This manuscript develops a hybrid forecast system by combining the CMA-GFS physical model and the FuXi ML model using the spectral nudging (SN) method. It takes FuXi's strength in large-scale circulation forecasts and CMA-GFS's strength in small-scale details. The hybrid system performs very well and offers useful references for operational NWP centers. I recommend acceptance after minor revisions.

The vertical nudging profile you used is similar to the ECCC approach. Why didn't you follow the ECMWF AIFS method and train the ML model directly on model levels instead of pressure levels?

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) .

Please explain your work's difference from ECMWF and ECCC. Highlight your improvements so readers can easily see your novelty.

Thank you for your suggestion, 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

You initialized FuXi with ERA5 reanalysis, which works for research but not for real-time operations. How do you plan to fix this when moving to operational runs with real-time analysis data?

Thank you for your suggestions, these are indeed key issues to address in our future work.

Our current work focuses on conceptual verification to confirm the feasibility of the SN method and the correctness of the system configuration. The ERA5 dataset is adopted here to initialize the FuXi model, ensuring optimal simulation performance.

To operationalize this system at CMA, we will run the FuXi model initialized with analysis fields from the GRAPES data assimilation cycle. Our tests show that initializing FuXi with GRAPES analysis instead of ERA5 reduces the model's predictable lead time by approximately 1–2 days across seasons and regions, a common issue in other ML models.

We plan to address this through two research directions: 1) Using Transformer-based neural networks to adjust GRAPES analysis fields to better align with ERA5 reanalysis data before applying them to FuXi. 2) Fine-tuning or retraining FuXi with GRAPES reanalysis data (derived from the GRAPES system) and GRAPES analysis fields to enhance its adaptability. Preliminary results from the first approach indicate that the predictable lead time can be extended by approximately one day, particularly over the Southern Hemisphere.

Relevant discussions have also been included in the first part of the future work plan.

Do you plan to extend this SN method to CMA regional models for better tropical cyclone simulation?

Thank you for your advice. They are highly meaningful for the development of regional models, particularly for the prediction of tropical cyclones.

Large-scale steering flow is crucial for typhoon track forecasting, and relevant work is currently underway. Specifically, we plan to extract large-scale circulations from FuXi outputs, and apply online correction to the 1 km-resolution CMA-MESO (national domain) system. In the future, we may also use outputs from the global SN system as initial and boundary conditions for regional model, to improve the performance of regional model in large-scale circulations.

Since a global SN system has already been established, constructing a regional SN system will be relatively straightforward—the overall workflow can be directly adapted from the global one. The main difference lies in: the global model employs a 4D-Var assimilation system, whose core routines (including transformations between lat-lon grids and Gaussian grids, spherical harmonic expansions, etc.) are already developed and do not need to be rebuilt. For regional models, however, corresponding modules must be newly developed. We plan to either reference relevant modules from the WRF model, or develop the code that directly performs expansion and truncation using the discrete cosine transform (DCT) on a regular lat-lon grid.