Physics-Incorporated Framework for Emulating Atmospheric Radiative Transfer and the Related Network Study

Yao Yichen[1], Zhong Xiaohui[2], Zheng Yongjun[3], and Wang Zhibin[1]

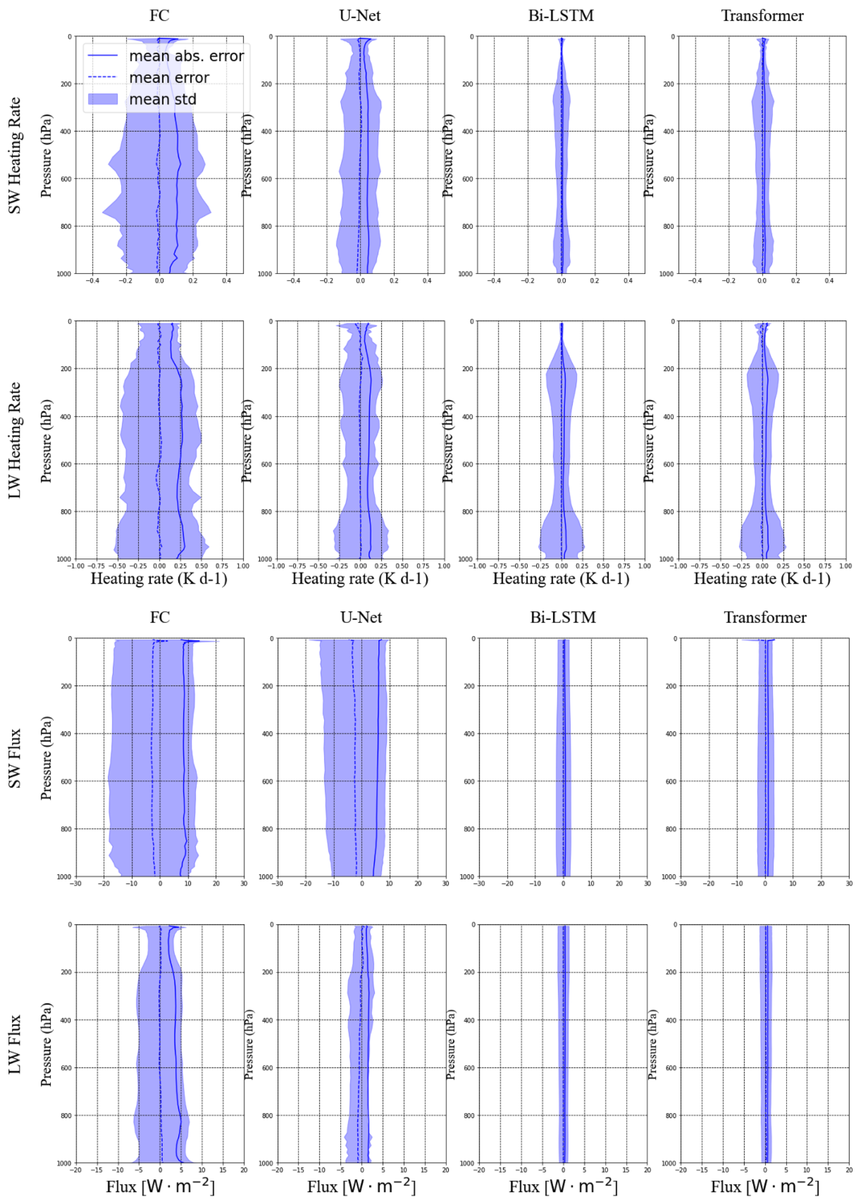
[1]Alibaba
[2]Damo Academy, Alibaba Group
[3]Nanjing University of Information Science and Technology

November 16, 2022

Abstract

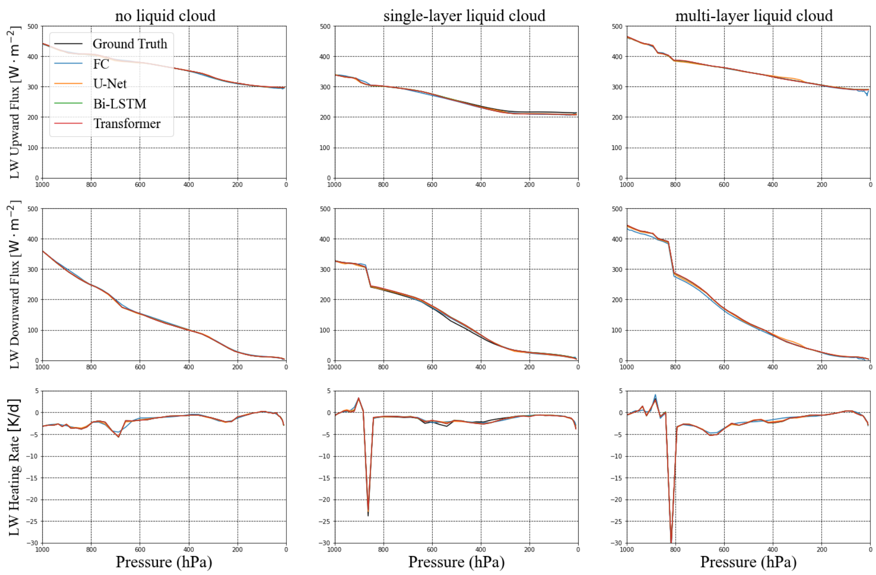
The calculations of atmospheric radiative transfer are among the most time-consuming components of the numerical weather prediction (NWP) models. Therefore, using deep learning to achieve fast radiative transfer has become a popular research direction. We propose a physics-incorporated framework for the radiative transfer model training, in which the thermal relationship between fluxes and heating rates is encoded as a layer of the network so that the energy conservation can be satisfied. Based on this framework, we compared various types of neural networks and found that the model structures with global receptive fields are more suitable for the radiative transfer problem, among which the Bi-LSTM model has the best performance.

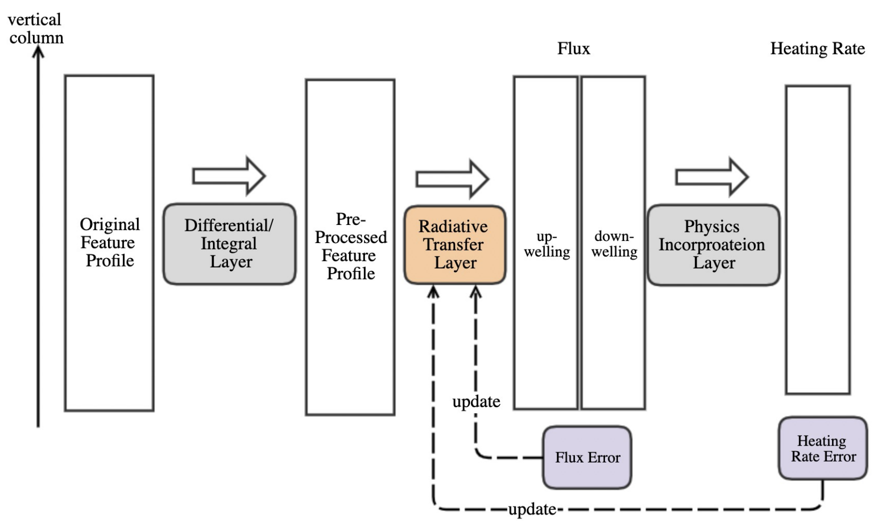
Posted on 16 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10512196.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

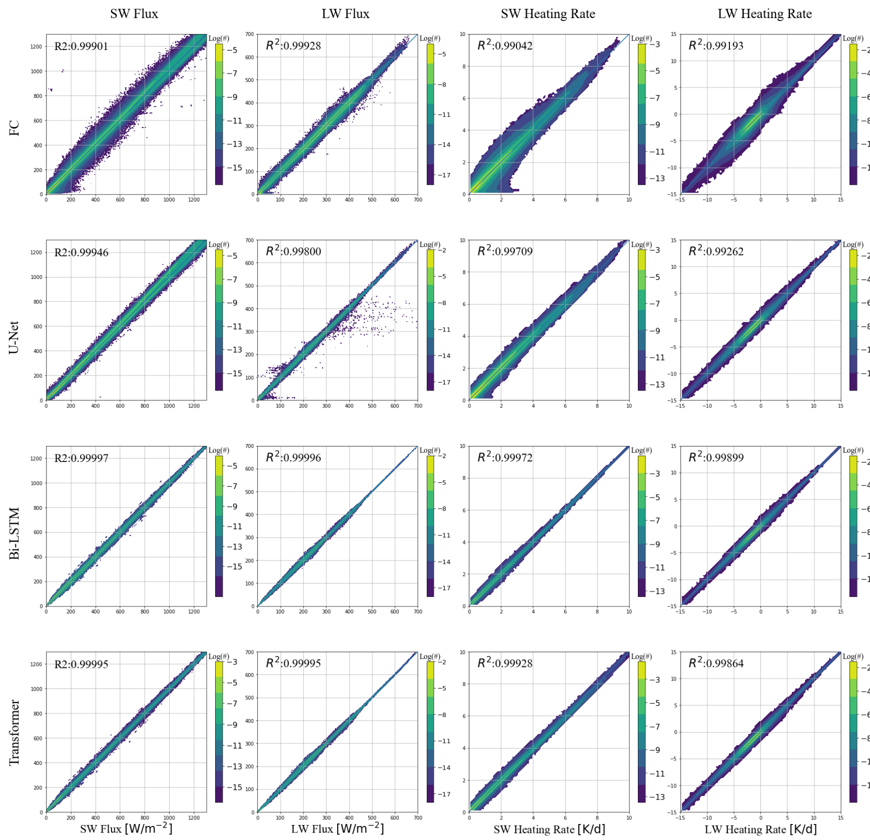


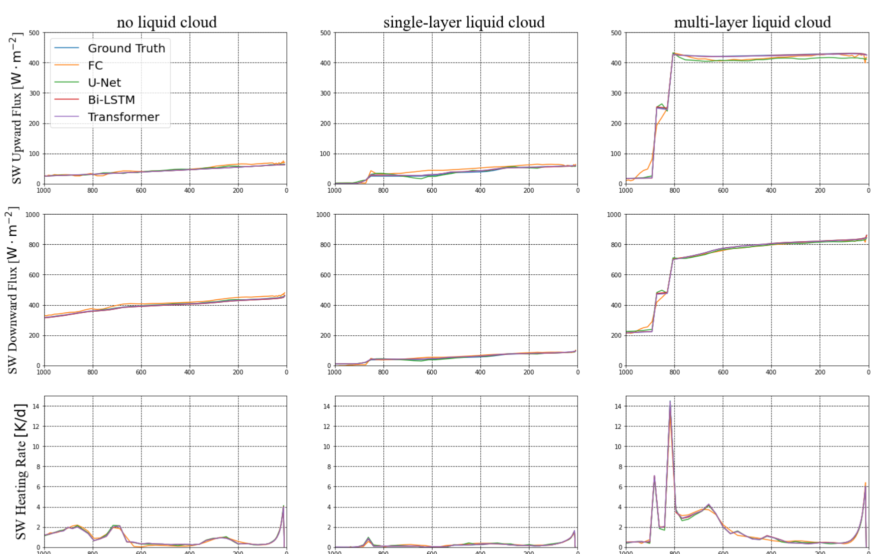
Posted on 16 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10512196.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 16 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10512196.1 — This a preprint and has not been peer reviewed. Data may be preliminary.









Posted on 16 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10512196.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Physics-Incorporated Framework for Emulating Atmospheric Radiative Transfer and the Related Network Study

Yichen Yao[1⋆], Xiaohui Zhong[1⋆], Yongjun Zheng[2], and Zhibin Wang[1]

[1]Damo Academy, Alibaba Group, Hangzhou 311121, China
[2]Nanjing University of Information Science and Technology, Nanjing 210044, China

Key Points:

- A physics-incorporated framework is proposed for the radiative transfer model training.
- The model structures with global receptive fields are more suitable for the radiative transfer problem.

* The two authors contributed equally to this paper.

Corresponding author: Yongjun Zheng, `zhengyongjun@gmail.com`

Abstract

The calculations of atmospheric radiative transfer are among the most time-consuming components of the numerical weather prediction (NWP) models. Therefore, using deep learning to achieve fast radiative transfer has become a popular research direction. We propose a physics-incorporated framework for the radiative transfer model training, in which the thermal relationship between fluxes and heating rates is encoded as a layer of the network so that the energy conservation can be satisfied. Based on this framework, we compared various types of neural networks and found that the model structures with global receptive fields are more suitable for the radiative transfer problem, among which the Bi-LSTM model has the best performance.

Plain Language Summary

Numerical weather prediction models require a lot of computational resources and time to run. Calculating the atmospheric radiative transfer processes is one of the most computationally expensive parts of the model. One alternative is to model the radiative transfer using deep learning models, but the deep learning models do not involve physical equations and may have physically inconsistent outputs. This paper proposes a model training framework to ensure the thermal equilibrium between fluxes and heating rates, which are outputs of radiative transfer models. Also, various neural network structures have been tested. The results demonstrate that model structures with global receptive fields work best for emulating radiative transfer calculations.

keywords

radiative transfer parameterization, neural networks, physics-incorporated

1 Introduction

Solar (shortwave, SW) and thermal radiation (longwave, LW) are the fundamental drivers of the atmospheric and oceanic circulation by creating the equator-versus-pole energy imbalance. The atmospheric radiative transfer processes are well understood and accurately represented by the line-by-line model LBLRTM (S. Clough et al., 2005; S. A. Clough et al., 1992). The LBLRTM requires unaffordable computational costs; thus, it is inappropriate for weather and climate modeling. Therefore, various parameterization methods are proposed to approximate radiative transfer calculations more efficiently for application in numerical models (Stephens, 1984).

Despite being simplified, the radiative transfer parameterization is still more computationally expensive than other dynamical or physical processes. Therefore, the radiative transfer parameterization is usually performed less frequently in time and on a coarser spatial grid. For example, in the European Centre for Medium-Range Weather Forecasts (ECMWF), the radiation scheme is run 8 times less frequently in time and 10.24 times coarser in spatial resolution than the high resolution deterministic forecast (HRES), which would degrade the precision compared to frequent calls in time and space (Hogan & Bozzo, 2018). While for the ECMWF ensemble forecast with 12 minutes time step, the radiation scheme is only called every 3 hours on a spatial grid 6.25 times coarser than the rest of the model.

To further speed up the radiation calculations in weather and climate models and make it feasible for more frequent calls of the radiation schemes, many researchers have investigated alternative approaches such as neural networks (NNs). Chevallier et al. (1998) and Chevallier et al. (2000) used shallow NNs with one hidden layer (called NeuroFlux) to simulate the LW radiative budget from the top of the atmosphere to the surface in a model with 31 vertical levels. The NeuroFlux achieved comparable accuracy to the accuracy of the ECMWF operational scheme and was also 22 times faster. However, NeuroFlux fails to maintain both accuracy and acceleration when applied to models with 60 vertical layers and above (Morcrette et al., 2008). Pal et al. (2019) developed two dense, fully connected, feed-forward deep NN (DNN) to emulate SW and LW radiative calculations. They replaced the original radiation parameterization in the Super-Parameterized Energy Exascale Earth System Model (SP-E3SM) with these DNN-based emulators and were able to run numerical simulations stably for up to a year. The DNN-based models achieved approximately 90-95% accuracy and were 8-10 times faster compared to the original parameterizations. Their results demonstrated the applicability of machine learning in modeling radiative transfer calculations in NWP models. Roh and Song (2020) found that the NN radiation model with high frequency call can perform better than the low frequency calls of the original radiation scheme with similar calculation costs. Moreover, Belochitski and Krasnopolsky (2021) showed that the shallow NN-based emulators of radiative transfer parameterization developed ten years ago for the general circulation model (GCM) are robust despite the structural change in the host model. Regarding model generalization, this model can generate realistic and stable radiation results when applied to numerical simulations for up to 7 months. Liu et al. (2020) compared feedforward NNs with the convolutional NNs for radiative transfer computations. Their results showed that the feed-forward NNs demonstrated a better trade-off between accuracy and computational performance.

However, the above methods and results were established using either incomprehensive datasets or non-common radiation schemes. Cachay et al. (2021) introduced ClimART, a dataset for applications of ML in radiative transfer problems. The ClimART dataset only took into account the pristine sky (no aerosols and no clouds) and clear sky conditions; thus, the NN models trained on the ClimART dataset would not be suitable for operational applications when the presence of clouds is inevitable. Dueben et al. (2021) established and published the MAELSTROM (MAchinE Learning for Scalable MeTeoROlogy and Climate) dataset, in which the dataset of A3 is generated using the input and output data from the ecRad Tripleclouds radiation scheme (Hogan & Bozzo, 2018). However, the ecRad radiation scheme is not widely used by other NWP models. For the NN-based radiative transfer schemes, if the training dataset contains more comprehensive weather conditions, it can have more practical value in the operational NWP simulations. Therefore, this paper build a dataset using the Model for Prediction Across Scales - Atmosphere (MPAS-A) that covers the entire globe and all months. The rapid radiative transfer model for general circulation models (RRTMG) is selected for radiative transfer calculations as the RRTMG model is widely used by many global and regional models.

With regards to the satisfaction of physical constraints, the previous studies (Krasnopolsky et al., 2010; Lagerquist et al., 2021; Liu et al., 2020; Roh & Song, 2020) trained NN-based emulators to output profiles of heating rates and fluxes at the surface and top-of-atmosphere directly, which causes the issues with energy conservation. Cachay et al. (2021) and Ukkonen (2022) chose to predict the radiative fluxes and compute heating rates from fluxes, which ensures physical consistency (Yuval et al., 2021). However, Ukkonen (2022) found that the heating rates are highly sensitive to the continuity in the fluxes profile, and small errors of fluxes lead to relatively large errors in heating rates. Based on the above research, the satisfaction of physical constraints has become a very critical issue in NN-based radiative transfer emulation. In this article, we will discuss this issue in detail from the aspect of framework design, and examine how to obtain accurate radiation emulation while satisfying the physical constraints.

In this paper, we use deep learning models to emulate radiative transfer calculations. We also propose a physically incorporated training scheme, where the energy conservation is encoded in the network in the form of constraints. Based on this framework, we apply and compare different network structures and analyze the advantages and disadvantages of each network structure in detail. Section 2 describes the dataset used for training and evaluation. The overall physics-incorporated solution, and various network structures are described in Section 3. The results related to each type of NNs and detailed error analysis are demonstrated in Section 4. Section 5 contains the conclusions and discussions.

2 Data

2.1 Data generation

The dataset was generated by running the Model for Prediction Across Scales - Atmosphere (MPAS-A) version 7.1 with initial conditions provided by the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS). MPAS employs an unstructured centroidal Voronoi mesh, which allows for variable horizontal resolution with higher resolution in a region of interest. In this study, we used the variable resolution ranging from 92 km to 25 km mesh containing 163842 horizontal grid cells and 57 vertical levels with a model top at 30 km.

The experiments used physics packages consisting of the “mesoscale reference” suite in MPAS-A. These packages include the new Tiedtke for cumulus convection (Zhang & Wang, 2017), RRTMG for SW and LW radiation (Iacono et al., 2008), Xu-Randall for subgrid cloud fraction (Xu & Randall, 1996), WRF Single-Moment 6-Class (WSM6) for microphysics (Hong & Lim, 2006), and Yonsei University (YSU) for planetary boundary layer mixing (Hong et al., 2006). The simulation was run for a total of 36 days in which a three consecutive days’ period was randomly selected from each of 12 months in the year 2021. The first two days of each three consecutive days are used for training, and the last day is used for testing. The model generates radiation inputs and outputs every 1 hour.

2.2 Input and output data

Table S1 lists all the input and output variables, where the input contains 29 original variables and the output contains 6 variables. Among the input variables, 11 variables are surface variables, and others are three-dimensional variables (either full layer or full level). To preprocess the data for the DL models, we pad the surface and layers variables to match the dimensions of the levels variables. The z-score normalization technique is applied to normalize all the input and output variables to ensure they have the same mean and variance. For three-dimensional variables, the mean and standard deviation (std) was determined from values of either all the vertical levels or layers.

3 Method

This section introduces the physics-incorporated model architecture and different network structures. The evaluation methods are described in the Text S1 in the supporting information.

3.1 Physics-Incorporated Framework

In the physics-based radiative transfer scheme, mapping between input and output variables is constructed column by column. The output comprises two parts: fluxes and heating rates. The flux is a measure of the energy being radiated per unit area, which has the unit of watts per meter square (W/m^2). The heating rate describes the temperature change per unit of time, and it has the units of Kelvin per day (K/d). These two types of variables are not independent of each other, and there is such a physical relationship:

$$HR_l = \frac{g}{c_p} \frac{(F_{l+1}^{up} - F_{l+1}^{down}) - (F_l^{up} - F_l^{down})}{{p_{l+1}}^{lev} - p_l^{lev}} \tag{1}$$

where g is the gravitational constant, c_p is the specific heat at constant pressure, F_l^{up}, F_l^{down}, and p_l^{lev} are the upward flux, downward flux, and pressure of level $l \in 1, \ldots, nlev$. As the full-level heating rates and the fluxes at the bottom and the top level will be used in the subsequent calculations of the NWP models, it is necessary to satisfy the conservation relationship described by Equation (1). Therefore, in designing NN structures, we focus on the satisfaction of this layer of physical relationship. Secondly, the change in atmospheric components of one layer/level has both local and global impacts on radiation along the entire vertical column. For example, the presence of clouds or liquid water at any layer affects the distribution of fluxes across all the vertical levels by producing local heating rates peaks.

Based on the above considerations, the structure is designed as shown in Figure 1 which includes three layers: the differential/integration layer, the radiative transfer layer, and the physics-incorporated layer.

The differential/integral layer is used as a data pre-processing module to preprocess input variables so that some prior knowledge can be fully utilized. As the cloud fraction (cldfrac in Table S1) and liquid water (qc) can affect fluxes far away from where they are present, these variables are integrated upward and downward along the vertical direction. The vertically accumulated cloud fraction and liquid water allow the models to learn vertically nonlocal effects. Meanwhile, calculating the heating rates requires the pressure difference between the two adjacent layers. Given the same values of fluxes, the smaller values of pressure difference result in larger values in heating rates. Therefore, the air pressure difference is obtained in advance by the differential module. The pre-processed features produced by the differential/integral layer are concatenated with the original features before being input to the models.

The radiative transfer layer is the most crucial part of the framework, and its output is fluxes only. The learnable parameters are only in this layer, as shown in the orange block in Figure 1. Through this layer, the mapping similar to that of the physics-based radiative transfer model is learned by NNs. A custom error function is designed as a weighted combination of the flux $\mathcal{L}_{flux}$ and heating rate $\mathcal{L}_{hr}$ as shown in Equation (2), in which λ is the weight of heating rate error. The flux error is defined as an average of the four groups of dimensionless values calculated as the mean square deviations divided by variance, as shown in Equation (3). Similarly, the heating rate error is an average of two groups of dimensionless values, as shown in Equation (4). In the forward propagation, the fluxes are first output by the selected networks, and then heating rates are derived by the physics-incorporated layer (third layer). The flux and heating rate error are combined, and then the network parameters of the radiative transfer layer will be updated accordingly. Many network structures can be implemented in this layer, and the details are described in the following subsection.

The last layer is the physics-incorporated layer, which constructs the relationship between fluxes and heating rates as shown in Equation (1). In order to make this relationship more strictly satisfied, the entire equation is treated as an independent layer and is encoded into the framework, avoiding the non-conservation of thermal equilibrium. Therefore, the gradient of heating rate error can be represented using the gradient of flux error and Equation (1), there are no learnable parameters within this layer.

$$\mathcal{L} = \mathcal{L}_{flux} + \lambda \mathcal{L}_{hr} \tag{2}$$

$$\mathcal{L}_{flux} = \frac{1}{4}\left[\frac{MSE_{F_{sw-up}}}{\sigma^2_{F_{sw-up}}} + \frac{MSE_{F_{sw-dn}}}{\sigma^2_{F_{sw-dn}}} + \frac{MSE_{F_{lw-up}}}{\sigma^2_{F_{lw-up}}} + \frac{MSE_{F_{lw-dn}}}{\sigma^2_{F_{lw-dn}}}\right] \tag{3}$$

$$\mathcal{L}_{hr} = \frac{1}{2}\left[\frac{MSE_{HR_{sw}}}{\sigma^2_{HR_{sw}}} + \frac{MSE_{HR_{lw}}}{\sigma^2_{HR_{lw}}}\right] \tag{4}$$

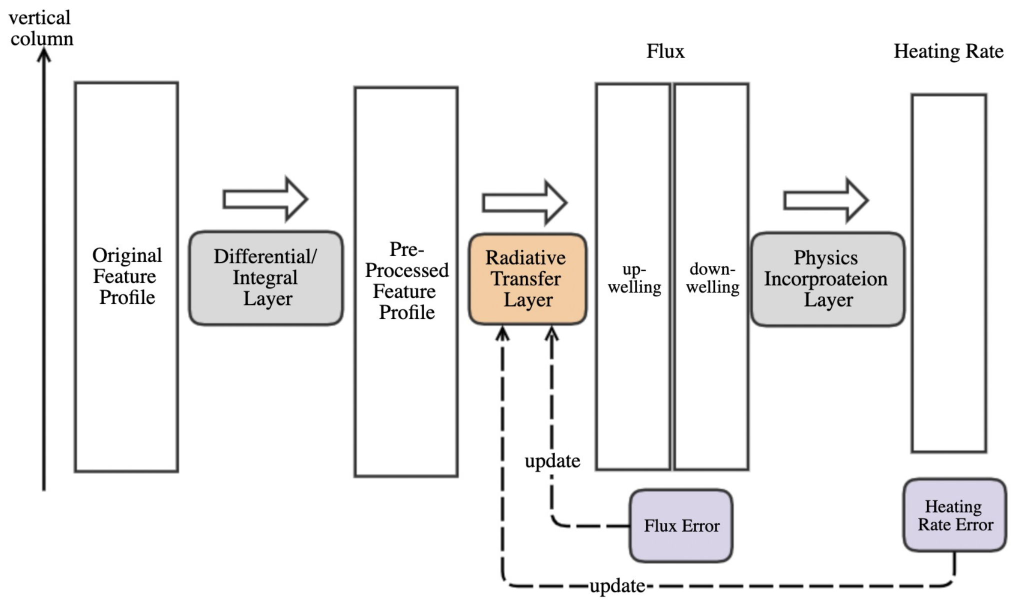


Figure 1. Physics-incorporated framework for emulating atmospheric radiative transfer

3.2 Network Structures

In this section, the detailed network structures in the radiative transfer layer are described. The layer realizes the mapping from input features ($W \times H$) to the fluxes outputs ($4 \times H$), in which W and H represent the number of features and vertical levels, respectively, and the four output variables are SW upward flux (SW_{up}), SW downward flux (SW_{dn}), LW upward flux (LW_{up}) and LW downward flux (LW_{dn}), respectively. In this paper, various network structures are tested, including fully connected networks, convolutional-based NNs, recurrent-based networks, Transformer-based NNs, and neural operator based NNs, respectively. For each group of network structures, we control the total number of parameters to be around 1 million. In this way, the influence of the number of the parameters can be ruled out, and the influence of the network structures on the radiative transfer modeling can be examined more clearly. As the fully connected networks and convolutional-based NN are studied by many researchers before (Krasnopolsky et al., 2010; Liu et al., 2020; Cachay et al., 2021; Lagerquist et al., 2021; Ukkonen, 2022), the details are described in Text S3 in the supporting information.

- Recurrent Type: Recurrent NNs (RNN) are widely used in natural language processing (NLP) tasks and are good at dealing with sequential problems. Here, the vertical direction is treated as the state transition direction, and the variable at a specific level is analogous to the word vector in the NLP tasks, which is represented by the feature vector at that level. In information transmission, a single-layer RNN can transmit information along the full vertical column, which is very similar to the propagation of radiative waves in the vertical direction. Also, a multi-layer RNN layer is used to mimic reflection in the radiative transfer processes. The long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) and gated recurrent units (GRUs) (Cho et al., 2014) are explicitly designed to avoid long-term dependency problems. They used gated units to retain useful information and remove irrelevant information. The LSTM selected in this paper is a 5-layer structure, each layer has 96 hidden layer units, and the number of network parameters is 1.12 million. For GRU, a 5-layer structure is used, with each layer having 128 hidden layer units, and the number of network parameters is 0.77 million. In addition, as the radiative transfer in the atmosphere involves both upward and downward processes, we implement the bidirectional LSTM and GRU to extract information from both directions.

- Transformer Type: Transformer (Vaswani et al., 2017) network has recently become a hot topic in the field of machine learning. It has global perception capabilities due to the attention mechanism. For the NN-based mapping of radiative transfer calculations, global dependencies exist between the input features and outputs. For example, when clouds occur, the fluxes at all levels are changed accordingly. Here, the self-attention mechanism is used so that the feature information is retrieved at all vertical levels, and the relevant information can be extracted and summarized. More specifically, the network initially superimposes the original feature and the position embedding of the vertical index. Then, the combined features are fed into seven layers of self-attention blocks. Each block contains one self-attention layer and two fully connected layers. The self-attention layer first maps the features into query, key, and value vectors and performs the dot product of vectors. All the query, key, and value vectors have a dimension of 128. At the end of the network, the embedding dimension is changed back to the output dimension through a 1×1 convolutional layer. The total number of trainable parameters in this Transformer network is 0.71 million.

- Neural Operator Type: The traditional radiative transfer parameterization approximates the full equations of radiative transfer by discretizing the atmosphere in the vertical direction. However, the discretization brings about a trade-off between speed and accuracy: low resolution is fast but less accurate, while high resolution is accurate but slower. Unlike traditional grid-dependent methods, the Fourier Neural Operators (FNO) can parameterize the radiative transfer modeling in function space instead of the discretized space. The output of FNO is the complete wave field solution, similar to the wavelike pattern of fluxes. The FNO (Li et al., 2020) we implement in this study includes four sequential modules, each composed of a frequency domain and a spatial domain. In the frequency domain, input features go through the Fourier transformation, low-pass truncation, and full connection operation. Lastly, the output is converted to the time-domain space through the inverse Fourier transform. The spatial domain is a simple fully connected network. This scheme allows a single layer operator to achieve a global perspective of the entire vertical column. The truncated wave number is set to 16, and the channel width in the module is 96. The channel width is mapped to the output dimension at the final output layer through a 1×1 convolution. The total number of trainable parameters in this Transformer network is 1.22 million.

All settings of the hyperparameters used for different NNs are the same. Each model is trained with 500 epochs using a batch size of 4096. Adam optimizer is used with the initial learning rate 1e-3. The plateau scheduler is applied to decrease the learning rate by a factor of 0.5 when the loss does not decrease for five consecutive epochs.

4 Results

4.1 Statistical results

Table 1 summarizes the error statistics of different NN-based emulators for fluxes and heating rates. The root mean square error (RMSE) of SW fluxes and heating rates predicted by the FC, ResNet, and U-Net models are higher than 10 W/m^2 and 0.1 K/day, respectively, across all the vertical layers and time. The RMSE of LW fluxes is greater than 2 W/m^2 and smaller than that of SW fluxes, which is due to the greater magnitude of SW fluxes than that of LW fluxes. The RMSE of LW heating rates is greater than 0.2 K/day and is also higher than the SW heating rates of each corresponding NN emulator, as LW heating rates are more sensitive to clouds and more difficult to predict (see Figure 2). FC and CNN

Table 1. Evaluation metrics (RMSE and MBE) of SW flux, LW flux, TOA net flux, SW heating rate and LW heating rate for NN emulators including FC, ResNet, U-Net, Bi-GRU, Bi-LSTM, Transformer and FNO for test data.

Model	SW Flux $W \cdot m^{-2}$		LW Flux $W \cdot m^{-2}$		TOA Net Flux $W \cdot m^{-2}$	SW Heating Rate $K \cdot d^{-1}$		LW Heating Rate $K \cdot d^{-1}$	
	RMSE	MBE	RMSE	MBE	MBE	RMSE	MBE	RMSE	MBE
FC	14.63	-2.31	5.28	0.182	-3.78	18.85e-2	-6.79e-3	3.94e-1	-1.19e-3
ResNet	38.97	-1.17	8.72	-0.38	-2.32e-1	22.89e-2	5.38e-3	4.14e-1	2.51e-3
Unet	10.92	-2.56	2.46	-0.314	-7.62	9.58e-2	-6.02e-3	2.17e-1	-7.06-3
Bi-GRU	2.334	7.31e-3	1.216	-8.20e-3	3.97e-1	3.29e-2	-4.87e-4	1.41e-1	-1.90e-3
Bi-LSTM	**2.315**	**-2.15e-3**	**1.205**	**-1.66e-3**	**4.91e-2**	**3.20e-2**	**7.02e-5**	**1.39e-1**	1.48e-4
Transformer	2.753	0.138	1.286	0.211	-5.61	4.06e-2	2.34e-3	1.46e-1	**6.85e-5**
FNO	3.755	-0.125	1.289	-0.0238	-6.77	4.20e-2	-1.90e-3	1.47e-1	5.92e-4

networks do not perform well in radiative transfer calculations, which can be explained by the structural properties of the two networks. For FC networks, the flattening operation erases the vertical distribution of all the features, leading to the loss of important information. CNN networks only have the local receptive fields in the vertical direction for each operation performed. Therefore, the overall performance of FC and CNN networks is not as good as RNN, Transformer, and FNO networks.

The Bi-GRU, Bi-LSTM, Transformer, and FNO achieve significant improvement with RMSE of SW and LW fluxes smaller than 2.5 and 1.3 W/m^2, respectively. In addition, the RMSE of SW and LW heating rates is reduced to less than 0.033 and 0.14 K/day, respectively. The advantage of these networks is that a global perspective of an entire atmospheric column can be obtained in single-layer operations. More specifically, the RNN networks allow the state to be transferred in the vertical direction through the recurrent mechanism. For the Transformer, it can query information at any level through the attention mechanism. The FNO networks encode the information into the Fourier function space, and each modal presents a wave function along the vertical direction. In summary, these networks enable complete information transfer in the vertical direction and show a considerable improvement in error statistics of the fluxes and heating rates. Overall, the RNN-type networks demonstrate the best performance, significantly outperforming the other structures in terms of both fluxes and heating rates. Among them, the Bi-LSTM model has the best performance. The RMSE of SW and LW fluxes are 2.315 and 1.205 respectively, and the RMSE of SW and LW heating rates are 3.20×10^{-2} and 1.39×10^{-1} respectively. Regarding mean bias error

(MBE) of fluxes and heating rates, Bi-GRU and Bi-LSTM also have the smallest values. In addition, the biases of the net fluxes at the top-of-atmosphere (TOA) directly determine the energy budget of the entire atmosphere. Therefore, if the MBE of net fluxes at the TOA tends to be 0, it represents a more consistent energy budget with the physics-based radiation schemes. It can be seen from Table 1 that the Bi-LSTM model has the highest accuracy in terms of net fluxes at TOA, with a value of 4.91×10^{-2}, which is at least one order of magnitude smaller than other schemes.

For a clearer analysis of the vertical distribution of errors, Figure 2 presents the vertical profiles of statistics for fluxes and heating rates. The FC and U-Net models generally have relatively higher variance, as shown by the vertical profiles of mean std of biases. The distribution of the error of the FC network is relatively uniform at different levels, while the U-Net shows some sawtooth distribution on the LW profile, and the error changes sharply with the vertical distribution. The Bi-LSTM and Transformer models are superior to the FC and U-Net models at all levels, which can be seen from the vertical profiles of mean absolute error (MAE). Overall, the error distributions of the Bi-LSTM and the Transformer are similar, with Bi-LSTM slightly better. The two models show a relatively uniform vertical distribution of error in fluxes. For heating rates, both models have relatively higher std of biases in the pressure layers between 800-1000 hPa and 200-400 hPa. Those two vertical regions are where liquid and ice clouds occur most frequently. Figure S1 illustrates the comparisons on scatter plots, and the conclusions are consistent with the vertical profiles shown above.

4.2 Benefits of introducing the physics-incorporated layer

In this subsection, we discuss the benefits of introducing the physics-incorporated layer. The physics-incorporated layer ensures the satisfaction of the thermal equilibrium between fluxes and heating rates as shown in Equation (1) by encoding it as part of network layers. We designed three groups of experiments: only supervising fluxes, only supervising heating rates, and a joint loss with the physics-incorporated layer imposed. For the case of joint loss, the weights of the heating rate and the flux are fixed 0.1 and 1, respectively. The RMSE of these experiments are summarized in Table S2 in the supporting information.

When only supervising the fluxes, we calculate the heating rates using Equation (1). As the vertical profiles of fluxes are often smooth and flat, the model is relatively easy to fit well. As a result, the RMSE of fluxes is only slightly worse than that using the physics-incorporated layer. However, the RMSE of SW and LW heating rates are 6 times and 1.5 times greater than using the physics-incorporated layer. When models are trained only to supervise the heating rates, fluxes cannot be derived accordingly. In this case, the heating rates are still less accurate than that with the physics-incorporated layer, and the RMSE of SW and LW heating rates are 1.5 and 1.25 times larger. In summary, the physics-incorporated layer demonstrates great superiority. Firstly, a physically consistent relationship between fluxes and heating rates can be ensured. Secondly, the heating rates and fluxes are also more accurate.

5 Conclusions

In this paper, we propose a physics-incorporated framework for emulating atmospheric radiative transfer processes. The physical relationship between fluxes and heating rates is considered in our framework, and it is encoded as a layer of the network. Based on this framework, we designed and compared various types of NN structures and found that the networks with a full receptive field in a single layer are more suitable for the radiative transfer problem, among which the Bi-LSTM model has the best accuracies for fluxes and heating rates. Furthermore, vertical profiles of heating rates and fluxes suggest the Bi-LSTM performs well at all vertical levels, although there are slightly larger errors and variances where clouds are present.

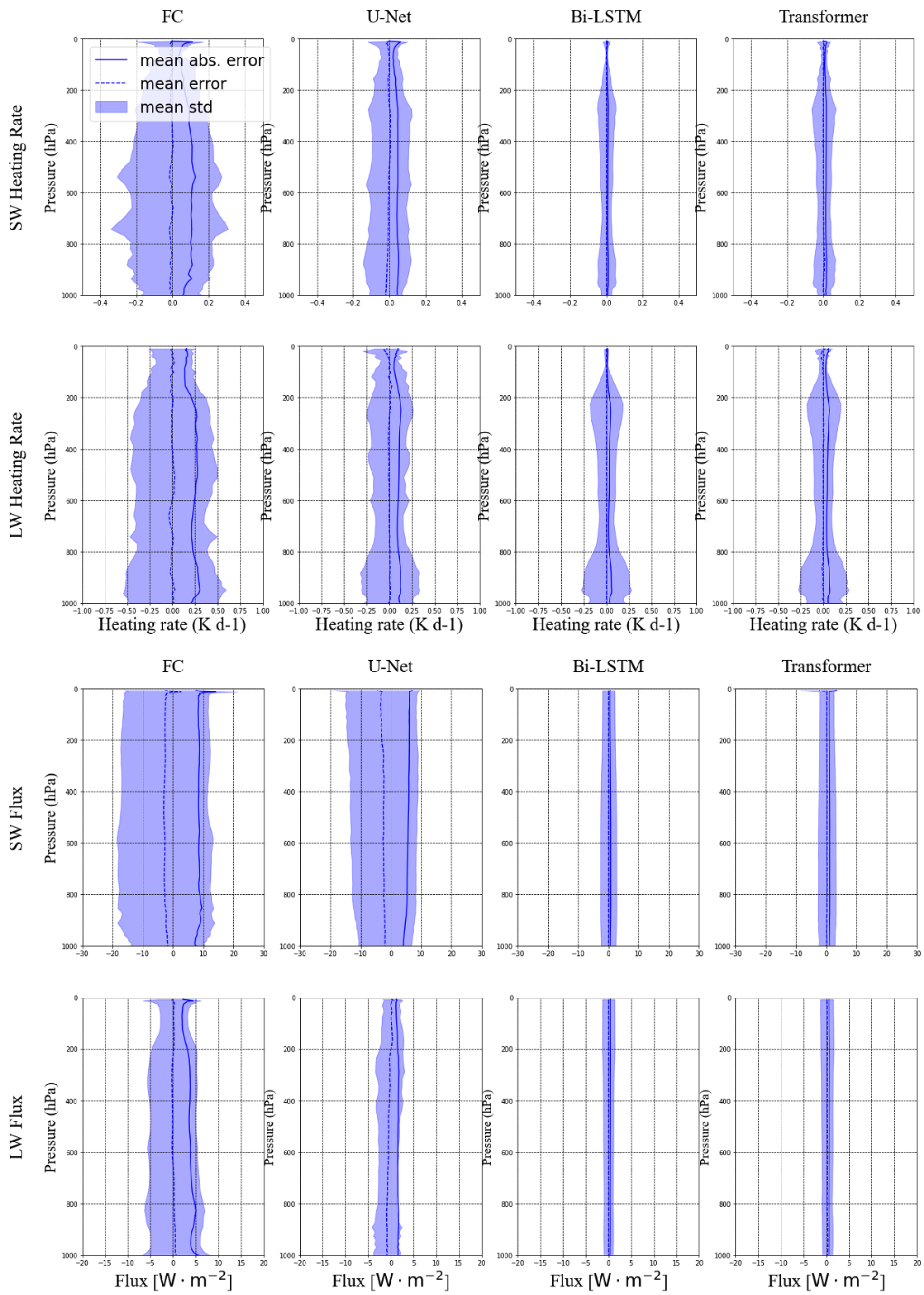


Figure 2. Vertical profiles of the statistics in SW heating rates (first row), LW heating rates(second row), SW fluxes(third row), and LW fluxes (fourth row) for the test data using different NN-based emulators: FC (first column), U-Net (second column), Bi-LSTM (third column), and Transformer (fourth column). The solid and dotted lines show the MAE and MBE profile, respectively, and the shaded area indicates the mean std relative to the bias.

Future work will investigate the online implementation of the DL-based emulators in an NWP model such as Weather Research and Forecasting (WRF) with different vertical levels. Besides, due to the nonlinearity of the radiative transfer models, there is no corresponding tangent-linear and adjoint model of radiative transfer models for WRF. Hatfield et al. (2021) demonstrated the feasibility of constructing the tangent-linear and adjoint models from the NN-based gravity wave drag parameterization scheme. They showed that the NN-derived tangent-linear and adjoint models successfully passed the standard test and were applied in four-dimensional variational data assimilation. Likewise, our future work includes developing the adjoint model of radiation schemes using NN-based radiation emulators to improve the four-dimensional variational data assimilation system.

Author contributions: Y.Y. trained the deep learning models and calculate the statistics of model performance. Y.Z. conducted the MPAS-A model simulations to provide dataset for training and evaluation, and offered valuable suggestions on the model training and paper revision. X.Z. and Y.Y wrote, reviewed and edited the original draft; Z.W. 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.

Acknowledgments

This work was supported in part by the Zhejiang Science and Technology Program under Grant 2021C01017.

Data Availability Statement

The source code and data used in this work are available at Github (https://github.com/yaoyichen/radiationNet).

References

Belochitski, A., & Krasnopolsky, V. (2021). Robustness of neural network emulations of radiative transfer parameterizations in a state-of-the-art general circulation model. *Geoscientific Model Development*, *14*(12), 7425–7437.

Cachay, S. R., Ramesh, V., Cole, J. N., Barker, H., & Rolnick, D. (2021). Climart: A benchmark dataset for emulating atmospheric radiative transfer in weather and climate models. *arXiv preprint arXiv:2111.14671*.

Chevallier, F., Chéruy, F., Scott, N., & Chédin, A. (1998). A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of applied meteorology*, *37*(11), 1385–1397.

Chevallier, F., Morcrette, J.-J., Chéruy, F., & Scott, N. (2000). Use of a neural-network-based long-wave radiative-transfer scheme in the ecmwf atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, *126*(563), 761–776.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., ... Brown, P. (2005). Atmospheric radiative transfer modeling: A summary of the aer codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *91*(2), 233–244.

Clough, S. A., Iacono, M. J., & Moncet, J.-L. (1992). Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor. *Journal of Geophysical Research: Atmospheres*, *97*(D14), 15761–15785.

Dueben, P., Chantry, M., Nipen, T., Denisenko, G., Ben-Nun, T., Gong, B., & Langguth. (2021). *D1.1 first version of datasets and cost functions to develop machine learning solutions for a1-a6 (version 1.0; machine learning for scalable meteorology and climate) retrieved from https://www.maelstrom-eurohpc.eu/content/docs/uploads/doc6.pdf.*

Hatfield, S., Chantry, M., Dueben, P., Lopez, P., Geer, A., & Palmer, T. (2021). Building tangent-linear and adjoint models for data assimilation with neural networks. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002521.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hogan, R. J., & Bozzo, A. (2018). A flexible and efficient radiation scheme for the ecmwf model. *Journal of Advances in Modeling Earth Systems*, *10*(8), 1990–2008.

Hong, S.-Y., & Lim, J.-O. J. (2006). The wrf single-moment 6-class microphysics scheme (wsm6). *Asia-Pacific Journal of Atmospheric Sciences*, *42*(2), 129–151.

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, *134*(9), 2318–2341.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13).

Krasnopolsky, V., Fox-Rabinovitz, M., Hou, Y., Lord, S., & Belochitski, A. (2010). Accurate and fast neural network emulations of model radiation for the ncep coupled climate forecast system: Climate simulations and seasonal predictions. *Monthly Weather Review*, *138*(5), 1822–1842.

Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., & Hagerty, V. (2021). Using deep learning to emulate and accelerate a radiative transfer model. *Journal of Atmospheric and Oceanic Technology*, *38*(10), 1673–1696.

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

Liu, Y., Caballero, R., & Monteiro, J. M. (2020). Radnet 1.0: Exploring deep learning architectures for longwave radiative transfer. *Geoscientific Model Development*, *13*(9), 4399–4412.

Morcrette, J.-J., Mozdzynski, G., & Leutbecher, M. (2008). A reduced radiation grid for the ecmwf integrated forecasting system. *Monthly weather review*, *136*(12), 4760–4772.

Pal, A., Mahajan, S., & Norman, M. R. (2019). Using deep neural networks as cost-effective surrogate models for super-parameterized e3sm radiative transfer. *Geophysical Research Letters*, *46*(11), 6069–6079.

Roh, S., & Song, H.-J. (2020). Evaluation of neural network emulations for radiation parameterization in cloud resolving model. *Geophysical Research Letters*, *47*(21), e2020GL089444.

Stephens, G. L. (1984). The parameterization of radiation for numerical weather prediction and climate models. *Monthly weather review*, *112*(4), 826–867.

Ukkonen, P. (2022). Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer. *Journal of Advances in Modeling Earth Systems*, *14*(4), e2021MS002875.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Xu, K.-M., & Randall, D. A. (1996). A semiempirical cloudiness parameterization for use in climate models. *Journal of the atmospheric sciences*, *53*(21), 3084–3102.

Yuval, J., O'Gorman, P. A., & Hill, C. N. (2021). Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters*, *48*(6), e2020GL091363.

Zhang, C., & Wang, Y. (2017). Projected future changes of tropical cyclone activity over the western north and south pacific in a 20-km-mesh regional climate model. *Journal of Climate*, *30*(15), 5923–5941.

Figure 1.

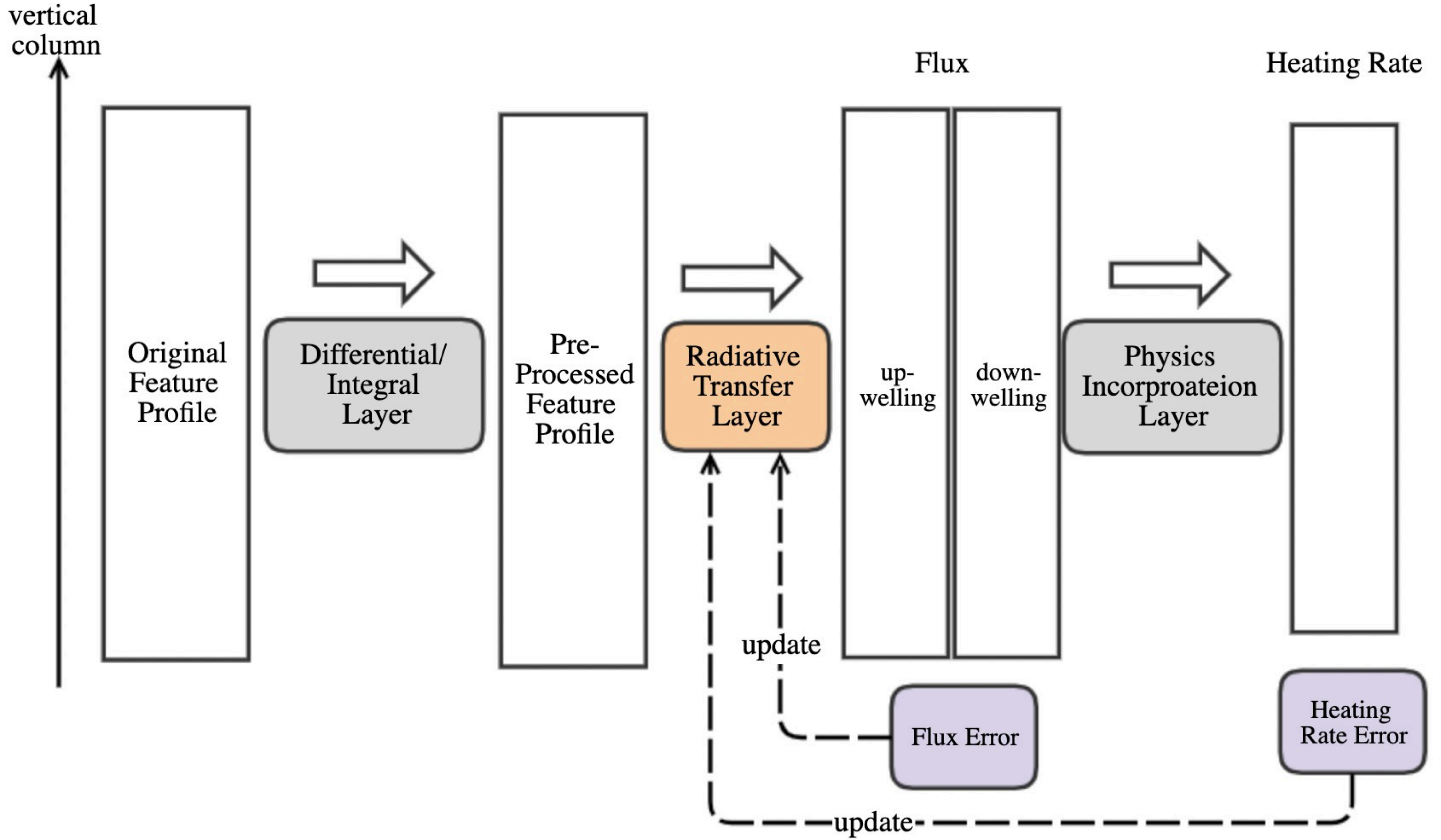
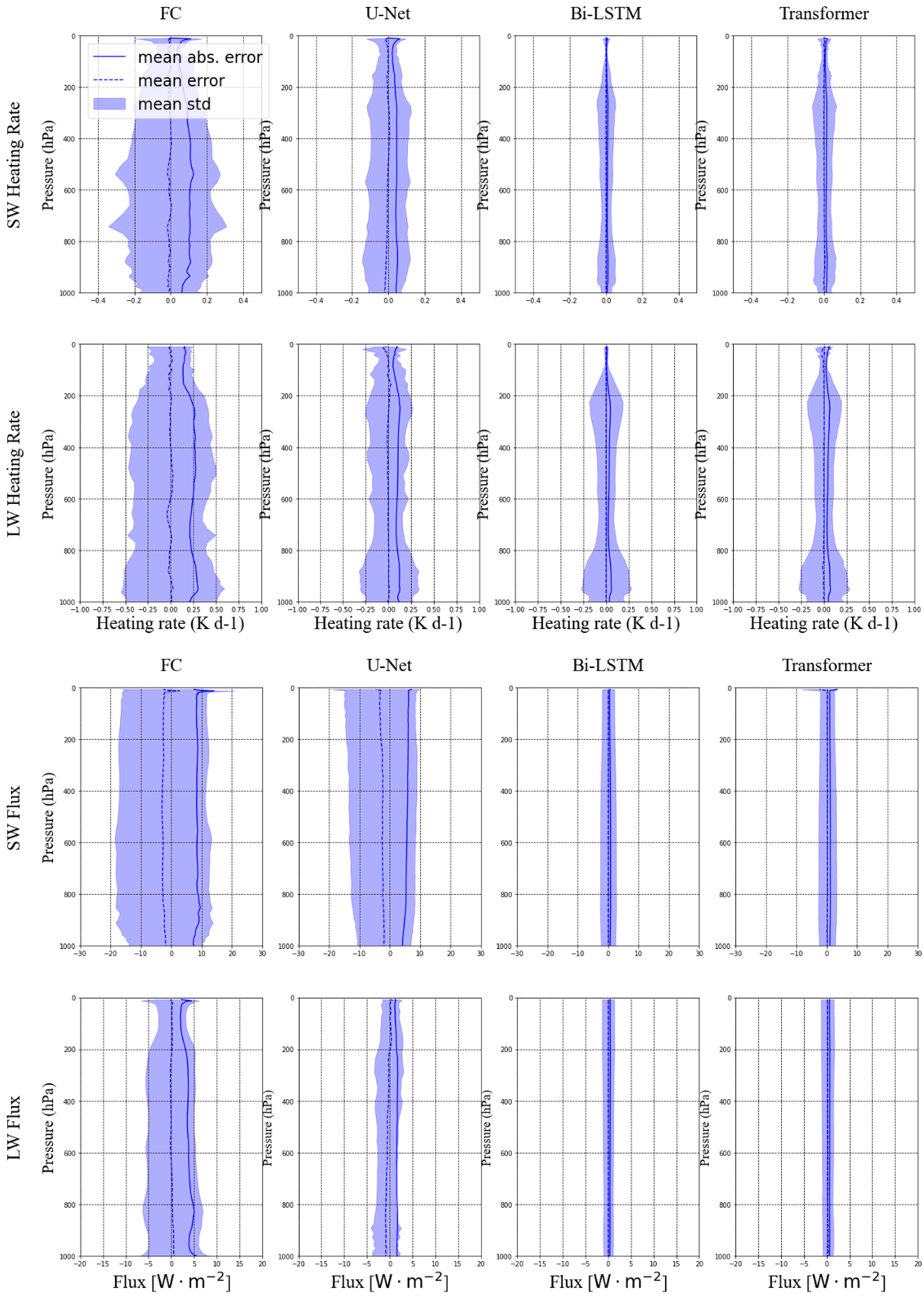

vertical
column
Flux
Heating Rate
Original
Feature
Profile
Differential/
Integral
Layer
Pre-
Processed
Feature
Profile
Radiative
Transfer
Layer
up-
welling
down-
welling
Physics
Incorproateion
Layer
update
Flux Error
Heating
Rate Error
update

Figure 2.



FC
U-Net
Bi-LSTM
Transformer
SW Heating Rate
LW Heating Rate
SW Flux
LW Flux
mean abs. error
mean error
mean std
Pressure (hPa)
Heating rate (K d-1)
Flux [W · m^-2]