



Toward Integrating E3SM with Machine Learning-enhanced Data Assimilation for Improved Earth System Predictability

Peyman Abbaszadeh¹, Ruby Leung², Yilin Fang², Hoang Tran², Fatemeh Gholizadeh¹ and Ali Gholinia¹

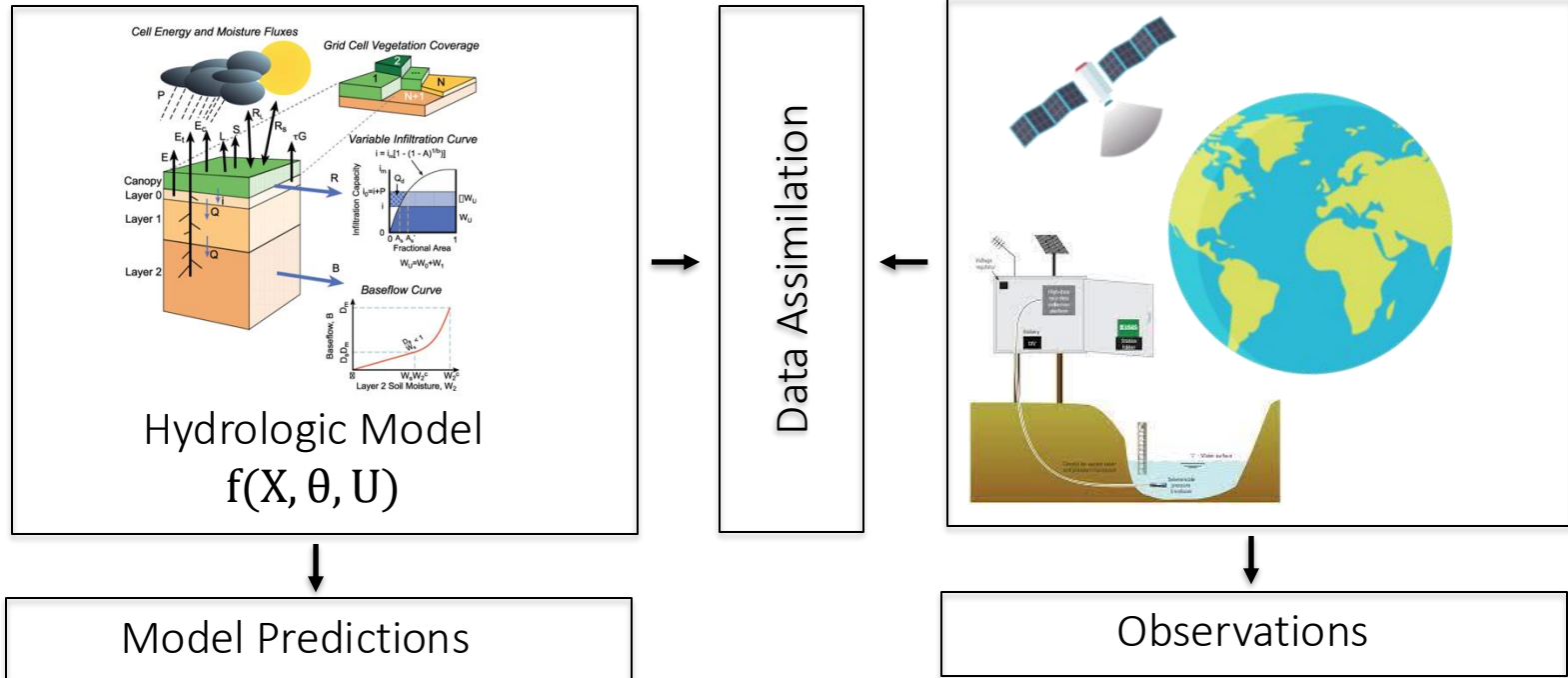
¹Hydrologic Modeling and Assimilation Lab

¹Department of Civil & Environmental Engineering, Portland State University

²Pacific Northwest National Laboratory

E3SM webinar series

Hydrologic Predictions vs. Observations

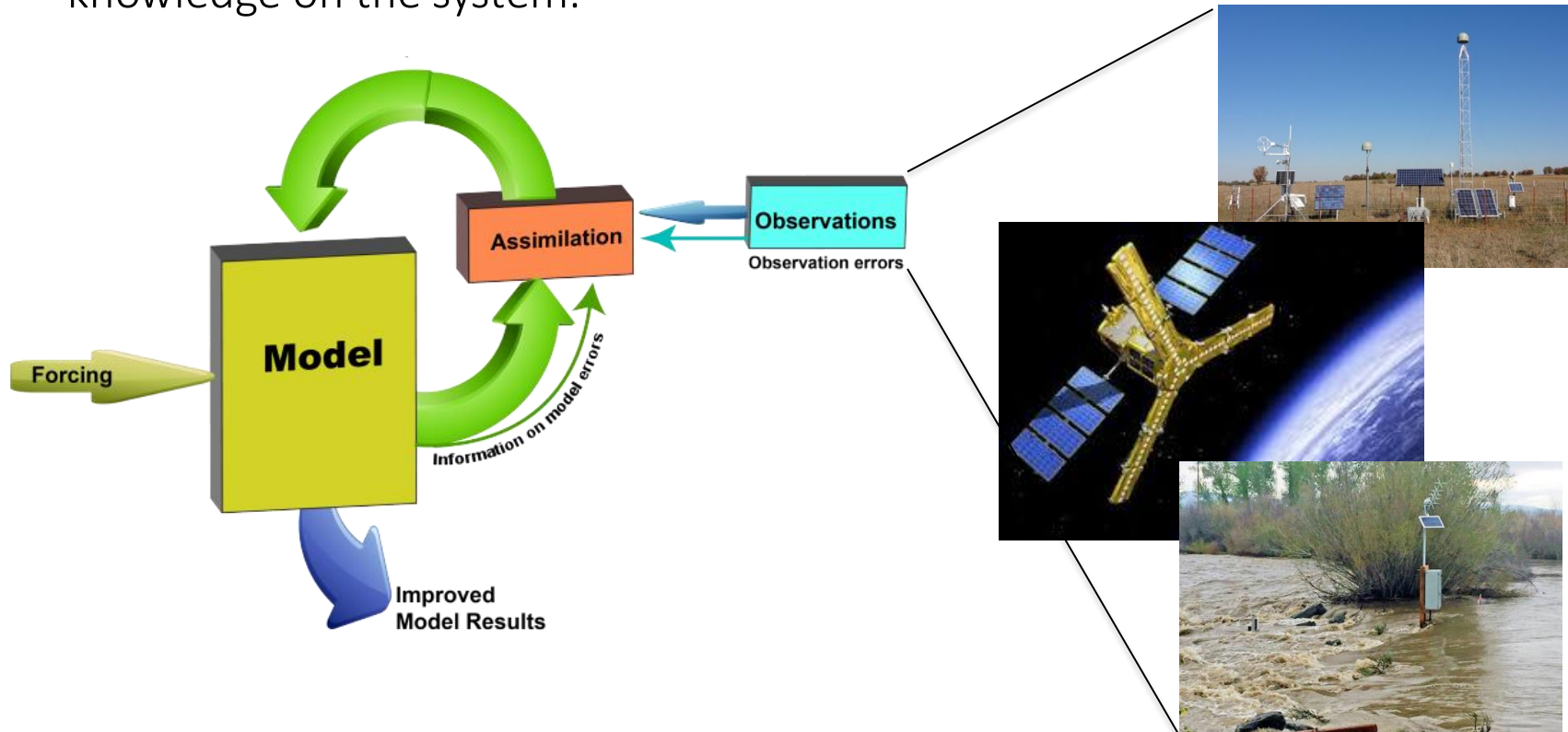


- limitation of measurements and spatiotemporal representativeness of the data
- conceptualization of the model and non-uniqueness of parameters
- imperfect representation of a real system
- Initial and boundary condition uncertainty

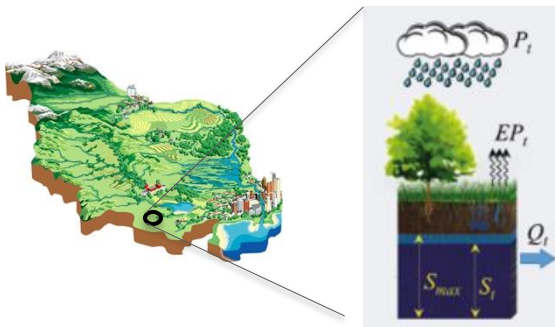
- Satellite retrieval error
- In situ measurement error
- Error in documenting and archiving data

Data Assimilation (DA)

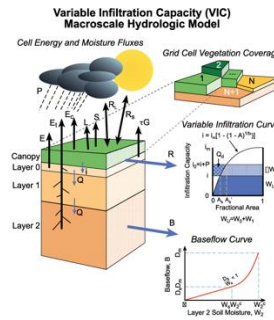
Is defined as the set of statistical techniques that enable to improve knowledge of the past, present or future system states and/or parameters, jointly using experimental data and the theoretical (a priori) knowledge on the system.



Hydrologic Data Assimilation Systems

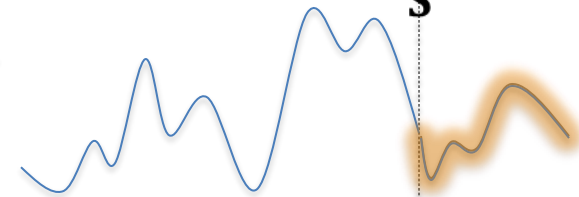


Model Conceptualization



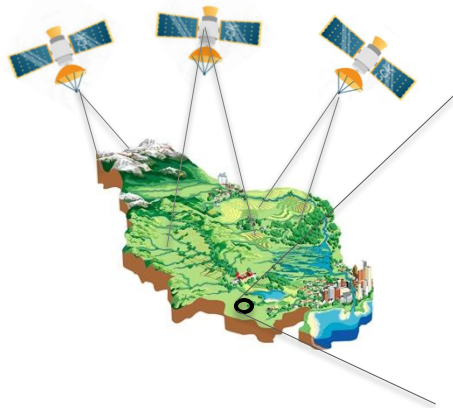
Hydrologic Model

Model calibration
using historical
observation data

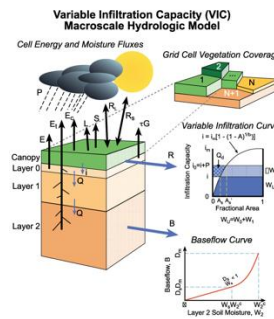


Prediction

Forecasting

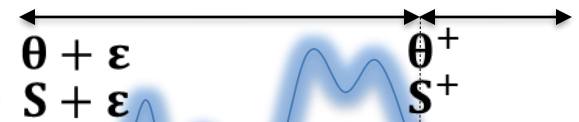


Model Conceptualization



Hydrologic Model

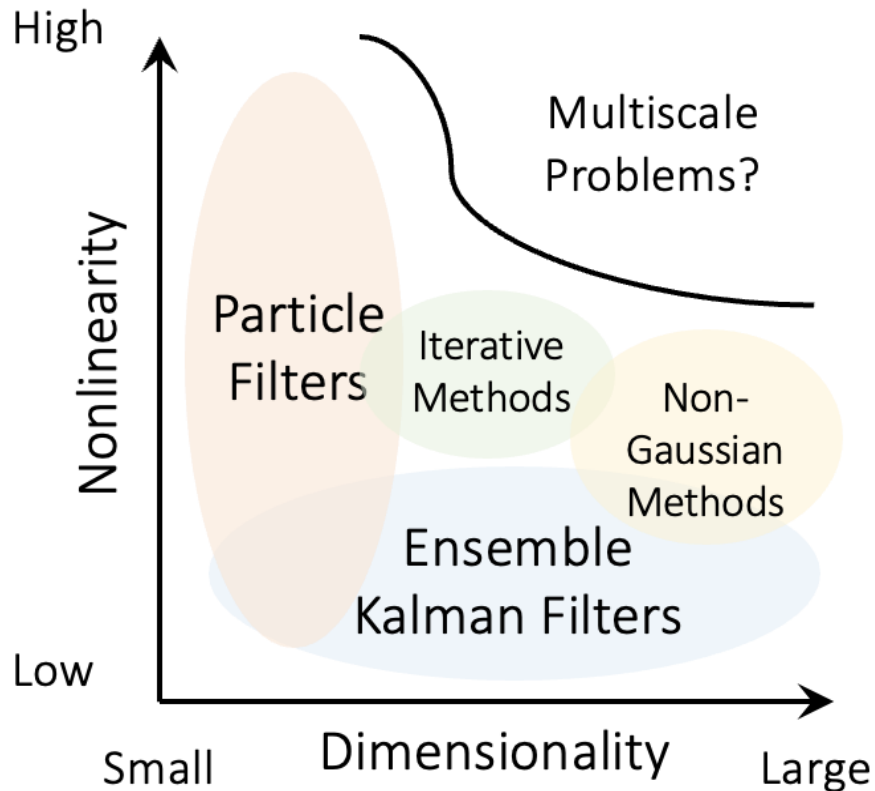
Dual state-parameter Data
Assimilation using satellite data



Prediction

Forecasting

Data Assimilation



Recent applications of DA often rely on the EnKF due to its straightforward implementation and efficiency in handling high dimensional problems.

EnKF has some inherent features which result in suboptimal performance.

These include the Gaussian assumption of errors, *linear updating rule within the EnKF* and violation of water balance that limit its usefulness

ML-Enhanced Data Assimilation (Evolutionary DA)

Hydrologic Data Assimilation with the Ensemble Kalman Filter

ROLF H. REICHLÉ,* DENNIS B. MCLAUGHLIN, AND DARA ENTEKHABI

Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 28 December 2000, in final form 22 March 2001)

WATER RESOURCES RESEARCH, VOL. 41, W05012, doi:10.1029/2004WR003604, 2005

Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter

Hamid Moradkhani and Kuo-Lin Hsu

Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, California, USA

Hoshin Gupta

Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA

Soroosh Sorooshian

Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, California, USA

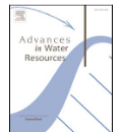
Advances in Water Resources 111 (2018) 192–204



Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



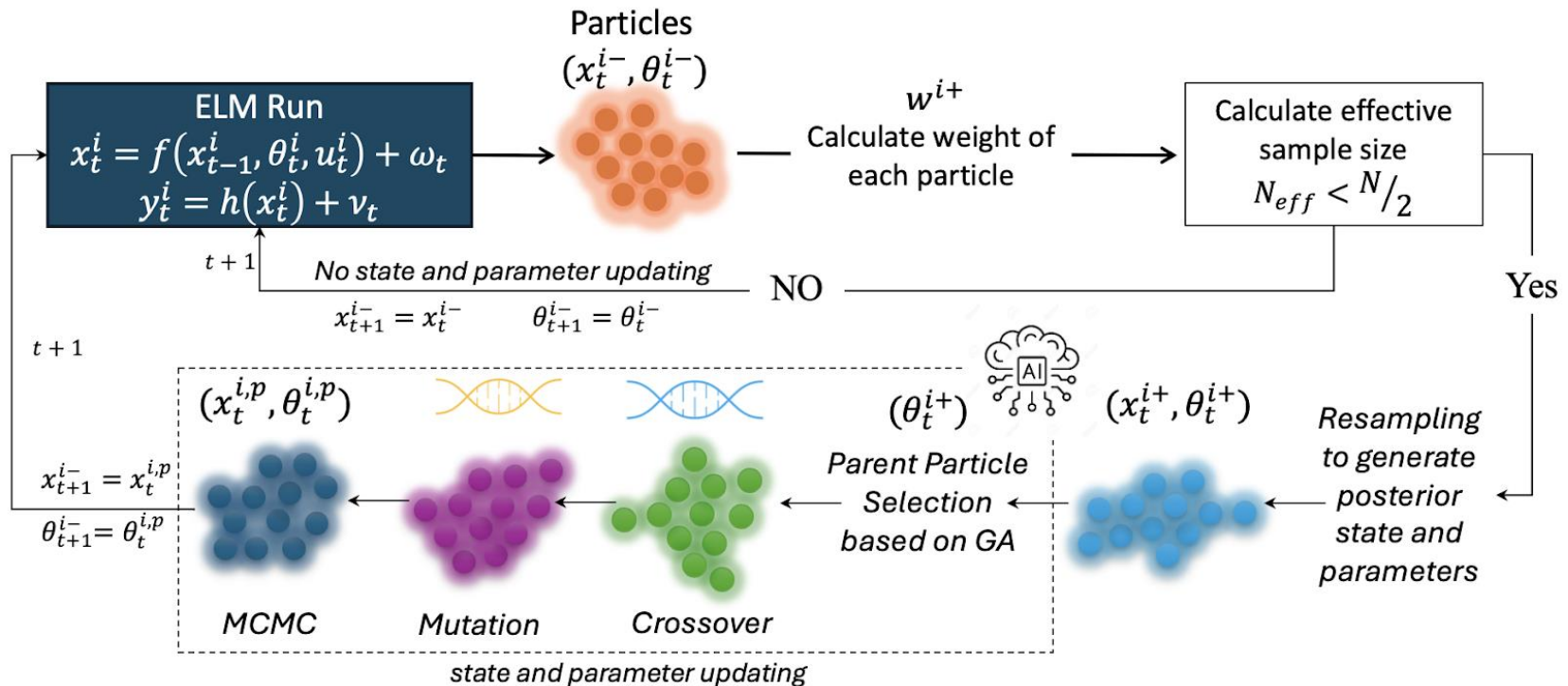
Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo

Peyman Abbaszadeh, Hamid Moradkhani*, Hongxiang Yan

Water Resources and Remote Sensing Lab, Department of Civil and Environmental Engineering, Portland State University, 1930 SW 4th Ave., suite 200, Portland, OR 97201, United States



ML-Enhanced Data Assimilation (Evolutionary DA)



Advances in Water Resources 111 (2018) 192–204

Contents lists available at ScienceDirect



Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo

Peyman Abbaszadeh, Hamid Moradkhani*, Hongxiang Yan

Water Resources and Remote Sensing Lab, Department of Civil and Environmental Engineering, Portland State University, 1930 SW 4th Ave., suite 200, Portland, OR 97201, United States



ML-Enhanced Data Assimilation (Evolutionary DA)

$$x_t = f(x_{t-1}, u_t, \theta_t) + \omega_t$$

$$y_t = h(x_t) + v_t$$

LIKELIHOOD

The probability of "B" being True, given "A" is True

PRIOR

The probability "A" being True. This is the knowledge.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

POSTERIOR

The probability of "A" being True, given "B" is True

MARGINALIZATION

The probability "B" being True.

Bayes' law

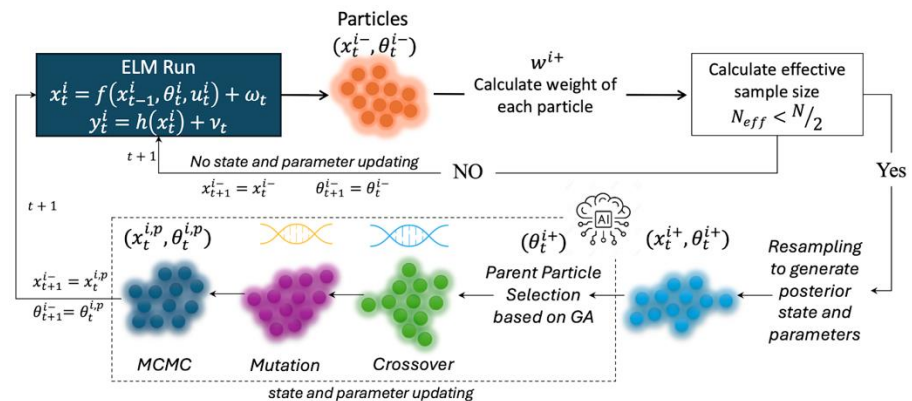
$$p(x_t, \theta_t | y_{1:t}) = p(x_t, \theta_t | y_{1:t-1}, y_t) = \frac{p(y_t | x_t, \theta_t) p(x_t, \theta_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} = \frac{p(y_t | x_t, \theta_t) p(x_t, \theta_t | y_{1:t-1})}{\int p(y_t | x_t, \theta_t) p(x_t, \theta_t | y_{1:t-1}) dx_t d\theta_t}$$

$$p(x_t, \theta_t | y_t) \approx \sum_{i=1}^N w^{i+} \delta(x_t - x_t^{i-}, \theta_t - \theta_t^{i-})$$

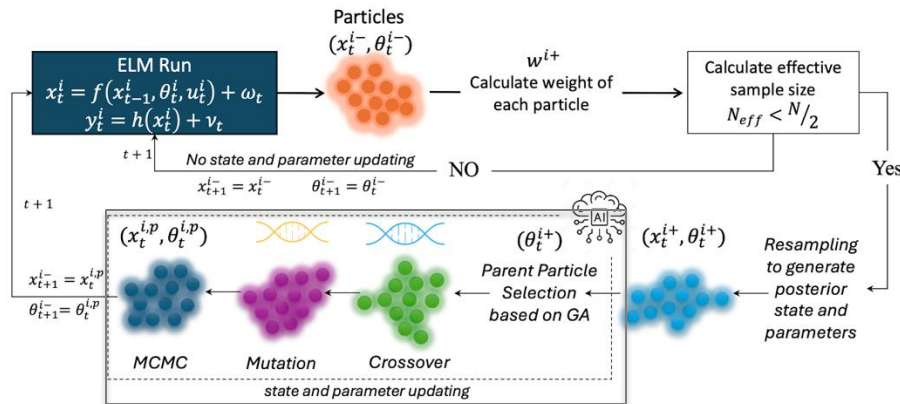
$$w^{i+} = \frac{w^{i-} \cdot p(y_t | x_t^{i-}, \theta_t^{i-})}{\sum_{i=1}^N w^{i-} \cdot p(y_t | x_t^{i-}, \theta_t^{i-})}$$

$$p(y_t | x_t^{i-}, \theta_t^{i-}) = \frac{L(y_t | x_t^{i-}, \theta_t^{i-})}{\sum_{i=1}^N L(y_t | x_t^{i-}, \theta_t^{i-})}$$

$$L(y_t | x_t^{i-}, \theta_t^{i-}) = \frac{1}{\sqrt{(2\pi R)}} \cdot \exp\left[-\frac{1}{2R} (y_t - y_t^i)^2\right]$$



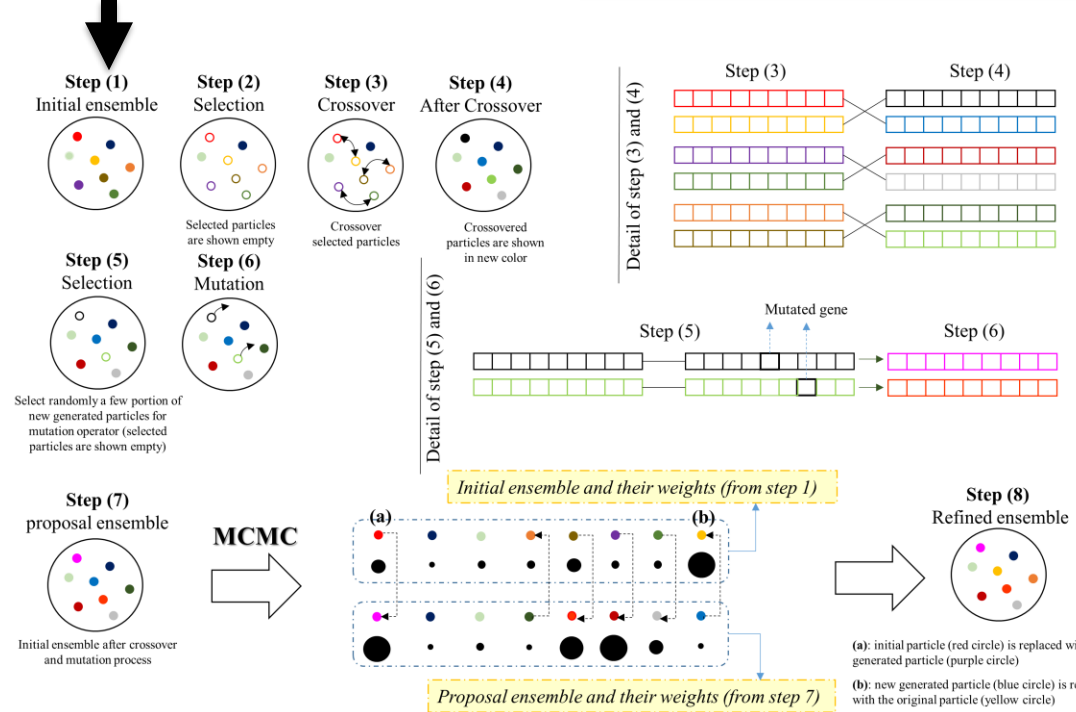
ML-Enhanced Data Assimilation (Evolutionary DA)



$$\alpha = \left(1, \frac{p(\theta_t^{i,p}, x_t^{i,p} | y_{1:t})}{p(\theta_t^{i+}, x_t^{i+} | y_{1:t})} \right)$$

$$p(\theta_t^{i,p}, x_t^{i,p} | y_{1:t}) \propto p(y_t^i | \theta_t^{i,p}, x_t^{i,p}) \cdot p(\theta_t^{i,p}, x_t^{i,p} | y_{1:t-1})$$

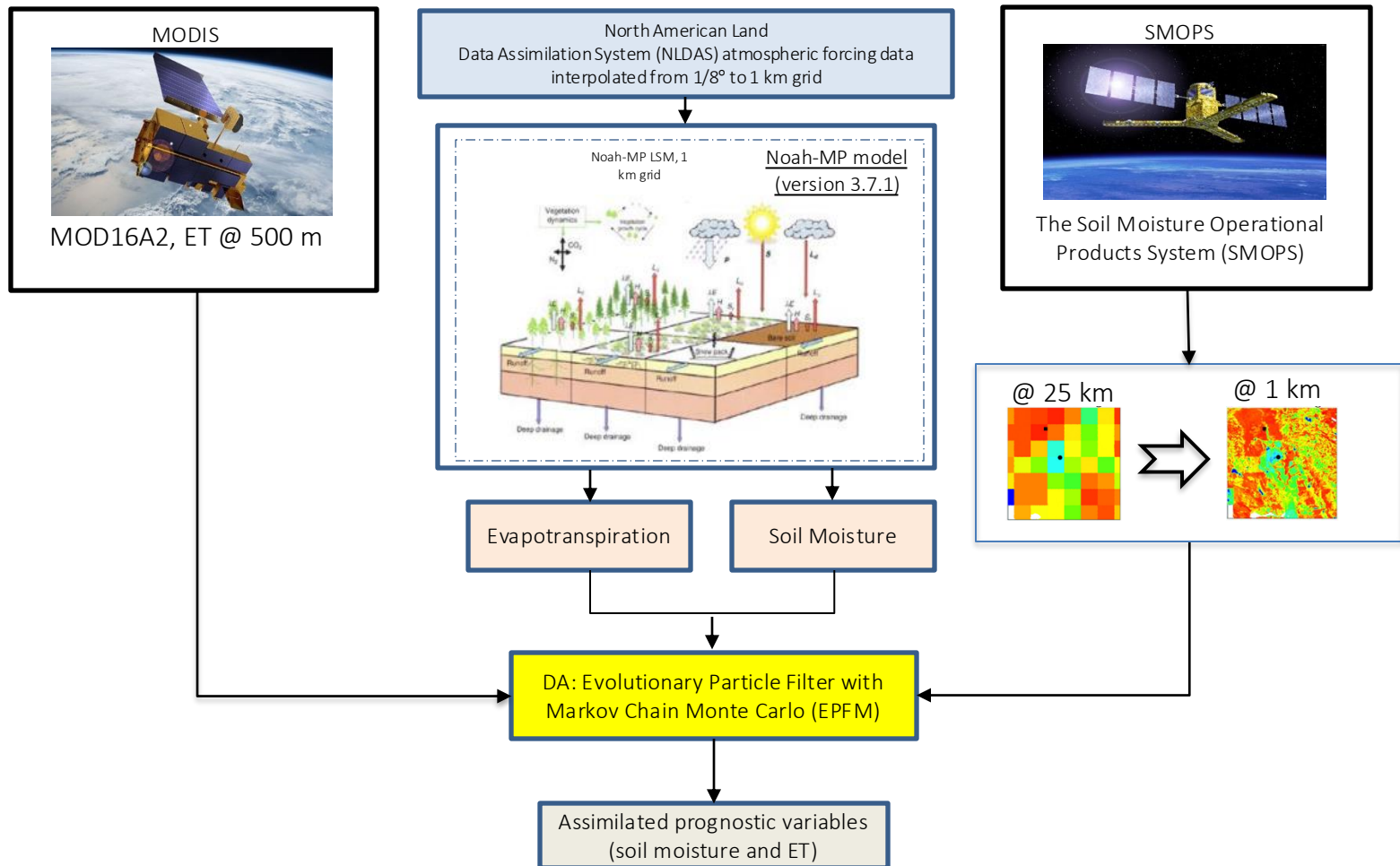
$$x_t^{i,p} = f(x_{t-1}^{i+}, u_t^{i+}, \theta_t^{i,p})$$



ML-Enhanced Data Assimilation (Evolutionary DA)

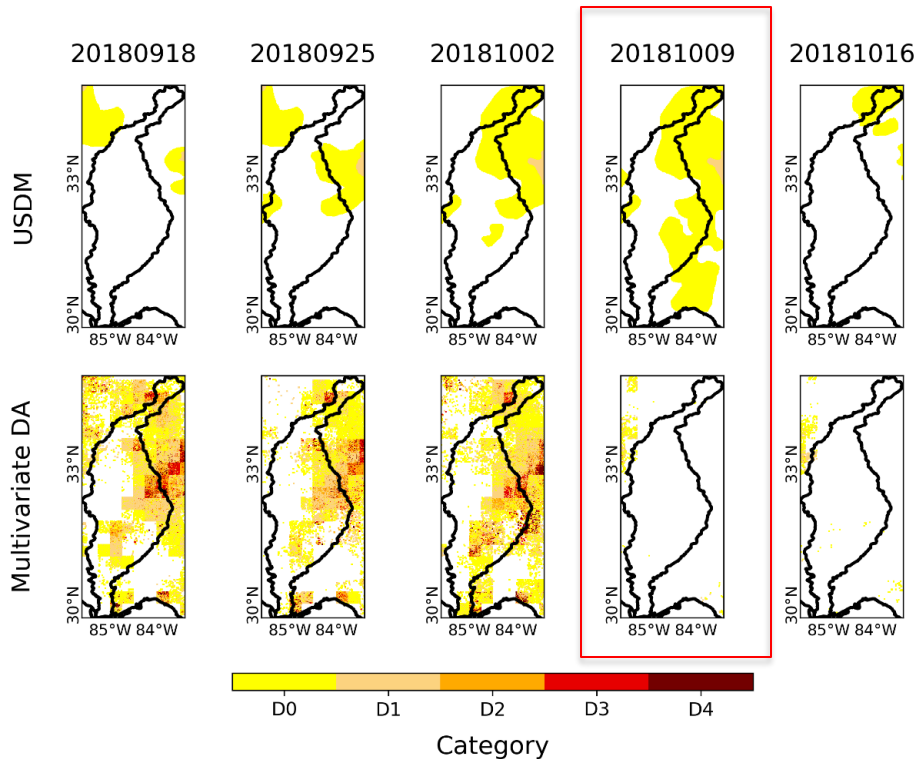
- More informative prior, and better estimation of the posterior distribution
- Minimizes the particle degeneracy and sample impoverishment
- A small ensemble size suffices to set up the DA
- Effective DA method to be used within high dimensional hydrologic and LSMs

Drought Predictability



Drought Predictability

How drought map could capture the footprint of the torrential rainfall of Hurricane Michael



Hurricane Michael

2018 tropical cyclone

Hurricane Michael was a very powerful and destructive tropical cyclone that became the first Category 5 hurricane to strike the contiguous United States since Andrew in 1992. [Wikipedia](#)

Highest wind speed: 162 mph

Date: October 7, 2018 – October 16, 2018

Category: Category 5 Hurricane (SS)

Damage: \$25.1 billion (2018 USD)

Location: Florida

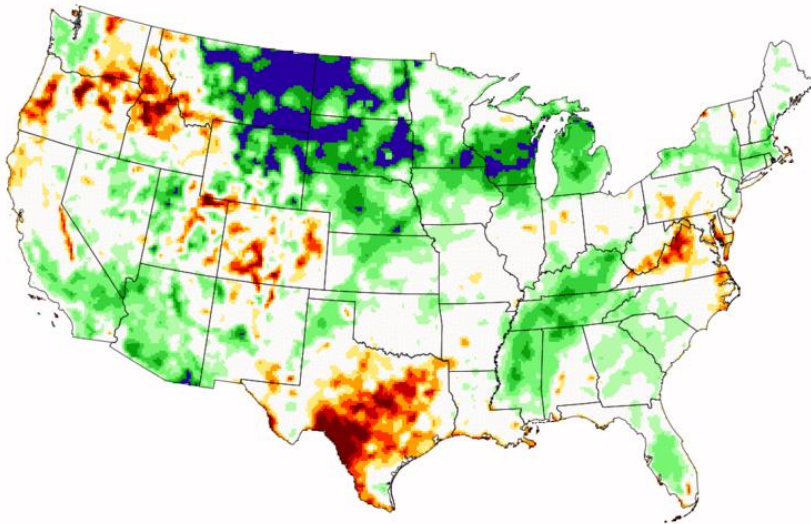
Affected areas: Florida, Cuba, Mexic



Drought Predictability

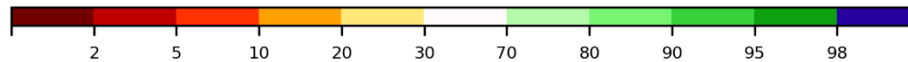
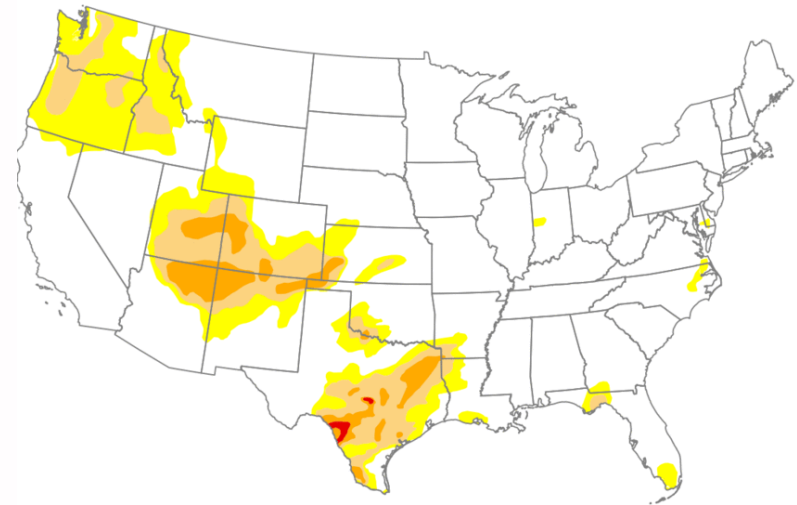
Integrated drought monitoring
system

01/07/2020



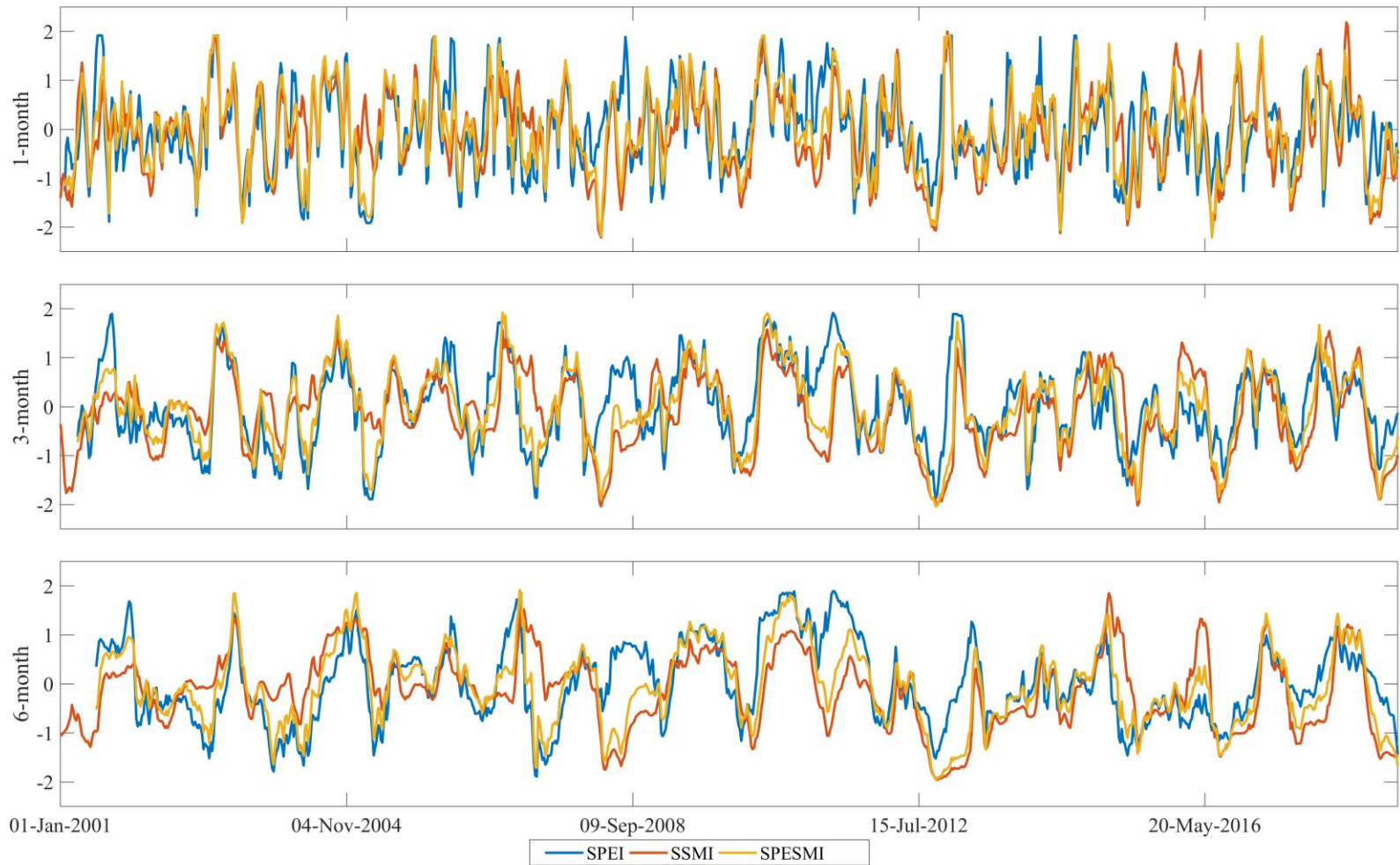
USDM system

01/07/2020

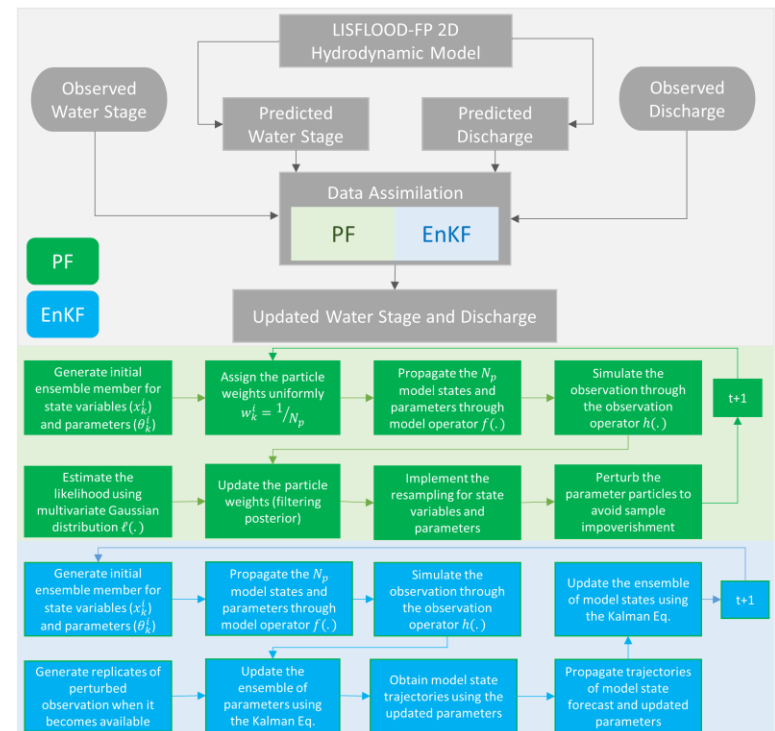
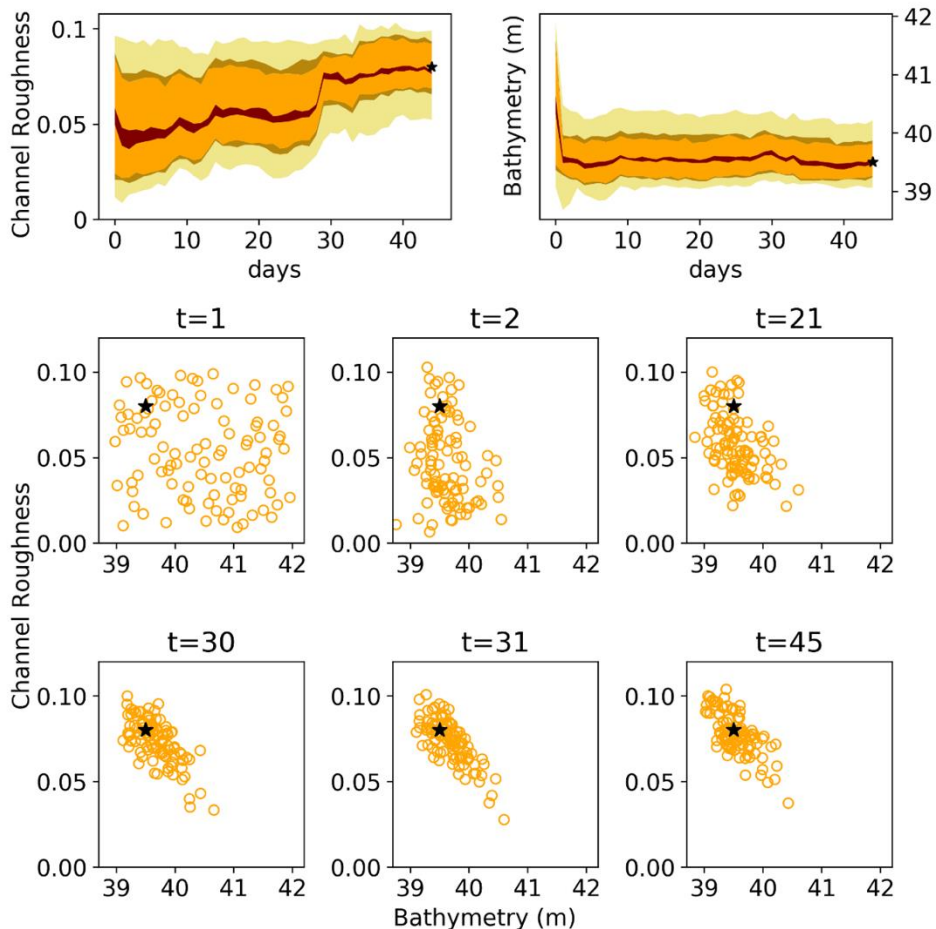


Soil Moisture Percentile

Drought Predictability

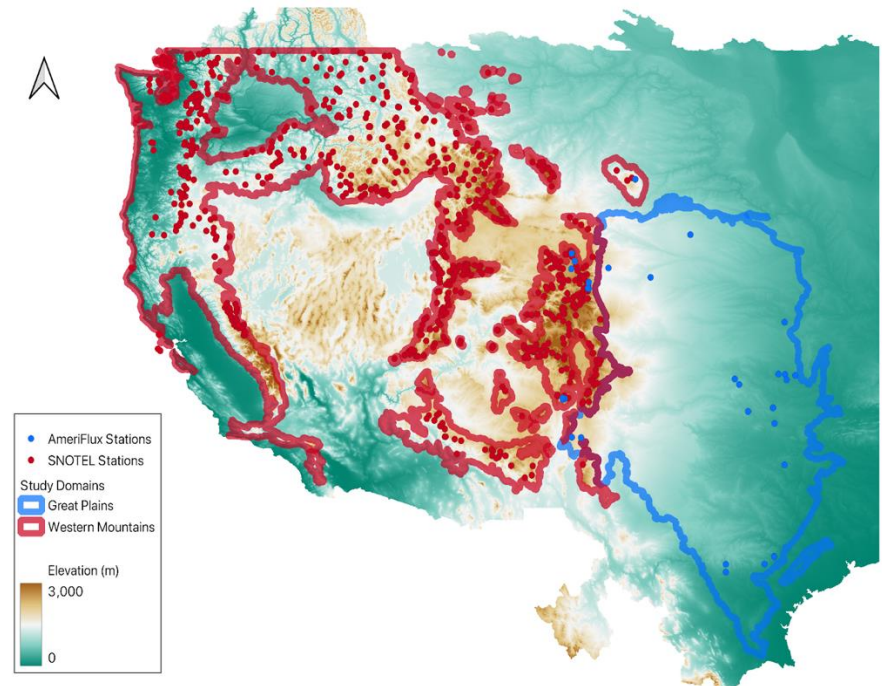


Flood Inundation Modeling



Coupling ELM with ML-enhanced Data Assimilation

Integrating E3SM Land Model (ELM), with a state-of-the-art data assimilation technology, ML-enhanced DA.



Developing a cyberinfrastructure that provides a capability to effectively assimilate satellite observations into ELM and improve E3SM predictability while accounting for uncertainties.

Contribution to E3SM S2D

Data assimilation is expected to provide **more accurate and reliable initial condition** for the ELM by integrating observations—such as soil moisture, snow and vegetation- related state variables into the model while accounting for **uncertainties**.

As a result, the **improved initialization** can translate into enhanced E3SM predictability, particularly on **seasonal to decadal timescales**, where land memory (e.g., soil moisture or snowpack anomalies) plays a crucial role in modulating longer-term variability.

Contribution to E3SM S2D

- Soil Moisture Impacts: E3SM simulations show soil moisture variability drives 67–82% of temperature differences and 85–91% of carbon flux differences between **El Niño** and **La Niña** years (Levine et al. 2019)
 - Accurate simulation of soil moisture will significantly enhance E3SM seasonal-to-decadal (S2D) predictions, particularly when employing online coupling techniques.
- SWE Biases: E3SM exhibits SWE biases across North America during **ENSO** years, due to limitations in capturing mountainous snowpack dynamics (Fang and Leung, 2023).
 - Enhanced DA modules developed in this project are expected to improve SWE and soil moisture representation in key study domains.

Coupling ELM with ML-enhanced Data Assimilation

Objectives

Assimilating satellite snow related variable observations into ELM *to improve the accuracy and reliability of SWE predictions along with other snow properties including snow cover fraction and snow depth.*

Assimilating the GLASS (Global LAnd Surface Satellite) LAI (Leaf Area Index) and SMAP (Soil Moisture Active Passive) satellite soil moisture into ELM *to improve land-atmosphere feedback in E3SM and therefore its drought predictability over the Great Plains.*

Coupling ELM with ML-enhanced Data Assimilation

Meteorological Forcing Perturbation:

- Generated 50 ensemble members of meteorological forcing data
- Applied to **Temperature, Precipitation, Longwave, Shortwave radiation**

Methodology:

- Perturbation framework follows Reichle et al. (2007) (used in NASA LIS)

Steps:

- Generate temporally correlated noise series (AR(1))
- Apply cross-correlations across variables
- Transform to physical distributions
- Apply perturbations to meteorological forcing inputs

Coupling ELM with ML-enhanced Data Assimilation

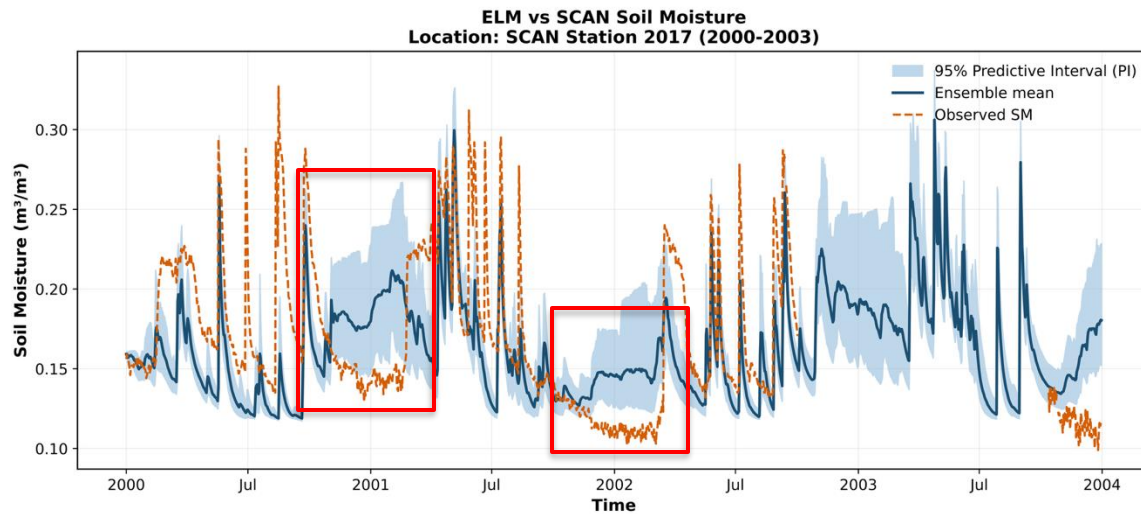
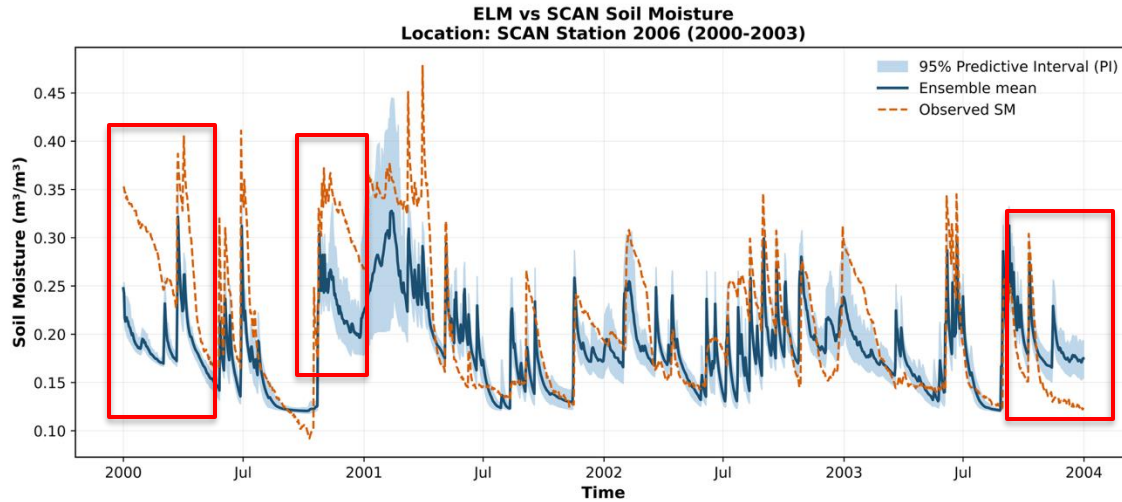
Reichle et al. (2007)

Variable	Type	Distribution	Perturbation Magnitudes
<i>TBOT (Temperature)</i>	Additive	Normal	$\sigma = 5.0 \text{ K}$
<i>FLDS (Longwave)</i>	Additive	Normal	$\sigma = 50.0 \text{ W/m}^2$
<i>PRECTmms (Precipitation)</i>	Multiplicative	Log-normal	CV = 25%
<i>FSDS (Shortwave)</i>	Multiplicative	Log-normal	CV = 30%

Variable	TBOT	PRECTmms	FLDS	FSDS
TBOT (Temperature)	1.0	-0.3	0.4	0.5
PRECTmms (Precipitation)	-0.3	1.0	0.5	-0.5
FLDS (Longwave)	0.4	0.5	1.0	-0.5
FSDS (Shortwave)	0.5	-0.5	-0.5	1.0

warm temperatures \rightarrow more sunlight \rightarrow more shortwave radiation

OL ELM Simulation (based on uncertain atmospheric forcing)



ELM SM & ET related parameters

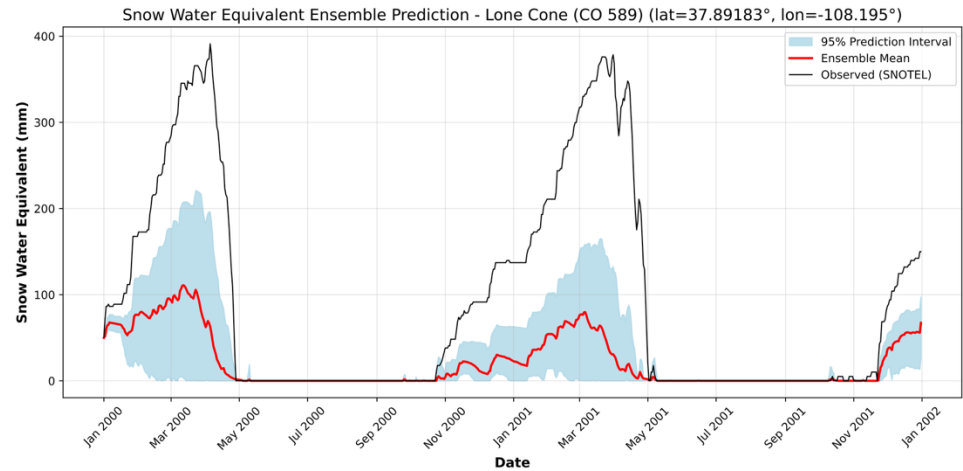
Number	Parameter	Long name	Description	Units	Min Value	Max Value	Ref
1	b (in paper), bbbopt (in model)	Ball-Berry intercept parameter	Sensitivity of stomatal conductance to photosynthesis/humidity in the Ball-Berry model	mmol H ₂ O m ⁻² s ⁻¹	0	275	Cai et al., 2019
2	leafcn (in paper), fleafcn (in model)	Leaf C:N ratio	Carbon to nitrogen ratio of leaf tissue	gC g N ⁻¹	12.5	70	Ricciuto et al., 2018
3	f_{nr} (in paper), flnr (in model)	Fraction of leaf nitrogen in RuBisco	Portion of leaf nitrogen allocated to the Rubisco enzyme	gN in Rubisco gN ⁻¹	0.0231	0.264	Cai et al., 2019
4	frootcn	Fine root C:N ratio	Carbon to nitrogen ratio in fine roots	gC g N ⁻¹	21	63	Ricciuto et al., 2018
5	leafcn	Leaf C:N ratio	Carbon to nitrogen ratio of leaf tissue	gC g N ⁻¹	12.5	70	Ricciuto et al., 2018
6	m (in paper), mbbopt (in model)	Ball-Berry slope parameter	Sensitivity of stomatal conductance to photosynthesis/humidity in the Ball-Berry model	dimensionless	4.5	13.5	Cai et al., 2019

ELM SM & ET related parameters

Number	Parameter	Long name	Description	Units	Min Value	Max Value	Ref
7	roota_par	Rooting depth distribution parameter a	hape parameters of the vertical root profile (how roots are spread with depth)	m ⁻¹	3.5	16.5	Ricciuto et al., 2018
8	rootb_par	Rooting depth distribution parameter b	hape parameters of the vertical root profile (how roots are spread with depth)	m ⁻¹	0.5	4.5	Ricciuto et al., 2018
9	slatop	Specific leaf area at canopy top	Leaf area per unit carbon for sunlit/top leaves (m ² gC ⁻¹)	m ² gC ⁻¹	0.002	0.03	Ricciuto et al., 2018
10	smpsc	Soil water potential at stomatal closure	Threshold where stomata are fully closed	mm	-642,000	-125,000	Ricciuto et al., 2018
11	smpso	Soil water potential at stomatal onset	Soil matric potential above which stomata can be fully open	mm	-125,000	-17,500	Ricciuto et al., 2018
12	soilpsi_off	Critical soil water potential for leaf offset/senescence	Water status threshold for ending leaf activity/senescence	MPa	-5	-0.75	Ricciuto et al., 2018
13	soilpsi_on	Critical soil water potential for leaf onset	Water status threshold to start leafing/active season	MPa	-5	-0.75	Ricciuto et al., 2018

OL ELM Simulation (based on uncertain atmospheric forcing)

The inadequate snow albedo treatments in CLM have been identified as a contributing factor to model biases in simulating surface albedo and snow-pack evolution (e.g., Chen et al., 2014; Thackeray et al., 2019; Toure et al., 2018).



Identifying the snow albedo parameters for assimilation purposes (snow albedo strongly influences the surface energy balance, melt rates, and the snowpack's evolution)

Parameter	Unit	Definition	Default value	Range
C_a	Unitless	Tunning parameter for top snow/soil layer thickness	0.34	[0.0, 1.0]
T_{fall}	K	Snowfall temperature	274.15	[273.15, 275.15]
K_{accum}	Unitless	Snow cover accumulation parameter	0.1	[0.02, 0.4]
N_{melt}	Unitless	Snow cover depletion curve shape parameter	200.0	[50, 400]
T_{freeze}	K	Snow freezing temperature	273.15	[271.15, 275.15]

Peyman Abbaszadeh
Assistant Professor
Hydrologic Modeling and Assimilation Lab
Department of Civil & Environmental Engineering
Portland State University



pabbaszadeh@pdx.edu
pabbaszadeh.github.io



Cannon Beach, Oregon