CAPS Convection-Allowing Model Forecasts and Ensemble Consensus Products for the Winter Weather Experiments

Keith Brewster & Nathan Snook
Andrew Berrington, Jun Park, Bryan Carroll,
Marshall Baldwin & Ming Xue

Center for Analysis and Prediction of Storms
University of Oklahoma
Norman, Oklahoma

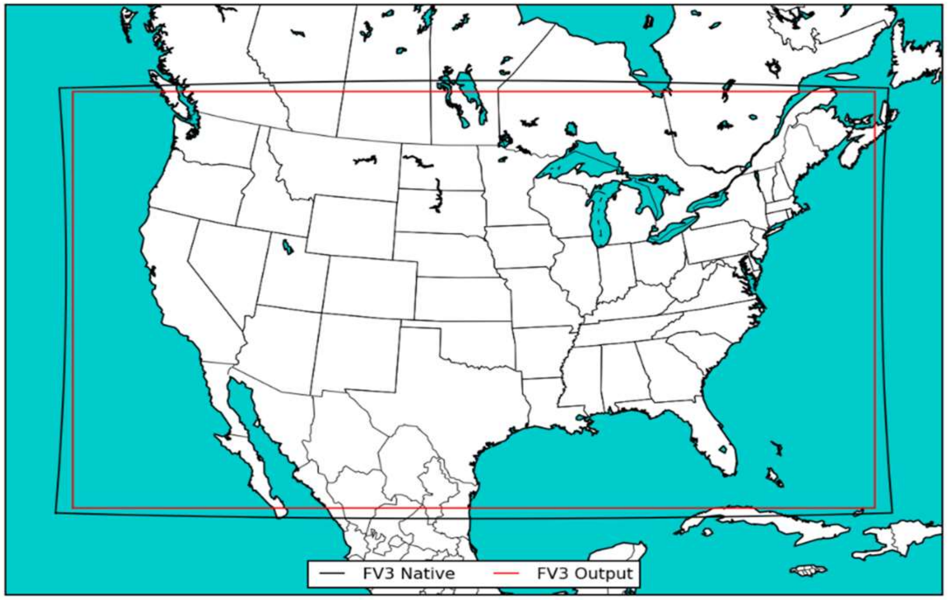
15th HMT Winter Weather Experiment
27 February 2025

CAPS Ensemble Experiment Goals

- Test FV3 CAM ensemble in quasi-operational winter setting: HMT Winter Weather Experiments – Add MPAS for 2025
- Generate CAM ensemble forecasts
- Test various physics combinations for possible operational use such as nascent Rapid Refresh Forecast System
- Evaluate ensemble consensus methods
- Develop machine learning (ML) algorithms to create quantitative rainfall and snowfall forecasts

CAPS Ensemble for 14th WWE (2023-2024)
FV3-LAM CAM Ensemble Configuration

- 11 FV3-LAM members
- 3 km grid spacing (GFDL grid)
- 64 vertical levels
- 84-hr forecasts initialized at 00 UTC
- Run at Texas Advanced Computing Center – Frontera
- Total of 30 days run for objective verification and ML training
- Results posted to web: https://caps.ou.edu/forecast/realtime/



TACC FRONTERA

14th HMT WWE (2023-24) CAPS Ensemble (11 Members)

Decoding member names:

M: Microphysics
- M0 = Thompson
- M1 = NSSL

B: Boundary Layer Scheme
- B0 = MYNN
- B1 = Shin-Hong
- B2 = TKE-EDMF

L: Land Surface Model
- L0 = NOAH
- L1 = NOAHMP
- L2 = RUC

P: Uses physics perturbations

I: Uses IC/LBC perturbations

Some members are configured similarly to operational or experimental models:

M0B0L2_P: Similar to RRFSm1

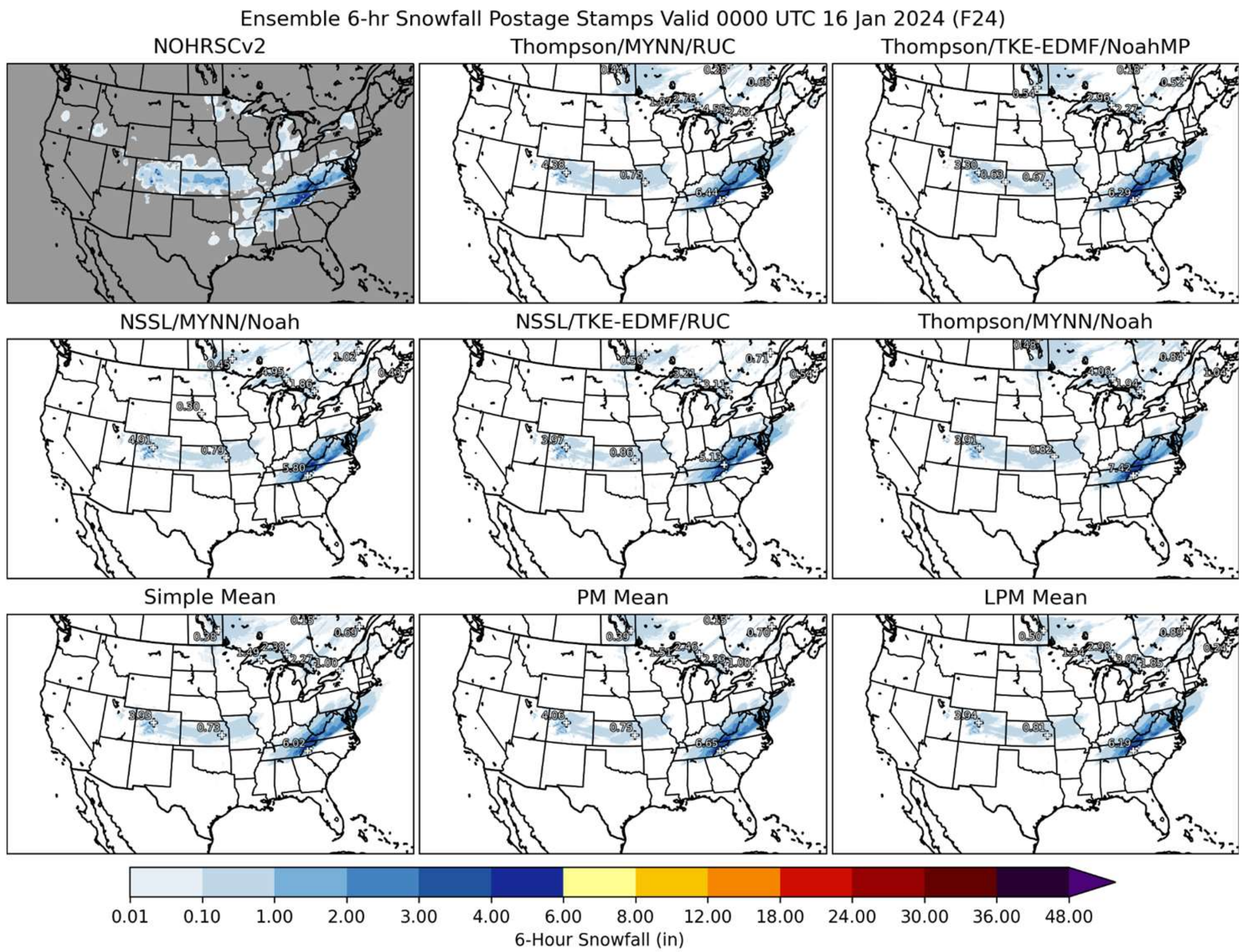
M1B2L2_P: Similar to RRFSmphys8

M0B2L1_P: Similar to GFSv16

Experiment	Microphysics	PBL	Surface	LSM	IC/LBC	AI/ML
Multi-Physics Core Configurations, Same IC/LBC						
M0B0L0_P	Thompson	MYNN	MYNN	NOAH	GFS	AI-1
M1B0L0_P	NSSL	MYNN	MYNN	NOAH	GFS	AI-2
M0B0L2_P	Thompson	MYNN	MYNN	RUC	GFS	
M1B2L2_P	NSSL	TKE-EDMF	GFS	RUC	GFS	
M0B2L1_P	Thompson	TKE-EDMF	GFS	NOAHMP	GFS	AI-3
Physics + IC Perturbation Ensemble						
M0B1L0_PI	Thompson	Shin-Hong	GFS	NOAH	GEFS_m1	
M0B2L1_PI	Thompson	TKE-EDMF	GFS	NOAHMP	GEFS_m2	
M0B2L2_PI	Thompson	TKE-EDMF	GFS	RUC	GEFS_m3	AI-4
M1B1L0_PI	NSSL	Shin-Hong	GFS	NOAH	GEFS_m4	
M1B2L1_PI	NSSL	TKE-EDMF	GFS	NOAHMP	GEFS_m5	
M1B2L2_PI	NSSL	TKE-EDMF	GFS	RUC	GEFS_m6	

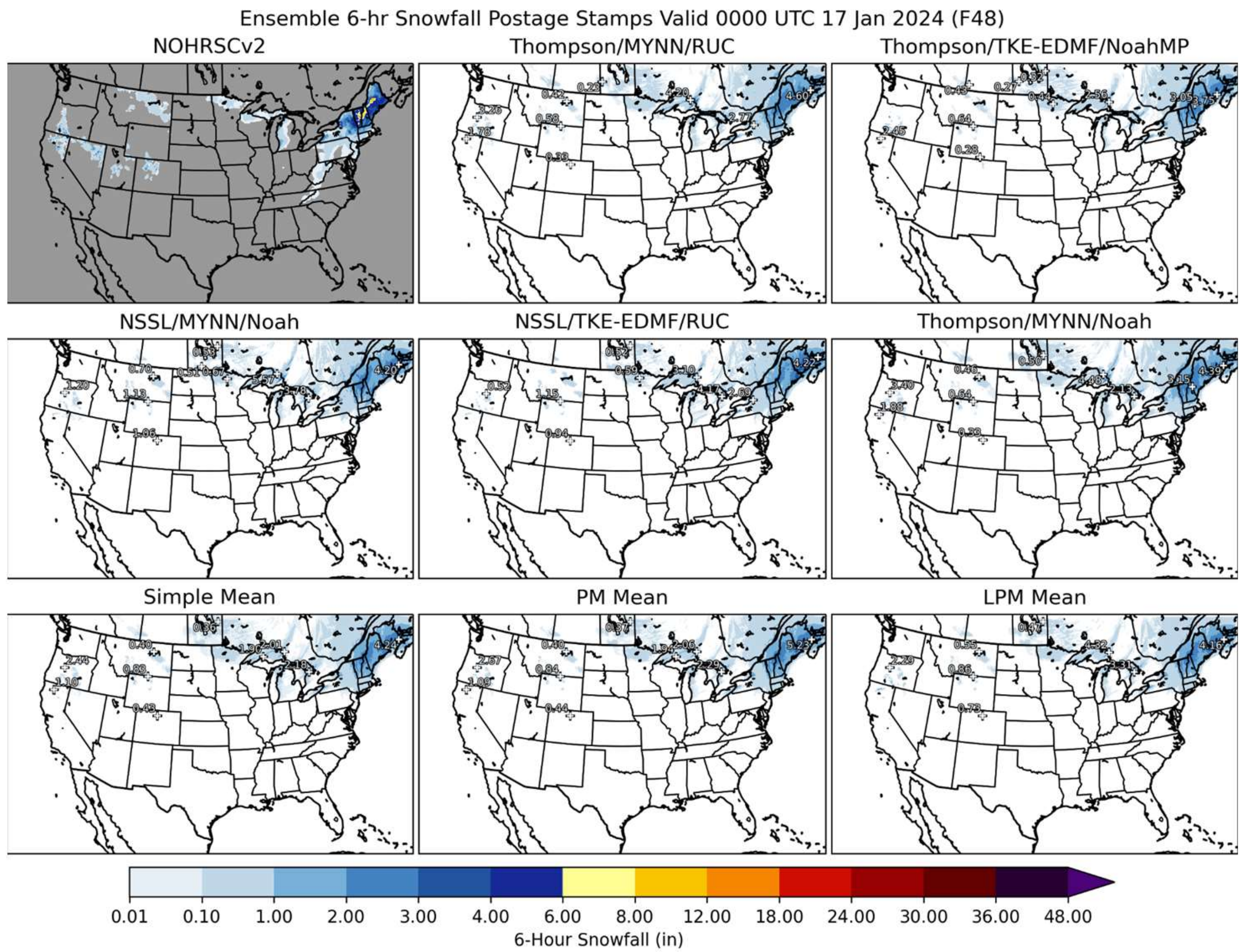
Sample Case (Core Configurations) – Jan. 15-17 2024

- 24-h forecast of 6-h accumulated snowfall forecasts valid at 00 UTC 16 Jan. 2024
- All core config. members capture the snowfall bands well, with slight variation in placement/intensity of heaviest snow
- Very little difference between ensemble consensus methods (simple mean, PM/LPM mean)
    - Close agreement between members
    - Broad, synoptically-driven features



Sample Case (Core Configurations) – Jan. 15-17 2024

- 48-h forecast of 6-h accumulated snowfall forecasts valid at 00 UTC 17 Jan. 2024
- All members underpredict intensity of heaviest snowfall in VT/NH/ME and fail to capture light snowfall extending south along the Appalachians.



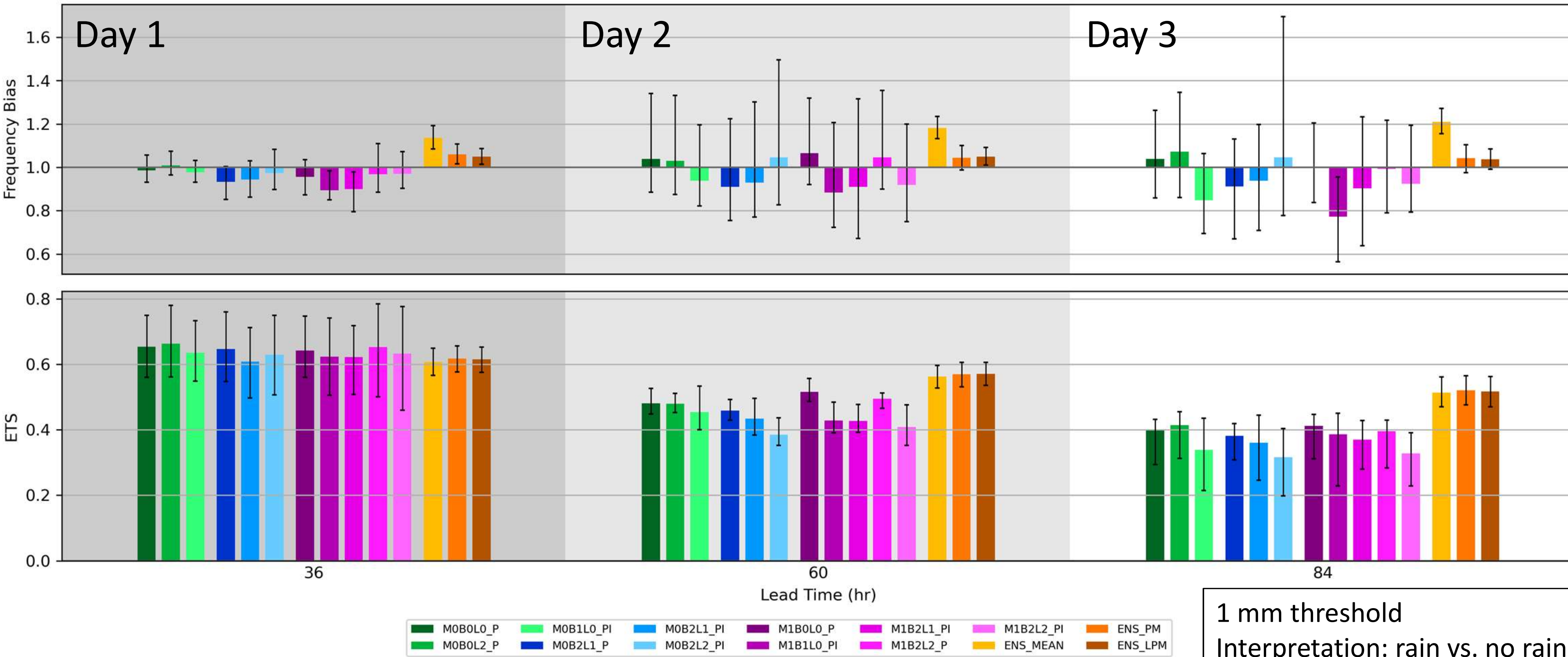
Forecast Verification (Seasonal Summary Statistics)

- Observations used:
  - Total Precipitation: Stage-4 precipitation accumulation
  - Snowfall: NOHRSC Snowfall Analyses

- Software package used: MET-Plus v11.1.0 from the Developmental Testbed Center)

- Metrics include frequency bias and equitable threat score (ETS)
  - Several intensity thresholds are considered to focus on light versus heavy rainfall/snowfall.
  - All verification metrics are calculated using a 30 km neighborhood radius.

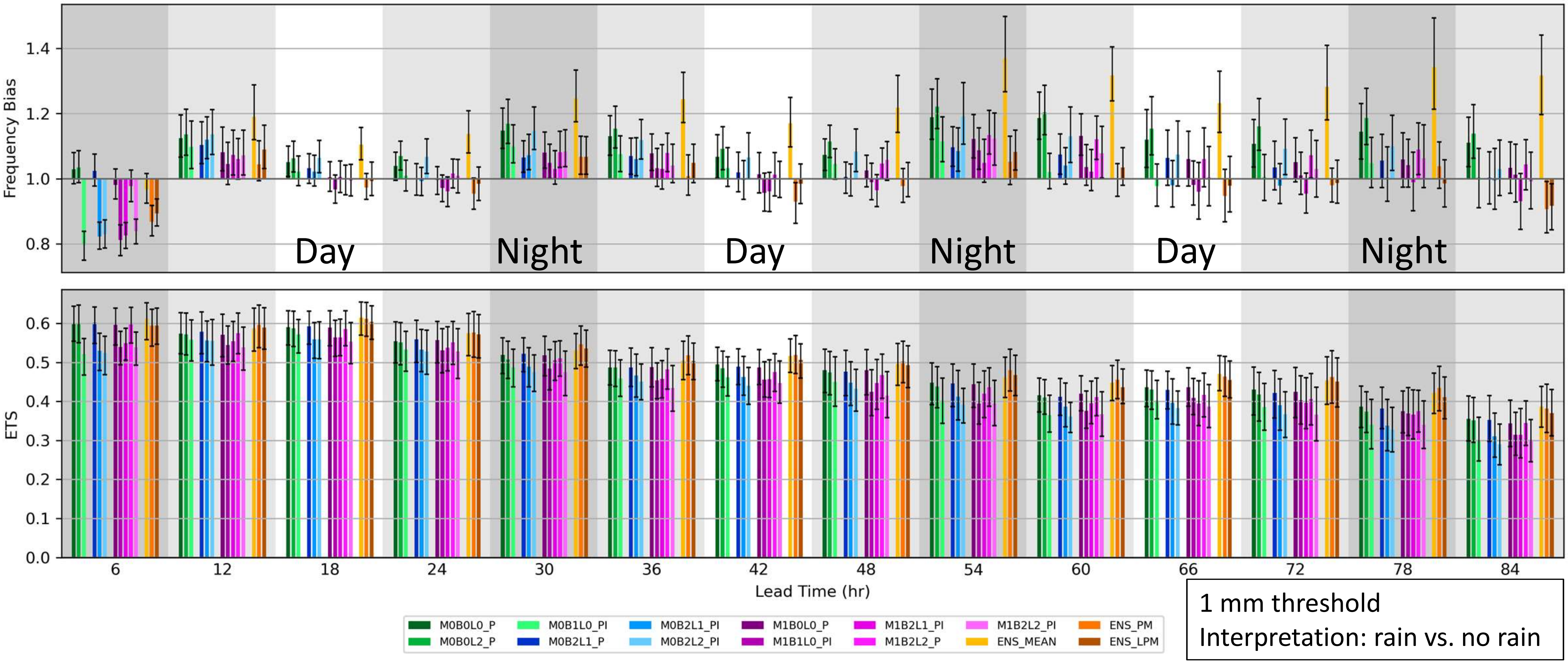
Verification: 24-h accumulated precipitation, 1 mm (precip/no-precip)



1 mm threshold
Interpretation: rain vs. no rain

- Individual member biases vary, but are generally near unbiased (0.8 – 1.1).
- Simple mean has an overall high bias at 1 mm threshold, as expected due to smoothing
- ETS for ensemble consensus products outperforms individual members for day 2 and especially day 3.

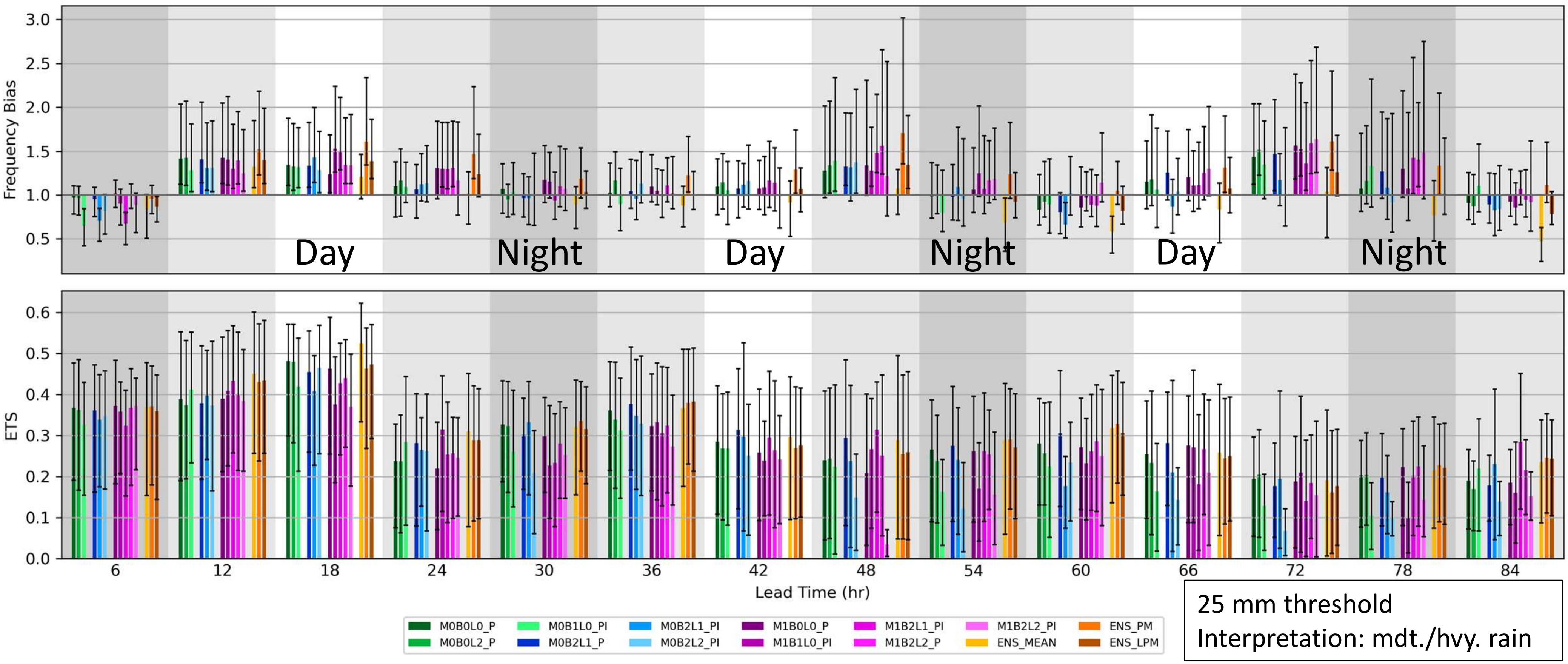
Verification: 6-h accumulated precipitation, 1 mm threshold



1 mm threshold
Interpretation: rain vs. no rain

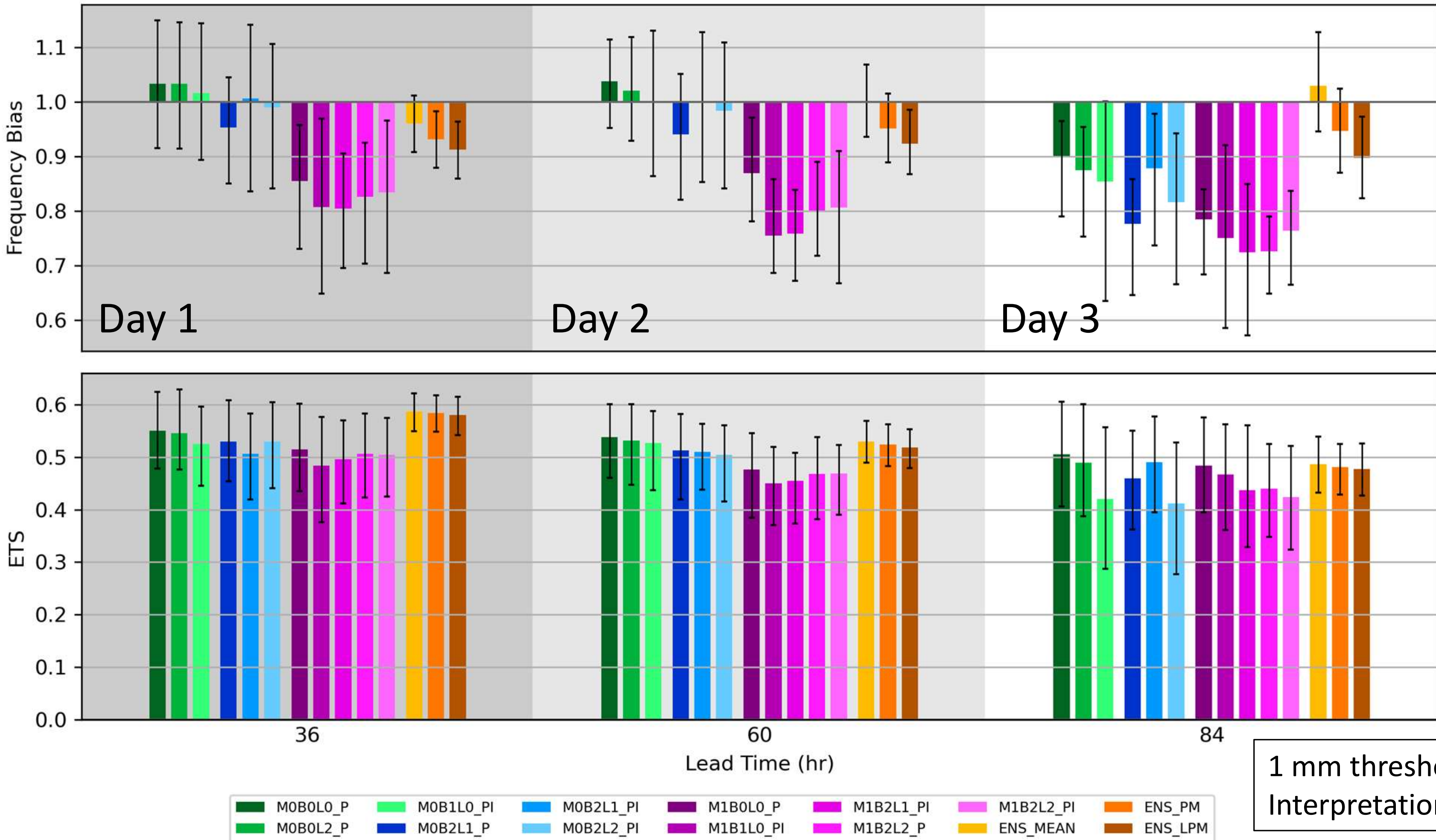
- Individual member biases still near unbiased (0.8 – 1.2), simple mean bias is higher, especially at night.
- PM and especially LPM exhibit very good bias characteristics, day and night.
- Notable diurnal cycle impacts, particularly for bias (high bias maximized during night and early morning).

Verification: 6-h accumulated precipitation, 25 mm threshold



- Spin up 0-6 h, then high bias during daytime and evening hours, near-neutral bias overnight into the morning hours.
- Diurnal cycle evident in both frequency bias and ETS (ETS highest in early morning hours, lowest in evening).
- Relative member performance varies with lead-time; M0B0L0 and M0B2L1 are among best performers.

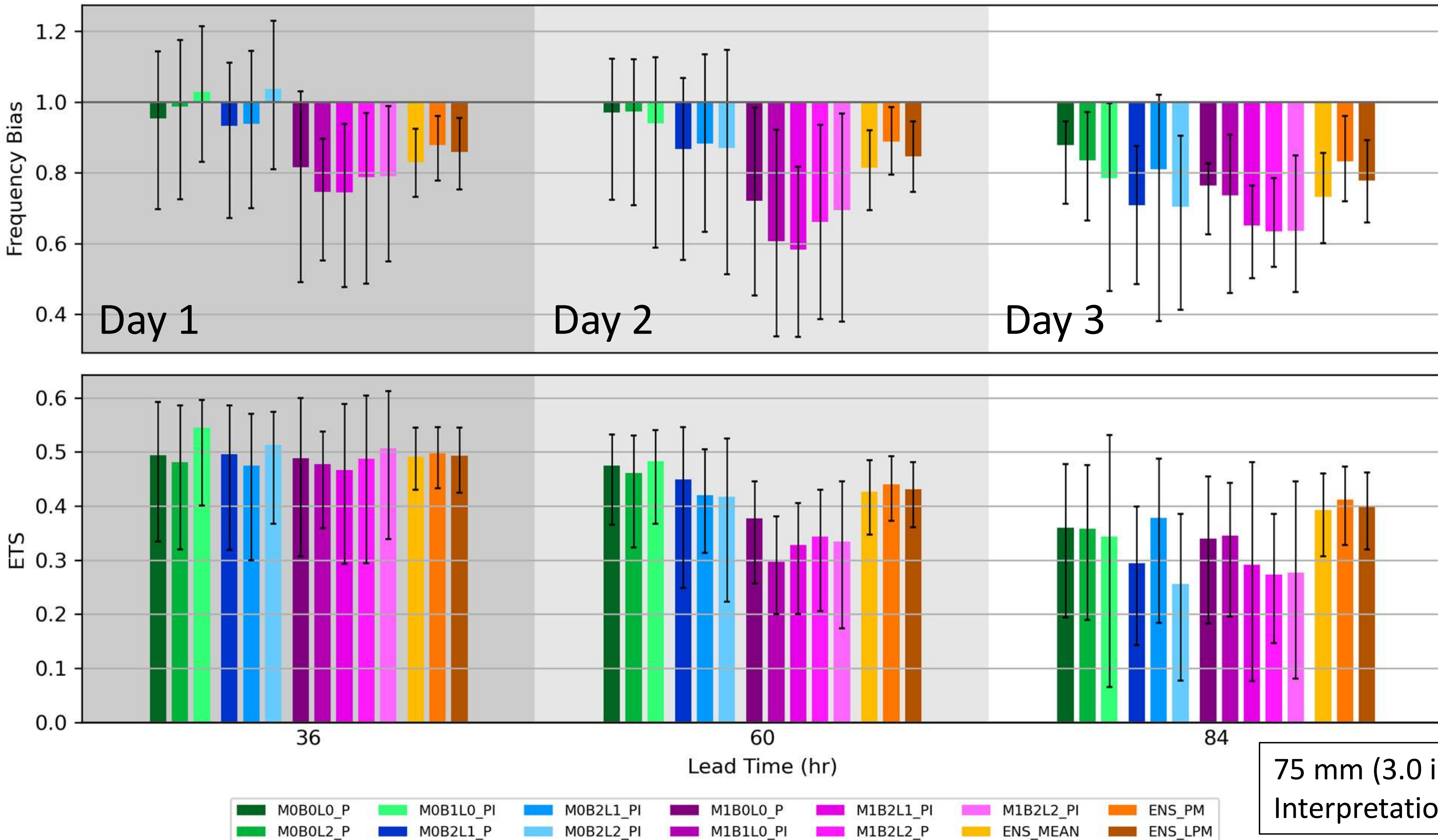
Verification: 24-h accumulated snowfall, 1 mm threshold



- Low bias in total area receiving snowfall, particularly for Day 3 and for members using NSSL microphysics (purple bars)
- Ensemble consensus forecasts are nearly unbiased (0.9 – 1.0).
- Best performing member: M0B0L0

1 mm threshold
Interpretation: snow vs. no snow

Verification: 24-h accumulated snowfall, 75 mm threshold



- NSSL microphysics members (purple bars) show low bias in accumulated mdt/hvy snow.
- Benefit of ensemble consensus on forecast skill only becomes evident late in forecast period (Day 3).

75 mm (3.0 in.) threshold
Interpretation: moderate snow

Planned 15th HMT WWE (2024-2025) CAPS Ensemble

Decoding member names:
M: Microphysics
- M0 = Thompson
- M1 = NSSL

B: Boundary Layer Scheme
- B0 = MYNN
- B1 = Shin-Hong
- B2 = TKE-EDMF

L: Land Surface Model
- L0 = NOAH
- L1 = NOAHMP
- L2 = RUC

C: Uses cumulus scheme
MP: MPAS member

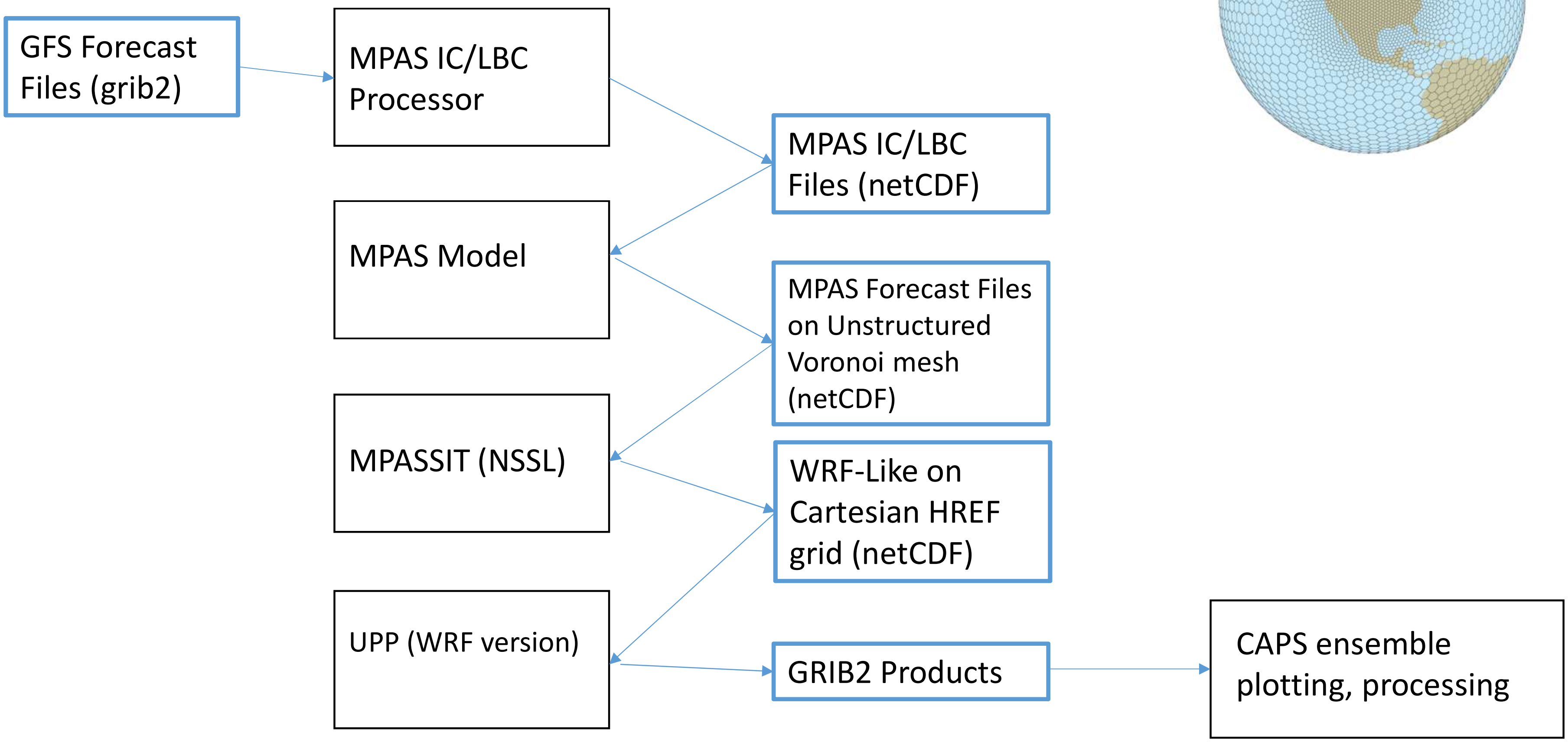
Experiment	Microphysics	PBL	Surface	LSM	IC/LBC	Cumulus	AI/ML
FV3-LAM Ensemble (Core Configurations)							
M0B0L0	Thompson	MYNN	MYNN	NOAH	GFS	None	AI-1
M1B0L0	NSSL	MYNN	MYNN	NOAH	GFS	None	AI-2
M1B0L2	NSSL	MYNN	MYNN	RUC	GFS	None	
M0B2L1	Thompson	TKE-EDMF	GFS	NOAHMP	GFS	None	AI-3
M0B0L2	Thompson	TKE-EDMF	MYNN	RUC	GFS	None	AI-4
Experimental MPAS Ensemble							
M0B0L2_MP	Thompson	MYNN	MYNN	RUC	GEFS_m1	None	
M1B0L2_MP	NSSL	MYNN	MYNN	RUC	GEFS_m2	None	
M0B0L0_MP	Thompson	MYNN	MYNN	NOAH	GEFS_m3	None	
M1B0L0_MP	NSSL	MYNN	MYNN	NOAH	GEFS_m4	None	
M1B0L2C_MP	NSSL	MYNN	MYNN	NOAH	GEFS_m5	SA-New-Tiedtke	

Some members are configured similarly to operational or experimental models:
M1B0L0_P: Similar to WoFS
M1B0L2: Similar to RRFSm1
M0B2L1_P: Similar to GFSv16

M0B0L2_MP: Similar to GSL-01
M1B0L2_MP: Similar to NSSL-01
M0B0L0_MP: Similar to NCAR-01

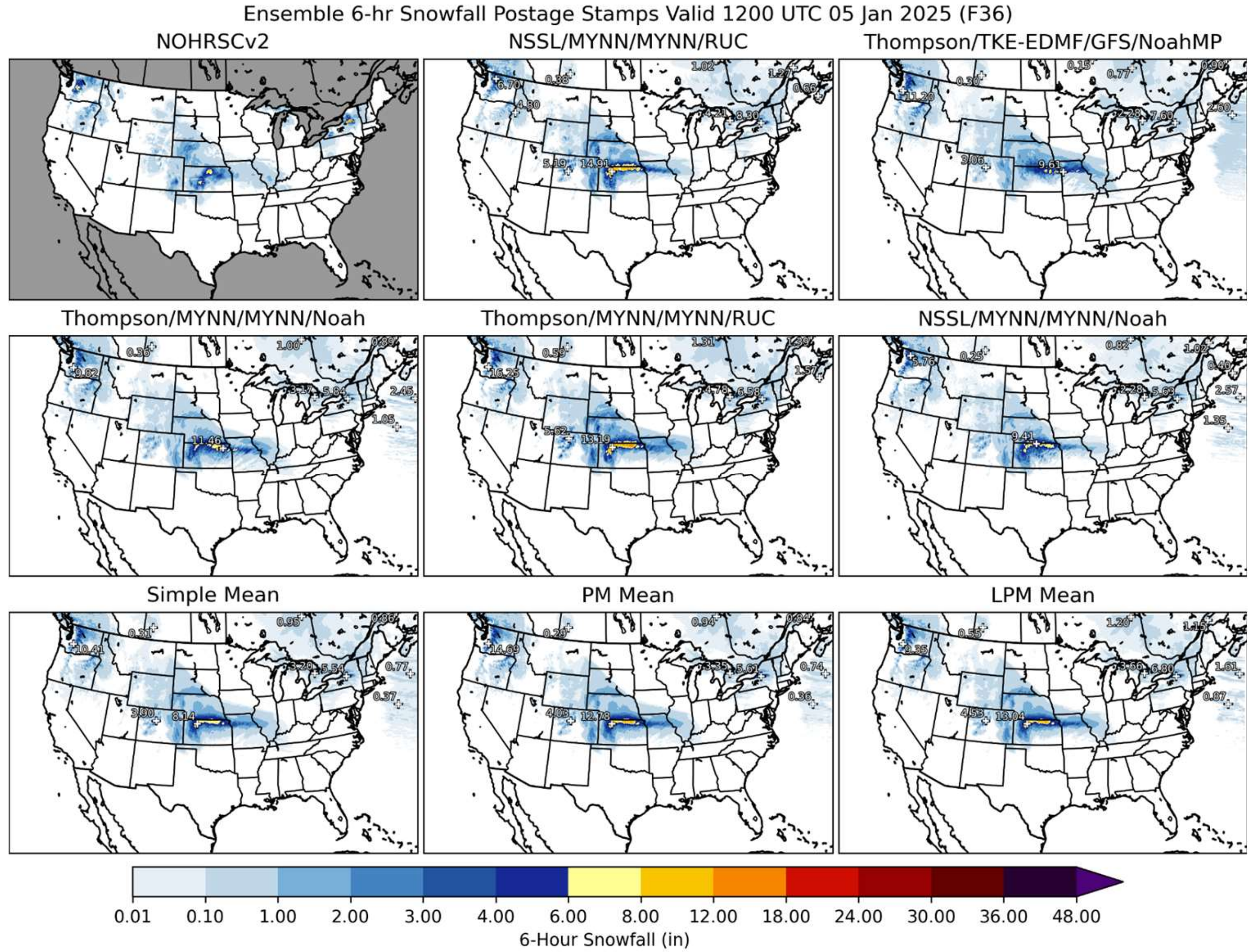
Near real-time forecast graphics are available online:
https://caps.ou.edu/forecast/realtime/

MPAS Workflow

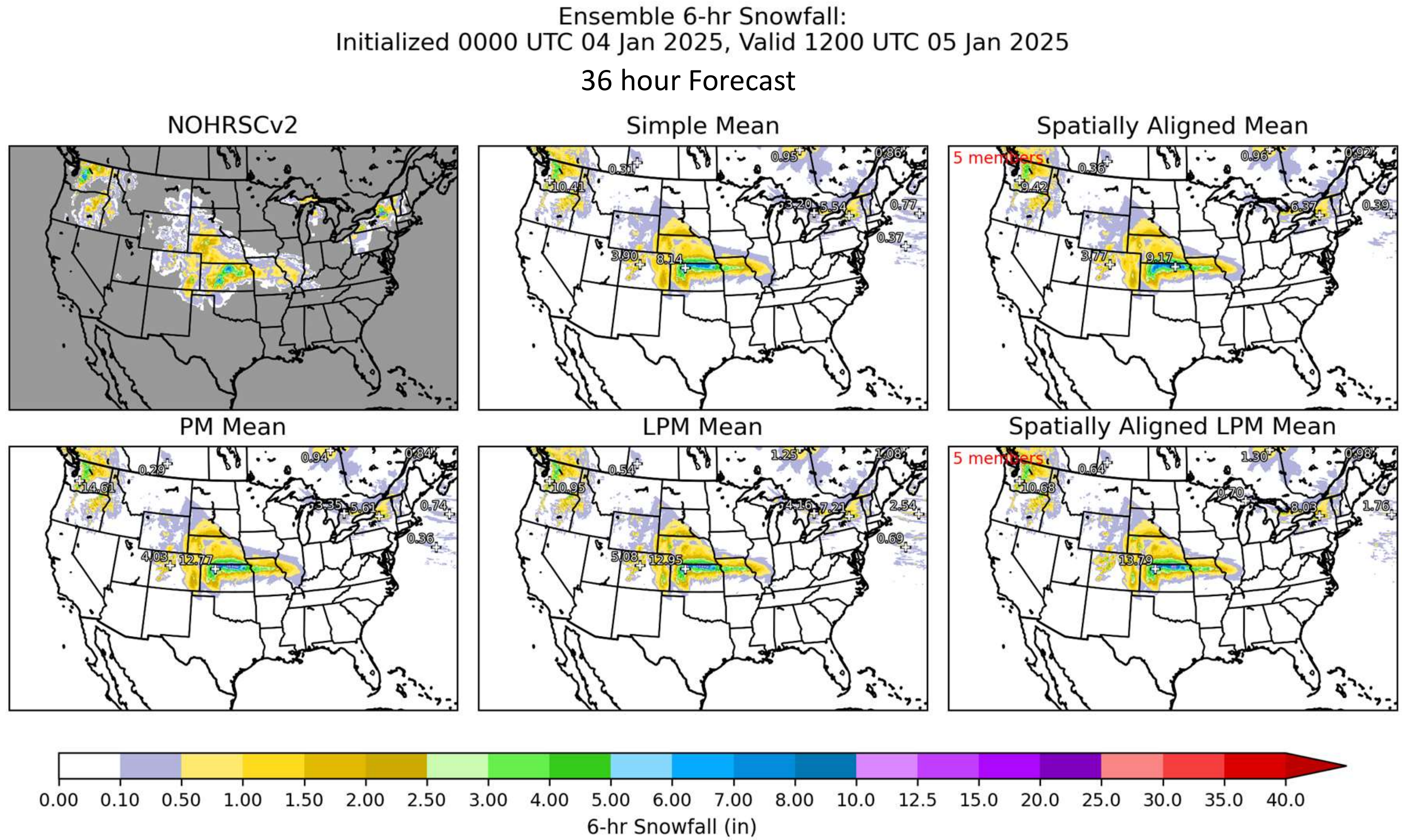


Sample Case (FV3-LAM Members) – Jan 5-7 2025

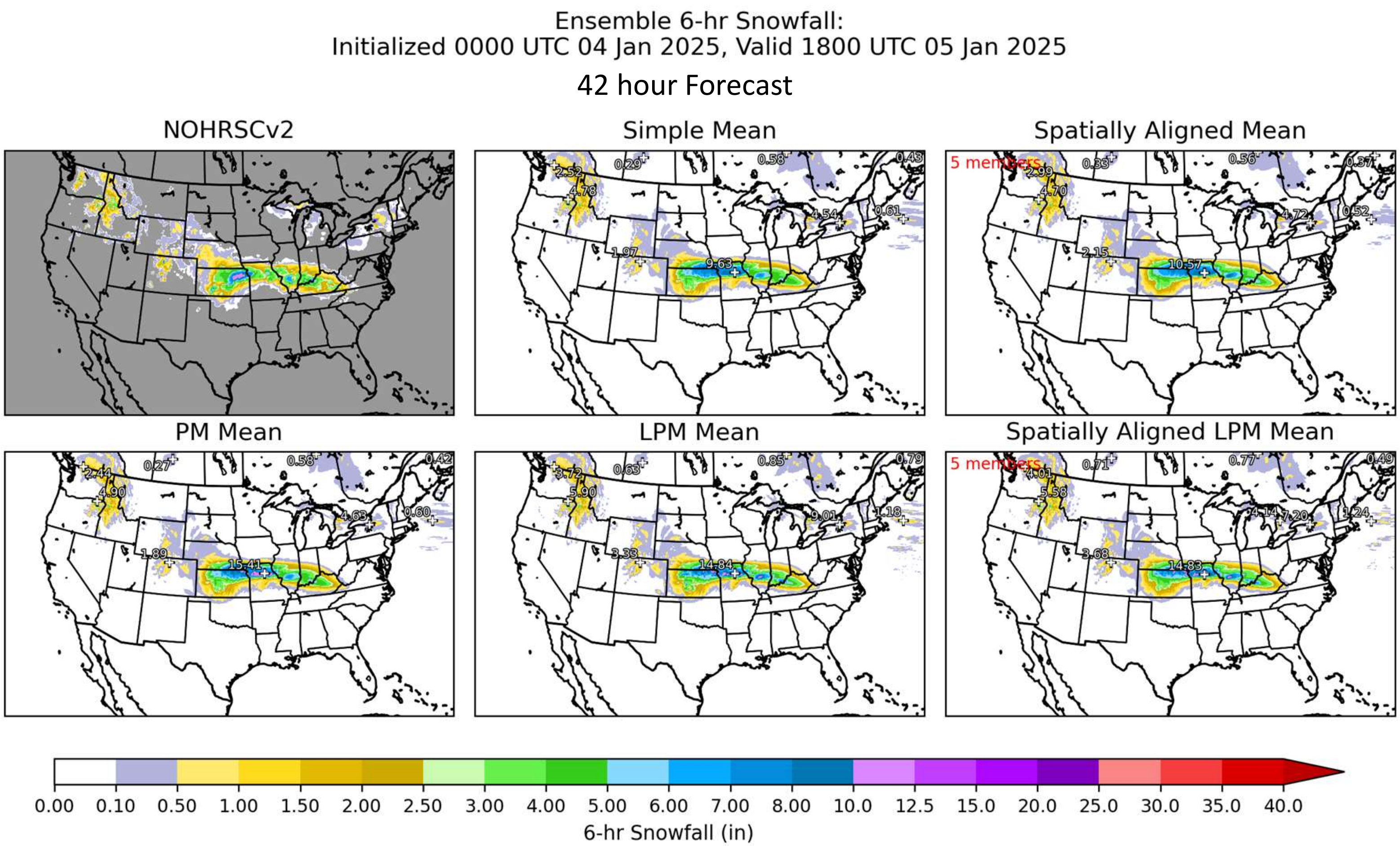
36 hour Forecast



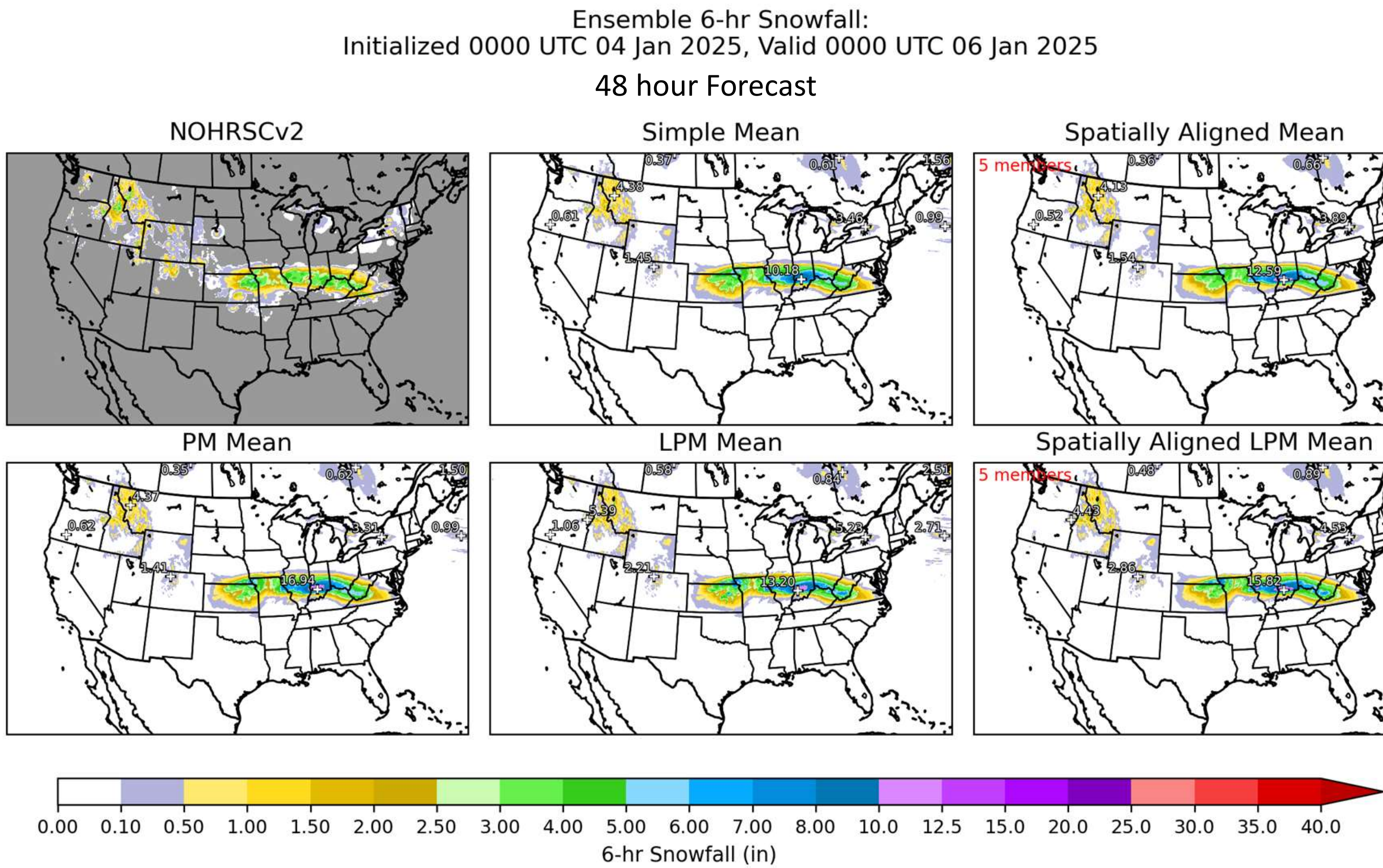
Sample Case (Ensemble Consensus Products) – Jan 5-7 2025



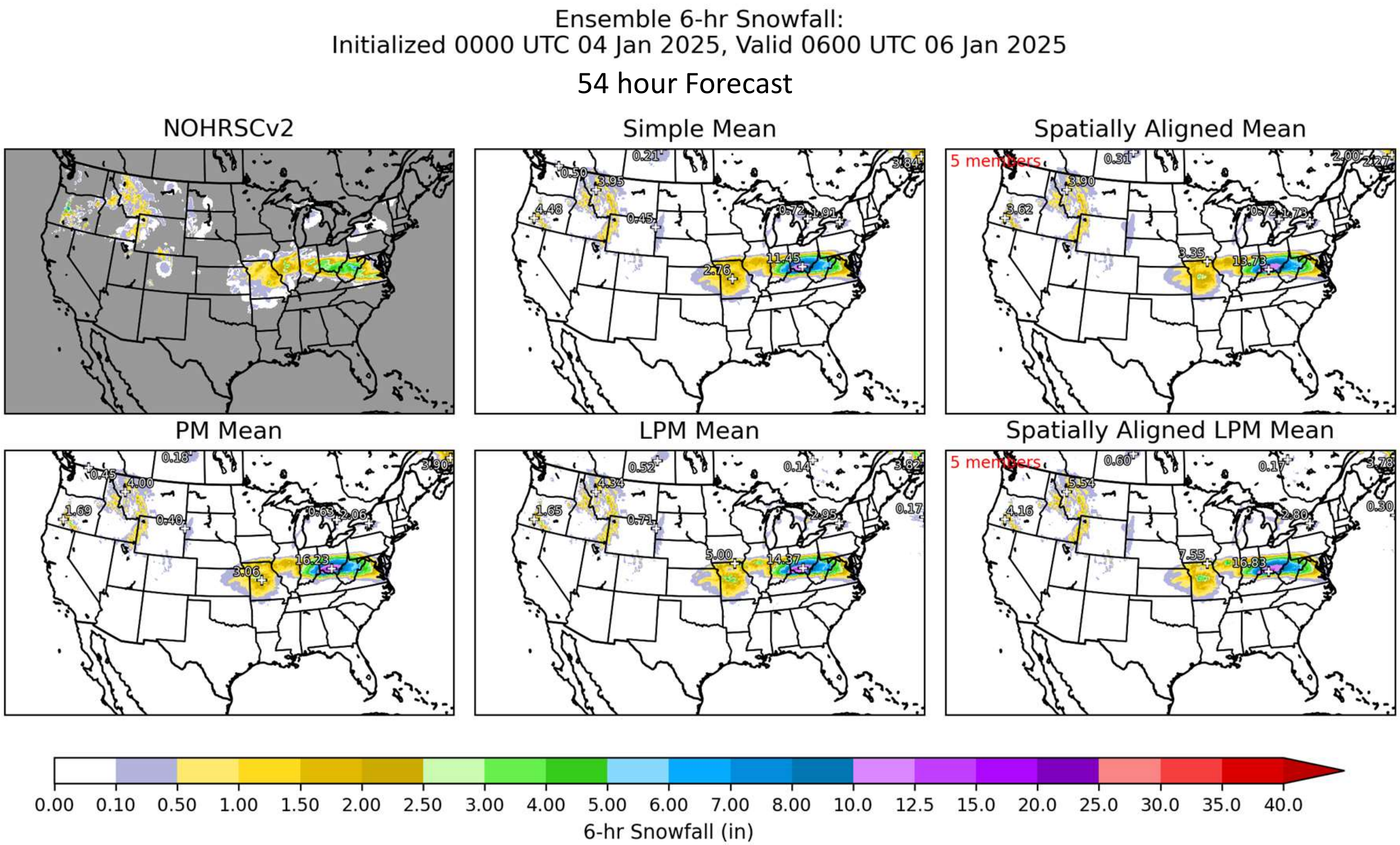
Sample Case (Ensemble Consensus Products) – Jan 5-7 2025



Sample Case (Ensemble Consensus Products) – Jan 5-7 2025



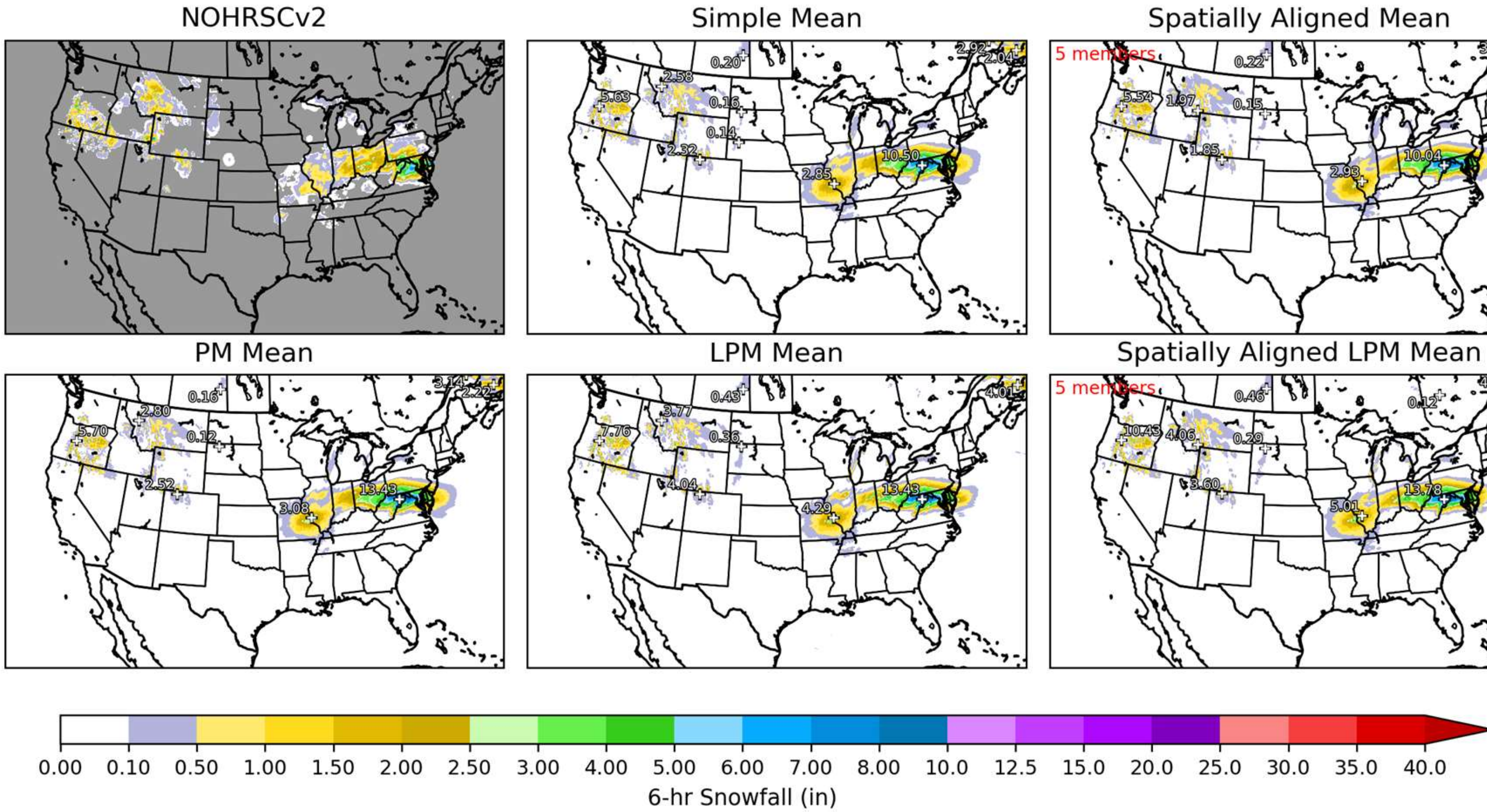
Sample Case (Ensemble Consensus Products) – Jan 5-7 2025



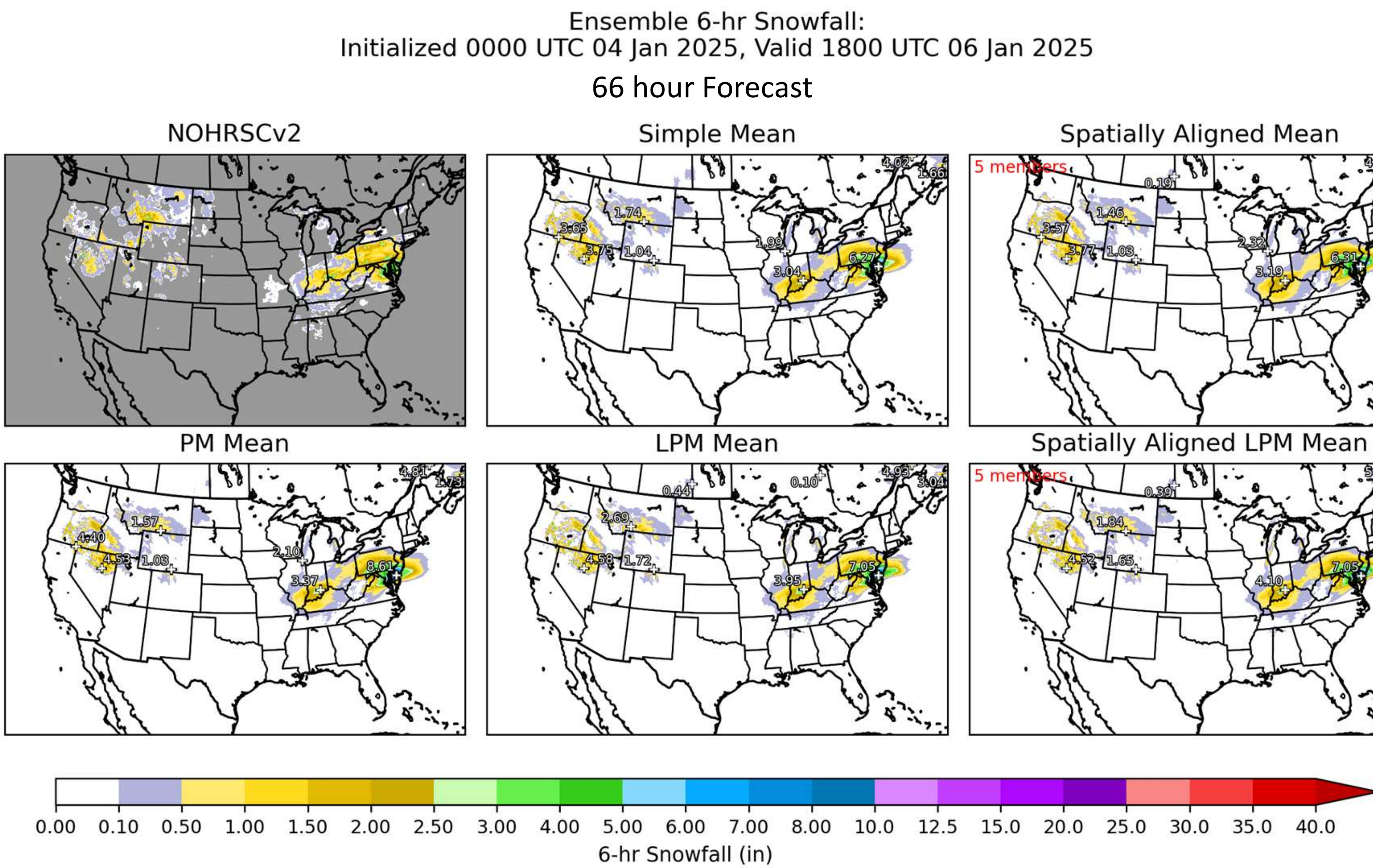
Sample Case (Ensemble Consensus Products) – Jan 5-7 2025

Ensemble 6-hr Snowfall:
Initialized 0000 UTC 04 Jan 2025, Valid 1200 UTC 06 Jan 2025

60 hour Forecast



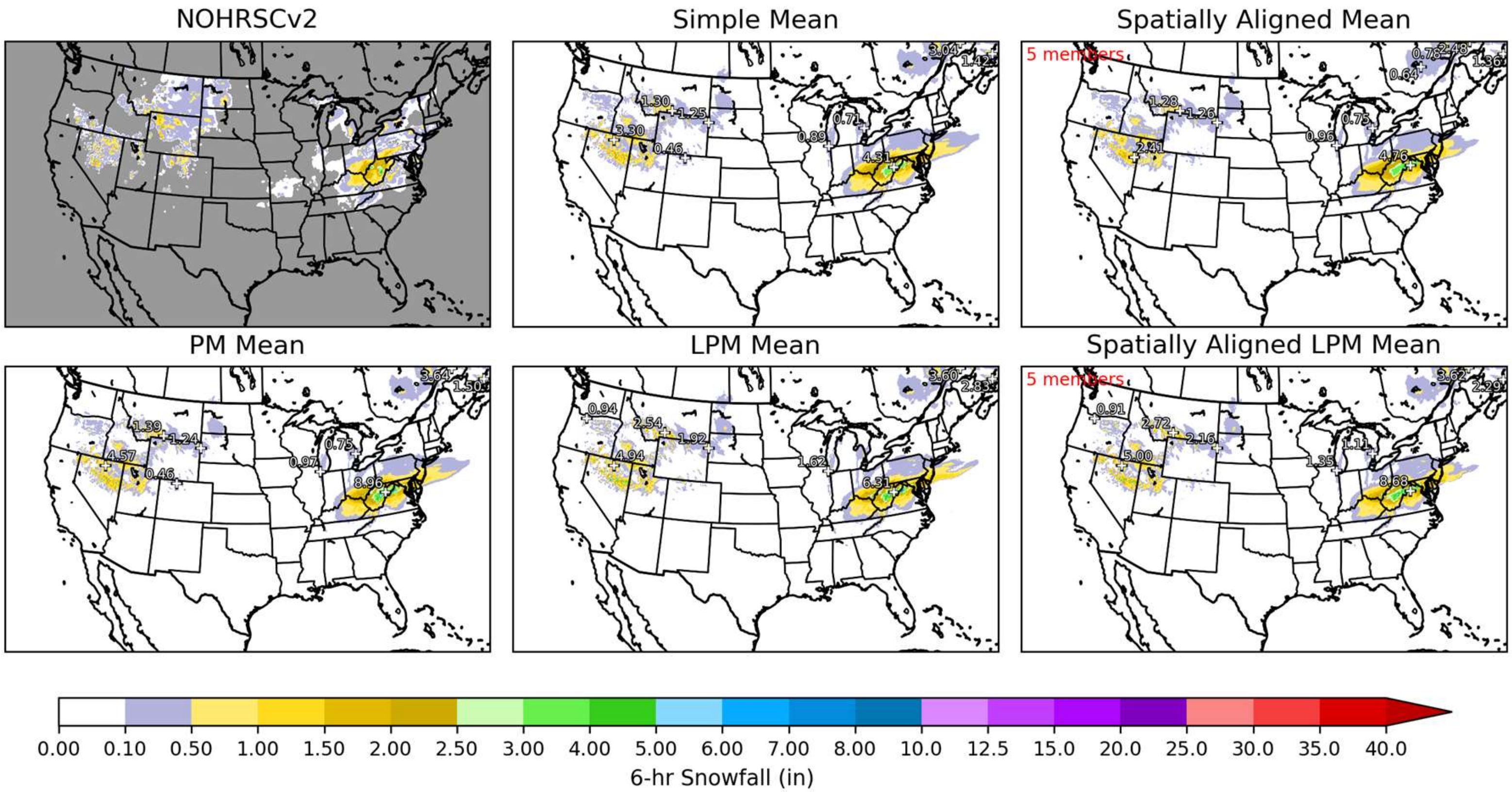
Sample Case (Ensemble Consensus Products) – Jan 5-7 2025



Sample Case (Ensemble Consensus Products) – Jan 5-7 2025

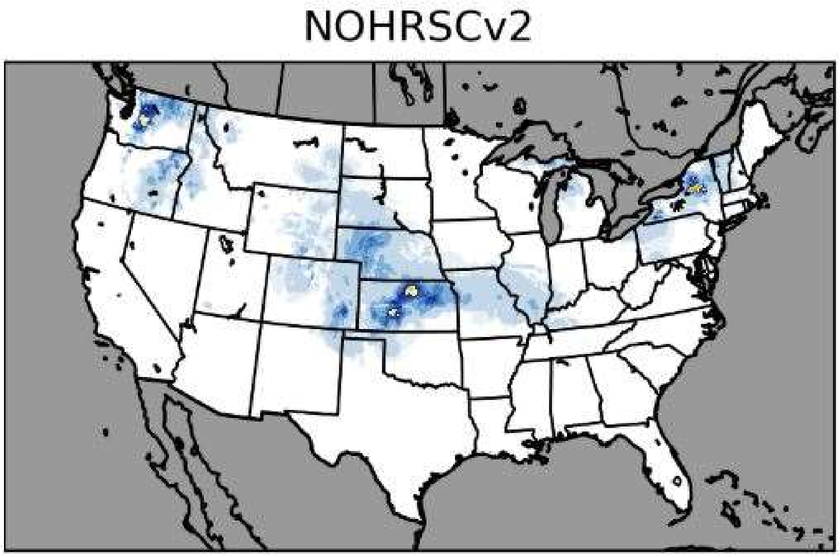
Ensemble 6-hr Snowfall:
Initialized 0000 UTC 04 Jan 2025, Valid 0000 UTC 07 Jan 2025

72 hour Forecast

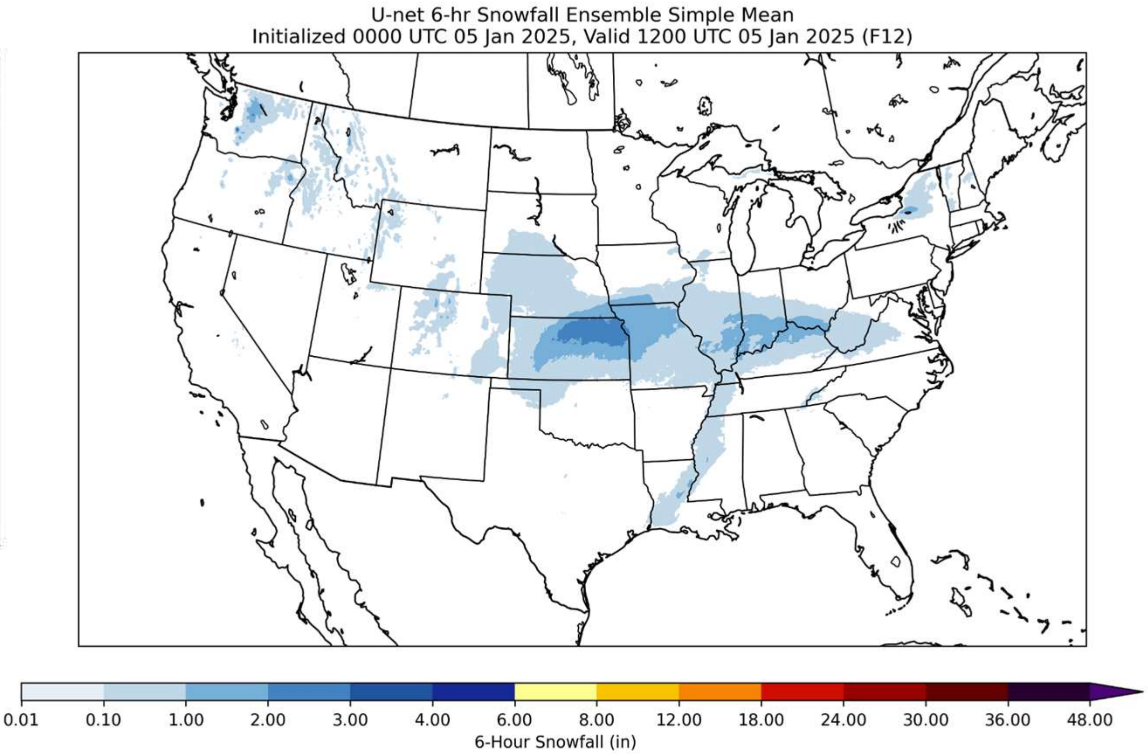


Sample Case (Machine Learning Products) – Jan 5-6 2025

6-h Snowfall 12z 05 Jan 2025

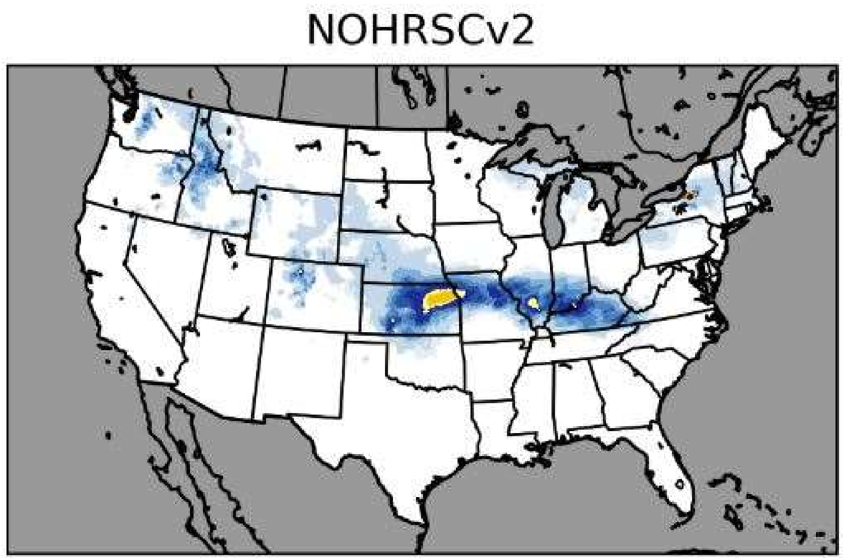


12-hour U-net Forecast

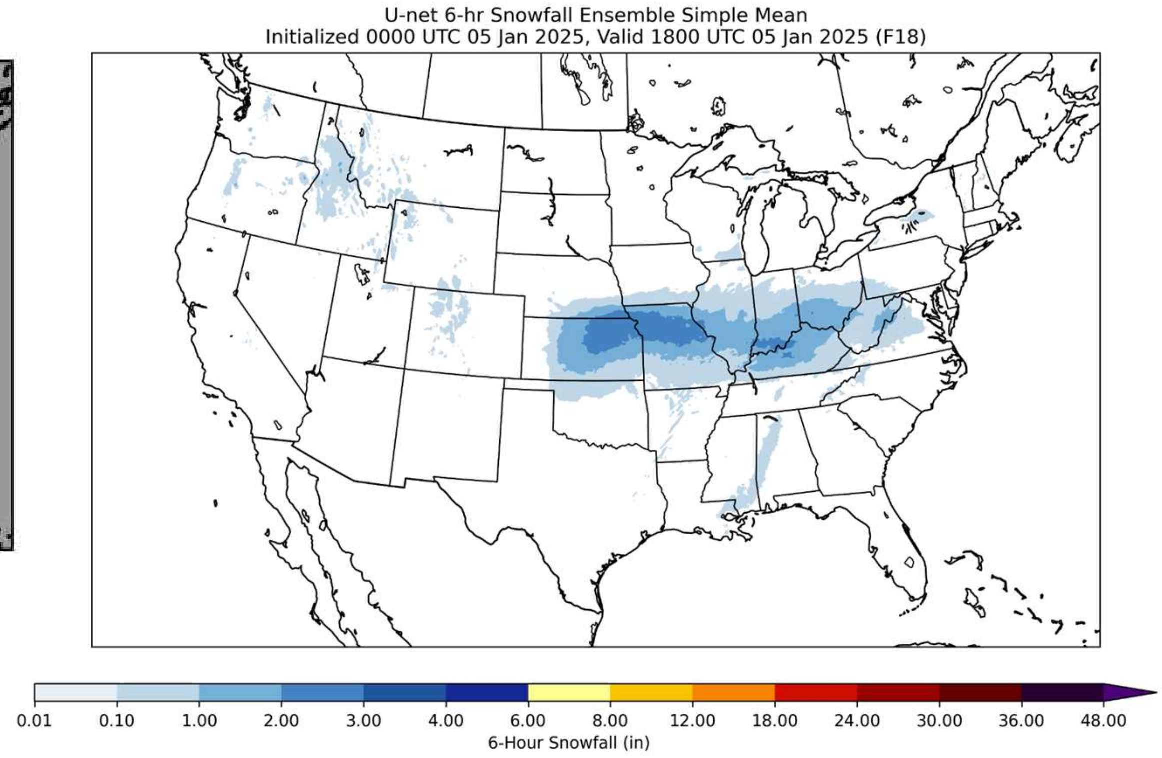


Sample Case (Machine Learning Products) – Jan 5-6 2025

6-h Snowfall 18z 05 Jan 2025

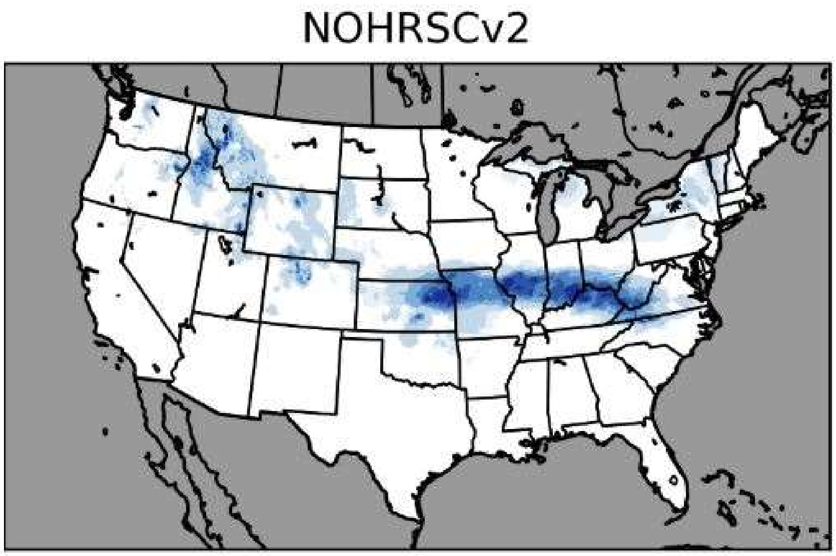


18-hour U-net Forecast

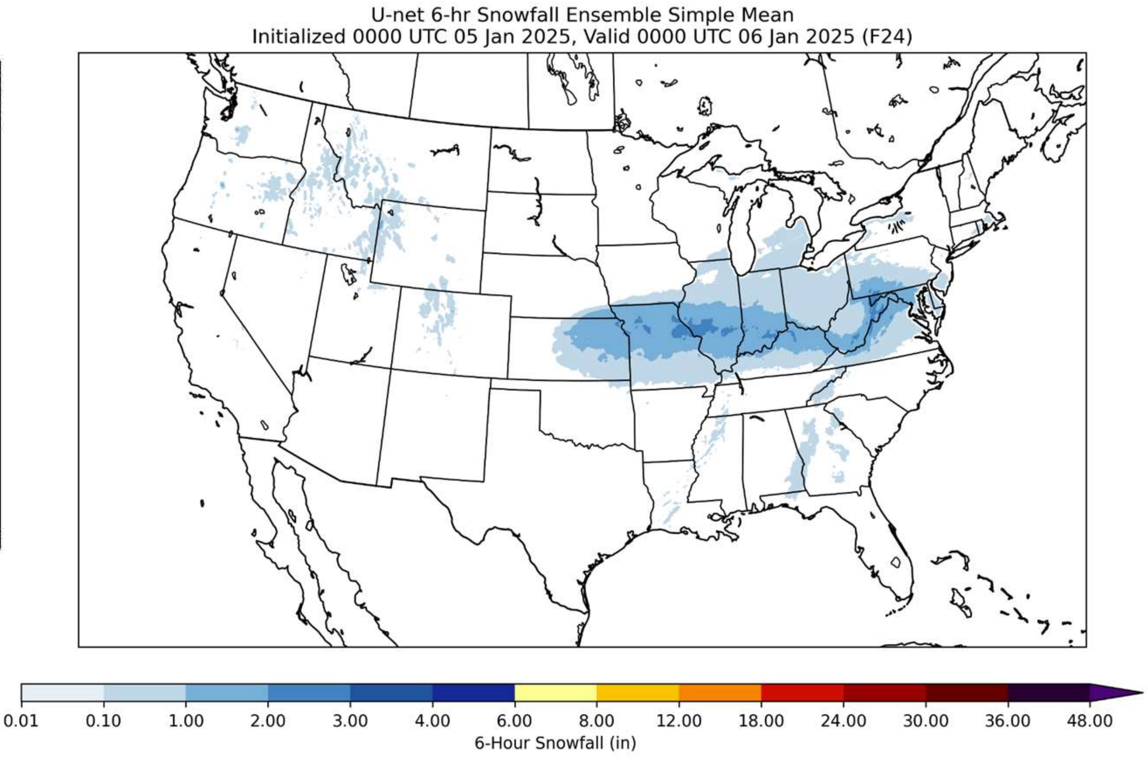


Sample Case (Machine Learning Products) – Jan 5-6 2025

6-h Snowfall 00z 06 Jan 2025

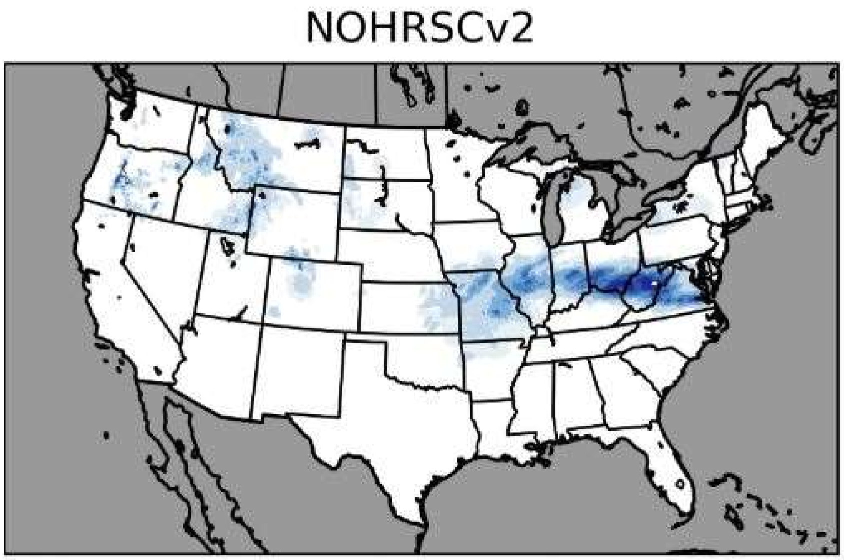


24-hour U-net Forecast

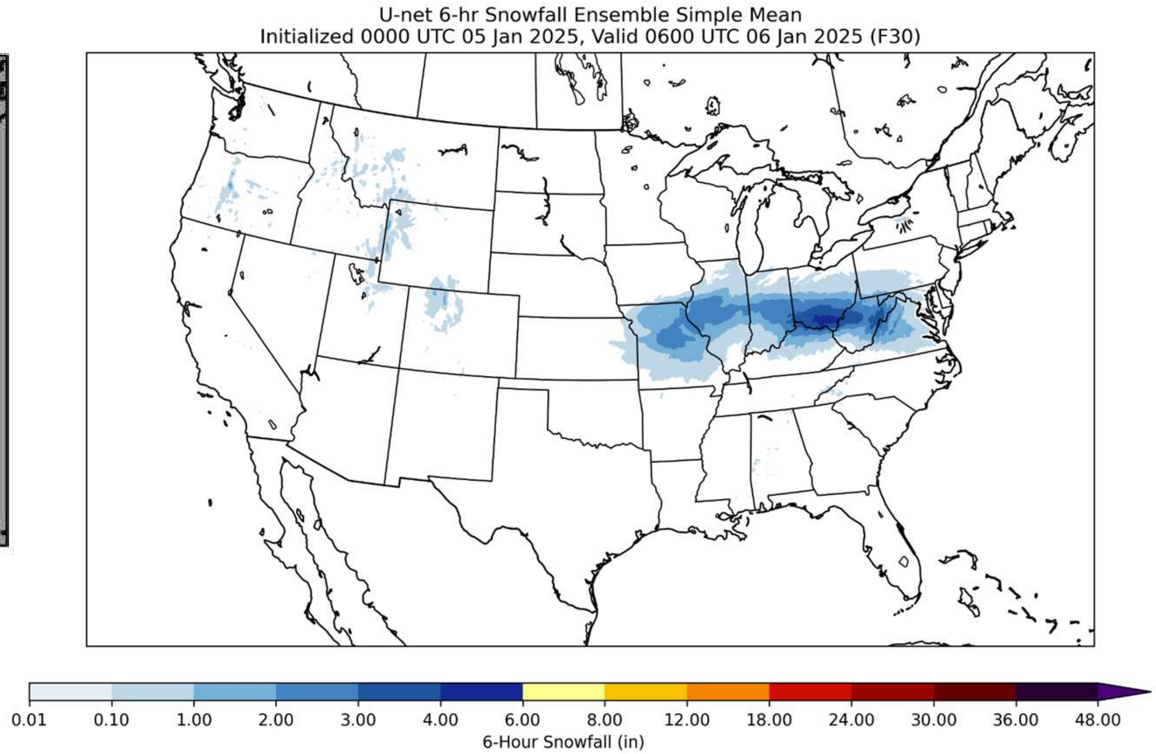


Sample Case (Machine Learning Products) – Jan 5-6 2025

6-h Snowfall 06z 06 Jan 2025

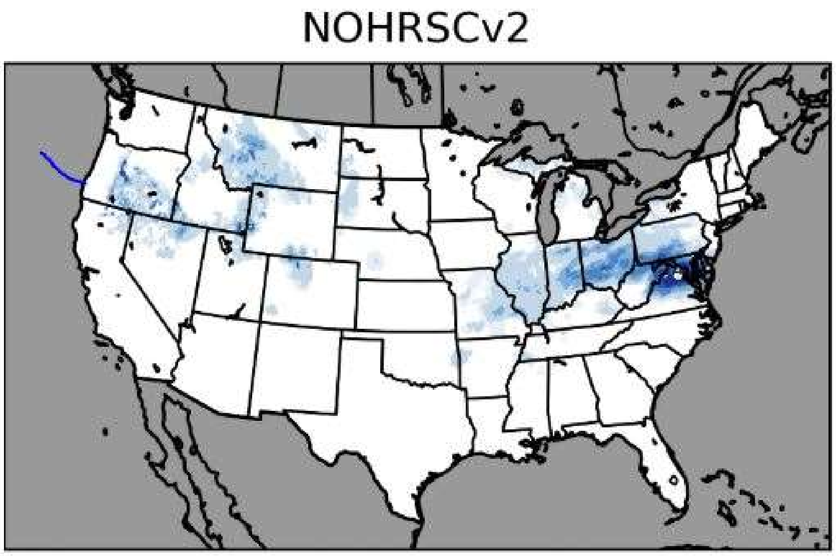


30-hour U-net Forecast

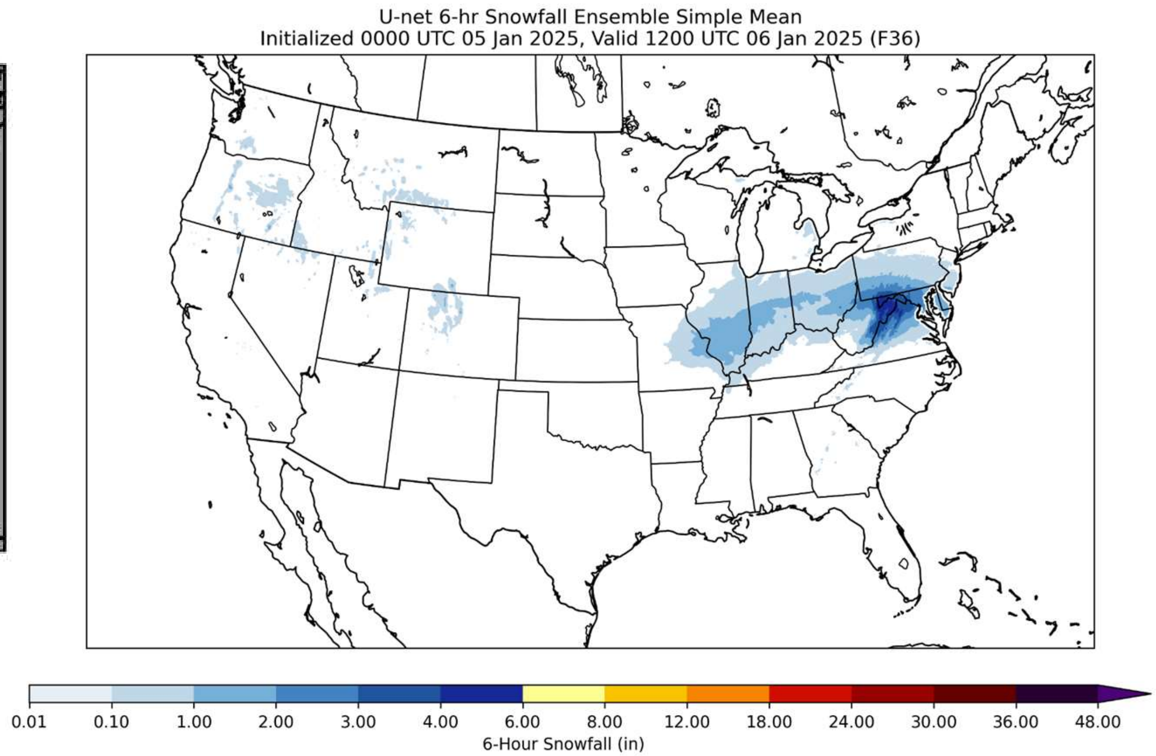


Sample Case (Machine Learning Products) – Jan 5-6 2025

6-h Snowfall 12z 06 Jan 2025



36-hour U-net Forecast



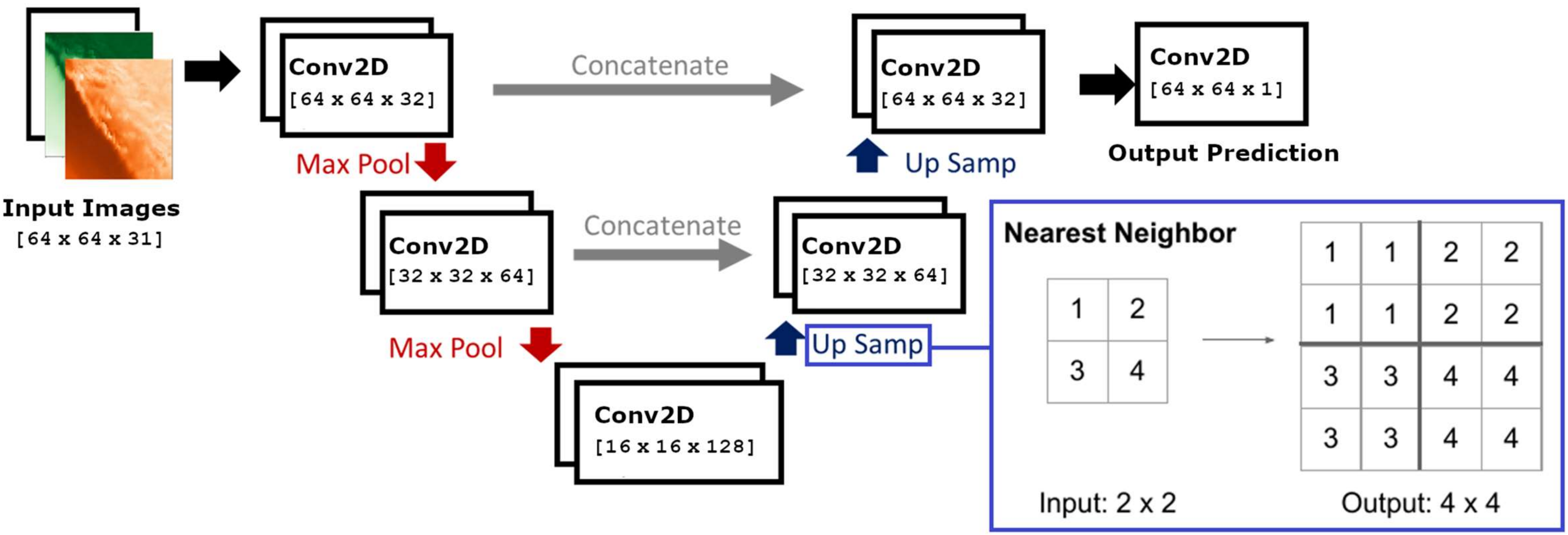
Machine Learning Component

- Performed in collaboration with NSF AI2ES Institute hosted at OU
- U-Net Convolutional Neural Network (Deep Learning)
- Builds upon earlier ML hail prediction for HWT (2017-2021) and ML rainfall prediction in HMT FFaIR
- Uses 8 HREF (4 each at 00, 12 UTC) and 4 CAPS FV3-LAM members.

ML Methods: U-Net Architecture

- CAPS FV3 Rainfall & Snowfall U-Nets use a collection of 2-D forecast images at different vertical levels as inputs for training.

- Patch size, number of connections, and number of layers are being evaluated as hyper-parameters (the exact details of the architecture shown below will likely change in later iterations).



ML Methods: Input Data (Training & Forecast Generation)

Current version of CAPS Snowfall U-Net uses 35 2-D NWP forecast variables relevant to snowfall prediction

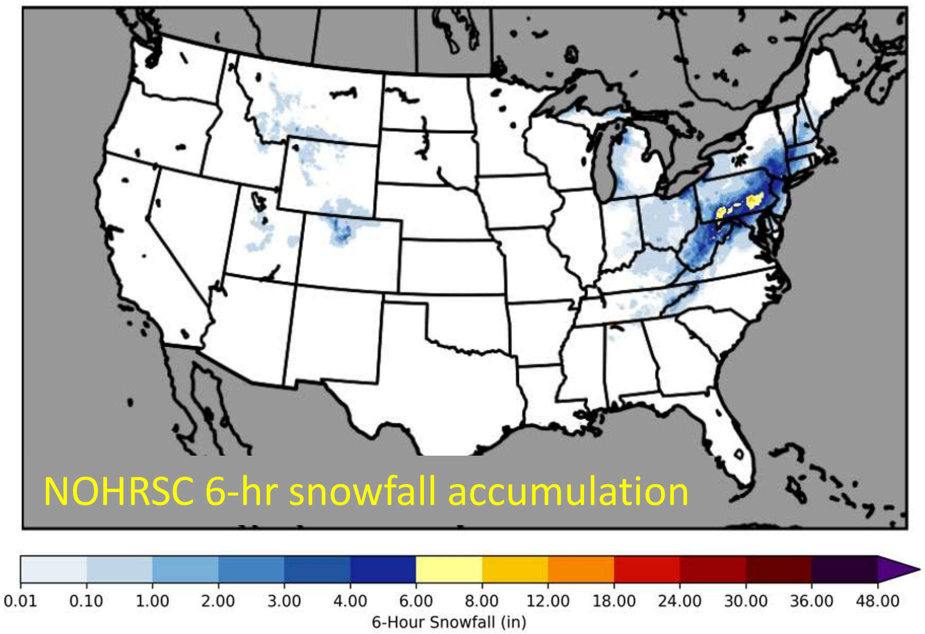
- Blue: Variable is used only for snowfall prediction (not for rainfall)
- Red: Variable is newly-added for 2023-2024 (not used in prior years)

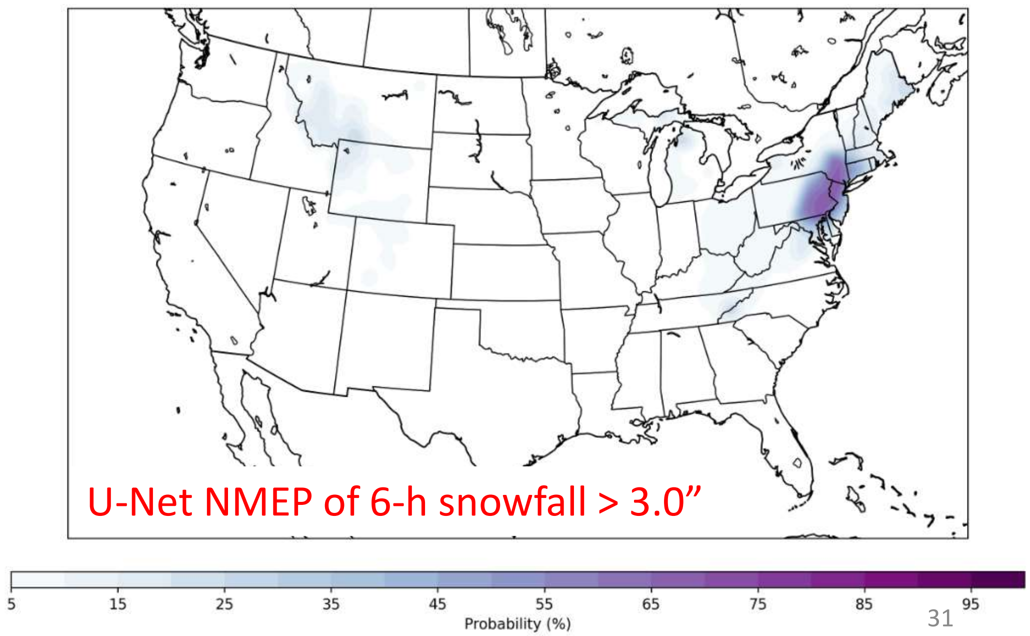
Variable	Level(s) Used (and/or other notes)
Geopotential height	500 hPa
Temperature	500, 700, 850, 925, 1000 hPa; 2 m AGL
Dewpoint	500, 700, 850, 925, 1000 hPa; 2 m AGL
u- and v- wind components	500 hPa; 10 m AGL
6-h maximum reflectivity	1 km AGL
Precipitable water	column-integrated
Hourly maximum updraft velocity	column maximum
6-h accumulated precipitation	
6-h accumulated snowfall	
Echo-top height	
Mean Sea Level Pressure	
Categorical SNOW, ICEP, FRZR, and RAIN	binary yes/no based on PTYPE at surface
Terrain Mean, Standard Deviation, Slope	Source: ASTER Global Digital Elevation Model
Vorticity	850, 500 hPa
Divergence	850, 500 hPa
Moisture Convergence	850 hPa; 10 m AGL
Land Use Classification	Classification source: WSSI Land Use Factor

ML Methods: Input Data (Training & Forecast Generation)

- Variables predicted: Probability of 6-h snowfall > 1, 2, and 3 inches, as well as ML ensemble simple mean ("ML best guess").
  - ML-predicted total snowfall (individual members and ensemble consensus) is being developed and evaluated internally, may be included in future year HMT WWE products.
- Observations (used for ML training and evaluation): NOHRSC snowfall analyses

NOHRSC observations and 24-h ensemble ML simple mean valid 0000 UTC on 20 Jan. 2025

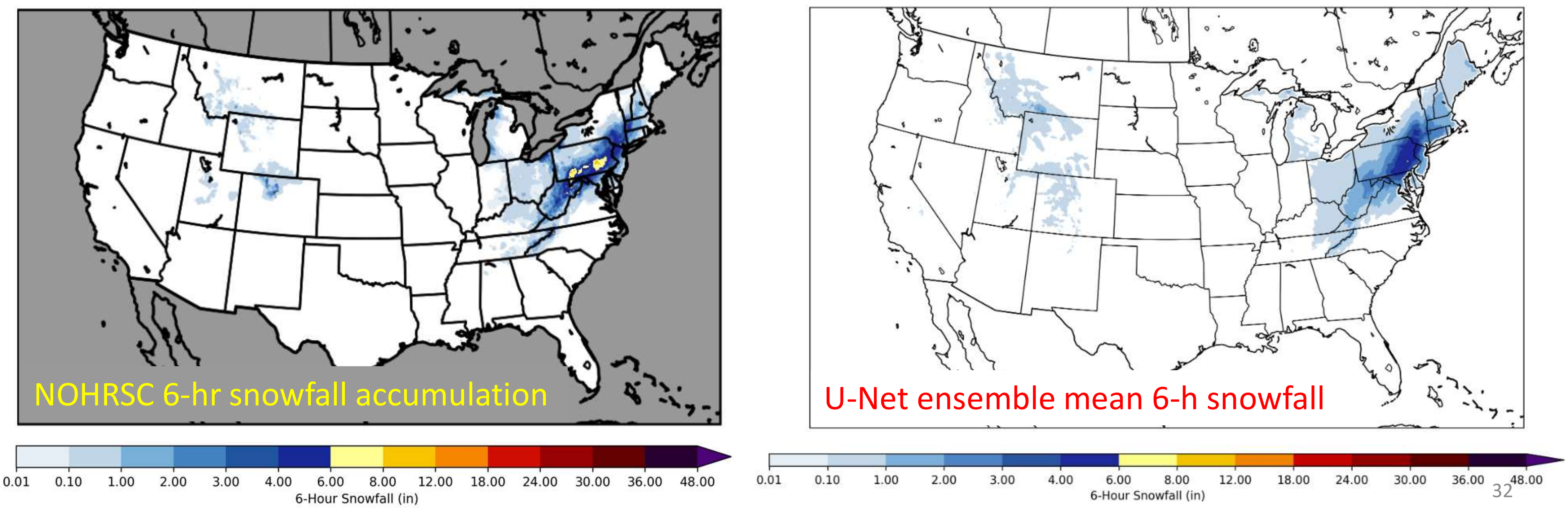




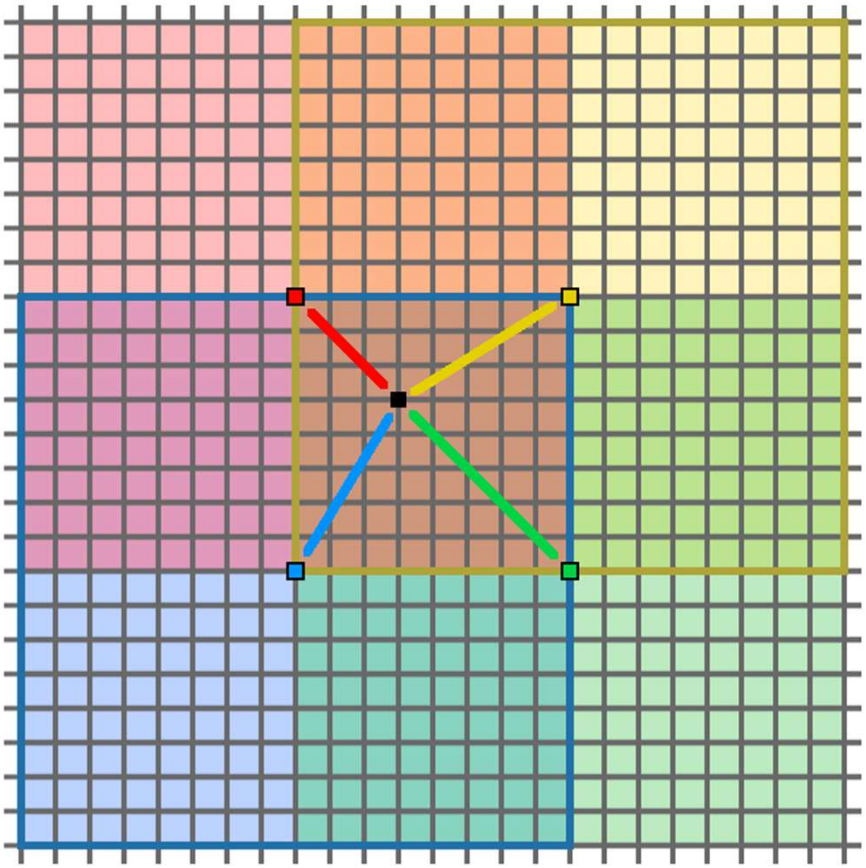
ML Methods: Input Data (Training & Forecast Generation)

- Variables predicted: Probability of 6-h snowfall > 1, 2, and 3 inches, as well as ML ensemble simple mean (“ML best guess”).
  - ML-predicted total snowfall (individual members and ensemble consensus) is being developed and evaluated internally, may be included in future year HMT WWE products.
- Observations (used for ML training and evaluation): NOHRSC snowfall analyses

NOHRSC observations and 24-h ensemble ML simple mean valid 0000 UTC on 20 Jan. 2025



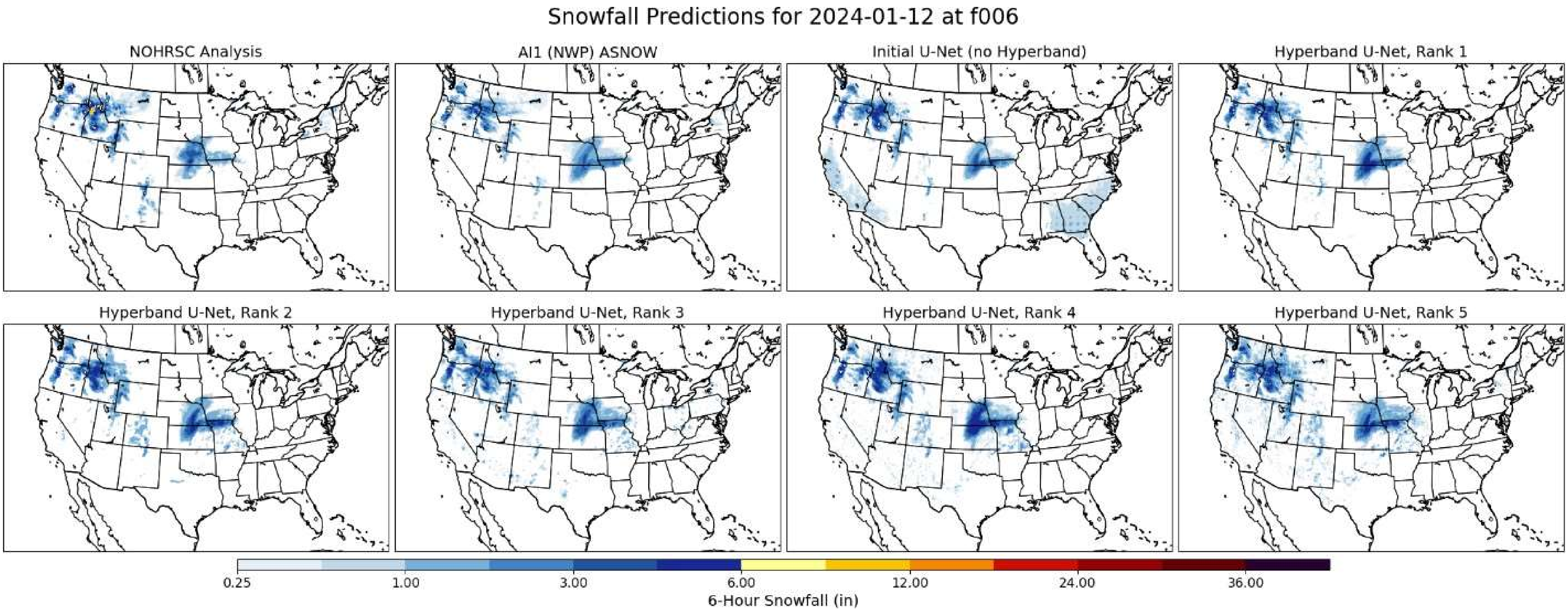
ML Methods: Patches, Training, and Forecast Generation



- Patch-wise U-Net predictions are generated using 64 x 64 overlapping grid square patches.
  - Patches are stitched together to form the full CONUS prediction
  - Weighted averaging of overlapping patches & applying light smoothing to the stitched forecast field minimizes discontinuities at patch boundaries
- Ensemble HREF+ probabilities are calculated from individual member predictions using a neighborhood maximum ensemble probability (NMEP) approach.
- A label offset (a modest, constant snowfall amount added to labels in regions of non-zero observed snowfall) is used.
  - Goal of label offset is to boost squared-error penalties and prevent the ML model from over-predicting regions of light snowfall.
  - The label offset is subtracted out from the final forecast products to prevent the introduction of a non-physical high bias.

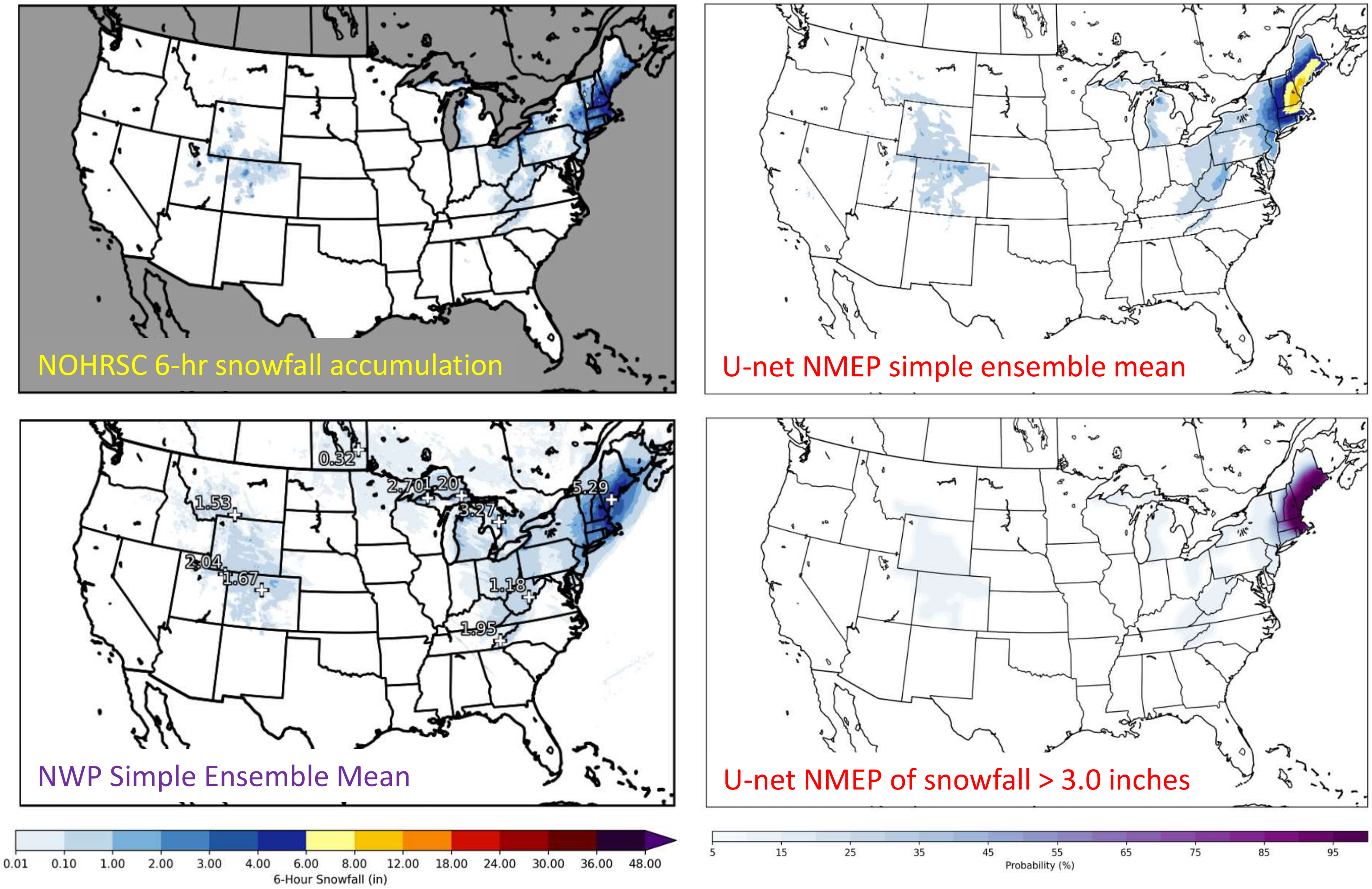
ML Methods: Hyperparameter Optimization

- Hyperband (Li et al. 2018) was used for ML hyperparameter optimization.
- Hyperparameters optimized include learning rate, depth of U-net, number of channels in hidden layers, and normalization approach.



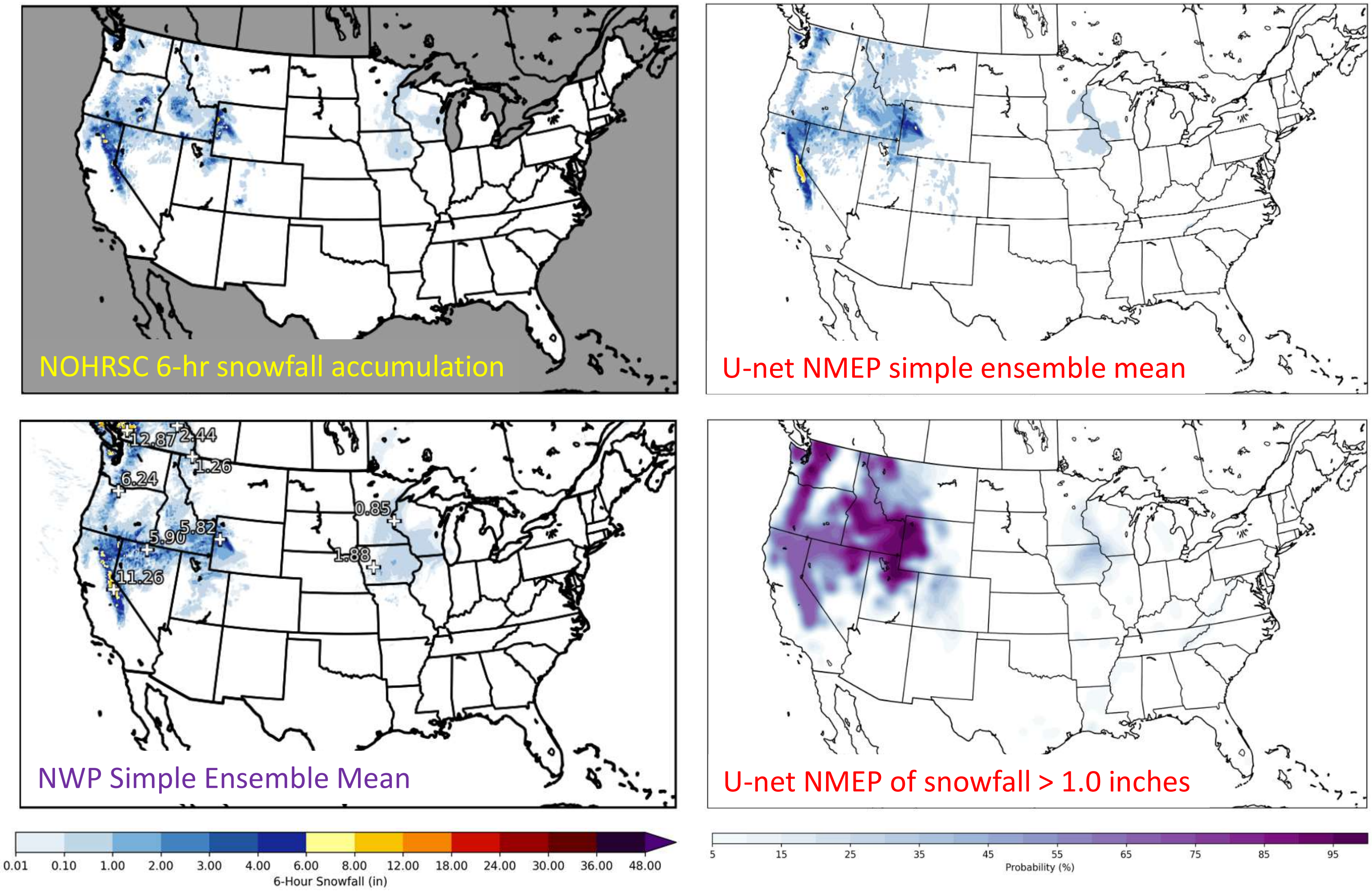
Reference: Li, L., K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, 2018: Hyperband: A novel bandit-based approach to hyperparameter optimization. *arXiv*, https://doi.org/10.48550/arXiv.1603.06560

ML Results: 30-h forecast valid 0600 UTC, 20 Jan. 2025



- 30-h U-net forecasts capture band of heavy snowfall over ME/NH/VT/MA with good timing/position accuracy
- ML simple mean often underforecasts—this is a rare instance where ML simple mean overforecasts snowfall.
- ML probability of snowfall > 3.0" performs quite well.
- ML simple mean does decent job with lighter snowfall in, e.g., MI, WV, CO, WY.

ML Results: 24-h forecast valid 0000 UTC, 15 Dec. 2024



- Both raw NWP and ML forecasts do a good job of capturing terrain-influenced snowfall in CA, CO.
- ML simple mean actually slightly better than NWP simple mean on peak snowfall amounts in heaviest bands!
- U-net NMEP forecasts produce overly-broad regions of high probability of snowfall exceeding 1.0", this is particularly notable over mountainous areas.
- Over-prediction of spatial coverage in U-net NMEP might be addressed by reducing/optimizing neighborhood radius – evaluation is ongoing.

Conclusions and Updates/Future Work

- All CAPS FV3-LAM ensemble members appear to accurately capture spatial patterns of precipitation/snowfall.
- No strong bias in precipitation forecasts; benefit of ensemble consensus most evident at longer lead-times.
- Forecast members using NSSL microphysics scheme (M1*) tend to under-forecast snowfall (low frequency bias) – snowfall ETS is also slightly lower for NSSL (M1*) members.
- Machine learning (ML) NMEP snowfall forecasts perform well, though NMEP sometimes results in spatial over-prediction.
- ML simple mean is performing quite well in many cases during 2024-2025 testing, in some cases outperforming CAPS FV3 NWP simple mean!
- Work is continuing during the 2024-2025 HMT WWE
  - Experimental MPAS ensemble is being tested
  - ML ensemble U-net continues to be optimized and evaluated—future version using MPAS is planned once sufficient training data have been collected.

Acknowledgements (Funding and Computing Resources)

Computing:

- NSF/Texas Advanced Supercomputing Center (TACC) *Frontera*

Funding:

- NOAA/OAR/OWAC Testbed Grants: NA19OAR4590141 & NA22OAR4590522
- UFS R2O Grant: NA16OAR4320115

Contacts: Nathan Snook (nsnook@ou.edu) Keith Brewster (kbrewster@ou.edu)