Impact of Time Lagging (TL) and Neighborhood Ensemble Probabilities (NEP) on Rapid Refresh Forecast System (RRFS) Ensemble Design

Gerard Ketefian[*1,2], J. Beck[1], W. Mayfield[4], M. Harrold[4],
B. Nelson[4], C. Kalb[4], C. Schwartz[5], J. K. Wolff[5], E. Kalina[6]
[1]NOAA/GSL/DTC, [2]CU/CIRES, [4]NCAR/DTC, [5]NCAR/RAL, [6]SciTech
*Gerard.Ketefian@noaa.gov

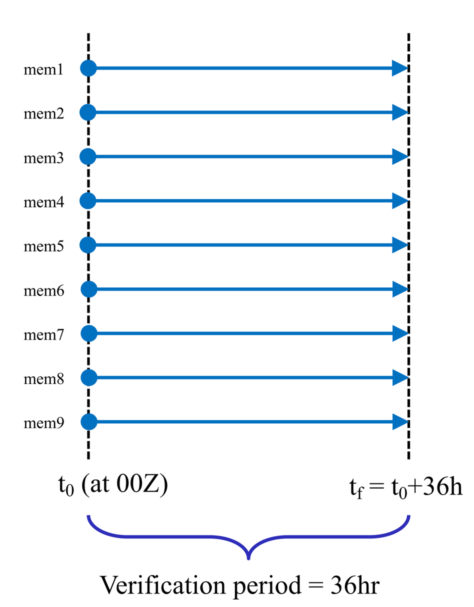
Background and Motivation

- NOAA is transitioning to the community-based **Unified Forecast System (UFS)** with the **Finite Volume on a Cubed-Sphere (FV3)** dynamical core for both global and regional applications.
- Regional systems in operations, both deterministic (NAM, RAP, HRRR) and ensemble (HREF) will be replaced by the UFS's FV3-based storm-scale ensemble system known as the **Rapid Refresh Forecast System (RRFS)**.
- HREF is a **multi-dycore** and **multi-physics** ensemble ⇒ members provide sufficient spread.
- RRFS is a **single-dycore** and (eventually) **single-physics** ensemble ⇒ may be **under-dispersive** (insufficient spread) if ensemble is not properly designed.
- The DTC's *Optimizing Ensemble Design for Use in the RRFS* project aims to improve the RRFS by exploring the following techniques to improve its ensemble spread and verification results:
  - **Initial condition perturbations** – Generate ensemble members by adding perturbations to the initial conditions.
  - **Stochastic physics perturbations** – Generate ensemble members by perturbing parameters in physics schemes.
  - **Time-lagging** – Generate ensemble members by combining forecasts of different lead times.
  - **Neighborhood ensemble probabilities** – Improve verification (vx) metrics by "relaxing the traditional requirement that forecast and observed events match at the grid scale" (Schwartz & Sobash, *MWR* (2017)).
- In this work, we:
  - Consider use of **time-lagging (TL)** and **neighborhood ensemble probabilities (NEP)** in the RRFS ensemble (but not initial condition and stochastic physics perturbations, which are still being evaluated).
  - Evaluate ability of TL and NEP to improve the spread-to-error ratio (aka spread-to-skill ratio), i.e., bring it closer to 1, without significantly increasing error/decreasing skill.
  - Verification metrics and diagrams considered include:
    - Spread-to-error (aka spread-to-skill) ratio vs. lead
    - Bias vs. lead (both ensemble mean and individual member biases are considered)
    - Brier score vs. lead
    - Reliability diagrams
    - ROC (Receiver Operating Characteristic) and AUC (Area Under the Curve) diagrams
    - Rank histograms

Comparison of Single Initialization (SI) and Time-Lagged (TL) Ensembles

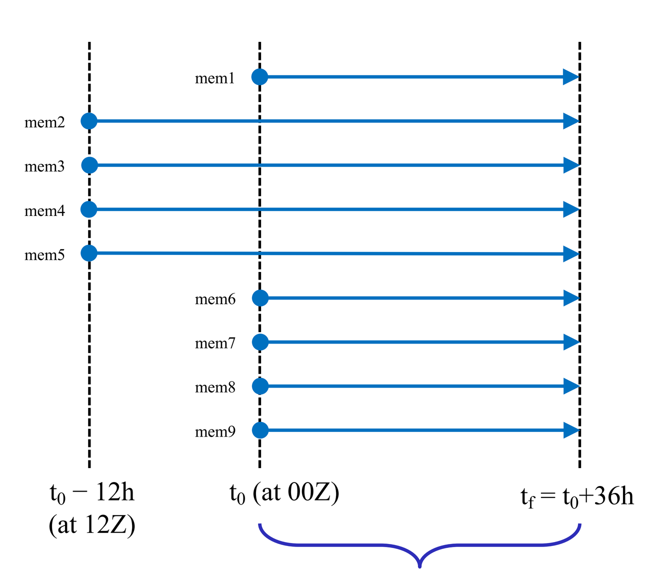
Single-Initialization (SI) Ensemble Experiments:

- **9 members**, all on the RRFS CONUS 3km grid.
- For forecasts, use RRFS output from the 2021 Hazardous Weather Testbed (HWT) Spring Forecast Experiment (SFE).
- Total of **21 initializations** between May 4th and June 4th, 2021:
  - 2021-05-[04-06, 08-09, 19-30]
  - 2021-06-[01-04]
- All forecasts start at **00Z**.
- All forecasts are **36hr** long.
- To perform verification, use the **UFS Short-Range Weather (SRW) App**, which in turn calls the DTC's **Model Evaluation Tools (MET)** verification package.

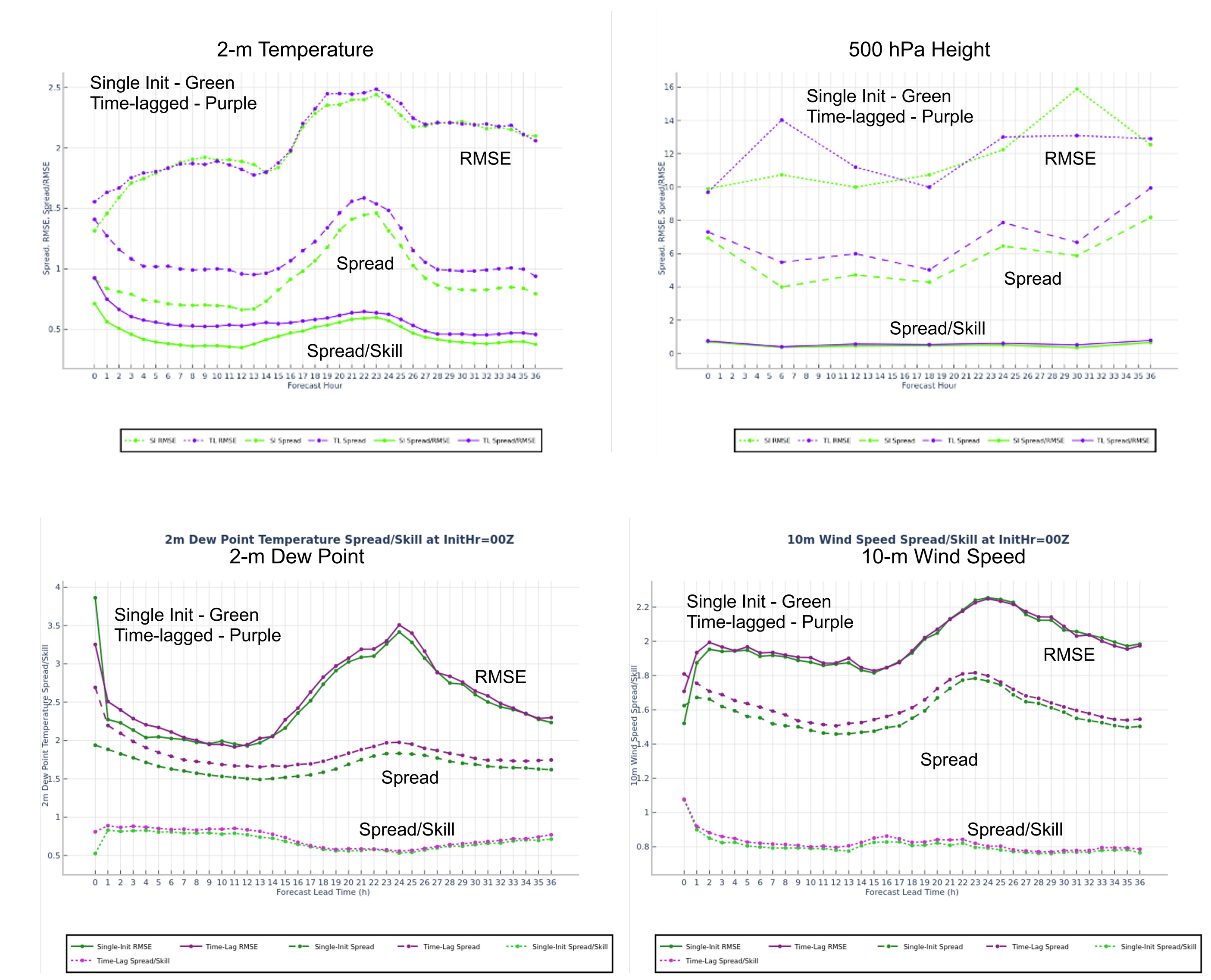


Time-Lagged (TL) Ensemble Experiments:

- **Time-lagging** combines forecasts of different lead times to create an ensemble with more members and **potentially more spread**.
- Replace 4 members (#2, #3, #4, and #5) of the single-initialization experiment with **48hr long** RRFS forecasts that start on **12Z** of the previous day. These are the time-lagged members.
- For these TL members, **run verification on only final 36hr** of forecast (i.e., drop the first 12 hours).
- Remaining 5 members (#1, #6, #7, #8, and #9) identical to their counterparts in the single-initialization ensemble (i.e., 00Z initialization, 36hr long).



Vx Results: Spread/Skill for SI vs. TL



- Increases in spread are generally statistically significant
- On average, no increase in RMSE when adding time-lagging

Comparison of Verification Using Ensemble Probabilities (EP; traditional approach) and Neighborhood Ensemble Probabilities (NEP)

- **Definition of Binary Probability (BP)**:
  Assume we have $f_{i,j}$ forecasts for $i = 1, \dots, M$ grid points and $j = 1, \dots, N$ ensemble members, and let q denote an event threshold, e.g., $q = 1.0$ mm h^{-1}. Then the binary probability BP of event occurrence at the ith grid point for the jth ensemble member is:

$$\mathrm{BP}_{i,j}(q) = \begin{cases} 1 & \text{if } f_{i,j} \geq 0 \\ 0 & \text{if } f_{i,j} < q \end{cases}$$

- **Definition of Ensemble Probability (EP)** (traditional metric without a neighborhood):
  Ensemble probability of event occurrence at the ith grid point is

$$\mathrm{EP}_i(q) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{BP}_{i,j}$$

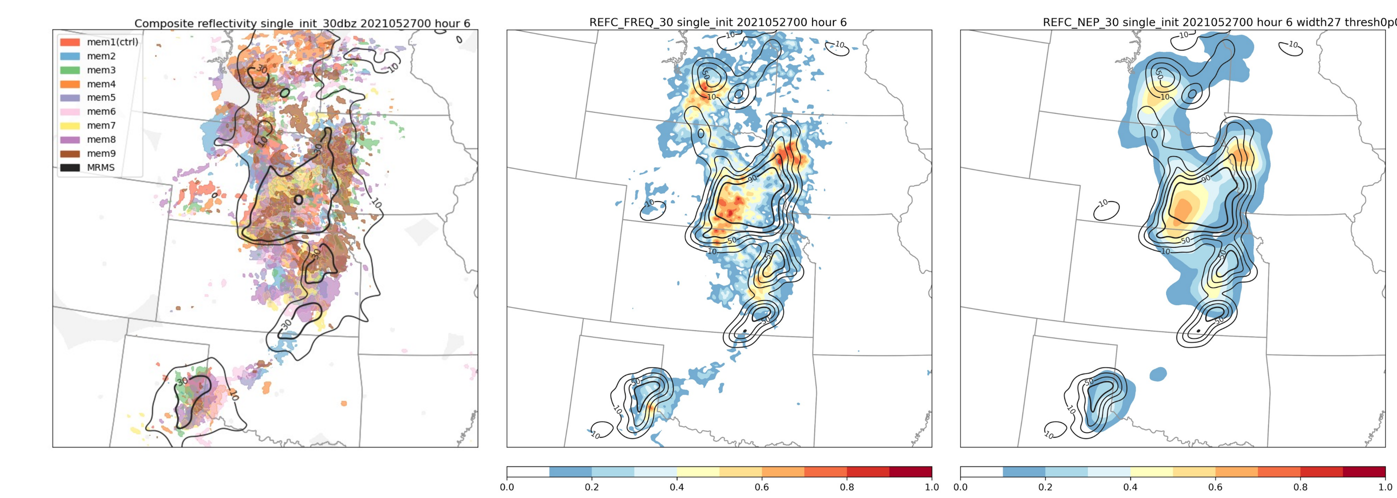
- **Definition of Neighborhood Ensemble Probability (NEP)**:
  Let S_i denote a unique set of N_b points within the neighborhood of the ith grid point. Then the neighborhood ensemble probability of event occurrence at the ith grid point is

$$\mathrm{NEP}_i(q) = \frac{1}{N_b}\sum_{k=1}^{N_b} \mathrm{EP}_k$$

- Notes:
  - Calculating $\mathrm{NEP}_i(q)$ from $\mathrm{EP}_i(q)$ is a **smoothing** operation (over the neighborhood S_i).
  - Neighborhood can be a square or a circle. In this study, we chose **circle with r = 40 km radius**, same as at Storm Prediction Center (SPC).

Effect of NEP (Smoothing) on Probabilities

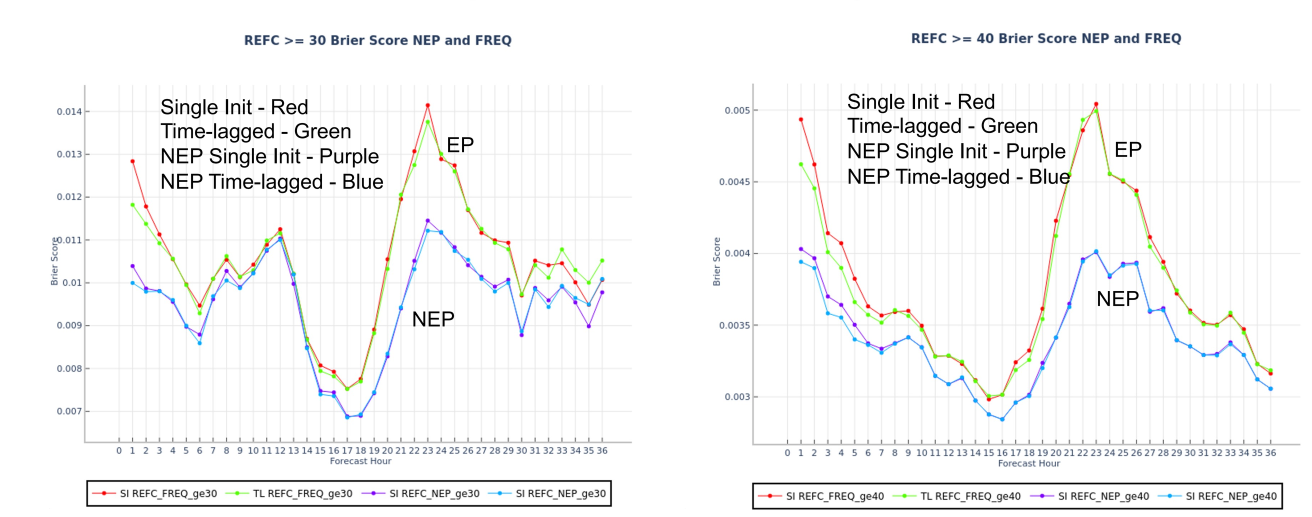
- RRFS 9-member forecast on 20210527 at 00Z; forecast hour 6
- Comparison of BP, EP, and NEP fields for reflectivity ≥ 20 dBZ.



Member Binary Probabilities (BPs; all 0 or 1) | Ensemble Probabilities (EPs) | Neighborhood Ensemble Probabilities (NEPs)

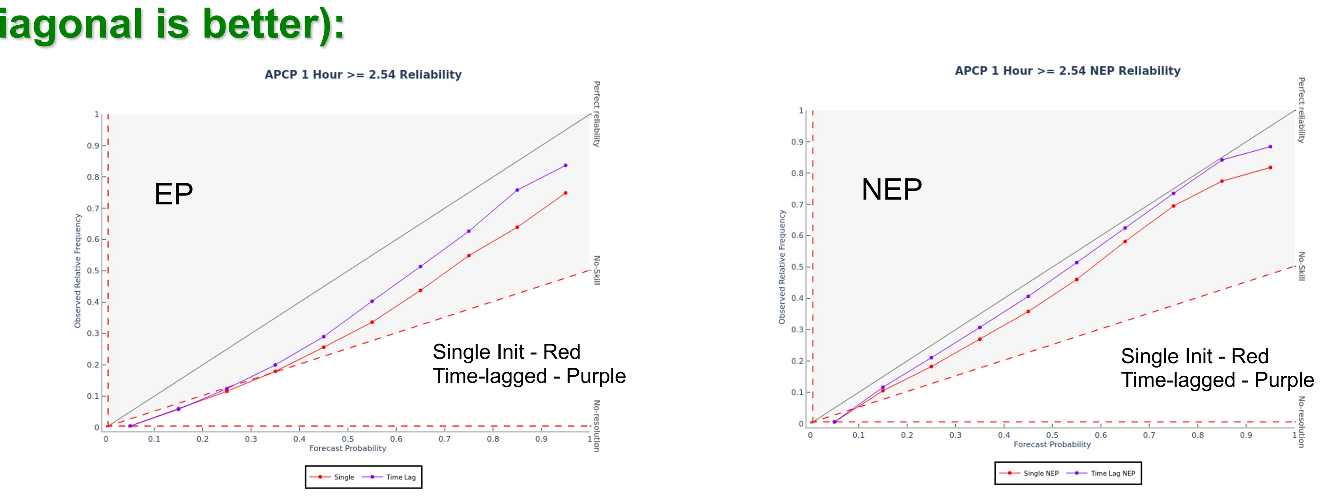
Vx Results: EP (traditional method) vs. NEP

Brier Score for Reflectivity >30 and >40 dBZ (lower values are better):



- Minor improvement when using time-lagging
- Larger improvement when using NEP, particularly for the convective period of the day

Reliability for APCP01 (1-hour accumulated precip) > 2.54 cm (closer to diagonal is better):



- NEP (right) improves precipitation reliability over EP (left) for all thresholds
- NEP improves forecasts compared to EP across all probabilities

Conclusions

- Time-lagging (TL) and NEP are cost-efficient and effective ensemble design techniques that can generally improve spread and probabilistic RRFS forecasts while introducing little to no error.
- Overall, both techniques provide improvements to the RRFS's ensemble vx metrics:
  - TL doesn't appear to modify the RRFS ensemble climatology, nor does it introduce large individual member or ensemble mean bias.
  - Use of NEPs provides significant improvements over traditional EPs in vx metrics of non-continuous fields.
- While inclusion of 12-hour time-lagged ensemble members resulted in improvements to RRFS vx metrics, inclusion of shorter-term time-lagged members (e.g., 1-, 3-, or 6-hour) will likely also provide benefits, although potentially smaller in magnitude.

Email: Gerard.Ketefian@noaa.gov