SPRING FORECASTING EXPERIMENT 2024

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

HWT Facility – National Weather Center

29 April – 31 May 2024

https://hwt.nssl.noaa.gov/sfe/2024/

Program Overview and Operations Plan

22 April 2024

Adam Clark[2,4], Israel Jirak[1], Tim Supinie[1], Kent Knopfmeier[2,3], David Harrison[1,3], Jake Vancil[1,3], David Jahn[1,3], Chris Karstens[1], Eric Loken[2,3], Nathan Dahl[1,3], David Imy[2], Andy Wade[1,3], Jeffrey Milne[1,3,4], Kimberly Hoogewind[2,3], Sean Ernst [1,3], Joey Picca[1,3], Matthew Flournoy[2,4], Michael Baldwin[1,3], Pamela Heinselman[2,4], Montgomery Flora[2,3], Joshua Martin[2,3], Brian Matilla[2,3], Patrick Skinner[2,3], and 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

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 2024 HWT Spring Forecasting Experiment (SFE 2024), a cornerstone of the EFP, will be conducted 29 April – 31 May. This will be the second hybrid experiment with both in-person and virtual participation. Relative to last year's hybrid experiment, SFE 2024 will have a similar format with all participants involved in morning and afternoon forecasting activities, as well as next-day model evaluation activities. Additionally, there will be a small evening activity in which 2-4 NWS forecasters will issue experimental 0-1 and 1-2 h lead time forecasts until 8pm CDT. 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 2024 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 2024 CLUE includes 34 members. The SFE 2024 will also continue testing of the Warn-on-Forecast System (WoFS, hereafter), which produces 18-member, 3-km grid-spacing forecasts, and will be used for the 8th year to issue very short lead-time outlooks and products.

With plans for operational implementation of the Rapid Refresh Forecast System (RRFS) in 2025, SFE 2024 will once again 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. Also, SFE 2024 will include more experimental configurations of the Model for Prediction Across Scales (MPAS), including an MPAS ensemble run by NOAA's Global Systems Laboratory and NSSL.

This document summarizes the core interests of SFE 2024 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 2024 and Section 3 describes the core interests and new concepts being introduced for SFE 2024. 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 other models, guidance, and products. The 2024 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 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 2024.

a) The 2024 Community Leveraged Unified Ensemble (CLUE)

The CLUE is a carefully designed ensemble with members contributed by NOAA units: NSSL, Environmental Modeling Center (EMC), GSL, and the 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, except the NSSL1, which covers the eastern 2/3 of the CONUS. CLUE members have 3-km grid-spacing, except NASA FV3 uses 2.2-km, NSSL1 uses 1-km, MPAS2 uses 2-km, and MPAS4 uses 4-km. Depending on the CLUE subset, forecast lengths range from 36 to 192 h. Table 1 summarizes all 2024 CLUE contributions. Subsequent tables provide details on members in each subset, as well as ensembles comprising different combinations of members that will be evaluated to test different configuration strategies.

Table 1 Summary of the 11 unique subsets that comprise the 2024 CLUE.

Clue Subset	# of mems	IC/LBC perts	Mixed Physics	Data Assimilation	Dynamical Core	Agency	Init. Times (UTC)	Forecast Length (h)	Domain
RRFS	1	none	no	Hybrid 3DEnVar	FV3	EMC/GSL	00, 06, 12, 18	60	CONUS
REFS	5	EnKF	yes	Hybrid 3DEnVar	FV3	EMC/GSL	00, 06, 12, 18	60	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, 12	48 or 60	CONUS
GSL-MPAS	5	EnKF	yes	RRFS ICs	MPAS	GSL	06 or 12	54 or 48	CONUS
MPAS2	1	none	no	HRRR ICs	MPAS	GSL	00	36	CONUS
MPAS4	1	none	no	HRRRICs	MPAS	GSL	00	36	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 Rapid Refresh Forecast System (RRFS). The RRFS is initialized from a hybrid 3DEnVar analysis and is the control member of the RRFS Ensemble Forecast System (REFS). The ensemble component of the 3DEnVar uses the RRFS Data Assimilation System (RDAS) ensemble Kalman filter. The RDAS uses a wide variety of conventional observations along with radar reflectivity. It also included a nonvariational cloud analysis. For gravity wave drag, the small scale and turbulence orographic form drag options are used. RRFS forecasts are initialized from 00, 06, 12, and 18 UTC with forecasts to 60 h.

Members: RRFS	ICs	LBCs	Micro-physics	PBL/SFC	LSM	Radiation	Cumulus	Dynamical Core
RRFS	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	GF-deep	FV3

Table 3 Specifications for the RRFS Ensemble Forecast System (REFS). REFS forecasts are initialized from 00, 06, 12, and 18 UTC with forecasts to 60 h. Schemes marked with an asterisk (*) include stochastically perturbed parameterizations (SPP) and those marked with a hashtag (#) include fixed parameter perturbations.

Members: REFS	ICs	LBCs	Micro-physics	PBL/SFC	LSM	Radiation	Cumulus	Dynamical Core
REFS01	RRFS enkf1	GEFS m1	Thompson*	TKE-EDMF/GFS	RUC*	RRTMG*	GF-deep*+sh	FV3
REFS02	RRFS enkf2	GEFS m2	Thompson*	MYNN*/MYNN*	RUC*	RRTMG*	saSAS deep	FV3
REFS03	RRFS enkf3	GEFS m3	NSSL#	MYNN*/MYNN*	RUC*	RRTMG*	GF deep*	FV3
REFS04	RRFS enkf4	GEFS m4	NSSL#	TKE-EDMF/GFS	RUC*	RRTMG*	GF-deep*+sh	FV3
REFS05	RRFS enkf5	GEFS m5	NSSL#	MYNN*/MYNN*	RUC*	RRTMG*	saSAS deep	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 (Mansell et al. 2010), "HT" = HRRR-Thompson, and "RT" = RRFS-Thompson). All NSSL-MPAS runs are initialized from 00 and 12 UTC; the NSSL-MPAS-HN and NSSL-MPAS-HT have forecast lengths of 48 h, while NSSL-MPAS-RT runs to 60 h.

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 GSL-MPAS CLUE members. These members use 3-km grid-spacing covering the CONUS. Members GSL-MPAS01-04 are initialized from 12 UTC REFS members and have 48 h forecasts; member GSL-MPAS05 is initialized from the 06 UTC RRFS and has 54 h forecasts.

Member: NSSL-MPAS	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Dynamical Core
GSL-MPAS01	RRFS enkf1	GEFS mem1	Thompson	MYNN	RUC	RRTMG	MPAS
GSL-MPAS02	RRFS enkf2	GEFS mem2	Thompson	MYNN	RUC	RRTMG	MPAS
GSL-MPAS03	RRFS enkf3	GEFS mem3	NSSL	MYNN	RUC	RRTMG	MPAS
GSL-MPAS04	RRFS enkf5	GEFS mem5	NSSL	MYNN	RUC	RRTMG	MPAS
GSL-MPAS05	RRFS	06Z RRFS	Thompson	MYNN	RUC	RRTMG	MPAS

Table 7 Specifications for the MPAS2 and MPAS4 CLUE members. The MPAS2 and MPAS4 configurations are initialized at 00 UTC and have the same specifications as NSSL-MPAS-HN (Table 5), except for 2-km and 4-km grid-spacing, respectively.

Member: NSSL-MPAS	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Dynamical Core
MPAS2	HRRR	HRRR	NSSL	MYNN	RUC	RRTMG	MPAS
MPAS4	HRRR	HRRR	NSSL	MYNN	RUC	RRTMG	MPAS

Table 8 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 9 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 00 and 12 UTC 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 10 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. This version of the C-SHiELD uses the EMC/UW TKE-EDMF PBL scheme (Han and Bretherton 2019) and version 2 of the GFDL Inline Microphysics (Zhou et al. 2019; Harris et al. 2020). 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 204 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 11 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 ICs. Initialization is cold-start from members 1–5 of real-time GEFS ICs. Simulations run daily at 00Z out to 132 hours.

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 2024 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/REFS vs. HRRR/HREF: With plans for operational implementation of RRFS and REFS in 2025, a critical evaluation activity for SFE 2024 will involve comparing RRFS and REFS 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.

REFS Configuration Strategies: The RRFS control (Table 2) and perturbed REFS members (Table 3) provide 60 h forecasts, which are the basis for four different REFS versions based at 1200 UTC: (1) EMC REFS uses the RRFS, all REFS members, and the HRRR initialized at 1200 and 0600 UTC; (2) SPC REFS has fewer members and a larger proportion of "control" members compared to EMC REFS by only using 1200 UTC REFS perturbed members and including older (*t-12*) time-lagged HRRR and RRFS members; (3) SPC REFS ARW is similar to SPC REFS, but replaces two of the REFS perturbed members with HRW ARW members initialized at 1200 and 0000 UTC; and (4) MPAS REFS mirrors SPC REFS, except all RRFS and REFS members are replaced by MPAS members initialized from the same respective analyses as the RRFS and REFS members (Table 12).

Table 12 Ensemble members comprising four versions of REFS based at 1200 UTC.

	EMC REFS		SPC REFS		SPC REFS ARW		MPAS REFS	
#	Member	Init. Time	Member	Init. Time	Member	Init. Time	Member	Init. Time
1	RRFS	12Z	RRFS	12Z	RRFS	12Z	NSSL-MPAS-RT	12Z
2	REFS01	12Z	REFS01	12Z	REFS03	12Z	GSL-MPAS01	12Z
3	REFS02	12Z	REFS02	12Z	REFS02	12Z	GSL-MPAS02	12Z
4	REFS03	12Z	REFS03	12Z	HRW ARW	12Z	GSL-MPAS03	12Z
5	REFS04	12Z	REFS05	12Z	HRRR	12Z	GSL-MPAS04	12Z
6	REFS05	12Z	HRRR	12Z	RRFS	06Z	HRRR	12Z
7	HRRR	12Z	RRFS	06Z	HRRR	06Z	GSL-MPAS05	06Z
8	RRFS	06Z	HRRR	06Z	RRFS	00Z	HRRR	06Z
9	REFS01	06Z	RRFS	00Z	HRRR	00Z	NSSL-MPAS-RT	00Z
10	REFS02	06Z	HRRR	00Z	HRW ARW	00Z	HRRR	00Z
11	REFS03	06Z						
12	REFS04	06Z						
13	REFS05	06Z						
14	HRRR	06Z						

Medium-Range CAM Ensembles: NCAR will be providing a 10-member, 0000 UTC FV3-based, 3-km grid-spacing ensemble with forecasts to 7 days (Table 10). Additionally, a 5-member, 3-km grid-spacing 0000 UTC MPAS ensemble will have forecasts to 5 days (Table 11). 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.

Model resolution sensitivities: NSSL will be running a version of WRF-ARW with 1-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. Additionally, GSL is running 2- and 4-km grid-spacing versions of MPAS (Table 7) that mirror the 3-km grid-spacing NSSL-MPAS-HN. These runs will be used to assess MPAS resolution sensitivities.

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 RAP-based 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 2024 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. 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 13.

Table 13 Model specifications for HREFv3.

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

c) NSSL cloud-based Warn-on-Forecast Experiments

Cloud-based Warn-on-Forecast (cb-WoFS) is the next WoFS iteration, upgraded in 2022 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 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-ARW-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 2024, WoFS will once again 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 NCEP Central Operations. 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 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 14 cb-WoFS configuration.

	WoFS
Model Version	WRF-ARW v3.9+
Grid Dimensions	300 x 300 x 50
Grid Spacing	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)

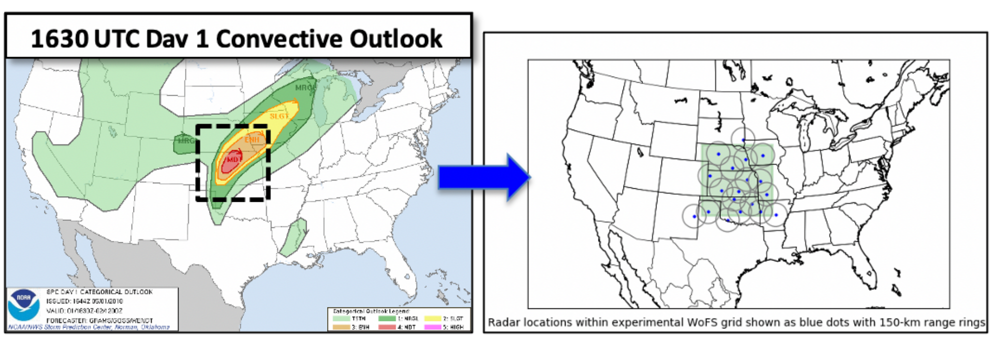

Figure 1 SPC 1630 UTC issued Day 1 convective outlook (left) and corresponding WoFS grid (right).

d) AI NWP Emulators

In the last two years, fully AI-based models (i.e., AI NWP emulators) have been developed by the private sector for global weather prediction. This area of research is advancing rapidly and has the potential to be an advancement in weather prediction since skill measures of the NWP emulators commonly exceed those of the ECMWF's Integrated Forecast System (IFS; ECMWF 2020), the world's most skillful global NWP system. Furthermore, the NWP emulators can produce forecasts in seconds, orders of magnitude faster and with fewer computational resources than traditional NWP systems. The algorithms are trained using large, global, multi-year reanalysis datasets like ERA5 (Hersbach et al. 2020). Several of these algorithms have been made public, and government agencies are beginning to run and train the models themselves. While objective skill measures have been impressive, these NWP emulators have yet to be tested for real-time operational forecasting applications. Thus, during SFE 2024, we will evaluate several of the publicly available algorithms that were trained using ERA5 data. The AI-based NWP emulators are being run experimentally at the Cooperative Institute for Research in the Atmosphere (CIRA) by providing GFS initial conditions to these AI models. The CIRA forecasts can be viewed at: https://aiweather.cira.colostate.edu/. Information on each AI NWP Emulator that will be evaluated during SFE 2024 is contained below.

i. FourCastNetv2

The AI NWP Emulator known as FourCastNet was developed by NVIDIA and uses adaptive Fourier neural operators to provide short to medium range global predictions at 0.25° resolution (Pathak et al. 2022). The code is publicly available at: https://github.com/NVlabs/FourCastNet. FourCastNet has been found to match the skill of IFS at short lead times for large-scale variables, while outperforming IFS for variables with complex fine-scale structure such as precipitation. FourCastNet can generate a week-long forecast in less than 2 seconds. A version of FourCastNet configured by CIRA, which provides GFS analyses as initial conditions, will be evaluated during SFE 2024.

ii. Pangu-Weather

Pangu-Weather is a deep learning-based system trained using 43 years of ERA5 data and was developed by Huawei Cloud (China) (Bi et al. 2022). The forecasts are produced with 0.25° resolution. At time ranges of 1 h to 1 week, Pangu-Weather was found to outperform the IFS in terms of RMSE and anomaly correlation coefficient (ACC) for fields like geopotential, specific humidity, wind speed, and temperature. Pangu-Weather is applied by designing a 3D Earth Specific Transformer architecture that formulates the pressure level information into cubic data, and applying a hierarchical temporal aggregation algorithm to alleviate cumulative forecast errors. The code is publicly available at https://github.com/198808xc/Pangu-Weather. A version of Pangu-Weather configured by CIRA, which inputs GFS analyses as initial conditions, will be evaluated during SFE 2024.

iii. GraphCast

GraphCast is a machine-learning algorithm developed by Google that is trained directly from ERA5 data (Lam et al. 2022). GraphCast predicts hundreds of weather variables over 10 days using 0.25° resolution and produces forecasts in under one minute. Objective verification found that GraphCast significantly outperformed the IFS on 90% of 1380 verification targets. The code for GraphCast is available publicly at https://github.com/deepmind/graphcast. A version of GraphCast run by CIRA will be evaluated during SFE 2024. GraphCast is pre-trained with ERA5 reanalysis data and forecasts are produced using GFS F00 initial conditions as input.

e) Calibrated Forecast Products

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

Similar to previous SFEs, the Colorado State University Machine Learning Probabilities (Hill et al. 2020; hereafter, GEFS Reforecast MLP) forecasts severe weather hazards through the application of random forests (RFs). The GEFS Reforecast MLP RFs are trained with about 9 years of daily 0000 UTC initializations from the FV3-based Global Ensemble Forecast System reforecast dataset (FV3-GEFS/R) along with reports of severe weather. For consistency with SPC outlooks as well as SFE activities, RFs are trained separately for individual hazards in the day 1-3 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 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 GEFS Reforecast 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

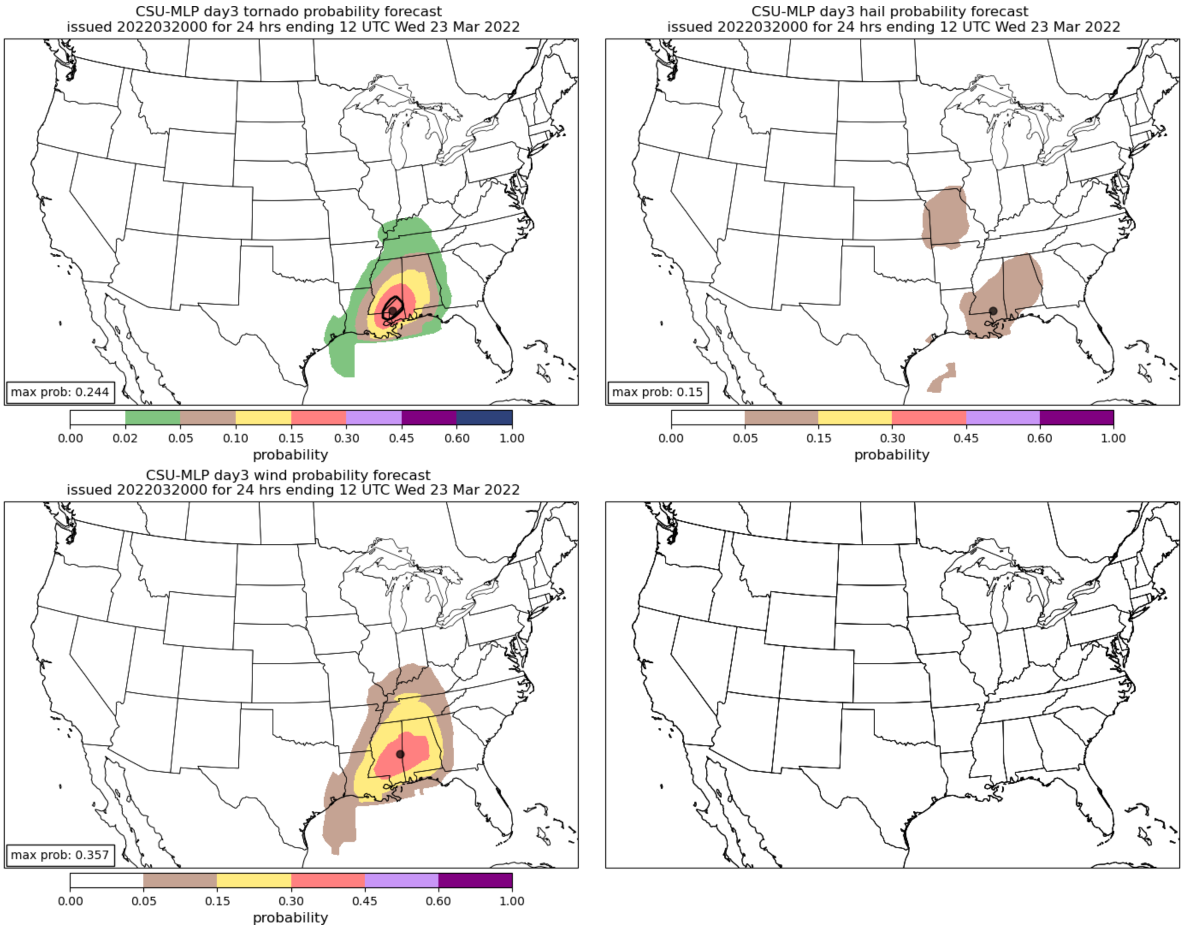

Figure 2 Probabilistic day-3 forecasts of (upper left) tornado, (upper right) hail, and (bottom left) wind hazards valid 1200 - 1200 UTC ending 23 March 2022. Hatched contours represent a 10% probability of significant severe hazards.

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

NSSL has formulated a similar RF model using archived operational GEFS data that provides probabilities of any severe weather at lead times of 1 to 15 days. The global GEFS fields are subset over a 0.5° by 0.5° grid covering the CONUS. In addition to the variables available within the GEFS forecasts, the pressure level data are used to derive additional diagnostics and indices commonly used for severe weather forecasting like bulk shear and the significant tornado parameter. After domain subsetting, the fields are interpolated to a coarser, 81-km grid that tightly encompasses the CONUS. After interpolation, only a set of 1,385 masked points covering CONUS land areas are used to train the RF. The fields are extracted at 3-hourly intervals from forecast hours 12 to 225 (Days 1-10) and 6-hourly intervals from forecast hours 228 to 372 (Days 11-15). In Table 16, the GEFS fields used for predictors in the new RF algorithm (GEFS Operational MLP, hereafter) are listed, with the fields that required additional post-processing marked with an asterisk.

There are several notable differences between GEFS Operational MLP and GEFS Reforecast MLP. First, GEFS Reforecast MLP uses 12 different variables as predictors and no additional diagnostics are computed from the pressure level data (aside from bulk vertical wind shear), while GEFS-ops RF uses 18 predictors and does include severe weather diagnostics computed from pressure level data. Both algorithms use the GEFS output at 3-hourly intervals, but GEFS Reforecast MLP uses 9 times per day: 12, 15, 18, 21, 00, 03, 06, 09, and 12 UTC, while GEFS Operational MLP uses 8 times: 12, 15, 18, 21, 00, 03, 06, and 09 UTC. Second, GEFS Reforecast MLP uses a higher resolution grid of about 55-km (0.5° x 0.5°), while GEFS-ops RF uses the 81-km NCEP 211 grid. Additionally, GEFS Reforecast MLP uses predictors at the grid-point being considered, as well as all points within a 7 x 7 point box surrounding the point. GEFS Operational MLP only uses predictors at the grid-point being considered. This means that for each point, GEFS Reforecast MLP uses: 49 surrounding points x 12 fields x 9 output times = 5292 predictors, while GEFS Operational MLP uses 1 point x 18 fields x 8 output times + 1 latitude coordinate + 1 longitude coordinate = 146 predictors for Days 1-10, and 74 predictors for Days 11-15 (since those lead times only contain 6-hourly GEFS output resulting in 4 output times per day). Third, for training, GEFS Operational MLP uses the ensemble *mean* of all 31 GEFS members, while GEFS Reforecast MLP is trained on the ensemble *median* of 5 GEFS reforecast members. Fourth, GEFS Reforecast MLP performs training over 4 distinct regions of the CONUS and stitches them together for a CONUS-wide forecast, while GEFS Operational MLP trains over the entire CONUS. Finally, for forecast input, GEFS Reforecast MLP uses the *median* of the first 21 GEFS members, while GEFS Operational MLP uses the *mean* of all 31 GEFS members. Table 17 summarizes the main differences between the algorithms.

Table 16 GEFS-based predictors used in GEFS-ops RF.

GEFS Operational MLP Predictors		
(1) Bulk Shear (0-1 km AGL)*	(8) Surface-based lifting condensation level (LCL) height	(15) u-wind (10-m)
(2) Bulk Shear (0-3 km AGL)*	(9) Significant tornado parameter (STP)*	(16) v-wind (10-m)
(3) Bulk Shear (0-6 km AGL)*	(10) Mean-sea-level pressure	(17) Wind magnitude (10-m)
(4) Surface-based convective available potential energy (CAPE)	(11) Precipitable water	(18) Most unstable CAPE*
(5) Surface-based convective inhibition	(12) Specific Humidity (2-m)	(19) Latitude
(6) Storm relative helicity (0-3 km)	(13) Temperature (2-m)	(20) Longitude
(7) Lape Rate (700-500 mb)*	(14) Precipitation (3-h accumulation)	

Table 17 Summary of differences between GEFS Operational MLP and GEFS Reforecast MLP.

	GEFS Operational MLP	**GEFS Reforecast MLP**
Grid-spacing	81-km (interpolated)	55-km (0.5° x 0.5°)
Training	Ensemble mean of 31 GEFS operational members	Ensemble median of 5 GEFS reforecast members
Lead time	Days 1-15	Days 1-8
Products	Total Severe	Total Severe (Days 1-8), Hazard probs & Sig Severe (Days 1-3)
Predictors	18	12
Forecast input	Mean of all 31 GEFS members	Median of first 21 GEFS members
Regional training?	No (CONUS land points only)	yes; 4 distinct regions over the CONUS
Neighboring points used for predictors?	no	yes; 7 x 7 point surrounding box
Latitude/longitude coordinate used for predictors?	yes	no

iii. Nadocast, HREF -based ML 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 outputs. Nadocast performs extensive feature engineering: each grid point from the HREF 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.

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

WoFS-PHI is an RF-based product designed to predict severe weather hazard probabilities within a given spatial radius in 60-minute windows out to 3-4 hours of lead time. This year, a 39-km radius is used to align with the verification scales used by the SPC. Predictors are from WoFS forecast fields as well as gridded, spatiotemporally-extrapolated ProbSevere Version 2 (PS2) data (predictors are summarized in more detail in Table 18).

While the overall preprocessing approach of WoFS-PHI is similar to last year (Fig. 3), multiple important changes have been made to WoFS-PHI based on user feedback. First, this year, WoFS-PHI has been redesigned to predict 60-, rather than 30-minute, time windows to facilitate a real-time forecasting activity. During preprocessing, less smoothing has been applied to the predictors at longer lead times, which should result in less over-smoothing of the final forecast at these times. To increase probability sharpness and provide greater context into the meaning of probabilities, initial RF output probabilities are remapped to success ratios based on verification from an independent validation set during the training process. Final WoFS-PHI probabilities are the larger of the raw RF probabilities and the corresponding success ratio at each probability level. Additionally, this year two separate RF models are created for each hazard and lead time: one trained using only local storm reports (LSRs), and one using LSRs plus warnings. Training on warnings can be valuable because it enhances the spatial density of the predictand during training, which should help the RF learn more accurate and spatially-precise relationships between predictors and severe weather occurrence. However, one limitation is that not all warnings are associated with observed severe weather; thus, both the LSR-only- and LSR-and-warnings-trained RFs are provided. Another notable change from last year is the introduction of two separate types of WoFS-PHI: forecast mode (Fig. 4) and warning mode (Fig. 5). Forecast mode is available under the “ML Products” tab of the WoFS Viewer and is run once per WoFS initialization. This mode allows users to look at severe weather hazard probabilities valid up to 4 hours after the current WoFS initialization. On the other hand, warning mode is available as an optional contour overlay (accessible on the right-hand-side of the WoFS Viewer). Unlike forecast mode, warning mode updates every 5-10 minutes using the most recent PS2 and WoFS data available; however, it always displays the hazard probability for the next 60 minutes (starting from the current real time). Thus, forecast mode allows users to look at longer, fixed lead times but takes longer to update, while warning mode updates more frequently but only allows users to view a single 60-minute forecast. Finally, this year WoFS-PHI uses a new color table designed to more clearly show the differences between probability levels.

During this SFE, a real-time forecasting activity will be held to test the impact of WoFS-PHI on (human-generated) real-time severe weather probability forecasts. During this activity, participants will be divided into two groups: one with access only to the traditional suite of (non-machine-learning) WoFS products, and one with access to the traditional WoFS products plus WoFS-PHI. Both groups will make two sets of 0-1- and 1-2-h severe weather hazard probability forecasts each afternoon. On Monday through Thursday, 2-4 National Weather Service forecasters will continue the activity into the evening so that data and feedback can be collected longer into each weather event. Participant forecasts will be subjectively evaluated during next-day evaluations.

Table 18 Summary of predictors used to create WoFS-PHI spatial hazard probabilities. Predictors are taken from WoFS at the point of prediction (column 1), WoFS at multiple spatial neighborhoods (column 2), and 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		
Latitude & Longitude		
WoFS x and y grid points		
WoFS initialization time		

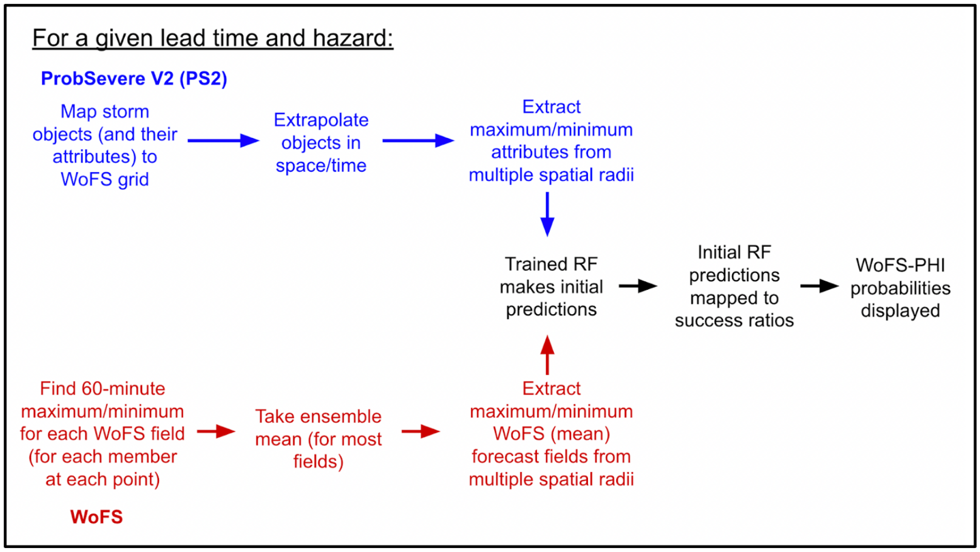

Figure 3 Flowchart showing the preprocessing steps used to create WoFS-PHI spatial hazard probabilities. Steps associated with the PS2, WoFS, and combined data, respectively, are colored in blue, red, and black.

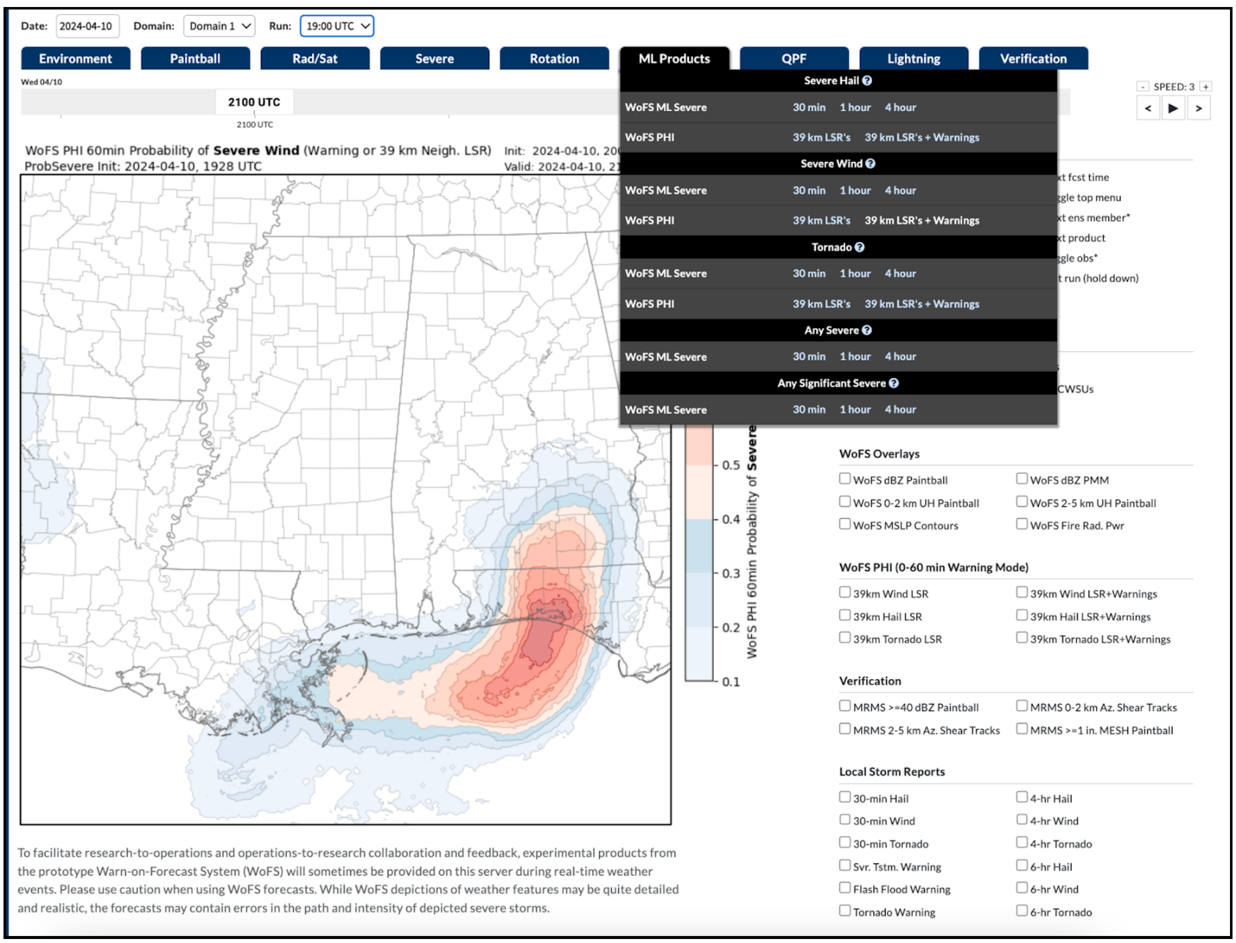

Figure 4 Example WoFS-PHI Forecast Mode severe wind probabilities, from the model trained on local storm reports plus warnings. WoFS-PHI Forecast Mode can be accessed under the "WoFS PHI" heading for each hazard under the "ML Products" tab in the WoFS Viewer. Users can select either "39 km LSR's" for the model trained only on LSRs or "39 km LSR's + Warnings" for the model trained on both LSRs and warnings.

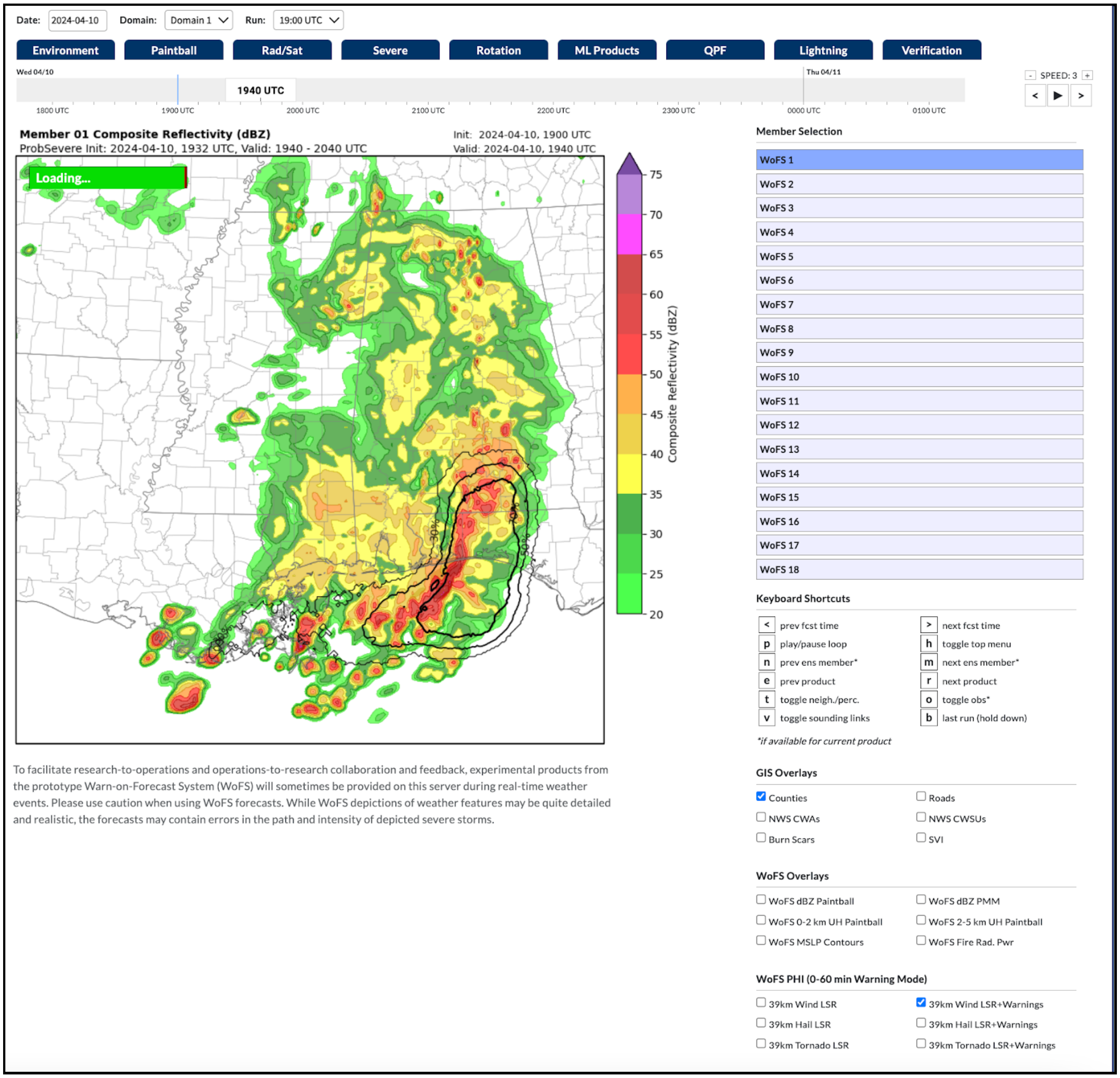

Figure 5 Example WoFS-PHI Warning Mode severe wind probabilities (contoured) overlaid on simulated reflectivity from WoFS member 1 (filled). Here, the WoFS-PHI probabilities are from the model trained on local storm reports plus warnings. WoFS-PHI warning mode contours are toggle-able based on the checkboxes in the lower right-hand corner of the WoFS Viewer.

v. 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/GEFS calibrated severe and Nadocast) that contain hourly 4-h timing information to distribute probabilities from the SPC outlooks across the 16 to 12 UTC time frame. Thus, they are a blend of the human forecast and various probabilistic calibrated guidance products.

vi. NSF NCAR ML-derived extended-range FV3-based convective hazard probabilities

For the 2024 HWT SFE, gridded machine learning-based probabilistic convective hazard guidance is being generated using neural networks (NNs) and the 10-member real-time FV3-based (C-SHiELD) ensemble forecasts generated at NSF NCAR (i.e., NCAR FV3; Table 10). The ML forecasts will extend to 8.5 days lead-time and were constructed to be similar to HRRR and WRF-based ML guidance evaluated in prior HWT SFEs. More specifically, the NNs were trained (Table 19) using a training dataset of 91 deterministic 00 UTC-initialized FV3-based (C-SHiELD) forecasts from 1 April 2022 – 30 June 2022. A set of 28 diagnostics (Table 20) from the training data were upscaled onto an 80-km grid while an additional 22 "neighborhood" predictors were constructed by taking the 5 environmental and 6 explicit predictors and computing means (for the environmental predictors) and maxima (for the explicit predictors) in space and time over 3x3x3 and 5x5x5 arrays of 80-km grid boxes.

Each grid box was labeled as a "hit" if a severe weather report occurred within a 4 h and 40-km or 120-km window of the grid box center point. The NNs were designed to output six independent probabilities: probability of hail, wind, tornado, significant hail, significant wind, or any storm report. The same NNs were applied to each ensemble member, with the 10-member average probability used for evaluation.

Table 19 Settings used to construct and train the NNs. Ten NNs with different initial weights were trained separately with their output probabilities averaged together.

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	10
Hidden layer activation function	Rectified Linear Unit
Output layer activation function	Sigmoid
Optimizer	Adam
Loss function	Binary Cross-entropy
Batch size	1024
Regularization	None
Batch normalization	On

Table 20 The 33 base predictors used to train the NNs (28 diagnostics and 5 static fields). 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 also constructed by taking larger spatial and temporal means and maximums of the environmental and explicit fields as described in the text.

Base Predictor	Type
Forecast Hour	Static
Local Solar Hour	Static
Latitude	Static
Longitude	Static
Day of Year	Static
Surface-based CAPE	Environment
Surface-based CIN	Environment
2-m Temperature	Environment
2-m Dewpoint Temperature	Environment
Mean Sea Level Pressure	Environment
Hourly-maximum 2–5km UH	Explicit
Hourly-minimum 2–5km UH (negative)	Explicit
Hourly-maximum updraft speed below 400 hPa	Explicit
Hourly-minimum downdraft speed below 400 hPa	Explicit
Hourly-maximum 10-m wind speed	Explicit
Hourly-maximum composite reflectivity	Explicit
700 hPa – 500 hPa lapse rate	Upper-air
925 hPa, 850 hPa, 700 hPa, and 500 hPa zonal wind	Upper-air
925 hPa, 850 hPa, 700 hPa, and 500 hPa meridional wind	Upper-air
925 hPa, 850 hPa, 700 hPa, and 500 hPa height	Upper-air
925 hPa, 850 hPa, 700 hPa, and 500 hPa temperature	Upper-air

vii. WoFS-based ML Hazard Guidance

The gridded machine-learning (ML) based guidance is a new WoFS-based product designed to provide advance notice for the occurrence of severe weather at lead times of 2-6 hours. For each six-hour forecast launched by the WoFS, ML guidance is generated for severe wind, severe hail, and tornadoes (e.g., Fig. 6). The guidance is valid over the full four-hour window (hours 2-6) and indicates the probability of the respective hazard occurring within 36 km of a grid point (similar to the SPC definition).

Separate histogram-based gradient boosting trees (HGBT) models were trained to predict the individual hazards, as well as whether any of the three hazards are possible ("any-severe"), using data exclusively from WoFS forecasts. These data include information from the WoFS ensemble about various storm and environmental fields at multiple spatial scales. During the training process, the W2W guidance was verified using NOAA storm data reports. For each hazard, the WoFS ML guidance is shown to be more skillful than non-ML baselines with skill increasing substantially after convective initiation has occurred and been assimilated into the WoFS.

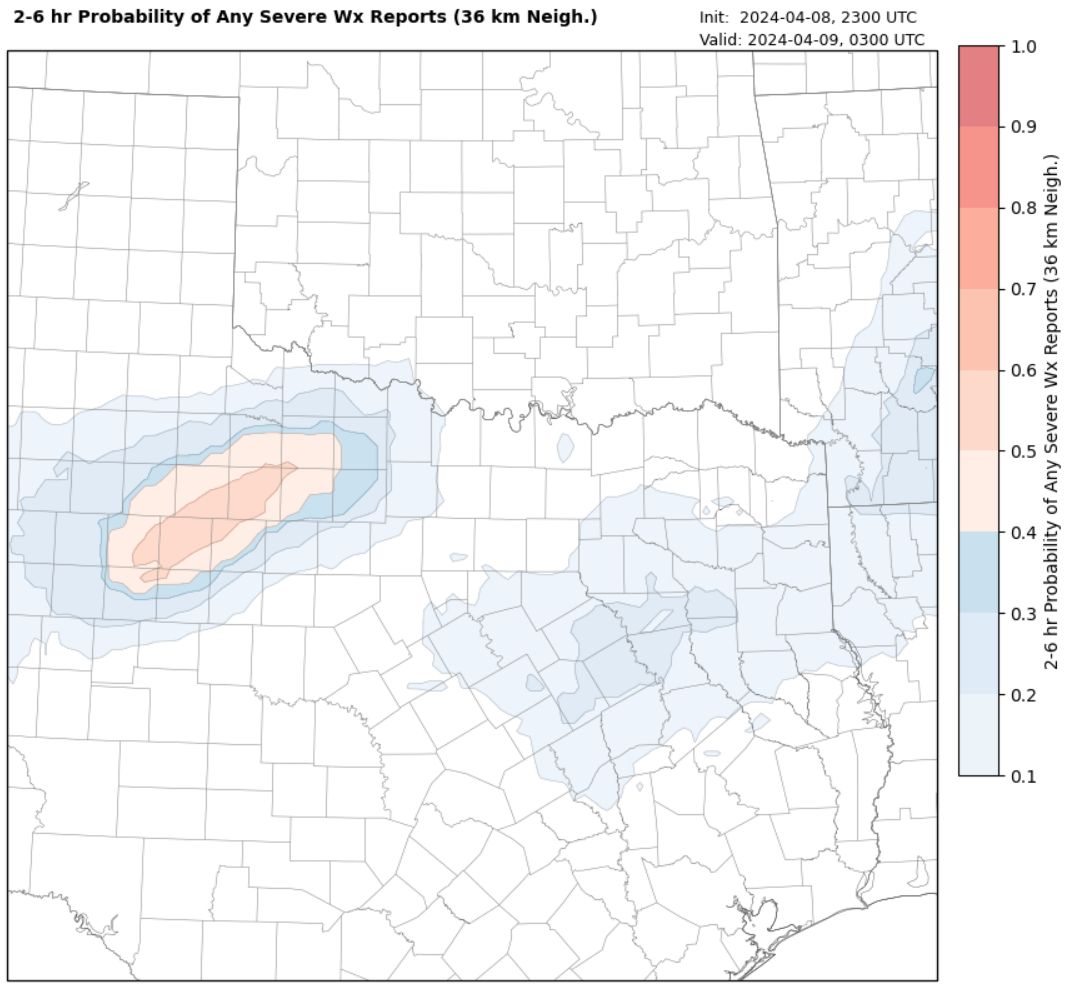

Figure 6 Example forecast of any severe weather report within 36-km of a point over a 4 h time period.

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.

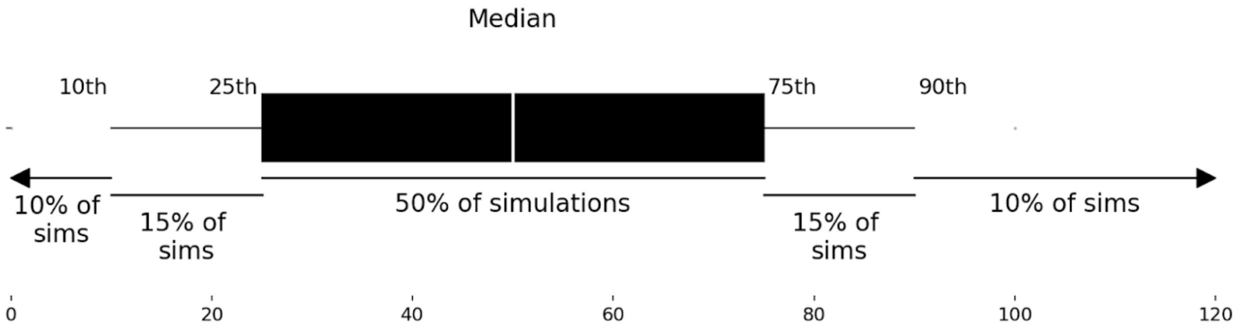

Figure 7 An example of tornado count and impact output from the IMPACTS system. Boxes represent the 25th-75th percentiles, while whiskers represent the 10th-25th and 75th-90th percentiles. The values of the 25th, 50th (median), and 75th percentiles are annotated on the individual box-and-whisker charts.

3. SFE 2024 Core Interests and Daily Activities

2024 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. Additionally, there will be an evening activity for which 2-4 NWS forecasters will issue 1-h experimental severe weather outlooks using WoFS-PHI from 12-8pm on Mondays and 2-8pm Tuesday-Thursday. The evening activity will not operate on Fridays. 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 21 and 22 provide a daily schedule for Monday, and Tuesday-Friday, respectively. Tables 23 and 24 provide the schedule for the evening forecasting activity. Further details are provided in subsequent sections.

Table 21 Schedule for Monday morning and afternoon.

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, Program CACs
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 PHI); *Virtual* (WoFS PHI)
3:15 PM – 4:00 PM	**Individual Forecasting Activity Continued** (MD & Day 1 Updates) *In-Person R2O* (Day 1 Update); *In-Person Innovation* (WoFS PHI); *Virtual* (WoFS PHI)

Table 22 Schedule for Tuesday – Friday morning and afternoon.

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 PHI); *Virtual* (WoFS PHI)
3:15 PM – 4:00 PM	**Individual Forecasting Activity Continued** (MD & Day 1 Updates) *In-Person R2O* (Day 1 Update); *In-Person Innovation* (WoFS PHI); *Virtual* (WoFS PHI)

Table 23 Schedule for Monday evening activity.

Time (CDT)	
12:00 PM – 2:00 PM	**WoFS-PHI Introduction and Training** *Evening Forecasters*
2:00 PM – 2:15 PM	**Update on Today's Weather** *Hybrid All* & *Evening Forecasters* (David Imy)
2:15 PM – 3:15 PM	**Individual Forecasting Activity** (WoFS PHI) *In-Person Innovation* (WoFS PHI); *Virtual* & *Evening Forecasters* (WoFS PHI)
3:15 PM – 4:00 PM	**Individual Forecasting Activity** (WoFS PHI) *In-Person Innovation* (WoFS PHI); *Virtual* & *Evening Forecasters* (WoFS PHI)
4:00 PM – 8:00 PM	**Individual Forecasting Activity** Evening Forecasters (WoFS PHI)

Table 24 Schedule for Tuesday – Thursday evening activity.

Time (CDT)	
2:00 PM – 2:15 PM	**Update on Today's Weather** *Hybrid All* & *Evening Forecasters* (David Imy)
2:15 PM – 3:15 PM	**Individual Forecasting Activity** (WoFS PHI) *In-Person Innovation* (WoFS PHI); *Virtual* & *Evening Forecasters* (WoFS PHI)
3:15 PM – 4:00 PM	**Individual Forecasting Activity** (WoFS PHI) *In-Person Innovation* (WoFS PHI); *Virtual* & *Evening Forecasters* (WoFS PHI)
4:00 PM – 8:00 PM	**Individual Forecasting Activity** Evening Forecasters (WoFS PHI)

a. Formal Evaluation Activities

SFE 2024 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, 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", "(C)alibrated Guidance", "(O)utlooks", and "(A)rtificial (I)ntelligence". 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 1 & 2 4-h SPC Tornado Timing Guidance

Timing guidance (TG) products are generated from (1) HREF/GEFS and (2) Nadocast calibrated tornado probabilities to produce tornado guidance valid in hourly 4-h time windows 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 16 to 12 UTC time frame on both Day 1 and Day 2. Both HREF/GEFS TG and Nadocast TG products are subjectively rated for Day 1 and Day 2.

C2. Day 1 & 2 4-h SPC Hail Timing Guidance

The same methods and products as in C1 are rated for Hail timing guidance.

C3. Day 1 & 2 4-h SPC Wind Timing Guidance

The same methods and products as in C1 are rated for Wind timing guidance.

C4. 4-h ML Hazard Guidance

This evaluation compares new gridded WoFS-based ML guidance for individual hazards to the HREF-based Nadocast guidance. Specifically, this evaluation will examine severe weather probabilities for two 4-h periods: 20-00Z (comparing 12Z HREF-based Nadocast and 18Z WoFS-based ML) and 00-04Z (comparing 12Z HREF-based Nadocast and 22Z WoFS-based ML).

C5. Medium Range 00Z Total Severe

Three different sets of extended range total severe probabilities for Day 3-7 lead times are subjectively rated. These methods include: (1) GEFS Reforecast MLP, (2) GEFS Operational MLP, and (3) NCAR CAM MLP.

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?

O4. Probabilistic 0-1 h Outlooks

Using WoFS-PHI and any other guidance, 0-1 h outlooks for tornado, wind, and hail, are produced by Forecaster #1, Forecaster #2, and a consensus of non-forecaster participants. These outlooks are subjectively rated and compared to another set of the same outlooks issued by participants without access to WoFS-PHI.

O5. Probabilistic 1-2 h Outlooks

This evaluation is the same as O4, but for 1-2 h lead times.

Primary Science Question(s): How does access to WoFS-PHI impact the quality of severe weather guidance issued by forecasters in the 0-2 h lead time range?

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 EMC REFS and HREF for Day 1 lead times. These comparisons will serve to unearth ways in which the currently operational CAM ensemble (i.e., HREF) differs from the candidate to replace it (i.e., EMC REFS), and whether the EMC REFS 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 EMC REFS compare to those of the HREF (e.g., spread and skill)? Are there systematic shortcomings or advantages of EMC REFS?

E2. CLUE: 12Z Day 1 Ensemble Flagships

The four different versions of REFS will be examined at Day 1 lead times to assess whether they can meet or exceed the skill of the operational baseline of HREF. Subjective ratings will be assigned to the following ensembles based at 12Z: (1) EMC REFS, (2) SPC REFS, (3) SPEC REFS ARW, (4) MPAS REFS, and (5) HREF.

E3. CLUE: Day 2 Ensemble Flagships

This evaluation is the same as E2, except Day 2 lead times are assessed.

Primary Science Question(s): What are the optimal configuration strategies for REFS? Does including a larger proportion of control members result in a more accurate and reliable ensemble? What is the impact of replacing FV3 members with MPAS members in REFS?

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 HT*), EMC/GSL (*RRFS*), and NASA (*NASA FV3*) – based on their skill and utility for severe weather forecasting at Day 1 lead times. 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: 12Z Day 2 Deterministic Flagships

Four deterministic, 12Z initialized, CAM configurations are subjectively evaluated for Day 2 lead times. These configurations include: (1) RRFS, (2) NSSL MPAS HT, (3) NASA FV3, and (4) HRRR.

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 Resolution Sensitivity

Similarly configured 2-, 3-, and 4-km grid-spacing versions of MPAS will be evaluated to assess MPAS resolution sensitivities. These configurations are: MPAS2, NSSL-MPAS-HN, and MPAS4, respectively.

Primary Science Question(s): How are the impacts of a slight increase and decrease in grid-spacing relative to the standard 3-km grid-spacing in MPAS? Does using 4-km or 2-km grid-spacing noticeable improve or degrade forecast quality relative to 3-km?

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 (3D-RTMA HRRR) will be compared to another version using RRFS as the background (3D-RTMA RRFS) to assess the role that the background first-guess plays on the final analysis. Comparisons will also be made to a 3-km grid-spacing version of the sfcOA that uses HRRR forecasts as the background (sfcOA HRRR) and applies a simple 2-pass Barnes objective analysis to incorporate the latest surface observations. 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. Upscaled Mesoscale Analysis Background

3D-RTMA HRRR, 3D-RTMA RRFS, and sfcOA HRRR will be upscaled to a 40-km grid and compared to the 40-km grid-spacing sfcOA that uses the RAP as the background.

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

A3. 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?

(A)rtificial (I)ntelligence Evaluations

AI1. First-Guess Watch Guidance

The first-guess watch guidance is a suite of machine learning models developed by CIWRO/SPC to predict where and when conditions will be favorable for a watch. This guidance is derived from multiple gradient-boosted classifiers trained on HREF environmental fields and storm-derived parameters, and the guidance is designed to provide at least 3 hours of lead time prior to the first occurrence of severe weather. Estimated watch counties are inferred at each 12z HREF forecast hour from the ML probabilistic output and masked such that a county must fall within at least a 13z D1 Marginal risk to qualify for a watch. New to this year, the watch guidance also estimates whether a tornado or severe thunderstorm watch would be more appropriate for the forecast conditions.

Participants will subjectively evaluate the first-guess watch guidance by comparing the placement, timing, and type of recommended watch to the true severe thunderstorm and tornado watches issued by the NWS. These evaluations may also consider any tornado or severe thunderstorm warnings that were issued during the evaluation period, as well as any local storm reports that were received to better assess the skill of the guidance.

Primary Science Question(s): Can ML be used to produce a skillful first-guess watch product that provides value to forecasters during and prior to severe weather events? How well do the first-guess watches capture the location and timing of severe weather warnings and reports? How well does the recommended watch type align with what is issued by the NWS? Is the recommended watch type appropriate for the observed hazards?

AI2. Global NWP Emulators

GraphCast, Pangu Weather, and FourCastNet AI-driven global weather predictions starting with GFS initial conditions will be assessed, compared, compared and subjectively rated alongside the GFS. The target lead time will be 7 days (i.e. forecast hours 156-180) and all 5 times (12, 18, 00, 06, and 12Z) that fall within the convective day will be considered. Participants will primarily consider the 500-mb height-wind patterns, but will use other available fields (e.g., 850 mb heights/winds, 2-m temperatures, etc.) to supplement their ratings. Participants will further assess the vertical coherency of the models by reviewing derived sounding profiles at select locations across the CONUS. Finally, the 7-day QPFs in the NWP emulators that have QPF available will also be subjectively rated alongside the GFS.

Primary Science Question(s): How do forecasts from NWP Emulators compare to traditional NWP forecasts? Is there value in the NWP emulators for extended range severe weather forecasting applications?

b. Forecast Products and Activities

There will be two periods of experimental forecast activities during SFE 2024. 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 2019-2023. 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 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, participants will split into two groups and generate experimental 0-1 and 1-2 h hazard probabilities for tornado, wind, and hail. The first set of forecasts is due at 3pm CDT and covers 3-4pm (0-1 h) and 4-5pm (1-2 h). The second set of forecasts is due at 4pm and covers 4-5pm (0-1 h) and 5-6pm (1-2 h). One group issuing these forecasts will have access to WoFS PHI and any other guidance. The other group will not have access to WoFS PHI. In both groups, two expert forecasters will be assigned whose forecasts will be evaluated the next day. All other participant forecasts will be combined into consensus forecasts, which will also be evaluated the next day. A small group of pre-determined “evening forecasters” will continue this activity into the evening. They will take a break from 4-5pm, then from 5-6pm issue 0-1 h and 1-2 h forecasts valid 6-7pm and 7-8pm, respectively. They will repeat the activity one more time from 6-7pm with all forecasts shifted one hour later, and finally from 7-8pm finish by completing a survey on their use of WoFS and WoFS PHI.

These WoF activities are the eighth year WoFS 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 2024 participants. Green denotes participants that are observing and not directly participating in activities.

Week 1	Week 2	Week 3	Week 4	Week 5
29 April - May 3	**6-10 May**	**13-17 May**	**20-24 May**	**28-31 May**
Monte Flora (CIWRO/NSSL)	Tom Galarneau (NSSL)	Harald Richter (BoM)	Harald Richter (BoM)	Cameron Homeyer (OU)
Youngsun Jung (NWS/OSTI)	Luke LeBel (PSU)	Bill Gallus (ISU)	Tony Lyza (NSSL)	Tyler Janoski (NSSL)
Tom Hultquist (NWS/AFS)	Derek Stratman (CIWRO/NSSL)	Sam Ritter (ISU)	Isabelle Jernigan (OU/CIWRO/NSSL)	Craig Schwartz (NCAR)
Vanderlai Vargas (DTC/GSL)	Corey Potvin (NSSL)	Aaron Hill (OU; M-Th)	Steven Cavallo (OU)	Will Mayfield (NCAR)
Miranda Silcott (CIWRO/NSSL)	Maria Madsen (OU/AI2ES)	Steve Willington (UK Met)	Gretchen Mullendore (NCAR; M-Th)	Jeff Beck (DTC)
Phil Pegion (PSL)	Nick Esposito (EMC)	Mike Silverstone (UK Met)	Steve Willington (UK Met)	Brice Coffer (NCSU)
Matt Morris (EMC)	Junjun Hu (GSL)	Marcel Caron (EMC)	Mike Silverstone (UK Met)	Alex Anderson-Frey (UW)
Christopher Melick (16WS)	Michael Hosek (OU)	Aaron Treadway (NWS/AFS)	Ryan Sobash (NCAR)	Miles Epstein (UW)
Michael Lee (NWS PHI)	Mike Kavulich (DTC)	Morgan Schneider (OU)	Ayesha Ghani (OU)	Henry Hua (UW)
Greg DeVoir (NWS CTP)	Ryan Barnes (SPC)	Clark Evans (UWM)	Ben Blake (EMC)	Eddie Wolff (U. Illinois)
Mateusz Taszarek (AMU; Poland)	Derek Hodges (NWS TSA)	Lee Hawkness-Smith (UK Met)	Steve Lack (16WS)	Aaron Hardin (NWS TWC)
Krzysztof Piasecki (AMU; Poland)	Ivan Gumbs (NWS FSD)	Seb Cole (UK Met)	Milind Sharma (TAMU)	Russell Danielson (NWS BOU)
	Michael Lowe (NWS WDTD)	Jeff Duda (GSL)	Devin Bissell (TAMU)	
	Bill Putman (NASA)	Geoffrey Heidelberger (NWS HUN)	Stan Czyzyk (NWS VEF)	
	Scott Rabenhorst (NASA)	Connor Belak (NWS LWX)	Kristian Oliver (NWS TAE)	
		Mike Jurewicz (NWS CTP)	Lexy Elizalde (NWS WDTD)	
Daniel Liota (ECCC; T)		Heather Pimiskern (ECCC; T)		
Mariane Peltier-Champigny (ECCC; T)		Rose Carlsen (ECCC; T)		
***** VIRTUAL PARTICIPATION *****				
Bob Rozumalski (NWS/FDTD)	Aurore Porson (UK Met)	Eric Aligo (EMC)	Jason Jordan (NWS/FDTD)	Jili Dong (EMC)
Marion Mittermaier (UK Met; a.m. only)	Michelle Harrold (DTC)	Carlo Cafaro (UK Met)	Gang Zhao (EMC)	Ed Colon (EMC)
Rob Hepper (NWS AWC)	Shun Liu (EMC)	Craig Hartsough (GSL)	Terri Adams (Howard)	Ruifang Li (GSL)
Donnie Lippi (EMC)	Gerard Ketefian (DTC/GSL)	Cansu Duzgun (FSU)	Dirk Petersen (WFO OAX; T-Th)	David Dowell (GSL)
Terra Ladwig (GSL)	David Ahijevych (NCAR)	David Zaff (NWS BUF)	Curtis Alexander (GSL)	Dan Baumgardt (NWS ARX)
Robin Tamamachi (Purdue)	Matt Pyle (EMC)		Ed Szoke (GSL)	
Jacob Bruss (Purdue)	Cass Shivers-Williams (OAR/WPO)		Burkely Gallo (16WS)	
Claiborne Wooton (M-W; UND)	Jana Houser (Ohio State)		John Boris (NWS APX)	
Nicholas Camp (UND)	Jake Mulholland (UND)		Adam Picard (NWS KEY)	
Connor Michael (OSU)	Ethan Weisberger (UND)			
Chandra Kondragunta (OAR/WPO)	Cole Hood (UND)			
Brian Tentinger (NWS BGM)	Levi Newel (UND)			
Mike Natoli (NWS CYS)	Mark Jarvis (NWS LMK)			
	Martin Mayeaux (NWS SHV)			
***** EVENING FORECASTING ACTIVITY *****				
Pierce Larkin (NWS CAE)	Tim Brice (NWS EPZ)	Brittany Whitlam (WFO REV)	Jeremy Buckles (NWS MRX)	Valerie Thaler (WFO OTX)
Julie Arthur (NWS BYZ)	Tony Hurt (NWS TBW)	Roman Miller (WFO AKQ)	Eswar Iyer (NWS AKQ)	Ryan Fucheck (WFO MHX)
	Zahaira Velez-Serrano (NWS EKA)	Matt Strauser (WFO CAR)		Matt Flanagan (WFO TOP)
				Colton Milcarek (WFO PBZ)
				Logan Poole (WFO JAN)

SFE Facilitators: Adam Clark (NSSL), Israel Jirak (SPC), Dave Imy (retired SPC), Tim Supinie (SPC), Sean Ernst (CIWRO/SPC), 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), 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

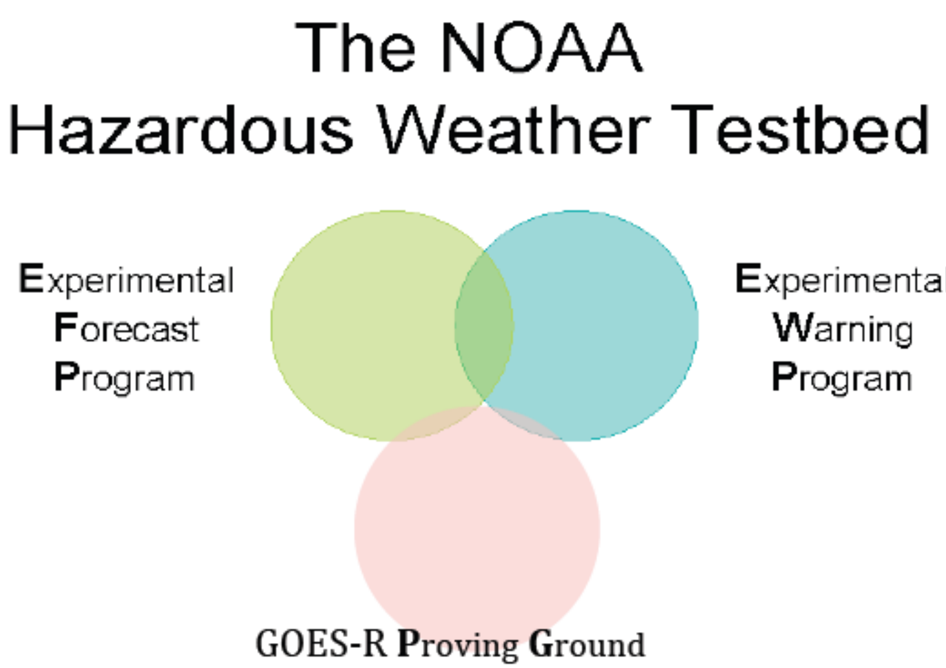

Figure B1: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

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 2024 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

Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2022: Pangu-Weather: A 3D High-resolution Model for Fast and Accurate Global Weather Forecast. https://doi.org/10.48550/arXiv.2211.02556.

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 Prediction 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

ECMWF, 2020: IFS Documentation CY47R1–Part V: Ensemble prediction system. ECMWF IFS Doc. 5, 23 pp., https://doi.org/10.21957/d7e3hrb.

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

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.

Hersbach, H., and Coauthors, 2020: The ERA5 Global Reanalysis. *QJRMS*, **146**, 1999-2049. https://doi.org/10.1002/qj.3803.

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.

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.

Lam, R., and Coauthors, 2022: GraphCast: Learning skillful medium-range global weather forecasting. https://doi.org/10.48550/arXiv.2212.12794.

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

Pathak, J., and Coauthors, 2022: FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. https://doi.org/10.48550/arXiv.2202.11214.

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.

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

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.