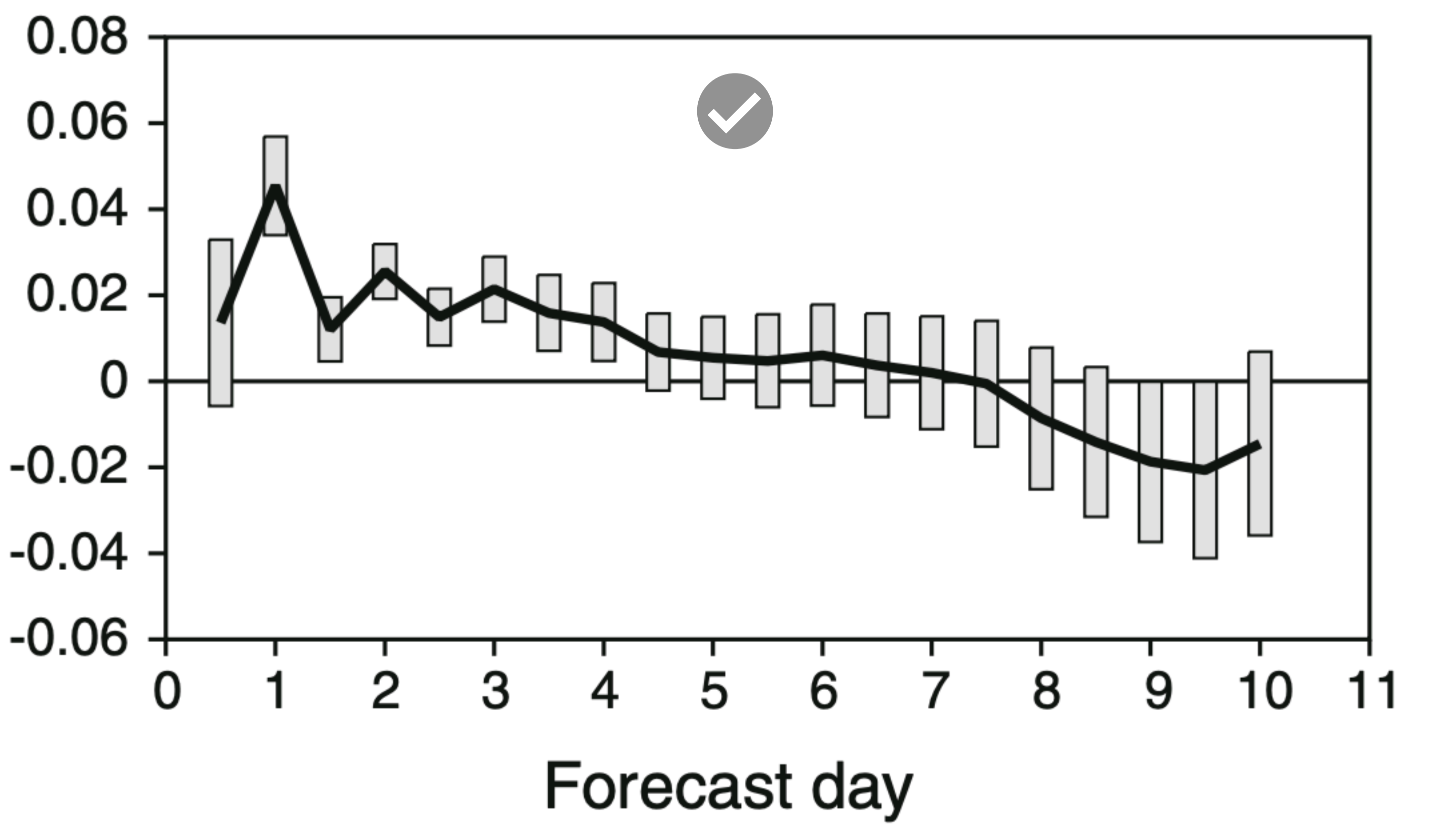
Progress Towards A State-Of-The-Art Land Data Assimilation System For NOAA's Global NWP System

Clara Draper, Sergey Frolov, and Jeff Whitaker (NOAA PSL),
Tseganeh Gichamo (NOAA PSL & CIRES),
Andy Fox (JCSDA),
Mike Barlage (NOAA EMC),
Jiarui Dong, Azadeh Gholoubi, and Youlong Xia (IMSG @ NOAA EMC)

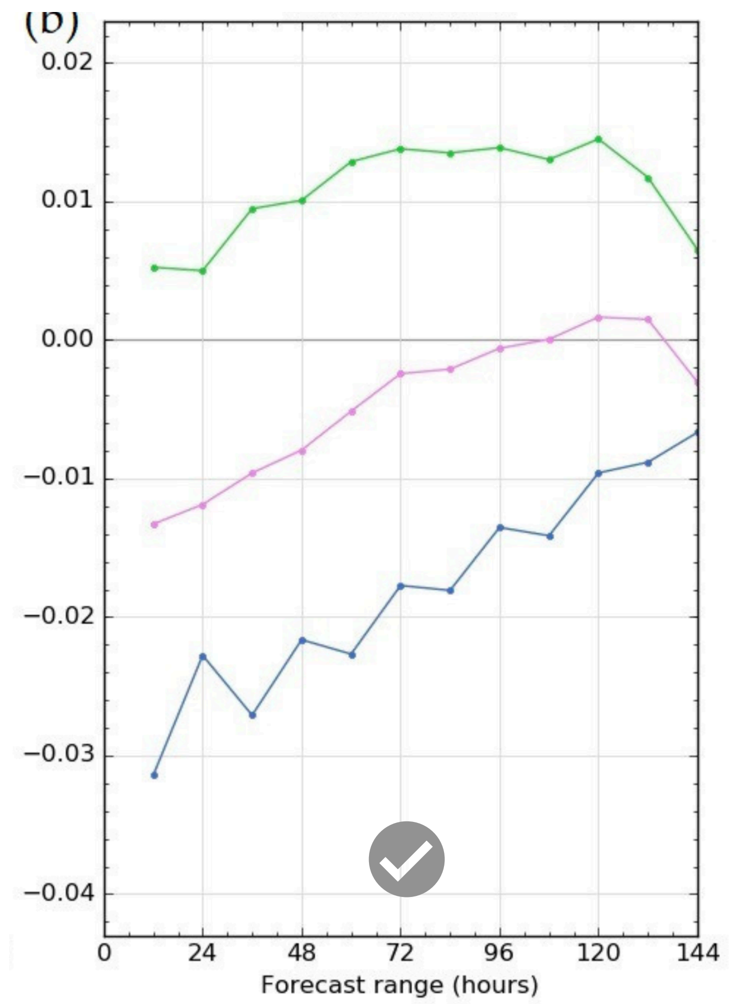
Background

- All major international NWP centers use land DA in their NWP systems to improve initialization of land states, leading to improved NWP forecasts
  - NOAA's land DA system is currently behind international practice
  - We are developing a new land DA system for initial implementation in GFSv17

- Today:
  - Outline the long term goals for land DA in the GFS, and review progress towards the initial implementation
  - Present examples to demonstrate the potential improvements to NWP forecast skill from land DA, and to highlight the current opportunity to better integrate land and atmospheric DA



Improvement in normalized root mean square forecast 1000 hPa geopotential error [-] at ECMWF, from updating the snow depth analysis (de Rosnay et al, 2014).



Change in forecast T2m RMSE [K] at UKMO due to the SEKF soil moisture analysis (three different versions shown; Gomez et al, 2020).

Land DA for Global NWP

Model State		Assimilated Observations	DA Method
Snow amount (snow depth/SWE)	NOAA	SNODEP snow depth analysis, Satellite snow cover (IMS)	Simple “window” method
	Internationally	Station snow depth, Satellite snow cover (IMS)	EnKF, OI, Cressman
Snow temperature	NOAA	-	-
	Internationally	Screen-level T	OI
Soil moisture	NOAA*	-	-
	Internationally	Screen-level T, q Satellite soil moisture (ASCAT retrieval, SMOS Tb)	Simplified EKF, OI
Soil temperature	NOAA	-	-
	Internationally	Screen-level T	OI

* No soil moisture DA, but is retrospectively corrected for precipitation errors.

Status of Land DA for Global NWP at NOAA

Model State		Assimilated Observations	DA Method
Snow amount (snow depth/SWE)	NOAA	SNODEP snow depth analysis, Satellite snow cover (IMS)	Simple “window” method
	Internationally	Station snow depth, Satellite snow cover (IMS)	EnKF, OI, Cressman
Snow temperature	NOAA	-	-
	Internationally	Screen-level T	OI
Soil moisture	NOAA*	-	-
	Internationally	Screen-level T, q Satellite soil moisture (ASCAT retrieval, SMOS) Tb)	Simplified EKF, OI
Soil temperature	NOAA	-	-
	Internationally	Screen-level T	OI

* No soil moisture DA, but is retrospectively corrected for precipitation errors.

The GFS Land DA System

- GFS Land DA system design:
  - Use modern software languages and design
  - Perform the land DA on the native model grid
  - Assimilate observations as directly as possible (limit unnecessary pre-processing and aggregation / use of gridded products)
  - Expand data sets being assimilated, and states being updated

- First priorities (targeted for GFSv17):
  - Update the snow depth analysis
  - Introduce a soil temperature and soil moisture analysis

- Associated system upgrades:
  - GFSv17 will use the Noah-MP land surface model in place of Noah
  - NOAA is adopting the JCSDA's JEDI platform for all DA components

Land DA Software Design: Adopting JEDI

- JEDI (Joint Effort for Data assimilation Integration) is a unified DA system for Earth system prediction, being developed by the JCSDA and partner organizations
  - Key feature is separation of concerns -> model and DA code are completely separate and transparent to each other
  - Specifically using the JEDI fv3-bundle (developed for atmos DA) to do the land DA

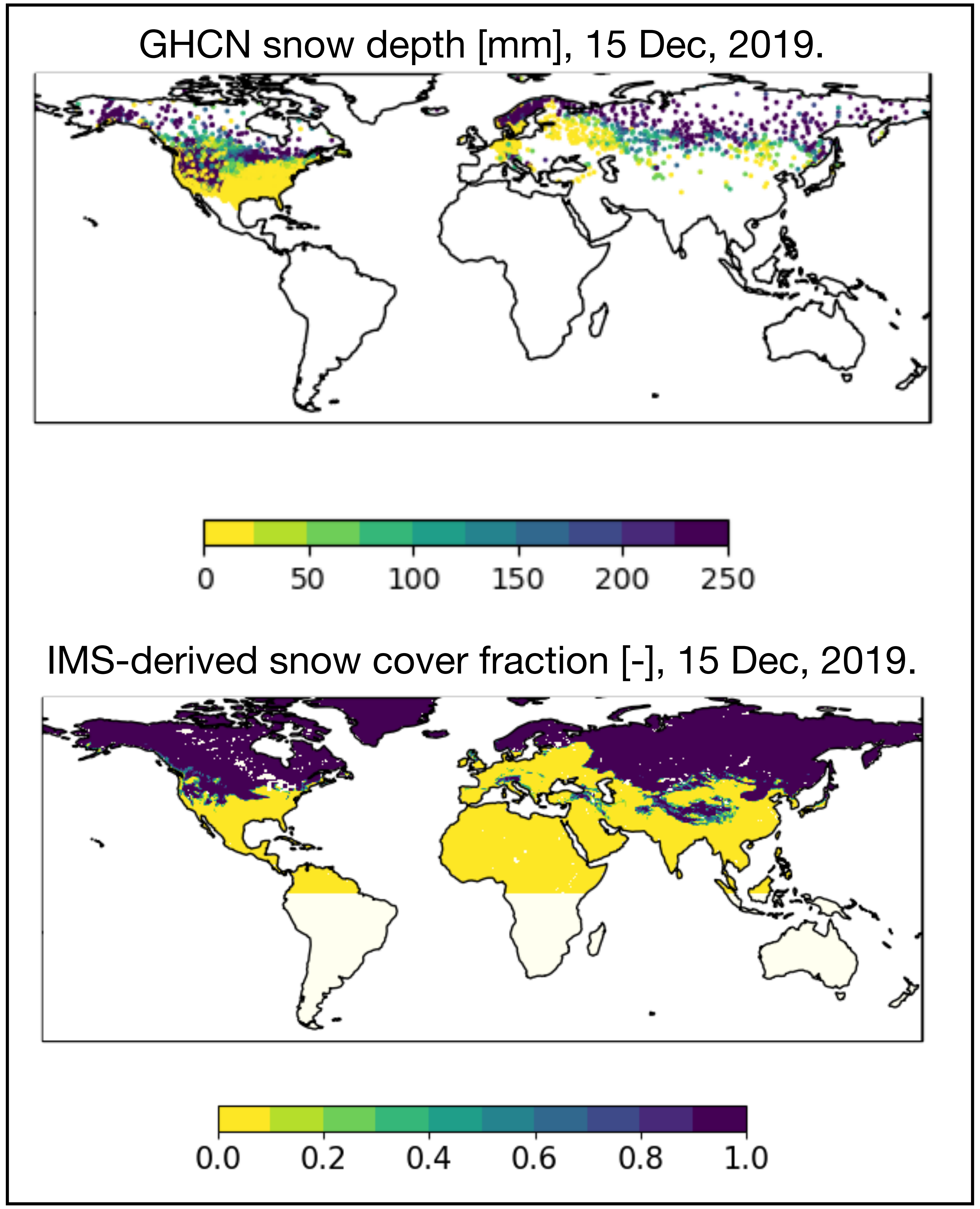
- Use of JEDI satisfies the land DA software design criteria:
  - Use modern coding languages and practices
  - Ability to perform the land DA on the native model grid

- Other advantages:
  - Land DA can leverage off the much larger atmospheric DA effort
  - DA code separated from model code -> flexibility to apply same land DA within atmospheric system or an offline model; or together with the atmospheric DA

Snow Analysis

Updating the Snow Depth Analysis

- NOAA's global NWP model currently has a very simple snow analysis, in which an externally produced snow depth analysis is merged with the forecast snow depth
- Most other national NWP centers use a 2D Optimal Interpolation (OI) or Cressman analysis of station snow depth and (often) remotely sensed snow cover
- For GFSv17, we have implemented 2D OI assimilation of station snow depth and satellite snow cover, based on schemes used elsewhere (EC, ECMWF)



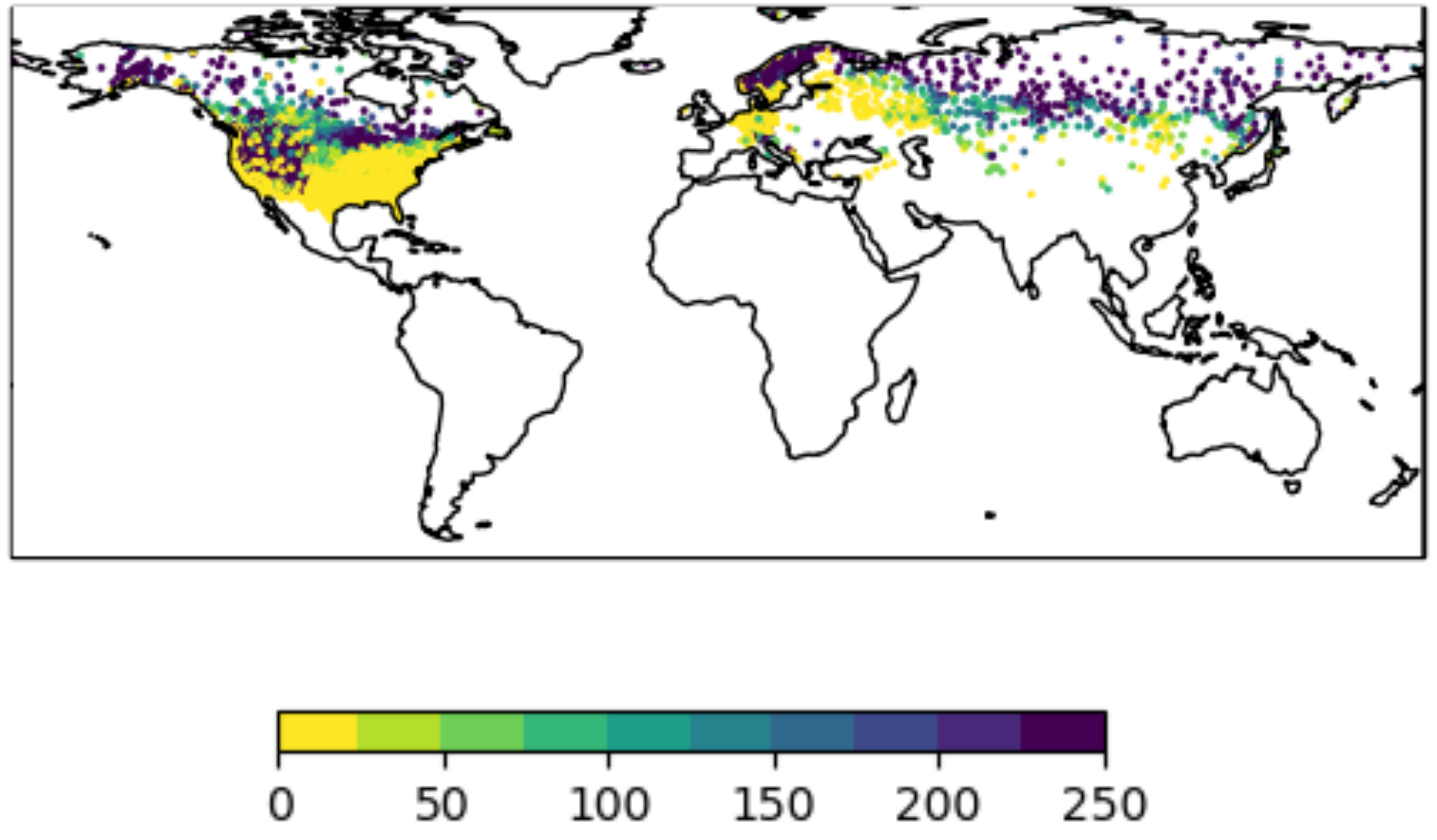
Gichamo and Draper, (sub), WAF.

Assimilated Snow Observations

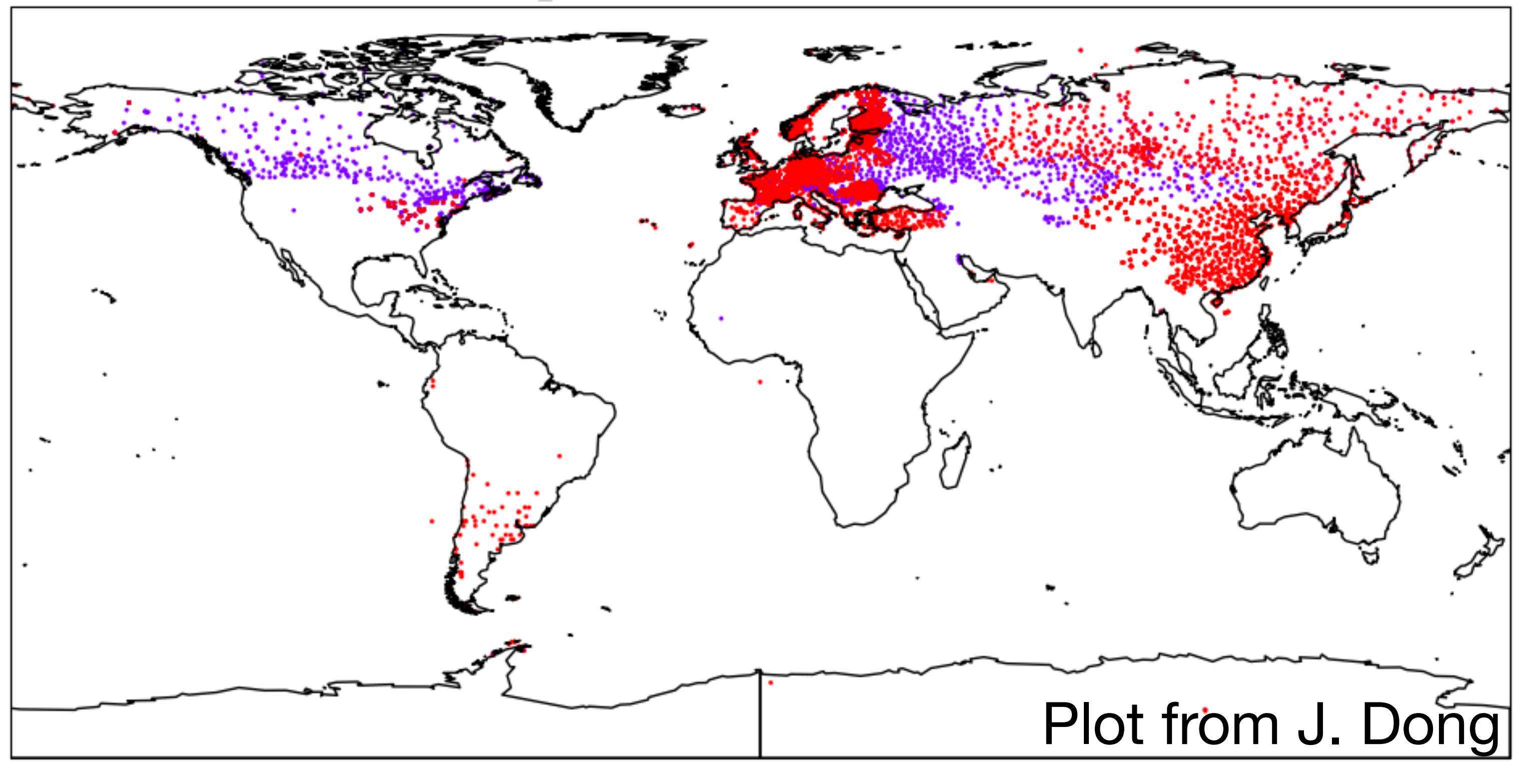
- Snow depth:
  - GTS snow depth observations for NRT NWP use not currently used or downloaded at NCEP
  - Initial testing of the snow depth OI is being done with GHCN observations (not available NRT)
  - EMC is working on observation processing and QC of the GTS observations, also adding US National Network observations

- Snow cover:
  - Assimilating NOAA's IMS snow cover product (4 km version)
  - NCEP (and other centers) currently assimilate IMS
  - IMS is a once daily product, aggregated from multiple satellites with input from analysts - does not satisfy design criteria to assimilate observations more directly
    - In future, will investigate replacement with VIIRS snow cover observations

GHCN snow depth [mm], 15 Dec, 2019.

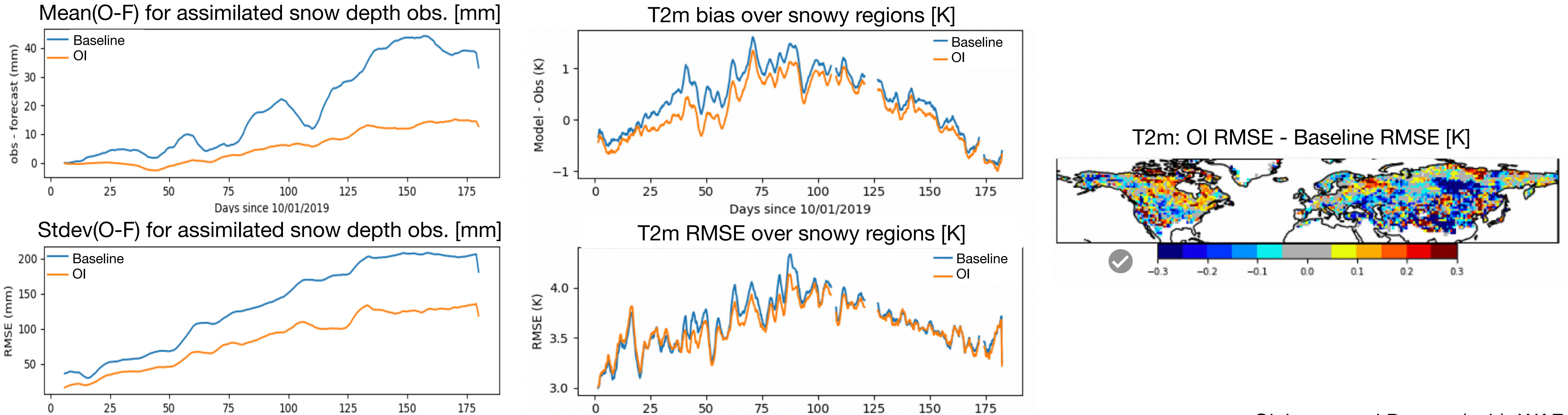


GTS snow depth observation locations (TAC, SYNOP), 1 Feb, 2021.



Improvements to GFS from Snow OI

- Evaluated the snow depth OI with GFSv15/GDAS (Noah land model) at ~1 degree, from October 2019 - April 2020, assimilating GHCN snow depth and IMS snow cover
- Compared to the current snow analysis, the snow depth OI:
    - Improved the snow depth background forecast
    - Improved the T2m over snow-affected land (largely due to improved biases)



Gichamo and Draper, (sub), WAF.

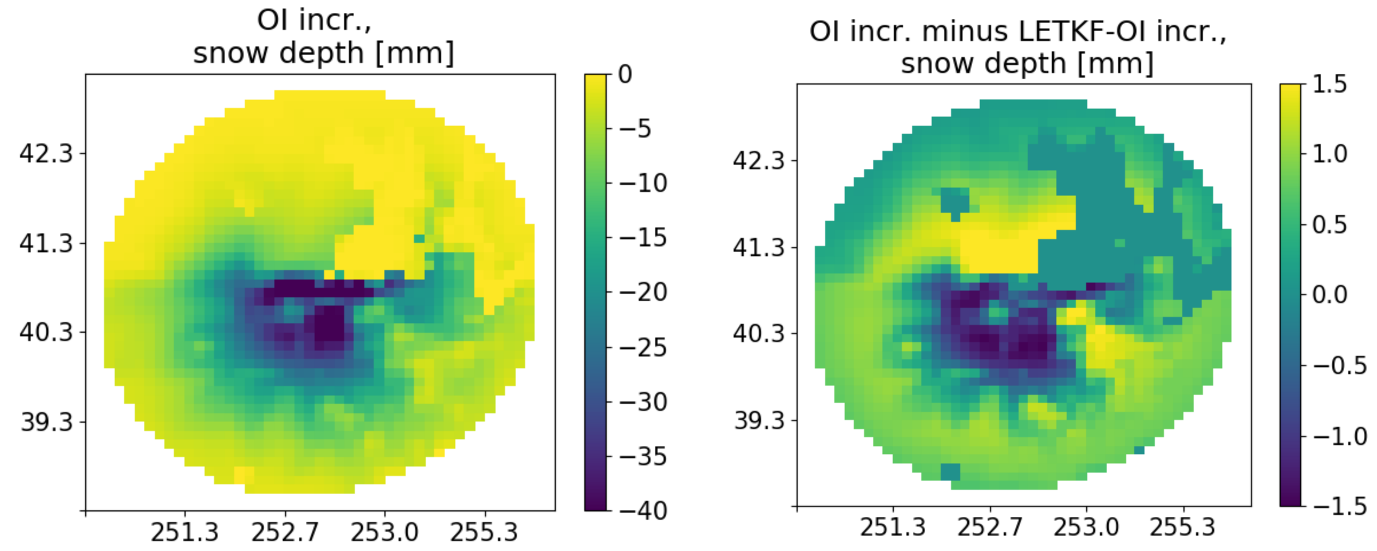
Transition to GFSv17

- Based on these results, the OI is being prepared for implementation in GFSv17
- Requires:
  - Conversion to JEDI
  - Update to Noah-MP land model used in GFSv17
  - For NWP application, replace once daily assimilation of GHCN station snow depth obs with 6-hrly assimilation of station obs from the GTS

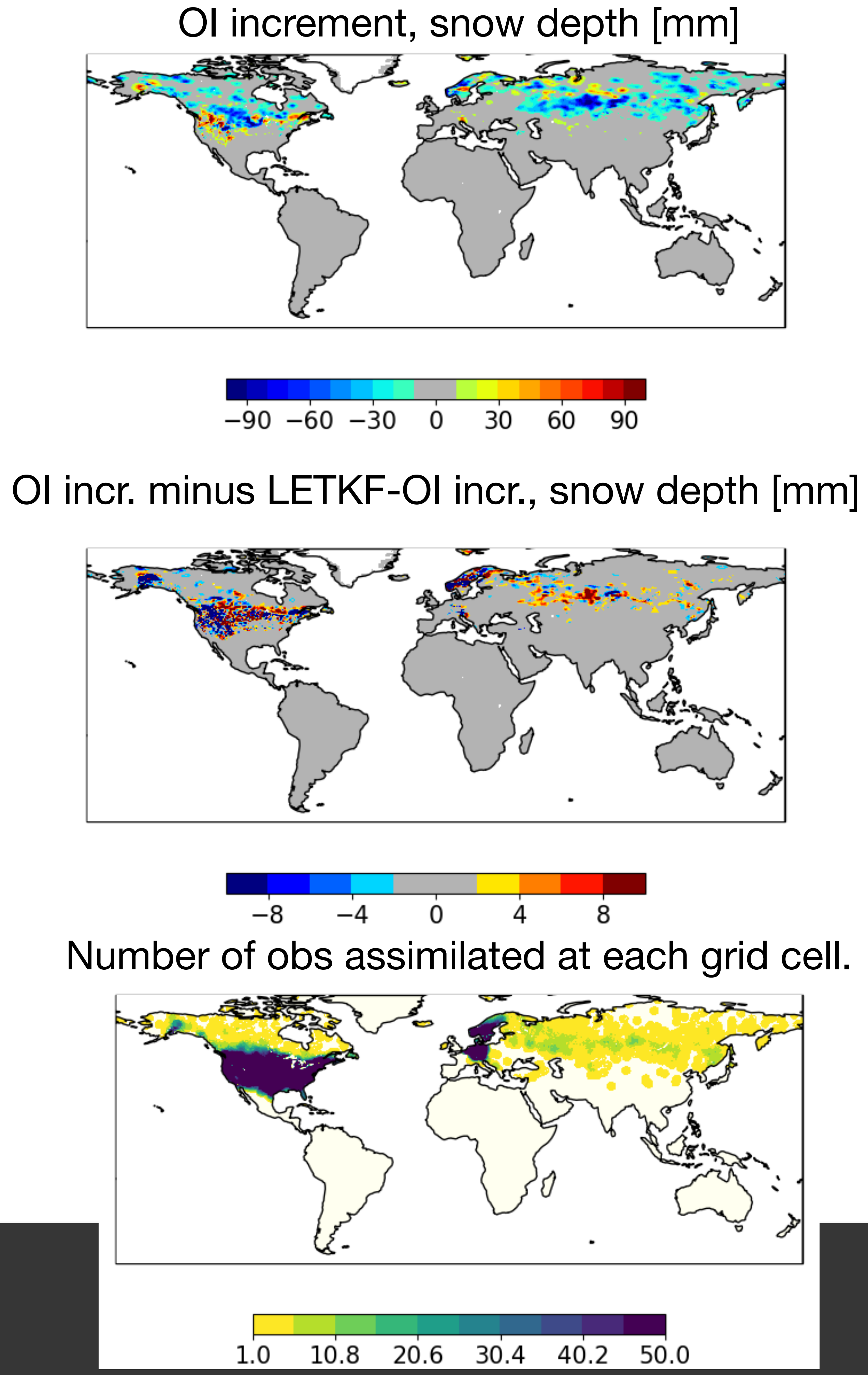
Conversion to JEDI

- OI algorithm not yet coded in JEDI, so have approximated OI by using LETKF (LETKF-OI):
  - Use a pseudo-ensemble and localization to approximate the error covariance functions used in the OI (Frolov et al (2022), QJRMS)
  - Approximation is very good for single observation experiments, but LETKF-OI has smaller increments where multiple observations are assimilated
  - Ideally, will investigate replacing the LETKF-OI with the OI in the future

Single observation experiment: RMSD = 0.8 mm, nRMSD = 8%



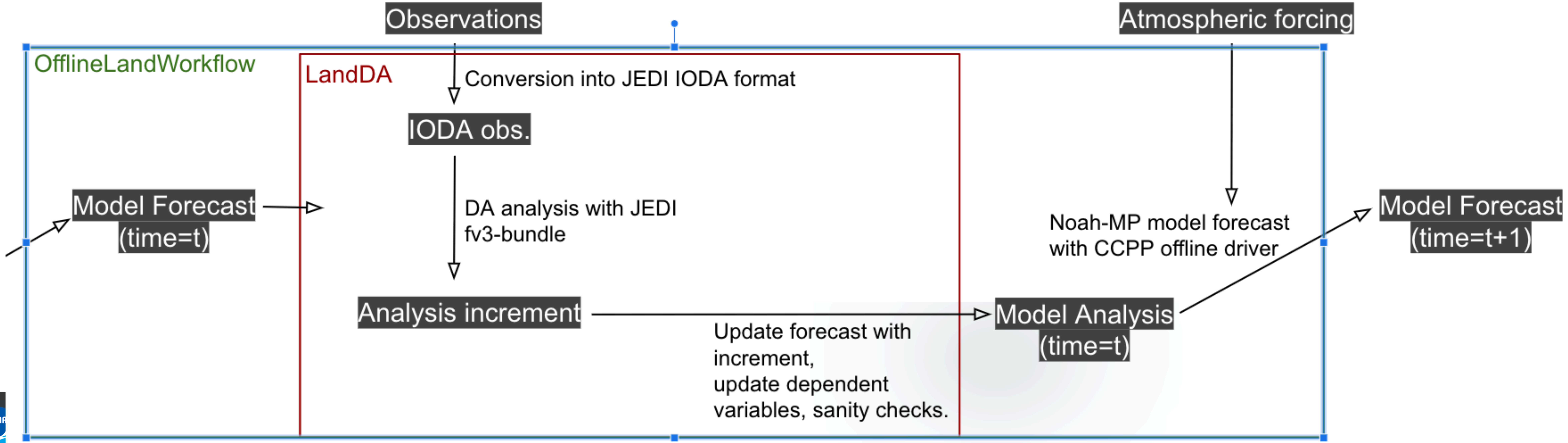
Global assimilation experiment: RMSD = 7 mm, nRMSD = 28%



The UFS Off-line Land DA System

- Developed an offline land DA workflow, to perform cycling model forecasts and DA, using same DA and land model code as in coupled GFS (land/atmosphere) DA system
  - Model: UFS code via Noah-MP CCPP code base
  - DA: JEDI fv3-bundle for land update

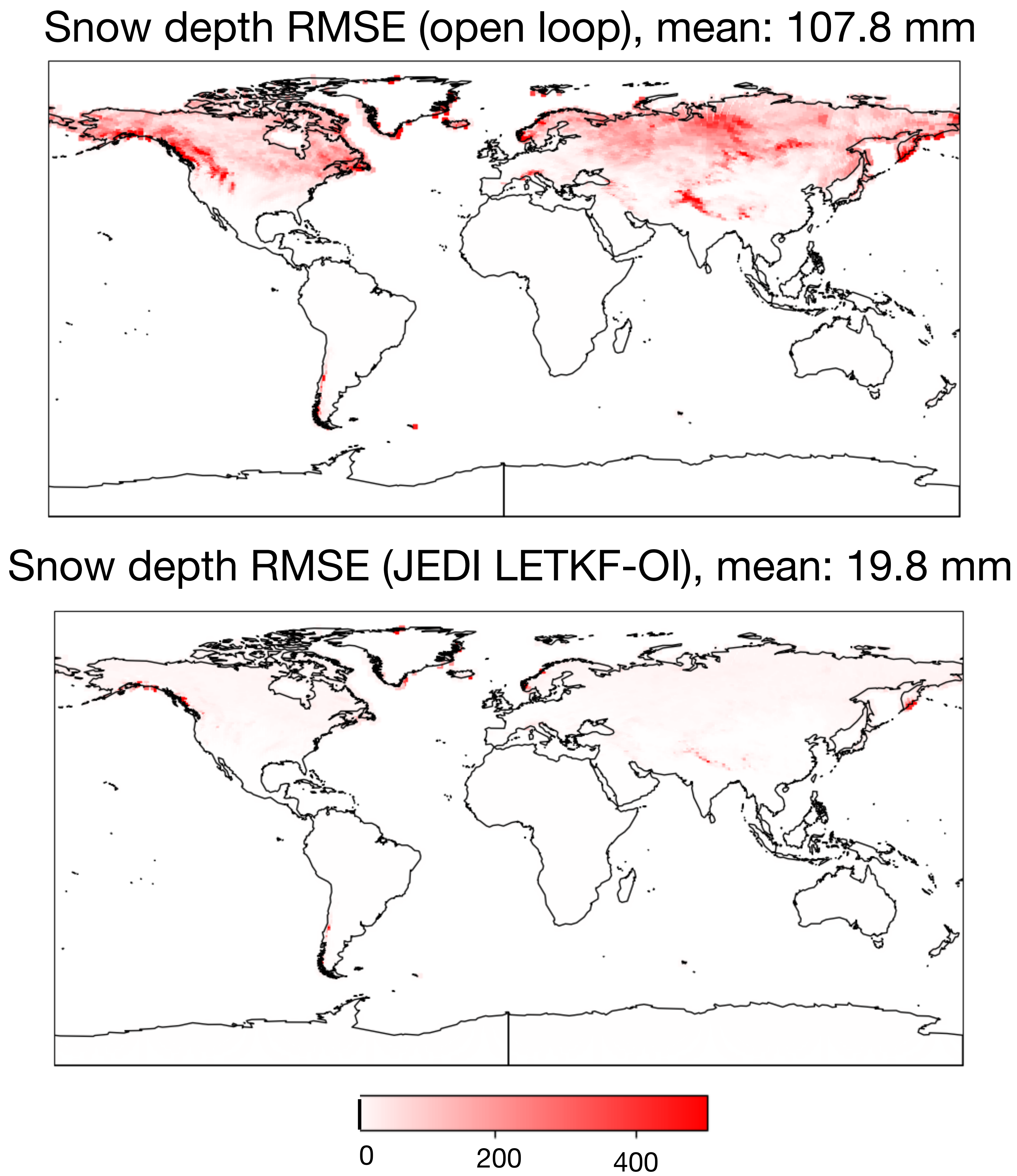
- Useful for development and testing of land model and DA
- In discussions with EPIC to make this available to research community to speed land research
- Currently includes snow depth LETKF-OI, soil moisture LETKF in development



Update to Noah-MP

- The snow model in Noah-MP is much more complicated than that in Noah:
  - Noah had a single snow layer, Noah-MP has a multi-layer snow model (many-to-one relationship between layer snow depths and the total snow depth)

- Aggregation of the total snow depth increment for Noah-MP:
  - Partition the total snow depth increment according to the fractional snow depth in each layer in the background, then update SWE in each layer by maintaining layer snow density
  - Synthetic twin experiments with offline system show good performance of DA / snow layer partitioning

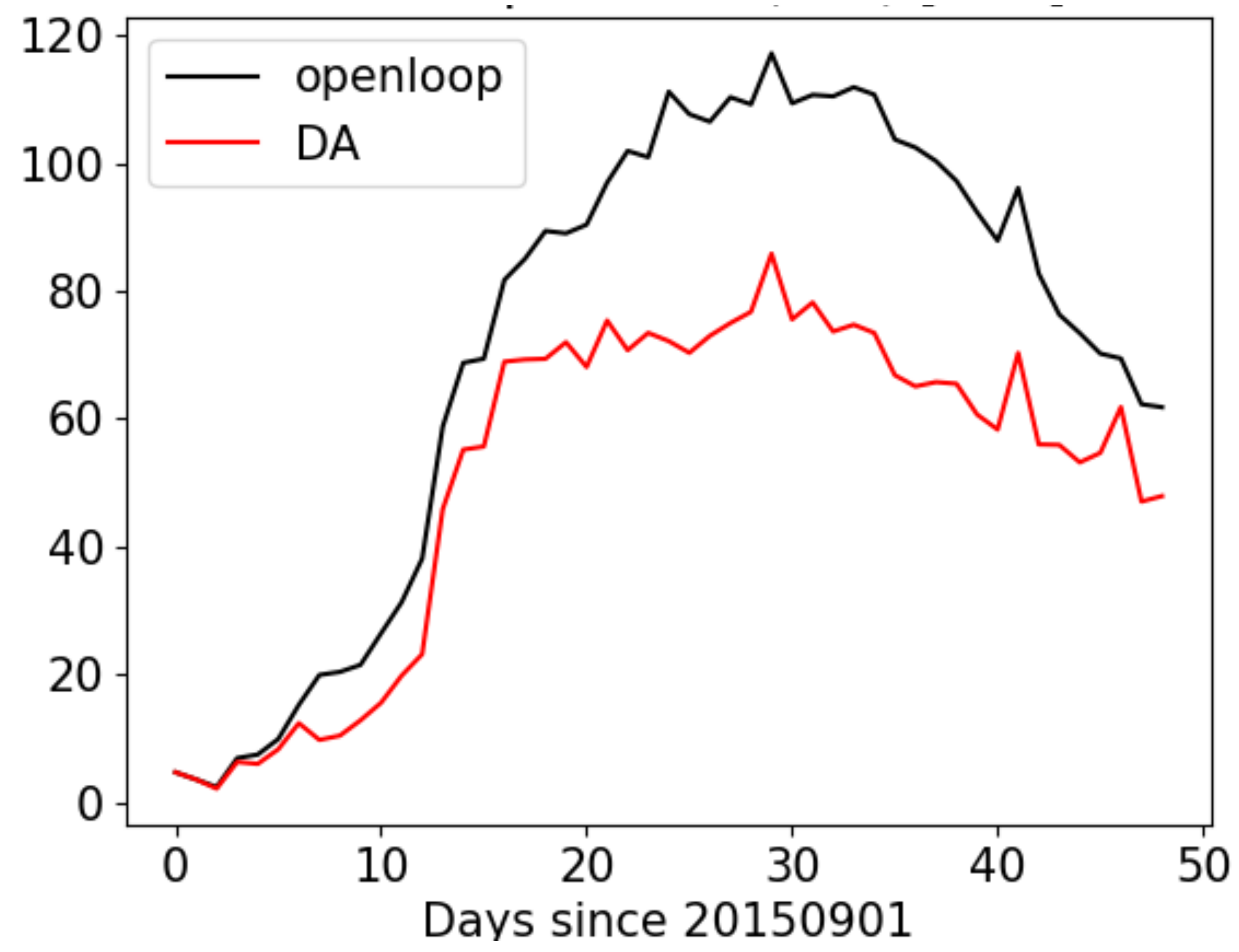


Synthetic twin experiment shows good performance of the LETKF-OI snow depth DA, from A. Gholoubi

Snow DA experiments with Noah-MP

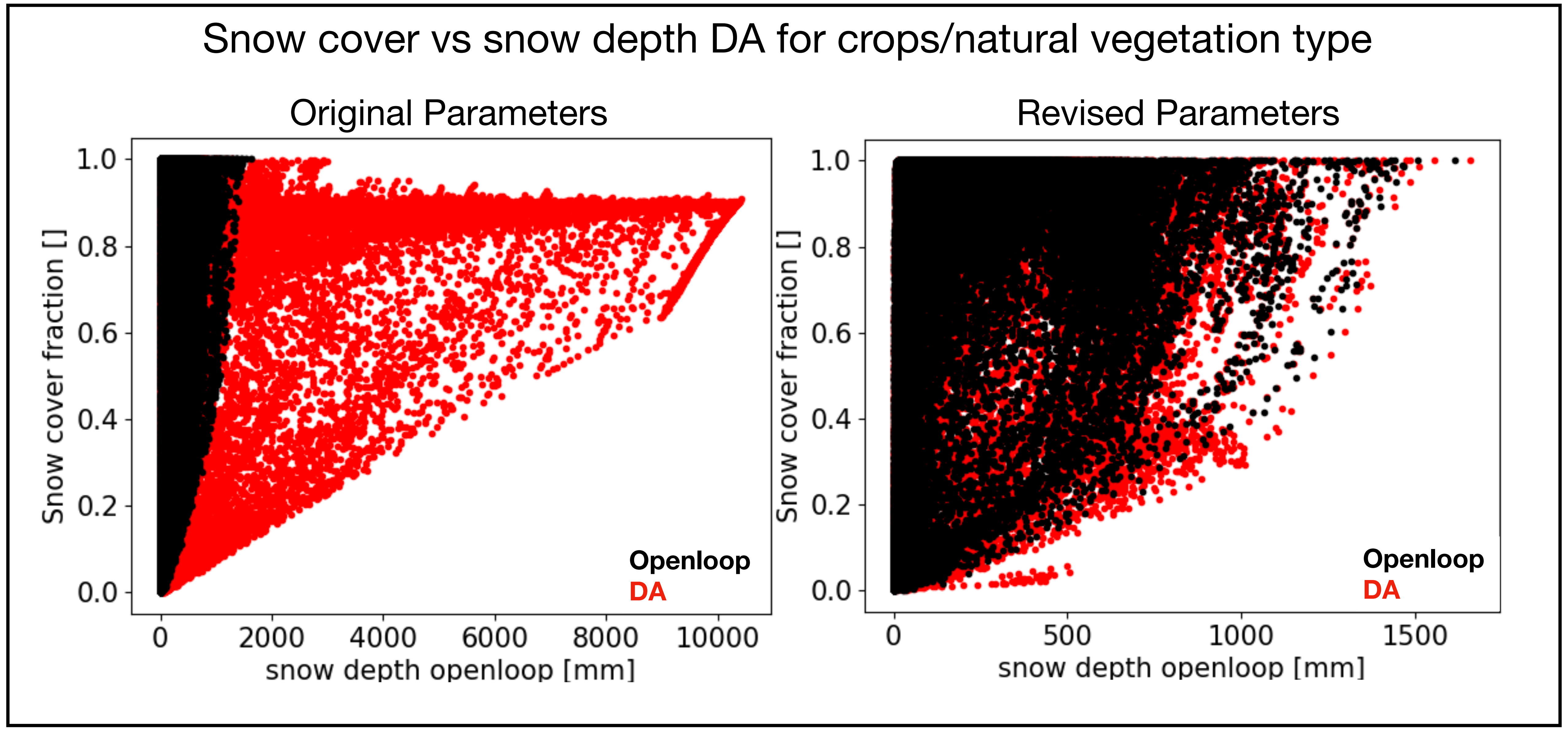
- Initial Noah-MP experiments conducted using the UFS offline land data assimilation system
  - Sep 2015 to May 2016, at 1 degree (coarse!), GDAS forcing
  - Assimilating GHCN station snow depth and IMS snow cover
  - Comparison of forecasts to assimilated observations (O-F) shows similar improvements to earlier Noah/GFSv15 experiments

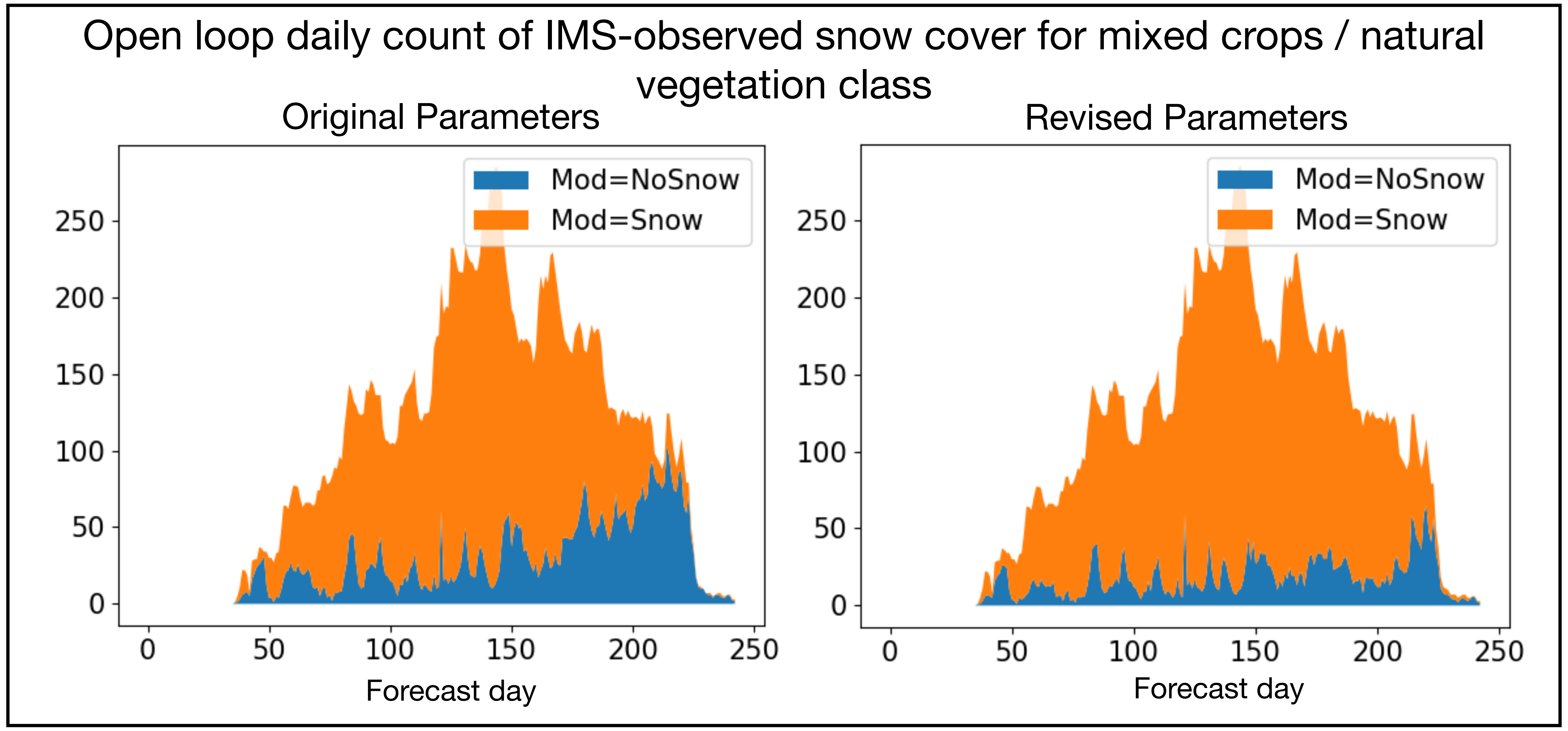
stdev(O-F) for the assimilation of GHCN snow depth observations [mm]



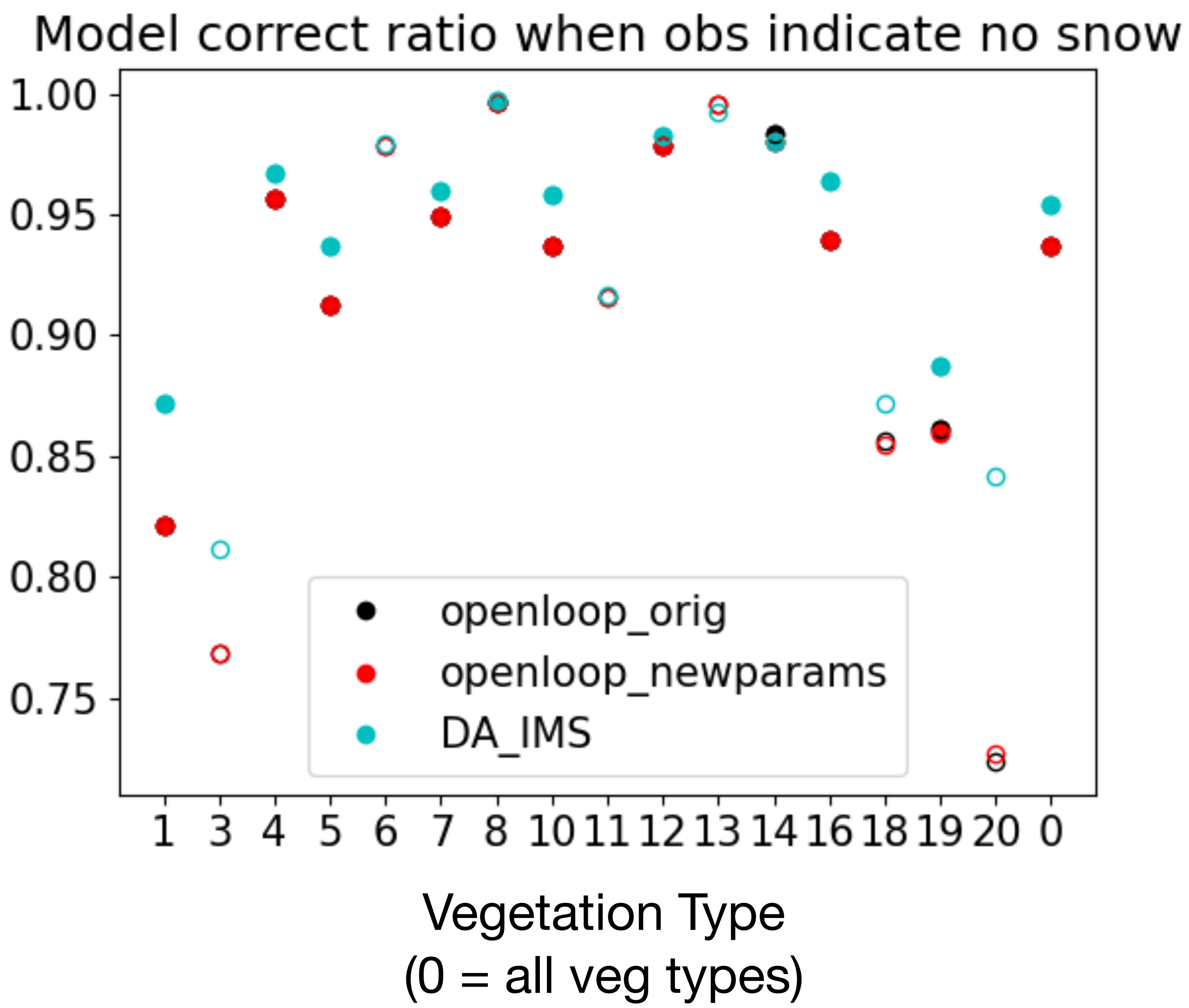
Revised Noah-MP Snow Cover Parameters

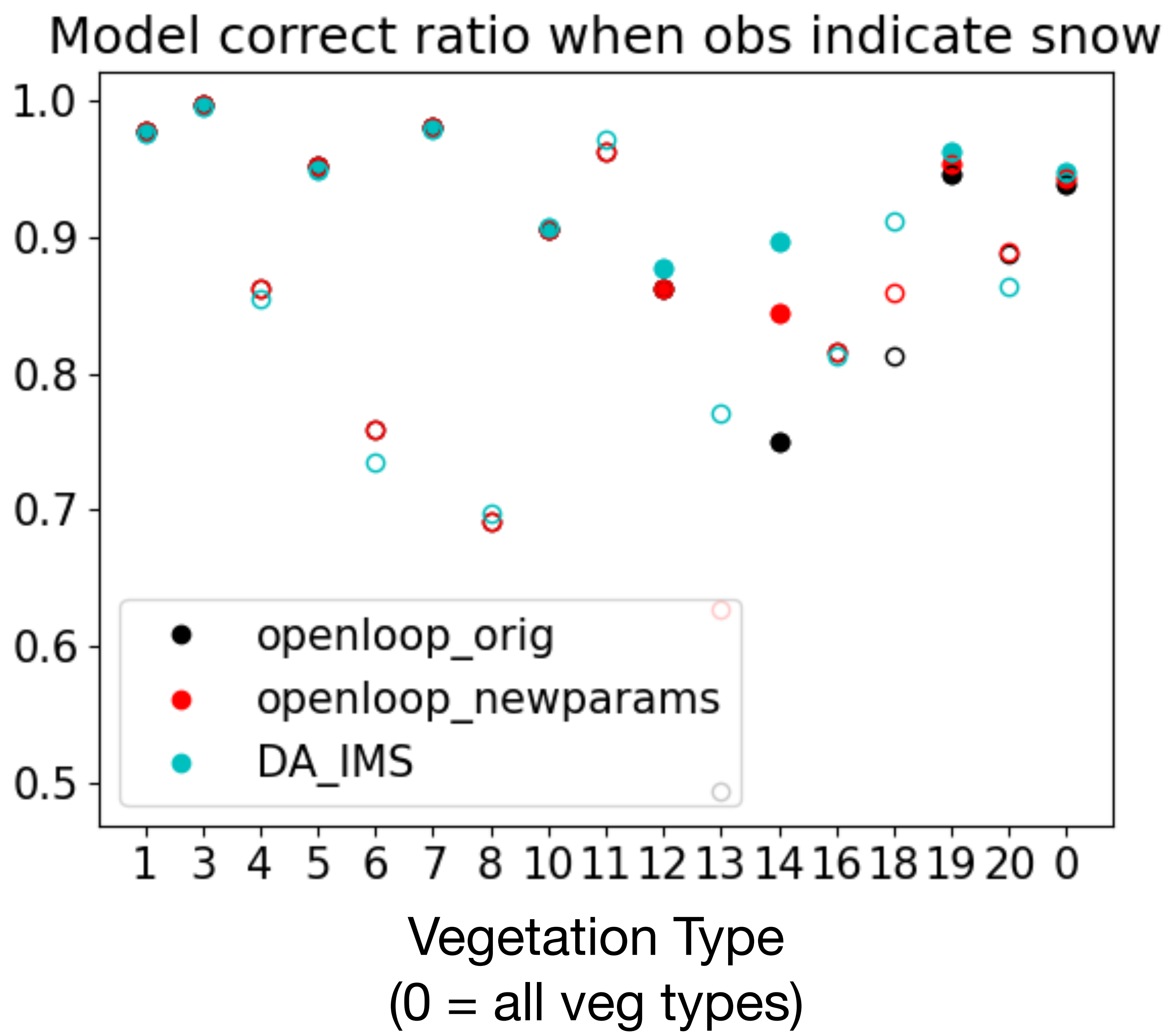
- Initial DA experiments revealed a significant bias in the model snow cover over several vegetation types, resulting in the DA adding an unrealistic amount of snow
- This bias was traced to the parameters used in the calculation of snow cover fraction
- Now using revised snow cover parameters:
  - DA no longer adds excessive snow mass
  - Model open loop (no DA) has better fit to snow cover obs.

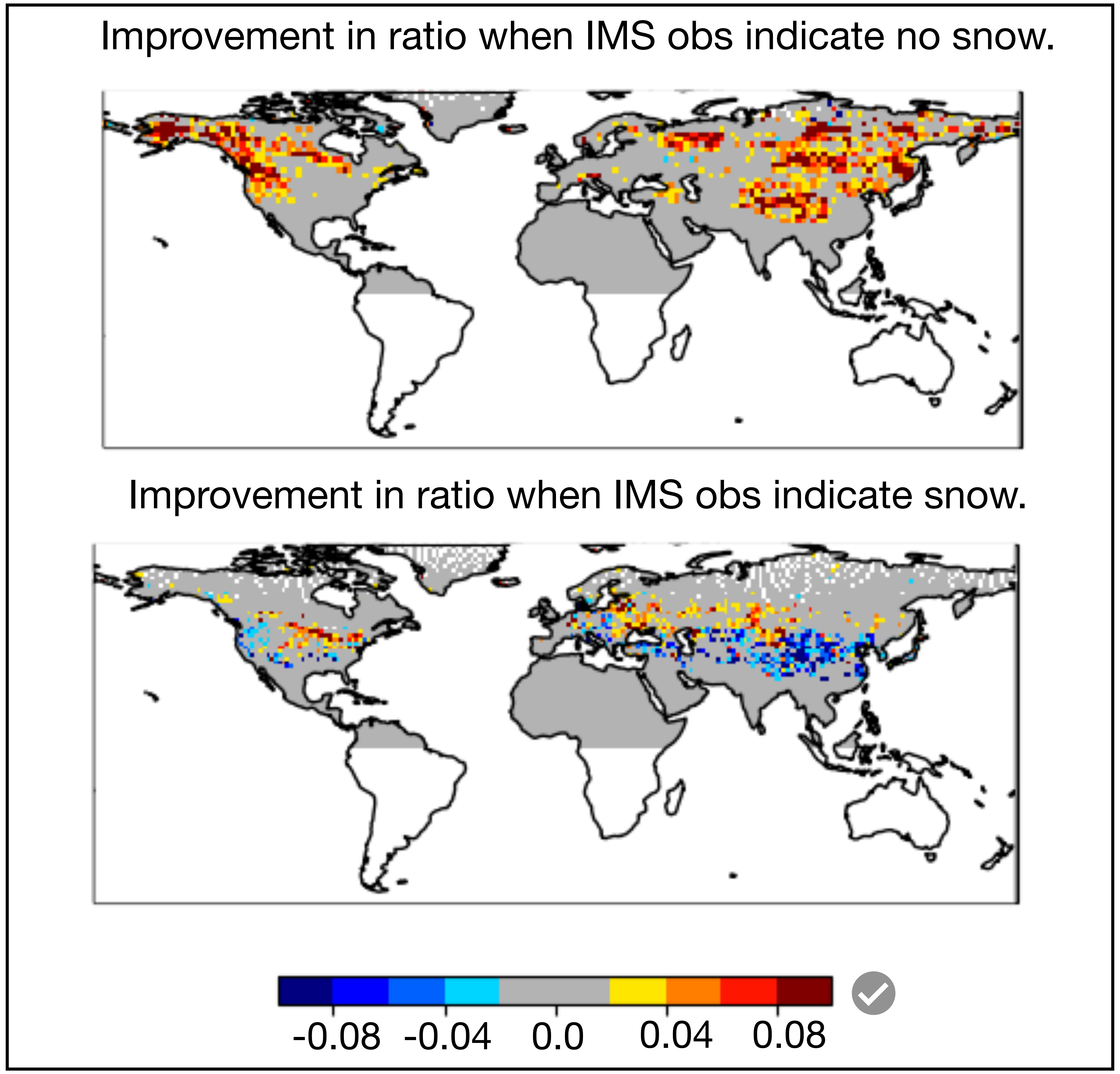




Snow Cover Assimilation







- Assimilation of snow cover observation effective when observations indicate no snow (upper left), much less effective when observations indicate snow (upper right)
- In many cases the added snow increments are not retained by the model; in southern China this makes the snow simulations worse - working on understanding why
- This problem probably exacerbated here by errors associated with coarse resolution.

Snow DA Summary

- Developed an Optimal (OI) Interpolation-based snow depth analysis for the GFS, based closely on schemes used elsewhere
- Early tests indicate improvements to the GFS snow fields and to near-surface atmospheric temperatures
- OI is being prepared for implementation
  - Updated to Noah-MP
  - Implemented in JEDI, and in an offline system

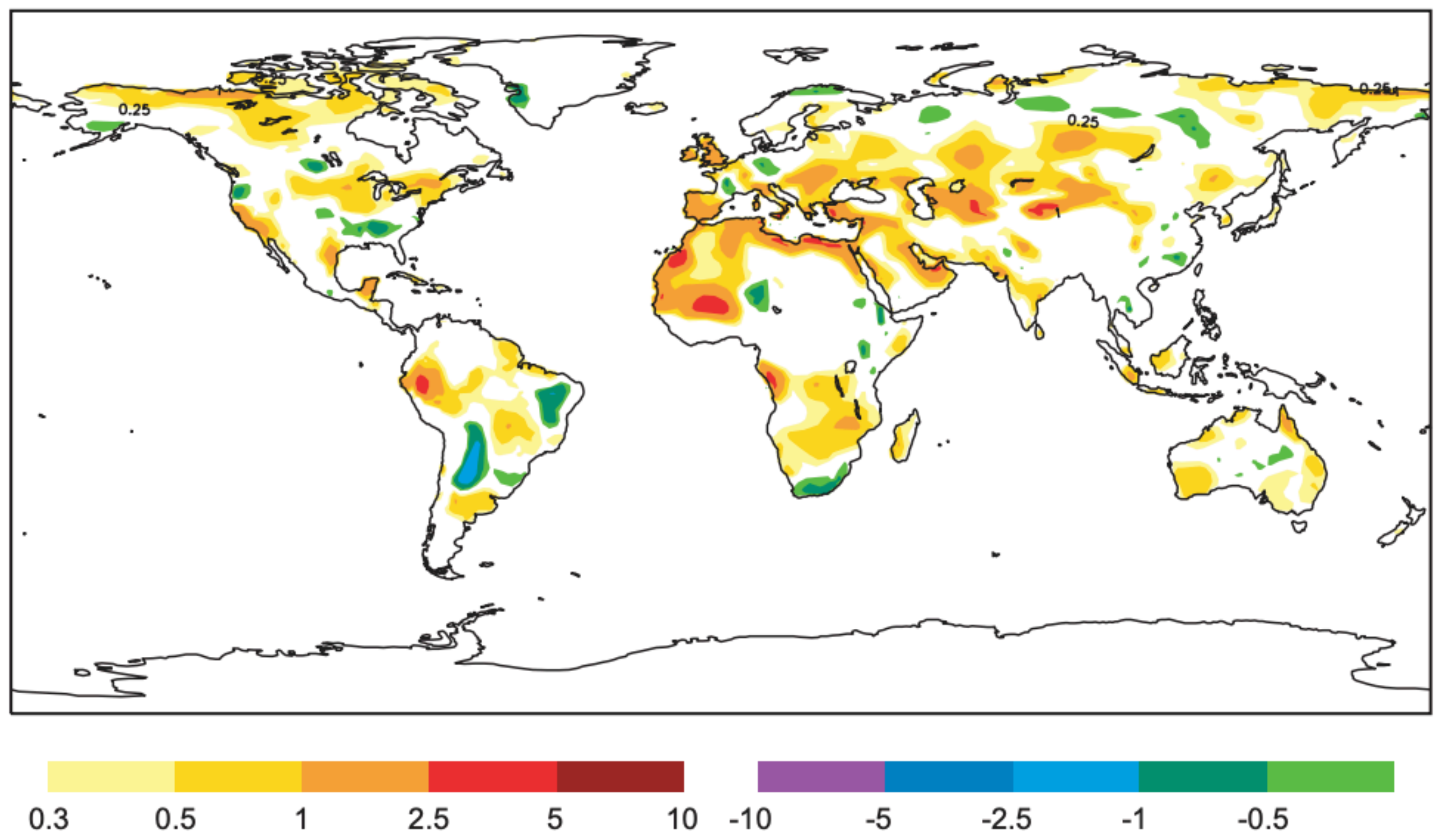
- Experiments with the OI demonstrate potential contribution of land DA to NWP:
  - Improved forecasts, due to improved initial conditions
  - Improved model physics, via model evaluation (may be the greater contribution)

- Next: fine-tune DA with Noah-MP offline at higher resolution, then conduct tests in the GFS atmos system

Soil Moisture and Temperature Analysis

Soil Moisture/Temperature Analysis for NWP

- NOAA currently has no soil moisture or temperature analysis in our global NWP
- Most other national NWP centers analyze soil moisture and soil temperature from screen-level (2m) station observations
  - First use was at Météo-France in 1985
  - UKMO, ECMWF now also assimilate satellite soil moisture / Tb (more direct observation of land states)
- Use relatively simple DA methods: empirical update coefficients, OI, Simplified EKF (SEKF)
  - Land update is an independent 1-D (vertical) update at each grid cell, using output from a 2-D (spatial) analysis of station observations
  - Computationally affordable, and is done on the model grid



Mean soil moisture increments (mm/day) from ECMWF's first soil analyses from screen-level observations, implemented in July 2000 (Mahfouf, 2000).

Soil Moisture DA in Hydrology

- In contrast to NWP, the hydrology community uses EnKF-type methods almost exclusively for land DA
  - More flexible (addition of new obs / updates states), more intuitive specification of model errors, account for errors of the day, more robust to non-linearities

- EnKF for land DA less common in atmospheric systems
  - Some LAM examples in literature
  - EC and NASA both use EnKF land DA, with the EnKF applied to separate land-only ensembles
  - Now that atmos DA is ensemble-based, no need to run a separate land-only ensemble

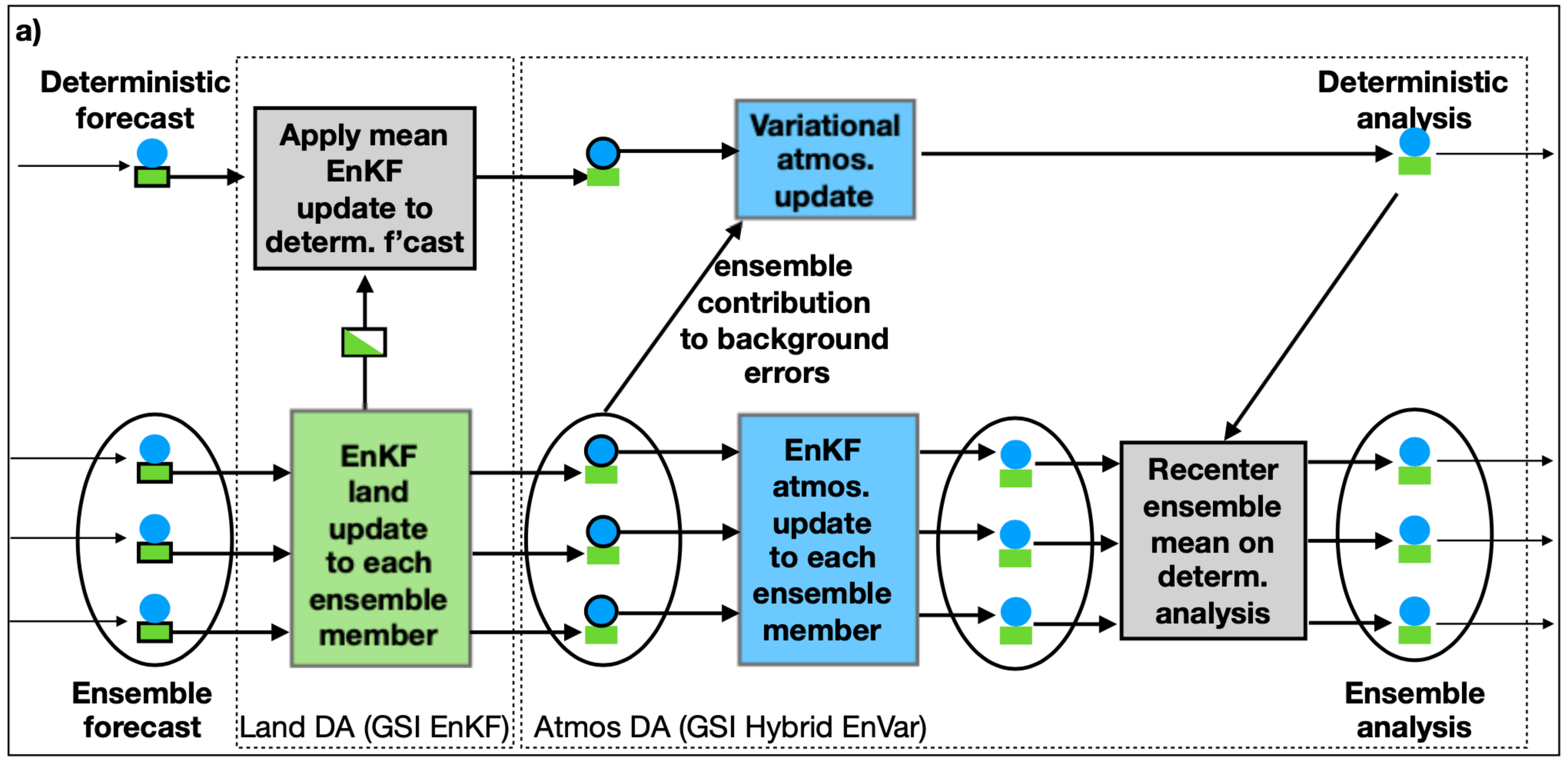
- Hydrology community assimilates satellite soil moisture observations (or associated Tb) rather than 2m observations
  - Assimilation of 2m observations is effective at improving low-level atmospheric forecasts, but *can* degrade the soil moisture and temperature states
  - NWP community moving towards assimilation of both 2m obs and satellite soil moisture information

The GFS soil analysis

- The GFSv17 soil moisture and soil temperature analysis is an EnKF assimilation, applied directly to the land states in the atmospheric ensemble
  - For now, development being done in the GSI (analysis not done at model resolution!)
  - Initially, using the EnKF (LETKF), but will likely upgrade to 4DEnVar in the future
  - Initially assimilate only T2m and q2m, then add satellite soil moisture observations later
    (Zofia Stanley working on SMAP soil moisture retrieval DA)

- GFSv17.1 will include the new land model (Noah-MP), plan to implement new soil analysis at next update

Schematic of GFSv17 LETKF Soil Analysis Workflow.

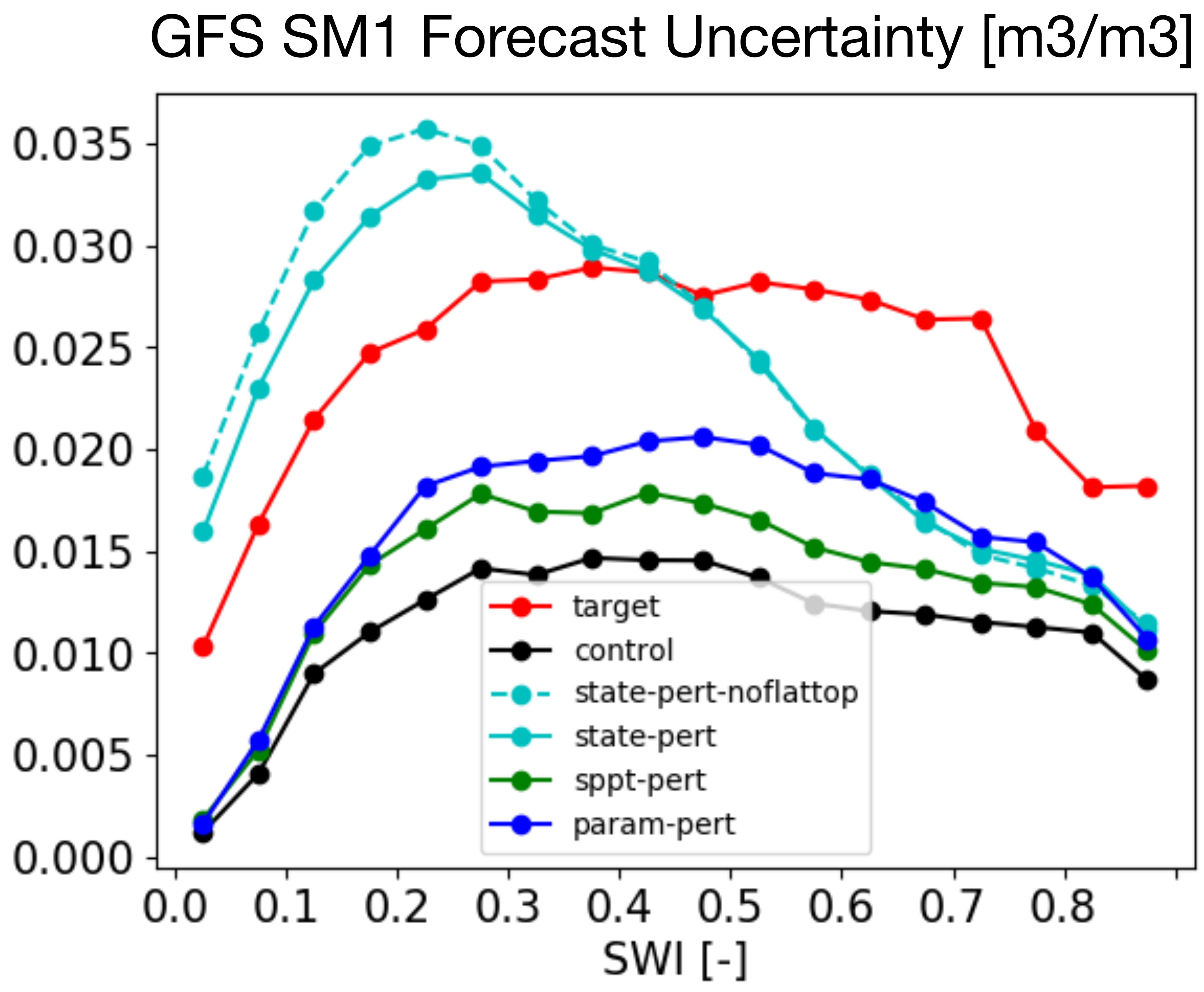


GFS Land Ensemble Spread Summary

- Using the GFS atmospheric ensemble for land DA requires reasonable ensemble spread at/near the land
  - NWP ensemble systems are under-dispersed near the land surface (ensemble generation methods do no account for land model uncertainty)

- Tested several methods to account for land model uncertainty in the GFS ensemble
  - SPPT can only generate a limited amount of soil moisture spread
  - Adding perturbations to land states enhances ensemble correlations when land is driving the land/atmosphere coupling, and degrades the correlations when the atmosphere drives the coupling
  - Perturbing model parameters used in land/atmosphere flux calculations creates ensembles in which the pre-existing correlations are enhanced (i.e., an ensemble representative of errors in the fluxes)

- Opted to perturb vegetation fraction
  - Generates reasonable spatial patterns in resulting ensemble spread



Draper, 2021 (JHM)

GFS LETKF Soil Temperature Analysis Experiments

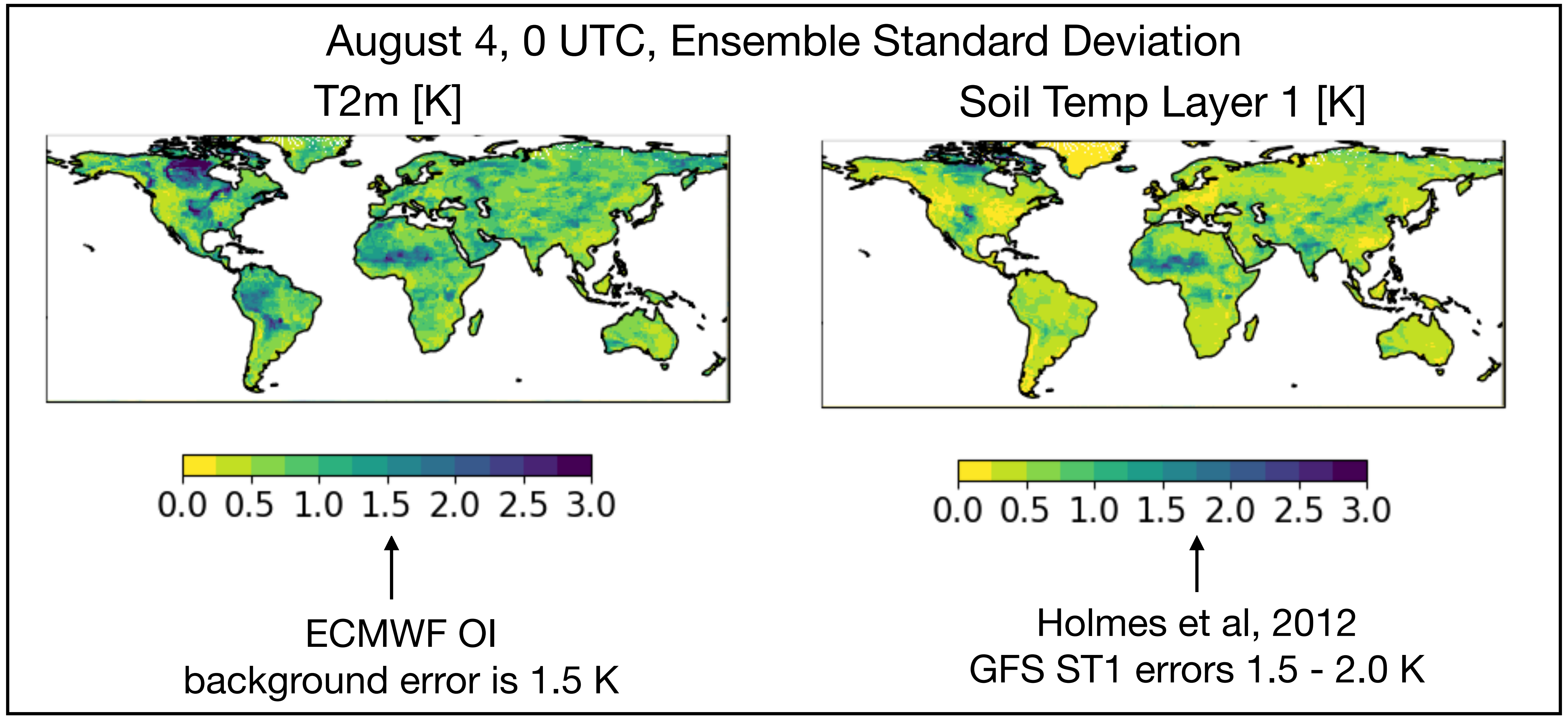
- With new land perturbation scheme, now have sufficient ensemble spread at/near land surface to do land DA in the GFS ensemble

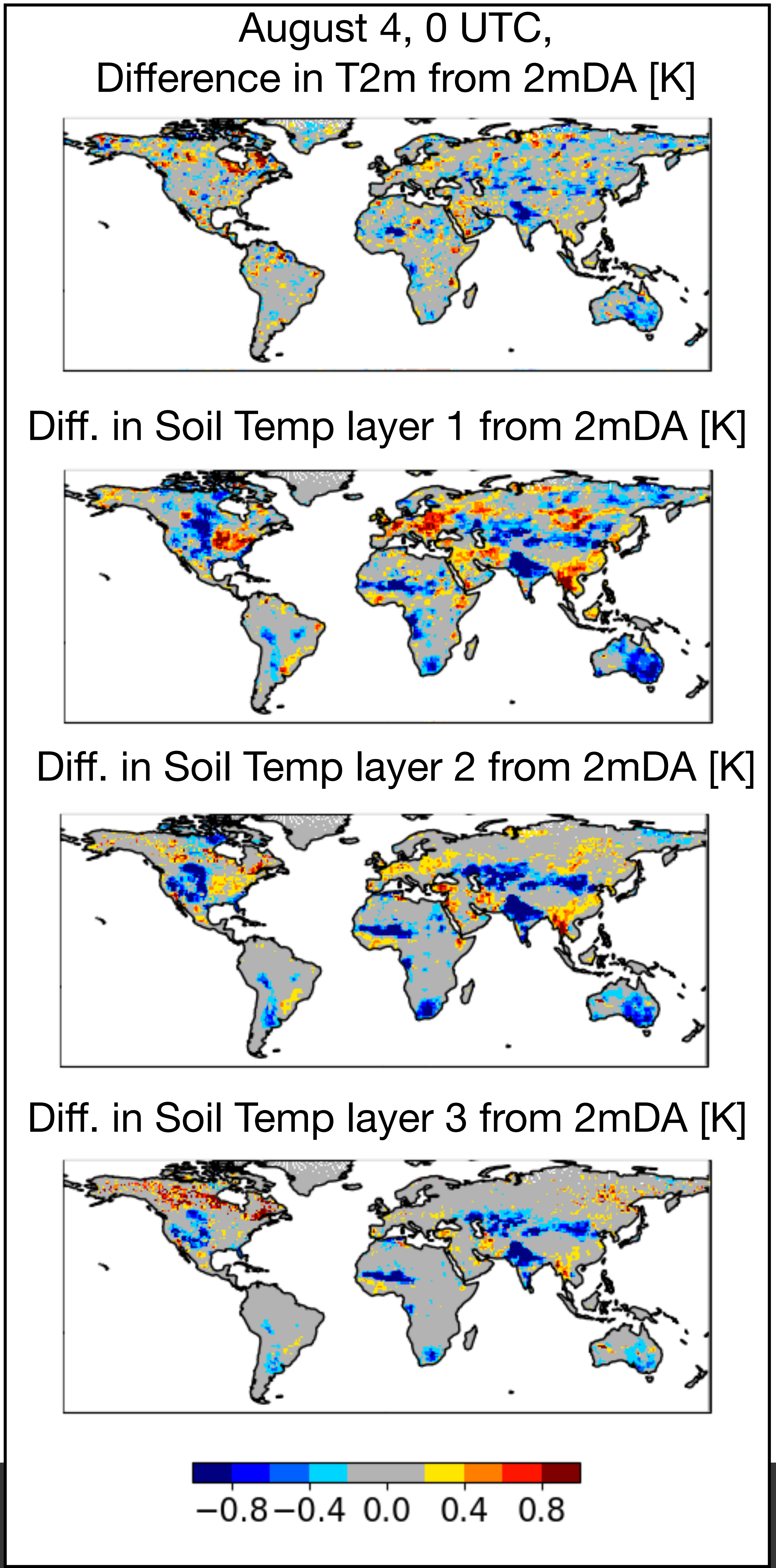
- Initially use LETKF only, and update soil temperature from T2m only
  - Model soil moisture / T2m relationship unrealistic

- LETKF experiment: 20 days from July 15 2020, at C192 (~0.5 degree), using GFSv16 model (Noah LSM), standard atmos stochastic physics plus perturbed vegetation fraction (stdev of pert: 10%).

- Experiments:
  - Control: LETKF of standard suite of atmospheric observations
  - 2mDA: Control + LETKF assimilation of T2m to update soil temperature in top 3 layers
    - Correcting T2m obs to model terrain
    - Minimal QC of T2m obs (gross error, bounds checks)

GFS LETKF Ensemble Spread

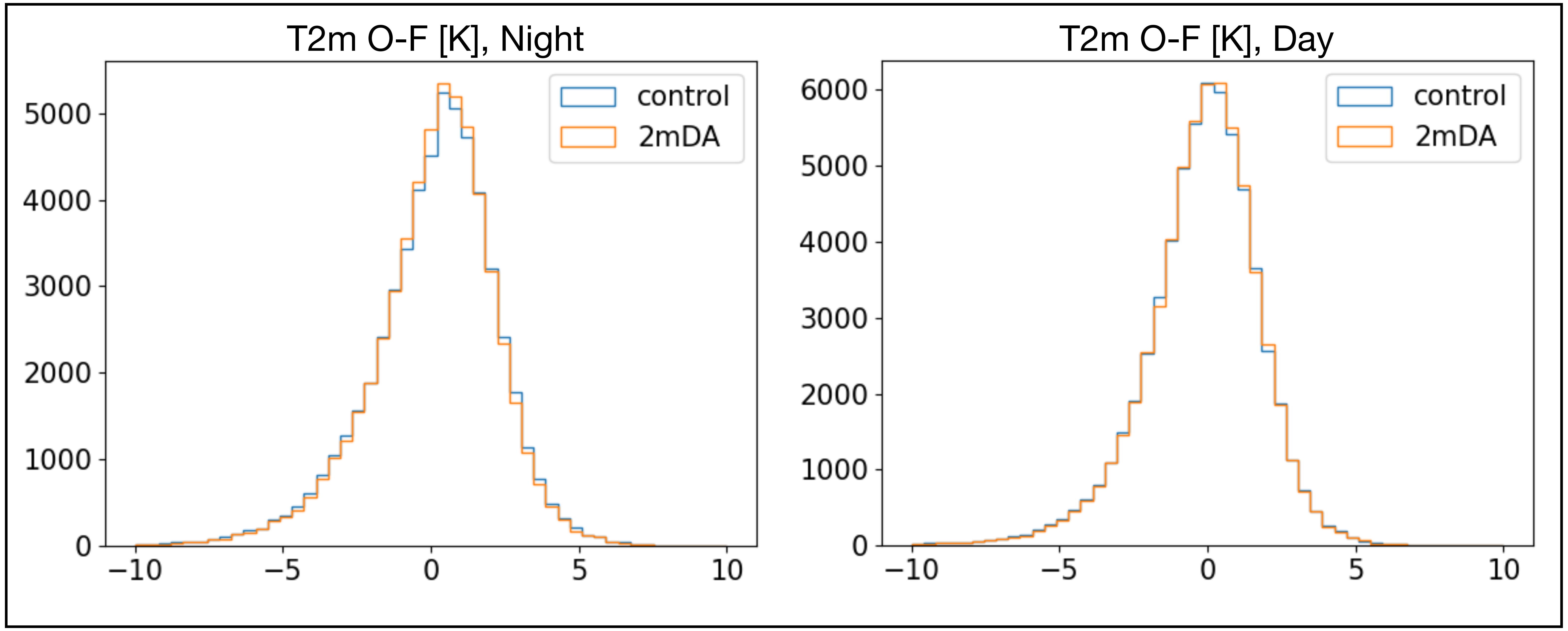




- Even with land perturbation scheme, ensemble is still under-dispersed in this experiment
- Second experiment with larger ensemble spread introduced a warm bias in the temperatures, due to non-linear response to applied perturbations
- In practice ensemble spread is usually limited by desire to avoid substantial changes in ensemble mean
- Will need to compensate for under-dispersed ensemble by applying some additional model error

T2m O-F Statistics

- Small reduction in T2m O-F over last 10 days of experiment, with larger reduction at night-time (6%) than day-time (2%)

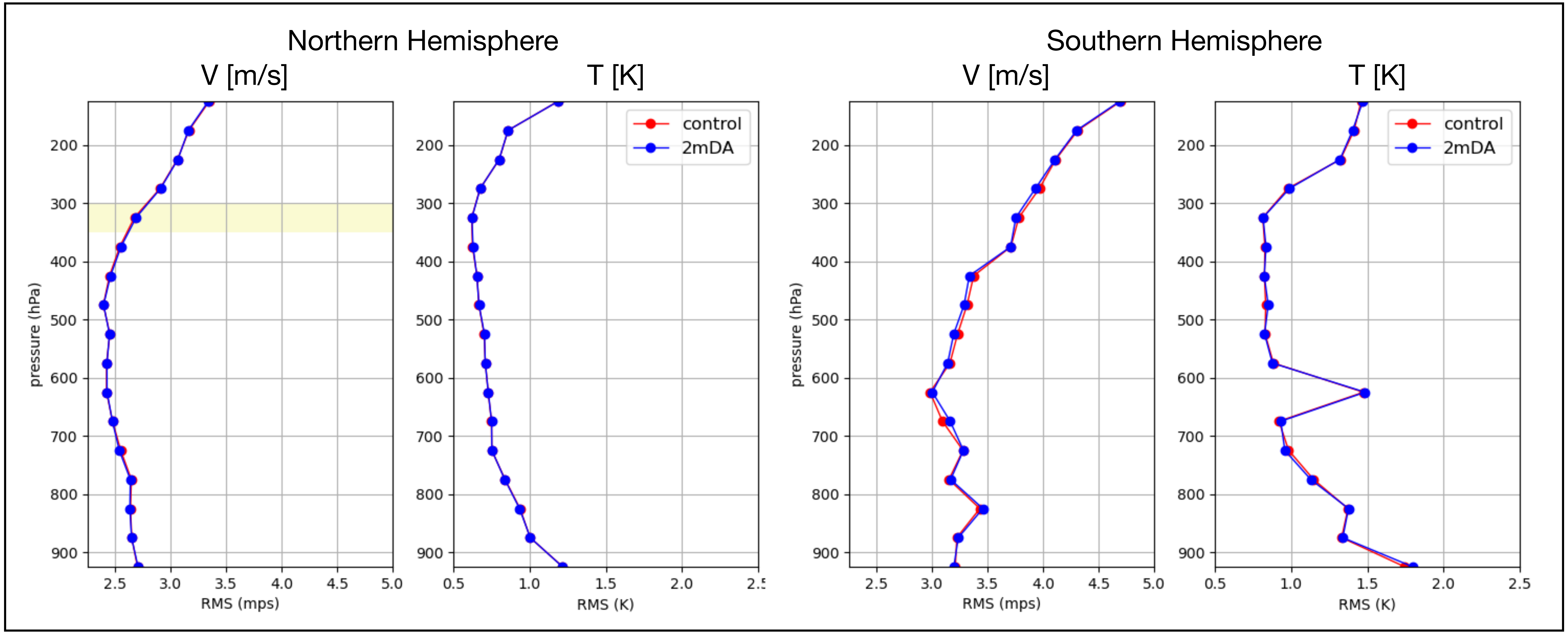


Standard deviation (mean) T2m O-F [K]

	CONTROL	2mDA
Night	4.85 (0.085)	4.56 (0.091)
Day	4.02 (-0.21)	3.95 (-0.19)

Atmospheric In Situ O-F Statistics

- Very limited impact on atmospheric profile over last 10 days of the experiment



Soil Analysis Conclusions

- EnKF-type DA has traditionally been preferred in the hydrology community, but is not as widely used by the atmospheric community
- Now have opportunity to do land DA using same DA code and ensemble forecasts as for atmosphere
  - Allows a much tighter integration between atmosphere and land (models, ensemble systems, DA, obs. operators, etc)
- Early results with LETKF demonstrate this approach is viable
- Attaining sufficient ensemble spread without introducing substantial changes in the model mean will be difficult
- Next steps: test additive inflation in the EnKF, update to use Noah-MP, and add q2m observations and soil moisture updates

Summary and Conclusions

- We are making progress towards a state-of-the-art land DA for the GFSv17 system
  - Initial focus on updating the snow DA system (relatively complete) and adding a soil moisture and soil temperature analysis (more work to be done)

- The new land DA system will:
  - Make use of modern software (JEDI)
  - Assimilate observations more directly
  - Do assimilation on model native grid
  - Addition of new model updates states and observation types

- Also develop an offline version of the GFS JEDI-based land DA system, using same DA code and same model code
  - Useful for development and evaluation within NOAA
  - Working with EPIC to share with broader research community

Summary and Conclusions

- Demonstrated that the new snow depth DA can improve the GFS through improved snow initial conditions, and through contributing to model evaluation/improvements

- Developing a new soil moisture and temperature analysis using the GSI (JEDI) LETKF
  - Initial experiments show very slightly improved T2m forecasts from updating the soil temperature
- Opportunity to better integrate atmospheric and land DA under one system
  - Given differing nature of land and atmosphere systems may be quite difficult to satisfy both using a single DA approach

Thank you for listening

Call for papers for JAMES special collection on Data assimilation for Earth system models.
Eds: Sergey Frolov and Clara Draper (NOAA/PSL), Lars Nerger (AWI).
Deadline: 31 December 2023

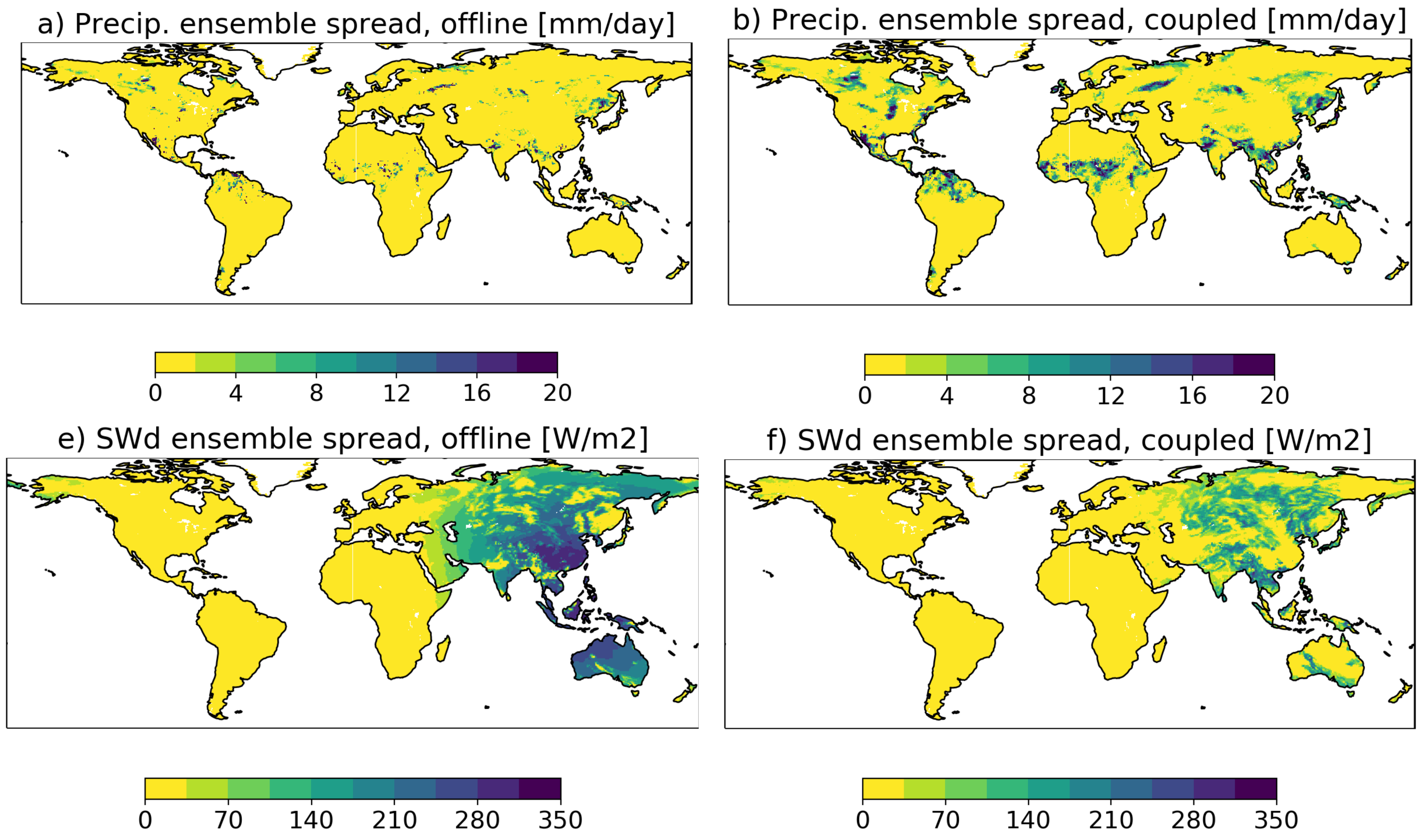
clara.draper@noaa.gov

The GFSv17 GSI Soil Analysis

- Advantages:
  - Closer integration with atmosphere, opportunity for more strongly coupled land/ atmos DA
  - Better representation of uncertainty in atmospheric forcing (compared to perturbing a single atmospheric realization)
  - Ease of implementation / code management / integration with atmos DA community
- Disadvantages:
  - Need to run full atmospheric ensemble for testing and development (until switch to JEDI)
- Note: ECMWF achieving similar outcome from EDA using ensemble to estimate the SEKF Jacobians

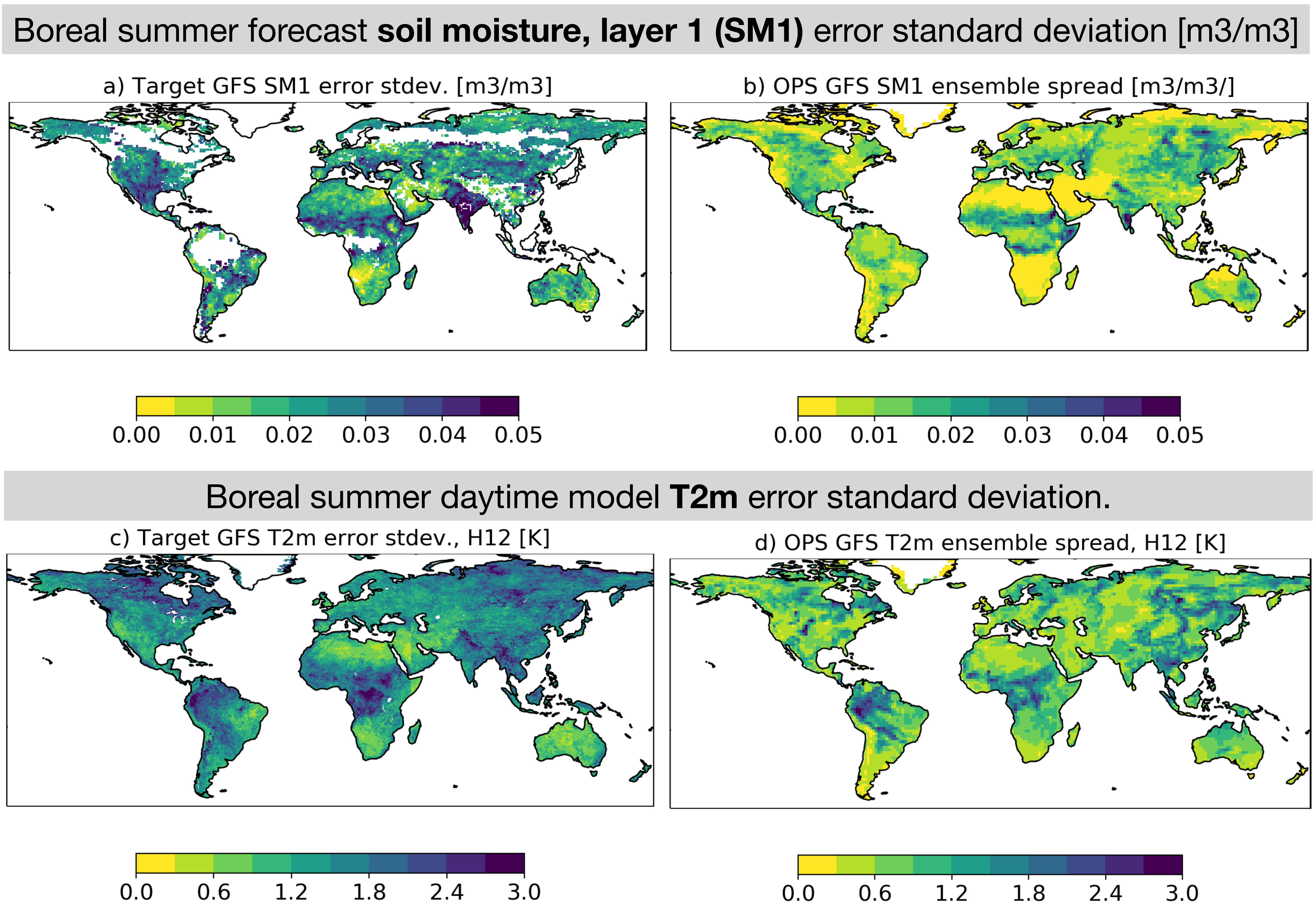
Ensemble spread from perturbing a single atmospheric realization

Ensemble Spread from GFS atmos ensemble



NWP Land Surface Ensemble Spread

- NWP ensemble systems (inc. the GFS) are under-dispersed at/near the land surface
  - Ensemble members are not currently perturbed to account for land model uncertainty
  - In hydrology community, typically perturb both atmospheric forcing and land states and/or parameters
  - First step to implementing the GFS land DA was to enhance GFS land ensemble spread by adding an ensemble perturbation scheme to account for land model uncertainty

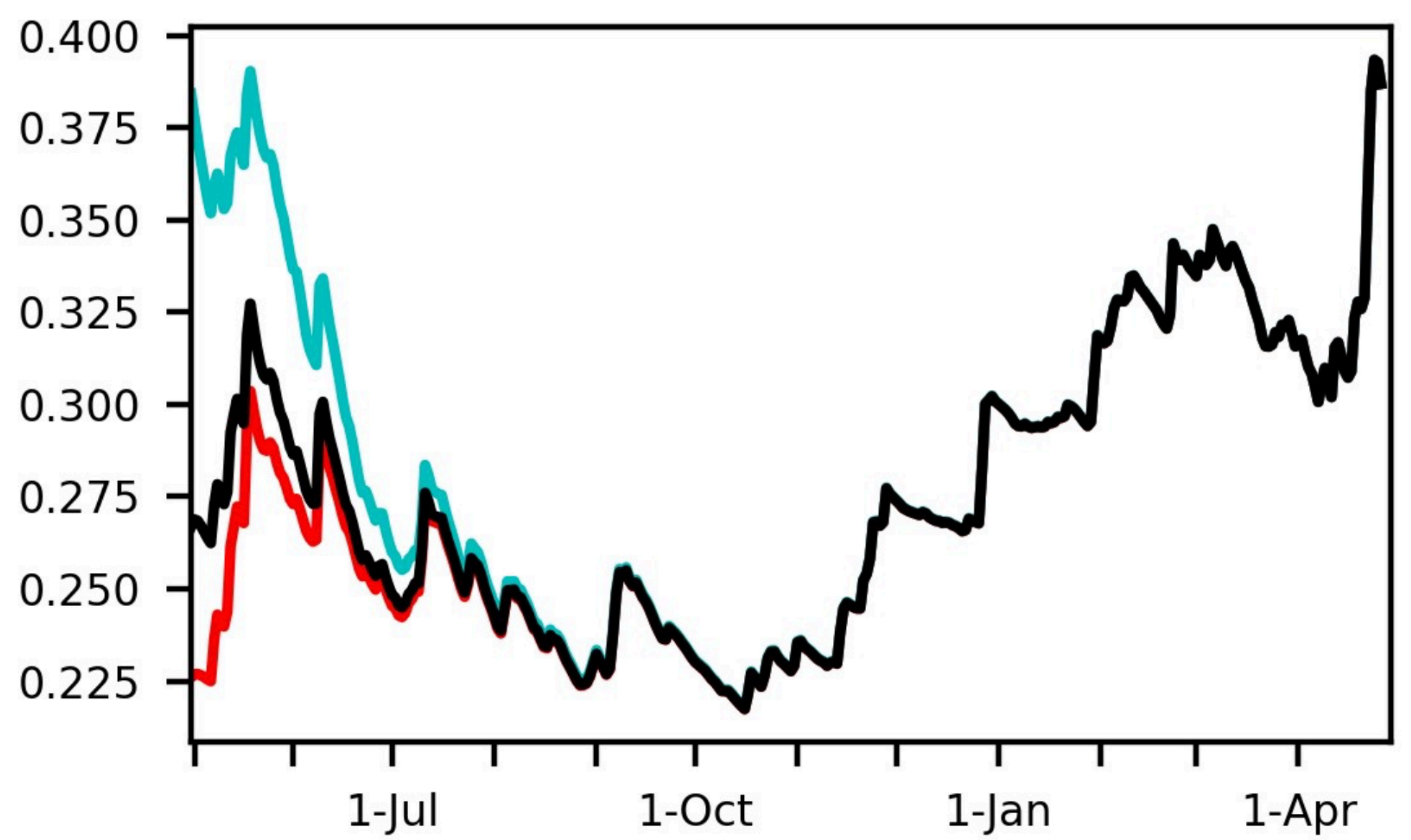


Target estimates, calculated using triple colocation (SM1), and comparison to ERA-5 analysis (T2m).

Ensemble standard deviation, from archived operational UFS output

Land Model Physics and Data Assimilation

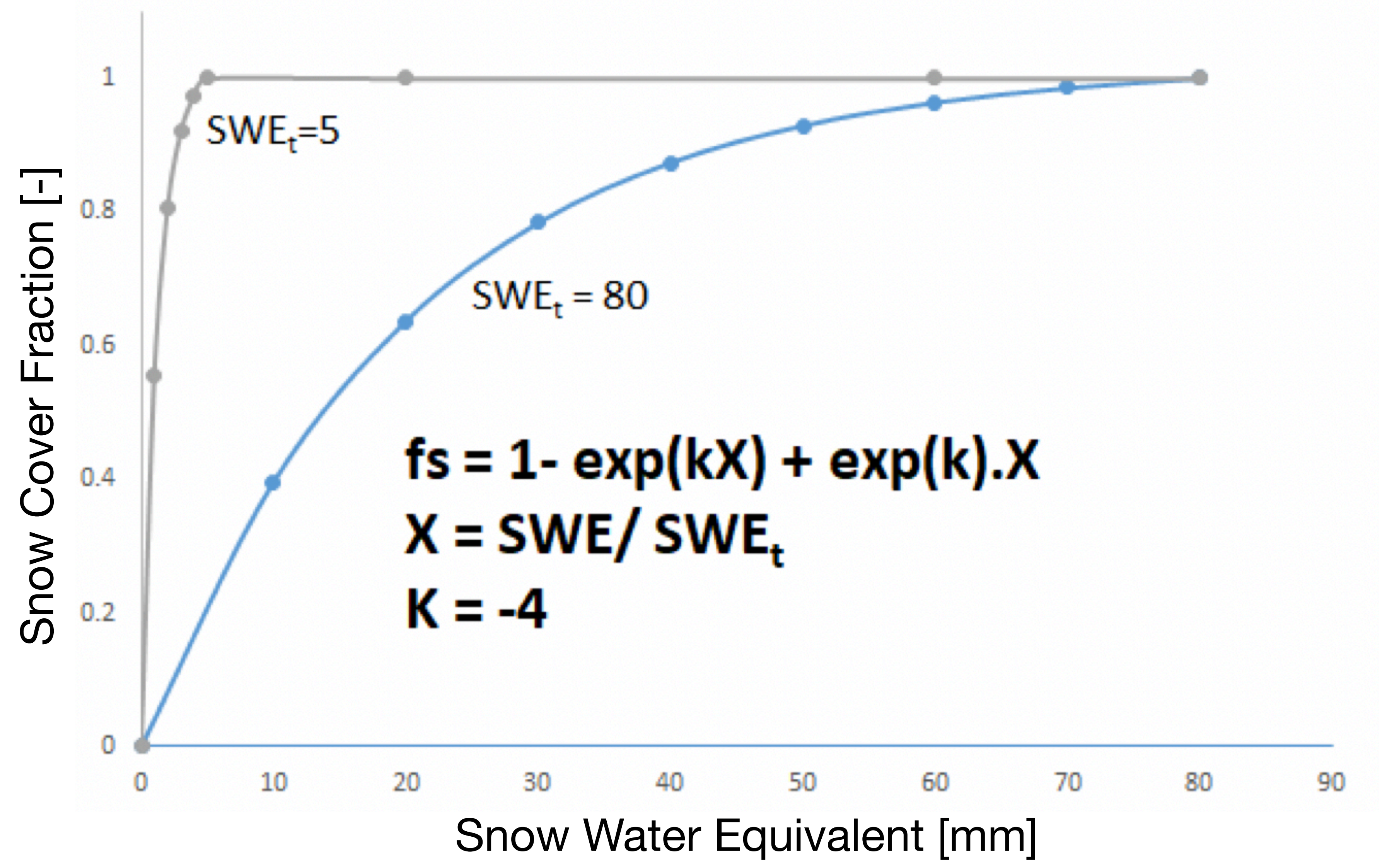
- Land models differ from atmospheric models, DA must be designed accordingly:
  - The land is strongly-forced (dissipative), and over time will converge to a state determined by its forcing
    - Not chaotic; error propagation and growth very different to atmosphere
  - Land models do not simulate horizontal flow between grid cells
    - No horizontal flow of errors, DA often done independently at each grid cell (1-D assimilation)
  - Land models are highly non-linear, and contain switches -> no LTM or adjoint
  - Land is highly heterogenous (intra- and inter- grid cell)



Catchment model rootzone soil moisture simulations for three different initial conditions and identical atmospheric forcing.

Assimilated Snow Cover

- Include satellite snow cover to compensate for limited spatial coverage of station snow depth observations
- Project IMS snow cover onto the model grid, to obtain a snow cover fraction for each grid cell
- Convert snow cover fraction to an equivalent snow depth by inverting the model's snow depletion curve
- Do not assimilate snow cover observations if both model background and observations indicate 100% snow cover

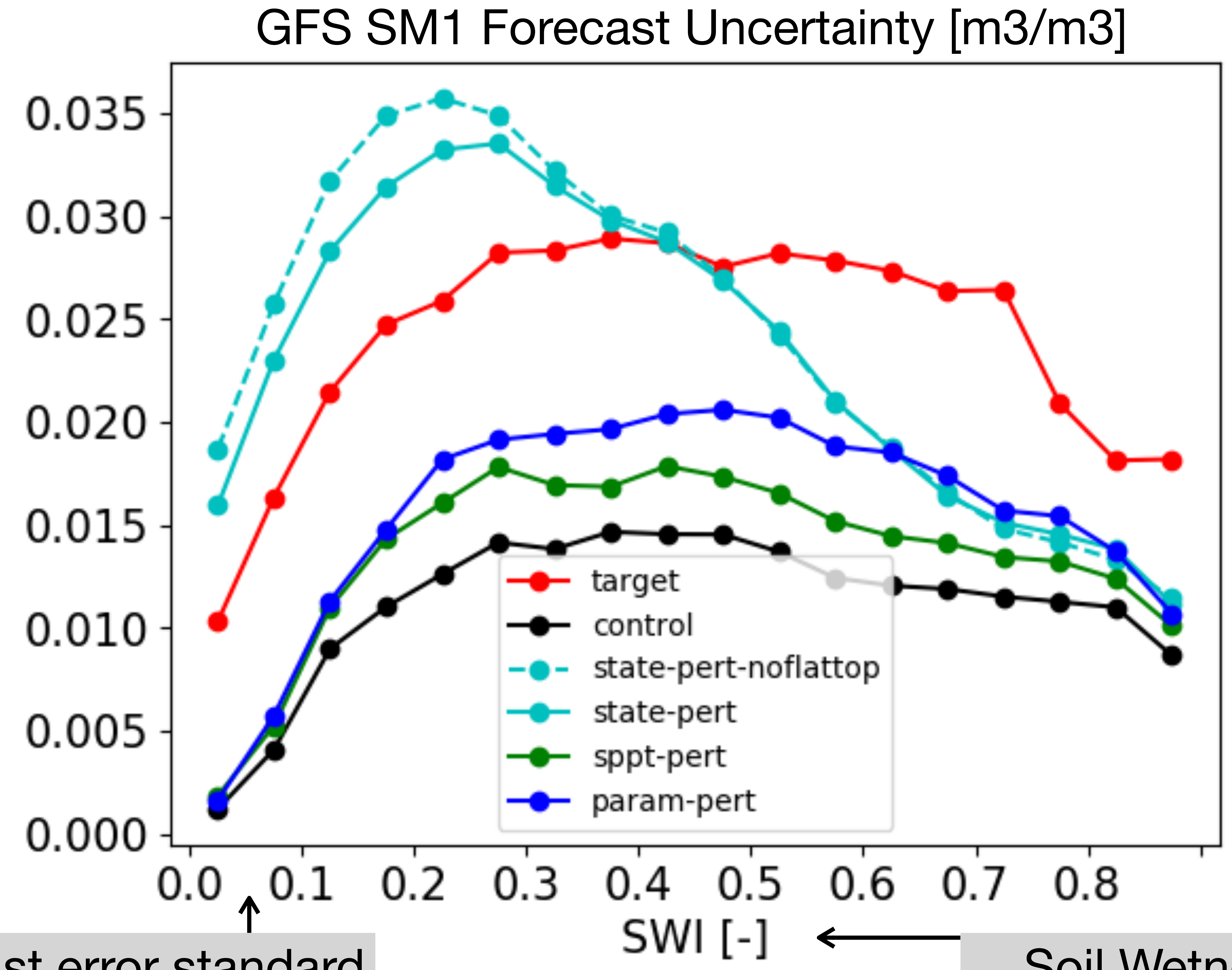


Noah areal snow depletion curve for two vegetation types, plot from T. Gichamo.

Ens. Spread, Soil Moisture Layer 1 (SMC1)

- State-pert induces too much spread in dry regions. Due to soil moisture memory being longer in dry conditions.

- SPPT-pert can induce only a small amount of spread. Inherent limitation of the method.



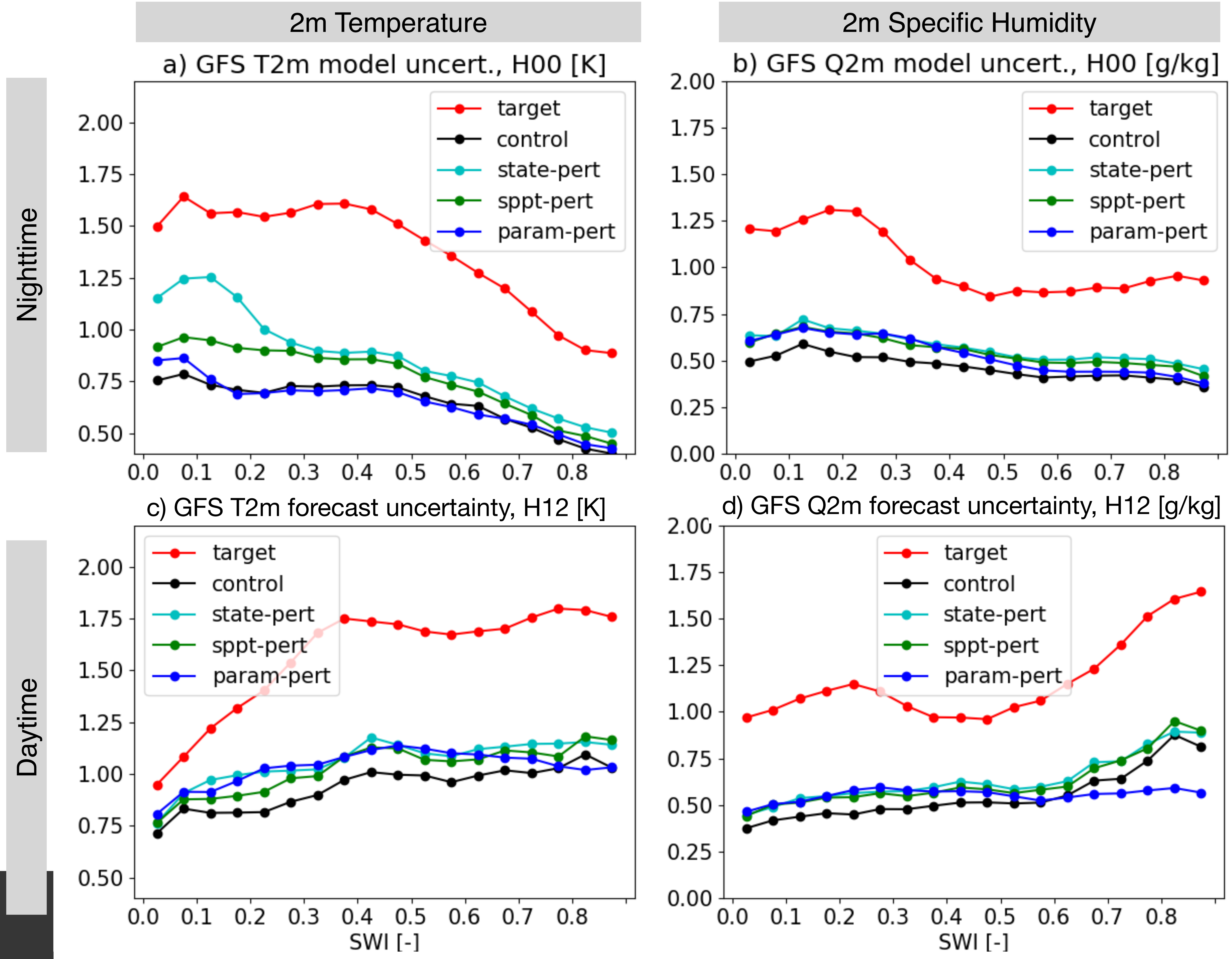
- Param-pert looks reasonable. Spread could be inflated by perturbing additional variables.

Target (red) is best estimate of forecast error standard deviation (c.f, independent obs). Others are ensemble-based estimates from each experiment.

Soil Wetness Index = Soil moisture, scaled between dry (0) and wet (1) limits.

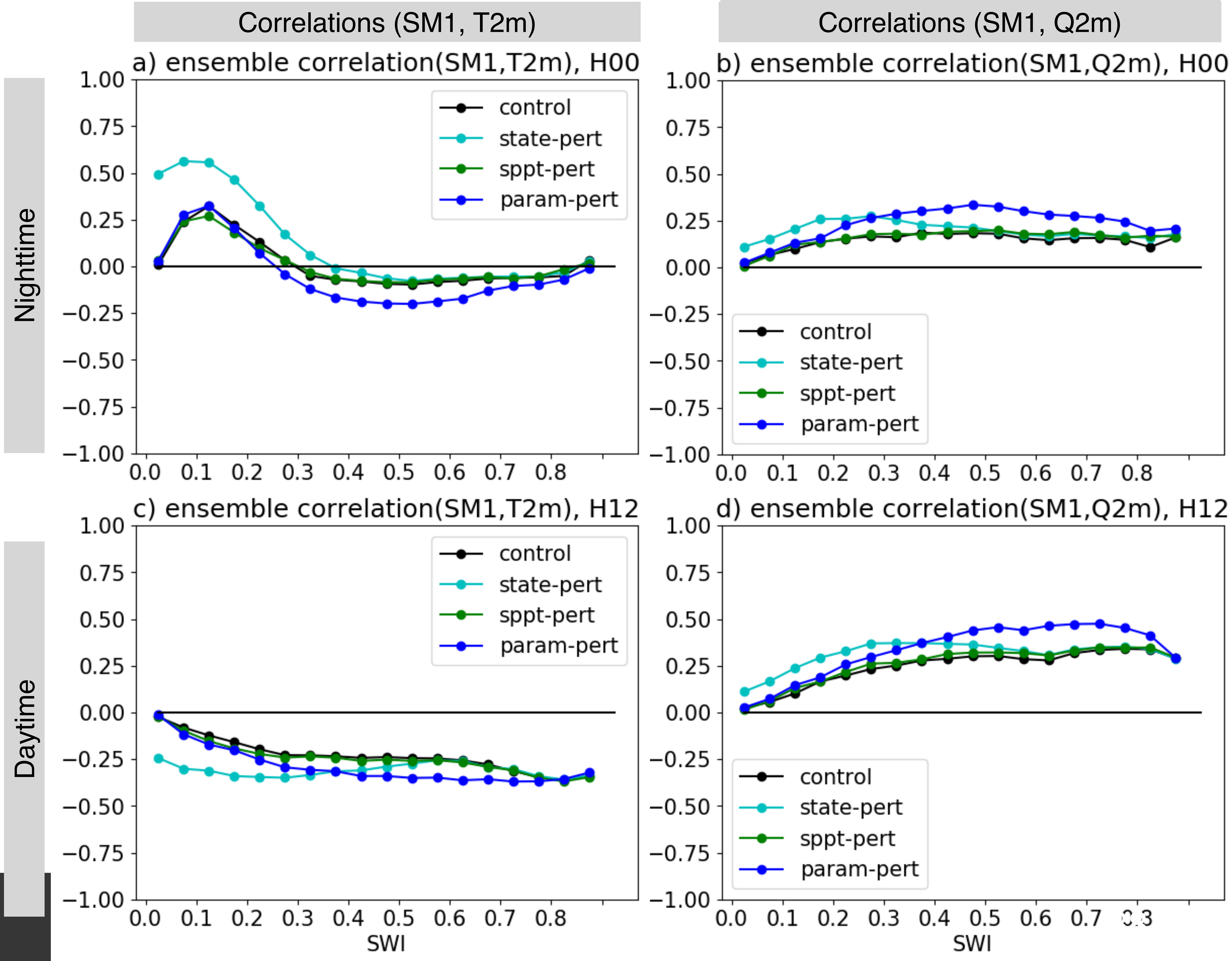
Ens. Spread, 2m Temperature and Specific Humidity

- Results binned into 6 hour local time windows
- Target estimates calculated by comparison to ERA-5 analysis
- Induced spread is generally limited in all experiments



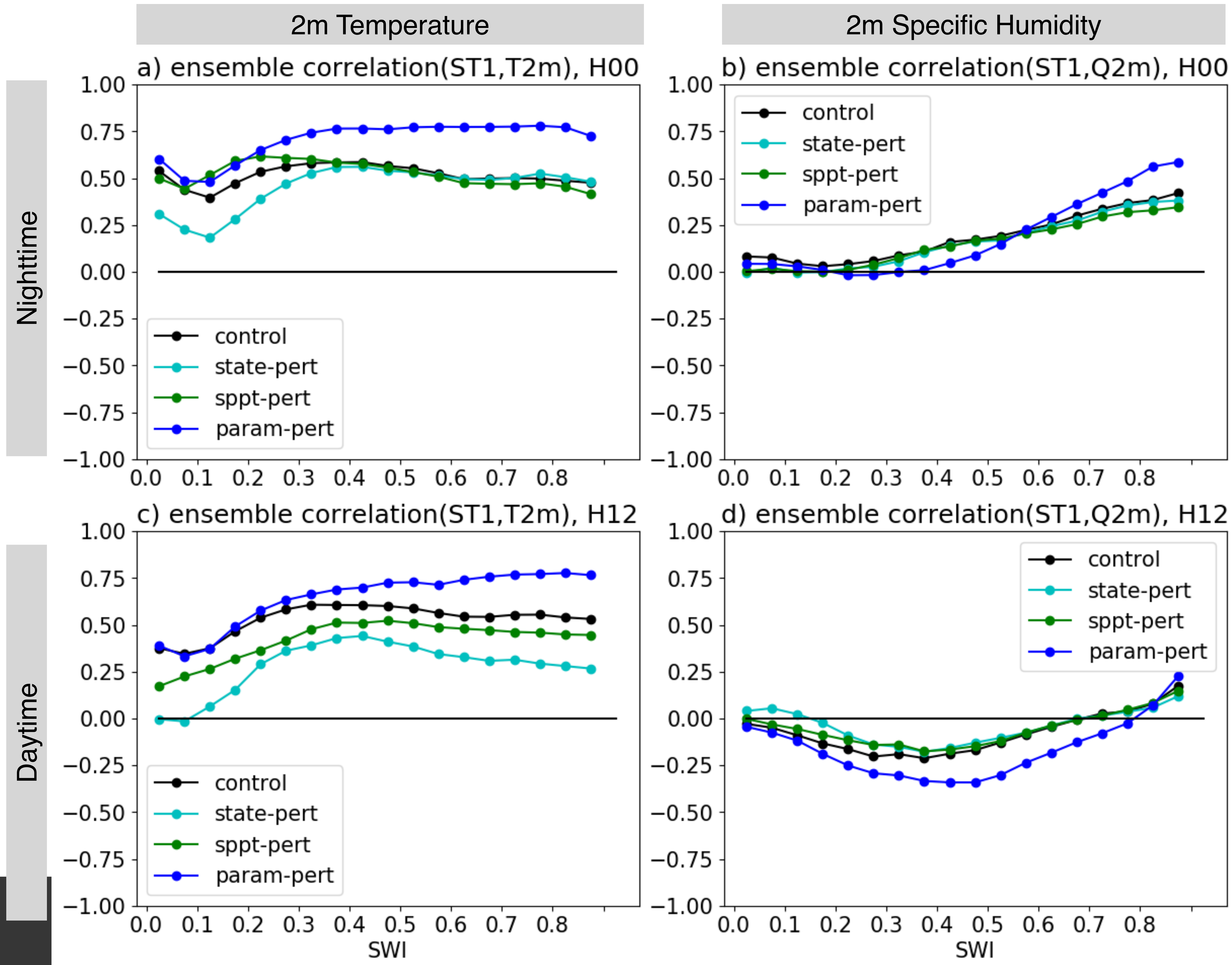
Land/Atmos Correlations, Soil Moisture Layer 1 (SM1)

- All experiments have incorrect positive SM1, T2m correlation in dry areas at night (problem in the model)
- State-pert strengthens correlations under dry conditions (when soil moisture drives land/ atmosphere coupling)
- Param-pert experiment generally strengthens the correlations

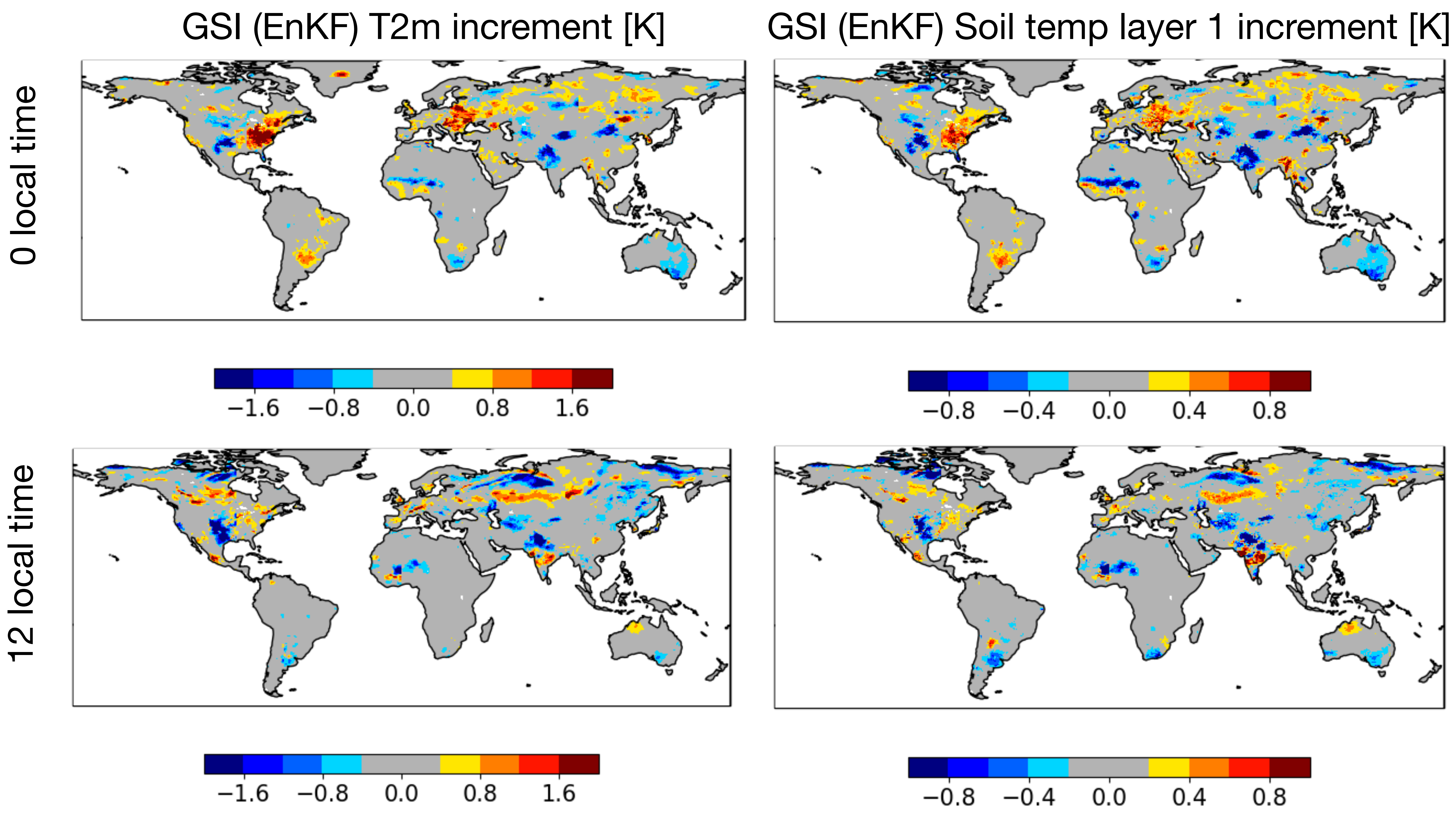


Land/Atmos Correlations, Soil Temperature Layer 1 (ST1)

- State-pert weakens the ST1, T2m correlations (atmosphere is driving the land/atmosphere coupling)
- Param-pert experiment again generally strengthens the correlations



Example Soil Temperature Increments



Note different color scales

Plots are for 202007015, and have been binned into night and day time windows