SPRING FORECASTING EXPERIMENT 2022

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

HWT Facility – National Weather Center

2 May – 3 June 2022

https://hwt.nssl.noaa.gov/sfe/2022/

Program Overview and Operations Plan

21 April, 2022

Adam Clark[2], Israel Jirak[1], Burkely T. Gallo[1,3], Brett Roberts[1,2,3], Kent Knopfmeier[2,3], Jake Vancil[1,3], David Jahn[1,3], Makenzie Krocak[3,4,5], Chris Karstens[1], Eric Loken[2,3], Nathan Dahl[1,3], David Harrison[1,3], David Imy[2], Andy Wade[1,3], Jeffrey Milne[1,3,4], Kimberly Hoogewind[2,3], Montgomery Flora[2,3], Joshua Martin[2,3], Brian Matilla[2,3], Joey Picca[1,3], Corey Potvin[2], Patrick Skinner[2,3], Patrick Burke[2]

(1) NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma
(2) NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma
(3) Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma
(4) School of Meteorology, University of Oklahoma, Norman, Oklahoma
(5) Institute for Public Policy Research and Analysis, 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 2022 HWT Spring Forecasting Experiment (SFE 2022), a cornerstone of the EFP, will be conducted 2 May – 3 June. Although we are transitioning back to in-person work, given continuing uncertainty related to the COVID-19 pandemic and the need for planning several months in advance, the 2022 SFE will be conducted virtually for the third consecutive year. However, it is expected that this will be the final year of completely virtual SFEs. Relative to SFE 2021, this year's virtual experiment will have a similar format with all participants participating in morning and afternoon forecasting activities, as well as next-day model evaluation activities. As in previous years, a suite of new and improved experimental CAM guidance contributed by our large group of collaborators will be central to these forecasting and model evaluation activities. These contributions comprise an ensemble framework called the Community Leveraged Unified Ensemble (CLUE; Clark et al. 2018). The 2022 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 2022 CLUE includes 60 members with 3-km grid-spacing, as well as a single member using 1-km grid-spacing. The SFE 2022 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 6th year to issue very short lead-time outlooks. This document summarizes the core interests of SFE 2022 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 2022 and Section 3 describes the core interests and new concepts being introduced for SFE 2022. A list of daily participants, details on the SFE forecasting, and more general information on NOAA's HWT are found in appendices.

2. Overview of Experimental Products and Models

Daily model evaluation activities will occur Tuesday through Friday from 9:15 – 11:00am (CDT) focusing on various CLUE subsets and calibrated guidance. The 2022 CLUE includes deterministic and ensemble forecasts using the most recent versions of the Finite Volume Cubed-Sphere Limited Area Model (FV3-LAM), and 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) and High Resolution Rapid Refresh version 4 (HRRRv4) will be examined as the operational modeling baselines. The rest of this section provides further details on each modeling system utilized in SFE 2022.

a) The 2022 Community Leveraged Unified Ensemble (CLUE)

The CLUE is a carefully designed ensemble with members contributed by NOAA units: NSSL, Environmental Modeling Center (EMC), Global Systems Laboratory (GSL), and Geophysical Fluid Dynamics Laboratory (GFDL); and research groups at the University of Oklahoma (OU): Multi-scale data Assimilation and Predictability (MAP) and Center for Analysis and Prediction of Storms (CAPS). CLUE members have 3-km grid-spacing and a CONUS domain, except for one member that has 1-km grid-spacing covering the eastern 2/3 of the CONUS. Depending on the CLUE subset, forecast lengths range from 18 to 126 h. Table 1 summarizes all 2022 CLUE contributions. Subsequent tables provide details on members in each subset.

Table 1 Summary of the 11 unique subsets that comprise the 2022 CLUE.

Clue Subset	# of mems	IC/LBC perts	Mixed Physics	Data Assimilation	Dynamical Core	Agency	Init. Times (UTC)	Forecast Length (h)	Domain
RRFSp1	1	none	no	Hybrid 3DEnVar	FV3	EMC/GSL	00, 12	60	CONUS
RRFSp2e	10	EnKF	no	Hybrid 3DEnVar	FV3	EMC/GSL	00	36	CONUS
MAP-VTS-rad	10	GFS, GEFS	no	GSI-EnVar	FV3	OU-MAP	00	36	CONUS
MAP-VTS-con	10	GFS, GEFS	no	GSI-EnVar	FV3	OU-MAP	00	36	CONUS
MAP-VTS-bot	10	GFS, GEFS	no	GSI-EnVar	FV3	OU-MAP	00	36	CONUS
NSSL-FV3-LAM	1	none	no	GFS cold start	FV3	NSSL	00	60	CONUS
NSSL3	1	none	no	GFS cold start	ARW	NSSL	00	30	CONUS
NSSL1	1	none	no	GFS cold start	ARW	NSSL	00	30	2/3 CONUS
GFDL-FV3	1	none	no	GFS cold start	FV3	GFDL	00	126	CONUS
RRFSp2eMP	10	EnKF	yes	Hybrid 3DEnVar	FV3	CAPS	00	84	CONUS
RRFSphys	6	none	yes	Hybrid 3DEnVar	FV3	CAPS	00	36	CONUS

Table 2 Specifications for the RRFSp1 (prototype 1) CLUE member. RRFSp1 is a 3-km convection-allowing, deterministic system featuring hourly analysis updates via a hybrid 3DEnVar data assimilation framework and uses a partial cycling capability similar to that employed by the North American Mesoscale Model (NAM) and its nests as well as RAP/HRRR. The hybrid 3DEnVar algorithm leverages the freely available EnKF members from the Global Data Assimilation System (GDAS) to provide flow-dependent information in the EnVar cost function; no 3-km ensemble information is used. However, a 3-km ensemble is used in RRFSp2. RRFSp1 is the control system for RRFSp2. RRFSp1 connects to RRFSp2 by providing high resolution, 3km central states at 1800 UTC to RRFSp2 around which Global Ensemble Forecast System perturbations are re-centered.

Member: RRFSp1	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Dynamical Core
RRFSp1	own	GFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3

Table 3 Specifications for the RRFSp2e ensemble. The RRFSp2e is a 10-member, 3-km grid-spacing CONUS FV3-LAM forecast system initialized 0000 UTC with forecasts to 36 h. The control member, RRFSp2, features hourly analysis updates via a hybrid 3DEnVar data assimilation framework using RRFSp1 as the initial state at 1800 UTC followed by six hours of cycling to 0000 UTC with the 3-km "RRFSDAS" ensemble that provides the flow-dependent information in the EnVar cost function during the hybrid analysis. The RRFSp2 hybrid analysis is used to recenter the EnKF ensemble mean each hour thereby forming the control member of the EnKF ensemble. The hybrid analysis also includes RAP/HRRR-like analysis components like adjustments to the soil temperature and moisture along with a non-variational cloud and precipitation hydrometeor analysis. RRFSp2e members 1-9 are perturbed forecasts initialized from the corresponding RRFSDAS members & also include stochastic parameter perturbations (SPP) applied to land-surface, PBL, and microphysics schemes.

Members: RRFSp2e	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Dynamical Core**
RRFSp2	18Z RRFSp1 central state, hourly 3km hybrid 3DEnVar	GFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e01	enkf_m01	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e02	enkf_m02	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e03	enkf_m03	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e04	enkf_m04	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e05	enkf_m05	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e06	enkf_m06	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e07	enkf_m07	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e08	enkf_m08	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3
RRFSp2e09	enkf_m09	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG	FV3

Table 4 Specifications for the MAP VTS-rad ensemble members. These 3-km grid-spacing ensemble forecasts are run with FV3-LAM and initialized by a multiscale GSI-based hybrid EnVar DA system directly assimilating both conventional and radar reflectivity observations (Wang and Wang 2017, 2021). The base ensemble size is 36 members, initialized daily at 1800 UTC from the GEFS. The EnVar control member is initialized from the 1800 UTC GFS control. LBCs are provided by re-centering GEFS around the GFS control. The system assimilates both operational RAP/HRRR in-situ data stream and MRMS radar reflectivity hourly during 1900-0000 UTC over the CONUS domain. This system includes valid time shifting (VTS; e.g. Gasperoni et al. 2022), which triples the ensemble size for radar EnVar component by including 36-member output 60-min before and after the each analysis time into the background error covariances. The 108-member VTS-expanded ensemble covariances not only mimics the effect of directly tripling ensemble size, but also includes information of model timing/phase uncertainty in convective systems. The base ensemble members are updated separately with the EnKF and recentered around the final control EnVar analysis. A 10-member 36-h ensemble free forecast is initialized at 0000 UTC from the final control analysis (MAP-VTS-rad_01) and 9 recentered ensemble members (MAP-VTS-rad_02-10). All members the "FV3_HRRR" CCPP physics suite. The 0000 UTC forecast also includes stochastic physics (SPPT/SKEB/SHUM/SPP). This configuration applies VTS ***only*** *for the radar observations DA component.*

Members: MAP-VTS-rad	**ICs**	**LBCs**	**VTS component**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Dynamical Core**
MAP-VTS-rad_01	EnVar	GFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_02	rEnKF_m1	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_03	rEnKF_m2	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_04	rEnKF_m3	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_05	rEnKF_m4	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_06	rEnKF_m5	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_07	rEnKF_m6	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_08	rEnKF_m7	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_09	rEnKF_m8	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-rad_10	rEnKF_m9	GEFS	Radar	Thompson	MYNN	RUC	RRTMG	FV3

*Table 5 Specifications for the MAP VTS-con ensemble members. This configuration applies VTS **only** for the conventional observations DA component with a 60-min shifting window to account for time-related model uncertainty of the mesoscale environment. It does not use VTS for the radar (storm-scale) DA component. All other aspects of this DA system match the description in Table 5. This ensemble will only be examined post-experiment.*

Members: MAP-VTS-con	ICs	LBCs	VTS component	Microphysics	PBL	LSM	Radiation	Dynamical Core
MAP-VTS-con_01	EnVar	GFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_02	rEnKF_m1	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_03	rEnKF_m2	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_04	rEnKF_m3	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_05	rEnKF_m4	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_06	rEnKF_m5	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_07	rEnKF_m6	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_08	rEnKF_m7	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_09	rEnKF_m8	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-con_10	rEnKF_m9	GEFS	Conventional	Thompson	MYNN	RUC	RRTMG	FV3

*Table 6 Specifications for the MAP VTS-bot ensemble members. The configuration of this system matches the descriptions in Tables 5 and 6, where the VTS approach is used for **both** radar and conventional. This tests the VTS for simultaneous multiscale application of radar (storm-scale) and conventional (mesoscale) DA components.*

Members: MAP-VTS-bot	ICs	LBCs	VTS component	Microphysics	PBL	LSM	Radiation	Dynamical Core
MAP-VTS-bot_01	EnVar	GFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_02	rEnKF_m1	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_03	rEnKF_m2	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_04	rEnKF_m3	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_05	rEnKF_m4	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_06	rEnKF_m5	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_07	rEnKF_m6	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_08	rEnKF_m7	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_09	rEnKF_m8	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3
MAP-VTS-bot_10	rEnKF_m9	GEFS	Conv & Radar	Thompson	MYNN	RUC	RRTMG	FV3

Table 7 Specifications for the NSSL FV3-LAM CLUE member. This member is configured the same as the RRFSp1 and RRFSp2 members (Tables 2 & 3), but with the NSSL microphysics scheme and cold start initialization from GFSv16 ICs/LBCs.

Member: NSSL FV3-LAM	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Dynamical Core
NSSL-FV3-LAM	GFS	GFS	NSSL	MYNN	NOAH	RRTMG	FV3

Table 8 Specifications for the NSSL3 CLUE member. This member uses 3-km grid-spacing covering a CONUS domain with forecasts to 30 h using WRF-ARW version 4.2. There are 41 vertical levels and the NSSL 2-moment microphysics scheme is used (Mansell 2010).

Member: NSSL3	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Dynamical Core
NSSL3	GFS	GFS	NSSL	MYNN	NOAH-MP	RRTMG	ARW

Table 9 Specifications for the NSSL1 CLUE member. This member uses 1-km grid-spacing covering the eastern 2/3 of the CONUS and is driven by NSSL3 using a one-way nest. For computational efficiency, the 1-km nest does not start integration until 12 h into the NSSL3 forecast (i.e., 1200 UTC), and forecasts to 30 h (i.e., 0600 UTC the next day) are provided.

Member: NSSL1	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Dynamical Core**
NSSL1	GFS	GFS	NSSL	MYNN	NOAH-MP	RRTMG	ARW

Table 10 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 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). 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**	**Dynamical Core**
gfdl-fv3	GFS	n/a	GFDL	TKE-EDMF	NOAH-MP	RRTMG	FV3

Table 11 Specifications for the RRFSp2eMP ensemble members. This ensemble run by OU-CAPS uses initial conditions from RRFSp2 and the RRFSp2e ensemble members. The FV3 model settings match the ones from RRFPp2e, except mixed physics are used.

Members: RRFSp2eMP	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**Sfc. Phys.**	**LSM**	**Radiation**	**Dynamical Core**
RRFSp2eMP_01	enkf_01	GEFS_m1	Thompson	MYNN	MYNN	NOAH	RRTMG	FV3
RRFSp2eMP_02	enkf_02	GEFS_m2	Thompson	Shin-Hong	GFS	NOAH	RRTMG	FV3
RRFSp2eMP_03	enkf_03	GEFS_m3	Thompson	TKE-EDMF	GFS	NOAH-MP	RRTMG	FV3
RRFSp2eMP_04	enkf_04	GEFS_m4	Thompson	MYNN	MYNN	NOAH-MP	RRTMG	FV3
RRFSp2eMP_05	enkf_05	GEFS_m5	Thompson	TKE-EDMF	GFS	RUC	RRTMG	FV3
RRFSp2eMP_06	enkf_06	GEFS_m6	NSSL	MYNN	MYNN	NOAH	RRTMG	FV3
RRFSp2eMP_07	enkf_07	GEFS_m7	NSSL	Shin-Hong	GFS	NOAH	RRTMG	FV3
RRFSp2eMP_08	enkf_08	GEFS_m8	NSSL	TKE-EDMF	GFS	NOAH-MP	RRTMG	FV3
RRFSp2eMP_09	enkf_09	GEFS_m9	NSSL	MYNN	MYNN	NOAH-MP	RRTMG	FV3
RRFSp2eMP_10	enkf_10	GEFS_m10	NSSL	TKE-EDMF	GFS	RUC	RRTMG	FV3

Table 12 Specifications for the RRFSphys ensemble members. This ensemble run by OU-CAPS uses initial conditions from the mean of the RRFSp2e analyses, as well as the GFS.

Members: RRFSphys	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**Sfc. Phys.**	**LSM**	**Radiation**	**Dynamical Core**
RRFSphys_01	RRFSp2e_mean	GFS	Thompson	MYNN	MYNN	NOAH	RRTMG	FV3
RRFSphys_02	RRFSp2e_mean	GFS	NSSL	MYNN	MYNN	NOAH	RRTMG	FV3
RRFSphys_03	RRFSp2e_mean	GFS	Thompson	MYNN	MYNN	NOAH-MP	RRTMG	FV3
RRFSphys_04	RRFSp2e_mean	GFS	NSSL	TKE-EDMF	GFS	RUC	RRTMG	FV3
RRFSphys_05	RRFSp2e_mean	GFS	Thompson	TKE-EDMF	GFS	NOAH-MP	RRTMG	FV3
RRFSphys_06_gfs	GFS	GFS	Thompson	MYNN	MYNN	NOAH	RRTMG	FV3

The configuration of the 2022 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. An ensemble with VTS applied to radar data only (Table 4) will be compared to an ensemble with VTS applied to both radar data and conventional observations (Table 6). Another ensemble with VTS applied only to conventional observations will also be examined post-experiment (Table 5).

RRFS Configuration Strategies: Several different ensembles will be contributed and evaluated against 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 RRFSp2e (Table 3), MAP-VTS-rad (Table 4), and RRFSp2eMP (Table 11).

FV3-LAM Physics: CAPS will run several configurations of FV3-LAM that are identical except for their physics packages (Table 12). This will allow an assessment of systematic differences and performance characteristics among the different physics suites.

FV3-LAM Data Assimilation: EMC and GSL are running two deterministic RRFS prototypes. Prototype 1 (Table 2) uses partially cycled (hourly) ensemble data assimilation with GDAS (Global Data Assimilation System). Prototype 2 starts from GDAS, but then engages an hourly cycled storm scale ensemble EnKF-based system that informs hybrid deterministic analyses from which a deterministic forecast is launched at 0000 UTC (Table 3). The goal here is to determine the impact of the more sophisticated DA approach (similar to RAP/HRRR, but in UFS framework), with an emphasis on the first 12 h of the forecast.

Enhanced resolution: NSSL is running two versions of WRF-ARW with 3- and 1-km grid-spacing (Tables 8 & 9, respectively) that will be compared to examine grid-spacing sensitivity and assess whether enhanced resolution can provide improved severe weather guidance. Particular attention will be given to the depiction of storm structure and mode, as well as low-level rotation diagnostics (e.g., 0-2 km AGL updraft helicity) for which recent research suggests the 1-km grid-spacing runs can provide improved tornado guidance.

3D-RTMA Background: Three hourly versions of the 3D-RTMA will be compared to assess the role that the background first-guess plays on the final analysis. 3DRTMA prototype 1 (3DRTMAp1) uses RRFSp1 1 h forecasts as background and the GDAS global ensemble for the background error covariances in the 3DEnVAR assimilation. 3DRTMA prototype 2 (3DRTMAp2) uses RRFSp2 1 h forecasts as the background and RRFSp2e for the background error covariances in the 3DEnVAR assimilation. Finally, 3DRTMA HRRR Baseline uses hourly 3-km analyses that rely on operational HRRR 1 h forecasts as the background and the GDAS global ensemble for the background error covariances in the 3DEnVAR assimilation.

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 the 2022 CLUE, all groups output fields to the 3-km CONUS grid used for the operational HRRRv4. 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 and a minimum set of 49 output fields is provided at hourly intervals. This list of mandatory CLUE fields is provided in Appendix C and includes fields that are relevant to a broad range of forecast needs, including aviation, severe weather, and precipitation.

b) High Resolution Ensemble Forecast (HREFv3) System

HREFv3 is a 10-member CAM ensemble that was implemented 11 May 2021. HREFv3 replaced 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 as the HREF in 2017. In HREFv3, the HRW NMMB simulations have been replaced with HRW FV3 and HRRRv3 has been upgraded to HRRRv4.

Table 13 Model specifications for HREFv3.

HREFv3	ICs	LBCs	Microphysics	PBL	dx (km)	Vertical Levels	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 cloud-based Warn-on-Forecast Experiments

Cloud-based Warn-on-Forecast (cb-WoFS) is the next WoFS iteration, upgraded to use current technologies in containerization and cloud computing. The entire WoFS application was rebuilt on top of multiple Platform-as-a-Service and Infrastucture-as-a-Service technologies on the Azure platform and the WRF model itself rebuilt to run in containers optimized for HPC. With the new cb-WoFS interface, administrators can easily configure the domain and dynamically create an HPC infrastructure for the run, and upon completion, tear it down, thereby reducing costs by only paying for used resources. Another benefit is that as Azure continues to add new, updated computer core types from chip manufacturers, these options are passed down to Azure customers, giving cb-WoFS operators the choice of running on the latest technologies. All parts of WoFS have been rebuilt for scalability: the containerized WRF can be executed on any node, the post-processing is built on high performance queues and containerized, so any number of post-processing jobs can run concurrently.

The cb-WoFS is a rapidly-updating 36-member, 3-km grid-spacing WRF-based ensemble data assimilation and forecast system. The cb-WoFS forecasts are initialized every 30 minutes and used to produce very short-range (0-6/0-3 h at top/bottom of the hour) probabilistic forecasts of individual thunderstorms and their associated hazardous weather phenomena such as supercell hail, high winds, flash flooding, and supercell thunderstorm rotation. The 900-km x 900-km daily cb-WoFS domain will target the primary region where severe weather is anticipated.

The starting point for each day's experiment will be the High-Resolution Rapid Refresh Data Assimilation System (HRRRDAS) and the 1200 UTC HRRR forecast provided by NCO/GSL. A 1-h forecast from the 1400 UTC, 36-member, hourly-cycled HRRRDAS analysis provides the ICs for cb-WoFS. Boundary conditions are perturbed HRRR forecasts, where perturbations from the 0600 UTC GEFS are added to the the 1200 UTC HRRR forecasts. The GEFS perturbations are scaled such that the ensemble spread at the lateral boundaries is similar to that provided previously by the experimental HRRR ensemble. Table 14 provides a summary of the model specifications for the cb-WoFS, and Figure 1 shows an example of a SPC Day 1 convective outlook and corresponding cb-WoFS domain with WSR-88D radars used for data assimilation overlaid. Further details on the cb-WoFS are included below.

The 36-member cb-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 cb-WoFS ensemble members use 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. 6-h (3-h) forecasts are initialized and launched from the first 18 members from the real-time cb-WoFS analyses on 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 cb-WoFS Forecast Viewer (https://cbwofs.nssl.noaa.gov).

Table 14 cb-WoFS configuration.

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

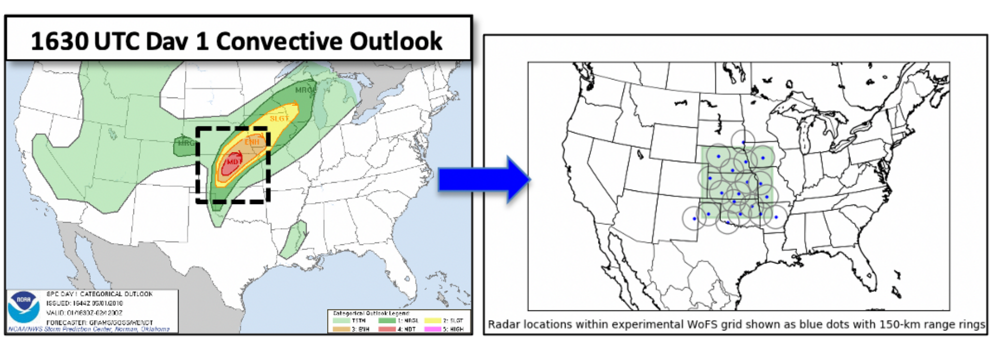

Figure 1 SPC 1630 UTC issued Day 1 convective outlook (left) and corresponding WoFS grid (right).

d) Iowa State University (ISU) Machine Learning-based Severe Wind Probabilities (credit: W. Gallus)

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 (39%) and 52 knot (60 mph; 25%) values. The 50 knot estimates often appear for reports involving tree damage, implying that many of these reports may be due to winds weaker than severe intensity.

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. In addition, algorithms were re-trained with an independent dataset of sub-severe thunderstorm wind measurements added. For both of these two training approaches, output from two different algorithms will be presented. One will be an ensemble model (average ensemble, stacked generalized linear model, random forest), while the other will be the best single model (gradient boosted machine) determined from objective measures in ongoing testing.

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, and land use data. Probabilities derived from each of these machine learning models will be available. An example is shown in Figure 2.

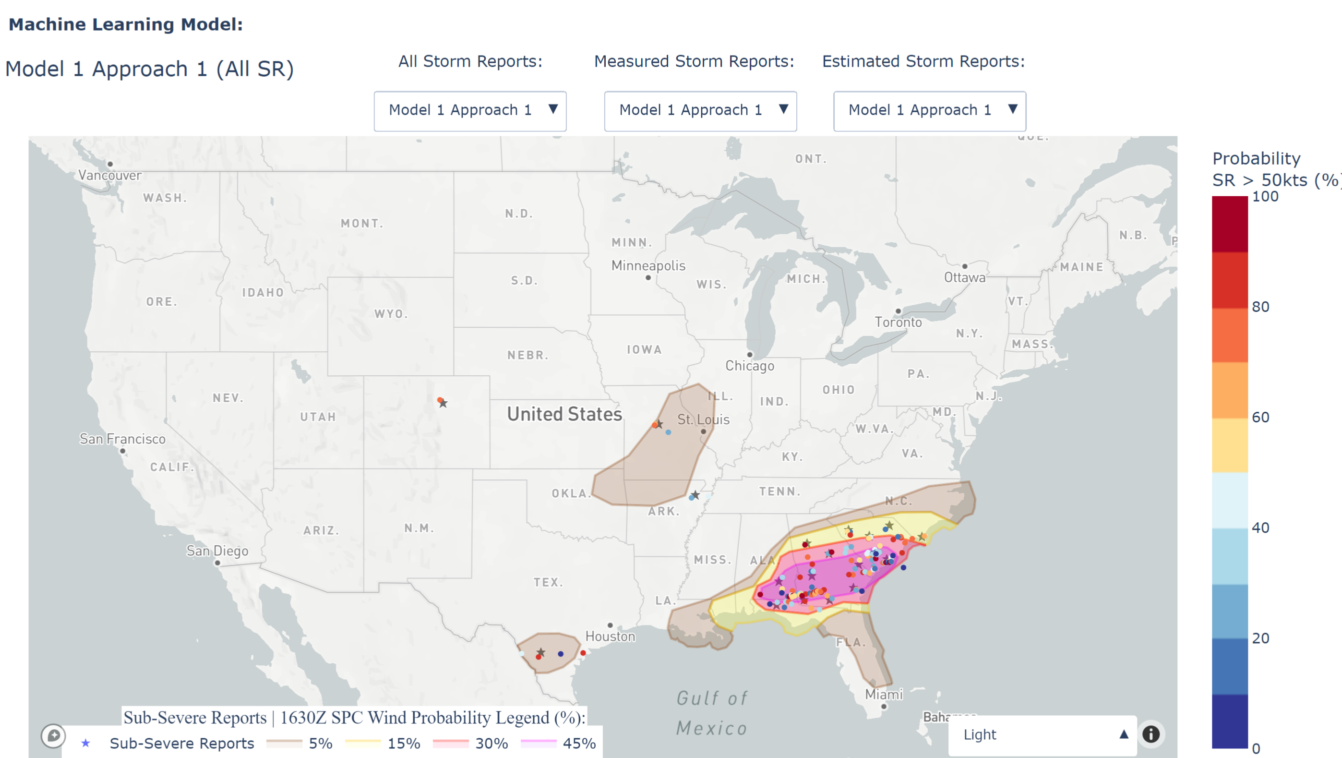

Figure 2 SPC Day 1 probabilities of damaging wind gusts (≥ 50 knots) within 40-km of a point (shaded)). The color of the points indicates the ML-based probability that the report was associated with an actual wind gust ≥ 50 knots. Points labeled with a star represent station measurements near in time and space to a storm report that did not reach 50kts.

e) Texas Tech University RRFS-based Ensemble Subsetting (credit: Brian Ancell)

Ensemble sensitivity is a statistical technique applied within an ensemble that identifies features in the flow at early forecast times that are related to the predictability of chosen severe storm characteristics later in the forecast. In other words, ensemble sensitivity reveals the flow features for which associated errors will grow rapidly to adversely affect the predictive skill of chosen severe storm aspects. It can thus be expected that ensemble members that have the least error in the most sensitive regions early in a forecast window will provide better forecasts than other members, allowing the generation of adjusted and improved probabilities well before the next extended forecast cycle. The goal of this SFE 2022 activity is to evaluate ensemble sensitivity-based subsets from a CLUE FV3-based RRFS ensemble suite to understand whether the subsetting technique can provide value in a real-time environment that includes both initial condition and physics variability. While traditional sensitivity-based subsetting procedure was designed initially for ensembles based on only initial condition variability, here the technique is tested with physics variability under the assumption that early forecast differences among physics members relates sufficiently to the evolution of convection to promote success. In turn, this evaluation represents the most relaxed set of constraints to date with regard to the subsetting technique at the HWT in an effort to discover its broader applicability.

A daily evaluation of subset probabilities from ensemble subsets against those based on 26 members that include RRFSp2e, RRFSp2eMP, and RRFSphys will be conducted. These forecasts are run to at least 36 hours and possess both stochastic parameter perturbations and mixed physics schemes. The subset will be composed of 6 members chosen from the full set of 26 members. Each day, a response function location and time will be chosen through a web-based graphical user interface that identifies areas of Day 1 severe convection. The Day 1 response function will be chosen over a 6-hr period between 1800 UTC (18-hr forecast) and 1200 UTC (36-hr forecast) in areas where better predictions of severe convection are desired (e.g. areas of high uncertainty with regard to convective parameters). Once the response time and location are chosen, the sensitivity of a single response function will be calculated: the number of grid points exceeding maximum hourly 100 m^2/s^2 2–5km updraft helicity. These sensitivities will be generated completely within the CLUE 20-member RRFS ensemble with respect to the 0600 UTC forecast hour (6-hr forecast). Members will then be chosen objectively based on their errors in the most sensitive regions using the 0600 UTC 3D-RTMAp2 analysis.

Probability fields (specifically exceedance probabilities of updraft helicity and simulated reflectivity) of Day 1 convection will be generated for the 6-member subset and will be compared against the full 20-member ensemble. SPC storm reports and the associated practically perfect probability field as well as MRMS data will serve as the observations against which the full and subset RRFS probabilities are evaluated.

f) Calibrated Forecast Products

i. NCAR HRRR-TL ML-based probabilistic convective mode guidance (credit: Ryan Sobash)

The goal of this evaluation is to subjectively evaluate probabilistic convective mode guidance generated from three different machine learning (ML) algorithms, using HRRR forecasts as input. The three algorithms include: 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, 2) a partially-supervised CNN system, that is trained using a "proxy" field related to convective mode (i.e., object size and updraft helicity) and clustered using a Gaussian mixture model (GMM), and 3) a new feedforward neural network (FNN) that predicts mode based on a set of convective storm properties, such as size, area, updraft helicity, reflectivity, etc. The CNN and GMM algorithms have been refined based on feedback from the 2021 HWT SFE, while the FNN algorithm is new for 2022.

Storm objects from every hourly HRRR initialization are generated and passed through the three ML algorithms to produce the probability that each HRRR storm object can be classified as a supercell, quasi-linear convective system, or disorganized mode. Gridded mode probabilities are generated by mapping these storm probabilities onto an 80-km grid, with the resulting binary field indicating grid boxes where each of the three modes is present at each forecast hour. The binary fields are then smoothed to produce a gridded probabilistic hazard guidance product. To improve the reliability of these gridded probabilistic prediction, the smoothed probabilistic mode fields are averaged from multiple overlapping HRRR initializations to produce a time-lagged HRRR (HRRR-TL) probabilistic convective mode guidance product that will be subjectively evaluated. Example HRRR-TL convective mode guidance is provided in Figure 3.

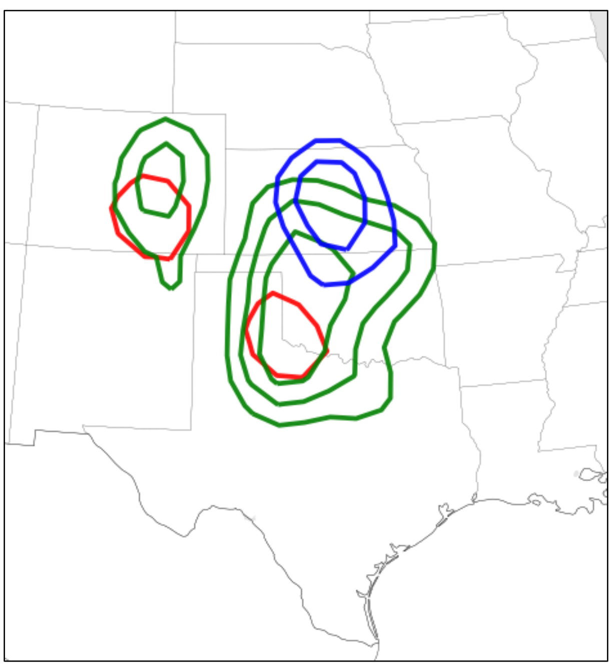

Figure 3 Example HRRR-TL smoothed neighborhood convective mode guidance using output from the CNN ML model. Contours (2, 5, and 10%) indicate the probability of HRRR storm objects being classified as a supercell (red), QLCS (blue), and disorganized (green) convective mode. Forecast valid at 18 UTC 20 May 2019.

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

As in the 2021 HWT SFE, gridded probabilistic convective hazard guidance is being generated with a neural network (NN) over the contiguous United States using the 00 UTC and 12 UTC operational HRRR. The 2021 version of the system (version 1; v1) produces probabilistic predictions for six hazards and was trained with 42 base diagnostics (Table 15) output from a set of ~300 experimental 00 UTC HRRRX forecasts for events between 1 October 2019 and 2 December 2020. The diagnostics are 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 and output probabilities include the three report types, two significant report types, and the occurrence of any storm report. The temporal neighborhood for reports was fixed at 2-h, to produce hazard guidance within 4-h windows, while two spatial neighborhoods are generated.

Table 15 The 49 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. Neighborhood predictors were constructed by taking larger spatial and temporal means and maximums of the environmental and explicit fields. Fields in red were added into the 2022 version (v2) of the system.

Base Predictor	Type	Base Predictor	Type
Forecast Hour	Static	700 hPa–500 hPa lapse rate	Environment
Day of Year	Static	**Freezing level height**	**Environment**
Local Solar Hour	**Static**	Hrly-max 2–5km UH	Explicit
Latitude	Static	Hrly-max 0–3km UH	Explicit
Longitude	Static	**Hrly-max 2–5km UH (negative)**	**Explicit**
Surface-based CAPE	Environment	**Hrly-max 0–2km UH**	**Explicit**
Most-unstable CAPE	Environment	Hrly-max 1 km relative vorticity	Explicit
Surface-based CIN	Environment	Hrly-max updraft speed below 400 hPa	Explicit
Mixed-layer CIN	Environment	Hrly-max downdraft speed below 400 hPa	Explicit
0–6km bulk wind difference	Environment	Hrly-max 10-m wind speed	Explicit
Surface-based LCL	Environment	**Hrly-max column-integrated graupel mass**	**Explicit**
0–1km bulk wind difference	Environment	Hourly precipitation accumulation	Explicit
0–1km storm-relative helicity	Environment	**Hrly-max lightning diagnostic**	**Explicit**
0–3km storm-relative helicity	Environment	**Hrly-max Thompson hail diagnostic**	**Explicit**
2-m temperature	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa zonal wind	Upper-air
2-m dew point temperature	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa meridional wind	Upper-air
Surface pressure	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa temperature	Upper-air
Most-unstable CAPE x 0-6km bulk wind diff.	Environment	925 hPa, 850 hPa, 700 hPa, and 500 hPa dew point	Upper-air
Significant tornado parameter	Environment		

A new version (version 2; v2) of the system is being tested in the 2022 HWT SFE that includes the following enhancements: 1) adding an additional year of operational 00 UTC HRRR training data from 3 December 2020 through 2 December 2021, 2) the inclusion of seven additional predictors (Table 15), 3) generating probabilities by averaging the output of 10 different NNs, 4) using a modified NN configuration with two layers (instead of one; Table 16), but with fewer neurons per layer, and 5) incorporating bug fixes to the pre-processing of certain diagnostics. Evaluations between the 2021 and 2022 NN systems will be facilitated through HWT-generated comparisons. A web-based visualization interface is also 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 4.

Table 16 Settings used to construct and train the NNs. Settings in parentheses are settings used in the 2022 version (v2) of the system.

Neural Network Hyperparameter	Value
Number of hidden layers	1 (2)
Number of neurons in hidden layer	1024 (16)
Dropout rate	0.1 (0)
Learning rate	0.001
Number of training epochs	10 (30)
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

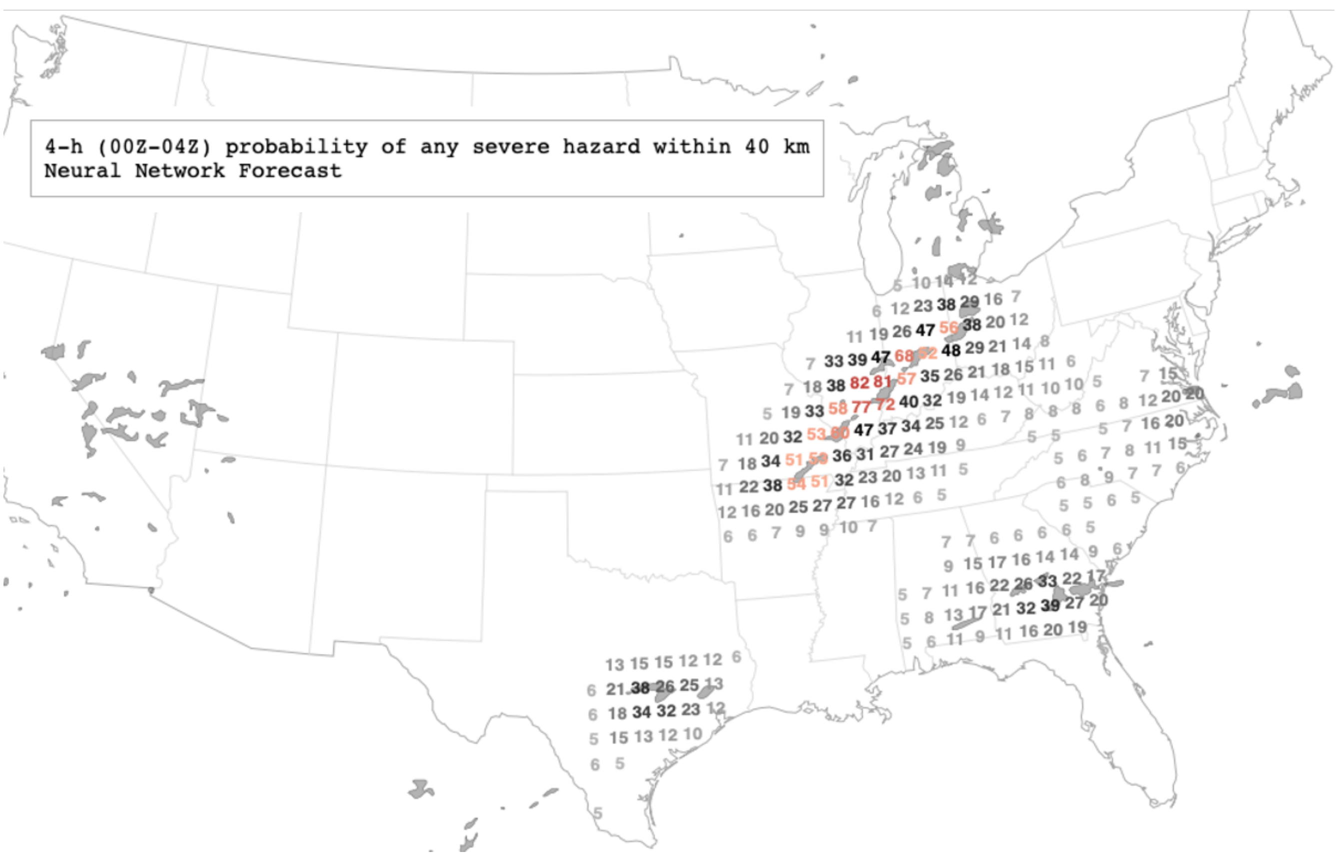

Figure 4 Neural network based probabilistic hazard forecast for the 4-h period between 00Z- 04Z 9 April 2020 based on a WRF forecast initialized at 00 UTC 8 April 2020. Numbers indicate the probability of any severe hazard occurring within 40-km of a grid point. Forecast reflectivity objects > 35 dBZ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

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

Automated "first guess" Day 1 (12-36 h lead times; 1200 UTC – 1200 UTC) and Day 2 (24-48 h lead times; 1200 UTC – 1200 UTC) hazard probabilities are created from random forests (RFs) trained on HREFv3 fields and observed SPC storm reports. Separate RFs predict the probability of tornadoes, severe hail, and severe wind. These predictions are made on an 80 km CONUS grid but are bilinearly interpolated to the native HREF grid for the SFE. The tornado-, severe hail-, and severe wind-predicting RFs use the same set of predictors for a given lead time. These predictors are derived from preprocessed HREFv3 variables. The procedure for creating the predictors is as follows:

1. Aggregate HREFv3 fields in time by computing a maximum or minimum over the forecast period.
2. Upscale the temporally-aggregated fields to an 80 km grid.
3. Compute the ensemble mean for each field on the 80 km grid.
4. Spatially smooth the daily maximum 2-5 km updraft helicity (UH2-5km) forecasts from each HREFv3 member using a 2-dimensional Gaussian kernel density function with σ = 90 km.
5. The final predictors include ensemble mean temporally-aggregated fields, spatially-smoothed individual-member UH2-5km, and 0-1 km relative vorticity from the two HRRR members. Predictors are taken from the point of prediction as well as the closest eight additional 80 km grid points. Latitude and longitude at the point of prediction are also included. Predictors are summarized in Table 17.

HREFv3 data and observed SPC storm reports from 392 days between March 2021 and April 2022 are used for training. The overall procedure for creating RFs is similar to that described in Loken et al. (2020). Key differences from past SFEs include the removal of most raw environmental field predictors (e.g., 2-m temperature) and the addition of: period-minimum storm and index predictors, individual-member smoothed UH2-5km, skewness of UH2-5km at the time of maximum (or minimum) UH2-5km, and 0-1km relative vorticity from the HRRR members.

Table 17 RF predictor fields, organized by the temporal aggregation strategy. Ensemble summary strategy (e.g., the use of an ensemble mean vs. individual members) is reported in parentheses.

Period Maximum	Period Minimum	Constant
2-5 km Updraft Helicity (Ens. mean)	2-5 km Updraft Helicity (Ens. mean)	Latitude
Upward Vertical Velocity (Ens. mean)	Downward Vertical Velocity (Ens. mean)	Longitude
0-3 km Storm Relative Helicity (Ens. mean)	0-3 km Storm Relative Helicity (Ens. mean)	
0-1 km Storm Relative Helicity (Ens. mean)	0-1 km Storm Relative Helicity (Ens. mean)	
0-3 km Energy Helicity Index (Ens. mean)	0-3 km Energy Helicity Index (Ens. mean)	
0-1 km Energy Helicity Index (Ens. mean)	0-1 km Energy Helicity Index (Ens. mean)	
1 km AGL Reflectivity (Ens. mean)		
0-3 km Updraft Helicity (Ens. mean)		
10 m Wind Speed (Ens. mean)		
(MUCAPE) x (10 m – 500 hPa wind shear magnitude) (Ens. mean)		
Supercell Composite Parameter (Ens. mean)		
Significant Tornado Parameter (Ens. mean)		
Significant Hail Parameter (Ens. mean)		
Smoothed Skewness of UH (2-5 km AGL) within a 39-km square radius at time of maximum or minimum UH (Ens. mean)		
Smoothed 2-5km Updraft Helicity (Individual members)		
0-1 km Relative Vorticity (HRRR and HRRR time-lagged members)		

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

In the 2022 SFE, the Colorado State University Machine Learning Probabilities (hereafter, CSU-MLP) prediction system is forecasting severe weather hazards through the application of RFs. The CSU-MLP RFs are trained with about 9 years of daily 0000 UTC initializations from the FV3 global ensemble forecast system reforecast dataset (GEFSv12) 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 1-3 timeframes, such that separate forecasts are issued for each hazard type (example in Figure 5).

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 18. 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 18 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

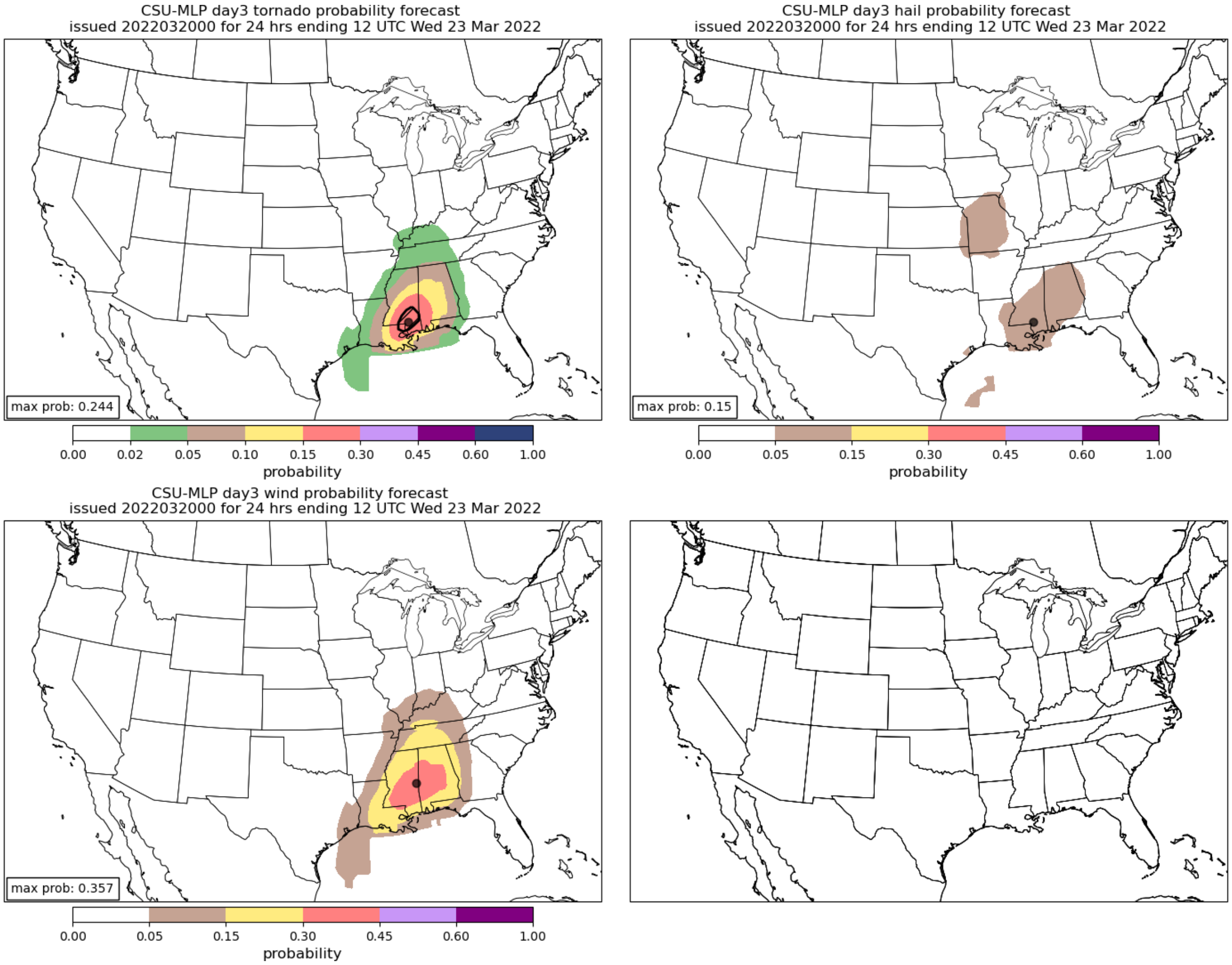

Figure 5 Probabilistic day-3 forecasts of (upper left) tornado, (upper right) hail, and (bottom left) wind hazards valid 1200 - 1200 UTC ending 23 March 2022. Hatched contours represent a 10% probability of significant severe hazards.

v. HREF/SREF, HREF/GEFS, and HREF/HREF Calibrated Severe Weather Probabilities (credit: Chris Karstens/Israel Jirak)

Calibrated probabilities for tornadoes, severe hail, and damaging winds valid over a 24-h time window corresponding to a convective day (i.e., 1200 – 1200 UTC) are produced for three ensembles (i.e., SREF, GEFS, and HREF) using the following procedure. At every grid-point for the valid forecast hour, two probabilities are paired (see Table 19 below):

1. Maximum neighborhood probability of HREF storm attribute variable(s). For all three convective hazards, UH ≥ 75/100/200 m^2s^{-2} for the ARW/NAM Nest/FV3 cores is used over all 4-h periods valid within the previous 24-h period. In addition, severe wind guidance considers two additional storm attribute-based fields: the operational HREF Calibrated Thunder probabilities and the neighborhood probability of 10- m AGL Wind Speeds ≥ 30 kt, which are masked by the aforementioned UH probabilities exceeding 5%.

2. Maximum probability of SREF, GEFS, or HREF environmental field(s) meeting a threshold over 3-h (SREF/GEFS) or 4-h (HREF) periods valid within the previous 24-h period. These fields include the Significant Tornado Parameter (STP), Most-Unstable CAPE (MUCAPE), and Effective Bulk Shear.

Note, the resulting calibrated wind probability field is the maximum of the three approaches listed in Table 19. The historical frequency of a hazard report (or MESH ≥ 29 mm for MESH-Hail calibrated hazard probabilities) occurring within 25 miles of that grid point and within the 24-h period for that forecast pair of probabilities is substituted as the 24-h calibrated hazard probability.

The HREF/SREF calibrated severe guidance is the current operational standard. The HREF/SREF version evaluated here has updated calibration, including a higher MUCAPE exceedance threshold. The HREF/GEFS and HREF/HREF versions are evaluated for comparison to examine the impact of the ensemble contributing the environmental information. This is especially relevant as the SREF is scheduled for retirement in a couple of years.

Table 19 Environmental fields for each hazard used in the HREF/SREF, HREF/GEFS, and HREF/HREF calibrated probabilities.

Hazard	HREF Storm-Attribute Variables	SREF/GEFS/HREF Environmental Variables
Tornado	Updraft Helicity ≥ Model/Core Threshold	STP ≥ 1
Hail	Updraft Helicity ≥ Model/Core Threshold	MUCAPE ≥ 1000 J/kg, Eff. Shear ≥ 20 kt
Wind (Max of 3 approaches)	1. Updraft Helicity ≥ Model/Core Threshold 2. Calibrated Thunder (UH ≥ 5% mask) 3. 10 m AGL Wind ≥ 30 kt (UH ≥ 5% mask)	1. MUCAPE ≥ 1000 J/kg, Eff. Shear ≥ 20 kt 2. MUCAPE ≥ 250 J/kg, Eff. Shear ≥ 20 kt 3. MUCAPE ≥ 1000 J/kg, Eff. Shear ≥ 20 kt

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

Calibrated tornado probabilities valid over 24 h periods valid for 1200 – 1200 UTC on Day 1 and Day 2 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 σ = 50 km is applied. 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-based tornado probabilities (STP Cal MCS-TF; credit: David Jahn)

This method is similar to the STP Cal Inflow method, except tornado probabilities are calculated based on tornado frequency vs. STP curves that are specifically tailored for mesoscale convective systems (MCSs) at grid points for which an MCS is identified. For all other grid points where UH exceeds the 99.985 percentile of climatology (indication of a rotating storm) the same tornado frequency vs. STP curve is used as with the Inflow method to calculate tornado probability (Jahn et al. 2020, Thompson et al. 2017). Storm mode, either MCS or supercell, is determined objectively at a grid point using pre-determined thresholds of either the skewness or standard deviation of the UH distribution within a 40-km radius (Jahn et al. 2022).

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

A series of ML models are being developed to provide rapidly updating probabilistic guidance to human forecasters for short-term (e.g., 0-4 h) severe weather forecasts. We generated the feature inputs into the ML models from 18-member WoFS forecasts. Rather than producing a gridded ML product as with next-day (i.e., 12-36 h lead times) 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 within a 30 minute time window—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.

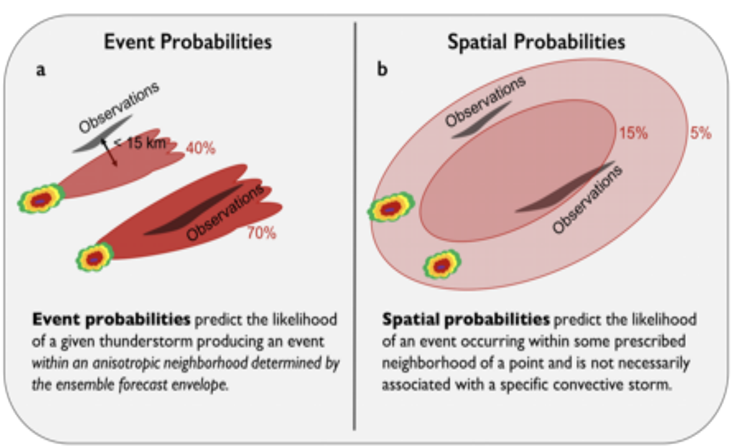

Figure 7 Illustration of the distinction between event and spatial probabilities (Fig. 2 of Flora et al. 2019).

The feature inputs were based on intra-storm and environmental variables from the WoFS and morphological variables describing the storm objects (Table 20).

Table 20 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 severe wind 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 a severe wind report will occur within the region. For example, this guidance suggests that multiple cells within the MCS in Southern MS and AL have 40-60% chance of producing severe wind in the next hour.

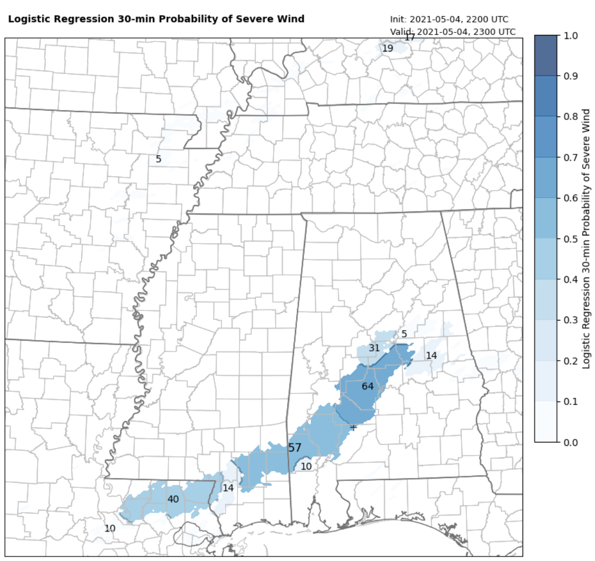

Figure 8 Example forecast from the severe-wind-based logistic regression. Number overlays indicate the probability of a severe wind within that region.

The product shown in Figure 8 is available every 5 min up to a lead time of 4 hrs. Due to chaotic evolution of thunderstorms, we also provide 1-hourly and 3-hourly summary products (Fig. 9). For example, in Figure 9, we compute the maximum tornado probability over the next hour. The guidance suggests a high tornado likelihood for the supercell in W TN and for the portion of the MCS over central MS, with a modest likelihood in southern MS.

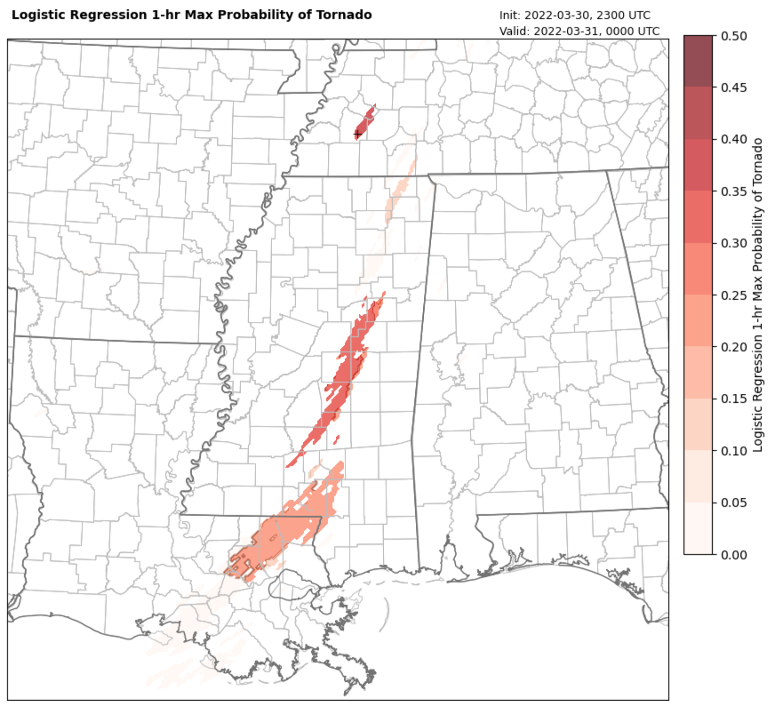

Figure 9 Example forecast of maximum tornado probability over the next hour.

x. Nadocast, HREF/SREF ML-based tornado probabilities (credit: Brian Hempel)

Nadocast is a machine learning system, initially focused on tornadoes, that aims to produce timely, calibrated, severe weather probabilities on the Day 1 time scale (2-35 hours). Probabilities are generated by gradient-boosted decision trees trained on 10,000+ storm and storm-adjacent hours of HREF and SREF outputs. Nadocast performs extensive feature engineering: each grid point from the HREF (or SREF) hourly output is supplemented by adding spatial blurs of various radii, spatial gradients, parameters from 1 h future and 1 h past, summary statistics over a 3 h window, and additional information such as climatology and an estimate of recent convective forcing. To provide rotational invariance, winds at each grid point are rotated relative to an estimate of the 500m-5000m shear vector. The result is over 10,000 features per grid point per hour, upon which the decision trees operate to produce hourly probabilities. To capture uncertainty at longer lead times, different models are trained for short- (2-13hr), medium- (13-24hr), and longer-range (24-35hr) forecasts. Hourly probabilities are pooled into day-long guidance on a 15km grid and rescaled to follow the historical characteristics of SPC thresholds. A preliminary objective comparison (n=260 days) suggests tornado-prediction performance that, on average, matches or slightly exceeds SPC 6Z Day 1 guidance. Evaluation in the SFE should shed light on Nadocast's subjective characteristics.

xi. Flow-dependent training of RF models for convective-outlook guidance (credit: A. Johnson)

Forecast errors, biases, and relationships between predictors and severe weather hazards can depend on the large-scale flow pattern, even within a given season. However, RF models do not inherently account for spatial patterns in the input predictors. Since the large-scale flow pattern, in general, can be forecast with high confidence at more than one-day lead time there may be opportunities to leverage such information that is known prior to consulting convection-allowing model guidance in the approach to training RF models used for Day 1 convective outlook guidance. For the 2022 SFE, OU MAP has implemented three sets of RF-based convective outlook guidance products in the post-processing of our FV3-based ensemble that uses VTS approaches with both conventional and radar data (MAP-VTS-bot). The convective outlook guidance includes probabilistic forecasts for severe wind, hail and tornado during the 24-hour period from 12 UTC to 12 UTC on Day 1. The first set of RF-based products (MAP_RF) are generated using a baseline configuration where predictors are interpolated onto an 80-km grid and trained over approximately 80 retrospective forecasts. The second set of RF-based products (MAP_RF_FD) includes flow-dependent training where the RF model that is applied on a given day is trained only on the subset of retrospective cases with a "similar" large scale flow pattern, where the similarity of the large-scale flow pattern is determined as described below. The third set of RF-based products (MAP_RF_MS) uses multi-scale predictors that includes the same predictors as MAP_RF, in addition to the same predictors smoothed over progressively larger circles with radii of 160, 320 and 800 km in order to implicitly account for characteristics of the larger scale flow.

For the explicitly flow-dependent RF training (MAP_RF_FD), permutation feature importance (PFI; Mecikalski et al. 2021) was used to identify the importance of each predictor overall, and on individual cases. PFI is the increase in Brier score error resulting from shuffling values of a given predictor on a given case, and it is here normalized by dividing by the error increase from the most important predictor on that case. Principal Component Analysis is then used to identify the leading modes of variability across the domain among the training cases for several forecast variables. For example, for severe wind prediction the normalized PFI of the 850 hPa U-component of wind predictor tends to be lower on cases that project negatively on the leading mode of variability for 850 hPa U-component of wind (Figure 10). Since the leading mode of variability for 850 hPa U-wind corresponds to large-scale anomalously strong westerly flow, this suggests that u850 does not provide the RF model much additional discrimination of localized severe wind threat in cases with weak large-scale westerly low-level flow. We thus hypothesize that for cases that project onto u850 PC 1 with a value below about -25 (see Figure 10), we can give the RF a head start on learning the relevant relationships among predictors for this case by only training on cases that also have a strongly negative projection onto u850 PC 1. This approach will be compared to the baseline approach and the approach of implicitly providing information about the large-scale flow pattern using multi-scale predictors during 2022 HWT SFE.

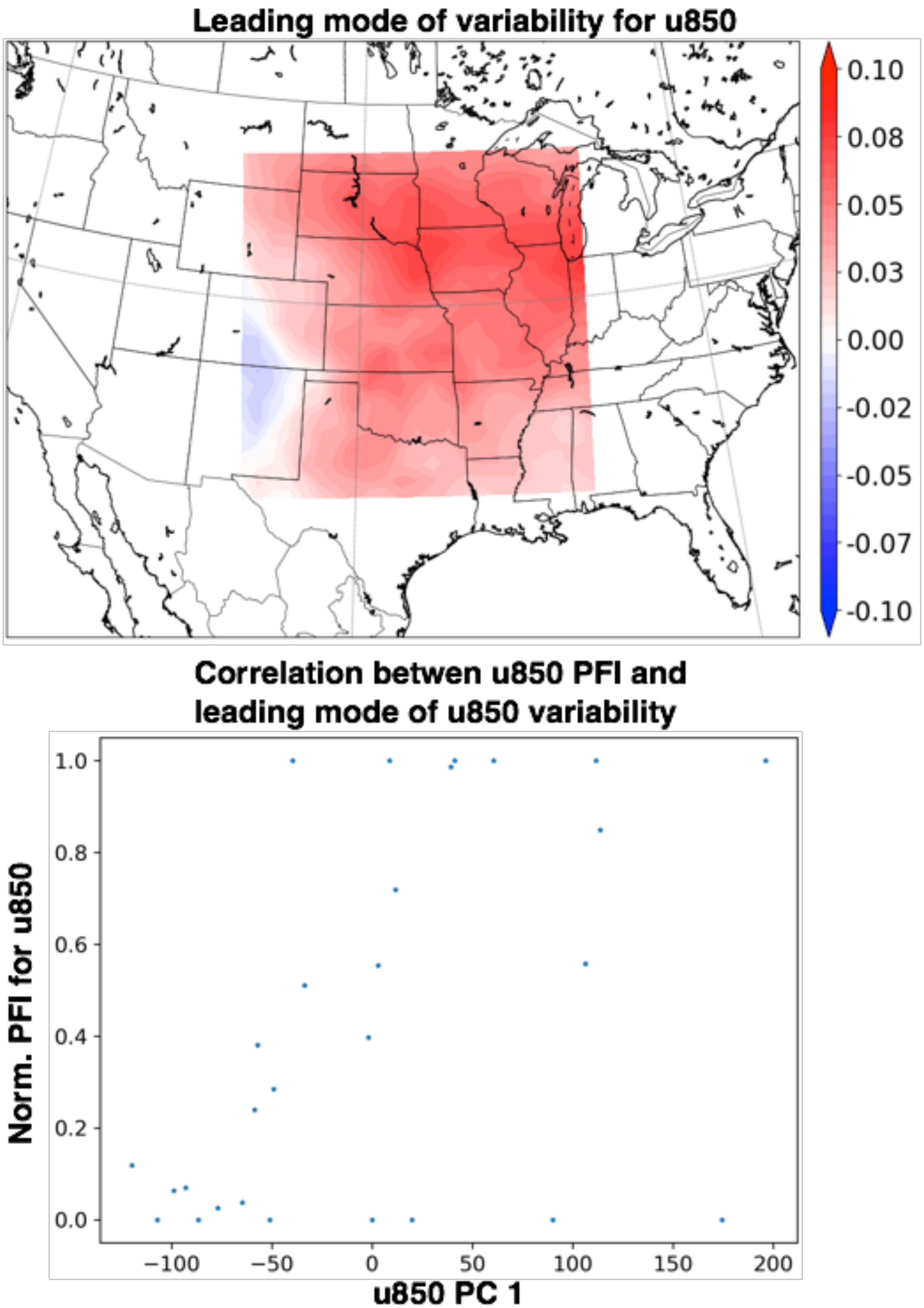

Figure 10 The top panel shows the leading principal component of 850 hPa u-wind (u850) at forecast hour 24 in the MAP forecasts from 2021 HWT, calculated by first recentering the u850 field based on that case's verification domain selected during the 2021 HWT SFE. The bottom panel shows a scatter plot of normalized u850 PFI vs. the projection for that case of the 24-hr u850 forecast onto the leading mode of u850 variability.

Table 21 Different classes of predictors used in the RF model. Storm environment, and "other", predictors are averaged over each 80-km grid box while storm attribute predictors are the 80-km grid box maximum.

Storm attributes		**Storm environment**		**Other**	
Maxuh25	Hourly maximum updraft helicity (UH) in 2-5 km layer	sbcape	Surface-based CAPE	Lat	Latitude
Maxuh03	Hourly maximum UH in 0-3 km layer	mucape	CAPE of must unstable parcel in lowest 255mb	lon	Longitude
Maxdbz10c	Hourly maximum reflectivity at 10 C level	sbcin	Surface-based CIN		
hrprecip	Hourly accumulated precipitation	mucin	CIN of must unstable parcel in lowest 255mb		
		sblcl	LCL of surface parcel		
		u250, u500, u850	u-component of wind at 250, 500 and 700 hPa		
		v250, v500, v850	v-component of wind at 250, 500 and 700 hPa		
		tmp500, tmp700, tmp850	Temperature at 500, 700 and 850 hPa		
		dew700, dew850	Dewpoint at 700 and 500 hPa		
		t2m, dew2m	Temperature and dewpoint at 2m AGL		
		hgt500	12-hour change in height at 500 hPa		
		hlcy1km, hlcy3km	Storm relative helicity in 0-1 and 0-3 km layers		
		pw	Precipitable water		

3. SFE 2022 Core Interests and Daily Activities

2022 SFE activities will occur from 9am-4pm CDT, with a lunch break from 12:30-2pm CDT. On Wednesdays there will be an optional science brown bag over a part of the lunch break. Tables 22 and 23 provide a schedule for Monday, and Tuesday-Friday, respectively. Further details are provided in subsequent sections.

Table 22 Schedule for Monday.

<table>
<tr><td>Time (CDT)</td><td colspan="2">R2O Group</td><td colspan="2">Innovation Group</td></tr>
<tr><td>9:00 AM – 9:45 AM</td><td colspan="4">Welcome and Introductions
Israel Jirak & Participants</td></tr>
<tr><td>9:45 AM – 10:30 AM</td><td colspan="4">HWT SFE Scientific Objectives and Goals
Israel Jirak & Adam Clark</td></tr>
<tr><td>10:30 AM - 11:00 AM</td><td colspan="4">Break
Fill out IRB Consent Form</td></tr>
<tr><td>11:00 AM - 11:15 AM</td><td colspan="4">Conditional Intensity Forecasting Overview
Israel Jirak</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 and Day 3 Hazards Coverage and Conditional Intensity Forecasts</td></tr>
<tr><td>No Cal. Guidance</td><td>Cal. Guidance</td><td>Day 2</td><td>Day 3</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 – 3:15 PM</td><td colspan="2">Issue MD Product</td><td colspan="2">Issue 1-h outlooks (21-22, 22-23Z)</td></tr>
<tr><td colspan="2">WoFS & obs</td><td>WoFS ML</td><td>WoFS No ML</td></tr>
<tr><td rowspan="2">3:15 PM – 4:00 PM</td><td>Update Day 1 Outlook</td><td>Focus Group Activity</td><td colspan="2">Issue 1-h outlooks (21-22, 22-23), End-of-Day WoFS ML Survey</td></tr>
<tr><td>WoFS & other guidance</td><td>Conditional Intensity Discussion</td><td>WoFS ML</td><td>WoFS No ML</td></tr>
</table>

Table 23 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 CAMs</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 and Day 3 Hazards Coverage and Conditional Intensity Forecasts (2 groups)</td></tr>
<tr><td>No Cal. Guidance</td><td>Cal. Guidance</td><td>Day 2</td><td>Day 3</td></tr>
<tr><td>12:30 PM – 2:00 PM</td><td colspan="4">Lunch/Break (Wed., Fri.)
Lunch/Science Brown Bag (Tues., Thurs.)</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 (21-22, 22-23Z)</td></tr>
<tr><td colspan="2">WoFS & obs</td><td>WoFS ML</td><td>WoFS No ML</td></tr>
<tr><td rowspan="2">3:00 PM – 4:00 PM</td><td>Update Day 1 Outlook</td><td>Focus Group Activity</td><td colspan="2">Issue 1-h outlooks (21-22, 22-23Z), End-of-Day WoFS ML Survey</td></tr>
<tr><td>WoFS & other guidance</td><td>Conditional Intensity Discussion</td><td>WoFS ML</td><td>WoFS No ML</td></tr>
</table>

a. Formal Evaluation Activities

SFE 2022 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, WoFS, and other products and guidance. 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 (on Tuesdays and Thursdays when participants join a new evaluation group), 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 facilitators 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. GEFS Tornado Guidance

Tornado probabilities for Days 1-3 from the CSU-MLP prediction system are subjectively rated.

b. Day 2 12Z HREF Calibrated Tornado Guidance

Four different methods based on HREF for deriving calibrated Day 2 tornado guidance are subjectively rated. These methods include: (1) HREF/GEFS Cal, (2) STP Cal Circle, (3) ML Random Forest, and (4) STP Cal MCS-TF.

c. Day 1 00Z HREF Calibrated Tornado Guidance

The same methods as in the Day 2 evaluation are rated for Day 1, with the addition of Nadocast.

d. HREF Tornado Guidance: Convective Mode

STP Circle and STP Cal MCS-TF are subjectively evaluated relative to STP Inflow for Day 1 tornado guidance. Note, the ratings for STP Cal MCS-TF and STP Cal Circle are carried over from the previous evaluation.

e. HREF Tornado Guidance: Environment

Four different methods for tornado guidance based on a combination of HREF and different sources of environmental information are subjectively rated. These methods include: (1) HREF/SREF Ops, (2) HREF/SREF Para, (3) HREF/GEFS Cal, and (4) HREF/HREF Cal.

f. OU-MAP Flow-Dependent Tornado Guidance

RF models for Day 1 tornado predictions contributed by the OU-MAP group are subjective evaluated. These include models that are non-flow dependent (MAP_RF), explicitly flow dependent (MAP_RF_FD), and implicitly flow dependent (MAP_RF_MS).

g. GEFS Hail Guidance

Hail probabilities for Days 1-3 from the CSU-MLP prediction system are subjectively rated.

h. Days 1 & 2 12Z HREF Calibrated Hail Guidance

Hail probabilities based on HREF for Days 1 & 2 from HREF/GEFS Cal and ML Random Forest are subjective evaluated.

i. HREF Hail Guidance: Environment

Four different methods for hail guidance based on a combination of HREF and different sources of environmental information are subjectively rated. These methods include: (1) HREF/SREF Ops, (2) HREF/SREF Para, (3) HREF/GEFS Cal, and (4) HREF/HREF Cal. Note, the ratings for HREF/GEFS Cal are carried over from the previous evaluation.

j. HREF Hail Guidance: MESH

HREF/GEFS MESH is subjectively evaluated relative to HREF/GEFS Cal. HREF/GEFS MESH is formulated the same as HREF/GEFS Cal, except MESH is used for calibration rather than severe hail reports. Note, the ratings for HREF/GEFS Cal are carried over from the previous evaluation.

k. OU-MAP Flow-Dependent Hail Guidance

RF models for Day 1 hail predictions contributed by the OU-MAP group are subjective evaluated. These include models that are non-flow dependent (MAP_RF), explicitly flow dependent (MAP_RF_FD), and implicitly flow dependent (MAP_RF_MS).

l. GEFS Wind Guidance

Wind probabilities for Days 1-3 from the CSU-MLP prediction system are subjectively rated.

m. Days 1 & 2 12Z HREF Calibrated Wind Guidance

Wind probabilities based on HREF for Days 1 & 2 from HREF/GEFS Cal and ML Random Forest are subjective evaluated.

n. HREF Wind Guidance: Environment

Four different methods for wind guidance based on a combination of HREF and different sources of environmental information are subjectively rated. These methods include: (1) HREF/SREF Ops, (2) HREF/SREF Para, (3) HREF/GEFS Cal, and (4) HREF/HREF Cal. Note, the ratings from HREF/GEFS Cal are carried over from the previous evaluation.

o. OU-MAP Flow-Dependent Wind Guidance

RF models for Day 1 wind predictions contributed by the OU-MAP group are subjective evaluated. These include models that are non-flow dependent (MAP_RF), explicitly flow dependent (MAP_RF_FD), and implicitly flow dependent (MAP_RF_MS).

p. 00Z HRRR NCAR NN Tor/Hail/Wind Guidance

Day 1 Tornado, Hail, and Wind predictions from last year's version (v1) of the NCAR NN hazard probabilities are compared to a new version (v2) that includes several enhancements.

Primary Science Question: What are the strengths and weaknesses of the various calibrated hazard guidance, and 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 ranking the primary deterministic CAMs provided by several SFE collaborators – GFDL (*GFDL-FV3*), NSSL (*NSSL-FV3-LAM*), and EMC/GSL (*RRFSp1 and RRFSp2*) – based on their skill and utility for severe weather forecasting. 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: RRFS vs. HRRR

This activity will feature a "deeper dive" into storm attribute and environmental fields in HRRRv4 and RRFSp2. These comparisons will serve to unearth ways in which the currently operational CAM (the HRRRv4) differs from a candidate to replace it (the RRFSp2), and whether the RRFSp2 improves upon or degrades forecasts of the HRRRv4 for fields relevant to forecasting severe weather. A greater number of fields will be available for this comparison relative to other comparisons, allowing for participants to examine more facets of the guidance and identify potential contributions to severe convective hazard forecast success or failure.

Primary Science Questions: How do forecasts of the RRFSp2 compare to those of the HRRRv4? Are there systematic shortcomings or advantages of the RRFSp2?

B3. CLUE: Data Assimilation

Five deterministic model configurations are examined in the first 12 hours of the forecast period that incorporate different data assimilation strategies: (1) HRRRv4, (2) RRFSp1, (3) RRFSp2, (4) MAP-VTS-rad Control, and (5) MAP-VTS-bot Control.

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

B4. CLUE: FV3 Physics Suites

The RRFSphys ensemble contains five different physics suites. 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. For the microphysics/PBL/LSM schemes, these members use (1) Thompson/MYNN/NOAH, (2) NSSL/MYNN/NOAH, (3) Thompson/MYNN/NOAH-MP, (4) NSSL/TKE-EDMF/RUC, and (5) Thompson/TKE-EDMF/NOAH-MP. 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. 1-km vs. 3-km

This comparison will focus on comparing the NSSL3 and NSSL1 configurations of WRF-ARW, which have 3- and 1-km grid-spacing, respectively. Particular attention will be given to unique storm attribute fields such as 0-1 km AGL UH and 0-2 km AGL maximum wind. It is hypothesized that for these fields, the enhanced resolution of NSSL1 will provide improved guidance for hazards like tornadoes, whose parent mesocyclones and associated low-level rotation are better resolved using 1-km grid-spacing, and wind, which is better resolved at higher resolutions.

Primary Science Question: *Does increasing horizontal grid-spacing from 3- to 1-km provide benefits when utilizing storm diagnostics that reflect the intensity of low-level rotation important for tornado prediction and the strength of convective wind gusts?*

Group C – CAM Ensembles

C1. CLUE: 00Z CAM Ensembles

This evaluation will compare three 00Z initialized, FV3-LAM CAM ensembles to HREFv3. Specifically, (1) RRFSp2e, (2) MAP-VTS-rad, and (3) RRFSp2eMP 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 (e.g., DA and physics) for FV3-LAM based CAM ensembles, and how do they compare to HREFv3?

C2. CLUE: RRFSp2e vs. HREF

This evaluation will feature an in-depth examination of several storm attribute and environment fields from 00Z and 12Z initialized versions of RRFSp2e and HREFv3. These comparisons will serve to unearth ways in which the currently operational CAM ensemble (the HREF) differs from a candidate to replace it (the RRFSp2e), and whether the RRFSp2e improves upon or degrades forecasts of the HREF for fields relevant to forecasting severe weather. A greater number of fields will be available for this comparison relative to other comparisons, allowing for participants to examine more facets of the guidance and identify potential contributions to severe convective hazard forecast success or failure. This comparison will also mark the first time that CAM ensemble environments will be compared during a next-day evaluation within the SFE

Primary Science Question: How do probabilistic forecasts of the RRFSp2e compare to those of the HREFv3 (e.g., spread and skill)? Are there systematic shortcomings or advantages of the RRFSp2e?

C3. CLUE: Data Assimilation

This evaluation will focus on the first 12 h of 00Z-initialized forecasts from three CAM ensembles that employ different data assimilation strategies and compare their forecasts to HREFv3. Specifically, (1) RRFSp2e, (2) MAP-VTS-rad, and (3) MAP-VTS-bot, will be compared.

Primary Science Question: Does valid-time-shifting (VTS) improve upon forecasts without VTS, and does incorporating both radar and conventional observation into VTS improve upon forecasts that only incorporate radar data in the VTS algorithm?

C4. CLUE: TTU Ensemble Subsetting

Severe weather probabilities formulated with updraft helicity and reflectivity for a subset of RRFSp2e and RRFSp2eMP members will be compared to those from the "full" ensemble of RRFSp2e, RRFSp2eMP, and RRFSphys members. The subset is composed of the members with the smallest errors in sensitive regions as determined by ensemble sensitivity analysis.

Primary Science Question: Can a sensitivity-based ensemble subsetting approach lead to improved guidance over the full ensemble for severe-weather forecasting?

C5. WoFS: Number of Members

In this evaluation, UH- and reflectivity-based probabilities and paintball plots from 21Z and 23Z WoFS initializations will be compared where the probabilities/paintballs are constructed from all 18 WoFS members, as well as 9- and 13-member subsets.

Primary Science Question: Is it possible to run WoFS with fewer members and get the same forecast quality of hourly probabilistic forecasts?

C6. WoFS: Time Lagging

In this evaluation, 18-member WoFS guidance will be compared where the 18 members come from a single initialization time versus different (i.e., time-lagged) initialization times. Specifically, one set of 18 members will come from 6 members drawn from 19, 20, and 21Z; another set of 18 members will come from 9 members drawn from 20 and 21Z; and the final set of 18 members will all come from 21Z. The same exercise will be repeated but for WoFS ensembles based at 23Z.

Primary Science Question: Does a time-lagging strategy benefit WoFS forecasts?

Group D – Medley

D1. ISU ML Severe Wind Probabilities

An evaluation will be conducted of different techniques 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. Additionally, Practically Perfect severe wind probabilities derived from LSRs will be compared to Practically Perfect severe wind probabilities where the LSRs have been assigned a probability that they were associated with wind ≥ 50 knots.

Primary Science Questions: Can machine-learning approaches provide useful information regarding the likelihood of wind damage reports being associated with gusts ≥ 50 knots? Do Practically Perfect wind probabilities utilizing the gusts ≥ 50 knots probabilities have more utility than the standard method used to compute the Practically Perfect probabilities?

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, three trained ML models will be tested: 1) a supervised ML system that trains a convolutional neural network (CNN) using a hand labeled dataset, 2) a partially-supervised CNN system, that is trained using a “proxy” field related to convective mode (i.e., object size and updraft helicity) and clustered using a Gaussian mixture model (GMM), and 3) a new feedforward neural network (FNN) that predicts mode based on a set of convective storm properties, such as size, area, updraft helicity, reflectivity, etc.. By using a time-lagged HRRR (HRRR-TL) approach, a probabilistic convective mode guidance product for each ML model will be subjectively evaluated.

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

D3. Mesoscale Analysis

a. Background

Three hourly versions of the 3D-RTMA will be compared to assess the role that the background first-guess plays on the final analysis. Specifically, 3DRTMAp1, 3DRTMAp2, and 3DRTMA HRRR

Baseline will be compared. The goal is to assess the utility of these analysis systems for situational awareness and short-term forecasting for convective-weather scenarios.

b. 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 prior to retiring 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?

D5. County-Based Watch Guidance – "Threats in Motion" (credit: David Harrison)

An HREF-based ML model has been developed to produce automated, non-static watch products that dynamically track with the predicted severe weather threat. This guidance is derived from a gradient boosted classifier trained on HREF ensemble updraft helicity, updraft vertical velocity, 10-m wind, and sfc-500 mb shear. Estimated watch counties are inferred at each 12z HREF forecast hour from the ML probabilistic output and masked such that a county must fall within at least a 13z D1 Slight risk to qualify for a watch. Automated watches produced by the ML guidance are designed to provide at least 3 hours of lead time prior to the first occurrence of severe weather.

An alternative automated watch product has also been derived from the SPC Severe Timing guidance. Estimated watch counties are inferred at each forecast hour from the temporally disaggregated individual hazard probabilities provided by the 13z Severe Timing guidance and the 12z HREF. A county is considered to be in a watch at a given forecast hour if the timing guidance produces individual hazard probabilities equivalent to at least a Slight risk at that location and time. As with the

ML guidance, these automated watches are designed to provide at least 3 hours of lead time prior to the first occurrence of severe weather.

These two approaches for generating county-based watch guidance will be compared and evaluated against counties in actual Severe Thunderstorm and Tornado Watches, as well as against Severe Thunderstorm and Tornado Warnings and preliminary Local Storm Reports (LSRs).

Primary Science Question: Can ML be used to produce automated, dynamic watch products that accurately track the severe weather threat? How well do these automated watches emulate the appearance and utility of human-generated watch products? Does ML add value over the simpler timing guidance approach for providing automated watch guidance?

b. Forecast Products and Activities

There will be two periods of experimental forecast activities during SFE 2022. The first will occur from 11:30am – 12:30pm CDT and will focus on generating probabilistic outlooks for individual hazards, as well as more precise information on the intensity of specific hazards. As in SFE 2021, we will split the participants into four groups: two R2O groups issuing products for Day 1 and two Innovation Groups issuing products for Days 2 & 3. 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. The Day 3 forecast is the only exception to the smaller domain, and will instead cover a full CONUS domain.

In all groups, the morning forecasts will be done collectively. 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 Days 2 & 3 outlooks will cover 1200 – 1200 UTC periods. 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 SFEs 2019-2021. 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 four conditional intensity probability distributions that can be forecast via examination of the atmospheric environment: "no severe", "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 all available operational and experimental guidance products for issuing their Day 1 individual hazard and conditional intensity forecasts, except for calibrated guidance, while another sub-group will use the same sets of guidance, along with the calibrated guidance products. Within the Innovation Group, one sub-group will issue Day 2 forecasts, while another sub-group will issue Day 3 forecasts.

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 create 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, the R2O group will be split into two sub-groups. In one sub-group, 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. In the other sub-group, a focus group activity will be conducted to gain insight on the conditional intensity products.

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 2100-2200 UTC and 2200-2300 UTC. Initial forecasts will be generated during the 2:15-3:15pm period and final forecasts will be generated during the 3:15-3:45pm period. After the final forecasts are issued, from approximately 3:45-4pm, participants will complete a survey to gain insight on the use of ML-based forecast products from WoFS. All of the Innovation Group afternoon forecasting activities will be conducted in two sub-groups. One group will have access to calibrated, WoFS-based ML guidance when issuing their forecasts, while the other will only use the uncalibrated WoFS products. For both sets of initial and final forecasts, two forecasters will be in the group that includes ML guidance (Forecaster CAL 1 & 2), while two other forecasters will be in the group without ML guidance (Forecaster NOCAL 1 & 2). Additionally, other participants in each group will issue forecasts with and without the ML guidance similarly to the expert forecasters, which will be combined into consensus forecasts (ConCAL and ConNOCAL, respectively).

These WoF activities are the sixth 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 2022 participants.

Week 1	Week 2	Week 3	Week 4	Week 5
2-6 May	**9-13 May**	**16-20 May**	**23-27 May**	**31 May - June 3**
Philippe Papin (NHC)	Tom Galarneau (CIWRO)	Bill Gallus (ISU)	Harald Richter (BoM)	Craig Schwartz (NCAR)
Maria Molina (NCAR)	Allie Mazurek (CSU)	Manda Chasteen (NCAR)	Brice Coffer (NC State)	Jordan Dale (WPO)
Kevin Thiel (CIWRO/SPC)	Robin Tamamachi (Purdue)	Trudy Kidd (EC-OSPC)	Noah Carpenter (OU SOM)	Becky Adams-Selin (AER)
Brad Vrolijk (EC-PASPC)	Allison LaFleur (Purdue)	Monica Vaswani (EC-OSPC)	Clark Evans (UWM)	Kelly Hobelman (EC-OSPC)
Andy Elliott (USAF)	Liz Tirone (ISU)	Heather Pimiskern (EC-PASPC)	Russ Schumacher (CSU)	Katrina Eyk (EC-OSPC)
Kelton Halbert (U. Wisc)	Leigh Orf (U. Wisc)	Jamie Foote (USAF)	Georgina Da costa Barradas (EC-QSPC)	Sherry Williams (EC-OSPC)
Victor Gensini (NIU)	Aaron Hill (CSU)	Andrew Winters (U. Colorado)	Kristin Corbosiero (U. of Albany)	Eric Van Lochem (EC-PASPC)
Keenan Eure (PSU)	Allie Brannan (CIWRO/SPC)	McKenzie Larson (U. Colorado)	Jen Henderson (TTU)	Nick Goldacker (NC State)
Will Mayfield (DTC)	Michelle Harold (DTC)	Casey Davenport (UNCC)	Roldolfo Hernandez (TTU)	Felicia Guarriello (WPO)
Ryan Sobash (NCAR)	Justin Spotts (TAMU)	Roger Riggin (UNCC)	Dave Ahijevych (NCAR)	Camille Hoover (USAF)
Geeta Nain (Purdue)	Charlie Becker (NCAR)	Kelly Lombardo (PSU)	Matthew Vaughan (St. Cloud)	Rob Hepper (AWC)
John Allen (CMU)	Marion Mittermaier (UK Met)	Alexandra Anderson-Frye (UW)	Tatiana Gonzalez (NWS AFS)	David Gagne (NCAR)
Dan Harris (UK Met)	Chris Smallcomb (NWS REV)	Zhanxiang Hua (UW)	Carlo Cafaro (UK Met)	Eric Guillot (NWS AFS)
Ka Yee Wong (GSL)	Steve Willington (UK Met)	Stephanie Avey (NWS AFS)	Aurore Porson (UK Met)	Brian Tang (Albany)
Tyler Hasenstein (NWS MPX)	Matt Bunkers (NWS UNR)	Justin Gibbs (NWS WDTD)	Matt Clark (UK Met)	David King (NWS MTR)
Binbin Zhou (EMC)	Marcel Caron (EMC)	Eswar Iyer (NWS AKQ)	Brian Tentinger (NWS BGM)	Jonathan Garner (NWS EKA)
Xiaoyan Zhang (EMC)	Shun Liu (EMC)	David Thomas (NWS BUF)	Mike Johnson (NWS MEG)	Kyle Pallozzi (NWS LWX)
Eric Aligo (EMC)	David Dowell (GSL)	Tony Wardle (UK Met)	Mike Dutter (NWS AKQ)	Jidong Gao (NSSL)
Scott Kleebauer (NWS MAF)	Justin Schultz (NWS DLH)	Sebastian Cole (UK Met)	Remington Lilya (St. Cloud)	Jeff Beck (GSL)
Jeff Duda (GSL)	John Boris (NWS APX)	Alyssa Clements (NWS ABQ)	Gang Zhou (EMC)	Chauncy Schultz (NWS BIS)
Terra Ladwig (GSL)	Austin Coleman (TTU)	Dylan Lusk (NWS FFC)	Matthew Pyle (EMC)	Logan Dawson (EMC)
Craig Evanego (NWS CTP)	Jason Frazier (NWS PBZ)	Justin Arnott (NWS GYX)	Ben Blake (EMC)	Geoff Manikin (EMC)
Chris Noles (NWS PAH)	Jay Engle (NWS OKX)	Jacob Carley (EMC)	Craig Hartsough (GSL)	Edward Colon (EMC)
Lee Britt (NWS DLH)	Stephen Travis (NWS CTP)	Matt Morris (EMC)	Kyle Pederson (GSL)	Linda Gilbert (NWS MQT)
Andrew Snyder (NWS LWX)	Tom Hultquist (NWS MPX)	Chris MacIntosh (EMC)	Jeffrey Hovis (NWS RLX)	Curtis Alexander (GSL)
Jonty Hall (BoM)	Melody Sturm (BoM; M-Th)	John Brown (GSL)	Pete Wolf (NWS JAX)	Harald Richter (BoM)
Sean Ernst (OU)	Aidan Kuroski (NWS MKX)	Ed Szoke (GSL)	Drew Shearer (OU)	Cameron Miller (NWS MKX)
	Aurora Bell (BoM; M-W)	Matthew Campbell (NWS ILN)	Dan Kubalek (OU)	Thomas Winesett (NWS JAN)
	Derrick Snyder (NWS PAH)	Adam Gill (NWS BGM)	Frank Alsheimer (NWS CAE)	Logan Poole (NWS JAN)
	Charles Smith (NWS MFR)	Christopher Kent (BoM)	Brendon Rubin-Oster (NWS LWX)	Sarah Trojniak (WPC)
			Alexander Majchrowski (BoM)	Tony Wedd (BoM)
			Dean Sgarbossa (BoM)	

SFE Facilitators: Adam Clark (NSSL), Israel Jirak (SPC), Dave Imy (retired SPC), Burkely Gallo (CIWRO/SPC), Kenzie Krocak (CIWRO/SPC/CRCM), Brett Roberts (CIWRO/SPC/NSSL), Kent Knopfmeier (CIWRO/NSSL), Chris Karstens (SPC), Eric Loken (CIWRO/NSSL), David Harrison (CIWRO/SPC), David Jahn (CIWRO/SPC), Jacob Vancil (CIWRO/SPC), Jeff Milne (CIWRO/SPC), Allie Brannan (CIWRO/SPC) and Nathan Dahl (CIWRO/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 https://hwt.nssl.noaa.gov/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 https://hwt.nssl.noaa.gov/ewp/. A key NWS strategic goal is to extend warning lead times through the "Warn-on-Forecast" concept (Stensrud et al. 2009), which involves using

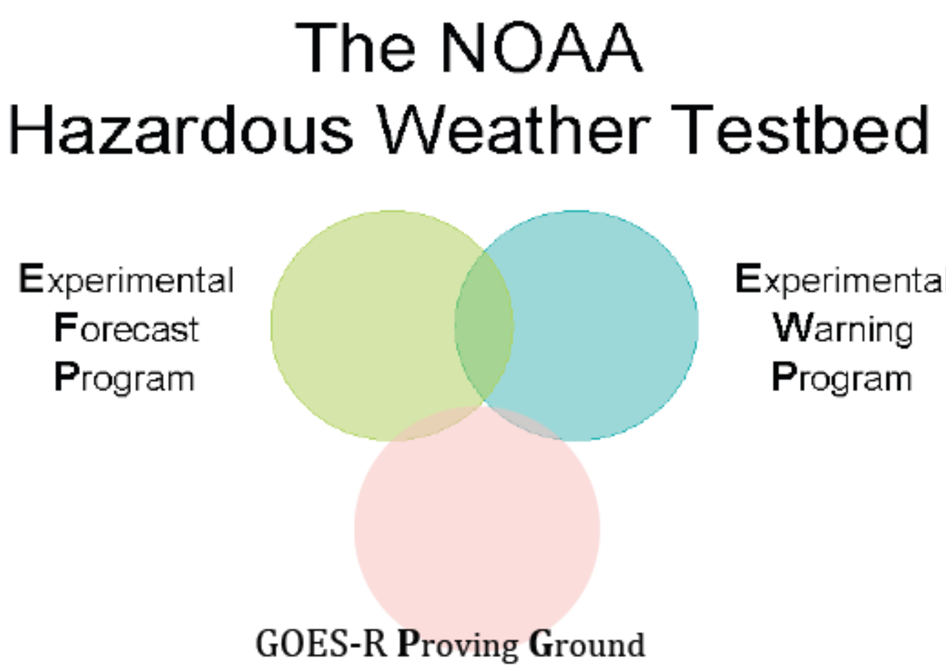

Figure B1: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

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 and communication 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 https://hwt.nssl.noaa.gov/. 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 Kain et al. (2003), 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: Mandatory 2022 CLUE Fields

1. Mean Sea Level Pressure	26. CIN (most unstable)
2. Composite reflectivity	27. CAPE (mixed layer)
3. Reflectivity at -10 C	28. CIN (mixed layer)
4. Maximum surface wind gust	29. 0-3 km AGL storm relative helicity
5. hrly-max upward motion 100-1000 hPa	30. 0-1 km AGL storm relative helicity
6. hrly-max downward motion 100-1000 hPa	31. 2-5 km AGL UH (instantaneous)
7. Reflectivity at 1-km AGL	32. Echo Top Height
8. Hrly-max reflectivity at 1-km	33. 300 hPa Height
9. Hrly-max reflectivity at -10 C	34. 300 hPa u-wind
10. Hrly-max 2-5 km AGL UH	35. 300 hPa v-wind
11. Hrly-min 2-5 km AGL UH	36. 300 hPa temperature
12. Hrly-max 0-3 km AGL UH	37. 500 hPa Height
13. Hrly-min 0-3 km AGL UH	38. 500 hPa u-wind
14. Surface Pressure	39. 500 hPa v-wind
15. Surface Height	40. 500 hPa temperature
16. 2-m temperature	41. 700 hPa Height
17. 2-m dewpoint	42. 700 hPa u-wind
18. 2-m relative humidity	43. 700 hPa v-wind
19. 10-m u-wind	44. 700 hPa temperature
20. 10-m v-wind	45. 850 hPa Height
21. Hrly-max 10-m Wind Speed	46. 850 hPa u-wind
22. Surface total precipitation (run total)	47. 850 hPa v-wind
23. CAPE (surface parcel)	48. 850 hPa temperature
24. CIN (surface parcel)	49. 850 hPa specific humidity
25. CAPE (most unstable)	

Appendix D: 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.

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, 2020: 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.

Gasperoni, N. A., X. Wang, and Y. Wang, 2022: Using a Cost-Effective Approach to Increase Background Ensemble Member Size within the GSI-Based EnVar System for Improved Radar Analyses and Forecasts of Convective Systems. *Mon. Wea. Rev.*, **150**, 667-689.

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.

Jahn, D. E., B. T. Gallo, C. Broyles, B. T. Smith, I. Jirak, J. Milne, 2020: Refining CAM-based tornado probability forecasts using storm-inflow and storm-attribute information. *26th Conf. on Numerical Weather Pred*., Boston, MA, Amer. Meteor. Soc.

Jahn, D. E., I. Jirak, A, Wade, J. Milne, 2022: Storm mode and tornado potential determination using statistical moments of updraft helicity distribution. *27th Conf. on Numerical Weather Pred.*, Houston, TX, Amer. Meteor. Soc.

Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between Forecasters and Research Scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc*., **12**, 1797-1806.

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*, **34**, 1955-1964).

Mansell, E. R., 2010: On sedimentation and advection in multimoment bulk microphysics. *J. Atmos. Sci*., **67**, 3084-3094.

Mecikalski, J. R., T. N. Sandmael, E. M. Murillo, C. R. Homeyer, K. M. Bedka, J. M. Apke, C. P. Jewett, 2021: A random-forest model to assess predictor importance and nowcast severe storms using high resolution radar-GOES satellite-lightning observations. *Mon. Wea. Rev*., **149**, 1725-1746.

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.

Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, J. C. Picca, A. E. Cohen, E. M. Leitman, A. M. Gleason, and P. T. Marsh, 2017: Tornado damage rating probabilities derived from WSR-88D data. *Wea. Forecasting*, **32**, 1509-1528.

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.

Wang, Y., and X. Wang, 2021: Development of Convective-Scale Static Background Error Covariance within GSI-Based Hybrid EnVar System for Direct Radar Reflectivity Data Assimilation. *Mon. Wea. Rev*., **149**, 2713-2736.

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.