

## A Neural-network-based Alternative Scheme to Include Nonhydrostatic Processes in an Atmospheric Dynamical Core

Yang XIA, Bin WANG, Lijuan LI, Li LIU, Jianghao LI, Li DONG, Shiming XU, Yiyuan LI, Wenwen XIA, Wenyu HUANG, Juanjuan LIU, Yong WANG, Hongbo LIU, Ye PU, Yujun HE, Kun XIA

**Citation:** Xia, Y., and Coauthors 2024: A Neural-network-based Alternative Scheme to Include Nonhydrostatic Processes in an Atmospheric Dynamical Core, *Adv. Atmos. Sci.*, 41, 1083–1099. doi: [10.1007/s00376-023-3119-1](https://doi.org/10.1007/s00376-023-3119-1).

View online: <https://doi.org/10.1007/s00376-023-3119-1>

## Related articles that may interest you

[A 3D Nonhydrostatic Compressible Atmospheric Dynamic Core by Multi-moment Constrained Finite Volume Method](#)

Advances in Atmospheric Sciences. 2019, 36(10), 1129 <https://doi.org/10.1007/s00376-019-9002-4>

[Three-dimensional Fusion of Spaceborne and Ground Radar Reflectivity Data Using a Neural Network-Based Approach](#)

Advances in Atmospheric Sciences. 2018, 35(3), 346 <https://doi.org/10.1007/s00376-017-6334-9>

[Modification of the SUNFLUX Solar Radiation Scheme with a New Aerosol Parameterization and Its Validation Using Observation Network Data](#)

Advances in Atmospheric Sciences. 2017, 34(11), 1301 <https://doi.org/10.1007/s00376-016-6262-0>

[Tower-Based Greenhouse Gas Measurement Network Design——The National Institute of Standards and Technology North East Corridor Testbed](#)

Advances in Atmospheric Sciences. 2017, 34(9), 1095 <https://doi.org/10.1007/s00376-017-6094-6>

[Melt Pond Scheme Parameter Estimation Using an Adjoint Model](#)

Advances in Atmospheric Sciences. 2021, 38(9), 1525 <https://doi.org/10.1007/s00376-021-0305-x>

[A Hybrid Statistical-Dynamical Downscaling of Air Temperature over Scandinavia Using the WRF Model](#)

Advances in Atmospheric Sciences. 2020, 37(1), 57 <https://doi.org/10.1007/s00376-019-9091-0>



AAS Website



AAS Weibo



AAS WeChat

Follow AAS public account for more information

• Original Paper •

# A Neural-network-based Alternative Scheme to Include Nonhydrostatic Processes in an Atmospheric Dynamical Core

Yang XIA<sup>1,4</sup>, Bin WANG<sup>2,1,3,6</sup>, Lijuan LI<sup>2</sup>, Li LIU<sup>1</sup>, Jianghao LI<sup>5</sup>, Li DONG<sup>2,6</sup>, Shiming XU<sup>1</sup>, Yiyuan LI<sup>2,6</sup>, Wenwen XIA<sup>1</sup>, Wenyu HUANG<sup>1</sup>, Juanjuan LIU<sup>2,6</sup>, Yong WANG<sup>1</sup>, Hongbo LIU<sup>2</sup>, Ye PU<sup>2</sup>, Yujun HE<sup>2</sup>, and Kun XIA<sup>2</sup>

<sup>1</sup>Ministry of Education Key Laboratory for Earth System Modeling, and Department of Earth System Science, Tsinghua University, Beijing 100084, China

<sup>2</sup>State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

<sup>3</sup>Innovation Group 311020008, Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai 519000, China

<sup>4</sup>Shanghai Ecological Forecasting and Remote Sensing Center, Shanghai 200030, China

<sup>5</sup>Key Laboratory of Earth System Modeling and Prediction, China Meteorological Administration, Beijing 100081, China

<sup>6</sup>College of Ocean Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

(Received 9 June 2023; revised 19 October 2023; accepted 6 November 2023)

## ABSTRACT

Here, a nonhydrostatic alternative scheme (NAS) is proposed for the grey zone where the nonhydrostatic impact on the atmosphere is evident but not large enough to justify the necessity to include an implicit nonhydrostatic solver in an atmospheric dynamical core. The NAS is designed to replace this solver, which can be incorporated into any hydrostatic models so that existing well-developed hydrostatic models can effectively serve for a longer time. Recent advances in machine learning (ML) provide a potential tool for capturing the main complicated nonlinear-nonhydrostatic relationship. In this study, an ML approach called a neural network (NN) was adopted to select leading input features and develop the NAS. The NNs were trained and evaluated with 12-day simulation results of dry baroclinic-wave tests by the Weather Research and Forecasting (WRF) model. The forward time difference of the nonhydrostatic tendency was used as the target variable, and the five selected features were the nonhydrostatic tendency at the last time step, and four hydrostatic variables at the current step including geopotential height, pressure in two different forms, and potential temperature, respectively. Finally, a practical NAS was developed with these features and trained layer by layer at a 20-km horizontal resolution, which can accurately reproduce the temporal variation and vertical distribution of the nonhydrostatic tendency. Corrected by the NN-based NAS, the improved hydrostatic solver at different horizontal resolutions can run stably for at least one month and effectively reduce most of the nonhydrostatic errors in terms of system bias, anomaly root-mean-square error, and the error of the wave spatial pattern, which proves the feasibility and superiority of this scheme.

**Key words:** neural network, nonhydrostatic alternative scheme, atmospheric model, dynamical core

**Citation:** Xia, Y., and Coauthors, 2024: A neural-network-based alternative scheme to include nonhydrostatic processes in an atmospheric dynamical core. *Adv. Atmos. Sci.*, **41**(6), 1083–1099, <https://doi.org/10.1007/s00376-023-3119-1>.

## Article Highlights:

- The concept of “nonhydrostatic alternative” is raised here to avoid complicated implicit nonhydrostatic integration.
- A layer-wise neural network algorithm with five selected leading input features is proposed to develop a nonhydrostatic alternative scheme.
- The scheme performs well in reproducing the nonhydrostatic tendency and correcting nonhydrostatic errors in the hydrostatic version of the WRF.

---

## 1. Introduction

The hydrostatic equilibrium approximation, which assumes that the upward pressure gradient force is balanced

---

\* Corresponding authors: Bin WANG, Hongbo LIU  
Emails: [wab@lasg.iap.ac.cn](mailto:wab@lasg.iap.ac.cn), [hongboliu@mail.iap.ac.cn](mailto:hongboliu@mail.iap.ac.cn)

by the downward gravitational pull of the Earth, is one of the most important numerical assumptions in the atmospheric dynamical core to simplify the equations for good computational efficiency and numerical stability. It can alleviate the restriction of timestep size and improve the computational efficiency of models by filtering vertically propagating acoustic waves, which carry little energy but seriously restrict the timestep size and computational efficiency (Staniforth and Wood, 2008). Therefore, the hydrostatic approximation is adopted in many current atmospheric general circulation models (AGCMs) that serve as the atmospheric components of global climate system models and earth system models. However, this approximation is reasonable only when horizontal scales are much larger than vertical scales, and no longer applicable at scales of a few kilometers following the rapid development of high-performance computing architectures (Smolarkiewicz et al., 2001). There is an emerging trend to consider nonhydrostatic effects for better overall accuracy, the convenience of local mesh refinement, and other benefits. The development of nonhydrostatic models mainly involves two aspects: the vertical coordinate and time-integration scheme.

The vertical coordinates that are popular in the mainstream include terrain-following or hybrid coordinates based on height  $z$  and pressure, as well as floating Lagrangian coordinates (Lin, 2004) and isentropic hybrid coordinates (Dowling et al., 2006; Toy and Randall, 2009) which are only adopted in a few operational atmospheric models (Kasahara, 1974). A pressure-based coordinate system is adopted in nearly all hydrostatic models (Wang et al., 2004) since it can largely simplify numerical equations and make theoretical analysis of large-scale motions easier (Sutcliffe, 1947). However, since the pressure-based coordinates are only suitable for hydrostatic models, most nonhydrostatic models (Tomita and Satoh, 2004; Skamarock et al., 2012; Zängl et al., 2015) have to resort to height-based coordinates. It is simple and straightforward to transform the vertical coordinates for the primitive atmospheric equations but relatively complicated to perform numerical integration. In addition, there is an emerging trend to use the hydrostatic-pressure or mass coordinate system proposed by Laprise (1992), because it can make nonhydrostatic equations take a form that very closely parallels the hydrostatic equations with pressure-based coordinates. Overall, for developing a nonhydrostatic model based on a hydrostatic one, a modification is often required to significantly alter the equation form for the vertical coordinate.

Regarding the time-integration scheme, the most prominent techniques adopted in the operational nonhydrostatic models of the atmosphere can be categorized into two distinct schemes, Eulerian-based and Lagrangian-based time-integration schemes (Mengaldo et al., 2019). For Eulerian-based schemes, which mainly include split-explicit schemes (Klemp and Wilhelmson, 1978) and horizontally-explicit and vertically-implicit schemes (Bao et al., 2015), the separation of scales between the vertical and horizontal directions is a key factor to speed up the model integration and to econo-

mize computer resources for communication. The main disadvantage concerns the difficulty in ensuring computational stability, which greatly limits the time step and requires additional artificial damping. In contrast, Lagrangian schemes, typified by the semi-implicit semi-Lagrangian (SISL) scheme (Robert, 1981), perform better in terms of the computational efficiency to maximize time-to-solution performance. A SISL scheme has been adopted in most operational global nonhydrostatic numerical weather prediction models since the semi-Lagrangian method is unconditionally stable for solving the transport equation and semi-implicit time discretization can mitigate high-speed gravity waves (Davies et al., 2005; Wedi and Smolarkiewicz, 2009; Wood et al., 2014). The main drawback is the lower parallel efficiency and huge computational cost of the iterative 3-dimensional (3D) Helmholtz solver (Zerroukat et al., 2002; Lauritzen et al., 2010).

In summary, the development of a nonhydrostatic model from a hydrostatic one not only requires changing the vertical coordinate and subsequent formulation but also necessitates the numerical control of vertical acoustic wave propagation by implicit integration, which is computationally expensive in both development and debugging. Notably, the nonhydrostatic impact begins to enter the scene but is not significantly remarkable when the horizontal resolution reaches a relatively high level ( $O\sim 10$  km), like a “grey zone”. Consequently, the hydrostatic equilibrium equation is used in many current operational climate and medium-range weather forecasting models, despite their horizontal resolutions falling into this grey zone.

The above discussion raises the question as to whether alternative schemes can be developed to replace the traditional nonhydrostatic solution approaches so that they can be easily incorporated into any hydrostatic models with horizontal resolutions in the grey zone, like a physical process parameterization. Such alternative schemes are expected to eliminate most of the nonhydrostatic errors in hydrostatic models without requiring a complicated implicit integration and also serve to extend the serving time of existing well-developed hydrostatic models, thereby saving the costs of developing a new nonhydrostatic model in the grey zone. Zhang et al. (2017) made an effort to develop an alternative scheme that incorporated the impacts of nonhydrostatic perturbations into a hydrostatic model based on a linear extrapolation operator. This scheme successfully reduced the nonhydrostatic errors of the hydrostatic model, which demonstrates the feasibility of a nonhydrostatic alternative scheme (NAS). However, two serious problems remain in the scheme even though it is a useful attempt in this aspect. First, the improvement can only last for several days since a linear extrapolation cannot maintain a nonlinear evolution of nonhydrostatic processes for a long time. Second, the linear extrapolation operator requires the nonhydrostatic state at the second step in addition to the initial condition (IC) at the beginning of model integration. Obviously, these two problems greatly decrease the practicability of the scheme. Therefore, a successful solution to the above problems is key to

developing a practical NAS, and the single-step multivariate nonlinear extrapolation may be a candidate solution.

As a kind of multivariate nonlinear function approximator, machine learning-like neural networks (NNs) have grown in popularity in recent years (Li et al., 2019; Liu et al., 2022). This method has been explored for applications in the development of atmospheric models in terms of physical process parameterizations and partial differential equation (PDE) solvers since it can allow for more degrees of freedom than the traditional polynomial or power-law methods.

Chevallier et al. (2000) and Krasnopolsky et al. (2005) first used NNs to emulate radiation parameterizations, which sped up calculations without large sacrifices in predictive accuracy. In the last several years, more recent studies have emerged that focused on ocean subgrid parameterization and atmospheric convection parameterization schemes based on convection-resolving simulations and super-parameterizations with deep NNs (Brenowitz and Bretherton, 2018; Gentine et al., 2018; Beucler et al., 2020; Gettelman et al., 2021) and convolutional NNs (Bolton and Zanna, 2019; Han et al., 2020). Additionally, the physics-informed neural network (PINN) (Raissi and Karniadakis, 2018; Raissi et al., 2019) and its improved versions (Dwivedi et al., 2019; Jin et al., 2021; Ranade et al., 2021) have been widely implemented in PDE solvers. These methods, which can learn the PDE solutions simultaneously, can alleviate the limitation of training data and can be extended to different conditions without training. However, replacing the total dynamical core with them greatly increases the costs related to the modification of the entire model structure; thus, most recent studies are only based on some simple cases such as incompressible fluids. Therefore, a reasonable combination of NNs and traditional approaches to improve or even develop atmospheric dynamic cores is more practical and efficient, for example, using NNs to solve the nonhydrostatic processes, as in this study.

This study aims to develop a practical NAS for a hydrostatic dynamical core based on NNs. The remainder of this is organized as follows. Section 2 introduces the atmospheric model used in this study and the model output data used for training. The training and performance assessment of the NN emulators are presented in section 3. The application and evaluation of the improved hydrostatic solver with the NN-based NAS are described in section 4. Conclusions and associated discussions are provided in section 5.

## 2. Methodology

### 2.1. Model and experimental design

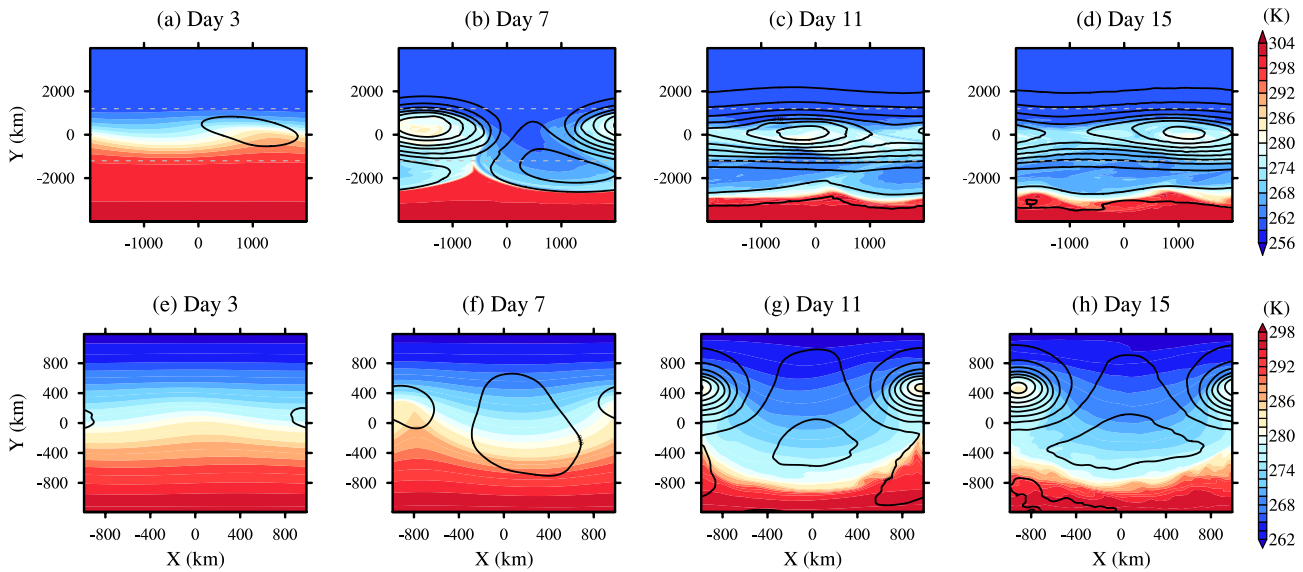
The atmospheric model used here is Version 3.9.1.1 of the Advanced Research Weather Research and Forecasting (WRF-ARW) model (Skamarock and Klemp, 2008) since it can switch from a hydrostatic solver (HDS) to a nonhydrostatic solver (NHDS) in a remarkably simple fashion. Here, an idealized 3D baroclinic-wave test case (Jablonowski and Williamson, 2006) in the idealized package was used as a rep-

resentative, which simulates the evolution of a 3D baroclinic wave within a baroclinically unstable jet in the northern hemisphere, under an  $f$ -plane approximation. This test was designed to target dry dynamical cores that are based on either the hydrostatic or non-hydrostatic primitive equations, which can represent the increased nonhydrostatic effect with horizontal resolution (Fig. S1 in the Electronic Supplementary Materials, ESM).

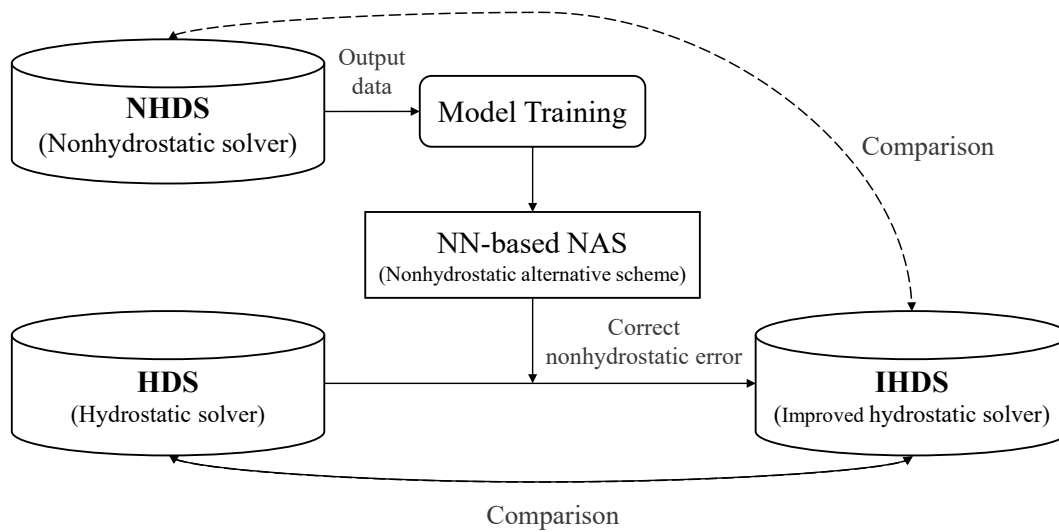
The original experimental design of this test in the idealized package for the domain size and resolution is fixed to a mesh of 41 points in the zonal direction and 81 in the meridional direction with a horizontal resolution of 100 km, which is difficult to directly adjust in the name list. To address the target horizontal resolution that falls in the grey zone, the ICs fixed in the idealized package should be interpolated to a 10-km resolution ( $x401 \times y801$ ). However, such a fine resolution may require high computational and storage costs. For this consideration, the 20-km resolution ( $x201 \times y401$ ) was chosen and only the central part ( $x101 \times y121$ ) of the domain, which basically covers the key centers of the strong westerly jet and temperature fronts (Fig. S2), was used to run the test case for the training of the NASs that were applied and evaluated at the resolutions of 20 km, 10 km, and 40 km. In addition, 50 vertical layers and a timestep of 60 s were adopted for all resolutions. Note that all physical parameterizations were switched off to eliminate the influence of any physical process, in an attempt to highlight the dynamics itself consistent with the original intention of Jablonowski and Williamson (2006) who proposed the idealized baroclinic-wave test. All model runs have a duration of one month.

To investigate the impact of reducing the domain size from  $4000 \text{ km} \times 8000 \text{ km}$  to  $2000 \text{ km} \times 2400 \text{ km}$ , Fig. 1 shows the wave evolutions of two experiments running in these two different domains with identical setups. In the original domain, a vortex system first closes at the surface layer on Day 3 (Fig. 1a), which is in great agreement with previous studies (e.g., Blázquez et al., 2013). Subsequently, the vortex gradually develops and expands with its movement from east to west. After an 11-day integration, the wave becomes stable and its magnitude stops growing, but the east-west movement still persists (Figs. 1c, d). Regarding the experiment in the reduced domain, the periodic boundary conditions in the  $x$ -direction and symmetric boundary conditions in the  $y$ -direction significantly slow down the wave evolution. Consequently, the vortex system does not close until Day 7 or later (Fig. 1f). Though the magnitude and phase speed have changed in the reduced domain relative to those in the original one, the overall wave pattern is still very similar to the original wave. Besides, the increase in the nonhydrostatic error with resolution (Fig. S3) is in accordance with that in the tests at the original domain (Fig. S1), jointly indicating that the experiment in the reduced domain can indeed represent the basic evolution of baroclinic waves.

The workflow is shown in Fig. 2. First, an NHDS was used to provide the samples of target variable and input features in the first 12 days for training and developing the



**Fig. 1.** The horizontal distributions of the surface pressure (hPa) (solid lines) and potential temperature (K) (color shades) from the baroclinic wave experiments by the nonhydrostatic dynamical core of the WRF using two different domains on (a, e) Day 3, (b, f) Day 7, (c, g) Day 11, and (d, h) Day 15. The top row shows the simulation results with the original domain (4000 km×8000 km), and the bottom row with the reduced domain (2000 km×2400 km). Both experiments adopted the same setups (horizontal resolution=20 km, timestep=60 s, etc.) except for the domain size. The two parallel dashed grey lines in each subfigure in the top row mark the meridional range of the reduction domain. The range of isobars is from 900 hPa to 1005 hPa, with intervals of 15 hPa for the top row and 5 hPa for the bottom row in the interest of a better view.



**Fig. 2.** The workflow of this study.

NAS. Then, the newly-developed NAS was applied to improve the HDS to correct the nonhydrostatic error. Finally, these three solvers were compared with each other for the evaluation of the new scheme. Therefore, all experiments included three runs by the nonhydrostatic, hydrostatic, and improved hydrostatic solvers (IHDS) of the WRF, which are named “NHDR”, “HDR” and “IHDR” for convenience, respectively.

**2.2. Input features and outputs**

According to the governing equations of the WRF model, the key difference between hydrostatic and nonhydro-

static dynamical core lies in the vertical momentum equation and the prognostic equation of geopotential height, which are written as:

$$\frac{\partial W}{\partial t} + (\nabla \cdot \mathbf{V}w) - g \left( \frac{1}{1+q} \frac{\partial p}{\partial \sigma} - \mu \right) = F_w, \quad (1)$$

$$\frac{\partial \phi}{\partial t} + \frac{1}{\mu} (\mathbf{V} \cdot \nabla \phi - gW) = 0, \quad (2)$$

where  $\sigma$  is the terrain-following coordinate based on HDP;  $W = \mu w$  is the mass-coupled vertical velocity in the height

coordinate;  $\mathbf{V} = (U, V, \Omega)$  represents the mass-coupled 3D flux-form velocities in the  $x, y, \sigma$  directions;  $\nabla = (\partial/\partial x, \partial/\partial y)$  is the horizontal gradient operator;  $g$  is the gravitational acceleration;  $p$  is the pressure;  $\mu$  is the mass per unit area;  $\phi$  is the geopotential height;  $q$  is the mixing ratio for moisture;  $F_w$  is forcing term that arises from model physics, turbulent mixing, spherical projections, and the Earth's rotation.

To solve the acoustic mode in the NHDS Eqs. (1) and (2) are combined to form a vertically implicit equation by replacing  $p$  in Eq. (1) with an expression for  $\phi$  in accordance with the hydrostatic relation and state equation, while in the HDS, Eq. (1) simplifies to:

$$g \left( \frac{1}{1+q} \frac{\partial p}{\partial \sigma} - \mu \right) = 0. \tag{3}$$

Within this simplification,  $p$  is first diagnosed using Eq. (3) and then the specific volume ( $\alpha$ ), geopotential height, and z-based vertical velocity ( $w$ ) are diagnosed in turn. Consequently, the elemental feature that determines the impact of the nonhydrostatic integration is whether the left-side term of Eq. (3) is equal to zero. Let  $S = \Delta\sigma(1/(1+q)\partial p/\partial\sigma - \mu)$ , which is the product of the vertical spacing ( $\Delta\sigma$ ), and the nonhydrostatic tendency (i.e., the difference between the vertical pressure gradient force and gravity) in the form of an HDP-based  $\sigma$  coordinate. The inclusion of vertical grid size in  $S$  aims to reduce the impact of the heterogeneous vertical layers on the magnitude of nonhydrostatic tendency at these layers. Thus,  $S$ , at time-level  $n+1$  (i.e.,  $S^{n+1}$ ), is the final target variable that the NN-based alternative scheme is going to obtain. For optimal performance, the forward time difference of  $S$ , i.e.,  $\Delta S = S^{n+1} - S^n$ , is also chosen as an intermediate target variable for the NAS. Note that all of the NN emulators discussed below adopt  $\Delta S$  as the target variable and that the final target,  $S^{n+1}$ , which works on the NAS directly, is obtained by the calculation from  $\Delta S$  as predicted by NN emulators.

Since there was little evidence concerning which variables significantly influence the variation of  $S$ , most of the basic variables of the WRF NHDR, including  $\phi, p, \mu, \Omega$ , and potential temperature  $\theta$ , were outputted to be examined by NN emulator methods. All of these variables are directly or indirectly related to the nonhydrostatic process. The ‘‘true’’

value of each variable,  $F$ , can be represented by the sum of its hydrostatic basal state  $\tilde{F}$  and a nonhydrostatic disturbing term  $F'$  (i.e.,  $F = \tilde{F} + F'$ ). The hydrostatic value at time-level  $n+1$  (i.e.,  $\tilde{F}^{n+1}$ ) and their forward time difference (i.e.,  $\Delta F = \tilde{F}^{n+1} - \tilde{F}^n$ ) constitute a list of candidates for input variables of NN models.

According to the integration sequence in the WRF, the solutions of  $\mu^{n+1}, \Omega^{n+1}$ , and  $\theta^{n+1}$  can be easily and directly obtained before the nonhydrostatic integration when the nonhydrostatic values of  $\phi^{n+1}$  and  $p^{n+1}$ , as well as  $S^{n+1}$  are calculated. Therefore, the nonhydrostatic disturbing terms of  $\mu^{n+1}, \Omega^{n+1}$ , and  $\theta^{n+1}$  are zero given a fixed initial value, i.e.,  $F = \tilde{F}$ . However, the hydrostatic values,  $\tilde{\phi}^{n+1}$  and  $\tilde{p}^{n+1}$ , which cannot be solved directly in the NHDS, requires an extra hydrostatic diagnosis to resolve them in every small time step before the nonhydrostatic integration. In addition to the above variables, the target variable at the last time step  $t^n$  ( $S^n$ ), was also included in the list of input variables. To consider vertical interaction, the vertical gradients of the above variables are calculated and included in the list, except for  $\mu$ , which does not vary in the vertical direction, and  $\tilde{p}^{n+1}$  whose vertical gradient equals  $\tilde{\mu}^{n+1}$  based on the definition of hydrostatic pressure. The complete list is shown in Table 1.

### 2.3. Training datasets and assessment metrics

The first 12-day outputs of all input and target variables from NDHR were used for training and assessment. In consideration of computation efficiency, independent vertical profiles of the input and target variables were randomly sampled from the entirety of the data with a varying number per day. The number of samples selected per day varied from 16,000 in the first 6 days to 100,000 in the last 6 days to maximize the representativeness of the samples. Because the surface vortex system in the baroclinic-wave test closes on Day 7 when the nonhydrostatic impact begins to become significant, this sampling method can increase the proportion of the samples that receive high nonhydrostatic impact from the last 6 days, while at the same time not ignoring the samples from the first 6 days when the system develops. To independently verify the NN emulator, the first 5 days and Days 7 to 11 were used for training (580,000 elements per vertical layer), and the remaining days (Days 6 and 12) (232,000 elements) were used for evaluation. All input and output values were

**Table 1.** The 19 candidate input features in the NN. For each variable  $F$ , the prefixes ‘‘ $\delta_\sigma$ ’’ and ‘‘ $\Delta$ ’’ indicate the vertical gradient and time difference, respectively. The superscript ‘‘ $n+1$ ’’ indicates the value at time-level  $n+1$ . Note that the vertical gradients of  $\mu$  and  $\tilde{p}^{n+1}$  are excluded here because  $\mu$  does not vary in vertical direction and the vertical gradient of  $\tilde{p}^{n+1}$  is equal to  $\tilde{\mu}^{n+1}$  based on the definition of hydrostatic pressure.

Variable		Input features		
Geopotential height: $\phi$	$\tilde{\phi}^{n+1}$	$\delta_\sigma \tilde{\phi}^{n+1}$	$\Delta \phi$	$\delta_\sigma \Delta \phi$
Pressure: $p$	$\tilde{p}^{n+1}$	–	$\Delta p$	$\delta_\sigma \Delta p$
Potential temperature: $\theta$	$\tilde{\theta}^{n+1}$	$\delta_\sigma \tilde{\theta}^{n+1}$	$\Delta \theta$	$\delta_\sigma \Delta \theta$
Mass per unit area: $\mu$	$\tilde{\mu}^{n+1}$	–	$\Delta \mu$	–
Vertical velocity in $\sigma$ directions: $\Omega$	$\tilde{\Omega}^{n+1}$	$\delta_\sigma \tilde{\Omega}^{n+1}$	$\Delta \Omega$	$\delta_\sigma \Delta \Omega$
Nonhydrostatic tendency: $S$	$S^n$	$\delta_\sigma S^n$	–	–

normalized to be dimensionless on each vertical layer.

The assessment score of the NN emulators adopted here was the coefficient of determination,  $R^2$ , which is defined as one minus the ratio of the mean squared error to the true variance, given as follows:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}, \quad (4)$$

where  $y$  represents the target fields from NHDR,  $\bar{y}$  is the mean of  $y$ , and  $\hat{y}$  is the estimated output of the NN. Except for the training score, i.e., the  $R^2$  calculated using the training dataset, the test score, the  $R^2$  calculated using the test dataset, represents a significant metric to assess the performance of NNs.

### 3. NN emulations

The fully connected NN consists of several interconnected levels of nonlinear nodes, which is capable of approximating arbitrary nonlinear functions (Rumelhart et al., 1986). We used the Python library Keras-Applications 1.0.8 (Chollet, 2015) for all NN experiments. The NN architecture consists of two fully connected levels with eight nodes for each experiment, as shown in Fig. 3. The internal activation algorithm is the nonlinear LeakyReLU activation function,

i.e.,  $\max(0.3x, x)$ , and that for the output level is a linear function. Each NN emulator was trained over 500 epochs with a batch size of 10,000 to avoid underfitting where each epoch is a complete cycle of NN-learning through the full training data set. Ultimately, the NN selects the best training result among the 500-epoch iterations that minimizes the loss function, i.e., the mean-squared-error.

The above hyperparameter choices were used to fit the target variable  $\Delta S$  with all 19 input features across all vertical layers in the whole-layer NN model. The vertical profiles of training and test scores are shown in Fig. 4a. The training scores decrease along with the vertical layers, especially among the top layers, which sharply turn to negative values

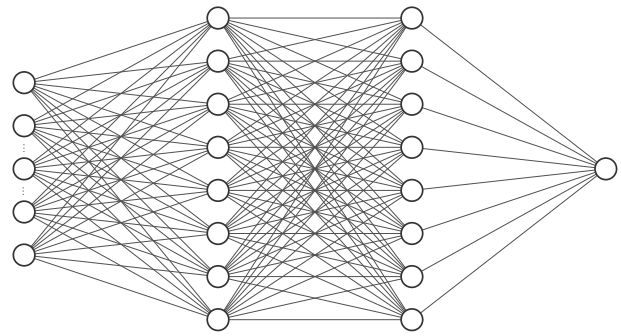


Fig. 3. Diagram showing the structure of NN.

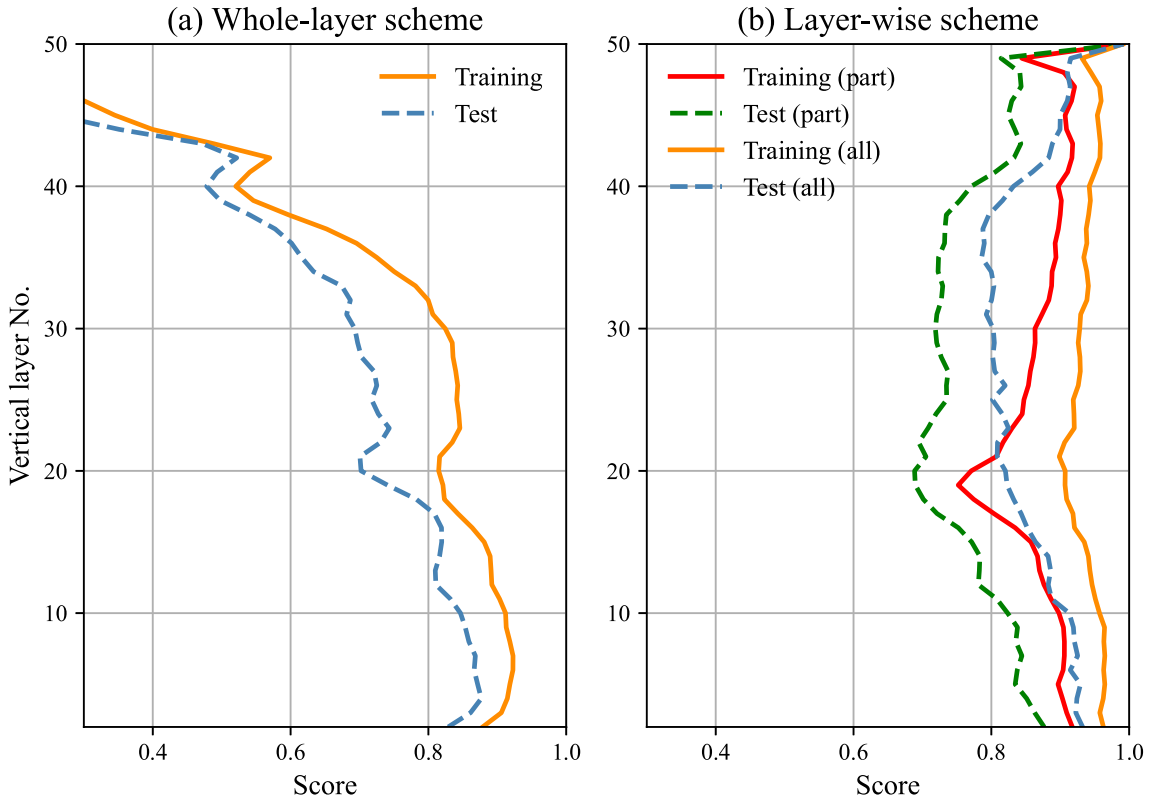


Fig. 4. The vertical profiles of the training score (solid lines) and the test score (dashed lines) of  $\Delta S$  from (a) the whole-layer NN scheme and (b) the layer-wise scheme. The orange and blue lines in both plots represent the score from the schemes that adopted all 19 input features. The red and green lines in (b) represent the score from the layer-wise scheme using the optimal choice of five input features selected ( $S^n$ ,  $\bar{\phi}^{n+1}$ ,  $\delta_\sigma \Delta p$ ,  $\Delta p$ , and  $\bar{\theta}^{n+1}$ ).

on the uppermost layer. The performance on test datasets is a bit inferior to those on training datasets but has a similar vertical distribution. The strong discrepancy in the vertical direction indicates that it is not a very reasonable way to train the same NN model on all vertical layers as a whole, which experiences difficulty in optimally fitting the target variables on each vertical layer. As a result, it is necessary to train NN emulators layer by layer, i.e., to train one emulator per vertical layer.

As shown in Fig. 4b, the layer-by-layer trained NN models obviously improve the training and test scores on all vertical layers, with values over 0.8 on nearly all layers. In addition, the strong discrepancy along the vertical direction in the whole-layer NN model has been also largely eliminated here. The models perform better on the bottom and top layers and slightly poorer on the middle layers. These improvements verify the superiority of the layer-by-layer training method.

However, such an approach is not economical and is easily overfitted when such a large number of input features are used in practice. To improve computational efficiency, it becomes necessary to select several leading input features which grasp the key components of nonhydrostatic information. Consequently, the “permutation importance” method was adopted to calculate feature importance and to find the leading features. This scheme basically works by randomly re-ordering the single feature series and assessing how the shuffled data affect the accuracy of predictions. Here, the importance is defined as the difference between the original metric and the metric from permutating the feature column (Breiman, 2001), which are normalized to a sum of 100% to conduct a fair comparison.

To match the layer-wise scheme, the feature importance

is also calculated layer by layer. On each vertical layer, all features were graded according to their rank in terms of feature importance. The first-, second-, and third-most important features on each layer score 3, 2, and 1 points, respectively, while the remaining features do not score. Figure 5a shows the vertical sum of all feature scores and their ranking. The most important feature is  $\Delta p$ , followed by  $\delta_\sigma \Delta p$ ,  $S^n$ , and  $\tilde{\phi}^{n+1}$ . All four features own a large value of importance and scores over 60, which differ from the others. Subsequently,  $\delta_\sigma \Delta \phi$  and  $\tilde{\theta}^{n+1}$  are two subdominant features, whose importance scores are over 30. However, upon considering their potential interaction and repeatability, the input features cannot be selected solely based on their ranks of importance in the NN, especially when some of them have close feature importance. Therefore, further analysis regarding this issue is necessary.

After removing most of the features with little importance, the above six features were left and used to train the new layer-wise NN models and to calculate feature importance again. The proportion of their contributions on each vertical layer is shown in Fig. 5b. The variables  $\Delta p$  and  $\delta_\sigma \Delta p$  nearly dominate the variation of the target variable together in most layers, especially the middle layers, which exactly explains why they are ranked at the top in Fig. 5a. Therefore, these two variables are very essential to the variation of the nonhydrostatic tendency and are worthy of inclusion into our NAS.

Among the remaining features,  $S^n$  shows its importance mainly on a few layers around the 6th and the top two layers.  $\tilde{\phi}^{n+1}$  appears more frequently among the secondary significant features below the 17th layer and several layers around the 40th layer.  $\tilde{\theta}^{n+1}$  becomes significant on the lower-

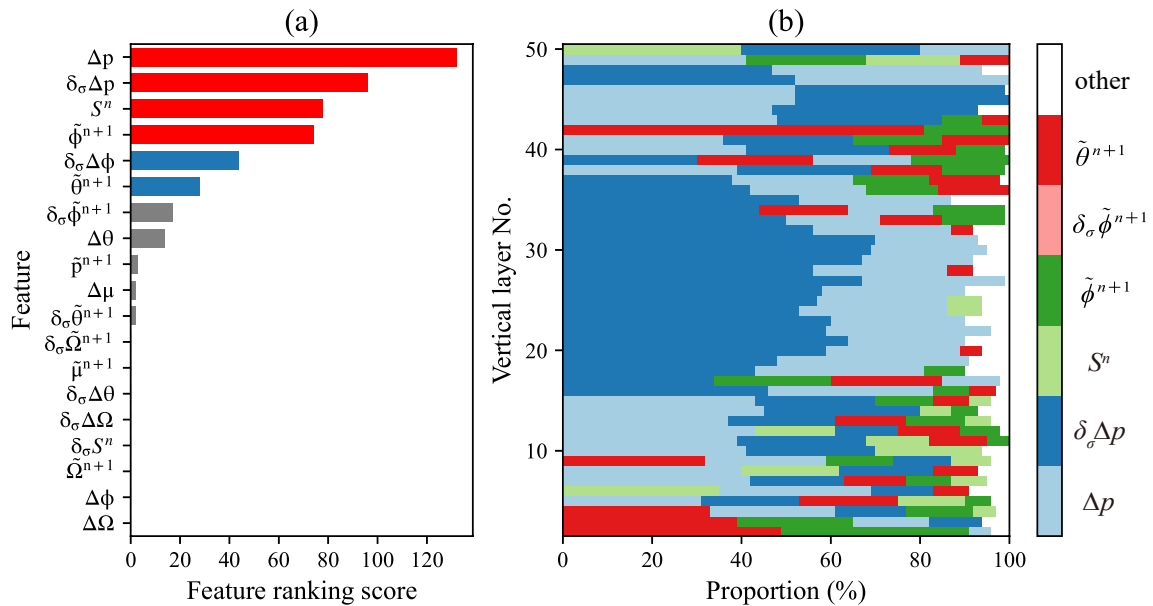


Fig. 5. (a) The vertical sum of feature ranking scores based on the importance of all 19 input features. The features that score over 60 are in red, those over 30 are in blue, and the others are in grey. Panel (b) shows the proportion of the contribution of the six leading features on each vertical layer. Only features whose importance values are not smaller than 5% are shown here, and other less important variables are filled with a blank.



most 3 layers and several layers around the 9th and 41st. Notably,  $\delta_\sigma \Delta \phi$ , which originally ranked fifth in Fig. 5a, fails to play an important role on any vertical layer. Its importance was previously overestimated under the interference of other unimportant features.

As a whole, the optimal choice of features on all vertical layers includes  $S^n$ ,  $\tilde{\phi}^{n+1}$ ,  $\delta_\sigma \Delta p$ ,  $\Delta p$ , and  $\tilde{\theta}^{n+1}$ , which are adopted uniformly but trained separately on each layer in an NN-based NAS for feature unification. Based on trial and error of different features, this particular scheme was also found optimal under the comprehensive consideration of all vertical layers. The vertical profiles of scores for this scheme are shown in Fig. 4b. The scheme performs closely to the one using all 19 features, only slightly worse in all layers within a reduction of below 0.1.

From the perspective of the physical significance of these leading features,  $S^n$  becomes significant because it is directly the value of the target variable at the previous step and represents the time memory of nonhydrostatic information. Other features like  $p$ ,  $\phi$ , and  $\theta$  are also critical in the training because these variables are involved directly in the implicit, nonhydrostatic integration process (Eqs. 1, 2). In particular, pressure is the elemental variable that directly determines the existence of the hydrostatic approximation. Geopotential height is the most relevant prognostic variable in the nonhydrostatic core but it turns into a diagnostic variable in the hydrostatic core. Potential temperature is the common prognostic variable in both the hydrostatic and nonhydrostatic cores, which carries a lot of information from the numerical integration.

#### 4. Application and evaluation of the NN-based NAS

This section focuses on the evaluation of the NN emulator in improving the WRF hydrostatic dynamical core as an alternative scheme of the nonhydrostatic solver at different horizontal resolutions of 20 km, 40 km and 10 km. The performance of this scheme is critical for its practicality. Here, the layer-wise NN emulators were used in the improved HDS (IHDS, i.e., the NN-based NAS that is incorporated into the HDS) to predict  $\Delta S$  and then  $S^{n+1}$ . Subsequently, the related nonhydrostatic disturbing variables, including  $p'$ ,  $\alpha'$ , and  $\phi'$ , were successively diagnosed based on the definition of  $S$ , the state equation, and the hydrostatic relation. Finally, all of the disturbances were added to their hydrostatic basic state so that the original HDS was corrected and improved to get close to the NHDS.

##### 4.1. Test at a 20-km resolution

Since the NNs were trained based on the output data from the 20-km test, the performance of the NN-based NAS should first be evaluated using the same resolution. The nonhydrostatic errors of the HDS and IHDS are calculated to be the difference between the HDR and NHDR and that between the IHDR and NHDR, respectively. The performances of the HDS and IHDS are evaluated mainly by two

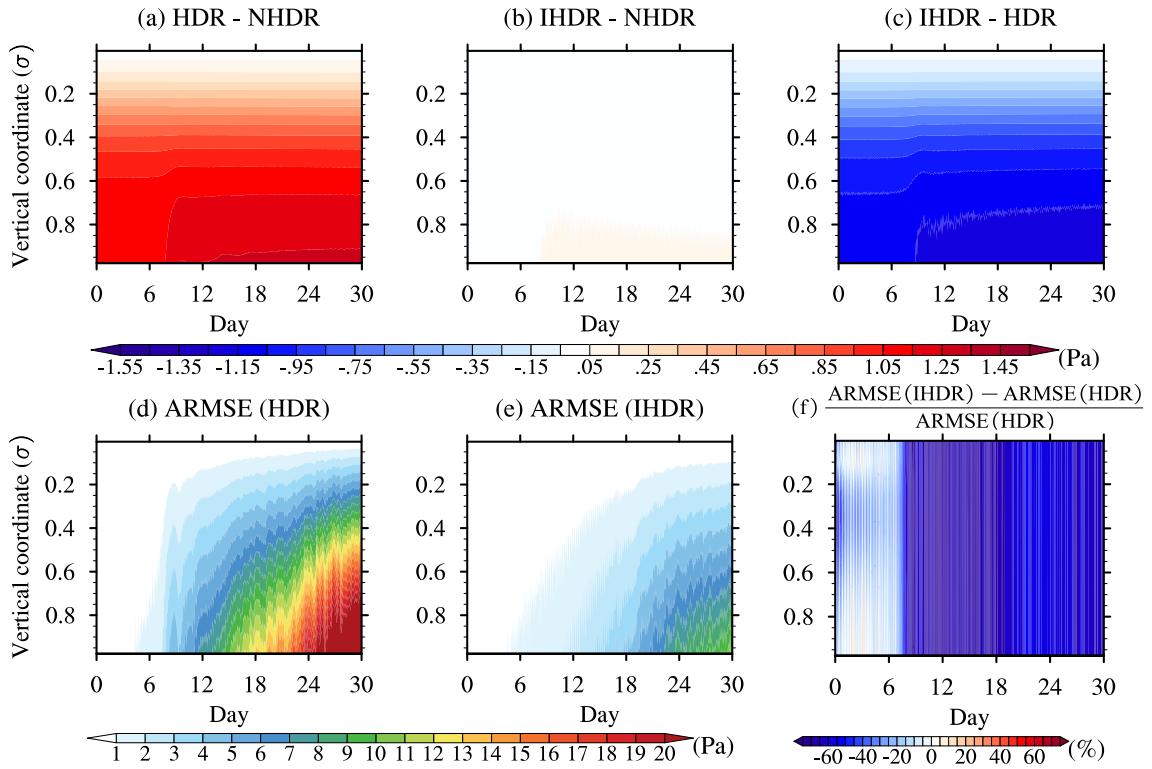
metrics: the nonhydrostatic systematic error (i.e., bias) and anomaly root-mean-square error (ARMSE). Both are calculated in horizontal dimensions on each vertical layer. Since  $p$  and  $\phi$  are two important diagnostic variables of the alternative scheme, their biases and ARMSEs are plotted (Figs. 6, 7).

As Fig. 6a shows, there are positive pressure biases in the HDR relative to the pressure in the NHDR, which decrease as the vertical height increases, consistent with the magnitude of the pressure on each layer. The temporal variations of these biases are relatively small. The IHDR significantly reduces the positive biases in HDR, whose biases are nearly invisible on all vertical layers in the figure with considerably small values between  $-0.05$  and  $0.05$  Pa except for a very few layers near the surface (Fig. 6b). The same patterns and close values with opposite sign in Figs. 6a and 6c that present the differences between the HDR and NHDR and those between the IHDR and HDR, respectively, also provide useful evidence for the reduction of systematic error in IHDR.

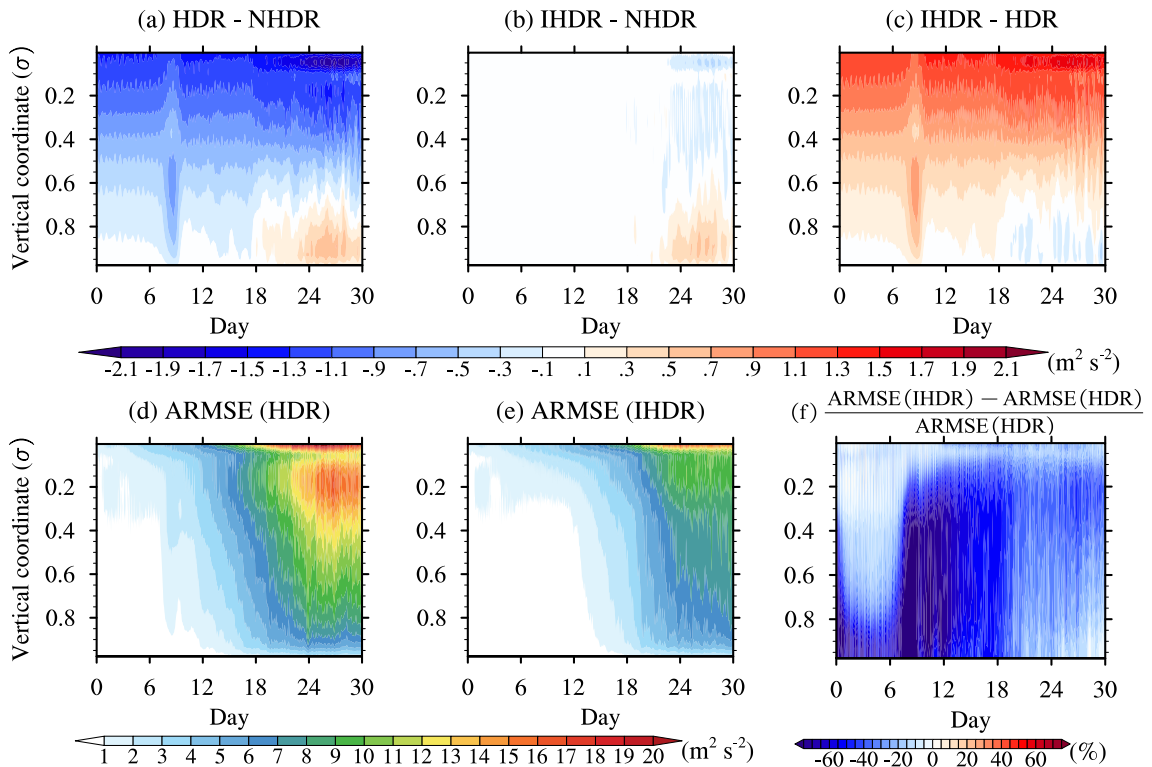
The ARMSEs of pressure in the HDR have a vertical distribution similar to that of the bias, larger in the lower layers and smaller in the upper layers, but with a much larger magnitude and a totally different temporal variation (Fig. 6d). It is indicative that the random errors dominate the nonhydrostatic errors of the HDR. These errors increase very slowly in the first 7 days, within a range of less than 1 Pa, before increasing sharply and rapidly thereafter on most of the vertical layers when the vortex closes and develops, consistent with the evolution of the baroclinic wave shown in Fig. 1. As stated above, the reduction of the integration domain slows down the wave evolution, causing the vortex system to finally close at the surface layer on approximately Day 7 or 8. Apparently, the nonhydrostatic effect begins to work only after the closure of the surface system. This is precisely the reason why the sampling number chosen in the first 6 days was set to be a smaller value in subsection 2.3.

With the correction of the NN-based NAS, the IHDR significantly reduces the ARMSEs of pressure in the HDR to a magnitude of under 12 Pa (Fig. 6e). The percentage of ARMSE reduction caused by the NN-based scheme in IHDR is shown in Fig. 6f. It can be observed that in the initial stage before Day 7 that the reduction percentage is pretty small due to the very small difference of the ARMSEs between the HDR and IHDR. However, after the evolution for 7 or 8 days, the originally sharply increased ARMSEs are reduced by over 60%, and similar reductions persist for the following 20 days, further noting that the magnitude of reduction varies little in the vertical direction.

Regarding  $\phi$ , negative system biases appear in the HDR on all vertical layers except for a few layers near the surface, which become larger after Day 20 (Fig. 7a). Different from the pressure, most of the time the geopotential height has larger biases on upper layers and smaller biases on lower layers. The IHDR greatly reduces these negative biases (Fig. 7b) and also decreases the positive biases near the surface after day 20 (Fig. 7c), whose biases are almost invisible in the figure with a narrow range of values from



**Fig. 6.** Time- $\sigma$  cross-section of the horizontally averaged systematic biases (the upper row) and anomaly root-mean-square error (the lower row) of pressure (Pa) (a and d) in the hydrostatic solver and (b and e) in the improved hydrostatic solver with the NN-based nonhydrostatic alternative scheme relative to the nonhydrostatic solver. Panel (c) shows the difference between the results of (a) and (b), and (f) is the percentage (%) of changes of (e) relative to (d), i.e.,  $\frac{\text{ARMSE (IHDR)} - \text{ARMSE (HDR)}}{\text{ARMSE (HDR)}}$ . The negative value (blue) in (f) means that the NN-based alternative scheme can reduce nonhydrostatic errors. The biases and errors at the initial time are zero.



**Fig. 7.** Same as in Fig. 6, but for the geopotential height (units:  $\text{m}^2 \text{s}^{-2}$ ).

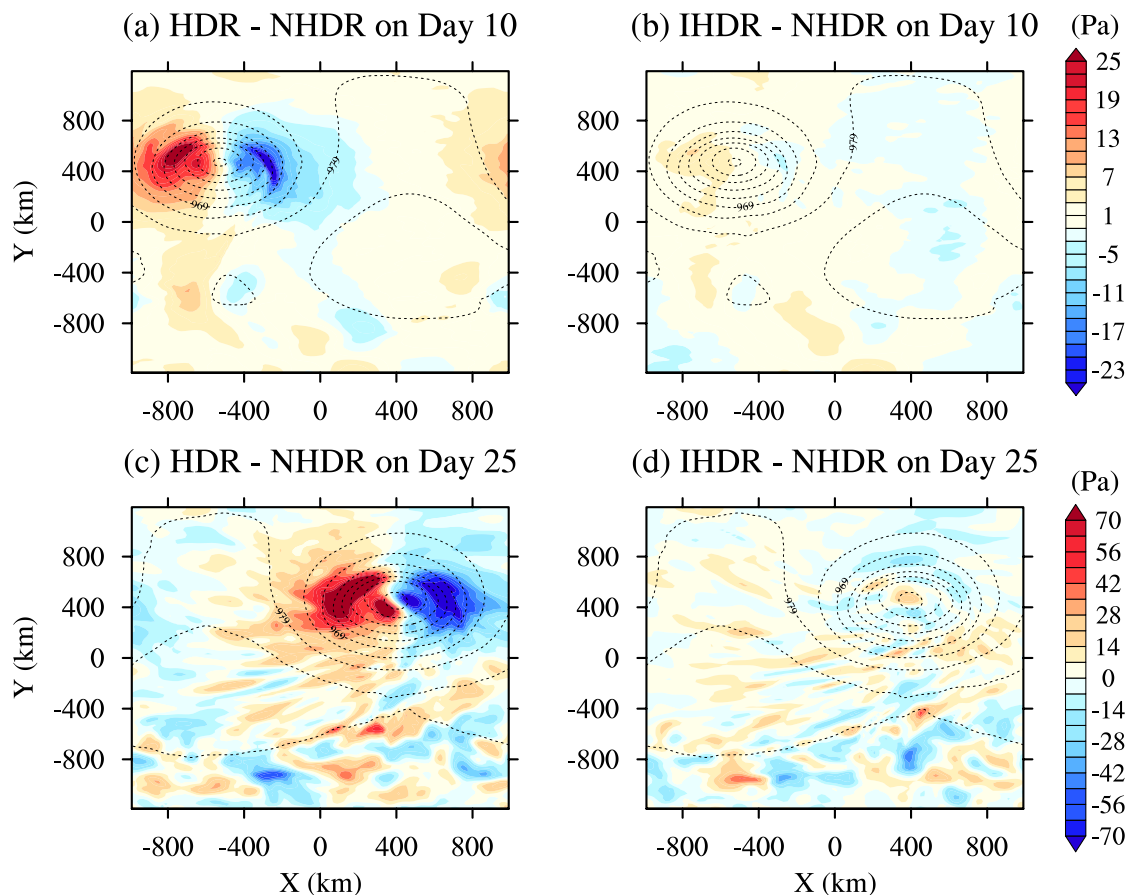
$-0.1$  to  $0.1 \text{ m}^2 \text{ s}^{-2}$  before Day 20.

The geopotential height also has a vertical distribution of ARMSEs (Fig. 7d) that differs from the pressure in the HDR (Figs. 6d), which are larger on upper layers and smaller on lower layers, consistent with that of the magnitude of geopotential height. The ARMSEs increase with time. They are obviously reduced in the IHDR but have a distribution similar to that in the HDR (Fig. 7e). Significant improvements appear on most vertical layers from Day 7 to Day 20 and near the surface before Day 7 (Fig. 7f). The overall improvements of geopotential height are smaller than those of pressure since pressure is the most essential variable to decide whether the hydrostatic assumption is tenable and the NN-based NAS is directly correct in the IHDS.

Since the closed vortex system that forms after Day 7 moves from east to west, the wave patterns of surface pressure and potential temperature in NHDR on Days 10 and 25 differ due to the different locations of the vortex, thus they are used to assess the performance of the IHDS. Relative to the NHDR, the significant nonhydrostatic errors of pressure in HDR appear along the eastern and western sides of the vortex center with negative and positive values, respectively (Figs. 8a, c). The errors on Day 25 are much larger than

those on Day 10. Compared to the HDR, the IHDR obviously reduces the pressure errors (Figs. 8b, d). On Day 10, the IHDR has an error distribution similar to the HDR, with negative errors along the eastern side and positive errors along the western side of the vortex. However, on Day 25, the error pattern in the IHDR is quite different from that in the HDR, where the positive and negative errors are more irregularly distributed.

Regarding potential temperature, the most significant nonhydrostatic errors in the HDR are mainly located in the zone where the potential temperature gradients are large, including along the northern side of the vortex and to the south of the domain (Figs. 9a, c). In particular, these errors increase and further extend to the south on Day 25, compared to those on Day 10. Apparently, the error distribution in the vortex has a totally reversed pattern compared to that of pressure, where negative and positive errors appear along the northeastern and northwestern sides of the vortex, respectively. Corrected by the NN-based alternative scheme, the errors in the IHDR are largely reduced, particularly inside the vortex on Day 10 (Figs. 9b, d). The improvements in the southern portion of the domain are less significant on Day 25 than on Day 10.



**Fig. 8.** The horizontal distributions of the surface pressure errors (Pa) of the (a and c) hydrostatic solver and (b and d) improved solver with the NN-based nonhydrostatic alternative scheme relative to the nonhydrostatic solver. The top row shows the simulation results on Day 10 (a and b) and the bottom row on Day 25 (c and d). The dashed lines represent the isobars (hPa) of the surface layer.

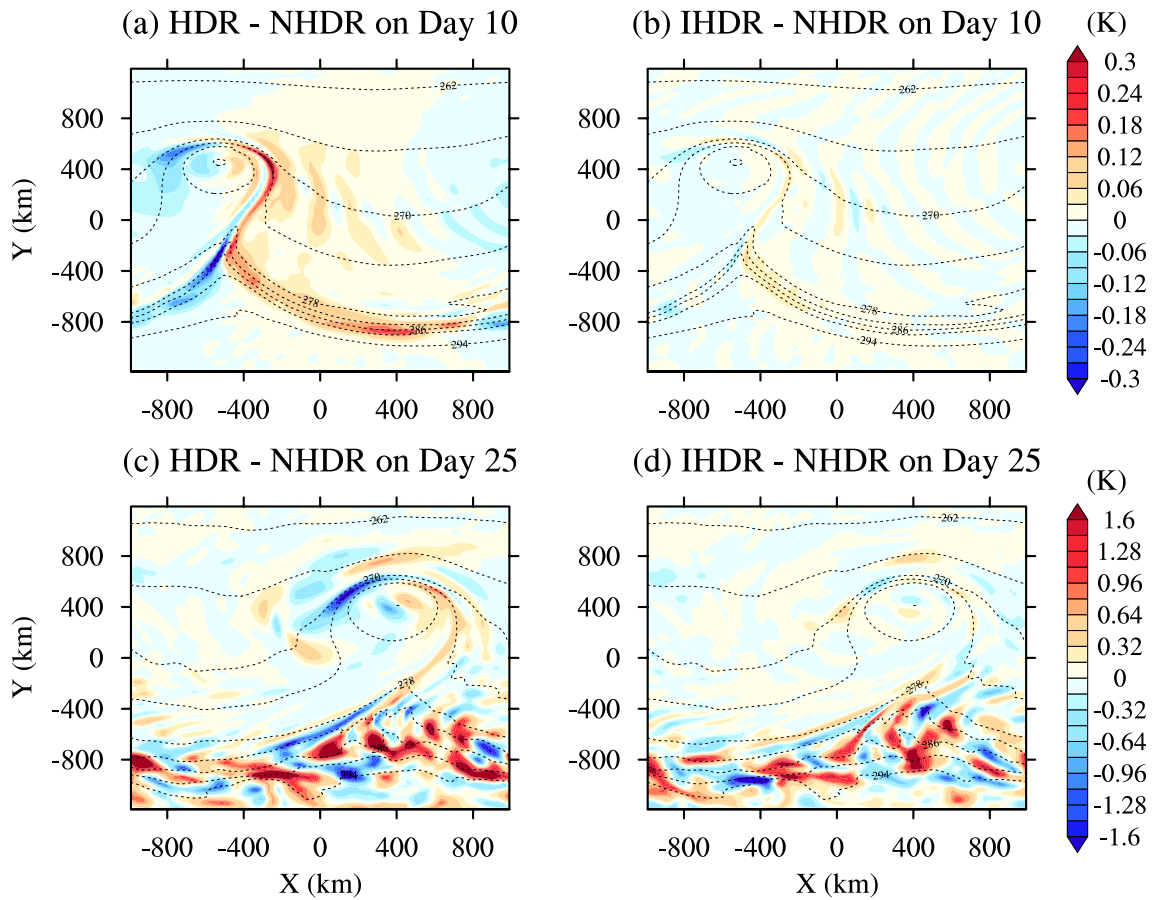


Fig. 9. Same as in Fig. 8, but for surface potential temperature (units: K).

4.2. Tests at different resolutions

To investigate the role of the scheme in improving the HDR in the tests at different resolutions, two experiments are conducted that apply the NAS trained at a horizontal resolution of 20 km to the solvers at 40 km and 10 km, respectively, and the performances of these two experiments are evaluated in this subsection. It is crucial to verify whether the scheme works well in different resolutions. For simplicity, the results of pressure (Figs. 10–13) are used as an example for the evaluation.

For the 40 km-resolution experiment, the overall patterns and magnitude of the biases and ARMSEs in the HDR are similar to those tested at a 20-km resolution, where the ARMSEs are slightly smaller in the early stage and larger in the last stage (Figs. 10a, d). After incorporating the NAS, the IHDR at a 40-km resolution also significantly reduces the positive biases in the HDR (Fig. 10c), where only very slight positive biases appear near the surface after Day 7 (Fig. 10b). The ARMSEs are also significantly reduced to within a value of 3 Pa in the first 24 days (Fig. 10e), and the improvement is much more obvious than that in the 20-km test, which varies little in the vertical direction. The overall reduction rates in the 40-km test are nearly 80%, with an overall increase of 20% compared to the 20-km test (Fig. 10f).

As for the wave patterns of surface pressure on Days 10 and 25, the 40-km experiment produces error distributions

in the HDR similar to those by the 20-km experiment, only with a smoother pattern of errors outside the vortex (Figs. 11a, c). After the correction by the NAS, the IHDR obviously reduces the pressure errors on both days. Compared to the HDR where positive and negative errors are distributed symmetrically inside the vortex, positive errors are more dominant in the IHDR, where only the first and fourth quadrant of the vortex is surrounded by negative errors on Day 10 and Day 25 (Figs. 11b, d), respectively. Overall, the performance of the NAS at a 40-km resolution is much better than that at a 20-km resolution, which indicates that the NAS can more effectively reduce the nonhydrostatic errors of a lower-resolution hydrostatic solver.

For the 10-km resolution experiment, the overall distribution of the system biases in the HDR and their significant correction by the NAS are slightly different from those at 20-km and 40-km resolutions (Figs. 12a–c). The ARMSEs in the HDR show a similar pattern and larger magnitude compared to the 20-km resolution experiment. After incorporating the NAS, the ARMSEs are also reduced (Fig. 12e). Similar to the 20-km experiment, the 10-km experiment achieves large improvements from Days 7 to 11 with reduction rates of over 60%. Subsequently, the improvements are gradually weakened after Day 12, but the error reduction rates rebound to a relatively high level in the last stage (Fig. 12f).

Regarding the wave patterns, the error distribution in

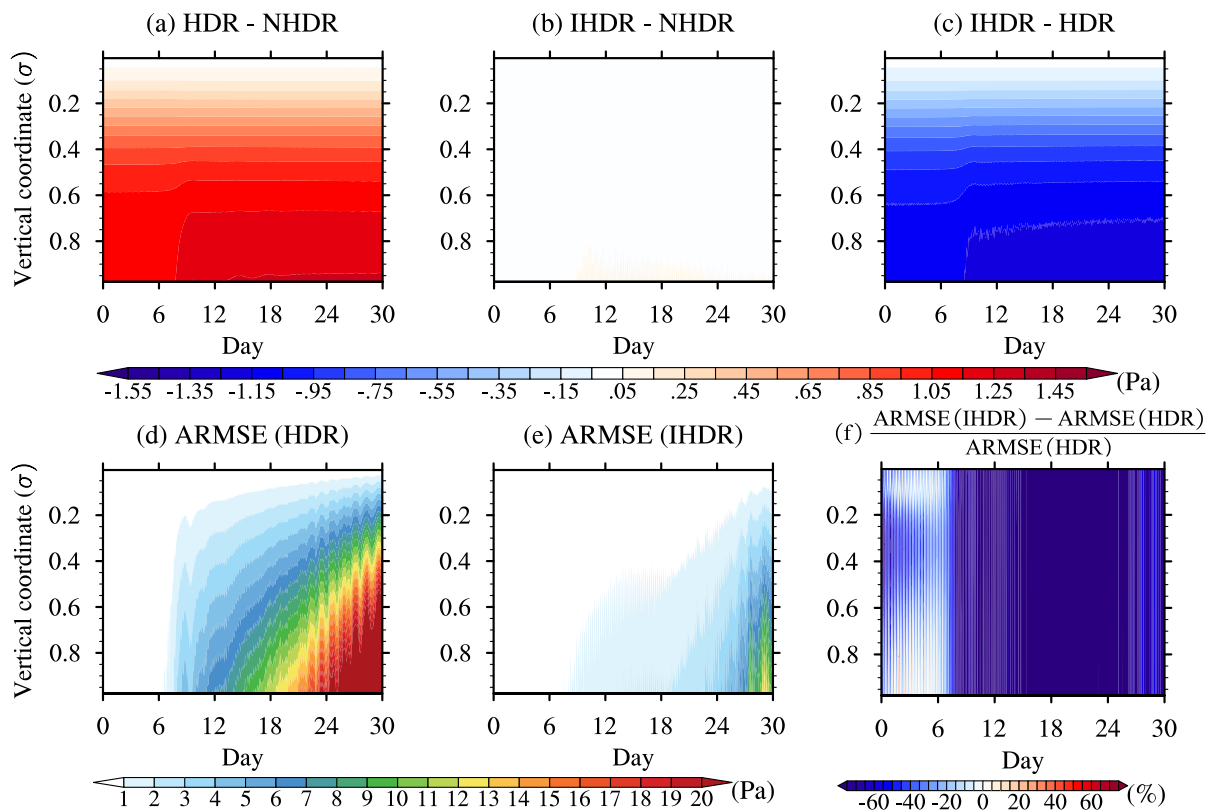


Fig. 10. Same as in Fig. 6, but for the 40-km resolution experiment.

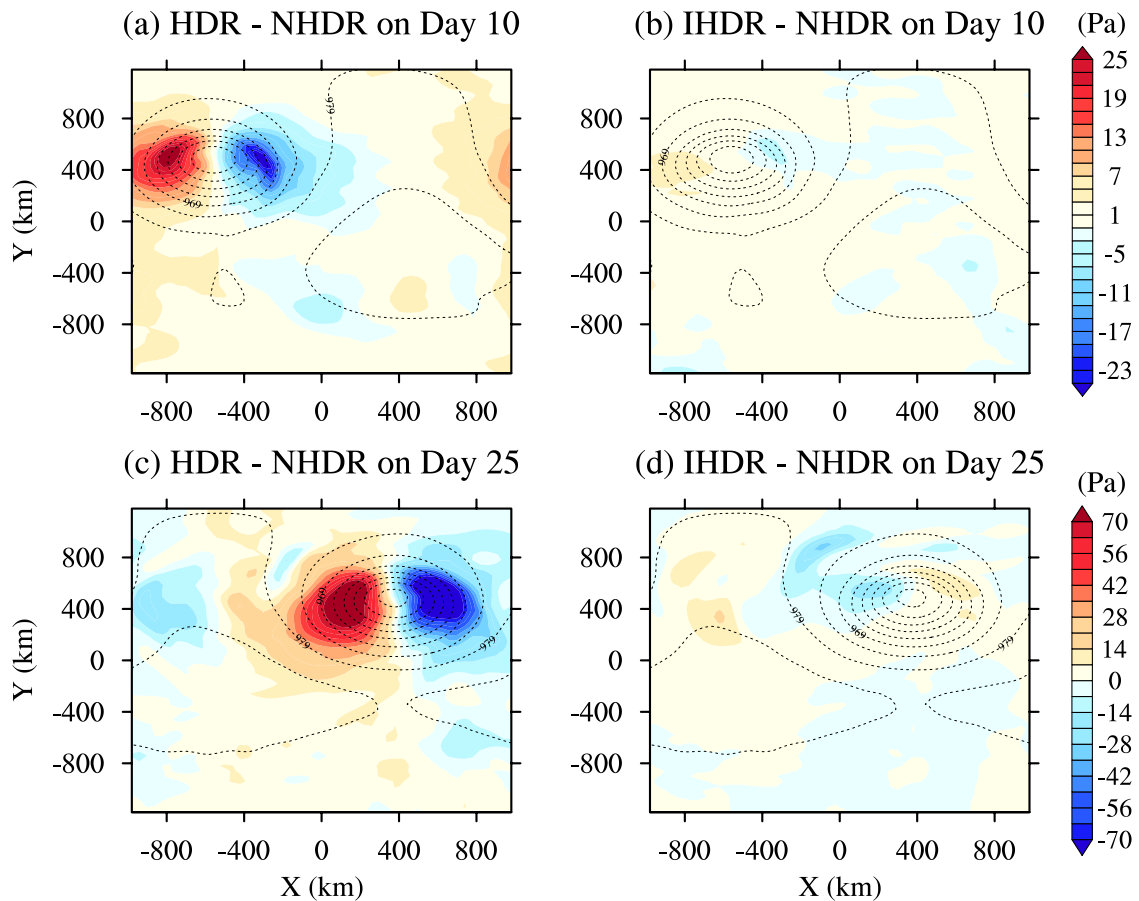


Fig. 11. Same as in Fig. 8, but for the 40 km-resolution experiment.

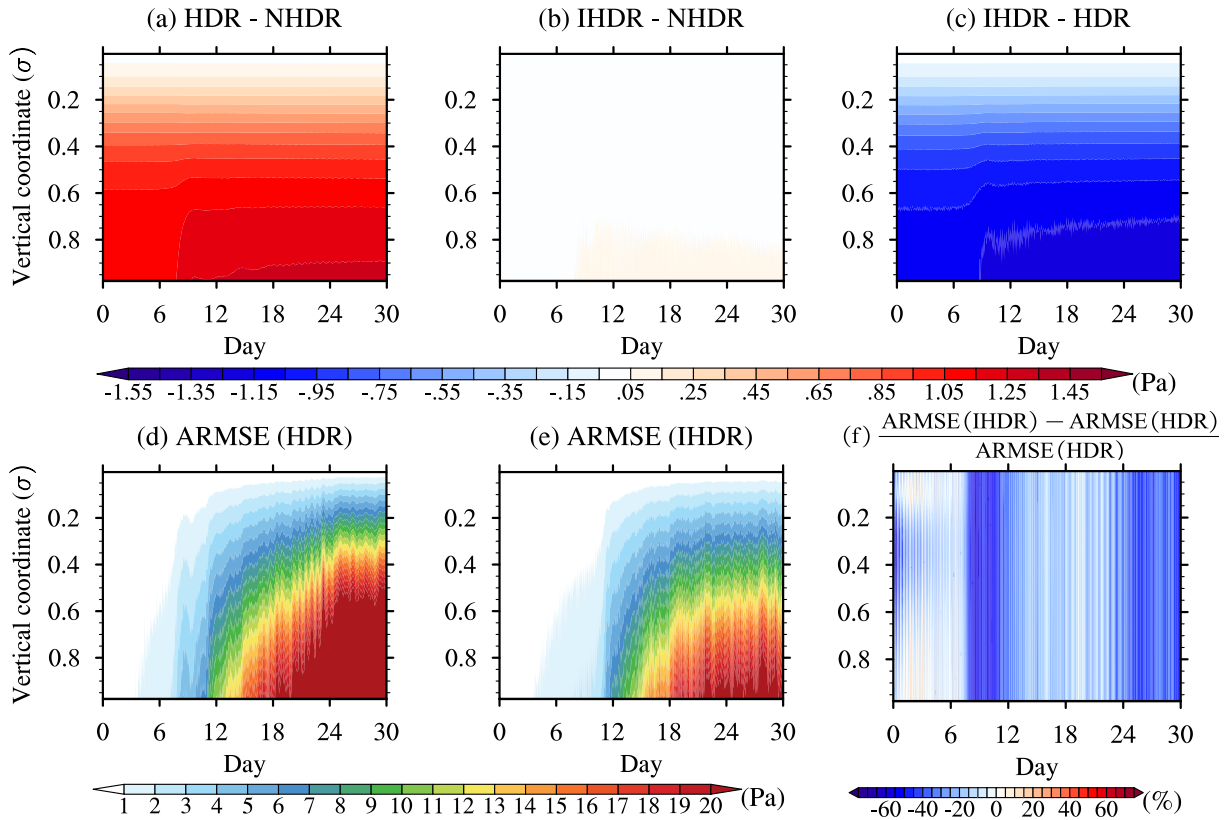


Fig. 12. Same as in Fig. 6, but for the 10-km resolution experiment.

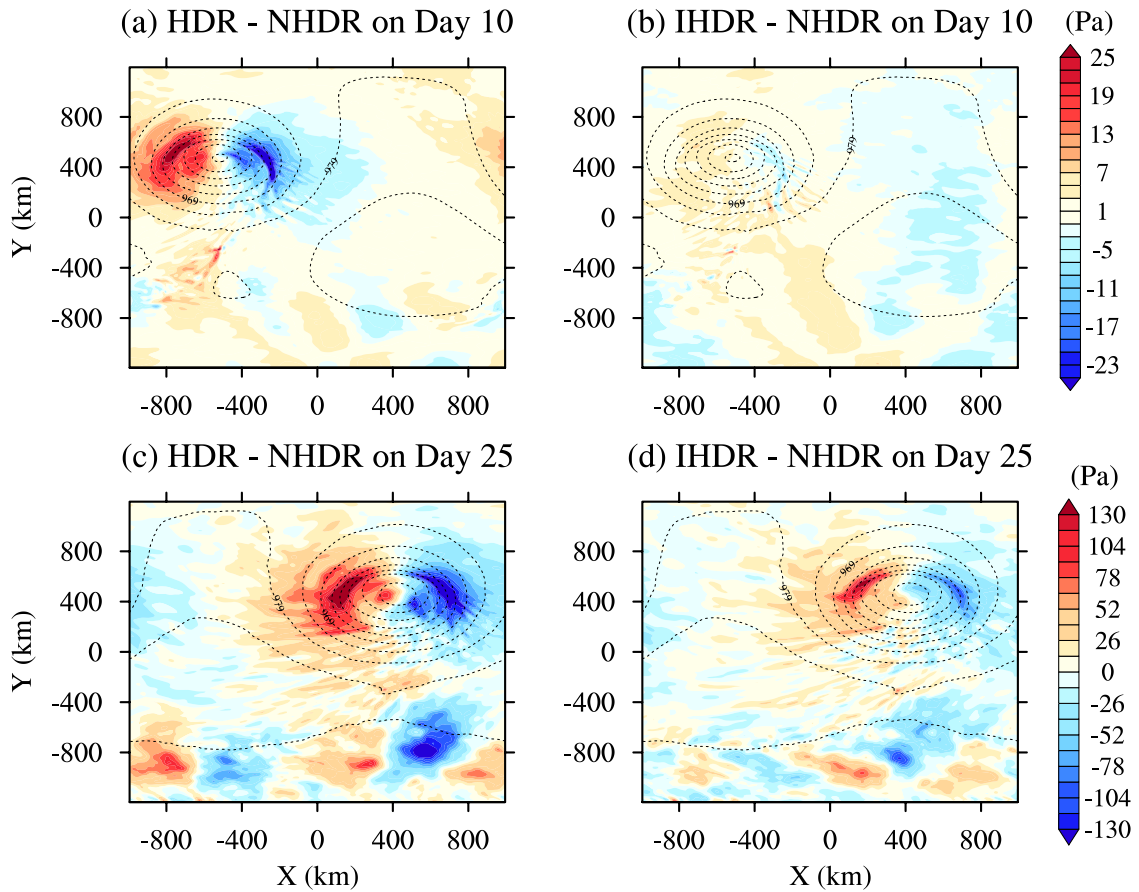


Fig. 13. Same as in Fig. 8, but for the 10-km resolution experiment.

the HDR on Day 10 is similar to that of the 20-km experiment (Fig. 13a). On Day 25, the overall magnitude of the errors is much larger and the pattern is more tanglesome than that in the 20-km experiment, particularly near the southern portion of the domain ( $y=-800$  km) (Fig. 13c). Compared to the HDR, the IHDR obviously reduces the pressure errors inside and outside the vortex, with a similar error pattern to that in the HDR (Figs. 13b, d). But the improvement on Day 25 is slightly weakened. Overall, the performance of the NAS in the 10-km experiment is encouragingly promising, especially during the development stage of the baroclinic wave, although it is not as good as those in the 20-km and 40-km experiments. This is to be expected because the NAS is trained at a coarser resolution and may inevitably lose some detailed information in a high-resolution information setting.

### 4.3. Comparisons between the NAS at 20 km and nonhydrostatic increments at different horizontal resolutions

To further validate the feasibility of this scheme, the corrections of the hydrostatic solver (IHDR-HDR) by the NAS at 20 km are also compared to the nonhydrostatic increments (NIs) that represent the difference between the nonhydrostatic

and hydrostatic solvers (NHDR-HDR) at both 40-km and 10-km resolutions in this subsection. Such comparisons are based on the fact that all nonhydrostatic information is included in NIs. The key point of these comparisons is to examine whether the NAS at 20 km can reproduce NIs at other resolutions. Note that for a uniform contrast to exist between two different resolutions, the simulations at the higher resolution are all interpolated to the lower resolution. Similar to the previous subsection, only the results of pressure (Figs. 14, 15) are used as an example for the evaluation.

For the experiment at a 40-km resolution, the NAS at 20 km can reproduce a correction very close to the NI at 40 km (Figs. 10a, 14a). The ARMSE of the NAS correction against the NI is also significantly reduced (Fig. 14b) compared to Fig. 10d. The reduction percentage in the first seven days, before the surface vortex system closes, is similar to the NAS at 40 km shown in Fig. 10f, where only some slight deterioration appears on several upper and bottom vertical layers. After evolving for 7 or 8 days, the magnitude of the reduction sharply increases with a value over 60% before Day 20. Subsequently, the reduction is slightly weakened for nearly a week before it rebounds in the last several days (Fig. 14c).

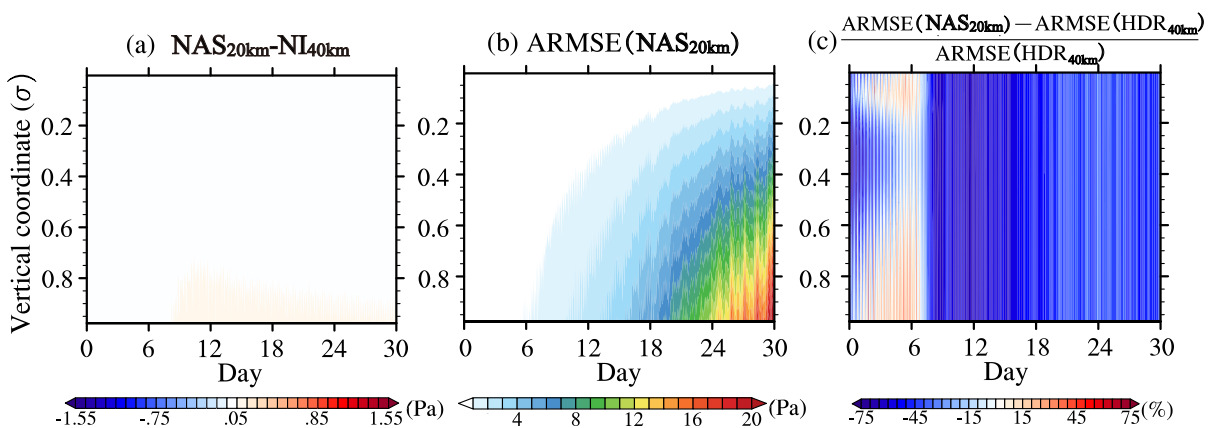


Fig. 14. Same as Figs. 10b, e, and f, but for  $\text{NAS}_{20\text{km}} - \text{NI}_{40\text{km}}$ . All simulation results at 20 km are uniformly interpolated to 40 km.

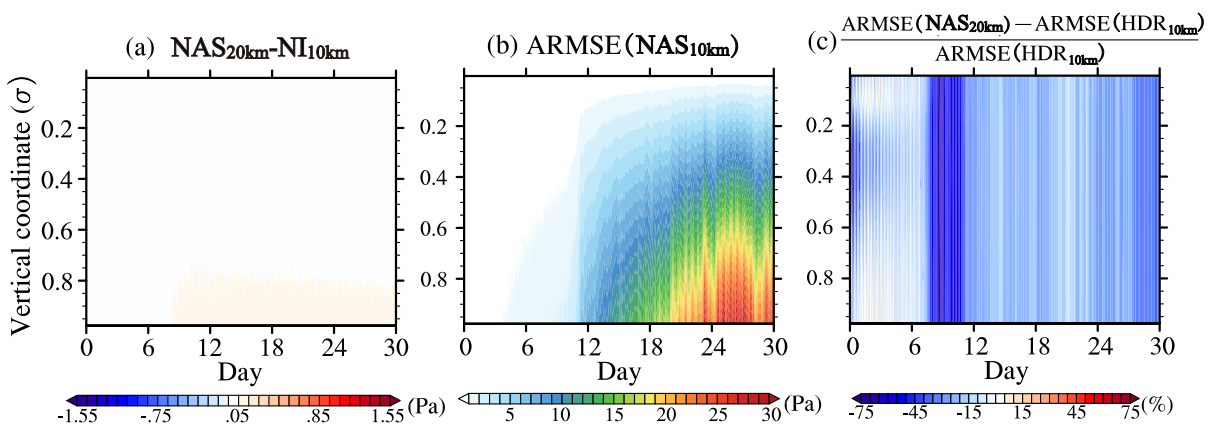


Fig. 15. Same as in Figs. 12b, e, and f, but for  $\text{NAS}_{20\text{km}} - \text{NI}_{10\text{km}}$ . All simulation results at 10 km are uniformly interpolated to 20 km.

Regarding the 10-km resolution experiment, the NAS at 20 km also achieves a correction close to the NI at a 10-km resolution with a reduced ARMSE against the NI shown in Figs. 12a and c, respectively. The reductions of the ARMSE also approached 60% during the developmental stage of the baroclinic wave from Day 7 to 11. The only exception is that the reduction rates gradually weakened from Day 12 before they rebounded slightly in the last several days (Fig. 15c), which is similar to the NAS at 10 km (Fig. 12f), as was explained in the last subsection. Overall, the corrections brought by the NAS at 20 km can well reproduce the corrections close to the NIs at both 10-km and 40-km resolutions, which demonstrates that the new scheme can grasp the key characteristic of nonhydrostatic processes across different resolutions.

In summary, the overall performances of the IHDS on pressure, geopotential height, and potential temperature in the multi-resolution experiments prove that this NN-based NAS can run stably for at least one month, and results in significant reductions of the nonhydrostatic errors in the HDR, no matter what the temporal and spatial dimensions. In particular, the good performance of the NAS at different resolutions suggests that ML approaches offer great potential and represent a bright prospect in improving and developing dynamic cores of atmospheric models.

## 5. Conclusions and discussion

An NN-based NAS was proposed so that the complicated process of implicit nonhydrostatic integration can be avoided, which is designed to improve any existing hydrostatic models. As such, this method can further extend the serving time of well-developed hydrostatic models in the “grey zone” where the horizontal resolution is not fine enough to consider the nonhydrostatic impact on model dynamics despite its visibility.

Fully connected NNs were adopted to perform feature engineering that selects leading input features and to develop an NN-based NAS to predict nonhydrostatic tendencies. They were trained and verified using historical data randomly sampled from the simulation results on the first twelve days of the baroclinic wave tests with the nonhydrostatic version of the WRF. The target variable was determined as the forward time difference of the product of the vertical spacing and nonhydrostatic tendency in the form under the HDP coordinate, i.e.,  $\Delta S$ . The five leading input features, selected by the NNs, include  $S^n$ ,  $\tilde{\phi}^{n+1}$ ,  $\delta_\sigma \Delta p$ ,  $\Delta p$ , and  $\tilde{\theta}^{n+1}$ . Finally, the NN was trained, layer by layer, to develop an NAS, with nearly 0.58 million samples on each vertical layer. On all vertical layers, the same five features are adopted with different weightings on various vertical layers.

The NAS was incorporated into the WRF hydrostatic dynamical core to correct its nonhydrostatic errors and this corrected core was applied to conduct the baroclinic-wave test which stably integrates for one month. It can well reproduce the nonhydrostatic tendency with high training and test scores on most vertical layers. As shown in the simulations

at multiple horizontal resolutions (10 km, 20 km, and 40 km), the scheme can correct the core so that it significantly and efficiently reduces the system bias, ARMSE, and error of the wave pattern.

The computational cost of the NAS is at the same level as that of the NHDS. The time costs of the three solvers demonstrate little significant distinction mainly because both the HDS and NHDS in the WRF are integrated within the unified framework using the same time-split integration scheme. Regarding the linkage between the NN and the model, there is still room for improvement. On the one hand, since the alternative schemes are independently trained layer by layer, there is great potential for improvement in terms of computational efficiency in simulations or predictions through parallel computing in the vertical direction. On the other hand, many methods exist which can be used to improve the computational efficiency of an NN. As a first step, this study mainly focuses on evaluating the feasibility and performance of the nonhydrostatic alternative schemes based on NNs. The optimization of the schemes and the prospects for increasing their computational efficiency require further study in the future.

In order to investigate the possibility of further increase of the NN performance, except for  $S^n$ ,  $S^{n-1}$  has been also included as an input feature, which represents the nonhydrostatic tendency at more previous time step. It did further increase the train and test scores of NNs, but failed to further improve or even degraded the performance of IHDS. As a consequence, the inclusion of  $S^{n-1}$  not only failed to further improve the performance of the NN-based NAS but also increased the computational cost. It's indicated that  $S^n$  is able to represent the most effective historical memory of the nonhydrostatic tendency.

This study used a relatively simple way to train an NN, which is a point-to-point mapping (i.e., using several features of a point to predict the single target variable of the same point). This scheme considers little spatial interaction, except that the vertical gradient of each variable at the same point was also included in the candidate list to represent part of the vertical interaction. The input features of an NN at each point only involve the output of basic variables from the WRF at this point and their post-processed variables (e.g., their vertical gradients and forward time differences) at the same point, which are fully independent of those on other points. In our opinion, using such a mapping method for the time being is more suitable for the following three reasons.

First, both the original nonhydrostatic implicit integration (Klemp and Wilhelmson, 1978) and the correction process of the NAS at each time step is completed in a single column, with large vertical exchanges but little horizontal ones. Consequently, the influence of horizontal interactions can be ignored in the early development of the NAS; thus, its input features do not include horizontal gradients of the basic variables but do include their vertical gradients. Second, more complicated NNs, like convolutional NNs (Han et al., 2020), can be used in the future to include horizontal



interactions in an NAS to further improve its accuracy. Although the cost of global communication would be increased, many methods can be used to reduce the overhead, such as using a GPU and overlapping communication and computation, which warrant further study in the future. Third, instead of the point-to-point mapping method, a column-to-column mapping method was also tested in which the vertical profile at each horizontal grid point is regarded as a whole as in many previous studies (e.g., Rasp et al., 2018); however, its training and test scores were remarkably low—smaller than 0.5. This test indirectly testifies to the practicability of the point-to-point mapping method in this case.

The good performance of the NASs is encouraging at different resolutions. However, more efforts are needed in the future to address how to extend the NN-based NAS to different cases (i.e., a real atmosphere experiment, etc.) and even with other hydrostatic models. Besides, the long-term numerical stability of a high-accuracy NAS is a challenging problem that warrants further study. Some efforts to improve the stability have been made (Beucler et al., 2020; Han et al., 2020; Jin et al., 2021) by adding physical constraints to the loss function of the NN. In addition, for regional tests on the real atmosphere, how to reasonably deal with the lateral boundary conditions in the NAS is also challenging.

**Acknowledgements.** This study is supported by the National Science Foundation of China (Grant No. 42230606). We acknowledge that the neural network codes used here are from Python Library Keras (Chollet, 2015). The training and testing data, associated codes, and application in the WRF are provided at Zenodo (<https://doi.org/10.5281/zenodo.6355086>). We also acknowledge Dr. Yilun HAN and Dr. Yan ZHANG for the useful discussion about machine learning and improvement in computational efficiency, respectively.

**Electronic supplementary material:** Supplementary material is available in the online version of this article at <https://doi.org/10.1007/s00376-023-3119-1>.

## REFERENCES

- Bao, L., R. Klöfkom, and R. D. Nair, 2015: Horizontally explicit and vertically implicit (HEVI) time discretization scheme for a discontinuous Galerkin Nonhydrostatic Model. *Mon. Wea. Rev.*, **143**, 972–990, <https://doi.org/10.1175/MWR-D-14-00083.1>.
- Beucler, T., M. Pritchard, P. Gentine, and S. Rasp, 2020: Towards physically-consistent, data-driven models of convection. Preprints, IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, IEEE, 3987–3990, <https://doi.org/10.1109/IGARSS39084.2020.9324569>.
- Blázquez, J., N. L. Pessacq, and P. L. M. Gonzalez, 2013: Simulation of a baroclinic wave with the WRF regional model: Sensitivity to the initial conditions in an ideal and a real experiment. *Meteorological Applications*, **20**, 447–456, <https://doi.org/10.1002/met.1307>.
- Bolton, T., and L. Zanna, 2019: Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, **11**, 376–399, <https://doi.org/10.1029/2018MS001472>.
- Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Brenowitz, N. D., and C. S. Bretherton, 2018: Prognostic validation of a neural network unified physics parameterization. *Geophys. Res. Lett.*, **45**, 6289–6298, <https://doi.org/10.1029/2018GL078510>.
- Chevallier, F., J. J. Morcrette, F. Chérury, and N. A. Scott, 2000: Use of a neural-network-based long-wave radiative-transfer scheme in the ECMWF atmospheric model. *Quart. J. Roy. Meteor. Soc.*, **126**, 761–776, <https://doi.org/10.1002/qj.49712656318>.
- Chollet, F., 2015: Keras. Accessed 30 August 2023, <https://keras.io>.
- Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **131**, 1759–1782, <https://doi.org/10.1256/qj.04.101>.
- Dowling, T. E., and Coauthors, 2006: The EPIC atmospheric model with an isentropic/terrain-following hybrid vertical coordinate. *Icarus*, **182**, 259–273, <https://doi.org/10.1016/j.icarus.2006.01.003>.
- Dwivedi, V., N. Parashar, and B. Srinivasan, 2019: Distributed physics informed neural network for data-efficient solution to partial differential equations. arXiv:1907.08967, <http://doi.org/10.48550/arXiv.1907.08967>.
- Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock?. *Geophys. Res. Lett.*, **45**, 5742–5751, <https://doi.org/10.1029/2018GL078202>.
- Gottelman, A., D. J. Gagne, C. C. Chen, M. W. Christensen, Z. J. Lebo, H. Morrison, and G. Gantos, 2021: Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems*, **13**, e2020MS002268, <https://doi.org/10.1029/2020MS002268>.
- Han, Y. L., G. J. Zhang, X. M. Huang, and Y. Wang, 2020: A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002076, <https://doi.org/10.1029/2020MS002076>.
- Jablonski, C., and D. L. Williamson, 2006: A baroclinic instability test case for atmospheric model dynamical cores. *Quart. J. Roy. Meteor. Soc.*, **132**, 2943–2975, <https://doi.org/10.1256/qj.06.12>.
- Jin, X. W., S. Z. Cai, H. Li, and G. E. Karniadakis, 2021: NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations. *J. Comput. Phys.*, **426**, 109951, <https://doi.org/10.1016/j.jcp.2020.109951>.
- Kasahara, A., 1974: Various vertical coordinate systems used for numerical weather prediction. *Mon. Wea. Rev.*, **102**, 509–522, [https://doi.org/10.1175/1520-0493\(1974\)102<0509:VVCUSUF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0509:VVCUSUF>2.0.CO;2).
- Klemp, J. B., and R. B. Wilhelmson, 1978: The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.*, **35**, 1070–1096, [https://doi.org/10.1175/1520-0469\(1978\)035<1070:tsotdc>2.0.co;2](https://doi.org/10.1175/1520-0469(1978)035<1070:tsotdc>2.0.co;2).
- Krasnopolsky, V. M., M. S. Fox-Rabinovitz, and D. V. Chalikov, 2005: New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of long-

- wave radiation in a climate model. *Mon. Wea. Rev.*, **133**, 1370–1383, <https://doi.org/10.1175/MWR2923.1>.
- Laprise, R., 1992: The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon. Wea. Rev.*, **120**, 197–207, [https://doi.org/10.1175/1520-0493\(1992\)120<0197:teomw>2.0.co;2](https://doi.org/10.1175/1520-0493(1992)120<0197:teomw>2.0.co;2).
- Lauritzen, P. H., R. D. Nair, and P. A. Ullrich, 2010: A conservative semi-Lagrangian multi-tracer transport scheme (CSLAM) on the cubed-sphere grid. *J. Comput. Phys.*, **229**, 1401–1424, <https://doi.org/10.1016/j.jcp.2009.10.036>.
- Li, H. C., C. Yu, J. J. Xia, Y. C. Wang, J. Zhu, and P. W. Zhang, 2019: A model output machine learning method for grid temperature forecasts in the Beijing Area. *Adv. Atmos. Sci.*, **36**, 1156–1170, <https://doi.org/10.1007/s00376-019-9023-z>.
- Lin, S. J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2).
- Liu, C., S. Yang, D. Di, Y. J. Yang, C. Zhou, X. Q. Hu, and B. J. Sohn, 2022: A machine learning-based cloud detection algorithm for the Himawari-8 spectral image. *Adv. Atmos. Sci.*, **39**, 1994–2007, <https://doi.org/10.1007/s00376-021-0366-x>.
- Mengaldo, G., A. Wyszogrodzki, M. Diamantakis, S. J. Lock, F. X. Giraldo, and N. P. Wedi, 2019: Current and emerging time-integration strategies in global numerical weather and climate prediction. *Archives of Computational Methods in Engineering*, **26**, 663–684, <https://doi.org/10.1007/s11831-018-9261-8>.
- Raissi, M., and G. E. Karniadakis, 2018: Hidden physics models: Machine learning of nonlinear partial differential equations. *J. Comput. Phys.*, **357**, 125–141, <https://doi.org/10.1016/j.jcp.2017.11.039>.
- Raissi, M., P. Perdikaris, and G. E. Karniadakis, 2019: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, **378**, 686–707, <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Ranade, R., C. Hill, and J. Pathak, 2021: DiscretizationNet: A machine-learning based solver for Navier-Stokes equations using finite volume discretization. *Computer Methods in Applied Mechanics and Engineering*, **378**, 113722, <https://doi.org/10.1016/j.cma.2021.113722>.
- Rasp, S., M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 9684–9689, <https://doi.org/10.1073/pnas.1810286115>.
- Robert, A., 1981: A stable numerical integration scheme for the primitive meteorological equations. *Atmosphere-Ocean*, **19**, 35–46, <https://doi.org/10.1080/07055900.1981.9649098>.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning representations by back-propagating errors. *Nature*, **323**, 533–536, <https://doi.org/10.1038/323533a0>.
- Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, <https://doi.org/10.1016/j.jcp.2007.01.037>.
- Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S. H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering. *Mon. Wea. Rev.*, **140**, 3090–3105, <https://doi.org/10.1175/MWR-D-11-00215.1>.
- Smolarkiewicz, P. K., L. G. Margolin, and A. A. Wyszogrodzki, 2001: A class of nonhydrostatic global models. *J. Atmos. Sci.*, **58**, 349–364, [https://doi.org/10.1175/1520-0469\(2001\)058<0349:ACONGM>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<0349:ACONGM>2.0.CO;2).
- Staniforth, A., and N. Wood, 2008: Aspects of the dynamical core of a nonhydrostatic, deep-atmosphere, unified weather and climate-prediction model. *J. Comput. Phys.*, **227**, 3445–3464, <https://doi.org/10.1016/j.jcp.2006.11.009>.
- Sutcliffe, R. C., 1947: A contribution to the problem of development. *Quart. J. Roy. Meteor. Soc.*, **73**, 370–383, <https://doi.org/10.1002/qj.49707331710>.
- Tomita, H., and M. Satoh, 2004: A new dynamical framework of nonhydrostatic global model using the icosahedral grid. *Fluid Dynamics Research*, **34**, 357–400, <https://doi.org/10.1016/j.fluidyn.2004.03.003>.
- Toy, M. D., and D. A. Randall, 2009: Design of a nonhydrostatic atmospheric model based on a generalized vertical coordinate. *Mon. Wea. Rev.*, **137**, 2305–2330, <https://doi.org/10.1175/2009MWR2834.1>.
- Wang, B., W. Hui, Z. Z. Ji, X. Zhang, R. C. Yu, Y. Q. Yu, and H. T. Liu, 2004: Design of a new dynamical core for global atmospheric models based on some efficient numerical methods. *Science in China Series A: Mathematics*, **47**, 4–21, <https://doi.org/10.1360/04za0001>.
- Wedi, N. P., and P. K. Smolarkiewicz, 2009: A framework for testing global non-hydrostatic models. *Quart. J. Roy. Meteor. Soc.*, **135**, 469–484, <https://doi.org/10.1002/qj.377>.
- Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quart. J. Roy. Meteor. Soc.*, **140**, 1505–1520, <https://doi.org/10.1002/qj.2235>.
- Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (ICO-sahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, <https://doi.org/10.1002/qj.2378>.
- Zerroukat, M., N. Wood, and A. Staniforth, 2002: SLICE: A Semi-Lagrangian Inherently Conserving and Efficient scheme for transport problems. *Quart. J. Roy. Meteor. Soc.*, **128**, 2801–2820, <https://doi.org/10.1256/qj.02.69>.
- Zhang, F. B., B. Wang, and L. J. Li, 2017: New approach to incorporating the impacts of non-hydrostatic perturbations in atmospheric models. *Atmospheric and Oceanic Science Letters*, **10**, 379–384, <https://doi.org/10.1080/16742834.2017.1348191>.