Ensemble Probabilistic Severe Weather Prediction Using Convection-Allowing Models and Deep Generative Models

Yingkai (Kyle) Sha[1], Ryan A. Sobash[1], David John Gagne II[1]

National Center for Atmospheric Research

Funding provided by NOAA grant JTTI grant NA19OAR4590128

Background and motivation

Convection Allowing Models (**CAMs**) have high spatial resolutions and they can partially resolve small-scale convection systems.

An important use of CAMs is to predict **severe weather**. However, when **ensemble predictions** are needed, it is a challenge to maintain a large number of CAM runs.

Therefore, we are seeking an alternative—using **deep generative models** to create “synthetic” CAM ensembles from a deterministic run, and prepare ensemble severe weather predictions.

Research overview

Research goal:

Post-process deterministic CAM forecasts and provide probabilistic ensemble predictions of severe weather (**tornadoes**, **hail**, and **wind gusts**) by using deep generative models and Convolutional Neural Networks (**CNN**s).

Research questions:

1. Can we generate ensembles of forecast fields and severe weather probabilities from deterministic CAM runs?

2. How well can CNN predict severe weather probabilities from high-resolution CAM forecasts?

Data: The CAM forecast and predictors

Name	Abbreviation	Type
Latitude	-	Static
Longitude	-	Static
Elevation	-	Static
Max/composite radar reflectivity	CREF	Explicit
Hourly maximum 0-2 km updraft helicity	0-2 km UH	Explicit
Hourly maximum 2-5 km updraft helicity	2-5 km UH	Explicit
Hourly accumulated precipitation	APCP	Explicit
Hourly maximum 10-m wind speed	10-m SPD	Explicit
Graupel mass	-	Explicit
Mean sea level pressure	MSLP	Environment
2-m air temperature	2-m Temp	Environment
2-m dewpoint temperature	2-m Dewpoint	Environment
Surface-based convective available potential energy	CAPE	Environment
Surface-based convective inhibition	CIN	Environment
0-1 km storm-relative helicity	0-1 km SRH	Environment
0-3 km storm-relative helicity	0-3 km SRH	Environment
0-6 km u component wind shear	0-6 km U shear	Environment
0-6 km u component wind shear	0-6 km V shear	Environment

High-Resolution Rapid Refresh version 3 and version 4 (**HRRR v3**, **v4**)

0000 UTC initializations

Hourly forecasts: 02-24Z

- Storm-scale explicit predictors (6)
- Environmental predictors (9)
- Static predictors (3)

**CIN considers magnitude only*

Data: The CAM forecast and predictors

Name	Abbreviation	Type
Latitude	-	Static
Longitude	-	Static
Elevation	-	Static
Max/composite radar reflectivity	CREF	Explicit
Hourly maximum 0-2 km updraft helicity	0-2 km UH	Explicit
Hourly maximum 2-5 km updraft helicity	2-5 km UH	Explicit
Hourly accumulated precipitation	APCP	Explicit
Hourly maximum 10-m wind speed	10-m SPD	Explicit
Graupel mass	-	Explicit
Mean sea level pressure	MSLP	Environment
2-m air temperature	2-m Temp	Environment
2-m dewpoint temperature	2-m Dewpoint	Environment
Surface-based convective available potential energy	CAPE	Environment
Surface-based convective inhibition	CIN	Environment
0-1 km storm-relative helicity	0-1 km SRH	Environment
0-3 km storm-relative helicity	0-3 km SRH	Environment
0-6 km u component wind shear	0-6 km U shear	Environment
0-6 km u component wind shear	0-6 km V shear	Environment

High-Resolution Rapid Refresh version 3 and version 4 (**HRRR v3**, **v4**)

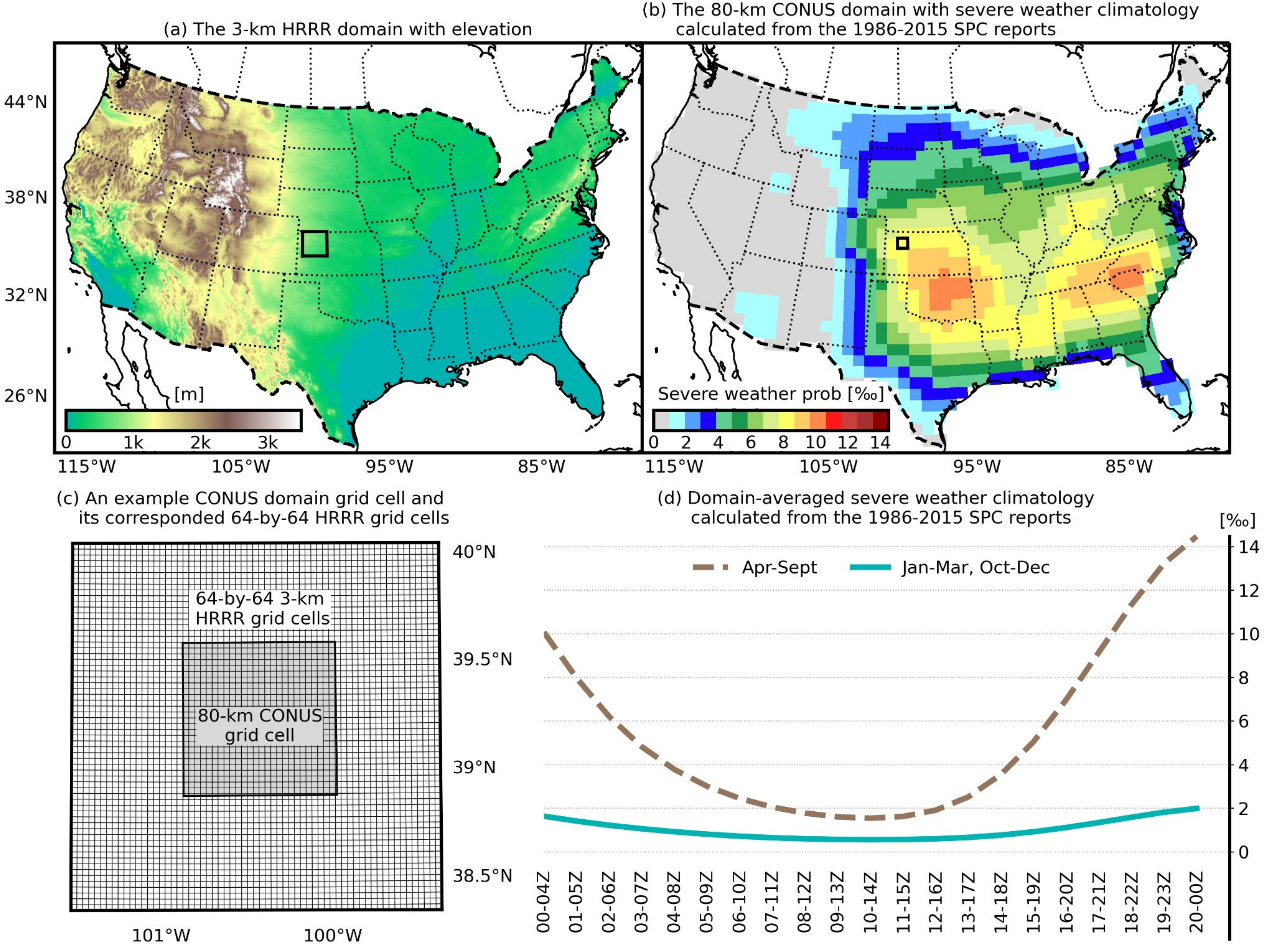
0000 UTC initializations

Hourly forecasts: 02-24Z

- Storm-scale explicit predictors (6)
- Environmental predictors (9)
- Static predictors (3)

**CIN considers magnitude only*

Data: Observations and region of interest



Input domain: 3-km grid spacing, HRRR domain

Output domain: 80-km grid spacing, ON388 211

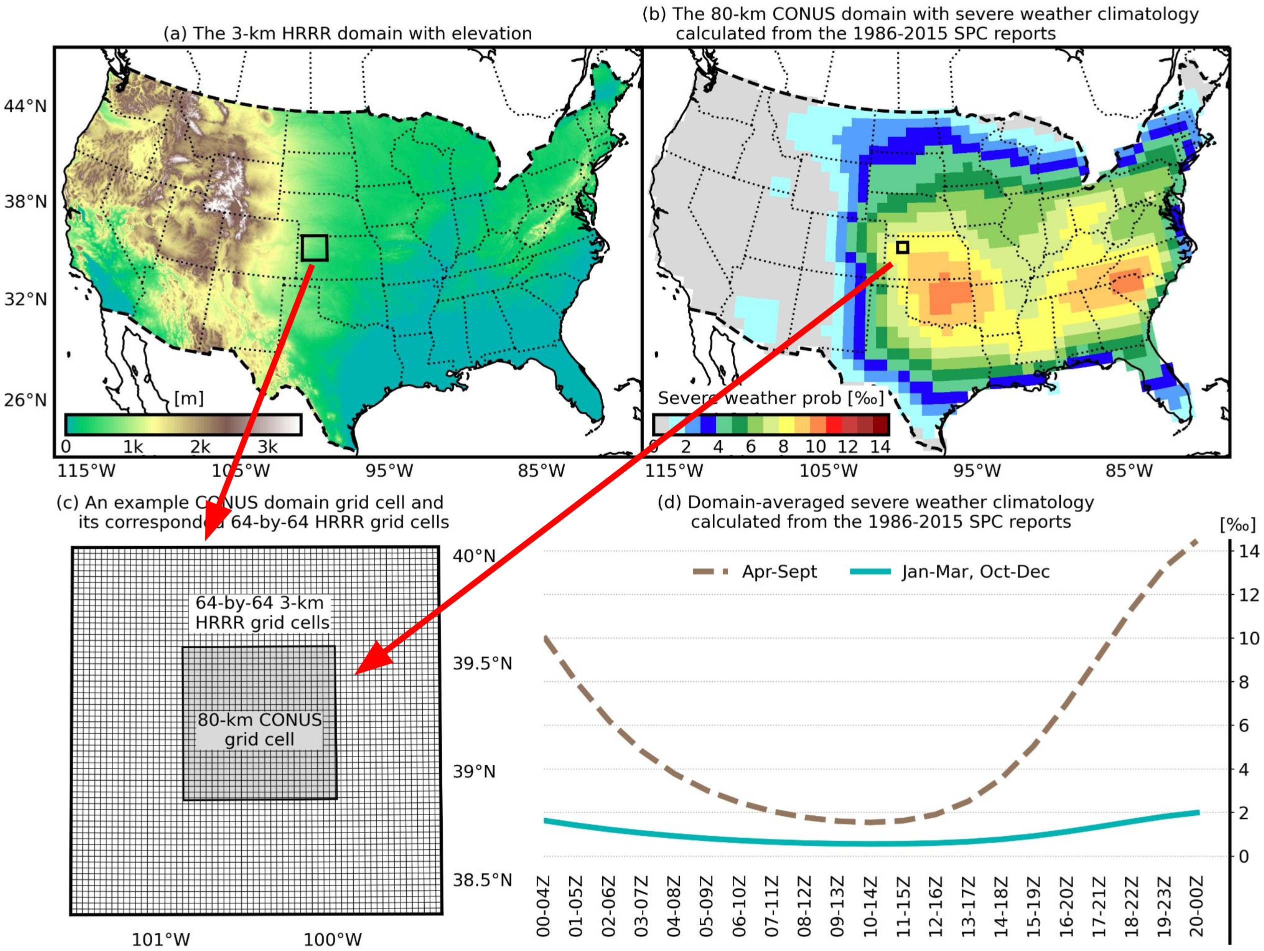
Observations:

SPC severe weather reports. Regridded to the output domain with **4-hr time window**, i.e., [-2, +1].

True: tornadoes, hail, or wind gusts

False: non-severe

Data: Observations and region of interest



Input domain: 3-km grid spacing, HRRR domain

Output domain: 80-km grid spacing, ON388 211

Observations:

SPC severe weather reports. Regridded to the output domain with **4-hr time window**, i.e., [-2, +1].

True: tornadoes, hail, or wind gusts

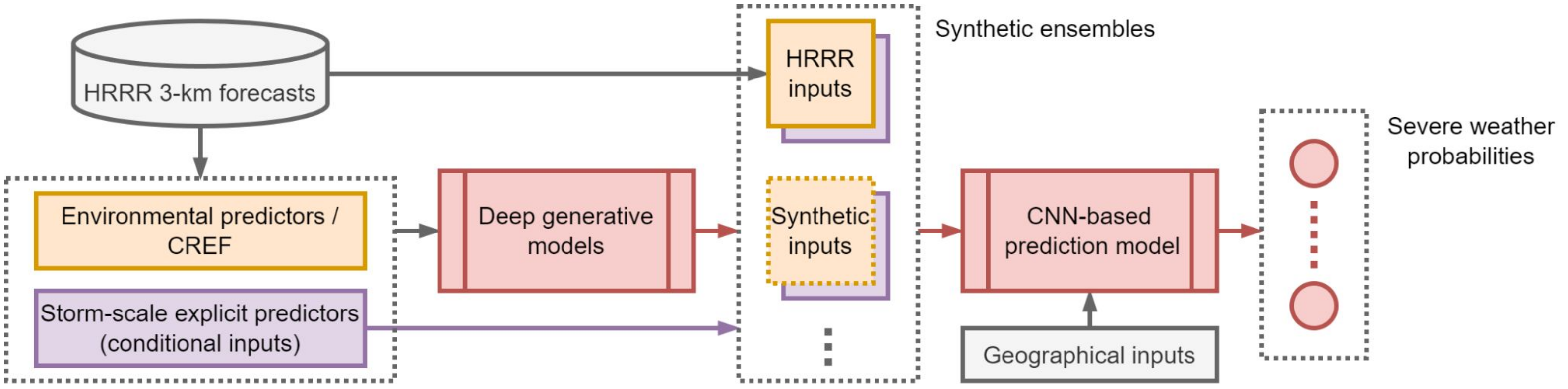
False: non-severe

Methods: roadmap

Step1: deep generative models are applied to create “synthetic ensembles” form deterministic HRRR forecasts.

- Generate: CREF, all environmental predictors
- Conditional inputs: UH, APCP, Graupel mass, 10-m SPD
- Let conditional inputs influence the generated fields to secure the output quality

Step2: CNN-based prediction model is used to produce severe weather probabilities on each synthetic member independently.



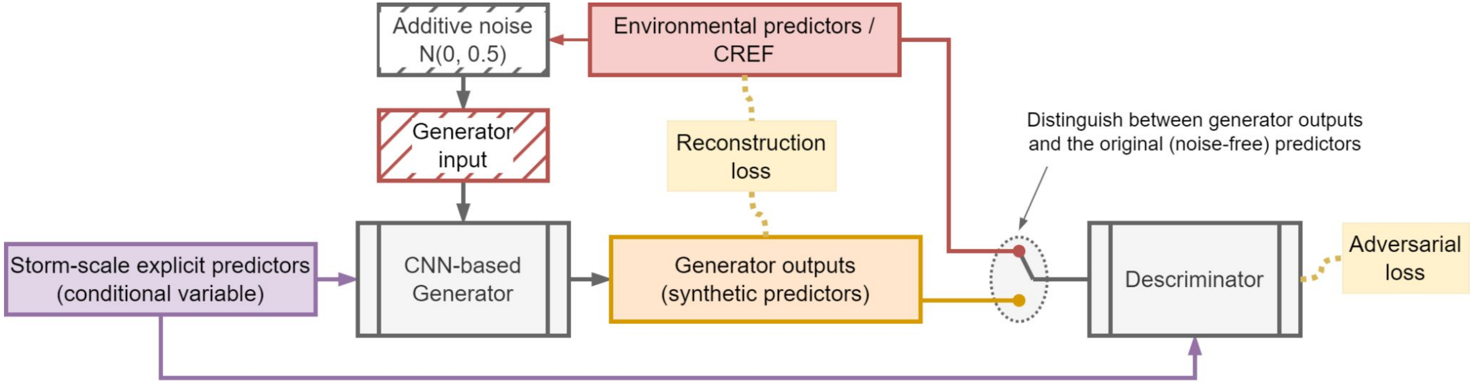
Conditional Generative Adversarial Network (CGAN)

Two CGANs are prepared, one for **[CREF, CAPE, and CIN]**, the other one for **[Other environmental predictors]**.

CGAN = **Generator** v.s. Discriminator (both are CNNs)

Training procedures *(200 epochs* + *training loss stabilizes)*:

1. Generator produce outputs to update the discriminator
2. Discriminator computes its loss function to update the generator



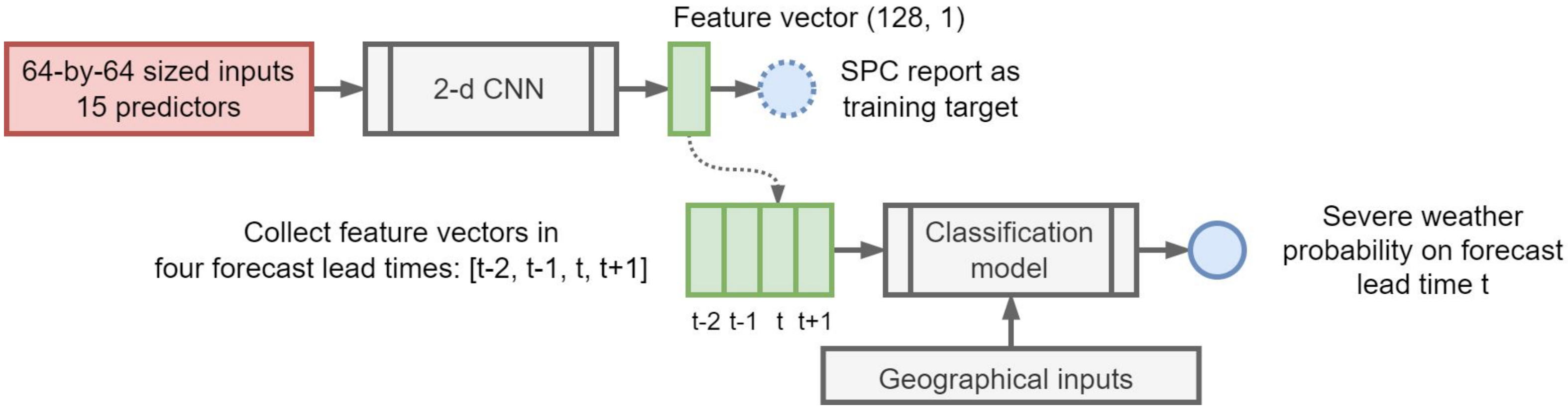
CNN-based severe weather prediction model

General design: decoupled **representation learning** and **classification**

- **Representation learning**: 2-d CNN + global max pooling → feature vector
- **Classification**: feature vec x4 → 1-d CNN + global max pooling → dense layer + Monte-Carlo (**MC**) dropout → Output

Benefits:

- Handle long-tailed data better
- Save computation (avoid the use of 3-d convolution kernels; faster training)



Methods: the post-processing experiment

Training: HRRR v3 v4, and SPC reports from **2018/07/15** to **2020/12/31**

Validation: 10% random-split from the training set.

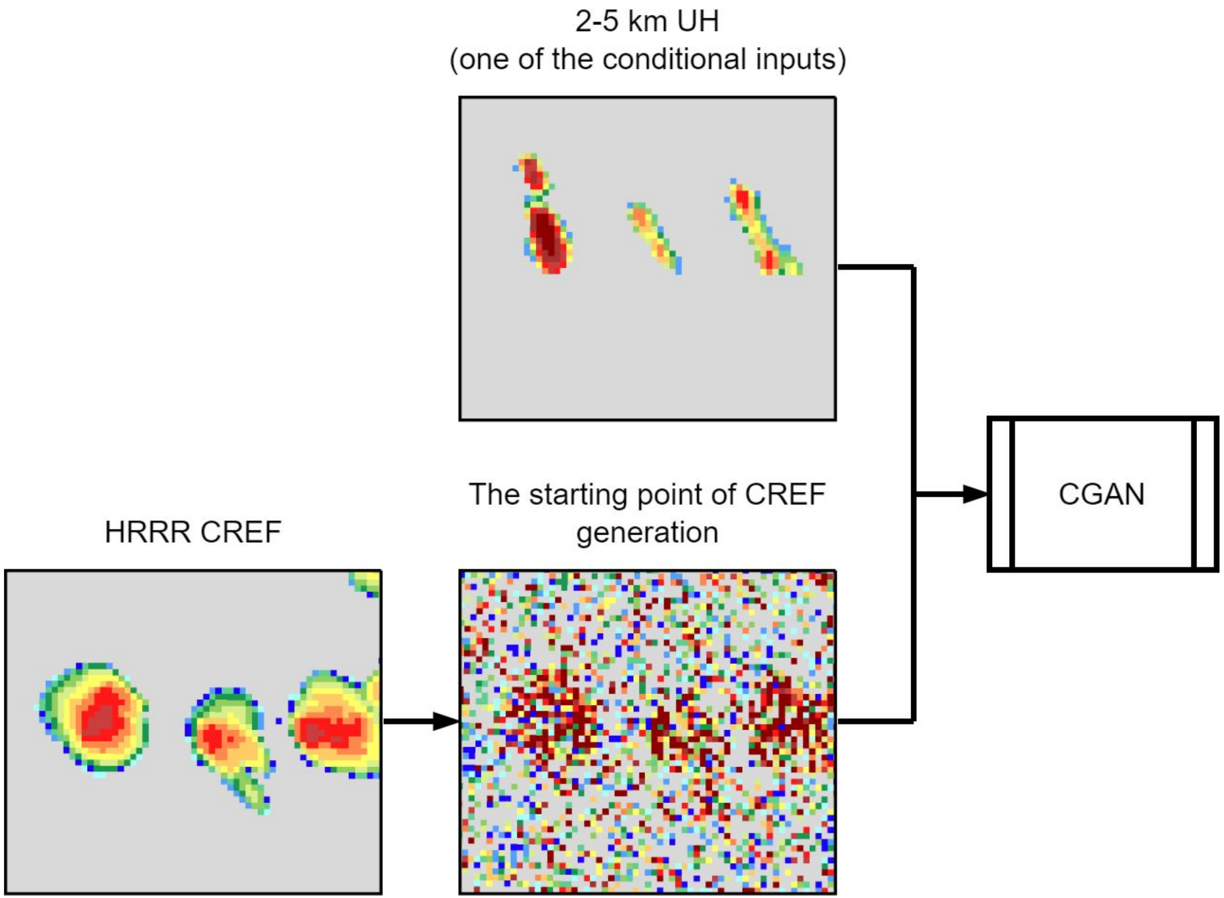
Verification: HRRR v4 and SPC reports from **2021/01/01** to **2021/12/31**

CGAN ensemble (ours): 50 members based on **CGAN** outputs and **MC dropout**

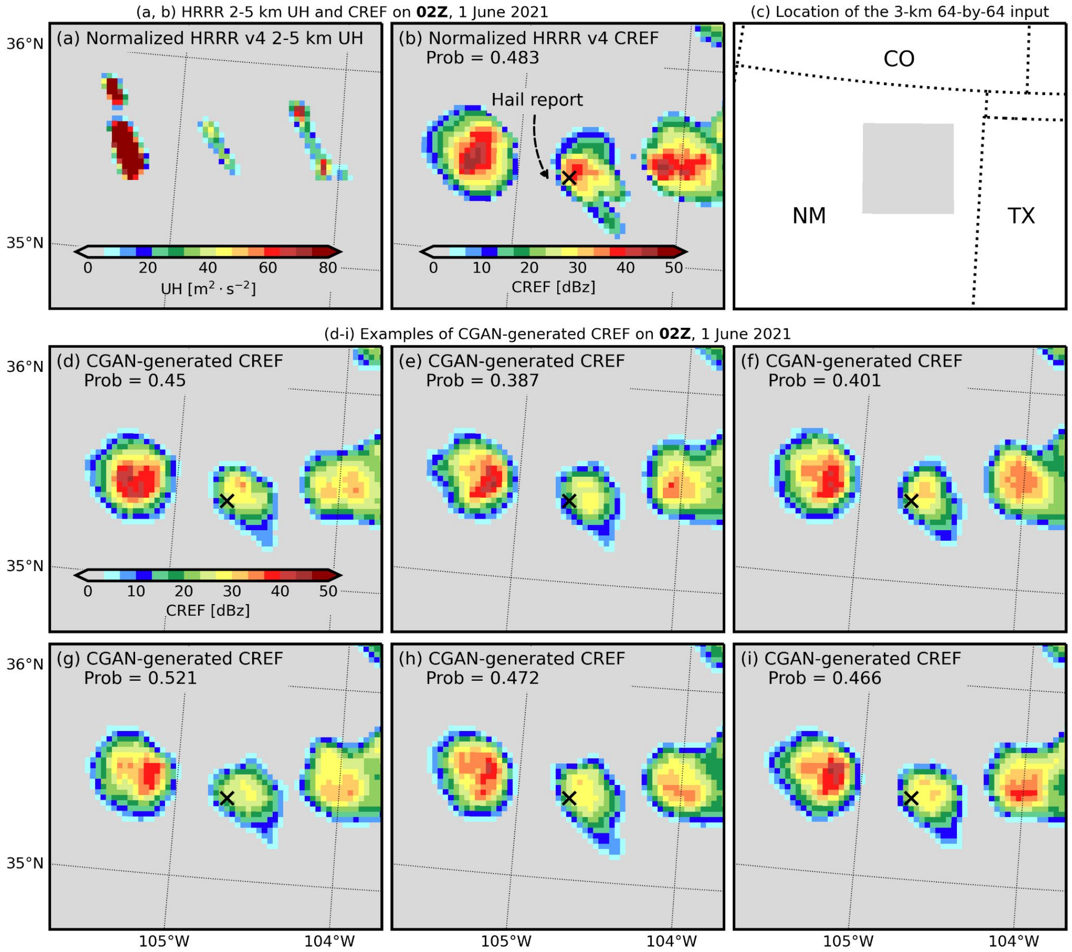
CNN ensemble: 50 members based on the same CNN prediction model as “ours”, but with **MC dropout** only, and without using CGANs

MLP ensemble: 50 members based on Multilayer Perceptrons (**MLP**) and **MC dropout**.

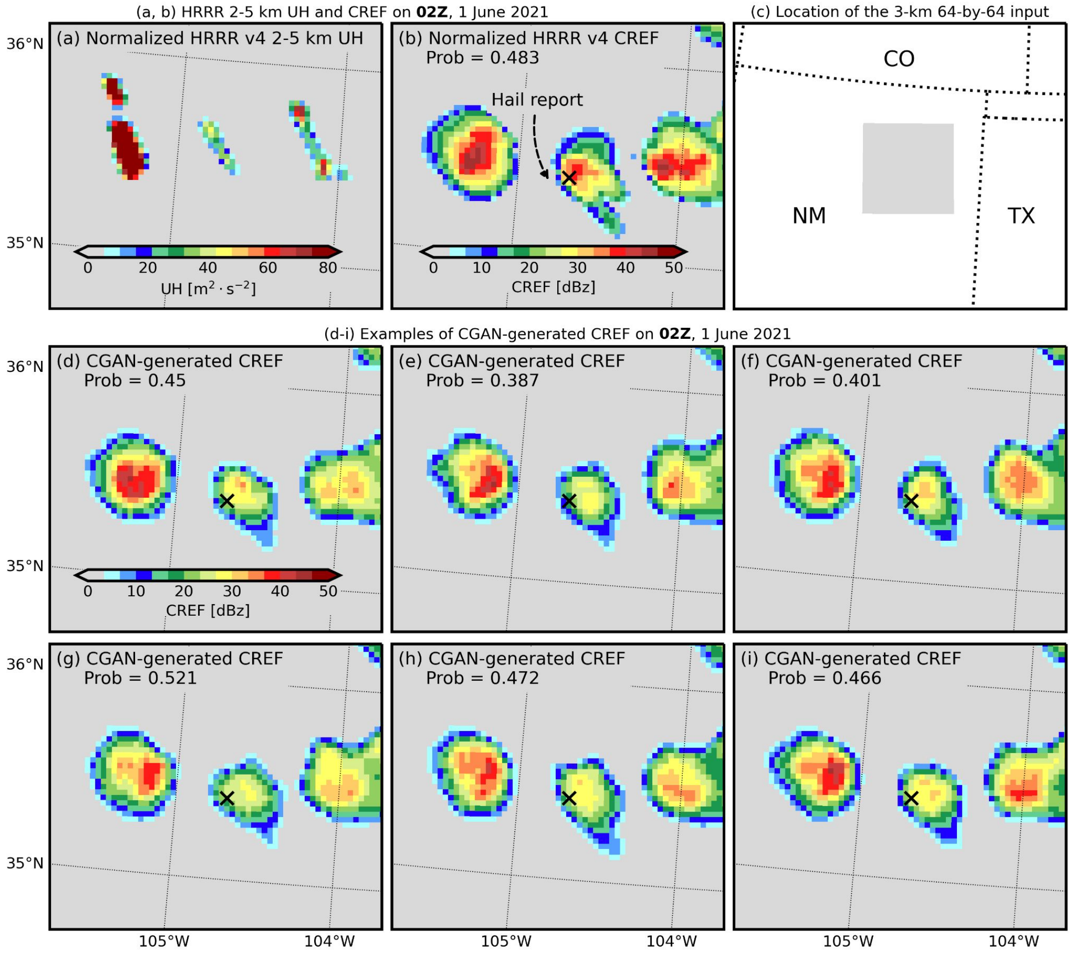
Result: examples of CGAN outputs



Result: examples of CGAN outputs

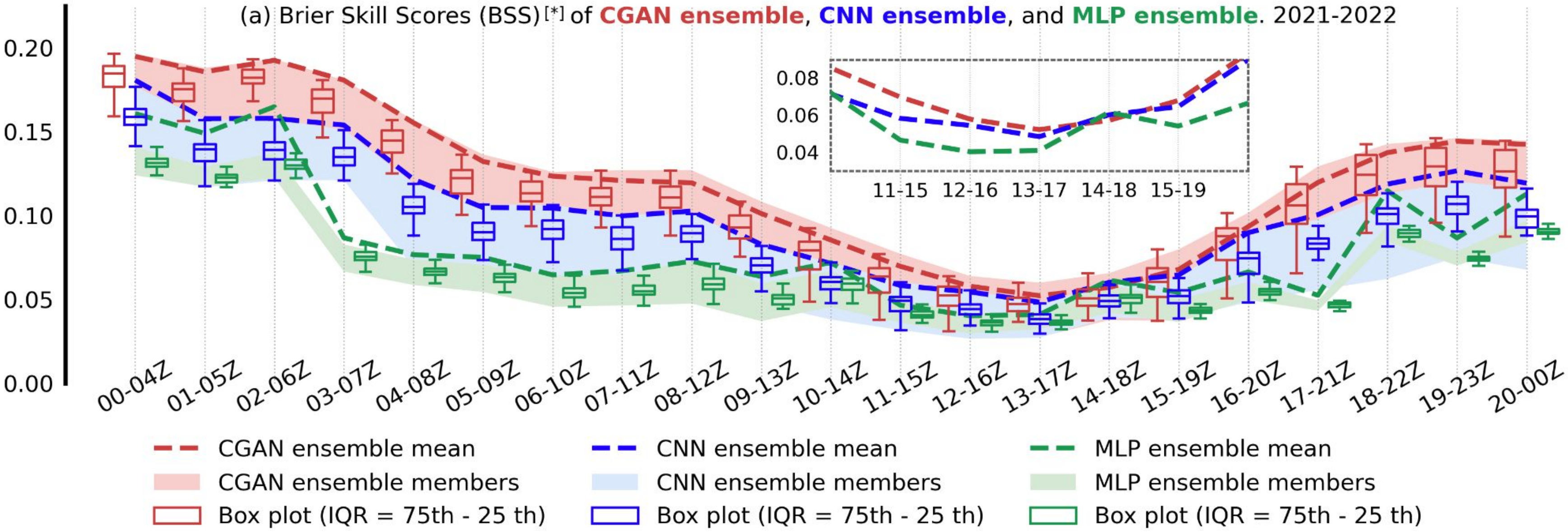


Result: examples of CGAN outputs



- CGAN outputs preserved the location and shape of the CREF patterns.
- CGAN outputs have variations on the intensity and edge of the patterns.
- These variations can impact the severe weather probabilities

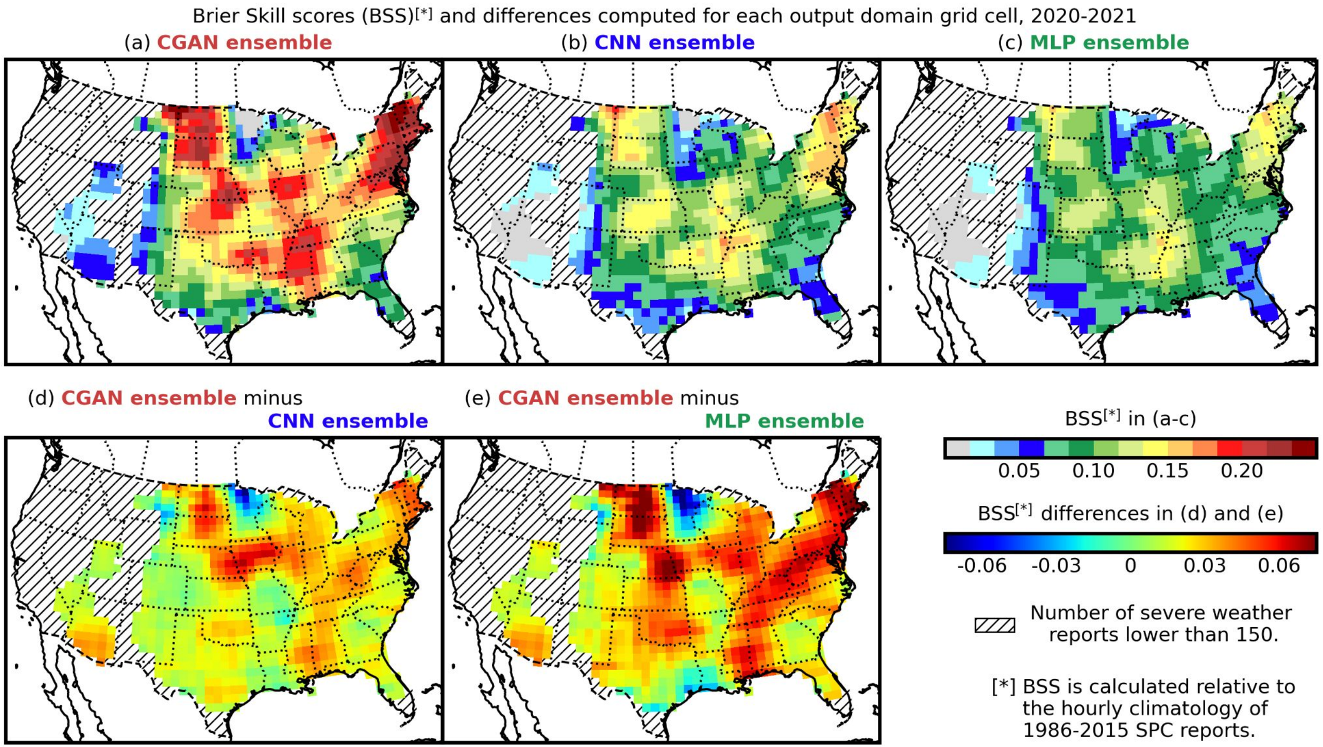
Result: Brier Skill Scores (BSSs)



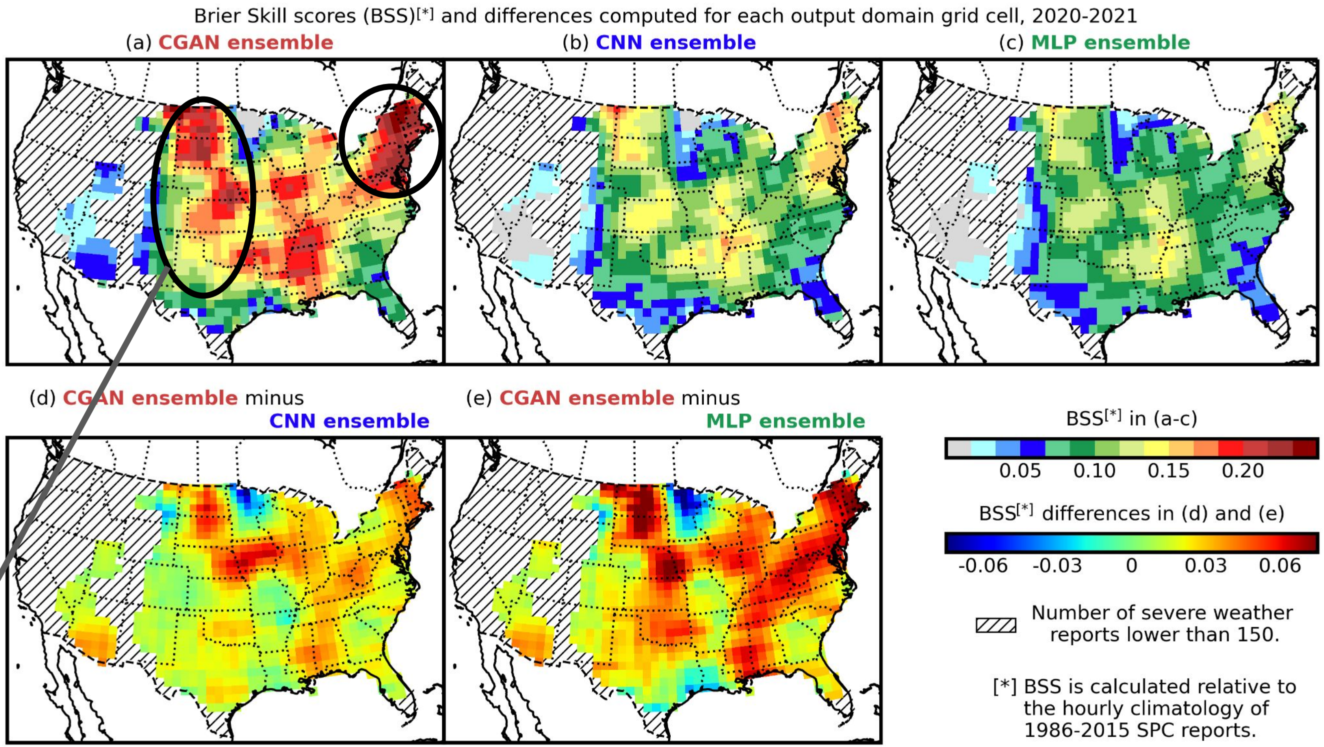
- Late night - early morning: all methods performed poorly.
  - HRRR forecasts are less skilful in these hours.
  - Fewer severe weather reports are available in these hours for training.

- Daytime hours: ours are clearly better than the two baselines.

Result: the spatial distribution of BSS



Result: the spatial distribution of BSS

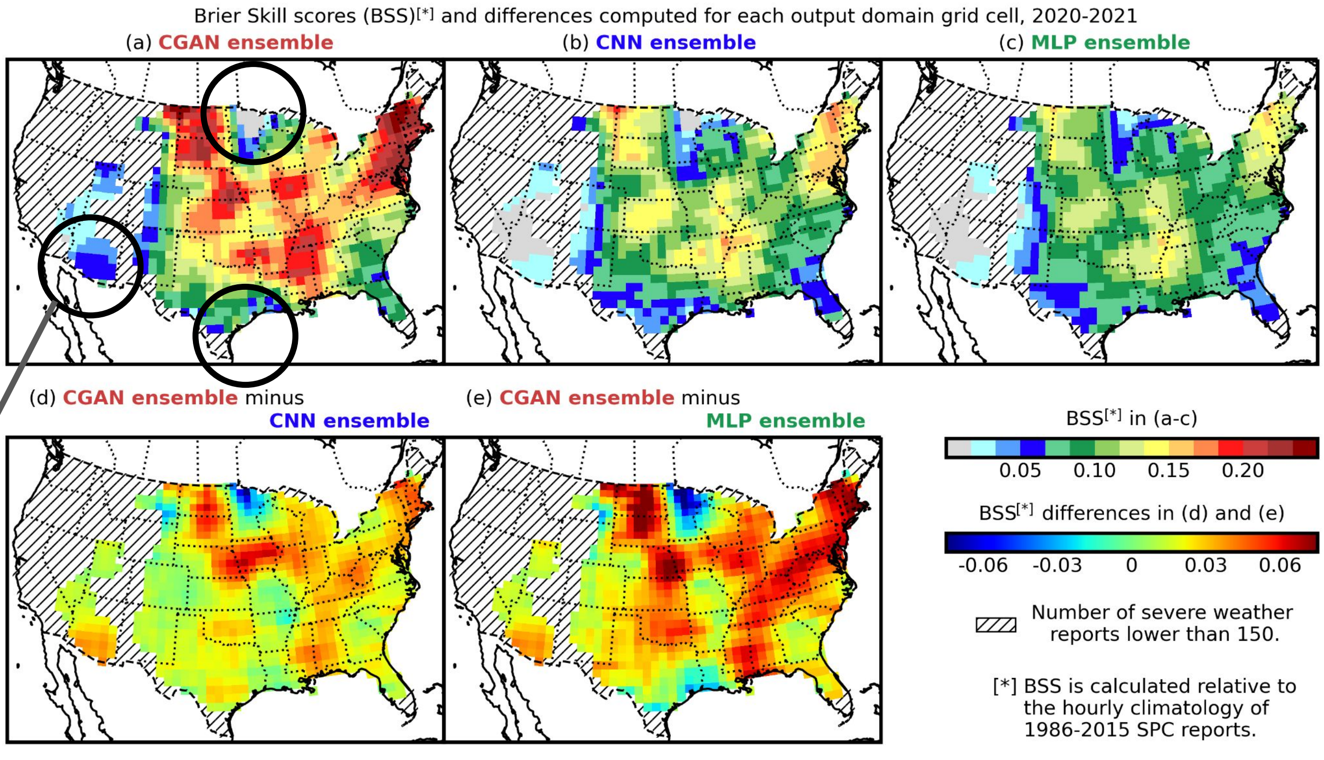


Good performance:

- Great Plain
- Northeastern US

- Predictors like UH are useful for severe weather events in the Great Plain
- Plenty severe weather cases to learn from the training data

Result: the spatial distribution of BSS



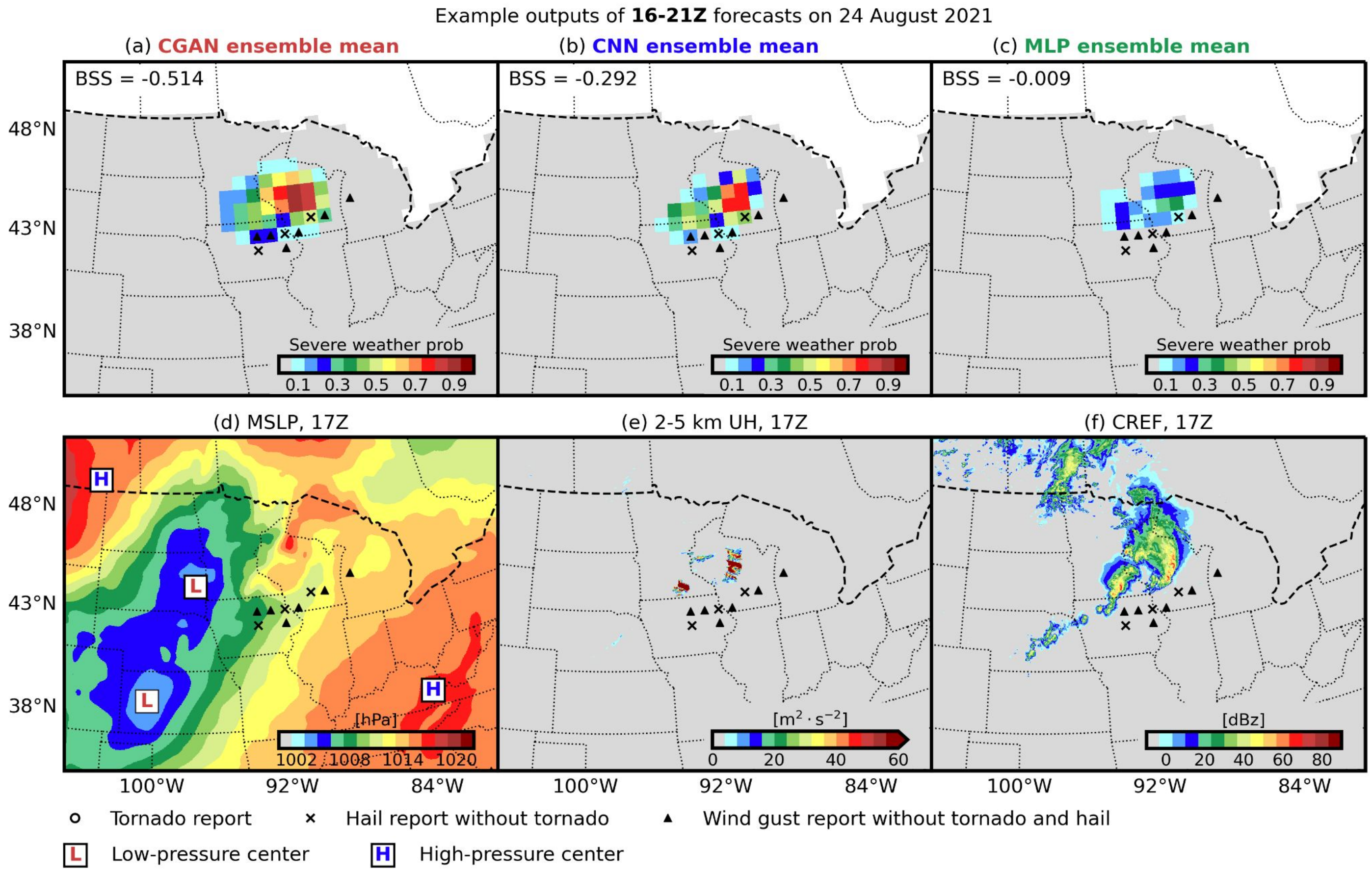
Bad performance:

- Some bordering areas
- Southern Arizona

Reason:

- Coastal environment; monsoon thunderstorms
- Positional errors

Result: The impact of HRRR positional errors



Result: successful predictions by severe weather categories

(f) The top 11.45% most successfully predicted severe weather cases[**] for each method and by categories

	Tornadoes only	Tornadoes & wind gusts	Tornadoes, wind gusts & hail	Wind gusts only	Wind gusts & hail	Hail only
Total number of cases	1510	711	126	15461	2311	5435
CGAN en. mean	100 (6.623%)	252 (35.443%)	35 (27.778%)	1927 (12.464%)	333 (14.409%)	279 (5.133%)
CNN en. mean	114 (7.550%)	235 (33.052%)	32 (25.397%)	1943 (12.567%)	320 (13.847%)	277 (5.097%)
MLP en. mean	106 (7.020%)	231 (32.489%)	23 (18.254%)	1986 (12.845%)	306 (13.241%)	269 (4.949%)

[**] We selected the top 15% most successful "true positive" predictions for each method and each forecast lead time based on Brier score rankings, and excluded predictions with prob < 0.45

Result: successful predictions by severe weather categories

(f) The top 11.45% most successfully predicted severe weather cases[**] for each method and by categories

	Tornadoes only	Tornadoes & wind gusts	Tornadoes, wind gusts & hail	Wind gusts only	Wind gusts & hail	Hail only
Total number of cases	1510	711	126	15461	2311	5435
CGAN en. mean	100 (6.623%)	252 (35.443%)	35 (27.778%)	1927 (12.464%)	333 (14.409%)	279 (5.133%)
CNN en. mean	114 (7.550%)	235 (33.052%)	32 (25.397%)	1943 (12.567%)	320 (13.847%)	277 (5.097%)
MLP en. mean	106 (7.020%)	231 (32.489%)	23 (18.254%)	1986 (12.845%)	306 (13.241%)	269 (4.949%)

[**] We selected the top 15% most successful "true positive" predictions for each method and each forecast lead time based on Brier score rankings, and excluded predictions with prob < 0.45

Bad performance:
- Tornadoes only
- Hail only

Good performance:
- Tornadoes and wind gusts

Conclusions

- A post-processing method was developed by integrating Conditional Generative Adversarial Networks (CGANs) and a Convolutional Neural Network (CNN) classifier to generate probabilistic forecasts for severe weather events using HRRR forecasts.
- The CGANs were trained to produce synthetic ensemble members based on the deterministic HRRR forecasts, while the CNN classifier utilized the outputs from the CGANs to generate forecasts of severe weather probabilities.
- The method is successful, it is overall more skillful than the MLP and CNN baselines, and it has achieved 0.2 BSS for short forecast lead times.
- The method performed well in the Great Plains and the Northeast CONUS. It is especially successful in predicting severe weather events that combine tornadoes and wind gusts.

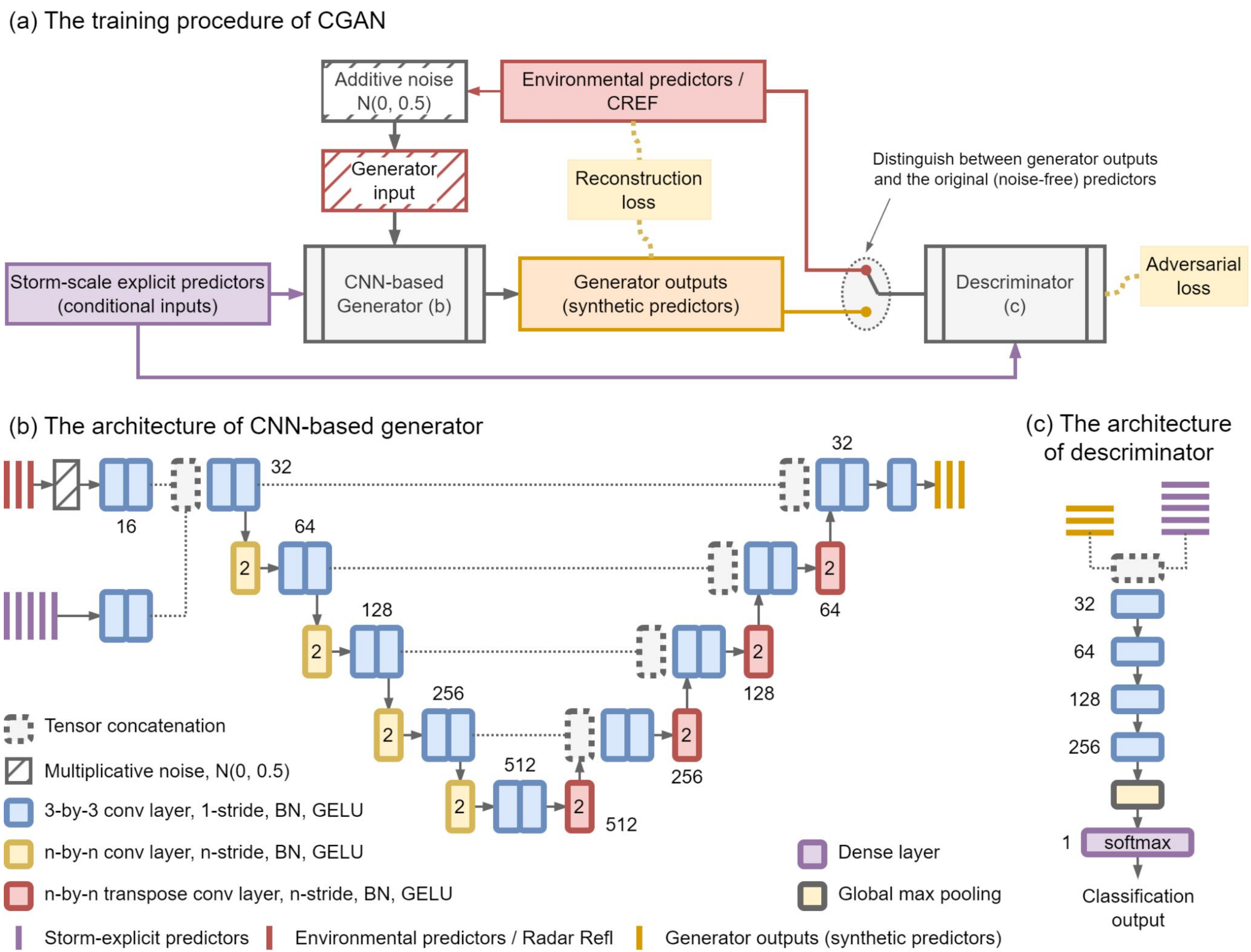
Thank you

Sha, Y., R. A. Sobash, D. J. Gagne II, 2023: Generative ensemble deep learning severe weather prediction from a deterministic convection-allowing model. in review: Artificial Intelligence for the Earth Systems. pre-print: http://arxiv.org/abs/2310.06045

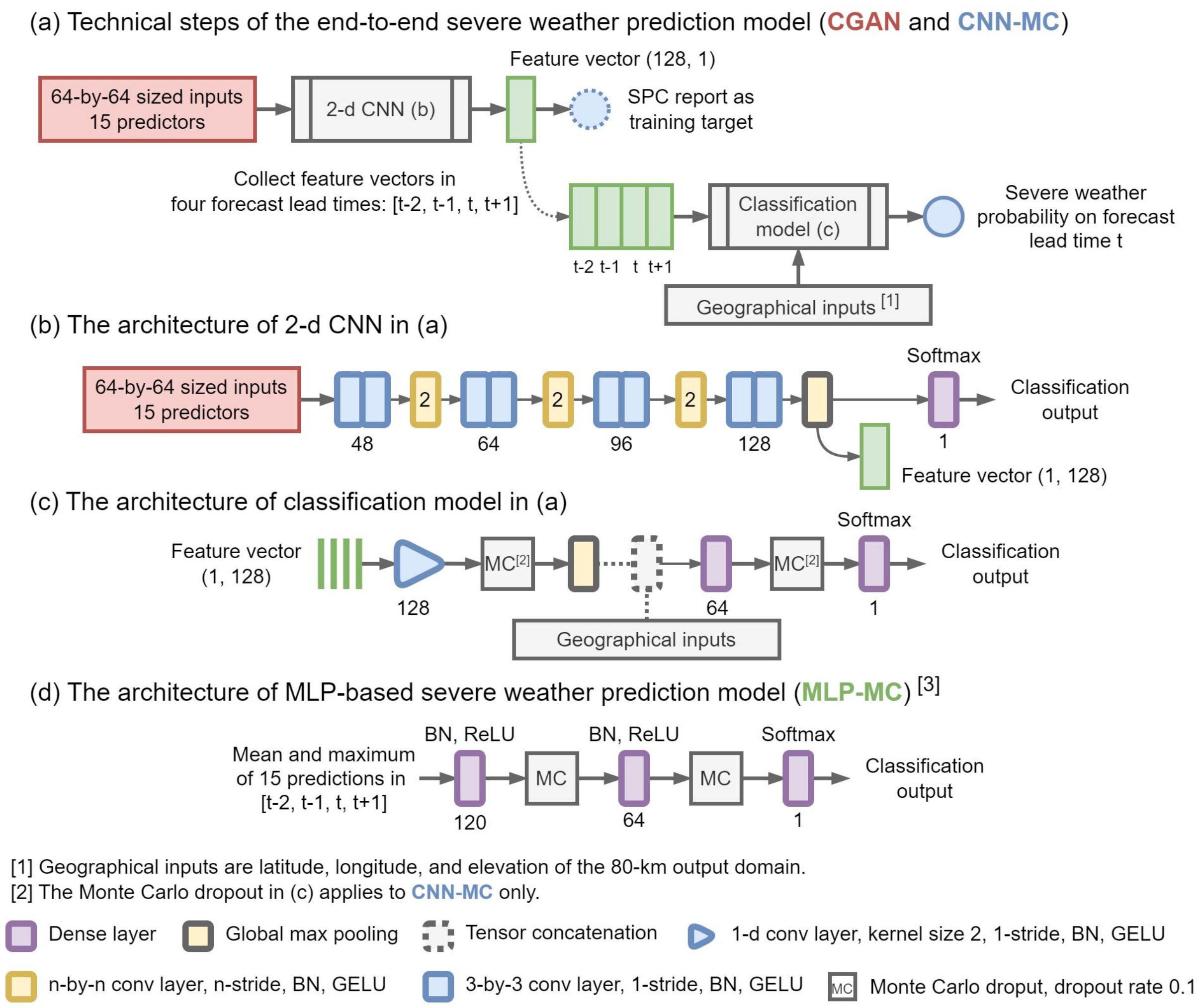
Contact: ksha@ucar.edu sobash@ucar.edu

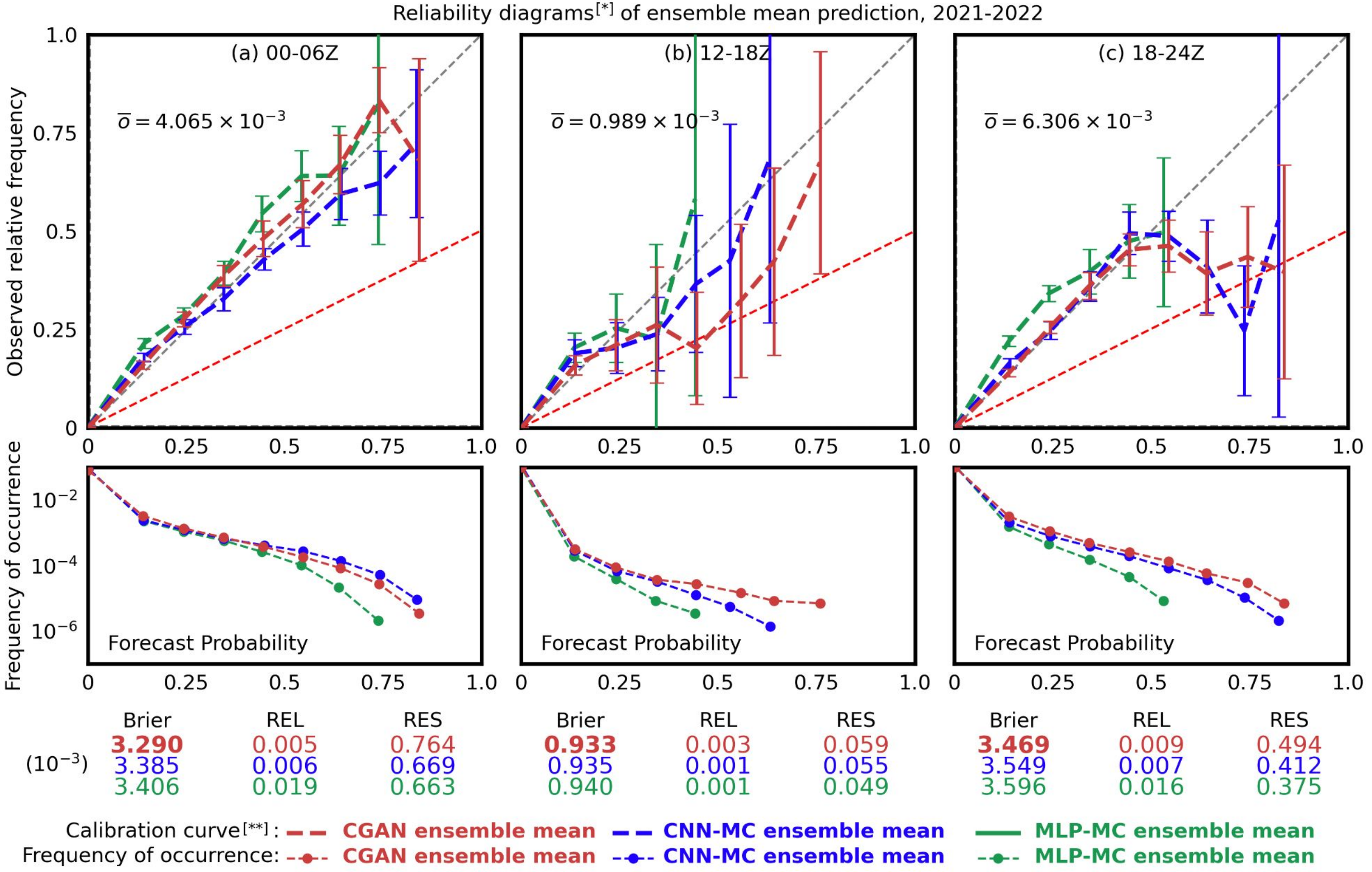
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Appendix



(a) The training procedure of CGAN
Additive noise N(0, 0.5)
Environmental predictors / CREF
Generator input
Reconstruction loss
Distinguish between generator outputs and the original (noise-free) predictors
Storm-scale explicit predictors (conditional inputs)
CNN-based Generator (b)
Generator outputs (synthetic predictors)
Descriminator (c)
Adversarial loss
(b) The architecture of CNN-based generator
16
32
64
128
256
512
Tensor concatenation
Multiplicative noise, N(0, 0.5)
3-by-3 conv layer, 1-stride, BN, GELU
n-by-n conv layer, n-stride, BN, GELU
n-by-n transpose conv layer, n-stride, BN, GELU
Dense layer
Global max pooling
Storm-explicit predictors
Environmental predictors / Radar Refl
Generator outputs (synthetic predictors)
(c) The architecture of descriminator
softmax
Classification output

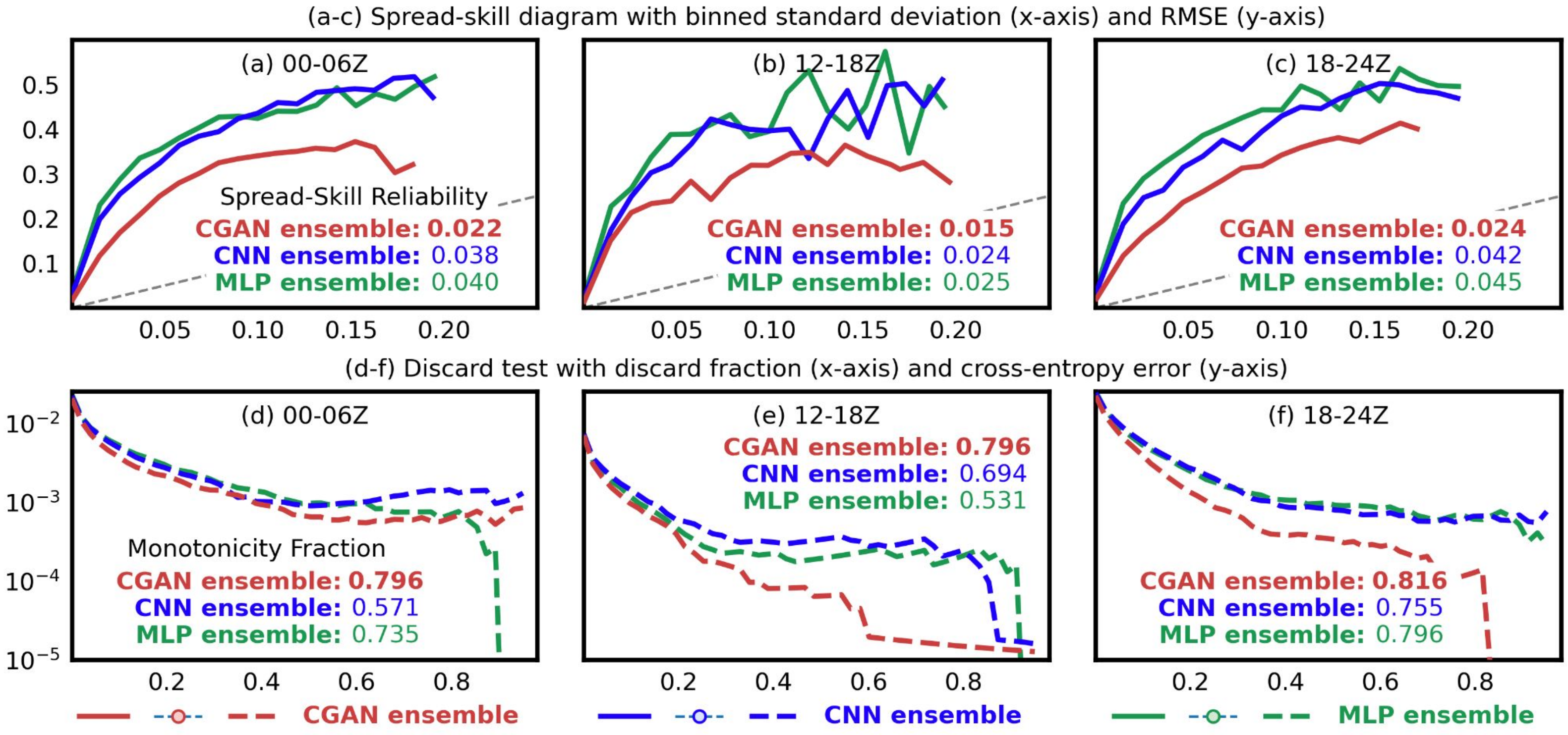

(a) Technical steps of the end-to-end severe weather prediction model (CGAN and CNN-MC)
64-by-64 sized inputs 15 predictors
2-d CNN (b)
Feature vector (128, 1)
SPC report as training target
Collect feature vectors in four forecast lead times: [t-2, t-1, t, t+1]
t-2 t-1 t t+1
Classification model (c)
Severe weather probability on forecast lead time t
Geographical inputs [1]
(b) The architecture of 2-d CNN in (a)
64-by-64 sized inputs 15 predictors
48
2
64
2
96
2
128
Softmax
1
Classification output
Feature vector (1, 128)
(c) The architecture of classification model in (a)
Feature vector (1, 128)
128
MC[2]
64
MC[2]
Softmax
1
Classification output
Geographical inputs
(d) The architecture of MLP-based severe weather prediction model (MLP-MC) [3]
Mean and maximum of 15 predictions in [t-2, t-1, t, t+1]
BN, ReLU
120
MC
BN, ReLU
64
MC
Softmax
1
Classification output
[1] Geographical inputs are latitude, longitude, and elevation of the 80-km output domain.
[2] The Monte Carlo dropout in (c) applies to CNN-MC only.
Dense layer
Global max pooling
Tensor concatenation
1-d conv layer, kernel size 2, 1-stride, BN, GELU
n-by-n conv layer, n-stride, BN, GELU
3-by-3 conv layer, 1-stride, BN, GELU
MC Monte Carlo droput, dropout rate 0.1



[*] Reliability diagrams and Brier score components are computed relative to the hourly climatology of 1986-2015 SPC reports.

[**] Calibration curves are averaged over 100 bootstrap replicates. Error bars represent the 95% confidence intervals.

Result: uncertainty quantification



Result: CGAN output v.s. HRRR (feature importance)