



Development and tests of machine-learned cloud droplet number concentration in a next-generation bulk microphysics scheme for operational applications

Anders A. Jensen¹, David D. Turner¹, Jorge Guerra³, Stanley G. Benjamin^{1,2}, Clark Evans¹, David J. Gagne³, Justin R. Minder⁴, and Joseph B. Olson¹

¹National Oceanic and Atmospheric Administration, Global Systems Laboratory, 325 Broadway, Boulder, Colorado

²CIRES, University of Colorado Boulder, Boulder, Colorado

³National Center for Atmospheric Research, Boulder, Colorado

⁴Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York

Correspondence: Anders A. Jensen (anders.jensen@noaa.gov)

Abstract. A machine-learning (ML) model was developed to predict cloud droplet number concentrations. This ML model was trained on data from simulations using the Model for Prediction Across Scales–Atmosphere (MPAS-A) coupled to a next-generation version of the Thompson-Eidhammer bulk microphysics scheme. The ML model was tested for two applications. First, the ML model was used to predict the number concentration, and hence the effective radii of subgrid-scale clouds from the MYNN boundary-layer scheme. These subgrid-scale clouds, necessary for accurate predictions of radiation, have historically been parameterized using constant values for their effective radii. The ML model predicted cloud droplet number concentrations for subgrid-scale clouds so that their effective radii could be calculated using a method physically consistent with those of fully resolved clouds predicted by the bulk microphysics scheme. These results were compared with observations using a novel approach to show that the ML model provided reasonable estimates of effective radii. Second, cloud droplet number concentrations were replaced in the bulk microphysics parameterization with values predicted by the ML model. Results from this test were compared with aircraft and surface observations from two IOPs during the WINTRE-MIX field campaign. Overall, the simulation with the ML prediction of cloud droplet number concentrations reproduced the 24-h total precipitation for the case and reasonably represented the microphysical details. The ML model performed better for relatively short forecasts, which was attributed to the lack of model spin-up required by physical parameterizations. The biases in the ML model for this case study resulted in reduced drizzle production, which slightly degraded the forecast of precipitation phase. Benefits of the ML method include lower computational cost, better coupling between fully resolved and unresolved (subgrid-scale) clouds, and more rapid development of realistic forecasts when initial conditions are not fully spun up.



1 Introduction

Bulk microphysics parameterizations (BMPs) are critical components of numerical weather prediction (NWP) models. BMPs predict clouds and precipitation and provide stakeholders with valuable information about weather hazards such as fog, aircraft icing conditions, snowstorms, heavy rain, flooding, and damaging hail. The prediction of clouds and precipitation is a complex challenge. BMPs contain an enigmatic amount of complexity—the amount of complexity needed varies with model spatial resolution and the types of clouds simulated—to differentiate liquid from ice and to represent a wide range of particle sizes from small cloud droplets to large hydrometeors. Thus, BMPs generally contain categories such as cloud water, cloud ice, rain, snow, and graupel/hail.

One-moment BMPs include predictions of only mass concentrations (e.g., cloud water mass) and are computationally faster than two-moment BMPs, which predict both mass and number concentrations. Computational cost scales linearly with the number of prognostic variables that are transported by the advection scheme. Despite the added computational cost, two-moment BMPs are popular because they have been shown to improve the representation of clouds and precipitation compared with one-moment schemes (Morrison et al., 2009; Igel et al., 2015). The prediction of both mass and number concentrations generally leads to improved constraints on particle sizes. This results in more physically realistic representations of microphysical processes that depend on size (e.g., rain evaporation or drizzle formation) and more realistic constraints on cloud and ice effective radii, which directly affect cloud radiative properties. Since one-moment schemes are computationally efficient and two-moment schemes generally produce better predictions, machine learning provides a natural method to attain the best of both worlds and predict a second moment (or other particle properties) without needing advection. This is especially true for BMPs used operationally, such as the Thompson schemes (Thompson et al., 2004, 2008; Thompson, 2013; Thompson and Eidhammer, 2014) in the High-Resolution Rapid Refresh (HRRR) model and data assimilation system (Benjamin et al., 2016; Dowell et al., 2022), which must balance complexity with computational efficiency to provide timely predictions.

The Thompson scheme originally predicted mass concentrations for cloud water, rain, cloud ice, snow, and graupel/hail, alongside number concentrations only for rain and cloud ice (Thompson et al., 2004, 2008). To enhance model performance without adding prognostic variables, complexity—guided when possible by observations—was incorporated into each of these categories to better represent particle sizes while predicting only mass concentrations (see Thompson, 2013). This scheme was shown to accurately predict an array of weather hazards, including severe convective storms (Clark et al., 2010). Continued development of the Thompson scheme into the Thompson-Eidhammer microphysics parameterization (TE14, Thompson and Eidhammer, 2014) added an “aerosol-aware” component and the prediction of cloud water number concentration, alongside water-friendly and ice-friendly aerosol number concentrations. The TE scheme was shown to produce remarkably accurate predictions of surface precipitation over most of the continental United States (Liu et al., 2017), aircraft icing environments (including droplet sizes) from supercooled cloud water (Thompson et al., 2017), and supercooled large drops (Tessendorf et al., 2021; Jensen et al., 2023b). The addition of a prognostic cloud droplet number concentration was proven to be critical for these applications.



Given its past success, the Thompson microphysics scheme continues to be developed for future research and operational forecast applications (including the Rapid Refresh Forecast System, or RRFS, Benjamin et al., 2026). As part of this development, the parameterization has been ported to a GitHub repository called the Thompson-Eidhammer Microphysics Parameterization for Operations, or TEMPO. TEMPO can be used as a submodule with various newly developed dynamical cores, such as the Model for Prediction Across Scales–Atmosphere (MPAS-A, Skamarock et al., 2012). Thus, microphysics development in TEMPO aims to be dynamical-core agnostic.

As part of the new repository, a machine-learning (ML) component has been incorporated that predicts cloud droplet number concentration. ML models are useful for improving or accelerating operationally run physical parameterizations (Lagerquist et al., 2021; Muñoz-Esparza et al., 2022), and they can accurately emulate the warm-rain process (Gettelman et al., 2021) and a full microphysics scheme (Seifert and Siewert, 2024). Additionally, ML models can improve microphysics diagnostics (Jensen et al., 2023b) and enhance the coupling between physics parameterizations, including microphysics, convection, and subgrid-scale clouds. Full replacement of an operational microphysics scheme with an ML model is not yet practical. Training an ML model to predict all microphysical processes is computationally expensive. Furthermore, post-training model parameter tuning (e.g., tuning collection efficiencies), which is often needed in operational forecast systems to account for interactions between the microphysics and other parameterizations or forecast components, is not possible (Seifert and Siewert, 2024). For these reasons, an ML model was developed to predict cloud droplet number concentration as a practical starting point for the implementation of ML components into the Thompson-based microphysics scheme. The ML model was tested for two use cases. First, the ML model was used to predict cloud droplet number concentrations for subgrid-scale clouds so that cloud effective radii were physically consistent between subgrid-scale and fully resolved clouds. Second, the ML model was used to predict cloud droplet number concentrations for fully resolved clouds. This second test emulated two-moment cloud water without needing number-concentration advection, which can be computationally expensive and produce errors (Wang et al., 2009).

In this paper, the next-generation Thompson microphysics scheme, or TEMPO, and the ML additions are tested in MPAS using a wintertime case study during the Winter Precipitation Type Research Multi-scale Experiment (WINTRE-MIX) field campaign. Aircraft and surface observations are used to evaluate the baseline TEMPO scheme (without ML additions) in MPAS, as well as the ML additions when applied separately to the subgrid-scale clouds to calculate effective radii and to the fully resolved clouds for number concentrations. A brief description of TEMPO is presented in Section 2. The ML model that predicts cloud droplet number concentration is described in Section 3. The case study and observations are described in Section 4. Results are presented in Sections 5 and 6, and Section 7 presents a summary.



80 2 TEMPO: The Thompson-Eidhammer Microphysics Parameterization for Operations

The Thompson-Eidhammer microphysics scheme has been developed in and for individual models and applications, such as the Weather Research and Forecasting model (WRF, Skamarock et al., 2019), which is the dynamical core used in the HRRR. This model-specific development approach can lead to code divergence among modeling frameworks such as WRF, MPAS, Cloud Model 1 (CM1, Bryan and Fritsch, 2002), and the Common Community Physics Package (CCPP, Heinzeller et al., 2023), all of which include versions of the Thompson microphysics scheme.

The Thompson-Eidhammer Microphysics Parameterization for Operations, or TEMPO (National Oceanic and Atmospheric Administration, 2026), is a version of the Thompson-Eidhammer BMP that resides in a public code repository for community use and development. TEMPO can connect to other repositories (e.g., dynamical cores, single-column models, or forecast systems) as a submodule. For example, TEMPO has been connected to the Unified Forecast System (UFS) Community fork of MPAS-A (UFS Community, 2026b) and the CCPP (UFS Community, 2026a). Additionally, TEMPO is being tested for release in the WRF model version 4.8.0 (National Center for Atmospheric Research, 2026b). Integration with the CCPP enables the use of TEMPO in all UFS applications as well as with the FV3 dynamical core (Putman and Lin, 2007). Similarly, integration with WRF makes TEMPO accessible to the extensive WRF research community. The main branch of TEMPO is being tested in the National Oceanic and Atmospheric Administration's Global Systems Laboratory (NOAA-GSL) MPAS-A-based convection-allowing Rapid Refresh Forecast System (RRFS), which is currently in development for future operational implementation.

TEMPO contains the original Thompson microphysics scheme (Thompson et al., 2004, 2008), an option to use the Thompson-Eidhammer "aerosol-aware" parameterization (Thompson and Eidhammer, 2014), and an option to use two-moment graupel with predicted density, or the "hail-aware" parameterization (Jensen et al., 2023a). The original Thompson scheme predicts seven variables: cloud water, rainwater, cloud ice, snow, and graupel/hail mass concentrations (q_C , q_R , q_I , q_S , q_G), alongside rain and cloud ice number concentrations (n_R , n_I). The graupel/hail hybrid category represents all rimed ice and is hereafter called graupel. Note that the prognostic variables are mass and number mixing ratios with units of kg kg^{-1} and kg^{-1} , which are related to mass and number concentrations (units of kg m^{-3} and m^{-3}) by the air density. The predictions of n_C alongside water-friendly (n_{wfa}) and ice-friendly (n_{ifa}) aerosol concentrations were added as "aerosol-aware" additions, allowing aerosols to affect clouds. Additionally, a "hail-aware" parameterization was added that changed the single-moment graupel category into a two-moment graupel category with predicted density through the additions of graupel number (n_G) and volume (b_G) concentrations. The "hail-aware" parameterization can differentiate low-density graupel from high-density hail and was shown to improve predictions of both hail size at the surface and reflectivity values in the upper levels of a convective storm system (Jensen et al., 2023a). In this study, both the "aerosol-aware" and "hail-aware" options were used.

2.1 Cloud droplet number concentration

The key benefit of the "aerosol-aware" parameterization is that it represents the effects of aerosols on cloud droplet properties. Aerosols affect cloud droplet size, which affects radiation, albedo, and warm-rain formation (Thompson and Eidhammer, 2014). The combined predictions of q_C and n_C also result in variability in the median volume diameter of cloud droplets



(MVD_C , Thompson and Eidhammer, 2014, their Fig. 6). This variability is a result of the dependence of MVD_C on both q_C and n_C , where

$$MVD_C \approx \frac{3.672 + \nu_C}{\lambda_C} \text{ and } \lambda_C = \left[\frac{n_C \pi \rho_w \Gamma(4 + \nu_C)}{6 q_C \Gamma(1 + \nu_C)} \right]^{1/3}, \quad (1)$$

115 where ρ_w is the density of water, Γ is the gamma function, ν_C is the shape parameter for cloud water that varies from 2 to 15 depending on n_C , λ_C is the slope parameter for cloud water, and π is the unending mathematical constant.

Note that when a constant value of n_C is used (as in the single-moment scheme), MVD_C depends only on q_C . In the two-moment scheme, a higher value of n_C produces smaller droplets (assuming constant values of q_C and ν_C). From Eq. 1, if $q_C = 0.2 \text{ g m}^{-3}$ and $\nu_C = 3$, then a value of $n_C = 250 \text{ cm}^{-3}$ yields $MVD_C = 16 \text{ }\mu\text{m}$, whereas a value of $n_C = 10 \text{ cm}^{-3}$ yields $MVD_C = 46 \text{ }\mu\text{m}$. One goal of using an ML model is to attempt to replicate variability in n_C without advection of n_C .

120

2.2 Cloud droplet effective radius

Cloud droplet effective radius (r_e) is a key variable for radiation and is calculated as a ratio of the volume to the cross-sectional area of cloud droplets, which for bulk size distributions ($n(D)$) can be written as half the ratio of the integrated cubed diameter (where D is the diameter) to the integrated squared diameter:

$$r_e = 0.5 \frac{\int_0^\infty D^3 n(D) dD}{\int_0^\infty D^2 n(D) dD} = 0.5 \frac{(3 + \nu_C)}{\lambda_C}. \quad (2)$$

125 Thus, r_e depends on both mass and number (Eq. 2) with a functional form similar to that of MVD_C .

In the HRRR, subgrid-scale clouds are diagnosed by the MYNN boundary-layer scheme (Nakanishi and Niino, 2009; Olson et al., 2019, 2026). Only the mass and cloud fraction of these unresolved or under-resolved clouds are diagnosed; therefore, r_e is set to a constant value—based on observations (Miles et al., 2000)—of $7.5 \text{ }\mu\text{m}$ over land and $10.5 \text{ }\mu\text{m}$ over water. This choice of r_e affects both radiation and model temperature biases (Benjamin et al., 2025). Thus, a second goal of using an ML model is to predict n_C for subgrid-scale clouds so that their r_e values are physically consistent with those of resolved clouds.

130



3 Neural network prediction of cloud droplet number concentration

Neural networks are commonly used ML tools that provide predictions based on learned statistical relationships in a dataset. A neural network is provided with training data that are used to characterize the relationships between user-defined input features and the target (predicted variable). New predictions can then be made using the trained model.

135 The target for this regression neural network was cloud droplet number concentration, n_C . The input features were local microphysical and environmental variables because microphysical processes are interdependent and modulated by the environment. Seven features were used: q_C , q_R , q_I , q_S , pressure (P), temperature (T), and vertical velocity (w). Water-friendly and ice-friendly aerosol number concentrations were not used as input features to ensure that the ML model could be used when the “aerosol-aware” parameterization is off. Furthermore, q_G was not used as an input feature. This simplifies the ML model
140 by reducing the number of input features and preventing the need for data from both the one-moment and two-moment graupel formulations in the training dataset. Given the four microphysical and three environmental input features, the addition of q_G would not be expected to significantly affect the results. Future work should build on this initial ML model development and explore the effects of additional input features.

3.1 Model setup to produce the ML training dataset

145 The training dataset was compiled from simulations using MPAS-A. MPAS-A is a fully compressible, non-hydrostatic model utilizing an unstructured mesh of mostly (but not necessarily exclusively) hexagons. Simulations were run using limited-area domains with 3-km horizontal mesh spacings and 59 vertical levels, with vertical spacings that increase with height. The MPAS-A vertical coordinate is terrain-following near the surface and transitions to smoothed levels higher in the atmosphere. A timestep of 20 s was used.

150 In addition to TEMPO microphysics with both aerosol-aware and hail-aware options set to true, the physics parameterizations used were similar to those used in the HRRR, including the MYNN surface-layer, boundary-layer, and subgrid-scale cloud schemes (Olson et al., 2021; Nakanishi and Niino, 2009; Olson et al., 2019); the Rapid Radiative Transfer Model for GCMs (RRTMG) longwave and shortwave radiation schemes (Mlawer et al., 1997; Iacono et al., 2008); the Rapid Update Cycle (RUC) land surface model (Smirnova et al., 2016; Smirnova and Benjamin, 2025); and NOAA’s unified gravity wave
155 physics (Toy et al., 2025).

Three simulations were run (using a fork of MPAS-A v8.2.2-2.0: Jensen, Anders A., 2025c) to produce a training dataset that encompassed multiple seasons and various types of clouds. The MPAS version numbering is broken into two parts: 8.2.2 tracks NCAR’s development, and 2.0 tracks development of the UFS Community fork. Additionally, two types of initial conditions (ICs, see Table 1) and two limited-area domains were used (Fig. 1). Various forecast lengths greater than 12 h were included
160 in the training data to ensure that the clouds and microphysics were sufficiently spun up. The three cases, described in Table 1, included a wintertime atmospheric river event, a springtime convective event, and Hurricane Debby.

These cases capture a wide range of the microphysical parameter space related to clouds and precipitation, including warm-rain processes in a tropical cyclone, warm-rain processes in an atmospheric river, liquid-ice interactions, supercooled clouds,



Table 1. Simulations used for the ML training dataset. ICs were from the Global Forecast System (GFS) and the Rapid Refresh Forecast System-A (RRFS-A). Domains are shown in Fig. 1. The “Training Dataset Forecasts” column shows the forecast hours included in the training data for each case.

Case	Initialization	ICs	Domain	Training Dataset Forecasts
Atmospheric River	0000 UTC 09 Jan 2023	GFS	Western US	12 h, 18 h
Springtime Convection	0600 UTC 15 May 2025	RRFS-A	CONUS	16 h, 21 h
Hurricane Debby	1200 UTC 05 Aug 2024	GFS	CONUS	13 h, 19 h



Figure 1. The two model domains used in this study. The blue box outlines the Western US domain, and the black box outlines the CONUS domain.

and deep convection. Cases with high aerosol concentrations (e.g., heavy pollution, significant smoke) were not explicitly included in this dataset because the current method targets the prediction of typical clouds and precipitation. Composite reflectivity from training dataset forecasts was compared with observations from the Multi-Radar Multi-Sensor (MRMS) dataset (Zhang et al., 2011) in Fig. 2. In general, the MPAS-A forecasts produced realistic representations of composite reflectivity compared with observations. The atmospheric river case (Fig. 2a, b) correctly represented the widespread stratiform precipitation across central California. The springtime convection case (Fig. 2c, d) produced the line of strongly forced precipitation across Minnesota and Wisconsin. The Hurricane Debby case (Fig. 2e, f) represented the approximate location of precipitation bands over South Carolina and scattered precipitation across the Western United States.

From the three simulations, a training dataset was created that contained 12 million data points with nonzero cloud water mixing ratios. The coexistence of cloud water in an MPAS-A cell volume with rain or snow occurs at nearly equal frequencies in this dataset: 27% for cloud water and snow and 36% for cloud water and rain. Thus, the dataset provides a balanced

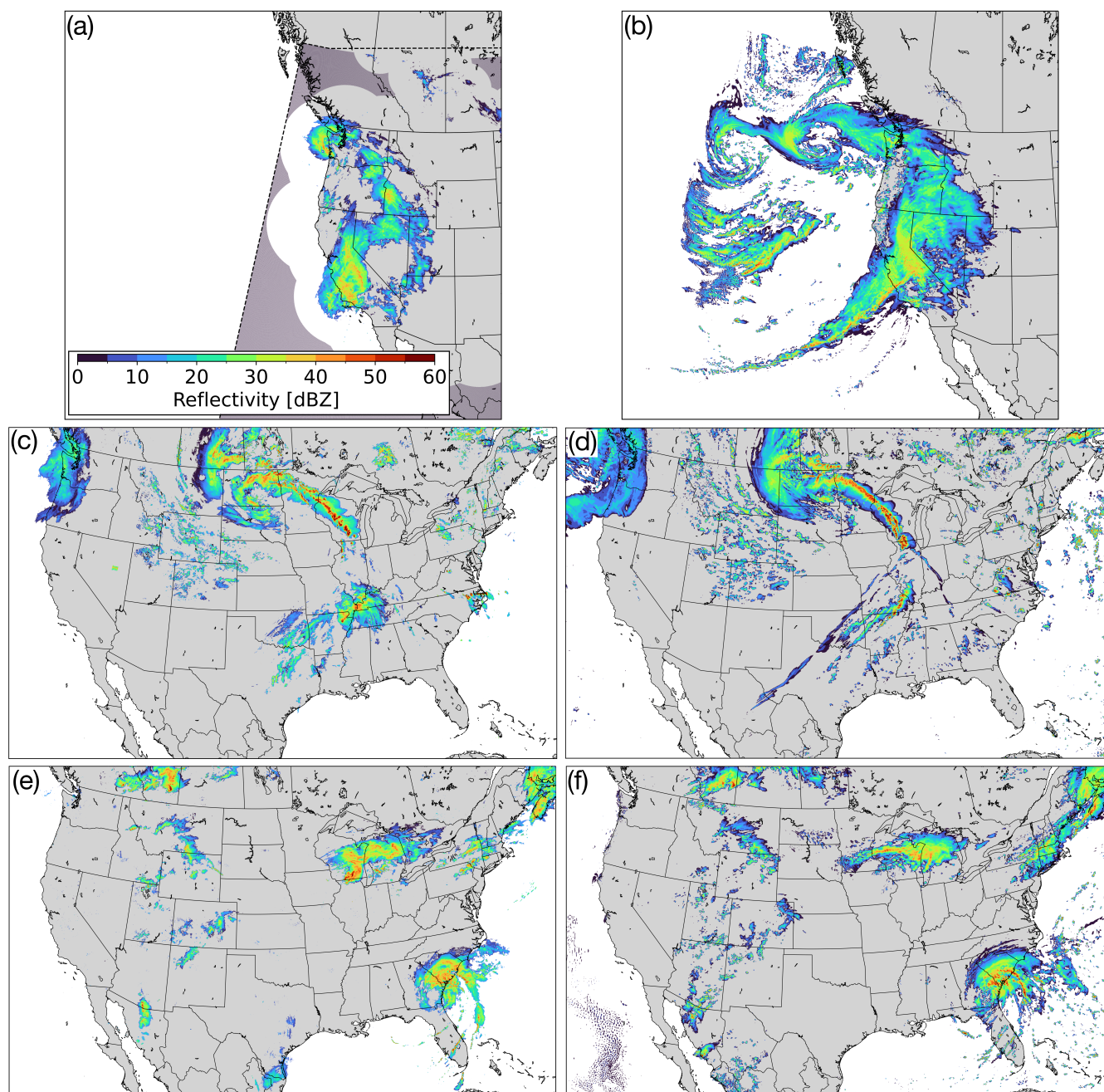


Figure 2. Composite reflectivity from the three cases in the training dataset. Panels (a), (c), and (e) are from MRMS, and (b), (d), and (f) are the MPAS-A forecasts for (from top to bottom) the atmospheric river (valid at 1800 UTC 09 Jan 2023), springtime convection (valid at 2200 UTC 15 May 2025), and Hurricane Debby (valid at 0700 UTC 06 Aug 2024) cases. The extent of the MRMS product in (a) is shown with the dashed black line, and areas without radar coverage are shaded gray, showing the extent of MRMS over the Pacific Ocean.



175 representation of mixed-phase and warm-rain conditions. The large dataset created from several types of forecasts provides confidence that the dataset appropriately represents most of the hydrometeor parameter space encountered by the microphysics model.

3.2 Training and evaluation

The neural network was trained using TensorFlow (Abadi et al., 2015). Each feature was first normalized by subtracting the mean and dividing by the standard deviation. A total of 20 ML models were trained using 3 epochs each and hyperparameter tuning that modified the learning rate, the kernel regularization, and the number of neurons. A rectified linear unit activation function was used. The model with the lowest mean squared error (MSE) for the target, $\log_{10}(n_C)$, was saved and used in this study. This model had a learning rate of 0.0009, 1 hidden layer, and 24 neurons.

Eighty percent of the dataset was used for training, and 20% was used to evaluate model performance (testing dataset). This evaluation is shown in Fig. 3. The ML model reasonably predicts n_C values across approximately six orders of magnitude (Fig. 3a), with a majority of the data falling between approximately 10 and 1000 cm^{-3} . The Pearson correlation coefficient between the ML-predicted and the actual (target) data is 0.92. The distribution of ML-predicted values has a higher normalized peak than that of the actual data (Fig. 3b). Thus, a correction with quantile mapping was applied. The quantile values from the ML-predicted data were mapped to the quantiles of the actual values, and the multiplicative mapping factor used for this correction is shown in Fig. 3c. A 5th-order polynomial was fit to this correction factor between the values of 1 and 1000 cm^{-3} . This limited range encompassed a large fraction of the microphysical parameter space and was used to ensure that a relatively low-order polynomial could be used to fit the quantile remapping factor robustly. Figures 3d, e show results after this correction factor was applied.

An evaluation of the ML model's ability to predict n_C for the effective radius calculation (see Eq. 2), which depends on both cloud droplet mass and number concentrations, along with correlations between key variables, is shown in Fig. 4. Figure 4a shows a comparison of r_e calculated using n_C predicted from the ML model and actual values of n_C . For this comparison, the mean bias (predicted minus actual) is $-0.05 \mu\text{m}$, and the root-mean-square error (RMSE) is $4.98 \mu\text{m}$. Importantly, the ML model captures similar parameter spaces to those of the actual data for q_C-r_e , q_C-n_C , and the spread in n_C across temperature values (Figs. 4b–d). Note that the ML model produced a stronger correlation between q_C and n_C than is present in the training dataset (Fig. 4c), likely due to some overfitting of the model.

3.3 Connection to TEMPO Microphysics

The ML model was trained offline, and the data necessary for the model, including input feature transformations and ML weights and biases, were saved and used in TEMPO microphysics. This approach is similar to the method used in the `mlmicrophysics` code (National Center for Atmospheric Research, 2026a) applied by Gettelman et al. (2021). For simplicity, the ML model data were saved in `module_mp_tempo_params.F90` (see e.g., Jensen, Anders A., 2025d). Additionally, an ML module (`module_mp_tempo_ml.F90`) was created that contains subroutines to initialize the ML model, perform variable transformations, apply activation functions, and generate predictions. Finally, the TEMPO driver (located in

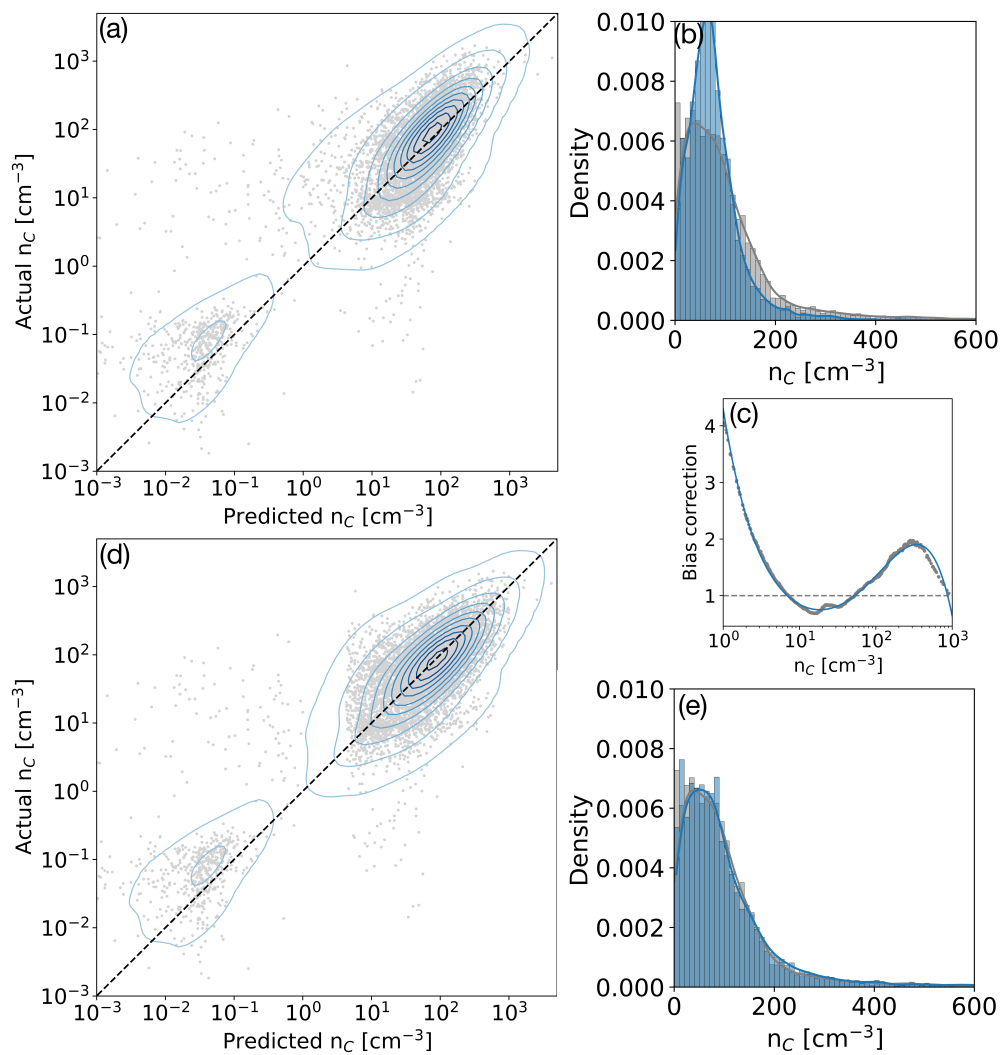


Figure 3. (a) ML-predicted versus actual (target) values of n_C . The points are from 10,000 random samples, and the contours are the kernel density estimates (KDE) starting at 5% and increasing by 10% intervals to 95%. The dashed black line is the one-to-one line. (b) Normalized distributions of n_C plotted for ML-predicted (blue) and actual (gray) values. (c) Quantile mapping correction factor (gray points) and a 5th-order polynomial fit to the correction factor (blue line). Panels (d) and (e) are the same as (a) and (b) but after using the quantile mapping correction.

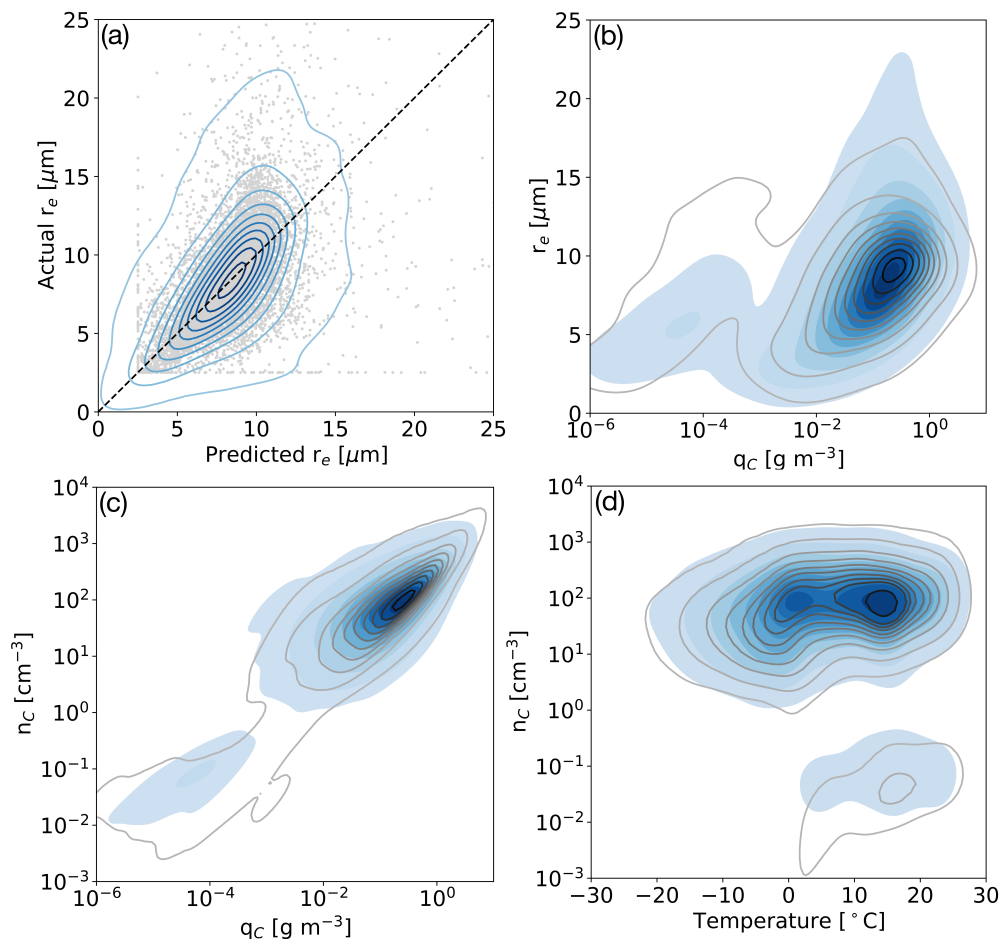


Figure 4. (a) ML-predicted versus actual (target) values of r_e . The points are from 10,000 random samples, and the contours are the kernel density estimates (KDE) starting at 5% and increasing by 10% intervals to 95%. The dashed black line is the one-to-one line. The lower limit of $2.5 \mu\text{m}$ is imposed by the radiation scheme and shown in this plot. KDE of (b) q_C versus r_e , (c) q_C versus n_C , and (d) temperature versus n_C . KDE are shown for the training dataset (filled blue contours) and the testing dataset (gray contour lines) using the same contour values as shown in (a).



drivers/mpas/), which is directly called by MPAS, was modified to include the subroutine that predicts n_C . When used to directly predict n_C , the ML model prediction occurred before the microphysics—to pass n_C to TEMPO—and also after the state was updated by the main microphysics routine. Subgrid-scale cloud water from the MYNN scheme was also incorporated into the TEMPO driver for r_e prediction. Future work will provide a more robust coupling of the ML model to MPAS by saving ML model parameters to external files readable by MPAS (or other models).



Table 2. Description of simulations.

Case	Description
CTRL	Control simulation
ML-RE	ML prediction of n_C to calculate r_e for subgrid-scale clouds
ML-NC	ML prediction of n_C in TEMPO microphysics

4 Observations and experiments

Three simulations, described in Table 2, were run to test TEMPO microphysics in MPAS-A and the ML model. These simulations coincided with the Winter Precipitation Type Research Multi-scale Experiment (WINTRE-MIX) field campaign. Additionally, during this time period, data from three ground-based infrared spectrometers (IRS) were available and used to retrieve r_e values. The control run (CTRL) was a baseline simulation without any ML additions. The ML-RE run used the ML model to predict n_C for subgrid-scale clouds to calculate r_e , which was compared with r_e retrieved from the IRS. This run used forks of MPAS-A v8.2.2-3.11 (Jensen, Anders A., 2025b) and TEMPO (Jensen, Anders A., 2025d). The ML-NC run used the ML model to predict n_C in TEMPO at every timestep, and the output was compared with airborne and surface observations from WINTRE-MIX. This run also used forks of MPAS-A v8.2.2-3.11 (Jensen, Anders A., 2025a) and TEMPO (Jensen, Anders A., 2025e). Note that ML-RE and ML-NC were independent; the ML-NC simulation did not include the ML-RE implementation.

4.1 Data from WINTRE-MIX

The WINTRE-MIX field campaign targeted sampling of wintertime precipitation using an array of observations from a Convair-580 aircraft (Bala and Wolde, 2022) and several surface-based platforms (Minder et al., 2023). WINTRE-MIX occurred from 01 February to 15 March 2022, with flights out of Ottawa, Canada. Observations used in this study from the aircraft were from the Cloud Droplet Probe (CDP, Nichman, 2022), which is a forward-scattering instrument that measures drop sizes between 2 and 50 μm . Liquid water content was also estimated from this probe and a Nevzorov sensor (Bliankinshtein, 2022). Because the Nevzorov probe measures total liquid water and the CDP measures liquid water specifically from cloud droplets, comparing the two can provide insight into how much liquid is partitioned between cloud water and rain. Surface-based precipitation and manual precipitation observations (Han and et al., 2022) were also used, including a Sentinel Hotplate at the Gault Canada Foundation for Innovation Climate Sentinels (CFICS) site (Girouard and et al., 2022) and a Parsivel Disdrometer at Sorel (Michelson, 2023).

4.2 Effective radius from ground-based infrared spectrometers

Two types of ground-based infrared spectrometers (IRS) were used in this analysis: the Atmospheric Emitted Radiance Interferometer (AERI, Knuteson et al., 2004) and the Atmospheric Sounder Spectrometer by Infrared Spectral Technology (ASSIST, Michaud-Belleau et al., 2025). These instruments were not associated with WINTRE-MIX. Instruments at three locations were used for this study: Columbia, Louisiana; a U.S. Department of Energy Southern Great Plains (SGP) observing site E39 (near



Morrison, Oklahoma); and Madison, Wisconsin. The basic measurements made by both of these instruments are essentially identical; both measure downwelling spectral infrared radiance between 3 and 19 μm at 1 cm^{-1} resolution approximately every 20 s. The radiometric accuracy of the observed radiance spectrum is better than 1% of the ambient radiance (Knuteson et al., 2004; Michaud-Belleau et al., 2025). These radiance observations provide information on both the thermodynamic profile, especially in the atmospheric boundary layer, and the properties of the clouds above the instrument. For this study, the focus is on the cloud liquid water path (LWP) and effective radius (r_e) of the overhead clouds, which were retrieved using the Mixed-phase Cloud Retrieval Algorithm (MIXCRA, Turner, 2005). Note MIXCRA needs thermodynamic profiles retrieved from the IRS using the Tropospheric Remotely Observed Profiling via Optimal Estimation (TROPOe) retrieval (Turner and Löhnert, 2014).

The retrieved LWP from IRS instruments, such as the AERI and ASSIST, is extremely accurate from 2 to 60 g m^{-2} ; above this threshold, the cloud is essentially opaque to infrared radiance (Turner, 2007). Comparisons of the retrieved LWP from multiple side-by-side IRS instruments demonstrated agreement in LWP to better than 2 g m^{-2} (Turner et al., 2026). The retrieved effective radius requires a significant signal from the cloud, yet the signal cannot be saturated; thus, the retrieved effective radius is accurate for LWP values between 5 and 45 g m^{-2} (Turner, 2007). Comparisons with collocated aircraft *in situ* observations demonstrated that the retrieved effective radius from IRS instruments agreed with the *in situ* observations within 1 μm (Turner, 2007).

4.3 Simulations and methods for comparisons with observations

The three simulations were run from 1200 UTC 06 March 2022 to 2000 UTC 07 March 2022 using initial conditions from the Rapid Refresh model (RAP, Benjamin et al., 2016). The RAP model was used because it resulted in robust clouds across several locations where the simulations were evaluated. Additionally, the RAP model contained initial conditions for n_C , which reduced the model spin-up time of that variable in the CTRL simulation. The simulations used the CONUS domain (Fig. 1). During the simulation period, two WINTRE-MIX Intensive Operation Periods (IOPs) occurred. IOP 8 occurred on 06 March 2022 from 1310 to 1712 UTC, and Flight 1 of IOP 9 (hereafter IOP 9) occurred on 07 March 2022 from 1611 to 1903 UTC. IOP 8 targeted clouds and precipitation associated with a warm-frontal passage, while IOP 9 targeted a warm-frontal passage with freezing rain (see Fig. 5d, e) and ultimately a transition to snow following a cold-frontal passage (Schiede, 2024). The simulations captured the composite reflectivity structure compared to MRMS observations including the broad region of precipitation along the cold front (Fig. 5a–c). Note that the MRMS product did not capture the broad region of reflectivity near and east of the Georgian Bay shown in the simulations, but those echos are apparent from the Environment Canada Radar Mosaic (not shown). Observations from the two research flights, along with surface observations, were used to evaluate CTRL and ML-NC. Ground-based IRS at three locations—Columbia, Louisiana; Madison, Wisconsin; and an Atmospheric Radiation Measurement Southern Great Plains (ARM–SGP) site in Oklahoma—were used to evaluate CTRL and ML-RE.

To compare the simulations with observations, several techniques were used to account for uncertainty in the timing and location of clouds. Point verification of clouds was used for comparisons with the IRS data. For this comparison, vertical profiles of model output were saved every 5 min at the MPAS-A mesh cell nearest to the observation location, alongside 10

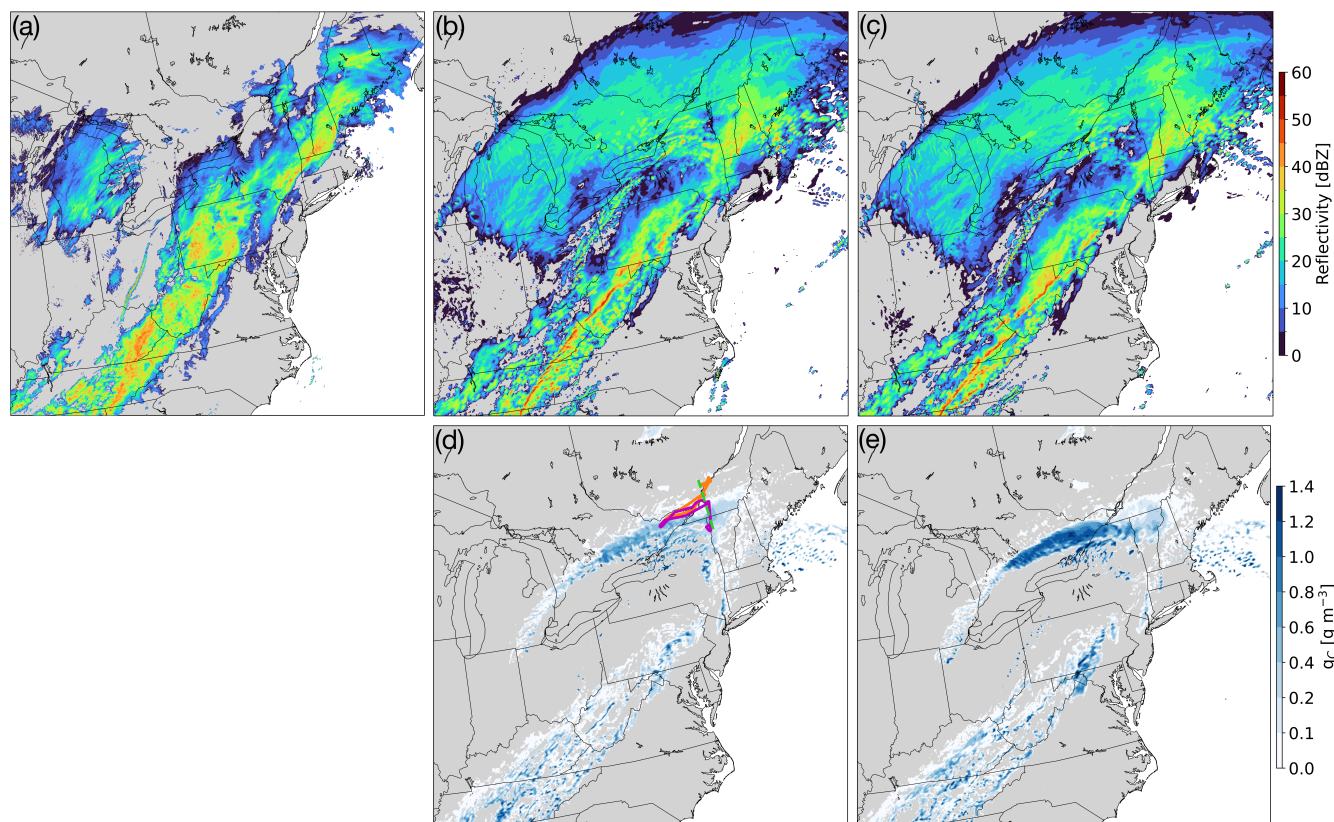


Figure 5. Composite reflectivity from (a) MRMS, (b) CTRL, and (c) ML-NC at 1800 UTC 07 March 2022, during IOP 9. Values of q_c at approximately 700 hPa and 1800 UTC 07 March 2022 from (d) CTRL and (e) ML-NC. IOP 8 (orange) and IOP 9 (magenta) flight paths and the cross section (dashed green line) along which model data are later plotted are shown in (d).

additional mesh cells within approximately 15 km of the observation location distributed as shown in Fig. 6a. For comparisons with aircraft observations, the mesh cell closest to the aircraft location at a given time, adjoining cells, and a second level of adjoining cells were used (Fig. 6b). For hexagons, the number of cells using this method is 19. Additionally, because the aircraft was generally located between model vertical levels, 19 mesh cells were used from two vertical levels—above and below the aircraft—resulting in a total of 38 cells in the comparison. For the evaluations in this paper, the aircraft observations were averaged to 30-s intervals, which, assuming an aircraft speed of 100 m s^{-1} , translates to a distance of approximately 3 km.

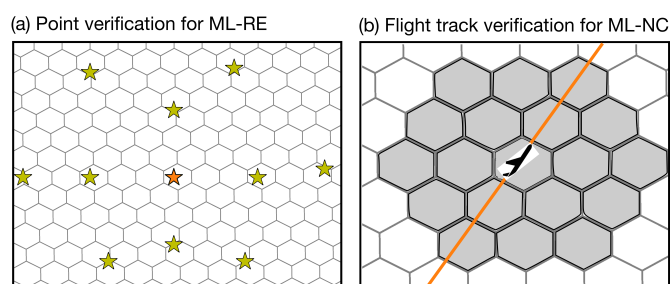


Figure 6. Methods used to compare model data with observations. (a) To compare model data with a point observation from the IRS at the location of the orange star, data were used from the model mesh cell nearest to the observation (orange star) and the 10 additional mesh cells nearest to the yellow stars, which surrounded the observation and accounted for uncertainty in the timing and location of clouds. (b) To compare model data with aircraft observations, data from the model mesh cell closest to the aircraft location at a given time, adjoining cells, and a second level of adjoining cells (shaded gray) were used.



280 5 Machine learning for subgrid-scale cloud effective radii

A map of r_e at approximately 1 km AGL is shown in Fig. 7 for CTRL and ML-RE. Variability in r_e occurs in CTRL where explicit clouds are predicted; otherwise, r_e is constant where subgrid-scale clouds reside, including over the Pacific Ocean, where r_e is set to $10.5 \mu\text{m}$ (the default over-water value), and across several Gulf states, where r_e is set to $7.5 \mu\text{m}$ (the default over-land value, Fig. 7a).

285 Figure 7b shows the impact of the ML model; r_e values vary across regions of subgrid-scale clouds, including the oceans and Gulf states. Generally, r_e values from ML-RE compared with CTRL are smaller at this altitude over the Pacific Ocean and slightly smaller over land, with some exceptions, such as over eastern Canada. Benjamin et al. (2025) showed that decreasing cloud effective radius values in the HRRR model reduced the downward shortwave bias; therefore, the reduction in r_e observed here would be expected to produce a positive overall impact on forecasts if results from this simulation generalize.

290 Values of r_e are compared with retrieved values of r_e at three locations in Fig. 8. Data from the simulations were extracted from 11 vertical profiles every 5 min at and around each location used for comparison (Fig. 6a). Values of r_e from these model profiles were “retrieved” in a manner consistent with the observational retrievals (see Appendix A).

Again, in the CTRL simulation (Fig. 8, blue), r_e values of $7.5 \mu\text{m}$ mean that subgrid-scale clouds are the dominant cloud type at a given location and time, and variability in r_e results from the existence of fully resolved clouds. At Columbia (Fig. 8a),
295 there was a relatively high percentage of subgrid-scale clouds in the model, as seen by the prevalence of r_e values of $7.5 \mu\text{m}$ across several hours of the simulation. This produced a narrow distribution of r_e . The ML-RE simulation resulted in an improved distribution of r_e (Fig. 8a, orange) compared with observations. At this location, the ML-RE simulation did a better job of capturing the relatively higher r_e values near $10 \mu\text{m}$, though the overall distribution was still narrower than observed.

Results were mixed at the other two locations. At Madison (Fig. 8b), the observed sizes were much smaller than predicted.
300 The ML-RE simulation exhibited a bump in the distribution near 5 to $6 \mu\text{m}$, capturing some of these smaller sizes. A higher percentage of clouds at this location were fully resolved by the model, since the CTRL simulation had relatively fewer $7.5\text{-}\mu\text{m}$ values. The ML-RE simulation had better variability at approximately 0000 UTC 07 March, also capturing the relatively larger r_e values observed about an hour later. At SGP (Fig. 8c), ML-RE resulted in a wider distribution of r_e in better agreement with the observations than that of CTRL. ML-RE captured the tail of small sizes below $5 \mu\text{m}$, but it did not capture the full tail of
305 larger values up to $15 \mu\text{m}$.

Overall, the ML-RE simulation revealed that the ML model could improve the representation of subgrid-scale clouds and produce reasonable r_e values, thus eliminating the need for self-imposed values of 7.5 and $10.5 \mu\text{m}$. A longer-term evaluation of the effects on radiation and temperature is left to future work. While the evaluation of the ML-RE simulation was limited to one case, the evaluation provides a proof of concept and insight into the appropriateness of the method.

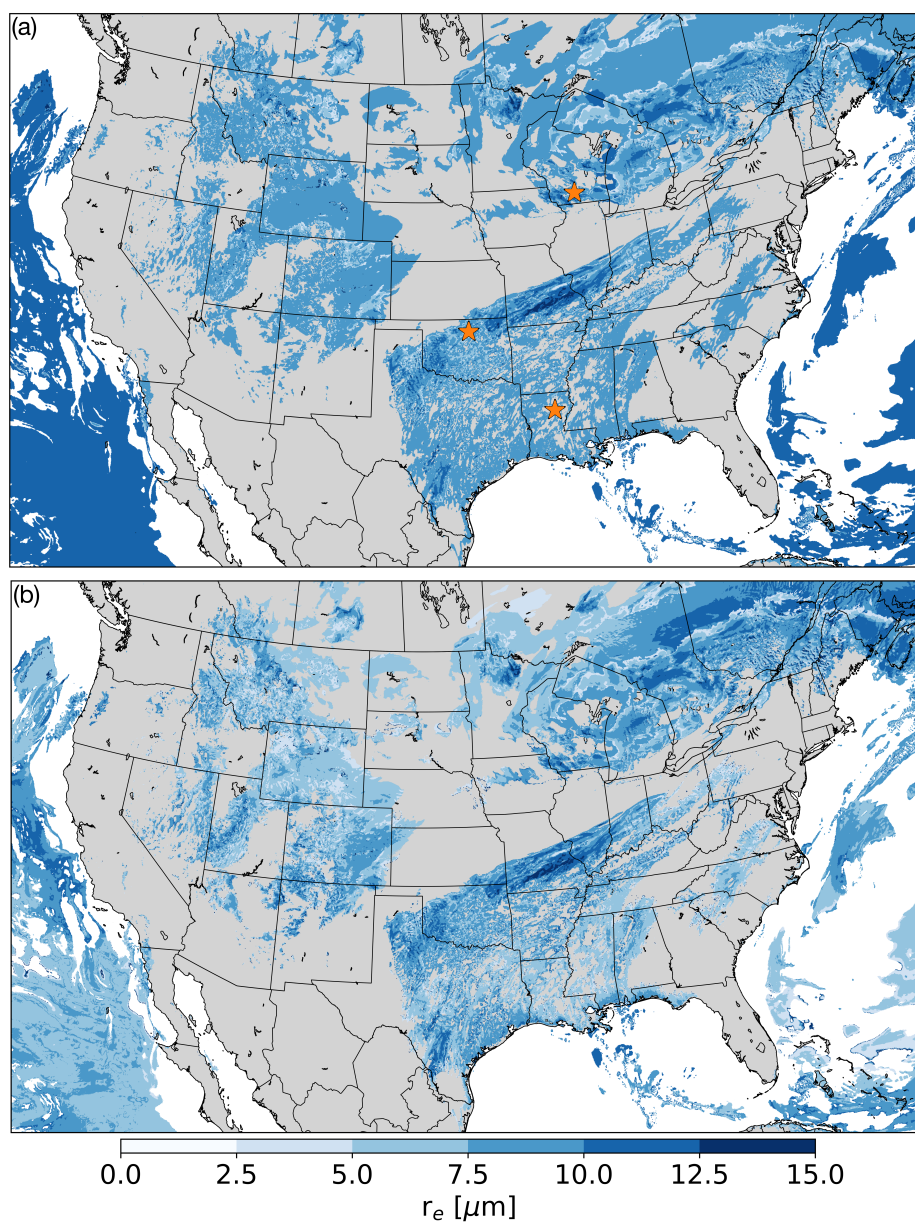


Figure 7. Values of r_e from (a) CTRL and (b) ML-RE. Values are shown at approximately 1 km AGL at 1800 UTC 06 March 2022. The orange stars in (a) mark the locations (from south to north) of the IRS instruments at Columbia, Louisiana; SGP E39 near Morrison, Oklahoma; and Madison, Wisconsin.

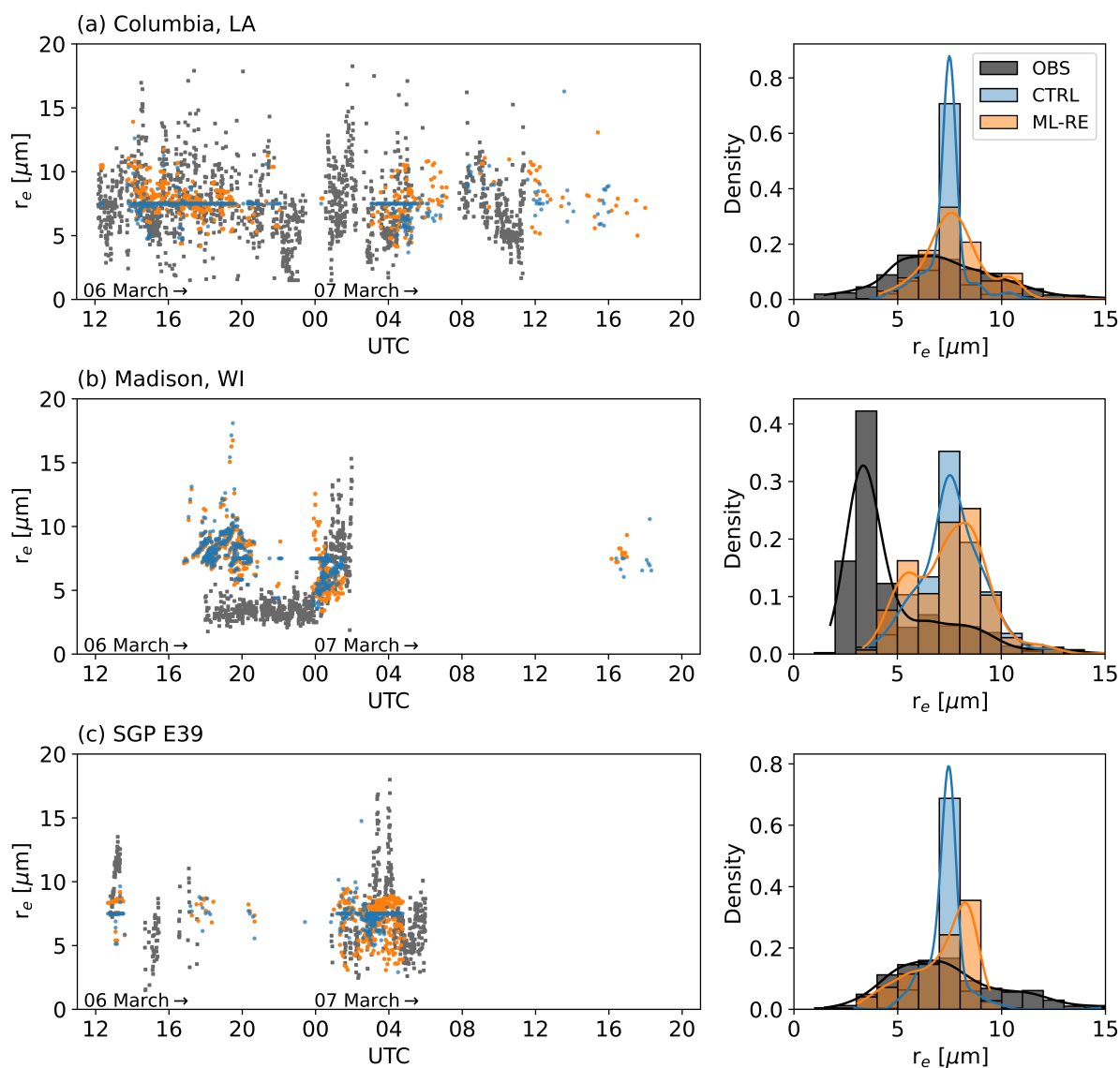


Figure 8. Time series of r_e at (a) Columbia, LA; (b) Madison, WI; and (c) SGP E39 (Morrison, OK). Retrieved values (gray), values from CTRL (blue), and values from ML-RE (orange) are shown. Normalized distributions and KDE (lines) are also shown for each location.



310 6 Machine learning for cloud droplet number concentrations

The CTRL and ML-NC simulations were evaluated for their ability to represent both cloud properties and precipitation compared with in-situ observations. The simulations were evaluated for both WINTRE-MIX IOP 8 and IOP 9 to provide insight into the effect of the ML model for two forecast lengths. IOP 8 coincided roughly with forecast hours 1.5 to 4, and IOP 9 coincided roughly with forecast hours 29 to 31. The CTRL simulation served as the benchmark. Furthermore, the evaluation of CTRL also functions as an important wintertime verification of TEMPO in MPAS-A.

Vertical profiles of temperature, q_C , and n_C were used to evaluate CTRL and ML-NC alongside aircraft observations. Figure 9 shows vertical profiles of model output and observations for IOP 8 across three arbitrary flight periods ranging in duration from 30 to 45 min. Vertical profiles from the model were output every 5 min from several model locations along the flight path (indicated by the white points on the maps in Fig. 9) for each time period. Note that time periods 1 and 3 included missed approaches; therefore, the observations extend lower in the atmosphere. The model output strategy was designed to capture both time and space variability. For example, time period 1 was 45 min (from 1345 to 1430 UTC) and included 15 model cell locations (Fig. 9a, white points), which resulted in 10 output times (at a 5-min output frequency) and 15 locations that provided a total of 150 vertical profiles available for that figure. q_C at 700 hPa and 1500 UTC from the CTRL simulation is shown in Fig. 9b, which is roughly the middle of the analysis time period.

Temperature and q_C profiles were similar between CTRL and ML-NC for all IOP 8 time periods (Fig. 9). Both simulations used the same initial conditions, so this result is expected, given that these are relatively short forecasts (initialization was 1200 UTC). Both simulations exhibited a slight warm bias below 800 hPa and a high q_C bias, although both simulations captured the vertical locations and extent of q_C compared with the observations. Differences arose when evaluating profiles of n_C . The ML-NC simulation produced vertical profiles of n_C that were slightly closer to the observations, albeit with a high bias, for time periods 1 and 2, whereas CTRL exhibited a larger high bias. By time period 3 (3.5 h into the forecast), CTRL matched the observations particularly well below 900 hPa.

Time series plots provide another method for understanding how the different simulations evolve compared with observations (Fig. 10). At each flight location (using the 30-s averaged flight data), an array of 38 model points surrounding the mesh cell closest to the aircraft was used for evaluation (this methodology is illustrated in Fig. 6b). The model output frequency was 30 min; therefore, each output file is valid for a 30-min period from 15 min before to 15 min after a given valid time. For example, flight data from 1415 to 1445 UTC were compared with model output valid at 1430 UTC.

Consistent with results from Fig. 9, the time series of q_C from the CTRL and ML-NC simulations were similar, and both exhibited a positive mean bias compared with CDP observations (Figs. 10a, b). Values of n_C were different between the simulations. ML-NC had a mean bias that was about half that of CTRL. However, both biases were positive since both simulations predicted cloud droplet number concentrations that were higher than observed (Figs. 10c, d). Comparing scatter plots of q_C and n_C , the ML-NC simulation did a reasonable job of representing the observations, but the spread in n_C as a function of q_C was notably smaller than that in the CTRL simulation (Figs. 10e, f).

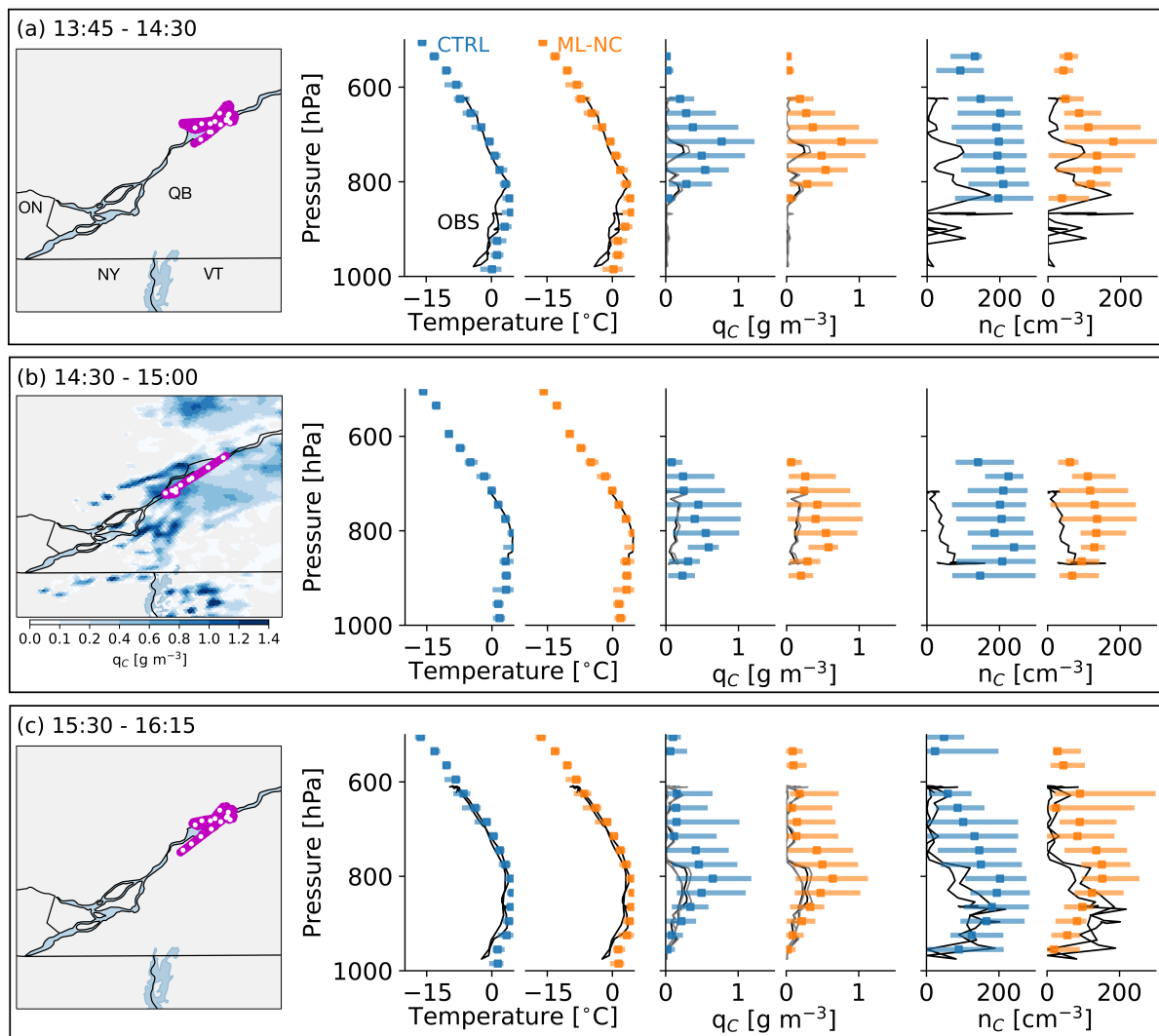


Figure 9. Vertical profiles of temperature, q_C , and n_C for IOP 8 along arbitrary flight time periods (a) time period 1: 1345 to 1430, (b) time period 2: 1430 to 1500, and (c) time period 3: 1530 to 1615. Values from CTRL are in blue, ML-NC in orange, and observations in black (Nevzorov liquid water content is in gray). The maps show the aircraft location (magenta) and the model cells where vertical profiles were extracted (white points). Time period 1 included 15 model locations, period 2 included 9 locations, and period 3 included 14 locations. Vertical profiles from the model were output every 5 min. q_C at 700 hPa from CTRL is shown on the map in (b). New York (NY), Vermont (VT), Ontario (ON), and Quebec (QB) are labeled on the map in (a).

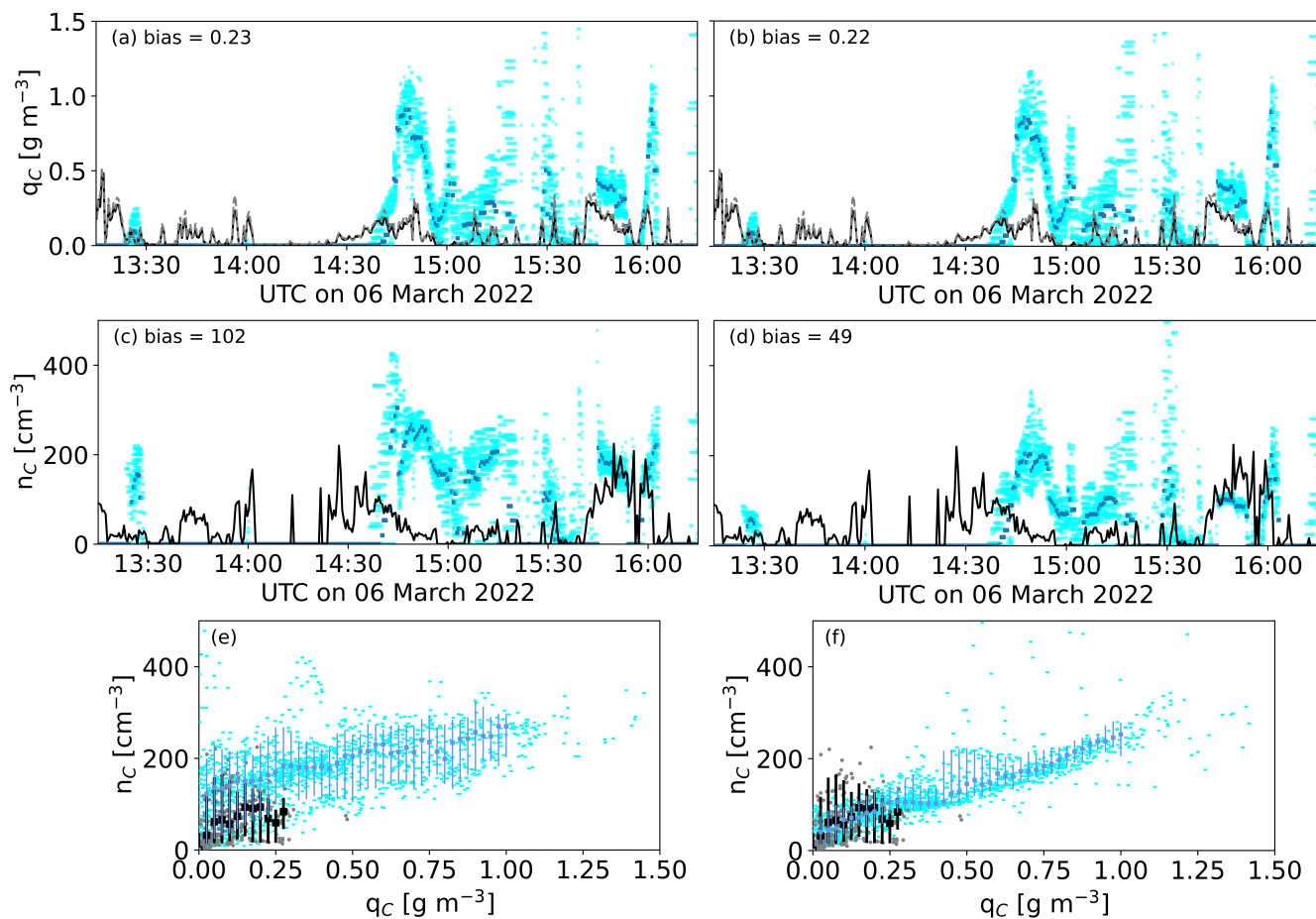


Figure 10. Comparison of q_C along the IOP 8 flight track for (a) CTRL and (b) ML-NC, and comparison of n_C for (c) CTRL and (d) ML-NC. The 38 model data points (cyan), model median values (blue), CDP observations (black), and Nevzorov observations (gray) are shown. Panels (e) and (f) show scatter plots of q_C – n_C for CTRL and ML-NC, respectively. Observations are the gray points, and model data are the blue points. Data are also binned in 0.025 g m^{-3} increments, and the 10th to 90th percentiles (bars) and median values (squares) are shown for observations (black) and the simulations (blue).



The same analysis was performed for IOP 9, which provided the ability to examine longer forecasts. Vertical profiles (Fig. 11) of temperature revealed that both simulations captured much of the complexity, although neither simulation fully represented the temperature inversion below 900 hPa during time period 3. At these longer forecast times, there was divergence between CTRL and ML-NC. For example, vertical profiles of q_C and n_C were different for time period 2 (Fig. 11b). ML-NC produced lower values of q_C at the flight levels (700 hPa) in better agreement with the observations, whereas CTRL produced lower values of n_C , better matching those observations. At 1830 UTC (Fig. 11c), both simulations produced too much q_C above 800 hPa, and ML-NC produced n_C values that were much too large. In this instance, the CTRL simulation better represented the vertically constant values of n_C , albeit the values were too high.

The ML-NC simulation produced a better prediction of q_C along the flight path before 1815 UTC, and overall, this simulation had a lower mean bias than CTRL (Figs. 12a, b). Clearly, by these longer forecast times, the differences in n_C resulted in different evolutions of clouds. ML-NC produced larger values of q_C to the west along the frontal boundary, which entered the flight path area later in the simulation just before 1830 UTC (see also Figs. 5d, e). These large q_C values correlated with a large increase in n_C that was much larger than observed. Overall, for this forecast period, CTRL had a lower mean bias for n_C (Figs. 12c, d). Importantly, both simulations represented the relatively low values of n_C when q_C values were low, early in the time period.

Scatter plots again revealed that CTRL produced a much larger spread in n_C , whereas ML-NC again exhibited less spread and a strong correlation between q_C and n_C (Figs. 12e, f). Thus, for ML-NC, high biases in q_C directly resulted in high biases in n_C , especially for $q_C > 0.1 \text{ g m}^{-3}$. Though note that the x-axis scales are different for both plots to show the extent of the q_C values for both simulations. Overall, the CTRL simulation was more representative of the observations.

Results from this analysis suggest that the ML model has some utility for improving short forecasts given that the model was trained on spun-up clouds; model spin-up of n_C is already baked in. For longer forecasts, the bias characteristics differed from those of the physical parameterization itself. The ML model produced less spread in n_C than CTRL, and the ML model was less representative of the observations for $q_C > 0.1 \text{ g m}^{-3}$.

Vertical cross sections from model output roughly along the flight path provide a deeper look into the microphysical differences between CTRL and ML-NC. While both simulations produced similar cloud fields in terms of spatial extent at this time, CTRL produced more rain farther north than ML-NC (Fig. 13). CTRL also produced lower values of n_C near cloud top, a trend that was reversed in ML-NC. Thus, the correlation between q_C and n_C in the ML model resulted in smaller drops at cloud top and less drizzle production (Fig. 13). Note that abundant freezing rain was observed by the aircraft for this IOP (Schiede, 2024). The ML model is not designed to predict physical processes, but the ML model undoubtedly affects physical processes that depend on n_C (or cloud droplet size).

The microphysical differences between CTRL and ML-NC resulted in some differences in precipitation (Fig. 14). The CTRL simulation yielded higher rain accumulations farther north in better agreement with observations of precipitation rates. Both simulations produced a transition to drizzle at Gault at 1930 UTC, about 1.5 h after rain was manually reported there, with CTRL producing higher rates in better agreement with observations. Overall, both simulations produced reasonable total precipitation rates at Gault compared with observations and similar transitions from snow before 1830 UTC to graupel (note

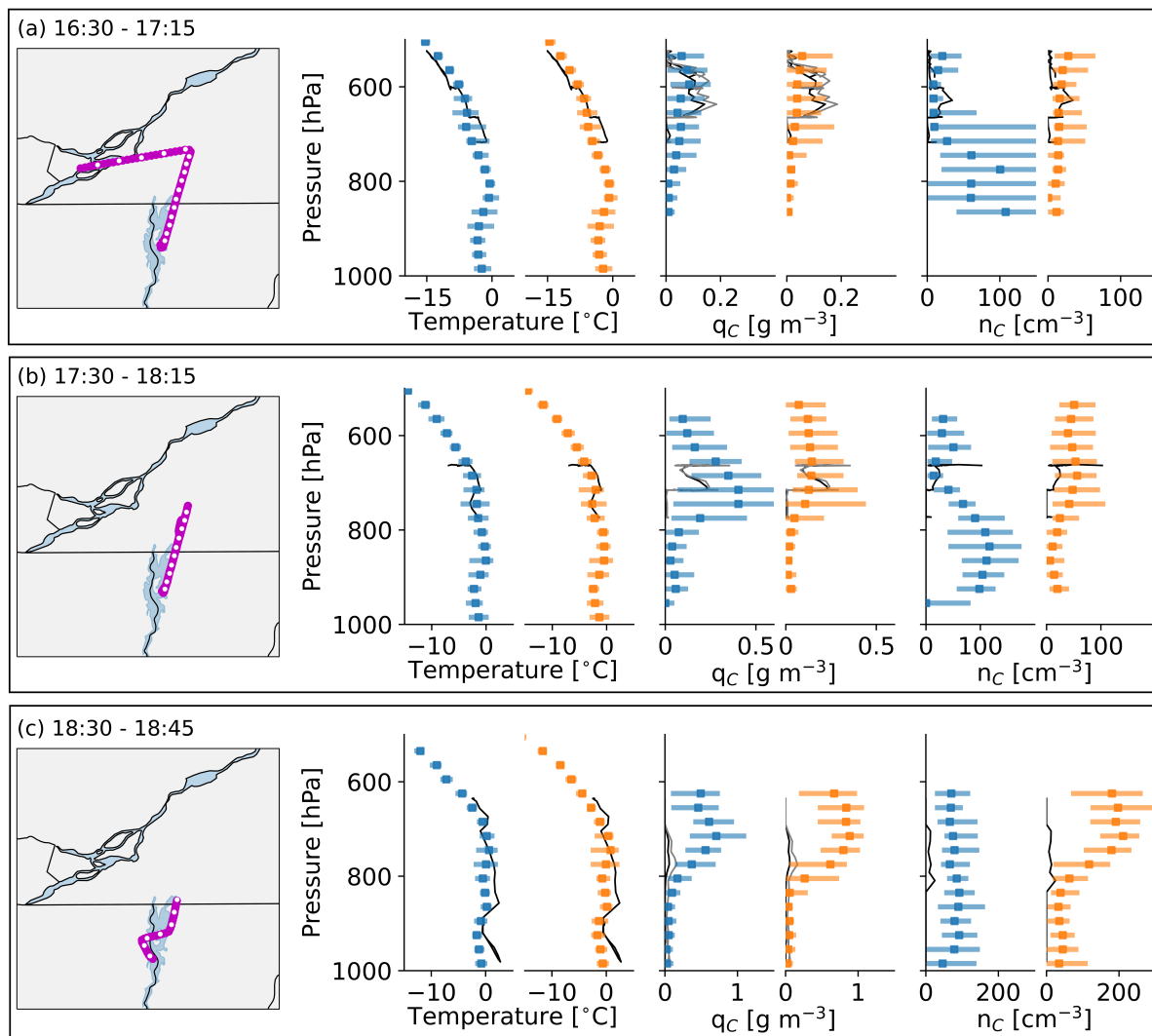


Figure 11. Same as Fig. 9 but for IOP 9. Time period 1 (1630 to 1715) included 16 model locations, period 2 (1730 to 1815) included 10 locations, and period 3 (1830 to 1845) included 7 locations. See Fig. 5d, e for q_c at 700 hPa.

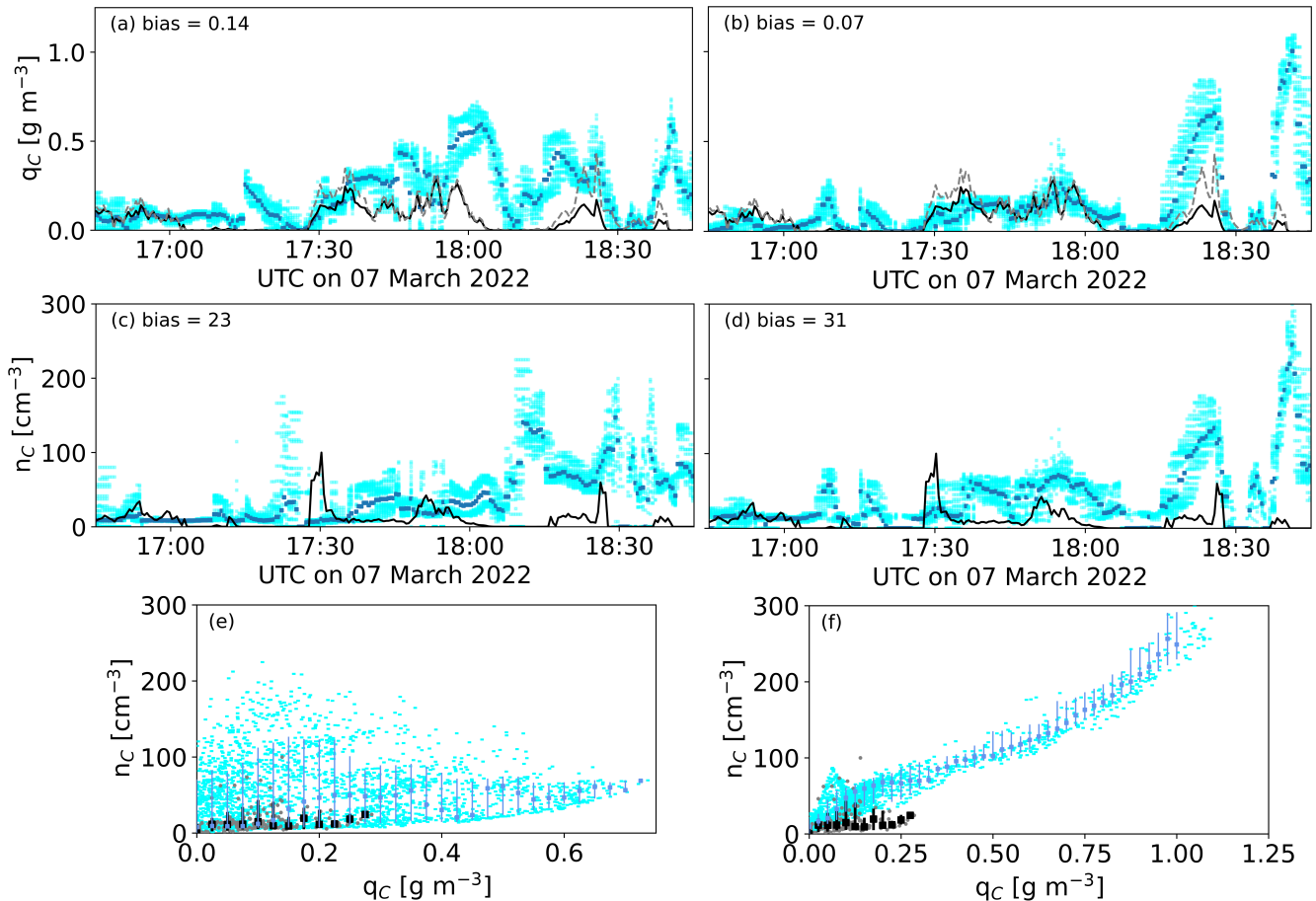


Figure 12. Same as Fig. 10 but for IOP 9.

that ice pellets are represented in the microphysics as graupel) and rain after 1930 UTC. Some of these phase transitions, like the transition from freezing rain to ice pellets, are sensitive to details of the microphysics parameterization (Han et al., 2026), and thus, feedbacks from changes to cloud droplet properties are not unexpected.

Both simulations reproduced total precipitation rates at Sorel, which is north of Gault, that roughly matched the observations. The ML-NC simulation produced slightly larger snowfall rates, but the onset of snow was similar between both simulations and matched the manual observations (Fig. 14). Both simulations also captured the transition from snow to ice pellets (graupel) at Sorel at approximately the correct time, about 1930 UTC.

The overall precipitation differences between CTRL and ML-NC were evaluated with the total 24-h precipitation from 1200 to 1200 UTC 06 to 07 March 2022. Both simulations produced similar overall precipitation patterns and generally captured the main precipitation features seen in NOAA's Stage IV observations (Fig. 15). There were some differences in the spatial distributions, with CTRL producing higher amounts across southern Illinois, which better matched the observations.

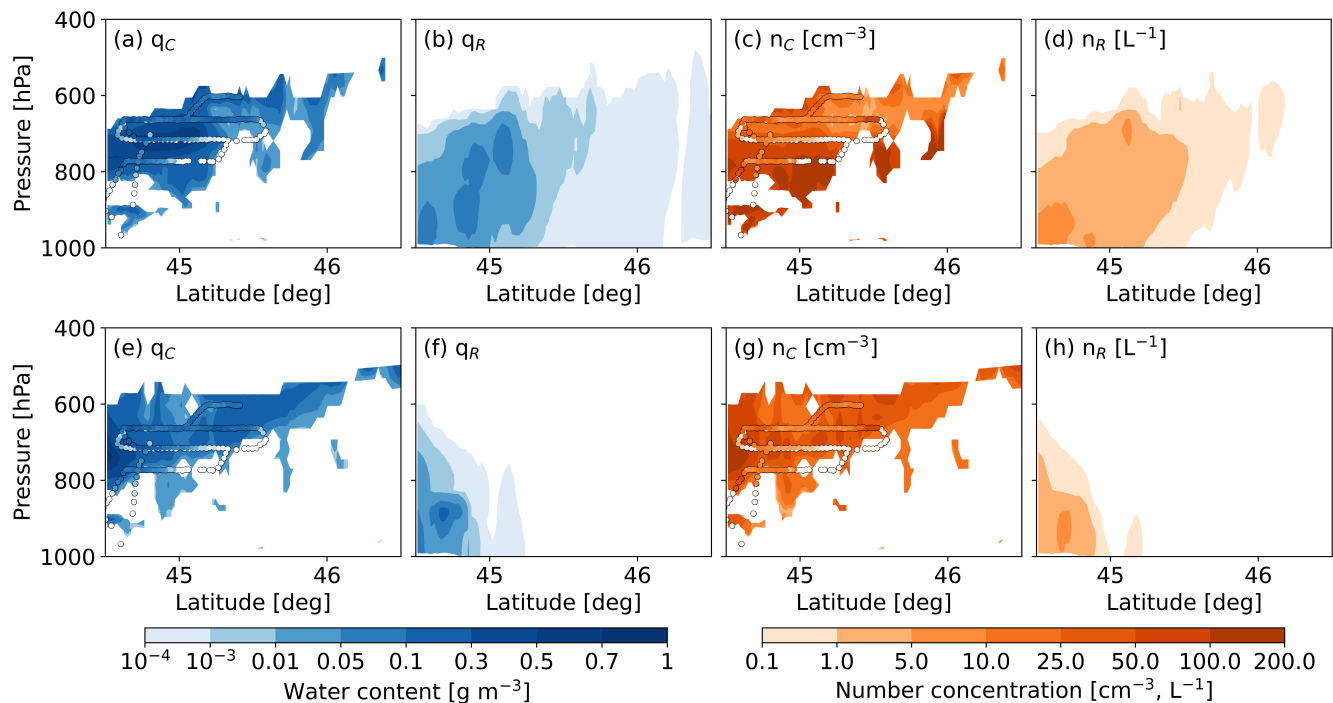


Figure 13. Cross section of (a) q_C , (b) q_R , (c) n_C , and (d) n_R from CTRL along the dashed green line shown in Fig. 5d at 1800 UTC 07 March. Panels (e), (f), (g), and (h) are the same but from ML-NC. Observations from the CDP are shown as circles, with the shading denoting the observed value.

Comparing the normalized distributions of precipitation (see Appendix B), the two simulations were closely matched, and both represented the observations well (Fig. 15d) with similar error distributions (Fig. 15e). For a bulk metric such as 24-h precipitation, the ML model closely matched the CTRL simulation.

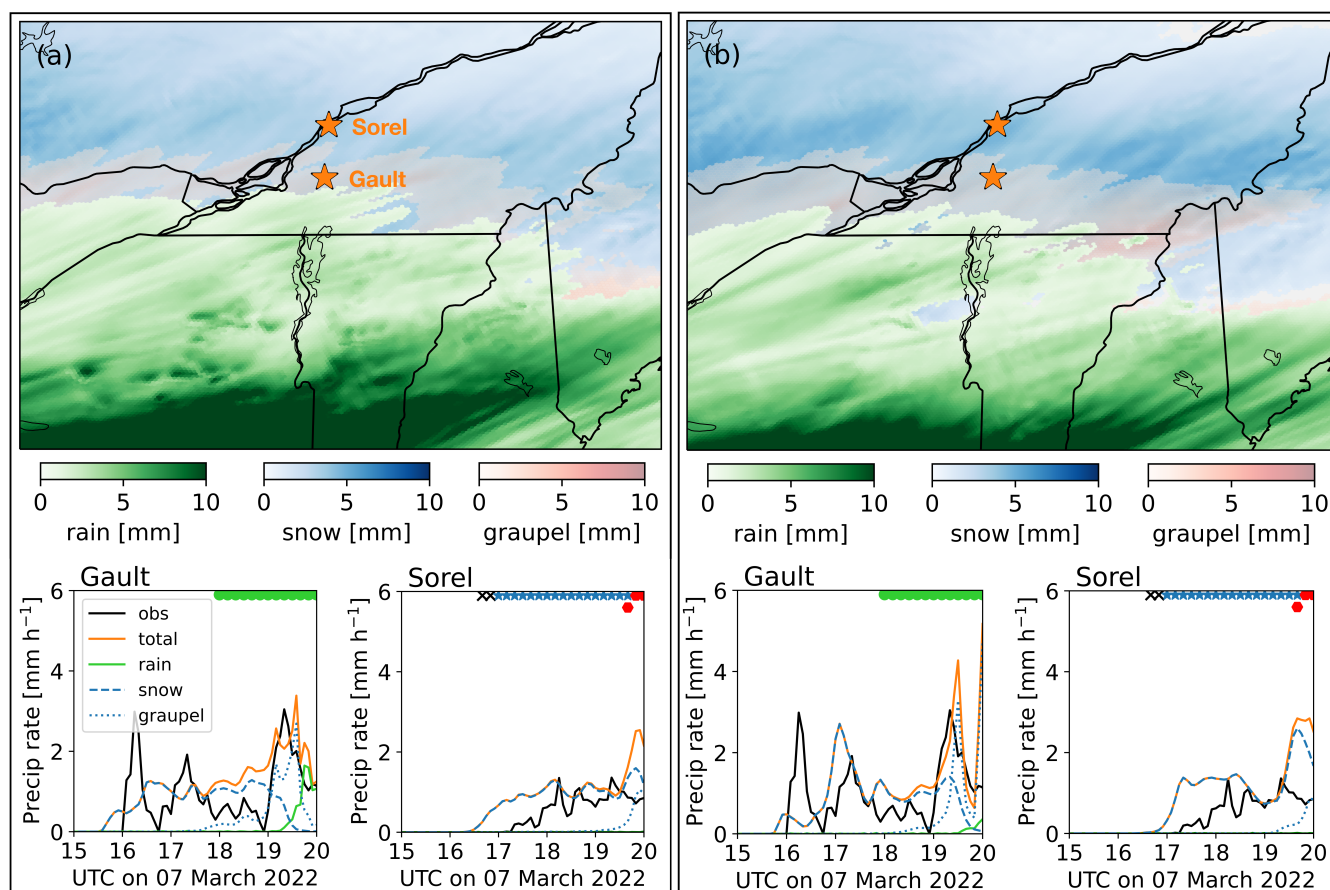


Figure 14. Precipitation from 1200–2000 UTC 07 March 2022 from (a) CTRL and (b) ML-NC. Accumulated amounts for rain (green), snow (blue), and graupel (red) are shown. Five-minute precipitation rates at Gault and Sorel are also shown. Locations are marked by the orange stars on the maps. Manually reported observations are also shown for rain (green circles), no precipitation (black exes), snow (blue stars), and ice pellets (red hexagons).

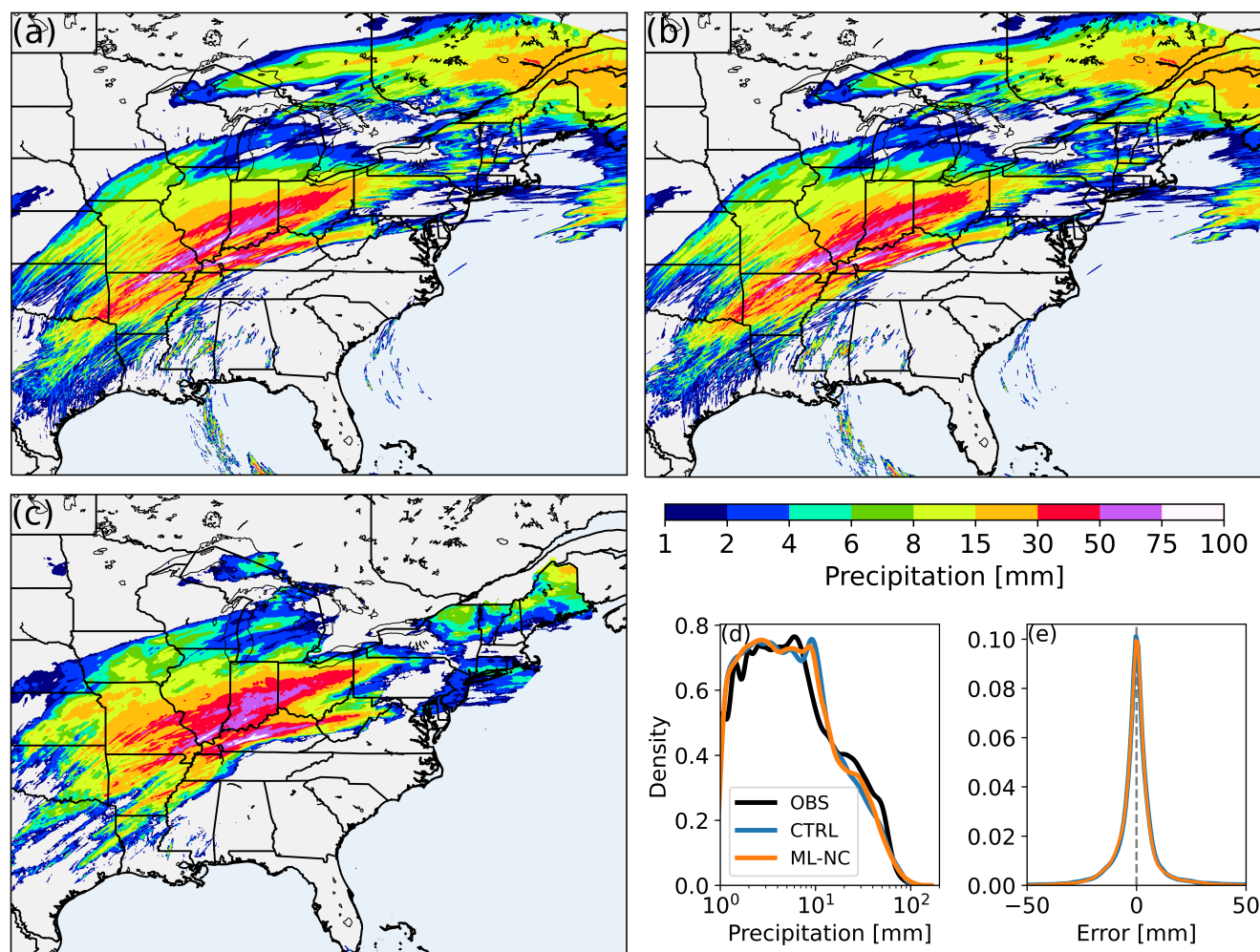


Figure 15. 24-h (1200 to 1200 UTC 06 to 07 March 2022) total liquid-equivalent precipitation from (a) CTRL, (b) ML-NC, and (c) Stage IV observations. Also shown are normalized (d) distributions and (e) error distributions (model minus observation). Model data plotted in (d) and (e) are regridded and masked to the Stage IV grid and area (see Appendix B).



7 Summary

A machine-learning model was developed for the Thompson-Eidhammer Microphysics Parameterization for Operations (TEMPO) to predict cloud droplet number concentration. The ML model was trained on data from several forecasts using MPAS-A. The ML model was tested for two use cases. First, the model was used to predict number concentrations for subgrid-scale clouds to compute r_e . The CTRL simulation used constant values of r_e , which is currently used in the operational HRRR. The ML model produced r_e values for subgrid-scale clouds that were physically consistent with values calculated for fully resolved clouds. These values were compared with retrieved values at three locations, using a new method to derive values of r_e from the model that were consistent with the observations. The key improvement demonstrated by the ML model was that the overall distributions of r_e were wider and more realistic. Further work to improve the representation of subgrid-scale clouds in models could leverage this method so that improvements to subgrid-scale parameterizations could be verified with observations.

Second, the ML model was tested for its ability to predict n_C in TEMPO microphysics. For this case, the ML model improved the prediction of n_C for short forecasts, in part, because the ML model was trained on longer forecasts with already spun-up clouds. Longer forecasts favored the CTRL simulation, which better maintained the relatively low values of n_C that were observed during IOP 9. CTRL better represented drizzle and rain production across the IOP region when compared with aircraft and surface observations. Furthermore, the CTRL simulation highlighted the strong performance of TEMPO microphysics in MPAS-A for a wintertime case at convection-allowing scales. While the ML model produced reasonable predictions of n_C , there was a high bias in n_C (due in part to a strong correlation between q_C and n_C from the ML model). This high bias in n_C that was more prevalent later during IOP 9 was coincident with less q_R aloft and at the surface at Gault compared with CTRL. Cloud droplet number concentration affects drizzle formation; these biases were likely linked. Overall, the 24-h total precipitation was nearly identical between CTRL and ML-NC. While the overall bulk precipitation statistics were appropriately reproduced with an ML model, it was important to understand some of the microphysical details and differences that resulted when using the ML model as shown here.

Machine-learning methods will continue to have utility in weather models. The push for higher spatial resolution—both globally and regionally—and the push for longer timesteps could benefit from fewer advected variables. Thus, this method could be extended to include number concentrations for additional hydrometeor categories. As additional model components are replaced with machine learning, models will become more efficient, allowing additional features and details to be included in operational forecast systems. As an additional advantage of this method, data assimilation systems could also be more closely coupled to microphysics using methods such as this one, with the goal of improving consistency between physics and data assimilation components. Updates to number concentrations from data assimilation are not well-constrained by observations; this method could help better link changes from data assimilation to changes in microphysical properties. While the method described here is relatively simple compared with other ML work, the benefits of this simplicity are that model tuning is still possible and ML model retraining is straightforward and efficient. In the future, observations could be incorporated into training datasets with the goal of improving the representation of targeted cloud types, such as stratocumulus.



425 *Code and data availability.* MPAS and TEMPO are publicly available through GitHub. Python code for simulations, analysis, figure gen-
eration, and IRS data are on Zenodo (<https://doi.org/10.5281/zenodo.21081016>, Jensen, 2026). Data from the WINTRE-MIX campaign is
available at https://data.eol.ucar.edu/master_lists/generated/wintre-mix/. Stage IV precipitation data are publicly available at <https://water.noaa.gov/about/precipitation-data-access>. MRMS data are available at <https://registry.opendata.aws/noaa-mrms-pds/>. RAP and GFS data are
also available at <https://registry.opendata.aws/noaa-rap/> and <https://registry.opendata.aws/noaa-gfs-bdp-pds/>, respectively. RRFS-A data will
430 be made available on request.



Appendix A: Calculation of vertically weighted effective radius from model output

Effective radius retrievals from the IRS depend on the cloud optical properties and the optical properties of the instrument, including the transmission wavelength (Turner, 2007). A single r_e value at a given MPAS-A mesh cell location is needed to compare with observations; this value is calculated as a transmission-weighted value to mimic the retrieved observation.

435 The method takes into account the transmission from the surface and the optical properties of the cloud layer and produces a normalized weighting function (Fig. A1e) that sums to one, is multiplied by the profile of r_e , and is summed to calculate a column-weighted average value of r_e . Values of r_e are only calculated for liquid water paths (LWPs) between 5 and 40 g m⁻² because this is approximately the valid range for the retrieval.

Appendix B: Precipitation regridding

440 To compare precipitation from MPAS-A simulations with Stage IV observations, two steps were taken. First, a tool called `convert_mpas` (Duda, Michael G., 2025) was used to convert the native MPAS-A grid to a latitude-longitude grid with grid center-point values defined. Second, `ESMPy` (ESMF Organization, 2025) was used to bilinearly regrid the latitude-longitude data to the Stage IV grid. This two-step regridding process was chosen to simplify the comparison with observations, even though it results in small changes to the precipitation characteristics (Fig. B1).

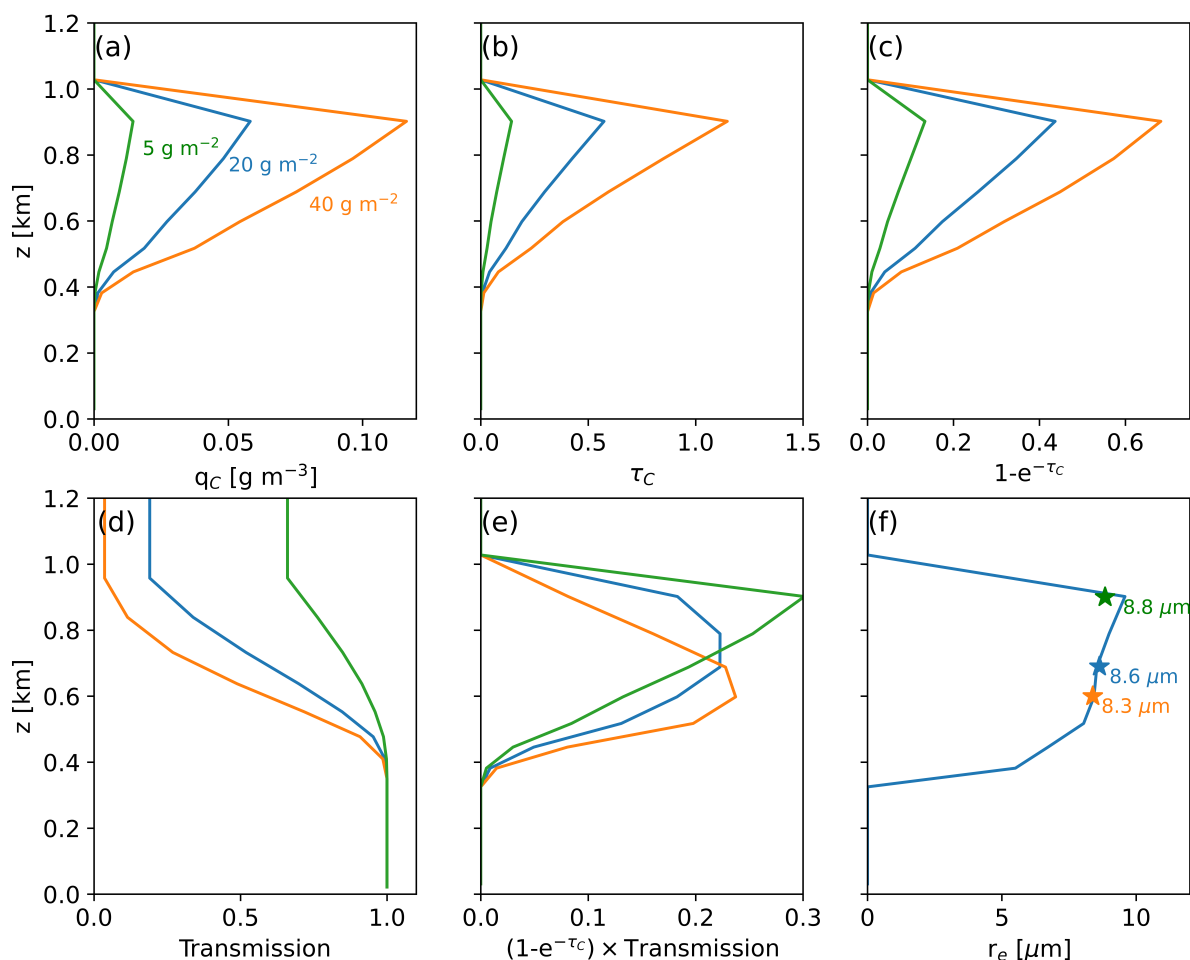


Figure A1. Example vertical profiles of (a) q_C , (b) cloud optical depth (τ_C), (c) cloud absorptivity, (d) approximate transmission at $11 \mu\text{m}$ (the center of the IRS spectral wavelength range used), (e) the weighting function, and (f) effective radius. The example profile from model data is shown in blue with an LWP of approximately 20 g m^{-2} . The q_C values from this profile were then divided by 4 (green) and multiplied by 2 (orange) to show results with LWPs of 5 and 40 g m^{-2} without changing the r_e profile. The final r_e values from the three LWPs are shown in (f).

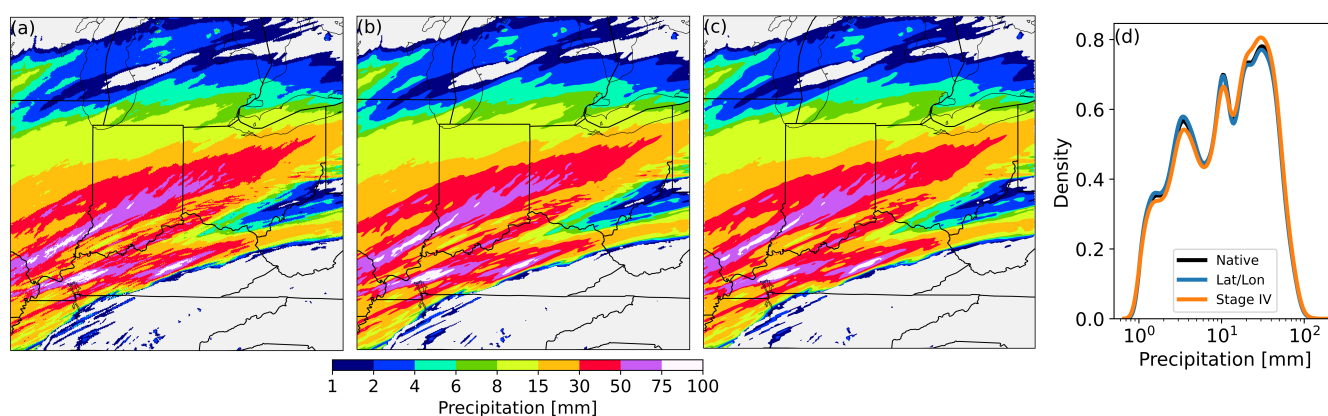


Figure B1. 24-h precipitation over a small sample area from (a) native MPAS-A output, (b) MPAS-A output converted to a latitude-longitude grid, and (c) MPAS-A output regridded to the Stage IV grid. Panel (d) shows the normalized distributions of precipitation from the regions shown in (a)–(c).



445 *Author contributions.* Anders A. Jensen developed the ML model, ran all simulations, performed the analyses, created figures, and led the writing and editing of the manuscript. David Turner provided guidance on the algorithm for calculating the weighted effective radius values from model output, processed the IRS observational dataset to derive retrieved effective radii, and wrote the section on the IRS instrumentation. David also provided feedback on figures and the analyses and edited the manuscript. Stanley Benjamin and Joseph Olson conceptualized the goals of this work with a particular focus on using the ML model for effective radius calculations. Both provided feedback on figures and analyses. Stanley Benjamin also provided manuscript edits. Clark Evans conceptualized the goals of this work and provided extensive manuscript edits and feedback on figures and analyses. Jorge Guerra and David Gagne worked with Anders A. Jensen to conceptualize and develop the machine learning methodology, and both edited the manuscript. Justin Minder assisted with interpretation of WINTRE-MIX precipitation measurements, edited the manuscript, and assisted with analysis of the WINTRE-MIX simulations and observations.

Competing interests. The authors have no competing interests.

455 *Disclaimer.* The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of OAR or the Department of Commerce. This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

460 *Acknowledgements.* The authors would like to thank the reviewers for their time. David Dowell provided useful comments on an earlier draft of this paper. The authors would like to thank the National Research Council (NRC) for their work to process aircraft observations. The authors thank Dr. Tim Wagner for providing the raw AERI data from University of Wisconsin - Madison site, and Drs. Bianca Adler and Laura Bianco for providing the ASSIST data from the NOAA Physical Sciences Laboratory site in Columbia. Gemini AI was used to write the piece of python code that added the bounding box and masked shaped region in Fig. 2a.



References

- 465 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/>, software available from tensorflow.org, 2015.
- 470 Bala, K. and Wolde, M.: WINTRE-MIX: NRC Convair 580 Atmospheric and Aircraft State Data. Version 1.0, <https://doi.org/https://doi.org/10.26023/BHS6-TEVC-GR04>, accessed 7 July 2025, 2022.
- Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., Olson, J. B., James, E. P., Dowell, D. C., Grell, G. A., Lin, H., Peckham, S. E., Smith, T. L., Moninger, W. R., Kenyon, J. S., and Manikin, G. S.: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh, *Monthly Weather Review*, 144, 1669 – 1694, <https://doi.org/10.1175/MWR-D-15-0242.1>,
475 2016.
- Benjamin, S. G., James, E. P., Turner, D. D., Balmes, K. A., Sedlar, J., Lantz, K. O., Jensen, A., Riihimaki, L. D., and Augustine, J. A.: Excessive downward shortwave radiation in the HRRR and RAP weather models and testing strategies for improvements, *Monthly Weather Review*, <https://doi.org/10.1175/MWR-D-25-0094.1>, 2025.
- Benjamin, S. G., James, E. P., Kenyon, J. S., Blake, B., and Turner, D. D.: Diagnostic fields for hourly updated NOAA weather models -
480 RRFsv1, HRRRv4, RAPv5, NOAA Technical Memorandum OAR GSL, <https://doi.org/https://doi.org/10.25923/937j-s495>, 2026.
- Bliankinshtein, N.: WINTRE-MIX: NRC Convair 580 Cloud Bulk Microphysics Data. Version 1.0, <https://doi.org/https://doi.org/10.26023/7FFB-FR7C-010C>, accessed 7 July 2025, 2022.
- Bryan, G. H. and Fritsch, J. M.: A Benchmark Simulation for Moist Nonhydrostatic Numerical Models, *Monthly Weather Review*, 130, 2917 – 2928, [https://doi.org/10.1175/1520-0493\(2002\)130<2917:ABSFMN>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2917:ABSFMN>2.0.CO;2), 2002.
- 485 Clark, A. J., Gallus, W. A., and Weisman, M. L.: Neighborhood-Based Verification of Precipitation Forecasts from Convection-Allowing NCAR WRF Model Simulations and the Operational NAM, *Weather and Forecasting*, 25, 1495 – 1509, <https://doi.org/10.1175/2010WAF2222404.1>, 2010.
- Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., Blake, B. T., Brown, J. M., Olson, J. B., Hu, M., Smirnova, T. G., Ladwig, T., Kenyon, J. S., Ahmadov, R., Turner, D. D., Duda, J. D., and Alcott, T. I.: The High-Resolution Rapid
490 Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description, *Weather and Forecasting*, 37, 1371 – 1395, <https://doi.org/10.1175/WAF-D-21-0151.1>, 2022.
- Duda, Michael G.: Convert MPAS, https://github.com/mgduda/convert_mpas, 2025.
- ESMF Organization: ESMPy, <https://github.com/esmf-org/esmf>, 2025.
- Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M. W., Lebo, Z. J., Morrison, H., and Gantos, G.: Machine Learning the Warm Rain
495 Process, *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002268, <https://doi.org/https://doi.org/10.1029/2020MS002268>, e2020MS002268 2020MS002268, 2021.
- Girouard, M. and et al.: WINTRE-MIX: CFI Climate Sentinels Hotplate Precipitation Gauge data. Version 1.0., <https://doi.org/https://doi.org/10.26023/BTZ8-RH82-PS12>, accessed 7 July 2025, 2022.
- Han, B. and et al.: WINTRE-MIX: Manual Hydrometeor Observation Reports. Version 1.0, <https://doi.org/https://doi.org/10.26023/68S9-0EBB-5A0D>,
500 0EBB-5A0D, accessed 7 July 2025, 2022.



- Han, B., Minder, J. R., Morrison, H., Milbrandt, J. A., and Cholette, M.: Simulating the Transition from Freezing Rain to Ice Pellets during WINTRE-MIX: The Role of Secondary Ice Production, *Monthly Weather Review*, 154, 835 – 855, <https://doi.org/10.1175/MWR-D-25-0191.1>, 2026.
- Heinzeller, D., Bernardet, L., Firl, G., Zhang, M., Sun, X., and Ek, M.: The Common Community Physics Package (CCPP) Framework v6, *Geoscientific Model Development*, 16, 2235–2259, <https://doi.org/10.5194/gmd-16-2235-2023>, 2023.
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *Journal of Geophysical Research: Atmospheres*, 113, <https://doi.org/https://doi.org/10.1029/2008JD009944>, 2008.
- Igel, A. L., Igel, M. R., and van den Heever, S. C.: Make It a Double? Sobering Results from Simulations Using Single-Moment Microphysics Schemes, *Journal of the Atmospheric Sciences*, 72, 910 – 925, <https://doi.org/10.1175/JAS-D-14-0107.1>, 2015.
- Jensen, A.: Software and data for TEMPO microphysics machine learning. Zenodo., <https://doi.org/10.5281/zenodo.21081016>, 2026.
- Jensen, A. A., Thompson, G., Ikeda, K., and Tessendorf, S. A.: Improving the Representation of Hail in the Thompson Microphysics Scheme, *Monthly Weather Review*, 151, 2307 – 2332, <https://doi.org/10.1175/MWR-D-21-0319.1>, 2023a.
- Jensen, A. A., Weeks, C., Xu, M., Landolt, S., Korolev, A., Wolde, M., and DiVito, S.: The Prediction of Supercooled Large Drops by a Microphysics and a Machine Learning Model for the ICICLE Field Campaign, *Weather and Forecasting*, 38, 1107 – 1124, <https://doi.org/10.1175/WAF-D-22-0105.1>, 2023b.
- Jensen, Anders A.: MPAS v8.2.2-3.11: Model for Prediction Across Scales for the ML-NC simulation, https://github.com/AndersJensen-NOAA/MPAS-Model/tree/pub/ml_nc, 2025a.
- Jensen, Anders A.: MPAS v8.2.2-3.11: Model for Prediction Across Scales for the ML-RE simulation, https://github.com/AndersJensen-NOAA/MPAS-Model/tree/pub/ml_re, 2025b.
- Jensen, Anders A.: MPAS v8.2.2-2.0: Model for Prediction Across Scales for ML training data, <https://github.com/AndersJensen-NOAA/MPAS-Model/commit/40ef3f32d5cd095c7fd3498be16636ec5c21b31f>, 2025c.
- Jensen, Anders A.: TEMPO: Thompson-Eidhammer Microphysics Parameterization for Operations, https://github.com/AndersJensen-NOAA/TEMPO/tree/exp/ml_pbl, 2025d.
- Jensen, Anders A.: TEMPO: Thompson-Eidhammer Microphysics Parameterization for Operations, https://github.com/AndersJensen-NOAA/TEMPO/tree/exp/ml_nc, 2025e.
- Knuteson, R. O., Revercomb, H. E., Best, F. A., Ciganovich, N. C., Dedecker, R. G., Dirks, T. P., Ellington, S. C., Feltz, W. F., Garcia, R. K., Howell, H. B., Smith, W. L., Short, J. F., and Tobin, D. C.: Atmospheric Emitted Radiance Interferometer. Part II: Instrument Performance, *Journal of Atmospheric and Oceanic Technology*, 21, 1777 – 1789, <https://doi.org/10.1175/JTECH-1663.1>, 2004.
- Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., and Hagerty, V.: Using Deep Learning to Emulate and Accelerate a Radiative Transfer Model, *Journal of Atmospheric and Oceanic Technology*, 38, 1673 – 1696, <https://doi.org/10.1175/JTECH-D-21-0007.1>, 2021.
- Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen, L., Clark, M., Dai, A., Dudhia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute, S., Li, Y., Thompson, G., and Yates, D.: Continental-scale convection-permitting modeling of the current and future climate of North America, *Climate Dynamics*, 49, 71–95, <https://doi.org/10.1007/s00382-016-3327-9>, 2017.
- Michaud-Belleau, V., Gaudreau, M., Lacoursière, J., Boisvert, E., Ravelomanantsoa, L., Turner, D. D., and Rochette, L.: The Atmospheric Sounder Spectrometer by Infrared Spectral Technology (ASSIST): instrument design and signal processing, *Atmospheric Measurement Techniques*, 18, 3585–3609, <https://doi.org/10.5194/amt-18-3585-2025>, 2025.



- Michelson, D.: WINTRE-MIX: Sorel, QC OTT Parsivel2 Disdrometer Data Version 0.1, <https://doi.org/https://doi.org/10.26023/2ATQ-PPKD-D80W>, accessed 01 May 2026, 2023.
- 540 Miles, N. L., Verlinde, J., and Clothiaux, E. E.: Cloud Droplet Size Distributions in Low-Level Stratiform Clouds, *Journal of the Atmospheric Sciences*, 57, 295 – 311, [https://doi.org/10.1175/1520-0469\(2000\)057<0295:CDSIDL>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<0295:CDSIDL>2.0.CO;2), 2000.
- Minder, J. R., Bassill, N., Fabry, F., French, J. R., Friedrich, K., Gultepe, I., Gyakum, J., Kingsmill, D. E., Kosiba, K., Lachapelle, M., Michelson, D., Nichman, L., Nguyen, C., Thériault, J. M., Winters, A. C., Wolde, M., and Wurman, J.: P-Type Processes and Predictability: The Winter Precipitation Type Research Multiscale Experiment (WINTRE-MIX), *Bulletin of the American Meteorological Society*, 104, E1469 – E1492, <https://doi.org/10.1175/BAMS-D-22-0095.1>, 2023.
- 545 Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *Journal of Geophysical Research: Atmospheres*, 102, 16 663–16 682, <https://doi.org/https://doi.org/10.1029/97JD00237>, 1997.
- Morrison, H., Thompson, G., and Tatarskii, V.: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes, *Monthly Weather Review*, 137, 991 – 1007, <https://doi.org/10.1175/2008MWR2556.1>, 2009.
- 550 Muñoz-Esparza, D., Becker, C., Sauer, J. A., Gagne II, D. J., Schreck, J., and Kosović, B.: On the Application of an Observations-Based Machine Learning Parameterization of Surface Layer Fluxes Within an Atmospheric Large-Eddy Simulation Model, *Journal of Geophysical Research: Atmospheres*, 127, e2021JD036 214, <https://doi.org/https://doi.org/10.1029/2021JD036214>, e2021JD036214 2021JD036214, 2022.
- 555 Nakanishi, M. and Niino, H.: Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer, *Journal of the Meteorological Society of Japan. Ser. II*, 87, 895–912, <https://doi.org/10.2151/jmsj.87.895>, 2009.
- National Center for Atmospheric Research: mlmicrophysics, <https://github.com/NCAR/mlmicrophysics>, 2026a.
- National Center for Atmospheric Research: WRF: Weather Research and Forecasting Model version 4.8.0, [https://github.com/wrf-model/](https://github.com/wrf-model/WRF/tree/release-v4.8.0)
- 560 [WRF/tree/release-v4.8.0](https://github.com/wrf-model/WRF/tree/release-v4.8.0), 2026b.
- National Oceanic and Atmospheric Administration: TEMPO: Thompson-Eidhammer Microphysics Parameterization for Operations, <https://github.com/NCAR/TEMPO>, 2026.
- Nichman, L.: WINTRE-MIX: NRC Convair 580 Scattering and Optical Array Probe Microphysics Data. Version 1.0, <https://doi.org/https://doi.org/10.26023/XNM4-MJP0-4Q0H>, accessed 7 July 2025, 2022.
- 565 Olson, J. B., Kenyon, J. S., Angevine, W. A., Brown, J. M., Pagowski, M., and Sušelj, K.: A Description of the MYNN-EDMF Scheme and the Coupling to Other Components in WRF–ARW, NOAA Technical Memorandum OAR GSD, <https://doi.org/https://doi.org/10.25923/n9wm-be49>, 2019.
- Olson, J. B., Smirnova, T., Kenyon, J. S., Turner, D. D., Brown, J. M., Zheng, W., and Green, B. W.: A Description of the MYNN Surface-Layer Scheme, NOAA technical memorandum OAR GSL, <https://doi.org/https://doi.org/10.25923/f6a8-bc75>, 2021.
- 570 Olson, J. B., Angevine, W. M., Turner, D. D., Sun, X., Simonson, J. M., Evans, C., Kenyon, J. S., Li, H., Schnell, J., Puhales, F. S., Cherubini, T., Li, W., and Zhang, M.: A Description of the MYNN-EDMF Turbulence Scheme, NOAA technical memorandum OAR GSL, <https://doi.org/https://doi.org/10.25923/rahr-sj70>, 2026.
- Putman, W. M. and Lin, S.-J.: Finite-volume transport on various cubed-sphere grids, *Journal of Computational Physics*, 227, 55–78, <https://doi.org/https://doi.org/10.1016/j.jcp.2007.07.022>, 2007.



- 575 Schiede, M.: Using Airborne Measurements to Evaluate Forecasts of Freezing Drizzle Aloft Results from the Wintre-Mix Field Campaign, Master's thesis, University at Albany, <https://scholarsarchive.library.albany.edu/etd/13>, 2024.
- Seifert, A. and Siewert, C.: An ML-Based P3-Like Multimodal Two-Moment Ice Microphysics in the ICON Model, *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004206, <https://doi.org/https://doi.org/10.1029/2023MS004206>, e2023MS004206 2023MS004206, 2024.
- 580 Skamarock, W., Klemp, J., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D., and Huang, X.-Y.: A Description of the Advanced Research WRF Model Version 4.1, <https://doi.org/https://doi.org/10.5065/1dfh-6p97>, 2019.
- Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S.-H., and Ringler, T. D.: A Multiscale Nonhydrostatic Atmospheric Model Using Centroidal Voronoi Tessellations and C-Grid Staggering, *Monthly Weather Review*, 140, 3090 – 3105, <https://doi.org/10.1175/MWR-D-11-00215.1>, 2012.
- 585 Smirnova, T. G. and Benjamin, S. G.: Advancements in the RUC land-surface model for use in the next-generation high-resolution weather prediction models, NOAA Technical Memorandum OAR GSL, <https://doi.org/https://doi.org/10.25923/55x8-cy36>, 2025.
- Smirnova, T. G., Brown, J. M., Benjamin, S. G., and Kenyon, J. S.: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) Available in the Weather Research and Forecasting (WRF) Model, *Monthly Weather Review*, 144, 1851 – 1865, <https://doi.org/10.1175/MWR-D-15-0198.1>, 2016.
- 590 Tessendorf, S. A., Rugg, A., Korolev, A., Heckman, I., Weeks, C., Thompson, G., Jacobson, D., Adriaansen, D., and Haggerty, J.: Differentiating Freezing Drizzle and Freezing Rain in HRRR Model Forecasts, *Weather and Forecasting*, 36, 1237 – 1251, <https://doi.org/10.1175/WAF-D-20-0138.1>, 2021.
- Thompson, G.: High-resolution winter simulations of winter precipitation over the Colorado Rockies, Ph.D. thesis, Shinfield Park, Reading, 2013.
- 595 Thompson, G. and Eidhammer, T.: A Study of Aerosol Impacts on Clouds and Precipitation Development in a Large Winter Cyclone, *Journal of the Atmospheric Sciences*, 71, 3636 – 3658, <https://doi.org/10.1175/JAS-D-13-0305.1>, 2014.
- Thompson, G., Rasmussen, R. M., and Manning, K.: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part I: Description and Sensitivity Analysis, *Monthly Weather Review*, 132, 519 – 542, [https://doi.org/10.1175/1520-0493\(2004\)132<0519:EFOWPU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2), 2004.
- 600 Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization, *Monthly Weather Review*, 136, 5095 – 5115, <https://doi.org/10.1175/2008MWR2387.1>, 2008.
- Thompson, G., Politovich, M. K., and Rasmussen, R. M.: A Numerical Weather Model's Ability to Predict Characteristics of Aircraft Icing Environments, *Weather and Forecasting*, 32, 207 – 221, <https://doi.org/10.1175/WAF-D-16-0125.1>, 2017.
- 605 Toy, M. D., Olson, J. B., Yang, F., Yang, B., and Hong, S.-Y.: A description of the Unified Gravity Wave Physics package, NOAA Technical Memorandum OAR GSD, <https://doi.org/https://doi.org/10.25923/v2kx-8e10>, 2025.
- Turner, D. D.: Arctic Mixed-Phase Cloud Properties from AERI Lidar Observations: Algorithm and Results from SHEBA, *Journal of Applied Meteorology*, 44, 427 – 444, <https://doi.org/10.1175/JAM2208.1>, 2005.
- Turner, D. D.: Improved ground-based liquid water path retrievals using a combined infrared and microwave approach, *Journal of Geophysical Research: Atmospheres*, 112, <https://doi.org/https://doi.org/10.1029/2007JD008530>, 2007.
- 610



Turner, D. D. and Löhnert, U.: Information Content and Uncertainties in Thermodynamic Profiles and Liquid Cloud Properties Retrieved from the Ground-Based Atmospheric Emitted Radiance Interferometer (AERI), *Journal of Applied Meteorology and Climatology*, 53, 752 – 771, <https://doi.org/10.1175/JAMC-D-13-0126.1>, 2014.

615 Turner, D. D., Adler, B., Bianco, L., Wilczak, J. M., Michaud-Belleau, V., and Rochette, L.: Intercomparison of seven collocated ground-based infrared spectrometer radiance observations and retrieved thermodynamic profiles, *Atmospheric Measurement Techniques*, 19, 1573–1586, <https://doi.org/10.5194/amt-19-1573-2026>, 2026.

UFS Community: CCpp: Common Community Physics Package, <https://github.com/ufs-community/ccpp-physics>, 2026a.

UFS Community: MPAS: Model for Prediction Across Scales, <https://github.com/ufs-community/MPAS-Model>, 2026b.

620 Wang, H., Skamarock, W. C., and Feingold, G.: Evaluation of Scalar Advection Schemes in the Advanced Research WRF Model Using Large-Eddy Simulations of Aerosol–Cloud Interactions, *Monthly Weather Review*, 137, 2547 – 2558, <https://doi.org/10.1175/2009MWR2820.1>, 2009.

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., Cooten, S. V., Kelleher, K., Kitzmiller, D., Ding, F., Seo, D.-J., Wells, E., and Dempsey, C.: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, Results, and Future Plans, *Bulletin of the American Meteorological Society*, 92, 1321 – 1338, <https://doi.org/10.1175/2011BAMS-D-11-00047.1>, 2011.