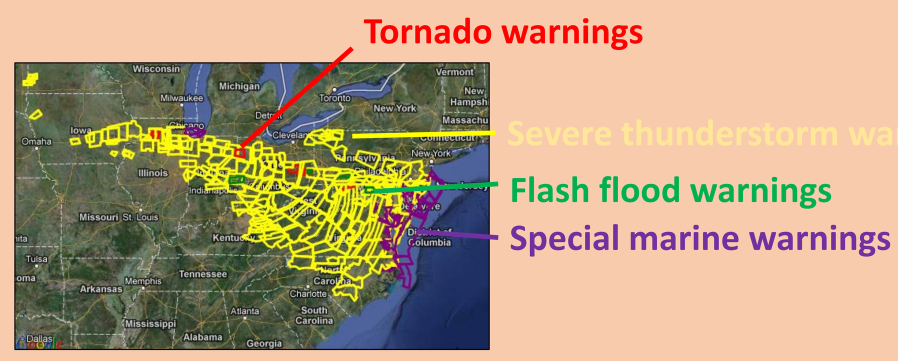
The 2012 North American Derecho: An evaluation of regional and global climate modeling systems at cloud-resolving scales

Weiran Liu[1] (wraliu@ucdavis.edu), Paul Ullrich[1], Jianfeng Li[2], Colin Zarzycki[3], Peter Caldwell[4], Ruby Leung[2], Yun Qian[2]

1. Department of Land, Air, and Water Resources, University of California-Davis, Davis, CA, USA
2. Pacific Northwest National Laboratory, Richland, WA, USA
3. Department of Meteorology and Atmospheric Science, Pennsylvania State University, University Park, PA, USA
4. Lawrence Livermore National Lab, Livermore, CA, USA

Introduction

- Global climate model evaluation has historically focused on climatological means or long-term averages. However, this doesn't necessarily convey information about how well the model capture the **most extreme and high-impact weather phenomena**, such as extreme mesoscale convective systems (MCSs).
- This suggests a need for **standardized and targeted evaluation** of specific weather features in one model or as part of an intercomparison across different models.
- We propose one such **extreme event testbed** for evaluating climate modeling systems that operate at cloud resolving scales. The testbed focuses on historical simulation of a single event, in this case the June 2012 North American derecho and accompanying MCS, with the intention of **providing a standard and comprehensive suite of metrics for model assessment and intercomparison.**

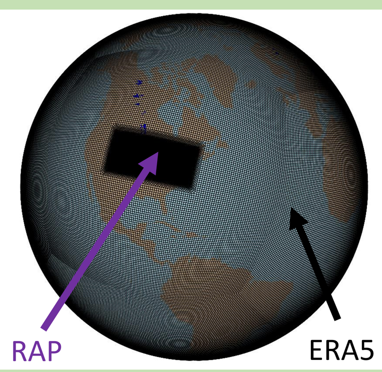


Warnings from the National Weather Service on June 29-30, 2012.

The June 2012 North American Derecho

- **Derechos**: Strong, quasi-linear wind-producing MCSs with a "bow echo" structure of intense convective features associated with **severe wind gusts** (> 25.7 m/s) and precipitation.
- One of the deadliest and fastest-moving derechos in US history.
- Caused widespread severe wind damage across the Northeastern US. Damage total of $2.9 billion USD.
- Began as a small storm cluster on 1400 UTC on 29 June 2012. Emerged into the Atlantic Ocean shortly before 0600 UTC on 30 June 2012.
- It was underforecasted days and hours ahead of time, as well as throughout much of the duration of the storm.

Models and Simulations



Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM) (Caldwell et al., 2021; Liu et al., 2022) **& WRF** v4.3.3 (Skamarock et al., 2019)

- Regionally Refined Mesh (**RRM**) generated by SQuadGen
- Global initial conditions (ICs) from ECMWF Reanalysis v5 (ERA5)
- North American ICs from NOAA Rapid Refresh (RAP)
- Sensitivity tests: RRM grid spacing (6.5 - 1.625 km), differences between hydrostatic and nonhydrostatic dynamical cores, low-resolution (72 vertical levels; deep convection scheme on) and high-resolution (128 vertical levels; deep convection scheme off) model configurations, initialization time, and data source for the ICs (ERA5, RAP, and ERAI).

SCREAM

Simulation Abbreviation	Fine Resolution	IC	Initialization Time (UTC)	LR/HR Configuration	Dynamical Core
SCREAM_6.5km	ne512 (6.5km)	ERA5+RAP	12:00 29 June 2012	HR	NH
SCREAM_3.25km	ne1024 (3.25km)	ERA5+RAP	12:00 29 June 2012	HR	NH
SCREAM_1.625km	ne2048 (1.625km)	ERA5+RAP	12:00 29 June 2012	HR	NH
SCREAM_ERA5	ne1024 (3.25km)	ERA5	12:00 29 June 2012	HR	NH
SCREAM_ERAI	ne1024 (3.25km)	ERAI	12:00 29 June 2012	HR	NH
SCREAM_06Z	ne1024 (3.25km)	ERA5+RAP	06:00 29 June 2012	HR	NH
SCREAM_LR	ne1024 (3.25km)	ERA5+RAP	12:00 29 June 2012	LR	NH
SCREAM_H	ne1024 (3.25km)	ERA5+RAP	12:00 29 June 2012	HR	H

WRF

Simulation Abbreviation	IC	Number of Vertical Levels	Microphysical Scheme
WRF_RAP	ERA5+RAP	45	Thompson
WRF_NARR	NARR	45	Thompson
WRF_HR	ERA5+RAP	72	Thompson
WRF_HR_P3	ERA5+RAP	72	P3

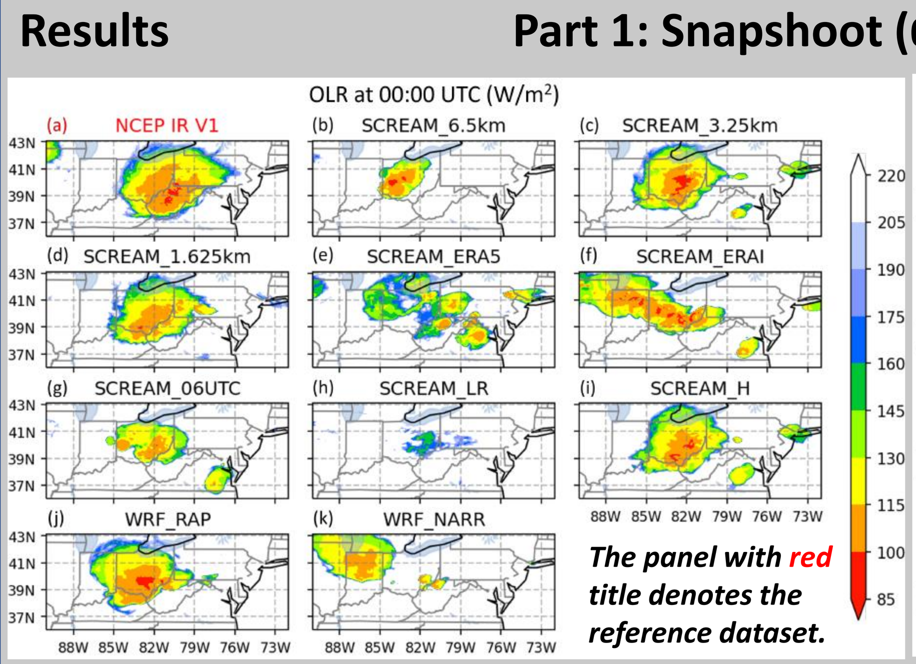
All above WRF simulations are 4 km.
Note: More WRF simulations (e.g., 1.6 km and 3.2 km resolutions) are conducted and will be included in the final paper (now shown).

Reference Datasets

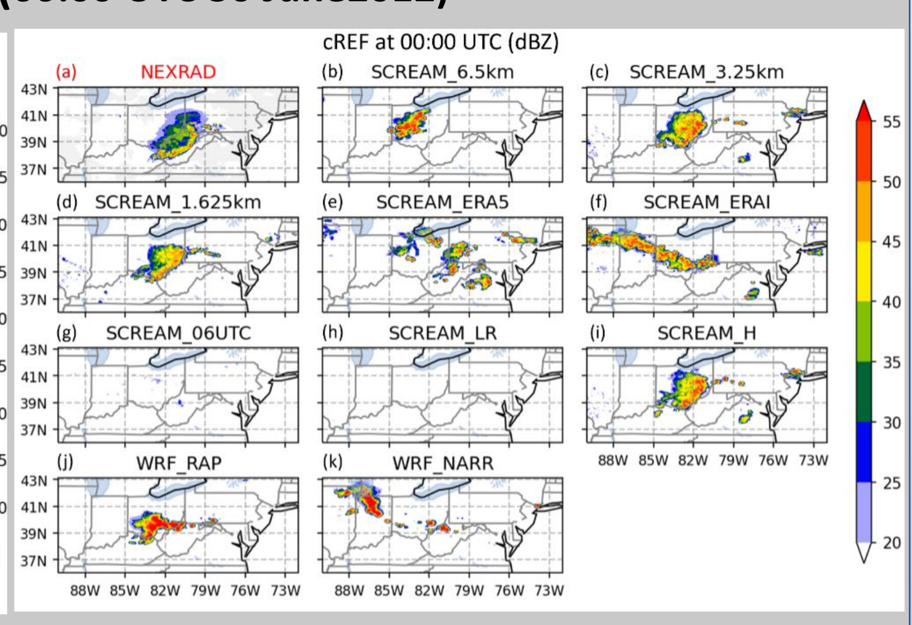
- **OLR:** NCEP IR
- **Composite Radar Reflectivity (cREF):** NEXRAD
- **Wind:** ASOS station (sustained and gust)
- **Precipitation:** 1. NCEP Stage IV (NEXRAD) 2. IMERG 2. CMORPH 4. ASOS station

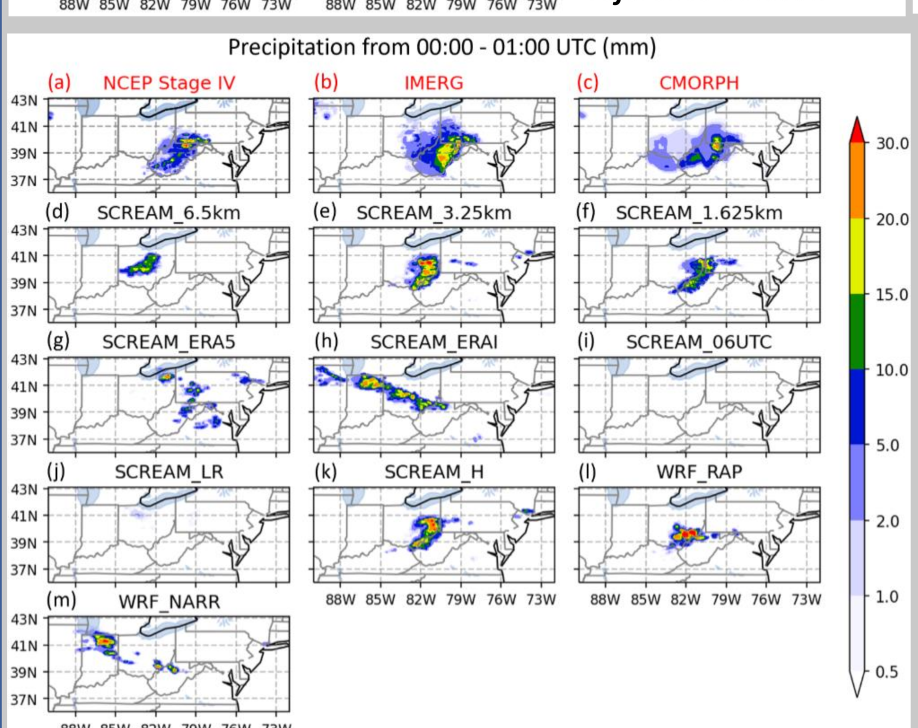
Results

Part 1: Snapshoot (00:00 UTC 30 June2012)

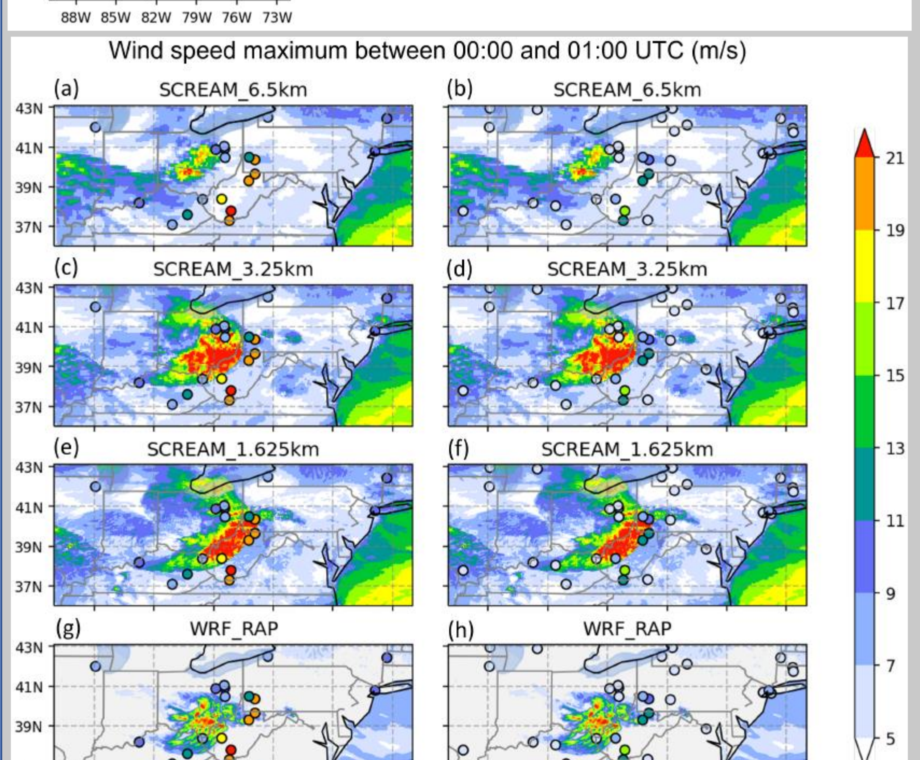


The panel with red title denotes the reference dataset.





- Higher **horizontal resolution** (in RRM grid) is associated with a better performance.
- Highly sensitive to **the IC source dataset and the initialization time**. RAP greatly improves the performance compared to ERA5 and ERAI.
- Both SCREAM and WRF **overestimate** the precipitation in NCEP Stage IV while the detailed structure of the derecho (e.g., the bow-shape echo) is better captured in SCREAM than WRF.

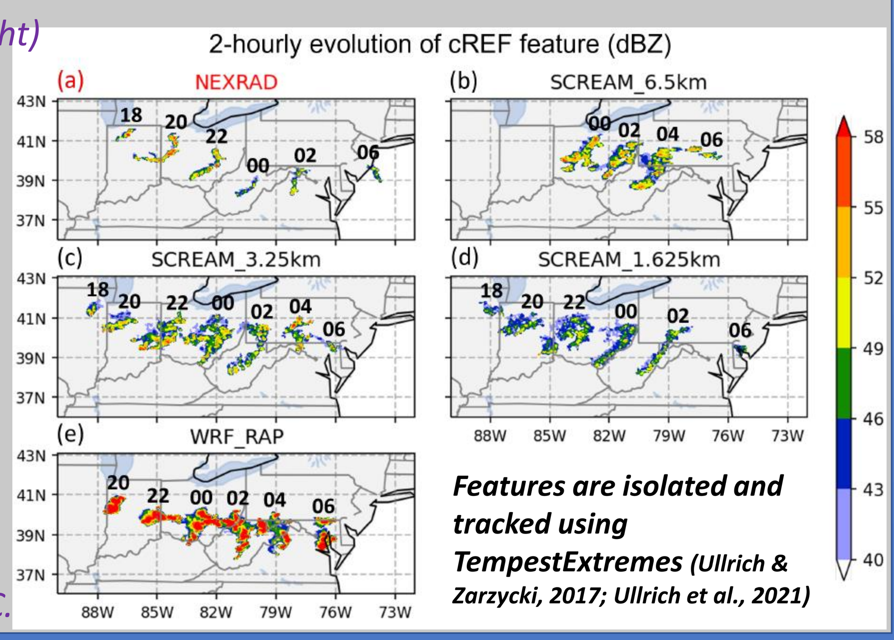


Dots: ASOS gust wind speed (left) and ASOS wind speed (right)

- This study places **additional emphasis on the 10-m wind speed evaluation**, which has not been thoroughly covered in previous studies.
- WRF simulation shows **lower 10-m wind speeds** than SCREAM. Further investigation (now shown) indicates that this underestimation in WRF could be partly reduced by finer vertical resolution (4.12%) or changing the analyzed output from the 15-minute instantaneous model result to the maximum during each 15-minute interval (9.83%).

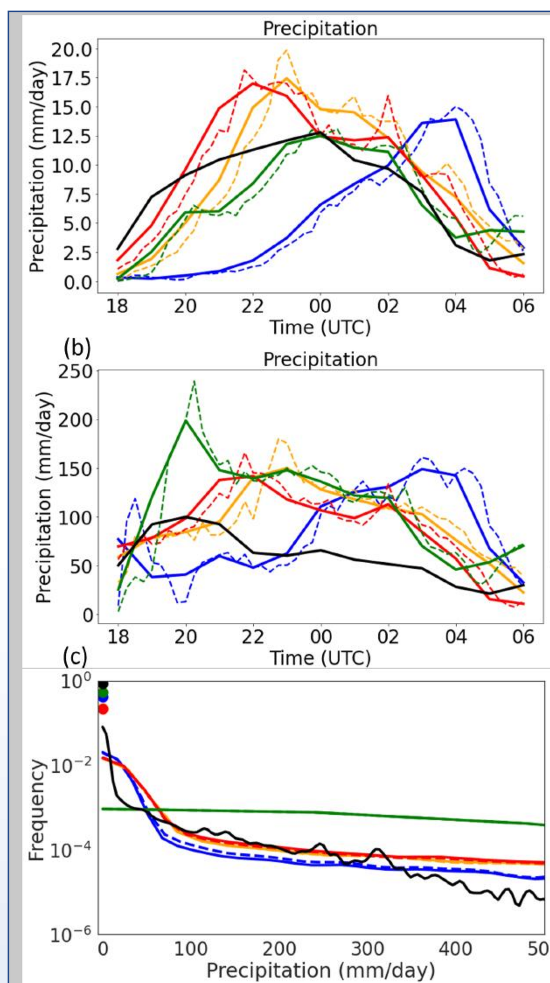
Part 2: Time Evolution

- The location of the derecho shows roughly **2-hour delay** in SCREAM_3.25km and WRF_RAP, and larger (~ 3-hour) delay in SCREAM_6.5km. Comparing to SCREAM_3.25km, SCREAM at 1.625 km resolution reduces the delay by ~ 0.5-1 hour.



Features are isolated and tracked using TempestExtremes (Ullrich & Zarzycki, 2017; Ullrich et al., 2021)

The black bold numbers mark the hours in UTC.



(a) Time series of regional-averaged precipitation rate (mm/day) in the analysis domain (76°-88°W, 36.5°-42°N). The dashed lines represent the simulation results at 15-minute frequency and the solid lines represent the hourly frequency.

(b) Same (a) but averaged only over precipitating grid points with rainfall rate > 1 mm/day.

(c) Frequency distribution of hourly precipitation rates of all grid points within the analysis domain in the period of 18:00 UTC 29 June to 06:00 UTC 30 June 2012 in the solid lines. The dashed lines are the same as the solid lines but apply a 2-hour forward shift. The dots on the y axis denote the frequencies of zero precipitation

- The observed regional-averaged precipitation peak time is captured by WRF accurately, but the averaged precipitation magnitude over the precipitating grid points is approximately twice as great as the observed (Figure b). WRF has **wet bias** over the precipitating grids, but the precipitating **area** (not shown) is reduced resulting in the similar magnitudes of precipitation peaks in the regional means (Figure a).
- While the maximum of averaged precipitation over the precipitating grid points is similar at three SCREAM resolutions, the time delay in the peak time is greatest in SCREAM_6.5km and declines at higher resolutions.

Part 3. Metrics

root-mean-square of the error (RMSE)

$$RMSE=\left[\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)^2\right]^{1/2}$$

mean error (ME)

$$ME=\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)$$

mean absolute error (MAE)

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|p_i-o_i|$$

bias score (BS)

$$BS=\frac{P}{O}$$

threat score (TS)

$$TS=\frac{H}{(P+O-H)}$$

pi: simulation
oi: reference

P: number of grid points > threshold in the simulations
O: number of grid points > threshold in the reference
H: number of grid points > threshold both simulated and observed

12-hour (18-06Z) averaged OLR

****numbers in parentheses are two-hour shifted***

Simulation Name	SCREAM_6.5km	SCREAM_3.25km	SCREAM_1.625km	SCREAM_H	WRF_RAP
RMSE	47.77 (39.35)	29.13 (26.43)	27.21 (25.29)	29.39 (27.69)	20.48 (17.65)
MAE	43.81 (34.61)	26.18 (22.61)	23.98 (21.46)	26.16 (23.89)	16.70 (14.15)
ME	43.81 (34.49)	22.31 (17.85)	18.86 (15.95)	21.66 (18.09)	8.18 (-0.11)
Pearson Correlation	0.89 (0.88)	0.89 (0.88)	0.88 (0.88)	0.87 (0.86)	0.88 (0.90)
Spearman Correlation	0.90 (0.86)	0.85 (0.80)	0.78 (0.78)	0.79 (0.74)	0.84 (0.82)
BS	0.20 (0.39)	0.58 (0.71)	0.73 (0.79)	0.61 (0.70)	0.83 (1.05)
TS	0.20 (0.39)	0.54 (0.64)	0.65 (0.71)	0.55 (0.61)	0.75 (0.87)

Red numbers are the best scores.

Analysis region: 76°-88°W 36.5°-42°N

12-hour (18-06Z) accumulated precipitation

Simulation/Observation Name	SCREAM_6.5km	SCREAM_3.25km	SCREAM_1.625km	SCREAM_H	WRF_RAP	CMORPH	IMERG
RMSE	6.99 (7.14)	8.50 (8.48)	7.73 (7.52)	8.17 (8.07)	9.26 (9.05)	5.78	8.65
MAE	4.32 (4.37)	4.66 (4.65)	4.49 (4.34)	4.81 (4.71)	4.87 (4.78)	3.88	5.47
ME	-1.73 (-1.46)	-0.70 (0.67)	1.03 (0.71)	1.16 (0.96)	-0.67 (-0.71)	2.89	4.56
Pearson Correlation	0.54 (0.53)	0.50 (0.51)	0.54 (0.55)	0.54 (0.54)	0.52 (0.53)	0.77	0.72
Spearman Correlation	0.70 (0.72)	0.73 (0.75)	0.71 (0.73)	0.71 (0.73)	0.73 (0.74)	0.87	0.86
BS	1.01 (1.07)	1.46 (1.48)	1.49 (1.41)	1.72 (1.67)	1.07 (1.08)	2.17	2.54
TS	0.28 (0.27)	0.23 (0.23)	0.23 (0.24)	0.24 (0.25)	0.25 (0.25)	0.30	0.28

- WRF_RAP produces smaller biases in OLR than SCREAM when compared to observations.
- OLR BS < 1 indicates an **underestimated cold cloud area** in the simulations.
- WRF_RAP shows the largest biases in precipitation RMSE and MAE, related to the **overestimates of precipitation**.
- The ranges of precipitation RMSE and MAE in the simulations are **comparable** to those in CMORPH and IMERG, suggesting reasonable model performance in line with obserational uncertainty.

Summary

- Hindcast simulations of poorly predicted but severe historic extreme weather events enable targeted evaluation of regional and global climate modeling systems. Standardized frameworks (test beds) for evaluation are useful for examining the representation of processes in these models.
- RRM-SCREAM demonstrates similar performance to WRF on the simulation of the historic 2012 North American Derecho (i.e., worse representation of OLR, better representation of precipitation, radar reflectivity, and winds). SCREAM and WRF simulations both capture the observed derecho track, but both produce a delay of approximately 2 hours in feature location and associated gust front timing.
- The representation of the derecho is shown to benefit from the finer horizontal resolutions in SCREAM, particularly at 1.625 km grid spacing, in a variety of ways including: cold cloud temperature and its coverage, radar reflectivity structure of the MCS, derecho position during the propagation, and simulated surface gust wind speed.
- The simulations exhibit high dependence on the IC source and the initialization time.
- SCREAM at 1.625km successfully reproduces severe surface gusts with wind speeds above 30 m/s. In comparison, WRF hardly produces surface wind over 25 m/s, and the derecho wind gust in WRF is 42-46% lower than in SCREAM.

*** This work was submitted to the Journal of Advances in Modeling Earth Systems in June 2022.***