Machine Learning Parameterization of the Multi-scale Kain-Fritsch (MSKF) Convection Scheme and stable simulation coupled in WRF using WRF-ML v1.0

Xiaohui Zhong[1], Xing Yu[2], and Hao Li[1]

[1]Fudan University, Shanghai, 200433, China
[2]Shenzhen Institute of Artificial Intelligence and Robotics for Society, Guangdong, 518000, China

Correspondence: Xing Yu (yuxing@cuhk.edu.cn)

Abstract. Warm-sector heavy rainfall often occurs along the coast of South China, and it is usually localized and long-lasting, making it challenging to predict. High-resolution numerical weather prediction (NWP) models are increasingly used to better resolve topographic features and forecast such high-impact weather events. However, when the grid spacing becomes comparable to the length scales of convection, known as the gray zone, the turbulent eddies in the atmospheric boundary layer are only partially resolved and parameterized to some extent. Whether using a convection parameterization (CP) scheme in the gray zone remains controversial. Scale-aware CP schemes are developed to enhance the representation of convective transport within the gray zone. The multi-scale Kain-Fritsch (MSKF) scheme includes modifications that allow for its effective implementation at a grid resolution as high as 2 km. In recent years, there has been an increasing application of machine learning (ML) models to various domains of atmospheric sciences, including the replacement of physical parameterizations with ML models. This work proposes a multi-output bidirectional long short-term memory (Bi-LSTM) model as a replace the scale-aware MSKF CP scheme. The Weather Research and Forecast (WRF) model is used to generate training and testing data over South China at a horizontal resolution of 5 km. Furthermore, the WRF model is coupled with the ML based CP scheme and compared with WRF simulations with original MSKF scheme. The results demonstrate that the Bi-LSTM model can achieve high accuracy, indicating the potential use of ML models to substitute the MSKF scheme in the gray zone.

1 Introduction

Warm-sector heavy rainfall often occurs during the pre-flood seasons in South China due to the influence of the East Asian summer monsoon (Ding, 2004). These rainfall events are localized and are characterized by high precipitation intensity but limited spatial coverage. However, despite being small size, unexpected and extreme warm-sector rainfall can cause significant damage, resulting in flooding homes and vehicles, destroying crop fields, and endangering lives, with costs amounting to millions or even billions of dollars (Tao, 1981; Zhao et al., 2007; Zhong et al., 2015). Accurately predicting warm-sector heavy rainfall using Numerical Weather Prediction (NWP) models is challenging due to the complex interaction between the influence of the low-level jet (LLJ), land-sea contrast, topography, and urban landscape (Zhong and Chen, 2017; Luo et al., 2017; Jian et al., 2002; Di et al., 2006; Xia and Zhao, 2009; Zhang and Ni, 2009). The land surface in the South China region is

characterized by complex terrain and heterogeneity, which play a crucial role in promoting more active convections. Previous
studies (Giorgi et al., 2016; Mishra et al., 2018; Schumacher et al., 2020; Onishi et al., 2023) have shown that higher spatial
resolution improves the performance of convective rainfall forecasts by resolving topographic characteristics more accurately.
Recognizing the importance of resolution in forecasting severe convective weather, both the Chinese government and the community have shown increased support for developing high-resolution operational forecast models for warm-sector rainstorms
and sudden local rainstorms. In early 2017, the China Meteorological Administration (CMA) initiated the development of a
robust framework to evaluate the forecast accuracy of all available models, including high-resolution regional models, and to
develop key technologies for high-impact weather forecasting.

With the increased availability computational resources, there has been a growing trend of using regional NWP models with
finer grid spacings, typically ranging from 1 to 10 km. However, when the model grid spacing becomes comparable to the
size of convection, which is known as the gray zone (Wyngaard, 2004; Hong and Dudhia, 2012), cumulus convection that was
previously unresolved becomes partially resolved. In theory, unless Direct Numerical Simulation (DNS) is used to accurately
capture the smallest turbulent scales with a resolution of millimeters to centimeters (Jeworrek et al., 2019), parameterization of
turbulence or convection remains necessary for weather modeling. Nevertheless, there is controversy about whether convection
parameterization (CP) should be used in the gray zone. Some previous studies (Chan et al., 2013; Johnson et al., 2013) reducing
the horizontal grid spacing to below 4 km while using CP scheme does not yield any improvement, and in some cases even
worsens precipitation forecast performance. In contrast, Schwartz (2014) showed that forecasts with a horizontal grid spacing
of 1 km provide a more accurate spatial representation of accumulated rainfall over 48 hours compared to forecasts with 4 km
grid spacing. These conflicting findings typically arise when the CP scheme is applied at scales outside of its designed scales or
when it is abruptly disabled at a resolution of approximately 3-5 km. Therefore, it remains unclear if utilizing any CP schemes
in the gray zone is effective for predicting localized warm-sector heavy rainfall.

To enhance simulations in the gray zone, researchers hsave developed scale-aware CP schemes. These schemes dynamically
parameterize convective processes based on the horizontal grid spacing and ensure a smooth transition across spatial scales.
A study conducted by (Jeworrek et al., 2019) demonstrated that two scale-aware CP schemes, namely Grell-Freitas (Grell and
Freitas, 2014) and multi-scale Kain-Fritsch (MSKF) (Zheng et al., 2016), outperform conventional CP schemes in terms of
precipitation timing and intensity over the Southern Great Plains of the United States. Despite the growing use of these scale-aware schemes due to their superior performance, it is important to note that they also depend on various empirical parameters
(Villalba-Pradas and Tapiador, 2022). Hence, the development of CP schemes specific to the gray zone in NWP models still
presents significant challenges.

In recent years, there has been a growing number of studies investigating the use of machine learning (ML) models as an
alternative to conventional physics-based CP schemes. ML-based parameterization schemes have the potential to be effective at various horizontal resolutions because they are trained using data generated from models with varying grid resolutions
(Yuval and O'Gorman, 2020). Unlike conventional CP schemes, which rely on assumptions like convective quasi-equilibrium
(Arakawa, 2004), ML-based parameterization schemes do not depend on such assumptions. Random forests (RFs) and fully-connected (FC) neural networks (NNs) have been the two most frequently employed ML models for convective parameter-

ization (CP) schemes in prior research. RFs automatically impose physical constraints, including energy conservation and
non-negative surface precipitation, which are crucial for stable simulations. O'Gorman and Dwyer (2018) demonstrated that
RFs can maintain stability and accurately reproduce essential climate statistics by training them to mimic the moist convection
of an aquaplanet general circulation model (GCM). More recently, Yuval and O'Gorman (2020) employed the coarse-grained
output of a high-resolution three-dimensional (3D) GCM model, simulated on an idealized equatorial beta plane, to train the
RF parameterization. They showed that the RF parameterization is capable of reproducing the climate of the high-resolution
simulation when used at coarse resolution. However, FC NNs offer several advantages over RFs, such as the potential for
higher accuracy and reduced memory requirements. In their pioneering work, Krasnopolsky et al. (2013) formulated a stochas-
tic CP scheme using an ensemble of 3-layer NNs, which were trained using data generated by a cloud-resolving model (CRM)
over a small area in the tropical Pacific Ocean during a four-month period of the TOGA-COARE [1]. The results revealed that
the NN-based parameterization yielded reasonable and promising decadal climate simulations across a broader tropical Pa-
cific region when incorporated into the National Center of Atmospheric Research (NCAR) Community Atmospheric Model
(CAM). Gentine et al. (2018) utilized data from idealized simulations performed using the SuperParameterized Community
Atmosphere Model (SPCAM) over an aquaplanet to train a deep NN (DNN). The DNN predicts temperature and moisture
tendencies influenced by convection and clouds, along with the cloud liquid and ice water contents. Additionally, Rasp et al.
(2018) successfully incorporated an NN-based parameterization into a global GCM on an aquaplanet. They conducted stable
prognostic simulations for multiple years, accurately reproducing the climatology of SPCAM and capturing crucial aspects of
variability, including extreme precipitation and realistic tropical waves. However, Rasp (2020) also found that minor modifi-
cations to the configuration rapidly led to unpredictable blow-ups in simulation. Therefore, addressing the instability of NN
parameterization in GCMs is necessary. Yuval et al. (2021) developed a FC NN to that predicts the subgrid fluxes instead of
tendencies, incorporating the physical constraints from coarse-grained high-resolution atmospheric simulation in an idealized
domain. Brenowitz and Bretherton (2018, 2019) proposed a novel loss function that minimizes the accumulated prediction
error over multiple time steps instead of a single one. They further ensured long-term stability and accuracy by excluding
upper atmospheric humidity and temperature from the inputs. However, the approach of removing particular variables from the
inputs is relatively rudimentary, demanding additional research to enhance the stability of NN-based parameterizations when
integrated into the model.

In addition, previous studies have mainly used FC NNs to emulate convection. However, there exist other advanced NN
structures that have the potential to achieve higher accuracy. In a recent study, Han et al. (2020) made an initial attempt to
employ a deep residual convolutional NN (ResNet) (He et al., 2016) for emulating convection and cloud parameterization
in the SPCAM model using a realistic configuration. They compared the performance of ResNet with other NN architectures,
such as a FC DNN, a DNN with skip connections, and a convolutional NN (CNN) without skip connections. The results of their
comparison demonstrated that both ResNet and CNN without skip connections outperformed the FC NN and the DNN with

---

[1]TOGA-COARE is an acronym for Tropical Ocean Global Atmospheres/Coupled Ocean Atmosphere Response Experiment. It is an international research
program that investigates the interaction or coupling of the ocean and atmosphere in the western Pacific warm pool region from November 1992 to February
1993, encompassing 120 days of field experiments involving the deployment of oceanographic ships, moorings, drifters, and Doppler radars (ship, land, air).

skip connections in terms of accuracy, with the performance of ResNet and CNN without skip connections exhibiting similar performance. This finding highlights the crucial role of convolutions in achieving higher accuracy. Another study conducted by Yao et al. (2023) compared various ML model structures for emulating atmospheric radiative transfer processes, including FC NNs, CNNs, bidirectional recurrent-based NNs (RNNs), transformer-based NNs (Vaswani et al., 2017), and Fourier Neural Operators (FNO (Li et al., 2020)). Their results revealed that models with ability to preceive global context of the entire atmospheric column outperformed FC NNs and CNNs. Particularly, the bidirectional long short-term memory (Bi-LSTM) achieved the highest accuracy. Similar to radiative transfer modeling, Han et al. (2020) also emphasized the importance of ML having a global perspective of the entire atmospheric column for ML models in convection modeling. Modifying the depths of CNNs from 4 to 22 layers led to an increase in model accuracy. This improvement is primarily attributed to the expansion of the receptive field in deeper CNN layers. Thus, ML models that integrate both global and local perception capabilities are better suited for developing ML-based CP schemes.

Furthermore, all previous studies have predominantly focused on using CP schemes in GCM models for climate forecasting. Moreover, the choice of CP schemes significantly influences the uncertainty in precipitation forecasts within weather forecasting models. The complexity of the CP schemes also surpasses those applied in climate models (Arakawa, 2004). This study employs ML models to simulate convective processes for weather forecasting. We generate our dataset by running the Weather and Research Forecasting (WRF) (Skamarock et al., 2021) model with the scale-aware MSKF scheme employed as the CP scheme. Developed as a means to mitigate the precipitation overestimation associated with the original KF scheme, the MSKF scheme serves as an improved version of the Kain-Fritsch (KF) scheme (Kain and Fritsch, 1990, 1993; Kain, 2004). The KF scheme often initiates convection prematurely, which leads to an overprediction of precipitation during summer. To address these issues, the MSKF incorporates a scale-dependent capability, such as modifying the formulation of the convective adjustment timescale. This vital parameter defines the intensity and duration of convection, and the MSKF scheme made it dynamic and grid resolution dependent (Zhang et al., 2021). The WRF model covers the South China region. Furthermore, we utilize a Bi-LSTM model to emulate the convective processes and couple it with the WRF model using the WRF-ML coupler developed by Zhong et al. (2023a). The performance of the ML-based CP scheme is evaluated in both offline and online settings.

The paper is structured as follows. Section 2 provides a description of the WRF model for data generation, as well as the input and output data of the ML model. In Section 3, the original and the ML-based MSKF scheme is introduced. The results for both offline and online testing of the ML-based MSKF scheme are presented in Section 4. Finally, Section 5 presents the summary and conclusion.

2 Data

2.1 Data generation

The dataset was generated by running the WRF model version 4.3 (Skamarock et al., 2019, 2021). The following subsections provide a comprehensive explanation of the WRF model configurations, as well as the input and output variables employed in the development of the ML-based CP scheme.

The WRF model is compiled using the GNU Fortran (gfortran version 7.5.0) compiler with the "dmpar" option. The WRF
125 model is run using the domain configuration illustrated in Figure 1. The WRF model is configured with a single domain
consisting of 44000 grid points, with a horizontal grid spacing of 5 km and dimensions of 220×200 grid points in the west-
east and north-south directions. The model consists of 45 vertical levels (i.e., 44 vertical layers), with a model top at 50 hPa.
Additionally, the WRF model is configured with physics schemes, including WSM 6-class graupel scheme (Hong and Lim,
2006) for microphysics, Bougeault-Lacarrère (BouLac) scheme (Bougeault and Lacarrère, 1989) for planetary boundary layer
130 (PBL) mixing, the Monin-Obukhov (Janjic) surface layer scheme (Janjic, 1996), the Unified Noah model (Livneh et al., 2011)
for land surface, RRTMG for both shortwave and longwave radiation (Iacono et al., 2008), and MSKF (Zheng et al., 2016) for
cumulus. The time step used for all WRF simulations is set to 15 seconds.

The initial and boundary conditions were derived from the ERA5 reanalysis dataset, which was provided by the European
Centre for Medium-range Weather Forecast (ECMWF) (Hersbach et al., 2020). The ERA5 reanalysis dataset used in this study
135 has a horizontal resolution of $0.25°$ and consists of 29 pressure levels below 50 hPa. To create a dataset for developing the ML
model, the WRF simulations were initialized at 12 UTC and conducted 9 times every 2 days, specifically from May 20th, 2022
to June 5th, 2022. Throughout the simulations, the MSKF scheme was called every 5 model minutes, generating outputs at each
call. The simulations ran for 36 hours each time, with the first 24 hours used for training and the last 12 hours for validation.
Therefore, the total number of training samples is 114,444,000 [2] while the offline validation set contains 57,024,000 [3] samples.
140 Furthermore, considering that the offline performance might not necessarily reflect its performance in online setting, we
conducted experiments by coupling the ML-based MSKF scheme with WRF model and comparing the results to the original
WRF simulations to evaluate the online performance of the ML-based MSKF scheme. The simulations were conducted 4 times
every 2 days, each lasting 36 hours. The initialization days spanned from June 12th, 2022 to June 18th, 2022.

---

[2] $114,444,000 = 44000 \times 9 \times (24 \times 60 / 5 + 1)$
[3] $57,024,000 = 44000 \times 9 \times 12 \times 60 / 5$
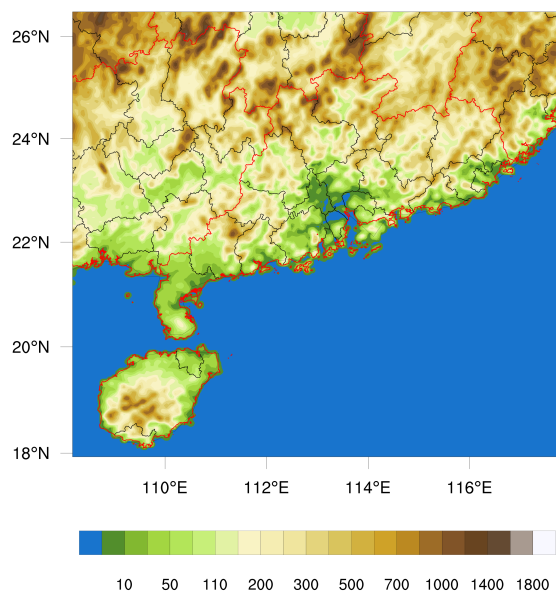
Figure 1. Digital evaluation data of the single WRF domain with horizontal resolution at 5. Red lines are the province borderlines, and black lines are the city borderlines.

2.2 Input and output data

145 Table 1 presents a comprehensive list of the input and output variables used in this study. There are 13 variables exclusively used as input, while 9 variables serve as both input and output. The output time-step precipitation due to convection ("raincv") is calculated through multiplying precipitation rate by the model time step. Out of all the variables, 5 are two-dimensional (2D) surface variables, while the remaining ones are 3D variables characterized by 44-layer vertical profiles. Additionally, the ML model used in this study incorporates 4 derived variables as input. These variables consist of a 2D Boolean variable indicating

150 convection triggering based on the values of "nca", pressure difference between each adjacent vertical levels, saturated water vapor mixing ratio, and relative humidity. Furthermore, the output "w0avg", which depends on vertical wind component (w) and input "w0avg", is also included as an input to model. In total, the ML model utilizes 27 input variables.

The variable "nca" represents the cloud relaxation time and must be an integer multiple of the model time step. For all WRF model simulations conducted in this study, a fixed time step of 15 seconds is used. Thus, "nca" is expected to be evenly divisible

155 by 15. To eliminate dependence on the specific model time step, "nca" is divided by the model time step before normalization is applied during model training. Moreover, within the MSKF scheme, "nca" plays a crucial role in determining the triggering of convection. Convection is triggered when "nca" is greater than or equal to half of the model time step.

To ensure consistency with the dimensions of the 3D variables, the surface variables are padded by duplicating the values of the surface layer for all layers before feeding them into the model. Prior to utilizing the variables in the Bi-LSTM model for

Table 1. Definition of all the input and output variables, and whether they are surface or 3D variables and their corresponding units. There are 44 model layers.

Type	Variable name	Definition	type	Unit
Input	u	meridional wind component	3D	m/s
	v	zonal wind component	3D	m/s
	w	vertical wind component	3D	m/s
	t	temperature	3D	K
	qv	Water vapor mixing ratio	3D	kg/kg
	p	pressure	3D	Pa
	th	potential temperature	3D	K
	dz8w	layer thickness	3D	m
	rho	air density	3D	kg/m^3
	pi	Exner function, which is dimensionless pressure and can be defined as: $\left(\frac{p}{p_0}\right)^{R_d/c_p}$		
	hfx	upward heat flux at surface	surface	W/m^2
	ust	u∗ in similarity theory	surface	W/m^2
	pblh	planetary boundary layer height	surface	m
Input and Output	rthcuten	potential temperature tendency due to cumulus parameterization	3D	K/s
	rqvcuten	water vapor mixing ratio tendency due to cumulus parameterization	3D	kg/kg/s
	rqccuten	cloud water mixing ratio tendency due to cumulus parameterization	3D	kg/kg/s
	rqrcuten	rain water mixing ratio tendency due to cumulus parameterization	3D	kg/kg/s
	rqicuten	cloud ice mixing ratio tendency due to cumulus parameterization	3D	kg/kg/s
	rqscuten	snow mixing ratio tendency due to cumulus parameterization	3D	kg/kg/s
	w0avg	average vertical velocity	3D	m/s
	nca	counter of the cloud relaxation time	3D	s
	pratec	precipitation rate due to cumulus parameterization	surface	mm/s
Output	raincv	precipitation due to cumulus paramterization	surface	mm

160   training and validation, normalization is applied to ensure uniformity in the magnitudes of all the variables. Each variable is divided by the maximum absolute value in the atmospheric column (for 3D variables) or at the surface (for surface variables).

3   Method

This section describes the flow chart of original MSKF scheme for determining convection trigger, ML model structures and training, and the evaluation methods.

3.1 Description of original MSKF module

The MSKF scheme is a scale-aware adaptation of the KF CP scheme, which was originally developed by Kain and Fritsch (1990, 1993) and later modified by Kain (2004). The flow chart illustrating the convection trigger process in the original MSKF scheme is illustrated in Figure 2. At the beginning of each call, the variable "nca" is examined to determine if exceeds or equals a threshold equal to half of the model time step (dt). If "nca" is greater than or equal to half of dt, the convective tendencies and precipitation activation rate do not require updating since convection is still active. If "nca" is smaller than the specified threshold, a 1-D cloud model within the MSKF scheme is used to calculate various variables related to cloud properties in order to determine whether convective should be triggered. These variables include the lifting condensation level (LCL), convective available potential energy (CAPE), cloud top and base heights, and entrainment rates. If convection is activated for a particular grid point, the MSKF scheme computes the convective tendencies and precipitation rate. On the other hand, if convection is not triggered, the convective tendencies and precipitation rate remain at 0. However, the variable "w0avg" is always updated regardless whether convection is triggered or not. As long as the convection remains active, "nca" is decremented by one model time step at each WRF model's time step.

3.2 Description of ML-based MSKF scheme

In the original MSKF scheme, the atmospheric column is processed sequentially, one at a time, until all horizontal grid points within the domain have been processed. In contrast, the ML-based MSKF scheme performs calculations on a batch (B on Figure 3) of data, consisting of 44 features and 27 vertical layers. As a result, the input data has dimensions of B × 27 × 44. Before being fed into the ML model, the input data data undergoes pre-processing through a module incorporating a 1-dimensional (1D) convolutional layer. This module increases the feature dimension from 27 to 64. The subsequent sections provide a comprehensive description of the structures of the ML model.

3.2.1 ML model structure

Predicting whether convection is triggered as well as modeling convective tendencies and precipitaion rate are two main tasks in conventional CP schemes. Therefore, the development of a ML-based CP scheme requires both a binary classification model for predicting convection trigger and a regression model for modeling convective tendencies. Accordingly, we propose a multi-output Bi-LSTM model that can simultaneously perform regression and classification predictions (Figure 3). Our proposed model consists of a shared Bi-LSTM layer for learning features, a classification subnetwork, and a regression subnetwork. The shared Bi-LSTM layer includes three repeated Bi-LSTM blocks, with each block containing a forward and a backward layer that have a feature dimension of 32. The classification subnetwork is composed of a 1×1 1D convolutional layer, a FC layer, and a Sigmoid activation layer. The output of the Sigmoid layer represents the probability distribution of the convection trigger. The binary cross-entropy loss function is employed as the cost function for this classification task. Meanwhile, the regression subnetwork incorporates a FC layer to output precipitation rate, "nca", and convective tendencies. Finally, the output
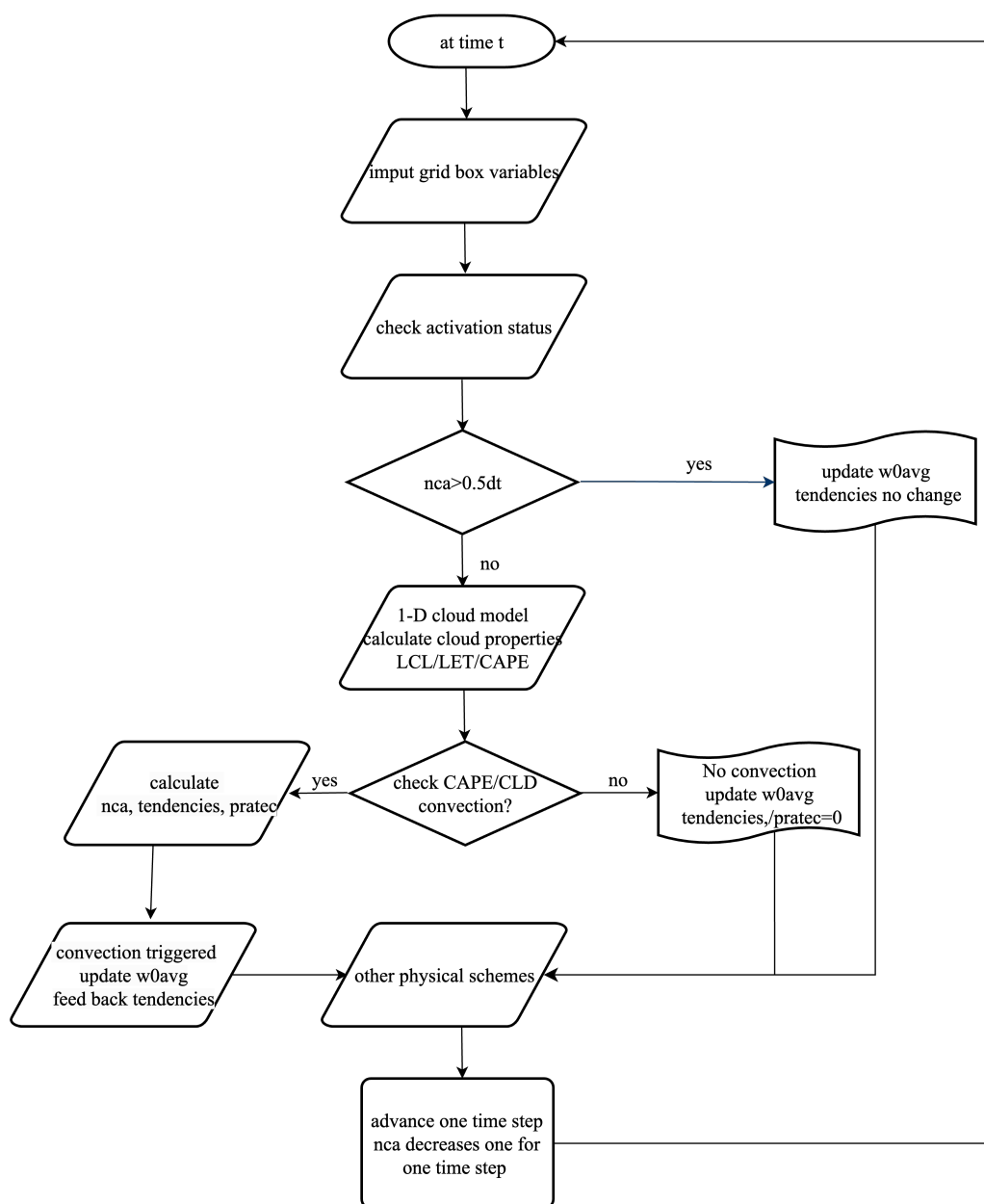
Figure 2. A flow chart outlining convection trigger process in the original MSKF scheme.
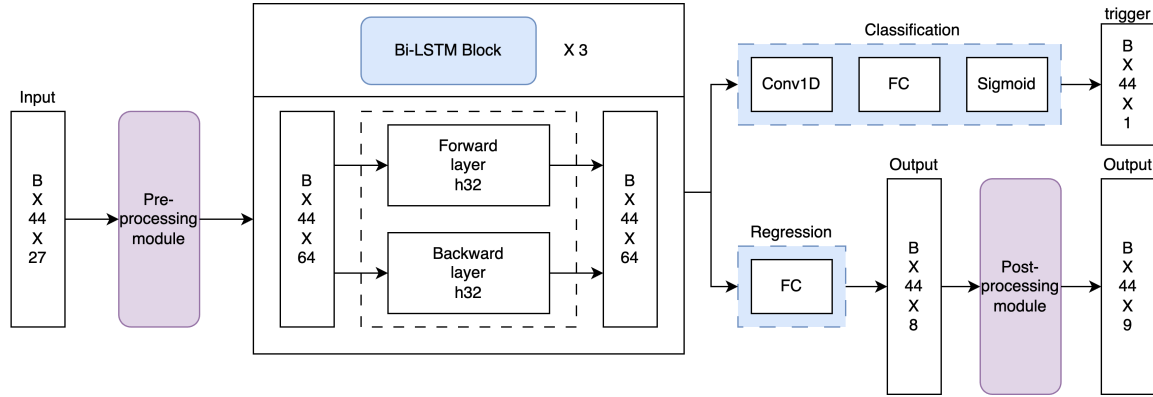
Figure 3. The architecture of the multi-output Bi-LSTM model for combined classification and regression predictions.

of both subnetworks passes through a post-processing module to ensure physical consistency. The post-processing module is introduced in more detail in the following subsection.

3.2.2 Post-processing module

The post-processing module is designed to ensure physical consistency of all variables. The following rules are applied: 1) For
200 grid points where the input "nca" is greater than or equal to half of dt, all other variables remain unchanged as they are still within the convection lifetime. 2) The output "nca" must be an integer. 3) For grid points where convection is predicted to be inactive, all other output variables are set to zero. In addition, the time-step convective precipitation (raincv) is calculated as described in the previous section 2.2.

3.2.3 Model training

205 As our model incorporates both classification and regression tasks, we optimize its performance by minimizing a multi-task loss function (Ren et al., 2016). The loss function is defined as the sum of the binary cross entropy loss for the convection trigger and a weighted combination of the $L1$ loss for all output variables from the regression subnetwork. The specific formulation of the loss function is as follows:

$$L = \frac{1}{N_{cls}} \sum_{i,j} L_{cls}(p_{i,j}, p_{i,j}^{gt}) + \sum_{c} \lambda_c \frac{1}{N_{reg}} \sum_{i,j} p_{i,j}^{gt} L1_c \tag{1}$$

210 Here, i and j denote the grid points in the domain. $p_{i,j}$ represents the probability of convection being triggered. The ground-truth label $p_{i,j}^{gt}$ takes a value of 1 if convection is triggered and 0 otherwise. The classification loss L_{cls} is calculated using the binary cross entropy loss. For the regression loss of different variables c, λ_c functions as a weight that balances the output variables by considering their respective magnitudes. The term $p_{i,j}^{gt} L1_c$ indicates that the $L1$ regression loss is activated only for triggered grid points ($p_{i,j}^{gt} = 1$) and is disabled otherwise ($p_{i,j}^{gt} = 0$). Both loss terms are normalized by N_{cls} and N_{reg},
215 which correspond to the total number of grid points and the number of triggered grid points, respectively.

Adam optimizer (Kingma and Ba, 2014) is used with an initial learning rate of 0.003 update the parameters of the model. Furthermore, the plateau scheduler is implemented to decrease the learning rate by a factor of 0.5 when the loss fails to decrease for five epochs. The model is trained for 150 epochs using a batch size of 44000.

3.3 Evaluation methods

220 The ML-based MSKF scheme is evaluated in both offline and online settings. The offline performance of the ML-based MSKF scheme is evaluated by comparing it against the outputs of the original MSKF scheme using validation dataset, including rthcuten, rqvcuten, rqccuten, rqrcuten, nca, and pratec. The overall model performance metrics include root mean squared error (RMSE) and correlation coefficient. The mean absolute error (MAE) and mean bias error (MBE) per vertical layer were were calculated using the equation below:

225
$$MAE_l = \frac{1}{N} \sum_{i=1}^{N} |Y_{ML}(i,l) - Y(i,l)| \tag{2}$$

$$MBE_l = \frac{1}{N} \sum_{i=1}^{N} Y_{ML}(i,l) - Y(i,l) \tag{3}$$

where $Y(i,l)$ and $Y_{ML}(i,l)$ represent the outputs from the original MSKF scheme and ML-based MSKF scheme, respectively. Here, i denotes the horizontal grid point of a vertical profile, N is the number of the horizontal grid points in the domain, l represents the vertical layer index.

230 ## 4 Results

4.1 Offline validation of the ML-based MSKF scheme

The offline validation was conducted using data that was not used during the training process. Figure 4 compares the cloud relaxation time (nca), precipitation rate (pratec), and convective tendencies predicted by both the original MSKF scheme and the ML-based MSKF scheme, respectively.To facilitate the comparison, the precipitation rate and temperature tendencies were
235 multiplied by a factor of 86,400 (24×3600) to convert from mm·s^{-1} and K·s^{-1} to mm·d^{-1} and K·d^{-1}, respectively. Similarly, the water vapor mixing ratio (rqvcuten), cloud water mixing ratio (rqccuten), and rain water mixing ratio (rqrcuten) resulting from convection were multiplied by 86,400,000 ($24 \times 3600 \times 1000$) to convert from kg·kg^{-1}·s^{-1} to g·kg^{-1}·d^{-1}. Among the output variables listed in Table 1, w0avg is excluded as it is calculated using an equation with the ground truth as input in this offline validation. Hence, evaluating w0avg in the offline evaluation is unnecessary.
240 Among all the variables illustrated in Figure 4, the variable "nca" exhibits a significantly higher RMSE of 4.32, and its data points appear scattered across a wide range of values. This suggests that accurately predicting convection poses a considerable challenge. Prior to plotting and performing statistical calculations, "nca" is divided by the model time step of 15 to eliminate

the time step dependence. The precipitation rate demonstrates the highest correlation coefficient and the smallest spread, as most data points cluster closely around the 1:1 line. In the case of temperature and the four moisture tendencies, the data also displays some dispersion, but a majority of the data points remain close to the 1:1 line. Overall, the ML-based MSKF scheme shows a strong correlation with the original MSKF scheme for all the variables, with correlation coefficients consistently higher than 0.91. This indicates that the the ML-based MSKF scheme has the potential to replace the original scheme.

To gain a comprehensive understanding of the vertical distribution of errors, Figure 5 illustrates the vertical profiles of statistics in convective tendencies. The solid and dotted lines in the figure represent the MAE and MBE of tendencies at each vertical layer, respectively. Additionally, the shaded area corresponds to the 5th and 95th percentiles of differences between tendencies predicted by the ML-based MSKF predicted scheme and the original MSKF scheme, respectively. The vertical error distribution in all tendencies is quite similar with higher variance observed among the pressure layers between 800 and 1,000 hPa. These pressure layers corresponds to the atmospheric layer where convection occurs most frequently. Due to the significantly lower cloud and rain content compared to water vapor in the atmosphere, the error magnitudes for rqccuten and rqrcuten are considerably smaller than those for rqvcuten.
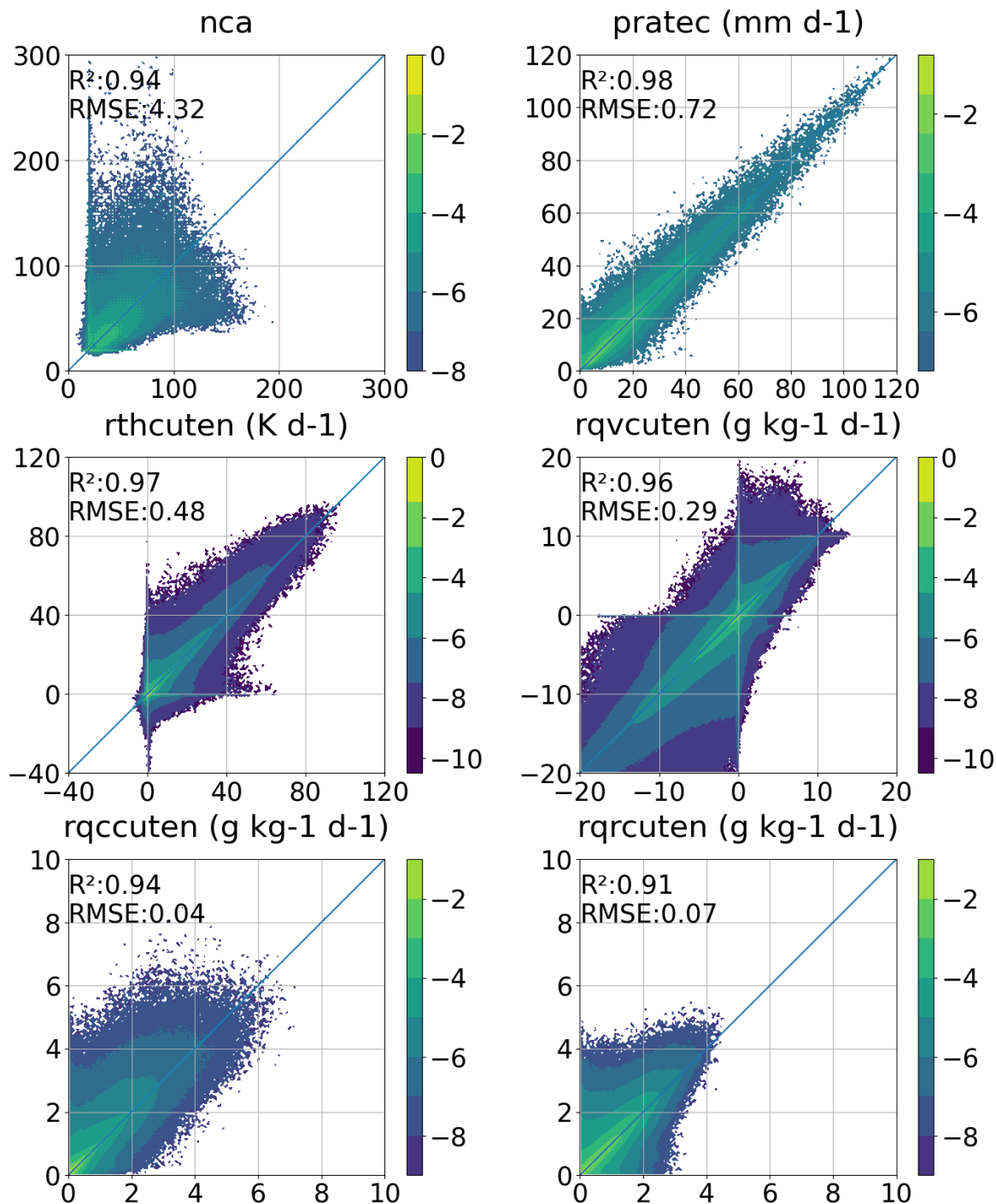
Figure 4. Comparison of the predicted (y axis) and true (x axis) nca, pratec, rthcuten (first column), rqvcuten (second column), rqccuten (third column), and rqrcuten for using validation data in the offline setting.
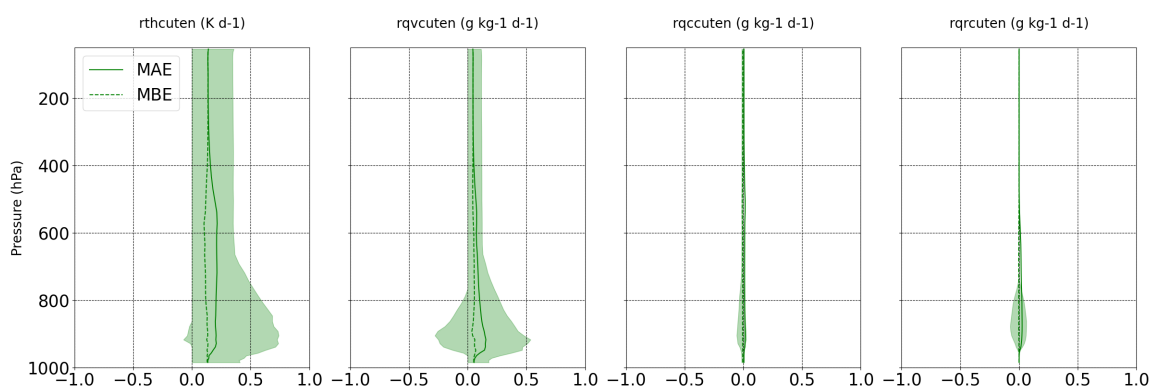
Figure 5. Vertical profiles of the statistics in rthcuten (first column), rqvcuten (second column), rqccuten (third column), and rqrcuten (fourth column) using validation data in the offline setting data using ML-based emulators. The solid and dotted lines show the MAE and MBE profile, respectively, and the shaded area indicates the 5th and 95th percentile of differences (prediction—target) at each layer.

4.2 Prognostic validation

This subsection presents the performance of the ML-based MSKF scheme in the online setting.

The ML-based MSKF scheme was integrated as a replacement for the original MSKF scheme in the WRF model to model convective effects. The WRF-ML coupler (Zhong et al., 2023a) was used to incorporate the ML-based MSKF scheme into the WRF model. Both the modified WRF model, incorporating the ML-based scheme, and the original WRF model were initialized on June 12, 14, 16, and 18, 2022, and run for 36 hours. It is worth mentioning that this runtime was completely independent of the training dataset.

In Figure 6, the averaged spatial forecasts of the 4-day prediction from the original WRF model are presented. These results include accumulation of convective and non-convective precipitation over a 12-hour period, as well as the 2-meter temperature at 12, 24, and 36 hours. The figure also illustrate the mean absolute difference (MAD) between the WRF simulations coupled with the ML-based MSKF scheme and those using the original MSKF scheme. The red and blue patterns in the spatial forecasts represent the magnitudes of the forecast values. In the spatial difference, red and blue patterns reveal the positive and negative biases of the ML-based simulations, respectively. while green patterns suggest little to no difference compared to the original WRF simulations. Additionally, a domain-averaged MAD is calculated to evaluate the overall performance of the ML-based scheme in prognostic simulations. Generally, the difference is small suggesting that the WRF simulations coupled with ML-based MSKF scheme agree well with the origianl WRF simulations. Also, the difference does not increase as the simulation time progresses, as there is a comparable domain-averaged MAD at 36 forecast hours compared to that at 12 forecast hours. These findings indicate that the ML-based MSKF scheme achieves stable prognostic simulations.
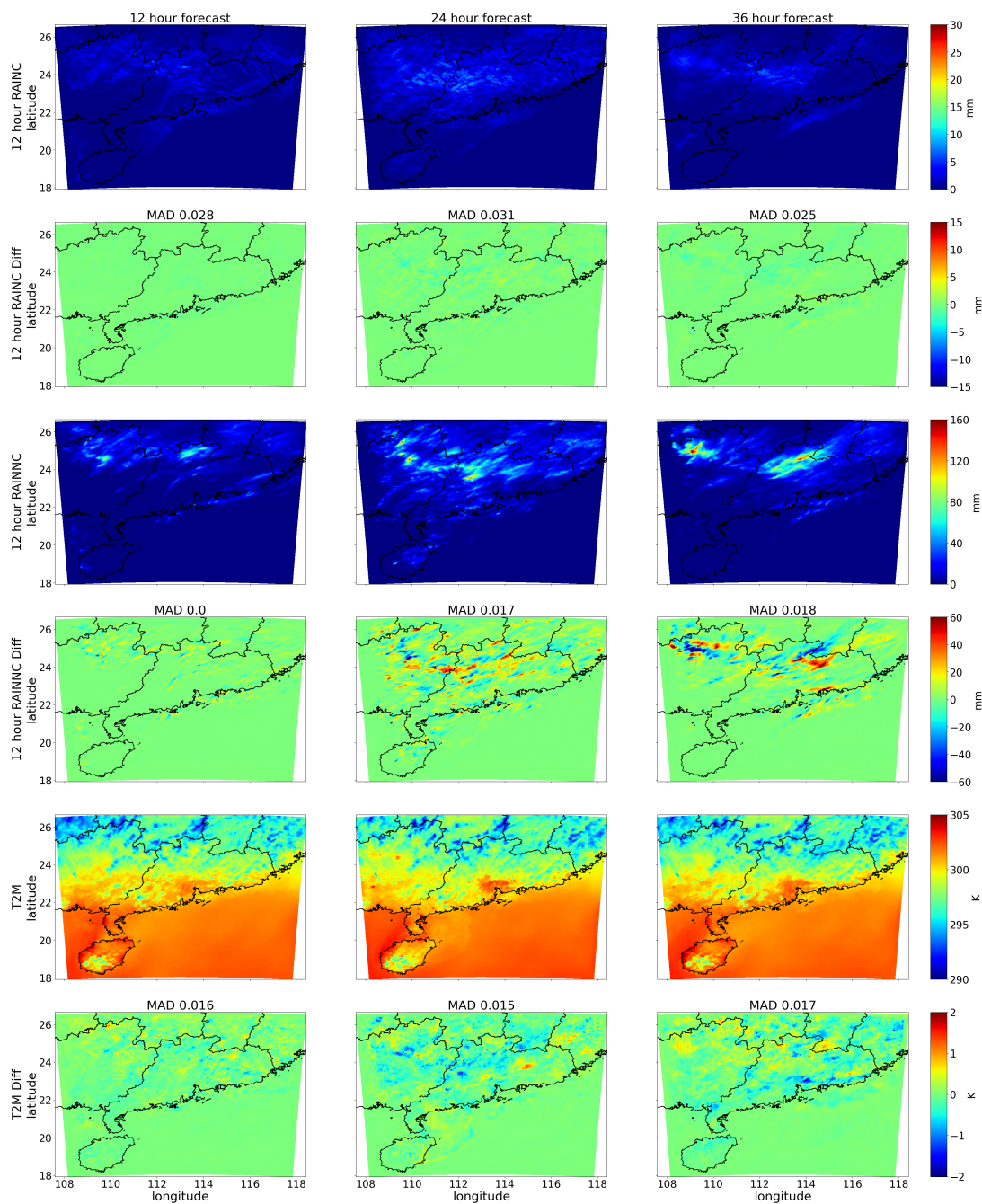
Figure 6. Spatial map of the average WRF simulations using the original MSKF scheme (in the first, third, and fifth rows) along with the average MAD between WRF simulations coupled with the ML-based MSKF scheme and WRF simulation with the original MSKF scheme (in the second, fourth and sixth rows). The simulations are shown for the 12-hour accumulated convective precipitation ($RAINC$) in the first and second rows, the 12-hour accumulated non-convective precipitation ($RAINNC$) in the third and fourth rows, and the 2-meter temperature ($T2M$) at forecast lead times of 12 hours (first column), 24 hours (second column), and 36 hours (third column).

5 Conclusions

In this paper, we proposed a multi-output Bi-LSTM model to developing a ML-based MSKF scheme for predicting convection trigger and reproducing the convective process in the gray zone. The model is trained using data generated from the WRF simulations at a spatial resolution of 5 km, covering the South China region. The output variables of the ML-based MSKF scheme are identical to those of the original MSKF scheme, which include cloud relaxation time ("nca"), precipitation rate ("pratec"), time-step convective precipitation ("raincv"), and convective tendencies. In the ML-based scheme, the physical consistency among all output variables is considered and encoded as a post-processing module to refine the output from the Bi-LSTM model. Offline validation demonstrates the excellent performance of the ML-based MSKF scheme. Furthermore, the ML-based MSKF scheme is coupled with the WRF model using WRF-ML coupler. The WRF simulations coupled with the ML-based MSKF scheme is compared against the WRF simulation with the original MSKF scheme. Results shows that the ML-based scheme can generate forecasts similar to the original ML scheme in online settings, showing the potential substitution of the MSKF scheme by ML models in gray-zone.

Code and data availability.   The source code for the WRF model version 4.3 used in this study is available at https://doi.org/10.5281/zenodo.10039053 (Skamarock et al., 2023). The source code and data used in this are available at https://doi.org/10.5281/zenodo.10032404 (Zhong et al., 2023b).

Author contributions.   X.Z. trained the deep learning models and calculate the statistics of model performance. X.Y. and X.Z. conducted the WRF simulations to provide dataset for training and evaluation, and offered valuable suggestions on the model training and paper revision. X.Z. and X.Y. wrote, reviewed and edited the original draft; X.Z., X.Y., and H.L. supervised and supported this research and gave important opinions. All of the authors have contributed to and agreed to the published version of the manuscript.

Competing interests.   The authors declare no conflict of interest.

Acknowledgements.   This work was supported by Basic and Applied Basic Research Foundation of Guangdong Province, under Grant No. 2021A1515012582. We are thankful for Mesoscale and Microscale Meteorology Laboratory (MMM) at NCAR for developing and sharing the WRF source codes.

References

Arakawa, A.: The Cumulus Parameterization Problem: Past, Present, and Future, Journal of Climate, 17, 2493 – 2525, https://doi.org/https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2, 2004.

Bougeault, P. and Lacarrère, P.: Parameterization of orographic induced turbulence in a mesobeta scale model, Monthly Weather Review, 117, 1872–1890, https://doi.org/10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2, 1989.

Brenowitz, N. D. and Bretherton, C. S.: Prognostic Validation of a Neural Network Unified Physics Parameterization, Geophysical Research Letters, 45, 6289–6298, https://doi.org/https://doi.org/10.1029/2018GL078510, 2018.

Brenowitz, N. D. and Bretherton, C. S.: Spatially Extended Tests of a Neural Network Parametrization Trained by Coarse-Graining, Journal of Advances in Modeling Earth Systems, 11, 2728–2744, https://doi.org/https://doi.org/10.1029/2019MS001711, 2019.

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Ferro, C. A. T., and Stephenson, D. B.: Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation?, Climate Dynamics, 41, 1475–1495, https://doi.org/10.1007/s00382-012-1568-9, 2013.

Di, X. R., Xiong, Z. S., and Beijing: A Study of Circumstances of Meso-β-Scale Systems of Strong Heavy Rainfall in Warm Sector Ahead of Fronts in South China, Chinese Journal of Atmospheric Sciences/Daqii kexue, 2006.

Ding, Y.: Seasonal march of the East-Asian summer monsoon, World Scientific, 2004.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could Machine Learning Break the Convection Parameterization Deadlock?, Geophysical Research Letters, 45, 5742–5751, https://doi.org/https://doi.org/10.1029/2018GL078202, 2018.

Giorgi, F., Torma, C., Coppola, E., Ban, N., Schär, C., and Somot, S.: Enhanced summer convective rainfall at Alpine high elevations in response to climate warming, Nature Geoscience, 9, 584–589, https://doi.org/10.1038/ngeo2761, 2016.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmospheric Chemistry and Physics, 14, 5233–5250, https://doi.org/10.5194/acp-14-5233-2014, 2014.

Han, Y., Zhang, G. J., Huang, X., and Wang, Y.: A Moist Physics Parameterization Based on Deep Learning, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 076, https://doi.org/https://doi.org/10.1029/2020MS002076, e2020MS002076 2020MS002076, 2020.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hong, S.-Y. and Dudhia, J.: Next-Generation Numerical Weather Prediction: Bridging Parameterization, Explicit Clouds, and Large Eddies, Bulletin of the American Meteorological Society, 93, ES6 – ES9, https://doi.org/https://doi.org/10.1175/2011BAMS3224.1, 2012.

Hong, S.-Y. and Lim, J.-O. J.: The WRF single-moment 6-class microphysics scheme (WSM6), Asia-Pacific Journal of Atmospheric Sciences, 42, 129–151, 2006.

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, 2008.

Janjic, Z.: The surface layer parameterization in the NCEP Eta Model, World Meteorological Organization-Publications-WMO TD, pp. 4–16, 1996.

Jeworrek, J., West, G., and Stull, R.: Evaluation of Cumulus and Microphysics Parameterizations in WRF across the Convective Gray Zone, Weather and Forecasting, 34, 1097 – 1115, https://doi.org/https://doi.org/10.1175/WAF-D-18-0178.1, 2019.

Jian, S., Ping, Z., and Xiuji, Z.: THE MESOSCALE STRUCTURE OF A SOUTH CHINA RAINSTORM AND THE INFLUENCE OF COMPLEX TOPOGRAPHY (in Chinese), Acta Meteorologica Sinica, pp. 333–342, https://doi.org/10.11676/qxxb2002.040, 2002.

Johnson, A., Wang, X., Kong, F., and Xue, M.: Object-Based Evaluation of the Impact of Horizontal Grid Spacing on Convection-Allowing Forecasts, Monthly Weather Review, 141, 3413 – 3425, https://doi.org/https://doi.org/10.1175/MWR-D-13-00027.1, 2013.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170 – 181, https://doi.org/https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Kain, J. S. and Fritsch, J. M.: A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization, Journal of Atmospheric Sciences, 47, 2784 – 2802, https://doi.org/https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2, 1990.

Kain, J. S. and Fritsch, J. M.: Convective Parameterization for Mesoscale Models: The Kain-Fritsch Scheme, American Meteorological Society,, 1993.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, 2014.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A.: Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model, Advances in Artificial Neural Systems, 2013, 5–5, 2013.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A.: Fourier neural operator for parametric partial differential equations, arXiv preprint arXiv:2010.08895, 2020.

Livneh, B., Restrepo, P. J., and Lettenmaier, D. P.: Development of a unified land model for prediction of surface hydrology and land–atmosphere interactions, Journal of Hydrometeorology, 12, 1299–1320, 2011.

Luo, Y., Zhang, R., Wan, Q., Wang, B., Wong, W. K., Hu, Z., Jou, B. J.-D., Lin, Y., Johnson, R. H., Chang, C.-P., Zhu, Y., Zhang, X., Wang, H., Xia, R., Ma, J., Zhang, D.-L., Gao, M., Zhang, Y., Liu, X., Chen, Y., Huang, H., Bao, X., Ruan, Z., Cui, Z., Meng, Z., Sun, J., Wu, M., Wang, H., Peng, X., Qian, W., Zhao, K., and Xiao, Y.: The Southern China Monsoon Rainfall Experiment (SCMREX), Bulletin of the American Meteorological Society, 98, 999 – 1013, https://doi.org/https://doi.org/10.1175/BAMS-D-15-00235.1, 2017.

Mishra, S. K., Anand, A., Fasullo, J., and Bhagat, S.: Importance of the Resolution of Surface Topography in Indian Monsoon Simulation, Journal of Climate, 31, 4879 – 4898, https://doi.org/https://doi.org/10.1175/JCLI-D-17-0324.1, 2018.

O'Gorman, P. A. and Dwyer, J. G.: Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events, Journal of Advances in Modeling Earth Systems, 10, 2548–2563, https://doi.org/https://doi.org/10.1029/2018MS001351, 2018.

Onishi, R., Hirai, J., Kolomenskiy, D., and Yasuda, Y.: Real-Time High-Resolution Prediction of Orographic Rainfall for Early Warning of Landslides, pp. 237–248, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-031-16898-7_17, 2023.

Rasp, S.: Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0), Geoscientific Model Development, 13, 2185–2196, https://doi.org/10.5194/gmd-13-2185-2020, 2020.

Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proceedings of the National Academy of Sciences, 115, 9684–9689, https://doi.org/10.1073/PNAS.1810286115, 2018.

Ren, S., He, K., Girshick, R., and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016.

Schumacher, V., Fernández, A., Justino, F., and Comin, A.: WRF high resolution dynamical downscaling of precipitation for the Central Andes of Chile and Argentina, Frontiers in Earth Science, 8, 328, 2020.

Schwartz, C. S.: Reproducing the September 2013 Record-Breaking Rainfall over the Colorado Front Range with High-Resolution WRF Forecasts, Weather and Forecasting, 29, 393 – 402, https://doi.org/https://doi.org/10.1175/WAF-D-13-00136.1, 2014.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J., Duda, M., Barker, D., et al.: A description of the advanced research WRF version 4, NCAR tech. note ncar/tn-556+ str, 145, https://doi.org/doi:10.5065/1dfh-6p97, 2019.

375 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., and Xiang-yu, H.: A Description of the Advanced Research WRF Model Version 4.3, National Center for Atmospheric Research: Boulder, CO, USA, https://doi.org/doi:10.5065/1dfh-6p97, 2021.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J., Duda, M., Barker, D., et al.: The source codes of WRF-V4.3 and WPS-V4.3 [Software]. Zenodo. https://doi.org/10.5281/zenodo.10039053, 2023.

380 Tao, S. Y.: Storm Rainfall in China (in Chinese)., Science Press, 1981.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Advances in neural information processing systems, 30, 2017.

Villalba-Pradas, A. and Tapiador, F. J.: Empirical values and assumptions in the convection schemes of numerical models, Geoscientific Model Development, 15, 3447–3518, https://doi.org/10.5194/gmd-15-3447-2022, 2022.

385 Wyngaard, J. C.: Toward Numerical Modeling in the "Terra Incognita", Journal of the Atmospheric Sciences, 61, 1816 – 1826, https://doi.org/https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2, 2004.

Xia, R. and Zhao, S.: Diagnosis and modeling of meso–scale systems of heavy rainfall in warm sector ahead of front in South China (middle part of Guangdong province) in June 2005, Chinese journal of atmospheric sciences/Daqi Kexue, 33, 2009.

Yao, Y., Zhong, X., Zheng, Y., and Wang, Z.: A Physics-Incorporated Deep Learning Framework for Parameteriza-
390 tion of Atmospheric Radiative Transfer, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 445, https://doi.org/https://doi.org/10.1029/2022MS003445, e2022MS003445 2022MS003445, 2023.

Yuval, J. and O'Gorman, P. A.: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions, Nature communications, 11, 3295, 2020.

Yuval, J., O'Gorman, P. A., and Hill, C. N.: Use of neural networks for stable, accurate and physically consistent parameterization of subgrid
395 atmospheric processes with good performance at reduced precision, Geophysical Research Letters, 48, e2020GL091 363, 2021.

Zhang, X. and Ni, Y.: A comparative study of a frontal and a non-frontal convective systems, Acta Meteorologica Sinica, pp. 108–121, https://doi.org/10.11676/qxxb2009.012, 2009.

Zhang, X., Bao, J.-W., Chen, B., and Huang, W.: Evaluation and Comparison of Two Deep Convection Parameterization Schemes at Convection-Permitting Resolution, Monthly Weather Review, 149, 3419 – 3432, https://doi.org/https://doi.org/10.1175/MWR-D-21-
400 0016.1, 2021.

Zhao, P., Zhang, R., Liu, J., Zhou, X., and He, J.: Onset of southwesterly wind over eastern China and associated atmospheric circulation and rainfall, Climate Dynamics, 28, 797–811, 2007.

Zheng, Y., Alapaty, K., Herwehe, J. A., Genio, A. D. D., and Niyogi, D.: Improving High-Resolution Weather Forecasts Using the Weather Research and Forecasting (WRF) Model with an Updated Kain–Fritsch Scheme, Monthly Weather Review, 144, 833 – 860,
405 https://doi.org/https://doi.org/10.1175/MWR-D-15-0005.1, 2016.

Zhong, L., Mu, R., Zhang, D., Zhao, P., Zhang, Z., and Wang, N.: An observational analysis of warm-sector rainfall characteristics associated with the 21 July 2012 Beijing extreme rainfall event, Journal of Geophysical Research, 120, 3274–3291, 2015.

Zhong, S. and Chen, Z.: The Impacts of Atmospheric Moisture Transportation on Warm Sector Torrential Rains over South China, Atmosphere, 8, https://doi.org/10.3390/atmos8070116, 2017.

410 Zhong, X., Ma, Z., Yao, Y., Xu, L., Wu, Y., and Wang, Z.: WRF–ML v1.0: a bridge between WRF v4.3 and machine learning parameterizations and its application to atmospheric radiative transfer, Geoscientific Model Development, 16, 199–209, https://doi.org/10.5194/gmd-16-199-2023, 2023a.

Zhong, X., Yu, X., and Li, H.: Machine Learning Parameterization of the Multi-scale Kain-Fritsch (MSKF) Convection Scheme and stable simulation coupled in WRF using WRF-ML v1.0 (Version 1.0) [Dataset] [Software]. Zenodo. https://doi.org/10.5281/zenodo.10032404, 415 2023b.