Assessing ensemble forecast performance for select members available in the Community Leveraged Unified Ensemble (CLUE) during 2017 Hazardous Weather Testbed Spring Forecast Experiment (HWT-SFE)

Lindsay Blank[1]

Jamie Wolff[1], Michelle Harrold[1], and Jeff Beck[2]

[1]National Center for Atmospheric Research (NCAR) and Developmental Testbed Center (DTC)

[2]Cooperative Institute for Research in the Atmosphere (CIRA)/Affiliated with NOAA/ESRL/GSD and Developmental Testbed Center (DTC)

lblank@ucar.edu

2018 Joint WRF/MPAS User's Workshop
Boulder, CO 14 June 2018

Overview

- Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE) collaborations
  - Excellent opportunity to evaluate experimental datasets and identify weaknesses that remain in NWP models.
- Community Leveraged Unified Ensemble (CLUE) dataset
  - Carefully coordinated set of 60+ members, contributed by multiple organizations.
  - Subjectively assessed during experiment; provide a great dataset for retrospective evaluation.
- Model Evaluation Tools (MET) verification software
  - Deterministic and probabilistic verification metrics

Help investigate significant scientific questions that remain regarding the best approach to constructing a convection-permitting ensemble system

Experiment Design

- **CLUE 2017 dataset:** Three 10-member subset ensembles
  - Multi-physics with IC/BC perturbations
  - Single physics with IC/BC perturbations
  - Stochastic physics with IC/BC perturbations
- **Test Period:** May 1 – June 2 2017
  - Single physics: May 10 – June 2
  - Stochastic physics: May 10 – May 24
  - 00 UTC initializations; 36 hour forecasts
- **Model Domain:** 3-km grid spacing, CONUS
- **Observations:** Gridded Multi-Radar/Multi-Sensor (MRMS) composite reflectivity and precipitation accumulation analyses

Multi-physics + IC/BC pert (10 members)			
Member	**MP**	**LSM**	**PBL**
core01	Thompson	Noah	MYJ
core02 (ctrl)	Thompson	RUC	MYNN
core03	P3	Noah	YSU
core04	MY	Noah	MYNN
core05	Morrison	Noah	MYJ
core06	P3	Noah	YSU
core07	MY	Noah	MYNN
core08	Morrison	Noah	YSU
core09	P3	Noah	MYJ
core10	Thompson	Noah	MYNN

Single physics + IC/BC pert (10 members)			
Member	**MP**	**LSM**	**PBL**
single-phys01	Thompson	RUC	MYNN

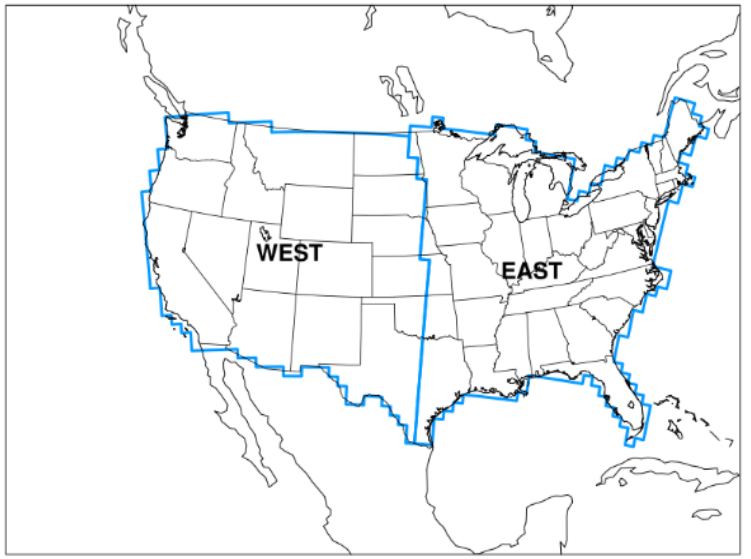
Stochastic physics + IC/BC pert (10 members)			
Member	**MP**	**LSM**	**PBL***
stoch-phys01	Thompson	RUC	MYNN

CLUE Subset Comparison: multi- vs. single vs. stochastic physics

Is there an advantage to using an ensemble with stochastic perturbations compared to using multiple microphysics/PBL parameterizations? How do these two ensemble subsets compare to an ensemble using one common physics suite?

Focus for this presentation:

- **Field:** Composite reflectivity (REFC) $\geq$ 30 dBZ
- **Verification domain:** EAST

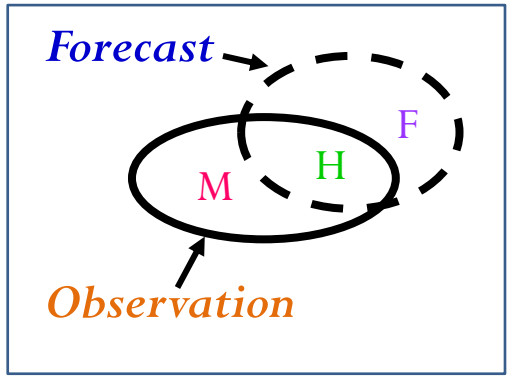


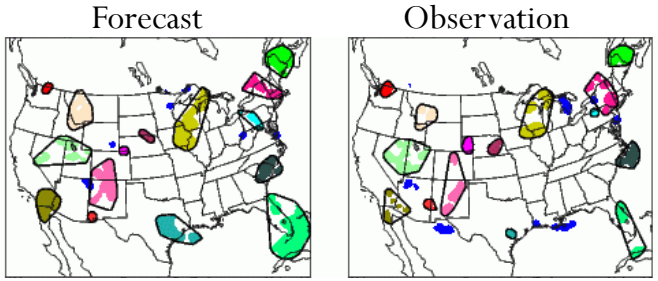
Verification sub-regions

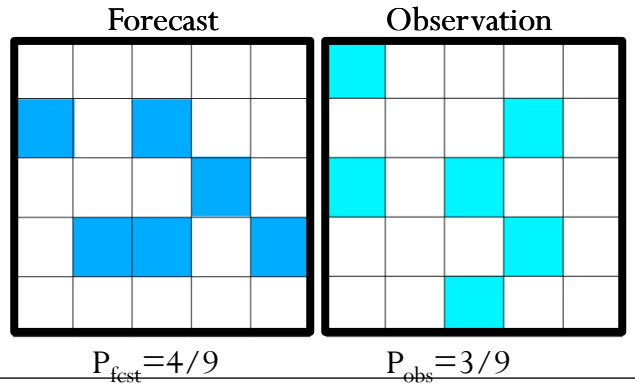
Model Verification

Deterministic

- **Traditional Verification**
  - **Gilbert Skill Score (GSS)**: Fraction of observed events that were correctly predicted; adjusted for random hits.
  - **Frequency Bias**: Ratio of the frequency of forecast events to observed events
- **Spatial Verification**
  - **Method for Object-based Diagnostic Evaluation (MODE)**: *Identify*, *merge* and *match* objects in forecast and observed fields
  - **Fractions Skill Score (FSS)**: Obtain a measure of how forecast skill varies with spatial scale







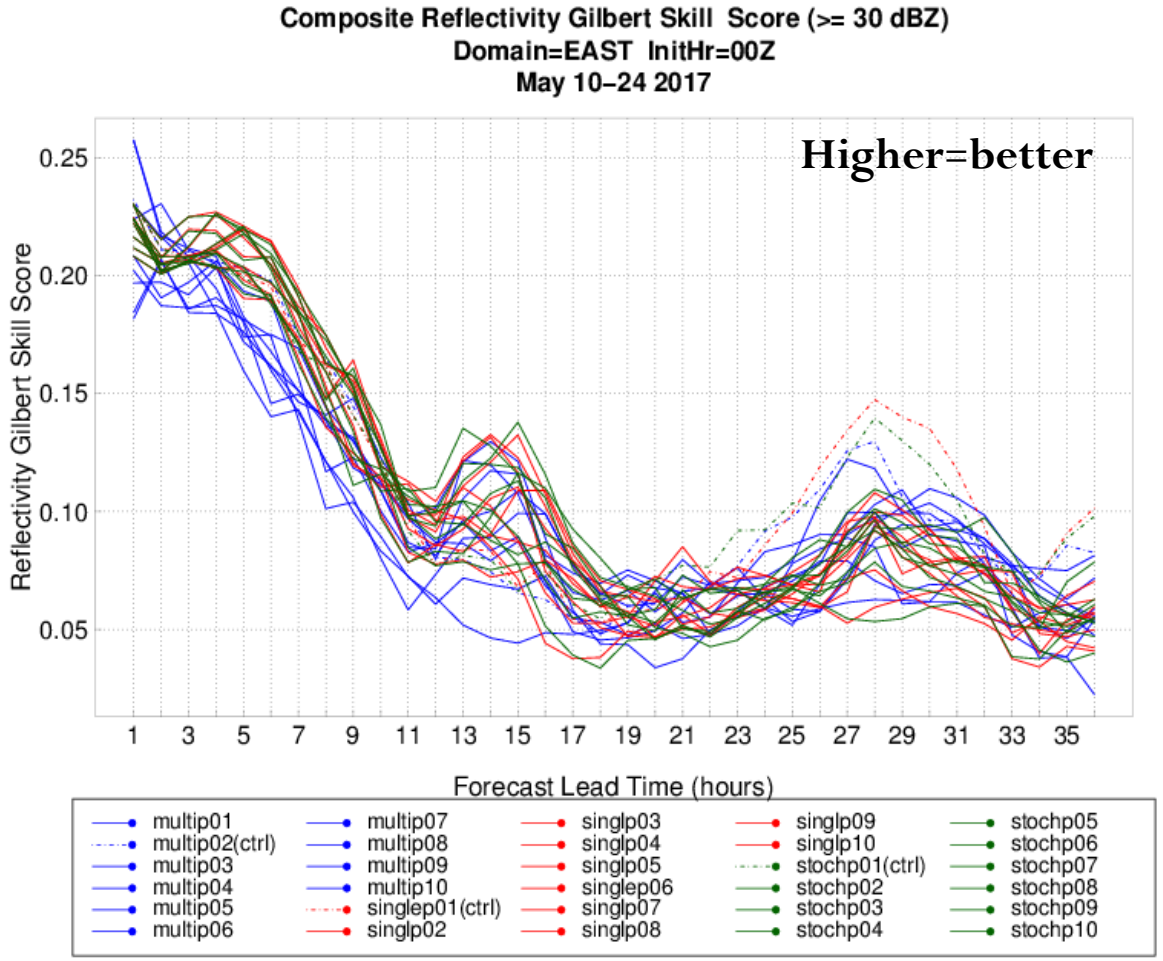
Model Verification Techniques

Probabilistic

- **Ensemble Verification**
  - **Spread:** Standard deviation of the ensemble member forecast values from the ensemble mean.
  - **Brier Score:** Magnitude of the probability forecast errors.
  - **Reliability diagram:** Observed frequency of events vs. forecast probability of those events (conditioned on forecasts).
  - **Receiver Operating Characteristic (ROC) curve:** Forecast discrimination power between events and non-events (measures resolution and is conditioned on observations).
  - **Rank Histogram:** Rank of the observation compared to all members of the ensemble forecast.

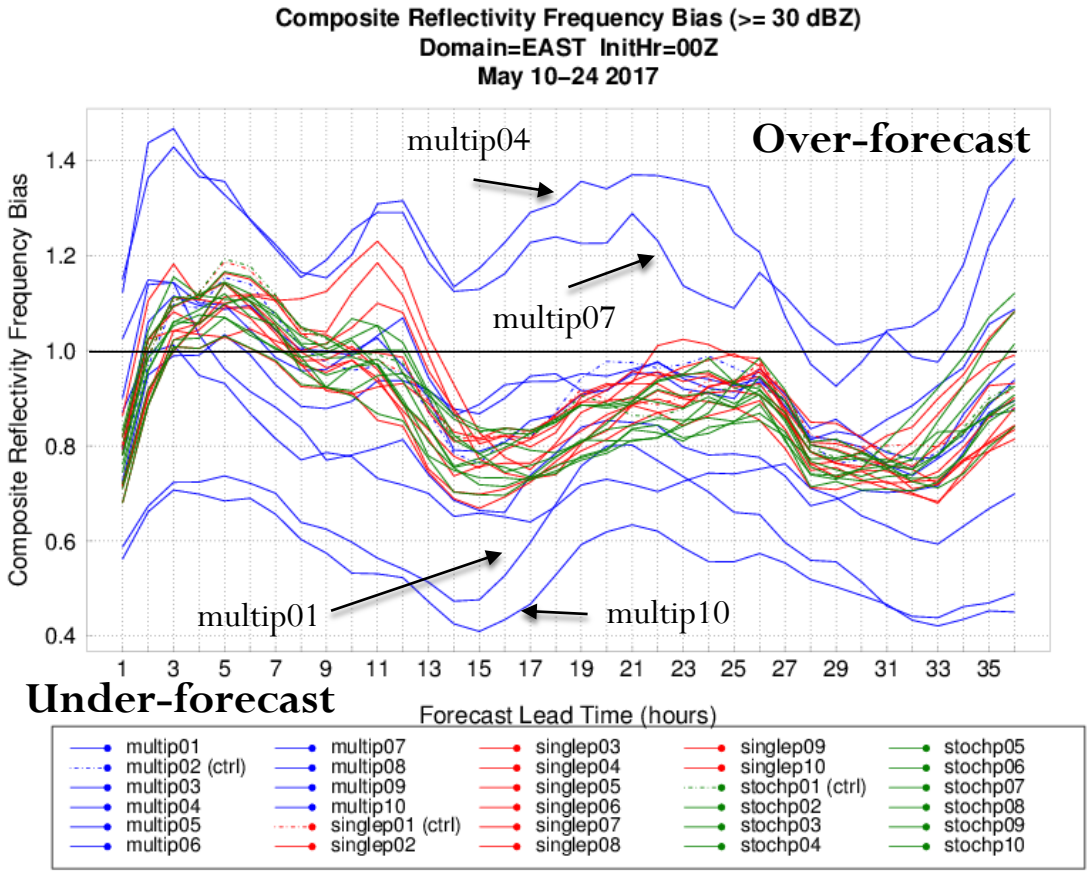
Composite Reflectivity ≥ 30 dBZ

Composite Reflectivity (≥ 30dBZ) GSS



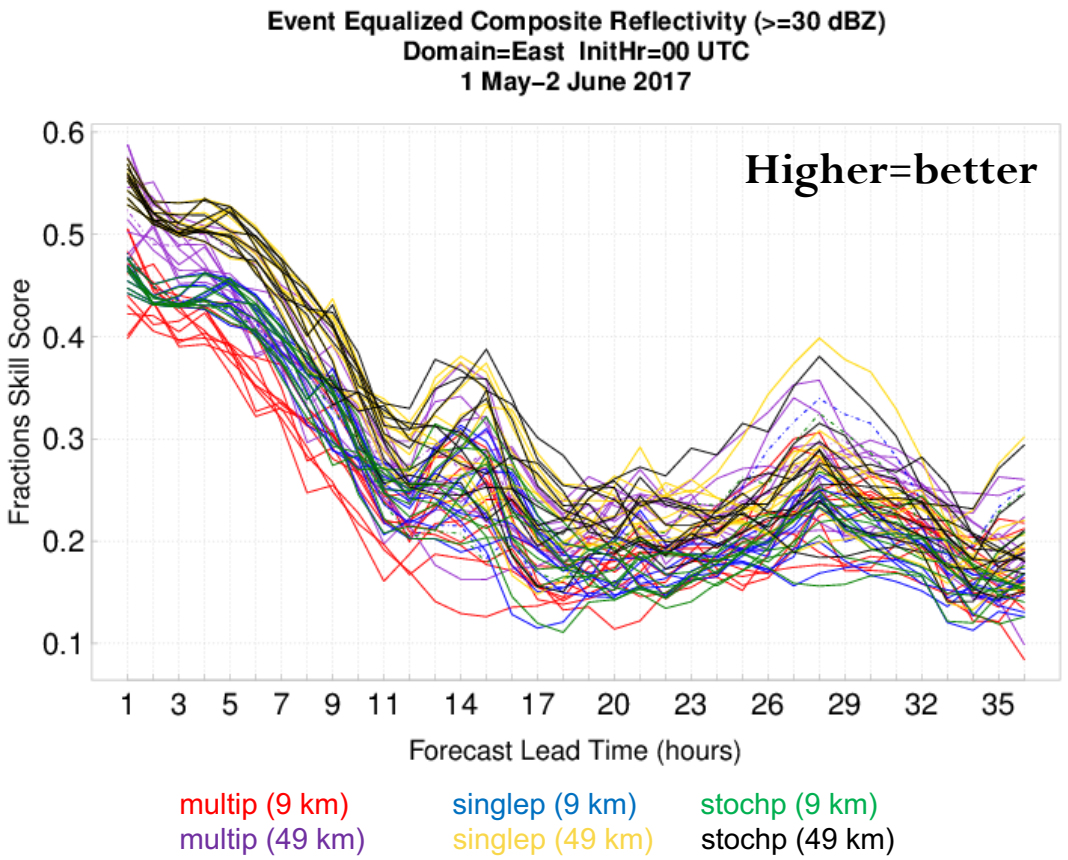
- Similar temporal trends across all subsets.
- **Multi-physics** displays lowest GSS values for approximately the first half of the forecast.
- **Single physics** and **stochastic physics** perform similarly across the forecast period.

Composite Reflectivity (≥ 30dBZ) Frequency Bias



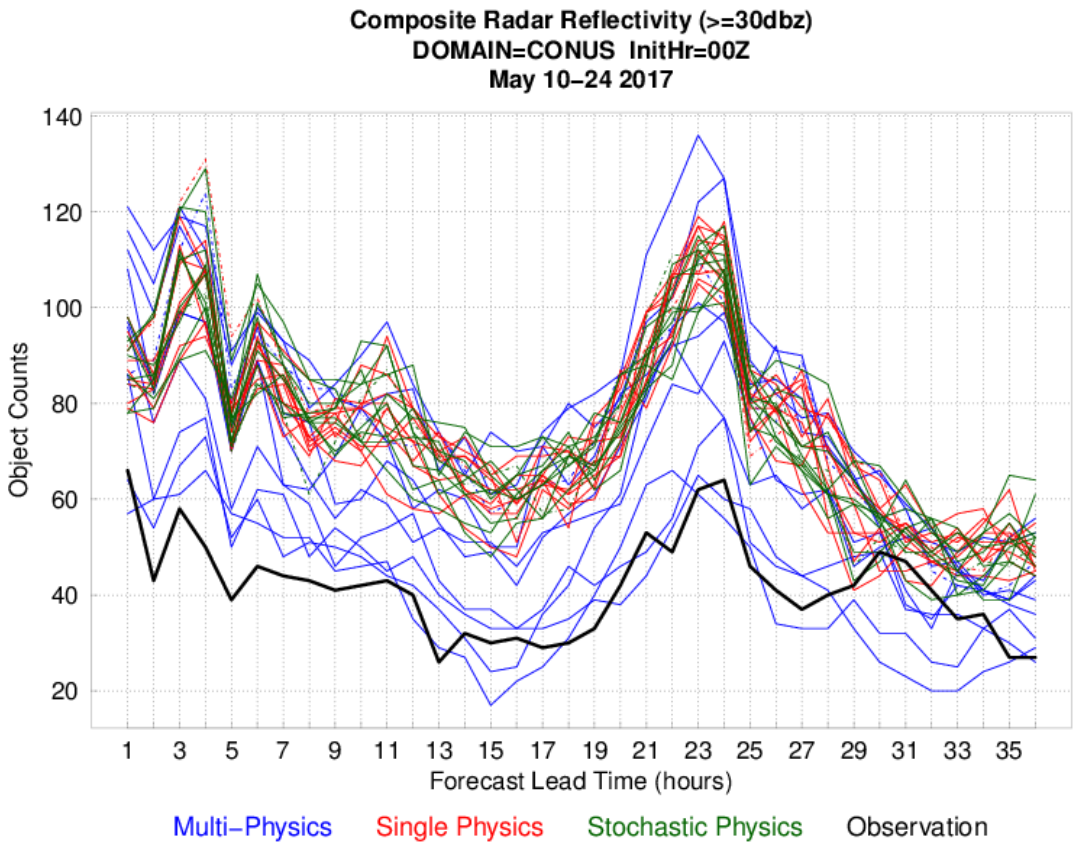
- Slight diurnal signal apparent in all subsets.
- Slight over-forecast after initial spin-up, then switching to under-forecast.
- **Multi-physics** ensemble has very large variability, both over- and under-forecasting throughout the forecast period.
- **Single physics** and **stochastic physics** perform similarly to one another.

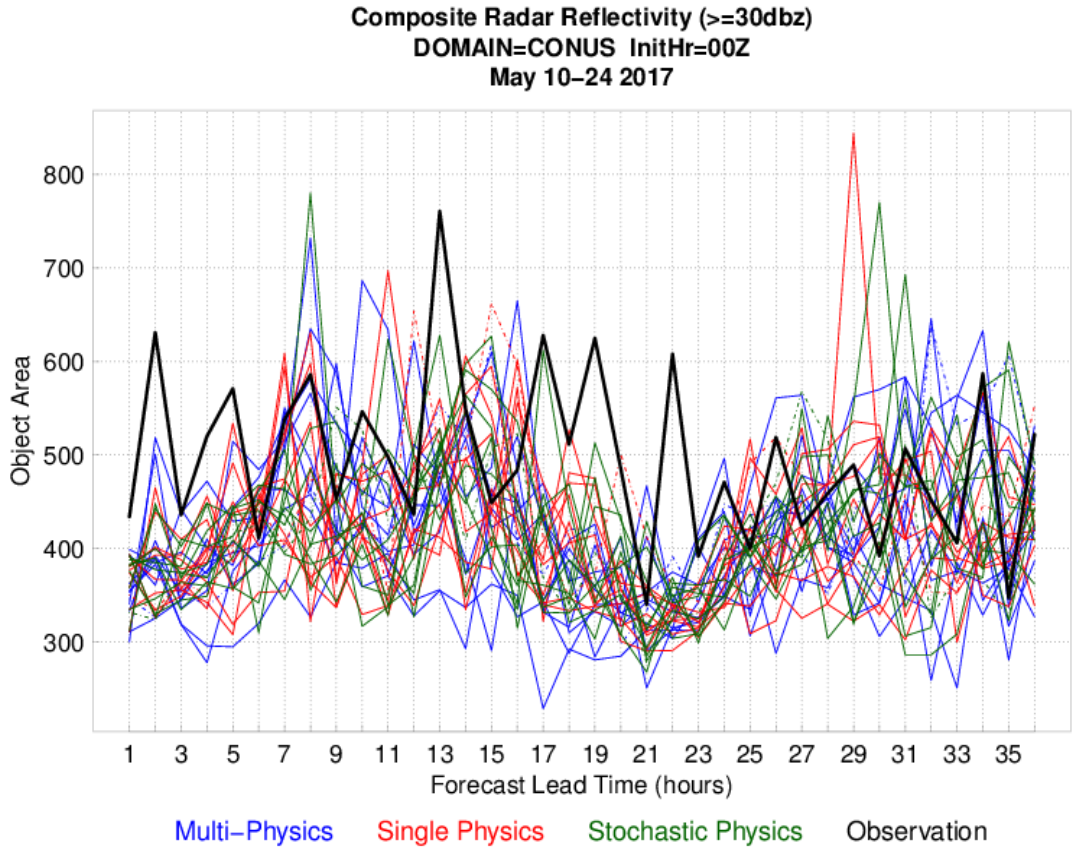
Composite Reflectivity (≥ 30dBZ) FSS



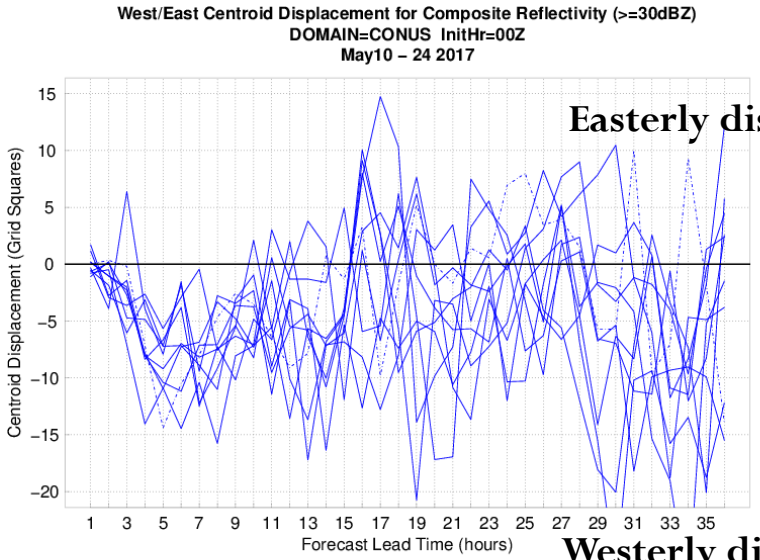
- Temporal trend is similar across the three subsets
- **Multi-physics** tends to perform the worst in the first half of the forecast.
- Skill is higher with increased neighborhood

Composite Reflectivity (≥ 30dBZ) MODE object counts and areas

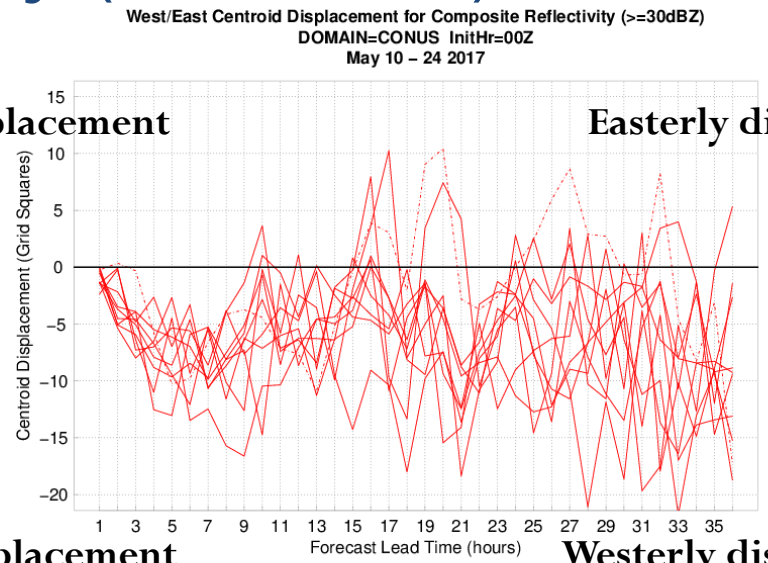




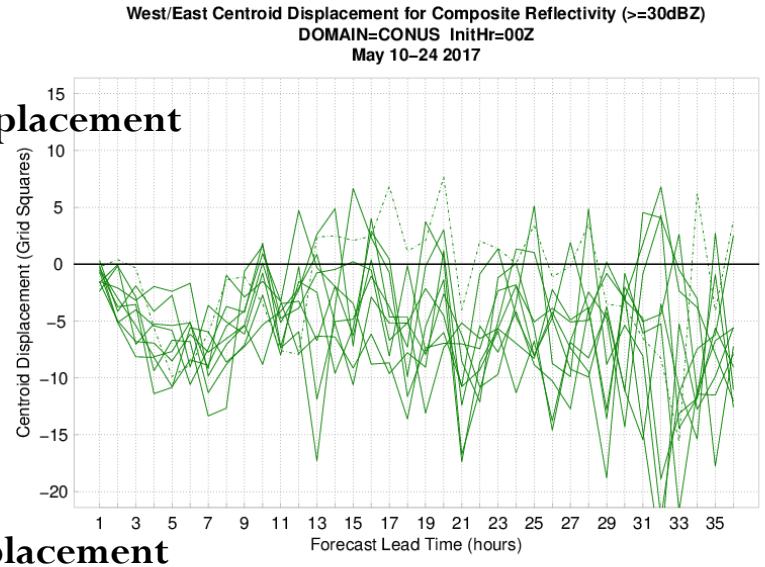
Composite Reflectivity (≥ 30dBZ) MODE object displacement



Easterly displacement



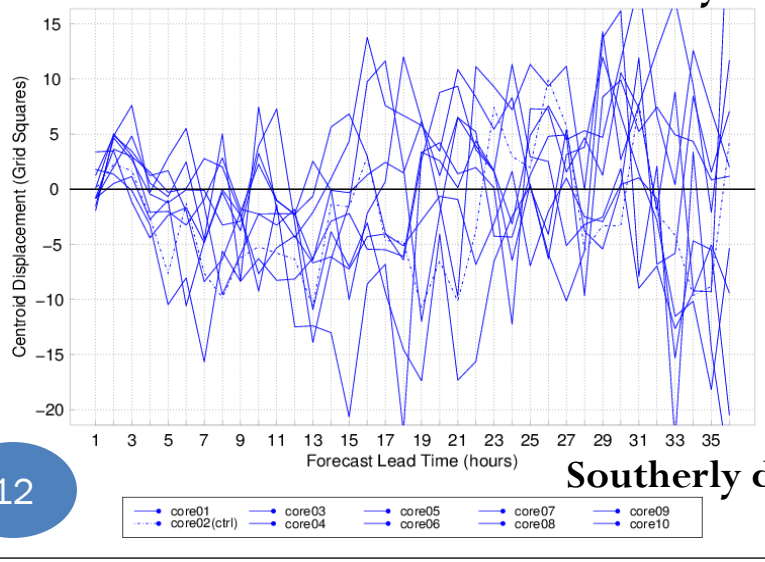
Easterly displacement



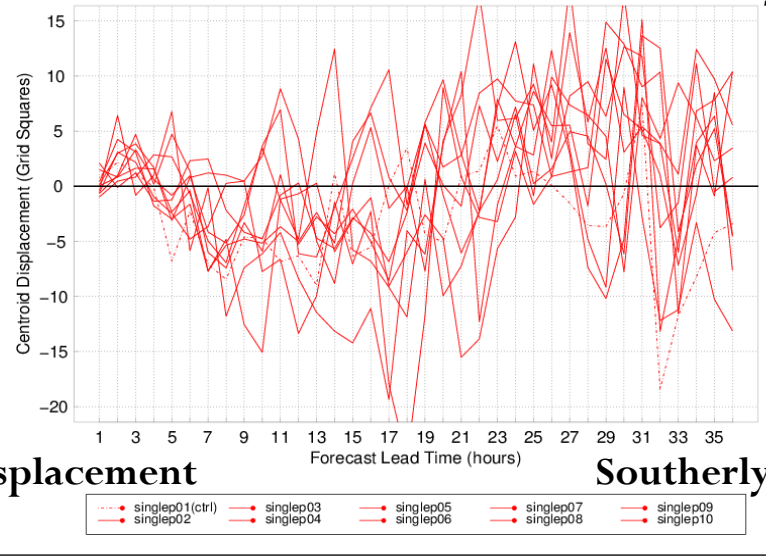
Westerly displacement

Westerly displacement

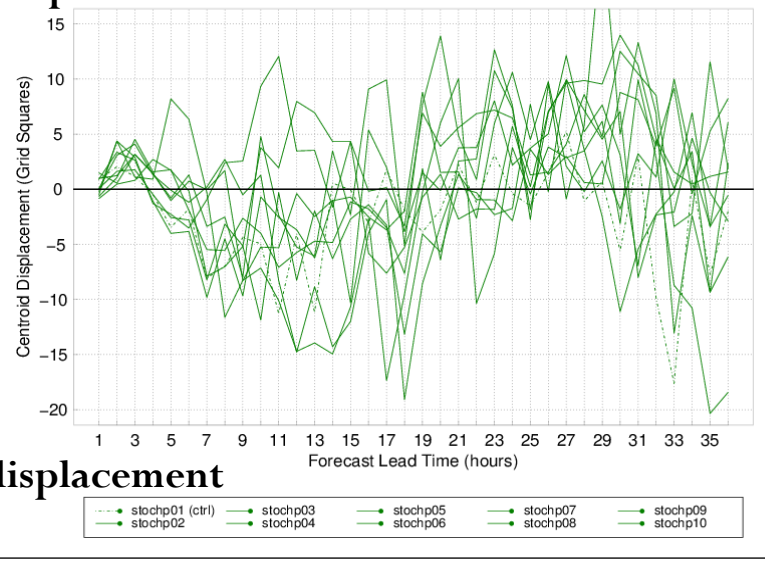
Northerly displacement



Northerly displacement



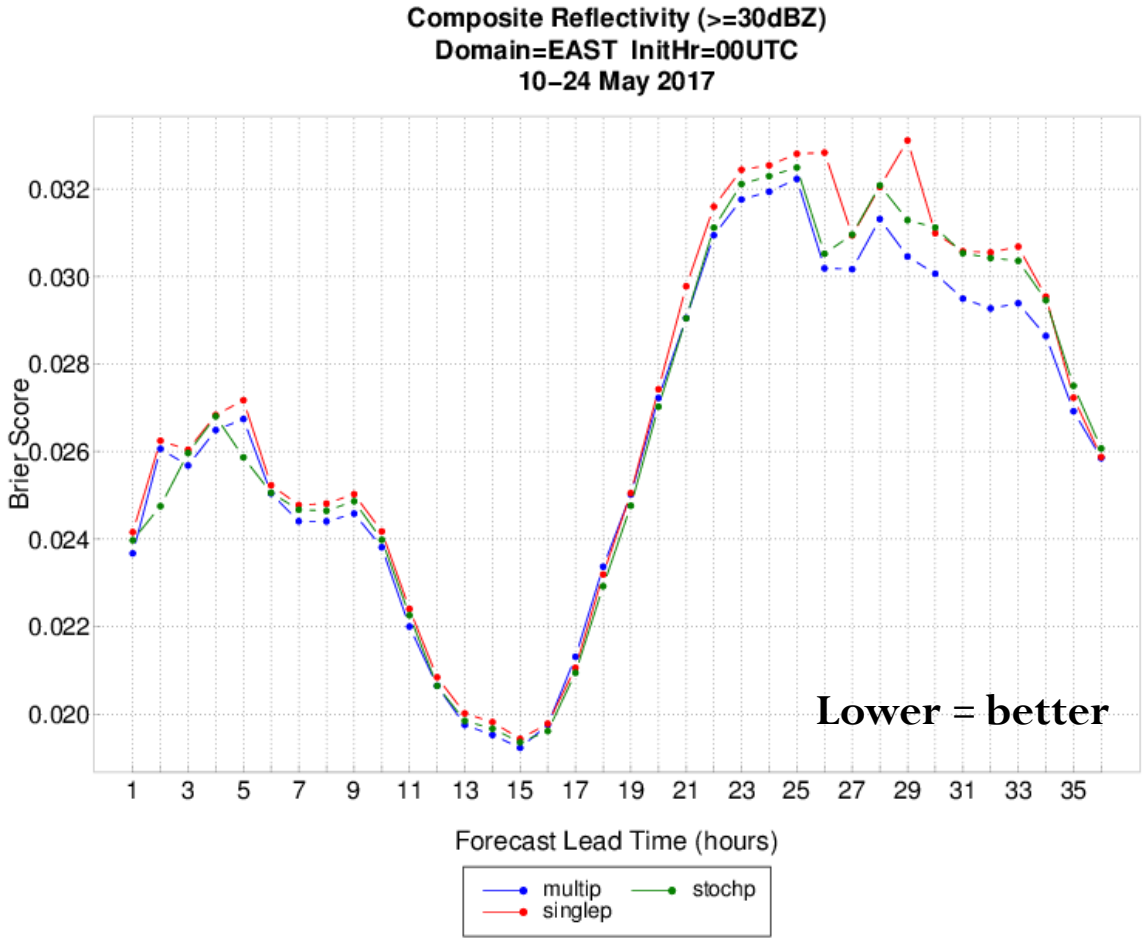
Northerly displacement

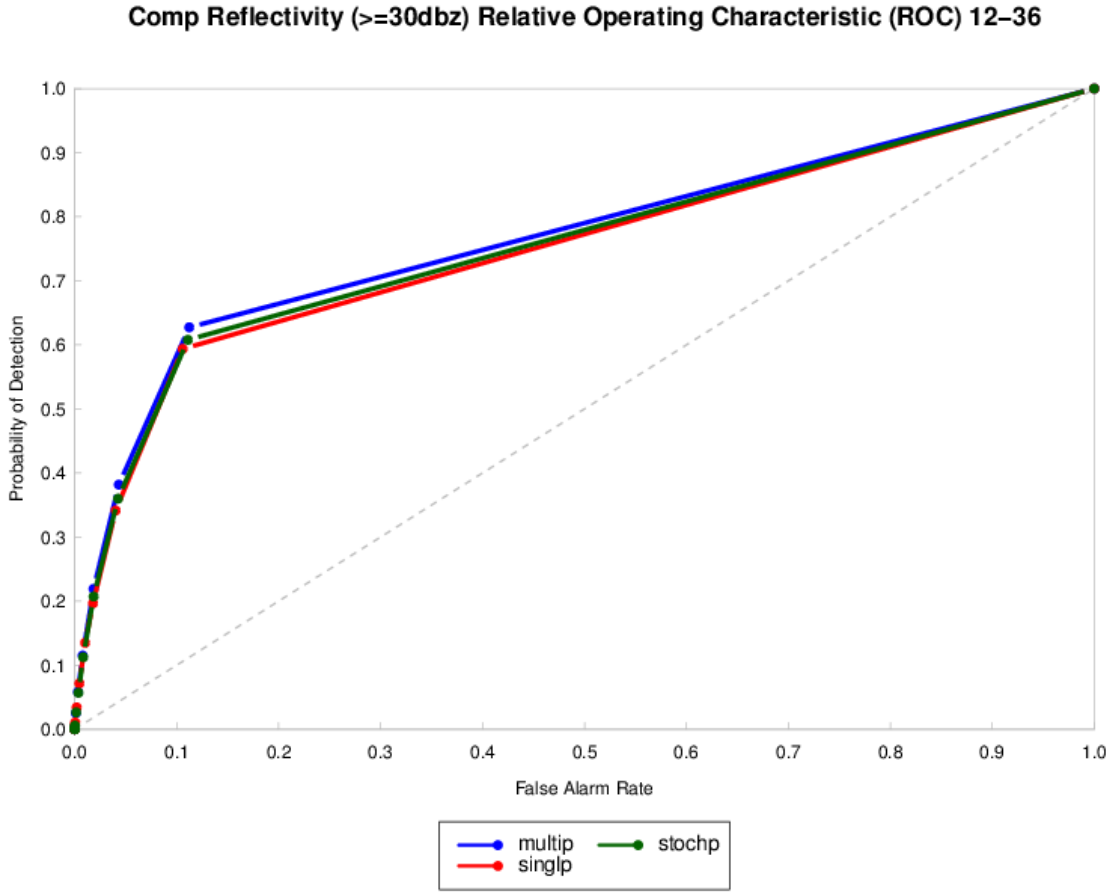


Southerly displacement

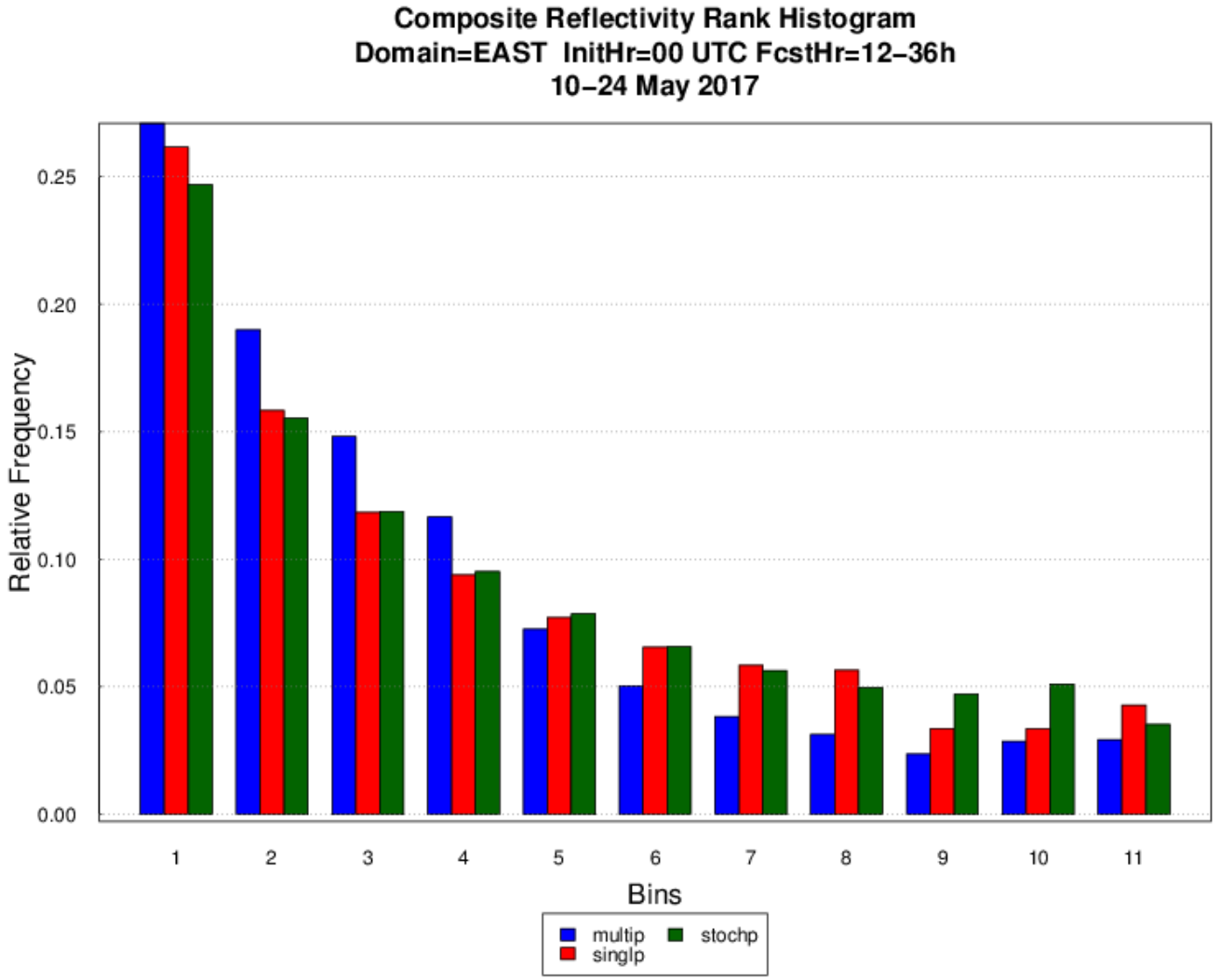
Southerly displacement

Composite Reflectivity (≥ 30dBZ) Brier Score and Components





Composite Reflectivity (≥ 30dBZ) Rank Histogram



Summary

- Many verification techniques are available to evaluate ensemble performance.
- **Multi-physics** ensemble has largest variability and members exhibiting the largest bias
- **Single physics** and **stochastic physics** perform comparably for most fields.
  - Promising result because stochastic physics ensemble is easier to maintain and less resource intensive.

Future Work

- Evaluate the CLUE 2018 dataset, with a focus on FV3 members.
- Evaluate echo top in addition to accumulated precipitation and composite reflectivity.
- Investigate additional verification techniques to assess performance.

Acknowledgements

- Thank you to all those who contributed to the CLUE 2017 ensemble subset:
  - NSSL
  - CAPS
  - UND
  - ESRL/GSD
  - NCAR
  - EMC
- This work could not have been conducted without collaborations with NSSL, SPC, and all of those involved with HWT-SFE.

The Developmental Testbed Center is funded by the National Oceanic and Atmospheric Administration (NOAA), the United States Air Force, the National Center for Atmospheric Research (NCAR), and the National Science Foundation (NSF). NCAR is sponsored by NSF.

Summary

- Many verification techniques are available to evaluate ensemble performance.
- **Multi-physics** ensemble has largest variability and members exhibiting the largest bias
- **Single physics** and **stochastic physics** perform comparably for most fields.
  - Promising result because stochastic physics ensemble is easier to maintain and less resource intensive.

Future Work

- Evaluate the CLUE 2018 dataset, with a focus on FV3 members.
- Evaluate echo top in addition to accumulated precipitation and composite reflectivity.
- Investigate additional verification techniques to assess performance.