Physics-Informed Recurrent Neural Networks for Land Model Surrogate Construction

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

[1] Sandia National Laboratories, Livermore CA

[2]Oak Ridge National Laboratory, Oak Ridge TN

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

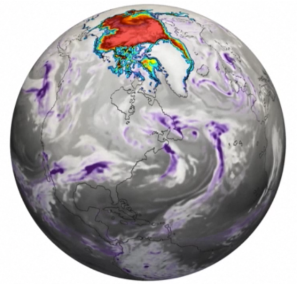
- Motivation
  - Energy Exascale Earth System Model (E3SM): Land component
- Need for Surrogate Models
  - Polynomial Chaos Surrogate
  - Neural Network Surrogates
    - Multilayer Perceptron
    - LSTM RNN
    - **Tree-LSTM RNN**
- Global Sensitivity Analysis (GSA)
- Preliminary Results

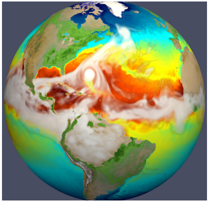
E3SM Model Overview

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 → hard to get training data

E3SM
Energy Exascale
Earth System Model

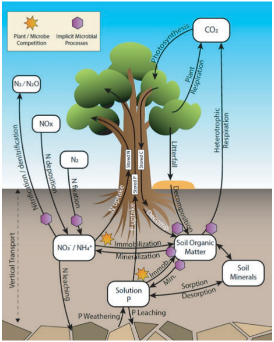


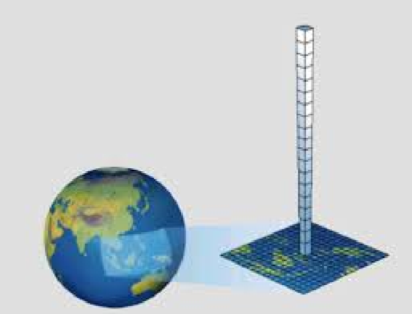


E3SM Land Model (ELM) Overview

Land model incorporates a set of biogeophysical processes:

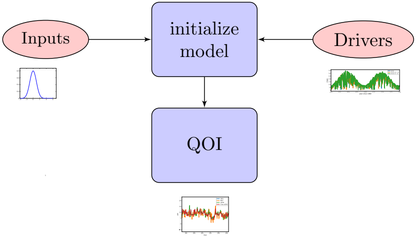
- Cheapest component, can run in single column mode
- Simplified python model available (sELM)
- Can evaluate model many times for various input parameters





ELM Produces Time Series given Input Parameters and Forcing Drivers

- Daily Forcings (site dependent)
  1. Min/Max Temperature
  2. Day of Year
  3. Solar radiation
  4. Water Availability
- Thirty year record (1980-2009) of forcings
- 47D Stochastic Space
- QOIs dependent on previous day's output



Surrogates are necessary for **expensive** computational models

- ... otherwise called supervised ML
- 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

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

- Represent QOIs as orthogonal expansion of random variables

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

- Germ: $\xi = (\xi_1, \xi_2, ..., \xi_d)$, e.g. uniform: $\lambda(\xi)$ is a linear scaling
- Multi-index $\alpha = \{\alpha_1, \ldots, \alpha_d\}$
- Orthogonal polynomials wrt $p(\xi)$, $\Psi_\alpha(\xi) = \prod_{i=1}^{d} \psi_{\alpha_i}(\xi_i)$
- Typical construction approach: regression to find PC modes c_α
- **Advantages of PC**
  - moment estimation, uncertainty propagation, global sensitivity analysis
- **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 $f(\lambda(\xi)) \simeq \sum_{\alpha \in \mathcal{I}} c_\alpha \Psi_\alpha(\xi)$

- Main effect sensitivity indices

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

  - $\mathcal{I}_i$ 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

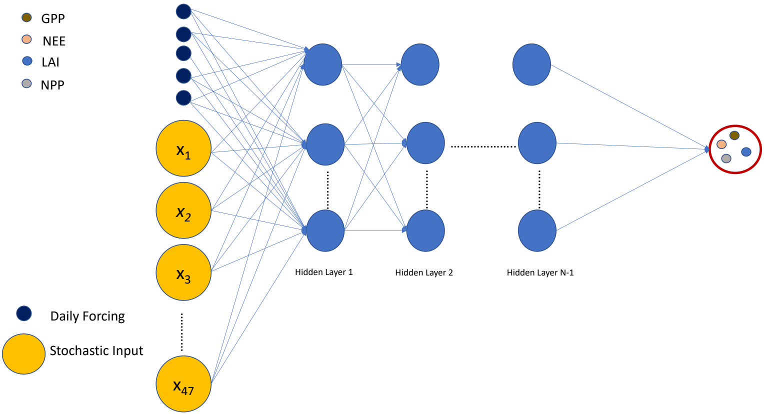
$$T_i = 1 - \frac{Var[\mathbb{E}(f(\lambda|\lambda_{-i})]}{Var[f(\lambda)]} = \frac{\sum_{\alpha \in \mathcal{I}_i^T} c_\alpha^2 ||\Psi_\alpha||^2}{\sum_{\alpha \in \mathcal{I} \setminus \{0\}} c_\alpha^2 ||\Psi_\alpha||^2}$$

  - $\mathcal{I}_i^T$ 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]

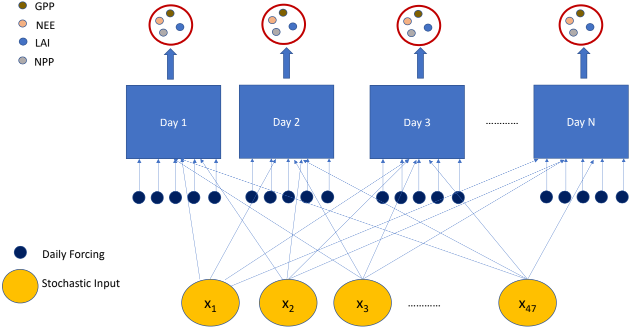
Surrogate Modeling: MLP Approach

- Benign feed-forward network
- Does not account for temporal aspect of model
  - Cannot propagate information of QOIs day to day (or month to month)
- Model dealt with as a black box



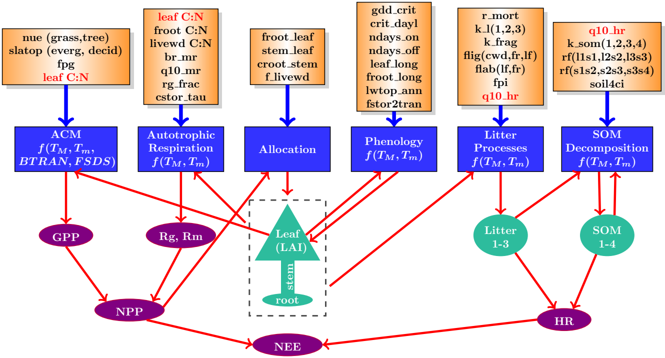
Surrogate Modeling: LSTM-RNN

- Vanilla LSTM Recurrent NN architecture
- Accounts for temporal aspect of model
- Model dealt with as a black box



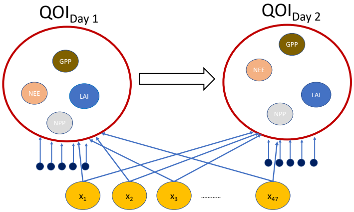
Graph Structure of ELM Land Model

Looking under the hood helps build physics-informed architecture

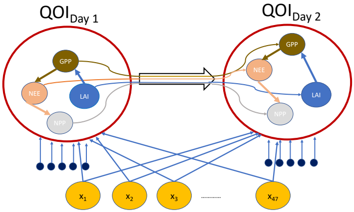


Surrogate Modeling: Tree LSTM RNN

- Retain similar structure to Gated RNN ([Tai, 2015])
- Incorporate physical connections into LSTM
- Each QOI has a LSTM gate



LSTM RNN
Vanilla



Tree-LSTM RNN
Physics-informed

For NN surrogates, GSA can be carried out with Monte-Carlo

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

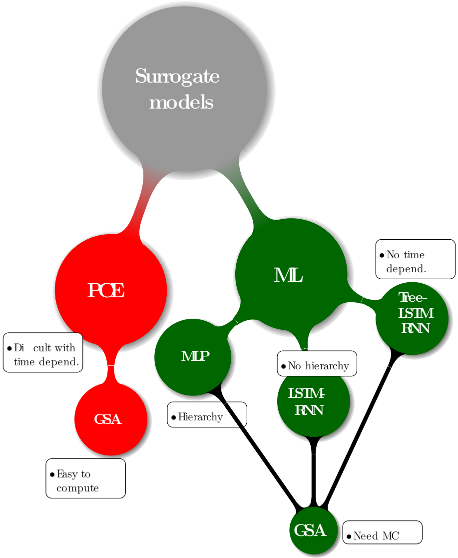
$$S_i = \frac{Var[\mathbb{E}(f(\lambda|\lambda_i)]}{Var[f(\lambda)]} \approx \frac{1}{Var[f(\lambda)]}\left(\frac{1}{N}\sum_{n=1}^{N} f(\lambda^{(n)}) f(\lambda_{-i}^{\prime(n)} \cup \lambda_i^{(n)}) - E[f(\lambda)]^2\right),$$

  where $\lambda_{-i}^{\prime(n)} \cup \lambda_i^{(n)}$ is a single-column swap sample given two sampling schemes $\lambda^{(n)}$ and $\lambda^{\prime(n)}$

- ... similar estimators for total sensitivity
- Inherits all the challenges of Monte-Carlo

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

Overview of Surrogate Models Implemented



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

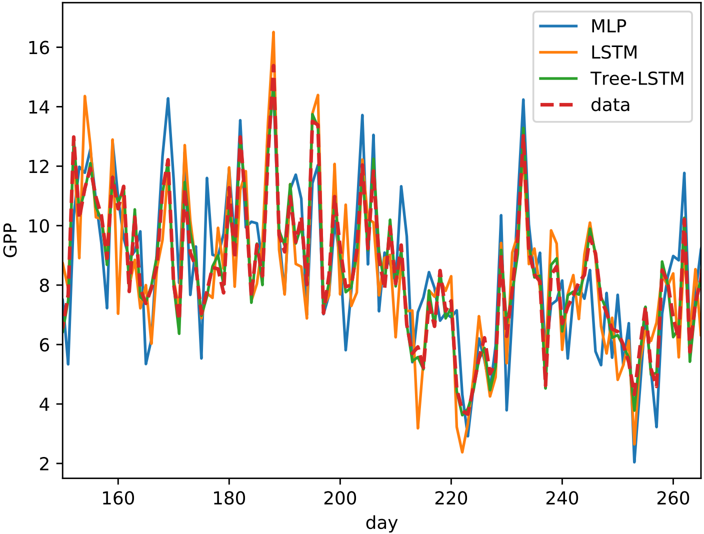
Tree-LSTM

- Time dynamics
- Physics

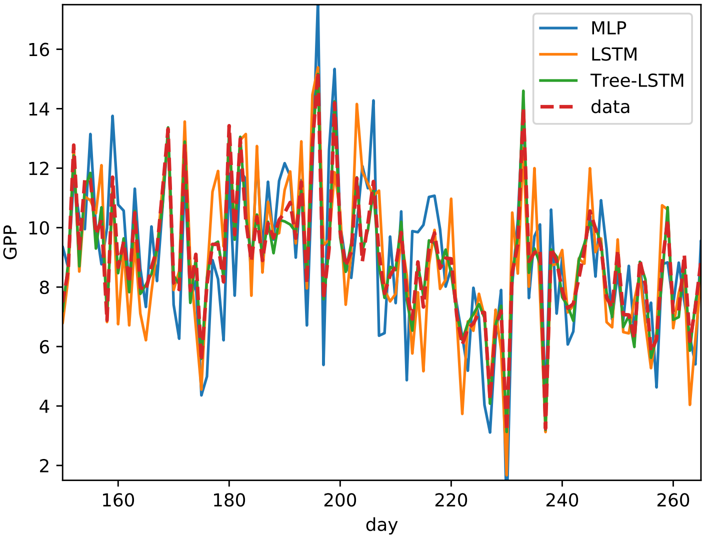
PC Details

- 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

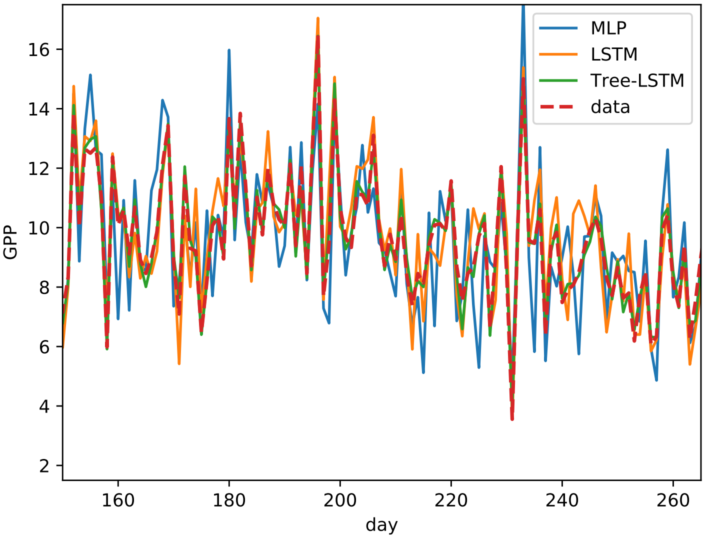
Snapshot of training sample 1



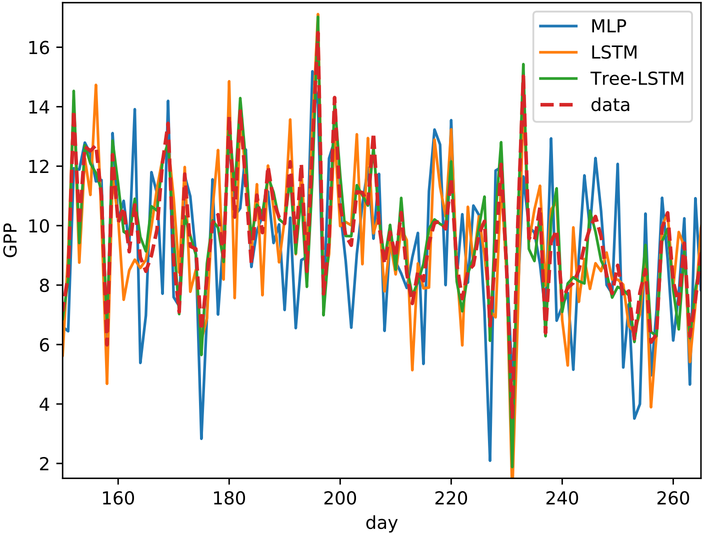
Snapshot of GPP training sample 2



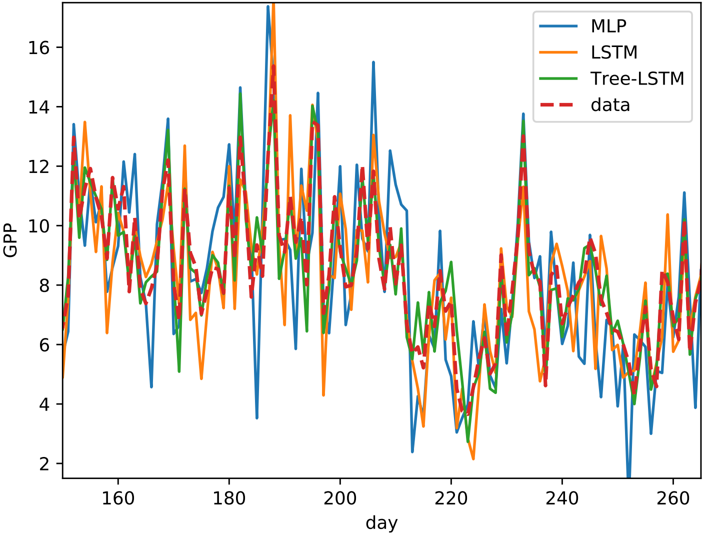
Snapshot of GPP training sample 3



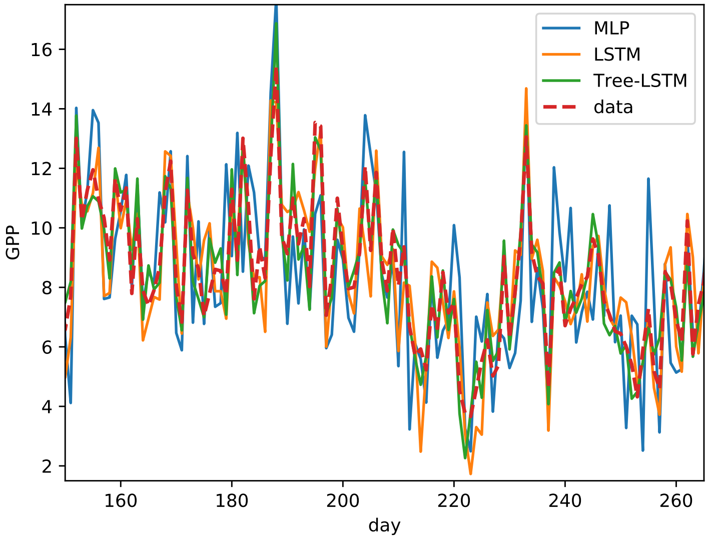
Snapshot of GPP validation sample 1



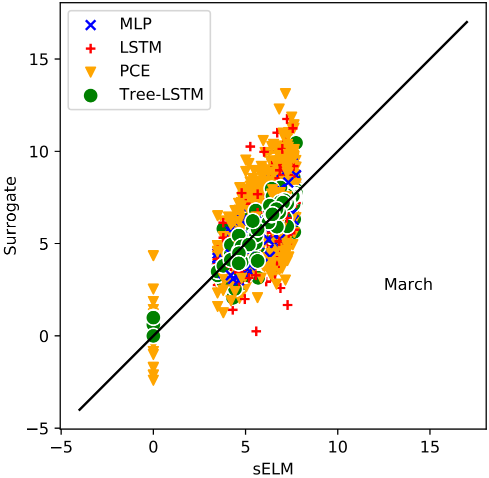
Snapshot of GPP validation sample 2



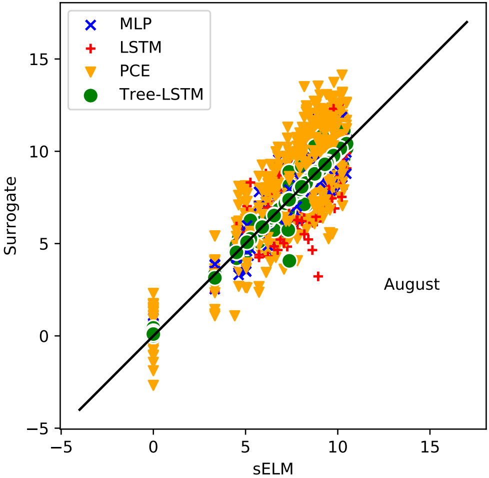
Snapshot of GPP validation sample 3



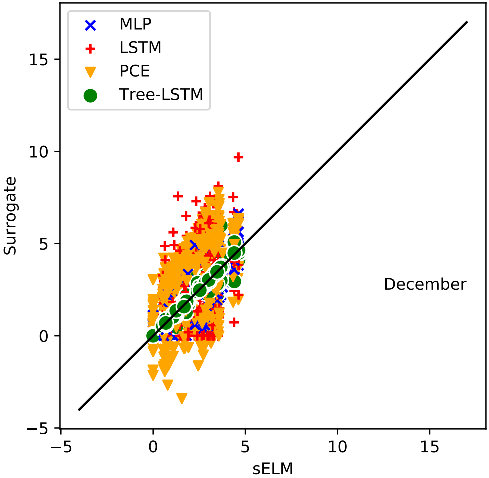
GPP Parity Plot comparisons for various months



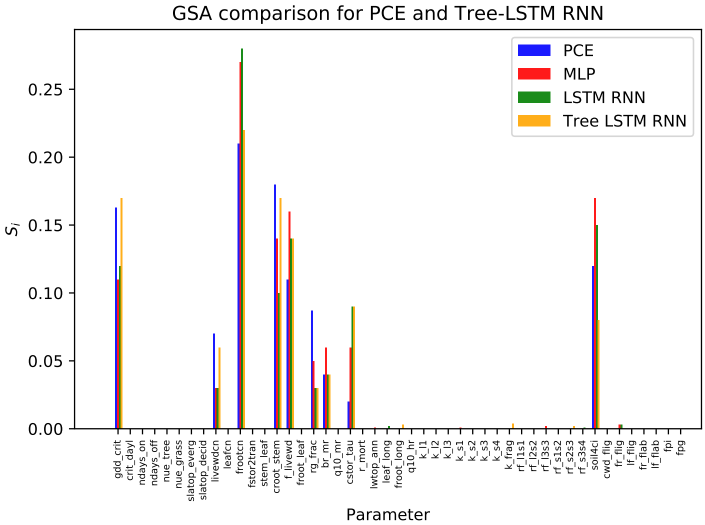
GPP Parity Plot comparisons for various months



GPP Parity Plot comparisons for various months



Global Sensitivity Analysis Comparison



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

Conclusions and Future Work **Thank you!**

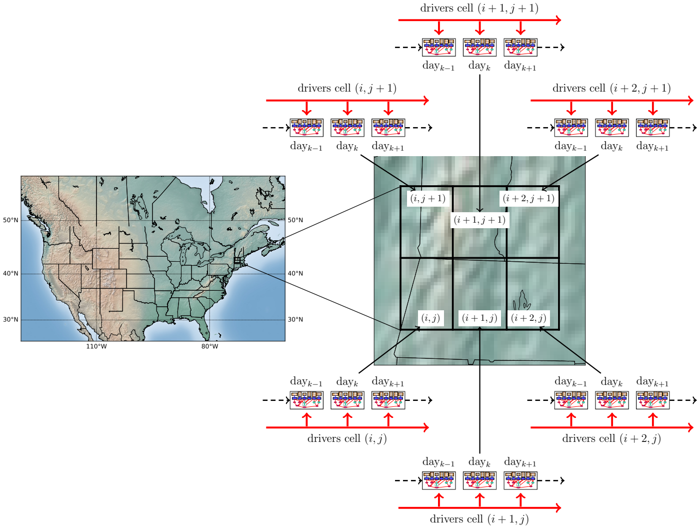
Conclusions

- Physics-based NN architecture outperforms traditional NN methods and PCE
- Using Tree-Structure LSTM gate can train on the noisy daily data
- Qualitatively similar sensitivity results compared to PCE

Future Work

- Employ the surrogate for calibration and optimal experimental design
- Analyze multiple sites
- Spatio-temporal physics-based RNN
- Incorporate low-rank tensor structure into RNN

ELM Simulation Details



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

$S_i = \frac{\mathrm{Var}_{X_i}(E(Y|X_i))}{Var(Y)}$ 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

Surrogate Modeling:Tree LSTM

LSTM Equations

$$
\begin{aligned}
f_{jk} &= \sigma(W_f x_j + \sum U_{kl} h_{jl}) \\
i_j &= \sigma(W_i x_j + \sum U_l h_{jl}) \\
o_j &= \sigma(W_o x_j + U_l h_{jl}) \\
u_j &= \tanh(W_c x_j + U_l h_{jl}) \\
c_j &= \sum f_{jl} \odot c_{jl} + i_j \odot u_j \\
h_j &= \sigma_j \odot \tanh(c_j)
\end{aligned}
\tag{1}
$$