This paper proposes an online correction framework that integrates a machine learning model (FuXi) with a physical numerical weather prediction model (CMA-GFS) through spectral nudging.

Overall, I find this work valuable as a careful and useful validation of an existing methodology. In particular, it demonstrates that the hybrid nudging framework can be successfully implemented within a different model system, which may be helpful for operational centers considering similar approaches. The manuscript is also clearly written and provides a detailed description of the workflow, which makes it easy to follow.

From a scientific perspective, however, the main contribution appears to be a system-specific implementation of an already established paradigm rather than a fundamentally new methodological development. The approach is largely consistent with prior work such as Husain et al. (2025), including scale-selective spectral nudging, handling of coarse vertical ML outputs, and the use of vertical weighting to mitigate inconsistencies. While this is not a limitation in itself, it may be helpful for the authors to more clearly position their work relative to these studies and clarify whether there are specific aspects in which their implementation provides advantages.

We thank the reviewer for the insightful comments, which have helped us improve the clarity and impact of our manuscript.

We would like to clarify the origin of the methodology. We independently proposed and developed this online correction framework that integrates a machine learning model (FuXi) with a physical numerical weather prediction model (CMA-GFS) via Spectral Nudging. Late in our research, we came across the work of Husain et al. (2025), which is largely consistent with ours in both concept and methodology. Their work, of course, was initiated earlier than ours. We fully acknowledge that Husain et al. (2025) have done an excellent job in formulating and validating their version of the method. Their work is rigorous and valuable, and we do not claim any priority over them. Instead, we view this independent convergence as strong evidence that the hybrid Spectral-Nudging framework is a timely and promising direction for integrating machine learning into operational NWP.

The process of my research is as follows: Our team have background in dynamical cores and variational data assimilation. Previously, We conducted some work on reference profiles and implemented 3-D and 4-D reference profile based on the CMA-GFS model (Su et al., 2025. doi: 10.1007/s13351-025-4114-5). Our initial idea was to introduce forecasts from the FuXi model as a time-varying 4-D reference profile into the dynamical solver, so that the reference state would stay close to the real atmosphere during integration and thereby improve the spatial discretization accuracy of the dynamical core. However, after implementing this method, it did not get significant improvements. Naturally, We then thought that direct nudging would certainly yield better effects. However, FuXi exhibits overly smoothed small-scale features and a rapidly decaying KES. This led me to the idea of using spectral methods to separate the large-scale components before applying nudging. Since the 4-Dvar module in CMA-GFS already contains spectral-grid transformation routines, the implementation was straightforward.

From a technical perspective: The inference module for FuXi and the preprocessing module connected to CMA-GFS were already completed during the development of 4DRef. For the vertical nudging coefficients, since FuXi output only contain 13 pressure levels, which are sparse near the surface and at the upper levels, applying a vertical profile and nudging only the middle levels became a necessary choice; The truncation wavenumber was determined through my own tests based on KES and real forecasts.

Therefore, objectively speaking, both the work of ECCC (Husain et al., 2025) and ECMWF (Polichtchouk et al.. 2024, 2026), as well as my own work, have developed similar forecast systems based on the Spectral Nudging (SN) method using their own physical and ML model. ECCC is the first center to implement this approach. ECMWF, by contrast, trained AIFS on model levels, thereby addressing the issue of sparse vertical levels in the ML model, and established an ensemble forecasting system using the SN method. My work indeed does not represent a novel scientific breakthrough. We did not summarize the limitations of the ECMWF and ECCC methods in the introduction, as doing so would imply an intent to solve these problems, which was not my objective.

After the methodology section, I have added a table comparing the key differences between the ECCC, ECMWF, and my own work across various aspects to facilitate readers' comparison, as following:

TABLE 2. the differences in technical details of the SN method.

	ECCC	ECMWF	CMA
Physics model	GEM	IFS/IFS-ENS	GRAPES-GFS
ML model	GraphCast	AIFS/AIFS-ENS	FuXi
Spectral Transform Method	Discrete cosine transform (DCT)	Spherical Harmonics expansion in IFS	Spherical Harmonics expansion in GRAPES-4Dvar
Truncation wavenumber or wavelength	Soft cutoff between 2750-2250km	T21	T21
Nudging variable	u,v,virtual temperature	vorticity, specific humidity, virtual temperature	u,v,exner pressure,potential temperature
Nudging vertical	850hPa-250hPa	Model levels 50-137	600hPa-200hPa

profiles		(approximately surface to 56hPa)	
Nudging relaxation time	12hour	12hour	6hour

One aspect that could benefit from further clarification is the preprocessing step used to map FuXi outputs from 13 pressure levels to the 87 model levels of CMA-GFS. This step is only briefly mentioned and not specified in detail. Since vertical interpolation is a critical component that can significantly influence the representation of atmospheric structure (especially gradients, stability, and boundary-layer processes), the lack of description raises concerns about reproducibility and scientific validity. It is unclear what interpolation scheme is used, how physical consistency is preserved, and to what extent this preprocessing step may introduce biases or damp important features before nudging is even applied.

We thank the reviewer for raising this important and technically critical point. We agree that the vertical interpolation from FuXi's 13 pressure levels to CMA-GFS's 87 model levels is a key preprocessing step, and our original manuscript did not provide sufficient detail. We substantially expanded the preprocessing procedure in the revised manuscript in line 209-214 as follows:

The detial of preprocessing procedure is as follows: First, in the horizontal direction, the geopotential height, temperature, zonal and meridional winds (h, t, u, v) forecast by FuXi on 13 pressure levels are bilinearly interpolated from 0.25° resolution to the model resolution of 0.125°. Then, in the vertical direction, h, t, u, v on pressure levels are interpolated to p, t, u, v on the 87 model levels using cubic spline interpolation, based on the height coordinates of pressure levels and model levels. Finally, the Π and θ on the 87 model levels are computed by $\Pi = \left(\frac{p}{p_0}\right)^{\frac{R}{C_p}}$ and $\theta = T\left(\frac{p}{p_0}\right)^{\frac{R}{C_p}}$, where p is pressure, p_0 is standard sea-level pressure. R is gas constan, C_p is specific heat capacity at constant pressure.

In the current work, the FuXi outputs are interpolated to the CMA-GFS vertical grid through this unspecified preprocessing step, and the resulting inconsistency is mitigated by applying a vertically varying nudging coefficient that limits the correction primarily to the mid–upper troposphere. However, this approach closely follows that of Husain et al. (2025), who employed a similar vertical weighting strategy to address the same issue. As such, it is unclear what methodological innovation is introduced here beyond adopting an existing workaround.

Thank you for your advice.

We thank the reviewer for this observation. We agree that our vertically varying nudging coefficient, which limits corrections primarily to the mid-upper troposphere, is technically similar to the strategy employed by Husain et al. (2025). However, we would like to clarify the following points.

As explained earlier, our work is not a replication or extension of Hussain's. We initially developed a 4D reference profile based on FuXi forecasts, and later shifted to building an SN-based hybrid system.

The vertical nudging profile was not introduced at the beginning. However, after applying SN, the deviations in the lower and upper levels increased. Since we are most concerned with the forecast leading time at 500 hPa, introducing a vertical coefficient was a natural and straightforward choice.

The vertical profile is only a temporary solution for constructing the hybrid system at the current stage and indeed lacks innovation. In the future, we will attempt to increase the vertical levels of the ML model or directly train FuXi on model levels.

At the same time, the manuscript acknowledges alternative approaches, such as Polichtchouk et al. (2024), who address this limitation more fundamentally by increasing the vertical resolution of the ML model (e.g., 137 levels), thereby reducing the need for ad hoc vertical weighting. Given this, it would be important for the authors to clarify why a similar strategy is not adopted in the present study. Is the choice driven by computational constraints, data availability, or compatibility with FuXi? Without such justification, the current approach appears as a pragmatic but potentially suboptimal solution rather than a deliberate methodological design.

Thank you very much for your advice. This is indeed one of the key issues in our study, and we would like to address it from the following two perspectives.

(1) For machine learning models, a denser vertical hierarchy does not necessarily lead to better forecast performance.

Husain et al. (2025) also discuss the influence of the vertical resolution of machine learning models on the Spectral Nudging system, as stated in the original text: "This study employs the 13-pressure-level version of GraphCast with pretrained weights (learned features of the GNNs) that are available from Google DeepMind. Although a 37-level version is available, only the 13-level variant has been subjected to additional fine-tuning with ECMWF's operational analyses (2016-21), making it more skillful than the 37-level version."

For ML developers, increasing the vertical levels poses no technical difficulties, and computing power is no longer an issue. Yet most ML models still only offer a 13-level version. After discussing with the developers of the ML models, they generally agree that, owing to the contribution of the 500hPa MSE within the overall loss function, the version with 13 pressure levels achieves higher ACC and better RMSE scores.

Therefore, We are also curious whether the large scale circulation forecasting capability (500hPa ACC and RMSE) of ECMWF's 137 model level AIFS can maintain the performance of the version with 13 pressure level, We have not found any relevant comparison in the literature.

(2) Regarding the copyright and model availability:

FuXi was developed by the Institute of Artificial Intelligence at Fudan University and does not belong to CMA. Currently, we are using the publicly released version of FuXi, which includes only the inference code but not the training code. Therefore, we are temporarily unable to retrain the FuXi model.

In the next phase, before the spectral nudging system is put into operational use at CMA, we will communicate with the FuXi team to obtain the training code. We will then use initial fields from CMA-GFS to fine-tune or retrain the model. At the same time, we plan to attempt increasing the number of isobaric levels, or even train the model directly on model levels, to see whether the physical model can be improved in a more comprehensive manner.

This is my preliminary understanding, which may not be fully accurate: If a higher ACC for the 500hPa geopotential height is required, the 13-pressure-level version should be adopted. If comprehensive improvements for the middle and lower troposphere are prioritized, the model level version is preferable, though this will compromise some of the 500hPa scores. Since it is uncertain whether this is correct and there has been no rigorous comparative validation, We have not included this viewpoint in the paper.

We thank the reviewer again for raising this important issue. The descriptions regarding copyright issues and the plan to train FuXi with more levels in future work have been incorporated into the manuscript (line 240-242) .

Regarding presentation, the introduction provides a broad survey of ML-based weather models, but the connection to the core contribution is not clearly articulated. The cited models appear more as a catalogue than as elements that directly motivate the proposed method. Given that the main idea can be summarized concisely — combining ML-derived large-scale circulation with physics-based small-scale consistency via spectral nudging — the introduction could be significantly streamlined. More generally, this issue extends beyond the introduction to the entire manuscript. While the detailed narrative of the research process is informative, the paper would benefit from a more concise and structured presentation. In particular, lengthy descriptive explanations of intermediate attempts or design choices could be reduced, and key ideas could instead be conveyed more effectively through tables, figures, or mathematical formulations.

Thank you for this constructive suggestion. We agree that the previous introduction was indeed somewhat broad and did not sufficiently focus on the core issues.

Following your advice, we have streamlined the overly general literature survey in the introduction and added or expanded content specifically related to the SN method. In addition, we have condensed the descriptions of the subsequent experimental procedures and scheme selections to improve clarity and conciseness.

We believe these revisions will make the manuscript more focused and easier to follow. We thank the reviewer again for helping us improve the presentation.

In addition, the paper does not sufficiently justify why spectral nudging is appropriate in this global modeling context. The evaluation is also limited to internal comparisons within a single modeling framework, without benchmarking against other state-of-the-art ML or hybrid systems, including closely related work such as Husain et al. (2025). This makes it difficult to assess the broader competitiveness or generality of the approach.

We thank the reviewer for this important comment.

The main purpose of out work is to establish a hybrid system of CMA-GFS and FuXi based on the SN method. In the introduction, we discussed the relative strengths and weaknesses of ML models and physical models, and then pointed out that the SN method can effectively combine the advantages of both.

During the evaluation, comparisons were mainly conducted among three systems: CMA-GFS, FuXi, and CMA-SN, where FuXi represents the state-of-the-art ML model. In the introduction, the description of the comparison platform of CMA and ECMWF (Table 1) demonstrates that the current FuXi model achieves world-class performance in large-scale circulation patterns in both winter and summer. In addition, we sincerely apologize that I am currently unable to run the hybrid model as described in Husain's paper, so a comparative evaluation with it cannot be performed.