https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

Transfer learning-based hybrid machine learning in single-column model of AFES v4

Yuya Baba[1]

[1]Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

Correspondence to: Yuya Baba (babay@jamstec.go.jp)

Abstract. The validity of a transfer learning-based hybrid machine learning (ML) in single-column model (SCM) of Atmospheric general circulation For the Earth Simulator (AFES) version 4 is examined. The results of the SCM with and without hybrid ML using transfer learning (i.e., the original and hybrid models) are compared against observational datasets and they are evaluated in tropical and midlatitude intensive observation periods. The hybrid model produces better results compared with the original model in all experiments, even when the period of training data is shifted from the target period. However, seasonality is more important for the midlatitude cases than the tropical cases, i.e., training data from the same month is necessary, even though the year of training data is different. The ML component of the hybrid model successfully corrects the model's bias, but the correction for temperature is greater than that for humidity, especially in the amplitude rather than in the phase. If the temporal and spatial variability is significant, the ML component fails to correct the biases. Analysis of the bias components reveals that the hybrid model can reduce the mean state bias, but it cannot reduce the high-frequency components of the biases. The hybrid model slightly improves precipitation depending on the cases but does not improve surface heat fluxes that cause biases in the low-level. This implies that further synchronisation is needed for surface heat fluxes. In conclusion, transfer learning-based hybrid ML can better simulate atmospheric variability by reducing mean state bias when the appropriate training data are used. Due to this advantage, the model has the potential to improve the prediction skill of numerical models over longer periods with limited training data.

1 Introduction

Atmospheric general circulation models (AGCMs) are used to study the mechanisms of weather and climate and have been employed for numerical weather prediction, seasonal prediction, and future projection. However, these numerical models, especially those related to unresolved physical processes such as cloud physics, are known to have large uncertainties (Long and Xie, 2016; Pathak et al., 2020; Schuh and Jacobson, 2023), and thus their faithfulness to nature as well as their ability to predict weather and climate remain imperfect. To reduce these uncertainties and improve performance, refinements to the numerical models have been developed, including new parameterisations and assimilation techniques (Baba, 2023; Baba et al., 2024, Baba and Ujiie, 2025). Nevertheless, there are still unresolved model biases, requiring deeper understandings of underlying mechanisms.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

Machine learning (ML) is becoming a useful tool for weather and climate prediction with low computational cost. The major ML methods are based on a data-driven approach which trains the ML model using historical or previous time series datasets and then attempts to predict the future states of weather or climate, without solving the underlying governing equations of the natural processes (Bouallegue et al., 2024). Although several studies have demonstrated the superiority of the methods compared with existing dynamical models, some limitations have been found, e.g., they are not effective for long-range prediction or extreme events. To accurately capture such phenomena, huge and comprehensive training datasets are required (De Burgh-Day and Leeuwenburg, 2023).

The limitations of the data-driven approach are partly originated from the fact that ML methods lack physical constraints, and this leads to a violation of physical laws in the predictions. To address this, recent studies have constrained ML methods by incorporating physical requirements, i.e., they adopted physics-informed ML (Karniadakis et al., 2021). However, deciding which constraints are appropriate for the ML-based model remains uncertain and dependent on the problem. Moreover, their applicability to climate and weather prediction, including extreme events, remains open question.

Therefore, to avoid relying on only data-driven models, some studies have developed hybrid ML methods that partly employ both governing equations and reservoir computing (Pathak et al. 2018; Wikner et al., 2020; Arcomano et al., 2022; 2023; Patel et al., 2025). Other hybrid models employed random forests, deep learning, and transformers (Watt-Meyers et al., 2021; Rasp et al., 2019; Kochkov et al., 2024). Although they employ different hybrid approaches, such methods commonly increase the accuracy of weather and climate simulations and reduce degradation in the performance of long-range prediction.

Therefore, hybrid ML is a powerful tool to enhance the performance of dynamical models for predicting weather and climate, avoiding the problems that stem from the data-driven approach. Despite this, there remain some questions about hybrid ML regarding practical application, for instance, the configurations of parameters used in ML, how many prognostic variables are synchronised in the ML, and the transfer of trained ML to different domains (called transfer learning, Weiss et al., 2016). Of these issues, the validity of transfer learning for a hybrid model is important for practical application. If transfer learning is practical for the hybrid model, then model can simulate various atmospheric conditions with less training data. Moreover, it enables longer-range prediction, since the model will be valid for a longer period with limited training data.

The purpose of this study is to examine the validity of the transfer learning-based hybrid ML model (hereafter referred to as the hybrid model). Hybrid ML is implemented in a single-column model (SCM) of Atmospheric general circulation For the Earth Simulator (AFES) version 4 to examine the above points in various configurations and situations. Here, the SCM is chosen because it enables us to analyse the parameterisation-derived error in detail (e.g., Wang and Zhang, 2013; Gettelman et al., 2019; Bogenschutz et al., 2020). Since the error of simulated results affects the performance of numerical weather and climate prediction, this study focuses on the reduction of bias by the ML component of the hybrid model. In the evaluation, only the time difference in the training data is considered in the transfer learning. The remainder of this manuscript is structured as follows. In Section 2, the hybrid ML used in the SCM is formulated. The model and

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

experimental setup are described in Section 3. Results and discussion are presented in Section 4. Summary and conclusions are provided in Section 5.

2 Hybrid ML

2.1 Formulation

To construct an SCM using hybrid ML, reservoir computing (RC hereafter, Lukoševičius and Jaeger, 2009) was chosen for the ML component of the hybrid model and implemented in the SCM. RC is a type of recurrent neural network (RNN) which has a relatively simpler structure compared with other RNNs. It consists of an input layer, a reservoir, and an output layer, the purpose of which is to predict targeted variables. In original RC, the reservoir state is updated using the following equation

$$\mathbf{r}(t+\Delta t) = \tanh[\mathbf{A}\mathbf{r}(t) + \mathbf{B}\mathbf{u}(t)], \tag{1}$$

where $\mathbf{r}(t)$ is the reservoir state vector with dimension D_r, $\mathbf{u}(t)$ is the reservoir input state vector formed from the original input vector $\mathbf{v}(t)$ (directly substituted in this study), and $\mathbf{A}$ and $\mathbf{B}$ are the input layer's weight matrices. Here, $\mathbf{A}$ and $\mathbf{B}$ are nonzero random matrices, which will be detailed later. The weight of the output layer is trained using the following equation

$$\mathbf{v}(t+\Delta t) = \mathbf{W}\mathbf{r}(t+\Delta t), \tag{2}$$

where $\mathbf{W}$ is the output layer's weight matrix. If the input and output state vectors are given by observed data with a time interval of Δt, then the output layer's weight is trained for the observation. In general, this training is performed for time-series data of $\mathbf{v}$ and $\mathbf{r}$ using a simple ridge regression. After the training, Eq. (2) gives an RC prediction model using $\mathbf{r}$ and Eq. (1).

In the hybrid ML model, Eq. (1) was also used, and Eq. (2) was modified to include the knowledge-based (i.e., physics-based) variables, e.g., prognostic variables developed using the governing equations. The modified equation is given by

$$\mathbf{v}^h(t+\Delta t) = \mathbf{W}\begin{pmatrix}\mathbf{v}^p(t+\Delta t)\\ \mathbf{r}(t+\Delta t)\end{pmatrix} = (\mathbf{W}_{mod} \quad \mathbf{W}_{res})\begin{pmatrix}\mathbf{v}^p(t+\Delta t)\\ \mathbf{r}(t+\Delta t)\end{pmatrix}, \tag{3}$$

where $\mathbf{v}^h$ is the hybrid output state vector, and $\mathbf{v}^p$ is the physics-based input state vector. $\mathbf{W}_{mod}$ and $\mathbf{W}_{res}$ are the output layer weights (model and reservoir components, respectively) for the hybrid model. For the training, Eq. (3) was solved for the output weight using a ridge regression and the time-series data of the matrices on both sides. This training is equivalent to minimising the following cost function $J(\mathbf{W})$, which is given by

$$J(\mathbf{W}) = \sum_{k=-K+1}^{0} \|\mathbf{v}^h(k\Delta t) - \mathbf{v}^o(k\Delta t)\|^2 + \beta_{mod}\|\mathbf{W}_{mod}\|^2 + \beta_{res}\|\mathbf{W}_{res}\|^2, \tag{4}$$

where $\mathbf{v}^o$ is the input state vector consisting of observational values, $\beta_{mod} = 10^0$ and $\beta_{res} = 10^{-4}$ are regularisation parameters for ridge regression, and $\|\cdot\|^2$ denotes the sum of squared entries in a matrix. The direct solution to minimising the cost function using ridge regression is given as (Arcomano et al., 2023)

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

$$\mathbf{W} = \begin{pmatrix} \mathbf{V}_o\mathbf{V}_p^T & \mathbf{V}_o\mathbf{R}^T \end{pmatrix} \begin{pmatrix} \mathbf{V}_p\mathbf{V}_p^T + \beta_{mod}\mathbf{I} & \mathbf{V}_p\mathbf{R}^T \\ \mathbf{R}\mathbf{V}_p^T & \mathbf{R}\mathbf{R}^T + \beta_{res}\mathbf{I} \end{pmatrix}^{-1} \tag{5}$$

where $\mathbf{V}_o$, $\mathbf{V}_p$, and $\mathbf{R}$ are vectors of which column k corresponds to $\mathbf{v}_o(k\Delta t)$, $\mathbf{v}_p(k\Delta t)$, and $\mathbf{r}(k\Delta t)$, respectively.

The flowcharts for training and prediction of the hybrid model are shown in Fig. 1. First, the model was trained following the flowchart in Fig. 1a. The input state vector $\mathbf{v}^o(t)$ is defined by the observations. This vector is used as the prognostic variable in the SCM and is converted to form the input state vector for the reservoir. When the vector is converted, a small amount of white noise $\epsilon(t)$ (random white noise, amplitude less than 1% of the prognostic value) is added to the values to obtain a stable solution for the ML component. After the computation of the SCM and updating the reservoir using Eq. (1), $\mathbf{v}^p(t+\Delta t)$ and $\mathbf{r}(t+\Delta t)$ are used as input vectors in Eq. (3), and $\mathbf{v}^h(t+\Delta t)$ is replaced with $\mathbf{v}^o(t+\Delta t)$, and finally the output layer weight $\mathbf{W}$ is obtained.

Secondly, the model is used for prediction as shown in Fig. 1b. The prognostic variables are input as $\mathbf{v}^h(t)$ in the SCM and are transformed into $\mathbf{u}^h(t)$ for the input layer of the reservoir. Then, $\mathbf{v}^p(t+\Delta t)$ and $\mathbf{r}(t+\Delta t)$ are outputs of the SCM and the reservoir, and using the trained output layer weight, $\mathbf{v}^h(t+\Delta t)$ is computed. This updated hybrid model value is used as the prognostic variable in the next timestep. By repeating these computations, a solution using the hybrid model is obtained.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

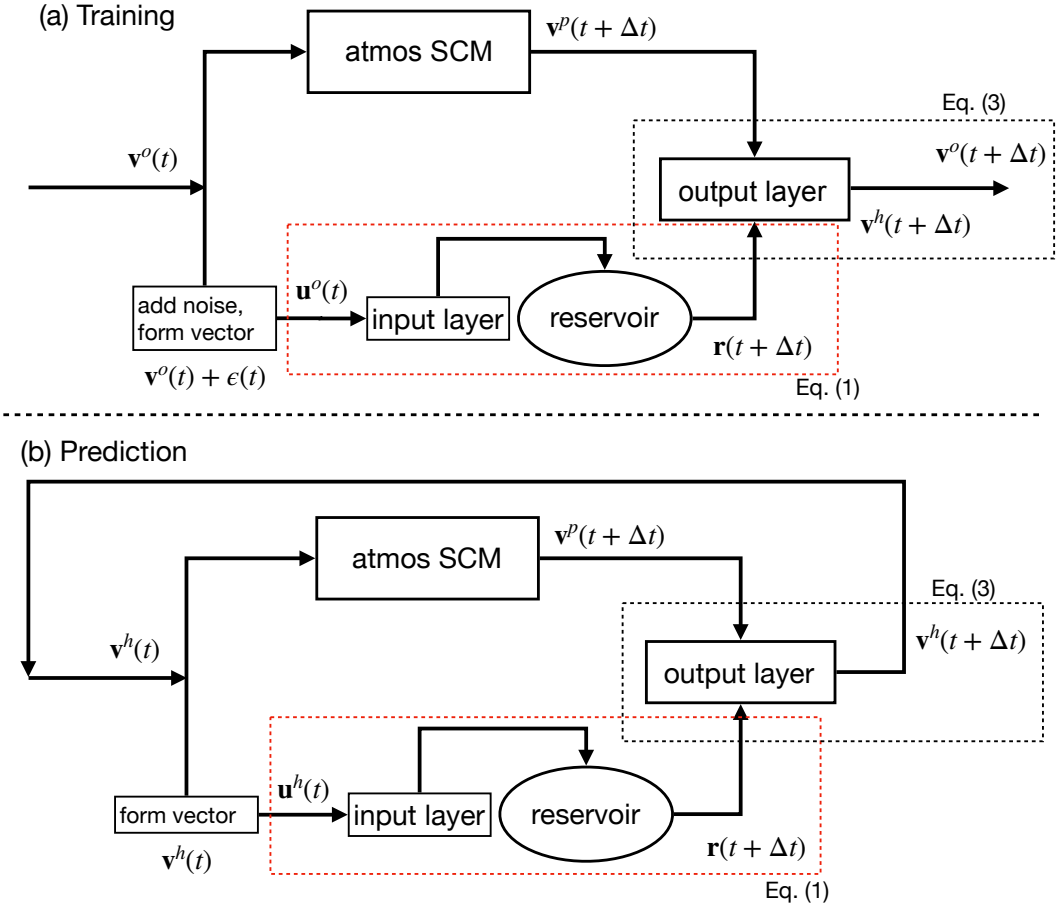

Figure 1: Flowcharts for training and prediction of the hybrid model. Red and black dotted boxes represent parts for Eqs. (1) and (3), respectively. The RC comprises of an input layer, reservoir, and output layer, and this behaves as the ML component of the hybrid model.

2.2 Hyperparameters and configuration

The hyperparameters used in RC which affect the performance of the model are configured based on previous studies (Pathak et al. 2018; Wikner et al., 2020; Arcomano et al., 2022). The time interval for training and synchronisation in the prediction is $\Delta t = 6$ hours. The dimension of the reservoir state vector is set to $D_r = 100$. A preliminary experiment using up to $D_r = 1000$ did not show clear improvements, so the smaller setting was used. Assuming that the average number of connections per node in the RC structure is $\kappa = 6$, the entries of $\mathbf{A}$ are randomly chosen so that they are nonzero with a probability of κ/D_r. The nonzero random number in this matrix is given range of $[-0.5,0.5]$. To determine the stability of the model, the spectral radius of RC is set to $\rho = 0.6$ by scaling $\mathbf{A}$ with its eigenvalue. The entries of $\mathbf{B}$ are simply set to

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

nonzero random numbers ranging from -1 to 1. The configuration of the hybrid ML in the SCM is based on that of Arcomano et al. (2022). While their study adopted synchronised points for all model levels, the present study adopted points limited to ten model levels corresponding to 1000 hPa to 100 hPa pressure levels with a 100 hPa interval. The temperature and humidity were chosen to be synchronised and so the input state vector has a dimension of 20. This is a relatively sparse synchronisation setting compared with that used by Arcomano et al. (2022). Preliminary experiments also showed that a further increase of the synchronised points (e.g., levels with 50 hPa interval) did not significantly enhance the model's performance. When the prognostic variables are used for the input state vector, they are normalised following minimum-maximum scaling so that the value can range from -1 to 1, which is considered suitable for the activation function employed in the present RC.

3 Model, experimental setup, and training data

3.1 Single-column model

The present model is an atmospheric SCM of AFES version 4 (Ohfuchi et al., 2004; Enomoto et al., 2008; Kuwano-Yoshida et al., 2010). The details of its parameterisation are summarized by Baba (2020). The vertical resolution is 48 sigma-levels which is identical to that used by Baba (2020). The default convection scheme of AFES is Emanuel scheme (Emanuel, 1991; Emanuel and Živković-Rothman, 1999) and this is used in this study, although a new convection scheme of Baba (2019) is available for this AGCM. In the SCM, only temperature and humidity are solved in the SCM using the governing equations with the fixed pressure levels (e.g., Randall and Cripe, 1999). The zonal and meridional winds are given by external forcing data. For simplicity, vertical advection of the prognostic variables is not computed by the model, but horizontal and vertical advection tendencies are given by the forcing data.

3.2 Experimental setup and training datasets

Several Intensive Observational Period (IOP) datasets were used to evaluate the SCM's performance. Their names, locations, periods, and references are summarised in Table 1. The IOP cases are broadly categorised into two types, i.e., tropical convection and midlatitude land convection cases (simply referred to as tropical and midlatitude cases, hereafter). These different experiments were used to evaluate the validity of the present hybrid model across the different latitudes. The IOP datasets also provide observed zonal winds, temperature, and humidity. The winds were used for the physical parameterisations, while the temperature and humidity were used to evaluate the performance of the models. The SCMs were forced using the horizontal and vertical advection tendencies of temperature and humidity at each pressure level. These forcings were assigned to the SCMs regardless of training and prediction periods. To obtain robust model performance regardless of the model's uncertainty, four ensemble members were considered for the original model run by adding a small perturbation to the initial condition (Hack and Pedretti, 2000; Davies et al., 2013; Bouttier et al., 2022). The preliminary

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

experiments showed that the initial perturbation caused only a small spread, even when a larger number of ensemble members was used.

For the training of the hybrid model, 6-hourly MERRA-2 (Gelaro et al., 2017) datasets were used. The datasets give time-dependent vertical one-dimensional temperature and humidity as the true profiles, horizontal and vertical tendencies for temperature and humidity, as well as sea surface temperature except for the midlatitude cases in which the skin temperature was determined by the land scheme of the model. To evaluate the hybrid model, training data from one month before the target month with two or five years shifted from the target year are used. For example, November 1987, 1990, 1994, and 1997 training data were used for the TOGA case (Table 1) which starts from December 1992, and the hindcast runs were conducted using the trained hybrid models. In addition to the year shift of the training data, unshifted or 1-month shifted training data were considered in the midlatitude cases (e.g., starting from June (1-month shifted) or July (unshifted) for ARM95). The reason for this setting will be explained later. Based on this setting, four ensemble members were considered for the hybrid model runs.

Table 1: List of IOP cases used for the model evaluations. Here, lat and lon denote the centre location of latitude and longitude of the SCM.

Name	Full name	Location (lat, lon)	Period (date, length)	Reference	Type
TOGA	Tropical Ocean Global Atmosphere	-2, 154	Dec 1992, 21 days	Webster and Lukas (1992)	Tropical convection
TWP-ICE	Tropical Western Pacific Convection	-12, 131	Jan 2006, 26 days	May et al. (2008)	Tropical convection
DYNAMO	Dynamics of the Madden Jullian Oscillation	-1, 73	Oct 2011, 90 days	Yoneyama et al. (2013)	Tropical convection
ARM95	ARM Southern Greate Plains	36, 263	Jul 1995, 18 days	Zhang and Lin (1997)	Midlatitude land convection
ARM97	ARM Southern Greate Plains	36, 263	Jun 1997, 30 days	Zhang and Lin (1997)	Midlatitude land convection

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

4 Results and discussion

4.1 Vertical profile of temperature and humidity biases

The model bias of the SCM was measured using normalised L2 norms for temperature and humidity for the tropical cases (Fig. 2). The L2 norms were measured using 6-hourly outputs during the IOP periods and they were normalised for the periods. The original model indicates a large model bias at higher altitude for temperature and a large model bias at lower altitude for humidity. The bias does not vary regardless of the initial small perturbation, because each member exhibits a similar bias profile. These features are common for all three tropical cases. The hybrid model reduces the model biases for temperature and humidity, although the degree of reduction is dependent on the case, and there are only small differences regardless of the year of the training data. This result indicates that the transfer learning-based hybrid model is useful for reducing model bias in the tropical cases and provides similar results even with different training data. Among the cases, the hybrid model was found to greatly reduce the temperature and humidity biases in DYNAMO compared with TOGA and TWP-ICE.

Figure 3 shows vertical profiles of the correlation coefficients and normalised standard deviation (STD) between the observed and model values. For both temperature and humidity, the hybrid model produces correlation coefficients comparable with the original models, but sometimes they are slightly degraded (TOGA and TWP-ICE), or they are improved in the low to mid-levels (DYNAMO). Focusing on the normalised STD, it was found that the hybrid model reduces the overestimated STD in the original model for all cases. Therefore, the results imply that the hybrid model does not work well for reducing phase (correlation) error with the observation but does work well to reduce the amplitude error in these cases.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

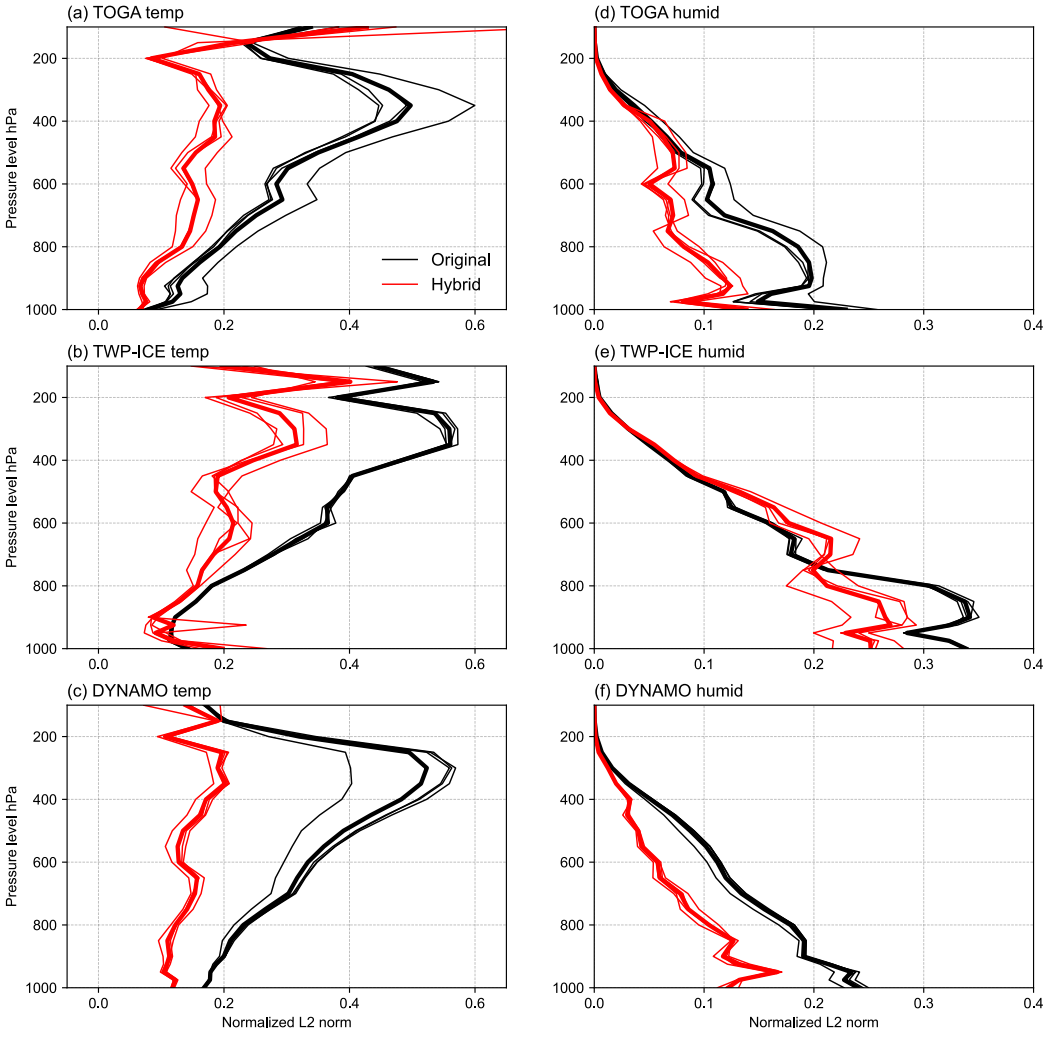

Figure 2: Vertical profiles of normalised L2 norms for (a)-(c) temperature and (d)-(f) humidity in the tropical cases. The black thin lines indicate the results of ensemble members (thick line is ensemble mean). In addition, the red thin lines indicate results of the hybrid model using different training data from a different year (thick line is a mean of these results).

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

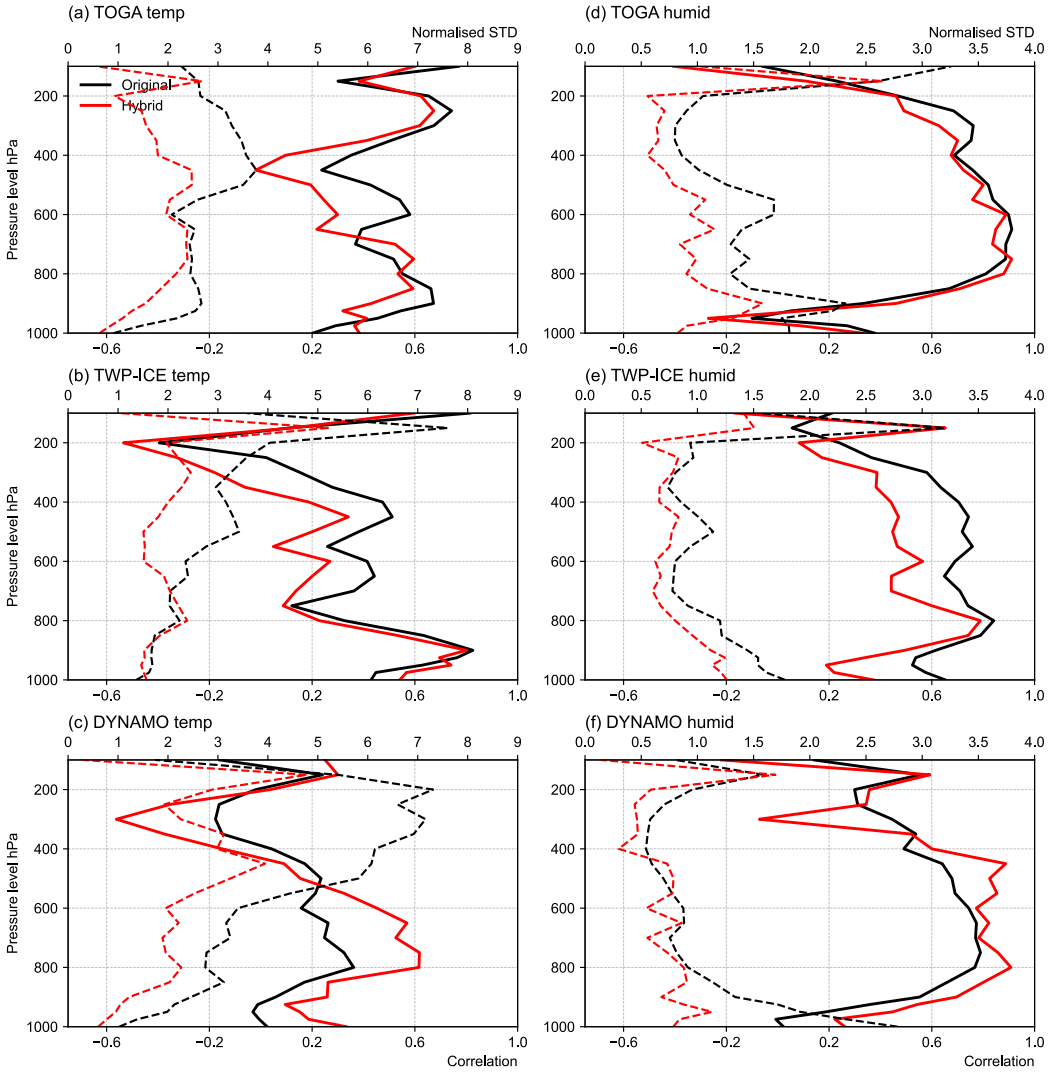

Figure 3: Vertical profiles of correlation coefficients (solid lines, bottom horizontal axis) and normalised standard deviation (STD, dashed lines, top horizontal axis) between observed and model values for (a)-(c) temperature and (d)-(f) humidity in the tropical cases. The black and red (solid and dashed) lines correspond to the original and hybrid models, respectively. The normalised STD is defined by the model value divided by the observed value. Only ensemble mean and mean of hybrid model results were used to estimate the correlation coefficient and normalised STD.

The model bias of SCM in the midlatitude cases is compared as done for the tropical cases (Fig. 4). The degree of bias reduction appears to be smaller than in the tropical cases for both temperature and humidity, but the hybrid model reduces the biases regardless of the training data. Notably, the bias reduction is much more sensitive to the training periods compared with the tropical cases. It is apparent from Fig. 4 that the hybrid model trained with unshifted training data exhibits

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

smaller errors than the model trained with 1-month shifted training data, especially in ARM97. This means that the hybrid model cannot reduce model bias if the training data are obtained from one month earlier. This also means that a simultaneous training period is necessary, even if the year of the training data is shifted from the target period. Since the hybrid model with unshifted data produces better results, the ensemble mean of this case will be used in the following analyses in both tropical and midlatitude cases.

The correlation coefficients and normalised STD were estimated for the midlatitude cases as done for the tropical cases (Fig. 5). The hybrid model presents slightly better correlation and better STD values than the original model for temperature and humidity. It should be noted here that all normalised STDs of the cases are smaller than those of the tropical cases, meaning that amplitude errors in these cases are initially smaller than those of the tropical cases. However, there are large biases represented by the L2 norm at high altitude (Figs. 4a and 4b). These biases in the midlatitude cases are larger than those appeared in the tropical cases at high altitudes.

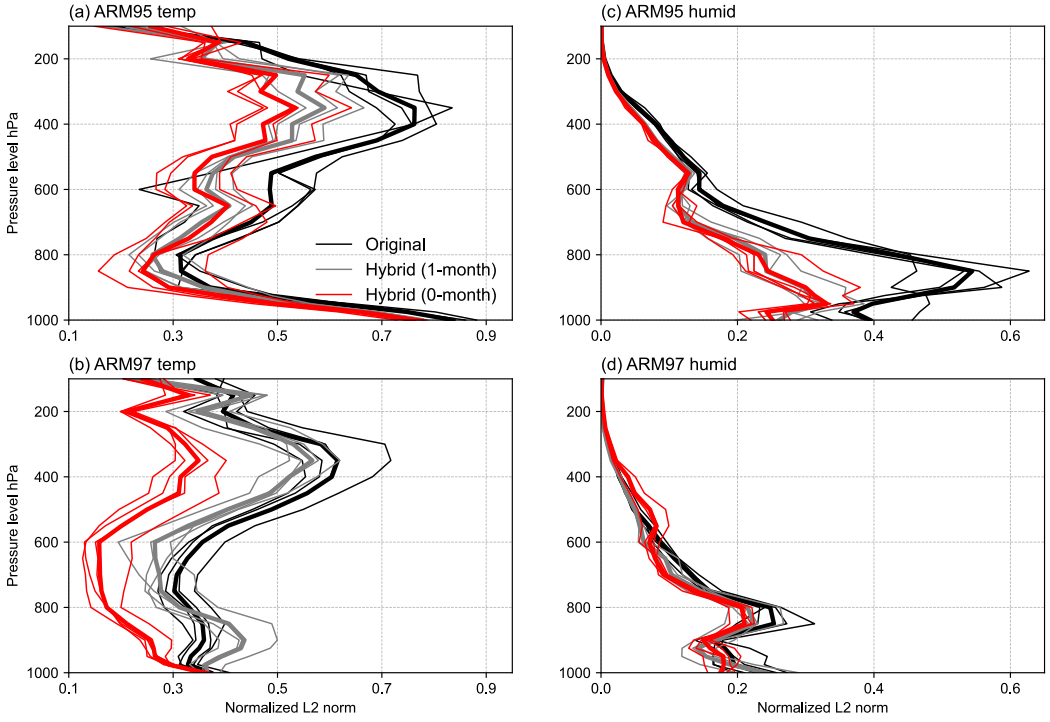

Figure 4: As Fig. 2 but for midlatitude cases. The black lines indicate original model results, and red and grey lines indicate hybrid model results, but their training data are different. The grey and red lines are hybrid model results using unshifted and 1-month shifted training data, respectively. The thick red and grey lines indicate means of respective results using the hybrid models with different training data.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

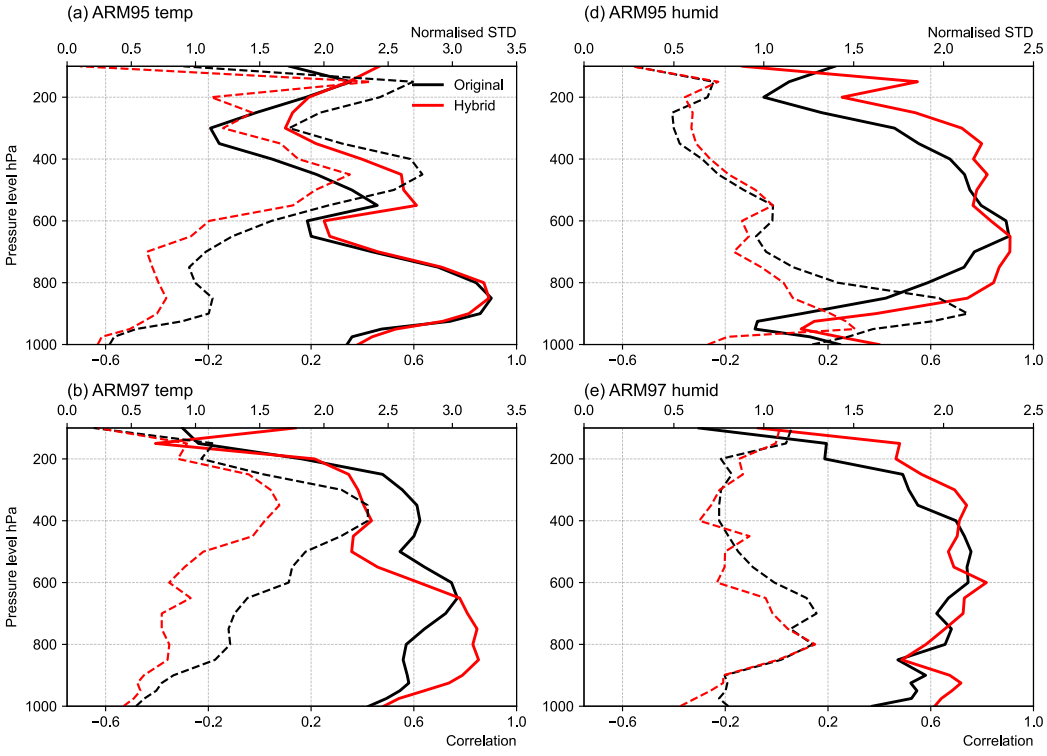

Figure 5: As Fig. 3 but for midlatitude cases.

220

4.2 Statistics of bias

A quantitative comparison of relative hybrid model bias with respect to the original model for the tropical cases is shown in Fig. 6a which indicates overall relative error between the original and hybrid models. All relative biases were found to be less than unity, meaning that the present hybrid model can successfully reduce the model bias compared with the
225 original model in the tropical cases. The bias reduction effect is the largest (smallest) in DYNAMO (TWP-ICE), and the bias reduction of temperature is the larger than that of humidity. Indeed, in the DYNAMO case, the hybrid model reduces excessive amplitude error compared with the original model (Fig. 3). The quantitative bias reductions for the midlatitude cases are compared in Fig. 6b. Although all hybrid models reduce the model biases, this result apparently shows that the model with unshifted training data is superior to that with 1-month shifted training data. This suggest that unshifted (but
230 different year) training data are more appropriate for learning the bias of midlatitude cases and can reduce the overall model bias better.

Figure 7 summarises the performance of both models using Taylor diagrams (Talor, 2001). In the tropical cases, the hybrid model greatly reduces the amplitude error for the temperature but not as much for the humidity. On the other hand, there are only small amplitude improvements in the midlatitude cases compared with the tropical cases for both temperature

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

and humidity. The reason for this is that the amplitude error of the original model is very significant in the tropics, but this is highly mitigated in the midlatitude in the original model (Baba, 2020; 2021). Therefore, the degree of bias reduction is strongly dependent on the features of the original model. In terms of correlation, the hybrid model slightly improves the correlation coefficients in both temperature and humidity for all cases. Therefore, it can be concluded that the hybrid model works well for reducing amplitude error and works slightly well for reducing the phase error.

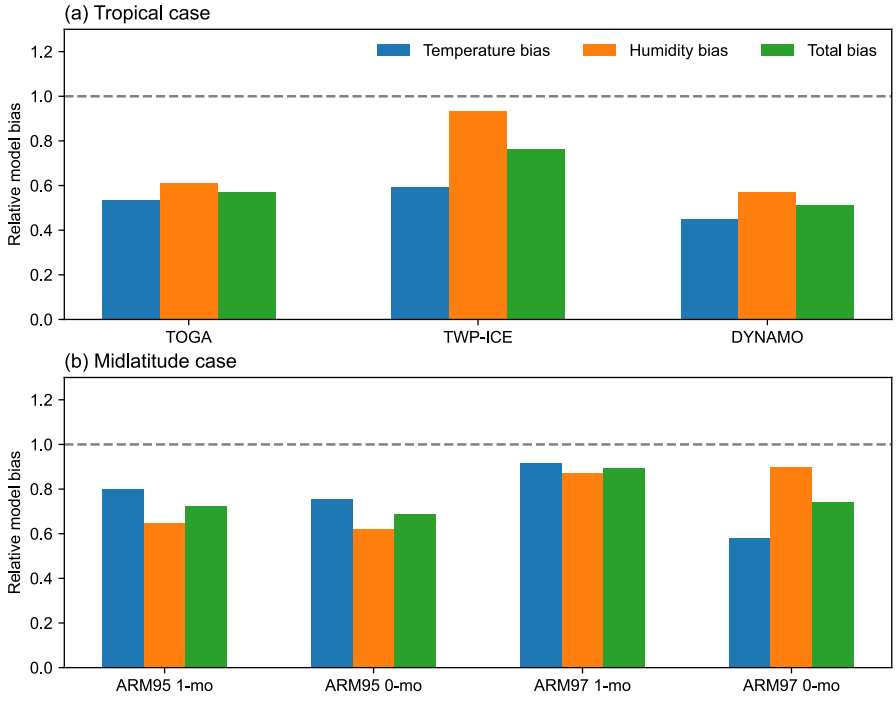

Figure 6: Quantitative comparison of relative hybrid model bias for temperature and humidity with respect to the original model in (a) tropical cases and (b) midlatitude cases. The model bias for each variable was estimated from the vertically integrated absolute difference between model and observed values, while the relative model bias is defined as the ratio between original and hybrid model biases (i.e., hybrid model bias divided by the original model bias). The total bias in the comparison was estimated by averaging the temperature and humidity biases.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

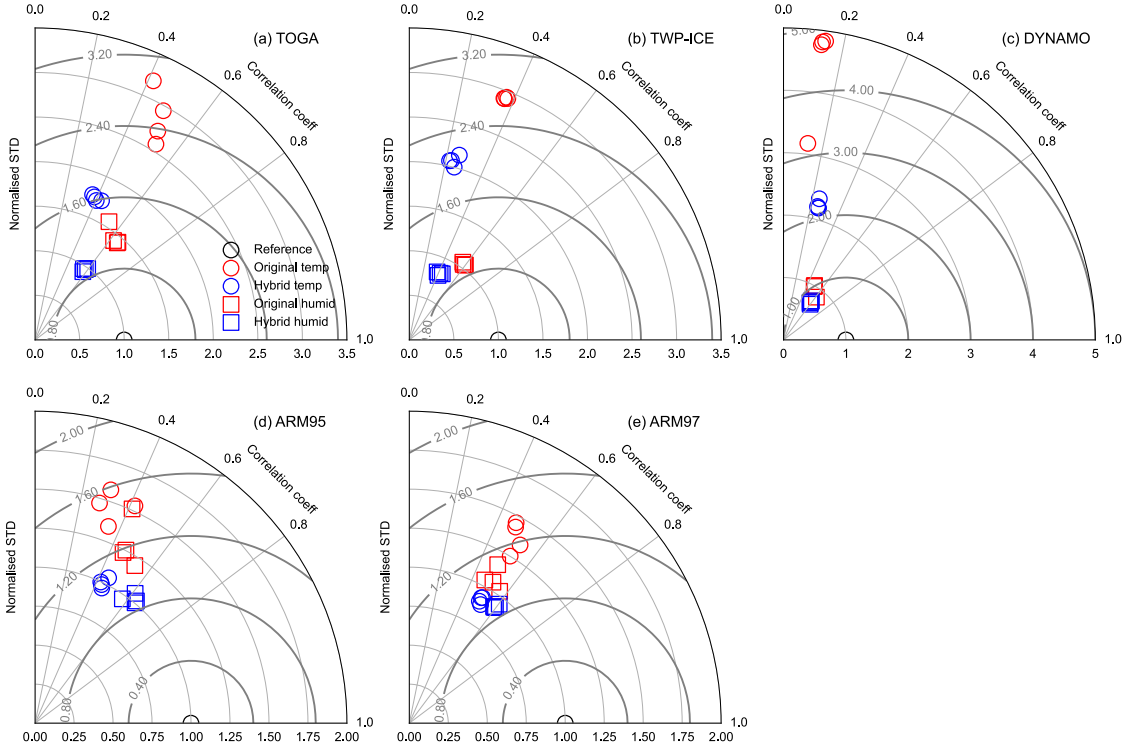

Figure 7: Taylor diagrams for temperature (circles) and humidity (squares) for all cases. Red and blue coloured circles (or squares) correspond to the original and hybrid model, respectively. The standard deviation (STD) of the model results was normalised using the observed values. The Taylor diagrams are drawn for the vertically averaged normalised STD and correlation coefficient.

4.3 Bias reduction effect

The above results revealed that the hybrid model can reduce model bias in both tropical and midlatitude cases. However, the degree of bias reduction varies with the cases. In particular, the effect is greater in tropical cases than in midlatitude cases especially in the amplitude error. To understand the bias reduction effect, time-dependent temperature and humidity biases are compared.

The temporal variations of the temperature and humidity biases for tropical cases are compared (Fig. 8). In all three cases, the hybrid model reduces temperature biases, especially in the mid to upper levels. The bias reduction effect is the greatest in the DYNAMO case because the strong cold bias in the original model is removed by the hybrid model. This feature is derived from the original cold model bias. The employed convection scheme in the model tends to cause cold bias in the upper levels in the tropics (Baba, 2020; 2021). The dry bias in humidity mainly appears in the lower levels in all cases and is also considered to stem from the original model bias (Baba, 2020; 2021). In the TOGA and DYNAMO cases, the

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

strong dry biases are highly mitigated compared with the original model. In contrast, the dry bias is not as reduced in the TWP-ICE case.

A similar comparison for the model biases was conducted for the midlatitude cases (Fig. 9). The original model also exhibits cold biases in the mid to upper levels also in these midlatitude cases and shows dry biases in the low levels. The hybrid model with 1-month shifted training data shows a clear bias reduction for the cold bias in both cases. The biases are further reduced if the unshifted training data are used in ARM97 but not significantly reduced in ARM95. These results indicate that the hybrid model with appropriate training data can further improve the model bias. Although there is no clear difference in ARM95 using different training data, the unshifted training data can reduce the overall error (Fig. 6), so this finding is consistent with the result of ARM95.

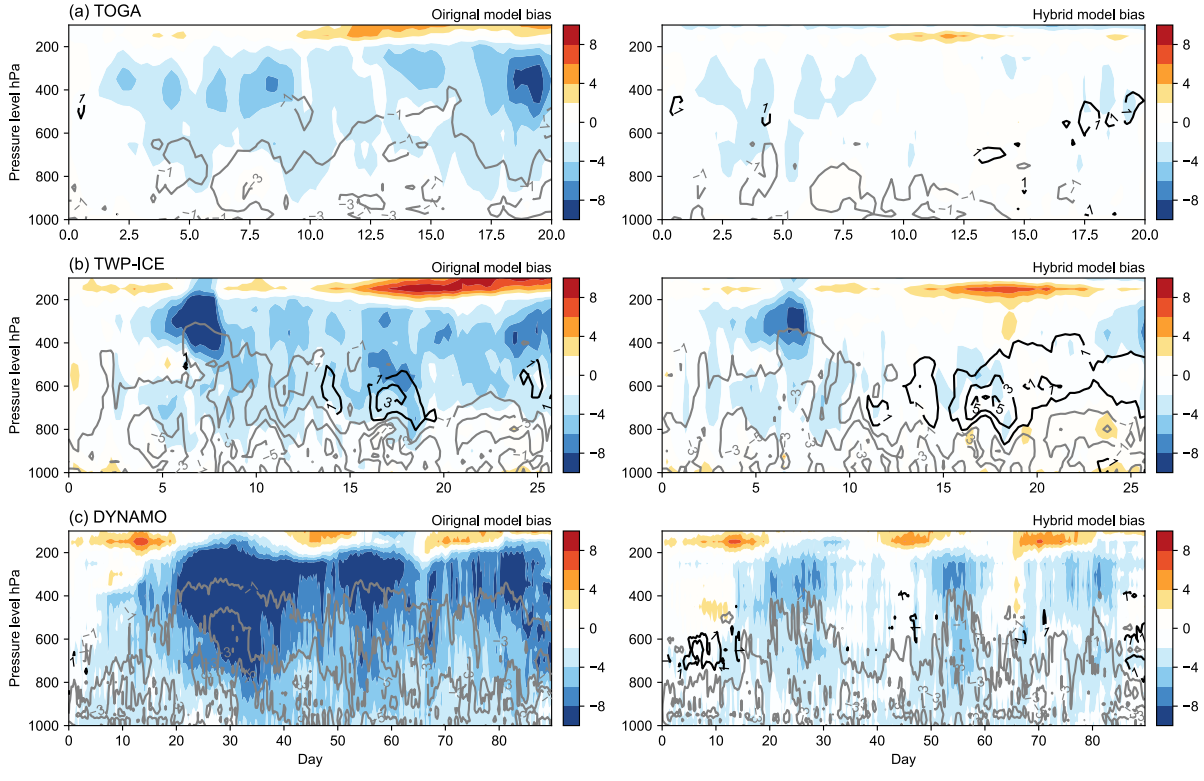

Figure 8: Time evolution of temperature bias (shading, unit: K) and humidity bias (contour line, unit: g kg^{-1}, positive and negative values are shown using black and grey lines) compared with the observed values in the tropical cases. The original and hybrid model biases were estimated from the ensemble means minus the observed values.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

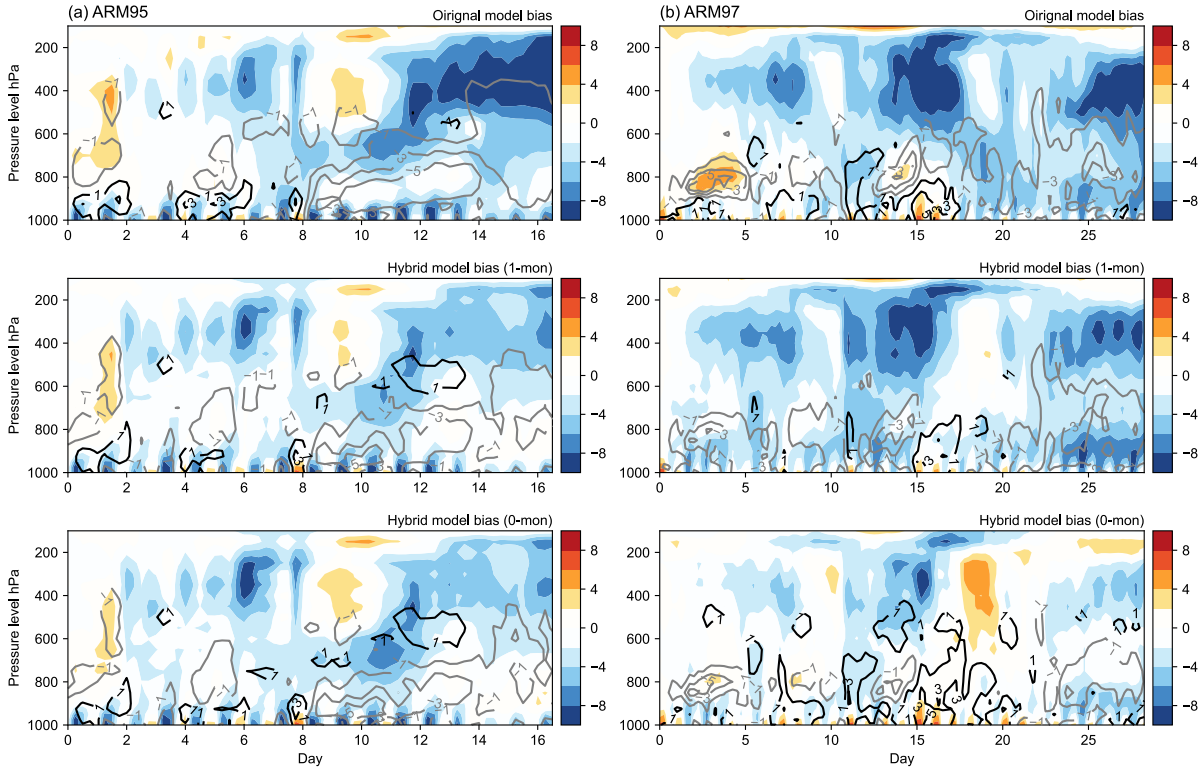

Figure 9: As Fig. 8 but for the midlatitude land convection cases. Note that left and right column panels are for (a) ARM95 and (b) ARM97, respectively, and each row corresponds to different models.

The reduction in bias is produced by the corrections made by the ML component of the hybrid model. Thus, it is worth investigating the time evolution of these corrections to understand how the ML component makes these improvements. Figure 10 shows the time evolution of the hybrid model's corrections for temperature and humidity in the tropical cases. The hybrid model provides a positive temperature tendency when the original model exhibits cold biases, meaning that the correction compensated for the original model's cold bias. Since the areas of cold bias and positive temperature tendency in the time-altitude coordinates are well correlated, the correction is considered to have a significant effect on the temperature bias. The hybrid model also provides a positive humidity tendency in the area where the original model simulated a dry bias. However, there are some discrepancies. For example, the hybrid model produces a negative humidity tendency despite there being no clear wet bias in the TOGA and TWP-ICE cases. In the DYNAMO case, the hybrid model computed a negative humidity tendency, despite the dry biases occurring in the low to mid-levels. Therefore, it appears that the bias correction effect by the hybrid model is less effective for humidity than for temperature.

Similar bias correction effects were also observed in the midlatitude cases (Fig. 11) with positive temperature corrections occurring in the low to upper levels in the ARM95 and ARM97 cases. Unlike the tropical cases, negative temperature corrections can coexist owing to the higher atmospheric variability compared to the tropical cases. The humidity

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

correction tendencies are more complicated than those for temperature, as seen in the tropical cases. Similar to the tropical cases, negative humidity correction appears even though the original model shows dry biases.

These results suggest that the hybrid model efficiently reduces temperature biases since the temperature bias profiles have relatively continuous large-scale temporal and spatial variation, while the humidity bias does not have such variation and the ML component appears to be poor at compensating for such bias. This leads to only a small reduction in bias for the humidity. However, the temperature bias profile varies depending on the latitude, so the reduction is also dependent on the latitude, with the reduction being smaller in the midlatitude than that in the tropics. These features strongly influence the effectiveness of bias reduction by the hybrid model.

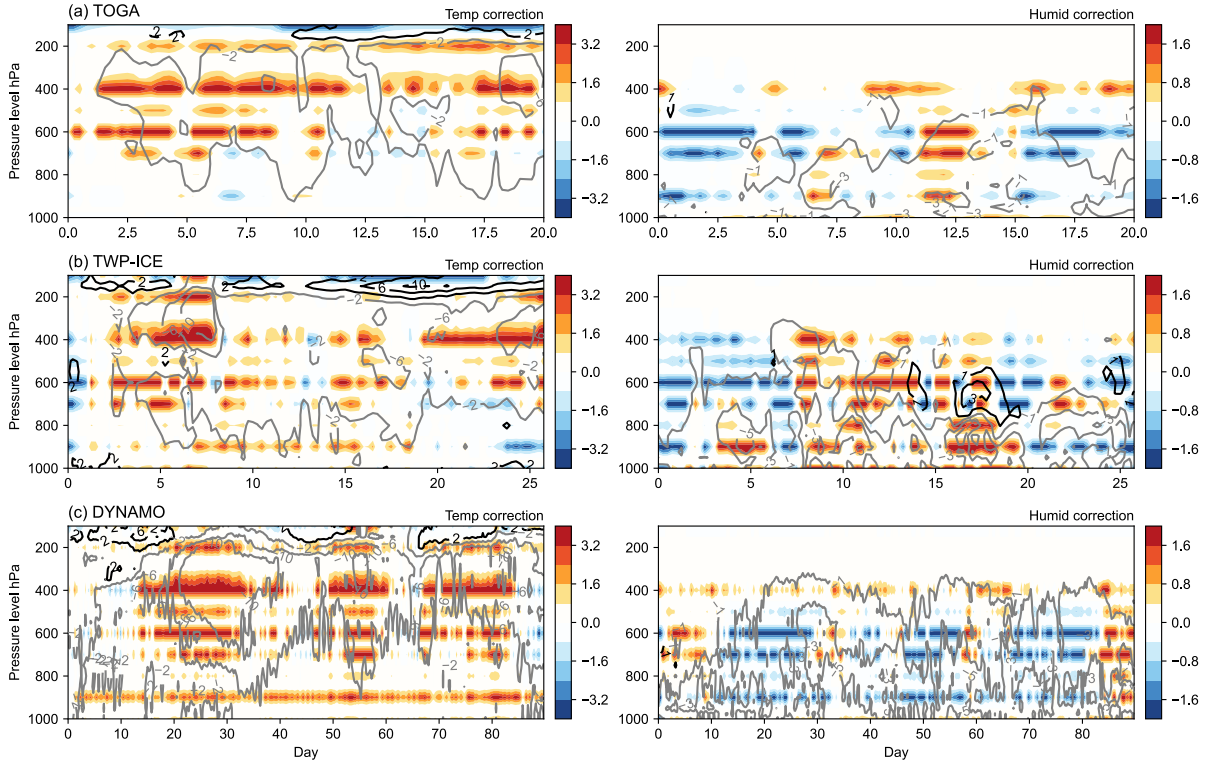

Figure 10: Temporal variation of the hybrid model's correction tendency for temperature (left, shading, unit: $\times 10^{-3}$ K s^{-1}) and humidity (right, shading, unit: $\times 10^{-3}$ g kg^{-1} s^{-1}) in the tropical cases. The temperature and humidity biases from the original model are shown by black (positive bias) and grey (negative bias) contour lines.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

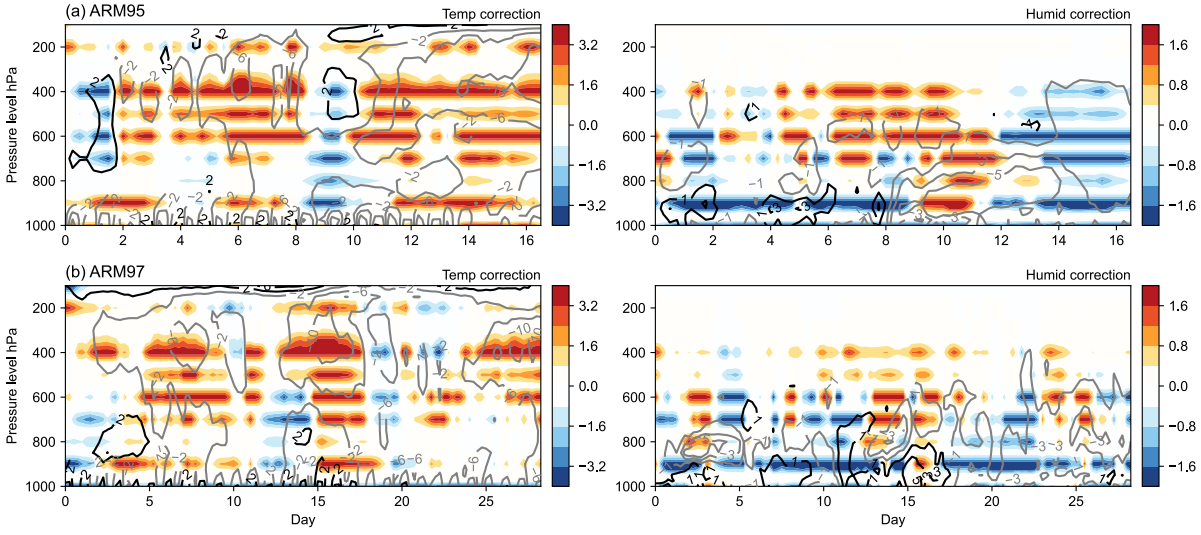

Figure 11: As Fig. 10 but for the midlatitude cases.

4.4 Reduced bias components

To evaluate which kind of bias is reduced by the hybrid model, the power spectra of temperature and humidity biases in the tropical cases are compared for the original and hybrid models (Fig. 12). In all cases, the hybrid model yields lower power spectra than the original model for periods longer than around 3 days (90% significance level in all levels for temperature but only low level for humidity). This means that the hybrid model reduces long-range biases more effectively than it does short-range biases. In addition, this bias reduction occurs in low, mid, and upper levels. These reduced long-term biases imply that systematic model bias is effectively reduced, i.e., it is effective at reducing the mean state bias. Since improving the mean state bias mitigates anomalous fields (e.g., Baba and Giorgetta, 2020), this feature of the hybrid model is beneficial for predicting weather and climate variability in the long term. In contrast to the other cases, the DYNAMO case does not display a clear power decrease in ranges shorter than 3 days, but does yields a large bias reduction (e.g., Fig. 2). This result is consistent with the above assumption that the bias reduction is due mainly to the mean state bias reduction. A similar comparison is conducted for the midlatitude cases (Fig. 13). Although the hybrid model reduces the overall power spectra longer than 3 days (90% significance level in all levels for temperature but upper and low or mid levels for humidity), the degree of reduction is smaller than observed in the tropical cases. There are slightly large power reductions in the long term, and this feature occurs for both temperature and humidity.

These results demonstrate that bias reduction by the hybrid model tends to result mainly from mean state bias reduction rather than reduction in the high-frequency components. Due to the temperature features, the hybrid model tends to reduce temperature bias better than it does humidity, but the effect is dependent on the latitude, as seen in Figs. 10 and 11. When the vertical profiles of STD for temperature and humidity are compared (Fig. 14), the midlatitude cases show an

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

apparently larger temporal deviation, especially for temperature, in contrast to the tropical cases. Such large variability in the prognostic variables causes decrease of bias reduction and this explains the poorer bias reduction in the midlatitude cases.

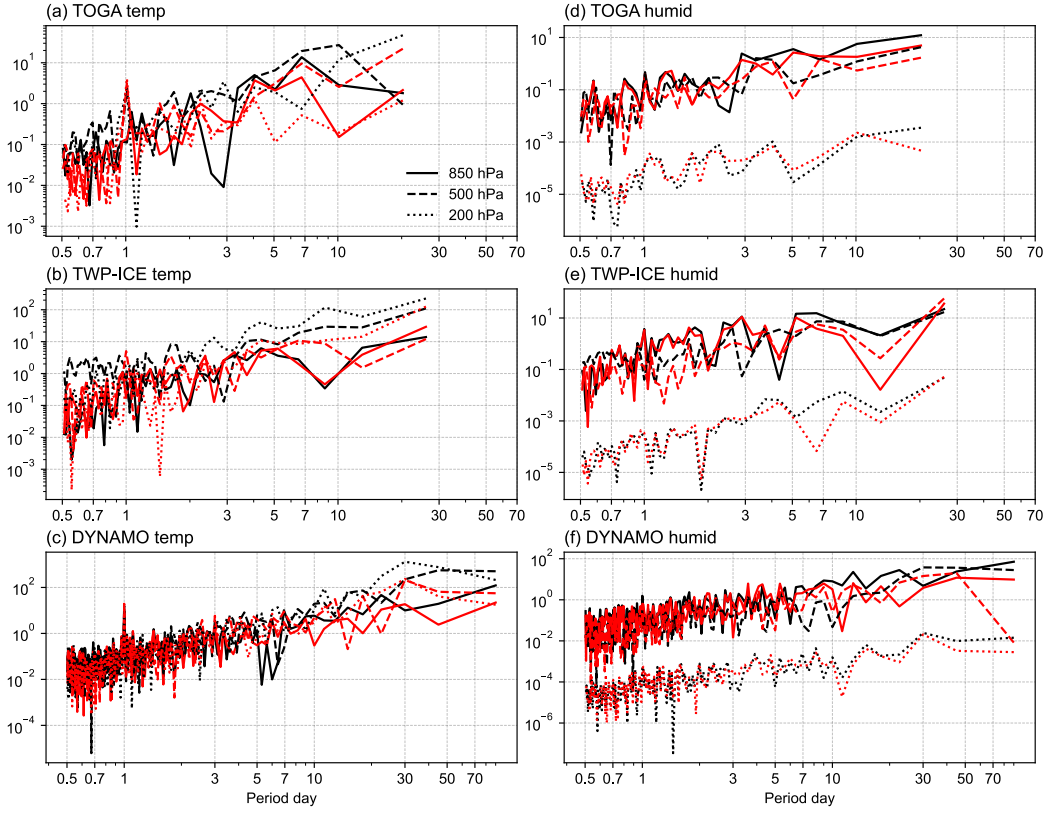

Figure 12: Power spectra in terms of period in the tropical cases (a)-(c) for temperature biases and (d)-(f) humidity biases. Black and red lines correspond to the power spectra of the original model's bias and the hybrid model's bias, respectively. The solid, dashed, and dotted lines represent the temperature biases at low (850 hPa), mid (500 hPa), and upper (200 hPa) levels, respectively.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

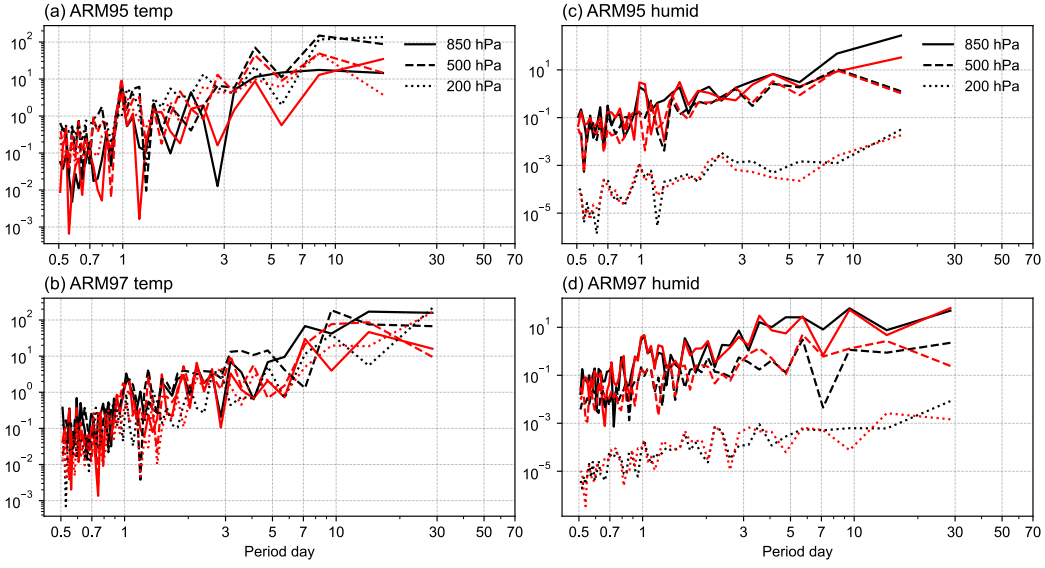

Figure 13: As Fig. 12 but for the midlatitude cases.

345

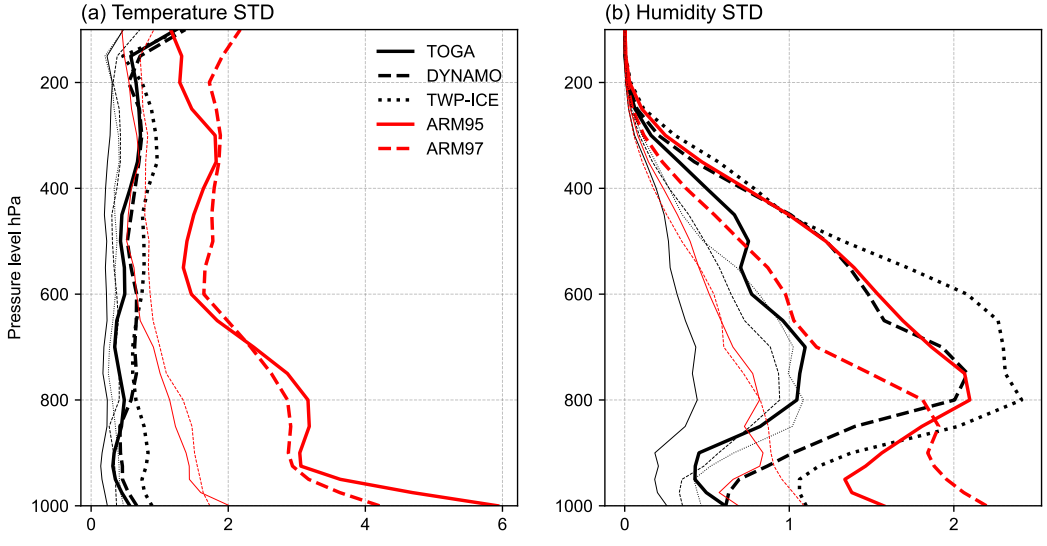

Figure 14: Comparison of vertical profiles of standard deviation (STD) for (a) temperature (K) and (b) humidity (g kg^{-1}) estimated from all IOP datasets with a 6-hourly interval (thick lines). The thin lines represent temperature or humidity difference between their maximum and minimum values during the IOP period (×10^{-1} K or ×10^{-1} g kg^{-1}).

350

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

4.5 Diagnostic variables

A model's ability to simulate precipitation is important for numerical weather and seasonal predictions since precipitation is closely related to the extreme events that cause significant losses or damage to infrastructure, economics, or human life. To evaluate the prediction ability of the hybrid model, 6-hourly accumulated precipitation is compared with the observed value for all cases (Fig. 15). It was found that, regardless of the models and cases, the simulated and observed precipitations agree well. This means that external forcing is the key factor in determining the accumulated precipitation rather than differences between the models. However, since the cases used here are limited to an SCM simulation, this finding may change in a multi-column (i.e., three-dimensional) simulation. In such a case, the ability of the hybrid model to simulate the horizontal and vertical advection tendencies will be important. Although the accumulated precipitation is well simulated and both models present comparable performance, there is a difference between the models. The correlation coefficients and root-mean square error (RMSE) between the model results and the observed values are slightly different depending on the model. The hybrid model generally shows higher value of the correlation coefficients while it shows larger RMSE than the original model depending on the cases.

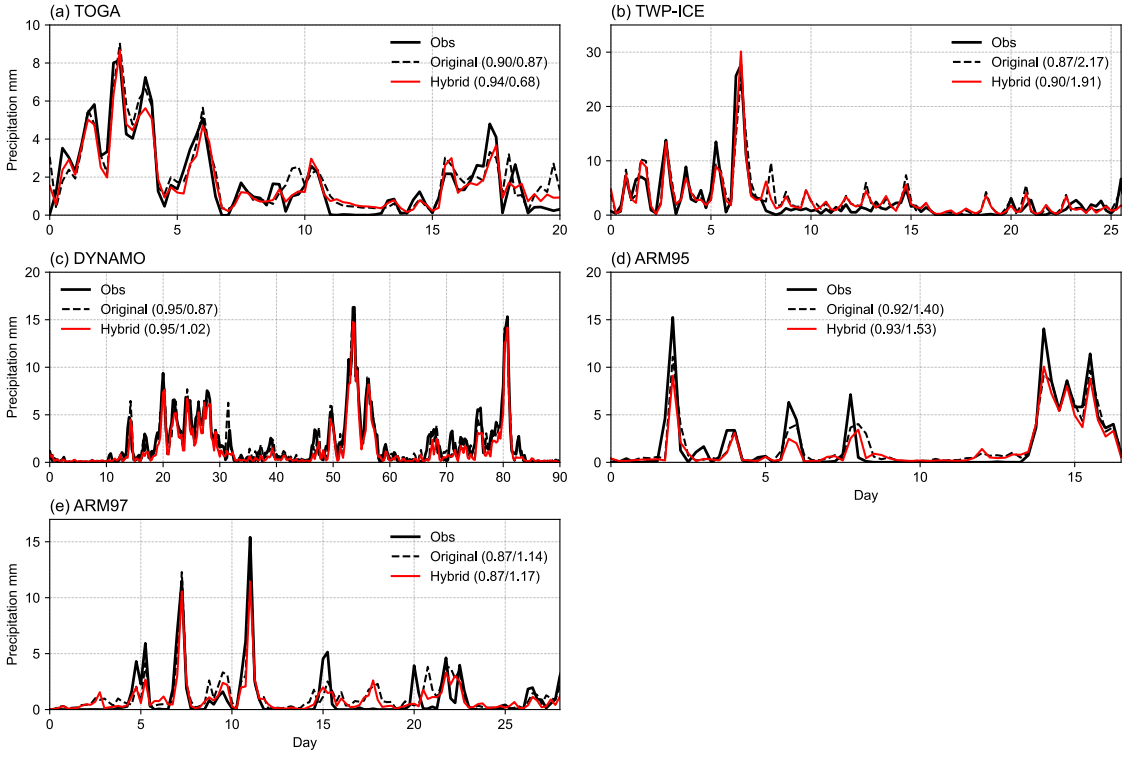

Figure 15: Comparison of temporal variations of 6-hourly accumulated precipitation (mm) for all cases. The value shown in parenthesis in the legends indicates the correlation coefficient (between observed and model results in time) and RMSE. The original and hybrid model results were obtained from their ensemble means.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

Since the total energy of the system is controlled not only by advection tendency but also by the model's top and bottom heat fluxes, it is worth investigating the features of these fluxes in terms of energy conservation (Baba, 2015). Figure 16 compares the temporal variations of the fluxes at the model's top and bottom. Generally, the original and hybrid models simulate similar features in the temporal variation, and they simulate outgoing longwave radiation (OLR) in good agreement with observations (Figs. 16a-16c). However, the normalised RMSE (nRMSE) especially for surface heat fluxes is different (Figs. 16f-16h). In TOGA, the hybrid model simulates surface heat fluxes slightly more accurately than the original model, as shown by the small error with respect to the observed values. However, TWP-ICE and DYNAMO show much larger nRMSE especially for the latent heat flux and show large deviation from the observed value. Therefore, it is speculated that the limited bias reduction in TWP-ICE (Fig. 6) may be derived from the latent heat flux error, which does not arise as a problem in DYNAMO since the error is mainly derived from the upper-level cold bias in this case. In the midlatitude cases, similar features are not seen, as the original and hybrid models show comparable and small nRMSEs with respect to the observed values (Figs. 16i-16j). This means that the error of surface heat flux is less significant in the midlatitude cases than in the tropical cases.

Consequently, in all cases, OLR is accurately simulated while the surface heat flux is generally not as accurately simulated, although there is clear reduction in bias in the temperature and humidity. This is believed to be because the OLR is greatly influenced by atmospheric conditions, while the surface heat flux is dominated by only the near-surface atmospheric conditions. If the surface heat flux is large and there is a large deviation from the observed, it does affect the bias reduction of the hybrid model. In such cases, the surface heat flux should be synchronized in the ML component of the hybrid model.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

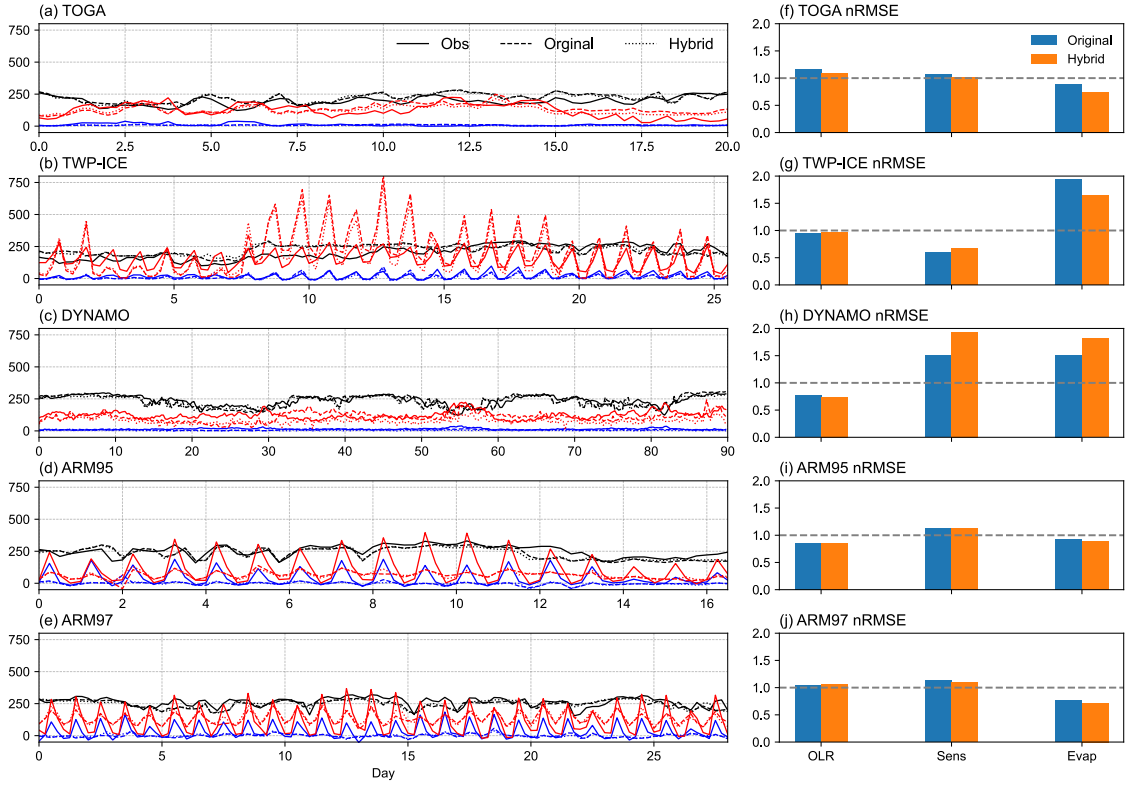

Figure 16: (a)-(e) Temporal variation of 6-hourly outgoing longwave radiation (OLR, black line), sensible heat flux (blue line), and latent heat flux (red line). Units for these values are W m^{-2}. Solid, dashed, and dotted lines correspond to IOP value, original model, and hybrid model. (f)-(j) Normalised root-mean square error (nRMSE) of original and hybrid models using the observed values for each variable (OLR: outgoing longwave radiation, Sens: sensible heat flux, Evap: latent heat flux). The RMSE was normalised using a STD of the observed value.

5. Summary and conclusions

The hybrid machine-learning (ML) model was utilised in an atmospheric single-column model (SCM) to examine the validity of a transfer learning-based hybrid model because transfer learning is useful for a hybrid model which lacks sufficient training data near the target period, thus enabling a longer-range prediction with limited training data. This was examined for both tropical convection and midlatitude land convection cases using intensive observational period (IOP) datasets. To evaluate transfer learning, two- or five-year shifted training data, i.e., similar period but different year to the target period, were used for the hybrid model. Bias reduction due to the ML component of the hybrid model was particular focus of the evaluations.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

In the tropical cases, the hybrid model using transfer learning generally made significant reductions in temperature and humidity biases, although the degree of reduction depended on the IOP cases. This result indicates that the hybrid model with transfer learning works well for tropical convection. On the other hand, although the biases are reduced by the hybrid model, degree of the bias reduction is smaller in the midlatitude cases than in the tropical cases. Specifically, the midlatitude cases required unshifted training data in order to learn the correct model bias during the target period, although the year of the training data is different from the target year. This implies that seasonality of the learning data is more important for the midlatitude than for tropics. The statistics of the bias showed that the bias reduction is more significant for the amplitude of temperature than for humidity, and the reduction is greater than that for phase error. It was also found that the reduction is greater in the tropical cases than in the midlatitude cases.

In the temporal variation of temperature and humidity biases, the hybrid model reduced the cold biases in the mid to upper levels, and the dry biases in the low levels better than the original model. These are known biases originating mainly from the employed convection scheme and are reduced by the ML component of the hybrid model, taking the form of a bias correction tendency. When the temporal variation of the correction tendency was analysed, it was found that the tendency is temporally and spatially continuous for the temperature bias but not continuous for the humidity bias. In the midlatitude cases, the bias has larger temporal and spatial variability, and the corresponding correction tendency is more complicated, so it is not as effective in areas where the model exhibits large biases.

Analysis of the power spectra of biases for the temporal component showed that bias reduction occurred mainly in the mean state, i.e., the low-frequency (long-range) part rather than the high-frequency (short-range) part. The hybrid model mainly reduces the biases which have periods longer than 3 days, but in some cases does not reduce the biases which have periods shorter than 3 days. These results demonstrate that the hybrid model can reduce the systematic model bias which originates from the mean state, but it is poor at reducing short-range variability.

Finally, ability of the hybrid model for simulating diagnostic variables, such as precipitation, OLR, and surface heat fluxes is evaluated. It is found that the hybrid model simulates slightly improved or comparable accumulated precipitation, while it does not improve the surface heat fluxes. Both original and hybrid models show large deviation in the surface heat fluxes depending on the cases, and this is assumed to cause the model biases in the low levels. Therefore, synchronization for the surface heat fluxes may be necessary for further improvements.

In conclusion, the transfer learning-based hybrid ML model was found to be practically useful even when the training period was shifted from the target period. The bias reduction was more notable for temperature than for humidity, owing to its relatively continuous temporal and spatial profiles. Moreover, the hybrid model reduced mean state bias, particularly for temperature, in the tropics more than that in the midlatitude. However, it should be noted that the effectiveness of bias reduction may be related to the degree of bias in the original model. The hybrid model is effective in reducing the mean state bias, so it is effective in reducing the biases in relatively long-range atmospheric variability. Since seasonal and intraseasonal predictions which predict anomalous fields greatly rely on the model's means state bias and climate drift (Baba and Giorgetta, 2020; Baba, 2023b), such features will contribute to enhance these prediction systems.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

The present findings are useful for enhancing the existing dynamical prediction models (e.g., Baba, 2023a; 2025a) if the hybrid ML is incorporated in the models (e.g., Patel et al., 2025). The performance evaluation for the prediction systems with the transfer learning-based hybrid ML should be conducted in a future study.

Code and data availability

The source code of the AFES v4 is available at https://gitlab.com/aosg_public/afes under a 2-clause BSD license. The exact version of the model used to produce the results in this paper is archived on Zenodo under DOI: 10.5281/zenodo.17060903 (Baba, 2025b), including input data and scripts to run the model. Figures, visualised data, and analysis scripts are archived on Figshare under DOI: 10.6084/m9.figshare.30060628 (Baba, 2025c).

Author contribution

Yuya Baba: Conceptualisation, formal analysis, investigation, methodology, software, validation, visualisation, writing (original draft preparation), writing (review and editing).

Competing interests

The author declares that he has no conflict of interest.

Acknowledgements

The code of hybrid ML model was developed, and all computations were performed on the Earth Simulator version 4 (ES4) of JAMSTEC. The author thanks Dr. Akira Yamazaki for providing useful information regarding hybrid ML models.

References

Arcomano, T., Szunyogh, I., Wikner, A. et al.: A hybrid approach to atmospheric modeling that combines machine learning with a physics-based numerical model. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002712, 2022.

Arcomano, T., Szunyogh, I., Wikner, A., et al.: A hybrid atmospheric model incorporating machine learning can capture dynamical processes not captured by its physics-based component. *Geophys. Res. Lett.*, **50**, e2022GL102649, 2023.

Baba, Y.: Sensitivity of the atmospheric energy budget to two-moment representation of cloud microphysics in idealized simulations of convective radiative quasi-equilibrium. *Quart. J. Roy Meteorol. Soc.*, **141**, 114-127, 2015.

Baba, Y.: Spectral cumulus parameterization based on cloud-resolving model. *Clim. Dyn.*, **52**, 309-334, 2019.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

Baba, Y.: Shallow convective closure in a spectral cumulus parameterization. *Atmos. Res.*, **233**, 104707, 2020.

Baba, Y., Influence of a spectral cumulus parameterization on simulating global tropical cyclone activity in an AGCM. *Quart. J. Roy. Meteorol. Soc.*, **147**, 1170-1188, 2021.

Baba, Y.: Improvements in SINTEX-F2 seasonal prediction by implementing an atmospheric nudging scheme. *Int. J. Climatol.*, **43**, 6900-6923, 2023a.

Baba, Y.: Impact of convection scheme on ENSO prediction of SINTEX-F2. *Dyn. Atmos. Ocn.*, **103**, 101385, 2023b.

Baba, Y.: Seasonal prediction of atmospheric rivers in the Western North Pacific using a seasonal prediction model. *Atmos. Sci. Lett.*, **26**, e1299, 2025a.

Baba, Y.: Model, input data, and run scripts for "Transfer learning-based hybrid machine learning in single-column model of AFES v4". [code], https://doi.org/10.5281/zenodo.17060903, 2025b.

Baba, Y.: Figures and scripts for "Transfer learning-based hybrid machine learning in single-column model of AFES v4". [data], https://doi.org/10.6084/m9.figshare.30060628, 2025c.

Baba, Y., Giorgetta, M. A.: Tropical variability simulated in ICON-A with a spectral cumulus parameterization. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001732, 2020.

Baba, Y., Ujiie, M., Ota, Y., Yonehara, H.: Implementation and evaluation of a spectral cumulus parameterization for simulating tropical cyclones in JMA-GSM., *Qaurt. J. Roy. Meteorol. Soc.*, **150**, 2045-2068, 2024.

Baba, Y., Ujiie, M.: Evaluation of JMA-GSM typhoon forecasts using a new spectral cumulus parameterization in Prapiroon (2018) and Hagibis (2019). *SOLA*, 2025, doi:10.2151/sola.2025-047.

Bogenschutz, P. A., Tang, S., Caldwell, P. M. et al.: The E3SM version 1 single-column model. *Geosci. Model. Dev.*, **13**, 4443-4458, 2020.

Bouallegue, Z. B., Clare, M. C. A., Magnusson, L., et al.: The rise of data-driven weather forecasting. *Bull. Amer. Meteorol. Soc.*, **105**, E864-E883, 2024.

Bouttier, F., Fleury, A., Bergot, T. et al.: A single-column comparison of model-error representations for ensemble prediction. *Boundary-Layer Meteorol.*, **183**, 167-197, 2022.

Davies, L., Jakob, C., Cheung, K., et al.: A single-column model ensemble approach applied to the TWP-ICE experiment. *J. Geophys. Res. Atmos.*, **118**, 6544-6563, 2013.

De Burgh-Day, O. C., Leeuwnburg, T.: Machine learning for numerical weather and climate modelling: a review. *Geosci. Model Dev.*, **16**, 6433-6477, 2023.

Emanuel, K. A.: A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci.*, **48**, 2313-2335, 1991.

Emanuel, K. A., Živković-Rothman, M.: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.*, **56**, 1766-1782, 1999.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

Enomoto, T., Kuwano-Yoshida, A., Komori, N., et al.: Description of AFES 2: Implements for high-resolution and coupled simulations. *In: Kevin, H., Wataru, O. (Eds.), High resolution numerical modelling of the atmosphere and ocean*. Springer, 77-97, 2008.

Gelaro, R., and coauthers: The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419-5454, 2017.

Gettelman, A., Truesdale, J. E., Bacmeister, J. T., et al.: The single column atmosphere model version 6 (SCAM6): Not a SCAM but tool for model evaluation and development. *J. Adv. Model. Earth Syst.*, **11**, 1381-1401, 2019.

Hack, J. J., Pedretti, J. A.: Assessment of solution uncertainties in single-column modeling frameworks. *J. Climate*, **13**, 352-365, 2000.

Karniadakis, G. E., Kevrekidis, I. G., Perdikaris, P. et al.: Physics-informed machine learning. *Nature Rev. Phys.*, **3**, 422-440, 2021.

Kochkov, D., Yuval, J., Langmore, I., et al.: Neural general circulation models for weather and climate. *Nature*, **632**, 1060-1066, 2024.

Kuwano-Yoshida, A., Enomoto, T., Ohfuchi, W.: An improved PDF cloud scheme for climate simulations. *Quart. J. Roy. Meteor. Soc.*, **136**, 1583-1597, 2010.

Long, S. M., Xie, S. P.: Uncertainty in tropical rainfall projections: Atmospheric circulation effect and the ocean coupling. *J. Climate*, **29**, 2671-2687, 2016.

Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.*, **3(3)**, 127-149, 2009.

May, P. T., Mather, J. H., Vaughan, G., et al.: The tropical warm pool international cloud experiment. *B. Am. Meoteorol. Soc.*, **89**, 629-646, 2008.

Ohfuchi, W., Nakamura, H., Yoshioka, M. K., et al.: 10-km mesh mesoscale resolving simulations of the global atmosphere on the Earth Simulator-preliminary outcomes of AFES (AGCM for the Earth Simulator). *J. Earth Simul.* **1**, 8–34, 2004.

Patel, D., Arcomano, T., Hunt, B. et al.: Prediction beyond the medium range with an atmosphere-ocean model that combines physics-based modelling and machine learning. *J. Adv. Model. Earth Syst.*, **17**, e2024MS004480, 2025.

Pathak, J., Wikner, A., Fussell, R., et al.: Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos*, **28**, 041101, 2018.

Pathak, R., Sahany, S., Mishra, S. K.: Uncertainty quantification based cloud parameterization sensitivity analysis in the NCAR community atmosphere model. *Sci. Rep.*, **10**, 17499, 2020.

Randall, D. A., Cripe, D. G.: Alternative methods for specification of observed forcing in single-column models and cloud system models. *J. Geophys. Res.*, **104**, 24,527-24,545, 1999.

Rasp, S., Pritchard, M. S., Gentine, P.: Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. U. S. A.*, **115 (39)**, 9684-9689, 2018.

https://doi.org/10.5194/egusphere-2025-4612
Preprint. Discussion started: 17 December 2025

Schuh, A. E., Jacobson, A. R.: Uncertainty in parameterized convection remains a key obstacle for estimating surface fluxes of carbon dioxide., *Atmos. Chem. Phys.*, **23**, 6285-6297, 2023.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram., *J. Geophys. Res.*, **106**, 7183-7192, 2001.

Wang, W., Zhang, M.: An analysis of parameterization interactions and sensitivity of single-column model simulations to convection schemes in CAM4 and CAM5. *J. Geophys. Res. Atmos.*, **118**, 8869-8880, 2013.

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K. et al.: Correcting weather climate models by machine learning nudged historical simulations. *Geophys. Res. Lett.*, **48**, e2021GL092555, 2021.

Webster, P. J., Lukas, R.: TOGA COARE: The coupled ocean-atmosphere response experiment. *B. Am. Meteorol. Soc.*, **73**, 1377-1416, 1992.

Weiss, K., Khoshgoftaar, T. M., Wang, D. D.: A survey of transfer learning. *J. Big Data*, **3**, 9, 2016.

Wikner, A., Pathak, J., Hunt, B. et al.: Combining machine learning with knowledge-based modeling for scalable forecasting and subgrid-scale closure of large, complex, spatiotemporal systems., *Chaos*, **30**, 053111, 2020.

Yoneyama, K., Zhang, C., Long, C. N.: Tracking pulses of the Madden-Julian Oscillation. *B. Am. Meteorol. Soc.*, **94**, 1871-1891, 2013.

Zhang, M. H., Lin, J. N.: Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements, *J. Atmos., Sci.*, **54**, 1503-1524, 1997.