Iowa State University
Department of Statistics

Conference on Applied Statistics in Agriculture and Natural Resources
17-20 May 2021

Luis Damiano*, Jarad Niemi

Automatic Dynamic Relevance Determination of soil properties over different soil layers for yield prediction using APSIM

APSIM[1] (deterministic mechanistic model)

AGRICULTURAL PRODUCTION SYSTEMS SIMULATOR

Soil | Weather | Land Mgmt → Yield

Emulator (AKA surrogate, or meta-model)

A statistical model that provides **a mechanistic model output estimate and an associated uncertainty measure** without running the computer model.

Why?[2-3]

- Scalability: large-scale yield prediction.
- Better understanding: model-based information to guide data acquisition, validation, and modeling efforts.
- Better prediction: adjust APSIM prediction at a sub-field level.
- Optimization: yield & NO3 leach trade-off.
- Portability: run it everywhere (e.g., online platform).

Gaussian Process[4]

Let $y_i \in \mathcal{R}$ be the APSIM output biomass at harvest date, $\mathbf{x_i} = \{x_{i,k} : k = 1, ..., K\}$ the corresponding root water extraction constant values across K soil layers indexed by depth $t_k \in [0, 1]$. Consider a **Gaussian process with mean zero and pairwise correlation**:

$$\rho(\mathbf{x_i}, \mathbf{x_j}) = \exp\left\{-\frac{1}{2}(\mathbf{x_i} - \mathbf{x_j})^\top \mathbf{L}(\mathbf{x_i} - \mathbf{x_j})\right\} \forall\, i, j \in \{1, \ldots, N\}$$

$$\mathbf{L}^{-1} = \operatorname{diag}\left(\{l_k^2 : k = 1, \ldots, K \text{ and } l_k > 0\}\right)$$

The role of lengthscale parameter

Smaller lengthscale ~ slower decay of the output correlation as a function of the input Euclidean distance **~ higher predictive relevance**

$$\omega_x(t_k) = \ell_x(t_k)^{-2}$$

$$\tilde{x}_{i,t_k} = \sqrt{\omega_x(t_k)}\; x_{i,t_k}$$

Vector-input lengthscale

It treats the input profile as an unstructured vector of measurements.

$$\{l_1, \ldots, l_K\}$$

Functional-input lengthscale[5-6]

It recognizes the intrinsic structure of the functional input and **captures how fast input predictive relevance transitions to a neutral state**.

$$l_x : \mathcal{R}^+ \to (0, 1]$$

$$l_x(t) = \exp\left\{-\sigma_{l_x}^2\, t^{\eta_{l_x}}\right\} \text{ where } \sigma_{l_x}^2 > 0,\ \eta_l \in (0, 2], \text{ and } t > 0.$$

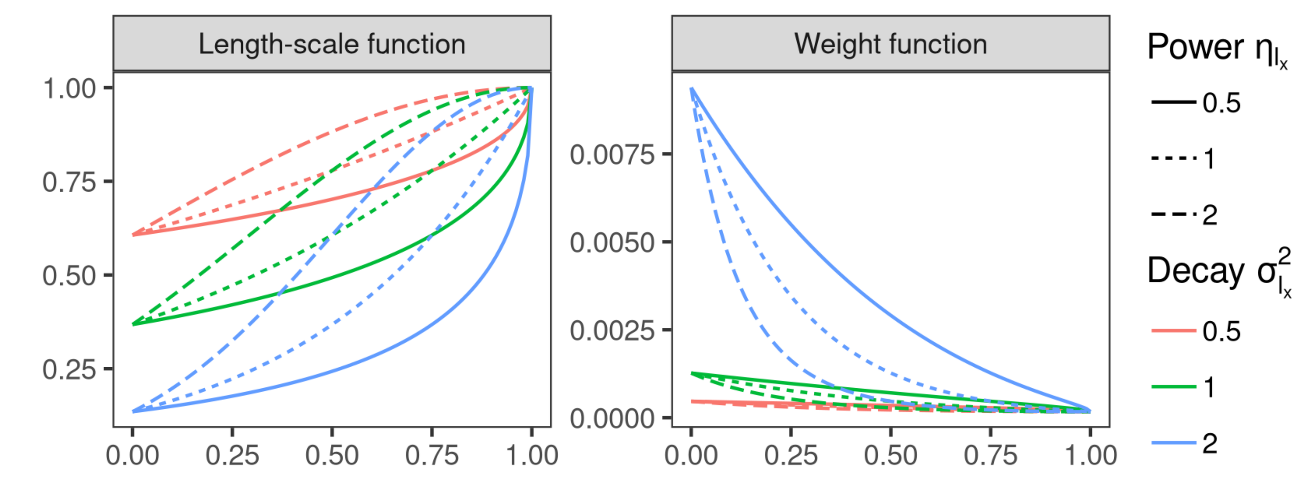

Figure 1: Functional input length-scale and weight functions.

Application

APSIM v7.10 with a soil, weather, and land management configuration representative of a Greene county, Iowa productive field on three years of continuous corn.

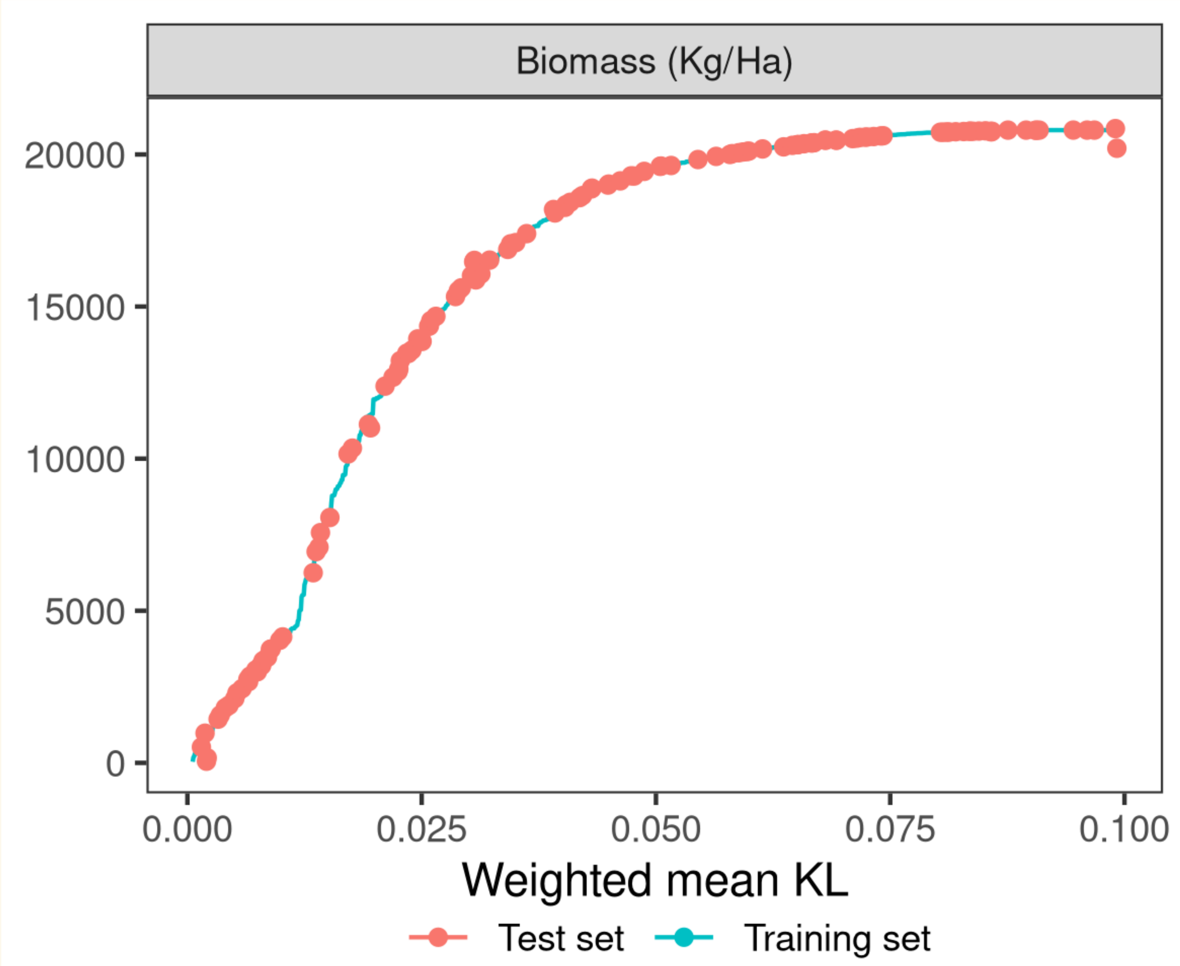

Figure 2: Biomass ~ Weighted mean KL relationship.

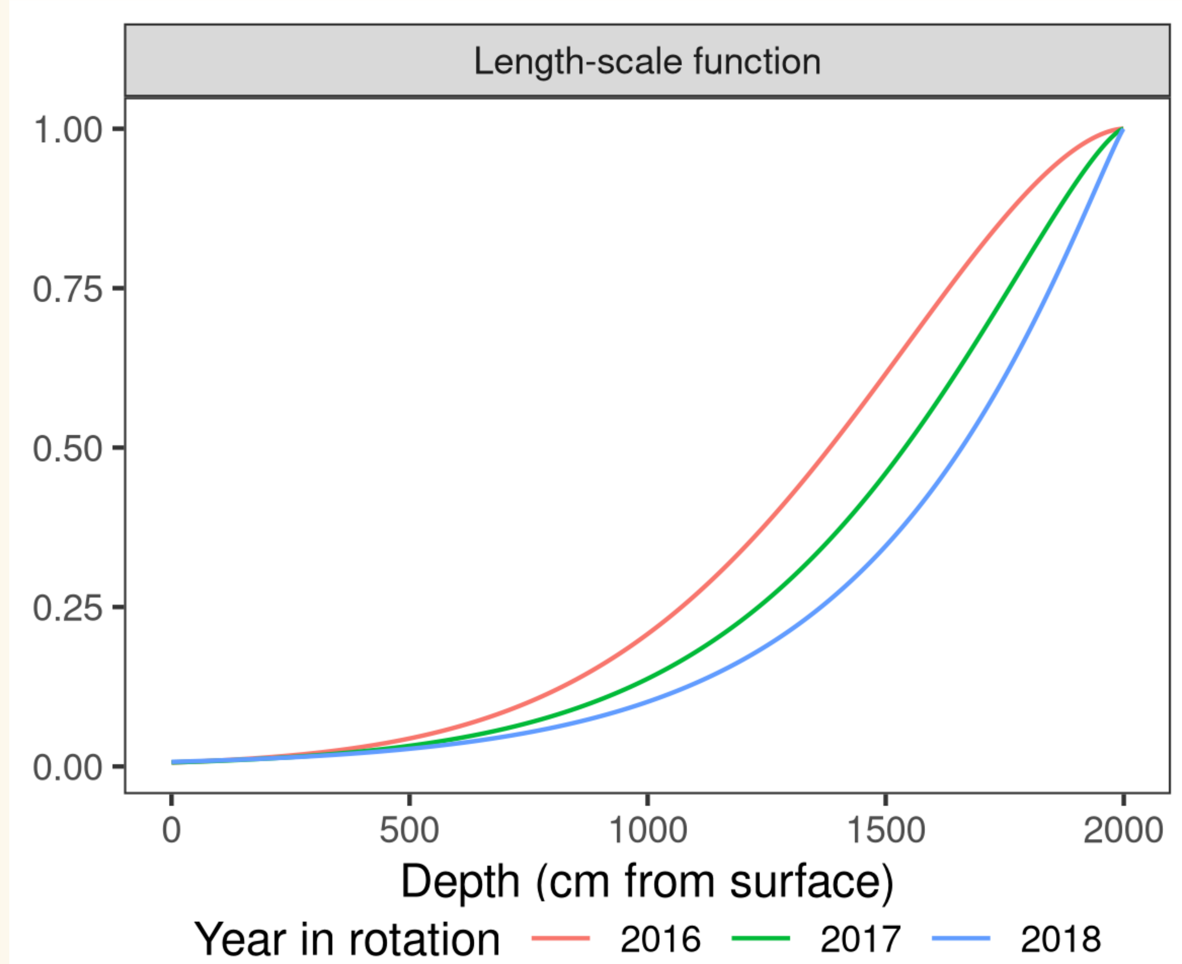

Figure 3: Length-scale function evaluated at the parameter estimates.

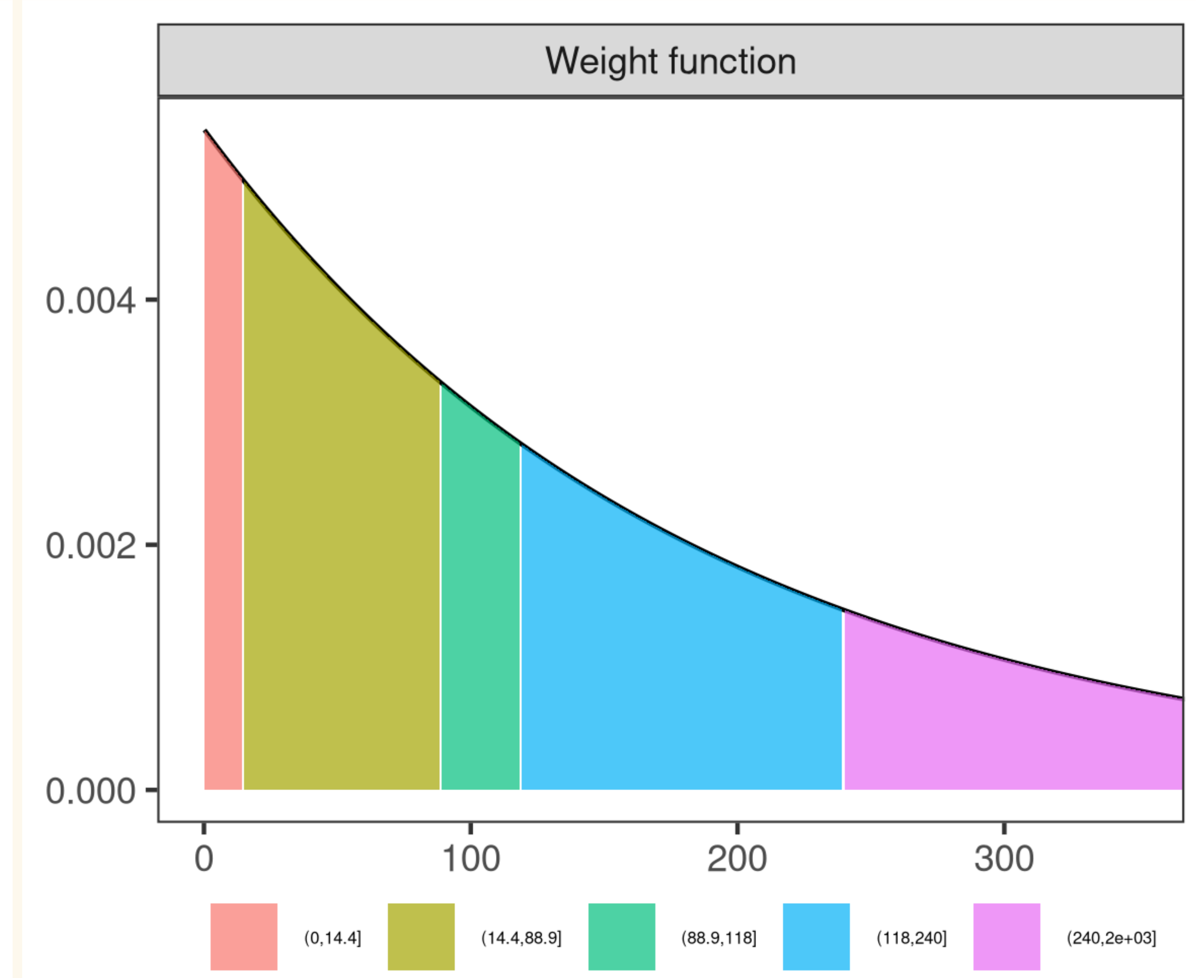

Figure 4: Weight function evaluated at the parameter estimates.

	2016	2017	2018
Vector Input GP	285.5	319.3	284.3
Functional Input GP	143.7	131.7	97.2
RMSE Reduction	49.7%	58.8%	65.8%

Table 1: Out of sample predictive RMSE (10-fold cross validation).

Depth interval (cm)	Cumulative root biomass[7]	Cumulative model weight
< 14.4	50%	7.8%
< 88.9	95%	38.1%
< 118.3	100%[a]	47.3%
< 240.0	100%[b]	72.4%
< 880.0	-	99.0%

[a] Value estimated by [7] [b] Value reviewed by [7]
Table 2: Cumulative root biomass and model weight.

[1] Holzworth, Dean P., Neil I. Huth, Peter G. deVoil, Eric J. Zurcher, Neville I. Herrmann, Greg McLean, Karine Chenu, et al. "APSIM – Evolution towards a New Generation of Agricultural Systems Simulation." Environmental Modelling & Software 62 (December 2014): 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009.
[2] Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and Analysis of Computer Experiments. Statistical Science, 4(4), 409–423. https://doi.org/10/fn276g
[3] Gramacy, R. B. (2020). Surrogates: Gaussian process modeling, design and optimization for the applied sciences. Chapman Hall/CRC. https://doi.org/10.1201/9780367815493
[4] Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001
[5] Morris, M. D. (2012). Gaussian Surrogates for Computer Models With Time-Varying Inputs and Outputs. Technometrics, 54(1), 42–50. https://doi.org/10.1080/00401706.2012.648870
[6] Kuttubekova, G. (2019). Emulator for Water Erosion Prediction Project computer model using Gaussian Processes with functional inputs. Creative Components.
[7] Fan, J., McConkey, B., Wang, H., & Janzen, H. (2016). Root distribution by depth for temperate agricultural crops. Field Crops Research, 189, 68-74. https://doi.org/10.1016/j.fcr.2016.02.013

Acknowledgements: Funding was provided by Iowa State University through the Presidential Interdisciplinary Research Initiative on C-CHANGE: Science for a Changing Agriculture and the Foundation for Food and Agriculture Research (FFAR).

* Correspondence to ldamiano@iastate.edu