Computer Model Emulation in Agriculture

Jarad Niemi

Iowa State University

June 16, 2022

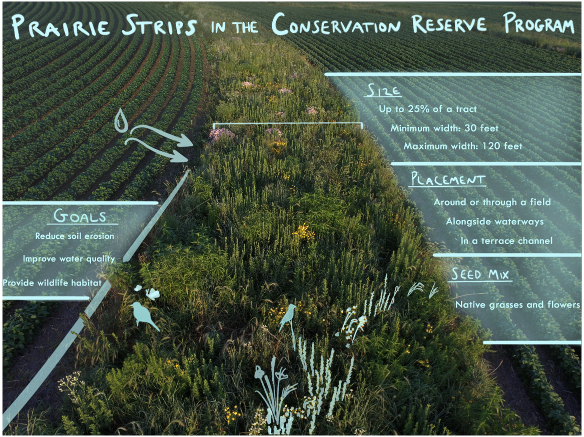
Funded, in part, by

- 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

C-CHANGE: Science for a changing agriculture

C-CHANGE

http://agchange.org



http://prairiestrips.org

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/

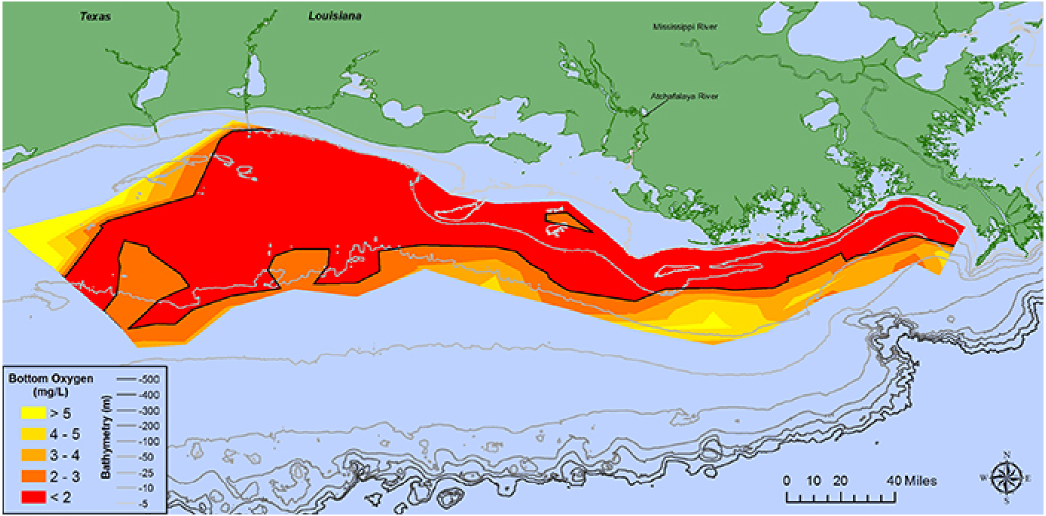
https://www.scientificamerican.com/article/meat-and-environment/

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

NIFA PRESS RELEASE USDA

NATIONAL INSTITUTE OF FOOD AND AGRICULTURE

USDA Scientific Research Program Funds Sustainable Agricultural Systems Projects

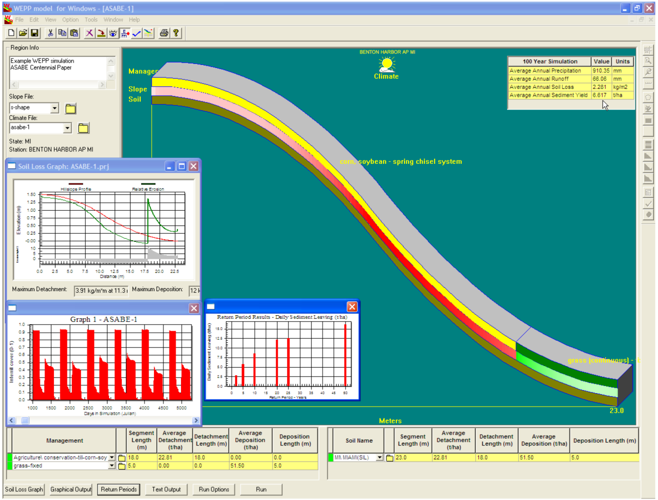
Feedstock Production | Anaerobic Digestion (AD) | Product Processing | Distribution and Markets | Society

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

Computer models

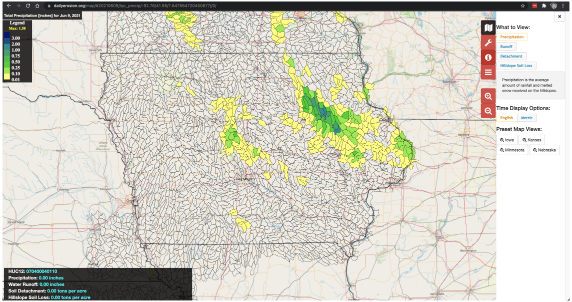
- Water Erosion Prediction Project (WEPP)
  - Daily Erosion Project (DEP)
- Agricultural Production Systems sIMulator (APSIM)
  - Foresite
- Cycles
- Iowa Biogas Assessment Model (IBAM)
- others...

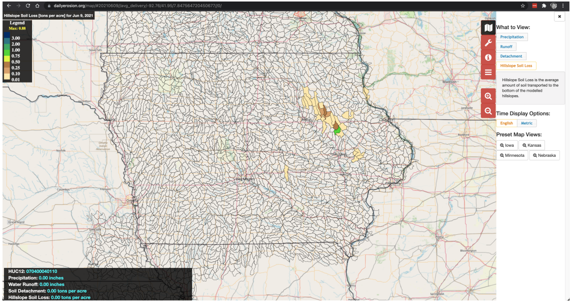
WEPP


WEPP model for Windows - [ASABE-1]
File Edit View Option Tools Window Help
Region Info
Example WEPP simulation
ASABE Centennial Paper
Slope File:
s-shape
Climate File:
asabe-1
State: MI
Station: BENTON HARBOR AP MI
Soil Loss Graph: ASABE-1.prj
Hillslope Profile
Relative Erosion
Elevation (m)
Distance (m)
Maximum Detachment:
3.91 kg/m^m at 11.3
Maximum Deposition:
Graph 1 - ASABE-1
Interrill cover (0-1)
Days in Simulation (Julian)
Return Period Results - Daily Sediment Leaving (t/ha)
Return Period - Years
BENTON HARBOR AP MI
Climate
Slope
Soil
corn, soybean - spring chisel system
100 Year Simulation
Value
Units
Average Annual Precipitation 910.35 mm
Average Annual Runoff 66.06 mm
Average Annual Soil Loss 2.281 kg/m2
Average Annual Sediment Yield 6.617 t/ha
23.0
Meters
Management
Segment Length (m)
Average Detachment (t/ha)
Detachment Length (m)
Average Deposition (t/ha)
Deposition Length (m)
Agriculture\ conservation-till-corn-soy
grass-fixed
Soil Name
Soil Loss Graph
Graphical Output
Return Periods
Text Output
Run Options
Run

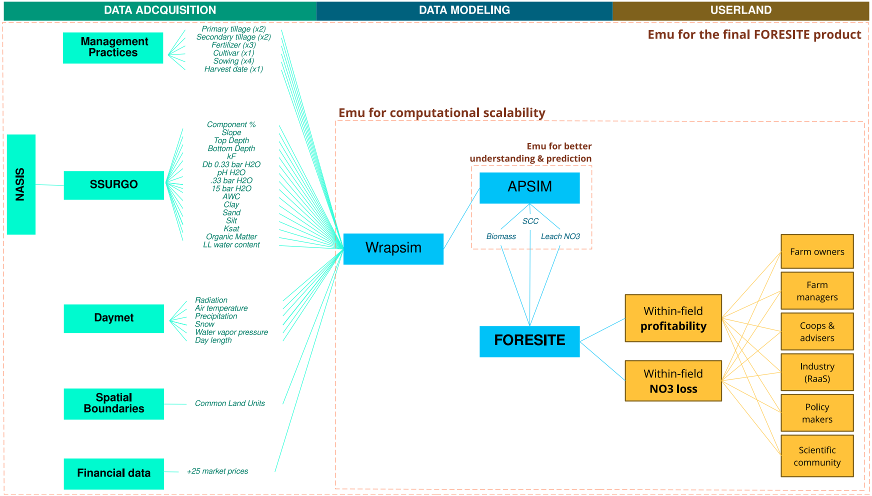
Daily Erosion Project (DEP)

using Water Erosion Prediction Project (WEPP)





APSIM


DATA ADCQUISITION
DATA MODELING
USERLAND
Emu for the final FORESITE product
NASIS
Management Practices
Primary tillage (x2)
Secondary tillage (x2)
Fertilizer (x3)
Cultivar (x1)
Sowing (x4)
Harvest date (x1)
SSURGO
Component %
Slope
Top Depth
Bottom Depth
kF
Db 0.33 bar H2O
pH H2O
.33 bar H2O
15 bar H2O
AWC
Clay
Sand
Silt
Ksat
Organic Matter
LL water content
Daymet
Radiation
Air temperature
Precipitation
Snow
Water vapor pressure
Day length
Spatial Boundaries
Common Land Units
Financial data
+25 market prices
Emu for computational scalability
Wrapsim
Emu for better understanding & prediction
APSIM
SCC
Biomass
Leach NO3
FORESITE
Within-field profitability
Within-field NO3 loss
Farm owners
Farm managers
Coops & advisers
Industry (RaaS)
Policy makers
Scientific community

APSIM

Input space

- 2,160 run-specific input values
  - Soil properties 22 functionals x 16 layers, 2 scalars.
  - Climate dynamics 4 functionals x 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.

Gaussian Process Emulators

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

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

where

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

We assume a Gaussian Process prior

$$f \sim \mathcal{GP}\left(m, k\right) \implies y \sim N(m_x, \Sigma_{x,x})$$

where $\Sigma_{x,x}(x, x') = k(x, x')$.

Two research directions:

- Computational tractability for large N
- Dealing with functional X_i

Training a GP

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

$$\hat{\theta} = \text{argmax}_\theta \, p(y|\theta) = \text{argmax}_\theta \, N\left(y; m_x, \tau^2 \mathrm{I} + \Sigma(\theta)\right)$$

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

$$\begin{aligned} \log \mathcal{N}(y; m_x, \tau^2 \mathrm{I} + \Sigma_\theta) = \quad & C \\ & -\tfrac{1}{2} \log |\tau^2 \mathrm{I} + \Sigma(\theta)| \\ & -\tfrac{1}{2} y^\top [\tau^2 \mathrm{I} + \Sigma(\theta)]^{-1} y \end{aligned}$$

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.

Fully Independent Conditional (FIC) Approximation

Introduce a set of knots $x^\dagger = \left\{x_1^\dagger, \ldots, x_K^\dagger\right\}$ with $K << N$, such that

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

where

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

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

This joint implies the following marginal distribution for f_x:

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

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

Train FIC Model

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

$$Y|x^\dagger, \theta \sim \mathcal{N}(m_x, \tau^2 \mathrm{I} + \Psi_{xx}).$$

Train the model by finding

$$\hat{x}^\dagger, \hat{\theta} = \text{argmax}_{x^\dagger,\theta} \mathcal{N}(y; m_x, \tau^2 I + \Psi_{xx}).$$

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

Appealing due to similarity with Full GP MLE approach and massively reduced computational complexity, but 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

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

Vector-input Gaussian Process (viGP)

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

Assume f is a zero-mean Gaussian process with

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

and

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

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

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

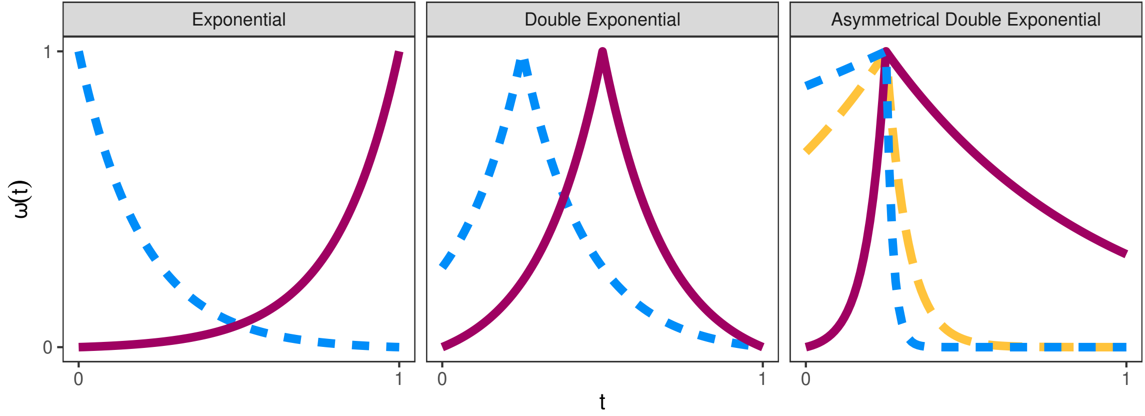
$$D(X_i, X_j, \omega) = \int_0^T \omega(t)(X_i(t) - X_j(t))^2 dt \approx \sum_{d=1}^{D} \omega(t_d)(X_i(t_d) - X_j(t_d))^2.$$

For the functional length-scale, we assume

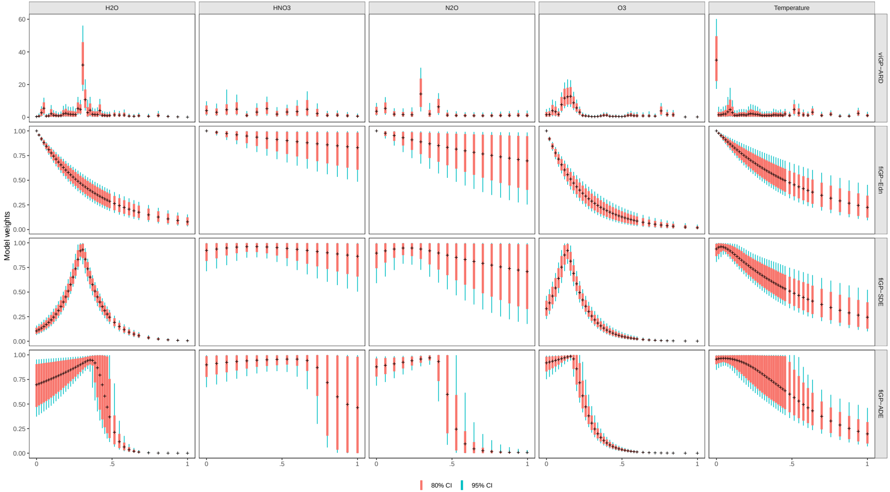
$$\omega(t) = \exp\left(2\sigma_\ell^2 t^\eta\right).$$

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/NCCC170_2022
- http://www.jarad.me/research/presentations.html

Thank you!

Other links:

- http://www.jarad.me/
- https://www.youtube.com/c/jaradniemi