SAND2016-0715C

Improving predictive capability of land surface models through robust statistical calibration techniques

J. Ray, L. Swiler (SNL), G. S. H. Pau, G. Bisht (LBL), F. M. Hoffman (ORNL), M. Huang, Z. Hou and X. Chen (PNNL)

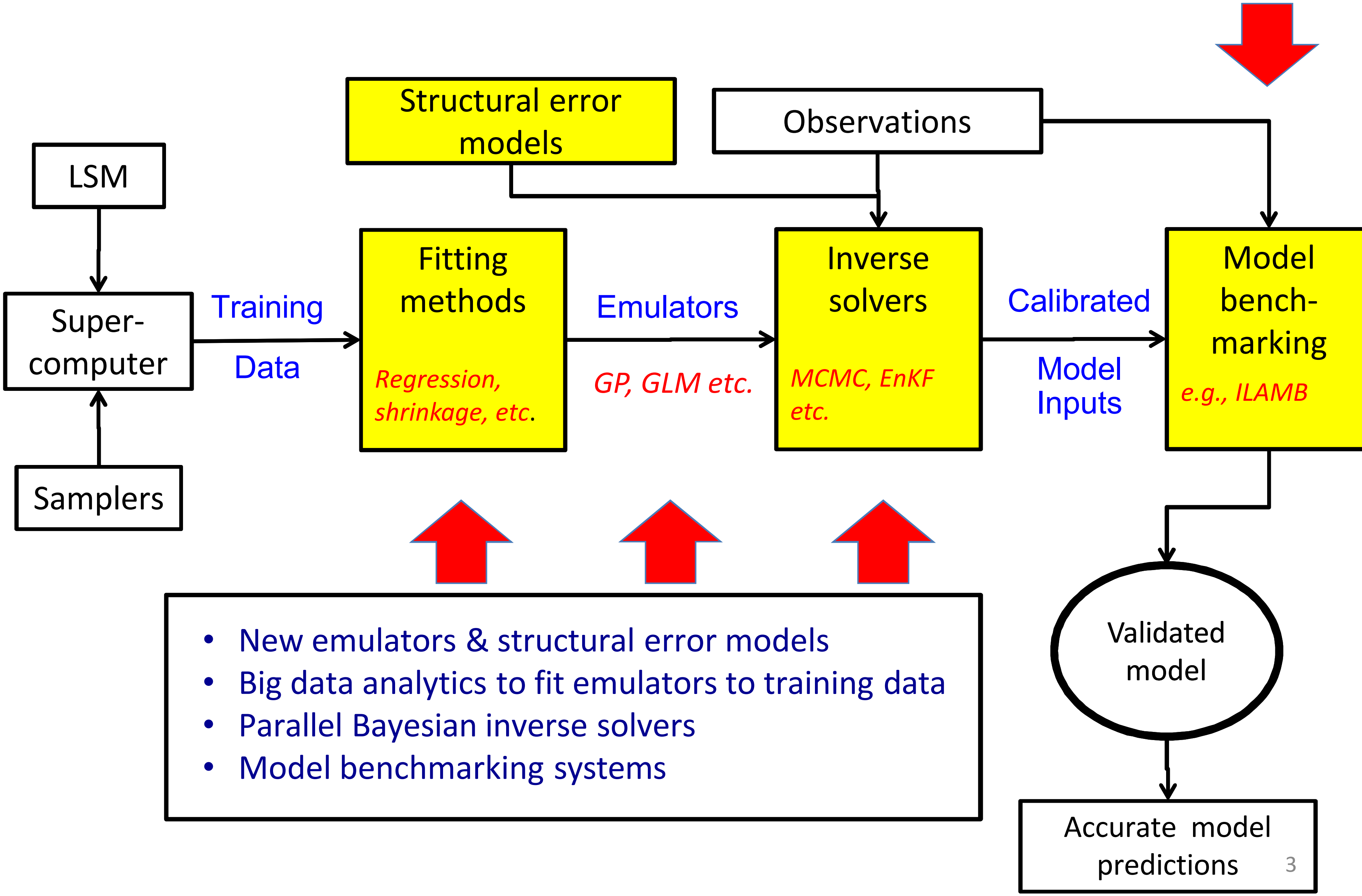
Advances in Mathematical and Computational Climate Modeling Workshop

January 26 – 27, 2016

The need for model calibration

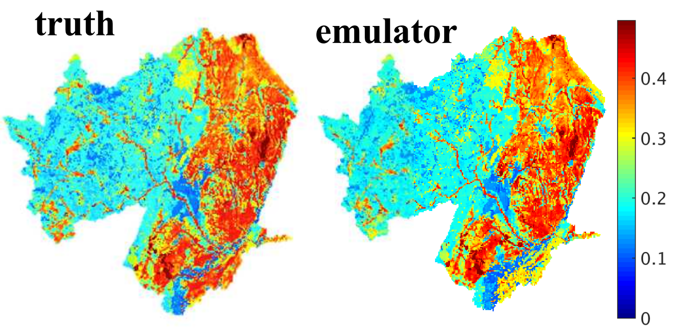
- Climate models, including Land Surface Models (LSMs) calibrated to perform well at the global scale
  - For practical planning, engineering and assessment of climate change, need predictions at a finer resolution e.g., county or provincial scale
- Many parameterizations in LSMs are scale-dependent
  - As spatial resolutions become finer, or when used at regional or site-level, need re-calibration
  - Calibration (observational) data is available at *site* scales
- Limited observational data and structural errors in LSMs hinder calibration
  - Have to continuously estimate uncertainty in calibrated parameters
  - Need to model structural errors in some form when performing calibration
  - Often have to consider vectors and fields as “parameters” being estimated, not just scalars E.g., CO_2 flux fields
- Some LSM calibration work performed; some idea of what might work

The calibration pipeline



Challenges in calibration

- Structural errors: In calibration, structural errors may be
  - Removed e.g., use better sub-model OR
  - Modeled and estimated from observations; need random field models
- Uncertainty quantification: Limited observational data demand that
  - We reduce dimensionality beforehand
  - Estimate model inputs along with their uncertainty



Including physics often improves emulators

- Emulators: curve-fits of model responses; used to address the computational expense of Bayesian methods
  - Not always possible to make accurate emulators
  - Emulators for fields (not scalars) tend to require huge training data
  - Rough model responses lack good parameterizations
- Scalability: Many statistical algos are parallelizable but tools are serial
  - Popular ones: R, Matlab, Dakota, UQTK

Solutions we need – 1/2

- Parallel MCMC methods for calibration
  - To calibrate using LSMs, not their emulators
  - Prototypes exist, but:
    - Scalability not proven (yet)
    - Don’t combine LSMs and their (perhaps inaccurate) emulators
    - No mature software base yet
- Emulators of different types
  - Many types of emulators exist when model response is *smooth*
    - But we need a “recipe” to decide what type fits a given training dataset
  - Need non-stationary RFMs for model responses that are *rough*
  - We also need classifiers to stay in the “physical” part of a parameter domain
    - Or at least, where an emulator is usable
  - Finally, we need scalable (> 1K cores) tools that can fit emulators to large datasets

Solutions we need – 2/2

- Non-stationary random field models
  - Fancy name for parameterizations that can capture rough fields
  - Needed when the length-scale of model response variation changes in perturbed parameter domain
    - Examples: wavelet-based, tree-based etc.
    - Basic idea is to chop the parametric domain into sub-domains (in a data-driven manner) and fit conventional emulators with compact support
  - They need specialized algorithms to fit to training data
- Need to use "big data analytics" for fitting emulators & RFMs
  - Fancy name for very scalable statistical tools
  - Contain scalable, statistical algos e.g., regression, shrinkage, NN, PCA, classifiers
  - Existing frameworks like Spark/MLlib and H20 scale to Google-sized problems - our emulator training dataset won't stretch them
- Finally, integration with a bench-marking system
  - Check the usefulness of calibration

Practicalities

- Prototypes of 3 / 5 solutions have been demonstrated (see 2 pager)
  - "big-data-analytics" in anything useful – not yet (2015 α-release)
  - No integration with a benchmarking system either
  - Scalable (MPI + R) tools to "do discovery" in climate datasets
    - U. Minnesota (NSF-funded); http://climatechange.cs.umn.edu;
    - No calibration or emulators
- Ongoing ASCR-funded work on developing RFMs and parallel MCMC
  - Prototypes/proof-of-concept; won't fund hardened tools
  - Have been demonstrated with LSMs
- Deliverables – to be useful in climate modeling, we need:
  - Libraries of scalable, multi-chain MCMC methods
  - Libraries of emulators that implement a pipeline:
    - Fit & simplify emulators to data using big-data-analytics (BDA) platforms
    - Select between them using ML model selection techniques
    - BDA infrastructure at supercomputing centers
  - Integration of calibration libraries with a benchmarking package

Synopsis

- We need a hardened software base for calibration of climate models
  - LSMs are the easiest target; there are preliminary results and prototype implementations of such a software base
- Calibration should yield the objects of interest as PDFs
  - Quantify the uncertainty due to limited data and/or structural errors
- The main constituents of this software base would be:
  - Methods to make and select emulators from very large training data
  - Use of big data analytics to fit emulators and RFMs to large data
  - Parallel calibration methods, in case we cannot make emulators
  - Integration with a benchmarking package
- In order to achieve this, we will need:
  - A use-case e.g., LSMs, to pose realistic calibration problems rather than theoretical ones
  - A team of climate scientists & ASCR-types to get this done
  - Funding would also be appreciated ☺

BACKGROUND

Calibration methods & shortcomings -1/2

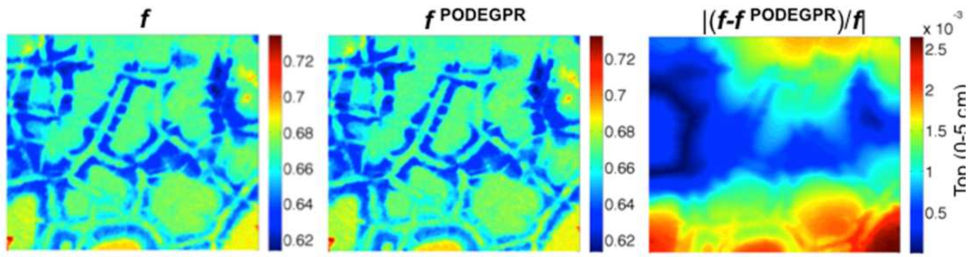
- During calibration, structural errors are incorporated in two ways
  - *Remove them:* e.g. use a detailed sub-surface simulator when using an LSM in calibration. E.g., use PFLOTRAN with CLM4
  - *Estimate them:* use a statistical parameterization for them and estimate them during calibration
- The effect of limited observational data is addressed in 2 ways
  - *Dimensionality reduction:* Estimate just the important model parameters
    - Use sensitivity analysis or shrinkage to discover them
  - *Bayesian calibration:* Use a Bayesian formulation and a method like Markov chain Monte Carlo (MCMC) or Ensemble Kalman Filters (EnKF) to estimate parameters as PDFs
    - PDFs provide you with a concise measure of estimation uncertainty
- Bayesian calibration is always more comp. expensive (10^2x-10^6x)
  - Sometimes, scalable methods exist e.g., EnKFs
  - Else, we need to use emulators (stat. curve-fits acting as proxies)

Calibration methods & shortcomings -2/2

- Emulators are **NOT** a silver bullet
  - Made by fitting response surfaces to training data from perturbed parameter runs
  - Limited to 1-10 parameters- huge training data requirements
- Often not possible to make *accurate* emulators (e.g., < 10%)
- Often need to model fields (not scalars) as a function of perturbed parameters
  - Both need parameterizations of surfaces in high dimensions (called random field models, RFMs)
- Tools to fit emulators usually serial (DAKOTA, Matlab, R)
  - Sometimes the training data *does not* fit in the memory (e.g., if emulating fields, not scalars)
  - We need:
    - Scalable, multi-processor statistical frameworks for fitting emulators
    - Calibration tools that will work if we fail to make emulators

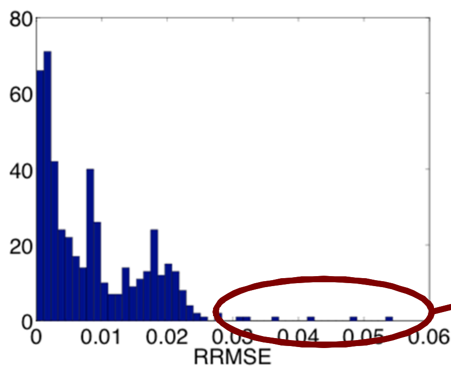
Representative results of emulators

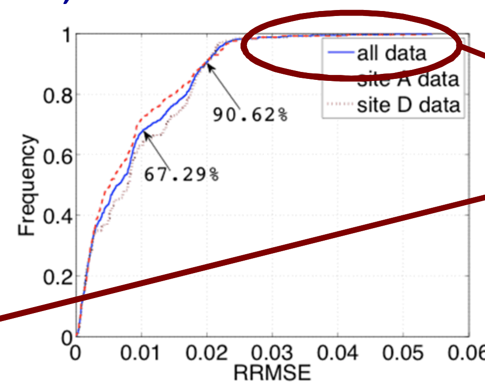
- Aim to reproduce moisture content distribution by combining data dimensional reduction technique (PCA) and Gaussian process regression.



Looks good for a single snapshot, BUT …

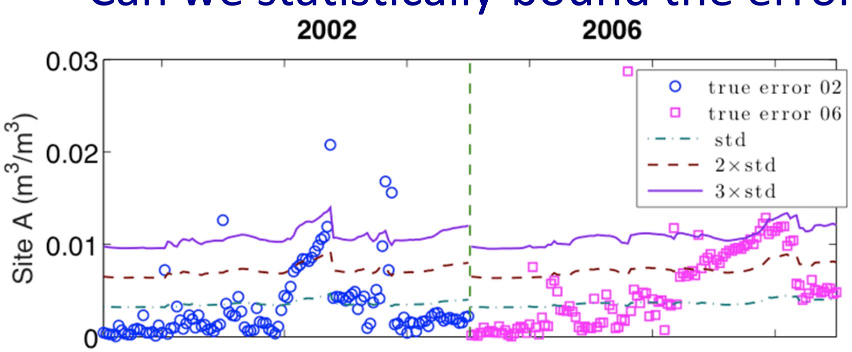
- Accuracy is not uniform, with some outliers.





What's the impact of these outliers on UQ analysis?

- Can we statistically bound the error?

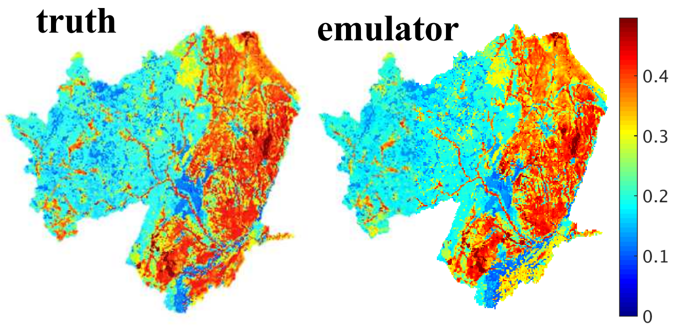


Somewhat. How do we use these estimates in an UQ analysis?

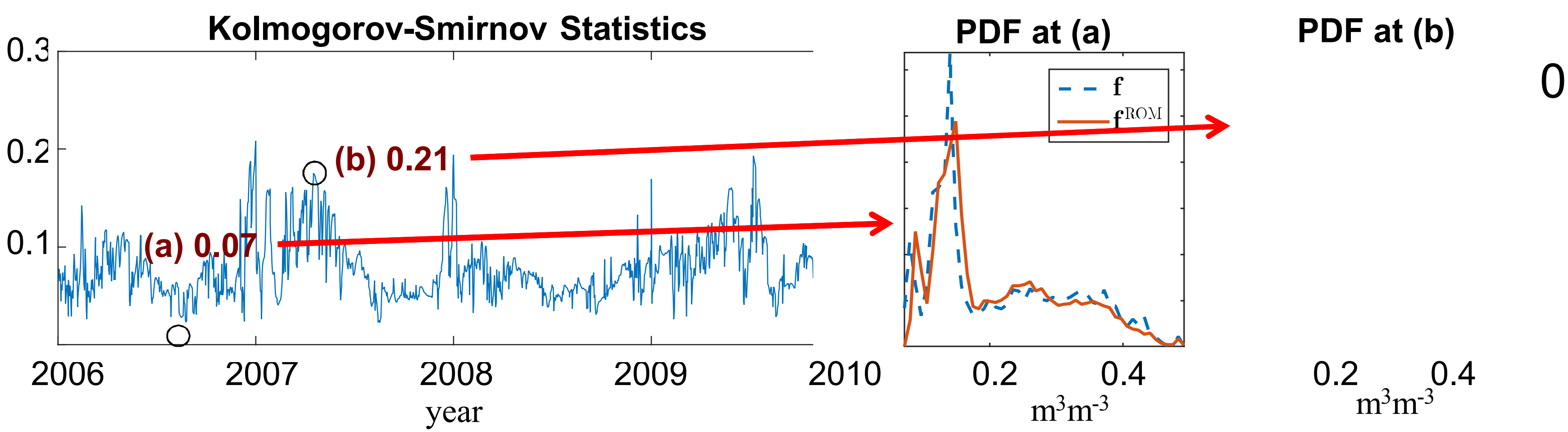
Source: Liu, Y., et al. (2015). "A hybrid reduced-order model of fine-resolution hydrologic simulations at a polygonal tundra site." Vadose Zone Journal. doi:10.2136/vzj2015.05.0068

Representative results of emulators

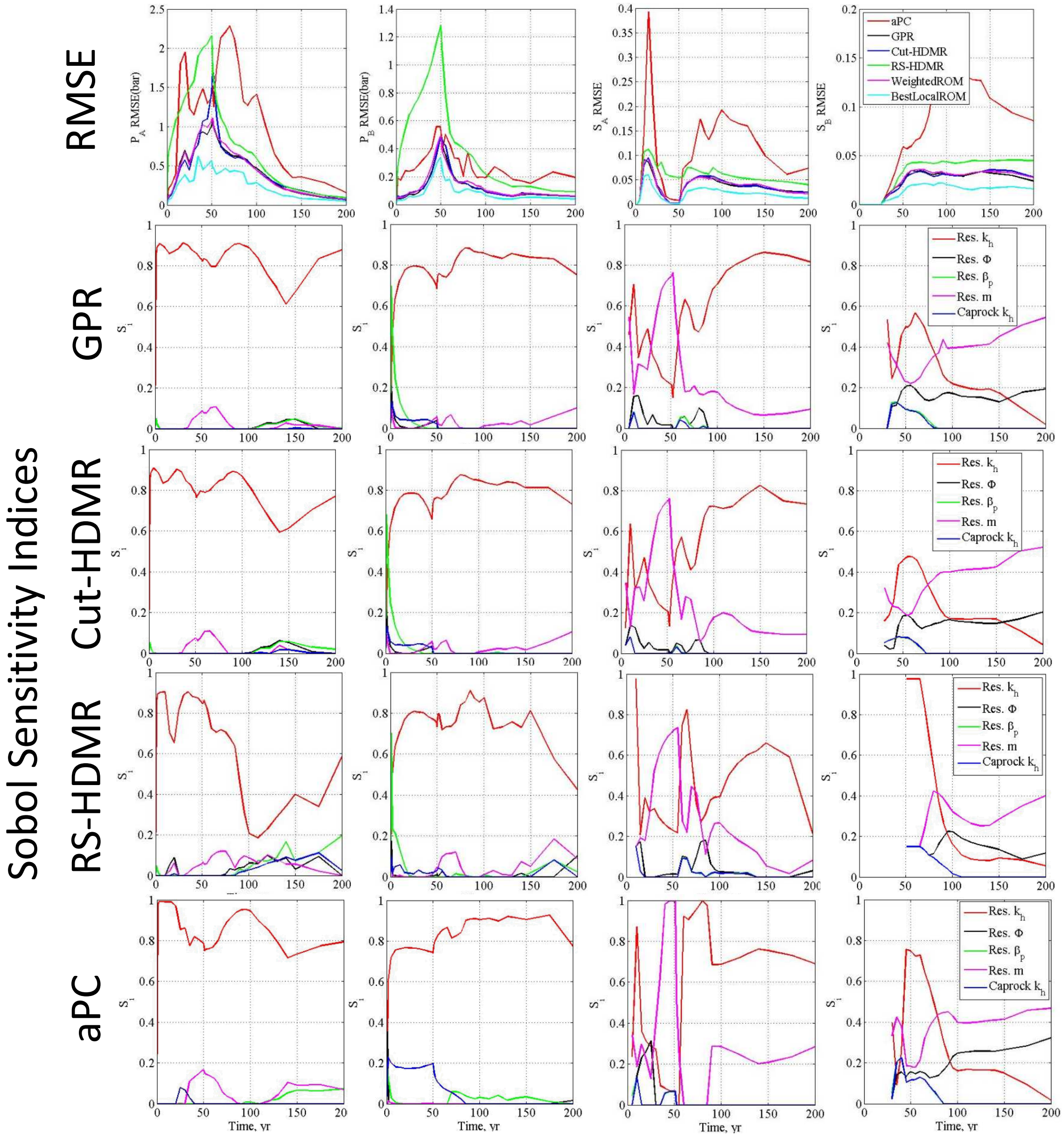
- A more heterogeneous problem require more advanced emulator technique: here based on coarse-grid simulation and PODMM downscaling technique.



Including physics-related information (through efficient coarse model) can improve the representation. Still, accuracy is not uniform.



Effects of ROM accuracy on Global SA



ROM studied:

- GPR: Gaussian Process Regression
- Cut-HDMR: Cut High Dimensional Model Reduction
- RS-HDMR: Random sampling HDMR
- aPC: arbitrary Polynomial Chaos

Accuracy of emulator depends

- type of ROMs: GPR and cut-HDMR have better accuracies.
- smoothness in the variables studied: pressure easier to estimate than saturation.

Accuracy of analysis

- Less sensitive to emulators if only qualitative analysis is needed.

Source: Zhang, Y, et al. (2016). "Evaluation of Multiple Reduced-Order Models to Enhance Confidence in Global Sensitivity Analyses." Submitted.