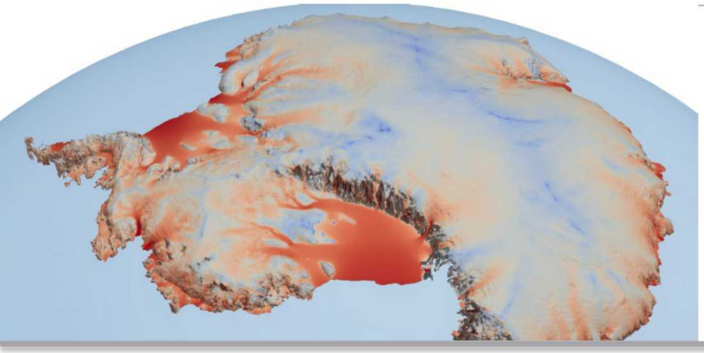
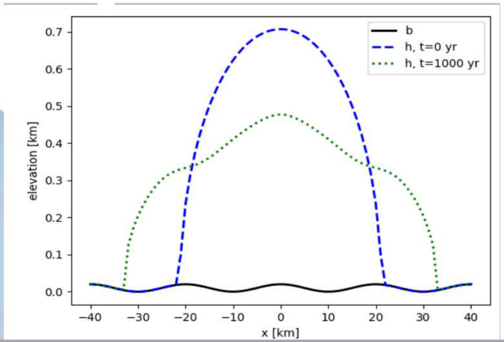
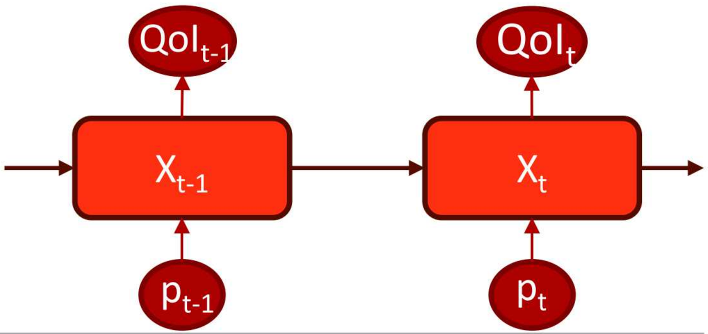
SAND2019-2255C

Exceptional service in the national interest

Sandia National Laboratories







Neural Networks Surrogates of PDE-based Dynamical Systems: Application to Ice Sheet Dynamics

John Jakeman, Mauro Perego, William Severa (SNL)

Lars Ruthotto (Emory Universty)

CCR Center for Computing Research

LDRD Laboratory Directed Research and Development

QTM, Emory

DOE ASCR PhILMs

U.S. DEPARTMENT OF ENERGY

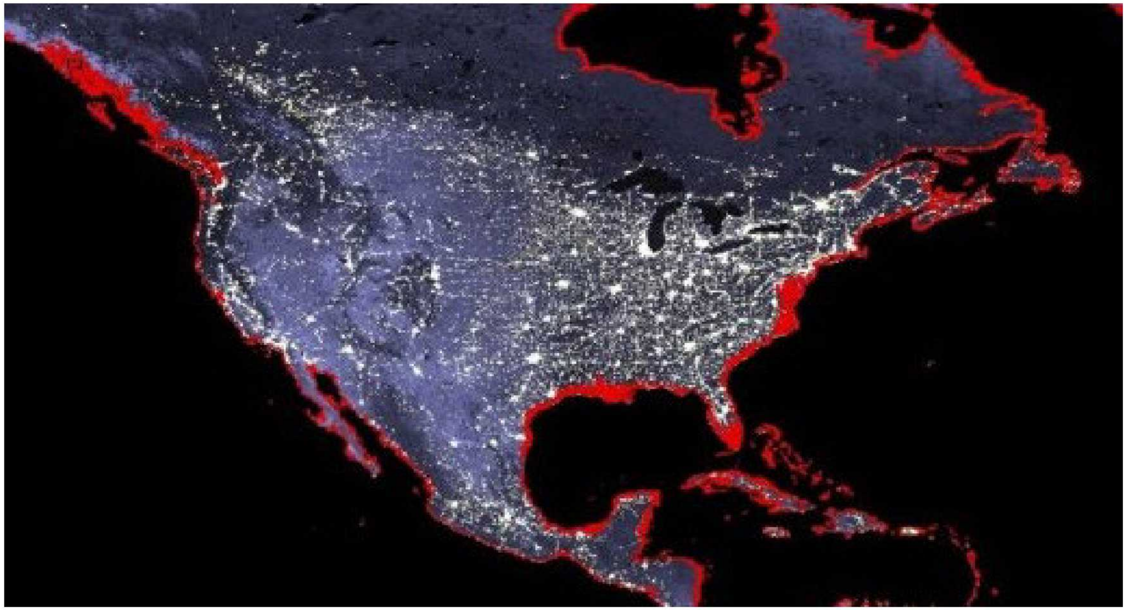
NNSA National Nuclear Security Administration

Siam CSE, Spokane (WA), Feb 27, 2019

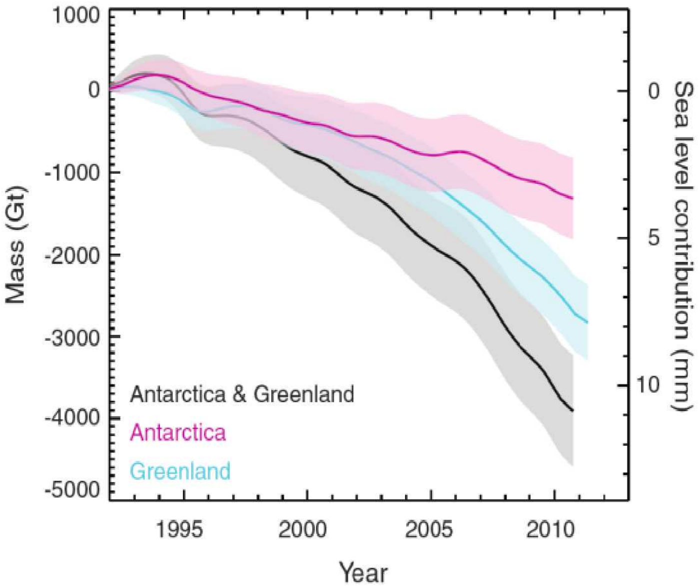
Motivation and Introduction

- **Sea-level change** is one of the most impactful consequences of climate change
- Greenland and Antarctica ice sheet are major contributors to the sea level*
- Global mean sea-level is rising at the rate of 3.2 mm/yr and the rate is increasing.
- Latest studies suggest possible increase of 0.3 – 2.5m by 2100
- Accurate probabilistic projections of sea level would be extremely useful to policy makers

Map with 6 meters sea-level rise in red (NASA).



total mass loss of ice sheets in 1992-2011 (sheperd et al. 2012)

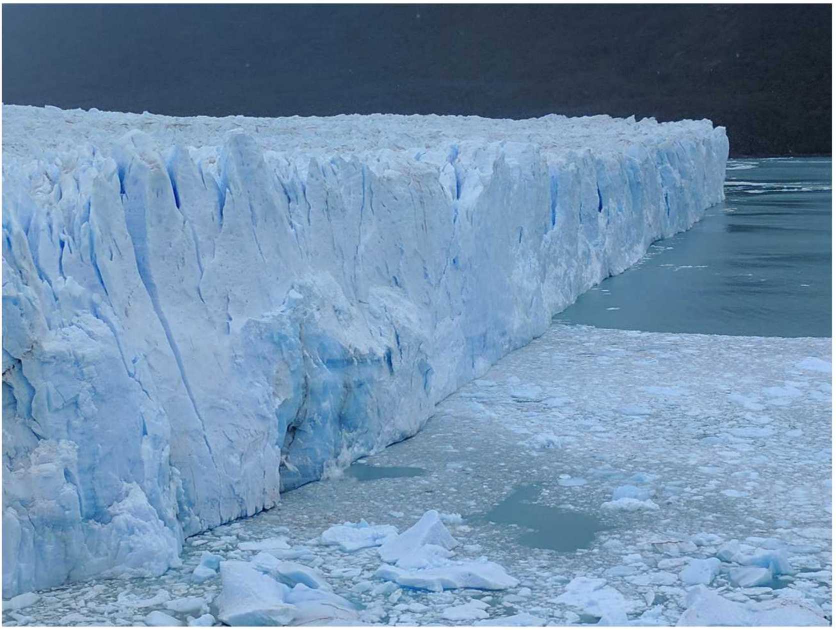


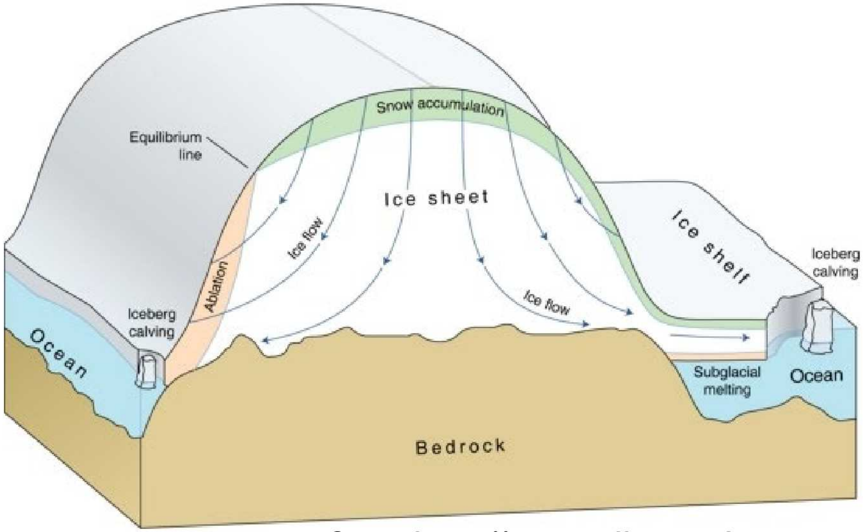
*DOE SciDAC project **ProSPect** (**Pro**babilistic **S**ea Level **P**roj**ect**ion from Ice Sheet and Earth System Models), Institutes: LANL, LBNL, SNL, ONL, NYU, UM

Motivation and Introduction

- Ice behaves like a very viscous shear-thinning fluid (similar to lava flow) driven by gravity. Source: snow packing/water freezing. Sink: ice melting / calving in ocean.
- There are several unknown or poorly known parmeters (e.g. basal friction, bed topography,rehogy)
- Simulating Ice sheet dynamics requires the solution of complex large scale computational models of Greenland and ice sheet

Perito Moreno glacier

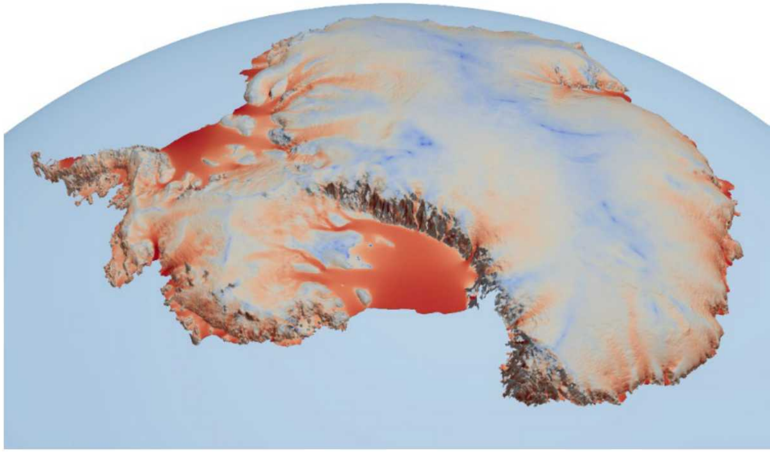




from http://www.climate.be

Uncertainty Quantification challenge

The total mass change of ice sheets is a proxy for the sea-level change and it is our Quantity of Interest (QoI) .
Despite the considered Quantity of Interest (QoI) is a **scalar** quantity, its computation requires the solution of complex multi-physics systems of partial differential equations, characterized by hundreds of million of unknowns and **a large number of parameters**



Accurate UQ analysis of Greenland and Antarctic ice sheets at high resolution is currently **unfeasible** due to the high-dimensional parameter space (curse of dimensionality) and the cost of running the physical model

Parameters:
basal friction
geothermal flux
precipitations

→

Expensive Computational model
Complex system of **PDEs**
O(10^8) unknowns
O(10^4) time-steps

→

scalar QoI:
ice sheets total mass change

Uncertainty Quantification, NN surrogates

Possible strategy: model reduction in *multifidelity** framework

Parameters:
basal friction
geothermal flux
precipitations

→ **Inexpensive** surrogate model →

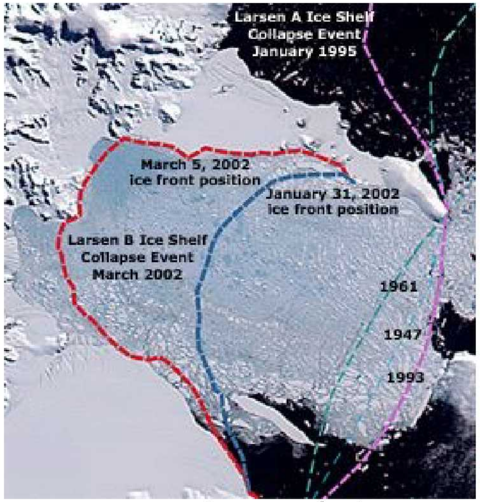
scalar QoI:
ice sheets total mass change

(Surrogate model is evaluated more often than the physical model to reduce costs while maintaining accuracy)

Issue: classic approaches like PCE are still too expensive especially in presence of nonlinear maps with interconnected parameter dimensions.

Idea: create surrogate models using Neural Networks (**NN**) **trained by model output** at different time instants

* *Peherstorfer, Willcox and Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Review, 2018*



surrogate needs to capture instabilities

Problem setting and methods

Problem setting:

- *Ice sheet model*: Shallow Ice Approximation

$$\boxed{h_t - \nabla \cdot (\mu(h,\beta)\, \nabla h) = f}$$

- QoI: total mass change in time

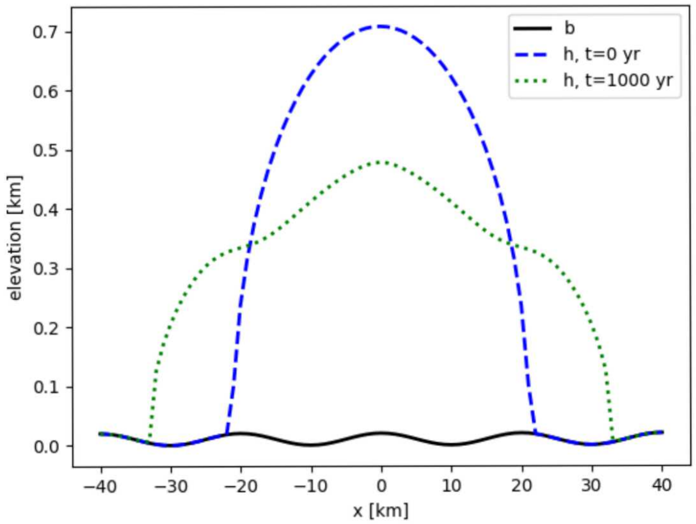
- *Parameters*: basal friction β, represented as a Karhunen–Loève Expansion (KLE) based on 20/100/500 independent uniformly distributed parameters:

$$\log(\beta(x,\omega)) = \log(\,\beta_0(x)) + \sigma \sum_{k=1}^{n_\omega} \sqrt{\lambda_k}\ \phi_k(x)\, \omega_k$$

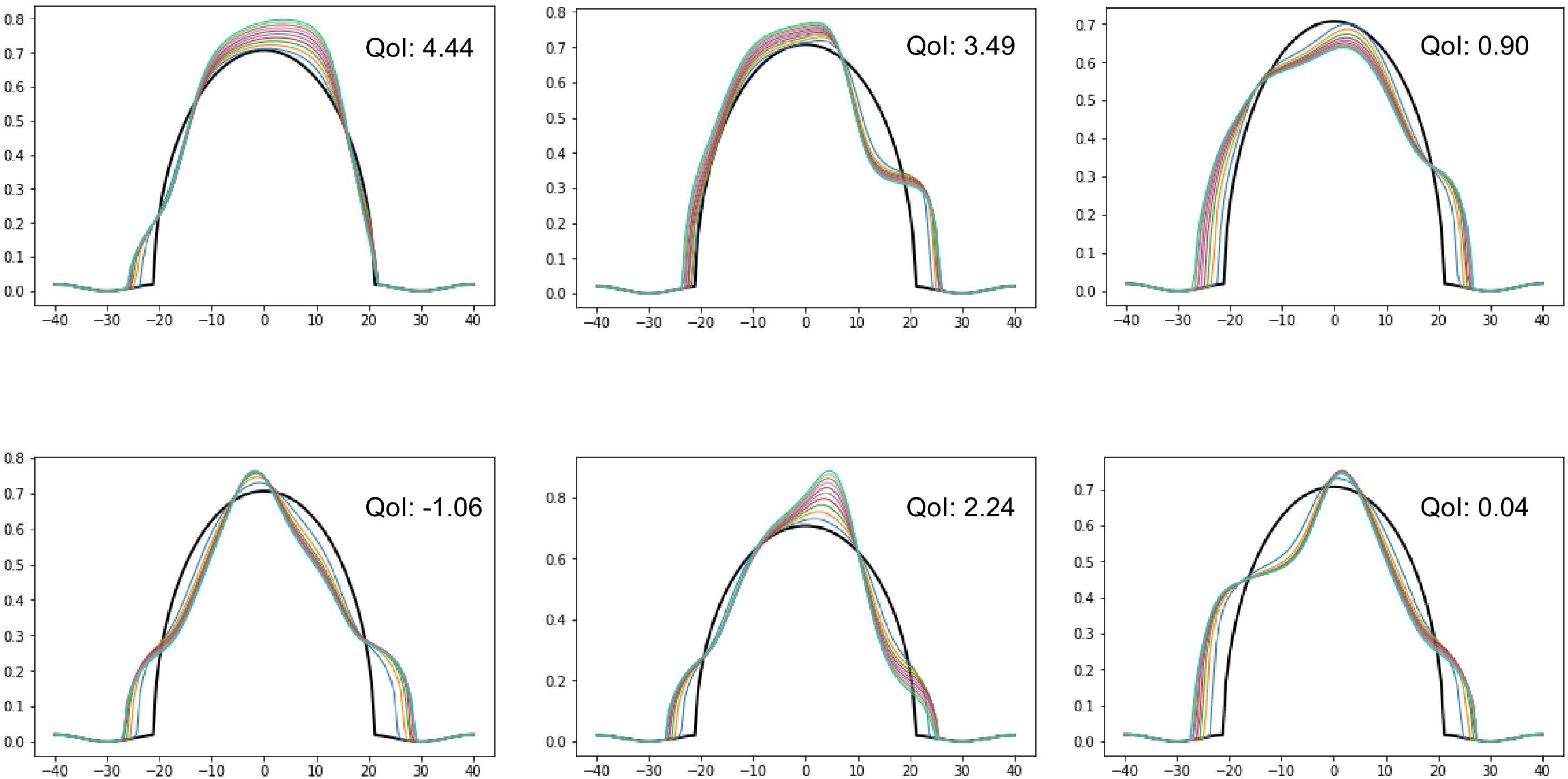
λ_k, ϕ_k are the eigenvalues and eigenvectors of $C = \exp\left(\frac{|x_1 - x_2|}{l}\right)$

- *Forcing*: ice accumulation/melt, sinusoidal in space/time

- Training set: 900 samples of N time instants
- Testing set: 100 sampls of N time instants

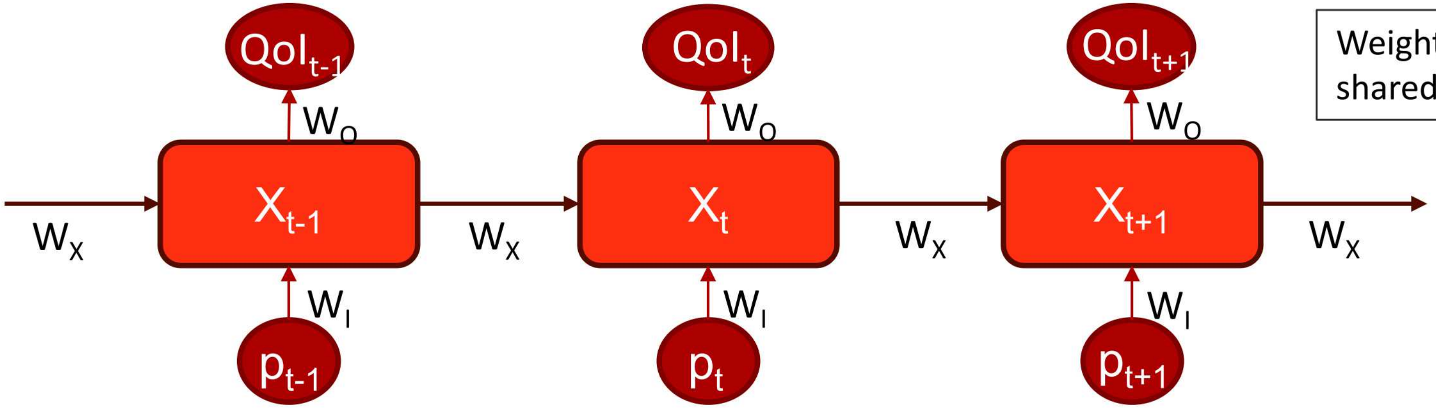


Ice sheet evolutions for different parameter samples



Ice sheet evolution for different samplings of β (20 parameters case).
Time instants [yr]: 0,50,100,…,500.

Recurrent Neural Network (RNN) surrogates



Weights W_X, W_O and W_I are shared by all units.

Why RNNs?

1. RNNs have been *effectively used to model dynamical systems**, and can provide prediction of the QoI at *different time instants*, whereas PCE would only provide the QoI at a given instant
2. Because different networks units share the same parameters, RNNs are relatively *fast to train*, especially given the low dimensionality of the QoI

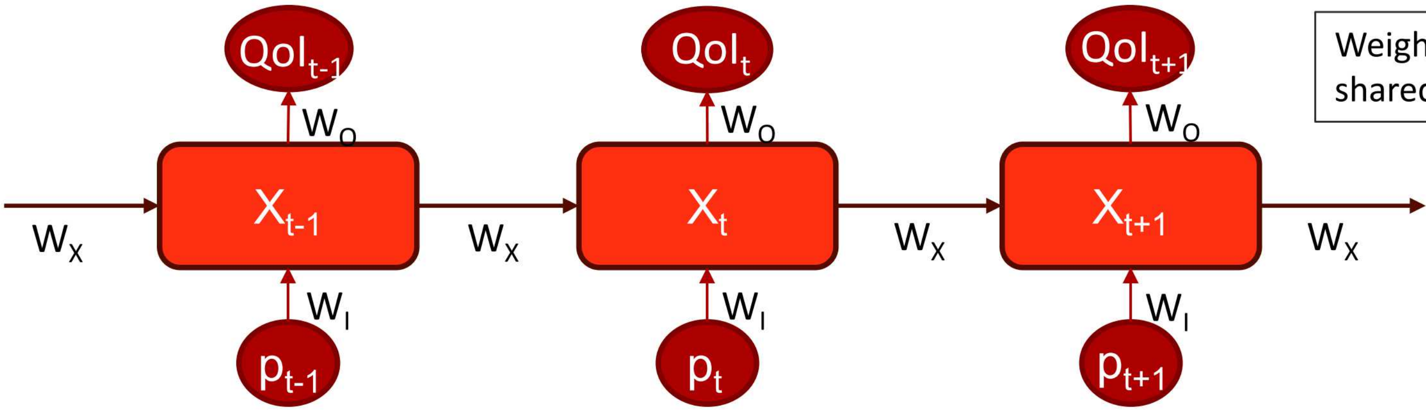
Implementation:

- Long-Short Term Memory (LSTM) networks** as implemented in Keras (built on top of TensorFlow)
- To better exploit the time dimension, we perform windowing (whether to split the temporal data in chunks of consecutive samples).

* Pathak, Hunt, Girvan and Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. Phys. Rev. Lett., 120, 2018

** S. Hochreiter, J. Schmidhuber, "Long short-term memory". Neural Computation. 9 (8): 1735–1780, 1997.

Recurrent Neural Network (RNN) surrogates



Weights W_X, W_O and W_I are shared by all units.

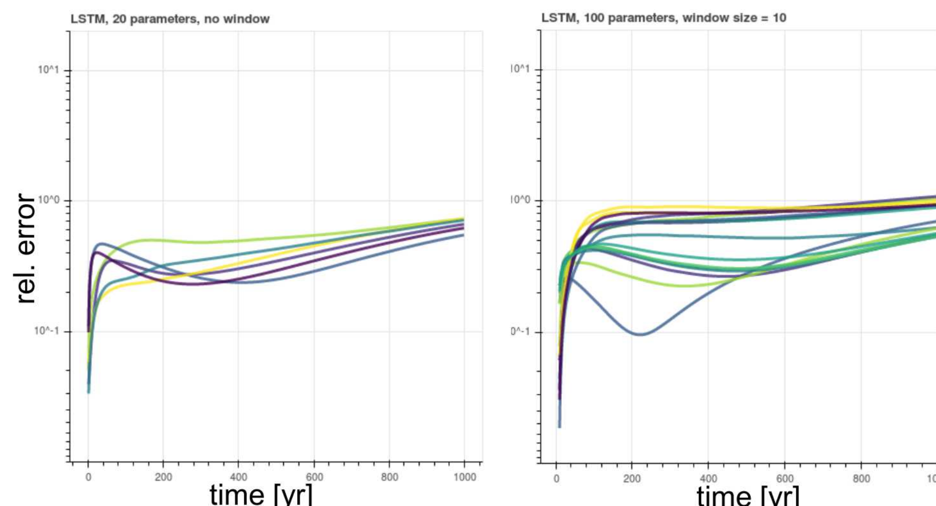
Results:

relative error: $$e_i = \frac{|QoI_{t_i} - \widetilde{QoI}_{t_i}|_{l_2}}{\max_i |QoI_{t_i}|_{l_2}}$$

Relative errors as a function of time
for different hyper-parameters
Left: window of size 10 [yr], 100 parameters
Right: no window, 20 parameters

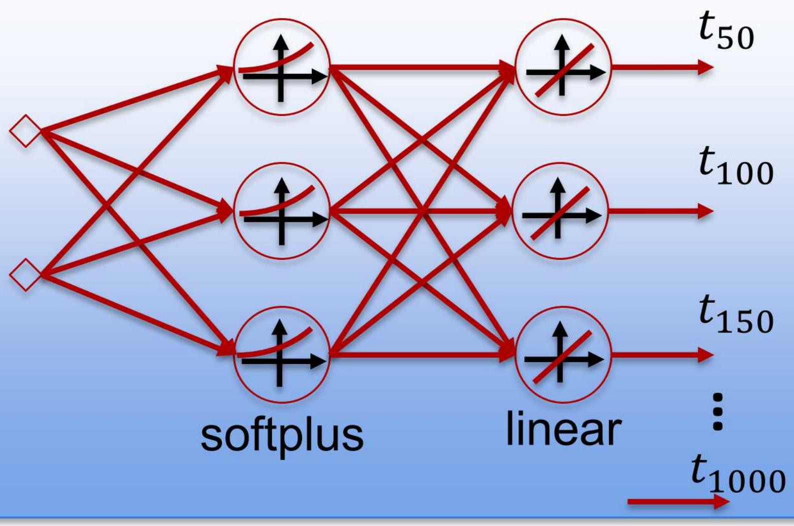
Most of networks have a relative error that is above 50%, which is insufficient for performing UQ

Simulations by W. Severa

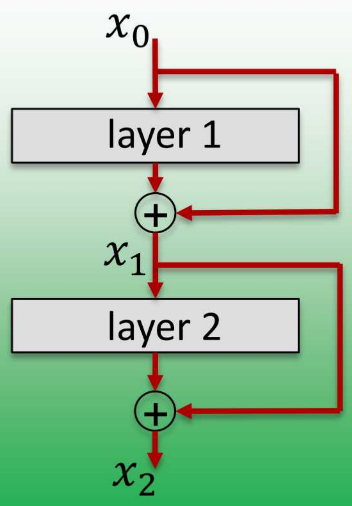


Non recurrent approaches: MLP, ResNet, PCE

Multi-Layers Perceptron (MLP)



Residual Network (ResNet)



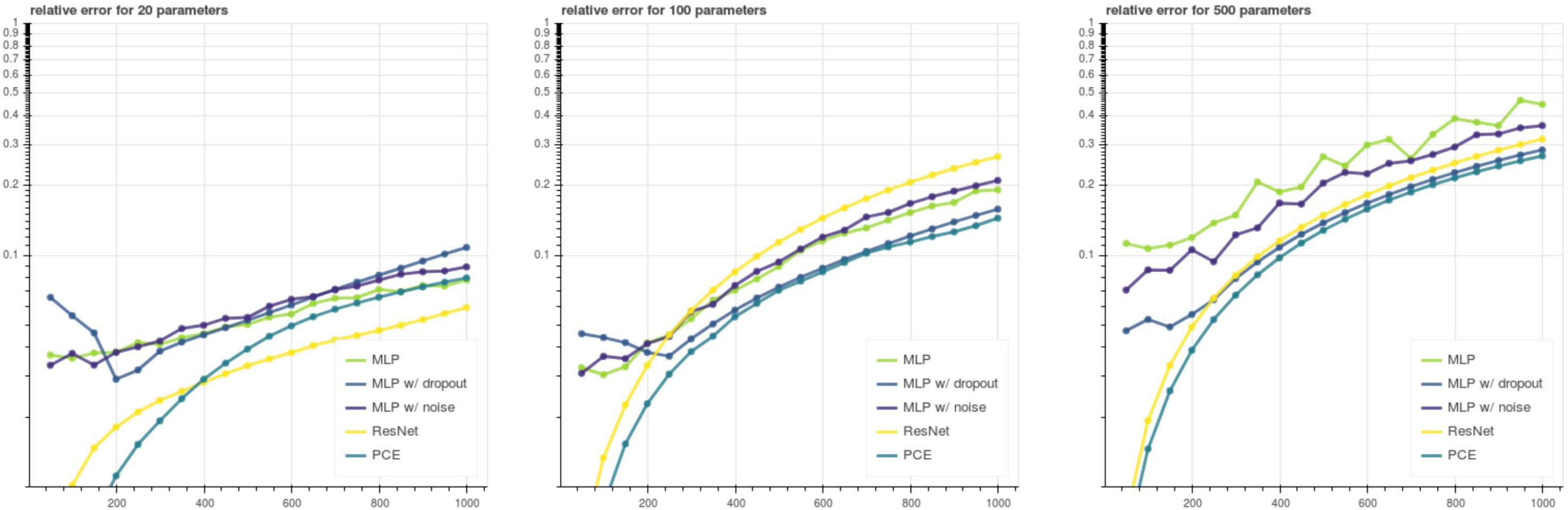
Polynomial Chaos Expansions (PCE)

$$\widetilde{QoI} = \sum_k \alpha_k \, \psi_k$$

ψ_k orthogonal polynomials in the parameter space

$$\underset{\alpha}{\operatorname{argmin}} \ \|\alpha\|_1 \quad \text{such that} \quad \left\| QoI - \sum_k \alpha_k \, \psi_k \right\|_2 < \varepsilon$$

Results: comparing MLP, ResNet and PCE



Comparison of relative errors of MLP, ResNet and PCE surrogates, for 20 (left), 100 (center) and 500 (right) parameters.

Details:

- MLP: Keras, single hidden layer of width 50, opt scheme: Adadelta, dropout: 0.3
- ResNet*: Matlab code by L. Ruthotto, 4 hidden layers of width 8, stepping scheme: RK4
- PCE*: python code by J. Jakeman, polynomial degree: 2/2/1, for 20/100/500 params

*E. Haber, L. Ruthotto, *Stable architecturs for deep neural networks*, Inverse Problems, 2017

** J.D. Jakeman et al., *Enhancing l1-minimization estimates of polynomial chaos expansions*, JCP, 2015

Conclusions
(is the glass half empty or half full?)

- the LSTM seem to lack the capability of capturing the global behavior of a dynamical system. They are also much more expensive than the other approach tried (ResNet, MLP, PCE)
- Compared to approaches like PCE, NNs have many more “knobs” (hyper-parameters) that are hard to tune for a specific application to get good performance
- We were able to devise MLP networks and Residual networks that were comparable to PCEs in term of accuracy
- Further directions include using NN to approximate the ice elevation at each time step, rather than the QoI and enforce mass conservation constrains

More general considerations:

- **Can we do drastically better than PCE?**
- **Is the parameter to prediction map intrinsically high-dimensional?**