SPRING FORECASTING EXPERIMENT 2021

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

HWT Facility – National Weather Center

3 May – 4 June 2021

https://hwt.nssl.noaa.gov/sfe/2021/

Program Overview and Operations Plan

26 April, 2021

Adam Clark[2], Israel Jirak[1], Burkely T. Gallo[1,3], Andy Dean[1], Kent Knopfmeier[2,3], Brett Roberts[1,2,3], Patrick Burke[2], Louis Wicker[2], Makenzie Krocak[3,4,5], Patrick Skinner[2,3], Pam Heinselman[2], Katie Wilson[2,3], Jake Vancil[1,3], Kimberly Hoogewind[2,3], Nathan Dahl[1,3], Gerry Creager[2,3], Thomas Jones[2,3], Jidong Gao[1], Yunheng Wang[2,3], Eric D. Loken[2,3,4], Montgomery Flora[2,3,4], Chris Kerr[2,3], Nusrat Yussouf[2,3], Joshua Martin[2,3], Jorge Guerra[2,3], Brian Matilla[2,3], David Jahn[1,3], David Harrison[1,3], and Dave Imy[2]

(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
(5) Center for Risk and Crisis Management, University of Oklahoma, Norman, Oklahoma

Table of Contents

The NOAA Hazardous Weather Testbed (photo credit: James Murnan, NSSL)

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 3 km grid-spacing) modeling systems.

The 2021 Spring Forecasting Experiment (SFE 2021), a cornerstone of the EFP, will be conducted 3 May – 4 June. Because of the COVID-19 pandemic, restrictions on travel and gatherings preclude an in-person experiment in the HWT for the second consecutive year. However, to maintain momentum in several areas of convection-allowing model (CAM) development, the EFP will once again conduct a virtual experiment. Relative to SFE 2020, this year's virtual experiment will have more forecasting activities, and all participants – not just National Weather Service (NWS) forecasters – will have the opportunity to participate in these forecasting activities. Additionally, participants will perform next-day evaluations of model performance. As in previous years, a suite of new and improved experimental CAM guidance contributed by our large group of collaborators will be central these forecasting and model evaluation activities. These contributions comprise an ensemble framework called the Community Leveraged Unified Ensemble (CLUE; Clark et al. 2018). The 2021 CLUE is constructed by using common model specifications (e.g., grid-spacing, model version, domain size, post-processing, etc.) wherever possible 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 deterministic CAMs and CAM ensembles. The 2021 CLUE includes 64 members with 3-km grid-spacing. SFE 2021 will also involve the continued testing of the Warn-on-Forecast System (WoFS, hereafter), which produces 18-member, 3-km grid-spacing forecasts, and will be used for the fourth year to issue very short lead-time outlooks. Additionally, a deterministic, 1.5-km grid-spacing simulation using dual resolution, hybrid data assimilation (WoFS-hybrid) will complement the full WoFS ensemble.

This document summarizes the core interests of SFE 2021 with information on experiment operations. 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 model and products being tested during SFE 2021 and Section 3 describes the core interests and new concepts being introduced for SFE 2021. 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

Daily model evaluation activities will occur from 9:15 – 11:00am (CDT) focusing on various CLUE subsets. The 2021 CLUE includes deterministic and ensemble forecasts using the most recent versions of the Finite Volume Cubed-Sphere Limited Area Model (FV3-LAM), as well as the Advanced Research Weather Research and Forecasting (WRF-ARW) model. In addition to the CLUE, the operational 3-km grid-spacing High-Resolution Ensemble Forecast system version 3 (HREFv3) will be examined. The rest of this section provides further details on each modeling system utilized in SFE 2021.

a) The 2021 Community Leveraged Unified Ensemble (CLUE)

The CLUE is a carefully designed ensemble with members contributed by NSSL, NOAA's Environmental Modeling Center (EMC), NOAA's Global Systems Laboratory (GSL), NOAA's Geophysical Fluid Dynamics Laboratory (GFDL), and the Multi-scale data Assimilation and Predictability (MAP) group at the University of Oklahoma. CLUE members have 3-km grid-spacing and either a CONUS or North America domain. Depending on the CLUE subset, forecast lengths range from 18 to 126 h. Table 1 summarizes all 2021 CLUE contributions. Subsequent tables provide details on members in each subset.

Table 1 Summary of the 14 unique subsets that comprise the 2021 CLUE.

Clue Subset	# of mems	IC/LBC perts	Mixed Physics	Data Assimilation	Model Core	Agency	Init. Times (UTC)	Forecast Length (h)	Domain
GSL RRFS	9	HRRRDAS/ GEFS	no	EnKF	FV3	GSL	00, 12	60, 48	CONUS
HRRRE-S	9	HRRRDAS/ GEFS	no	EnKF	ARW	GSL	12	24	CONUS
HRRRE-M	9	HRRRDAS/ GEFS	yes	EnKF	ARW	GSL	12	24	CONUS
GSL FV3-LAM	1	none	no	Hybrid 3DEnVar (GDAS Ensemble)	FV3	GSL	00-23 (hourly)	20x18h, 4x48h	CONUS
GSL FV3-LAM-NA	1	none	no	cold start from GFS	FV3	GSL	00, 12	60, 60	N. America
EMC FV3-LAM	1	none	no	cold start from GFS	FV3	EMC	00, 12	60, 60	CONUS
EMC FV3-LAMX	1	none	no	cold start from GFS	FV3	EMC	00, 12	60, 60	N. America
EMC FV3-LAMDAX	1	none	no	Hybrid 3DEnVar (GDAS EnKF)	FV3	EMC	00, 12	60, 60	CONUS
HRRRv4	1	none	no	GSI-EnVar	ARW	EMC	00-23 (hourly)	20x18h, 4x48h	CONUS
RRFS Cloud	9	GFS, GEFS	yes	cold start from GFS, GEFS	FV3	EMC/GSL	00	60	N. America
MAP RRFS	10	GFS, GEFS	no	GSI-EnVar	FV3	OU-MAP	21, 00	39, 36	CONUS
MAP RRFS VTS	10	GFS, GEFS	no	GSI-EnVar	FV3	OU-MAP	21, 00	39, 36	CONUS
NSSL FV3-LAM	1	none	no	cold start from GFS	FV3	NSSL	00	60	CONUS
GFDL FV3	1	none	no	cold start from GFS	FV3	GFDL	00	126	CONUS

Table 2 Specifications for the GSL RRFS CLUE members. The GSL RRFS is a 9-member 3-km CONUS FV3-LAM ensemble forecast. Initial conditions come from the operational hourly-cycled, 3-km ensemble ("HRRR Data-Assimilation System" or "HRRRDAS"; https://rapidrefresh.noaa.gov/internal/pdfs/2020_Spring_Experiment_HRRRE_Documentation.pdf) that is a component of HRRRv4. RRFS forecast member 1 is an unperturbed "control member" initialized from the analysis mean of the HRRRDAS. GSL RRFS members 2-9 are perturbed forecasts initialized from the corresponding members in the HRRRDAS. The perturbed members 2-9 also include stochastic parameter perturbations (SPP) applied to the land-surface, PBL, and microphysics schemes plus stochastically perturbed parameterization tendencies (SPPT). This ensemble will be initialized at 00z and 12z each day with forecasts to 60 and 48 hrs respectively.

Members: GSL RRFS	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
gsl-rrfs01	enkf_mean	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs02	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs03	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs04	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs05	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs06	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs07	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs08	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3
gsl-rrfs09	HRRRDAS	GEFS	Thompson	MYNN	RUC	RRTMG	FV3

Table 3 Specifications for the HRRRE-S ("single-physics") CLUE members. The legacy HRRR Ensemble is a 9-member, WRF-ARW forecast. Initial conditions come from the operational hourly-cycled, 3-km ensemble ("HRRR Data-Assimilation System" or "HRRRDAS") that is a component of HRRRv4. HRRRE forecast member 1 is an unperturbed "control member" initialized from the analysis mean of the HRRRDAS. HRRRE members 2-9 are perturbed forecasts initialized from the corresponding members in the HRRRDAS. The perturbed members 2-9 include stochastic parameter perturbations (SPP) applied to the land-surface, PBL, and microphysics schemes plus stochastically perturbed parameterization tendencies (SPPT). The 2021 HRRRE-S configuration is very similar to the 2020 configuration; the most important difference is that member 1 is now an unperturbed forecast initialized from the ensemble mean. As in previous years, the HRRRDAS analyses and HRRRE-S forecasts provide initial conditions and boundary conditions for the experimental WoFS. This HRRRE ensemble will be initialized at 12z each day with forecasts to 24 hrs.

Members: HRRRE-S	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
hrrre-s01	enkf_mean	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s02	enkf_m02	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s03	enkf_m03	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s04	enkf_m04	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s05	enkf_m05	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s06	enkf_m06	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s07	enkf_m07	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s08	enkf_m08	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-s09	enkf_m09	GEFS	Thompson	MYNN	RUC	RRTMG	ARW

Table 4 Specifications for the HRRRE-M ("mixed-physics") CLUE members (WRF-ARW). Motivated by results from previous SFEs, the potential value that physics-scheme diversity adds to ensemble forecasts will be tested in the HRRRE framework this year by running a mixed-physics ensemble forecast to complement the standard single-physics ensemble forecast. In the HRRRE-M, members 1-5 run with HRRR physics (including stochastic physics for members 2-5) while members 6-9 run with a physics configuration that has been used previously in the NSSL WRF. Since members 1-5 of HRRRE-M and HRRRE-S are identical, any differences between the two ensemble forecasts overall depend specifically on differences in how members 6-9 are configured. This HRRRE ensemble will be initialized at 12z each day with forecasts to 24 hrs.

Members: HRRRE-M	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
hrrre-m01	enkf_mean	GFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-m02	enkf_m02	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-m03	enkf_m03	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-m04	enkf_m04	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-m05	enkf_m05	GEFS	Thompson	MYNN	RUC	RRTMG	ARW
hrrre-m06	enkf_m06	GEFS	WSM6	MYJ	NOAH	RRTM, Dudhia	ARW
hrrre-m07	enkf_m07	GEFS	WSM6	MYJ	NOAH	RRTM, Dudhia	ARW
hrrre-m08	enkf_m08	GEFS	WSM6	MYJ	NOAH	RRTM, Dudhia	ARW
hrrre-m09	enkf_mean	GEFS	WSM6	MYJ	NOAH	RRTM, Dudhia	ARW

Table 5 Specifications for the GSL FV3-LAM CLUE member. GSL in collaboration with EMC, NSSL and other organizations will be providing experimental deterministic configurations of the FV3-based limited area model (LAM) system running on a 3-km grid. This system is under development and testing towards a future operational implementation as part of the Unified Forecast System (UFS) Convection Allowing Model (CAM) application known as the Rapid Refresh Forecast System (RRFS). This configuration uses a recent version of the RAP/HRRR physics suite along with hourly-cycled data assimilation using a hybrid 3DEnVar (GDAS ensemble) analysis and periodically drawing initial and boundary conditions from a 13-km North American FV3LAM hourly-cycled configuration using the same RAP/HRRR physics suite. The 3-km hourly-updating forecasts are initialized over CONUS with forecasts to 60 hrs once every six hours and 18 hrs otherwise.

Member: GSL FV3-LAM	**ICs**	**LBCs**	**DA (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
GSL FV3-LAM	Cycled	13-km FV3-LAM	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3

Table 6 Specifications for the GSL FV3-LAM-NA CLUE member. This configuration uses a recent version of the RAP/HRRR physics suite matching the other GSL FV3-LAM configuration but with initial and boundary conditions taken from the GFSv16. No data assimilation is executed with this configuration. The 3-km forecasts are initialized over North America at 00z and 12z with forecasts to 60 hrs.

Members: GSL FV3-LAM-NA	**ICs**	**LBCs**	**DA (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
GSL FV3-LAM-NA	GFSv16	GFSv16f	no	N. America	Thompson	MYNN	RUC	RRTMG	FV3

Table 7 Specifications for the EMC FV3-LAM CLUE member. This member is the control version of EMC's FV3-LAM contributions. It uses a cold start and runs over the CONUS.

Members: EMC FV3-LAM	**ICs**	**LBCs**	**DA (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
EMC FV3-LAM	GFSv16	GFSv16f	no	CONUS	Thompson	MYNN	NOAH	RRTMG	FV3

Table 8 Specifications for the EMC FV3-LAMX CLUE member. This configuration runs over an expanded North American domain, but is otherwise identical to the EMC FV3-LAM configuration. The EMC FV3-LAMX will underpin the RRFS, a rapid-update, convection-allowing ensemble data assimilation and prediction system planned for implementation in the NCEP production suite during the Q4FY2023 timeframe.

Members: EMC FV3-LAMX	**ICs**	**LBCs**	**DA (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
EMC FV3-LAMX	GFSv16	GFSv16f	no	N. America	Thompson	MYNN	NOAH	RRTMG	FV3

Table 9 Specifications for the EMC FV3-LAMDAX CLUE member. This configuration is similar to the EMC FV3-LAM, but it is initialized from a 6-h data assimilation cycle with the 3-km regional FV3 model with hourly analysis updates using RAP observations. The DA cycle is cold-started from a 6-h GDAS forecast valid at 6-h prior to forecast start time.

Members: EMC FV3-LAMDAX	**ICs**	**LBCs**	**DA (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
EMC FV3-LAMDAX	GFSv16	GFSv16f	no	CONUS	Thompson	MYNN	NOAH	RRTMG	FV3

Table 10 Specifications for the HRRRv4 CLUE member. The final update to the deterministic HRRRv4 was implemented operationally in June 2020. The physics suite for HRRRv4 continues to use actively-developed versions of Thompson et al. (2014) aerosol-aware microphysics, MYNN PBL scheme, RUC land surface model and RRTMG SW/LW radiation schemes. Enhancements have been made to the MYNN PBL scheme to further improve both representation of sub-grid-scale clouds and their effects on the local environment (reducing model bias of incoming radiation and temperature/moisture fields). Gravity-wave drag enhancements have been made to improve representation of the effects of sub-grid terrain on the horizontal flow. Land surface model and state changes include installation of an inland lake model for improved lake-temperature prediction, higher-resolution MODIS albedo and inland lake datasets, use of fractional sea-ice data and FVCOM dynamic specification of temperature and ice concentrations for the Great Lakes. Enhancements to numerics in HRRRv4 include a reduction in magnitude of the 6th order filter for momentum, thermodynamic and hydrometeor fields to improve depiction of weaker small-scale cloud and precipitation features. A new implicit-explicit vertical advection scheme in HRRRv4 permits larger vertical motion in intense convection to facilitate improved diagnosis of rotational features such as mesocyclones. For data assimilation, The HRRRv4 uses an updated version of GSI and includes assimilation of additional datasets including lightning data from GOES (GLM), aircraft and RAOB moisture observations above 300 mb. A 36-member, hourly-cycled, storm-scale ensemble data assimilation system (HRRRDAS) provides a background deterministic state estimate (ensemble mean) and background ensemble for initialization of the CONUS HRRRv4. This system is designed to improve use of conventional and radar observations during data assimilation with better representation of meso-to-storm scale covariances when compared with the comparatively coarse global ensemble (GDAS) used in HRRRv3. More accurate retention and evolution of meso-to-storm scale features, particularly in the early forecast hours, are intended benefits of HRRRDAS use. The HRRRDAS, while intended to improve deterministic HRRRv4 forecasts, also forms the basis for HRRR ensemble forecasts described in the HRRRE section.

Member: HRRRv4	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
HRRRv4	HRRRDAS	RAP	Thompson	MYNN	RUC	RRTMG	ARW

Table 11 Specifications for the RRFS Cloud CLUE members. As a part of a collaborative effort between EMC, GSL, and NSSL a prototype 9-member RRFS ensemble will be run for the first time using cloud-based high performance computing. Three physics suites will be utilized: Thompson MP / MYNN PBL, GFDL MP / TKE-EDMF PBL, and NSSL MP / Hybrid-EDMF PBL. With each physics suite, one member will not contain any stochastic perturbations (GFS mem), one member will use SPPT (mem 1), and the remaining member will have SPPT/SHUM/SKEB perturbations (mem 2). Initial condition and lateral boundary perturbations are provided via 6-h forecasts from the 18Z GFS and GEFS. This system will provide one 00Z ensemble forecast per day over North America out to 60-h (12-h longer than the upcoming HREFv3).

Members: RRFS Cloud	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
rrfs-cloud01	18ZGFS	GFS	Thompson	MYNN	NOAH	RRTMG	FV3
rrfs-cloud02	18ZGEFS mem1	GEFS mem1	Thompson	MYNN	NOAH	RRTMG	FV3
rrfs-cloud03	18ZGEFS mem2	GEFS mem2	Thompson	MYNN	NOAH	RRTMG	FV3
rrfs-cloud04	18ZGFS	GFS	GFDL	TKE-EDMF	NOAH	RRTMG	FV3
rrfs-cloud05	18ZGEFS mem1	GEFS mem1	GFDL	TKE-EDMF	NOAH	RRTMG	FV3
rrfs-cloud06	18ZGEFS mem2	GEFS mem2	GFDL	TKE-EDMF	NOAH	RRTMG	FV3
rrfs-cloud07	18ZGFS	GFS	NSSL	hybrid-EDMF	NOAH	RRTMG	FV3
rrfs-cloud08	18ZGEFS mem1	GEFS mem1	NSSL	hybrid-EDMF	NOAH	RRTMG	FV3
rrfs-cloud09	18ZGEFS mem2	GEFS mem2	NSSL	hybrid-EDMF	NOAH	RRTMG	FV3

Table 12 Specifications for the MAP RRFS CLUE members. These 3-km grid-spacing ensemble forecasts are run with FV3LAM and initialized by a GSI-based hybrid EnVar DA system directly assimilating both conventional and radar reflectivity observations (Johnson et al. 2015, Wang and Wang 2017). The ensemble for data assimilation has 36 members. The LBCs are provided by re-centering GEFS around the GFS control, with external ICs provided at 1800 UTC by the GFS control member and GEFS ensemble members. The system assimilates both operational RAP/HRRR in-situ data stream and MRMS radar reflectivity hourly during 1900-0000 UTC over the CONUS domain. Two 10-member ensemble forecasts are initialized at 2100 and 0000 UTC and advanced for 39 and 36 hours, respectively, including a control forecast member (map-hybrid01) initialized from the GSI based EnVar control analysis and 9-members from the GSI EnKF analyses re-centered around the control member. The same physics schemes as listed below are adopted for all members in both data assimilation and ensemble free forecasts.

Members: MAP RRFS	**ICs**	**LBCs**	**VTS (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
map-rrfs01	EnVar	GFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs02	rEnKF_m1	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs03	rEnKF_m2	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs04	rEnKF_m3	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs05	rEnKF_m4	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs06	rEnKF_m5	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs07	rEnKF_m6	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs08	rEnKF_m7	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs09	rEnKF_m8	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs10	rEnKF_m9	GEFS	no	CONUS	Thompson	MYNN	RUC	RRTMG	FV3

Table 13 Specifications for the MAP RRFS VTS CLUE members. The configuration of the valid time shifted (vts; Gasperoni et al. 2021, Huang and Wang 2018) ensemble forecasts follow the "hybrid" DA configuration of Table 12, but includes vts-expanded ensemble covariances for the control member (vts01) radar EnVar analysis during hourly cycling. By including 36-member output 30-min before and after each central analysis time to the original ensemble, the 108-member vts-expanded ensemble covariances mimic the effects of directly increasing ensemble size by a factor of 3 and includes information of model timing/phase uncertainty in convective systems. The remaining 9 members (vts02-vts10) are updated using the same EnKF procedure as in Table 12. Although vts is only directly applied to the radar EnVar update of the control member, its effects may further transfer to members 2-10 via recentering around the vts-enabled analyses.

Members: MAP RRFS VTS	**ICs**	**LBCs**	**VTS (yes/no)**	**Domain**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Model**
map-rrfs-vts01	vts-enabled EnVAR	GFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts02	rEnKF_m1	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts03	rEnKF_m2	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts04	rEnKF_m3	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts05	rEnKF_m4	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts06	rEnKF_m5	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts07	rEnKF_m6	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts08	rEnKF_m7	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts09	rEnKF_m8	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3
map-rrfs-vts10	rEnKF_m9	GEFS	yes	CONUS	Thompson	MYNN	RUC	RRTMG	FV3

Table 14 Specifications for the NSSL FV3-LAM CLUE member. This member is configured the same as the EMC FV3-LAM member, but with the NSSL microphysics scheme.

Member: NSSL FV3-LAM	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
NSSL FV3-LAM	GFSv16	GFSv16	NSSL	MYNN	NOAH	RRTMG	FV3

Table 15 Specifications for the GFDL FV3 CLUE member. GFDL's C-SHiELD (Harris et al., 2019) is an FV3-based model that uses a 13-km global grid and a 3-km factor-of-five CONUS nest, coupled to a modified form of the GFS Physics. C-SHiELD uses the GFDL In-line Microphysics (Zhou et al. 2019; Harris et al. 2020) and the EMC/UW TKE-EDMF PBL scheme (Han and Bretherton 2019). The deep convective scheme is disabled on the nested grid. On the CONUS nest the Noah-MP LSM is used; the global domain uses the GFS Noah LSM. Initialization is cold start from regridded GFS real-time analyses. GFDL will provide simulations run daily at 00Z out to 126 hours to demonstrate the potential for medium-range prediction of convective-scale events.

Member: GFDL FV3	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Model**
GFDL FV3	GFS	n/a	GFDL	TKE-EDMF	NOAH-MP	RRTMG	FV3

The configuration of the 2021 CLUE will allow for several unique experiments that have been designed to examine issues immediately relevant to the design of a NCEP/EMC operational CAM-based ensemble. Some of the major themes are listed below:

Valid Time Shifting Data Assimilation: The OU MAP group has a project to test the impact of a data assimilation approach known as Valid Time Shifting (VTS). This approach is a cost-effective way to increase the membership (by a factor of three) for the background ensemble in convective scale, hybrid EnVar data assimilation. The increased membership is achieved by populating the background ensemble with analyses valid at slightly different lead times. Ensembles with and without (Tables 13 & 14, respectively) VTS initialized at 2100 and 0000 UTC will be examined.

RRFS Configuration Strategies: Several different ensembles will be contributed and evaluated against the HREFv3. The goal is to identify a strategy within the UFS framework (i.e., single-model, FV3-LAM) that performs as good as or better than HREFv3, so that it can serve as a replacement in NCEP's production suite. These ensembles include GSL RRFS (Table 2), RRFS Cloud (Table 11), MAP RRFS (Table 12), and MAP RRFS VTS (Table 13) that vary in data assimilation and physics strategies.

CAM Ensemble Physics: Two configurations of the 12Z HRRRE will be compared to assess the role of a mixed-physics approach for increasing the spread and diversity of CAM ensemble forecasts. One ensemble (HRRRE-S) uses a single physics scheme with SPP and SPPT stochastic perturbations while the other ensemble (HRRRE-M) replaces four of the members from HRRRE-S with a different physics configuration based on the NSSL-WRF.

FV3-LAM Configurations: GSL, NSSL, EMC, and GFDL will run various configurations of FV3 (Tables 5-9, 11, & 14-15). These coordinated runs will allow for the assessment of many aspects of FV3-LAM configuration including physics, data assimilation, initial conditions, domain, and impact of stochastic physics.

Day 2 FV3-LAM performance: New to SFE 2021 will be evaluation of FV3-LAM configurations for the Day 2 forecast period (i.e., forecast hours 36-60), which will focus on the pairs of EMC and GSL configurations that use CONUS and North America domains. Specifically, EMC FV3-LAM (Table 7), EMC FV3-LAMX (Table 8), GSL FV3-LAM (Table 5), and GSL FV3-LAM-NA (Table 6).

3D-RTMA Background: Two hourly versions of the 3D-RTMA will be compared to assess the role that the background first-guess plays on the final analysis. One version from EMC uses the operational HRRRv4 as the background while the other version from GSL uses the FV3-LAM as the background.

To ensure consistent post-processing, visualization, and verification, post-processing is standardized as much as possible, so that a consistent set of model output fields are output on the same grid. For both WRF-ARW and FV3-LAM, the Unified Post-Processor software (UPP; available at http://www.dtcenter.org/upp/users/downloads/index.php) is used. For the WRF-ARW runs (HRRRv4, HRRRE-S, & HRRRE-M), there are about 150 fields output that are relevant to a broad range of forecasting needs, including aviation, severe weather, and precipitation. The number of FV3-LAM output fields is smaller, about 50, but includes sensible weather fields, storm diagnostics, and environmental indices important for severe weather forecasting.

b) High Resolution Ensemble Forecast (HREFv3) System

HREFv3 is a 10-member CAM ensemble scheduled for operational implementation 11 May 2021. HREFv3 will replace HREFv2.1. The design of HREFv3 originated from the SSEO, which demonstrated skill for six years in the HWT and SPC prior to operational implementation. In HREFv3, the HRW NMMB simulations have been replaced with HRW FV3 and HRRRv3 has been upgraded to HRRRv4.

Table 16 Model specifications for HREFv3.

HREFv3	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**dx (km)**	**Vertical Levels**	**Included in HREF hours**
HRRRv4	HRRRDAS	RAP -1h	Thompson	MYNN	3.0	50	0 – 48
HRRRv4 -6h	HRRRDAS	RAP -1h	Thompson	MYNN	3.0	50	0 – 42
HRW ARW	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 48
HRW ARW -12h	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 36
HRW FV3	GFS	GFS -6h	GFDL	EDMF	3	50	0 – 60
HRW FV3 -12h	GFS	GFS-6h	GFDL	EDMF	3	50	0 – 48
HRW NSSL	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 48
HRW NSSL -12h	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 36
NAM CONUS Nest	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 60
NAM CONUS Nest -12h	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 48

c) NSSL Warn-on-Forecast Experiments

The NSSL Warn-on-Forecast System (WoFS) is a rapidly-updating 36-member, 3-km grid-spacing WRF-based ensemble data assimilation and forecast system. The WoFS is initialized every 30 minutes and used to produce very short-range (0-6 h) probabilistic forecasts of individual thunderstorms and their associated hazardous weather phenomena such as supercell hail, high winds, flash flooding, and supercell thunderstorm rotation. In addition, a dual-resolution hybrid data assimilation and forecast system, WoFS-Hybrid is used to produce a single 1.5-km deterministic forecast. The 900-km x 900-km daily WoFS domain will target the primary region where severe weather is anticipated.

The starting point for each day's experiment will be the operational High-Resolution Rapid Refresh Data Assimilation System (HRRRDAS) provided by NCO/GSL and the HRRRE-S (Table 3) provided by GSL. A 1-h forecast from the 1400 UTC, 36-member, hourly-cycled HRRRDAS analysis provides the initial conditions for both the WoFS and WoFS-Hybrid. Boundary conditions are provided by 1200 UTC HRRRE-S forecasts, initialized from the 1200 UTC HRRRDAS analysis and valid for the period 1500 UTC Day 1 – 0300 UTC Day 2. Table 17 provides a summary of model specifications for HRRRE, WoFS, and WoFS-Hybrid, and Figure 1 shows an example of a SPC Day 1 convective outlook and corresponding WoFS domain with WSR-88D radars used for data assimilation overlaid. Further details on WoFS and WoFS-hybrid are included below.

i) WoFS

The 36-member WoFS, run from 1500 UTC Day 1 to 0300 UTC Day 2, cycles its data assimilation every 15 minutes by GSI-EnKF assimilation of MRMS radar reflectivity and radial velocity data, cloud water path retrievals and clear-sky radiances from the GOES-16 imager, and Oklahoma Mesonet observations (when available). Conventional (i.e. prepbufr) observations are also assimilated at 15 minutes past each hour. All WoFS ensemble members utilize the NSSL 2-moment microphysics parameterization and the RUC land-surface model; however, the PBL and radiation physics options are varied amongst the ensemble members to increase ensemble spread, given the fact that the EnKF may underrepresent model physics errors. Six-hour (three-hour) forecasts are initialized and launched from the first 18 ensemble members from the real-time WoFS analyses at the top of each hour (half-hour). The first available forecast is launched at 1700 UTC Day 1 and the last at 0300 UTC Day 2. These forecasts will be viewable using the web-based WoFS Forecast Viewer (https://wof.nssl.noaa.gov/realtime/).

ii) WoFS-Hybrid

In parallel with the baseline WoFS, an efficient, weather-adaptive, hybrid three-dimensional variational and Ensemble Kalman Filter analysis and forecast system was built and implemented for the WoF project (WoFS-Hybrid, Gao et al. 2013; Wang et al. 2019). The system incorporates flow-dependent background error covariances estimated from the ensemble forecasts of the baseline WoFS, but provides a high-resolution deterministic analysis and forecast component that can be regarded as a complement to the baseline WoFS. One can think of this as mirroring the way that other coupled model systems attempt to provide one skillful, control forecast member to complement the associated ensemble (e.g., GFS and GEFS).

In the WoFS-Hybrid, WSR-88 radar data, GOES-16 GLM Lightning-Derived Water Vapor, and surface observations will be used through rapid DA and forecast cycles (every 15 minutes), though some of these data will be used in different formats from those used in WoFS baseline. A forecast launched from 1200 UTC with HRRRE member 1 is used to provide boundary conditions. Similarly, a 1-h forecast launched from the 1400 UTC HRRRDAS member mean is used to provide initial conditions for the WoFS-Hybrid analysis. The WoFS-Hybrid system will run from 1500 UTC Day 1 to 0300 UTC Day 2. A 6-h forecast will be launched from the analysis each hour from 1700 UTC during this period. The daily domain will be the same as WoFS.

Table 17 HRRRE, WoFS, and WoFS-Hybrid configuration comparison.

	HRRRE	**WoFS**	**WoFS-Hybrid**
Model Version	WRF-ARW v3.9+	WRF-ARW v3.9+	WRF-ARW v3.9+
Grid Dimensions	1800 x 1060 x 50	300 x 300 x 50	600 x 600 x 50
Grid Resolution	3 km	3 km	1.5 km
EnKF cycling	36-mem. w/ GSI-EnKF **every 1 hr**	36-mem. w/ GSI-EnKF **every 15 min**	36-mem. w/ GSI-EnKF **every 15 min**
Observations	- Prepbufr conventional observations - GOES-16 ABI radiances - MRMS radar reflectivity	- Prepbufr conventional observations - Oklahoma Mesonet (when available) - MRMS reflectivity ≥ 15 dBZ; radar 'zeroes' - MRMS radial velocity - GOES-16 cloud-water path - GOES-16 clear sky radiances - GOES-16 atmospheric motion vectors	- Prepbufr conventional observations - Oklahoma Mesonet (when available) - Raw radar reflectivity ≥ 15 dBZ; radar 'zeroes' - Raw radial velocity - GOES-16 cloud-water path - GOES-16 clear sky radiances - GOES-16 atmospheric motion vectors - GOES-16 GLM data
Radiation LW/SW	RRTMG/RRTMG	Dudhia/RRTM, RRTMG/RRTMG	RRTMG/RRTMG
Microphysics	Thompson (aerosol aware)	NSSL 2-moment	NSSL 2-moment
PBL	MYNN	YSU, MYJ, or MYNN	MYJ
LSM	RUC (Smirnova)	RUC (Smirnova)	RUC (Smirnova)

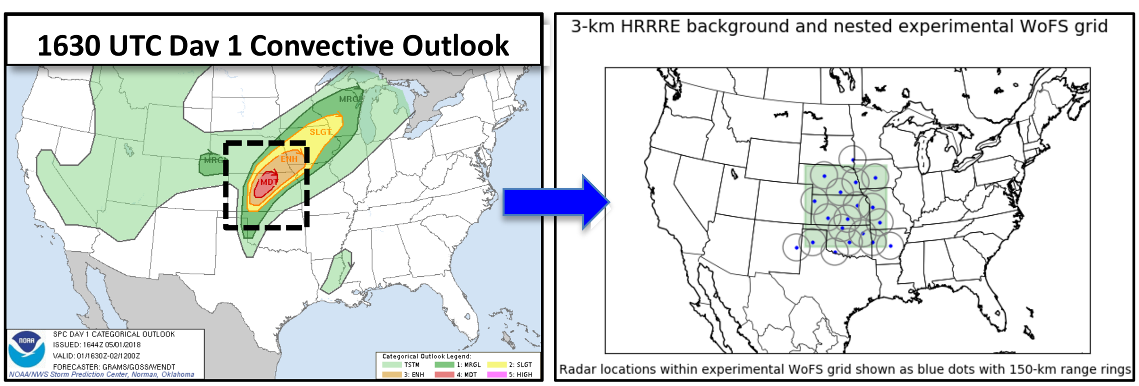

Figure 1 SPC 1630 UTC issued Day 1 convective outlook (left) and corresponding WoFS grid (right).

d) Iowa State University (ISU) Machine Learning-based Severe Wind Probabilities

Machine-learning-based tools will be used to derive probabilities that thunderstorm wind damage reports were truly due to severe intensity winds (50 knots or more). It is well-known that there are deficiencies in the way that estimated wind values are currently assigned to thunderstorm wind damage reports. Roughly 90% of all reports do not have a measured value, and instead are given an estimate, with an artificial spike in the frequency of 50 knot and 52 knot (60 mph) values. The 50 knot estimates often appear for reports involving tree damage, implying that many of these reports are not actually due to severe intensity winds.

Several machine learning algorithms were trained on thunderstorm wind damage reports that had a measured wind value assigned to them during the 2007-2017 period. Three approaches have been used, with one including radar data, one not using radar data, and one using algorithms trained on two different regions of the US (along with radar data). For each of the three approaches, output from two different algorithms will be presented. One will be an average ensemble, while the other will be the best single model determined from objective measures in ongoing testing (either gradient boosted machine, a generalized linear model, an artificial neural network, or a random forest). The training of these models utilized information from the Storm Report database, including textual damage reports, along with SPC mesoanalysis output for 31 weather parameters over a 200 x 200 km box centered on the storm reports at the nearest hour prior to the report occurrence, population density, elevation, land use data, and in some cases, bulk radar statistics within a 66 x 66 km box centered on the storm report every 15 minutes within the hour centered on the report time. Probabilities derived from each of these machine learning models will be available. An example is shown in Figure 2.

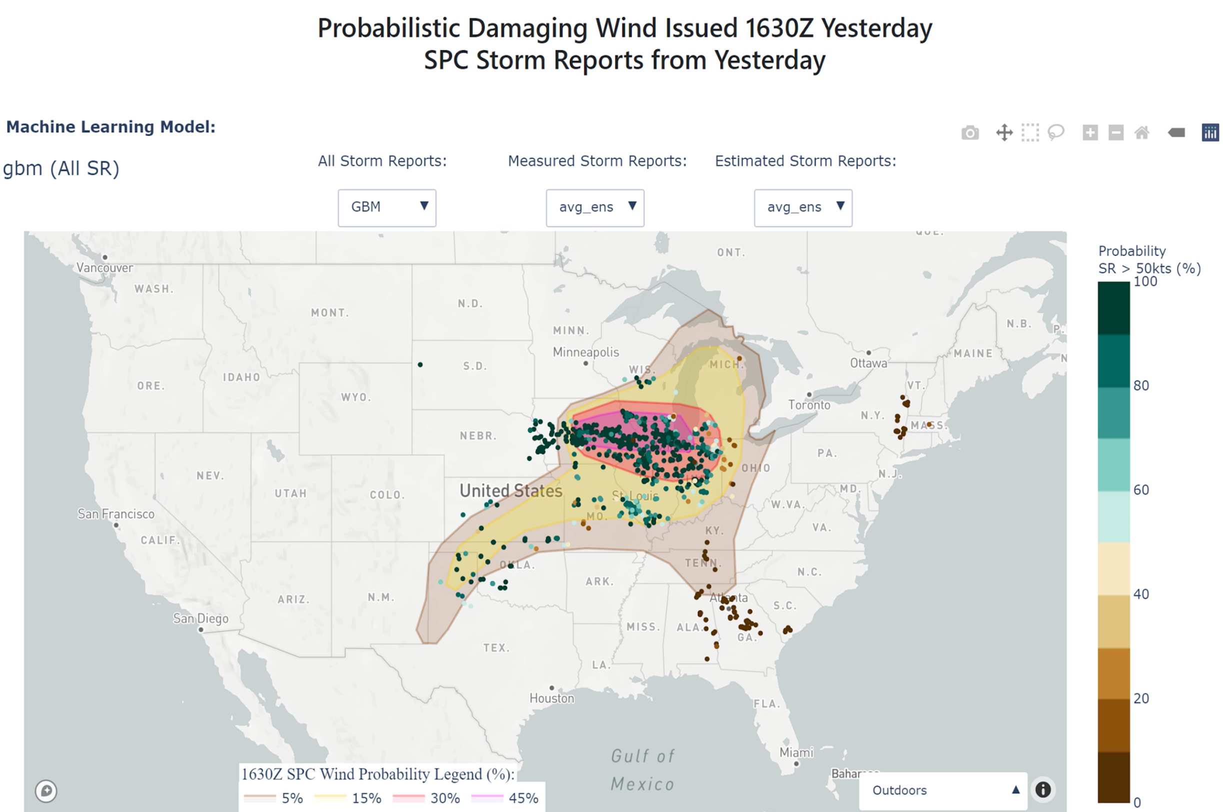

Figure 2 SPC Day 1 probabilities of damaging wind gusts (≥ 50 knots) within 40-km of a point (shaded)). The color of the points indicates the probability that the report was associated with an actual wind gust ≥ 50 knots.

e) Calibrated Forecast Products

i. NCAR CAM ML-derived convective mode probabilistic guidance (credit: Ryan Sobash)

The goal of this evaluation is to assess the utility of ML algorithms trained to provide probabilistic guidance of simulated storm mode using CAM output. Specifically, two trained ML models will be tested in 2021: 1) a supervised ML system that trains a convolutional neural network (CNN) to predict the mode of CAM storms using a hand labeled dataset of ~2000 CAM storms (CNN-labeled), and 2) a partially-supervised CNN system, that is trained with UH and clustered using a Gaussian mixture model (CNN-GMM). Both systems will be trained to provide probabilistic predictions of supercells, quasi-linear convective systems, and disorganized modes in the CAM output. The trained systems will ingest CAM storms from both a 3-km, 36-hr, deterministic, 00 UTC-initialized WRF forecast generated locally at NCAR, as well as simulated storms present within the 00 UTC HRRRv4. Evaluations will focus on the ability of the CNN and CNN-GMM to correctly classify storm modes based on subjective impressions by HWT participants, as well as assess differences in the two systems' predictions when using the local NCAR WRF vs. the HRRRv4 forecasts. Example output from CNN-labeled and CNN-GMM is provided in Figs. 3 & 4.

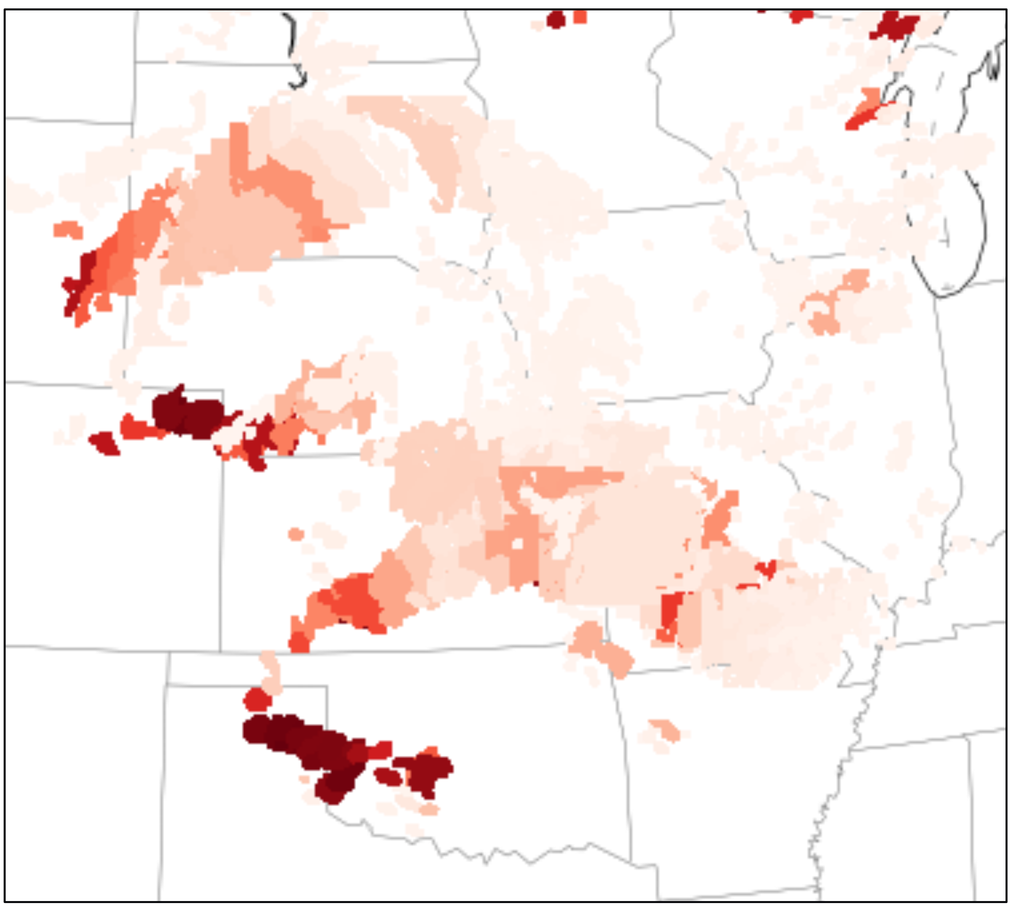

Figure 3 WRF CAM storm objects shaded according to the probability that each storm is a supercell generated by the CNN-labeled model described in the text (dark red shading indicates higher probability).

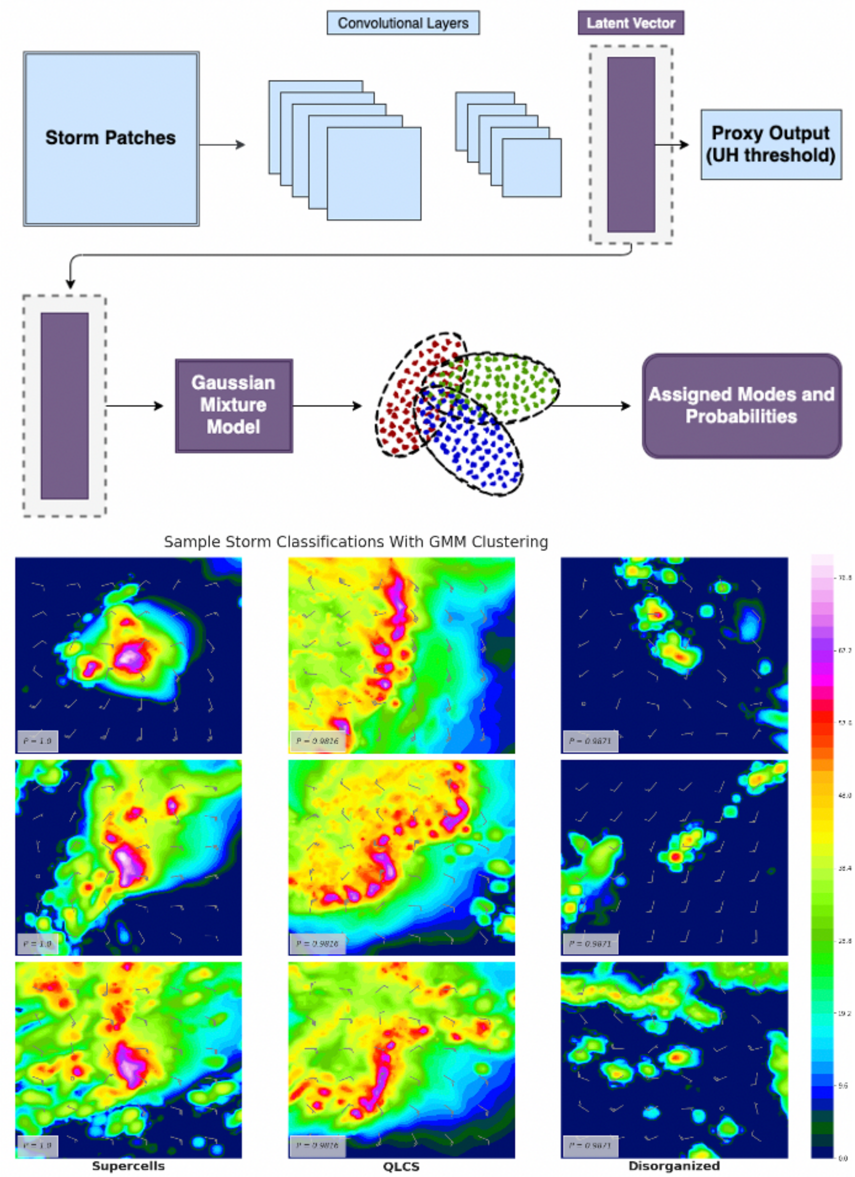

Figure 4 Summary of CNN-GMM system configuration and associated mode classifications.

ii. NCAR ML-derived HRRR-based convective hazard probabilities (credit: Ryan Sobash)

For the 2021 SFE, a neural network [NN] is being used to produce gridded probabilistic convective hazard guidance over the contiguous United States using the 00 UTC and 12 UTC HRRRv4. The NNs were trained with 42 base diagnostics (Table 18) output from a set of ~300 experimental 00 UTC HRRRX forecasts for events between 1 October 2019 and 2 December 2020. The diagnostics were upscaled to an 80-km grid and each grid point was labeled as a "hit" if a severe weather report occurred within a spatial and temporal neighborhood. Storm reports include the three report types, two significant report types, and a label if any report occurred. The temporal neighborhood for reports was fixed at 2-h, to produce hazard guidance within 4-h windows, while two spatial neighborhoods were tested (40 km and 120 km). The configuration details of the trained NNs are provided in Table 19. For comparison, a smoothed mid-level UH-based forecast will also be produced using a UH threshold of 75 m2/s2 and Gaussian smoother with $\boldsymbol{\sigma}$ = 160 km. Evaluation of the forecasts will be facilitated through HWT-generated comparisons, as well as a web-based visualization interface available here: https://www2.mmm.ucar.edu/projects/ncar_ensemble/camviewer/. An example 4-h all severe hazard forecast from 8 April 2020 is provided in Figure 5.

Table 18 The 42 base predictors used to train the NNs. The mean of the environmental and upper-air fields, and the maximum of the explicit fields, within each 80-km grid box, was used as input into the NNs. In addition, 132 neighborhood predictors were constructed by taking larger spatial and temporal means and maximums of the 15 environmental and 7 explicit fields resulting in a final set of 174 predictors used as input into the ML models.

Base Predictor	Type	Base Predictor	Type
Forecast Hour	Static	Surface pressure	Environment
Day of Year	Static	Most-unstable CAPE x 0-6km bulk wind difference	Environment
Latitude	Static	Significant tornado parameter	Environment
Longitude	Static	700 hPa–500 hPa lapse rate	Environment
Surface-based CAPE	Environment	Hrly-max 2–5km UH	Explicit
Most-unstable CAPE	Environment	Hrly-max 0–3km UH	Explicit
Surface-based CIN	Environment	Hrly-max 1 km relative vorticity	Explicit
Mixed-layer CIN	Environment	Hrly-max updraft speed below 400 hPa	Explicit
0-6km bulk wind difference	Environment	Hrly-max downdraft speed below 400 hPa	Explicit
Surface-based lifted condensation level	Environment	Hourly-max 10-m wind speed	Explicit
0-1km bulk wind difference	Environment	Hourly precipitation accumulation	Explicit
0-1km storm-relative helicity	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa zonal wind	Upper-air
0-3km storm-relative helicity	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa meridional wind	Upper-air
2-m temperature	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa temperature	Upper-air
2-m dew point temperature	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa dew point	Upper-air

Table 19 Settings used to construct and train the NNs.

Neural Network Hyperparameter	Value
Number of hidden layers	1
Number of neurons in hidden layer	1024
Dropout rate	0.1
Learning rate	0.001
Number of training epochs	10
Hidden layer activation function	Rectified Linear Unit
Output layer activation function	Sigmoid
Optimizer	Stochastic Gradient Descent
Loss function	Binary Cross-entropy
Batch size	1024
Regularization	L2
Batch normalization	On

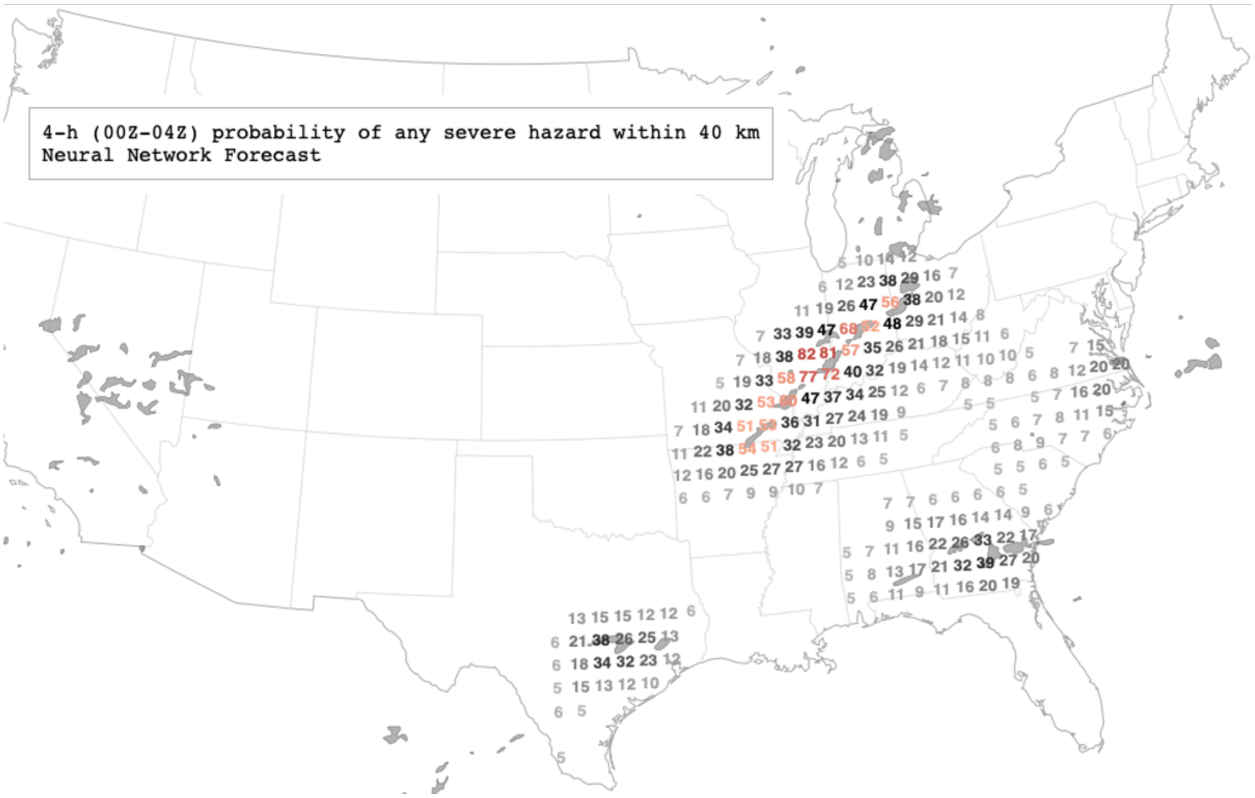

Figure 5 Neural network based probabilistic hazard forecast for the 4-h period between 00Z- 04Z 9 April 2020 based on a WRF forecast initialized at 00 UTC 8 April 2020. Numbers indicate the probability of any severe hazard occurring within 40-km of a grid point. Forecast reflectivity objects > 35 dBZ are overlaid.

iii) NSSL ML Random Forest Hazard Probabilities (credit: Eric Loken)

Automated “first guess” Day 1 (1200 UTC – 1200 UTC) forecast hazard probabilities are generated using random forests (RFs). Separate RFs predict the probability of tornadoes, severe wind, and severe hail, respectively, on an approximately 80 km grid over the CONUS.

The three RFs use the same set of predictors, which are derived from temporally-aggregated (i.e., daily maximum, minimum, or mean) ensemble mean forecast variables from the HREFv2.1. Predictors include simulated storm-, index-, and environment-related fields at the point of prediction and closest eight 80 km grid points as well as latitude and longitude (Table 20). Each RF is trained on ensemble mean HREFv2.1 data and observed SPC storm reports from 653 (non-continuous) days between April 2018 and May 2020. Although the RFs are trained on ensemble mean HREFv2.1 data, they use ensemble mean HREFv3 data for real-time prediction. The methods are similar to those described in Loken et al. (2020).

*Table 20 RF predictor fields. The temporal aggregation strategy for each variable is noted in parentheses. * denotes a derived quantity.*

Simulated Storm	Simulated Environment		Simulated Index	Lat/Lon
1 km Reflectivity (24h max.)	0-3 km Storm Relative Helicity (24h max.)	MUCAPE (24h mean)	Supercell Composite Parameter* (24h max.)	Latitude
Echo Top (24h max.)	0-1 km Storm Relative Helicity (24h max.)	MUCIN (24h mean)	Significant Tornado Parameter* (24h max.)	Longitude
Upward Vertical Velocity (24h max.)	2-m Temperature (24h mean)	SB/MUCAPE ratio* (24h mean)	Significant Hail Parameter* (24h max.)	-
Downward Vertical Velocity (24h min.)	2-m Dewpoint Temperature (24h mean)	700 – 500 hPa Lapse Rate* (24h mean)	0-1 km Energy Helicity Index* (24h max.)	-
2-5 km Updraft Helicity (24h max.)	2 m, 925 hPa, 850 hPa, 700 hPa, 500 hPa Dewpoint Depression* (24h mean)	Critical Angle Proxy* (At time of max. STP)	0-3 km Energy Helicity Index* (24h max.)	-
0-3 km Updraft Helicity (24h max.)	10 m – 500 hPa wind shear magnitude* (24h mean)	Max 10 m Wind Speed (24h max.)	Product of (MUCAPE) x (10 m – 500 hPa wind shear magnitude) * (24h max.)	-
Number of Grid Points With At Least 30 dBZ Simulated Reflectivity (At time of max. 2-5 km Updraft Helicity [if non-zero] or Upward Vertical Velocity)	10 m – 925 hPa wind shear magnitude* (24h mean)	10 m Wind Direction (At time of maximum 10 m wind speed)	Lifted Index (24h min.)	-

iv. Colorado State University (CSU) GEFS-based, ML-derived Hazard Probabilities (credit: Aaron Hill)

In the 2021 SFE, the Colorado State University Machine Learning Probabilities (hereafter known as CSU-MLP) system is forecasting severe weather hazards through the application of random forests (RFs). The CSU-MLP RFs are trained with approximately nine years of daily 0000 UTC initializations from the FV3 global ensemble forecast system reforecast dataset (FV3-GEFS/R) along with reports of severe weather. For consistency with SPC outlooks as well as SFE activities, RFs are trained separately for individual hazards in the day 1 and 2 timeframes, such that separate forecasts are issued for each hazard type (example forecast in Figure 6). The forecasts are analogous to operational SPC outlooks making them useful as guidance for experimental outlooks generated by SFE participants.

Predictors from the FV3-GEFS/R correspond to parameters expected to be related to severe weather occurrence, including bulk wind shear, convective available potential energy, low-level wind and thermodynamics, as well as derived quantities like lifting condensation level; all predictors are listed in Table 21. To be consistent across variables and times, all predictors are gridded to a 0.5 degree grid for preprocessing. Severe weather reports (i.e., storm data) are similarly gridded over the training period, where each point is labeled a 0, 1, or 2 for the occurrence of no severe report, a severe report, and a significant severe report. For every gridded event of severe weather across the contiguous United States, predictors are selected around the training point with spatiotemporal dimensions to capture any pre-existing dynamical model biases from the FV3-GEFS/R, which allows the RFs to learn predictor biases during training. Spatially, predictors are gathered within a latitudinal and longitudinal radius (set to 3 in these models) around the training point so each grid point represents a separate predictor. Temporally, this procedure is followed at each model output time over the forecast window; the new FV3-GEFS/R has 3-hourly output through day 10. For example, during the day-1 period, predictors are gathered 3-hourly from forecast hour 12 through hour 36, totaling nine predictor times. The predictor assembly results in approximately 6,500 predictors for each training point in which to build the RFs.

Table 21 Short-hand notation (left) and long description (right) of predictor variables used to train CSU-MLP severe weather RFs. Derived variables from FV3-GEFS/R output are denoted with an asterisk (*).

Predictor Acronym	Predictor Description
APCP	3-hourly accumulated precipitation
CAPE	Convective available potential energy
CIN	Convective inhibition
U10	10 m latitudinal wind speed
V10	10 m longitudinal wind speed
T2M	2 m temperature
Q2M	2 m specific humidity
MSLP	Mean sea level pressure
PWAT	Precipitable water
UV10	10 m wind speed
SRH03	0 - 3km storm relative helicity
SHEAR850*	0 - 850 hPa bulk wind shear
SHEAR500*	0 - 500 hPa bulk wind shear
ZLCL*	Height of lifting condensation level
RH2M*	2 m relative humidity

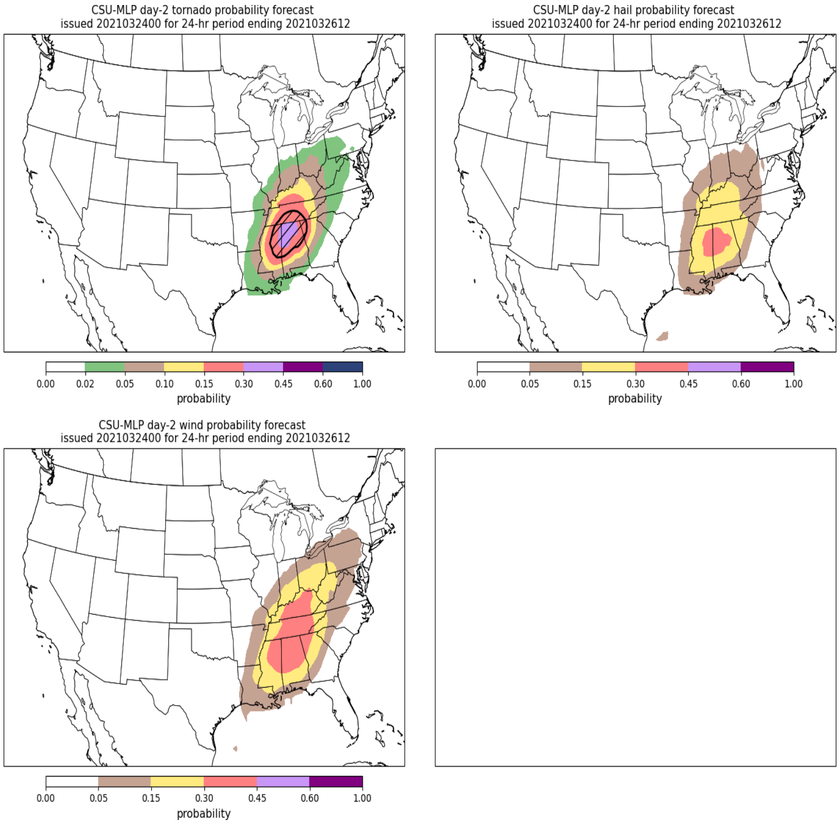

Figure 6 Probabilistic day-2 forecasts of (upper left) tornado, (upper right) hail, and (bottom left) wind hazards valid 1200 - 1200 UTC ending 26 March 2021. Hatched contours represent a 10% probability of significant severe hazards.

v. HREF/SREF Calibrated Severe Weather Probabilities (credit: Israel Jirak)

Probabilities valid over 4-h time windows are produced using the following procedure. At every grid-point for the valid forecast hour, two probabilities are paired: (1) Probability of UH $\geq$ 75/100/200 m^2/s^2 for the ARW/NMMB/FV3 cores over the previous 4 h (from the HREF), (2) Probability of environmental field(s) meeting a threshold over the previous 4 h (from the SREF; see Table 22 below). The historical frequency of a hazard report occurring within 25 miles of that grid point and within the 4 h period for that forecast pair of probabilities is substituted as the 4 h calibrated hazard probability.

Table 22 Environmental fields for each hazard used in the HREF/SREF calibrated probabilities.

Hazard	**SREF Conditions**
Tornado	STP ≥ 1
Hail	MUCAPE ≥ 1000 J/kg, Eff. Shear ≥ 20 kt
Wind	MUCAPE ≥ 250 J/kg, Eff. Shear ≥ 20 kt

To construct 24 h time window probabilities, the 4 h hazard probability forecasts that cover the 24 h convective day are used (i.e., 1200 – 1200 UTC). At every grid point, the cumulative sum of the 4 h probabilities and the maximum 4 h probability are paired. The historical frequency of a hazard report occurring within 25 miles of that grid point and within the 24 h period for those 4 h calibrated hazard probabilities is substituted as the 24 h calibrated hazard probability.

vi. STP-based tornado probabilities (STP Cal Circle; credit: Burkely Gallo)

Automated “first guess” tornado probabilities valid over 24 h periods are produced using the following procedure: A distribution of the significant tornado parameter (STP) is formed for each grid point from points where UH in the following hour exceeds the 99.985th percentile (within each HREF member's climatology) within a 40 km radius. The 10th percentile of STP from that distribution is then assigned to each point at each hour, and then the maximum daily STP value for each point is used to assign a probability based on the climatological frequency of a tornado given a right-moving supercell and an STP value for each ensemble member. The mean probability at each point is taken across the members, and then a Gaussian smoother with $\sigma = 50$ km is applied. A similar procedure is used to derive tornado probabilities valid over 4 h time periods. For further details, see Gallo et al. (2018).

vii. STP-based tornado probabilities (STP Cal Inflow; credit: David Jahn)

This alternative approach for deriving 24-h tornado probabilities follows the STP Cal Circle methodology except uses the 50th percentile of the STP distribution that is formulated from points within the inflow region relative to a point, rather than over the surrounding 40-km circular region. The inflow area is defined as a quadrant region of 40-km radius that is centrally oriented along the direction of the environmental wind at 1 km AGL.

viii. STP and UH ML-based tornado probabilities (ML NN; credit David Jahn)

Tornado probabilities are calculated with a ML model using as predictors the 10th, 50th, and 90th percentiles from the STP distributions as described above for the STP Cal Inflow product. Predictors also include the 96th, 98th, and 99th values of the UH distribution from a 40-km circular region about a given point as well as the UH (2-5 km) and UH (0-3 km) values at the same point. This ML system is unique because it uses training data of spatial resolution consistent with the native HREF 3-km grid as opposed to a 40-km or 80-km grid used by other ML methods. This system trains separate neural network classification models for each of the 10 HREF ensemble members and combines the results using a neural network regression technique.

ix. ML-based, Random Forest Hail Probabilities (credit: Amanda Burke)

RF hail forecast products are produced hourly, from 12-36 h of forecast time, for the HREF ensemble. The products predict probability of severe hail (diameter > 1") and probability of significant-severe hail (diameter > 2"). In addition, the products provide an estimate of maximum predicted hail size to be used in the display of the products during the HWT SFE.

The RF hail forecast products use a multi-step machine learning process. First, storm objects are identified within the HREF ensemble forecast by examining the maximum updraft velocity. For each storm object, input data are extracted from approximately 25 2-D input fields in the HREF forecast, including a mixture of storm-related variables (e.g., hourly maximum radar reflectivity) and environmental variables (e.g. 500 hPa wind, 850 hPa dewpoint temperature). These statistics are used to predict hail associated with the storm objects using a multi-step machine learning process.

In the first step, a random forest classification model predicts whether each storm object will produce hail. Those storm objects predicted to produce hail move to the second step--a random forest regression model which predicts the distribution of hail diameters within each hail-producing storm object. Finally, an isotonic regression step is applied to calibrate output hail probabilities to those estimated from radar-observed MESH in training data. This process is performed for each member of the HREF, and the resulting ensemble of ML hail predictions is used to generate probabilistic forecast products.

The training data used include HREF forecasts and MRMS MESH observations from the spring and summer months of 2017-2020. Temporal weighting is applied to more heavily weight hail events occurring during the month of May to account for seasonal variation in hail (this type of temporal weighting has been found to increase forecast skill).

x. ML-based, Deep Learning Hail Probabilities (credit: Nate Snook)

The UNET hail forecasts are an alternate kind of machine learning hail forecast product produced using the HREF. The UNET forecast products produce hourly predictions of maximum hail size for 12-36 h of forecast time. As for the RF hail forecast products, a separate machine learning model is run for each member, and the resulting ensemble of machine learning forecasts is used to generate the final forecast products.

Unlike the RF hail forecast products, which use a random forest machine learning architecture which relies upon an ensemble of decision trees, the UNET hail forecast products use the recently-developed UNET architecture. UNET is a deep learning architecture which uses a combination of convolution, pooling, and upscaling layers to generate point-by-point predictions (as opposed to the object-based predictions of the RF hail forecast products). The UNET architecture allows the model to learn to identify structures over a range of spatial scales, as well as to predict areas of potential hail threat outside of areas where the HREF ensemble predicted storms (because UNET does not rely upon storm objects).

To produce UNET hail forecasts, a set of approximately 10 2-D input fields are used from the HREF ensemble forecasts, focusing primarily on environmental fields (including 500 hPa and 850 hPa wind, temperature, and moisture). As with the RF hail forecast products, the UNET products are trained using 2017-2020 spring and summer HREF ensemble forecast data and MRMS MESH observations.

Input data from the HREF are considered in overlapping 64 by 64 grid-point patches (using 8 grid-points of overlap to avoid discontinuities at patch boundaries). These patches are stitched together to produce the final CONUS-wide forecast product. This is the first year that UNET forecast products have been produced; as such they have not been refined and calibrated as extensively as the RF hail forecast products have been.

xi. SPC Timing Guidance (credit: Israel Jirak)

SPC Timing Guidance products (valid over 4 h time windows) are generated for tornadoes, wind, and hail using a temporal disaggregation method with HREF/SREF calibrated guidance as applied to the operationally issued SPC convective outlooks at 0600 and 1300 UTC (Jirak et al. 2012, 2020). Thus, they are a blend of the human forecast and the first-guess calibrated guidance.

xii. Machine-Learning calibrated WoFS probabilities (credit: Monte Flora)

A series of machine learning (ML) models are being developed to provide rapidly updating probabilistic guidance to human forecasters for short-term (e.g., 0-3 h) severe weather forecasts. We generated the feature inputs into the ML models from Warn-on-Forecast System (WoFS) forecasts. Rather than producing a gridded ML product as with next-day (12-36 hr) convection-allowing model (CAM) products (e.g., Burke et al. 2019; Loken et al. 2020; Sobash et al. 2020; Hill et al. 2020), the current method produces object-based predictions that are interpreted in an event-based framework—What is the likelihood that a given storm will produce a hazard—as opposed to spatial probabilities (what is the likelihood of a hazard occurring within some prescribed distance of a point?; Fig. 1). The objects in this case are ensemble storm tracks which—conceptually—are regions bounded by the ensemble forecast uncertainty in storm location (determined by 30-min updraft tracks). An ensemble storm track can be composed of a single ensemble member's storm track or some combination of up to all 18 ensemble members. We trained random forests, gradient-boosted trees, and logistic regression algorithms to predict which WoFS 30-min ensemble storm tracks will overlap a tornado, severe hail, and/or severe wind report. For the SFE, we will be solely highlighting the logistic regression model as it performed the best on an independent dataset.

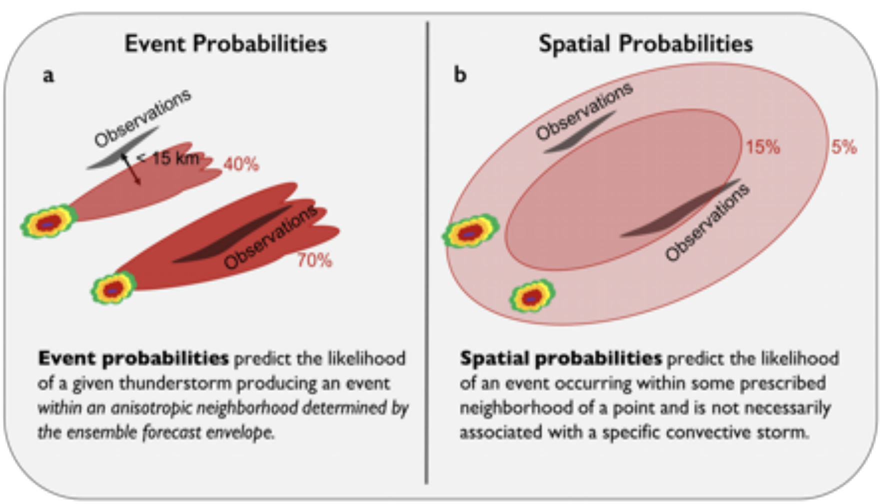

Figure 7 Illustration of the distinction between event and spatial probabilities (Fig. 2 of Flora et al. 2019).

The feature inputs were based on intra-storm and environmental variables from the WoFS and morphological variables describing the storm objects (Table 23).

Table 23 Input variables from the WoFS. The asterisk (*) refers to negatively oriented variables. Values in the parentheses indicate those variables that are extracted from different vertical levels or layers.

Intra-storm	Environment	Object Properties
Updraft Helicity (0-2 km, 2-5 km)	Storm-Relative Helicity (0-1 km, 0-3 km)	Area
Cloud Top Temperature*	75 mb Mixed-layer CAPE	Eccentricity
0-2 km Avg. Vertical Vorticity	75 mb Mixed-layer CIN	Orientation
Composite Reflectivity	75 mb Mixed-Layer LCL	Minor axis length
1-3 km Maximum Reflectivity	75 mb Mixed-Layer Equivalent Potential Temperature	Major axis length
3-5 km Maximum Reflectivity	U Shear (0-6 km, 0-1 km)	Extent
80-m wind speed	V Shear (0-6 km, 0-1 km)	Initialization Time
10-500 m Bulk Wind Shear	10-m U	
10-m Divergence*	10-m V	
Column-maximum Updraft	Mid-Level Lapse Rate	
Column-minimum Downdraft*	Low-level Lapse Rate	
Low-level updraft (1 km AGL)	Temperature (850, 700, 500 mb)	
HAILCAST maximum hail diameter	Dewpoint Temperature (850, 700, 500 mb)	
Cold Pool Buoyancy*	Geopotential Height (850, 700 500 mb)	

From these variables, we computed ensemble statistics as input features (more details in Flora et al. 2021). We show an example hail forecast from the logistic regression model in Fig. 8. Each object is a composite of ensemble member forecast tracks of a storm, colored according to the probability of the storm producing a severe hail report. It is best to think of these objects like warning polygons—most likely the event would occur within these bounds—, but with an associated probability of occurrence. For example, the ML model predicts there is a nearly 50% chance that the supercell over the Western Red River Valley will produce severe hail in the next hour.

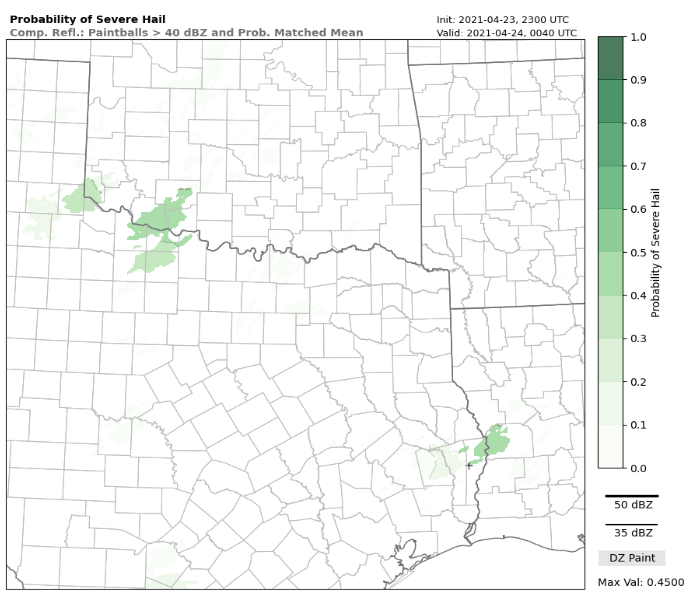

Figure 8 Example forecast from the severe-hail-based logistic regression.

3. SFE 2021 Core Interests and Daily Activities

2021 SFE activities will occur from 9am-4pm CDT, with a lunch break from 12:30-2pm CDT. Tables 24 and 25 provide a schedule for Monday, and Tuesday-Friday, respectively. Further details are provided in subsequent sections.

Table 24 Schedule for Monday.

Time (CDT)	**R2O Group**		**Innovation Group**	
9:00 AM – 9:45 AM	**Welcome and Introductions** Israel Jirak & Participants			
9:45 AM – 10:30 AM	**HWT SFE Scientific Objectives and Goals** Israel Jirak & Adam Clark			
10:30 AM - 11:00 AM	***Break*** Fill out IRB Consent Form and distribute survey link			
11:00 AM - 11:15 AM	**Conditional Intensity Forecasting Overview** Israel Jirak			
11:15 AM – 11:30 AM	**Weather Briefing** David Imy			
11:30 AM – 12:30 PM	**Issue *Day 1* Hazards Coverage and Conditional Intensity Forecasts (2 groups)**		**Issue *Day 2* Hazards Coverage and Conditional Intensity Forecasts (2 groups)**	
	12z HREF	*12z GSL RRFS*	*No CAMs*	*All data (incl. CAMs)*
12:30 PM – 2:00 PM	***Lunch/Break***			
2:00 PM – 2:15 PM	**Update on Today's Weather** David Imy			
2:15 PM – 3:15 PM	**Issue MD Product**		**Issue 1-h outlooks (22-23, 23-00Z)**	
	WoFS & obs		*WoFS*	*No WoFS*
3:15 PM – 4:00 PM	**Update Day 1 Outlook**		**Issue 1-h outlooks (22-23, 23-00, 00-01Z)**	
	WoFS & other guidance		*WoFS*	*No WoFS*

Table 25 Schedule for Tuesday – Friday.

<table>
<tr><td>Time (CDT)</td><td colspan="2">R2O Group</td><td colspan="2">Innovation Group</td></tr>
<tr><td>9:00 AM – 9:15 AM</td><td colspan="4">Overview of Yesterday’s Severe Weather
David Imy</td></tr>
<tr><td rowspan="2">9:15 AM – 11:00 AM</td><td colspan="4">Evaluation Orientation, Individual Working Time, and Discussion</td></tr>
<tr><td>Group A: Calibrated Guidance</td><td>Group B: Deterministic CAM</td><td>Group C: CAM Ensembles</td><td>Group D: Medley</td></tr>
<tr><td>11:00 AM - 11:15 AM</td><td colspan="4">Break</td></tr>
<tr><td>11:15 AM – 11:30 AM</td><td colspan="4">Weather Briefing
David Imy</td></tr>
<tr><td rowspan="2">11:30 AM – 12:30 PM</td><td colspan="2">Issue Day 1 Hazards Coverage and Conditional Intensity Forecasts (2 groups)</td><td colspan="2">Issue Day 2 Hazards Coverage and Conditional Intensity Forecasts (2 groups)</td></tr>
<tr><td>12z HREF</td><td>12z GSL RRFS</td><td>No CAMs</td><td>All data (incl. CAMs)</td></tr>
<tr><td>12:30 PM – 2:00 PM</td><td colspan="4">Lunch/Break</td></tr>
<tr><td>2:00 PM – 2:15 PM</td><td colspan="4">Update on Today’s Weather
David Imy</td></tr>
<tr><td rowspan="2">2:15 PM – 3:00 PM</td><td colspan="2">Issue MD Product</td><td colspan="2">Issue 1-h outlooks (22-23, 23-00Z)</td></tr>
<tr><td colspan="2">WoFS & obs</td><td>WoFS</td><td>No WoFS</td></tr>
<tr><td rowspan="2">3:00 PM – 4:00 PM</td><td colspan="2">Update Day 1 Outlook</td><td colspan="2">Issue 1-h outlooks (22-23, 23-00, 00-01Z)</td></tr>
<tr><td colspan="2">WoFS & other guidance</td><td>WoFS</td><td>No WoFS</td></tr>
</table>

a. Formal Evaluation Activities

SFE 2021 will feature one period of formal evaluation from 9:15-11:00am CDT Tuesday-Friday. The evaluations will be done virtually and involve comparisons of different ensemble diagnostics, CLUE ensemble subsets, HREF, and WoFS. Additionally, the evaluations of yesterday’s experimental forecast products will be conducted during this time. Participants will be split into Groups A, B, C, & D, which will each conduct a separate set of evaluations. In each group, from 9:15-9:25am CDT, a short tutorial will be presented to instruct and familiarize participants with the evaluations in their respective groups, and then from 9:25-10:15am CDT, participants will conduct the evaluations independently while facilitator remain available for questions. Finally, from 10:15-10:45am CDT, each group will reconvene in a virtual meeting to discuss various aspects of the just-completed evaluations (e.g., interesting observations, notable differences in performance, etc.), and from 10:45-11am CDT the evaluations of yesterday’s forecasts will be discussed. The four different sets of evaluations are summarized below:

Group A – Calibrated Guidance

A1. Calibrated Guidance

a. Day 2 Calibrated Tornado Guidance

Four different methods for deriving calibrated Day 2 tornado guidance are subjectively rated. These methods were described in the previous section and include: (1) HREF/SREF calibrated (12Z), (2) CSU MLP (00Z), (3) HRRR NCAR, and (4) STP Cal Circle.

b. Day 1 Calibrated Tornado Guidance

The same methods as in the Day 2 evaluation are rated for Day 1, except for STP Cal Circle.

c. 00Z HREF 24-h Calibrated Tornado Guidance

Five different methods based on HREF for generating 24-h calibrated tornado guidance are subjectively evaluated: (1) HREF/SREF Calibrated, (2) STP Cal Circle, (3) STP Cal Inflow, (4) ML Random Forest, and (5) ML NN.

d. 00Z HREF 4-h Calibrated Tornado Guidance

Four different sets of calibrated tornado guidance valid in 4-h time windows are subjectively rated. This guidance includes: (1) HREF/SREF Calibrated, (2) STP Cal Circle, (3) 06Z Day 1 SPC Timing Guidance, and (4) 12Z Day 1 SPC Timing Guidance.

e. Day 2 Calibrated Hail Guidance

Three different methods for deriving calibrated Day 2 hail guidance are subjectively rated. These methods include: (1) HREF/SREF Calibrated (12Z), (2) CSU MLP (00Z), and (3) HRRR NCAR (12Z).

f. Day 1 Calibrated Hail Guidance

The same methods as in the Day 2 evaluation are rated.

g. 00Z HREF 24-h Calibrated Hail Guidance

Four different methods based on HREF for generating 24-h calibrated hail guidance are subjectively evaluated: (1) HREF/SREF Calibrated, (2) ML Deep Learning, (3) ML Random Forest (Loken), and (4) ML Random Forest (Burke).

h. 00Z HREF 4-h Calibrated Hail Guidance

Five different sets of calibrated hail guidance valid in 4-h time windows are subjectively rated. This guidance includes: (1) HREF/SREF Calibrated, (2) ML Deep Learning (Snook), (3) ML Random Forest (Burke), (4) 06Z Day 1 SPC Timing Guidance, and (5) 13Z Day 1 SPC Timing Guidance.

i. Day 2 Calibrated Wind Guidance

Three different methods for deriving calibrated Day 2 wind guidance are subjectively rated. These methods include: (1) HREF/SREF Calibrated (12Z), (2) CSU MLP (00Z), and (3) HRRR NCAR (12Z).

j. Day 1 Calibrated Wind Guidance

The same methods as in the Day 2 evaluation are rated.

k. 00Z HREF 24-h Calibrated Wind Guidance

Two methods based on HREF for generated 24-h calibrated wind guidance are subjectively evaluated: (1) HREF/SREF Calibrated, and (2) ML Random Forest (Loken).

l. 00Z HREF 4-h Calibrated Wind Guidance

Three different sets of calibrated wind guidance valid in 4-h time windows are subjectively rated. This guidance includes: (1) HREF/SREF Calibrated, (2) 06Z Day 1 SPC Timing Guidance, and (3) 12Z Day 1 SPC Timing Guidance.

Primary Science Question: What are the best approaches and techniques to develop calibrated hazard probabilities?

Group B – Deterministic CAMs

B1. CLUE: Deterministic Flagships

This activity will focus on assigning ratings to gauge the skill and utility of the primary deterministic CAMs provided by each SFE collaborator – GFDL (*GFDL FV3*), NSSL (*NSSL FV3-LAM*), EMC (*EMC FV3-LAM*), and GSL (*GSL FV3-LAM*). These runs will be compared to the operational HRRRv4, which was developed by GSL. 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.

Primary Science Question: How do the deterministic CAM runs using the FV3 dynamic core compare to the operational standard for convective forecasting (i.e., WRF-ARW-based HRRRv4)?

B2. CLUE: FV3-LAM Expanded North American Domain

a. Day 2

Two pairs of runs are evaluated for the Day 2 forecast period (i.e., hours 36-60) that are similar except one uses a CONUS domain (EMC FV3-LAM & GSL FV3-LAM) and the other uses an expanded North American domain (EMC FV3-LAMX & GSL FV3-LAM-NA), which is planned for future versions of the RRFS. 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.

b. Day 1

This evaluation is the same as B2.i., except for the Day 1 forecast period.

Primary Science Question: Are there benefits, or noticeable differences, in FV3-LAM forecasts at Days 1 & 2 when using an expanded North American domain (i.e., lateral boundary conditions farther from forecast area of interest) compared to a CONUS domain?

B3. CLUE: FV3-LAM Data Assimilation

Five deterministic FV3-LAM configurations are examined that incorporate different data assimilation strategies: (1) EMC FV3-LAM, (2) EMC FV3-LAMDAX, (3) GSL FV3-LAM, (4) MAP RRFS Control, and (5) MAP RRFS VTS Control. The EMC FV3-LAM and EMC FV3-LAMDAX are similarly configured, except EMC FV3-LAM uses a cold start while EMC FV3-LAMDAX is initialized from a 6-h data assimilation cycle with hourly analysis updates from RAP observations (the DA cycle is cold-started from a 6-h GDAS forecast). The GSL FV3-LAM uses hourly-cycled data assimilation using a hybrid 3DEnVar (GDAS ensemble) analysis and periodically drawing initial and boundary conditions from a 13-km North American FV3-LAM hourly-cycled configuration using the same RAP/HRRR physics suite. Finally, the two MAP runs both use hybrid 3DEnVAR, but MAP RRFS VTS Control uses valid time shifting while MAP RRFS Control does not use valid time shifting. Particular attention will be given to simulated storm structure, convective evolution, and location/coverage of storms, especially at the beginning of the forecast to directly assess the impact of the data assimilation. Storm surrogate fields, like hourly maximum updraft helicity, will also be examined to gauge their utility for forecasting severe storms.

Primary Science Question: What are the optimal data assimilation strategies in FV3-LAM configurations for convective weather forecasting?

B4. CLUE: FV3 Physics Suites

Three different physics suites are used in the membership of the RRFS Cloud ensemble. Since each suite uses the same set of ICs/LBCs, this allows a controlled comparison in which we can evaluate the impact of the differences in physics. Members 1, 4, & 7 of the RRFS Cloud ensemble are compared, which use Thompson/MYNN, GFDL/TKE-EDMF, and NSSL/hybrid-EDMF for the microphysics/PBL schemes, respectively. 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.

Primary Science Question: What is the optimal physics package in FV3-LAM for convective weather forecasting?

B5. CLUE: FV3 Stochastic Physics

Within each physics package of the RRFS Cloud membership, three different strategies for stochastic physics perturbations are implemented: (1) no stochastic perturbations, (2) SPPT perturbations, and (3) SPPT/SHUM/SKEB perturbations. These perturbation strategies correspond to RRFS Cloud members 1, 2, & 3, respectively. 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. Note, this is not technically a "controlled" comparison since these members also have differences in their ICs/LBCs, but the IC/LBCs for members 2 & 3 (initialized from GEFS members) should be statistically indistinguishable over a sufficiently large sample.

Primary Science Question: Are there obvious differences and/or advantages/disadvantages that can be attributed to the different stochastic physics perturbation strategies?

Group C – CAM Ensembles

C1. CLUE: 00Z CAM Ensembles

This evaluation will compare four 00Z initialized, FV3-LAM CAM ensembles to HREFv3. Specifically, (1) GSL RRFS, (2) RRFS Cloud, (3) MAP RRFS, and (4) MAP RRFS VTS will be compared. Each of these datasets has a unique configuration strategy, so the primary goal is to find which strategy is optimal and how it performs relative to HREFv3.

Primary Science Question: What are the best ensemble configuration strategies for FV3-LAM based CAM ensembles, and how do they compare to HREFv3?

C2. CLUE: 12Z CAM Ensembles

Three 12Z initialized ensembles are compared to HREFv3: (1) GSL RRFS, (2) HRRRE-S, and (3) HRRRE-M.

Primary Science Question: How do stochastic physics and multi-physics approaches compare in the HRRRE, and what effect does that choice have on their performance relative to the HREFv3 and the GSL RRFS?

C3. Hourly Updating CAM Ensembles

In this evaluation, various strategies for producing CAM ensemble guidance after 12Z, but before 00Z, are examined. Specifically, five different strategies are compared to 12Z HREF: (1) 12Z HRRR-TL,

(2) 15Z HRRR-TL, (3) 18Z HRRR-TL, (4) 18Z HREF/HRRR-TL Blend, and (5) 18Z Error-Weighted Blend. The HRRR-TL comprises the four most recent hourly runs of the HRRR, weighted equally. In the time-based blend, the HREF's weight is the ratio of the HRRR-TL forecast's lead time to that of the HREF forecast, so that each new run of the HRRR-TL receives linearly increasing weight as the HREF ages. For example, the 18Z blend valid at 00Z is 50% 12Z HREF, 50% 18Z HRRR-TL. The error-based blend combines the 10 HREF members and 4 HRRR-TL members into a single ensemble. Each member receives weight based on the sum of its normalized domain-wide RMSEs in 2-m temperature, 2-m dewpoint, and 10-m wind component fields using RTMA at the blend initialization time as truth. The member with the largest errors on the SFE domain receives no weight and the member with the smallest errors receives maximum weight. This is motivated by a set of test cases in which these short-term errors were weakly negatively correlated with convective forecast skill at later times.

Primary Science Question: Are there optimal ways to produce updated and improved CAM ensemble guidance within the Day 1 forecast period in between HREF updates (i.e., 15Z to 03Z)?

C4. CLUE: VTS DA

In this evaluation, ensembles initialized at 2100 and 0000 UTC with and without the valid-time-shifting data assimilation strategy are compared for the first 12 hours of the forecasts.

Primary Science Question: Does the valid-time-shifting data assimilation strategy improve ensemble performance within the first 12 hours of the forecast?

C5. WoFS evaluations

a. WoFS vs. HRRR-TL

WoFS initializations at 2100 and 2300 UTC are compared to HRRR-TL ensembles based at the same times.

Primary Science Question: How does the WoFS perform relative to systems that are currently available operationally?

b. Deterministic WoFS

These comparisons will examine the WoFS deterministic 1.5 km grid-spacing hybrid data assimilation runs initialized at 2100 and 2300 UTC, which will be compared to a random member from the 3-km baseline WoFS configuration with the same physic configuration. The WoFS forecast viewer will be used for additional comparisons between the WoFS and WoFS-Hybrid systems.

Primary Science Questions: Does the WoFS-Hybrid, on average, perform better than the individual members of WoFS? Does the WoFS-Hybrid provide additional value relative to the WoFS?

c. Machine-Learning calibrated WoFS probabilities

Hazard probabilities are derived using predictors from the WoFS output. The activity will examine the utility of these probabilities, and participants will be given the opportunity to comment on the visualization strategy within the WoFS web-viewer.

Primary Science Question: Do WoFS-derived, machine-learning calibrated hazard probabilities provide value on top of the already available WoFS guidance products?

Group D – Medley

D1. ISU ML Severe Wind Probabilities

An evaluation will be conducted of six different techniques (i.e., two ML models, with and without radar data, and regional or CONUS training) to produce ML-based probabilities to estimate the likelihood that a damaging wind report was caused by wind ≥ 50 knots. The evaluations will focus on perceived usefulness of the output via comparison with SPC forecasts of severe wind probability, best methods to display the information, and subjective evaluation of three different ML techniques. The evaluation will be conducted on an external web page hosted by Iowa State University.

Primary Science Question: Can machine-learning approaches provide useful information regarding the likelihood of wind damage reports being associated with gusts ≥ 50 knots?

D2. NCAR ML Mode

This evaluation will assess the utility of ML algorithms trained to provide probabilistic guidance of simulated storm mode using CAM model output. Specifically, two trained ML models will be tested: 1) a supervised ML system that trains a convolutional neural network (CNN) to predict the mode of CAM storms using a hand labeled dataset of ~2000 CAM storms (CNN-labeled), and 2) a partially-supervised CNN system, that is trained with UH and clustered using a Gaussian mixture model (CNN-GMM). Evaluations will focus on the ability of the CNN and CNN-GMM to correctly classify storm modes based on subjective impressions by HWT participants, as well as assess differences in the two systems' predictions when using the local NCAR WRF vs. the HRRRv4 forecasts. The evaluation will be conducted on an external web page hosted by NCAR.

Primary Science Question: Can machine-learning be used to provide automated guidance on convective mode, and which machine-learning techniques work best?

D3. Mesoscale Analysis

a. Background

Mesoscale analyses using the HRRRv4 as background (EMC 3D-RTMA) are compared to analyses using the GSL FV3-LAM as background (GSL 3D-RTMA). Both systems use background error covariance

(BEC) information from the GDAS. In addition, both systems are upscaled to a 40-km grid for comparisons to the SPC surface objective analysis (sfcOA). The goal is to assess the utility of these analysis systems for situational awareness and short-term forecasting for convective-weather scenarios.

b. DA Frequency

Two mesoscale analysis systems that use different data assimilation frequencies are compared: (1) EMC 3D-RTMA with GDAS for BEC, and (2) EMC 15-min 3D-RTMA with HRRRDAS for BEC.

c. Storm Scale

WoFS-based "analyses" (actually 15-minute maximum forecasts) of 10-m and 80-m wind are compared to preliminary local storm reports, including gust measurements and estimates.

Primary Science Question: What are the optimal methods for producing quality mesoscale analyses for convective forecasting applications and can a high resolution, rapidly updating ensemble DA system serve as a verification source for severe winds?

D4. GEFS vs. SREF

a. Day 3

Several sets of environmental parameters (2-m Td, CAPE, shear) and ensemble fields (mean, spread, and probabilities), as well as calibrated thunder and severe thunderstorm guidance are compared between the GEFS and SREF systems for the Day 3 forecast period. As NOAA moves toward a more unified model production suite, the SREF is planned for retirement, but the GEFS must be able to demonstrate forecast skill comparable or better than the SREF.

b. Day 2

This evaluation is the same as for Day 3, but for the Day 2 forecast period.

Primary Science Question: Can the GEFS provide similar or improved forecast quality as the SREF during the Day 2 & 3 forecast period for severe weather applications?

b. Forecast Products and Activities

There will be two periods of experimental forecast activities during SFE 2021. The first will occur from 11:30am – 12:30pm CDT and will focus on providing individual hazard guidance, as well as more precise information on the intensity of specific hazards. As in previous years, we will split the participants into two groups, with those in the R2O group issuing products for Day 1 and those in the Innovation Group issuing products for Day 2. The experimental forecasts will cover a limited-area domain typically covering the primary severe threat area with a center-point selected base on existing SPC outlooks and/or where interesting convective forecast challenges are expected.

In both groups, the forecasts will be done as a group activity. The individual hazard forecasts will mimic the SPC operational Day 1 & 2 Convective Outlooks by producing individual probabilistic coverage forecasts of large hail, damaging wind, and tornadoes within 25 miles (40 km) of a point. The Day 1 outlooks will cover the period 1800 UTC to 1200 UTC the next day, while the Day 2 outlooks will cover the following 1200 – 1200 UTC period. Additionally, each group will issue conditional intensity forecasts of tornado, wind, and hail, in which areas are delineated with reports that are expected to follow a “normal”, “hatched”, or “double-hatched” distribution. These conditional intensity forecasts are similar to those issued during SFE 2020. When generating Day 1 Convective Outlooks, SPC forecasters draw probabilities that represent the chance of each hazard occurring within 25 miles of a point. Forecasters can also delineate “hatched” areas, which represent regions with a 10% chance or greater of significant severe weather (EF-2 or greater tornadoes, winds ≥ 65 kts, or hail ≥ 2-in.) within 25 miles of a point. Research by the SPC has shown that, as the forecast coverage of a hazard increases, the expected intensity of the verifying reports also increases. For instance, on days where a “hatched” area is drawn and the maximum tornado coverage is 10 or 15%, 17% of the observed tornadoes are significant. When a “hatched” area is drawn and the maximum tornado coverage is 30% or higher, 32% of observed tornadoes are significant. In other words, as the forecast tornado coverage increases, the observed tornadoes grow progressively more intense, regardless of how many tornadoes occur; preliminary results show a similar pattern for wind and hail. Therefore, current coverage forecasts include intensity information that is not explicitly communicated to users, so coverage forecasts and intensity forecasts could be better labeled/communicated. These results have been used to identify three conditional intensity probability distributions that can be forecast via examination of the atmospheric environment: “normal”, “hatched”, and “double-hatched”. In plain language, “normal” refers to a typical severe weather day, where significant severe weather is unlikely, “hatched” areas indicate where significant severe weather is possible, and “double-hatched” areas indicate where high impact significant severe weather is expected.

Within the R2O Group, one sub-group will use 12Z HREF guidance for issuing their Day 1 individual hazard and conditional intensity forecasts, while another sub-group will use 12Z GSL RRFS guidance. Within the Innovation Group, one sub-group will use only non-CAM-based guidance for their Day 2 forecasts, while another sub-group will use all available CAM and non-CAM guidance available within the Day 2 time period.

The second period of experimental forecasting activities will occur during the 2-4pm CDT time period. In the R2O group, the 2:15-3pm CDT time period will be devoted to an activity in which each participant will issue their own Mesoscale Discussion (MD) Product using WoFS and other available CAM guidance within the SFE Drawing Tool. Then, during the 3-4pm time period, each R2O group participant will use WoFS and other available guidance to update the Day 1 individual hazard coverage and conditional intensity forecasts for the period 2100 – 1200 UTC.

During the 2:15-4pm CDT time period in the Innovation Group, participants will generate severe hazard probabilities valid over 1-h time windows covering 2200-2300 UTC, 2300-0000 UTC, and 0000-0100 UTC. Two initial forecasts will be generated during the 2:15-3:15pm period, which will cover the 22-23Z and 23-00Z time windows. Then, during the 3:15-4pm period, the 22-23Z and 23-00Z periods will be updated, and one more outlook covering 00-01Z will be generated. For both sets of initial and final forecasts, two forecasters will use all available datasets including WoFS (Forecaster WOF 1 & 2), while two other forecasters will use all available datasets except for WoFS (Forecaster NOWOF 1 & 2). The No-WoFS forecasters will use the SFE viewer (https://hwt.nssl.noaa.gov/sfe_viewer/2021/forecast_tool) to generate forecasts, while the WoFS forecasters will use the WoFS viewer (https://wof.nssl.noaa.gov/realtime/). Forecasters using the SFE viewer will have access to the WoFS domain bounds, so that the forecast domain will be the same between the two groups. Additionally, two other groups of non-expert forecasters will issue forecasts with and without WoFS similarly to the expert forecasters, which will be combined into consensus forecasts (ConWoFS and ConNoWoFS, respectively).

These WoF activities are the fifth year the WoF Ensemble has been tested in the SFE to explore the potential utility of WoF products for issuing guidance between the watch and warning time scales (i.e. 0.5 to 6-h lead times). These activities 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; Rothfusz et al. 2018) initiatives. These efforts also support the transition to higher temporal resolution forecasts at the SPC.

Appendix A: List of scheduled SFE 2021 participants.

Week 1 3-7 May	Week 2 10-14 May	Week 3 17-21 May	Week 4 24-28 May	Week 5 1-4 June
Rick Garuckas (WFO MRX)	Nick Hampshire (WFO EWX)	Anna Lindeman (WFO BOI)	Heather Kenyon (WFO BUF)	Andrew Zimmerman (WFO AKQ)
John Wetenkamp (WFO ARX)	Francis Kredensor (WFO TFX)	Kevin Huyck (WFO DLH)	Emily McGraw (WFO CHS)	Michael Sporer (WFO RNK)
Pat Spoden (WFO PAH)	Steve Zubrick (WFO LWX)	Chad Entremont (WFO JAN)	Nate McGinnis (WFO ILN)	Tara Dudzik (WFO IND)
Kristen Cassady (WFO ILN)	Matthew Brady (WFO EWX)	Jack Settlemaier (NWS SRH)	Linda Gilbert (WFO MQT)	Nick Vertz (WFO BYZ)
Lee Robertson (WFO PHI)	Keith Sherburn (WFO UNR)	Nicholas Fenner (WFO JAN)	James Wood (WFO MKX)	Aidan Kuroski (WFO MKX)
Dirk Peterson (WFO OAX)	Eswar Iyer (WFO AKQ)	Jaclyn Anderson (Ritzman) (WFO MKX)	Eric Bunker (M-Th; WFO TAE)	Keith White (WFO EWX)
Stephen Harrison (WFO SJT)	Brian Carcione (WFO HUN)	Lizzie Tirone (ISU)	Maria Molina (NCAR)	Austin Coleman (TTU)
Evan Kuchera (USAF)	Ed Shimon (WFO ILX)	Clark Evans (UWM)	Lance Bosart (SUNY-Albany)	Chris Melick (USAF)
Greg Stumpf (NSSL)	Bill Gallus (ISU)	Dillon Blount (UWM)	Steve Weiss (ret. SPC)	Craig Schwartz (NCAR)
Chris Karstens (SPC)	Becky A.-Selin (AER)	Felicia Guarriello (WPO)	Harald Richter (BOM)	Mike Coniglio (NSSL)
Jamie Wolff (DTC)	Russ Schumacher (CSU)	Casey Davenport (UNCC)	Reid Strickler (USAF)	Derek Stratman (CIMMS)
Aaron Hill (CSU)	John Peters (Naval P.-Grad)	Roger Riggin (UNCC)	Gary Lackmann (NCSU)	Jidong Gao (NSSL)
Leigh Orf (Wisc)	J. Peters student #1	Kyle Struckmann (NWS NAM)	Trevor Campell (NCSU)	Lewis Kanofsky (AWC)
Gabrielle Gantos (NCAR)	J. Peters student #2	Nick Goldacker (NCSU; M. Parker student)	Jacob Radford (NCSU)	Kai-Chih Tseng (GFDL/Princeton)
Rob Hepper (AWC)	Dave Ahijevych (NCAR)	Andrew Winters (CU)	Jeff Beck (CIRA/GSL/DTC)	Tim Marchok (GFDL)
Chris Nowotarski (TAMU)	Ty Higginbotham (AWC)	Rebecca Baiman (CU)	Nat Johnson (GFDL)	Kelly Lombardo (PSU)
Matt Brown (TAMU)	Tomas Pucik (ESSL)	Alexandra A.-Frey (UW)	Kelton Halbert (Wisc)	Geoff Manikin (EMC)
Brice Coffer (NC State)	Francesco Battaglioli (ESSL)	Rohan Jain (UW)	Jacob Carley (EMC)	Matthew Pyle (EMC)
Chris MacIntosh (EMC)	Binbin Zhou (EMC)	Charlie Becker (NCAR)	Gang Zhao (EMC)	Kendall Junker (CAPS/OU)
Shun Liu (EMC)	Shannon Shields (EMC)	Logan Dawson (EMC)	Ben Blake (EMC)	Jana Houser (U. Ohio)
Xiaoyan Zhang (EMC)	Matthew Morris (EMC)	Annette Gibbs (EMC)	Nigel Roberts (UK Met)	Darby Johnson (U. Ohio)
Nick Silkstone (UK Met)	Stephen Gallagher (UK Met)	Travis Elless (EMC)	Matt Lehnert (UK Met)	Curtis Alexander (GSL)
Aurore Porson (UK Met)	Adrian Semple (UK Met)	Sebastian Cole (UK Met)	Steve Willington (UK Met)	Dan Dawson (Purdue)
David Dowell (GSL)	Chris Bulmer (UK Met)	Steve Willington (UK Met)	Eric James (GSL)	Allie Mazurek (CSU)
Sarah Trojniak (WPC)	Aaron Johnson (OU/MAP)	Nate Snook (CAPS)	John Brown (GSL)	Ben Henry (Princeton undergrad)
	Jeff Duda (GSL)	Xuguang Wang (OU/MAP)	Mike Baldwin (Purdue)	
	John Allen (CMU)	Terra Ladwig (GSL)	Geeta Nain (Purdue)	
		Ed Szoke (GSL)		
		Jordan Dale (WPO)		

SFE Facilitators: Adam Clark (NSSL), Israel Jirak (SPC), Dave Imy (retired SPC), Burkely Gallo (CIMMS/SPC), Kenzie Krocak (CIMMS/SPC/CRCM), Brett Roberts (CIMMS/SPC/NSSL), Kent Knopfmeier (CIMMS/NSSL), Andy Dean (SPC), Eric Loken (CIMMS/NSSL), David Harrison (CIMMS/SPC), David Jahn (CIMMS/SPC), Jacob Vancil (CIMMS/SPC), Jeff Milne (CIMMS/SPC), and Nathan Dahl (CIMMS/SPC).

Appendix B: 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. B1). 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),

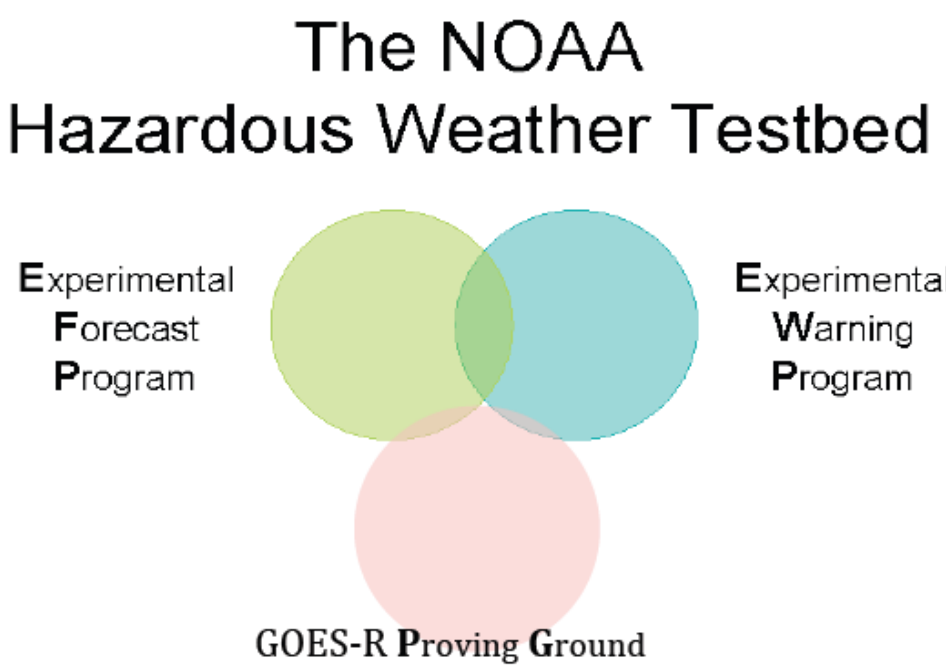

Figure B1: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

which involves using frequently updated short-range forecasts (≤ 1h lead time) from convection-resolving ensembles. This provides a natural overlap between the EFP and EWP activities.

The GOES-R Proving Ground (established in 2009) exists to provide demonstration of new and innovative products as well as the capabilities available on the next generation GOES-16 satellite. 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; 2018; 2020; 2021), and Gallo et al. (2017).

Appendix C: References

Burke, A., N. Snook, D. J. Gagne, S. McCorkle, and A. McGovern, 2019: Calibration of Machine Learning-Based Probabilistic Hail Predictions for Operational Forecasting. *Wea. Forecasting*, doi:10.1175/waf-d-19-0105.1.

Burke, A., N. Snook, D. J. Gagne, S. McCorkle, A. McGovern, 2020: Calibration of Machine Learning-Based Probabilistic Hail Predictions for Operational Forecasting, *Wea. Forecasting*, **35**, 149-168.

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. and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.,* **99**, https://doi.org/10.1175/BAMS-D-16-0309.1

Clark, A. J., and Coauthors, 2019: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Amer. Meteor. Soc.*, 0, (https://doi.org/10.1175/BAMS-D-19-0298.1).

Clark, A. J., and Coauthors, 2021: A Real-Time, Virtual Spring Forecasting Experiment to Advance Severe Weather Prediction. *Bull. Amer. Meteor. Soc.*, (In Press).

Flora, M. L., Potvin, C. K., Skinner, P. S., Handler, S., & McGovern, A., 2021: Using Machine Learning to Generate Storm-Scale Probabilistic Guidance of Severe Weather Hazards in the Warn-on-Forecast System, *Mon. Wea. Rev*., (In Press).

Flora, M. L., P. S. Skinner, C. K. Potvin, A. E. Reinhart, T. A. Jones, N. Yussouf, and K. H. Knopfmeier, 2019: Object-based verification of short-term, storm-scale probabilistic mesocyclone guidance from an experimental warn-on-forecast system. *Wea. Forecasting*, **34** (6), 1721–1739, doi:10.1175/WAF-D-19-0094.1.

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.

Gao, J., T. M. Smith, D. J. Stensrud, C. Fu, K. Calhoun, K. L. Monross, J. Brogden, V. Lakshmanan, Y. Wang, K. W. Thomas, K. Brewster, and M. Xue, 2013: A Real-Time Weather-Adaptive 3DVAR Analysis System for Severe Weather Detections and Warnings. *Wea. Forecasting*, **28**, 727-745.

Han, J., and C. S. Bretherton, 2019: TKE-based moist Eddy-Diffusivity Mass-Flux (EDMF) parameterization for vertical turbulent mixing, *Wea. Forecasting*, **34**, 869-886.

Harris, L. M., S. L. Rees, M. J. Morin, L. Zhou, and W. F. Stern, 2019: Explicit prediction of continental convection in a skillful variable-resolution global model. *Journal of Advances in Modeling Earth Systems*, **11(6)**, DOI:10.1029/2018MS001542.

Harris, L. M., and Coauthors, 2020: GFDL SHiELD: A Unified System for Weather-to-Seasonal Prediction. *Journal of Advances in Modeling Earth Systems*, **12(10)**, DOI:10.1029/2020MS002223.

Hill, A. J., G. R. Herman, and R. S. Schumacher, 2020: Forecasting Severe Weather with Random Forests. *Mon. Wea. Rev*.,148 (5), 2135–2161, doi:10.1175/MWR-D-19-0344.1.

Huang, B., and W. Wang, 2018: On the use of cost-effective valid-time-shifting (VTS) method to increase ensemble size in the GFS Hybrid 4DEnVar System. *Mon. Wea. Rev*., **146**, 2973-2998.

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.

Loken, E. D., A. J. Clark, and C. D. Karstens, 2020: Generating Probabilistic Next-Day Severe Weather Forecasts from Convection-Allowing Ensembles Using Random Forests. *Wea. Forecasting* (In review).

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. D. Karsten, G. J. Stumpf, and t. M. Smith, 2018: FACETs: A Proposed Next-Generation Paradigm for High-Impact Weather Forecasting. *Bull. Amer. Meteor. Soc.,* **99**, 2025-2043.

Sobash, R. A., G. S. Romine, and C. S. Schwartz, 2020: A Comparison of Neural-Network and Surrogate-Severe Probabilistic Convective Hazard Guidance Derived from a Convection-Allowing Model. *Wea. Forecasting*, **35** (5), 1981–2000, doi:10.1175/WAF-D-20-0036.1.

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., J. Gao, P. S. Skinner, K. Knopfmeier, T. Jones, G. Creager, P. L. Heinselman, and L. J. Wicker, 2019: Test of a Weather-Adaptive Dual-Resolution Hybrid Warn-on-Forecast Analysis and Forecast System for Severe Severe Weather Events. *Wea. Forecasting*, **34**, 1807-1827.

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.

Weiss, S. J., A. J. Clark, I. L. Jirak, C. J. Melick, C. Siewert, R. A. Sobash, P. T. Marsh, A. R. Dean, J. S. Kain, M. C. Coniglio, M. Xue, F. Kong, K. W. Thomas, J. Du, D. R. Novak, F. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, R. S. Schneider, and T. L. Jensen, 2010: An Overview of the 2010 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. *25th Conf. on Severe Local Storms*, Amer. Meteor. Soc., 7B.1.

Zhou, L., S.-J. Lin, J.-H. Chen, L. Harris, X. Chen, and S. L Rees, 2019: Toward Convective-Scale Prediction within the Next Generation Global Prediction System. *Bull. Amer. Meteor. Soc*. DOI:10.1175/BAMS-D-17-0246.1.