SPRING FORECASTING EXPERIMENT 2023

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

HWT Facility – National Weather Center

1 May – 2 June 2023

https://hwt.nssl.noaa.gov/sfe/2023/

Program Overview and Operations Plan

26 April 2023

Adam Clark[2,4], Israel Jirak[1], Tim Supinie[1], Kent Knopfmeier[2,3], Jake Vancil[1,3], David Jahn[1,3], David Harrison[1,3], Allie Brannan[1,3], Chris Karstens[1], Eric Loken[2,3], Nathan Dahl[1,3], Makenzie Krocak[3,4,5], David Imy[2], Andy Wade[1,3], Jeffrey Milne[1,3,4], Kimberly Hoogewind[2,3], Pamela Heinselman[2,4], Montgomery Flora[2,3], Joshua Martin[2,3], Brian Matilla[2,3], Joey Picca[1,3], 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 2023 HWT Spring Forecasting Experiment (SFE 2023), a cornerstone of the EFP, will be conducted 1 May – 2 June. After three years of virtual experiments, this will be the first experiment with in-person participation since 2019. Additionally, we'll continue to have virtual participation in the 2023 SFE, making 2023 the first hybrid experiment. Relative to the last two virtual experiments, this year's hybrid 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 2023 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 2023 CLUE includes 40 members with 3-km grid-spacing, as well as a single member using 1-km grid-spacing. The SFE 2023 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 7th year to issue very short lead-time outlooks.

With tentative plans for operational implementation of the Rapid Refresh Forecast System (RRFS) in late 2024, SFE 2023 will make it a point of emphasis to evaluate the RRFS against the operational High-Resolution Rapid Refresh (HRRR) and High-Resolution Ensemble Forecast (HREF) systems for severe weather forecasting applications. New testing and evaluation activities for SFE 2023 will include examinations of regional and global-with-nest versions of the Model for Prediction Across Scales and extended range (Days 3-8) CAM ensemble evaluations.

This document summarizes the core interests of SFE 2023 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 2023 and Section 3 describes the core interests and new concepts being introduced for SFE 2023. 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 9:00 – 11:00am (CDT) focusing on various CLUE subsets and calibrated guidance. The 2023 CLUE includes deterministic and ensemble forecasts using the most recent versions of the Finite Volume Cubed-Sphere Model (FV3), the

Advanced Research Weather Research and Forecasting (WRF-ARW) model, and the Model for Prediction Across Scales (MPAS). 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 2023.

a) The 2023 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 National Aeronautics and Space Administration (NASA) and the National Center for Atmospheric Research (NCAR). All CLUE members cover a CONUS domain with 3-km grid-spacing, except the NASA FV is run at 2-km grid-spacing and the NSSL1 is run at 1-km grid-spacing over the eastern 2/3 of the CONUS. Depending on the CLUE subset, forecast lengths range from 18 to 192 h. Table 1 summarizes all 2023 CLUE contributions. Subsequent tables provide details on members in each subset.

Table 1 Summary of the 8 unique subsets that comprise the 2023 CLUE. For the RRFS and RRFSphys CLUE Subsets, 00, 06, 12, & 18 UTC initializations have 60 h forecast lengths and the entire 10-member ensemble is run; for all other RRFS and RRFSphys initialization times, only the control member is initialized with forecast lengths of 18-h.

Clue Subset	# of mems	IC/LBC perts	Mixed Physics	Data Assimilation	Dynamical Core	Agency	Init. Times (UTC)	Forecast Length (h)	Domain
RRFS	10	EnKF	no	Hybrid 3DEnVar	FV3	EMC/GSL	00-23	60/18	CONUS
RRFSphys	9	EnKF	yes	Hybrid 3DEnVar	FV3	EMC/GSL	00-23	60/18	CONUS
NSSL1	1	none	no	HRRR ICs	ARW	NSSL	00	36	2/3 CONUS
NSSL-MPAS	3	none	no	HRRR or RRFS ICs	MPAS	NSSL	00	48	CONUS
GFDL-FV3	1	none	no	GFS cold start	FV3	GFDL	00	126	CONUS
NASA-FV3	1	none	no	GEOS-DA	FV3	NASA	00	120	CONUS
NCAR-FV3	10	GEFS	no	GEFS cold start	FV3	NCAR	00	192	CONUS
NCAR-MPAS	5	GEFS	no	GEFS cold start	MPAS	NCAR	00	132	CONUS

Table 2 Specifications for the RRFS ensemble. The RRFS is a rapidly-updated, convection-allowing (3 km) ensemble forecast system. Ensembles are initialized using 3-km ensemble perturbations drawn directly from the RRFS Data Assimilation System's (RDAS) ensemble Kalman filter analysis members. The control member forecast is initialized from the hybrid 3DEnVar analysis. The RDAS uses a wide variety of conventional observations along with radar reflectivity. It also includes a nonvariational cloud analysis. For gravity wave drag, the small scale and turbulence orographic form drag options are used in all members. Stochastically perturbed parameterization tendencies (SPPT) are applied to all perturbed members (i.e., RRFS01-09), and stochastic parameter perturbations (SPP) are applied to the microphysics, PBL, surface layer, radiation, and gravity wave drag parameterizations in the perturbed members. The RRFS will be under active development throughout the HWT SFE period and it may undergo changes to its underlying scientific package that impact results. This is distinct from prior years when every effort was made to minimize changes to the real-time test system during the SFE. The extent of changes and expected impact will be communicated to the HWT/SFE facilitators if and when changes are made that will significantly impact results.

Members: RRFS	**ICs**	**LBCs**	**Micro-physics**	**PBL/SFC**	**LSM**	**Radiation**	**Shallow Cumulus**	**Dynamical Core**
RRFS (ctl)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS01	enkf1	GEFS m1	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS02	enkf2	GEFS m2	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS03	enkf3	GEFS m3	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS04	enkf4	GEFS m4	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS05	enkf5	GEFS m5	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS06	enkf6	GEFS m6	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS07	enkf7	GEFS m7	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS08	enkf8	GEFS m8	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS09	enkf9	GEFS m9	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3

Table 3 Specifications for the RRFSphys ensemble. Note, RRFS and RRFSphys share the same control member. SPPTs are applied to all perturbed members. SPP is applied to Thompson in members RRFSphys02-05; MYNN PBL & surface layer physics in members RRFSphys04, 06, and 09; and LSM, radiation, and gravity wave drag parameterizations in all perturbed members. For the NSSL microphysics (Mansell 2010) members, SPP is applied using a parameter perturbation following a Latin hypercube sampling with multidimensional uniformity technique.

Members: RRFSphys	**ICs**	**LBCs**	**Micro-physics**	**PBL/SFC**	**LSM**	**Radiation**	**Shallow Cumulus**	**Dynamical Core**
RRFS (ctl)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFSphys01	enkf1	GEFS m1	Thompson	H-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys02	enkf2	GEFS m2	Thompson	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys03	enkf3	GEFS m3	Thompson	MYNN/MYNN	RUC	RRTMG	saSAS Shal	FV3
RRFSphys04	enkf4	GEFS m4	Thompson	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys05	enkf5	GEFS m5	NSSL	MYNN/MYNN	RUC	RRTMG	saSAS Shal	FV3
RRFSphys06	enkf6	GEFS m6	NSSL	H-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys07	enkf7	GEFS m7	NSSL	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys08	enkf8	GEFS m8	NSSL	MYNN/MYNN	RUC	RRTMG	saSAS Shal	FV3
RRFSphys09	enkf9	GEFS m9	NSSL	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3

Table 4 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 the HRRR. For computational efficiency, the 1-km nest does not start integration until 6 h into the 00Z-initialized HRRR forecast (i.e., 0600 UTC), and forecasts to 36 h (i.e., 1200 UTC the next day) are provided.

Member: NSSL1	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Dynamical Core**
NSSL1	HRRR	HRRR	NSSL	MYNN	RUC	RRTMG	ARW

Table 5 Specifications for the NSSL-MPAS CLUE members. These members use 3-km grid-spacing covering the CONUS and are driven by the HRRR or RRFS. The last two letters of each member denote the ICs and microphysics ("HN" = HRRR-NSSL, "HT" = HRRR-Thompson, and "RT" = RRFS-Thompson).

Member: NSSL-MPAS	**ICs**	**LBCs**	**Microphysics**	**PBL**	**LSM**	**Radiation**	**Dynamical Core**
NSSL-MPAS-HN	HRRR	HRRR	NSSL	MYNN	RUC	RRTMG	MPAS
NSSL-MPAS-HT	HRRR	HRRR	Thompson	MYNN	RUC	RRTMG	MPAS
NSSL-MPAS-RT	RRFS	RRFS	Thompson	MYNN	RUC	RRTMG	MPAS

Table 6 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 version 3 of the GFDL In-line Microphysics (Zhou et al. 2022) 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 7 Specifications for the NASA-FV3 CLUE member. The NASA-FV3 is also known as the NASA GEOS model and will run an FV3-based stretched global grid. The target resolution is a c2160 grid with 137 vertical levels, the stretching will produce a 2-km domain over CONUS with the coarsest global resolution of 12-km over the Indian Ocean. We will be running this case in a replay mode using an incremental analysis update (IAU) to our GEOS-FP 12-km production data assimilation system. The IAU approach permits our higher resolution model to evolve dynamically with time and avoids having to cold start forecasts each day. The NASA FV3 model will produce 5-day forecasts at 00Z daily to provide medium range prediction of convective events over CONUS.

Member: NASA-FV3	**ICs**	**LBCs**	**Micro-physics**	**PBL**	**LSM**	**Radiation**	**Dynamical Core**
NASA-FV3	GEOS-FP	None	GEOS-GFDL	Lock-Louis & UW	Nasa Catchment	RRTMG	FV3

Table 8 Specifications for the NCAR-FV3 ensemble members. These 10-member ensemble forecasts are based on GFDL's C-SHiELD (Harris et al., 2020), 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 Inline 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 while the global domain uses the GFS Noah LSM. The Scale-aware Simplified (SAS) Arakawa-Schubert cumulus parameterization is also used; both shallow and deep schemes are employed on the 13-km global grid but only a shallow scheme is employed on the 3-km nest. All members use identical physics, with ensemble diversity solely provided by initial conditions. No stochastic physics are used. Initialization is cold-start from members 1–10 of real-time GEFS initial conditions. Simulations run daily at 00Z out to 192 hours to demonstrate the potential for medium-range prediction of convective-scale events.

Members: NCAR-FV3	ICs	LBCs	Micro-physics	PBL	LSM	Radiation	Cumulus	Dynamical Core
NCAR-FV3-01	GEFS m1	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-02	GEFS m2	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-03	GEFS m3	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-04	GEFS m4	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-05	GEFS m5	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-06	GEFS m6	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-07	GEFS m7	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-08	GEFS m8	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-09	GEFS m9	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3
NCAR-FV3-10	GEFS m10	n/a	GFDLv2	TKE-EDMF	NOAH-MP	RRTMG	SAS Arakawa-Schubert	FV3

Table 9 Specifications for the NCAR-MPAS ensemble members. All 5 ensemble members use NCAR's MPAS model and identical physics, with ensemble diversity solely provided by initial conditions. No stochastic physics are used. Initialization is cold-start from members 1–5 of real-time GEFS initial conditions. Simulations run daily at 00Z out to 132 hours to demonstrate the potential for medium-range prediction of convective-scale events.

Members: NCAR-MPAS	ICs	LBCs	Micro-physics	PBL	LSM	Radiation	Cumulus	Dynamical Core
NCAR-MPAS01	GEFS m1	n/a	Thompson	MYNN	NOAH	RRTMG	Scale-aware New Tiedtke	MPAS
NCAR-MPAS02	GEFS m2	n/a	Thompson	MYNN	NOAH	RRTMG	Scale-aware New Tiedtke	MPAS
NCAR-MPAS03	GEFS m3	n/a	Thompson	MYNN	NOAH	RRTMG	Scale-aware New Tiedtke	MPAS
NCAR-MPAS04	GEFS m4	n/a	Thompson	MYNN	NOAH	RRTMG	Scale-aware New Tiedtke	MPAS
NCAR-MPAS05	GEFS m5	n/a	Thompson	MYNN	NOAH	RRTMG	Scale-aware New Tiedtke	MPAS

The configuration of the 2023 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:

RRFS vs. HRRR/HREF: With plans for operational implementation of the RRFS in late 2024, a critical evaluation activity for SFE 2023 will involve comparing the deterministic and ensemble components of RRFS to their operational counterparts HRRR and HREF, respectively. Comparisons will be made for Day 1 & 2 lead times. Additional comparisons will be made during the first 12 h of the forecasts to evaluate the effectiveness of the data assimilation systems in each system.

RRFS Configuration Strategies: With the possibility that computational limitations may only allow 6 members at a single initialization time when RRFS is implemented operationally, 6-hour and 12-hour time-lagging will be tested as a cost-effective strategy to increase membership in the RRFS (Table 2). Furthermore, RRFS with mixed-physics (Table 3) will be examined to assess and compare performance characteristics compared to the single physics RRFS (Table 2) to help determine the optimal operational configuration. In the 6- and 12-h time lagged configurations, members RRFS (ctl), RRFS01, RRFS02, RRFS03, RRFS04, and RRFS05 will be used for the single physics ensemble, and members RRFS (ctl), RRFSphys01, RRFSphys02, RRFSphys05, RRFSphys06, and RRFSphys07 will be used for the mixed-physics ensemble.

Medium-Range CAM Ensembles: NCAR will be providing a 10-member, FV3-based, 3-km grid-spacing ensemble with forecasts to 7 days (Table 8). Additionally, a 5-member, 3-km grid-spacing MPAS ensemble will have forecasts to 5 days (Table 9). Days 3-7 will be evaluated in the NCAR-FV3 ensemble for the same valid time to assess the evolution of forecast quality with increasing lead time. For lead times of 3-5 days, a 5-member subset of NCAR-FV3 will be compared to the 5-member NCAR-MPAS ensemble to assess differences in forecast quality between the two ensembles at each lead time.

Enhanced resolution: NSSL will be running a version of WRF-ARW with 3-km grid-spacing initialized from the HRRR (NSSL1; Table 4). The NSSL1 forecasts will be compared to the HRRR 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 and Storm-Scale Analyses: Hourly versions of the 3D-RTMA will be compared to assess the role that the background first-guess has on the final analysis. One version uses the HRRR for the background while the other uses the RRFS. Versions of the analyses upscaled to 40-km will also be examined and compared with SPC's surface objective analysis (sfcOA). Finally, 15-minute WoFS forecasts of hourly maximum 80-m winds, UH, and updraft speed will be compared to Multi-Radar, Multi-Sensor (MRMS) products to gauge whether these 15-minute WoFS forecasts are a viable proxy for observed hazards.

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 2023 CLUE, all groups output fields to the 3-km CONUS grid used for the operational HRRRv4. For WRF-ARW, FV3-LAM, and MPAS 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. HREFv3 specifications are listed in Table 10.

Table 10 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. For SFE 2023, WoFS will have the capability to run over two different domains. A second domain will only be implemented when there are two separate regions where severe weather is expected (e.g., Midwest and East Coast), or when there is a very large single area for which two domains are needed to cover the entire risk area.

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 11 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 11 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)

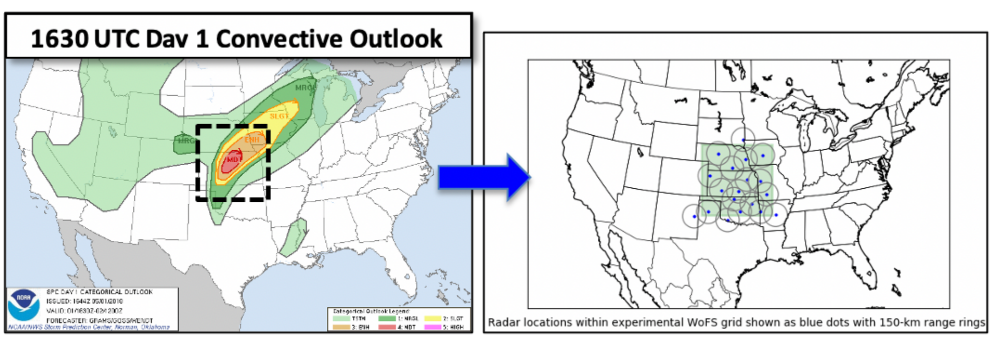

Figure 1 SPC 1630 UTC issued Day 1 convective outlook (left) and corresponding WoFS grid (right).

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. Based on past evaluations, it appears two algorithms, the gradient boosted machine (a single model) and the stacked generalized linear model (an ensemble approach) performed best, and both were most skillful when the sub-severe data were used in training. Therefore, output from both of these algorithms will be evaluated.

The training of these models utilized information from the Storm Events 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, and elevation data. Probabilities derived from the two machine learning models will be available. An example is shown in Figure 2.

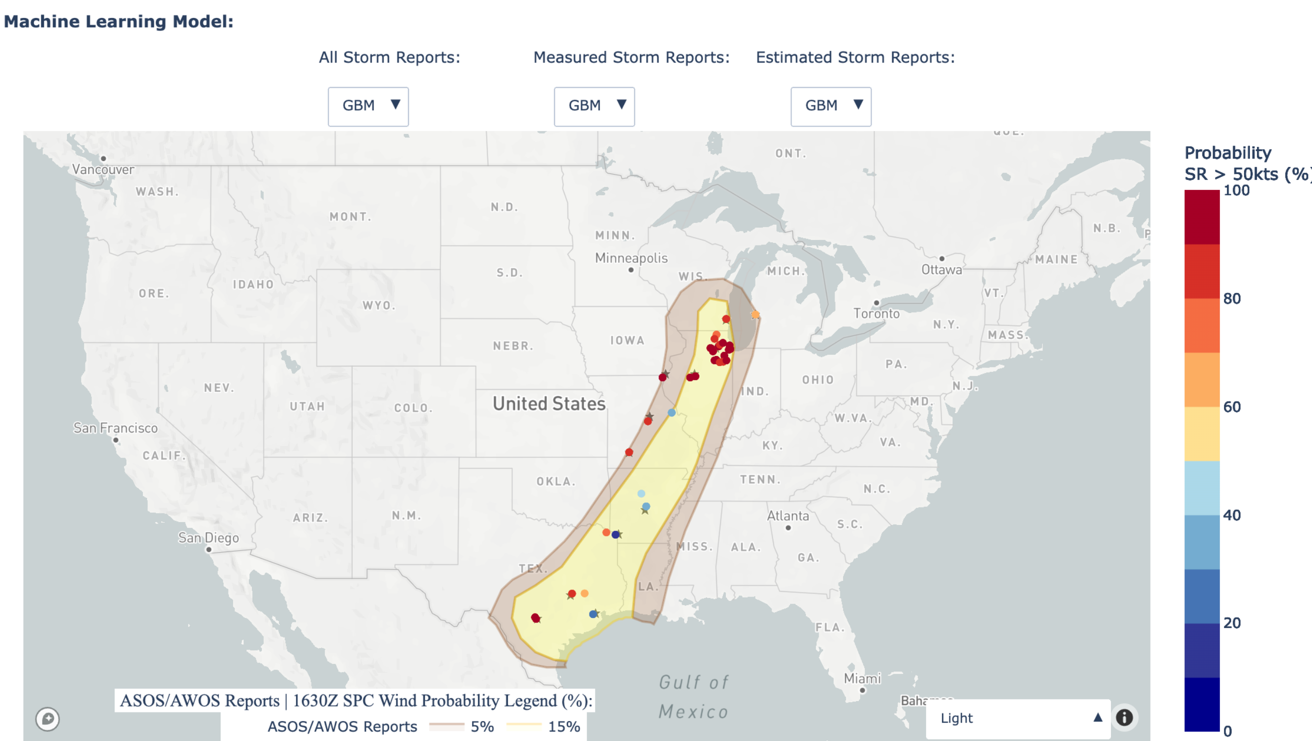

Figure 2 SPC Day 1 probabilities of damaging wind gusts (≥ 50 knots) within 40-km of a point (shaded)). The color of the points indicates the ML-based probability that the report was associated with an actual wind gust ≥ 50 knots. Points labeled with a star represent station measurements near in time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

e) Calibrated Forecast Products

i. NCAR ML-derived HRRR-based convective hazard probabilities using convective mode information (credit: Ryan Sobash)

For the 2023 HWT SFE, gridded probabilistic convective hazard guidance is being generated using a neural network (NN) and the 00 UTC operational HRRR. The NNs generate probabilistic predictions for six hazards and were trained with 42 base diagnostics (Table 11) output from a set of 365 00 UTC HRRRX forecasts in 2021. The settings used to construct and train the NN are provided in Table 12. The HRRR 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 system was trained to predict the occurrence of these report types within 2-hr (i.e., a 4-hr window) and 40-km of each grid point.

A new version of the system is being tested in 2023 that includes 6 new predictors related to convective mode. The convective mode predictors are derived from output from two machine learning algorithms that classify each HRRR storm into one of three convective modes (supercell, QLCS, or disorganized). These storm-based mode classifications are then gridded and added to the 42 base predictors from the HRRR. Evaluations will take place between the two versions of the system, one with and one without the convective mode predictors to determine where and when differences emerge.

Table 12 The 55 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 version of the system using convective mode information.

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		
Conv. neural network supercell prob.	**Mode**	**Dense neural network supercell prob.**	**Mode**
Conv. neural network QLCS prob.	**Mode**	**Dense neural network QLCS prob.**	**Mode**
Conv. neural network disorganized prob.	**Mode**	**Dense neural network disorganized prob.**	**Mode**

Table 13 Settings used to construct and train the NNs.

Neural Network Hyperparameter	Value
Number of hidden layers	1
Number of neurons in hidden layer	16
Dropout rate	0
Learning rate	0.001
Number of training epochs	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

ii) 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 14.

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 14 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)		

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

Similar to the 2022 SFE, for the 2023 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 3). Then, for days 4-7, forecasts are issued for any hazard type.

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 15. 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 15 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

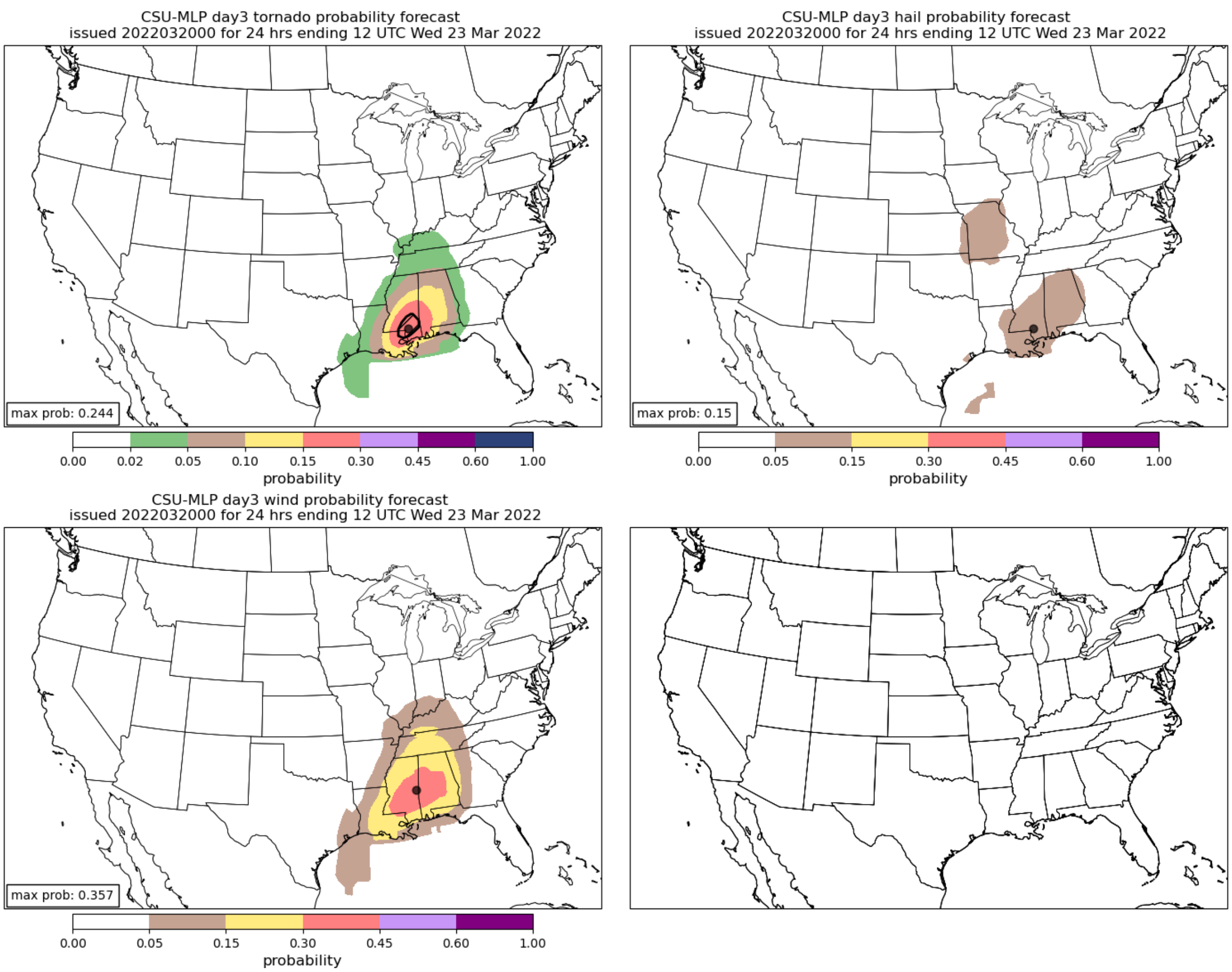

Figure 3 Probabilistic day-3 forecasts of (upper left) tornado, (upper right) hail, and (bottom left) wind hazards valid 1200 - 1200 UTC ending 23 March 2022. Hatched contours represent a 10% probability of significant severe hazards.

iv. NSSL GEFS-based, ML-derived Hazard Probabilities (credit: A. Clark)

NSSL has formulated a similar GEFS-based, random forest model to provide probabilities of any severe weather focused on lead times of 3 to 7 days. The approach is similar to CSU, but the one key difference is that training for the NSSL algorithm uses operational GEFS forecasts from the period March 2021 – February 2023, rather than GEFS reforecasts. Using the operational GEFS forecasts allows us to utilize all 31 GEFS members in training, rather than 5 members, which are used in the CSU algorithm. However, the CSU-MLP is able to leverage a longer training period since the reforecasts are available over a longer time period.

v. NSSL GEFS-based, Environment-derived Hazard Probabilities (credit: K. Hoogewind)

For this product, 20 years of GEFS reforecasts are used to develop 3- and 24-h calibrated thunderstorm and severe thunderstorm probabilities for lead times of 3-7 days. These products mirror counterpart products from SREF developed by the SPC. During the 2021 and 2022 SFEs comparisons were made to the SREF calibrated products finding that the GEFS calibrated forecasts could provide similar or improved guidance for thunderstorm and severe weather hazards relative to SREF.

vi. HREF/GEFS 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 using the following procedure. At every grid-point for the valid forecast hour, two probabilities are paired (see Table 16 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 GEFS environmental field(s) meeting a threshold over 3-h 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 16. 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.

Table 16 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

vii. 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).

iix. STP-based tornado probabilities (STP Cal MCS-TF; credit: David Jahn)

This method is similar to the STP Cal Circle method, except the 50th percentile of the STP distribution that is formulated from points within the inflow region (rather than a 40-km circular neighborhood). 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. In addition, 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. Nadocast, HREF/SREF ML-based hazard 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 performance that, on average, matches or slightly exceeds SPC 6Z Day 1 guidance for all three hazards. Only tornado probabilities were subjectively evaluated during SFE 2022, so SFE 2023 will be the first time that hail and wind probabilities will also be subjectively evaluated for Nadocast. For wind, two sets of probabilities will be provided. One set will simply use the observed severe wind reports, and in another set of probabilities the observed wind reports used in training will be weighted to match the observed climatology of measured severe wind gusts, which is meant to account for the abundance of wind reports over eastern areas of the US that are sub-severe.

x. Calibrated Ensemble tornado probabilities (Cal Ens; credit: David Jahn)

In an effort to summarize all of the HREF-based calibrated guidance products for tornado forecasting, five probabilistic products (NSSL ML Random Forest, HREF/GEFS, STP Cal Circle, STP Cal MCS-TF, and Nadocast) are processed as an ensemble. Various ensemble products (mean, median, 90th percentile) can be derived from combining these independent calibrated products.

xi. 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. 4). 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.

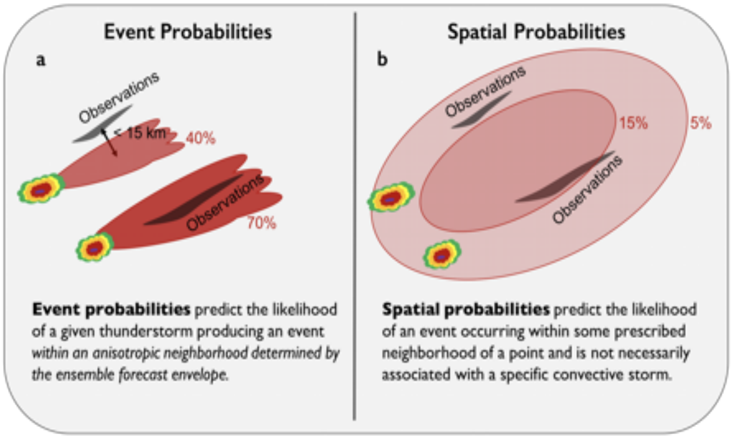

Figure 4 Illustration of the distinction between event and spatial probabilities (Fig. 2 of Flora et al. 2019).

The feature inputs were based on intra-storm and environmental variables from the WoFS and morphological variables describing the storm objects (Table 17).

Table 17 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. 2022). We show an example severe wind forecast from the logistic regression model in Fig. 5. 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.

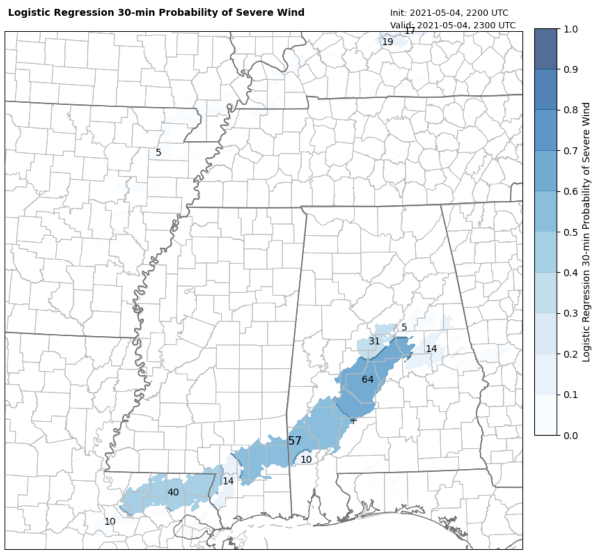

Figure 5 Example forecast from the severe-wind-based logistic regression. Number overlays indicate the probability of a severe wind within that region.

The product shown in Figure 5 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. 6). For example, in Figure 6, 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.

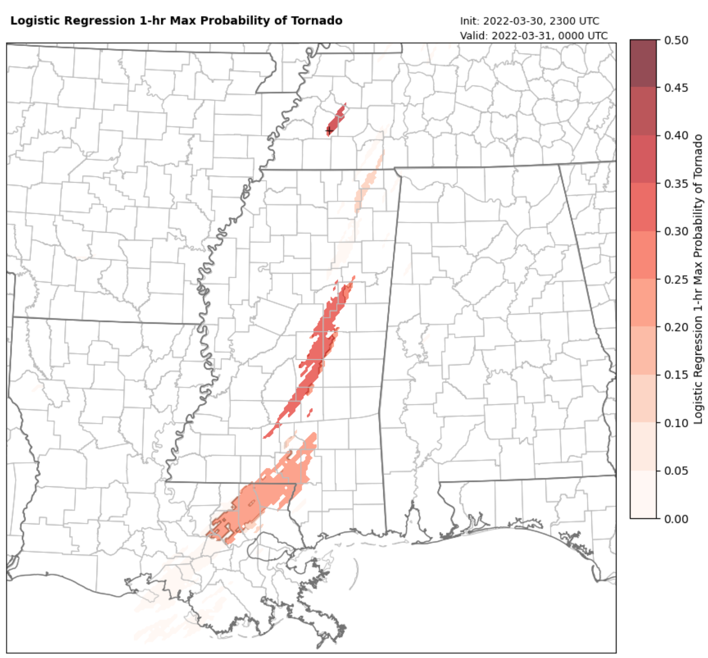

Figure 6 Example forecast of maximum tornado probability over the next hour.

xii. Machine Learning WoFS-PHI Spatial Hazard Probabilities (credit Eric Loken)

Multiple random forest (RF) machine learning models are trained to predict the probability of severe hail, wind, and tornadoes within varying spatial radii (e.g., 7.5, 15, 30, and 39km) and time windows (e.g., 30, 60, 90, and 180 minutes) at lead times between 30 minutes and 3 hours. The RFs are trained on forecast data from the Warn-on-Forecast System (WoFS) and gridded storm object attributes from ProbSevere Version 2 (PS2). MRMS-adjusted observed storm reports from the National Centers for Environmental Information Storm Events Database are used as the RF targets. The preprocessing steps are summarized in Figure 7, and the full set of predictors is summarized in Table 18.

There are two key differences between these WoFS-PHI probabilities and WoFS machine learning hazard probabilities that have been developed previously (Flora et al. 2022): first, the WoFS-PHI probabilities are valid within specific spatial radii as opposed to for a given WoFS storm object; and second, the WoFS-PHI probabilities consider observation-based PS2 data in addition to WoFS forecasts. As a result, the WoFS-PHI probabilities rapidly account for the properties of both existing and future storms, and they are designed to update frequently (e.g., every 5 minutes) to consider the most recent observations. These updates will occur even if no new WoFS run is available. Feedback is solicited on the preferred spatial scale of these probabilities at varying lead times for the different hazards as well as on the product itself.

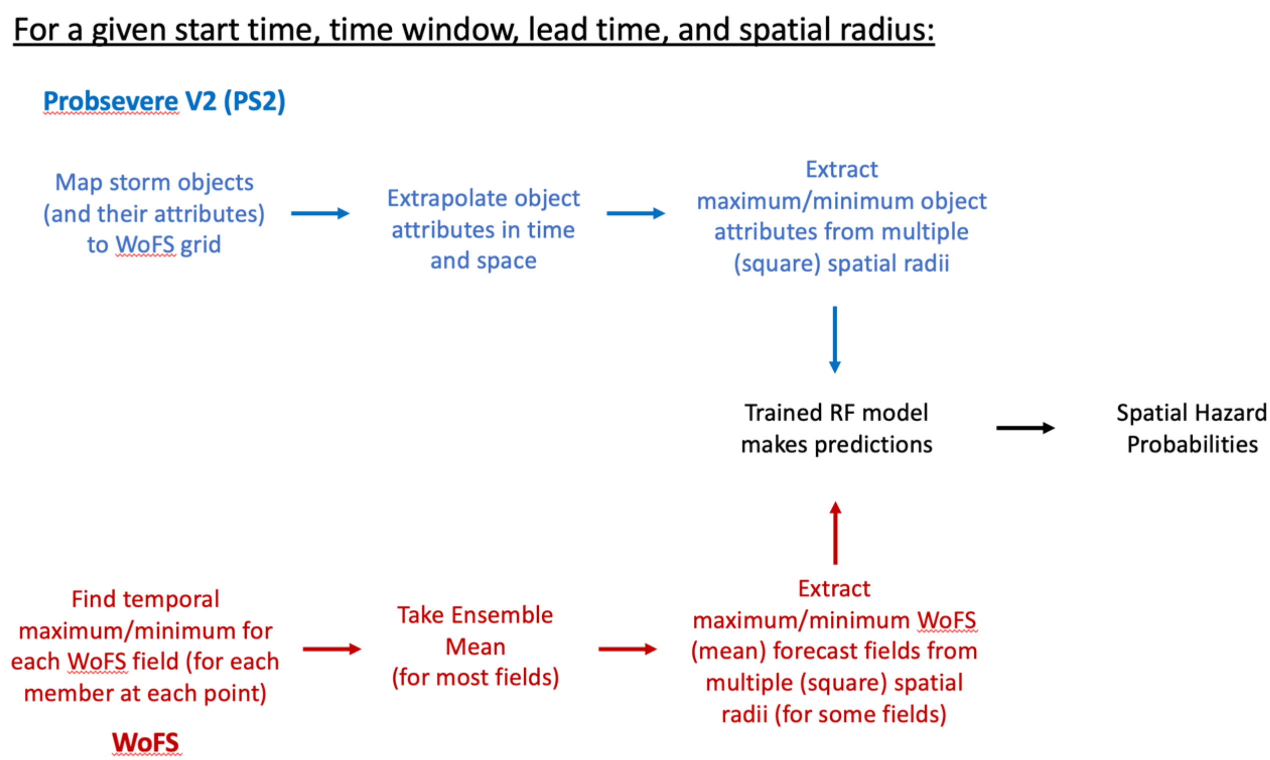

Figure 7 Flowchart showing the preprocessing steps to create WoFS-PHI spatial hazard probabilities. Steps associated with the PS2, WoFS, and combined data, respectively, are colored in blue, red, and black.

Table 18 Summary of predictors used to create WoFS-PHI spatial hazard probabilities. Predictors are taken from the WoFS at the point of prediction (column 1), from the WoFS at multiple spatial neighborhoods (column 2), and from PS2 at multiple spatial neighborhoods (column 3).

Single-Point WoFS (Values taken from point of prediction only)	**Multiple-Point WoFS (Values taken from within 0, 15, 30, 45, and 60km of point of prediction)**	**Multiple-Point ProbSevere (Values taken from within 0, 15, 30, 45, and 60km of point of prediction)**
10-500m bulk wind shear	80m wind speed	Raw expanded, extrapolated PS2 probability of hail, wind, and tornadoes
10m wind components	1km Simulated reflectivity	Spatially-smoothed expanded, extrapolated PS2 probability of hail, wind, and tornadoes
2m temperature and dewpoint	0-2km vertical vorticity	Age of storm object
0-1, 0-3, and 0-6 km wind shear components	0-2 and 2-5km updraft helicity	Extrapolation lead time (minutes)
0-500m, 0-1km, and 0-3km storm relative helicity (SRH)	Updraft speed (1km and column-maximum)	14- and 30-minute hail, wind, and tornado PS2 probability changes
Surface-based CAPE	Flash extent density (FED)	
Significant Tornado Parameter (STP); Traditional and using 0-500m SRH	Downdraft speed	
Supercell Composite Parameter (SCP)	Mean sea level and surface pressure	
Cloud Top Temperature	Ensemble probability of reflectivity exceeding 40 dBZ	
Surface-based LCL	Individual-member 2-5km updraft helicity	
Predicted Hail		
Freezing Level		
Binary proxy variable indicating if storm was present in the WoFS initial conditions		
Latitude & Longitude		

xiii. HREF Calibrated Thunder (credit: D. Harrison)

The HREF calibrated thunder (HREFCT; Harrison et al. 2022) guidance is a suite of probabilistic forecast products designed to predict the likelihood of at least one cloud-to-ground (CG) lightning flash within 20 km of a point during a given 1-, 4-, and 24-h time intervals. This guidance takes advantage of a combination of storm attribute and environmental fields produced by the HREF to objectively improve upon lightning forecasts generated by the SREF.

xiv. SPC Timing Guidance (credit: Israel Jirak)

SPC Timing Guidance products (valid over 4-h time windows throughout the convective day) are generated for tornadoes, wind, and hail. These products are created by using different datasets (i.e., HREF/SREF calibrated severe, HREF/GEFS calibrated severe, HREFCT, and Nadocast) that contain hourly 4-h timing information to distribute probabilities from the SPC outlooks across the 20 to 12 UTC time frame. Thus, they are a blend of the human forecast and various probabilistic calibrated guidance products.

f) SPC Impacts System

SPC maintains an internal analytics system (e.g., Clark et al 2019) for quantifying the number of tornadoes and their potential impacts to society, using the Day 1 tornado forecast (both coverage and conditional intensity forecasts) as the initial input. From this input, the system runs a series of Monte-Carlo-like simulations (currently set to n=10000 simulations) that draw from a number of historical distributions (tornado frequency per unit area, tornado rating, path length, direction) to produce n possible realizations of a tornado day. Each of these realizations are then overlaid on 1-km gridded societal data (e.g., population, schools) from the 2020 Census, such that the potential impact to society from each tornado (and thus each realization) can also be quantified. Additionally, a machine-learning/regression workflow (trained on historical tornadoes from 1999 to 2021) is used to predict a number of injuries and fatalities associated with each tornado.

With impact numbers across each of these realizations, the resultant distributions of potential impacts can be used to calculate descriptive statistics (e.g., 25th percentile, median, 75th percentile). Box-and-whisker plots are generated for the number of tornadoes (organized by rating), the number of injuries and fatalities, and the number of potential population, schools, and mobile homes impacted (e.g., Fig. 8). Thus, this system can be used to convert human-generated tornado forecasts into quantifiable impact data that can be communicated to partners in emergency management, etc. for improved preparedness.

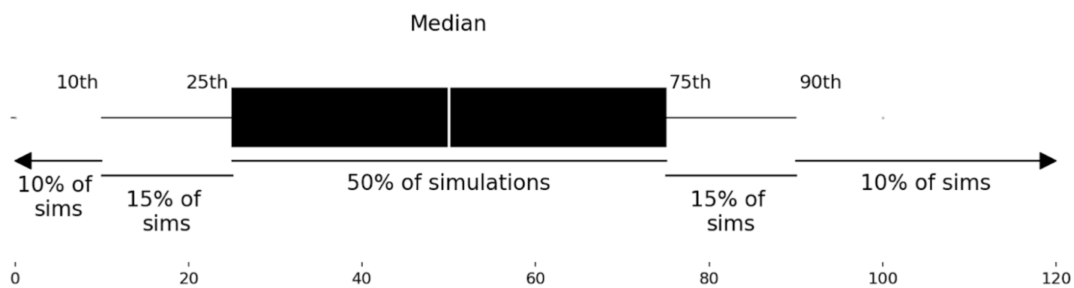

Figure 8 An example of tornado count and impact output from the IMPACTS system. Boxes represent the 25th-75th percentiles, while whiskers represent the 10th-25th and 75th-90th percentiles. The values of the 25th, 50th (median), and 75th percentiles are annotated on the individual box-and-whisker charts.

3. SFE 2023 Core Interests and Daily Activities

2023 SFE activities will occur from 9am-4pm CDT on Mondays, and 8:45am-4pm Tuesday-Friday with an optional 8-8:45am period for map analysis, data loading, and networking. Each day will have a lunch break from 12:30-2pm CDT. On Wednesdays there will be an optional science panel discussion from 1:15-2pm. Tables 19 and 20 provide a schedule for Monday, and Tuesday-Friday, respectively. Further details are provided in subsequent sections.

Table 19 Schedule for Monday.

Time (CDT)	
9:00 AM – 9:45 AM	**Welcome and Introductions** *Hybrid All* (Israel Jirak & Participants)
9:45 AM – 10:30 AM	**HWT SFE Scientific Objectives and Goals** *Hybrid All* (Israel Jirak & Adam Clark)
10:30 AM – 10:45 AM	***Break*** Fill out IRB Consent Form
10:45 AM – 11:00 AM	**Conditional Intensity Forecasting Overview** *Hybrid All* (Israel Jirak)
11:00 AM – 11:15 AM	**Weather Briefing** *Hybrid All* (David Imy)
11:15 AM – 12:30 PM	**Group Forecasting Activity** (Coverage and Conditional Intensity Outlooks) *In-Person R2O* (Day 1); *In-Person Innovation* (Days 3 & 4); *Virtual* (Day 2)
12:30 PM – 2:00 PM	***Lunch/Break***
2:00 PM – 2:15 PM	**Update on Today's Weather** *Hybrid All* (David Imy)
2:15 PM – 3:15 PM	**Individual Forecasting Activity** (Mesoscale Discussions and Discussion) *In-Person R2O* (Meso-beta MD); *In-Person Innovation* (WoFS MD); *Virtual* (WoFS MD)
3:15 PM – 4:00 PM	**Individual Forecasting Activity Continued** (MD & Day 1 Updates) *In-Person R2O* (Day 1 Update); *In-Person Innovation* (WoFS MD); *Virtual* (WoFS MD)

Table 20 Schedule for Tuesday – Friday.

Time (CDT)	
8:00 AM – 8:45 AM	*(Optional) Map Analysis, Data Loading, and Networking* *In-Person (Optional)*
8:45 AM – 9:00 AM	**Overview of Yesterday's Severe Weather** *Hybrid All* (David Imy)
9:00 AM – 10:30 AM	**Model & Outlook Evaluation** (Orientation, Surveys, and Discussion) *Hybrid Groups* (Group 1; Group 2; Group 3)
10:30 AM – 10:45 AM	***Break***
10:45 AM – 11:00 AM	***Evaluation Highlights*** *Hybrid All* (Group 1; Group 2; Group 3)
11:00 AM – 11:15 AM	**Weather Briefing** *Hybrid All* (David Imy)
11:15 AM – 12:30 PM	**Group Forecasting Activity** (Coverage and Conditional Intensity Outlooks) *In-Person R2O* (Day 1); *In-Person Innovation* (Days 3 & 4); *Virtual* (Day 2)
12:30 PM – 2:00 PM	***Lunch/Break*** ***Science Discussion (Wednesdays @ 1:15)***
2:00 PM – 2:15 PM	**Update on Today's Weather** *Hybrid All* (David Imy)
2:15 PM – 3:15 PM	**Individual Forecasting Activity** (Mesoscale Discussions and Discussion) *In-Person R2O* (Meso-beta MD); *In-Person Innovation* (WoFS MD); *Virtual* (WoFS MD)
3:15 PM – 4:00 PM	**Individual Forecasting Activity Continued** (MD & Day 1 Updates) *In-Person R2O* (Day 1 Update); *In-Person Innovation* (WoFS MD); *Virtual* (WoFS MD)

a. Formal Evaluation Activities

SFE 2023 will feature one period of formal evaluation from 9-11:00am CDT Tuesday-Friday. The evaluations will be done in three hybrid groups (i.e., each group will have in-person and virtual participants) and involve comparisons of different ensemble diagnostics, CLUE ensemble subsets, HREF, 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 1, 2, & 3, which will each conduct a separate set of evaluations. In each group, for each set of evaluations, a short tutorial will be presented and then participants will conduct the evaluations independently while facilitators remain available for questions. Following each set of evaluations, there will be a short discussion period during which participants can discuss noteworthy aspects of the evaluations, evaluation philosophy, questions, or any other topics related to the evaluations. The evaluations will end at 10:30am, followed by a 15-minute break, and from 10:45-11:00am each evaluation group will have 5 minutes to discuss highlights from their group with all participants. The evaluations are categorized as "CAM (E)nsembles", "(D)eterministic CAMs", "(A)nalyses", "Funded (P)rojects", "(C)alibrated Guidance", or "(O)utlooks". The letter in parentheses combined with a number is used to label the individual evaluations in each category (e.g., E1 refers to the first CAM Ensemble evaluation). Each evaluation group will conduct a mix of evaluations from each category. The evaluations in each category are summarized below:

(C)alibrated Guidance

C1. Day 2 12Z HREF Calibrated Tornado Guidance

Five different methods based on HREF for deriving calibrated Day 2 tornado guidance are subjectively rated along with an ensemble product derived from those methods. These methods include: (1) HREF/GEFS Cal, (2) STP Cal Circle, (3) Nadocast, (4) ML Random Forest, (5) STP Cal MCS-TF, and (6) Cal Ens Mean.

C2. Day 1 12Z HREF Calibrated Tornado Guidance

The same methods as in the Day 2 evaluation are rated for Day 1.

C3. 1630Z 4-h SPC Tornado Timing Guidance (hourly 20-12Z)

Four different sets of timing guidance products are used to produce calibrated tornado guidance valid in hourly 4-h time windows between 20 and 12 UTC that is consistent with the operational SPC outlook. For each of the timing guidance products, timing information is used to distribute probabilities from the SPC outlooks across the 20 to 12 UTC time frame. Timing guidance (TG) products are derived from four different inputs: (1) HREF/SREF TG, (2) HREF/GEFS TG, (3) HREFCT TG, and (4) Nadocast TG, each of which will be subjectively rated.

C4. Day 2 12Z HREF Calibrated Hail Guidance

Three different methods based on HREF for deriving calibrated Day 2 hail guidance are subjectively rated. These methods include: (1) HREF/GEFS Cal, (2) Nadocast, and (3) ML Random Forest.

C5. Day 1 12Z HREF Calibrated Hail Guidance

The same methods as in the Day 2 evaluation are rated for Day 1.

C6. Day 1 12Z HREF Calibrated Hail Guidance: MESH (Maximum Estimated Size of Hail)

HREF/GEFS Cal (rating carried over previous evaluation) is compared to HREF/GEFS MESH for Day 1. For this comparison, both the practically perfect probabilities derived from severe hail reports and the Multi-Radar, Multi-Sensor (MRMS) derived MESH are available.

C7. 1630Z 4-h SPC Hail Timing Guidance (hourly 20-12Z)

The same approach used for Tornado Timing Guidance (C3) is used for hail.

C8. Day 2 12Z HREF Calibrated Wind Guidance

Four different methods based on HREF for deriving calibrated Day 2 wind guidance are subjectively rated. These methods include: (1) HREF/GEFS Cal, (2) Nadocast, (3) ML Random Forest, and (4) Nadocast Adj.

C9. Day 1 12Z HREF Calibrated Wind Guidance

The same methods as in the Day 2 evaluation are rated for Day 1.

C10. 1630Z 4-h SPC Wind Timing Guidance (hourly 20-12Z)

The same approach used for Tornado Timing Guidance (C3) is used for wind.

C11. Medium Range 00Z GEFS Total Severe

Three different sets of extended range total severe probabilities for Day 3-7 lead times are subjectively rated. These methods include: (1) CSU-MLP, (2) NSSL, and (3) NSSL ML.

C12. 00Z HRRR NCAR NN Tor/Hail/Wind Guidance

Day 1 tornado, hail, and wind predictions from v2 of the NCAR NN hazard probabilities are compared in versions with and without convective mode information in the predictors.

Primary Science Question(s): 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?

(O)utlook Evaluations

O1. Day 1/2/3/4 Outlooks

The experimental Day 1-3 outlooks for tornado, wind, and hail, and Day 4 outlook for total severe produced by SFE teams are subjectively rated and compared.

O2. Day 1 Outlook Update (w/ WoFS)

The Day 1 outlooks for tornado, wind, and hail are compared to the Day 1 outlook updates, which are produced utilizing WoFS by Forecaster #1, Forecaster #2, and a consensus of non-forecaster participants.

Primary Science Question(s): How does the skill for tornado, hail, and wind severe outlooks vary with increasing lead time? How skillful are the Day 4 total severe outlooks and was CAM guidance useful at this lead time?

O3. SPC Impacts System: Day 1 Outlook Tornado Counts and Impacts

The SPC Impacts System is run on the Day 1 tornado outlooks with conditional intensity information to estimate the number of tornadoes by EF scale and the potential societal impacts.

Primary Science Question(s): Would this information be helpful in communicating the potential severe weather impacts on a given day? What is the best way to visualize this information?

CAM (E)nsembles

E1. CLUE: 00Z RRFS vs. HREF

This evaluation will feature an in-depth examination of severe storm attribute and environmental fields from 00Z initialized versions of RRFS and HREF for Day 1 lead times. These comparisons will serve to unearth ways in which the currently operational CAM ensemble (the HREF) differs from a candidate to replace it (the RRFS), and whether the RRFS 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.

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

E2. CLUE: 12Z Day 1 RRFS Physics & Time-Lagging vs. HREF

Because of computational constraints, 6-hourly RRFS ensemble initializations may only have 6 members when implemented operationally. Thus, various time-lagging strategies will be tested to assess whether they can meet or exceed the skill of a 10-member RRFS ensemble initialized from a single time (*t*), as well as the operational baseline of HREF. Additionally, the impact of mixed physics will be compared to a single-physics approach (with stochastic perturbations). Subjective ratings will be assigned to the following RRFS configurations based at 12Z: (1) RRFS (10 members at *t*), (2) RRFS-TL6 (6 members at *t* and 6 members at *t-6h*), (3) RRFS-TL12 (6 members at *t* and 6 members at *t-12h*), (4) RRFSphys (10 members at *t*), (5) RRFSphys-TL6 (6 members at *t* and 6 members at *t-6h*), and (6) HREF. In the 6- and 12-h time lagged configurations, members RRFS (ctl), RRFS01, RRFS02, RRFS03, RRFS04, and RRFS05 will be used for the single-physics ensemble, and members RRFS (ctl), RRFSphys01, RRFSphys02, RRFSphys05, RRFSphys06, and RRFSphys07 will be used for the mixed-physics ensemble.

Primary Science Question(s): Can time-lagging strategies help meet or exceed the skill of non-time-lagged ensembles? What are optimal time-lagging strategies? Does a mixed-physics ensemble improve the probabilistic forecasts compared to a single-physics ensemble?

E3. CLUE: 12Z Day 2 RRFS Physics and Time-Lagging vs. HREF

This evaluation is similar to E2, except for Day 2 lead times.

Primary Science Question(s): Can time-lagging strategies help meet or exceed the skill of non-time-lagged ensembles? What are optimal time-lagging strategies? Does a mixed-physics ensemble improve the probabilistic forecasts compared to a single-physics ensemble?

E4. CLUE: Medium-Range Lead Time/Core/Members

Subjective ratings of forecast skill are assigned to CAM ensemble guidance from a 5-member subset of the NCAR FV3 and 5-member MPAS ensemble at lead times of 3-5 days. Additionally, subjective ratings are assigned to the 10-member NCAR FV3 for lead times of 3-7 days.

Primary Science Question(s): How does CAM ensemble forecast skill vary with increasing lead time? What is the maximum lead time at which CAM ensembles have value? What are the differences in forecast quality and characteristics between the FV3 and MPAS model cores for lead times of 3-5 days?

(D)eterministic CAMs

D1. CLUE: 00Z Day 1 Deterministic Flagships

This activity will focus on rating the primary deterministic CAMs provided by several SFE collaborators – GFDL (*GFDL FV3*), NSSL (*NSSL MPAS RT*), EMC/GSL (*RRFS ctl*), and NASA (*NASA FV3*) – based on their skill and utility for severe weather forecasting. These runs will be compared to the operational HRRR, 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(s): How do various deterministic CAMs compare to the operational standard for convective forecasting (i.e., WRF-ARW-based HRRRv4)?

D2. CLUE: 00Z Day 2 Deterministic Flagships

Five deterministic CAM configurations are subjectively evaluated for Day 2 lead times. These configurations include: (1) NAM Nest, (2) RRFS, (3) NSSL MPAS RT, (4) GFDL FV3, and (5) NASA FV3.

Primary Science Question(s): What strategies for CAM configurations perform the best at Day 2 lead times, and what are their forecast characteristics at Day 2 lead times for severe weather forecasting applications?

D3. CLUE: RRFS vs. HRRR

This activity will feature a "deeper dive" into storm attribute and environmental fields in HRRR and RRFS. These comparisons will serve to unearth ways in which the currently operational CAM (the HRRR) differs from a candidate to replace it (the RRFS), and whether the RRFS improves upon or degrades forecasts of the HRRR 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 Question(s): How do forecasts of the RRFS compare to those of the HRRR? Are there systematic shortcomings or advantages of the RRFS?

D4. CLUE: RRFS vs. HRRR DA

The HRRR and RRFS are examined in the first 12 hours of the forecast period for 21 and 00Z initializations to evaluate the impact of their data assimilation.

Primary Science Question(s): How do the data assimilation strategies in HRRR and RRFS impact short-term convective weather forecasts?

D5. CLUE: 00Z MPAS

The three configurations of MPAS run by NSSL are assigned subjective ratings. These configurations include: (1) NSSL MPAS HT, (2) NSSL MPAS HN, and (3) NSSL MPAS RT. "HT" refers to HRRR initialization with Thompson microphysics, "HN" is HRRR initialization with NSSL microphysics, and "RT" is RRFS initialization with Thompson microphysics (Table 5).

Primary Science Question(s): What is the impact of Thompson vs. NSSL microphysics parameterizations in MPAS configurations initialized from HRRR for severe weather forecasting applications? In similarly configured MPAS runs, what is the impact of initializing with the HRRR vs. the RRFS?

D6. CLUE: 1-km vs. 3-km

This comparison will focus on comparing the NSSL1 and HRRR configurations of WRF-ARW, which have 1- and 3-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(s): Does decreasing 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?

(A)nalyses

A1. Mesoscale Analysis Background

Hourly versions of the 3D-RTMA using HRRR as the background will be compared to another version using RRFS as the background to assess the role that the background first-guess plays on the final analysis. Comparisons will be conducted using the analyses on their native 3-km grids, as well as upscaled to 40-km grids, so that equitable comparisons can be made to the sfcOA. The goal is to assess the utility of these analysis systems for situational awareness and short-term forecasting for convective-weather scenarios.

Primary Science Question(s): What are the optimal methods for producing quality mesoscale analyses for convective forecasting applications?

A2. Storm Scale Analysis

WoFS-based "analyses" (actually 15-minute maximum forecasts) of 80-m wind are compared to preliminary local storm reports, including gust measurements and estimates. Additionally, similar WoFS-based, 15-minute maximum 2-5 km AGL UH and updraft speed are compared to MRMS Mid-Level Rotation Tracks (MLRT) and MRMS MESH, respectively.

Primary Science Question(s): Can a high resolution, rapidly updating ensemble DA system serve as a verification source for severe winds, mesocyclone tracks, and hail?

Funded (P)roject Evaluations

P1. 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 $\geq$ 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 two different ML techniques.

Primary Science Questions: Can machine-learning approaches provide useful information regarding the likelihood of wind damage reports being associated with gusts ≥ 50 knots? Is it preferred to provide the probabilities from only the algorithm determined to be the best by objective skill metrics, or instead probabilities from several machine-learning algorithms?

P2. WoFS-PHI Spatial Hazard Probabilities

This evaluation will assess the utility of grid-based, machine-learning probabilities that use input from ProbSevere and WoFS to produce short-term calibrated severe hazard guidance at lead times up to 3 hours. With probabilities produced for 7.5-, 15-, 30-, and 39-km radii, participants will be asked to identify the spatial scale of probabilities that are most useful for providing skillful guidance.

Primary Science Question(s): What are the minimum spatial scales for probabilities that provide useful guidance? Does this minimum spatial scale vary according to lead time? Does this minimum spatial scale vary depending on whether the regime is pre- or post-CI?

b. Forecast Products and Activities

There will be two periods of experimental forecast activities during SFE 2023. The first will occur from 11:00am – 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. Participants will be split into three groups: (1) In-Person R2O, (2) In-Person Innovation, and (3) Virtual. As the naming convention suggests, in-person participants will be in R2O and Innovation groups, while all virtual participants will be in the Virtual group. The In-Person R2O group will issue products for Day 1, the Virtual group will issue products for Day 2, and the In-Person Innovation group will issue products for Days 3 & 4. 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 & 4 forecast is the only exception to the smaller domain, and will instead cover a full CONUS domain. Also, the Day 4 outlooks will only cover total severe (i.e., no individual hazards or conditional intensity forecasts).

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, & 4 outlooks will cover 1200 – 1200 UTC periods. Additionally, for experimental outlooks covering Days 1, 2, & 3, conditional intensity forecasts of tornado, wind, and hail will be issued, in which areas are delineated with reports that are expected to follow intensity distributions defined by conditional intensity groups (see more information below). These conditional intensity forecasts are similar to those issued during SFEs 2019-2022. 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 groups (CIG) that can be forecast via examination of the atmospheric environment: no CIG, CIG 0, CIG 1, and CIG 2. In plain language, CIG 0 refers to a typical severe weather day, where significant severe weather is unlikely, CIG 1 areas indicate where significant severe weather is possible, and CIG 2 areas indicate where high impact significant severe weather is expected. All groups will have access to all available operational and experimental guidance products for issuing their outlooks.

The second period of experimental forecasting activities will occur during the 2-4pm CDT time period. From 2-2:15pm CDT, a weather briefing led by Dave Imy will be conducted for all participants during which an update on current weather will be given. In the In-Person R2O group, the 2:15-3:15pm 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 3:15-4pm CDT time period, each In-Person R2O participant will use WoFS and other available guidance to update the Day 1 individual hazard coverage and conditional intensity forecasts for the period 2100 – 1200 UTC.

During the 2:15-4pm CDT time period in the In-Person Innovation Group and Virtual Group, another activity will be devoted to issuing short-term, meso-beta to meso-gamma scale predictions of severe weather. In this activity, each participant will issue a forecast consisting of two parts: (1) a geographic threat area (i.e., graphic) and (2) a text discussion. The geographic threat area will be created using the WoFS web viewer drawing tool and may take one of three formats: (1) A single contour highlighting a region of expected severe weather along the track of an individual storm, (2) two contours, one encompassing a broader region where severe weather is expected and the second, smaller contour outlining what is perceived as the corridor of greatest risk, or (3) A single contour that highlights a broader region where severe weather is expected. Each participant will issue their first set of predictions during the 2:15-3pm CDT time period, and then from 3-3:15pm CDT each participant will have an opportunity to present and discuss their product. Then, from 3:15-3:45pm CDT the outlooks and text discussions will be updated with a focus on how more recent observations and more up-to-date WoFS guidance is influencing the perceived threat and confidence in the forecast. For example, does WoFS indicate increasing or decreasing likelihood of an event relative to previous guidance, or does the more recent guidance simple reinforce earlier guidance? Finally, from 3:45-4pm each participant will participate in a short survey with some targeted questions on WoFS products used, changes in forecasts between 1st and 2nd hours, and overall confidence.

These WoF activities are the seventh 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 2023 participants. Green denotes participants that are observing and not directly participating in activities.

Week 1	Week 2	Week 3	Week 4	Week 5
1-5 May	**8-12 May**	**15-19 May**	**22-26 May**	**30 May - 2 June**
Lizzie Tirone (ISU)	Marion Mittermaier (UK Met)	Will Mayfield (NCAR/DTC)	Austin Coleman (WPC)	Cindy Wang (GFDL)
Terra Ladwig (GSL)	Ryan Sobash (NCAR)	Steve Willington (UK Met)	Steve Willington (UK Met)	Curtis Alexander (GSL)
Thomas Galarneau (NSSL)	Mike Kavulich (DTC)	Allie Mazurek (CSU)	Ed Szoke (GSL)	Brice Coffer (NCSU)
Mark Jarvis (WFO LMK)	Lauren Pounds (OU)	Matt Morris (EMC)	Craig Schwartz (NCAR)	Marcel Caron (EMC)
Derek Williams (WFO LKN)	Lizzie Tirone (ISU)	Bob Rozumalski (FDTD)	Kelly Lombardo (PSU)	Cameron Nixon (CMU)
Andew Loconto (WFO BOX)	Jeff Duda (GSL)	Brian Haynes (MDL)	Matt Kumjian (PSU)	Jack Lind (AWC)
Ted Mansell (NSSL)	James Sullivan (WFO CLE)	Nicholas Goldacker (NCSU)	Ben Blake (EMC)	Lizzie Tirone (ISU)
Matt Flournoy (SPC)	Sarah McCorkle (WFO MTR)	Harald Richter (BoM)	John Allen (CMU)	Bill Gallus (ISU)
Evan Bentley (SPC; M-W)	Matthew Gropp (WFO CAE)	Lizzie Tirone (ISU)	Lizzie Tirone (ISU)	Montgomery Flora (CIWRO/NSSL)
Brian Squitieri (SPC; Th-F)	Jaret Rogers (WFO BOI)	Jeff Beck (GSL)	Harald Richter (BoM)	Brett Borchardt (WFO LOT)
Chang Jae Lee (KMA)	Chris Kerr (CIWRO/NSSL)	John Boris (WFO APX)	Sam Shamburger (WFO OHX)	Matt Anderson (WFO TBW)
Derek Stratman (CIWRO/NSSL)	Matt Mosier (SPC; M-W)	Michael Bowlan (WFO TSA)	Justin McReynolds (WFO KEY)	Jared Schadler (AWC)
Jacob Widanski (OU/SoM)	Brian Squitieri (SPC; W-Th)	Sam Childs (USAF)	Bill Putman (NASA; M-W)	Mike Dutter (WFO AKQ)
Greg Mcfarquhar (CIWRO; Th-F)	Bryan Smith (SPC; T,F)	Andy Dean (SPC; M-T)	Andrew Bufalino (BoM)	Corey Potvin (NSSL)
Rebecca Cyr (EC; T)	Patrick Burke (NSSL)	Matt Mosier (SPC; W)	Brad Vrolijk (EC; T)	Ryan Rozinskis (EC; T)
Ray Houle (EC; T)	Jordan Dale (WPO; T)	Aron Stanton (EC; T)		
***** VIRTUAL PARTICIPATION *****				
Jana Houser (OSU)	Michelle Harold (DTC)	Matt Pyle (EMC)	Alex Anderson-Frey (UW)	Ben Price (Ohio)
Geoff Manikin (EMC)	Shun Liu (EMC)	Gang Zhou (EMC)	Abigail King (UW)	Logan Dawson (EMC)
Donald Lippi (EMC)	Eric Aligo (EMC)	Aaron Hill (CSU)	Sho Yokota (EMC)	Ed Colon (EMC)
Jason Jordan (FDTD)	Jacob Carley (EMC)	Rosie Jones (M-W; UK Met Office)	Michael Sessa (Illinois)	Matt Eastin (UNCC)
Becky Adams-Selin (AER)	Russ Schumacher (CSU)	Sean Whelan (OSU)	Jason Puma (WFO IND)	Shawn Murdzek (GSL)
Mark Antolik (MDL)	Dave Rudack (MDL)	Kendall Parks (JSU)	Derek Hodges (WFO TSA)	Jeff Makowski (WFO ARX)
Binbin Zhou (EMC)	Nicholas Price (JSU)	David Byers (WFO GJT)	Brian Mejia (WFO BRO)	Keith Sherburn (WFO UNR)
Sarah Trojniak (WPC)	Jimmy Correia (WPC)	David Zaff (WFO BUF)	Chris Melick (USAF)	Eswar Iyer (WFO AKQ)
Dave Ahijevych (NCAR)	Christina Anderson (WFO RAH)	Michael Laczko (BoM)	Callum Stuart (BoM)	Alyssa Bates (WDTD; PM)
Brad Carlberg (WFO CAE)	Robert Setzenfand (WFO BYZ)	Alyssa Bates (WDTD; PM)	Alyssa Bates (WDTD; PM)	

SFE Facilitators: Adam Clark (NSSL), Israel Jirak (SPC), Dave Imy (retired SPC), Tim Supinie (SPC), Kenzie Krocak (CIWRO/SPC/CRCM), 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), Andy Wade (CIWRO/SPC), Allie Brannan (CIWRO/SPC), Pamela Heinselman (NSSL), Joey Picca (CIWRO/SPC), Patrick Skinner (CIWRO/NSSL), Patrick Burke (NSSL), 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

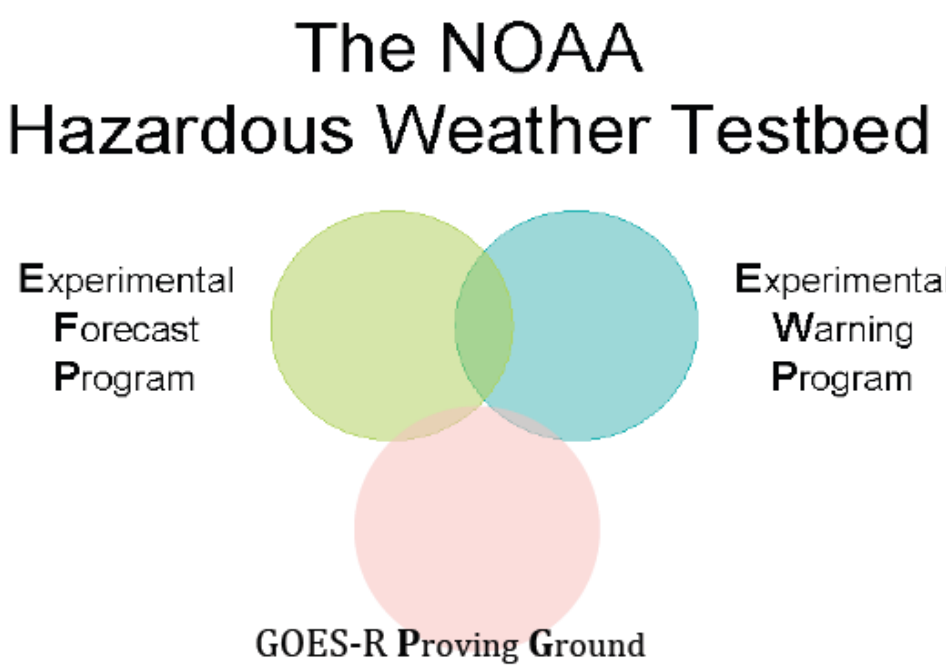

Figure B1: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

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; 2022; 2023), and Gallo et al. (2017).

Appendix C: Mandatory 2023 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, R., P. T. Marsh, R. S. Schneider, and S. A. Erickson, 2019: Using NOAA/NWS Storm Predictino Center Forecasts to Estimate Potential Societal Impacts from Severe Weather. *35th Conf. on Environmental Information Processing Technologies*, Amer. Meteor. Soc., 2.1 (https://ams.confex.com/ams/2019Annual/webprogram/Paper352385.html).

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.*, E814-816. https://doi.org/10.1175/BAMS-D-20-0268.1

Clark, A. J., and Coauthors, 2022: The Second Real-Time, Virtual Spring Forecasting Experiment to Advance Severe Weather Prediction. *Bull. Amer. Meteor. Soc.*, E1114-1116. https://doi.org/10.1175/BAMS-D-21-0239.1

Clark, A. J., and Coauthors, 2023: The Third Real-Time, Virtual Spring Forecasting Experiment to Advance Severe Weather Prediction Capabilities. Bull. Amer. Meteor. Soc., E456-458. https://doi.org/10.1175/BAMS-D-20-0268.1

Flora, M. L., Potvin, C. K., Skinner, P. S., Handler, S., & McGovern, A., 2022: Using Machine Learning to Generate Storm-Scale Probabilistic Guidance of Severe Weather Hazards in the Warn-on-Forecast System, *Mon. Wea. Rev*., **149**, 1535-1557.

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.

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.

Harrison, D. R., M. S. Elliott, I. L. Jirak, and P. T. Marsh, 2022: Utilizing the High-resolution Ensemble Forecast System to Produce Calibrated Probabilistic Thunderstorm Guidance. *Wea. Forecasting*, **37**, 1103-1115.

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.

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, 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.

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.

Zhou, L., Harris, L., Chen, J.-H., Gao, K., Guo, H., Xiang, B., et al. (2022). Improving global weather prediction in GFDL SHiELD through an upgraded GFDL cloud microphysics scheme. Journal of Advances in Modeling Earth Systems, 14, e2021MS002971.