Dimensionality Reduction and Physics-Informed Neural Networks for Climate Land Models

Khachik Sargsyan [1] Cosmin Safta [1] Vishagan Ratnaswamy [1]
Daniel M. Ricciuto [2]

[1] Sandia National Laboratories, Livermore CA

[2]Oak Ridge National Laboratory, Oak Ridge TN

Virtual SIAM CSE
March 1-5, 2021

Acknowledgements

- DOE, Office of Science,
  - Advanced Scientific Computing Research (ASCR)
  - Scientific Discovery through Advanced Computing (SciDAC) program
  - Biological and Environmental Research (BER)
  - National Energy Research Scientific Computing Center (NERSC)

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.

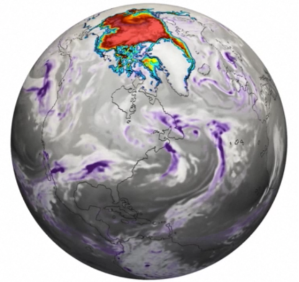
Outline

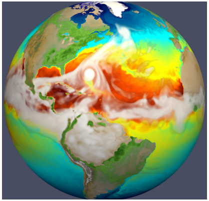
- Motivation
  - Energy Exascale Earth System Model (E3SM): Land component
- Need for Surrogate Models
  - Polynomial Chaos Surrogate
  - **Karhunen-Loève expansions for field quantities**
  - Neural Network Surrogates
    - Multilayer Perceptron (MLP)
    - Long short term memory (LSTM)
    - **Physics-based LSTM**
- Global Sensitivity Analysis (GSA)
- Preliminary Results

E3SM Model Overview

Energy Exascale Earth System Model (E3SM) is a coupled earth model

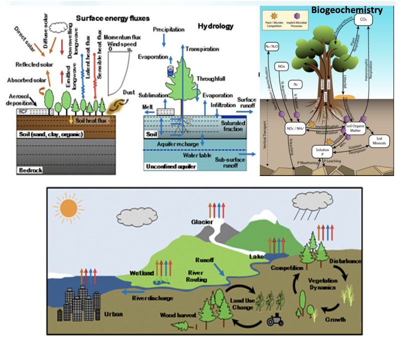
- US Department of Energy (DOE) sponsored Earth system model https://e3sm.org
- Ocean, Atmosphere, Sea ice and **Land** Components
- Computationally expensive to run the coupled mode (including ocean and atmosphere)
- Can only do a few global simulations of coupled E3SM models $\rightarrow$ hard to get training data

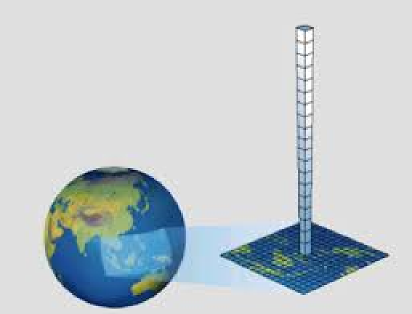




Land model incorporates a set of biogeophysical processes:

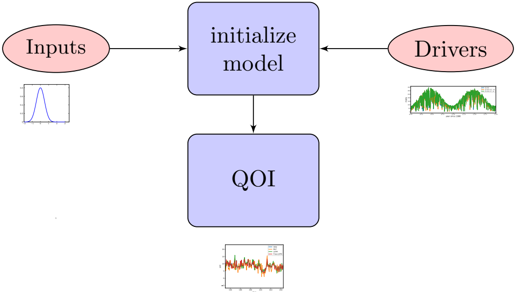
- Cheapest component, can run in single column mode
- Simplified python model available (sELM)
- Can evaluate model many times for various input parameters

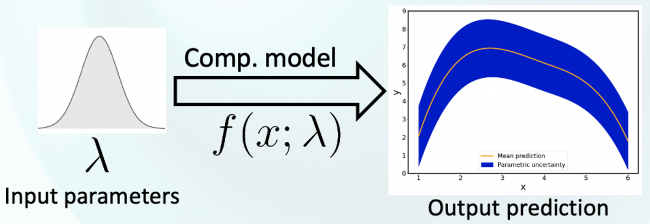




ELM Produces Time Series given Input Parameters and Forcing Drivers

- $\mathcal{O}(10) - \mathcal{O}(100)$ uncertain inputs
- Daily Forcings/Drivers
  1. Min/Max Temperature
  2. Day of Year
  3. Solar radiation
  4. Water Availability





Surrogates are necessary for **expensive** computational models

- ... otherwise called supervised ML, metamodel, emulator, proxy, response surface.

$$f(x; \lambda) \approx f_s(x; \lambda)$$

- Surrogates are required for ensemble-intensive studies, such as
  - parameter estimation
  - uncertainty propagation
  - global sensitivity analysis
  - optimal experimental design

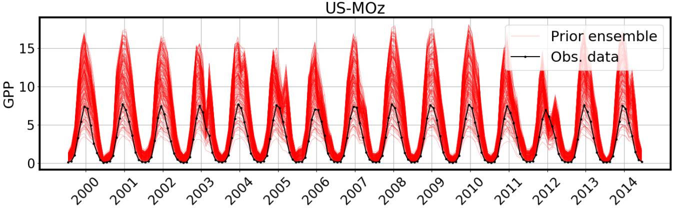
This talk

Investigating surrogate construction approaches for ELM to enable all of the above

Curse of dimensionality hits twice

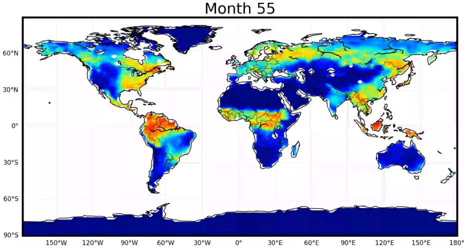
$$f(x; \boldsymbol{\lambda}) \approx f_s(x; \boldsymbol{\lambda})$$

- Input : large number of input parameters



$$f(\boldsymbol{x}; \lambda) \approx f_s(\boldsymbol{x}; \lambda)$$

- Output : spatio-temporal resolution



Polynomial chaos (PC) surrogate for black-box $f(\boldsymbol{\lambda})$

- Represent QOIs as orthogonal expansion of random variables

$$f(\boldsymbol{\lambda}) \approx f_c(\boldsymbol{\lambda}(\boldsymbol{\xi})) = \sum_{\alpha \in \mathcal{I}} c_\alpha \Psi_\alpha(\boldsymbol{\xi})$$

- Germ: $\xi = (\xi_1, \xi_2, ..., \xi_d)$, e.g. uniform: $\boldsymbol{\lambda}(\boldsymbol{\xi})$ is a linear scaling
- Multi-index $\alpha = \{\alpha_1, \ldots, \alpha_d\}$
- Orthogonal polynomials wrt $p(\boldsymbol{\xi})$, $\Psi_\alpha(\boldsymbol{\xi}) = \prod_{i=1}^d \psi_{\alpha_i}(\xi_i)$
- Typical construction approach: regression to find PC modes c_α
- **Advantages of PC**
  - moment estimation, uncertainty propagation, global sensitivity
- **Expensive model challenge**
  - Use Bayesian regression, helps to quantify lack of simulation data
- **High-d challenge** $d \gg 1$
  - Number of terms in expansion of order p and dimension d: $|\mathcal{I}| = \frac{(d+p)!}{d!p!}$
  - Use sparse regression
- We employ Bayesian compressed sensing (BCS): iterative algorithm for Bayesian sparse learning [Babacan, 2010; Sargsyan, 2014; Ricciuto, 2018]

‘Free’ Global Sensitivity Analysis with PC

$$f(\boldsymbol{\lambda}(\boldsymbol{\xi})) \simeq \sum_{\alpha\in\mathcal{I}} c_\alpha \Psi_\alpha(\boldsymbol{\xi})$$

- Main effect sensitivity indices

$$S_i = \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_i)]}{Var[f(\boldsymbol{\lambda})]} = \frac{\sum_{\alpha\in\mathcal{I}_i} c_\alpha^2 ||\Psi_\alpha||^2}{\sum_{\alpha\in\mathcal{I}\setminus\{0\}} c_\alpha^2 ||\Psi_\alpha||^2}$$

  - $\mathcal{I}_i$ is the set of bases with only ξ_i involved
  - S_i is the uncertainty contribution that is due to i-th parameter only

- Total effect sensitivity indices

$$T_i = 1 - \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_{-i})]}{Var[f(\boldsymbol{\lambda})]} = \frac{\sum_{\alpha\in\mathcal{I}_i^T} c_\alpha^2 ||\Psi_\alpha||^2}{\sum_{\alpha\in\mathcal{I}\setminus\{0\}} c_\alpha^2 ||\Psi_\alpha||^2}$$

  - $\mathcal{I}_i^T$ is the set of bases with ξ_i involved, including all its interactions.
  - T_i is the total uncertainty contribution due to i-th parameter

[Sudret, 2008; Crestaux, 2009; Sargsyan, 2017]

Spatio-temporal surrogate model via Karhunen-Loève expansions

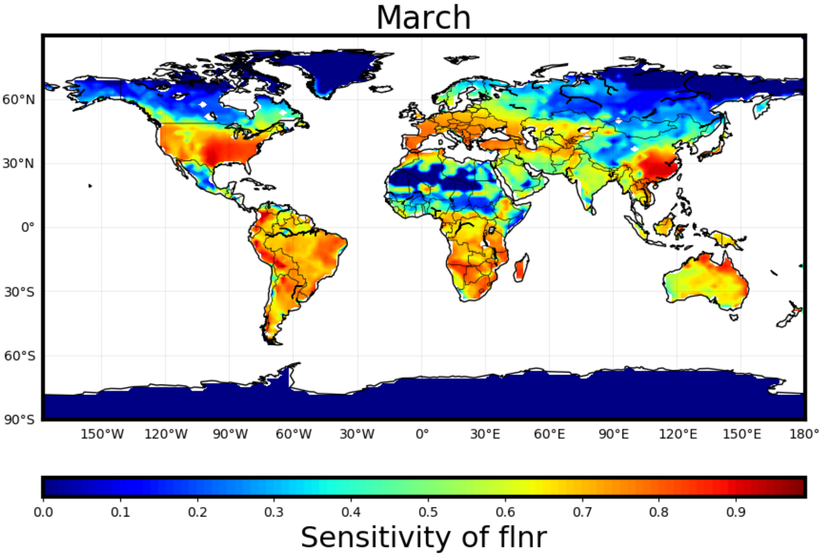
- 3183 active land cells over 180 months is $> 500,000$ outputs
- Karhunen-Loève expansions help reduce dimensionality due to strong spatio-temporal correlations

$$f(x;\lambda) = \bar{f}(\lambda) + \sum_{j=1}^{J} f_j(\lambda)\sqrt{\mu_j}\phi_j(x)$$

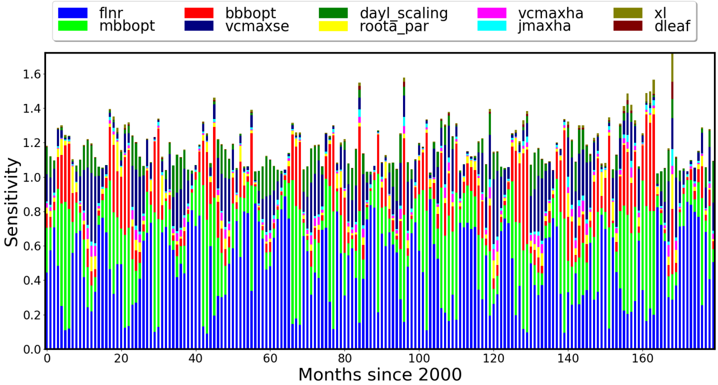
- Instead of 500,000 surrogates, we build about $J = 2,000$ surrogates, one for each coefficient $f_j(\lambda)$

$$\begin{aligned} f(x;\lambda) &= \sum_{j=0}^{J}\sum_{k=0}^{K} f_{jk}\Psi_k(\lambda)\sqrt{\mu_j}\phi_j(x) \\ &= \sum_{k=0}^{K}\left(\sum_{j=0}^{J} f_{jk}\sqrt{\mu_j}\phi_j(x)\right)\Psi_k(\lambda) \end{aligned}$$

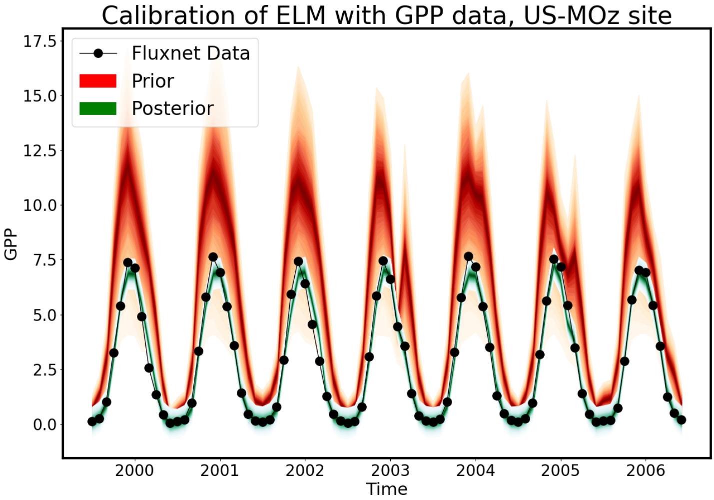
Spatially resolved GSA



Time-resolved GSA

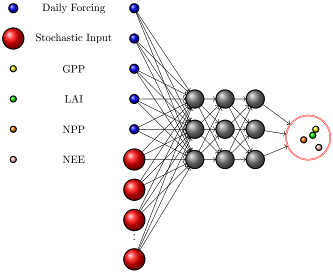


Bayesian inference and prediction with surrogate



Transitioning from UQ technologies to ML

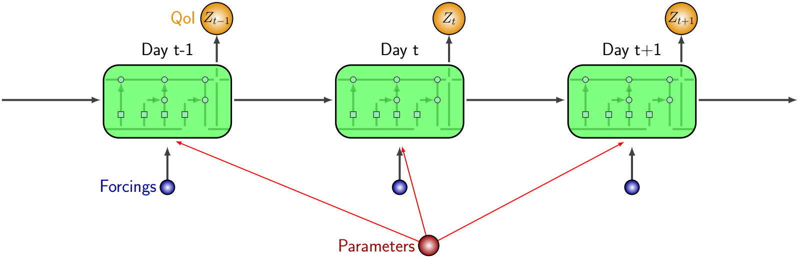
- Polynomial chaos requires smooth QoIs
- Exploring transient behavior and daily dynamics requires more accurate surrogates
- Benign MLP, feed-forward network did not do too well



- Does not account for temporal aspect of model
- Cannot propagate information of QOIs day to day (or month to month)

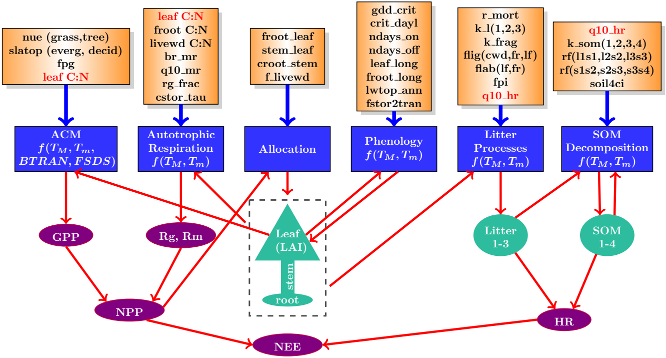
Long short term memory (LSTM) accounts for temporal evolution

- Vanilla LSTM Recurrent NN architecture
- One network per QoI
- Much better than PC, better than MLP

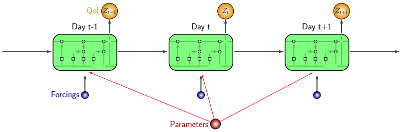


Graph Structure of ELM Land Model

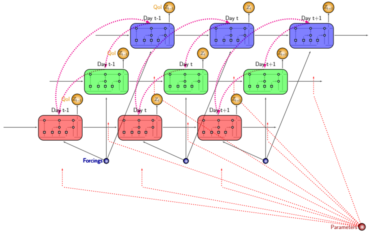
Looking under the hood helps build physics-informed architecture



Physics-driven architecture incorporates known connections into LSTM

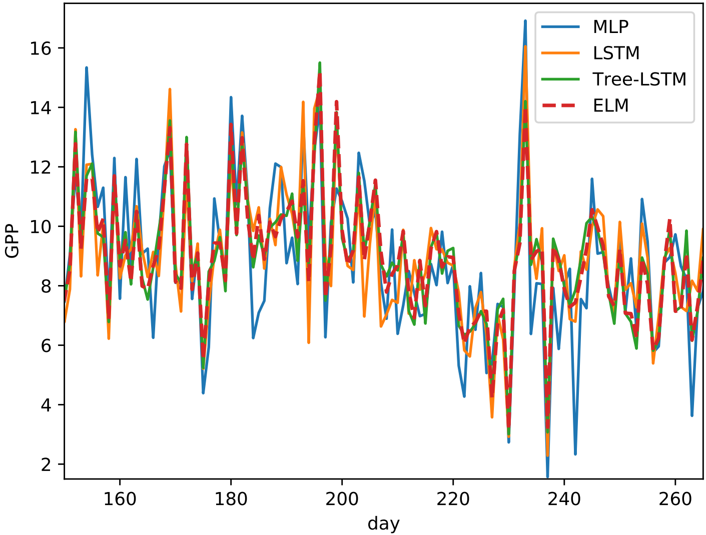


Vanilla LSTM

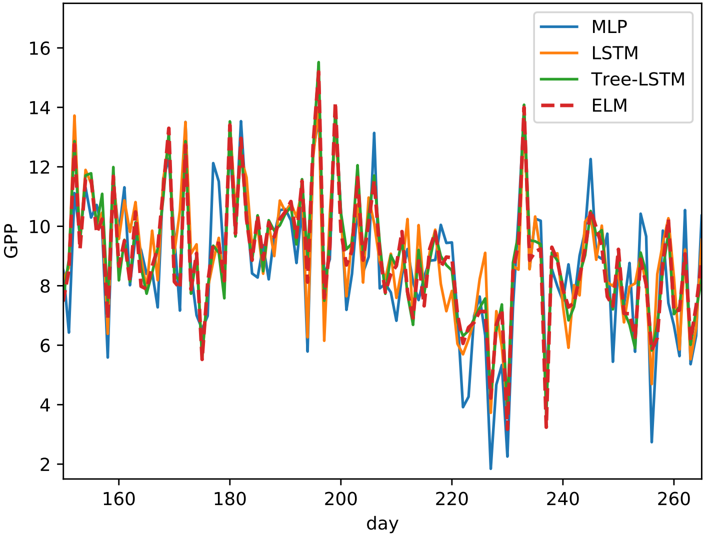


Physics-informed LSTM

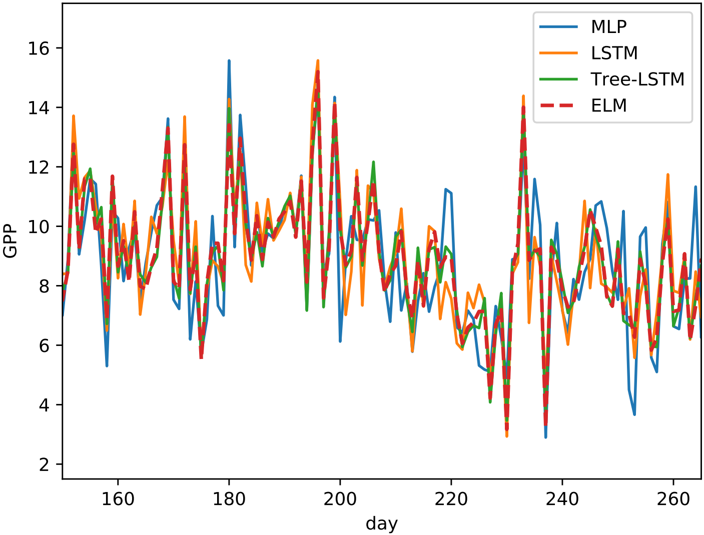
NN fits to ELM



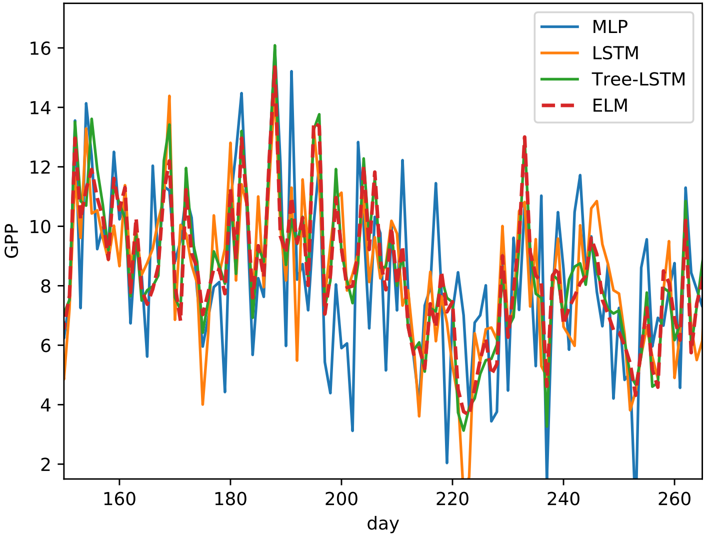
NN fits to ELM



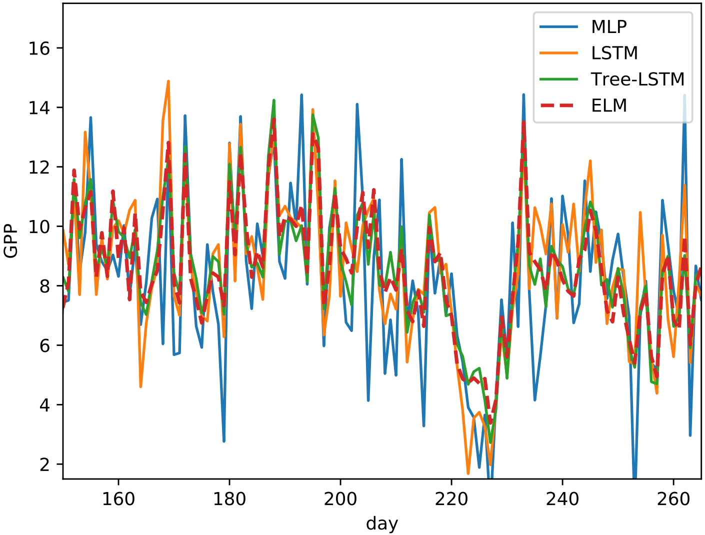
NN fits to ELM



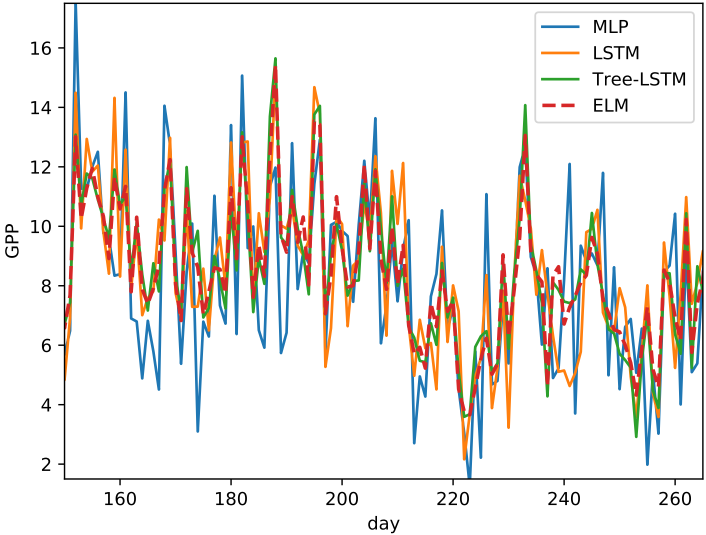
NN fits to ELM



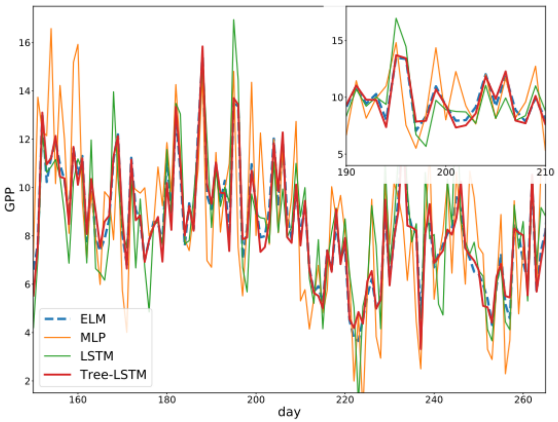
NN fits to ELM



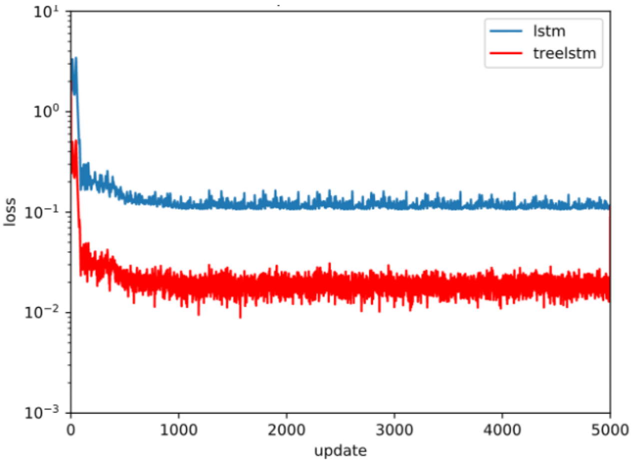
NN fits to ELM



NN fits to ELM



NN fits to ELM



Price to pay: for NN surrogates, GSA should be carried out with Monte-Carlo

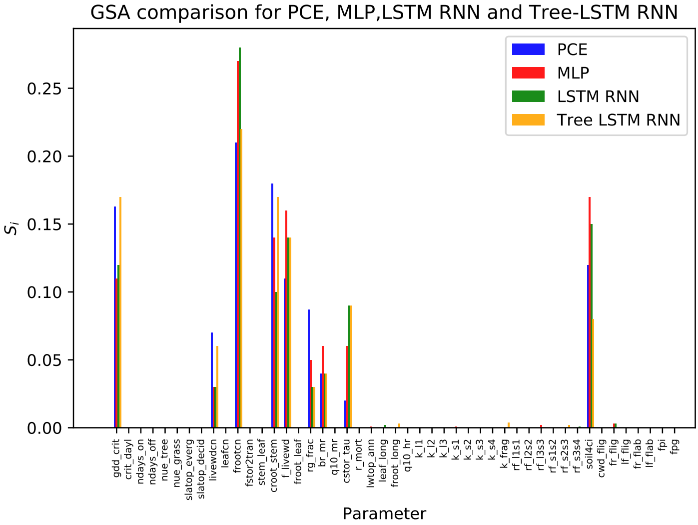
- Mean estimate: $E[f(\boldsymbol{\lambda})] \approx \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{\lambda}^{(n)})$
- Variance estimate: $Var[f(\boldsymbol{\lambda})] \approx \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{\lambda}^{(n)})^2 - E[f(\boldsymbol{\lambda})]^2$
- Main sensitivity:

$$S_i = \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_i)]}{Var[f(\boldsymbol{\lambda})]} \approx \frac{1}{Var[f(\boldsymbol{\lambda})]} \left(\frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{\lambda}^{(n)}) f(\boldsymbol{\lambda}_{-i}^{'(n)} \cup \boldsymbol{\lambda}_i^{(n)}) - E[f(\boldsymbol{\lambda})]^2\right.$$

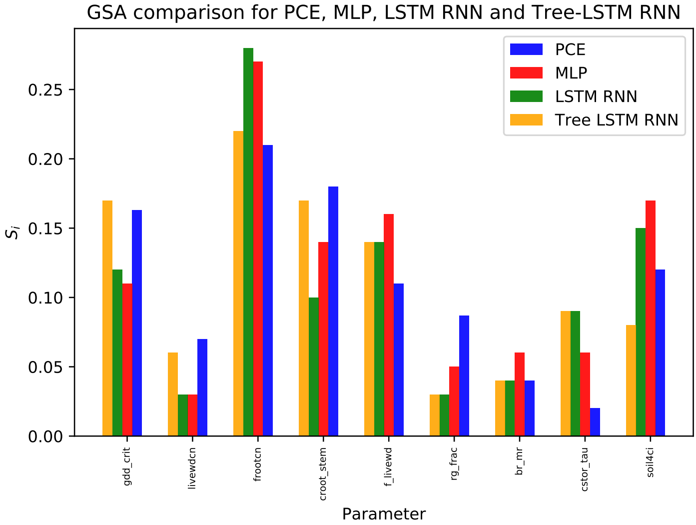
  where $\boldsymbol{\lambda}_{-i}^{'(n)} \cup \boldsymbol{\lambda}_i^{(n)}$ is a single-column swap sample given two sampling schemes $\boldsymbol{\lambda}^{(n)}$ and $\boldsymbol{\lambda}^{'(n)}$
- ... similar estimators for total sensitivity
- Inherits all the challenges of Monte-Carlo

[Jansen, 1999; Sobol, 2001; Saltelli, 2002]

Global Sensitivity Analysis Comparison



Global Sensitivity Analysis Comparison



Summary **Thank you!**

Overview

- Key UQ step, surrogate construction $==$ supervised ML
- Dimensionality reduction via Karhunen-Loève expansions (aka autoencoder)
- Physics-based LSTM architecture outperforms traditional NN methods (MLP) and traditional UQ methods (PCEs)
- Qualitatively similar sensitivity results compared to PCE

Current:

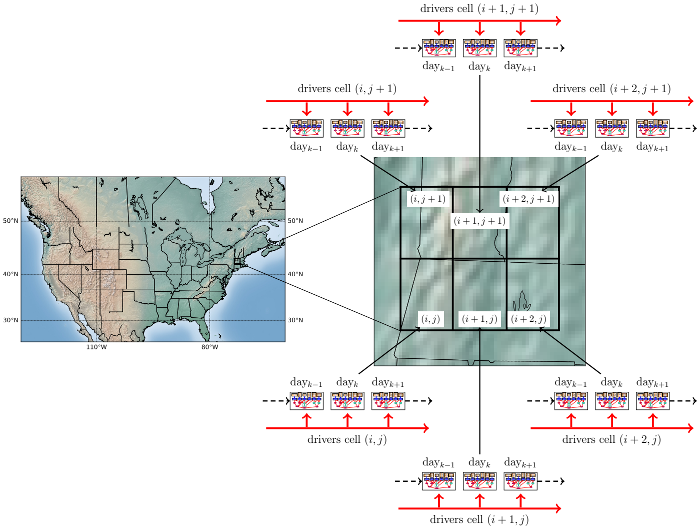
- Employing the reduced-dimensional spatio-temporal surrogate for calibration and optimal experimental design

Shameless Plug:

- Postdoc position(s) available at SNL-California at the intersection of UQ/ML
- www.sandia.gov/careers $\rightarrow$ 'View All Openings', e.g. job id 675390
- ... or email me at ksargsy@sandia.gov

Additional Material

ELM Simulation Details



Tree RNN more accurate than PCE and traditional ML methods

- Computed Mean RMS for each Surrogate

Method	**Train (Daily/Month)**	**Val (Daily/Month)**
PCE	$(N/A)/35\%$	$(N/A)/46\%$
MLP	$19/14\%$	$32/20\%$
LSTM	$14/10\%$	$21/16\%$
Tree-LSTM	$6/2\%$	$9/5\%$

- Tree-LSTM outperforms PCE, MLP and LSTM-RNN

Global Sensitivity Analysis

- $Y = f(X_1, X_2, X_3, ..., X_N)$
- Total Variance decomposition (normalized)

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$$

- If X_i are independent: $V(Y) = \sum_{i=1}^{N} V_i + \sum_{1 \leq i \leq j \leq N} V_{ij} + .. + V_{1,..,p}$
- Use Sobol indices $\rightarrow$ QOI's variance to be decomposed based on variances of inputs

Sobol Indices

$S_i = \frac{\text{Var}_{X_i}(E(Y|X_i))}{Var(Y)}$ First Sobol Indices

$S_{ij} = \frac{Var_{ij}}{Var(Y)}$ Second Order Sobol Indices

Method

- PCE allows for analytical computation of S_i
- ML needs Monte Carlo Integration to compute S_i

Summary of Case Study

Training Details

- Generate samples from sELM model: 30 years (1980-2009)
- Each training set (time history) has $10,950$ data points (daily)
- Simulation at University of Michigan Biological Station site
- 500 training samples, 500 validation samples

NN Details

- Train on daily QoIs
- 500 Epochs of SGD
- 2 layers, 150 neurons
- L_2 loss, dropout regularization

MLP

- No time dynamics
- No physics

LSTM

- Time dynamics
- No physics

Tree-LSTM

- Time dynamics
- Physics

PC Details

- Hard to train on daily averages (noisy)
- Train on monthly averages
- Use Bayesian compressive sensing to compute coefficients
- Build surrogate for each average month, i.e. 30×12 surrogates