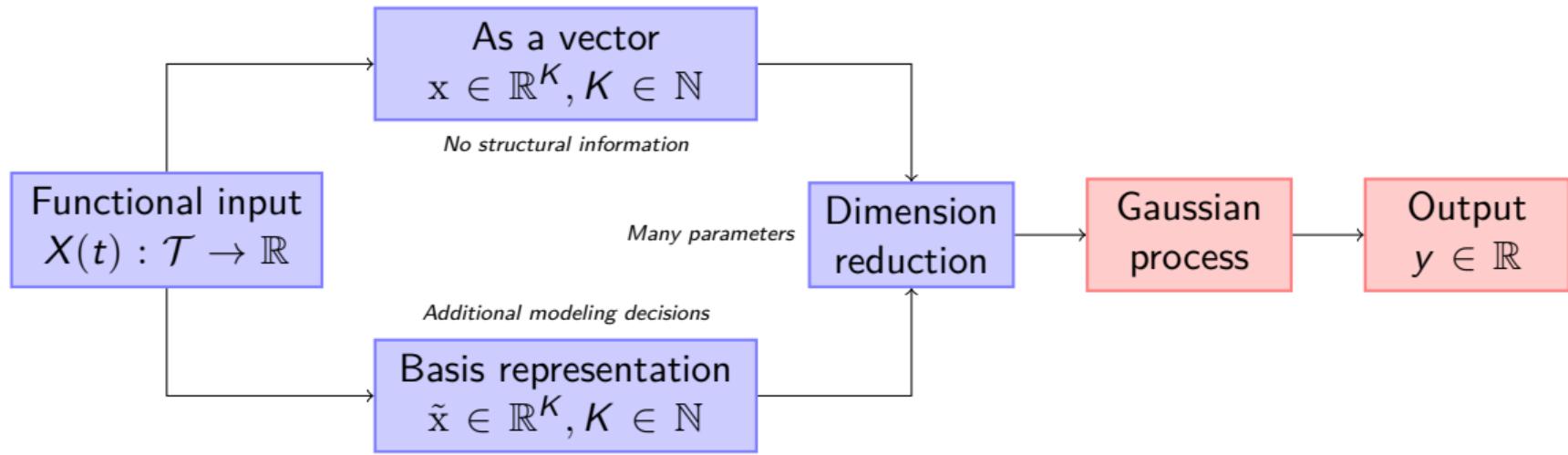
# Gaussian process with functional input



# Gaussian process with functional input

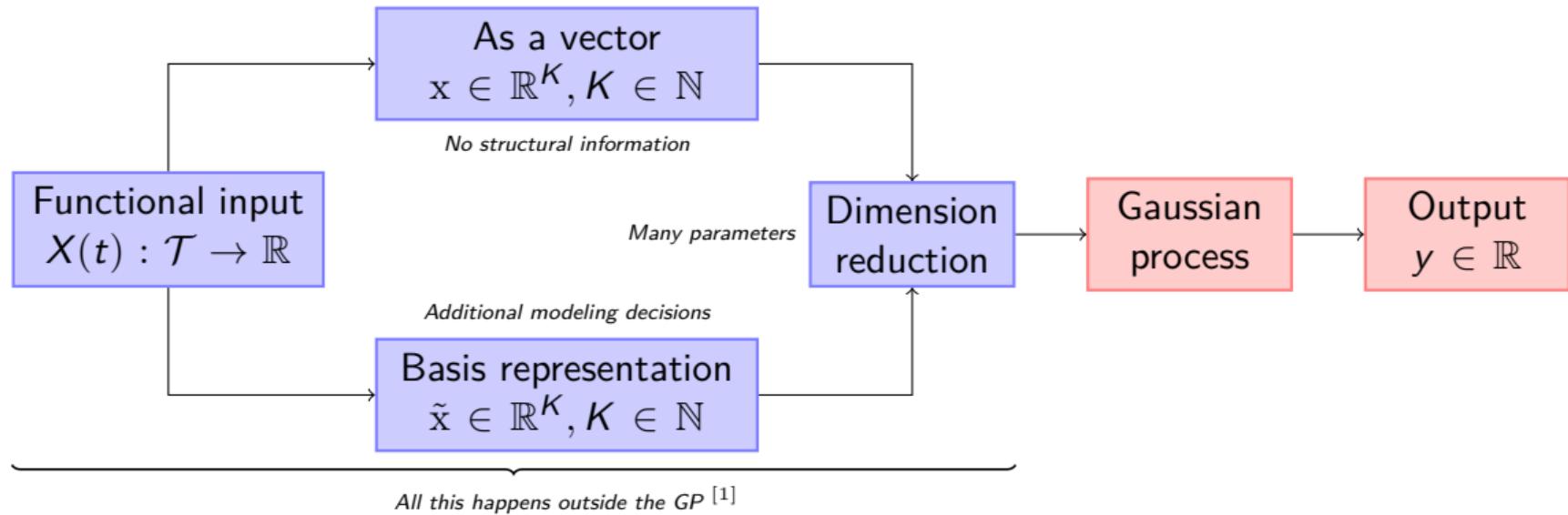


# Gaussian process with functional input



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

# Gaussian process with functional input

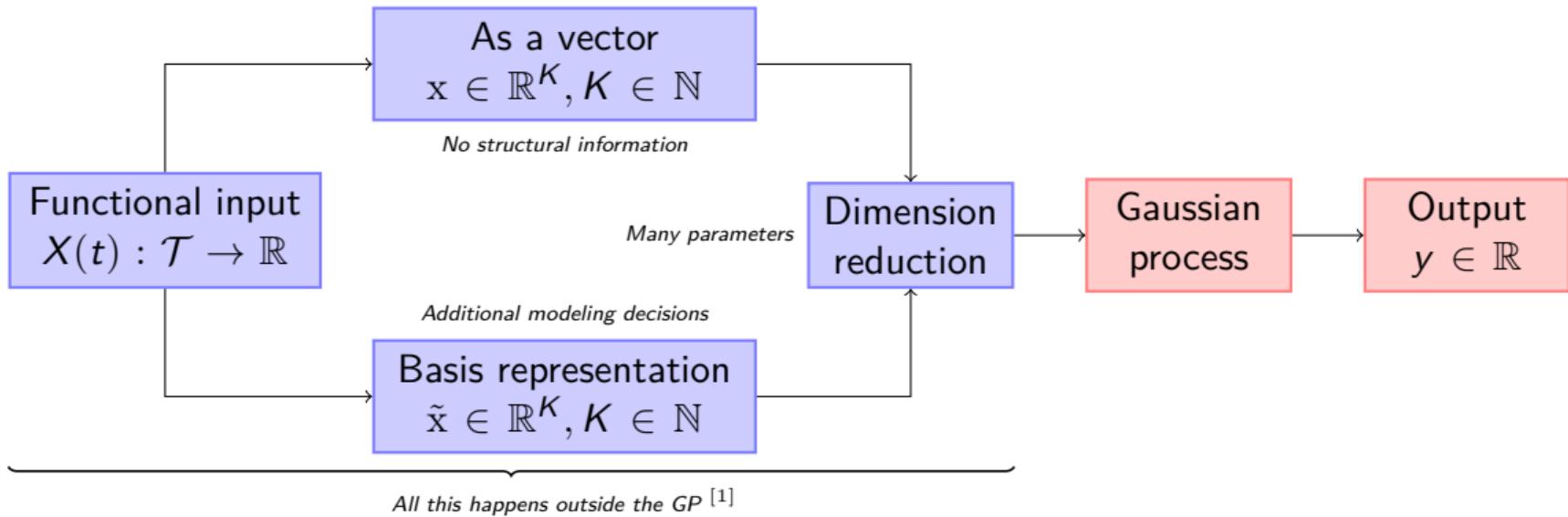


- Can we connect the functional input structure to a physical mechanism?
- Can we incorporate the functional input structure into the GP? [2]

[1][1, 2, 3, 4, 5, 6, 7, 8]

[2][9, 10]

# Gaussian process with functional input



- Can we connect the functional input structure to a physical mechanism?
- Can we incorporate the functional input structure into the GP? [2]
- Can we circumvent input dimension reduction?

[1][1, 2, 3, 4, 5, 6, 7, 8] [2][9, 10]

# 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?*

# From relevance to *dynamic* relevance

# Extending relevance

Some inputs  
matter more than others

$x_1$  vs  $x_2$

Screening  
*(exploration  
& diagnostics)*

Permutation  
Feature  
Importance [1]

Model tuning  
*(learning)*

Automatic  
Relevance  
Determination [2]

---

[1] [11, 12, 13, 14, 15, 16]

[2] [17, 18]

# 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 [1]

Model tuning  
(*learning*)

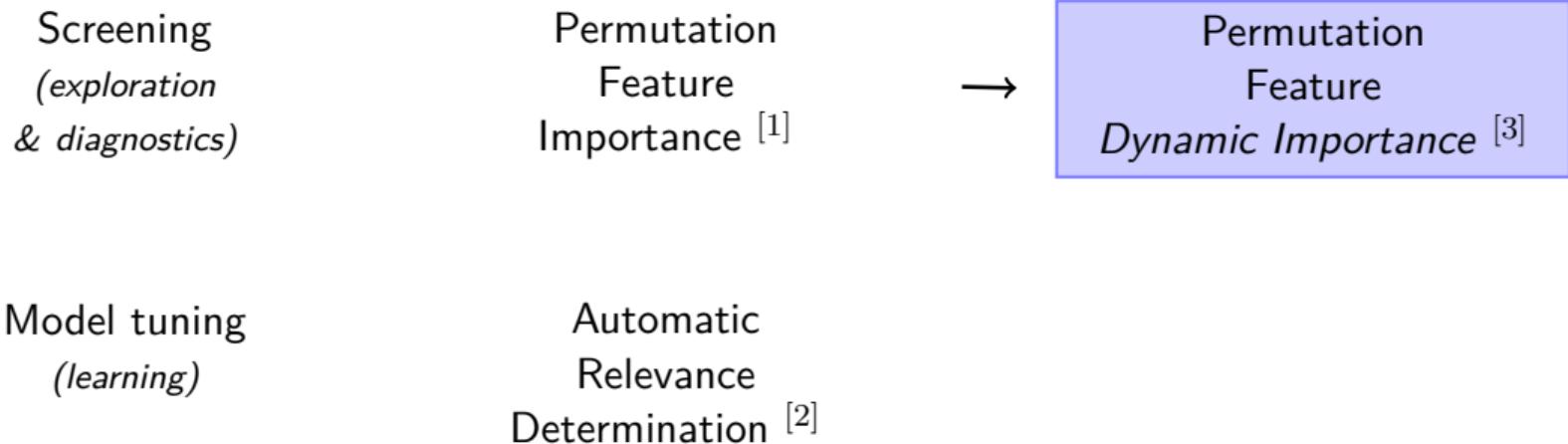
Automatic  
Relevance  
Determination [2]

[1] [11, 12, 13, 14, 15, 16]

[2] [17, 18]

# Extending relevance

Some inputs  
matter more than others  
 $x_1$  vs  $x_2$  → Some index subspaces  
matter more than others  
 $X(t_1)$  vs  $X(t_2)$



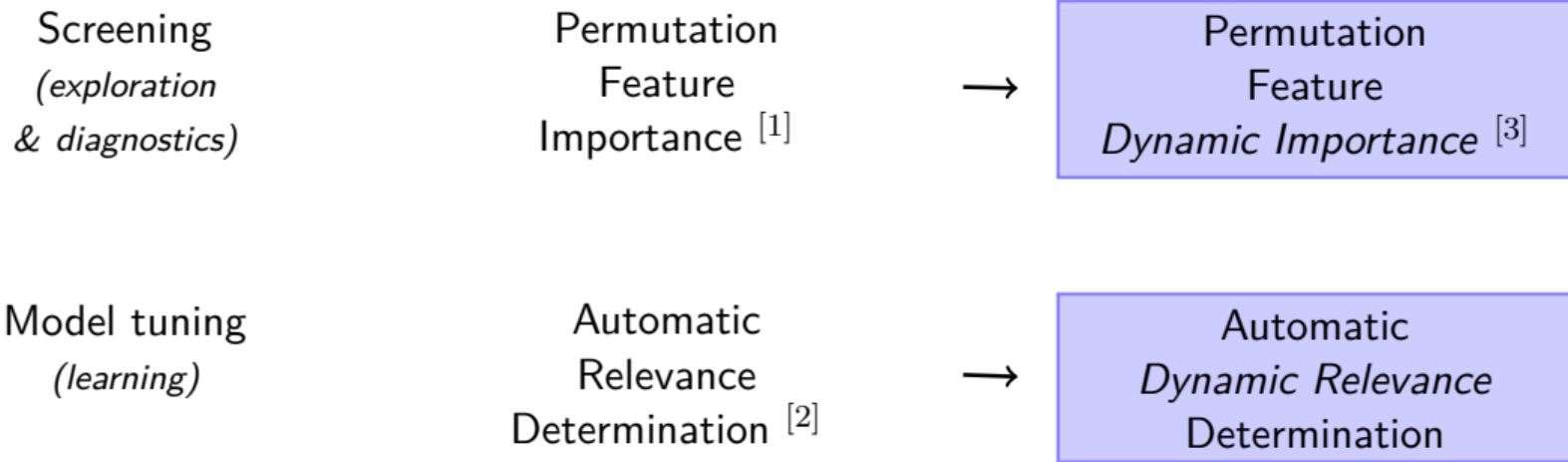
[1] [11, 12, 13, 14, 15, 16]

[2] [17, 18]

[3] 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  $X(t_2)$



[1] [11, 12, 13, 14, 15, 16]

[2] [17, 18]

[3] 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)*

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

# Modeling *dynamic* relevance

Model tuning  
(*learning*)

Automatic  
Relevance  
Determination



Automatic  
*Dynamic Relevance*  
Determination

Distance  
 $d(X_i, X_j)$

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

Weights  
(*relevance*)

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

# Modeling *dynamic* relevance

Model tuning  
(*learning*)

Automatic  
Relevance  
Determination



Automatic  
*Dynamic Relevance*  
Determination

Distance  
 $d(X_i, X_j)$

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

$$\int_T \omega(t) (X_i(t) - X_j(t))^2 dt$$

Weights  
(*relevance*)

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

# Modeling *dynamic* relevance

Model tuning  
(*learning*)

Automatic  
Relevance  
Determination



Automatic  
*Dynamic Relevance*  
Determination

Distance  
 $d(X_i, X_j)$

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

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

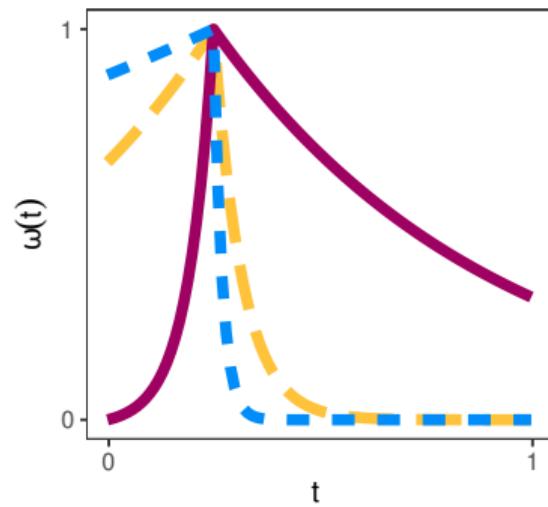
Weights  
(*relevance*)

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

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

# Asymmetric Laplace function

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



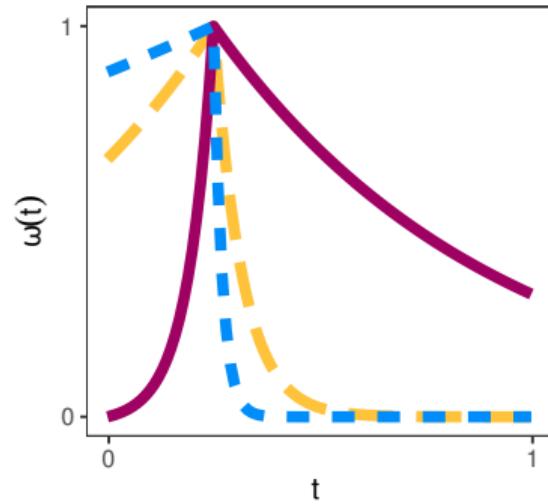
---

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

# Asymmetric Laplace function

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

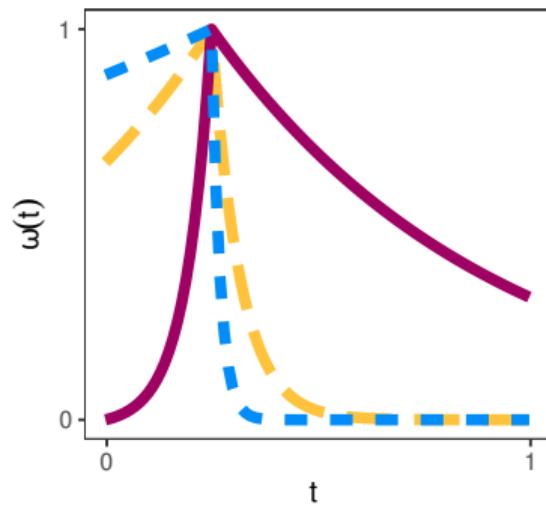
- The input is most relevant at  $\tau$  (relevance peak)



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

# Asymmetric Laplace function

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

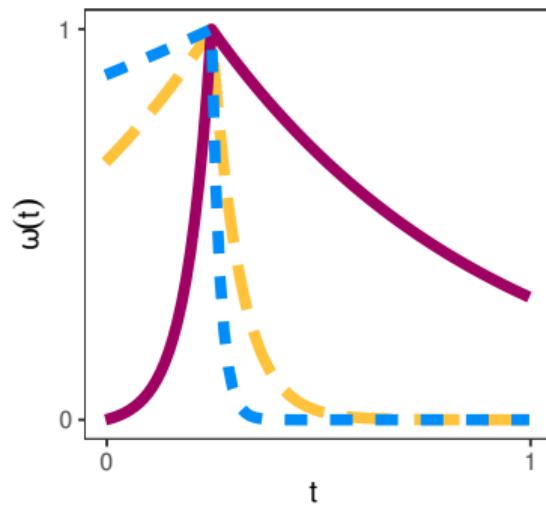


- The input is most relevant at  $\tau$  (relevance peak)
- Relevance increases at a  $\lambda_1 = \lambda\kappa^{-1}$  rate from  $t = 0$  to the peak

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

# Asymmetric Laplace function

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

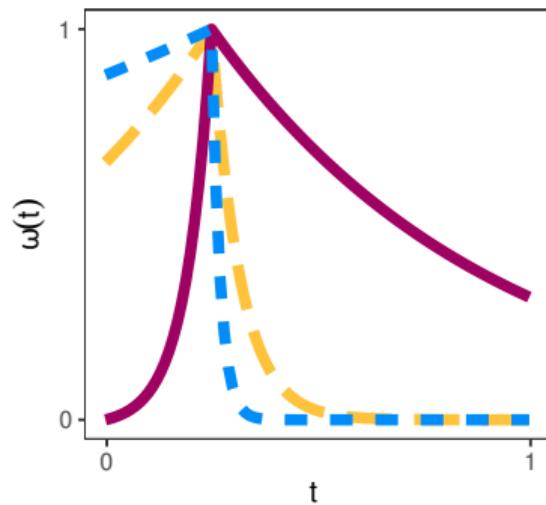


- The input is most relevant at  $\tau$  (relevance peak)
- Relevance increases at a  $\lambda_1 = \lambda\kappa^{-1}$  rate from  $t = 0$  to the peak
- Relevance decreases at a  $\lambda_2 = \lambda\kappa$  rate from the peak to  $t = 1$

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

# Asymmetric Laplace function

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

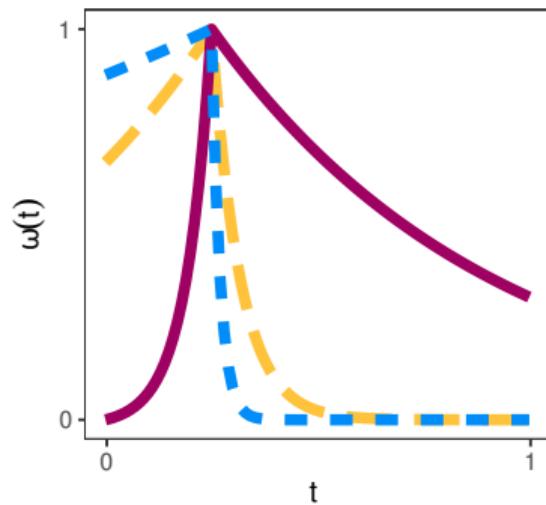


- The input is most relevant at  $\tau$  (relevance peak)
- Relevance increases at a  $\lambda_1 = \lambda\kappa^{-1}$  rate from  $t = 0$  to the peak
- Relevance decreases at a  $\lambda_2 = \lambda\kappa$  rate from the peak to  $t = 1$
- To predict the output, look for input profiles that are similar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

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

# Asymmetric Laplace function

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



- The input 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

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

# Functional Input Gaussian Process (fiGP)

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

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

(3)

(4)

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

# Functional Input Gaussian Process (fiGP)

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

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

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

(4)

---

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

# Functional Input Gaussian Process (fiGP)

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

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

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

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

---

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

# Functional Input Gaussian Process (fiGP)

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

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

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

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

Weakly informative priors  $\phi \sim \text{INVGAMMA}$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim N^+$ ,  $\log(\kappa) \sim N$

---

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

# Functional Input Gaussian Process (fiGP)

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

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

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

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

Weakly informative priors  $\phi \sim \text{INVGAMMA}$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim N^+$ ,  $\log(\kappa) \sim N$

Multiple inputs e.g., correlation product

---

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

# Functional Input Gaussian Process (fiGP)

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

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

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

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

Weakly informative priors  $\phi \sim \text{INVGAMMA}$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim N^+$ ,  $\log(\kappa) \sim N$

Multiple inputs e.g., correlation product

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

---

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

# Functional Input Gaussian Process (fiGP)

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

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

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

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

Weakly informative priors  $\phi \sim \text{INVGAMMA}$ ,  $\tau \sim \text{BETA}(\cdot, \cdot)$ ,  $\lambda \sim N^+$ ,  $\log(\kappa) \sim N$

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

---

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

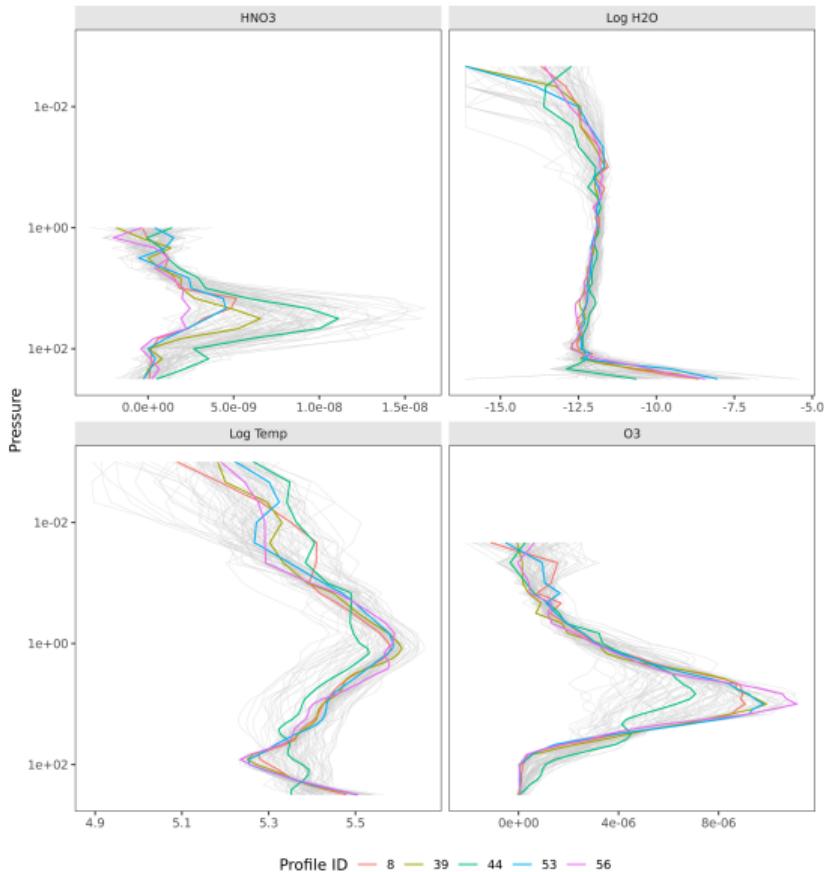
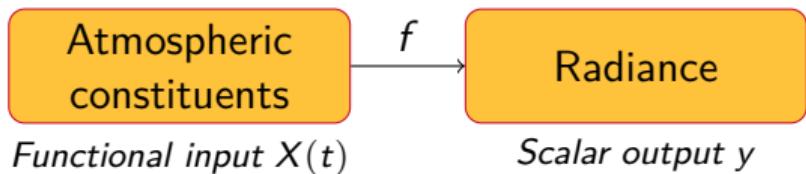
# NASA's Microwave Limb Sounder

*a case study*

# Data structure



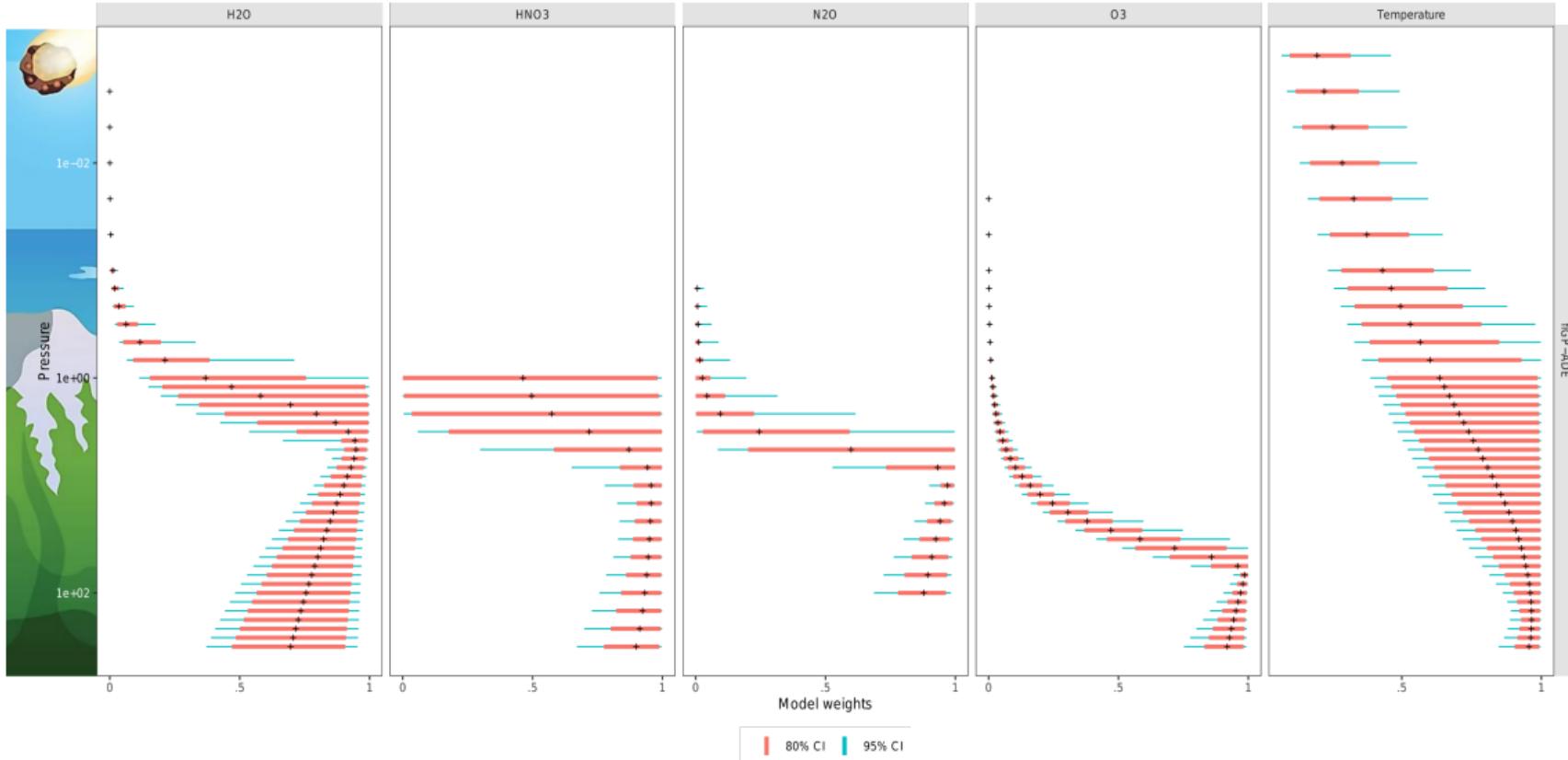
Credit: NASA Aura



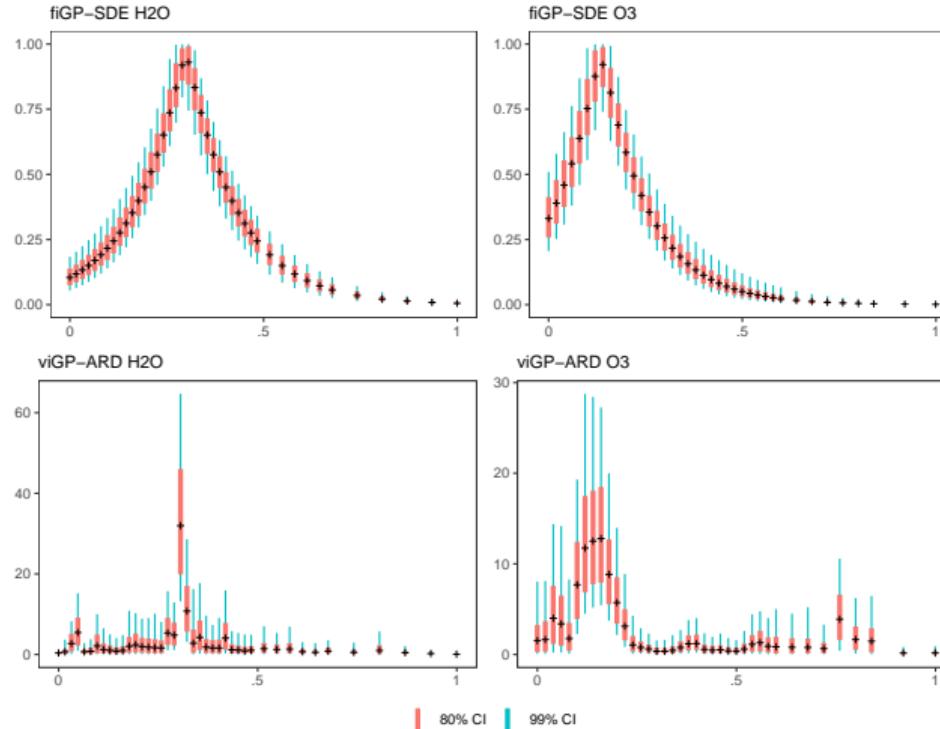
# Implementation

- 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



In this slide only, we fix  $\kappa = 1$  so that  $\omega(t)$  is symmetrical.

# Why a fiGP?

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

# 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

# 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 in the case study <sup>[1]</sup>

---

[1] Appendix slides and forthcoming paper

# 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 in the case study [1]
- ++ Extensible to **complex index spaces**, e.g., spatio-temporal spectral inputs [2]

---

[1] Appendix slides and forthcoming paper [2] Future research

# Acknowledgments

The MLS team at JPL, California Institute of Technology

# Thank you!

References and extra slides on the back

▶ mail [ldamiano@iastate.edu](mailto:ldamiano@iastate.edu)

▶ repo <https://github.com/luisdamiano/SIAMUQ22>

# Appendix

# References I

- [1] 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.
- [2] Simon Nanty, Céline Helbert, Amandine Marrel, Nadia Pérot, and Clémentine Prieur.  
Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs.  
*SIAM/ASA Journal on Uncertainty Quantification*, 4(1):636–659, January 2016.
- [3] Bo Wang, Tao Chen, and Aiping Xu.  
Gaussian process regression with functional covariates and multivariate response.  
*Chemometrics and Intelligent Laboratory Systems*, 163:1–6, April 2017.
- [4] Matthias H. Y. Tan and Guilin Li.  
Gaussian Process Modeling Using the Principle of Superposition.  
*Technometrics*, 61(2):202–218, April 2019.
- [5] Bo Wang and Aiping Xu.  
Gaussian process methods for nonparametric functional regression with mixed predictors.  
*Computational Statistics & Data Analysis*, 131:80–90, March 2019.
- [6] José Betancourt, François Bachoc, Thierry Klein, Déborah Idier, Rodrigo Pedreros, and Jérémie Rohmer.  
Gaussian process metamodeling of functional-input code for coastal flood hazard assessment.  
*Reliability Engineering & System Safety*, 198:106870, June 2020.
- [7] José Daniel Betancourt, François Bachoc, and Thierry Klein.  
Gaussian process regression for scalar and functional inputs with funGp - the in-depth tour.  
April 2020.

## References II

- [8] Zhaohui Li and Matthias Hwai Yong Tan.  
A Gaussian Process Emulator Based Approach for Bayesian Calibration of a Functional Input.  
*Technometrics*, pages 1–13, October 2021.
- [9] Max D. Morris.  
Gaussian surrogates for computer models with time-varying inputs and outputs.  
*Technometrics*, 54(1):42–50, February 2012.
- [10] Gulzina Kuttubekova.  
Emulator for water erosion prediction project computer model using gaussian processes with functional inputs.  
*Creative Components*, January 2019.
- [11] Leo Breiman.  
Random forests.  
*Machine Learning*, 45(1):5–32, 2001.
- [12] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn.  
Bias in random forest variable importance measures: Illustrations, sources and a solution.  
*BMC Bioinformatics*, 8(1):25, December 2007.
- [13] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis.  
Conditional variable importance for random forests.  
*BMC Bioinformatics*, 9(1):307, December 2008.
- [14] Kristin K Nicodemus, James D Malley, Carolin Strobl, and Andreas Ziegler.  
The behaviour of random forest permutation-based variable importance measures under predictor correlation.  
*BMC Bioinformatics*, 11(1):110, December 2010.

# References III

- [15] Aaron Fisher, Cynthia Rudin, and Francesca Dominici.  
All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously.  
*Journal of machine learning research: JMLR*, 20:177, 2019.
- [16] Giles Hooker, Lucas Mentch, and Siyu Zhou.  
Unrestricted permutation forces extrapolation: Variable importance requires at least one more model, or there is no free variable importance.  
[arXiv:1905.03151 \[cs, stat\]](https://arxiv.org/abs/1905.03151), October 2021.
- [17] Radford M. Neal.  
*Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes in Statistics*.  
Springer New York, New York, NY, 1996.
- [18] Juho Piironen and Aki Vehtari.  
Projection predictive model selection for gaussian processes.  
In [2016 IEEE 26th International Workshop on Machine Learning for Signal Processing \(MLSP\)](#), pages 1–6, September 2016.
- [19] Margaret Johnson.  
Forward model emulation for NASA's microwave limb sounder, August 2020.

# Notation

# Notation

Index vector  $t$  with the vertical pressure level

## Notation

**Index vector**  $t$  with the vertical pressure level

**State vector**  $x_i$  characterizing an atmospheric vertical profile

# Notation

**Index vector**  $t$  with the vertical pressure level

**State vector**  $x_i$  characterizing an atmospheric vertical profile

**Output scalar**  $y_i$  is the radiance first functional principal component [19]

# Notation

**Index vector**  $t$  with the vertical pressure level

**State vector**  $x_i$  characterizing an atmospheric vertical profile

**Output scalar**  $y_i$  is the radiance first functional principal component [19]

**Sounding** a collection of an observed radiance, retrieved state and pressure vectors

# Trapezoidal approximation

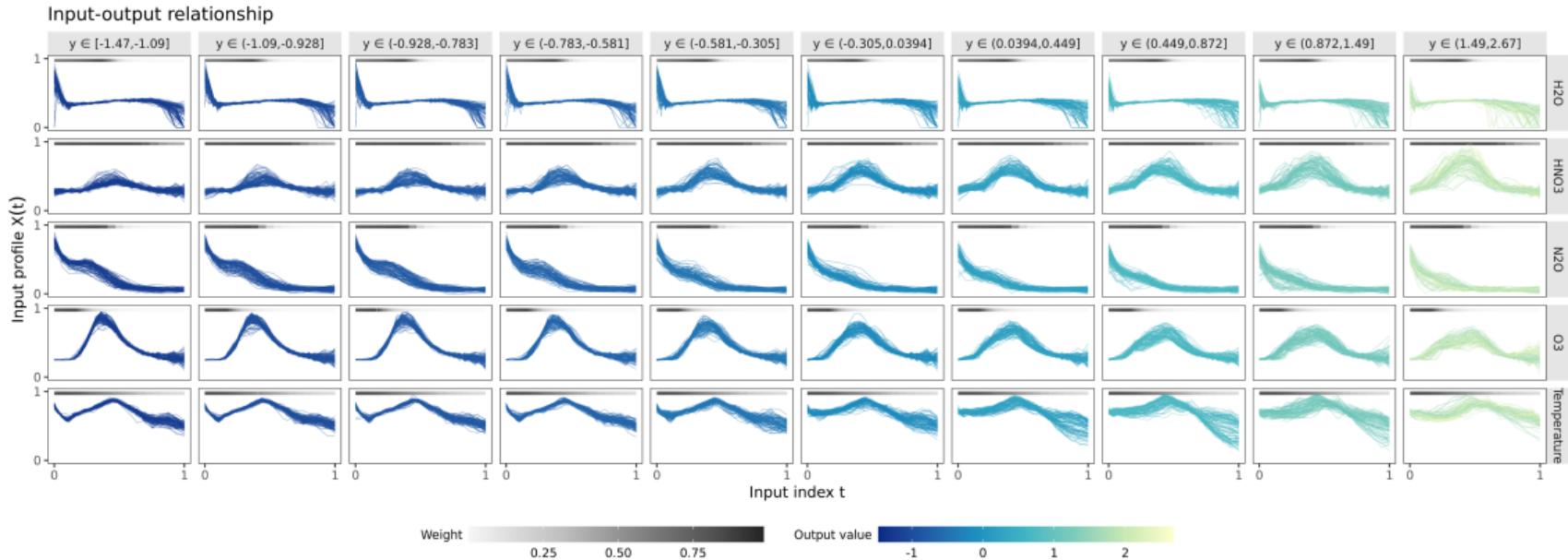
$$\int_{\mathcal{T}} \omega(t) (X_i(t) - X_j(t))^2 dt \approx \sum_{k=2}^K (t_k - t_{k-1}) \frac{\Delta_{i,j,k} + \Delta_{i,j,k-1}}{2} \quad (5)$$

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

---

See [1] for a B-spline approach

# Weights, inputs and output



Panes: profiles grouped by input variable (row) and output decile (column). Color: dark (light) for low (high) output values. Background: dark (light) for large (small) weights.

# 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	-7	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>	2	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.