SPRING FORECASTING EXPERIMENT 2018

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

HWT Facility – National Weather Center

30 April - 1 June 2018

https://hwt.nssl.noaa.gov/sfe/2018/

Program Overview and

Operations Plan

Burkely T. Gallo[2,3], Adam Clark[2], Israel Jirak[1], Steven J. Weiss[1], Andy Dean[1,3], Kent Knopfmeier[2,3], Brett Roberts[2,3], Louis Wicker[2], Makenzie Krocak[3,4], Nathan Wendt[1,3], Patrick Skinner[2,3], Jessica Choate[2,3], Pam Heinselman[2], Katie Wilson[2,3], Robert Hepper[1,3], James Correia[1,3], Gerry Creager[2,3], Thomas Jones[2,3], Jidong Gao[1], Yunheng Wang[2,3], Scott Dembek[2,3]

(1) NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma

(2) NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma

(3) Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

(4) School of Meteorology, University of Oklahoma, Norman, Oklahoma

1. Introduction

Each spring, the Experimental Forecast Program (EFP) of the NOAA/Hazardous Weather Testbed (HWT), organized by the Storm Prediction Center (SPC) and National Severe Storms Laboratory (NSSL), conducts a collaborative experiment to test emerging concepts and technologies designed to improve the prediction of hazardous convective weather. The primary goals of the HWT are to accelerate the transfer of promising new tools from research to operations, to inspire new initiatives for operationally relevant research, and to identify and document sensitivities and the performance of state-of-the art experimental convection-allowing (1 to 4 km grid-spacing) modeling systems (CAMs).

The 2018 Spring Forecasting Experiment (SFE 2018), a cornerstone of the EFP, will be conducted 30 April – 1 June with participation expected from about 100 forecasters, researchers, and model developers from around the world. Building upon successful experiments of previous years, a main emphasis of SFE 2018 will be the generation of probabilistic forecasts of severe weather valid over shorter time periods than current SPC operational products. This will be an important step toward addressing a strategy within the National Weather Service of providing nearly continuous probabilistic hazard forecasts on increasingly fine spatial and temporal scales, consistent with the NOAA Forecasting a Continuum of Environmental Threats (FACETs) vision. As in previous experiments, a suite of new and improved experimental CAM guidance contributed by our large group of collaborators will be central to the generation of these forecasts. Furthermore, for the third year, these contributions have been coordinated into a single ensemble framework called the Community Leveraged Unified Ensemble (CLUE; Clark et al. 2018). The 2018 CLUE is constructed by using a set of common model specifications (e.g., grid-spacing, vertical levels, domain size, etc.) so that the simulations contributed by each group can be used in carefully designed controlled experiments. This design will once again allow us to conduct several experiments geared toward identifying optimal configuration strategies for CAM-based ensembles. The 2018 CLUE includes 81 members using 3-km grid-spacing that will allow a set of five unique experiments. An additional feature of SFE 2018 will involve the continued testing of a Warn-on-Forecast prototype system called the NSSL Experimental Warn-on-Forecast System for Ensembles (NEWS-e), which will be used for the second year to issue very short lead-time outlooks during the afternoon. Additionally, surveys and real-time analytics will be used to mine information on how the NEWS-e products are used and interpreted.

This operations plan summarizes the core interests of SFE 2018 and provides information on the operations of the experiment. Detailed information on the organizational structure of the HWT and information on various forecast tools and diagnostics can also be found in this document. The remainder of the operations plan is organized as follows: Section 2 provides details on a number of new models and products being introduced during SFE 2018 and Section 3 describes the core interests and new concepts being introduced for SFE 2018. A list of daily participants, details on the SFE forecasting, and more general information on the HWT are found in appendices.

2. Overview of Experimental Products and Models

A primary goal of the SFE 2018 forecasting activities will be to test methods for generating probabilistic forecasts of severe weather that are valid over shorter time windows than current SPC operational products. Two separate groups led by SPC and NSSL staff, named the Severe Hazards and Innovation Desks, respectively, will issue different sets of convective outlooks for this testing. The Severe Hazards desk will issue Day 1 full-period outlooks (valid 1600 to 1200 UTC for Day 1) for individual severe weather hazards (tornado, wind, and hail), along with three 4-h period outlooks within the Day 1 period for each hazard covering the periods 1700 to 2100 UTC, 1900 to 2300 UTC, and 2100 to 0100 UTC. The Innovation Desk will be issuing Day 1 full-period outlooks for total severe (i.e., outlook for combined hazards of severe hail, wind, or tornadoes), as well as potential severe timing (PST) areas, which will indicate when the combined hazard probability will be ≥ 15% during the outlook day. These 4-h periods can occur at any time within the full-period outlook. These PSTs are designed to highlight areas and timing of severe weather occurrence as in the 2016 and 2017 SFEs, but take a different approach to the isochrones featured in those experiments. However, the goals of the activities are the same – namely to explore the feasibility of issuing a timing product to supplement current categorical forecast products (e.g., SPC Mesoscale Discussions and Severe Thunderstorm/Tornado Watches). Finally, for the second year the Innovation desk will conduct a short-term forecasting activity during the afternoon using the NEWS-e to issue two probabilistic total severe outlooks valid 2100–2200 UTC and 2200–2300 UTC. The Severe Hazards desk will use these forecasts to update their hazard forecasts for the 4-hour period valid 2100–0100 UTC. These activities are the second time a WoF-prototype has been tested in the EFP, and explores the potential utility of WoF products for issuing guidance between the watch and warning time scales (i.e. 0.5 to 4-h lead times). These activities represent efforts to explore ways of seamlessly merging probabilistic severe weather outlooks with probabilistic severe weather warnings as part of NOAA's Warn-on-Forecast (WoF; Stensrud et al. 2009) and Forecasting a Continuum of Environmental Threats (FACETs; http://www.nssl.noaa.gov/projects/facets/) initiatives. These efforts also support efforts to transition to higher temporal resolution forecasts at the SPC.

Generating the forecasts described above will be intensive and will thus rely on deterministic and ensemble CAM output for guidance and to generate first guesses for the severe weather probabilities. Most of the CAMs will be based on recent versions of the Advanced Research Weather Research and Forecasting (WRF-ARW) model. In addition, an eleven-member ensemble of experimental convection-allowing (3-km grid-spacing) versions of the Finite-Volume Cubed-Sphere model (FV3) developed at NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) will be examined. The members have varying microphysics and planetary boundary layer (PBL) schemes, to explore the sensitivity of forecasts of severe convection to different parameterization schemes, and to build on the two versions of the FV3 that were run during SFE 2017. The FV3 was selected to replace the GFS as part of the Next Generation Global Prediction System (NGGPS) program. Furthermore, NOAA plans for the FV3 to be the foundation of a unified modeling suite encompassing all prediction time and space scales currently under the purview of NOAA's Environmental Modeling Center (EMC). Thus, FV3 may eventually be the operational dynamical core for regional deterministic and ensemble systems. However, the FV3 has thus far only been tested in real-time for high resolution forecasting applications during SFE 2017, and continued testing during SFE 2018 will provide further insights into the model behavior. In addition to examining two-dimensional forecast fields over the CONUS, an evaluation of forecast soundings will also be performed.

In addition to the ensemble subsets contained within the 81-member CLUE system, the United Kingdom Meteorological (UK Met) Office will provide 0000 and 1200 UTC initialized 120-h forecasts from a 2.2-km grid-spacing regional version of their Unified Modeling System and from their 10-km grid-spacing global model. Additionally, several versions of the High Resolution Ensemble Forecast system Version 2 (HREFv2) will be examined, which is a formalized implementation of the Storm Scale Ensemble of Opportunity (SSEO) with all members run at EMC using 3-km grid-spacing. The HREFv2 became operational in November of 2017. SFE 2018 will compare the 8-member version of the HREFv2 from the SPC's website with versions that include multiple runs of the High-Resolution Rapid Refresh (HRRR) and those that remove time-lagged members. Results from this evaluation will be used to make recommendations for future versions of the HREF system.

For the generation of first-guess guidance forecasts from the CAM ensembles, it is important to extract explicit and proxy variables in the forecasts that track the potential of severe weather in the models. Previous SFEs and operational experience have shown that fields like hourly-maximum updraft helicity (UH) and hourly-maximum wind speed near the surface can be effective for highlighting the likelihood of severe weather in CAMs (Kain et al. 2010, Sobash et al. 2011, Clark et al. 2013; Gallo et al. 2016, 2018; Sobash et al. 2016a, b). To support the goal of SFE 2018 to generate forecasts of individual hazards, there will be further efforts to explore the ability of new model fields to delineate individual hazards, particularly for the size of hail. The hail size fields will be formally evaluated within the mixed-physics CLUE subset provided by the University of Oklahoma (OU) Center for Analysis and Prediction of Storms (CAPS), and will include hail size forecasts derived by (1) the HAILCAST algorithm (Adams-Selin and Ziegler 2016), which predicts maximum hail size using a hail growth model coupled to WRF, (2) the Thompson method, which estimates hail size directly from the microphysics size distribution by finding the largest graupel or hail hydrometeor diameter that exceeds a specified number concentration, (3) a method developed at SPC that estimates hail size based on the hourly maximum updraft speed (i.e., a spherical hailstone grows until its terminal velocity exceeds the maximum forecast updraft speed), and (4) a machine-learning-based method that provides probabilistic hail size forecasts (Gagne et al. 2017). An ensemble-subsetting method developed by researchers at Texas Tech University will also be tested during SFE 2018, in which ensemble sensitivity to desired fields will be used to eliminate ensemble members that are performing poorly in sensitive areas. Probabilities from the full ensemble and the subset of ensemble members will then be compared to determine the effect of the subsetting method.

Finally, new methods of real-time verification will be taking place during SFE 2018. One method will generate an experimental, CAM scorecard being developed jointly with scientists at the Developmental Testbed Center (DTC), NSSL, and SPC. The purpose of this scorecard is to follow recommendations to unify verification systems between NOAA partner labs and the DTC where possible. The scorecard is based on the Model Evaluation Tools (MET) software package, and includes metrics specific to CAM ensembles, such as surrogate severe probabilities generated using UH (following Sobash et al. 2011). This verification will be applied to a subset of deterministic and ensemble forecasts, and will update in real-time throughout the experiment. The rest of this section provides further details on each modeling system utilized in SFE 2018.

a) The 2018 Community Leveraged Unified Ensemble (CLUE)

The CLUE is a carefully designed ensemble with subsets of members contributed by NSSL, the OU CAPS and Multi-scale data Assimilation and Predictability (MAP) groups, NOAA's Earth Systems Research Laboratory/Global Systems Division (ESRL/GSD), the National Center for Atmospheric Research (NCAR), and GFDL. All members are initialized weekdays at 0000 UTC with 3-km grid-spacing covering a CONUS domain. All

WRF-ARW members have 1620 grid-points in the east-west direction, 1120 north-south, and 51 vertical levels[1] with a model top of 50 hPa and use RRTMG short and long wave radiation. Depending on the CLUE subset, forecast lengths range from 36 to 120 h. Specifications for the members within each subset are detailed in the tables below.

Table 1. Specifications for the "mixed physics" members of the CLUE that use the ARW dynamic core, mixed physics, perturbed ICs/LBCs, and ARPS 3DVAR radar assimilation. NAMa and NAMf refer to 12 km NAM analysis and forecast, respectively. RAPa refers to 13 km RAP analysis and GFSf refers to 18 UTC initialized GFS forecasts (note, 18UTC GFS initializations are used for LBCs because the 00 UTC forecasts are not available in time). 3DVAR refers to ARPS 3DVAR and cloud analysis. Under the IC column, the model names appended with "pert" refer to perturbations extracted from a 16 km grid-spacing SREF member. For members core03-10 under the BC column, names refer to SREF member forecasts. The starred member (mixed-phys11) is identical to mixed-phys01 but uses a different strategy for the vertical levels within WRF.

(1) CAPS: mixed phys						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
mixed-phys01	NAMa+3DVAR	NAMf	Thompson	NOAH	MYJ	arw
mixed-phys02	RAPa+3DVAR	GFSf	Thompson	RUC	MYNN	arw
mixed-phys03	mixed-phys01+arw-p1_pert	arw-p1	NSSL	NOAH	YSU	arw
mixed-phys04	mixed-phys01+arw-n1_pert	arw-n1	NSSL	NOAH	MYNN	arw
mixed-phys05	mixed-phys01+nmmb-p1_pert	nmmb-p1	Morrison	NOAH	MYJ	arw
mixed-phys06	mixed-phys01+nmmb-n1_pert	nmmb-n1	P3	NOAH	YSU	arw
mixed-phys07	mixed-phys01+arw-p2_pert	arw-p2	NSSL	NOAH	MYJ	arw
mixed-phys08	mixed-phys01+arw-n2_pert	arw-n2	Morrison	NOAH	YSU	arw
mixed-phys09	mixed-phys01+nmmb-p2_pert	nmmb-p2	P3	NOAH	MYNN	arw
mixed-phys10	mixed-phys01+nmmb-n2_pert	nmmb-n2	Thompson	NOAH	MYNN	arw
mixed-phys11*	NAMa+3DVAR	NAMf	Thompson	NOAH	MYJ	arw

Table 2. Specifications for the "te14" CLUE member that uses the Thompson-Eidhammer (TE14; Thompson and Eidhammer [2014]) microphysics scheme, which includes a stochastic perturbation technique.

(2) CAPS: TE14						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
te14	RAPa+3DVAR	GFSf	TE14	RUC	MYNN	arw

[1] WRF-ARW sigma levels are set to: 1.0, 0.998, 0.994, 0.987, 0.975, 0.959, 0.939, 0.916, 0.892, 0.865, 0.835, 0.802, 0.766, 0.727, 0.685, 0.64, 0.592, 0.542, 0.497, 0.4565, 0.4205, 0.3877, 0.3582, 0.3317, 0.3078, 0.2863, 0.267, 0.2496, 0.2329, 0.2188, 0.2047, 0.1906, 0.1765, 0.1624, 0.1483, 0.1342, 0.1201, 0.106, 0.0919, 0.0778, 0.0657, 0.0568, 0.0486, 0.0409, 0.0337, 0.0271, 0.0209, 0.0151, 0.0097, 0.0047, 0.0.

Table 3. Specifications for the "single-phys" members of the CLUE that use the ARW dynamic core, single physics, perturbed ICs and LBCs, and radar data assimilation using ARPS 3DVAR. The blue shaded member is repeated from Table 1 because it is also the control member of the core ensemble subset.

(3) CAPS: single phys						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
single-phys01	NAMa+3DVAR	NAMf	Thompson	RUC	MYNN	arw
single-phys02 (mixed-phys02)	RAPa+3DVAR	GFSf	Thompson	RUC	MYNN	arw
single-phys03	core01+arw-p1_pert	arw-p1	Thompson	RUC	MYNN	arw
single-phys04	core01+arw-n1_pert	arw-n1	Thompson	RUC	MYNN	arw
single-phys05	core01+nmmb-p1_pert	nmmb-p1	Thompson	RUC	MYNN	arw
single-phys06	core01+nmmb-n1_pert	nmmb-n1	Thompson	RUC	MYNN	arw
single-phys07	core02+arw-p2_pert	arw-p2	Thompson	RUC	MYNN	arw
single-phys08	core02+arw-n2_pert	arw-n2	Thompson	RUC	MYNN	arw

*Table 4. Specifications for the "stoch-phys" CLUE members that use ARPS 3DVAR radar assimilation and cloud analysis, single physics, and stochastic physics perturbations (*indicates components with stochastic perturbations). For more information on the physics perturbations see the description of experiments below.*

(4) CAPS: stochastic physics + radar						
Members	**IC**	**BC**	**Microphysics**	**LSM***	**PBL***	**Model**
stoch-phys01	NAMa+3DVAR	NAMf	Thompson	RUC	MYNN	arw
stoch-phys02 (mixed-phys02)	RAPa+3DVAR	GFSf	Thompson	RUC	MYNN	arw
stoch-phys03	core01+arw-p1_pert	arw-p1	Thompson	RUC	MYNN	arw
stoch-phys04	core01+arw-n1_pert	arw-n1	Thompson	RUC	MYNN	arw
stoch-phys05	core01+nmmb-p1_pert	nmmb-p1	Thompson	RUC	MYNN	arw
stoch-phys06	core01+nmmb-n1_pert	nmmb-n1	Thompson	RUC	MYNN	arw
stoch-phys07	core02+arw-p2_pert	arw-p2	Thompson	RUC	MYNN	arw
stoch-phys08	core02+arw-n2_pert	arw-n2	Thompson	RUC	MYNN	arw

Table 5. Specifications for the "caps-enkf" members of the CLUE, which are run as follows: The 3-km GSI-EnKF system will be initialized at 1800 UTC each day, and will assimilate the RAP/HRRR GSI data stream hourly (except satellite data) from 1800-0000 UTC over the CONUS domain. Radar data will be assimilated every 15 minutes from 2300-0000 UTC using the CAPS EnKF system. The ensemble consists of 40 ARW members with initial perturbations and mixed physics options to provide input for the EnKF ensemble analyses. Each member uses Thompson microphysics, although with varied graupel density among members. A 12-member ensemble forecast (run for 48 h) follows using the final EnKF analyses at 0000 UTC using the same multi-physics configurations. The specifications for the 12 forecast members are listed below. The starred member is experimental and not yet optimally configured. As such, configuration of this member may shift during the testbed and it will not be viewed during SFE 2018.

(5) CAPS: EnKF						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
caps-enkf01	enkf_m01a	NAMf	Thompson	NOAH	MYJ	arw
caps-enkf02	enkf_m02a	arw-p1	NSSL	NOAH	YSU	arw
caps-enkf03	enkf_m15a	arw-n1	NSSL	NOAH	MYNN	arw
caps-enkf04	enkf_m40a	nmmb-p1	Morrison	NOAH	MYJ	arw
caps-enkf05	enkf_m08a	nmmb-n1	P3	NOAH	YSU	arw
caps-enkf06	enkf_m26a	arw-p2	NSSL	NOAH	MYJ	arw
caps-enkf07	enkf_m39a	arw-n2	Morrison	NOAH	YSU	arw
caps-enkf08	enkf_m12a	nmmb-p2	P3	NOAH	MYNN	arw
caps-enkf09	enkf_nm	NAMf	Thompson	NOAH	MYJ	arw
caps-enkf10	enkf_nm	NAMf	NSSL	NOAH	MYJ	arw
caps-enkf11	enkf_m25a	nmmb-n2	Thompson	NOAH	MYNN	arw
caps-enkf12*	3Dvar	NAMf	Thompson	NOAH	MYJ	arw

Table 6. Specifications for the FV3 ensemble. This ensemble will test different microphysics, PBL parameterizations, and cumulus parameterizations (outside of the high-resolution nest) within the FV3. SA-SAS refers to the scale-aware simplified Arakawa Schubert parameterization. Further description of FV3 is provided in the summary of experiments in the next section.

(6) CAPS: FV3 Physics							
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**	**Cumulus**
fv3-phys01	GFS	n/a	Thompson	NOAH	MYNN-SA	fv3	Tiedtke
fv3-phys02	GFS	n/a	Thompson	NOAH	MYNN	fv3	Tiedtke
fv3-phys03	GFS	n/a	Thompson	NOAH	YSU-SA	fv3	Tiedtke
fv3-phys04	GFS	n/a	Thompson	NOAH	YSU	fv3	Tiedtke
fv3-phys05	GFS	n/a	Thompson	NOAH	EDMF	fv3	Tiedtke
fv3-phys06	GFS	n/a	NSSL	NOAH	MYNN-SA	fv3	Tiedtke
fv3-phys07	GFS	n/a	NSSL	NOAH	MYNN	fv3	Tiedtke
fv3-phys08	GFS	n/a	NSSL	NOAH	YSU-SA	fv3	Tiedtke
fv3-phys09	GFS	n/a	NSSL	NOAH	YSU	fv3	Tiedtke
fv3-phys10	GFS	n/a	NSSL	NOAH	EDMF	fv3	Tiedtke
fv3-phys11	GFS	n/a	Thompson	NOAH	MYNN-SA	fv3	SA-SAS

Table 7. Specifications for the "map-hybrid" members of the CLUE. These 3-km grid-spacing ensemble forecasts are run with WRF ARW and initialized by a GSI-based EnKF-Var hybrid DA system assimilating both conventional and radar observations (Johnson et al. 2015, Wang and Wang 2017). The ensemble for data assimilation has 41 members. The LBCs are provided by re-centering GEFS and SREF around the GFS control. The system assimilates the operational RAP/HRRR in-situ data stream hourly during 1800-0000 UTC and radar observations (reflectivity and velocity) every 20-min from 23-00Z over the CONUS domain. The control member is updated by GSI-based EnKF-Var hybrid DA system; the remaining 40 members are updated by EnKF and recentered around the EnKF-Var hybrid control analysis (Wang et al. 2018a,b). A 10-member ensemble forecast is initialized at 00Z and made out to 36 hours, including one forecast(map-hybrid01) initialized from the GSI EnKF-Var hybrid control analysis and 9-member re-centered GSI EnKF analyses.

(7) MAP-hybrid: WRF-ARW GSI-EnKF						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
map-hybrid01	EnVar	GFS-control	Thompson	RUC	MYNN	arw
map-hybrid02	rEnKF_m1	GEFS	Thompson	RUC	MYNN	arw
map-hybrid03	rEnKF_m2	GEFS	Thompson	RUC	MYNN	arw
map-hybrid04	rEnKF_m3	GEFS	Thompson	RUC	MYNN	arw
map-hybrid05	rEnKF_m4	GEFS	Thompson	RUC	MYNN	arw
map-hybrid06	rEnKF_m5	GEFS	Thompson	RUC	MYNN	arw
map-hybrid07	rEnKF_m6	GEFS	Thompson	RUC	MYNN	arw
map-hybrid08	rEnKF_m7	GEFS	Thompson	RUC	MYNN	arw
map-hybrid09	rEnKF_m8	GEFS	Thompson	RUC	MYNN	arw
map-hybrid10	rEnKF_m9	GEFS	Thompson	RUC	MYNN	arw

Table 8. Specifications for the "ncar" members of the CLUE. This ensemble provides forecasts to 48 h and uses NCAR's DART (Data Assimilation Research Testbed) software with ARW version 3.8 with the same horizontal domain as the HRRRE CLUE members. The mesoscale analysis system is comprised of 80 members that are continuously cycled using the ensemble adjustment Kalman filter (EAKF). New analyses are produced every 1 h with 15-km grid-spacing. Additional analyses are downscaled twice daily on the 3-km forecast grid, also with hourly cycling, for a window of 6 hours. These 3-km analyses are used to initialize CLUE forecasts, nested within forecasts initialized from the 15-km analysis domain. Other specifications include: 51 vertical levels with a 50 hPa top, a horizontal localization of 1270 km and vertical localization of 1.5 scale heights, relaxation to prior spread posterior inflation, sampling error correction, spread restoration, and freely-evolving soil states. The following observational sources are utilized: PREPBUFR ACARS, METARs, radiosondes, profilers and marine, CIMMS cloud-track winds, Oklahoma Mesonet, and GPS radio occultation. All analysis members have constant physics, which include Tiedtke cumulus parameterization, aerosol aware Thompson microphysics, MYNN PBL, RUC land-surface model, and RRTMG shortwave and longwave radiation with aerosol and ozone climatologies. The 10-member forecasts are initialized daily at 0000 and 1200 UTC with ICs provided by the first ten ensemble analysis members of the WRF/DART EAKF analyses (described above). Perturbed LBCs from GFS forecasts are used. The physics are the same as from the data assimilation system, but without cumulus parameterization (detailed in the HRRRE table caption).

(8) NCAR: EnKF						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
ncar01	anal01	GFS	Thompson	NOAH	MYNN	arw
ncar02	anal02	GFS	Thompson	RUC	MYNN	arw
ncar03	anal03	GFS	Thompson	RUC	MYNN	arw
ncar04	anal04	GFS	Thompson	RUC	MYNN	arw
ncar05	anal05	GFS	Thompson	RUC	MYNN	arw
ncar06	anal06	GFS	Thompson	RUC	MYNN	arw
ncar07	anal07	GFS	Thompson	RUC	MYNN	arw
ncar08	anal08	GFS	Thompson	RUC	MYNN	arw
ncar09	anal09	GFS	Thompson	RUC	MYNN	arw
ncar10	anal10	GFS	Thompson	RUC	MYNN	arw

Table 9. The experimental HRRR ensemble (HRRRE) is initialized at 0300 UTC each day from a combination of atmospheric RAPv4 mean and GFS data assimilation ensemble (GDAS) perturbations along with HRRRv3 land surface data. A total of 36 3-km HRRR members are initialized and then cycled hourly through 0000 UTC using an Ensemble Kalman filter to assimilate conventional and radar observations each hour followed by the application of the HRRR cloud analysis and soil adjustment to each member. At 0000 UTC, nine members produce 36 h forecasts. Stochastic soil moisture perturbations are introduced across all members at 0300 UTC along with lateral boundary perturbations at 0000 UTC and inflation during the cycled data assimilation to promote spread and represent both initial condition and model forecast uncertainties. The HRRRE uses WRF-ARW version 3.8 with the same physics configuration as the HRRRv3.

(9) GSD: HRRRE						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
hrrre01	enkf_m01b	GFS	Thompson	RUC	MYNN	arw
hrrre02	enkf_m02b	GFS	Thompson	RUC	MYNN	arw
hrrre03	enkf_m03b	GFS	Thompson	RUC	MYNN	arw
hrrre04	enkf_m04b	GFS	Thompson	RUC	MYNN	arw
hrrre05	enkf_m05b	GFS	Thompson	RUC	MYNN	arw
hrrre06	enkf_m06b	GFS	Thompson	RUC	MYNN	arw
hrrre07	enkf_m07b	GFS	Thompson	RUC	MYNN	arw
hrrre08	enkf_m08b	GFS	Thompson	RUC	MYNN	arw
hrrre09	enkf_m09b	GFS	Thompson	RUC	MYNN	arw

Table 10. Specifications for the "HRRR36" experimental member of CLUE that uses the ARW dynamic core with HRRR physics and data assimilation. For further details, see Section 2f.

(10) GSD: 36 h HRRR						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
hrrr36	RAP	GFSf	Thompson	RUC	MYNN	arw

Table 11. Specifications for the deterministic FV3 members of CLUE. Further description of FV3 is provided in the summary of experiments in the next section.

(11) NSSL/GFDL: FV3 preliminary tests						
Members	**IC**	**BC**	**Microphysics**	**LSM**	**PBL**	**Model**
nssl-fv3	GFS	n/a	Thompson	NOAH	MYNN	fv3
gfdl-fv3	GFS	n/a	GFDL-6cat	NOAH	YSU	fv3

The configuration of the 2018 CLUE will allow for five unique experiments that have been designed to examine issues immediately relevant to the design of a NCEP/EMC operational CAM-based ensemble. These experiments are listed below:

(1) Physics perturbation experiment: Three ensembles will be compared that all have perturbed ICs/LBCs - one will use single physics (*single-phys01-08*; Table 3), one mixed-physics (*mixed-phys01-08*; Table 1), and one single physics with stochastic perturbations (*stoch-phys01-08;* Table 4). The ultimate goal is for the stochastic perturbations to be as good as or better than the mixed-physics in terms of forecast skill and reliability. The single physics will be used as a baseline against which both mixed-physics and stochastic physics strategies can be compared. CAPS, NSSL and DTC have worked together to configure the *stoch-phys* members. The stochastic perturbations are created by applying Stochastic Parameter Perturbation (SPP) within the MYNN PBL scheme and Thompson microphysics, which are a part of the HRRR/RAP physics suite. The SPP approach is based on stochastic pattern generator that produces 2D perturbation fields with spatial and temporal correlations. The approach is analogous to the one used at European Center for Medium-Range Weather Forecasts (ECMWF) to perturb physics tendencies (Palmer 2009). The pattern is fully determined by four parameters specified by the user (with the namelist setting shown in parentheses): grid point standard deviation (gridpt_stddev and stddev_cutoff), length scales (lengthscale) and de-correlation time (timescale). One example of 2-D patterns for different scales is presented in Figure 1.

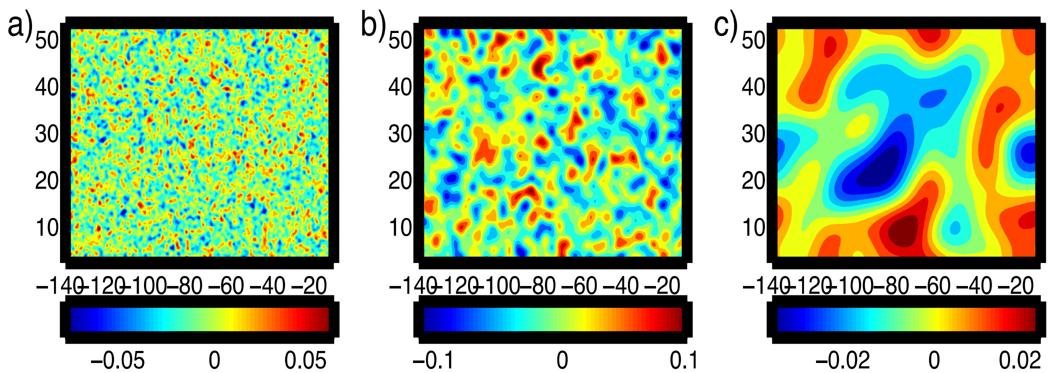

Figure 1. Example perturbation patterns for different spatial correlations a) convective, b) meso- and c) large-scale.

Within MYNN, SPP was applied to sub-grid cloud fraction, mixing length, roughness lengths (aerodynamic, thermal and moisture), mass fluxes and Prandtl number (only for use in stable conditions). For the Thompson microphysics, SPP was applied to the relationship used to specify the Y-intercept parameter of the assumed inverse exponential size distribution within the hybrid graupel/hail category.

(2) Data assimilation comparisons: Similar to the 2017 CLUE, there are several ensemble subsets that use various data assimilation strategies with an EnKF component. The *map-hybrid* (Table 7*)*, *ncar (*Table 8), and *hrrre* (Table 9) are similarly configured except for their data assimilation systems, allowing for a comparison between strategies. Additionally, the *caps-enkf* ensemble utilizes EnKF data assimilation with a set of mixed physical parameterizations, providing another complementary configuration for this experiment, albeit less controlled than the comparison between the other three configurations. From this experiment, we hope to be able to speak to the relative performance and advantages/disadvantages of each data assimilation method.

(3) Microphysics sensitivities: One member of the 2018 CLUE being run by CAPS examines the effects of using a stochastically perturbed microphysics scheme (*te14*, Table 2). The perturbations are applied to the shape parameter of the gamma distribution describing the distribution of cloud water and to the grid-resolved vertical velocity field during cloud condensation nuclei and ice nuclei activation. Applying these perturbations is hoped to help account for large uncertainty in the spatial distribution of water vapor and the strength of sub-grid scale vertical motions.

(4) FV3 (Credit: Lucas Harris)**:** Building upon the FV3 activities from SFE 2017, three different versions of FV3 (Table 6; *fv3-phys01*, Table 11) will be examined and compared to current well-known models (e.g., HRRRv3) to gauge performance at convective scales. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms. Storm surrogate fields like hourly maximum updraft helicity and updraft speed will also be examined to gauge their utility for forecasting severe storms. FV3 refers to the GFDL Finite-Volume Cubed-Sphere dynamical core (Putman and Lin 2007). This dynamical core uses the forward-in-time scheme and Lagrangian vertical coordinate of Lin (2004), based on the Lagrangian dynamics of Lin and Rood (1997) and the finite-volume pressure gradient force of Lin (1997). Fast vertically-propagating sound and gravity wave processes are handled by a semi-implicit solver. Advection of scalars uses the positive-definite two-dimensional advection scheme of Lin and Rood (1996) based on the piecewise-parabolic method. In the dynamics, grid-scale noise is dissipated through the use of an eighth-order divergence damping, Smagorinsky damping, and weak sixth-order damping of the non-monotonic vorticity and potential temperature fluxes.

The gfdl-fv3 member uses the Yonsei University PBL scheme (Hong et al. (2006) and the six-category single-moment GFDL microphysics parameterization (Chen and Lin 2013) with a cold-start initialization from 0000 UTC GFS analyses and forecasts that extend to 120-h. The gfdl-fv3 grid uses a combination of grid nesting (Harris and Lin, 2013) and stretching (Harris et al. 2016) to refine a 13-km global grid to a 3-km nested grid covering the CONUS region, which is displayed in Figure 2.

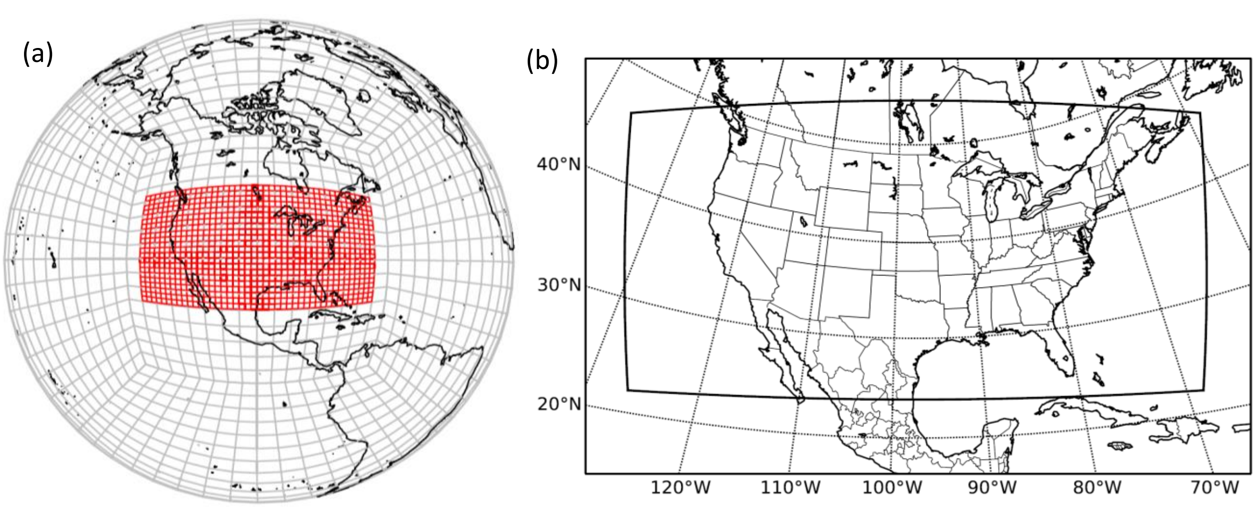

Figure 2. The grid used for the GFDL configuration of the FV3, from (a) a global and (b) a zoomed-in perspective.

The *nssl-fv3* member will use the latest release of NEMSfv3gfs in the NOAA VLab, which is from 27 February 2018. Furthermore, the CAPS-implemented Thompson microphysics scheme and the MYNN PBL scheme have been incorporated into the *nssl-fv3*. The *nssl-fv3* will use a similar dynamic configuration to *gfdl-fv3* and *caps-fv3* except for NEMSfv3gfs-specific settings. However, the *nssl-fv3* configuration will use a different global grid (about 25-km grid-spacing) to save computer power over regions outside the main area of forecast interest for SFE 2018 (i.e., outside of the CONUS). Correspondingly, the stretch factor for the tile over North America and the refinement ratio for nesting over the CONUS are adjusted to produce a 3.3-km CONUS grid (Figure 3). The *nssl-fv3* is initialized from 0000 UTC GFS analyses with forecasts to 36 h.

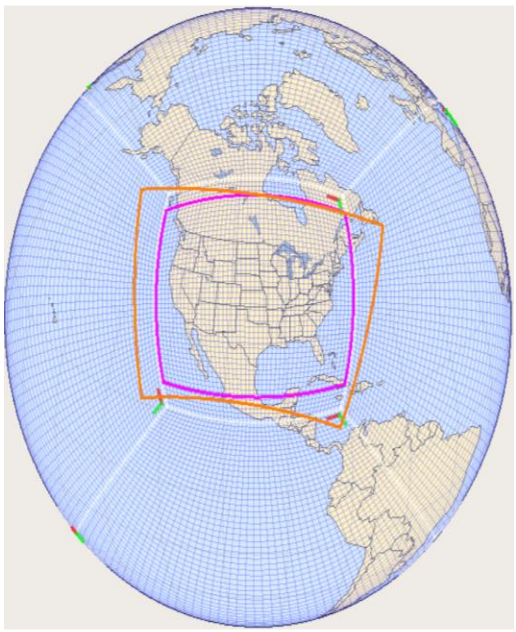

Figure 3. NSSL-fv3 domain configuration within the FV3 C384 global grid (~25 km). The stretch factor is 2.0 for the tile over North America (resolution ~13 km). The refinement ratio for nesting is 4, which gives a model domain (purple) similar as CONUS (orange) at ~3 km

The CAPS-run *fv3-phys01* member uses the MYNN-SA, which is an updated, "scale-aware" version of MYNN, and Thompson microphysics. These runs are initialized with 0000 UTC GFS analyses and forecast lengths extending to 84 h. The fv3-phys01 grid is configured differently relative to the FV3 runs provided by CAPS in SFE2017, or the runs from NSSL and GFDL of this year. The global grid will have an essentially uniform 13 km grid spacing, instead of using grid stretching. A 4:1 ratio is used for the two-way interactive nested grid over CONUS, resulting in a grid spacing of 3.25 to 3.5 km on the nested grid (see Fig. 4). The uniform global grid provides an opportunity to evaluate global forecasts using different physics.

Though the eventual goal is for FV3 to produce the same standard output as the remainder of the CLUE members, the initial focus is on commonly used parameters for forecasting severe weather. The output from the CAPS FV3 configurations, which serves as a rough template for all of the FV3 configurations, is presented in Table 12.

Table 12. Output variables for the CAPS FV3 runs

Variable	Level(s)/Layer(s)
Geopotential Height	Surface; 1000, 925, 850, 700, 500, 250 mb
Air Temperature	2 m; 1000, 925, 850, 700, 500 mb
Dewpoint Temperature	2 m; 1000, 925, 850, 700, 500 mb
U and V Winds	10m; 1000, 925, 850, 700, 500, 250 mb
Specific Humidity	2 m
Pressure	Surface
Sea-Level Pressure	Mean Sea Level
1-hr Maximum Updraft	Column-Max
1-hr Maximum Downdraft	Column-Max
Vertical Velocity	700 mb
Storm-Relative Helicity	0-1, 0-3 km AGL
CAPE	Surface-Based
CIN	Surface-Based
U and V Components of Bulk Shear	0-1, 0-6 km AGL
U and V Components of Storm Motion	-
1-hr Maximum Updraft Helicity	0-3, 2-5 km AGL
1-hr Minimum Updraft Helicity	0-3, 2-5 km AGL
Simulated Reflectivity	1, 4 km AGL; -10 C; Column-Max (Composite)
1-hr Maximum Simulated Reflectivity	1 km AGL
1-hr Precipitation Accumulation	Surface
Precipitable Water	Column-Total
Upwelling Longwave Radiation Flux	Top-of-Atmosphere
Downwelling Shortwave Radiation Flux	Surface
Height of the PBL Top	-
Cloud Cover	Column-Total, Low Cloud Level, Mid Cloud Level, High Cloud Level

(5) FV3 physics parameterizations: New to SFE 2018, CAPS will be providing an ensemble of FV3 members to test the impact of different parameterization schemes (Table 6). Two microphysics schemes, five PBL schemes, and two cumulus parameterizations (applicable outside of the high resolution nest) will be examined with this ensemble. This work will build upon the comparisons first made in SFE 2017 between two deterministic configurations of the FV3 run by CAPS and GFDL which used different microphysics schemes, and help identify sensitivities to specific parameterizations. The nesting configuration of the CAPS FV3 members is the same as *fv3-phys01* described in the previous section (Figure 4).

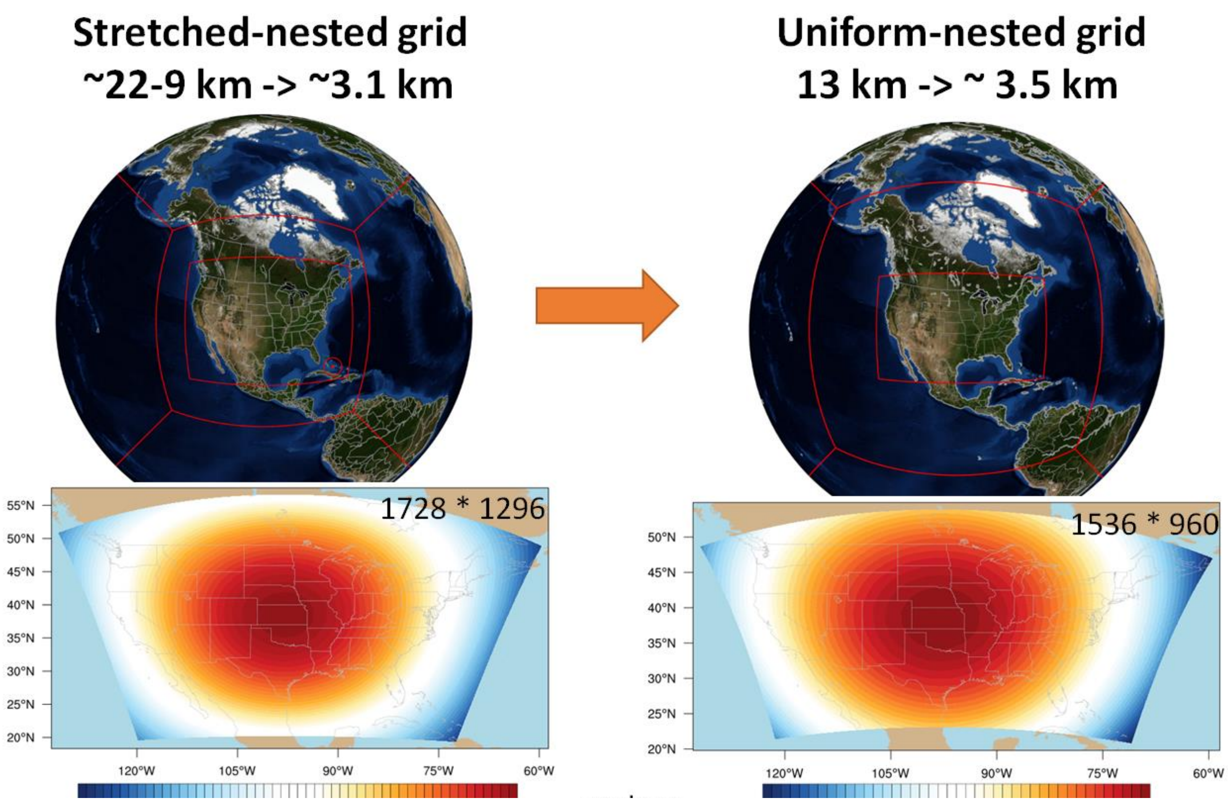

Figure 4. The current grid configuration for SFE 2018 (right) compared to the grid configuration used in SFE 2017 (left).

To ensure consistent post-processing, visualization, and verification for subsets of CLUE ensemble members contributed by different collaborators, all groups will utilize the same post-processing software to output the same set of model output fields on the same grid. Thus, NSSL worked closely with scientists at the Developmental Testbed Center (DTC) and EMC to modify the most recent version of Unified Post-Processor software (UPPv3.2; available at http://www.dtcenter.org/upp/users/downloads/index.php) to output a set of 123 output fields from each CLUE member (Table 13). These fields (output in grib2 format) are the same as the 2D fields output by the experimental HRRR (which will become operational in June 2018), and were chosen because of their relevance to a broad range of forecasting needs, including aviation, severe weather, and precipitation. All CLUE collaborating groups are providing this set of 123 fields, but they can also add additional diagnostics based on their own research interests. Furthermore, ensembles run by CAPS will output additional experimental hail diagnostics (Table 14).

Table 13. The set of 123 required output diagnostics for the CLUE members, which are output at hourly intervals.

Number	Level/Layer	Parameter	Description
001	entire atmosphere	REFC	Composite reflectivity [dB]
002	cloud top	RETOP	Echo Top [m]
003	entire atmosphere	VIL	Radar-Simulated Vertically Integrated Liquid [kg/m^2]
004	surface	VIS	Visibility [m]
005	1000 m above ground	REFD	Reflectivity [dB]
006	4000 m above ground	REFD	Reflectivity [dB]
007	surface	GUST	Wind Speed (Gust) [m/s]
008	500 mb	HGT	Geopotential Height [gpm]
009	500 mb	TMP	Temperature [K]
010	500 mb	DPT	Dew Point Temperature [K]
011	500 mb	UGRD	U-Component of Wind [m/s]
012	500 mb	VGRD	V-Component of Wind [m/s]
013	700 mb	HGT	Geopotential Height [gpm]
014	700 mb	TMP	Temperature [K]
015	700 mb	DPT	Dew Point Temperature [K]
016	700 mb	UGRD	U-Component of Wind [m/s]
017	700 mb	VGRD	V-Component of Wind [m/s]
018	850 mb	HGT	Geopotential Height [gpm]
019	850 mb	TMP	Temperature [K]
020	850 mb	DPT	Dew Point Temperature [K]
021	850 mb	UGRD	U-Component of Wind [m/s]

022	850 mb	VGRD	V-Component of Wind [m/s]
023	925 mb	TMP	Temperature [K]
024	925 mb	DPT	Dew Point Temperature [K]
025	925 mb	UGRD	U-Component of Wind [m/s]
026	925 mb	VGRD	V-Component of Wind [m/s]
027	1000 mb	TMP	Temperature [K]
028	1000 mb	DPT	Dew Point Temperature [K]
029	1000 mb	UGRD	U-Component of Wind [m/s]
030	1000 mb	VGRD	V-Component of Wind [m/s]
031	100-1000 mb above ground	MAXUVV	Hourly Max upward Vertical Velocity - lowest 100hPa [m/s]
032	100-1000 mb above ground	MAXDVV	Hrly Max downward Vertical Velocity - lowest 100hPa [m/s]
033	0.5-0.8 sigma layer	DZDT	Vertical Velocity (Geometric) [m/s]
034	mean sea level	PRMSL	Pressure Reduced to MSL [Pa]
035	1000 mb	HGT	Geopotential Height [gpm]
036	1000 m above ground	MAXREF	Hourly Max of Simulated Reflectivity at 1 km AGL [dB]
037	5000-2000 m above ground	MXUPHL	Hrly Max Updraft Helicity - 2km to 5 km AGL [m^2/s^2]
038	entire column	TCOLG	Total Column Integrated Graupel [kg/m^2]
039	surface	LTNG	Lightning [non-dim]
040	80 m above ground	UGRD	U-Component of Wind [m/s]
041	80 m above ground	VGRD	V-Component of Wind [m/s]
042	surface	PRES	Pressure [Pa]

043	surface	HGT	Geopotential Height [gpm]
044	surface	TMP	Temperature [K]
045	0 m underground	MSTAV	Moisture Availability [%]
046	surface	WEASD	Water Equivalent of Accumulated Snow Depth [kg/m^2]
047	surface	SNOWC	Snow Cover [%]
048	surface	SNOD	Snow Depth [m]
049	2 m above ground	TMP	Temperature [K]
050	2 m above ground	SPFH	Specific Humidity [kg/kg]
051	2 m above ground	DPT	Dew Point Temperature [K]
052	10 m above ground	UGRD	U-Component of Wind [m/s]
053	10 m above ground	VGRD	V-Component of Wind [m/s]
054	10 m above ground	WIND	Wind Speed [m/s]
055	surface	CPOFP	Percent frozen precipitation [%]
056	surface	PRATE	Precipitation Rate [kg/m^2/s]
057	surface	APCP	Total Precipitation [kg/m^2]
058	surface	WEASD	Water Equivalent of Accumulated Snow Depth [kg/m^2]
059	surface	APCP	Precipitation [kg/m^2] – hourly total
060	surface	WEASD	Water Equivalent of Accumulated Snow Depth [kg/m^2]
061	surface	CSNOW	Categorical Snow [-]
062	surface	CICEP	Categorical Ice Pellets [-]
063	surface	CFRZR	Categorical Freezing Rain [-]

064	surface	CRAIN	Categorical Rain [-]
065	surface	VGTYP	Vegetation Type [Integer(0- 13)]
066	500-1000 mb	LFTX	Surface Lifted Index [K]
067	surface	CAPE	Convective Available Potential Energy [J/kg]
068	surface	CIN	Convective Inhibition [J/kg]
069	entire column	PWAT	Precipitable Water [kg/m^2]
070	low cloud layer	LCDC	Low Cloud Cover [%]
071	middle cloud layer	MCDC	Medium Cloud Cover [%]
072	high cloud layer	HCDC	High Cloud Cover [%]
073	entire atmosphere	TCDC	Total Cloud Cover [%]
074	cloud base	PRES	Pressure [Pa]
075	cloud base	HGT	Geopotential Height [gpm]
076	cloud ceiling	HGT	Geopotential Height [gpm]
077	cloud top	PRES	Pressure [Pa]
078	cloud top	HGT	Geopotential Height [gpm]
079	top of atmosphere	ULWRF	Upward Long-Wave Rad. Flux [W/m^2]
080	surface	DSWRF	Downward Short-Wave Radiation Flux [W/m^2]
081	3000-0 m above ground	HLCY	Storm Relative Helicity [m^2/s^2]
082	1000-0 m above ground	HLCY	Storm Relative Helicity [m^2/s^2]
083	0-6000 m above ground	USTM	U-Component Storm Motion [m/s]
084	0-6000 m above ground	VSTM	V-Component Storm Motion [m/s]
085	0-1000 m above ground	VUCSH	Vertical U-Component Shear [1/s]

086	0-1000 m above ground	VVCSH	Vertical V-Component Shear [1/s]
087	0-6000 m above ground	VUCSH	Vertical U-Component Shear [1/s]
088	0-6000 m above ground	VVCSH	Vertical V-Component Shear [1/s]
089	180-0 mb above ground	4LFTX	Best (4 layer) Lifted Index [K]
090	180-0 mb above ground	CAPE	Convective Available Potential Energy [J/kg]
091	180-0 mb above ground	CIN	Convective Inhibition [J/kg]
092	surface	HPBL	Planetary Boundary Layer Height [m]
093	lifted condensation level	HGT	Geopotential Height [gpm]
094	90-0 mb above ground	CAPE	Convective Available Potential Energy [J/kg]
095	90-0 mb above ground	CIN	Convective Inhibition [J/kg]
096	255-0 mb above ground	CAPE	Convective Available Potential Energy [J/kg]
097	255-0 mb above ground	CIN	Convective Inhibition [J/kg]
098	equilibrium level	HGT	Geopotential Height [gpm]
099	255-0 mb above ground	PLPL	Pressure of level from which parcel was lifted [Pa]
100	surface	LAND	Land Cover (0=sea, 1=land) [Proportion]
101	surface	ICEC	Ice Cover [Proportion]
102	250 mb	UGRD	U-component of wind [m/s]
103	250 mb	VGRD	V-component of wind [m/s]
104	250 mb	HGT	Geopotential Height [gpm]
105	250 mb	TMP	Temperature [K]
106	700 mb	VVEL	Vertical Velocity [m/s]
107	-10 C	REFD	Reflectivity [dB]

108	-10 C	REFD	Hourly maximum of -10C reflectivity
109	5000-2000 m above ground	MNUPHL	Hrly Min Updraft Helicity - 2km to 5 km AGL [m^2/s^2]
110	2000-0 m above ground	MXUPHL	Hrly Max Updraft Helicity - 0km to 2 km AGL [m^2/s^2]
111	2000-0 m above ground	MNUPHL	Hrly Min Updraft Helicity - 0km to 2 km AGL [m^2/s^2]
112	3000-0 m above ground	MXUPHL	Hrly Max Updraft Helicity - 0km to 3 km AGL [m^2/s^2]
113	3000-0 m above ground	MNUPHL	Hrly Min Updraft Helicity - 0km to 3 km AGL [m^2/s^2]
114	2000-0 m above ground	RELV	Hrly Max Rel. Vort. – 0km to 2km AGL [1/s]
115	1000-0 m above ground	RELV	Hrly Max Rel. Vort. – 0km to 1km AGL [1/s]
116	entire column	HAIL	Hrly Max of Hail/Graupel Diameter [m]
117	0.1 sigma	HAIL	Hrly Max of Hail/Graupel Diameter [m]
118	5000-2000m AGL	UPHL	Updraft Helicity (instantaneous) [m^2/s^2]
119	6000-1000m AGL	UPHL	Updraft Helicity (instantaneous) [m^2/s^2]
120	top of atmos	SBT123	Simulated Brightness T for GOES 12 Ch. 3 [K]
121	top of atmos	SBT124	Simulated Brightness T for GOES 12 Ch. 4 [K]
122	top of atmos	SBT113	Simulated Brightness T for GOES 11 Ch. 3 [K]
123	top of atmos	SBT114	Simulated Brightness T for GOES 11 Ch. 4 [K]

Table 14. The set of six experimental hail diagnostics for the CLUE members contributed by CAPS, which are output at hourly intervals.

Number	Level/Layer	Parameter	Description
001	surface	HAIL1	Maximum hail size from HAILCAST [mm]
002	surface	HAIL2	Maximum hail size based on updraft speed [mm]
003	surface	HAIL3	Maximum hail size from machine learning method [mm]

b) High Resolution Ensemble Forecast (HREFv2) System

The HREFv2 is an 8-member CAM ensemble currently running at EMC that was implemented operationally on 1 November 2017 with forecasts that can be viewed at: http://www.spc.noaa.gov/exper/href/. Half of the membership of HREFv2 consists of 12-h time-lagged runs (Table 15). The design of HREFv2 closely follows that of the SSEO, which demonstrated skill during the previous six years in the HWT and SPC prior to HREFv2 operational implementation. All members, except for the NAM CONUS Nest, are initialized with a "cold-start". Forecasts to 36 h, including storm-attribute hourly maximum fields (HMFs), are produced at 0000 and 1200 UTC.

Table 15. HREFv2 specifications.

Members	**ICs/LBCs**	**Microphysics**	**Grid-spacing**	**PBL**
EMC HRW ARW2	NAM/NAM-6h	WSM6	3.2 km	MYJ
EMC HRW ARW2; -12h	NAM/NAM-6h	WSM6	3.2 km	MYJ
EMC HRW ARW	RAP/GFS-6h	WSM6	3.2 km	YSU
EMC HRW ARW; -12h	RAP/GFS-6h	WSM6	3.2 km	YSU
EMC HRW NMMB	RAP/GFS-6h	Ferrier-Aligo	3.2 km	MYJ
EMC HRW NMMB; -12h	RAP/GFS-6h	Ferrier-Aligo	3.2 km	MYJ
EMC NAM CONUS NEST	NAM/NAM	Ferrier-Aligo	3 km	MYJ
EMC CONUS NAM NEST; -12h	NAM/NAM	Ferrier-Aligo	3 km	MYJ

c) UK-Met Office convection allowing models (credit: Humphrey Lean)

The Met Office Unified Model (UM) is the name given to the suite of numerical modelling software used by the Met Office. One fully operational, nested limited-area high-resolution version of the UM at 2.2 km running twice per day will be supplied to SFE 2018. This operational nested hi-res version will incorporate the latest UM settings that are used over the UK.

The 2.2-km version has 70 vertical levels (spaced between 5m and 40 km) across a slightly sub-CONUS domain. The model takes its initial and lateral boundary conditions from the 0000 UTC and 1200 UTC runs of the 10-km horizontal grid-spacing global configuration of the UM and initializes without data assimilation running out to T+120. This model configuration uses a 3D turbulent mixing scheme using a locally scale-dependent blending of Smagorinsky and boundary layer mixing schemes, stochastic perturbations are made to the low-level resolved-scale temperature field in conditionally unstable regimes (to encourage the transition from sub-grid to resolved scale flows) and the microphysics is single moment. Partial cloudiness is diagnosed assuming a triangular moisture distribution with a width that is a universally specified function of height only. There is no convection parameterization in this UM configuration. The model will use the very latest UK configuration internally designated as Parallel Suite 41 (PS41), which has seen the RA1-M (Regional Atmosphere version 1 for Mid-latitudes) configuration implemented as part of this upgrade. This has been extensively tested with parallel running and is anticipated to become the UK operational model configuration in September 2018.

The RA configuration intends to provide standardized, portable, versions (mid-latitude and tropical versions) for use in other parts of the world on a longer (annual) development cycle than the internal, UK model. For RA1 the main emphases were on reducing the excessive precipitation bias especially in convective situations, improving convective initiation, re-tuning of boundary layer mixing in cumulus boundary layers and improving the diurnal modelling of temperature by using a more accurate representation of land surface properties. In addition, data from the Met Office global model will also be provided to allow for comparison against the 2.2 km to gain more insight into the source of the errors in the convective scale model. The global data is provided from the OS40 version of the Met Office global UM, currently running at a horizontal resolution of approximately 10 km in the mid-latitudes and using 70 vertical levels up to 80 km. It makes use of 4D-VAR hybrid data assimilation with the main scientific differences against the 2.2km being, use of a mass flux convection scheme (based on Graham Rowntree), a Prognostic Cloud Scheme (PC2), a purely 1D non-local boundary layer scheme and schemes for parametrizing the effects of gravity wave drag and sub-grid orographic drag.

d) ESRL High Resolution Rapid Refresh (HRRRv3) model (Credit: David Dowell and Curtis Alexander)

The 3-km grid-spacing HRRR model developed by the NOAA/Earth Systems Research Laboratory (ESRL) will continue to be examined in SFE 2018. Specifically, the ESRL experimental HRRR (HRRRv3) will be examined, as it is slated to become operational in June 2018. The HRRR uses GSI hybrid data assimilation (instead of 3D-VAR), is initialized with the latest 3-D radar reflectivity and features a WRF-ARW core version 3.9, Thompson microphysics, and is fully convection allowing. The background ensemble for this assimilation is the 80-member GDAS (GFS) ensemble. The HRRRv3 runs every hour on a 3-km grid with output to 18 h (01z, 02z, 04z, 05z, ….) or 36 hours (00z, 03z, 06z…). The HRRRv3 is initialized with an hour of 3-D radar reflectivity using a latent-heating specification technique including some refinements in this latent-heating from the parent RAPv4 model. The HRRRv3 uses grid-point statistical interpolation (GSI) hybrid GFS ensemble-variational data assimilation of conventional observations. Building upon the advancements in the operational HRRRv2 at NCEP, HRRRv3 includes assimilation of TAMDAR aircraft observations, refines assimilation of surface observations for improved lower-tropospheric temperature, dewpoint (humidity) winds and cloud base heights and places more weight on the ensemble contribution to the data assimilation. HRRRv3 also adds assimilation of lightning flash rates as a complement to radar reflectivity observations through a similar conversion to specified latent heating rates during a one-hour spin-up period in the model. Numerous model changes within the HRRRv3 include an update to WRF-ARW version 3.9 utilization of Thompson microphysics, transition to a hybrid sigma-pressure vertical coordinate for improved tropospheric temperature, dewpoint and wind forecasts along with a higher resolution (15 second) land use dataset. Physics enhancements have also been made to the MYNN PBL scheme and RUC land surface model along with additional refinements to shallow cumulus/sub-grid-scale cloud parameterizations including enhanced interactions with the radiation and microphysics schemes for greater retention of cloud features.

e) High Resolution Rapid Refresh Ensemble (HRRRE; credit: David Dowell)

In addition to the 0000 UTC initialized HRRRE runs that are part of the CLUE, HRRRE forecasts will also be provided at 1200, 1800, and 2100 UTC. These forecasts will be run across approximately the eastern 55% of the CONUS, and will be available to 18 h for the 1800 and the 2100 UTC runs. The 1200 UTC run will extend to 48 h across the same domain. These ensembles will be initialized from 3-km analyses in their data assimilation process rather than 15-km analyses, but are otherwise configured similarly to the 0000 UTC initialized HRRRE runs (Table 9).

f) NCAR Ensemble (credit: Glen Romine and David Dowell)

Additional NCAR ensemble runs will also be available at 1200 UTC and forecasts extending to 48 h. The HRRRE and NCAR ensembles being used in SFE 2018 have been developed in coordination, with shared features such as hourly cycling, a large outer analysis grid with 15-km grid spacing, and a nested grid with 3-km grid spacing. Forecast-model and data-assimilation codes will also be made as similar as possible. *One primary difference between the two systems is continuous cycling in the NCAR Ensemble versus once-daily partial cycling in the HRRRE.* This difference between the two systems will be analyzed to help determine best practices for future systems.

g) NSSL Experimental Warn-on-Forecast System for ensembles (NEWS-e)

The NSSL Experimental Warn-on-Forecast System for ensembles (NEWS-e) is a 36-member WRF-based ensemble data assimilation system used to produce very short-range (0-6 h) probabilistic 18-member forecasts of supercell thunderstorm rotation, hail, high winds, and flash flooding. The starting point for each day's experiment will be the experimental HRRRE (Table 16) provided by ESRL/GSD. The full ensemble is updated by hourly EnKF assimilation of conventional observations and Multi-Radar/MultiSensor (MRMS) radar reflectivity from 0300 UTC to 1800 UTC Day 1. A 48-h ensemble forecast launched from the 1200 UTC HRRRE analysis is used to provide boundary conditions for the NEWS-e system for the period 1800 UTC Day 1 – 0300 UTC Day 2. Similarly, a 1-h ensemble forecast launched from the 1700 UTC HRRRE analysis is used to provide initial conditions for the NEWS-e at 1800 UTC.

The daily NEWS-e domain location will target the primary region where severe weather is anticipated and cover a 750-km wide region with very frequent 15-min updates. All ensemble members utilize the NSSL 2-moment microphysics parameterization and the RAP land-surface model, but the PBL and radiation physics options are varied amongst the ensemble members to address uncertainties in model physics. Multi-Radar/Multi-Sensor (MRMS) radar reflectivity and Level II radial velocity data, cloud water path retrievals from the GOES-16 imager, and Oklahoma Mesonet observations (when available) will be assimilated every 15 min using an EnKF approach, beginning at 1800 UTC each day. ASOS data will also be assimilated at 15 minutes past each hour. A 6-h (5-h) ensemble forecast will be initialized from the 1900 (2000) UTC NEWS-e analysis for HWT product evaluation from 2000 – 2100 UTC. Beginning at 2030 UTC, a 180-min ensemble forecast with 5-min output will be launched every 30 minutes through 0300 UTC the next day. These forecasts will be viewable using the web-based NEWS-e Forecast Viewer (https://www.nssl.noaa.gov/projects/wof/news-e/realtime/). Table 15 shows the differences in model specifications between HRRRE and NEWS-e, and Figure 5 shows an example of a SPC Day 1 convective outlook and corresponding NEWS-e grid with WSR-88D radars used for data assimilation overlaid.

Table 16. HRRRE and NEWS-e configuration comparison.

	HRRRE	NEWS-e
Model Version	WRF-ARW v3.8+	WRF-ARW v3.8+
Grid Points	1150 x 960 x 50	251 x 251 x 50
Grid Spacing	3 km	3 km
EnKF Cycling	36 mem. w/ GSI-DART or GSI-EnKF **every 1 hr**	36 mem. w/ DART **every 15 min**
Observations	**-Conventional obs: T, q_v, u, v, and p from rawinsonde, aircraft, surface (land and marine), profiler;** **-MRMS radar reflectivity**	**-Surface: ASOS, Oklahoma Mesonet (when available);** **-Doppler velocity from ~20-25 WSR-88D sites;** **-MRMS reflectivity > 20 dBZ; radar 'zeroes';** **-Cloud-water path (GOES-16)**
Radiation LW/SW	RRTMG/RRTMG	Dudhia/RRTM, RRTMG/RRTMG
Microphysics	**Thompson (aerosol aware)**	**NSSL 2-moment**
PBL	**MYNN**	**YSU, MYJ, or MYNN**
LSM	RUC (Smirnova)	RUC (Smirnova)

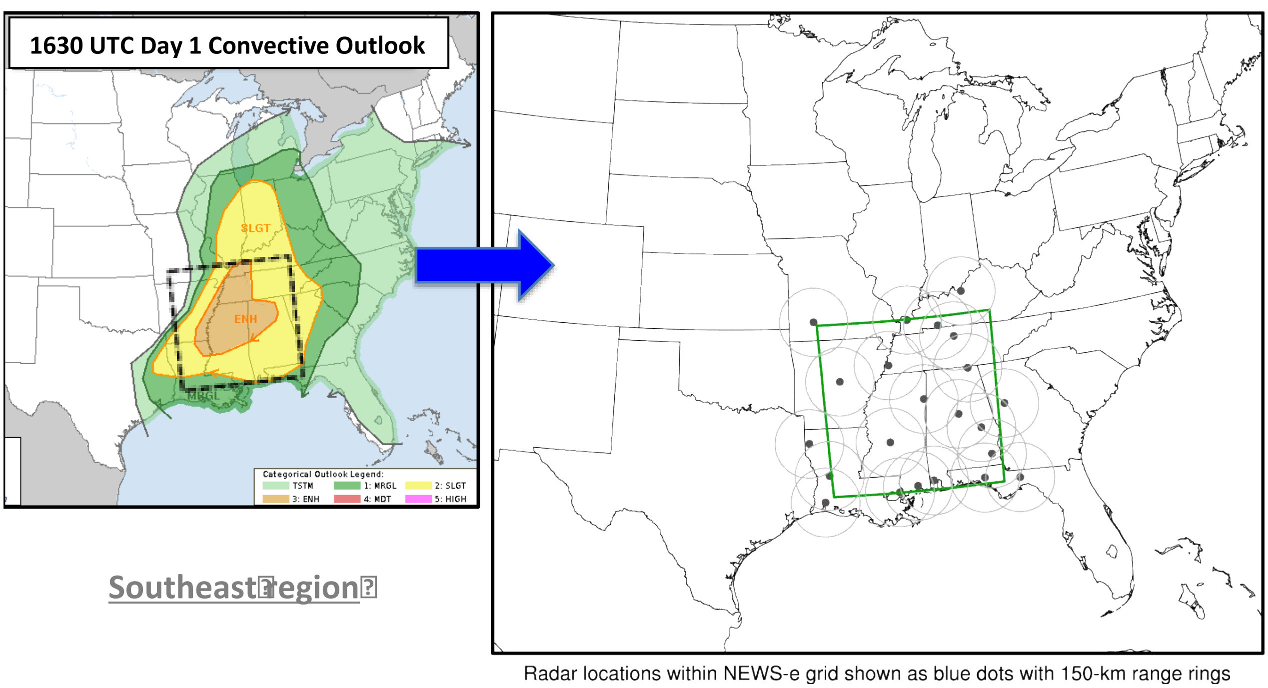

Figure 5. SPC 1630 UTC issued Day 1 convective outlook (left) and corresponding NEWS-e grid (right).

3. SFE 2018 Core Interests/Daily Activities

a. Forecast products and activities

Similar to previous years, the experimental forecasts this year will focus on our ability to add temporal specificity to longer-term convective outlooks. As such, all of the forecast activities this year will focus on periods within the Day 1 time frame. We will continue to split the participants into two desks, with those at the Innovation Desk forecasting total severe threat (combining hail, wind, and tornado hazards) and those at the Severe Hazards Desk forecasting individual severe hazards. For the Severe Hazards Desk, the first forecast will mimic the SPC operational Day 1 Convective Outlooks by producing individual probabilistic forecasts of large hail, damaging wind, and tornadoes within 25 miles (40 km) of a point valid 1600 UTC to 1200 UTC the next day. The first forecast on the Innovation Desk will also cover the 1600 to 1200 UTC time period, but cover total severe (combined tornado, hail and wind). The experimental forecasts cover a limited-area domain typically covering the primary severe threat area with a center-point selected based on existing SPC outlooks and/or where interesting convective forecast challenges are expected.

Each desk will then manually stratify the experimental Day 1 outlooks into periods with higher temporal resolution. The Severe Hazards Desk will generate separate probability forecasts of large hail, damaging wind, and/or tornadoes for two 4-h periods: 1700–2100 UTC and 2100–0100 UTC. As an alternative way of stratifying the Day 1 outlook, the Innovation Desk will create a product aimed toward the emergency management community, designating areas and 4-h periods where severe convective hazard occurrence is expected throughout the day. These potential severe timing (PST) areas will occur within areas of 15% probability as indicated by the Day 1 full-period outlook previously generated by the Innovation Desk. This approach builds upon the isochrones approach taken during SFE 2016 and SFE 2017, with the timing and areal information both available on the final figure. Despite the different end products, the goals of the activities are the same as in prior years – namely to explore different ways of introducing probabilistic severe weather forecasts on time/space scales that are not currently addressed with categorical forecast products (e.g., SPC Mesoscale Discussions and Severe Thunderstorm/Tornado Watches), and to begin to explore ways of seamlessly bridging probabilistic severe weather outlooks and probabilistic severe weather warnings as part of the NOAA WoF and FACETS initiatives.

During previous experiments, calibrated probabilistic severe guidance from the SREF/SSEO (Jirak et al. 2014) was used to temporally disaggregate a 1600–1200 UTC period human forecast. A scaling factor was formulated by matching the full-period calibrated severe SREF/SSEO guidance to the human forecast, and then this scaling factor (unique at every grid point) was applied to the calibrated severe guidance for each individual period. Finally, consistency checks were conducted to arrive at the final temporally disaggregated forecasts (Jirak et al. 2012). These automated forecasts from SFE 2012 to SFE 2017 fared favorably both in terms of objective metrics (e.g., CSI, FSS) and subjective impressions when compared to manually drawn forecasts. Similarly for SFE 2018, the 1600–1200 UTC human forecasts for the individual hazards will be temporally disaggregated into the 4-h periods (1700–2100 UTC and 2100–0100 UTC) using HREF/SREF calibrated hazard guidance to provide a first guess for the two forecast periods.

The first set of short-time-window forecasts and timing forecasts will be issued in the morning by both desks. At both the Severe Hazards Desk and the Innovation Desk, the lead forecaster will generate the short-time-window forecasts on the N-AWIPS machines. However, the participants will split into five groups for each desk and use a web interface to generate their own short-time window probability forecasts using Google Chromebooks. The redesigned web interface is similar to the Probabilistic Hazard Information (PHI) tool used in past experiments, but has been specifically designed to incorporate data from CLUE subsets and other experimental CAM ensemble guidance. Each Chromebook will be associated with a specific ensemble or CLUE subset; participants will be asked to rely on that ensemble or CLUE subset to guide their forecast generation. Additionally, the web interface will have other important observational and model fields for participants to utilize in the forecast generation process. After issuing the high-temporal resolution individual forecasts based on model subsets and reporting the anonymous demographics of the group (e.g., how much forecasting experience the participants in each group have), the desks will regroup and discuss the forecasts and behavior of the CLUE subsets. This approach is planned to engage the participants more directly with the CLUE subsets, since in prior years participants only interacted with CLUE subsets through facilitator-led discussions. After the teams issue and discuss the high-temporal resolution forecasts with their Desk, there will be a map discussion summarizing forecast challenges and highlighting interesting findings from the previous day open to all tenants of the National Weather Center. Each day of the week will also feature a brief discussion of a special topic, which can be found on the daily schedule of activities (Table 17).

After lunch, the Innovation Desk will update their higher temporal resolution forecasts. Teams will once more examine operational guidance as a group. Of the five ensemble subsets, three will update at 1200 UTC with new information, allowing testing of the impact of updated CAM ensemble guidance on timing forecasts of severe weather. Participants using the non-updating subsets will have to update their forecasts based solely on observations and updated deterministic CAM information (e.g., more recent HRRR runs). Since the forecast process for these updates will begin in the afternoon, participants will be instructed to only update their ESPs valid between 1900 UTC and 1200 UTC. Forecasters at the Severe Hazards Desk will follow a similar process, but generate a new forecast valid from 1900–2300 UTC using their CAM ensemble subsets.

Later in the afternoon, scientific evaluations will take place (summarized in the next section). For the final activity of the day on Tuesday through Friday, forecasting activities using the WoF-prototype system, NEWS-e, will be conducted on both desks. For the Innovation Desk activity, the 1900 UTC initialized NEWS-e with 6-h forecast products available at the website https://www.nssl.noaa.gov/projects/wof/news-e/realtime/ will be used to issue two 1-h time window forecasts of total severe valid 2100-2200 and 2200-2300 (i.e., 4–5PM and 5–6 PM CDT). Then, these forecasts will be updated using 2000 UTC initialized NEWS-e products. Forecasts will be drawn by facilitators (Adam Clark and Burkely Gallo) and informed by small groups of participants interrogating NEWS-e data on their Chromebooks, as well as by the forecast lead (Jack Hales). On the Severe Hazards Desk, participants will use the NEWS-e data to update their probabilistic hail, wind, and tornado forecasts valid from 2100–0100 UTC.

To prepare participants for the NEWS-e activity, a training session will be provided 3–4pm on Monday of each week (in NWC 5930). This training session will give a description of NEWS-e and provide an overview of how to navigate the NEWS-e website and view forecast products. Participants will be asked to use a Google Chromebook or their personal laptop during this session. Following the presentation portion of the training, facilitators will work with smaller groups (of ~5 participants) and walk through a test case to become familiar with the process of the NEWs-e activity. After practicing the issuance of a 1-hour outlook and update, participants will be asked to view, answer, and ask for clarification on a set of survey questions that will be completed following each NEWS-e activity session (these questions should take less than 10 min to complete).

Questions will be available in a Google survey form, and will consist of multiple-choice, ranking, and open-ended questions designed to capture participants' perceptions of the NEWS-e products specific to the forecast challenge presented in the activity. Finally, participants will be made aware of additional NEWS-e specific survey questions that will be asked during the verification evaluation activity scheduled first thing on Tue-Fri mornings. These questions will be appended to the Google survey form that will be used for the verification evaluation activity that will evaluate all experimental forecasts made the previous day.

The training session will also be used to obtain participants' consent (per IRB protocol) to take part in this activity and answer survey questions. Additionally, we will ask participants to provide their subjective rating of forecasting experience on a scale of 1–3 (none/minimal, some, and extensive). Participants will be given examples of what these different rating levels mean. The ratings will be used to assign participants each day to either group 1 or group 2 of the activity. The purpose of this group assignment is to try and ensure a balance of forecasting experience for each of the outlooks that are issued, as well as to encourage discussion between participants of varying professional backgrounds (i.e., operational and research oriented).

A summary of the SFE 2018 daily activities schedule is shown below in Table 17.

Table 17. Summary of SFE 2018 Daily Schedule.

<table>
<tr><th>Severe Hazards Desk</th><th>Innovation Desk</th></tr>
<tr><td colspan="2">0800 – 0845: Evaluation of Experimental Forecasts & Guidance
Subjective rating relative to radar evolution/characteristics, warnings, preliminary reports, and MRMS MESH and rotation tracks</td></tr>
<tr><td>• Day 1 full-period probabilistic forecasts of tornado, wind, and hail
• Day 1 4-h period forecasts and guidance for tornado, wind, and hail</td><td>• Days 1 full-period probabilistic forecast of total severe
• Day 1 4-h areas for severe weather timing
• Day 1 1-h total severe outlooks</td></tr>
<tr><td colspan="2">0845 – 0915: Map Analysis
Hand analysis of 12Z upper-air and surface maps, discussion, and domain selection (from two areas)</td></tr>
<tr><td colspan="2">0915 – 1130: Convective Outlook Generation</td></tr>
<tr><td>• Day 1 full-period probabilistic forecasts of tornado, wind, and hail valid 16-12Z over mesoscale area of interest
• Day 1 4-h probabilistic forecasts of tornado, wind, and hail valid 17-21Z and 21-01Z using CLUE subsets*</td><td>• Day 1 full-period probabilistic forecast of total severe valid 16-12Z over mesoscale area of interest
• Day 1 4-h timing areas (16-12Z) for full-period total severe ≥15% using CLUE subsets*</td></tr>
<tr><td colspan="2">1130 – 1200: Map Discussion
Brief discussion of today’s forecast challenges and products
Topic of the day: 3D Vis, Met Office, FV3, NEWS-e, CAM scorecard</td></tr>
<tr><td colspan="2">1200 – 1300: Lunch</td></tr>
<tr><td colspan="2">1300 – 1345: Convective Outlook Generation</td></tr>
<tr><td>• Day 1 4-h probabilistic forecasts of tornado, wind, and hail valid 19-23Z using 12Z CAM ensembles*</td><td>• Update Day 1 4-h timing areas (19-12Z) for full-period total severe ≥15% using 12Z CAM ensembles*</td></tr>
<tr><td colspan="2">1345 – 1500: Scientific Evaluations</td></tr>
<tr><td>• HREF Configurations
• CLUE: HRRRE
• Hail Guidance
• Deterministic CAMs (FV3, UM, HRRR)
• TTU Sensitivity-Based Ensemble Subsetting</td><td>• CLUE: Physics Experiment
• CLUE: FV3 Physics
• Met Office UM Evaluation
• CLUE: Microphysics
• Ensemble Object-Based Visualization</td></tr>
<tr><td colspan="2">1500 – 1600: Short-term Outlook Update</td></tr>
<tr><td>• Update 4-h probabilistic forecasts of tornado, wind, and hail valid 21-01Z using SPC Short-Term Hazard Guidance and NEWS-e*</td><td>• Utilize NEWS-e to generate preliminary and final hourly probabilistic forecasts of total severe valid 21-22 and 22-23*</td></tr>
<tr><td colspan="2">* Denotes forecasts also made by participants using the web drawing tool on Chromebooks.</td></tr>
</table>

b. Formal Evaluation Activities

There will be two periods of formal evaluations during SFE 2018. The first will occur during the morning on Tuesday through Friday, when experimental outlooks from the previous day generated by both forecast teams will be examined. In these next-day evaluations, the team forecasts and the first-guess guidance will be compared to observed radar reflectivity, reports of severe weather (LSRs), NWS warnings, and Multi-Radar Multi-Sensor (MRMS) radar-estimated hail sizes over the same time periods. Both raw LSRs and "practically perfect" fields (Hitchens et al. 2013) will be used. The SFE participants will provide their subjective evaluations of the strengths and weaknesses of each of the forecasts. This evaluation will include examining and comparing calibrated guidance, temporal disaggregation first guess, and human initial and updated forecasts. The goal is to determine the relative skill of the first-guess guidance and the human-generated forecasts over all periods, in part to assess the feasibility of issuing operational high-temporal resolution severe weather forecasts. Objective verification metrics will also be computed for some of the experimental outlooks and first-guess guidance.

The afternoon evaluation period will involve comparisons of different ensemble diagnostics and CLUE ensemble subsets. The Innovation and Severe Hazards Desks will conduct two different sets of evaluations.

Evaluations at the Innovation Desk

(1) Single vs. Mixed vs. Stochastic Physics Ensembles

This evaluation activity will compare various ensemble products (e.g., probability, mean, maximum from any member) from the single physics, mixed physics, and stochastic physics ensembles that are a part of the CLUE. The ensembles will be assigned ratings based on their perceived skill, and participants will be encouraged to provide comments on noticeable differences or systematic biases observed between the ensemble subsets.

(2) FV3 Physics Evaluation

The contribution of eleven different FV3 runs by CAPS allows for testing of the forecast sensitivity to different parameterization schemes. Two microphysics schemes and five PBL schemes will be compared in this evaluation via examination of CAM output such as reflectivity and updraft helicity. However, environmental fields such as temperature, dewpoint, and CAPE will also be examined. In addition to field forecasts, soundings from the FV3 ensemble subsets will also be compared amongst the ensemble subsets and with observations at sounding sites. Features of particular interest include position of the dryline and strength and sharpness of the capping inversion.

(3) Met Office UM Evaluation

The UK Met Office is providing both the global UM and a high-resolution UM configuration to the HWT for SFE 2018, to gather a dataset for diagnosing forecast error source. Comparisons will occur between the rain rate and reflectivity fields from the global and high-resolution model, respectively. Participants will also rate the degree of differences between the two configurations, ranging from "little to no difference" to "major differences in magnitudes and/or placement of important features". This will be repeated with several other environmental fields. A sounding comparison element will also take place during this evaluation, with a particular focus on convective parameters such as CIN.

(4) Microphysics Comparison

Since 2010, one component of model evaluation activities during annual SFEs has involved subjectively examining sensitivity to microphysics parameterizations used in the WRF model by comparing various forecast fields including simulated reflectivity, simulated brightness temperature, low-level temperature and moisture, and instability for a set of ensemble members with identical configurations except for their microphysical parameterization. During SFE 2018, this evaluation will focus on a new implementation of stochastic processes within the Thompson-Eidhammer microphysics scheme (Thompson and Eidhammer 2014). Along with the aforementioned fields being compared, simulated satellite imagery is also planned for this comparison.

(5) Ensemble Object-Based Visualization (Credit: Aaron Johnson)

A new object-based approach to ensemble probability forecasts has been developed and is being introduced into HWT activities this year for the first time. Potentially important convective scale details, such as storm mode and morphology, can be lost when generating grid point or neighborhood based ensemble mean or probability plots. Manually evaluating such details in each available ensemble member can be time consuming, and even then might be difficult to synthesize into actionable quantitative guidance. The OBPROB product is intended to mimic forecaster evaluations of the ensemble predicted distribution of storm modes and morphologies in an automated and quantitative manner. The goals of introducing this product into the HWT are to (1) determine the usefulness of the OBPROBs in the operational severe weather forecasting process, and (2) tune the technique, if needed, to properly reflect how the CAM ensembles are used by forecasters so that ensemble verification research can use the technique to determine ensemble design improvements that have the greatest potential to translate into improved operational convective outlooks and watches.

To identify objects, a 2*dx (6 km) Gaussian convolution is first applied to composite reflectivity forecasts to remove grid scale noise, and a threshold of 35 dBZ is applied to the convolved field to identify a discrete set of storm objects. For each object, the attributes defined in Table 18 are calculated if the area of the object is at least 7*7=49 grid points (i.e. above the model's effective resolution of ~7*dx).

Table 18. Description of object attributes

Attribute Name	Description
Centroid Location	Center of "mass" of the binary field of the object.
Area	Number of grid points contained in the object.
Aspect Ratio	W/L, where L=length of longest axis, W=length of axis perpendicular to L.
UH	90^{th} percentile of updraft helicity values within the object
10mspd	As in UH, except for 10m wind speed values
HAIL	As in UH, except for hail size values
PRECIP	As in UH, except for hourly accumulated precipitation

A Total Interest (*I*) value is calculated to determine the overall similarity of any pair of objects (e.g., objects from different ensemble members). For simplicity, and ease of physical interpretation, we here use

$$I = f_{a_1} * f_{a_2} * f_{a_3}$$

where f_{a_i} is the similarity function in terms of each of Centroid Location difference, Area ratio, and Aspect Ratio difference. These attributes quantify the similarity of the storm objects' location, size and shape according to specific interest thresholds (Figure 6).

The probability that each forecast object will be matched by a similar observed object is taken to be the percentage of ensemble members with a matching object. In order to include the possibilities from the entire ensemble "envelope", the OBPROB product is constructed as follows:

Step 1: Calculate probability for every object from every ensemble member.

Step 2: Sort the objects based on probability.

Step 3: Add the highest probability object to the plot, shaded by probability.

Step 4: Remove this object, and all matching objects that contributed to its probability, from the list.

Step 5: Repeat from step 1 until the list is empty.

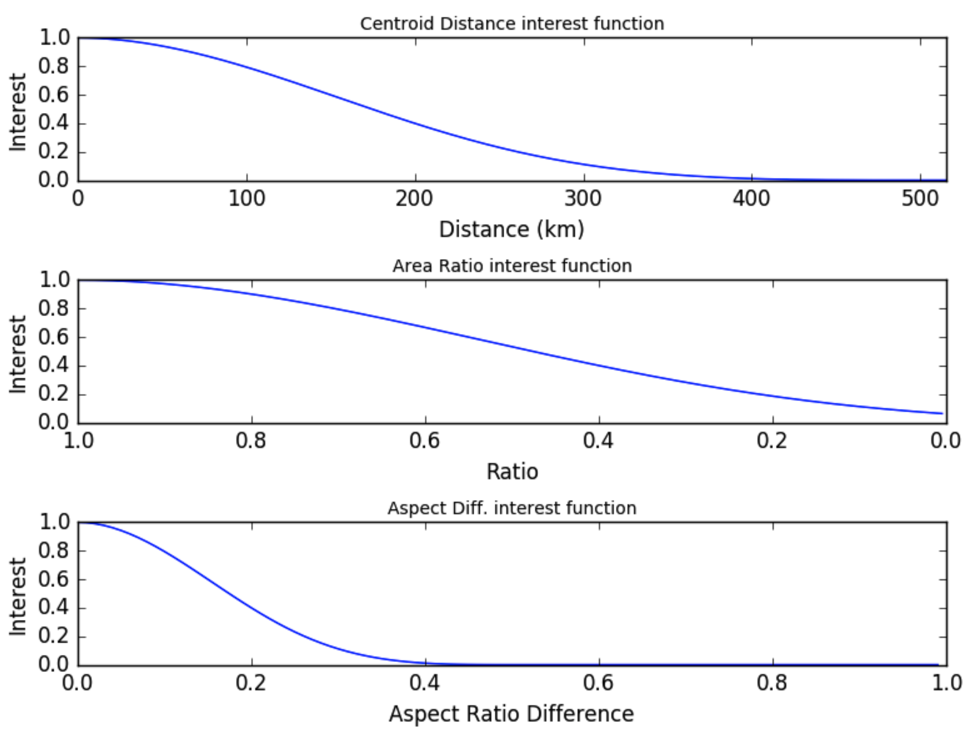

Figure 6. Interest functions to define similarity (i.e., interest) of the centroid location, area and aspect ratio attributes between two objects.

The resulting OBPROB product is conceptually similar to a paintball plot since every storm object is included. The difference is that many similar storms from different ensemble members can be represented by a single high-probability object. This makes the plot less busy, and easier to interpret. As an example, the ensemble forecast in the Fig. 7 could be interpreted to indicate a much higher probability that the storm(s) in

Oklahoma will still have a cellular mode at this time, although there is a slight chance of upscale growth having occurred. Similarly, while both linear and cellular modes are likely in central Kansas into southern Nebraska, the linear systems have greater ensemble agreement. This information could be gathered qualitatively from the ensemble by evaluating each member individually or quantitatively from the OBPROB product (Fig. 7k).

The OBPROB products for the OUMAP ensemble can be viewed at http://weather.ou.edu/~map/dev/interface.php where forecasters will also have the option of experimenting with the e-folding distance of the Gaspari and Cohn functions shown in Fig. 6. Furthermore, the forecaster can focus on storm mode as demonstrated here, or can focus on severe objects only, high wind producing objects only, large hail producing objects only, or heavy rain producing objects only. When these options are selected on the web interface, a minimum threshold of 100 $m^2 s^{-2}$, 20 $m s^{-1}$, 2.54 mm, or 12.7 mm h^{-1}, respectively, is enforced when identifying objects.

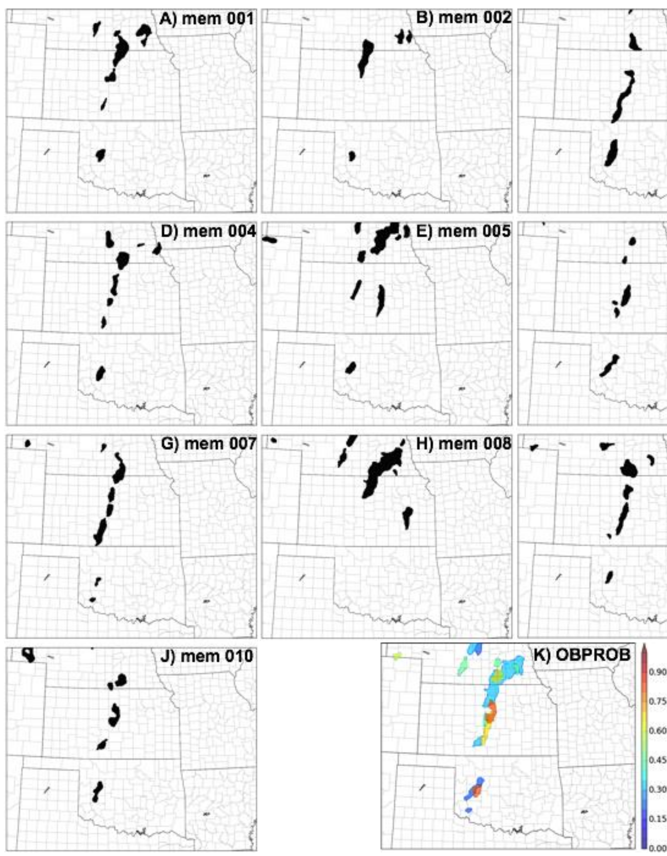

Figure 7. Objects forecast by the (a)-(j) each ensemble member, and (k) the OBPROB product. Forecast was initialized at 0000 UTC 16 May 2017, and valid at 0100 UTC 17 May 2017.

Evaluations at the Severe Hazards Desk

(1) HREF Configurations

Various ensemble subsets from the CLUE will be compared to HREFv2, which is essentially an operational version of the SSEO. This experiment should highlight the current status of ensemble development and inform the design of subsequent configurations for operational convection-allowing ensembles. Various configurations of the HREF will be evaluated, including an HREF which includes extended-length HRRRv3 forecasts and an HREF without time-lagged members. The objective component of this evaluation will focus on ensemble neighborhood probability forecasts of simulated reflectivity compared to observed radar reflectivity while the subjective component will examine ensemble forecasts (ensemble maximum and neighborhood probabilities) of HMFs of UH relative to LSRs of hail, wind, and tornadoes. As in prior years, the evaluation will be facilitated via webpages using spatial plots for distinct time frames as well as a table, which summarizes statistical results for some ensemble subsets. For these statistical analyses, time-series plots will also be constructed for a more graphical representation of trends in the scores. This approach has proved effective since introduced during SFE 2012 (Melick et al. 2012) for the participants to quickly evaluate the forecast verification metrics and provide feedback on the comparison between the objective results and their own subjective impressions.

(2) HRRRE Comparisons

This comparison encompasses multiple elements. One aspect focuses on the three ensembles within the CLUE that are configured with EnKF or hybrid data assimilation: the HRRRE, the NCAR ensemble, and the MAP hybrid system. A second aspect compares the 1200 UTC cycle of these three systems to the time-lagged HRRR ensemble (HRRR-TLE). All of the prior ensembles will be compared to the operational HREFv2, which was shown in SFE 2017 to be skillful and will be used as a baseline for this comparison.

(3) Hail Guidance

Similar to the 2014–17 SFEs, there is interest in evaluating the ability of CAMs to predict hail size because of the need to forecast individual thunderstorm hazards, which are included in the SPC operational Day 1 Convective Outlooks. Thus, for the fifth year, the most recent version of the HAILCAST algorithm implemented in ARW will be used to predict hail size (Adams-Selin and Ziegler 2016), which is based on the algorithm in Brimelow (2002) and Jewell and Brimelow (2009). The HAILCAST model uses convective cloud and updraft attributes to determine the growth of hail from initial embryos. Additionally, a hail size diagnostic derived directly from the microphysics parameterizations will be examined, which was implemented by Greg Thompson of NCAR. A method based on the updraft speed developed by Nathan Wendt of CIMMS/SPC will also be examined. Finally, probabilistic hail size forecasts derived from a machine-learning algorithm developed by David Gagne of NCAR and Nate Snook of OU will be examined. These forecasts are derived as follows: a hail size forecasting model generates hail size forecasts from a combination of image processing and machine learning models. Potential hailstorms are identified and tracked in the modeled hourly-max updraft field and the MRMS MESH product using the enhanced watershed object identification method. Model and observed tracks are matched using a multidimensional distance function. The distribution of MESH values within an object is approximated by a gamma distribution. The gamma distribution parameters are simultaneously predicted by a single random forest machine learning model given information about each model storm and its environment. Another random forest model predicts whether or not an observed hailstorm will occur in the vicinity of the forecast hailstorm, which determines whether or not a particular hail object is kept. Finally, hail sizes are sampled from each predicted gamma distribution and are applied in rank order to each grid point within each hailstorm object. This process is repeated for each ensemble member, and ensemble neighborhood probabilities and other derived products can be produced with this information.

As part of the evaluation activity, the utility of probabilistic hail size forecasts using all four methods will be compared to neighborhood probabilities of UH to determine the value added by these specific hail-size prediction methods. The predictions of hail size will be evaluated against storm reports and MRMS MESH.

(4) Deterministic CAMs (e.g., FV3, UM)

This activity will focus on assigning ratings to gauge the skill and utility of several deterministic CAMs. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms. Storm surrogate fields, like hourly maximum updraft helicity, will also be examined to gauge their utility for forecasting severe storms. One of the main purposes of this evaluation will be to quantify how well the three versions of FV3 (*caps-fv3, nssl-fv3, and gfdl-fv3*) perform compared to other CAMs with well-known performance characteristics.

(5) TTU Sensitivity-Based Ensemble Subsetting (Credit: Brian Ancell)

Ensemble sensitivity is a statistical technique applied within an ensemble that identifies features in the flow at early forecast times that are related to the predictability of chosen severe storm characteristics later in the forecast. In other words, ensemble sensitivity reveals the flow features for which associated errors will grow rapidly to adversely affect the predictive skill of chosen severe storm aspects. It can thus be expected that ensemble members that have the least error in the most sensitive regions will provide better forecasts than other members, allowing the generation of adjusted and improved probabilities based on ensemble subsets. *The ultimate goal of this HWT 2018 evaluation is to understand whether ensemble sensitivity-based subset probabilities provide any forecast skill over that of the full ensemble*. This technique has been successfully tested for synoptic-scale events such as mid-latitude cyclones (showing improvement by the subset), but needs to be evaluated at convective scales in a real-time environment where nonlinearity and physics errors, two characteristics that may degrade the sensitivity-based method, are more prominent. The planned activity follows an evaluation at SFE 2016 of ensemble sensitivity fields alone that showed day-to-day coherent early forecast-time sensitivity signals at 500 hPa and above for severe convection. Feedback from that evaluation supported the use of ensemble sensitivity "behind the scenes" given the operational difficulty of interpreting sensitivity fields. The activity planned for SFE 2018 is in direct response to that feedback.

Texas Tech University personnel will conduct a daily evaluation of probabilities from a full 42-member ensemble against those based on 10-member ensemble subsets chosen objectively through the use of ensemble sensitivity information. Each day, a response function location and time will be chosen through a web-based graphical user interface that identifies areas of Day 1 severe convection within that day's 0000 UTC Texas Tech ensemble run. The Day 1 response function will be chosen over a 6-hr period between 1800 UTC and 1200 UTC (next day) in areas where better predictions of severe convection are desired (e.g. areas of high uncertainty). Once the response time and location are chosen, the sensitivity of three independent response functions will be calculated: 1) maximum 2–5km updraft helicity, 2) number of grid points exceeding 25 m^2/s^2

2–5km updraft helicity, and 3) number of grid points exceeding 40 dBZ near-surface simulated reflectivity. The sensitivities of the three response functions will be calculated with respect to 300- and 500-hPa temperature, winds, and geopotential height, and 700-hPa temperature all with respect to the 6-hr forecast (valid 0600 UTC). Ensemble members from the 0000 UTC run will then be chosen objectively based on their errors in the most sensitive regions (greatest 50% of sensitivity magnitudes) using the 0600 UTC analysis from either the 0600 UTC Texas Tech ensemble initialization or the Rapid Refresh system.

Probability fields (specifically maximum 6-hourly 20-mile neighborhood exceedance probabilities of 25 m^2/s^2 2–5km updraft helicity and 40 dBZ simulated near-surface reflectivity) of Day 1 convection will be generated for the ensemble subset and will be compared against full ensemble probabilities the following day after the severe event has occurred (Figure 8c,d; differences between the full and subset probabilities will also be displayed). 6-hourly paintball plots of large maximum hourly 2–5km updraft helicity and simulated reflectivity will also be evaluated for both the full and subset ensemble. SPC storm reports will serve as the observations against which both the full and subset ensemble probabilities and paintball plots are judged. The Texas Tech 42-member ensemble system within which the fully automated subsetting technique will run is a

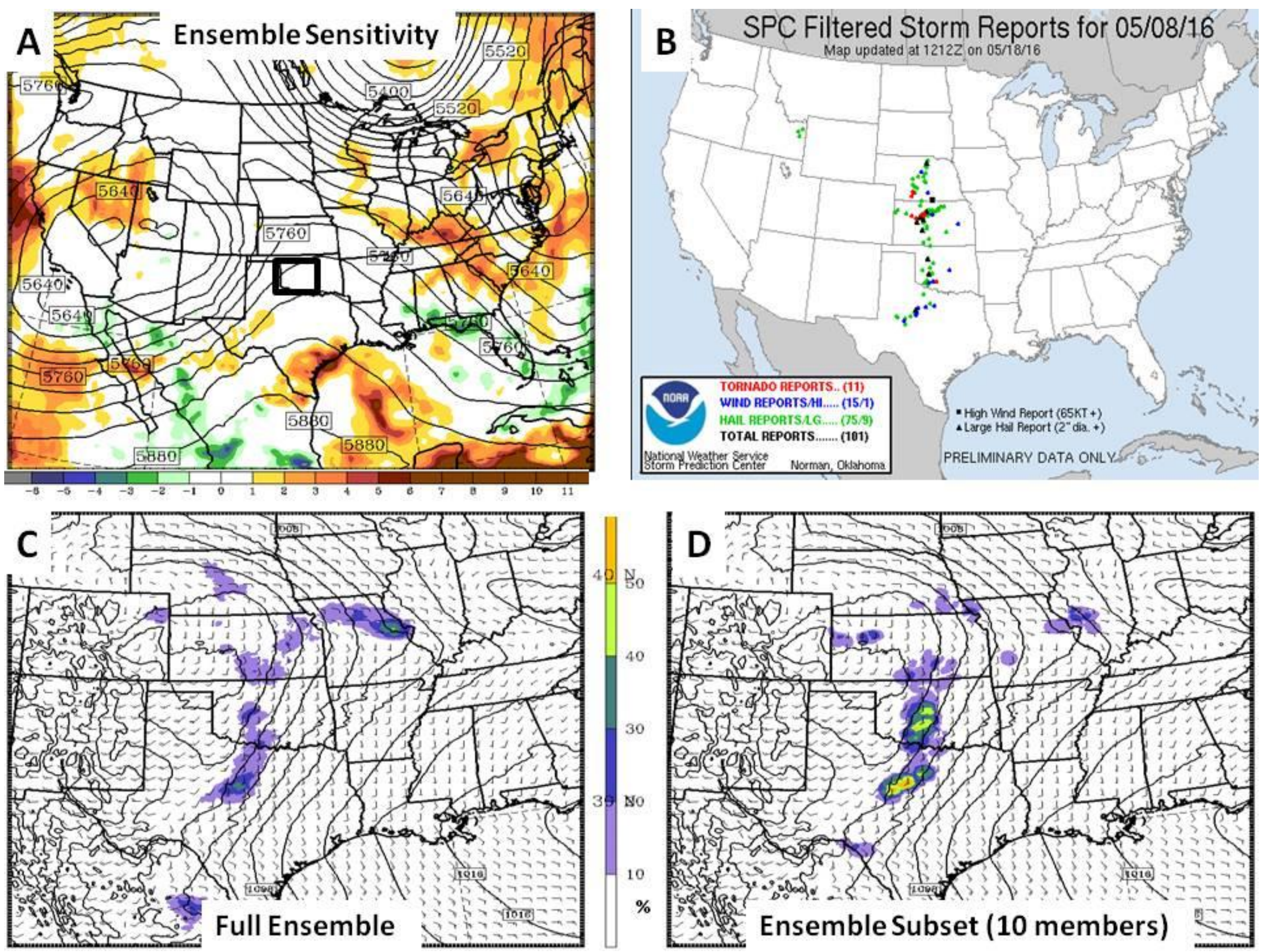

Figure 8. Example sensitivity-based subsettting case for improving severe convection in Oklahoma (verifying storm reports shown in panel B). Panel A shows the sensitivity of maximum hourly 35-hr 2-5km updraft helicity in the black box to 6-hr 500-hPa geopotential height used to choose the ensemble subset (RAP analysis used as truth). Panel C shows the full ensemble neighborhood probability of exceeding 25 m^2/s^2 updraft helicity, and panel D the 10-member subset probabilities, which revealed significantly higher chances of large updraft helicity in the response function area where severe reports occurred.

DART WRF ensemble Kalman filter that assimilates numerous surface and upper-air observations on a 6-hr assimilation cycle. Assimilation is performed over a 12-km CONUS domain with downscaled 48-hr WRF forecasts run twice daily on a 4-km domain across the U.S. Midwest and South Plains. Real-time output from the Texas Tech ensemble can be viewed at http://www.atmo.ttu.edu/bancell/real_time_ENS/ttuenshome.php.

c. Other specialized activities

CAM output in three-dimensional (3D) displays will be presented again in real-time. CAPS will provide selected 3D model fields over a daily mesoscale region of interest at 10-min output frequency for 18 – 30 h forecasts to interrogate using the VAPOR software. The goal is to explore CAM storm characteristics like vertical vorticity, graupel mixing ratio, simulated reflectivity, and cold pools in 3D to learn more about how simulated storms are structured on convection-allowing grids (see Figure 9 for an example display).We will also examine characteristics of the storm environments in CAM forecasts like depth of water vapor mixing ratio in the PBL and depictions of low-level convergence boundaries and how they may play a role in the initiation of convection in the model.

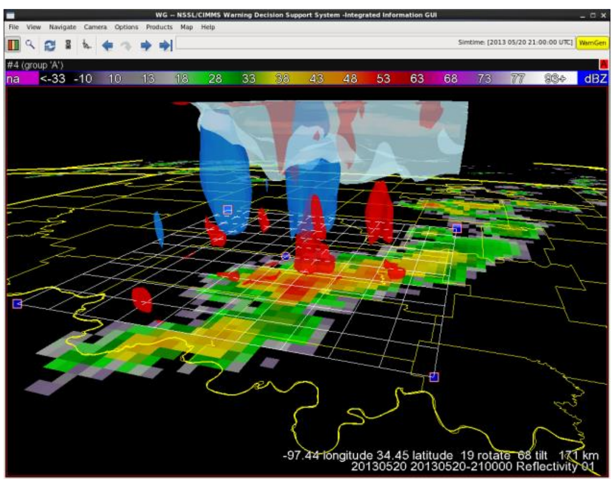

Figure 9. Example of how CAM forecasts will be interrogated for a select few runs from the CAPS SSEF system. The 2D field is the simulated reflectivity on the lowest model level (color scale near the top of the image) with 3D isosurfaces of vertical velocity x vertical vorticity ($w*\zeta$) > 2 m s^{-2} (red areas), graupel mixing ratio > 5 g kg^{-1} (dark blue areas), and snow mixing ratio > 2 g kg^{-1} shown within the box outlined in white.

Finally, blog posts will be made 2-3 times weekly during the experiment, as in the past two years. These blog posts will highlight interesting case studies, preliminary results from evaluations, and feature activities and discussions taking place during the SFEs. The blog is hoped to supplement the formal results produced during the 2018 SFE by providing a more informal look into the questions and discussions that take place within the framework of the formal forecasting and evaluation activities. The blog can be found at: https://blog.nssl.noaa.gov/efp/.

Appendix A: List of scheduled SFE 2018 participants. Facilitators/leaders for SFE 2018 include: Adam Clark (NSSL), Israel Jirak (SPC), Burkely Gallo (CIMMS/NSSL), Kent Knopfmeier (CIMMS/NSSL), Jack Hales (retired SPC), James Correia Jr. (CIMMS/SPC), Andy Dean (SPC), and Steve Willington (UKMO).

Week 1	Week 2	Week 3	Week 4	Week 5
April 30-May 4	**May 7-11**	**May 14-18**	**May 21-25**	**May 29-June 1**
Eric Loken (OU)	Nate Snook (OU/CAPS)	Steve Willington (Met Office)	Steve Willington (Met Office)	Steve Willington (Met Office)
Christina Kalb (DTC)	Brad Grant (WDTD)	Sarah Bull (Met Office)	Sarah Bull (Met Office)	Sarah Bull (Met Office)
Brian Ancell (TTU)	Shannon Rees (GFDL)	Matthew Lewis (Met Office)	Matthew Lewis (Met Office)	Matthew Lewis (Met Office)
Aaron Hill (TTU)	Andy Hazelton (GFDL)	Jason Otkin (CIMSS)	Harald Richter (BoM)	Justin Gibbs (WDTD)
Victor Gensini (NIU)	Bill Gallus (ISU)	Greg Thompson (NCAR)	Lance Bosart (SUNYA)	Tara Jensen (DTC)
Jamie Wolff (DTC)	Nicholas Vertz (ISU)	Amanda Burke (OU)	Massey Bartolini (SUNYA)	Clark Evans (UWM)
Terra Ladwig (GSD)	Ryan Sobash (NCAR)	Brian Ancell (TTU)	Marshall Pfahler (SUNYA)	David Nevius (UWM)
Dave Turner (GSD)	Brian Ancell (TTU)	Austin Coleman (TTU)	Craig Schwartz (NCAR; M-W)	Austin Coleman (TTU)
Tracy Dorian (EMC)	Aaron Hill (TTU)	Brian Kolts (FirstEnergy)	Austin Coleman (TTU)	Pete Wolf (NWS JAX)
Scott Rentschler (557WW)	Eric James (GSD)	Becky Adams-Selin (AER)	Ed Szoke (GSD)	Jeff Beck (DTC/GSD)
Dan Leins (NWS TWC)	Trevor Alcott (GSD)	John Brown (GSD)	Curtis Alexander (GSD)	Michelle Harrold (DTC)
Austin Harris (WDTD)	Alicia Bentley (EMC)	Jeff Duda (GSD)	Logan Dawson (EMC)	Ed Strobach (EMC)
Brittany Peterson (NWS FGF)	Geoff Manikin (EMC)	Eric Aligo (EMC)	Ben Blake (EMC)	Hugh Morrison (NCAR, T-W)
Michael Strickler (NWS RAH)	Robert Hart (NWS CRP)	Glen Romine (NCAR)	Andy Hatzos (NWS ILN)	Matthew Jackson (NWS TFX)
Dave Imy (retired SPC)	Darren Van Cleave (NWS SLC)	Jaret Rogers (NWS PSR)	Keith Sherburn (NWS UNR)	Jeff Milne (OU/CIMMS/SPC)
Colby Neuman (NWS PQR)	John Allen (CMU)	Matthew Friedlein (NWS LOT)	Brian Squitieri (SPC)	Ryan Solomon (AWC)
Caleb Grunzke (CIMMS/SPC)	John Gagan (NWS MKX)	Jason Davis (NWS BMX)		
	Mike Evans (WFO ALY)	Brendon Ruben-Oster (WPC)		
	Nathan Wendt (SPC)			

Appendix B: Experimental Severe Thunderstorm Forecasts

Severe weather graphics for the full-period Day 1 (1600-1200 UTC) individual hazard probabilities will be in the same format as that used for the operational SPC day 1 outlooks (categorical and general thunderstorm outlooks will not be made). For reference, the Probability-to-Categorical conversion for individual hazards used for the SPC Day 1 Outlook is shown below. These same probabilities will be used for generating the individual hazard forecasts in the four-hour periods.

Day 1 Probability to Categorical Outlook conversions

Day 1 Outlook Probability	TORN	WIND	HAIL
2%	MRGL	Not Used	Not Used
5%	SLGT	MRGL	MRGL
10%	ENH	Not Used	Not Used
10% with Significant Severe	ENH	Not Used	Not Used
15%	ENH	SLGT	SLGT
15% with Significant Severe	MDT	SLGT	SLGT
30%	MDT	ENH	ENH
30% with Significant Severe	HIGH	ENH	ENH
45%	HIGH	ENH	ENH
45% with Significant Severe	HIGH	MDT	MDT
60%	HIGH	MDT	MDT
60% with Significant Severe	HIGH	HIGH	MDT

Total severe weather probabilities for the full period Day 1 (1600-1200 UTC) total severe storm hazards will be in the same format as that used for the operational SPC Day 2 outlooks (5, 15, 30, 45, and 60 %). An area delineating potential for significant severe storms will be included when the probability for significant severe is 10% or greater. For reference, the Probability-to-Categorical conversion for total severe used for the SPC Day 2 Outlook, and is shown below. The same probabilities will be used for generating the hourly forecasts based on NEWS-e output.

Day 2 Probability to Categorical Outlook conversions

Day 2 Outlook Probability	Combined TOR, WIND, HAIL
5%	MRGL
15%	SLGT
15% with Significant Severe	SLGT
30%	ENH
30% with Significant Severe	ENH
45%	ENH
45% with Significant Severe	MDT
60%	MDT
60% with Significant Severe	HIGH

Appendix C. Organizational structure of the NOAA/Hazardous Weather Testbed

NOAA's Hazardous Weather Testbed (HWT) is a facility jointly managed by the National Severe Storms Laboratory (NSSL), the Storm Prediction Center (SPC), and the NWS Oklahoma City/Norman Weather Forecast Office (OUN) within the National Weather Center building on the University of Oklahoma South Research Campus. The HWT is designed to accelerate the transition of promising new meteorological insights and technologies into advances in forecasting and warning for hazardous mesoscale weather events throughout the United States. The HWT facilities are situated between the operations rooms of the SPC and OUN. The proximity to operational facilities, and access to data and workstations replicating those used operationally within the SPC, creates a unique environment supporting collaboration between researchers and operational forecasters on topics of mutual interest.

The HWT organizational structure is composed of three overlapping programs (Fig. 10). The Experimental Forecast Program (EFP) is focused on predicting hazardous mesoscale weather events on time scales ranging from hours to a week in advance, and on spatial domains ranging from several counties to the CONUS. The EFP embodies the collaborative experiments and activities previously undertaken by the annual SPC/NSSL Spring Experiments. For more information see http://www.nssl.noaa.gov/projects/hwt/efp/.

The Experimental Warning Program (EWP) is concerned with detecting and predicting mesoscale and smaller weather hazards on time scales of minutes to a few hours, and on spatial domains from several counties to fractions of counties. The EWP embodies the collaborative warning-scale experiments and technology activities previously undertaken by the OUN and NSSL. For more information about the EWP see http://www.nssl.noaa.gov/projects/hwt/ewp/. A key NWS strategic goal is to extend warning lead times through the "Warn-on-Forecast" concept (Stensrud et al. 2009), which involves using frequently updated short-range forecasts (≤ 1h lead time) from convection-resolving ensembles. This provides a natural overlap between

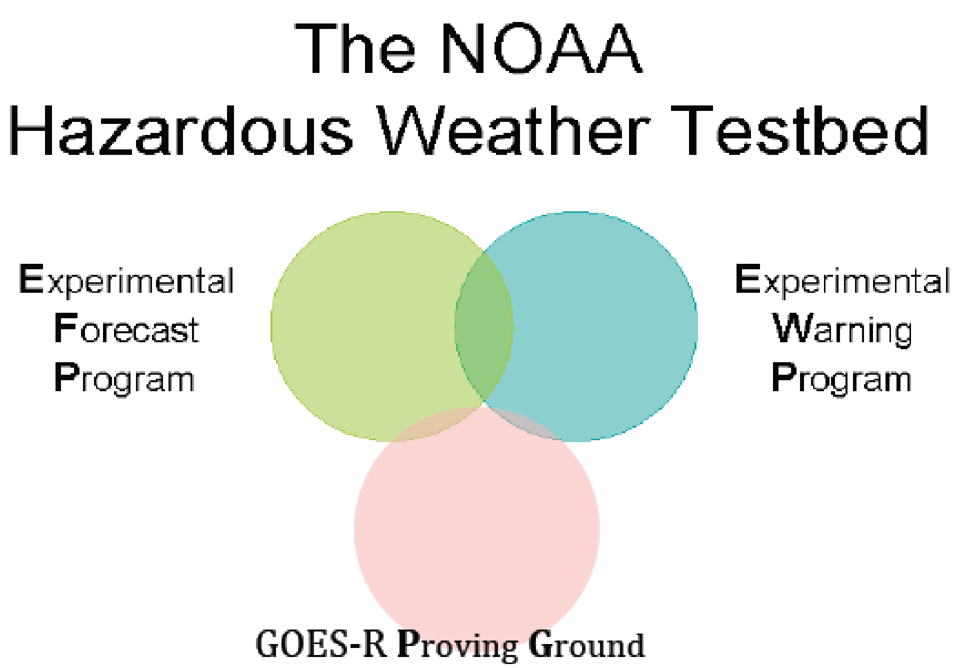

Figure 10: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

the EFP and EWP activities.

The GOES-R Proving Ground (established in 2009) exists to provide pre-operational demonstration of new and innovative products as well as the capabilities available on the next generation GOES-R satellite. The overall goal of the Proving Ground is to provide day-1 readiness once GOES-R launches in late 2015. The PG interacts closely with both product developers and NWS forecasters. More information about GOES-R Proving Ground is found at http://cimss.ssec.wisc.edu/goes_r/proving-ground.html.

Rapid science and technology infusion for the advancement of operational forecasting requires direct, focused interactions between research scientists, numerical model developers, information technology specialists, and operational forecasters. The HWT provides a unique setting to facilitate such interactions and allows participants to better understand the scientific, technical, and operational challenges associated with the prediction and detection of hazardous weather events. The HWT allows participating organizations to:

- Refine and optimize emerging operational forecast and warning tools for rapid integration into operations
- Educate forecasters on the scientifically correct use of newly emerging tools and to familiarize them with the latest research related to forecasting and warning operations
- Educate research scientists on the operational needs and constraints that must be met by any new tools (e.g., robustness, timeliness, accuracy, and universality)
- Motivate other collaborative and individual research projects that are directly relevant to forecast and warning improvement

For more information about the HWT, see http://www.nssl.noaa.gov/hwt/. Detailed historical background about the EFP Spring Experiments, including scientific and operational motivation for the intensive examination of high resolution NWP model applications for convective weather forecasting, and the unique collaborative interactions that occur within the HWT between the research and operational communities, are found in Weiss et al. (2010 – see http://www.spc.noaa.gov/publications/weiss/hwt-2010.pdf), Clark et al. (2012), and Gallo et al. (2017).

Appendix D: References

Adams-Selin, R.D. and C.L. Ziegler, 2016: Forecasting Hail Using a One-Dimensional Hail Growth Model within WRF. *Mon. Wea. Rev.,* **144**, 4919–4939,https://doi.org/10.1175/MWR-D-16-0027.1

Brimelow, J.C., 1999: Modeling maximum hail size in Alberta thunderstorms. *Wea. Forecasting*, **17**, 1048-1062.

Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. of Climate*, **26:2**, 380-398.

Clark, A. J., and Coauthors, 2012: An Overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Clark, A.J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia, Jr., M. Xue, and F. Kong, 2013: Tornado path length forecasts from 2010-2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387-407.

Clark, A.J., I.L. Jirak, S.R. Dembek, G.J. Creager, F. Kong, K.W. Thomas, K.H. Knopfmeier, B.T. Gallo, C.J. Melick, M. Xue, K.A. Brewster, Y. Jung, A. Kennedy, X. Dong, J. Markel, G.S. Romine, K.R. Fossell, R.A. Sobash, J.R. Carley, B.S. Ferrier, M. Pyle, C.R. Alexander, S.J. Weiss, J.S. Kain, L.J. Wicker, G. Thompson, R.D. Adams-Selin, and D.A. Imy, 0: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.,* **0**, https://doi.org/10.1175/BAMS-D-16-0309.1

Gagne, D. J., A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-Based Probabilistic Hail Forecasting with Machine Learning Applied to Convection-Allowing Ensembles. *Wea. Forecasting*, **32**, 1819-1840.

Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting Tornadoes Using Convection-Permitting Ensemble. *Wea. Forecasting*, **31**, 273-295.

Gallo, B.T., and Coauthors, 2017: Breaking New Ground in Severe Weather Prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting,* **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1

Gallo, B.T., A.J. Clark, B.T. Smith, R.L. Thompson, I. Jirak, and S.R. Dembek, 2018: Blended Probabilistic Tornado Forecasts: Combining Climatological Frequencies with NSSL–WRF Ensemble Forecasts. *Wea. Forecasting,* **33**, 443–460, https://doi.org/10.1175/WAF-D-17-0132.1

Han, J., Wang, W., Kwon, Y.C., Hong, S.Y., Tallapragada, V. and Yang, F., 2017: Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Wea. Forecasting*, **32**, 2005-2017.

Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev*., **141**, 283-306.

Harris, L.M., S.-J. Lin, and C.-Y. Tu, 2016: High-Resolution Climate Simulations Using GFDL HiRAM with a Stretched Global Grid. *Journal of Climate*, **29**, 4293-4314.

Hitchens, N. M., Harold E. Brooks, M. P. Kay, 2013: Objective Limits on Forecasting Skill of Rare Events. *Wea. Forecasting*, **28**, 525–534.

Hong, S.Y., Noh, Y. and Dudhia, J., 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.,* **34**, 2318-2341.

Jewell, R., and J. Brimelow, 2009: Evaluation of Alberta Hail Growth Model using severe hail proximity soundings from the United States. *Wea. Forecasting*, **24**, 1592-1609.

Jirak, I. L., C. J. Melick, A. R. Dean, S. J. Weiss, and J. Correia, Jr., 2012: Investigation of an automated temporal disaggregation technique for convective outlooks during the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. Preprints, *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 10.2.

Jirak, I. L., C. J. Melick, and S. J. Weiss, 2014: Combining probabilistic ensemble information from the environment with simulated storm attributes to generate calibrated probabilities of severe weather hazards. Preprints, *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 2.5.

Johnson, A., X. Wang, J.R. Carley, L.J. Wicker, and C. Karstens, 2015: A Comparison of Multiscale GSI-Based EnKF and 3DVar Data Assimilation Using Radar and Conventional Observations for Midlatitude Convective-Scale Precipitation Forecasts. *Mon. Wea. Rev.,* **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1

Kain, J.S., S.R. Dembek, S.J. Weiss, J.L. Case, J.J. Leit, and R.A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536-1542.

Lin, S.-J., and R.B. Rood, 1996: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes. *Mon. Wea. Rev.*, **124**, 2046-2070.

Lin, S.-J., and R.B. Rood, 1997: An explicit flux-form semi-lagrangian shallow-water model on the sphere. *QJRMS*, **123**, 2477-2498.

Lin, S.-J., 1997: A finite-volume integration method for computing pressure gradient force in general vertical coordinates. *QJRMS*, **123**, 1749-1762.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293-2307.

Melick, C. J, I. L. Jirak, A. R. Dean, J. Correia Jr, and S. J. Weiss, 2012: Real Time Objective Verification of Convective Forecasts: 2012 HWT Spring Forecast Experiment. Preprints, *37th Natl. Wea. Assoc. Annual Meeting*, Madison, WI, Natl. Wea. Assoc*.,* P1.52.

Palmer, T.N., R. Buizza, F. DoblasReyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer,
and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp. [Available online at http://www.ecmwf.int/publications/.]

Putman, W M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, **227**, 55-78.

Sobash, R. A., J.S. Kain, D.R. Bright, A.R. Dean, M.C. Coniglio, and S.J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714-728.

Sobash, R. A., C. S. Schwartz, G. S. Romine, K. R. Fossess, and M. L. Weisman, 2016a: Severe Weather Prediction Using Storm Surrogates from an Ensemble Forecasting System. *Wea. Forecasting*, **31**, 255-271.

Sobash, R. A., G. S. Romine, C. S. Schwartz, D. J. Gagne, and M. L. Weisman, 2016b: Explicit Forecasts of Low-Level Rotation from Convection-Allowing Models for Next-Day Tornado Prediction. *Wea. Forecasting*, **31**, 1591-1614.

Stensrud, D. J., and Coauthors, 2009: Convective-Scale Warn-on-Forecast System. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Thompson, G. and T. Eidhammer, 2014: A Study of Aerosol Impacts on Clouds and Precipitation Development in a Large Winter Cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1

Wang, Y. and X. Wang, 2017: Direct Assimilation of Radar Reflectivity without Tangent Linear and Adjoint of the Nonlinear Observation Operator in the GSI-Based EnVar System: Methodology and Experiment with the 8 May 2003 Oklahoma City Tornadic Supercell. *Mon. Wea. Rev.,* **145**, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1

Wang Y., X. Wang, and J. Carley, 2018a: GSI-based EnKF-Variational Hybrid Data Assimilation for NCEP NAMRR: System Development and Initial Testing. *Mon. Wea. Rev.* to be submitted.

Wang Y., X. Wang, and D. Dowell, 2018b: GSI-based EnKF-Variational Hybrid Data Assimilation for HRRR. *Mon. Wea. Rev. in preparation*.