

Gaussian Processes with Functional Length-Scales for Agricultural Models

Jarad Niemi

Iowa State University

June 11, 2021

Collaborators:

- ▶ Luis Damiano, PhD Candidate
- ▶ Gulzina Kuttubekova, ISU MS
- ▶ Aditya Ranade, ISU MS
- ▶ Matt Nowatzke, ISU PhD Candidate
- ▶ Andy VanLoocke, ISU Associate Professor
- ▶ Emily Heaton, Illinois Professor

Funded, in part, by

C-CHANGE: Science for a changing agriculture



C-CHANGE

<http://agchange.org>



<http://prairiestrips.org>

Prairie STRIPS

PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or E

Home Articles Front Matter News Podcasts Authors

NEW RESEARCH IN

Physical Sciences

Social Sciences

RESEARCH ARTICLE



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

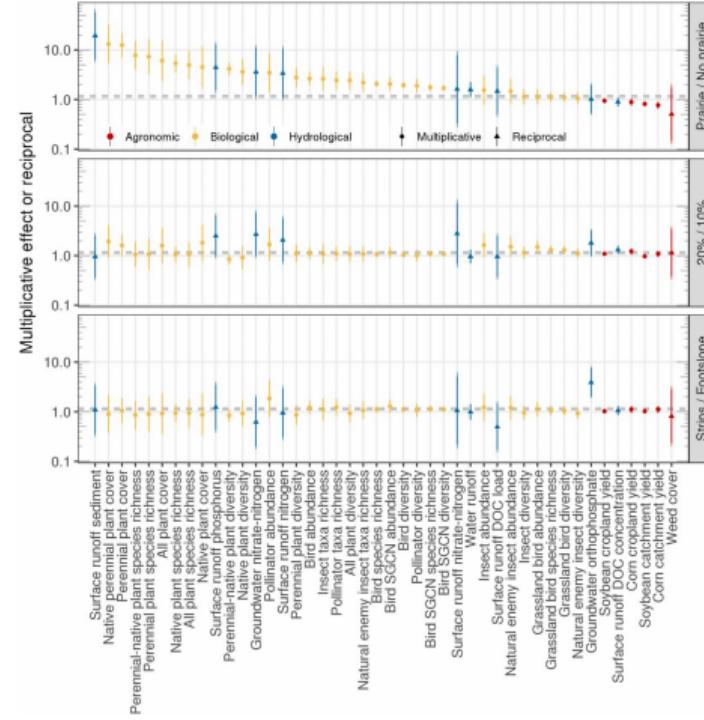
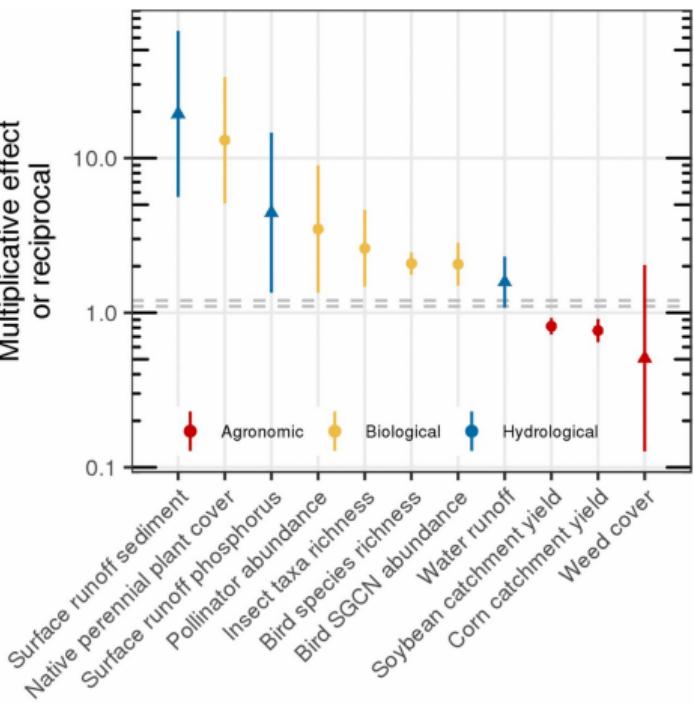
✉ Lisa A. Schulte, Ⓛ Jarad Niemi, Matthew J. Helmers, Matt Liebman, Ⓛ J. Gordon Arbuckle, David E. James, Randall K. Kolka, Matthew E. O’Neal, Mark D. Tomer, John C. Tyndall, Heidi Asbjørnsen, Pauline Drobney, Jeri Neal, Gary Van Ryswyk, and Chris Witte

PNAS October 17, 2017 114 (42) 11247-11252; first published October 2, 2017 https://doi.org/10.1073/pnas.1620229114

Edited by David Tilman, University of Minnesota, St. Paul, MN, and approved August 1, 2017 (received for review December 9, 2016)

<https://www.pnas.org/content/114/42/11247.short>

Prairie STRIPS results



<https://www.pnas.org/content/114/42/11247/tab-figures-data>

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/>

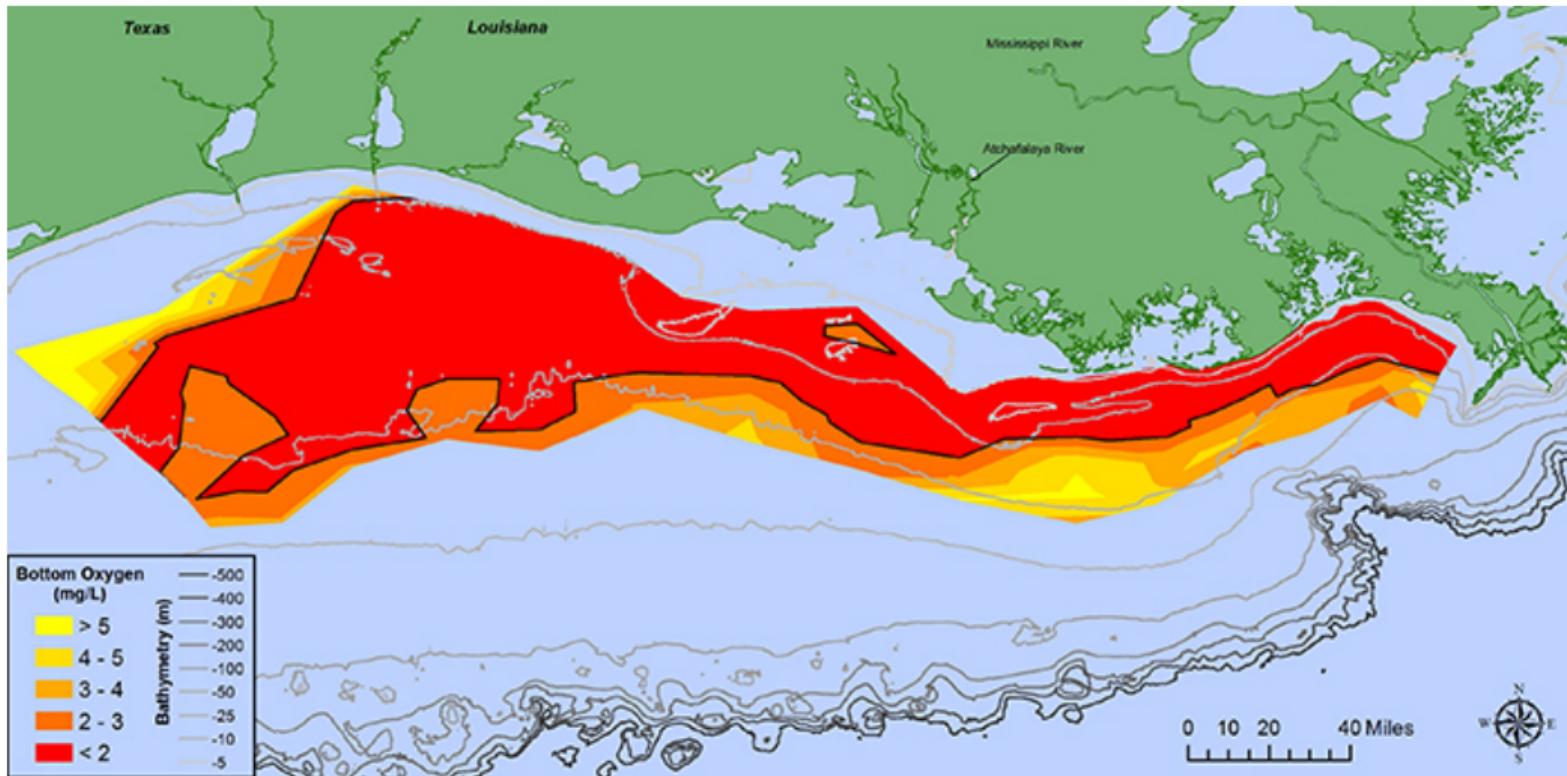
Soil loss

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>

Des Moines Water Works Lawsuit



<https://www.lwvumrr.org/blog/des-moines-water-works-lawsuit-update>

USDA NIFA - Biogas production from manure and herbaceous biomass



USDA Scientific Research Program Funds Sustainable Agricultural Systems Projects

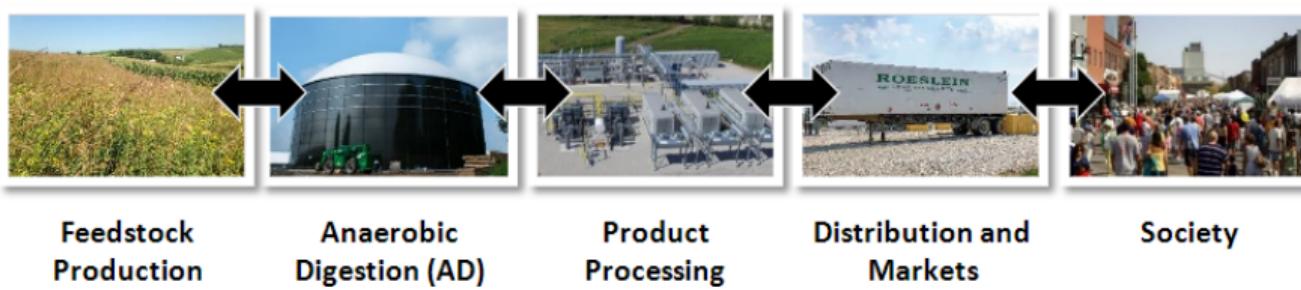
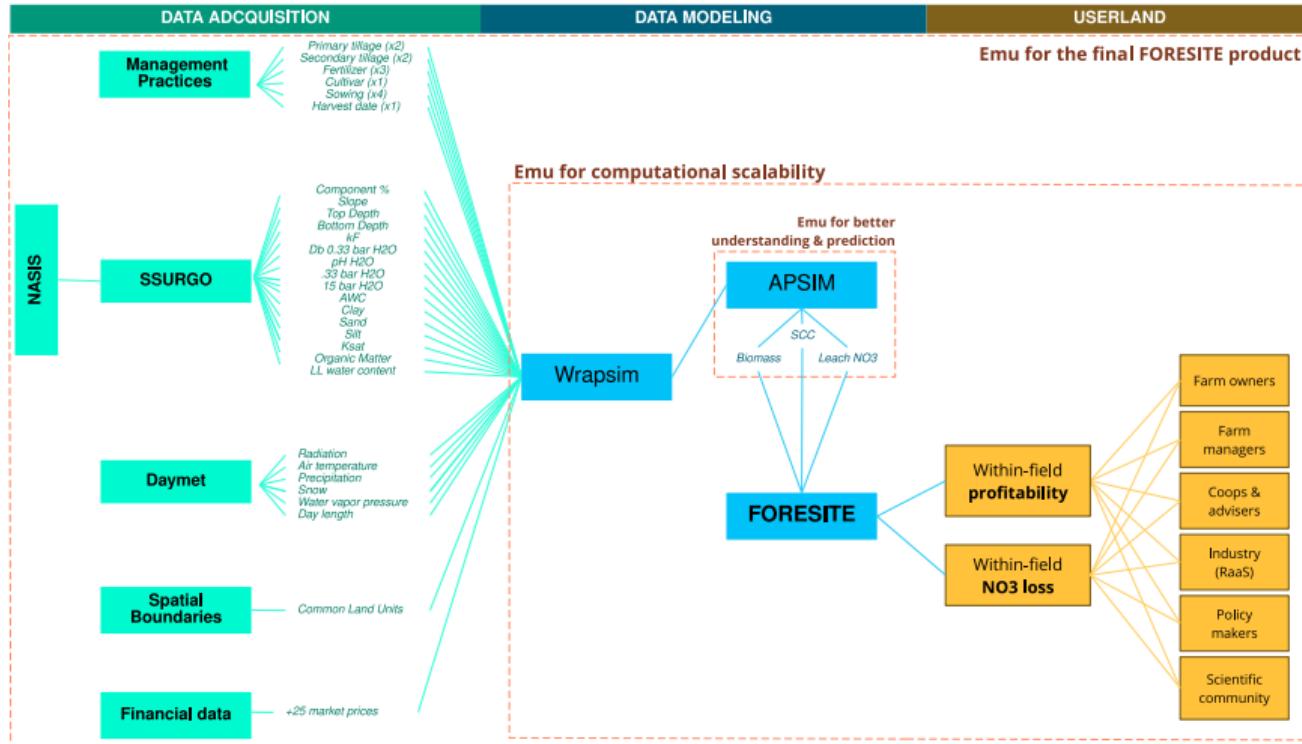


Fig. 1. The agricultural value chain developed through C-CHANGE.

Computer models

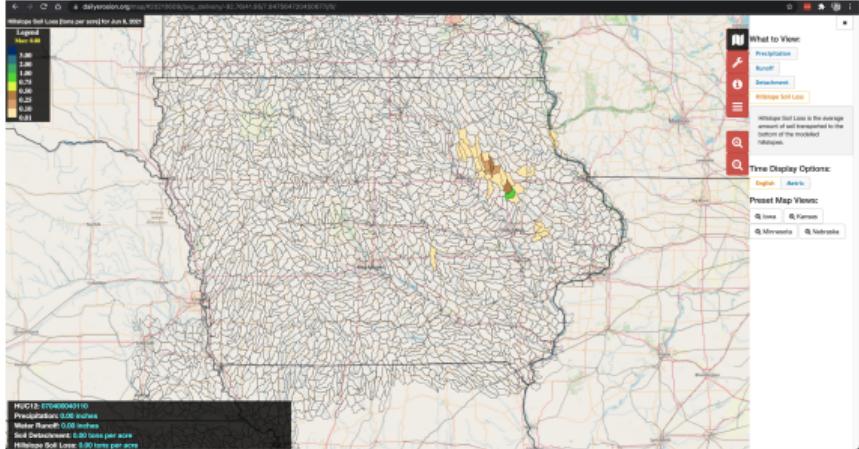
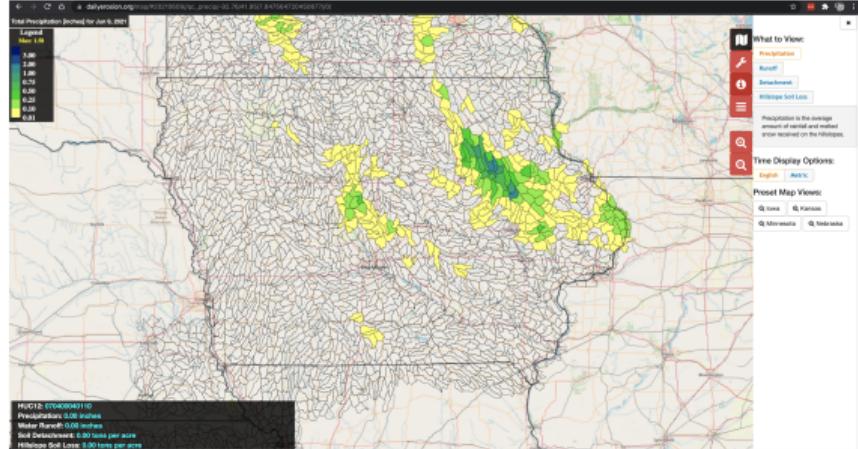
- ▶ Water Erosion Prediction Project (WEPP)
 - ▶ Daily Erosion Project (DEP)
- ▶ Agricultural Production Systems sIMulator (APSIM)
 - ▶ Foresite
- ▶ Cycles
- ▶ Cycles3
- ▶ Iowa Biogas Assessment Model (IBAM)
- ▶ others...

APSIM



Daily Erosion Project (DEP)

using Water Erosion Prediction Project (WEPP)



<https://www.dailyerosion.org/>

Input files

```

4.30
1 1 2006.2
Static comments: soil file
Latitude 2 1
41.53 'Pershing' 'SICL' 4 97.5 # 98.4
Observed 180 19.7 32.5 2.5 #
-2.0 250 20.0 31.0 0.5 # from slope #
Observed 790 7.9 41.5 0.25 #
-11.5 1520 20.0 32.0 0. # 2 # number of OEE's
Observed 0 0.000000 0 2 180.000 1.000 7 # (total) years in simulation
176.0 23 'Rinda' 'SICL' 3 0.2 3 48.939999
Observed 230 20.0 31.0 2.5 0.000000, 0.0 #####
26.2 360 17.3 35.0 0.2 2 48.720005 #####
da mo ye 1520 22.1 50.0 0. 0.000000, 0.0 #####
0 0.000000 0

1 1 2007 0 3.5 0.0 242 0.2 0 4.0 2 # Number of plant scenarios
2 1 2007 0 4.0 -7.2 240 3.6 0 -5.6
3 1 2007 0 11.0 0.0 245 7.2 0 -2.5
4 1 2007 0 13.0 5.6 212 7.6 0 4.0
5 1 2007 0 9.0 0.0 210 4.2 0 3.0
bromegr1
'Bromegrass-High Fertilization Level'
(from WEPP distribution database)
6 1 2007 0 7.2 -3.0 253 3.8 0 -2.5
7 1 2007 0 5.0 -3.0 173 5.5 0 -4.3
8 1 2007 0 8.0 -4.0 211 6.1 0 -5.0
9 1 2007 0 3.0 -8.0 188 6.9 0 -6.4
10 1 2007 0 8.3 -8.0 255 7.0 0 -6.6
11 1 2007 0 11.0 -2.0 165 5.8 0 1.2 # mfo - <non fragile>
12 1 2007 0 -2.8 -11.0 136 7.4 0 -9.0
13 1 2007 4 -5.0 -11.1 107 4.9 0 -10.0
10.53 0.00
12.43 2.02
14.83 4.02
23.98 5.00
ALFALFA
Alfalfa
J. M. Laflein
Set minimum temperature to 0.5 degrees C - def

```

WEPPR

jarad / WEPPR

Code Issues 14 Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags Go to file Add file ▾ Code ▾

Gulzina Kuttubekova and Gulzina Kuttubekova Added write_cli function f26f53e 4 days ago 54 commits

R	Added write_cli function	4 days ago
data-raw	Add data-raw/README.md	17 days ago
data	Fix Roxygen documentation	17 days ago
inst/extdata	Get read_cli() up and running	16 months ago
man	Fix url in calculate_Ke	17 days ago
tests	Add basic type/class tests	17 days ago
.Rbuildignore	Start package	17 months ago
.gitignore	Add functions to calculate Ke and Kb	16 months ago
DESCRIPTION	Add basic type/class tests	17 days ago
LICENSE	Initial commit	17 months ago
NAMESPACE	Fix Roxygen documentation	17 days ago

About

An R package to run WEPP

Readme

GPL-3.0 License

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Contributors 2

Jarad Niemi (ISU)

Emulation of APSIM/WEPP

June 11, 2021

14 / 22

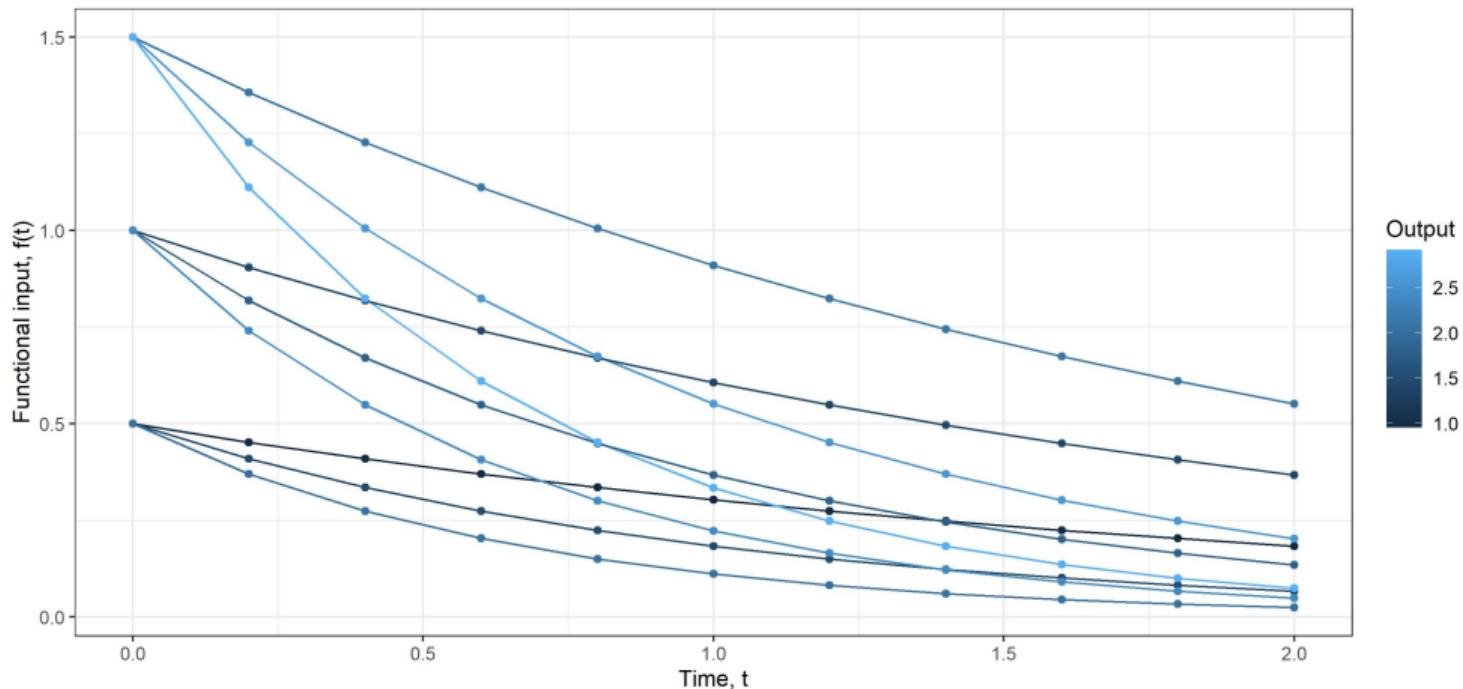
APSIM

Input space

- ▶ 2,160 run-specific input values
 - ▶ Soil properties 22 functionals \times 16 layers, 2 scalars.
 - ▶ Climate dynamics 4 functionals \times 365 daily values, 2 scalars.
 - ▶ Land management practices, mostly categorical scalars.
- ▶ The input space is...
 - ▶ High dimensional: computational and modeling challenging.
 - ▶ Structured: complex to capture hierarchy.
 - ▶ Vast: large number of runs to explore it.

Functional input - scalar output

Example 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

$$\text{Cov}(Y_i, Y_j) = \sigma^2 k(X_i, X_j)$$

$$k(X_i, X_j) = e^{-\frac{1}{2}D(X_i, X_j, \omega)}$$

and

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

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

Vector-input Gaussian Process using Principal Component Analysis

For observation i , we have $Y_i \in \mathbb{R}$ and $X_i = (X_{i,1}, \dots, X_{i,D})$ with PCA-rotated inputs $\tilde{X}_i = (\tilde{X}_{i,1}, \dots, \tilde{X}_{i,D})$. Our computer model is $f()$ with $Y_i = f(\tilde{X}_i)$.

Assume f is a zero-mean Gaussian process with

$$\text{Cov}(Y_i, Y_j) = \sigma^2 k(\tilde{X}_i, \tilde{X}_j)$$

$$k(\tilde{X}_i, \tilde{X}_j) = e^{-\frac{1}{2}D(\tilde{X}_i, \tilde{X}_j, \omega)}$$

and

$$D(\tilde{X}_i, \tilde{X}_j, \omega) = \sum_{d=1}^D \omega_d (\tilde{X}_{i,d} - \tilde{X}_{j,d})^2.$$

We'll refer to this as **viGP (PCA)**.

Functional-input Gaussian Process

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

The functional-input Gaussian Process has

$$k(X_i, X_j) = e^{-\frac{1}{2}D(X_i, X_j, \omega)}$$

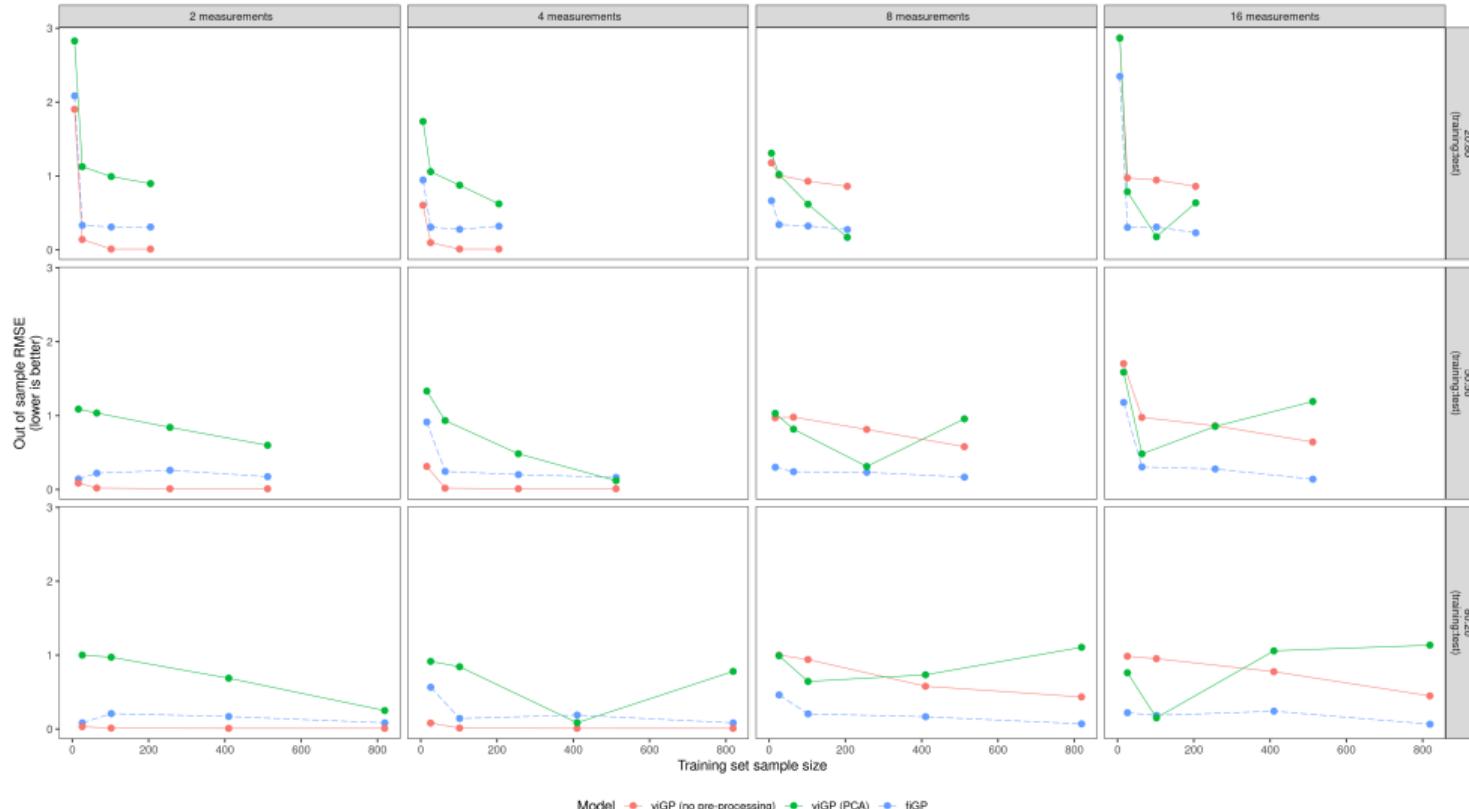
$$\begin{aligned} D(X_i, X_j, \omega) &= \int_0^T \omega(t)(X_i(t) - X_j(t))^2 dt \\ &\approx \sum_{d=1}^D \omega(t_d)(X_i(t_d) - X_j(t_d))^2. \end{aligned}$$

For the (inverse) **functional length-scale**, we assume

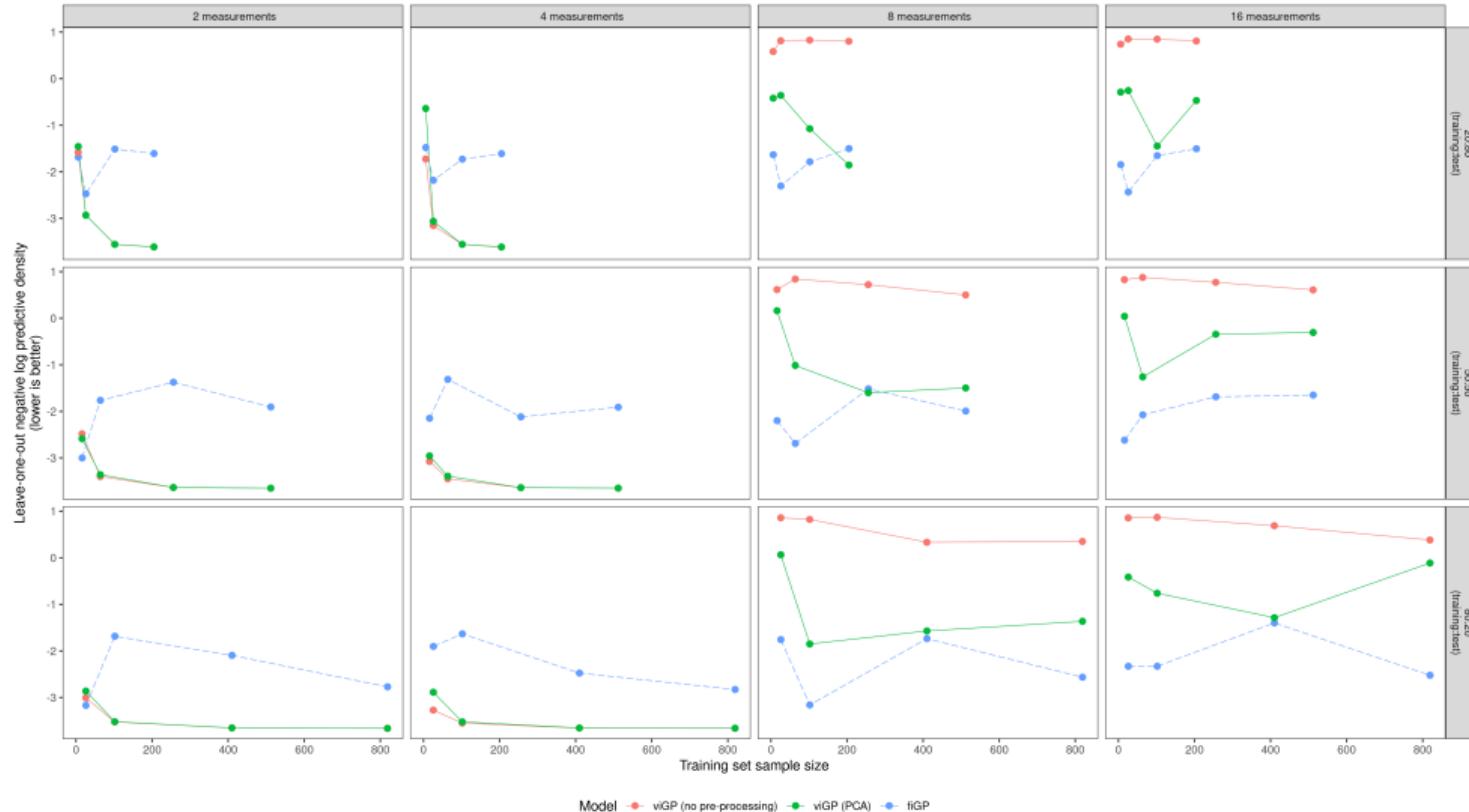
$$\omega(t) = \exp(2\sigma_\ell^2 t^\eta).$$

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

Out-of-sample RMSE



Leave-one-out – $\log p(\tilde{y})$



Summary

Functional length-scales

- ▶ Additional simulations
- ▶ Expanding to high dimensional functional inputs
- ▶ More flexible functional length scales
- ▶ Combining with additional inputs

These slides are available at

- ▶ <https://github.com/jarad/SSC2021>
- ▶ <http://www.jarad.me/research/presentations.html>

Thank you!

Other links:

- ▶ <http://www.jarad.me/>
- ▶ <https://luisdamiano.github.io/>