REDUCED ORDER TECHNIQUES FOR TERRESTRIAL MODELS

Presented by Zack Subin (LBL)
Led by George Pau (LBL)
Collaborators: Yaning Liu, Bill Riley, Gautam Bisht, Charlie Koven (LBL), Chaopeng Shen (PSU),…

BERKELEY LAB
Lawrence Berkeley National Laboratory

ESD
EARTH SCIENCES DIVISION

CESM Land Model Working Group, Boulder, CO Mar 3, 2015.

Motivation: many-query applications

Parameters $\mu \in \mathcal{D}$ → **Models** → **Output of interest** $f(\mu)$

Parameters:
- Geophysical properties
- Fluid properties
- Source/sink terms
- Boundary conditions

Models:
- Experimental fit
- Small conceptual models
- **High-resolution numerical models**

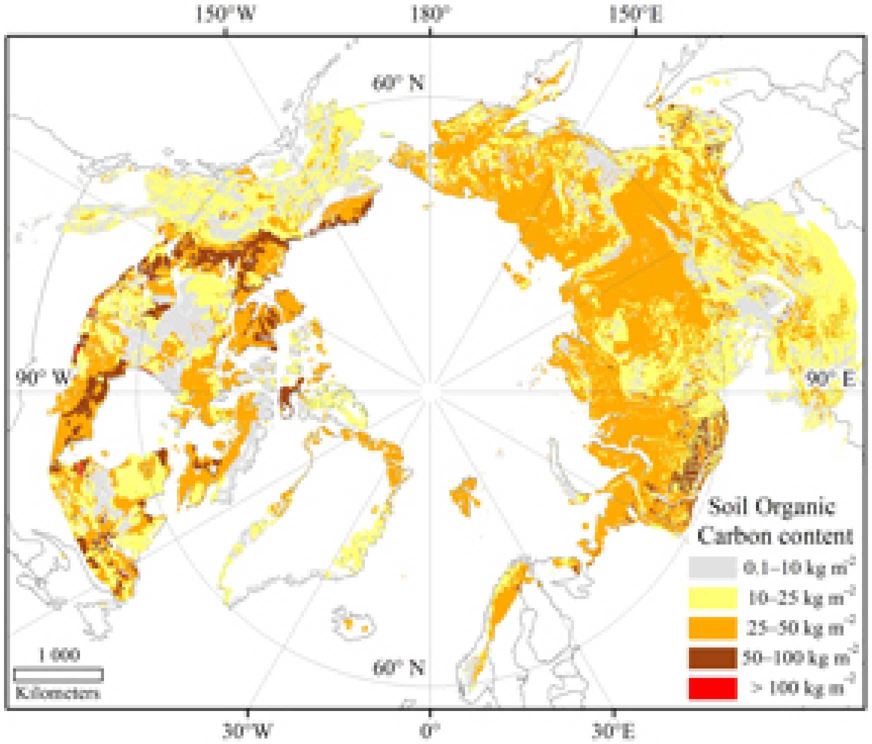
Output of interest:
- Saturation field
- Temperature field
- Change in water table
- Net carbon flux

Computationally expensive
to evaluate many times: needed for uncertainty quantification, data assimilation and sensitivity analysis.

Why do we need high-resolution models?

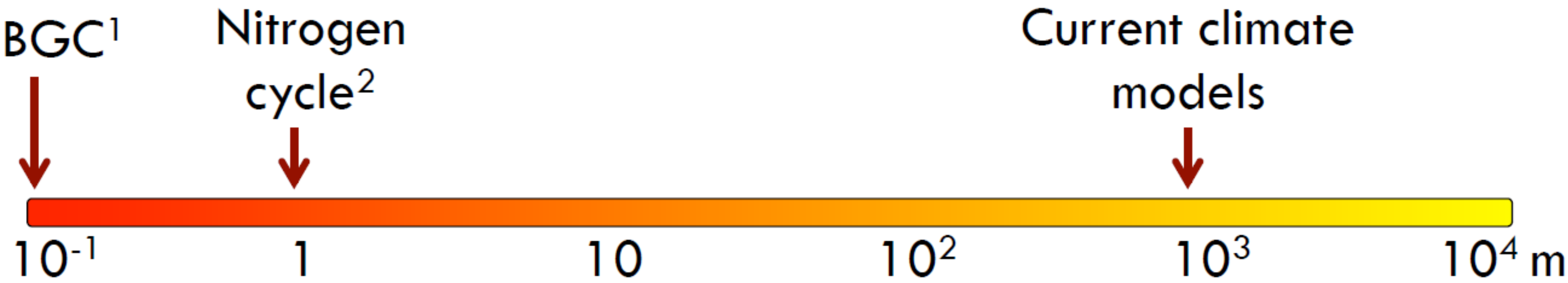
- (e.g.) High-latitude ecosystems (permafrost tundra and peatlands) illustrate nonlinear dependence on environmental state variables

Yedoma, from http://dgrnewsservice.org/



c/o Permafrost Carbon Network

Motivation: Multiscale simulations



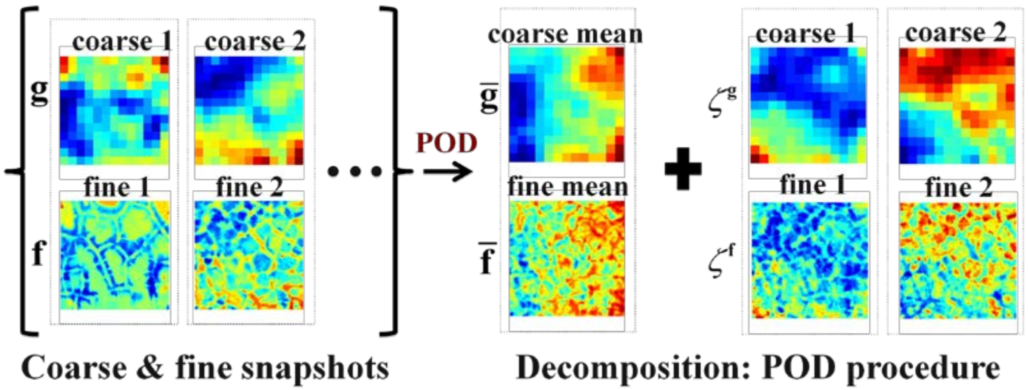
- Accurate description of BGC processes, e.g. methanogenesis, requires preservation of subgrid heterogeneity.
- A global model discretized at the BGC scale would be too computationally expensive.
- If we model processes on difference scales, how do we bridge the scale differences?

[1]Frei et al., J Geophys Res-Biogeo, 117, 2012. [2]McClain et al., Ecosystems, 6, 301-312, 2003.

Approach for Scaling: Reduced-Order Model (ROM)

1. Perform simulations with fine-scale, process model that sample the parameter space
   - PFLOTRAN (<1 m), CLM-PAWS (~100 m)
2. Train ROM: numerical surrogate
3. Couple ROM to large-scale model

Example: POD (Principle Orthogonal Decomp.) = EOF

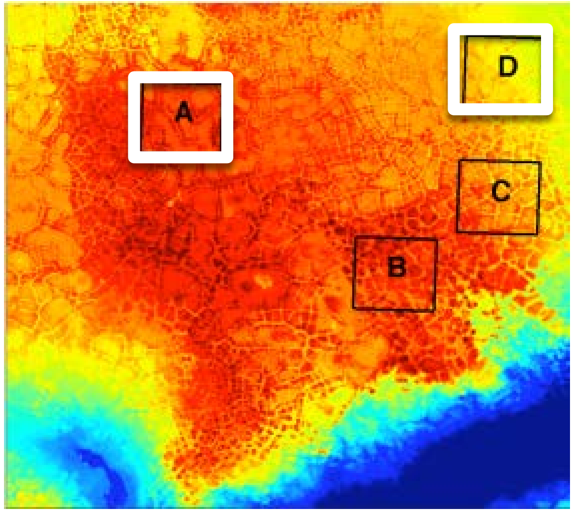


Large existing ROM literature in other fields…

Terrestrial Ecosystem Examples

- POD-MM applied to polygonal tundra and temperate watersheds
  - Maps coarse-grid solutions to fine-grid solutions
- GPOD-EIP applied to global soil carbon
  - Reproduces spatial field based on results on sparse sample
- POD-GPR applied to temperate watersheds
  - Develops functional relationship between inputs (environmental forcings and model parameters) and outputs (hydrology and BGC)

Barrow Polygonal Tundra Experimental Setup



LIDAR DEM of the sites.

- Simulation setup
  - Performed 5 years (summer) of surface-subsurface isothermal flow simulation using PFLOTRAN
  - Horizontal grid spacing: 0.25 to 8 m
  - Boundary conditions: Offline CLM simulations

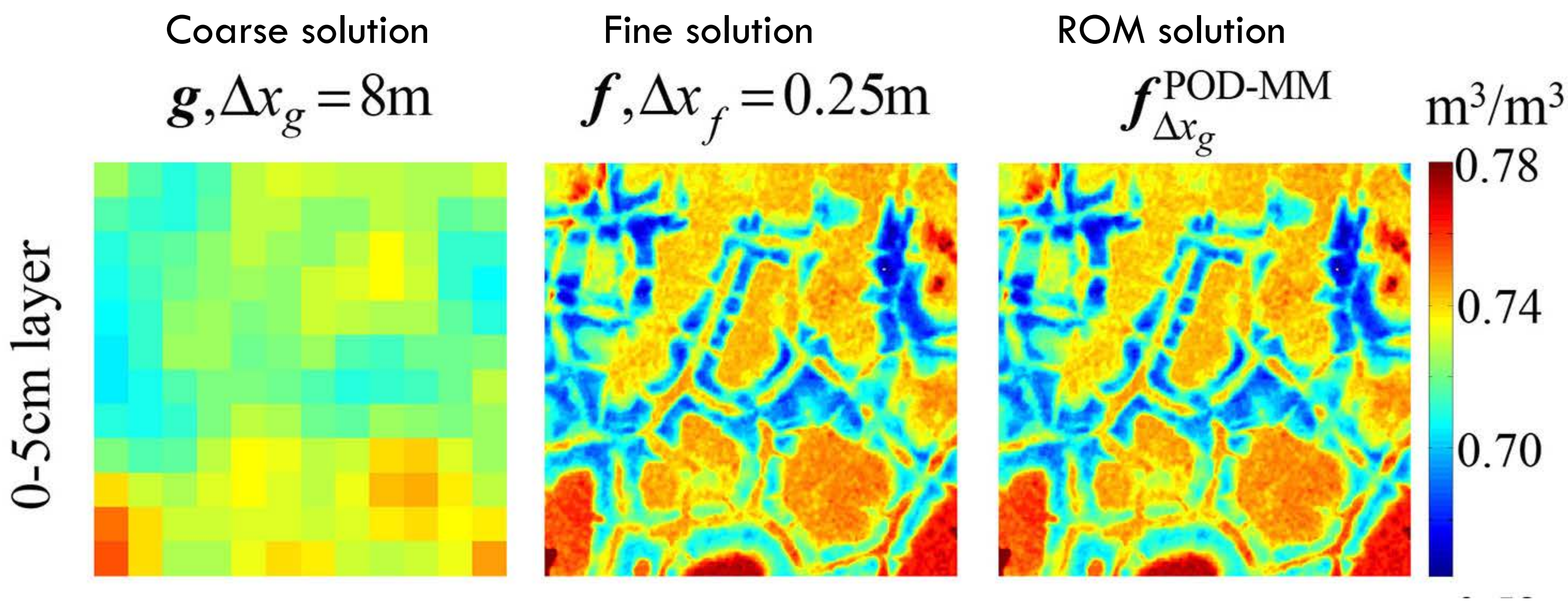
Principal Orthogonal Decomposition-Mapping Method

- POD-MM Setup
  - Trained using the first 3 years using soil moisture solutions.
  - Validated using soil moisture solutions from the last 2 years.
  - The 0.25 m grid space is the “truth” solution we want to reproduce.

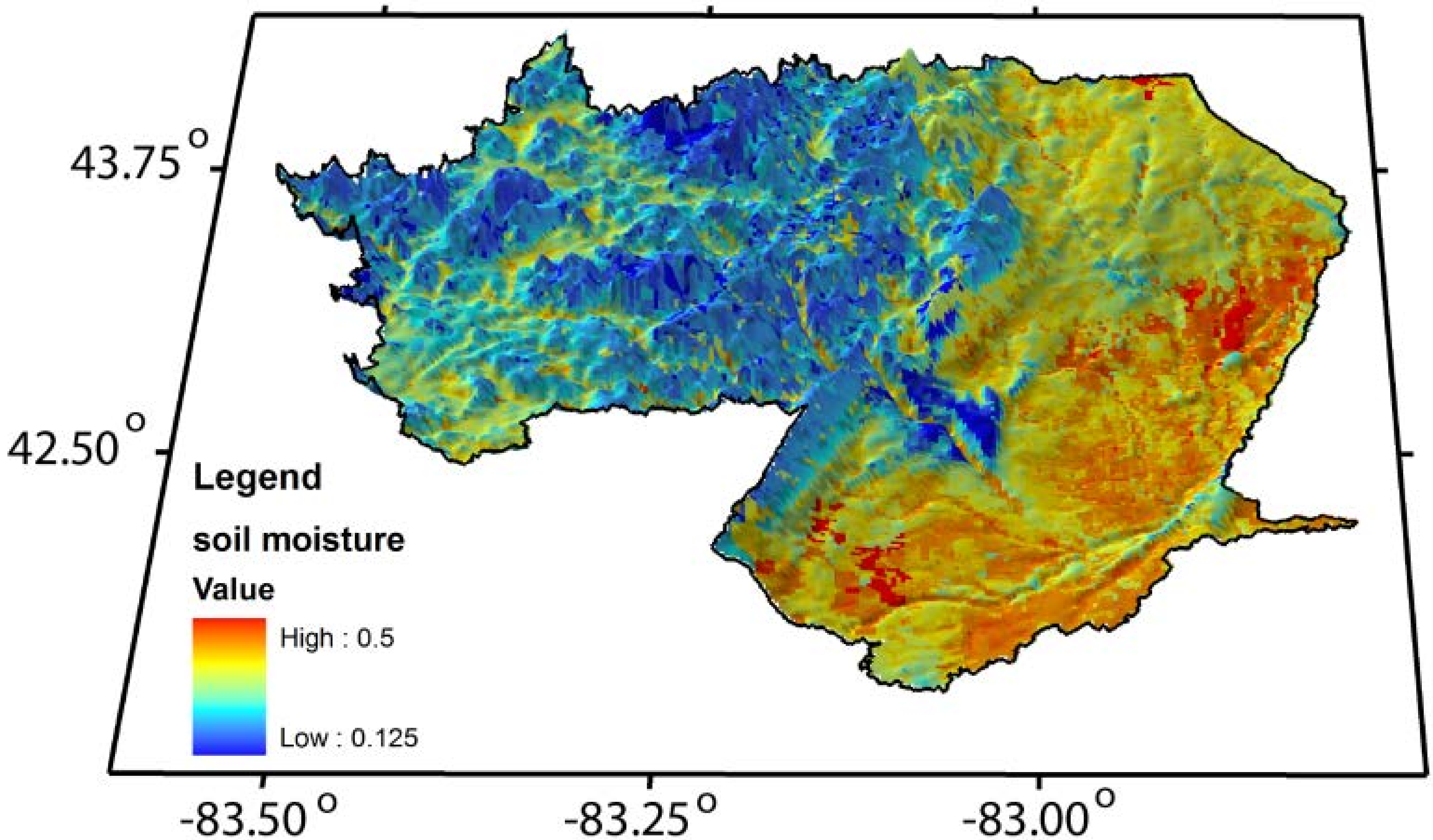
Results Summary

Pau et al. (2014, GMD)

- Reproduces soil moisture at 32-times finer resolution with coarse information
  - 1000 times computational speedup
  - Relative error ~ 10^{-4}



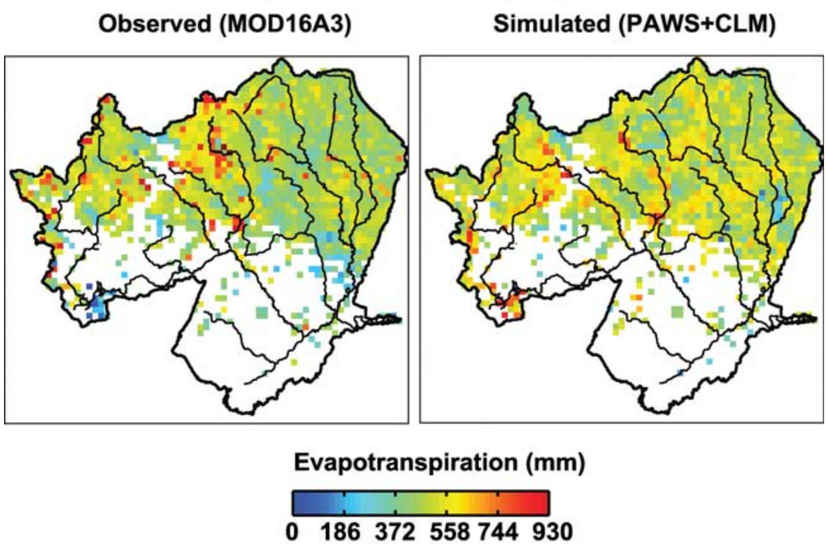
Clinton River (MI) Watershed



CLM4-PAWS simulation at 220 m (Riley and Shen, HESS 2014)

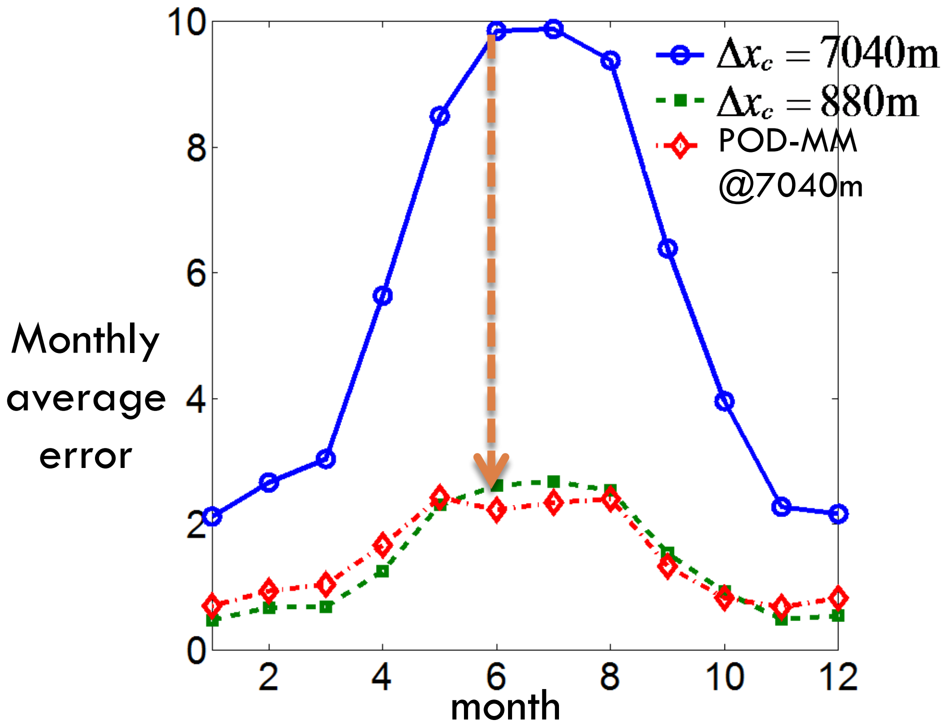
Clinton River Setup

- CLM4-PAWS
  - PAWS: Shen & Phanikumar (2010, WRR): a quasi-3-D saturated groundwater domain and 1-D Richards equation.
  - Resolutions ranging from 220 to 7040m.

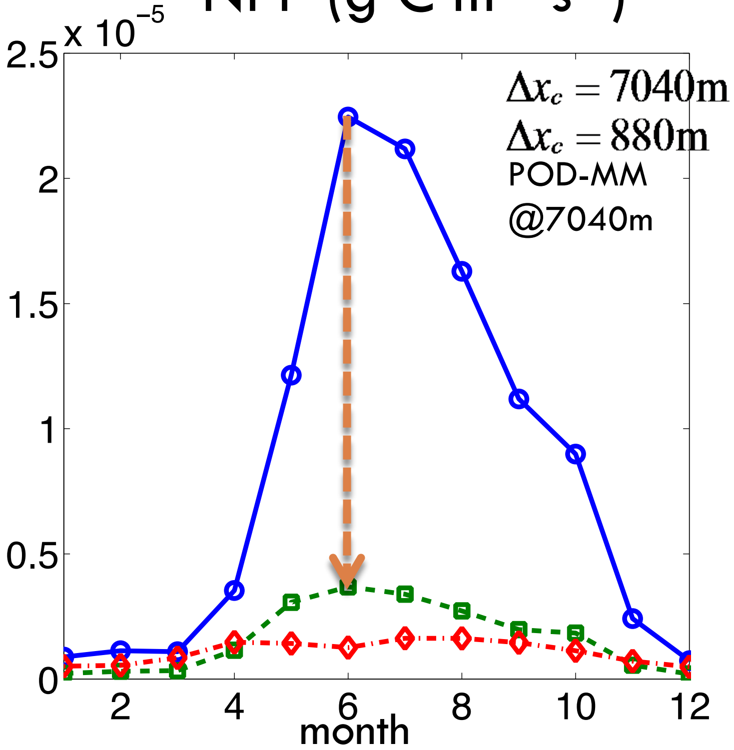


Results for Latent heat & NPP

Latent heat (W m^{-2})

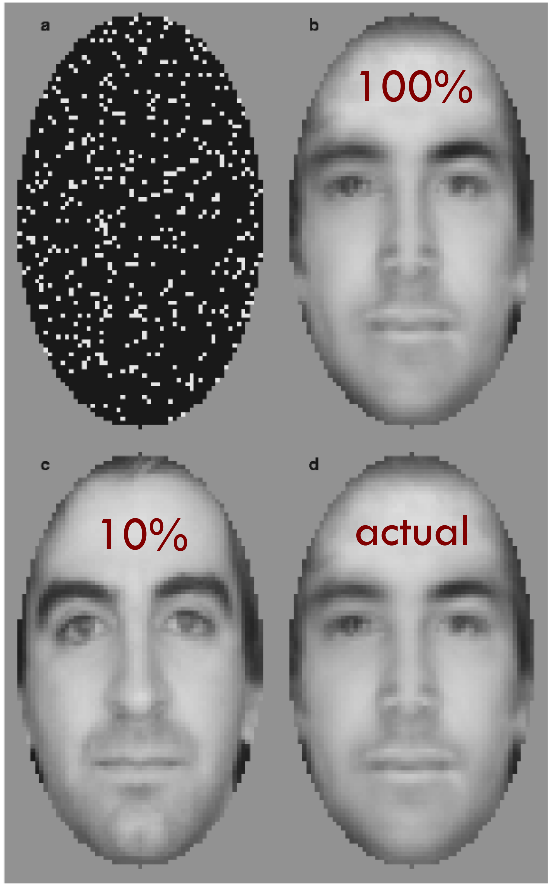


NPP (g C m^{-2} s^{-1})



Monthly absolute error on the coarse grid (using fine grid as truth) is significantly reduced.

Gappy POD (GPOD) for CLM4.5BGC



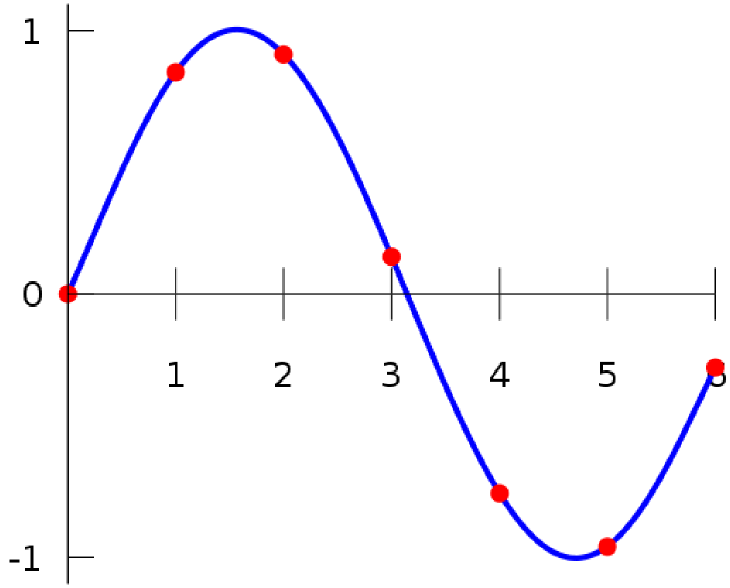
- Utilizes a subset of the fine-resolution solution to reconstruct the whole
  - Everson and Sirovich (1995, J Opt Soc Am)
  - Works well with embarrassingly-parallel models like CLM.
- Uncertainty Quantification:
  - Exploring full parameter space with CLM is expensive
  - Can we capture the characteristic behavior with only ~50-500 representative gridcells?
- Another application: faster spinup

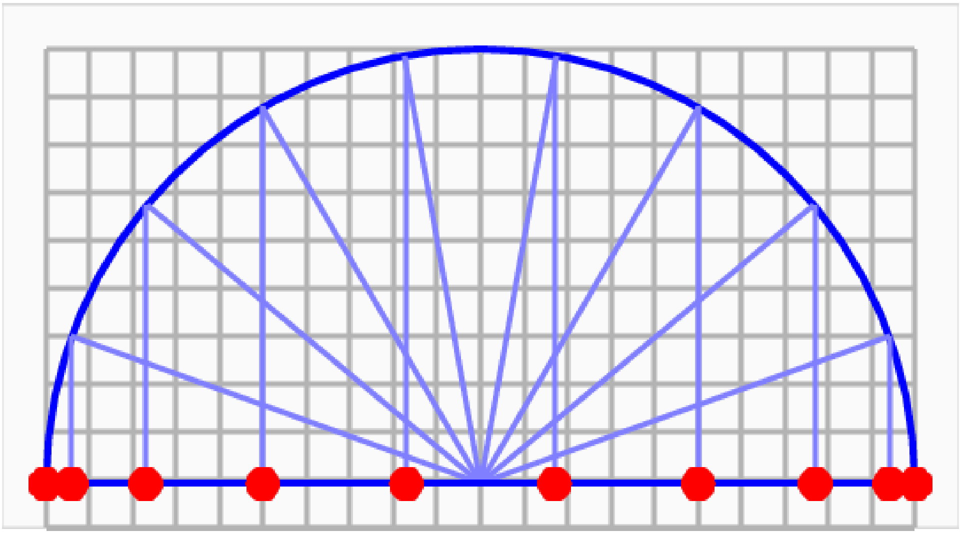
GPOD + EIP

What is a good (enough) subset of points to use in the GPOD procedure?

- Unlike the image-processing problem, we get to choose the subset of points to be representative.
- Finding optimal set is intractable so use a heuristic
  - EIP proved superior to several alternatives

EIP (Maday et al., Commun. Pur. Appl. Anal., 2009)



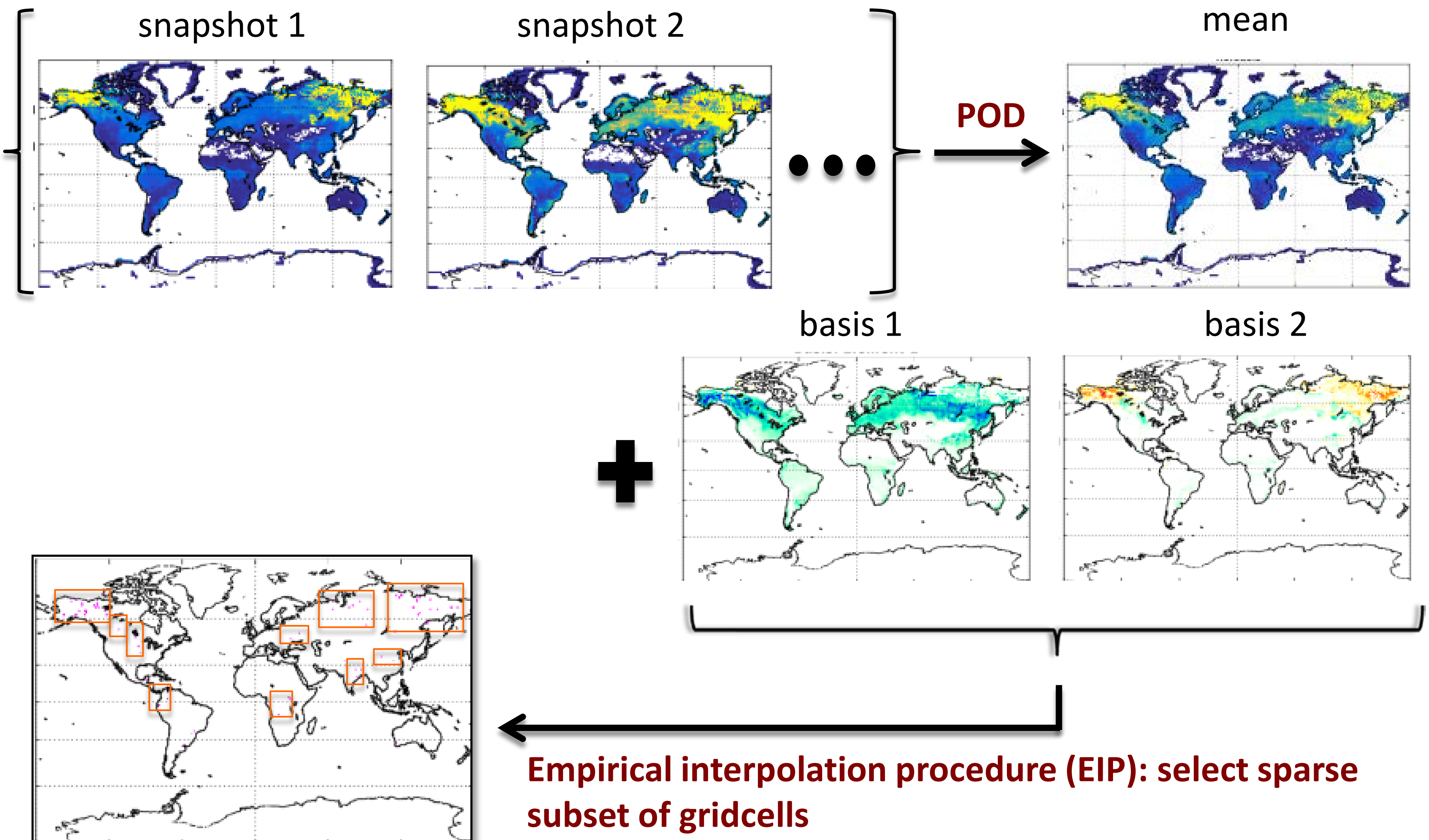


Classical analogue is polynomial interpolation.

Chebyshev points are good points for polynomial interpolation.

Similarily, we pick “good” points for GPOD using EIP on the basis.

GPOD: Train



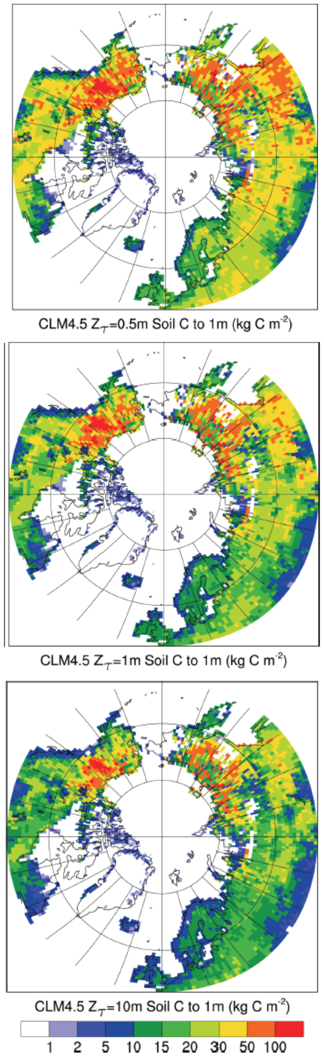
Empirical interpolation procedure (EIP): select sparse subset of gridcells

GPOD: Predict

1. Perform simulation on sparse grid.
2. Choose M POD basis elements (i.e., EOFs)
3. Find the linear combination of the M that minimizes the error on the sparse grid.
4. Project onto the full grid

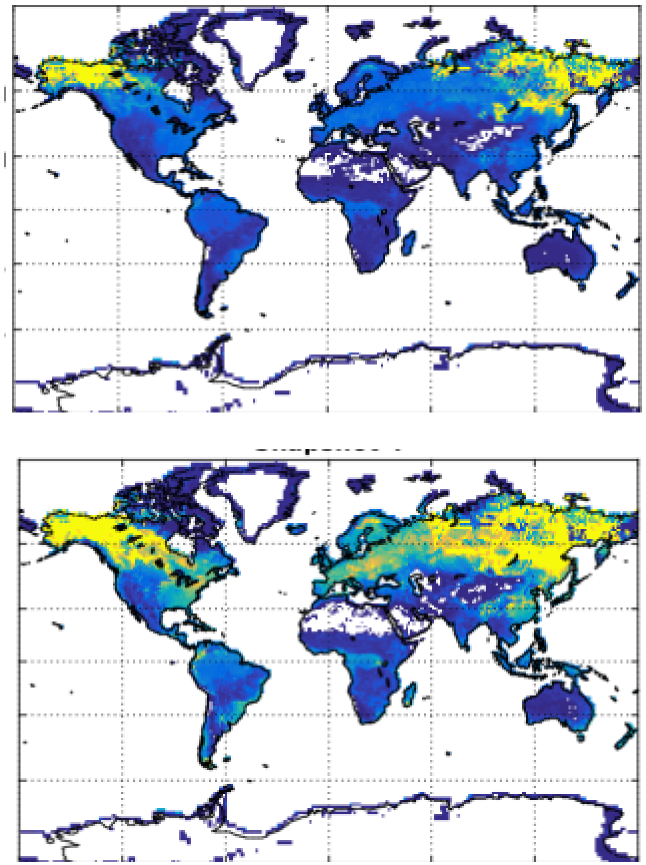
Permafrost Carbon Feedback

Soil carbon for different z_τ



- Setup
  - Koven, Lawrence, & Riley (in revision, PNAS)
  - Predict future soil and vegetation carbon distributions.
  - 31 CLM4.5 simulations differed with respect to:
    - Forcing (historical, RCP8.5…)
    - CO_2 concentration
    - Active N cycle
    - Depth-dependent decomposition (z_T) parameter: surrogate for missing processes (priming, mineral interactions, etc.).

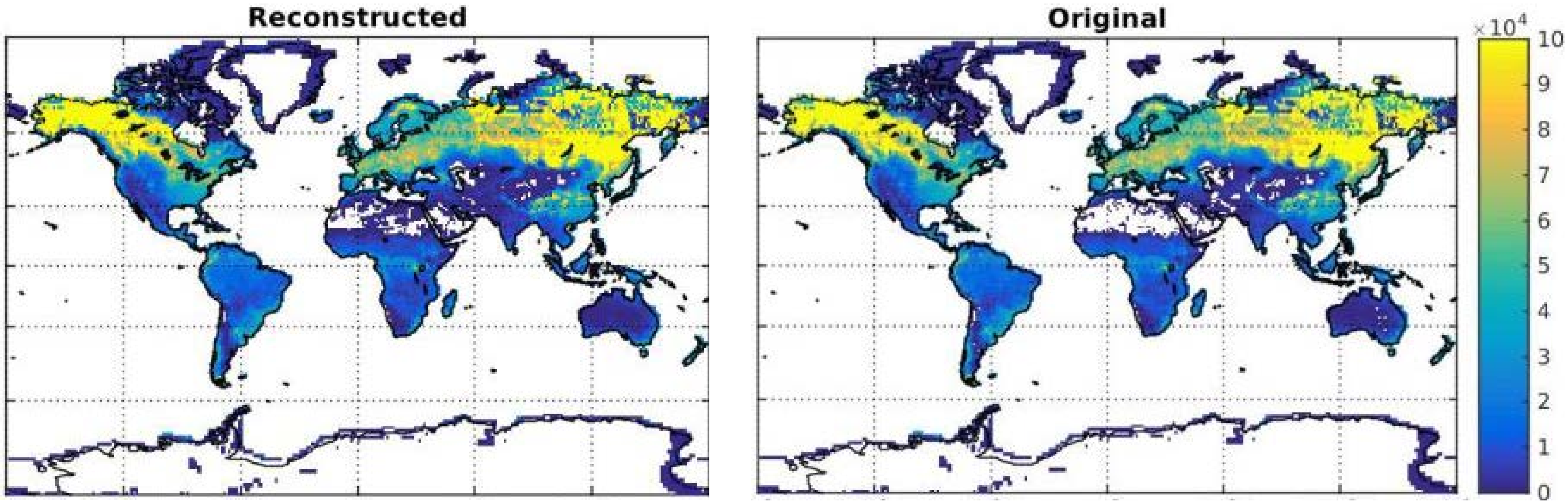
Permafrost Carbon Feedback



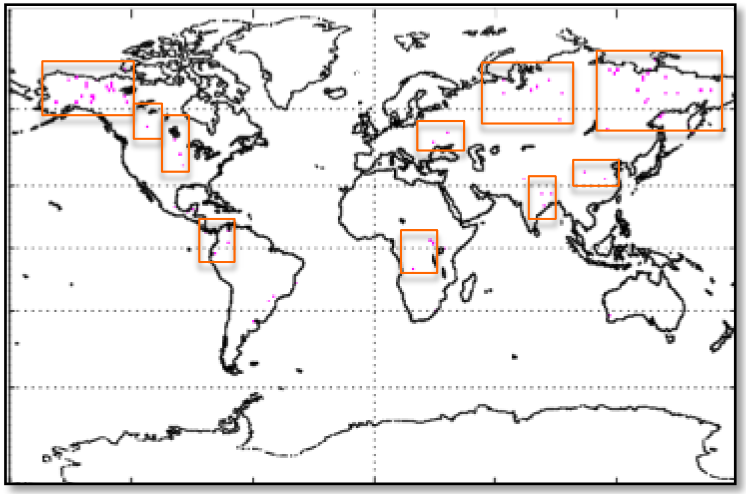
Two most different soil carbon snapshots

- ROM Setup
  - Choose 10 simulations out of the 31 simulations for training.
    - Based on an adaptive sampling procedure.
    - Monthly or annual data.
  - The rest of the 31 simulations are used for validation.

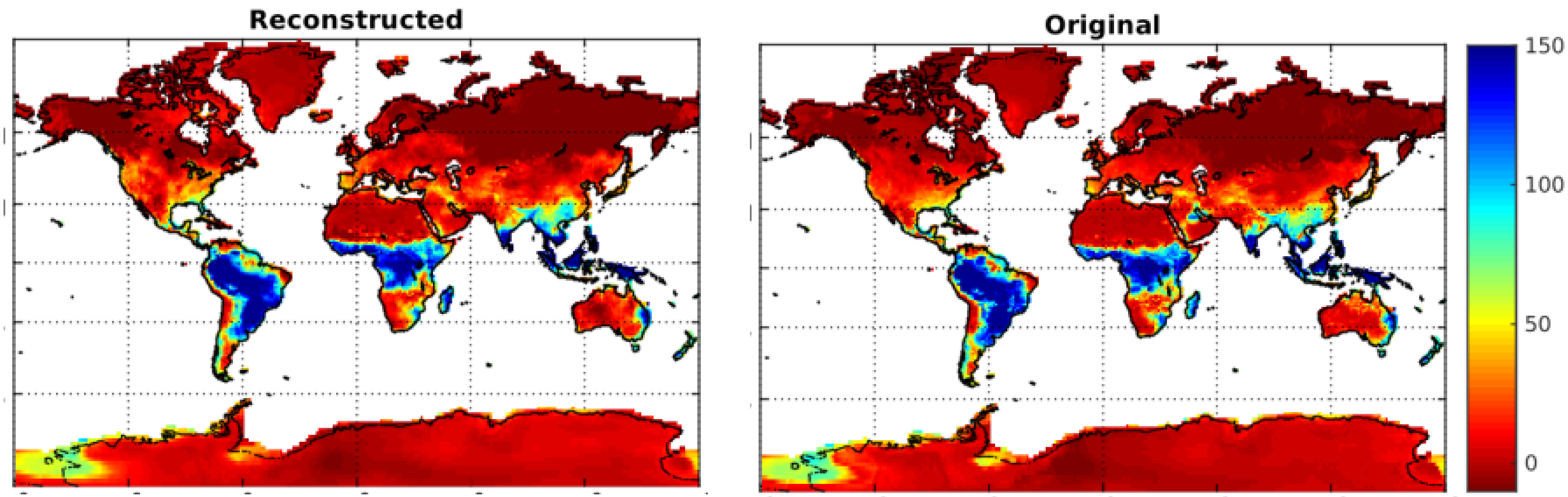
Soil carbon results



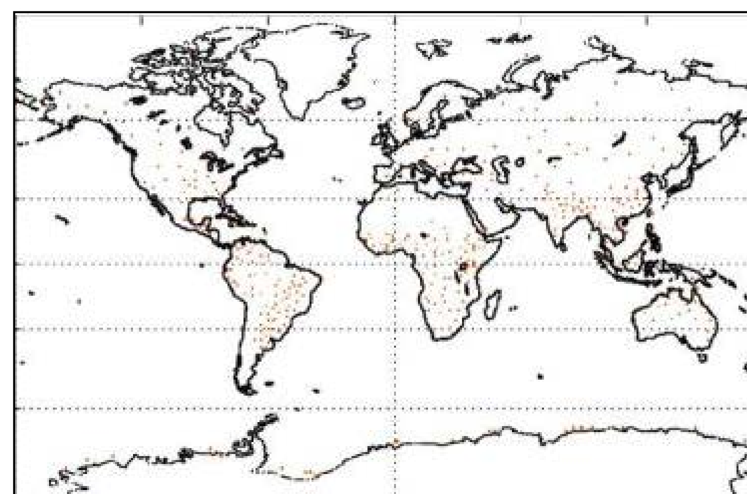
- Worst reconstruction (0.8% error) for annual-mean soil carbon (g m^{-2}).
- 80 gridcells (0.4%)



Latent heat flux results



- Monthly latent heat flux ($W\ m^{-2}$).
- 500 gridcells (2.5%)
- Worst mean relative error = 11%.
  - Annual: 6%, Decadal: 3%

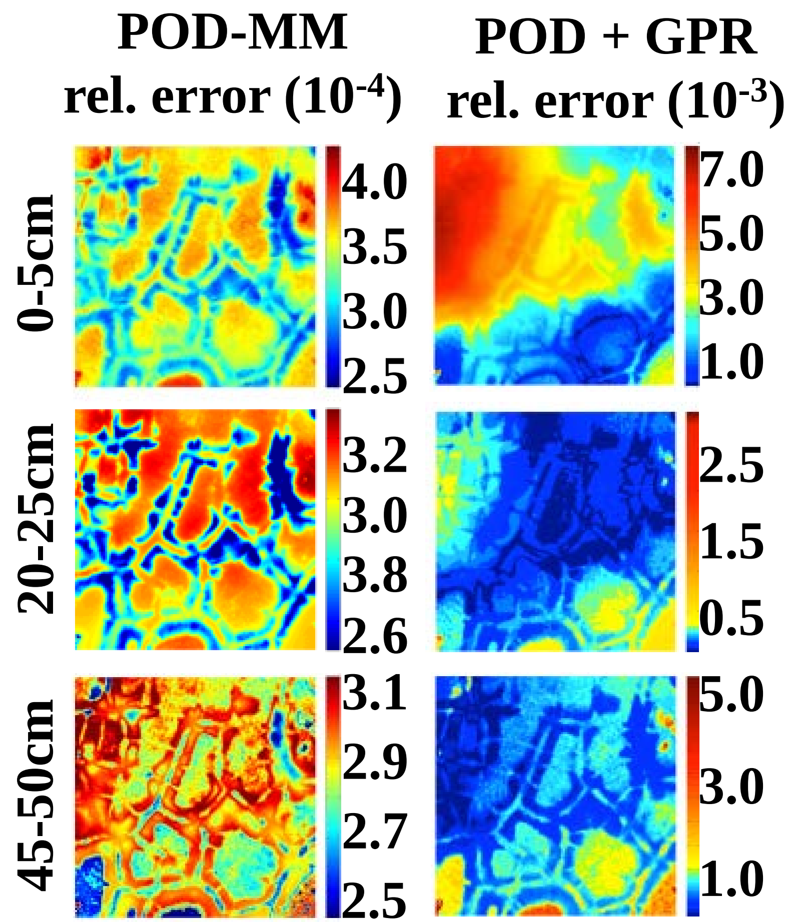


Gaussian Process Regression (GPR)

- A machine learning algorithm: generalization of kriging
  - Input → output function modeled as Gaussian process
- Implicitly determines a parsimonious representation of heterogeneity in parameter space
  - Input data + forcing

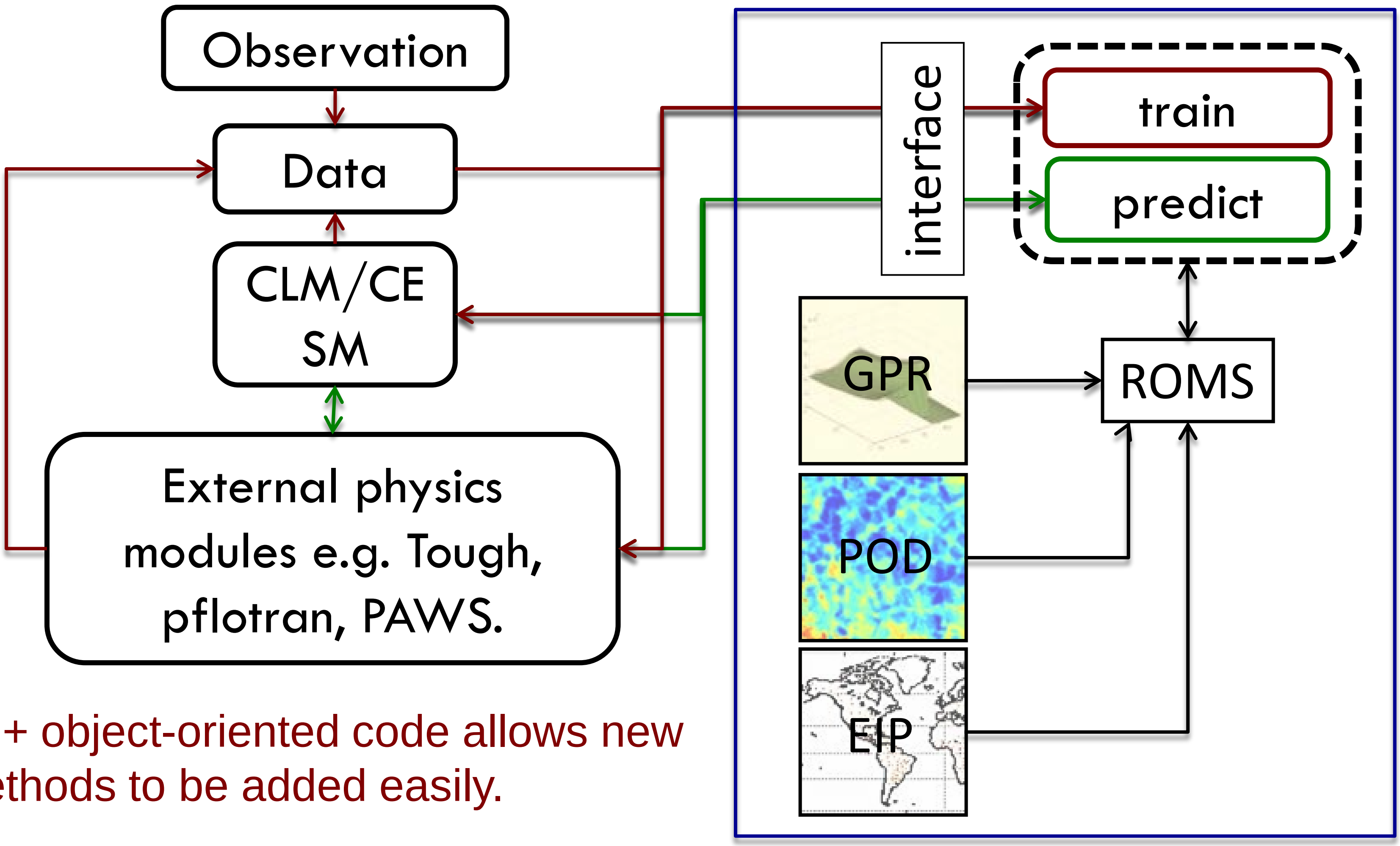
POD+GPR: Barrow

- 3 parameters:
  - Time (day)
  - Precip.
  - ET
- Relative error ~ 10^{-3}.
  - Larger than POD-MM but faster.



pROME

Parallel Reduced Order Models for Earth Systems



C++ object-oriented code allows new methods to be added easily.

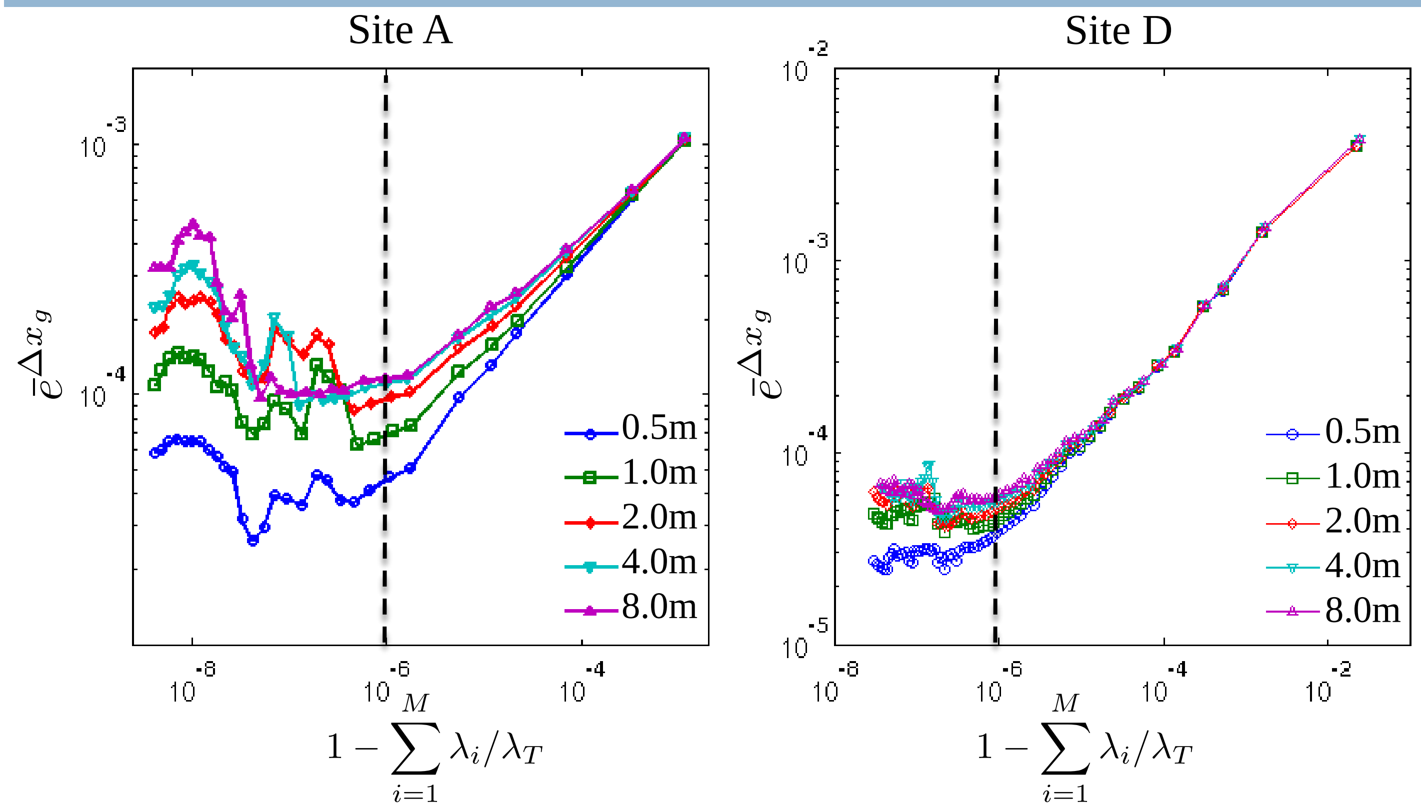
Summary

- We demonstrated examples in which ROM maintains accuracy while reducing computation
- No one-size-fits-all solution.
  - Trial & error with different problems.
- **Working on pROME**: standardize ROM workflow and couple to CLM.
- **Lacking**: comprehensive high-fidelity models for ecosystems of interest (i.e., permafrost / peatlands)

Acknowledgement

- Additional collaborators: Yingqi Zhang, Stefan Finsterle.
- Funding for this study was provided by the US Department of Energy, BER Program, contract # DE-AC02-05CH11231 under the Early Career Research Program.

Barrow: Optimal M

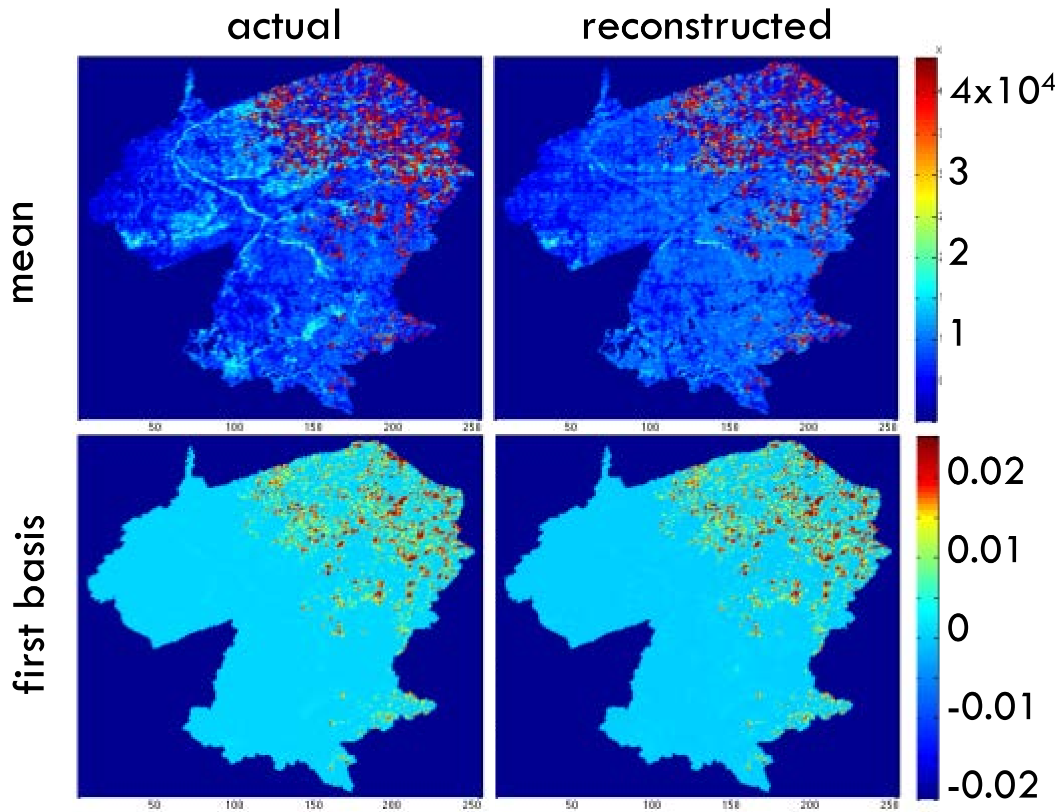


- Determine an optimal M without knowing actual error.
- A good criterion:

$$1-\sum_{i=1}^{M}\lambda_i/\lambda_T \le 10^{-6}, \quad \lambda_T=\sum_{i=1}^{N}\lambda_i$$

POD+GPR at Clinton River

- For constructing POD bases at any site.
  - As a way to deal with site-independent ROM.



GPOD: PREDICT

1. Perform simulation on sparse grid.
2. Choose M (≤N) POD bases

$$1-\sum_{i=1}^{M}\lambda_i/\sum_{i=1}^{N}\lambda_i \leq$$ Tolerance for neglected variance

3. Solve a least-square minimization problem (a MxM linear system):

$$\alpha = \arg\min_{\gamma} \| \mathbf{f}(\mathbf{x}_s) - \overline{\mathbf{f}}(\mathbf{x}_s) - \sum_{i=1}^{M} \gamma_i \zeta_i^{\mathbf{f}}(\mathbf{x}_s) \|_2$$

4. Reconstruct the ROM solution:

$$\mathbf{f} \approx \mathbf{f}^{\mathrm{GPOD}} = \overline{\mathbf{f}} + \sum_{i=1}^{M} \alpha_i \zeta_i^{\mathbf{f}}$$