

Advancing Scientific Practice through Agricultural Statistics

Jarad Niemi

Iowa State University

September 25, 2022

Funded, in part, by

- the Iowa State University Presidential Interdisciplinary Research Initiative on C-CHANGE: Science for a Changing Agriculture
- USDA NIFA CAP: Consortium for Cultivating Human And Naturally reGenerative Enterprises (C-CHANGE)
- Foundation for Food and Agriculture Research: Prairie Strips for Healthy Soils and Thriving Farms

Collaborators

Prairie STRIPS Collaborators: <http://prairiestrips.org/people>

Gaussian Process Emulators:



Jarad Niemi (ISU)



Advancing Scientific Practice through Agricultural Statistics



September 25, 2022

Agriculture

Iowa Agricultural Production

<https://www.iadg.com/iowa-advantages/target-industries/>

Iowa is the largest producer of corn, pork and eggs in the United States and second in soybeans and red meat production.



<https://www.britannica.com/plant/corn-plant>

<https://www.nationalhogfarmer.com/marketing/total-pork-production-2014-down-slightly>

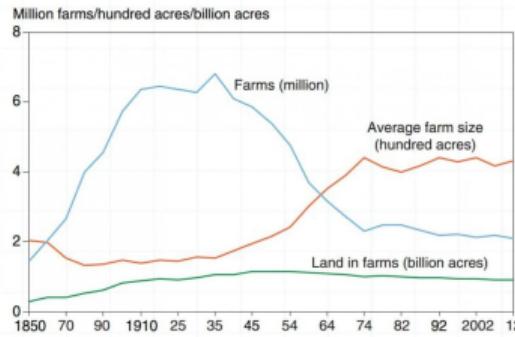
<https://www.medicalnewstoday.com/articles/283659>

<https://www.midwestfarmreport.com/2019/12/11/state-soybean-yield-contest-entries-announced/>

<https://www.scientificamerican.com/article/meat-and-environment/>

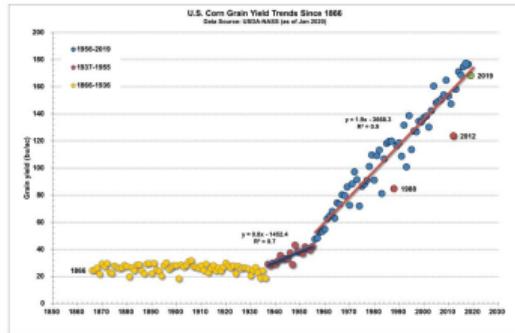
Corn production

Farms, land in farms, and average acres per farm, 1850–2012



Source: USDA, Economic Research Service using data from USDA, National Agricultural Statistics Service, Census of Agriculture.

<https://www.iowaagliteracy.org/Article/Family-Farms-Then-and-Now>



<https://extension.entm.purdue.edu/newsletters/pestandcrop/article/historical-corn-grain-yields-in-the-u-s/>

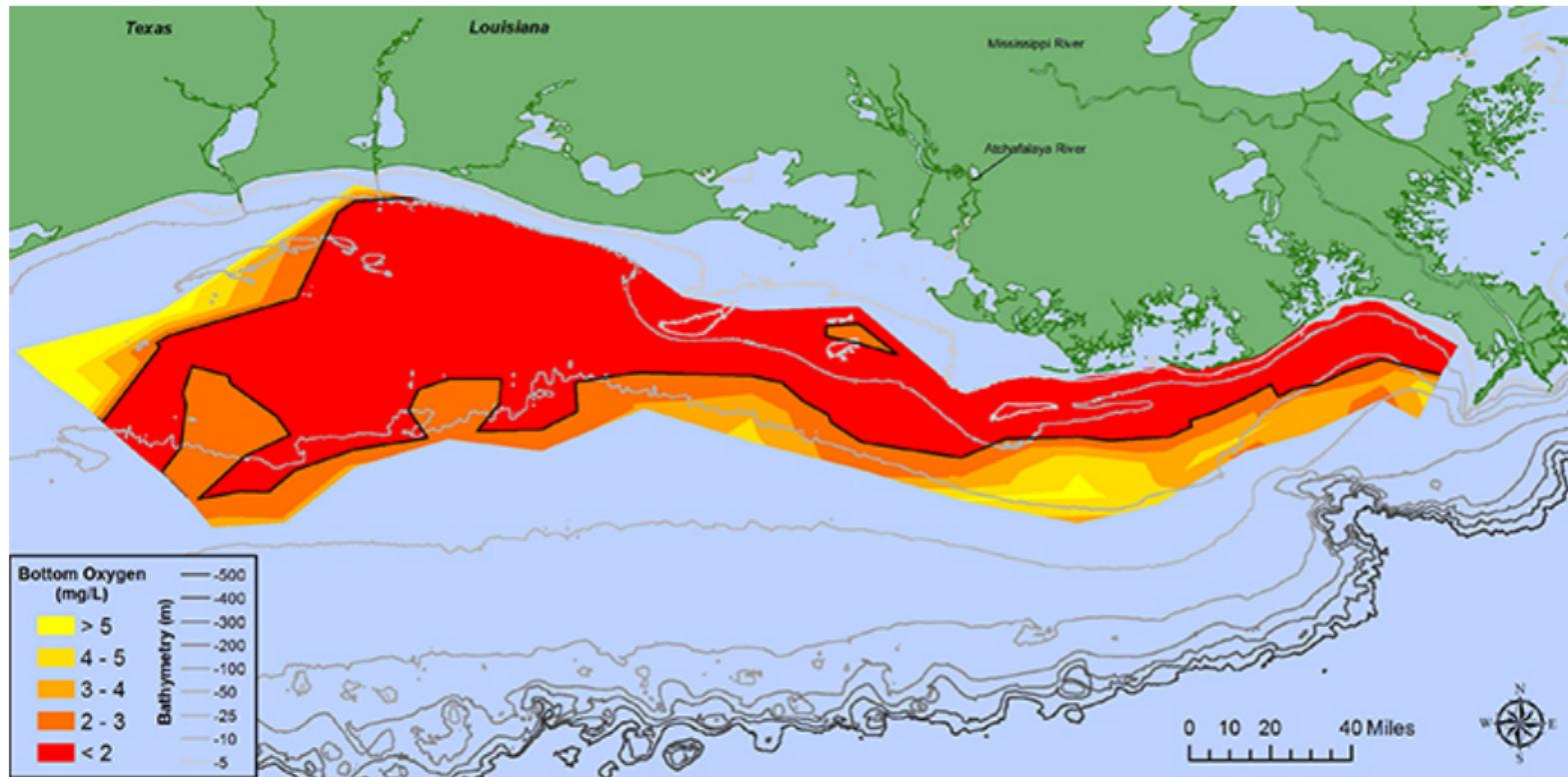
Soil loss

Iowa loses \$1,000,000,000/year in soil



<https://www.desmoinesregister.com/story/money/agriculture/2014/05/03/erosion-estimated-cost-iowa-billion-yield/8682651/>

Gulf of Mexico Dead Zone



<https://www.noaa.gov/media-release/gulf-of-mexico-dead-zone-is-largest-ever-measured>

C-CHANGE: Science for a changing agriculture



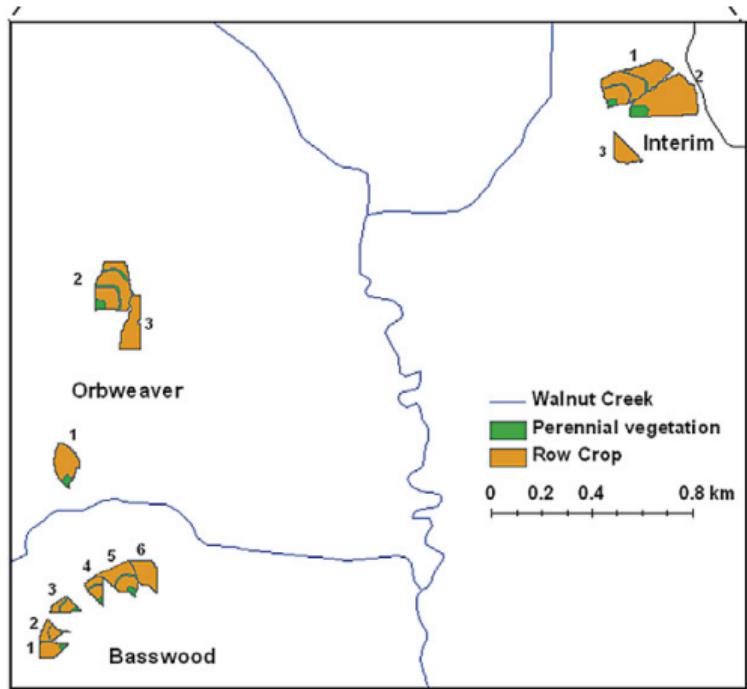
C·CHANGE

<http://agchange.org>

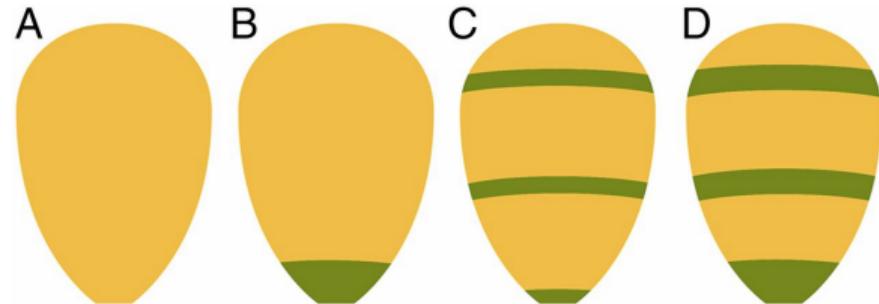
Prairie STRIPS
C-CHANGE PIRIR
C-CHANGE Grass2Gas
Climate Smart Ag

Prairie STRIPS

STRIPS1 - Neal Smith National Wildlife Refuge



Find on Google Maps



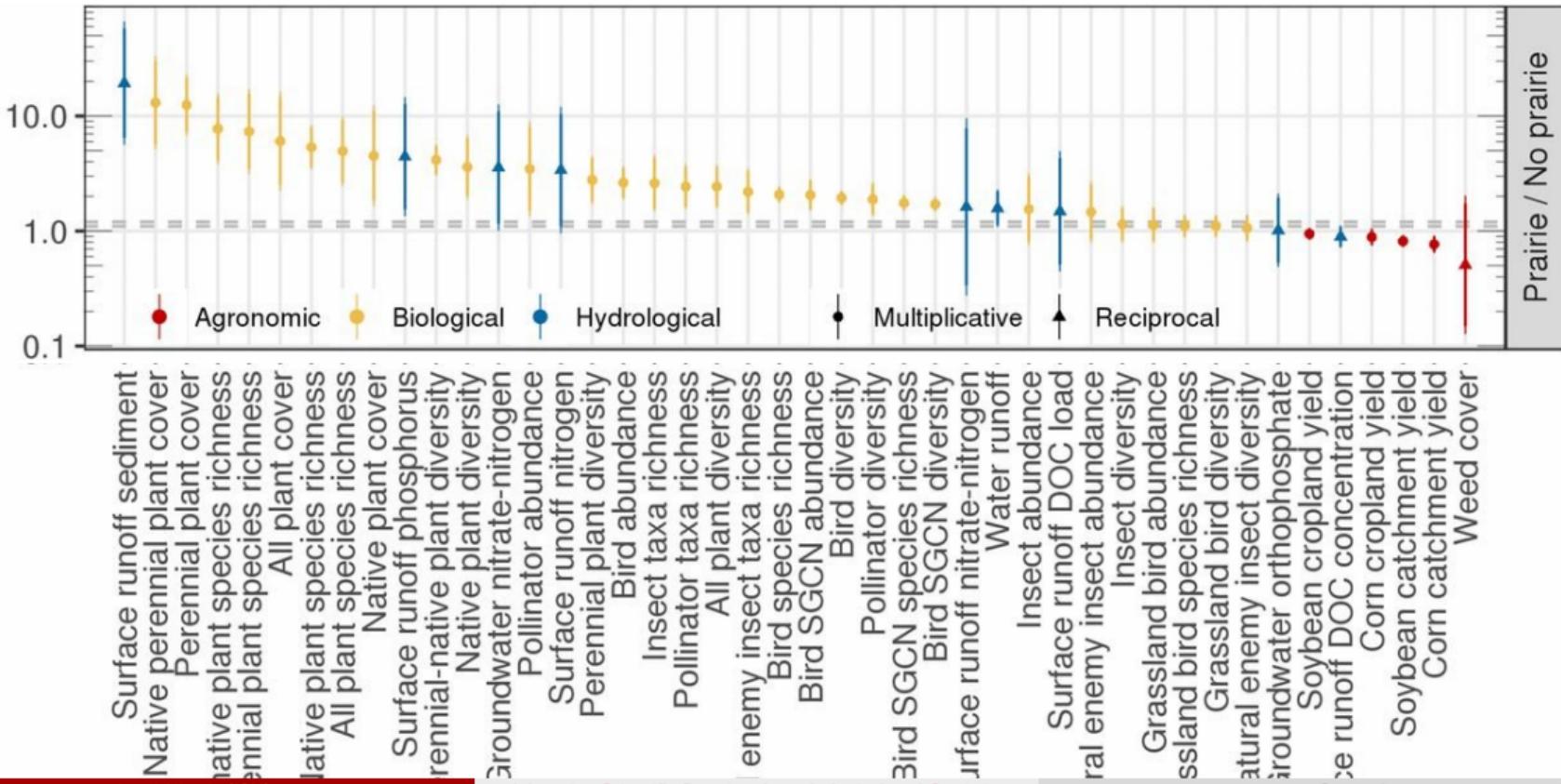
| Field | Treatment | | | |
|----------------|-----------|-------|-------|-----|
| | Control | 10% B | 10% S | 20% |
| Interim | X | X | X | |
| Orbweaver | X | X | | X |
| Basswood-South | | X | X | X |
| Basswood-North | X | | X | X |

STRIPS1 - Science



STRIPS1 - PNAS

procal



USDA CRP: CP-43

Prairie strips improve biodiversity and the delivery of multiple ecosystem services from corn–soybean croplands

Lisa A. Schulte   , Jarad Niemi  , Matthew J. Helmers  , and Chris Witte [Authors Info & Affiliations](#)



United States Department of Agriculture

Farm Service Agency
Conservation Reserve Program

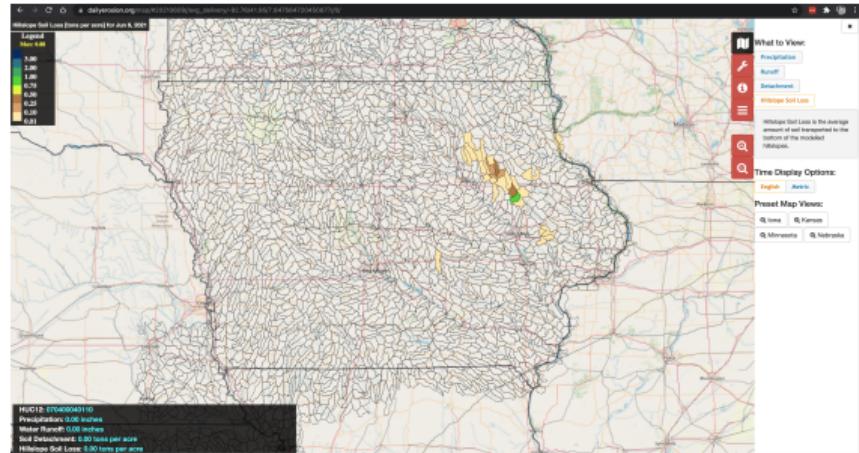
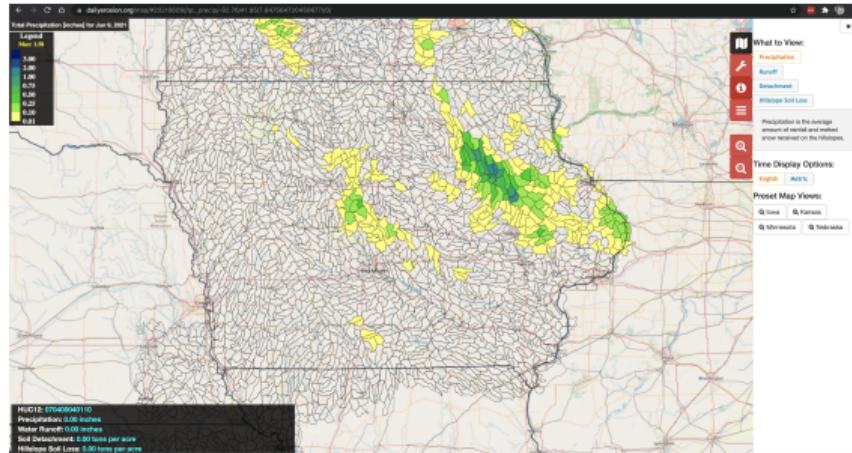
CLEAN LAKES, ESTUARIES AND RIVERS (CLEAR) INITIATIVE
Prairie Strip Practice (CP-43)

FACT SHEET
December 2019

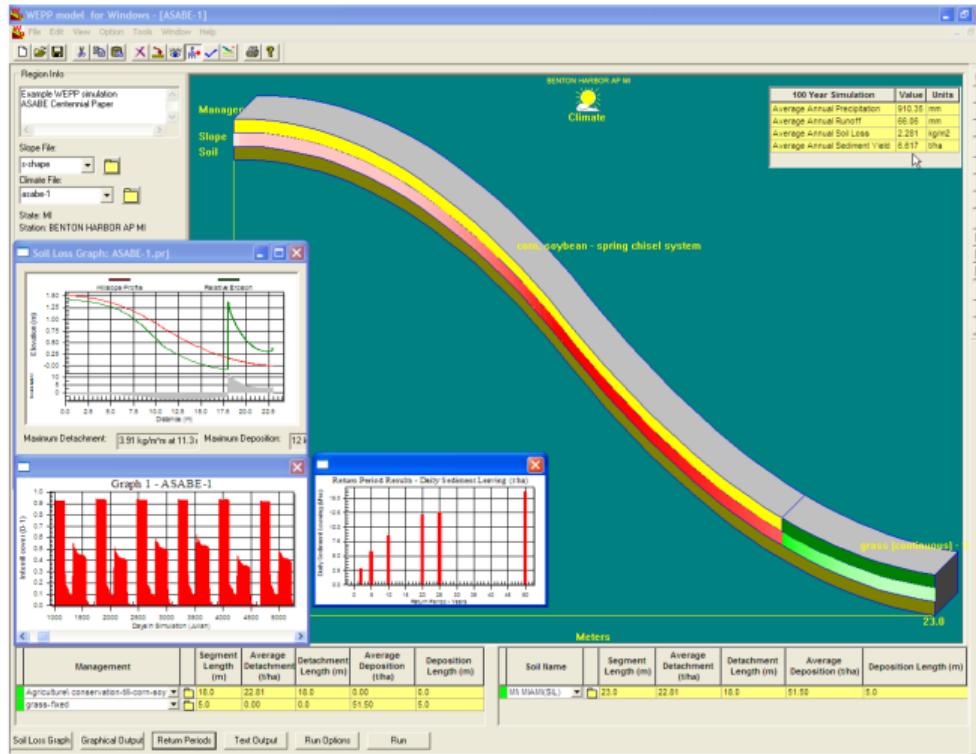
Gaussian Process Emulators

Daily Erosion Project (DEP)

using Water Erosion Prediction Project (WEPP)



Water Erosion Prediction Project (WEPP)



Gaussian Process Emulators

Consider a deterministic computer model $f(\cdot)$ with

$$Y_i = f(X_i, \psi) = f(X_i), \quad i = 1, \dots, N$$

where

- ▶ ψ are model parameters,
- ▶ X_i are inputs, and
- ▶ Y_i are outputs.

We assume a Gaussian Process prior

$$f \sim \mathcal{GP}(m, k) \implies Y \sim N(m_x, \Sigma)$$

where (often) $m_x = 0$ and, for scalar inputs, $\Sigma_{ij} = k(x_i, x_j) = \sigma^2 e^{-(x_i - x_j)^2 / \phi}$.

Two research directions:

- ▶ Computational tractability for large N
- ▶ Dealing with functional inputs

Training a GP

Find the maximum likelihood estimator (MLE) for $\theta = (\sigma^2, \phi)$,

$$\hat{\theta} = \operatorname{argmax}_{\theta} p(y|\theta) = \operatorname{argmax}_{\theta} N(y; 0, \Sigma(\theta))$$

where $y = (y_1, \dots, y_N)$. The log-likelihood is

$$\log \mathcal{N}(y; 0, \Sigma(\theta)) \propto C - \log |\Sigma(\theta)| - y^\top \Sigma(\theta)^{-1} y$$

If there are N observations, $\Sigma(\theta)$ is an $N \times N$ covariance matrix and thus the computational time scales as $\mathcal{O}(N^3)$.

This is doable if $N \approx 1,000$ but not when you start getting larger and larger data sets.

OAT Algorithm

Fully Independent Conditional (FIC) Approximation

Introduce a set of knots $x^\dagger = \{x_1^\dagger, \dots, x_K^\dagger\}$ with $K \ll N$, such that

$$p(f_x, f_{x^\dagger} | \theta) = p(f_x | f_{x^\dagger}, \theta)p(f_{x^\dagger} | \theta)$$

where

$$\begin{aligned} f_{x^\dagger} | \theta &\sim \mathcal{N}(0, \Sigma_{x^\dagger x^\dagger}) \\ f_x | f_{x^\dagger}, \theta &\sim \mathcal{N}(\Sigma_{xx^\dagger} \Sigma_{x^\dagger x^\dagger}^{-1} (f_{x^\dagger}), \Lambda) \end{aligned}$$

with $\Lambda = \text{diag}(\Sigma_{xx} - \Sigma_{xx^\dagger} \Sigma_{x^\dagger x^\dagger}^{-1} \Sigma_{x^\dagger x})$.

This joint implies the following marginal distribution for f_x :

$$f_x | \theta \sim \mathcal{N}(0, \Lambda + \Sigma_{xx^\dagger} \Sigma_{x^\dagger x^\dagger}^{-1} \Sigma_{x^\dagger x})$$

which has the correct marginal means and variances, but the covariances are controlled by the knots.

Train FIC Model

Let $\Psi(x^\dagger, \theta) \equiv \Lambda(\theta) + \Sigma_{xx^\dagger}(\theta)\Sigma_{x^\dagger x^\dagger}(\theta)^{-1}\Sigma_{x^\dagger x}(\theta)$, then

$$Y|x^\dagger, \theta \sim \mathcal{N}(0, \Psi(x^\dagger, \theta)).$$

Train the model by finding

$$\hat{x}^\dagger, \hat{\theta} = \operatorname{argmax}_{x^\dagger, \theta} \mathcal{N}(y; 0, \Psi(x^\dagger, \theta)).$$

which has computational complexity of $\mathcal{O}(NK^2)$.

There are a number of questions:

- ▶ how many knots are needed?
- ▶ where should the knots be?

One-at-a-time (OAT) selection

We developed a one-at-a-time (OAT) knot selection that

- ▶ Begins with a small number of knots
- ▶ Optimizes the knot locations according to the marginal likelihood or variational objective function
- ▶ Iteratively adds knots until no improvement is seen in the objective function

Summary of results:

- ▶ Prediction is equivalent to Full GP and simultaneous knot optimization
- ▶ Formally, OAT has computational complexity of $\mathcal{O}(NK^2)$
- ▶ Practically, OAT is much faster than Full GP and simultaneous knot optimization

Manuscripts:

- ▶ Nate Garton, Jarad Niemi, and Alicia Carriquiry. (2020) "Knot Selection in Sparse Gaussian Processes with a Variational Objective." *Statistical Analysis and Data Mining.* 13(4): 324-336.
- ▶ Nate Garton, Jarad Niemi, and Alicia Carriquiry. "Knot Selection in Sparse Gaussian processes." *arXiv:2002.09538*

Functional inputs

Vector-input Gaussian Process (viGP)

For observation i , we have response $Y_i \in \mathbb{R}$ and input $X_i = (X_{i,1}, \dots, X_{i,D})$. Our computer model is $f()$ with $Y_i = f(X_i)$.

Assume f is a zero-mean Gaussian process with

$$\text{Cov}(Y_i, Y_j) = \sigma^2 e^{-\frac{1}{2}D(X_i, X_j, \omega)}$$

and

$$D(X_i, X_j, \omega) = \sum_{d=1}^D \omega_d (x_{i,d} - x_{j,d})^2.$$

Some call this **automatic relevance determination**.

Functional-input Gaussian Process

For observation i , we have $Y_i \in \mathbb{R}$ and $X_i(t)$ for $t \in [0, T]$. Our computer model is $f()$ with $Y_i = f(X_i(t))$.

The functional-input Gaussian Process has

$$\text{Cov}(Y_i, Y_j) = \sigma^2 e^{-\frac{1}{2}D(X_i(t), X_j(t), \omega)}$$

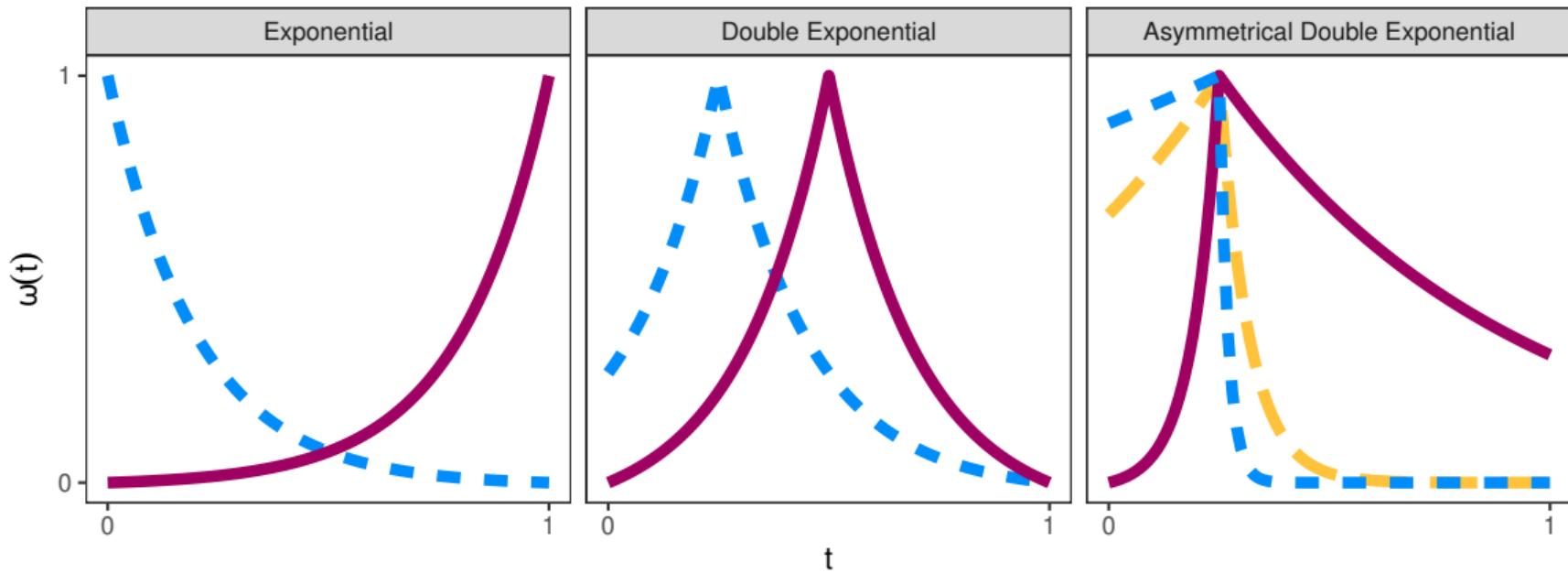
$$D(X_i, X_j, \omega) = \int_0^T \omega(t)(X_i(t) - X_j(t))^2 dt \approx \sum_{d=1}^D \omega(t_d)(X_i(t_d) - X_j(t_d))^2.$$

For the **functional length-scale**, we assume

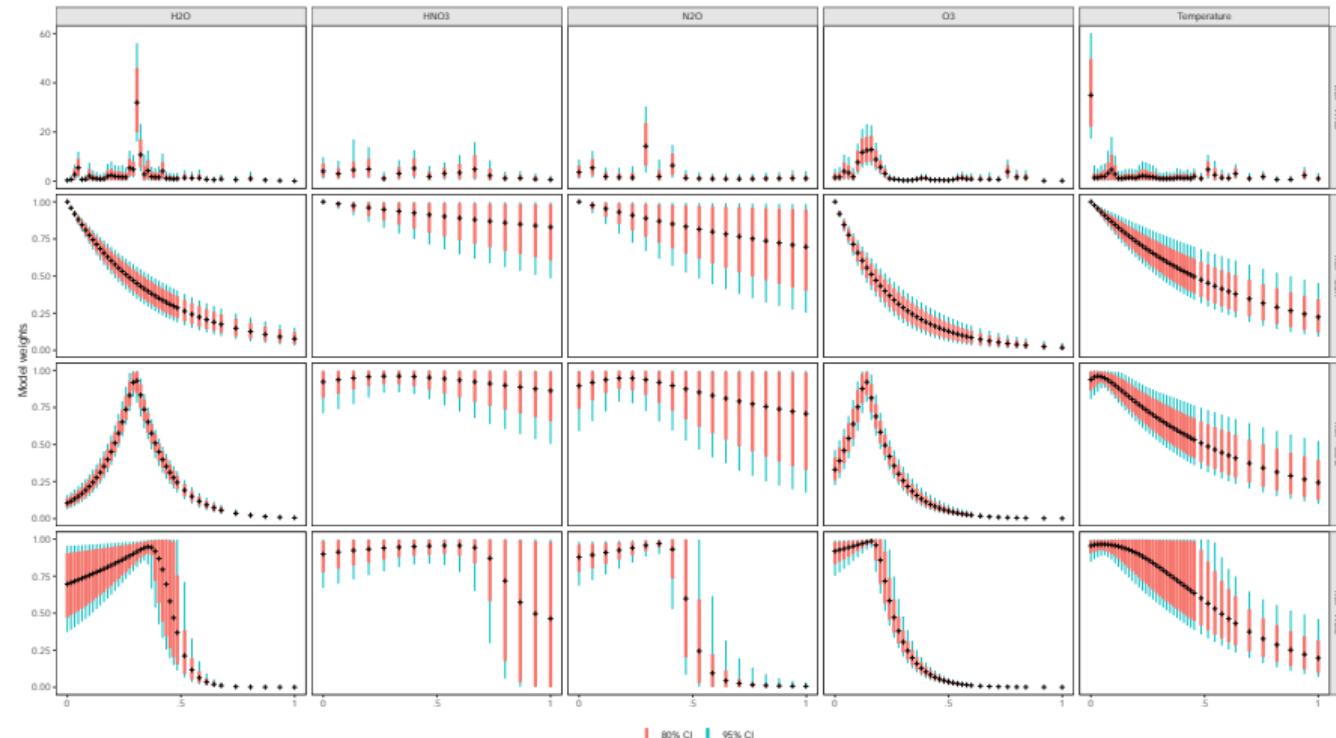
$$\omega(t) = \exp(2\sigma_\ell^2 t^\eta).$$

We'll refer to this as **fiGP**.

Theoretical fiGP functional length-scale



Estimated fiGP functional length-scale



Summary

Research advances

- ▶ OAT algorithm for knot selection to deal with computational intractability due to large n
- ▶ fiGP model for functional inputs

These slides are available at

- ▶ <https://github.com/jarad/ISU2022>
- ▶ <http://www.jarad.me/research/presentations.html>

Thank you!

Other links:

- ▶ <http://www.jarad.me/>
- ▶ <http://www.youtube.com/c/jaradniemi>