Machine Learning Techniques for Global Sensitivity Analysis in Earth System Models

Cosmin Safta, Khachik Sargsyan, Daniel Ricciuto

Sandia National Laboratories, Livermore, CA & Oak Ridge National Laboratory, Oak Ridge, TN

AGU Fall Meeting 2017 https://climatemodeling.science.energy.gov/projects/optimization-sensor-networks-improving-climate-model-predictions Email: csafta@sandia.gov

Goal

Tackle high-dimensionality and computational expense in Earth System Models via Global Sensitivity Analysis and Machine Learning.

Introduction

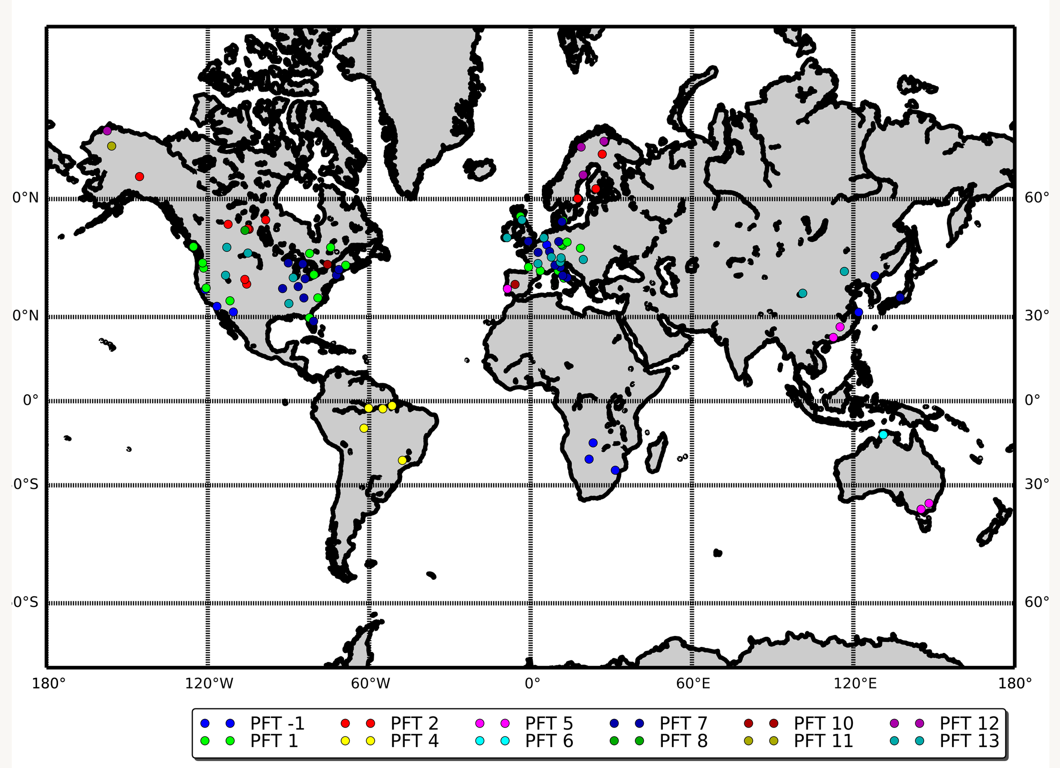
We explore a combination of techniques to extract relevant parameters for each QoI and subsequently construct surrogate models with quantified uncertainty necessary to future developments, e.g. model calibration and prediction studies. In the first step, we will compare the skill of machine-learning models (e.g. deep neural networks) with sparse learning techniques to identify the optimal number of parameters that are influential for the selected QoIs.

E3SM Land Model @ 96 FLUXNET Sites

Parametric uncertainty analysis at FLUXNET sites

Selected 96 FLUXNET sites across several PFTs

- Vary 68 input parameters over selected ranges
- Analyze 2 steady state outputs: Gross Primary Production (GPP) & Total Leaf Area Index (TLAI)



PFT ID & Name	#
-1 mixed	9
1 Boreal evergreen needleleaf tree	22
2 Temperate evergreen needleleaf tree	11
3 Boreal deciduous needleleaf tree	0
4 Tropical evergreen broadleaf tree	5
5 Temperate evergreen broadleaf tree	5
6 Tropical deciduous broadleaf tree	1
7 Temperate decidous broadleaf tree	20
8 Boreal deciduous broadleaf tree	1
9 Boreal evergreen shrub	0
10 Temperate decidous broadleaf shrub	2
11 Boreal deciduous broadleaf shrub	1
12 C3 ArcEc grass	4
13 C3 non-ArcEc grass	15
14 C4 grass	0

- Variance-based decomposition (Sobol indices) of uncertainties into fractional input contributions from each parameter
- Dimensionality reduction and subsequent focus on fewer parameters
- Accurate surrogate construction for input-output maps to enable optimization and efficient calibration

Global Sensitivity Analysis: Variance-based Indices

Variance-based decomposition:

$$f(x_1,x_2,\dots,x_d)=f_0+\sum_{1\le i\le d} f_i(x_i)+\sum_{1\le i<j\le d} f_{i,j}(x_i,x_j)+\sum_{1\le i<j<k\le d} f_{i,j,k}(x_i,x_j,x_k)+\dots$$

- f_i, $f_{i,j}$, $f_{i,j,k}$, … are mutually orthogonal

Sobol sensitivity indices correspond to variance-based decomposition, as they measure fractional contributions of each parameter or group of parameters towards the total variance of selected QoIs

- *Main effect sensitivities,* also called first-order sensitivities, measure variance contribution due to i-th parameter only, defined as

$$S_i=\frac{\mathrm{V}_{\boldsymbol{x}_i}\left[\mathrm{E}_{\boldsymbol{x}_{-i}}[f(\boldsymbol{x})]\,|\boldsymbol{x}_i\right]}{\mathrm{V}f(\boldsymbol{x})},$$

  where $\mathrm{V}_{\boldsymbol{x}_i}$ and $\mathrm{E}_{\boldsymbol{x}_{-i}}$ indicate variance with respect to the i-th parameter and expectation with respect to the rest of the parameters, respectively.
- *Total effect sensitivities* measure total variance contribution of the i-th parameter, *i.e.* including interactions with other parameters, and are defined as

$$S_i^T=\frac{\mathrm{E}_{\boldsymbol{x}_{-i}}\left[\mathrm{V}_{\boldsymbol{x}_i}[f(\boldsymbol{x})|\boldsymbol{x}_{-i}]\right]}{\mathrm{V}\left[f(\boldsymbol{x})\right]}=1-\frac{\mathrm{V}_{\boldsymbol{x}_{-i}}\left[\mathrm{E}_{\boldsymbol{x}_i}[f(\boldsymbol{x})|\boldsymbol{x}_{-i}]\right]}{\mathrm{V}\left[f(\boldsymbol{x})\right]},$$

  where $\mathrm{E}_{\boldsymbol{x}_i}$ and $\mathrm{V}_{\boldsymbol{x}_{-i}}$ indicate expectation with respect to the i-th parameter and variance with respect to the rest of the parameters, respectively.
- Joint sensitivity indices for groups of two or more parameters can also be estimated.

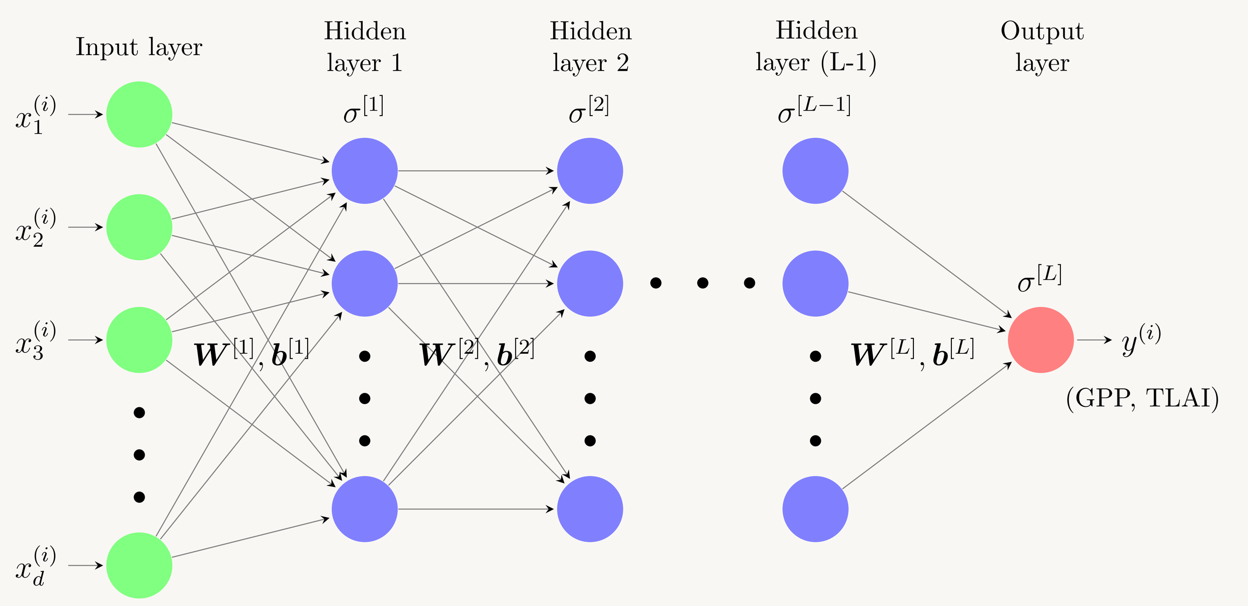
Sobol indices estimates:

- Random Sampling ([5]) $\rightarrow$ need computationally cheap (surrogate) models
- Polynomial Chaos Expansions $\rightarrow$ exploit orthogonality of basis terms

Surrogate Models via Deep Neural Networks

Deep neural network (DNN):

$$y(x;\boldsymbol{W},\boldsymbol{b})=\sigma^{[L]}\left(W^{[L]}\sigma^{[L-1]}\left(W^{[L-1]}\sigma^{[L-2]}\left(\dots\sigma^{[1]}\left(W^{[1]}x+b^{[1]}\right)\dots\right)+b^{[L-1]}\right)+b^{[L]}\right)$$



DNN training using a quadratic objective function augmented with Tikhonov regularization

$$J(\boldsymbol{W},\boldsymbol{b})=\frac{1}{m}\sum_{i=1}^{m}\left(\mathrm{QoI}_i-y^{(i)}\right)^2+\frac{\lambda}{2m}\sum_{j=1}^{L}\|W^{[j]}\|_F^2\rightarrow(\boldsymbol{W},\boldsymbol{b})=\mathrm{argmin}_{\boldsymbol{W},\boldsymbol{b}}J(\boldsymbol{W},\boldsymbol{b})$$

- m - number of training samples; L - number of layers in the network; $\sigma^{[j]}$ - activation function for the j-th layer (relu, sigmoid, etc)
- QoI_i - quantity of interest (GPP & TLAI) corresponding to the ELM simulation using the i-th parameter sample $x^{(i)}$; $y^{(i)}=y(x^{(i)};\boldsymbol{W},\boldsymbol{b})$
- $\|W^{[j]}\|_F$ - Frobenius norm of the j-th layer weight matrix
- λ - regularization parameter, value selected through cross-validation

Surrogate Models via Sparse Polynomial Chaos Expansions

Consider $f(\boldsymbol{x})$ that maps the input parameter vector $\boldsymbol{x}=(x_1,\dots,x_d)$ to an output QoI.

- The input parameter set $\boldsymbol{x}$ is in general viewed as a jointly distributed random vector, but for surrogate construction over ranges $x_i\in[x_{i,\min},x_{i,\max}]$, for $i=1,2,\dots,d$, can be written component-wise as

$$x_i=0.5\left(x_{i,\min}+x_{i,\max}+(x_{i,\max}-x_{i,\min})\right)\xi_i$$

- $\boldsymbol{\xi}\in[-1,1]^d$ - vector of d independent and identically distributed (*i.i.d.*) uniform random variables

The output QoI is viewed as a random variable induced by the uniform random input $\boldsymbol{x}$, and is written as a Polynomial Chaos Expansion [4] with respect to standard multivariate polynomials $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$,

$$\mathrm{QoI}=f(\boldsymbol{x}(\boldsymbol{\xi}))\approx\sum_{\boldsymbol{\alpha}\in\mathcal{I}}c_{\boldsymbol{\alpha}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}),$$

where $\mathcal{I}$ is a multiindex set with size $K=|\mathcal{I}|$.

- Multivariate polynomial $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ corresponds to a multtindex vector $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_d)$ as $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})=\psi_{\alpha_1}(\xi_1)\psi_{\alpha_2}(\xi_2)\cdots\psi_{\alpha_d}(\xi_d)$, and $\psi_{\alpha_i}(\xi_i)$ - univariate polynomial of order α_i in ξ_i.
- The standard polynomials are ortho-normal with respect to the PDF of $\boldsymbol{\xi}$,

$$\langle\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\Psi_{\boldsymbol{\alpha}'}(\boldsymbol{\xi})\rangle\equiv\int_{\boldsymbol{\xi}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\Psi_{\boldsymbol{\alpha}'}(\boldsymbol{\xi})\pi(\boldsymbol{\xi})d\boldsymbol{\xi}=0\quad\text{if}\quad\boldsymbol{\alpha}\neq\boldsymbol{\alpha}',$$

- For uniform inputs - employ Legendre polynomials that are orthogonal with respect to uniform measure $\pi(\boldsymbol{\xi})=2^{-d}$. Other polynomials are available depending on the expected behavior of the QoIs.

Sparse regression in a Bayesian framework - Bayesian Compressive Sensing

f (ELM) is high-dimensional and computationally expensive $\rightarrow$ standard regression approaches are underdetermined $\rightarrow$ Compressive Sensing ([2, 3])

$$\boldsymbol{c}^{CS}=\arg\min_{\boldsymbol{c}}\sum_{i=1}^{m}\left(f(\boldsymbol{x}(\boldsymbol{\xi}^{(i)}))-\sum_{\boldsymbol{\alpha}\in\mathcal{I}}c_{\boldsymbol{\alpha}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}^{(i)})\right)^2+\lambda\sum_{\boldsymbol{\alpha}\in\mathcal{I}}|\boldsymbol{c}_{\boldsymbol{\alpha}}|=\arg\min_{\boldsymbol{c}}\left[\|\boldsymbol{f}-\boldsymbol{G}\boldsymbol{c}\|_2+\lambda\|\boldsymbol{c}\|_1\right]$$

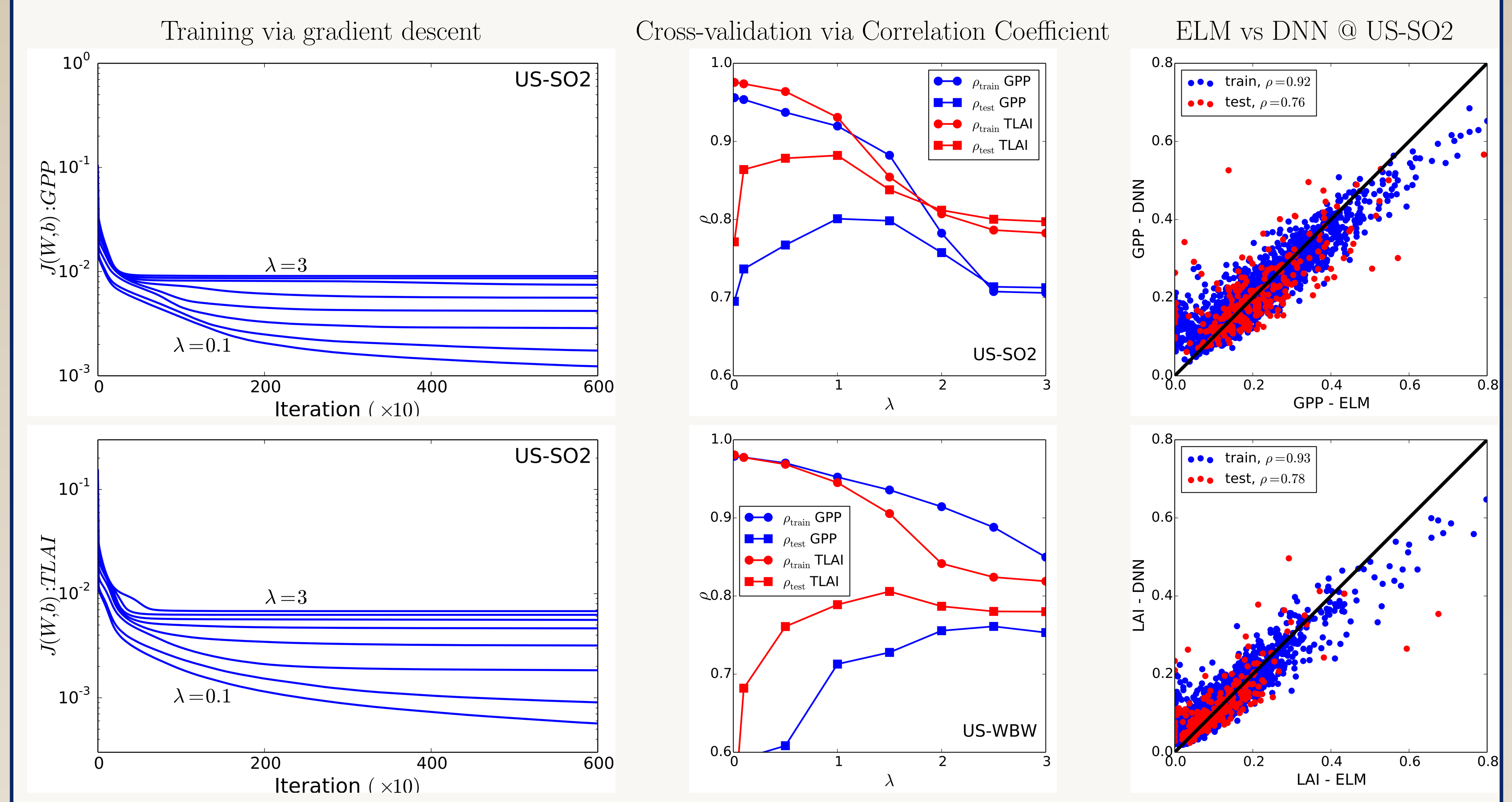
Compressive Sensing in a Bayesian framework [1, 6]:

$$\overbrace{p(\boldsymbol{c}|\mathcal{D})}^{\text{Posterior}}\propto\overbrace{p(\mathcal{D}|\boldsymbol{c})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{c})}^{\text{Prior}}\longrightarrow\boldsymbol{c}^{MAP}=\arg\max_{\boldsymbol{c}}\log p(\boldsymbol{c}|\mathcal{D})=\arg\max_{\boldsymbol{c}}\left[\log L_{\mathcal{D}}(\boldsymbol{c})+\log p(\boldsymbol{c})\right]$$

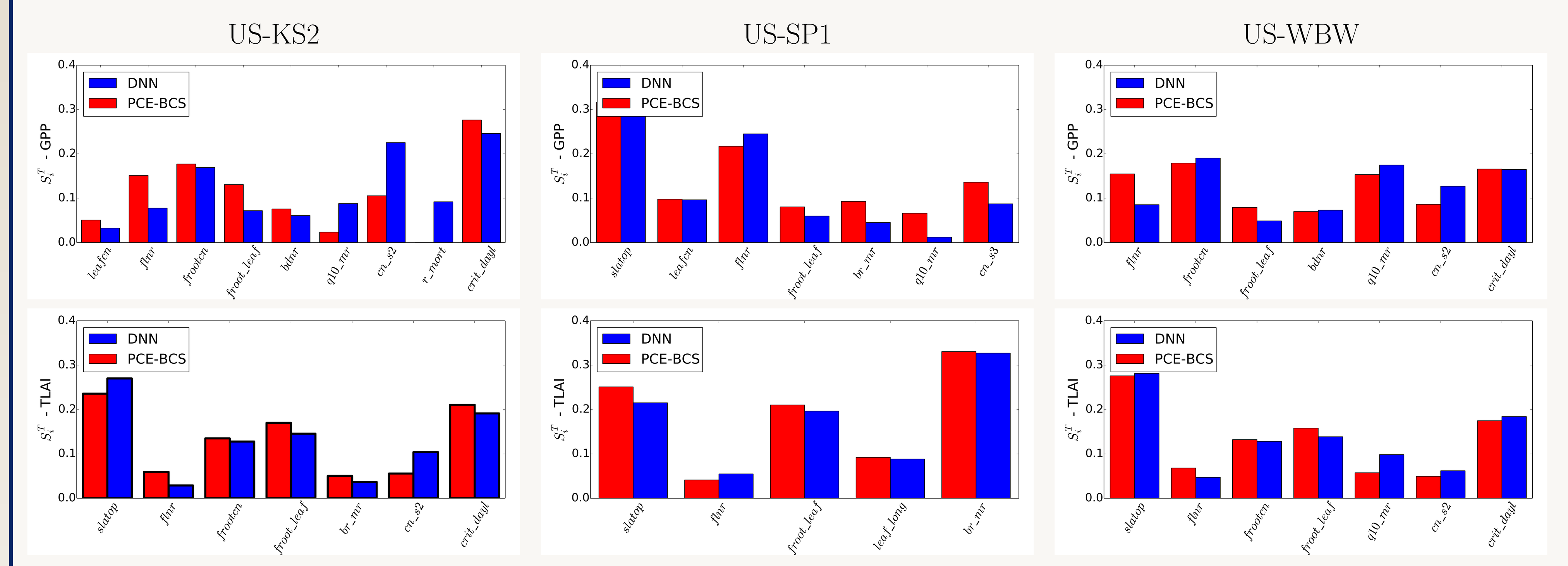
Laplace sparsifying prior

$$p(\boldsymbol{c})=\left(\frac{\lambda}{2}\right)^K\exp\left(-\lambda\sum_{\boldsymbol{\alpha}\in\mathcal{I}}|\boldsymbol{c}_{\boldsymbol{\alpha}}|\right)$$

Neural Network Model Training



Global Sensitivity Analysis Results - DNN vs PCE-BCS



- DNN Model - based on optimal regularization parameter λ
- PCE-BCS Model - based on the Maximum A-Posteriori (MAP) estimate of its coefficients

Dominant ELM parameters:

- TLAI: slatop, flnr, frootcn, froot_leaf, q10_mr, cn_s2, crit_dayl, leaf_long, br_mr.
- GPP: flnr, frootcn, froot_leaf, bdnr, q10_mr, cn_s2, crit_dayl, slatop, leafcn, br_mr, q10_mr, cn_s3, r_mort.

Summary

- GSA results obtained via alternative formulations (DNN & PCE-BCS) are in qualitative agreement.
- Identified a subset of 10-15 parameters that are influential to the two QoIs (GPP & TLAI) selected for this study.

[1] Babacan, S., R. Molina, and A. Katsaggelos (2010), Bayesian compressive sensing using Laplace priors, *IEEE Transactions on Image Processing*, *19*(1), 53–63.

[2] Candès, E., J. Romberg, and T. Tao (2006), Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, *52*(2), 489–509.

[3] Donoho, D. (2006), Compressed sensing, *IEEE Trans. Inform. Theory*, *52*(4), 1289–1306.

[4] Ghanem, R., and P. Spanos (1991), *Stochastic Finite Elements: A Spectral Approach*, Springer Verlag, New York.

[5] Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola (2010), Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications*, *181*(2), 259–270.

[6] Sargsyan, K., C. Safta, H. Najm, B. Debusschere, D. Ricciuto, and P. Thornton (2014), Dimensionality reduction for complex models via Bayesian compressive sensing, *International Journal of Uncertainty Quantification*, *4*(1), 63–93.

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-NA-0003525. Oak Ridge National Laboratory is managed by UT-BATTELLE for DOE under contract DE-AC05-00OR22725.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Offices of Biological and Environmental Research and Advance Scientific Computing Research. This study used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.