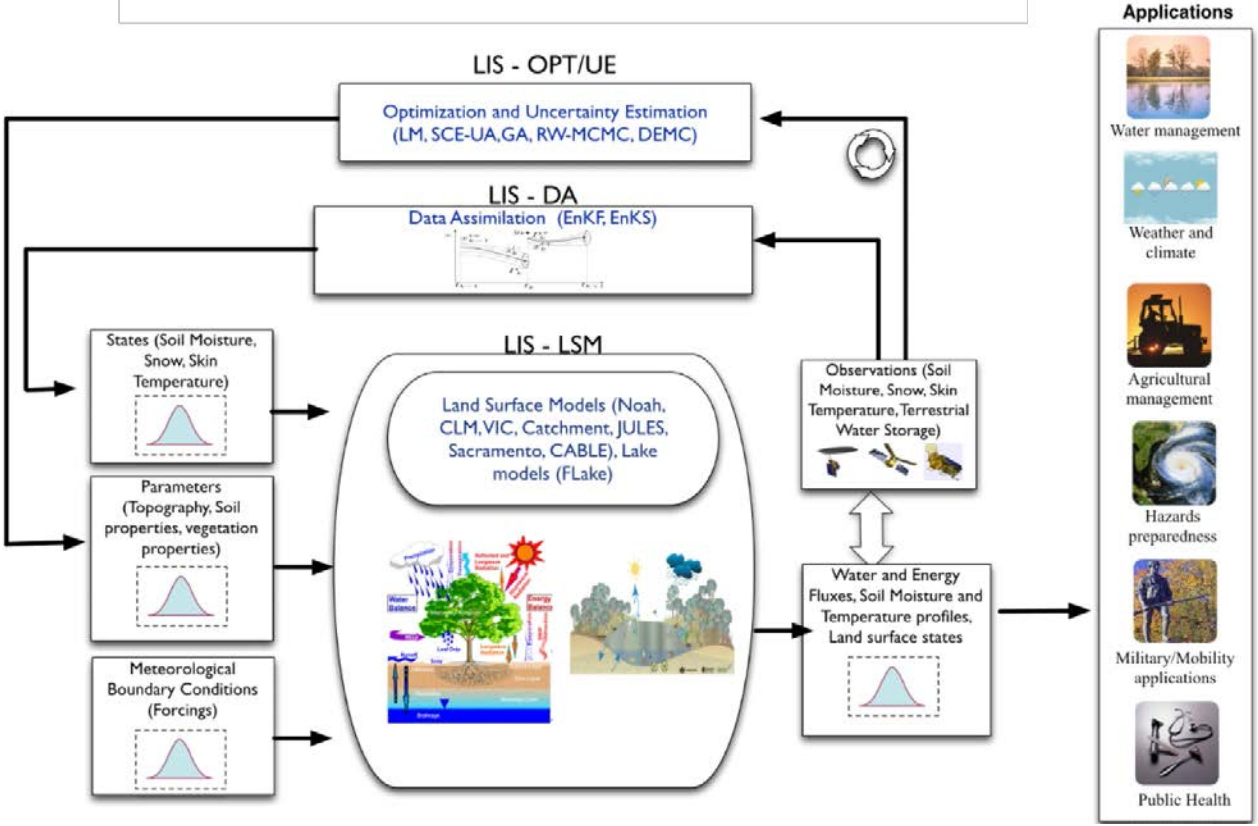
Learning Model Structural Uncertainty with Gaussian Processes

[1,2]Jules Kouatchou, [1,2]Craig Pelissier, [3]Grey Nearing, [1]Dan Duffy
[1]Christa D Peters-Lidard and [1]Jim Geiger

[1]NASA Goddard Space Flight Center, Maryland, USA
[2]Science Systems and Applications Incorporated, Maryland, USA
[3]University of Alabama, Alabama, USA

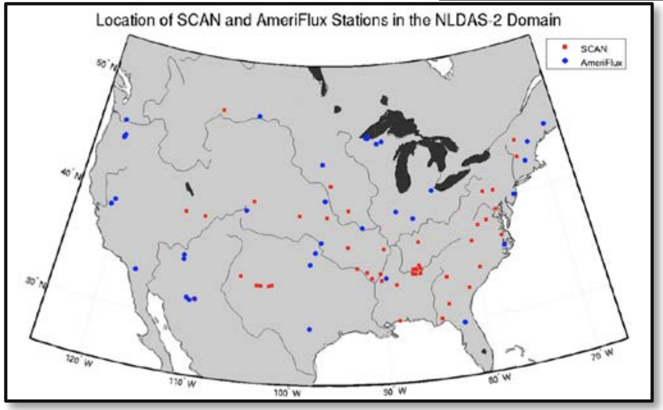
LIS Model



LIS - OPT/UE
Optimization and Uncertainty Estimation (LM, SCE-UA,GA, RW-MCMC, DEMC)
LIS - DA
Data Assimilation (EnKF, EnKS)
LIS - LSM
States (Soil Moisture, Snow, Skin Temperature)
Parameters (Topography, Soil properties, vegetation properties)
Meteorological Boundary Conditions (Forcings)
Land Surface Models (Noah, CLM,VIC, Catchment, JULES, Sacramento, CABLE), Lake models (FLake)
Observations (Soil Moisture, Snow, Skin Temperature, Terrestrial Water Storage)
Water and Energy Fluxes, Soil Moisture and Temperature profiles, Land surface states
Applications
Water management
Weather and climate
Agricultural management
Hazards preparedness
Military/Mobility applications
Public Health

Structural Uncertainty in the Models

NLDAS-2

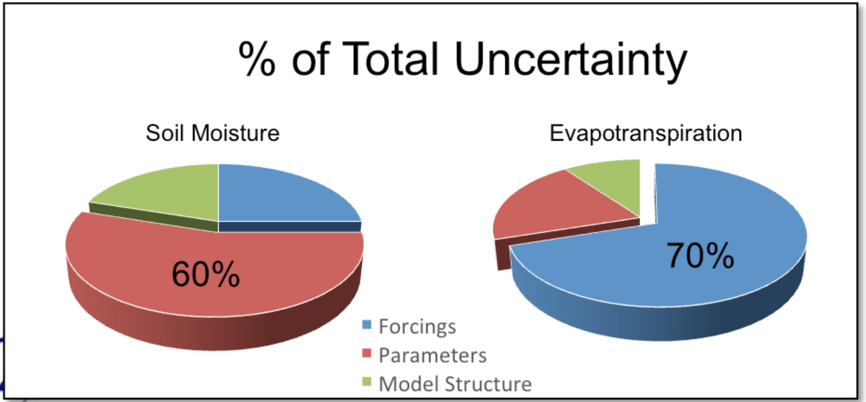


VIC

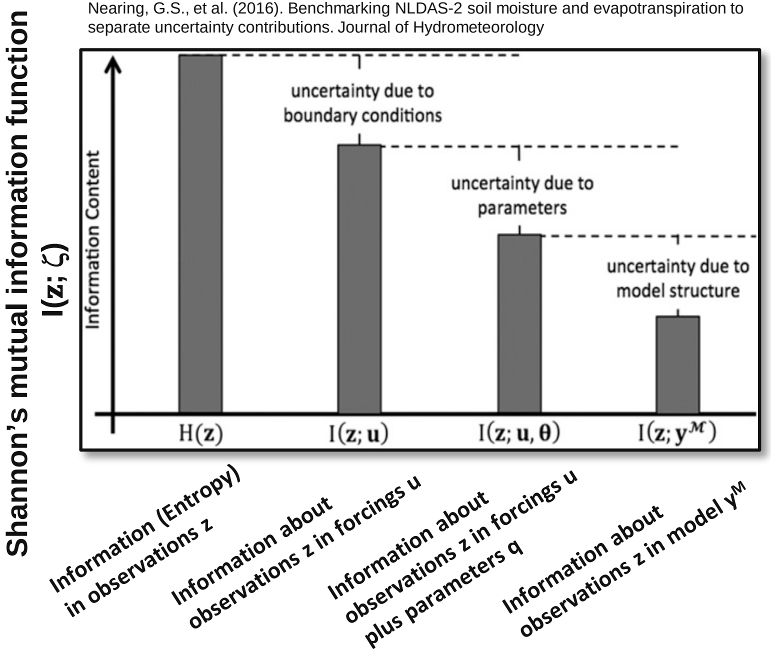
Mosaic

Noah

SAC-SMA



Nearing, G.S., et al. (2016). Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions. Journal of Hydrometeorology



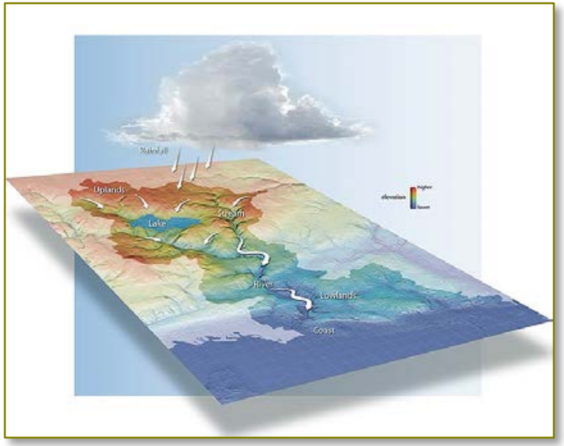
Dynamic Post-Processing with ML

$$dx = f(x,u,\theta_f)\,dt + \mathcal{GP}_{\mu}(f,x,u,\theta_{\mathcal{GP}}) + \mathcal{GP}_{\sigma}(f,x,u,\theta_{\mathcal{GP}})$$

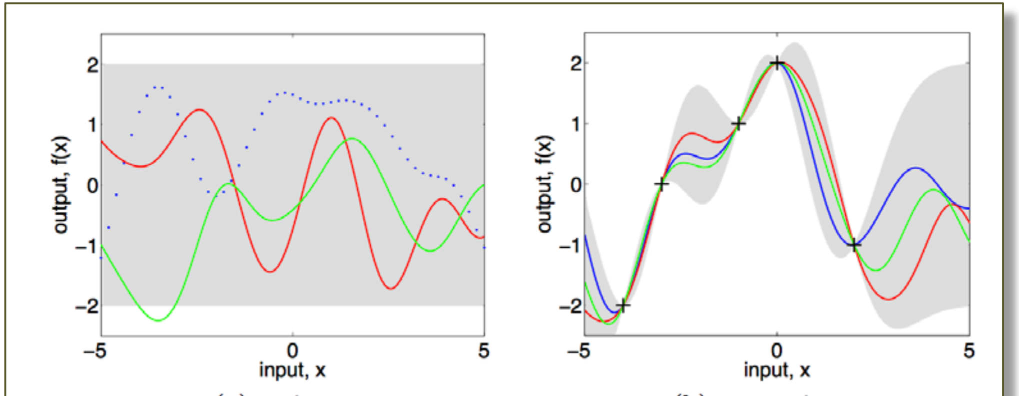
Biogeophysics lives here

Structured model error is here

Uncertainty lives here



+



(a), prior (b), posterior

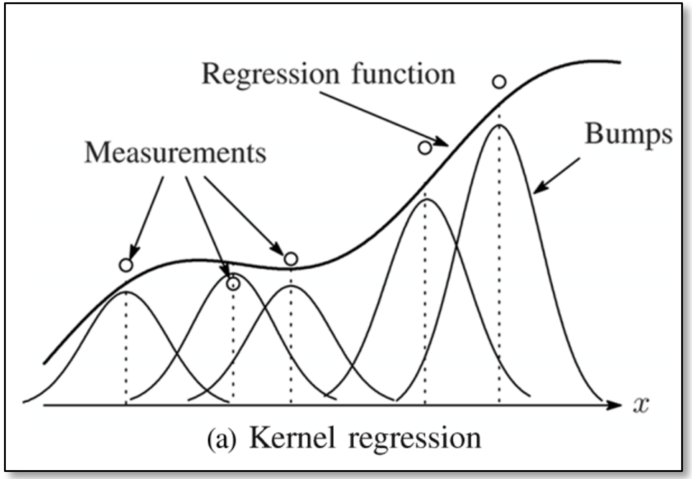
Gaussian Processes Regression

Gaussian Process Regression with ARD Kernel

$$K(x, x', \theta) = \sigma_f^2 \exp[-\frac{1}{2}(x - x')^T \Sigma^{-1} (x - x')]$$

$$\Sigma_{ij} = \sigma_i^2 \delta_{ij}$$

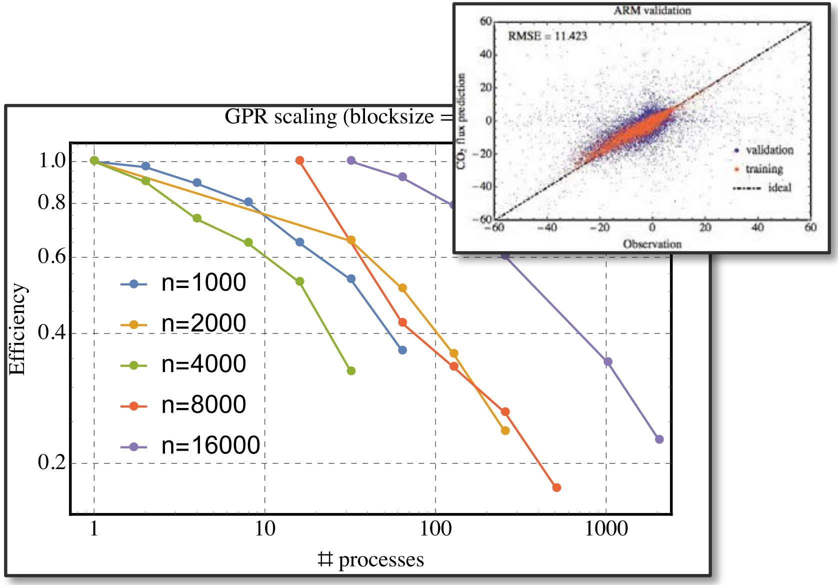
$$p(y|f) = \mathcal{N}(f, \sigma_y^2)$$



(a) Kernel regression

Efficient MPI-enabled c++ GPR

- Uses ScaLapack.
- Can do 50k-100k samples.
- Kernels: ARD w/wo noise and NIGP.

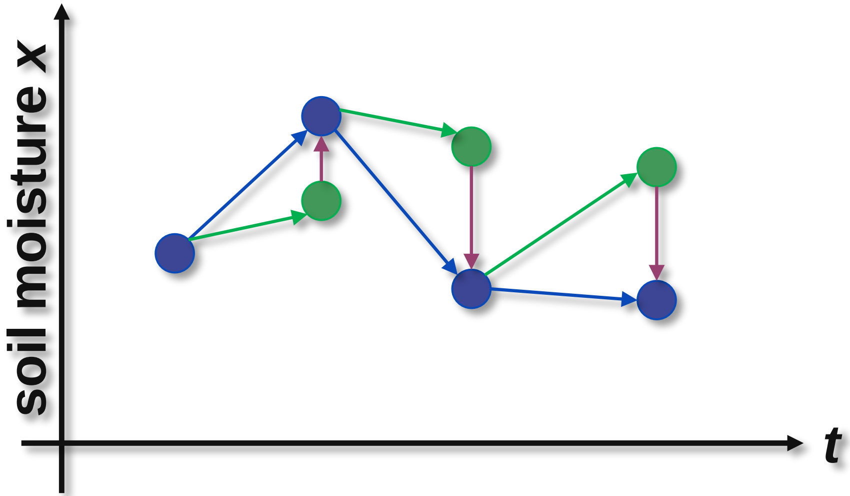


Training a Corrective Model

$$dy = f(x,u,\theta_f)\,dt + \Delta$$

Truth — Physical model — ML correction

If everything were perfect …

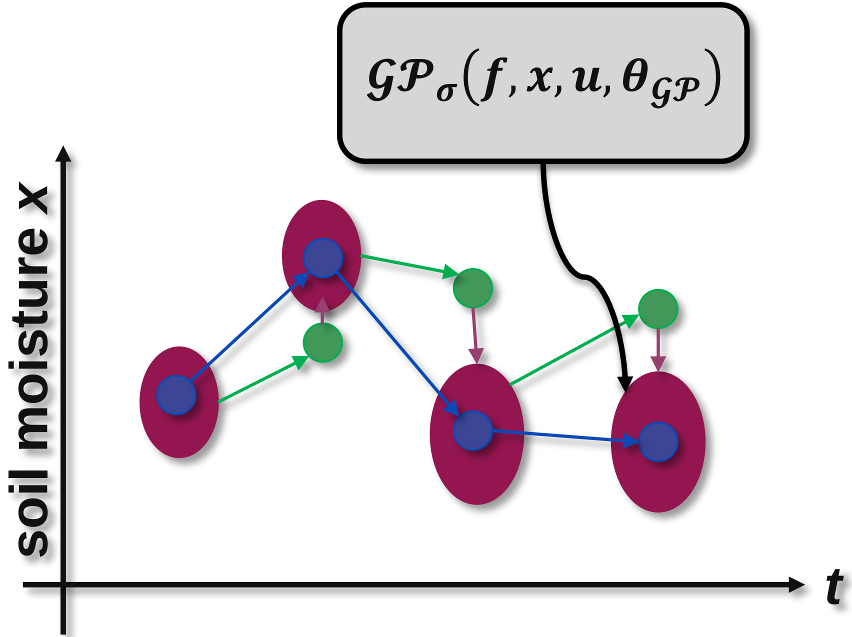


Creating a training set

1. Equilibrate Noah.
2. Start from observation "truth" at time ***t***.
3. Make prediction $f(x,u,\theta_f)dt$ (Noah).
4. Record training target = Δ, and and training input input = (fdt,x,u)
5. Repeat steps 2-4 for ***t+1***

Noisy Training Sets

Uncertainty in correction needs to be accounted for …

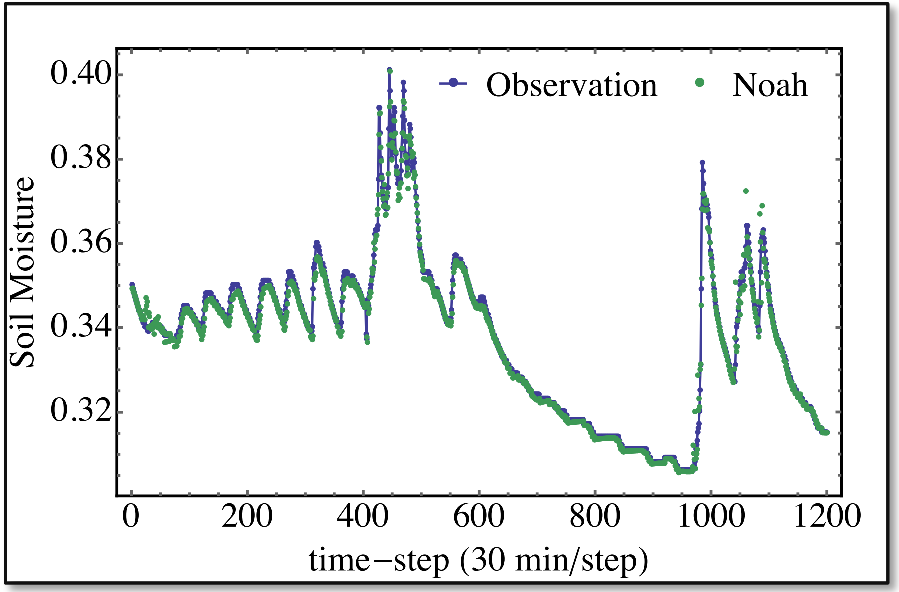


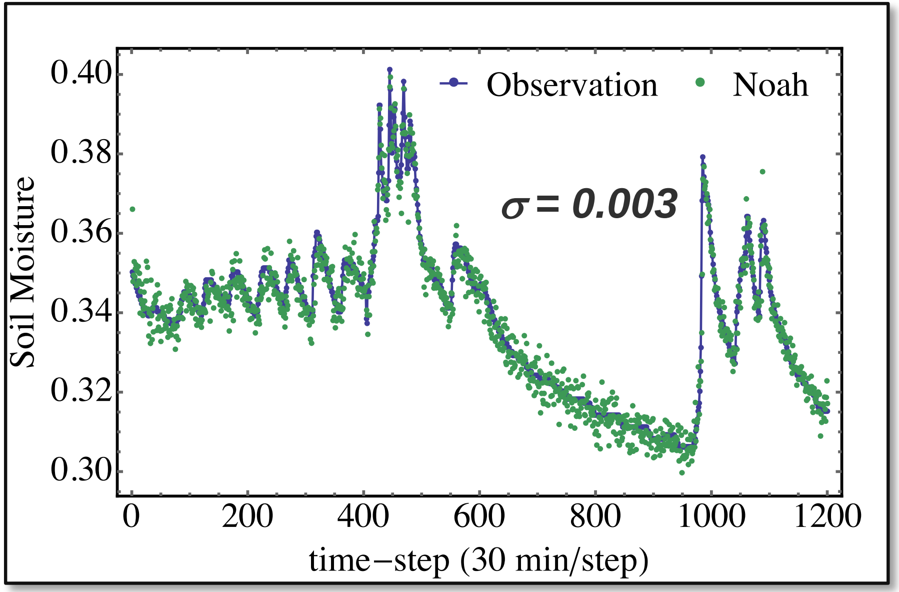
Modify procedure

Creating a training set

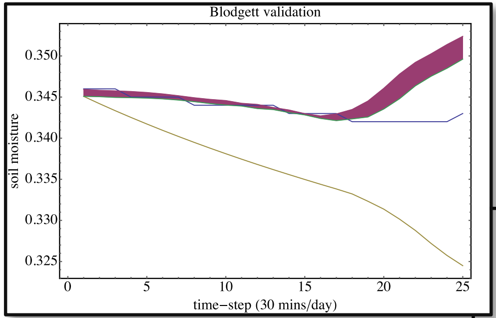
1. Equilibrate Noah.
2. Start from observation "truth" at time ***t***.
3. *Add **N(0,σ)** noise to lagged value to propagate uncertainty to the prediction **$f(x,u,\theta_f)dt$*** (Noah).
4. Make prediction ***$f(x,u,\theta_f)dt$*** (Noah).
5. Record training target = *Δ*, and and training input = ***(fdt,x,u)***
6. Repeat steps 2-5 for ***t+1***

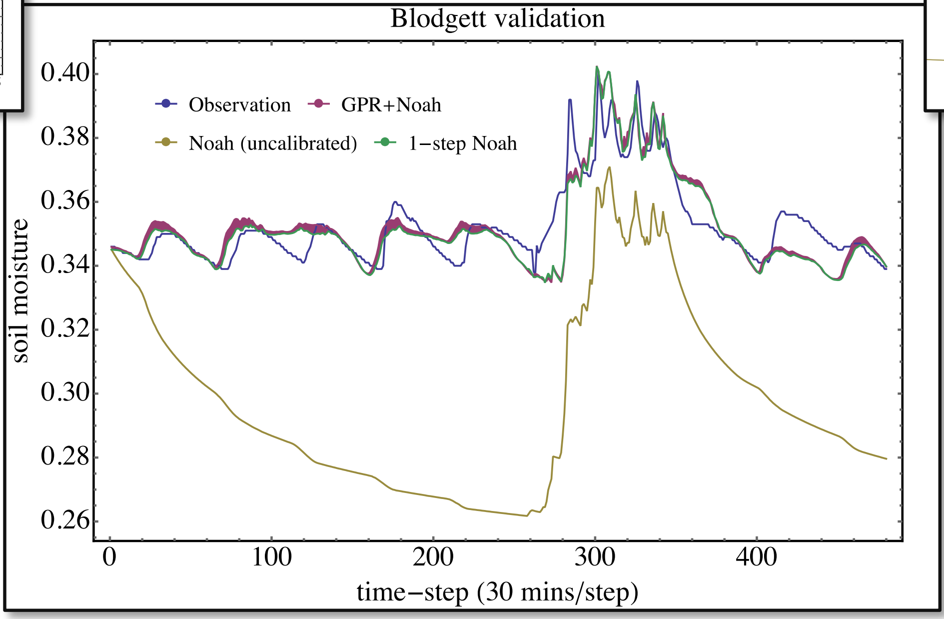
Blodgett Training Sets

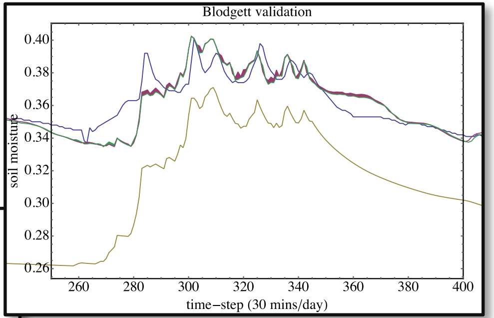




Blodgett AmeriFlux Tower







Noisy Input GPs

$$y = f(x - \epsilon_x) + \epsilon_y \approx f(x) + \epsilon_x^T \nabla f_{\mu}(x) + \epsilon_y$$

Input uncertainty

Observation uncertainty

Noisy Input GPs with ARD Kernel

$$\mathrm{K}(x, x', \theta) = \sigma_f^2 \exp[-\frac{1}{2}(x - x')^T \Sigma^{-1} (x - x')]$$

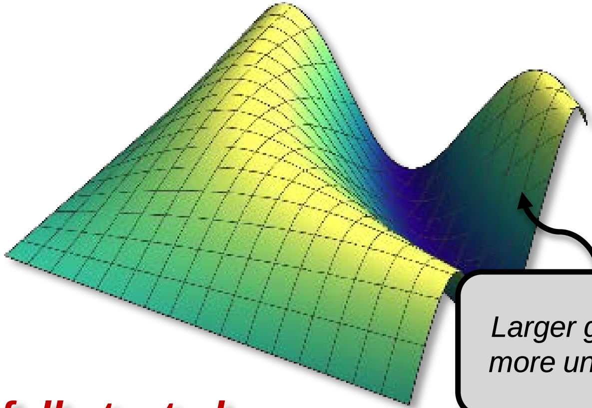
$$\Sigma_{ij} = \sigma_i^2 \delta_{ij}$$

$$p(y|f) = \mathcal{N}(f, \sigma_y^2 + \nabla f^T \Sigma^{noise} \nabla f)$$

$$\Sigma_{ij}^{noise} = \sigma_i^2 \delta_{ij}$$

Heteroscedastic output noise

Tunable hyperparameters



Larger gradients more uncertainty

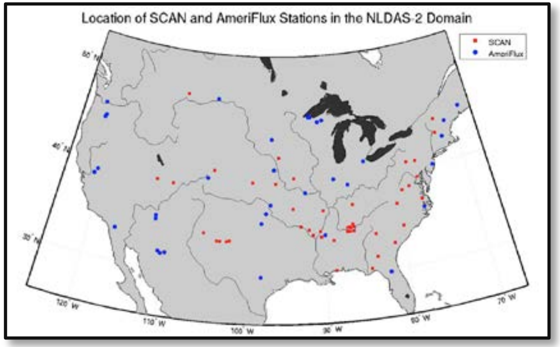
COMING SOON!

Implemented but not fully tested

Outlook

Coming soon

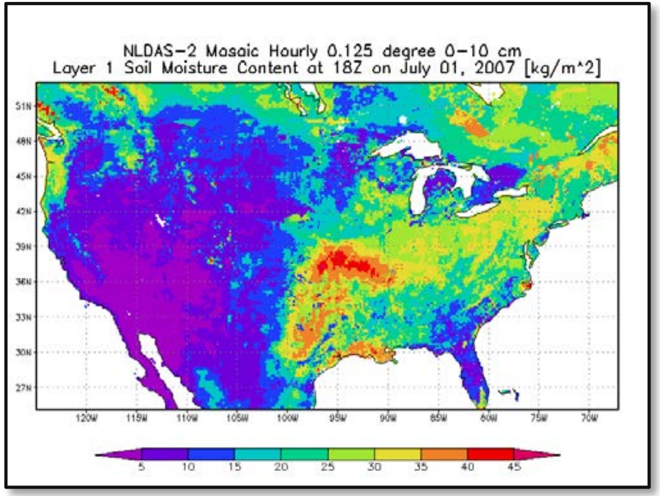
Validate spatially and temporally over 1 year at the flux towers sites.



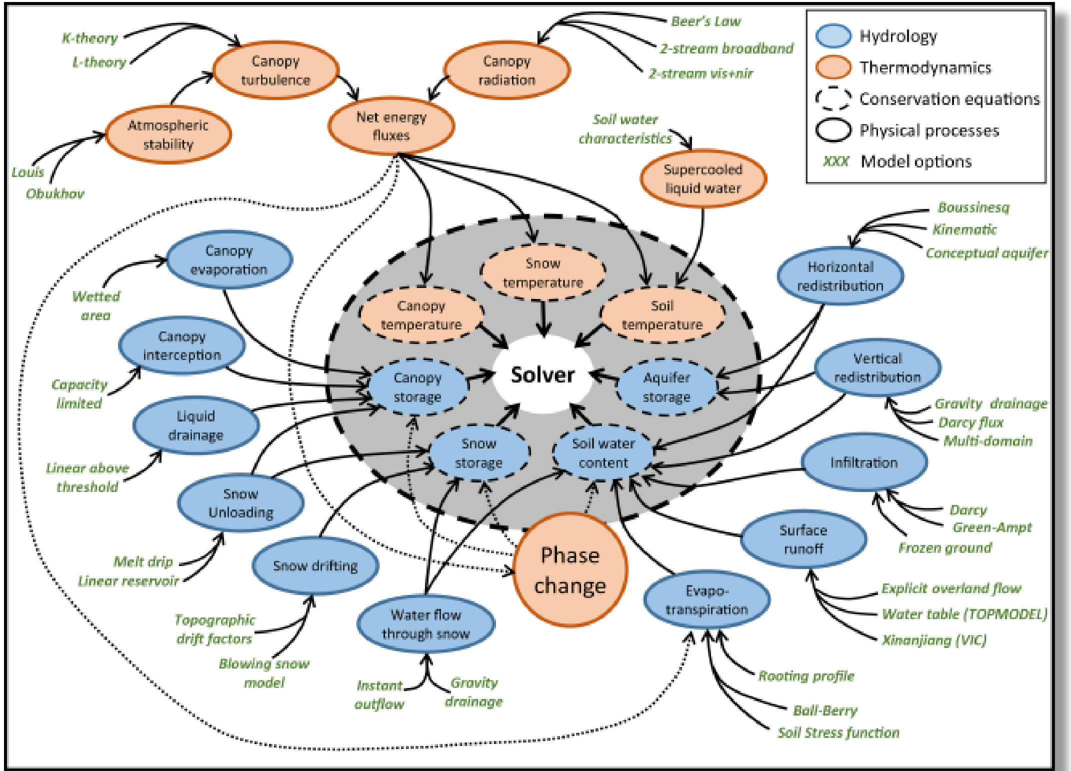
1. Incorporate NIGPs to reduce uncertainty.
2. Validate at each tower individually for 1 year with at least 10 day forecasting (temporal).
3. Perform LOO validation over the towers (spatial).

In the near future

Create NLDAS maps with hourly 0.125 degree resolutio



An Emerging Vision



M. Clark et al. (2015) "A unified approach for process-based hydrologic modeling: 1. Modeling concept" *Water Resources Research*

We need radically novel strategies for merging complex process models with the powerful ability of machine learning to extract information from data.

SSAI

Thanks!