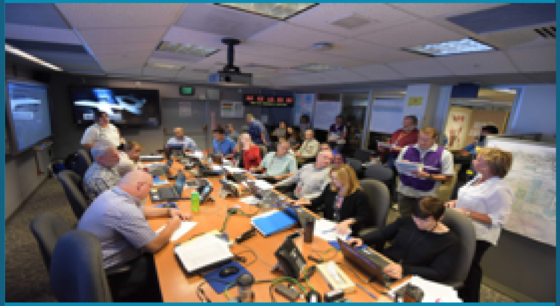
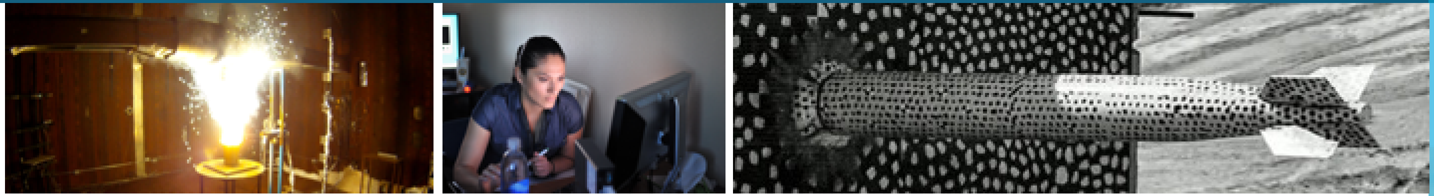
SAND2021-2479C

Low-Rank Tensor Network Approximations for Earth System Model





Cosmin Safta[1], Alex Gorodetsky[2], John Jakeman[1], Khachik Sargsyan[1]
Daniel Ricciuto[3]

[1]Sandia National Laboratories
[2]University of Michigan
[3]Oak Ridge National Laboratory

SIAM Computational Science and Engineering
Virtual Conference March 1-5, 2021

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Acknowledgements

- work supported by the U.S. Department of Energy, Office of Science
  - Advanced Scientific Computing Research (ASCR)
  - Scientific Discovery through Advanced Computing (SciDAC) program through the FASTMath Institute
  - Biological and Environmental Research (BER)
  - National Energy Research Scientific Computing Center (NERSC)

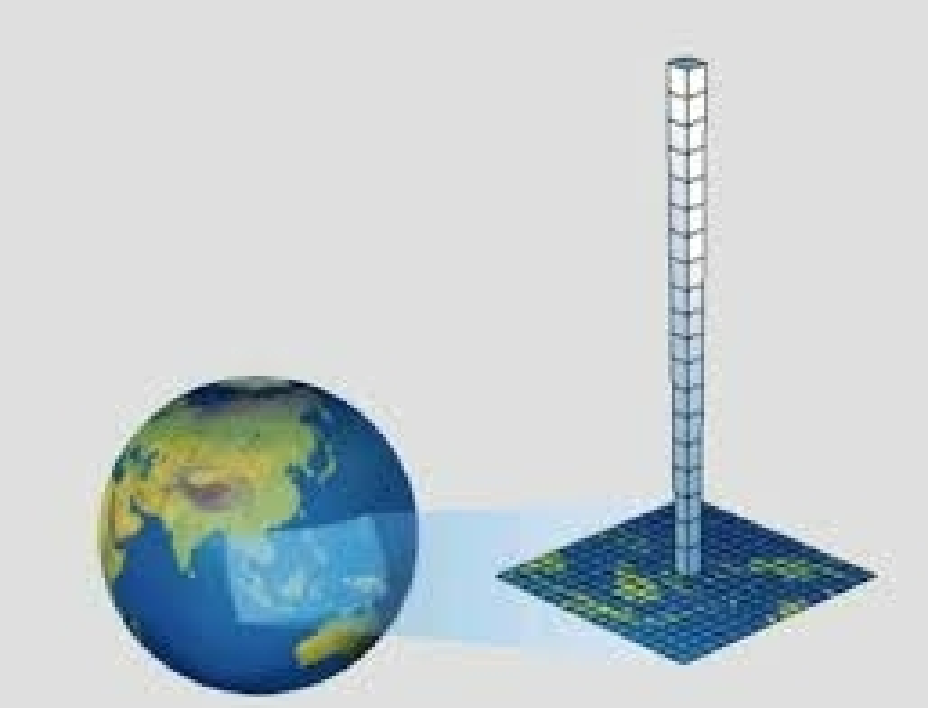
Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

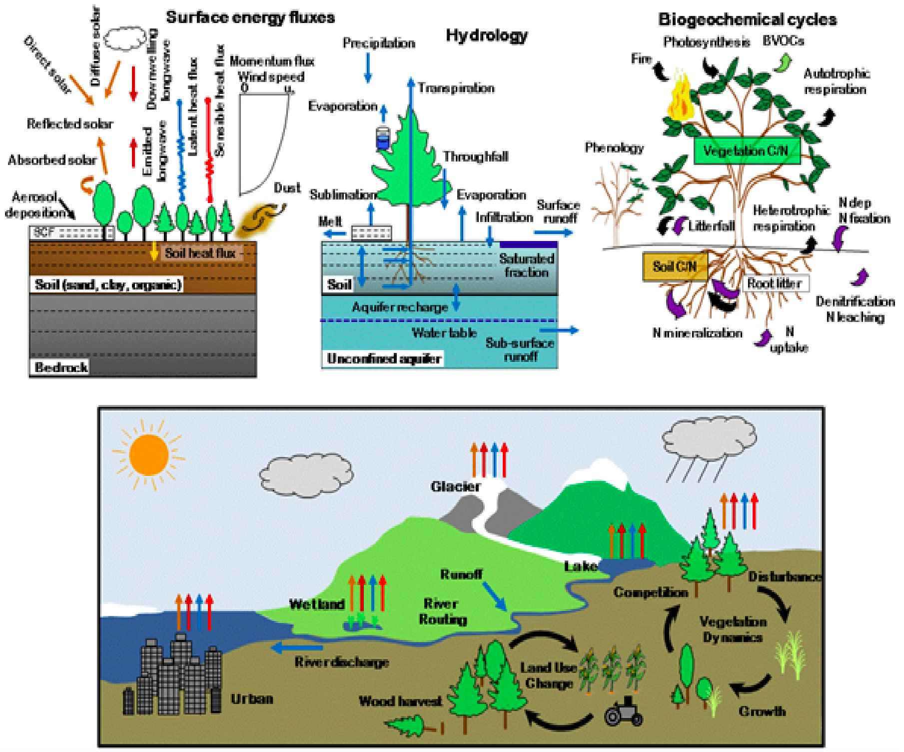
Outline

- Application Driver: Energy Exascale Earth System Model (E3SM) – Land Component
  - Motivation for surrogate model construction
- Low-rank Functional Tensor Network Models
  - Background
  - Efficient function and gradient evaluations
- Results: global sensitivity analysis

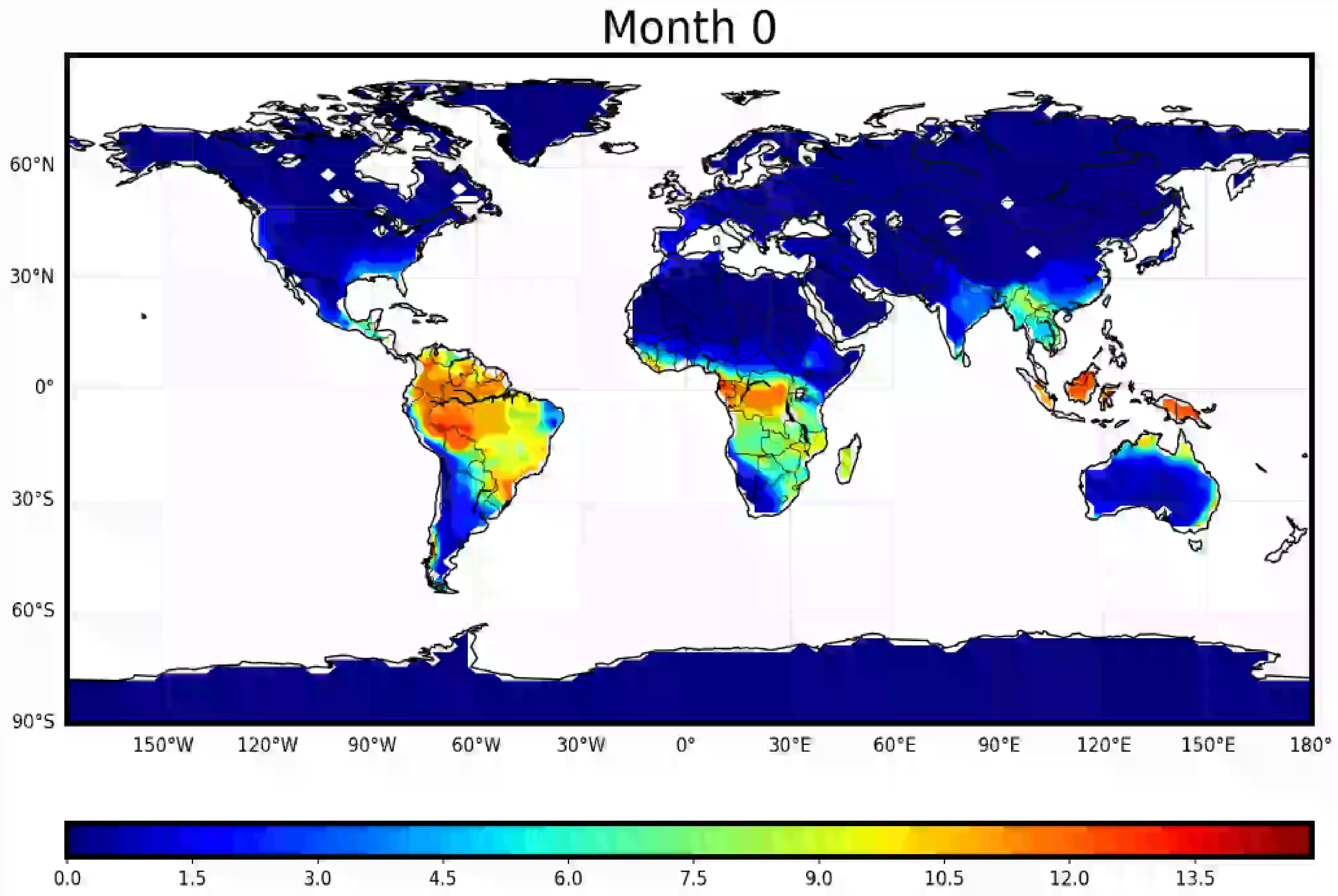
Energy Exascale Earth System Model (E3SM) – Land Component

- The Land Model (ELM) Component of the Energy Exascale Earth System Model (E3SM) is increasingly complex with many processes
    - Large ensembles are needed for uncertainty quantification… but computationally infeasible
    - Focus on surrogate models based on small ensembles to increase the efficiency of sensitivity analysis and model calibration studies





Quantity of Interest: Gross Primary Production (GPP) [gC m^{-2} day^{-1}]



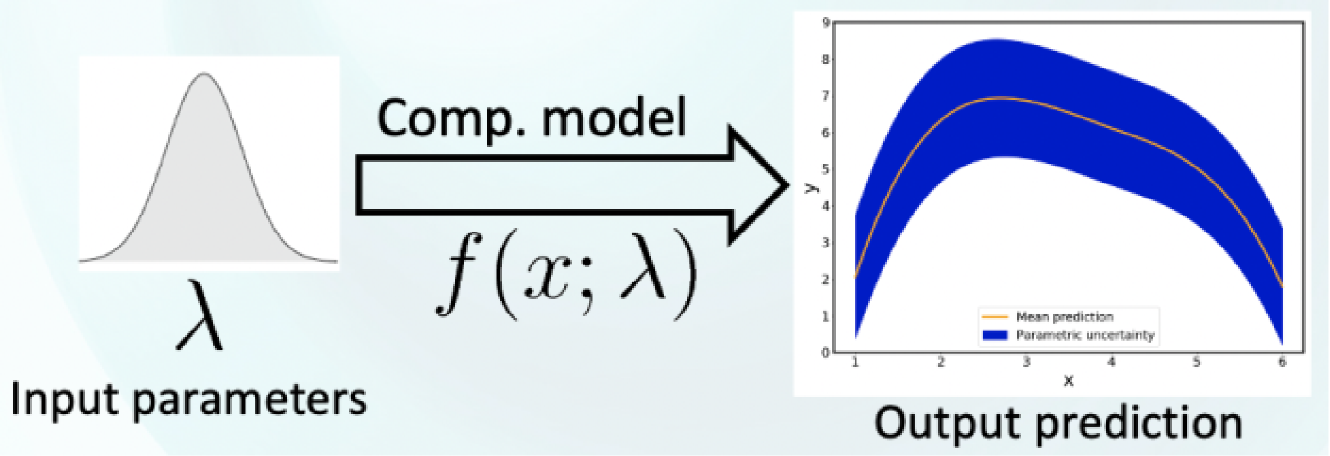
GPP (gC m^{-2} day^{-1})

UQ Challenges in E3SM

- What processes drive uncertainty ?
- What accounts for differences among models ?
- Can we improve predictive capabilities though calibration using available observations ?

<u>E3SM Land Model (ELM) Produces Time Series given Input Parameters and Forcing Drivers</u>

- O(10)-O(100) uncertain inputs
- Daily forcings/drivers
  - Min/max temperatures
  - Solar radiation
  - Water availability



Cheaper **Surrogates** are Necessary to Replace Expensive Computational Models for UQ Assessments

- functional approximations
- non-parametric models, e.g. Gaussian processes
- neural networks and other supervised learning techniques

Requirements:

- expressivity with a limited number of parameters
- cheap – analyses often requiring $O(10^6)$ evaluations with limited computational resources

Functional Approximations

Tensor-product basis approximations:

$$f(\boldsymbol{\lambda}) = \sum_{i_1}^{N_1} \sum_{i_2}^{N_2} \dots \sum_{i_d}^{N_d} \phi_1^{i_1}(\lambda_1; \boldsymbol{\theta}) \phi_2^{i_2}(\lambda_2; \boldsymbol{\theta}) \dots \phi_d^{i_d}(\lambda_d; \boldsymbol{\theta})$$

- use orthogonal polynomials, radial basis functions, …
- the curse of dimensionality $O(N^d)$ typically limits the polynomial order/no. of functions
  - this places limits on the surrogate model capacity to adapt to non-linear behavior

Functional Tensor-Train Models

- Analogous to tensor-train models [Oseledets, 2013]: approximate multivariate functions instead of multidimensional arrays

$$
\begin{aligned}
f(\boldsymbol{\lambda}) &= \sum_{i_0=1}^{r_0} \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} f_1^{(i_0 i_1)}(\lambda_1; \boldsymbol{\theta}_1) f_2^{(i_1 i_2)}(\lambda_2; \boldsymbol{\theta}_2) \cdots f_d^{(i_{d-1} i_d)}(\lambda_d; \boldsymbol{\theta}_d) \\
&= \mathcal{F}_1(\lambda_1; \boldsymbol{\theta}_1) \mathcal{F}_2(\lambda_2; \boldsymbol{\theta}_2) \cdots \mathcal{F}_d(\lambda_d; \boldsymbol{\theta}_d)
\end{aligned}
$$

$$
\mathcal{F}_k(\lambda_k; \boldsymbol{\theta}_k) = \begin{bmatrix}
f_k^{(11)}(\lambda_k; \boldsymbol{\theta}_k^{(11)}) & f_k^{(12)}(\lambda_k; \boldsymbol{\theta}_k^{(12)}) & \cdots & f_k^{(1 r_k)}(\lambda_k; \boldsymbol{\theta}_k^{(1 r_k)}) \\
f_k^{(21)}(\lambda_k; \boldsymbol{\theta}_k^{(21)}) & f_k^{(22)}(\lambda_k; \boldsymbol{\theta}_k^{(22)}) & \cdots & f_k^{(2 r_k)}(\lambda_k; \boldsymbol{\theta}_k^{(2 r_k)}) \\
\vdots & \vdots & \ddots & \vdots \\
f_k^{(r_{i-1} 1)}(\lambda_k; \boldsymbol{\theta}_k^{(r_{k-1} 1)}) & f_k^{(r_{i-1} 2)}(\lambda_k; \boldsymbol{\theta}_k^{(r_{k-1} 2)}) & \cdots & f_k^{(r_{i-1} r_k)}(\lambda_k; \boldsymbol{\theta}_k^{(r_{k-1} r_k)})
\end{bmatrix}
$$

- Model evaluation/gradient computation consists of a sequence of matrix-vector multiplications

[Gorodetsky & Jakeman, 2020]

Functional Representations – Univariate Functions

- Linear Representations

$$f_k^{(ij)}(\lambda_k(\xi_k);\boldsymbol{\theta}_k^{(ij)}) = \sum_{l=0}^{p_k} \theta_{k,l}^{(ij)} \Psi_l^{(ijk)}(\xi_k)$$

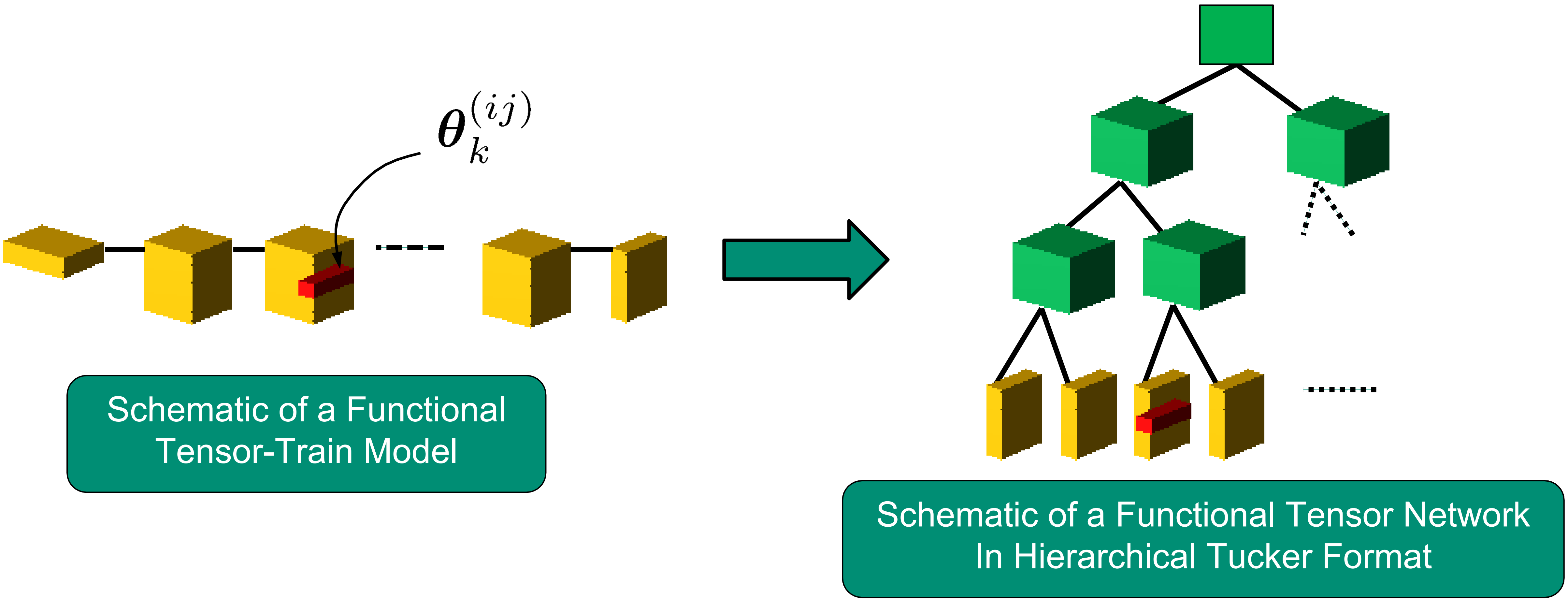
(e.g. Polynomial Chaos Expansions)

- Non-Linear Representations

$$f_k^{(ij)}(\lambda_k;\boldsymbol{\theta}_k^{(ij)}) = \sum_{l=0}^{p_k} \theta_{k,l,1}^{(ij)} \exp(-\theta_{k,l,2}^{(ij)}(\lambda_k - \theta_{k,l,3}^{(ij)})^2)$$

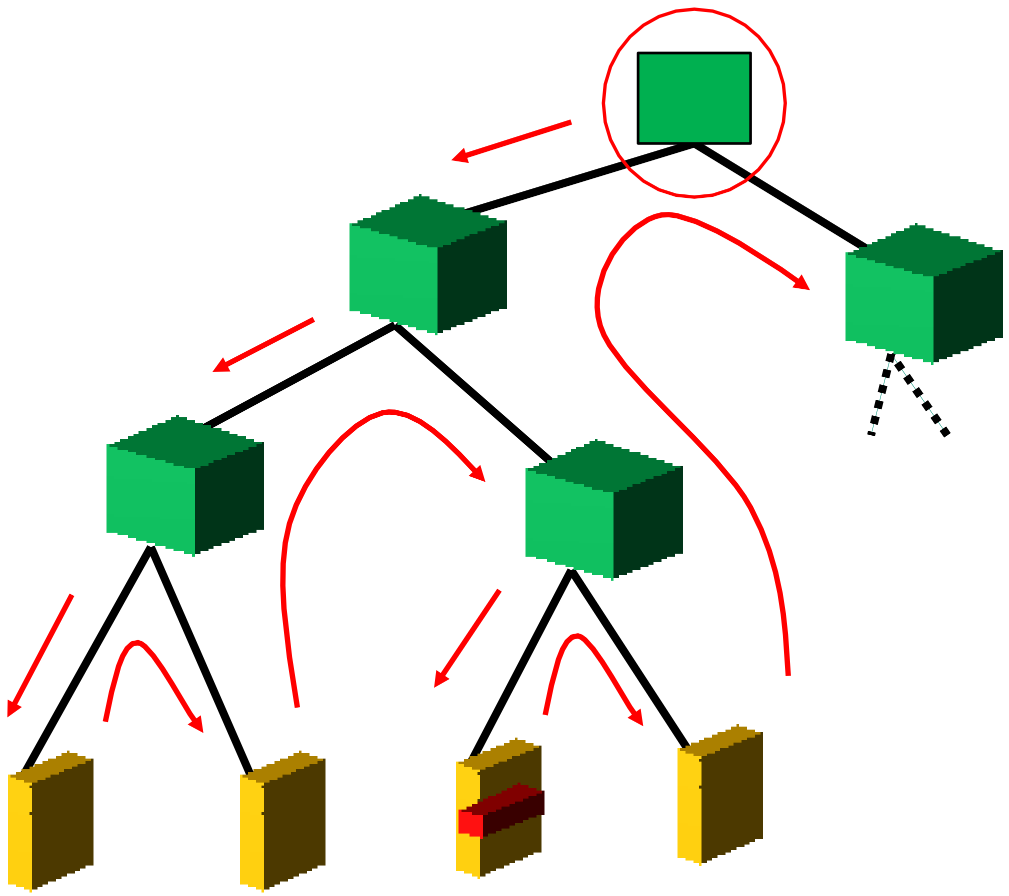
(Radial Basis Functions)

Functional Tensor Network Models



- **Black connectors** represent contractions between adjacent tensors
- **Red vectors** represent coefficients for the corresponding univariate functions

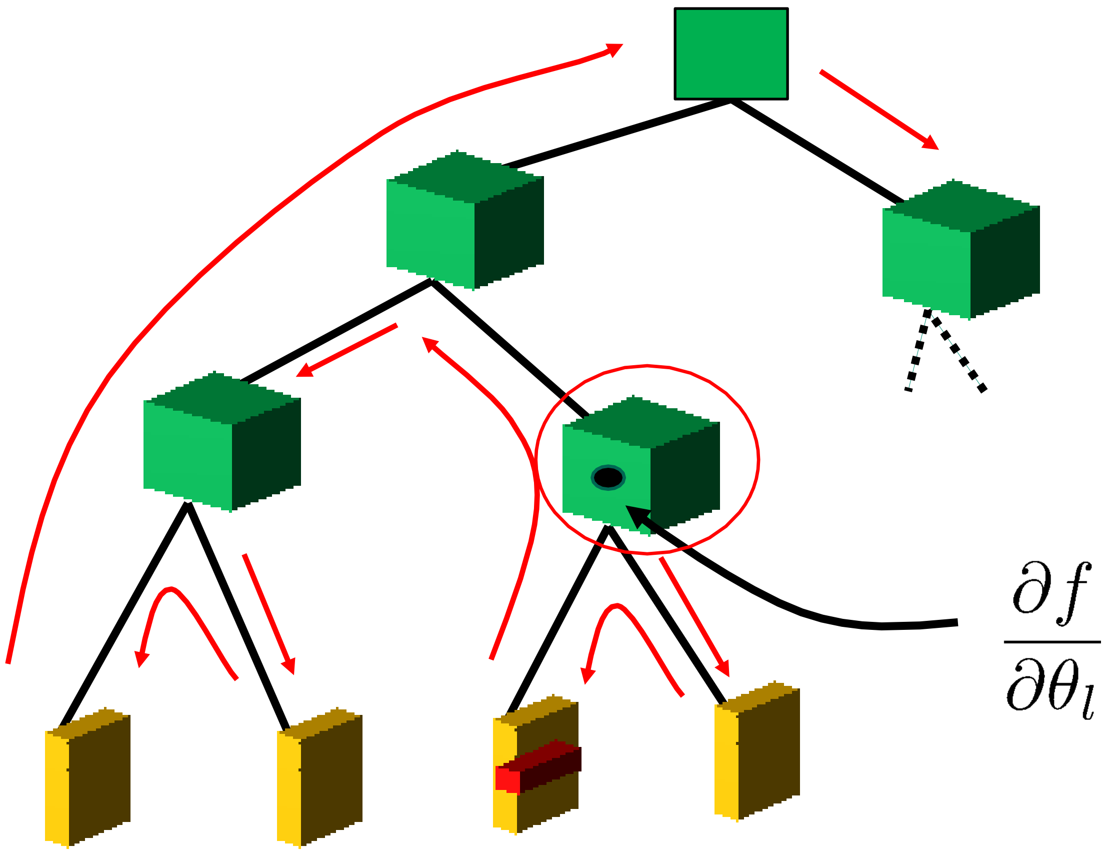
Functional Tensor Network Models – Recursive Contractions



(red arrows represent search paths)

1. Evaluate all open nodes for the set of training parameters
2. Depth first search (DFS) starting from one of the nodes to recursively contract tensors along graph edges
3. Store intermediate results, to be re-used for gradient evaluations

Functional Tensor Network Models – Gradient Evaluations

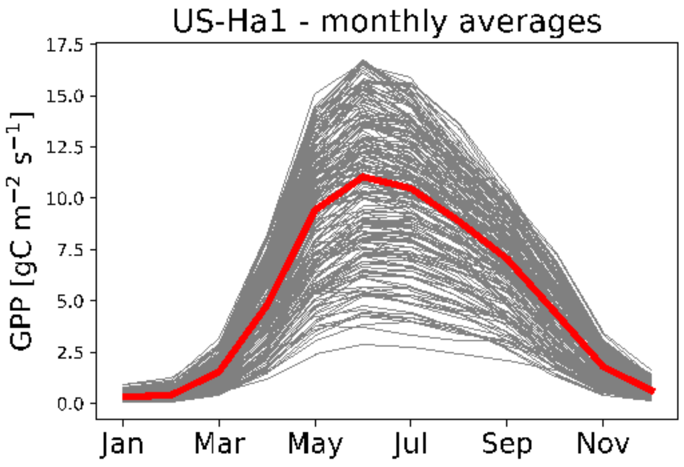


(red arrows represent search paths)

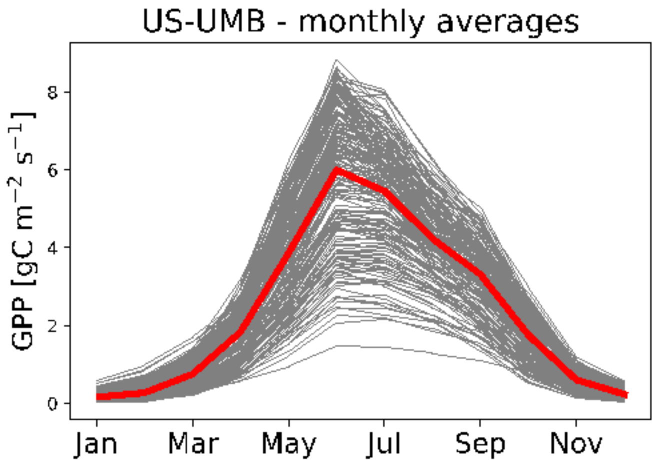
- A similar DFS process starting from each node
- Store partial contractions and re-use paths that were evaluated already.
  - Exploits model structure to reduce the computational expense

ELM Results – Simulations Corresponding to Select Observation sites (fluxnet.org)

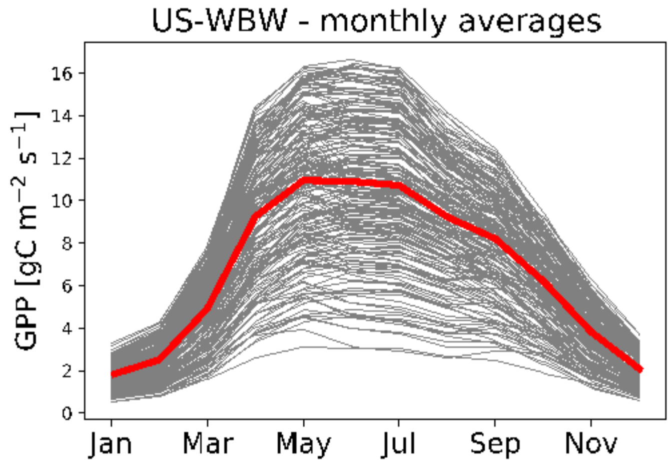
Harvard Forest EMS Tower



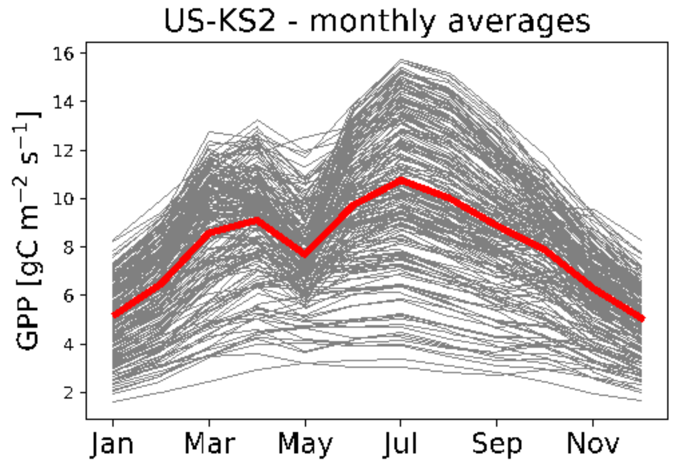
U. of Michigan Biological Station



Walker Branch Watershed



Kennedy Space Center



- **200 runs** corresponding to uniformly randomly sampled parameters over a **10D** parameter space
  - 160 training runs/40 validations runs
  - 8-fold cross validation over 160 training runs

Functional Tensor Network Models – Training

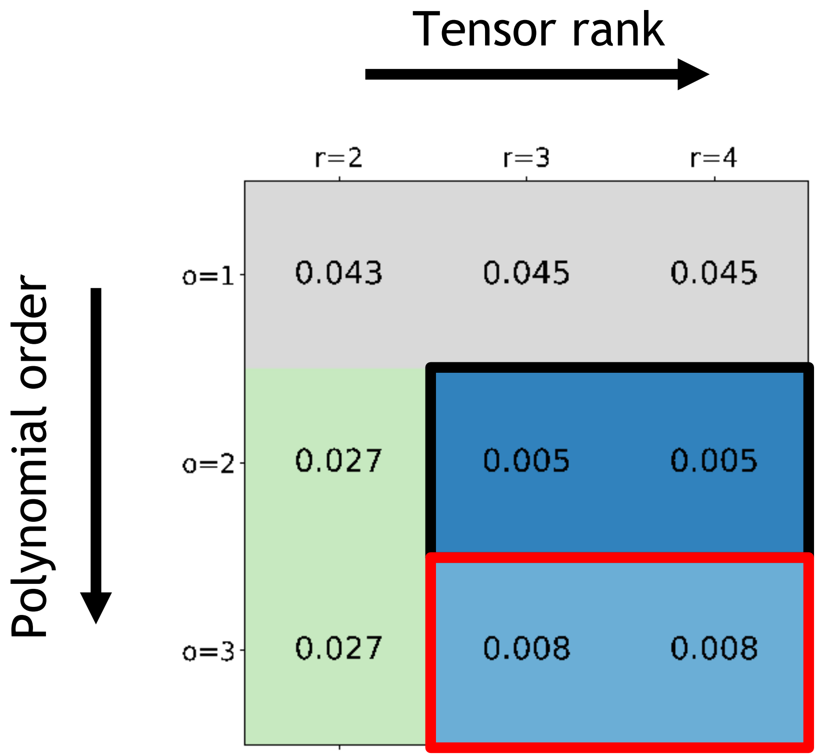
- Data split into 160 training runs / 40 validations runs
- Non-linear least squares with 8-fold cross validation over the training runs
- Univariate functions represented as polynomial expansions based on Legendre polynomials
  - Cross-validation to pick optimum regularization parameter, tensor rank, and polynomial order

$$\theta^* = \arg\min_{\theta} \left(\frac{1}{2} \sum_{i=1}^{N} \left(f(\lambda^{(i)}; \theta) - y^{(i)} \right)^2 + \alpha ||\theta||_2^2 \right)$$

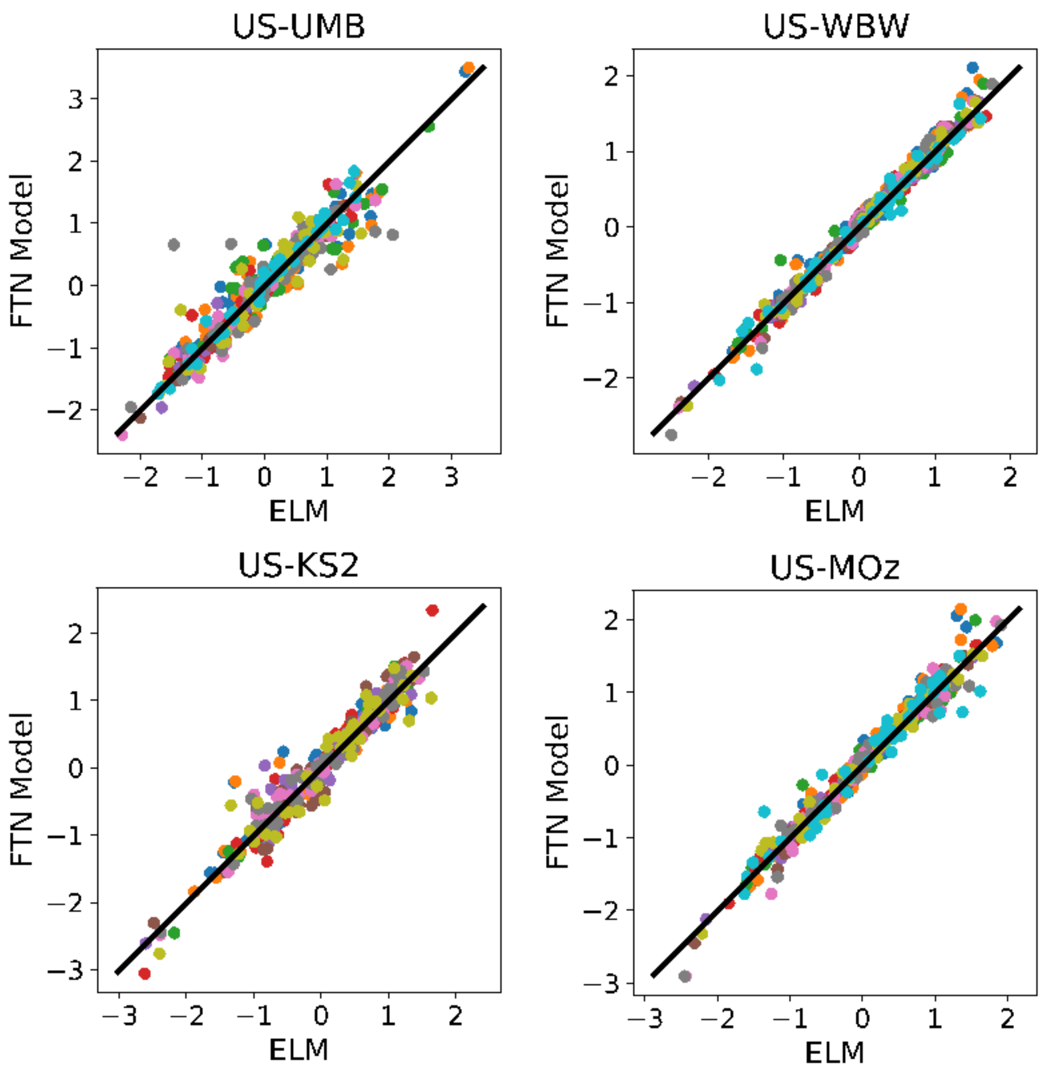
  - Quality of fit assessed via mean-squared error (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(f(\lambda^{(i)}; \theta^*) - y^{(i)} \right)^2$$

ELM Fit Results – FTN Models (in Hierarchical Tucker Format)



Site US-Ha1/June: Validation MSE



Parity plots collect data for all months (with different colors) in the same frame

Validation data centered and normalized by the monthly standard deviation

ELM Results –Variance-based GSA

Main Effect Sobol Index

$$S_i = \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_i)]}{Var[f(\boldsymbol{\lambda})]}$$

Total Effect Sobol Index

$$S_i^T = 1 - \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_{-i})]}{Var[f(\boldsymbol{\lambda})]}$$

Parameter	March		June		September		October	
	S_i	S_i^T	S_i	S_i^T	S_i	S_i^T	S_i	S_i^T
flnr	0.70	0.72	0.80	0.83	0.84	0.86	0.76	0.77
mbbopt	0.01	0.02	0.09	0.13	0.04	0.06	0.02	0.02
vcmaxse	0.13	0.15	0.02	0.02	0	0	0.02	0.02
dayl_scaling	0.06	0.07	0	0	0.04	0.05	0.14	0.14

- fnlr (fraction of N in RuBisCO – CO_2 conversion process)
- mbbopt (stomatal conductance slope – net CO_2 flux)
- vcmaxse (entropy for photosynthetic parameters)
- dayl_scaling (day length scaling parameter)

Summary

- Extended functional tensor train models to accommodate generic tensor network configurations
  - Expanded flexibility in capturing the structure of the original model
  - Efficient gradient computations through tensor network contractions

[Gorodetsky, Safta, Jakeman, submitted, 2021]

- Functional tensor network models constructed via ridge regression are in good agreement with validation data for the driver application
  - Global Sensitivity Analysis results match subject matter expertise given the training runs available for this study