Extreme Learning Machines for Variance-Based Global Sensitivity Analysis

John Darges, Alen Alexanderian, and Pierre Gremaud

Department of Mathematics, North Carolina State University

Abstract

Variance-based global sensitivity analysis (GSA) can provide a wealth of information when applied to complex models. A well-known Achilles' heel of this approach is its computational cost which often renders it unfeasible in practice. An appealing alternative is to analyze instead the sensitivity of a surrogate model with the goal of lowering computational costs while maintaining sufficient accuracy. Should a surrogate be "simple" enough to be amenable to the analytical calculations of its Sobol' indices, the cost of GSA is essentially reduced to the construction of the surrogate. We propose a new class of sparse weight Extreme Learning Machines (SW-ELMs) which, when considered as surrogates in the context of GSA, admit analytical formulas for their Sobol' indices and, unlike the standard ELMs, yield accurate approximations of these indices. The effectiveness of this approach is illustrated through both traditional benchmarks in the field and on a chemical reaction network.

Background

Motivations and Goals

1. Consider model $f : [0,1]^d \to \mathbb{R}$ with uniformly distributed, independent input $\mathbf{x}$
2. Sobol' indices are invaluable tools for GSA:

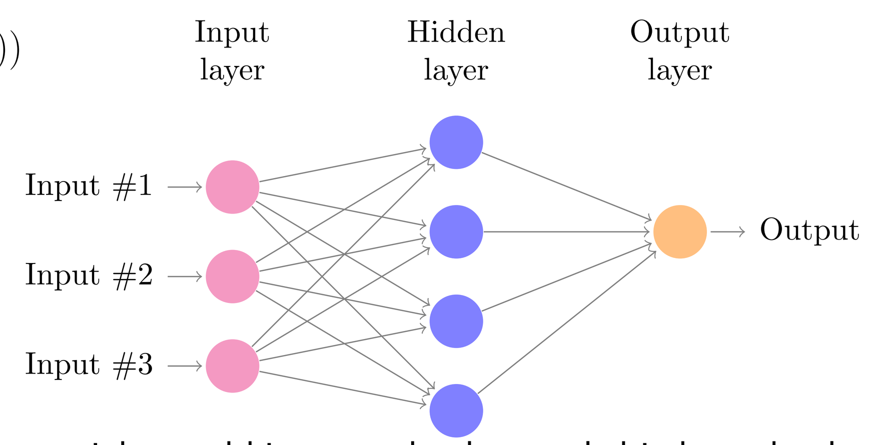
$$S_k = \frac{\mathrm{var}[f_k(x_k)]}{\mathrm{var}[f(\mathbf{x})]}, \quad S_k^{\mathrm{tot}} = 1 - \frac{\mathrm{var}[f_{-k}(\mathbf{x}_{-k})]}{\mathrm{var}[f(\mathbf{x})]}$$

3. Approximation using Monte Carlo (MC) methods is **intractable when f is expensive to evaluate**
4. Use surrogates with analytically known Sobol' indices (e.g. polynomial chaos, Gaussian processes) to avoid sampling
5. Can neural networks (NN) work as surrogates with analytic formulas for Sobol' indices?

Extreme learning machines

ELM has the form $\hat{f}(\mathbf{x}) = \beta^T(\phi(\mathbf{W}\mathbf{x}+\mathbf{b}))$

1. $\mathbf{W}$ - inner layer weight matrix
2. $\mathbf{b}$ - inner layer biases
3. β - output weights
4. ϕ activation function (acts component-wise)



Extreme learning machines (ELM): hidden layer weights and biases randomly sampled independently, no output bias used. Only need to train output weights β via linear least squares. For M training points:

1. Sample $\mathbf{W}$ and $\mathbf{b}$ (e.g. from standard normal distribution)
2. Assemble $\mathbf{H}_{i,*} = ((\phi(\mathbf{W}\mathbf{x}_i+\mathbf{b}))^T$ and $\mathbf{y} = (y_1 \dots y_M)^T$
3. Solve the L_2 regularized least squares problem

$$\arg\min_{\beta} \frac{1}{2}\|\mathbf{H}\beta - \mathbf{y}\|_2^2 + \frac{\alpha}{2}\|\beta\|_2^2$$

4. Regularization parameter α determined by L-curve method or generalized cross validation

Supported by NSF through awards DMS-1745654 and DMS 1953271

SIAM 2023 AMSTERDAM Conference on Computational Science and Engineering

Variance-Based GSA with ELMs

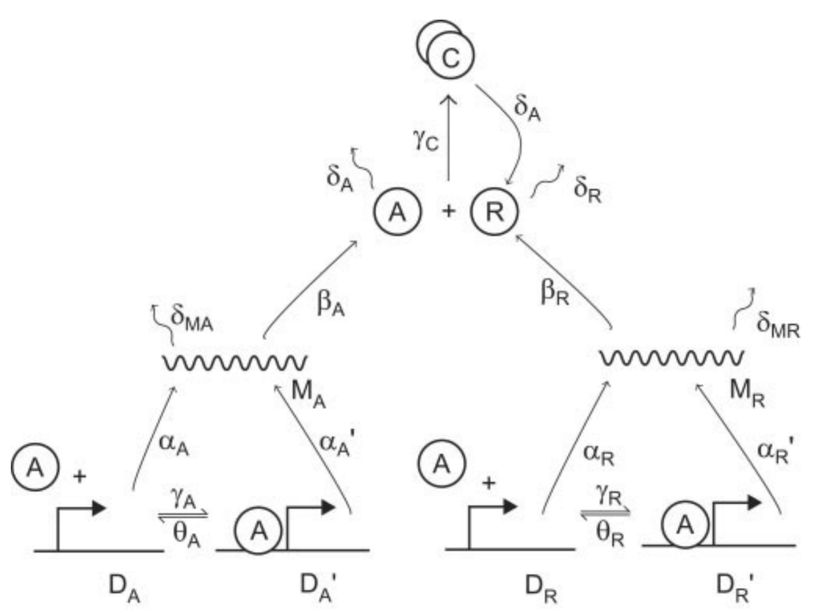
1. To get formulas for Sobol' indices, integrating surrogate should be easy
2. Common machine learning activation functions (e.g. sigmoid, ReLU) do not make integration easy
3. From theory, activation function can be any smooth non-polynomial function

1. Why not use $\phi(t) = e^t$ to make integration as easy as possible?
2. Setting $\phi(t) = e^t$, we can derive analytic formulas for Sobol' indices in terms of $\mathbf{b}$, $\mathbf{W}$, and β
3. After training ELM, we can obtain Sobol' indices for free:

$$S(\hat{f}) = S(\mathbf{b}, \mathbf{W}, \beta), \quad S^{\mathrm{tot}}(\hat{f}) = S^{\mathrm{tot}}(\mathbf{b}, \mathbf{W}, \beta)$$

Numerical Experiment

Application to Genetic Oscillator



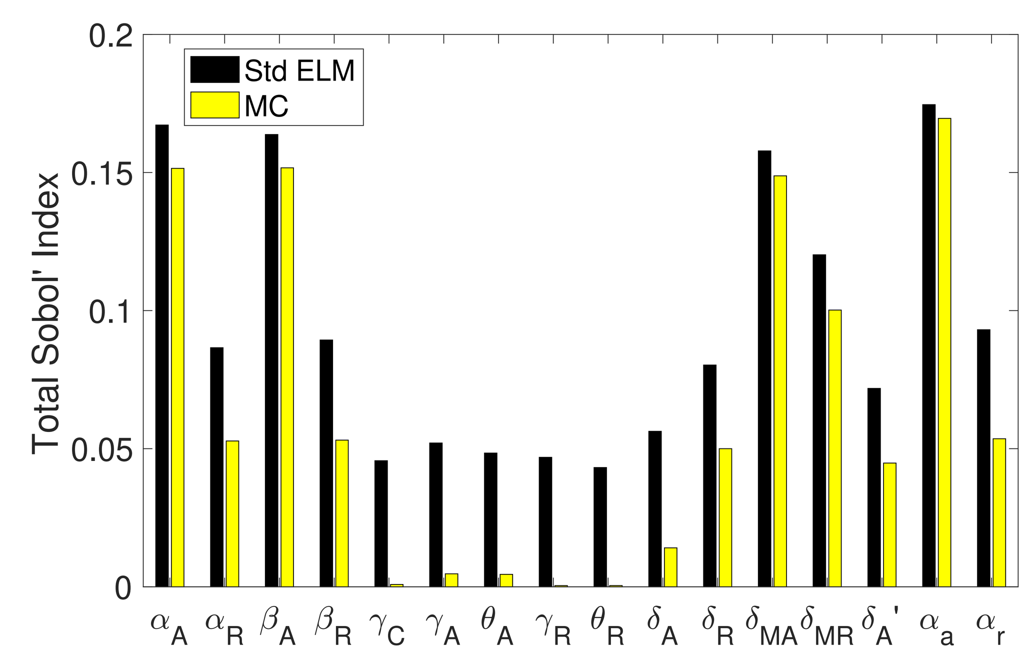
1. Biochemical model describing molecular species involved in circadian rhythm regulation
2. Model is stiff ODE system (expensive to solve)
3. Uncertain parameters $\mathbf{x}$ are 16 reaction rate parameters
4. Each parameter uniformly distributed in interval given by $\pm 5\%$ of respective nominal value
5. We study the average concentration in time of species R as our QoI f

$$f(\mathbf{x}) = \frac{1}{T}\int_0^T R(t;\mathbf{x})\,dt$$

Initial GSA

1. 3000 training size
2. Points sampled via Latin hypercube sampling
3. 1000 hidden layer neurons

1. Regularization parameter $\alpha = 10^{-4}$ from L-curve method
2. Compare to indices computed using 10^5 Monte Carlo samples



ELM surrogate overestimates total indices and underestimates first order indices

Sparse-Weight ELMs

Sparsification

1. **Issue:** ELM may overestimate the influence of higher order ANOVA terms
2. **Idea:** We can reduce influence of higher order terms by making inner weight matrix sparse

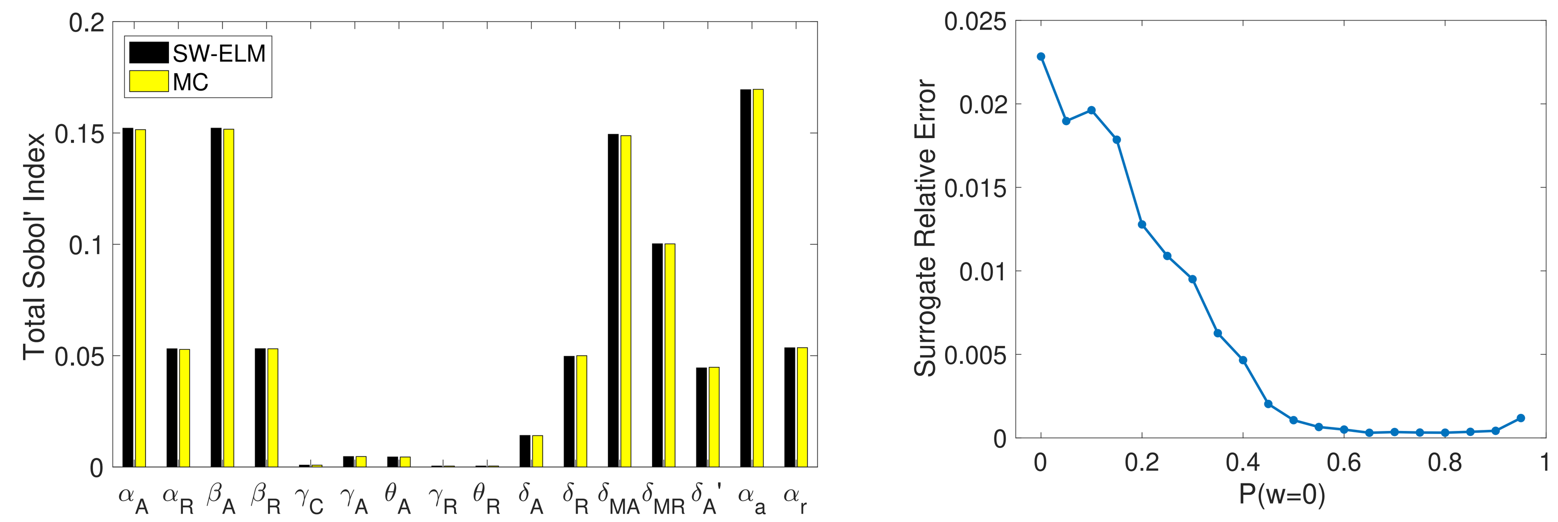
1. Sparse weight matrix $\mathbf{W}_s = \mathbf{B} \circ \mathbf{W}$, $\circ$ denotes component-wise multiplication
2. $\mathbf{B}$ randomly sparsifies entries of weight matrix
3. How do pick what p to use?

$$B_{ij} = \begin{cases} 0 \text{ with probability } p, \\ 1 \text{ with probability } 1-p \end{cases},$$

Proposed Method

1. Training ELMs is cheap
2. Create different SW-ELMs for different choices of p
3. Use SW-ELM with smallest approximation error for GSA
4. **Main Idea:** If sparsifying gives a better surrogate, it should give better Sobol' indices

GSA with SW-ELM



SW-ELM gives **better** GSA with **fewer** training points

Future Directions

1. Strategies for improving sparsification method
2. Connect the "optimal" sparsity of the weight matrix to total Sobol' indices
3. Can sensitivity analysis inform neural network architecture?

References

[1] J. Darges, A. Alexanderian, and P. Gremaud. Extreme learning machines for variance-based global sensitivity analysis [preprint]. 2022.

[2] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006.

[3] M. Merritt, A. Alexanderian, and P. A. Gremaud. Multiscale Global Sensitivity Analysis for Stochastic Chemical Systems. *Multiscale Modeling & Simulation*, 19(1): 440–459, 2021.

[4] J. M. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9): 5988–5992, 2002.