The authors have created a high-resolution, dynamically downscaled dataset over the Yangtze River Delta called YRD1km, based on the WRF model, dynamically downscaled from the ERA5 reanalysis. The dataset is then further refined using a hybrid nudging approach, combining observational nudging and analysis nudging, to produce a generally skillful 3D wind dataset. The authors verified the dataset against surface observations primarily for June 2022, as well as against radiosonde observations, and for a convective case study event, also in June 2022. In general, the manuscript is well written and presented, but the analysis, methodology, and conclusions drawn from this work (and particularly the verification) are incomplete. I do believe this work is valid and important, but I recommend major revisions and inclusion of additional detail + verification work prior to publication.

We sincerely thank the reviewer for the careful assessment of our manuscript and for recognizing the scientific value and relevance of the YRD1km dataset. We appreciate the constructive comments regarding the completeness of the analysis, methodology, and validation, particularly with respect to the verification strategy.

In response to these concerns, we have substantially revised the manuscript by adding new experiments, expanding the validation period and scope, and clarifying key methodological details. The revised version includes more comprehensive and transparent verification analyses, as well as clearer descriptions of the model configuration and evaluation framework. Below, we provide detailed, point-by-point responses to each comment and describe the corresponding revisions made to the manuscript.

Scientific Comments:

- Though the dataset is valid for the summer months of 2021-2023, most of the verification was conducted in June 2022, with the first few figures and tables focusing solely on June 1, 2022. Additionally, the time period for the daily verification statistics for Table 3 is unclear; also, the time series for Figure 5 wasn't clear (are those averaged MAE/NSE across all stations?). Though I don't doubt that 1-km WRF would generally outperform the 0.25-deg ERA5 for winds, the shown verification statistics should be more transparent and comprehensive; otherwise it feels like "picking and choosing".

Response: Thank you for this important comment regarding the representativeness, transparency, and temporal coverage of the validation strategy.

To clarify the experimental design, June 2022 was selected a priori at the beginning of the study rather than after comparing multiple periods. In particular, 1 June 2022 was chosen as a representative early-summer day to illustrate spatial wind field characteristics, and not because it exhibited especially favorable performance.

To address the concern about transparency and to avoid any impression of selective sampling, the temporal scope of the validation analyses has been substantially expanded. For the methodology and sensitivity experiments (Section 3), all statistics originally based on a single day are now evaluated over a continuous four-day period from 1 to 4 June 2022, resulting in more than 30,000 samples for Table 1 and Table 2. For Section 4.1, the validation period has been further extended to a one-week interval from 1 to 7 June 2022, yielding nearly 60,000 samples for Figure 3. The daily statistics reported in Table 3 correspond explicitly to the same 1–7 June period and are aggregated from the hourly evaluations shown in Figure 3. The time series in Figure 5 (now Figure 6) represent spatially averaged MAE, RMSE, and NSE computed over all available surface stations across the YRD region for every hour of June 2022. These clarifications have been added to the revised text, table captions, and figure captions.

As the validation period was extended, the numerical values of the statistics changed slightly; however, the overall performance differences between YRD1km and ERA5 remain highly consistent. These revisions ensure that the presented verification results are comprehensive, transparent, and representative rather than case-specific.

- The model setup is not particularly clear, e.g. the justification of the physics parameterizations based on sensitivity testing is dubious. Those schemes are the baseline "default" schemes for the WRF model; though I understand that extensive testing of physics schemes would be unfeasible and beyond the scope of this work, many other dynamical downscaling related articles have discussed this and have either used other schemes and/or have provided justification for why they chose the schemes that they did. Additionally, the nesting and initialization setups are unclear, i.e. was two-way nesting used? Did all three domains start at the same time, meaning that the ERA5 initial conditions would be interpolated directly to the 1-km grid? Is a 1 hour spinup really adequate? Are there discontinuities/"jumps" in the data every 6 hours?

Response: Thank you for this constructive comment regarding the clarity of the model configuration, the justification of the physical parameterization schemes, and the nesting and initialization strategy.

Regarding the physical parameterizations, we acknowledge that several schemes adopted in this study represent commonly used baseline configurations within the WRF modeling community. For microphysics, the Thompson scheme (Thompson et al., 2008) was selected because it is widely regarded as well suited for high-resolution simulations and has demonstrated robust performance in representing cloud microphysical processes, which is consistent with the objectives of this study.

For the planetary boundary layer (PBL) scheme, we conducted comparative tests among several commonly used options, including MYJ, MYNN, ShinHong, and YSU. As shown in Figure S2, the YSU scheme exhibits a slight but consistent advantage in near-surface wind performance relative to the other schemes. Although the differences are relatively modest and therefore were not elaborated upon in the main text, this outcome supports the selection of YSU for the final configuration. In addition, following your suggestion, we have cited recent comprehensive sensitivity studies (e.g., Sahu et al., 2025), which evaluated a large number of physics scheme combinations and confirmed that the Thompson–YSU pairing performs favorably for convective and high-resolution applications. To avoid overstatement, we have revised the manuscript wording from “Through sensitivity testing” to “Based on previous studies” when describing the selection of physical parameterizations.

Concerning the model configuration, two-way nesting was employed for all three domains. All domains were initialized simultaneously, with ERA5 initial conditions interpolated directly onto each domain, including the 1-km grid. Wind is a prognostic model variable that adjusts rapidly to the model dynamical framework, in contrast to variables such as precipitation that are indirectly generated and typically may require longer spin-up periods to equilibrate. Accordingly, a 1-hour spin-up was adopted to balance model adjustment and error accumulation. The hourly evaluation results shown in Figure 6 further indicate that no evidence of artificial discontinuities or abrupt “jumps” associated with the 6-hour cycling strategy is observed.

To improve clarity and address this comment, we have revised the Methods section to explicitly describe the nesting strategy, initialization procedure, and the rationale underlying the selection of physical parameterization schemes.

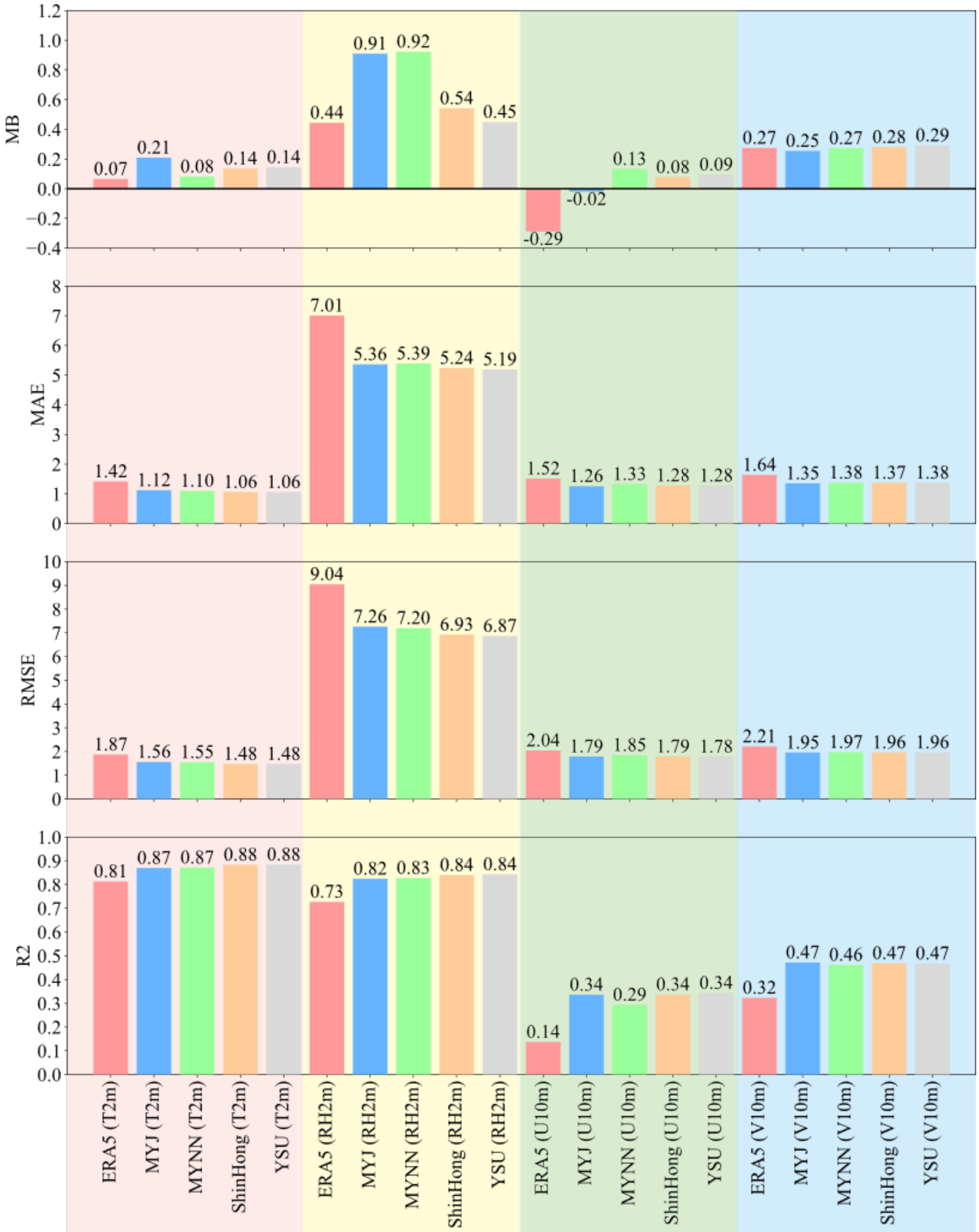

Figure S2. Sensitivity comparison of near-surface temperature and wind simulations using different planetary boundary layer (PBL) schemes in the WRF model. Shown are statistical performance metrics for 2-m temperature (T2m), 2-m relative humidity (RH2m), and 10-m wind components (U10m and V10m), including Mean Bias (MB), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Nash–Sutcliffe Efficiency (NSE). Results are compared among ERA5 and four commonly used PBL schemes (MYJ, MYNN, ShinHong, and YSU). The shaded background indicates different variable groups. Overall, the YSU scheme exhibits slightly better performance for near-surface wind variables, supporting its selection in the final model configuration.

- How much of the skillful verification statistics are due to the WRF model/enhanced spatial resolution, rather than an "edit" of grid point values based on observation nudging? If WRF grid points were nudged on the hour, and verified on the hour (and on the same grid points as the nudging? unclear in the manuscript), then there would be a reduction in error metrics via that "edit".

It would have been nice to see a baseline configuration with WRF being run without any nudging at all, to actually determine how much the nudging improved skill.

Additionally, it wasn't clear how the obs nudging was fully conducted, i.e. what variables were nudged (was it just wind?). Would the other variables (e.g. 2-m temperature, precipitation) be compromised because of this nudging, i.e. is this dataset only useful for wind, and not for other variables?

Response: We thank you for this important and insightful comment regarding the attribution of the skill improvements, the respective roles of enhanced spatial resolution and observational nudging, and the potential implications of the nudging strategy for other variables.

As discussed in the Methods section, we quantitatively assessed the impacts of both observational nudging and high-resolution land-use updates on model performance. When WRF is applied solely as a free-running dynamical downscaling system, its higher spatial resolution can indeed yield improved accuracy relative to ERA5 during the initial forecast hours. However, as forecast lead time increases, model errors tend to accumulate, and performance can degrade relative to ERA5, which is continuously constrained by data assimilation. By contrast, the hourly observational nudging applied in this study directly constrains the model state and therefore contributes to improved verification statistics.

To address the concern that the reported skill gains might arise primarily from a local "editing" effect at nudged grid points, we introduced an explicit independent validation design (Section 4.1.2). In this experiment, approximately 10% of the AWS stations were randomly withheld from the nudging procedure and used exclusively for independent evaluation. These stations provided no direct observational constraints to the model. Validation over a continuous four-day period (1–4 June 2022) shows that YRD1km still consistently outperforms ERA5 at these independent locations, with systematic improvements in MAE, RMSE, and NSE for both wind components (Figure 4). The closer alignment of modeled and observed winds with the 1:1 relationship further confirms that the skill gains are not confined to assimilated grid points, but extend spatially across the domain.

These findings indicate that the improved performance of YRD1km cannot be attributed solely to pointwise nudging "edits." Instead, it reflects the combined effects of enhanced spatial resolution, physically consistent dynamical downscaling, refined land-surface representation, and the hybrid observational–analysis nudging framework. Moreover, the spatial wind field patterns shown in Figure 5 and the multi-level analyses of the representative convective case in Figure 10 further demonstrate that YRD1km captures meteorologically important mesoscale and convective features more effectively than ERA5.

Regarding the nudging configuration, we clarify that observational nudging was applied not only to wind, but also to standard near-surface thermodynamic variables, including 2-m temperature, relative humidity, and surface pressure, following conventional FDDA practice. Precipitation was not nudged, as it is not a prognostic model variable and cannot be directly constrained through nudging. While the present manuscript focuses on wind fields in line with the primary objective of constructing a high-resolution 3D wind dataset, the inclusion of thermodynamic variables in the nudging process helps maintain physical consistency within the boundary layer and does not compromise other fields. A more comprehensive evaluation of temperature and humidity fields is planned for future work.

Overall, through the addition of the independent validation experiment and clarification of the nudging strategy, the revised manuscript more clearly demonstrates that the skill improvements in YRD1km arise from physically meaningful enhancements rather than from localized nudging artifacts alone.

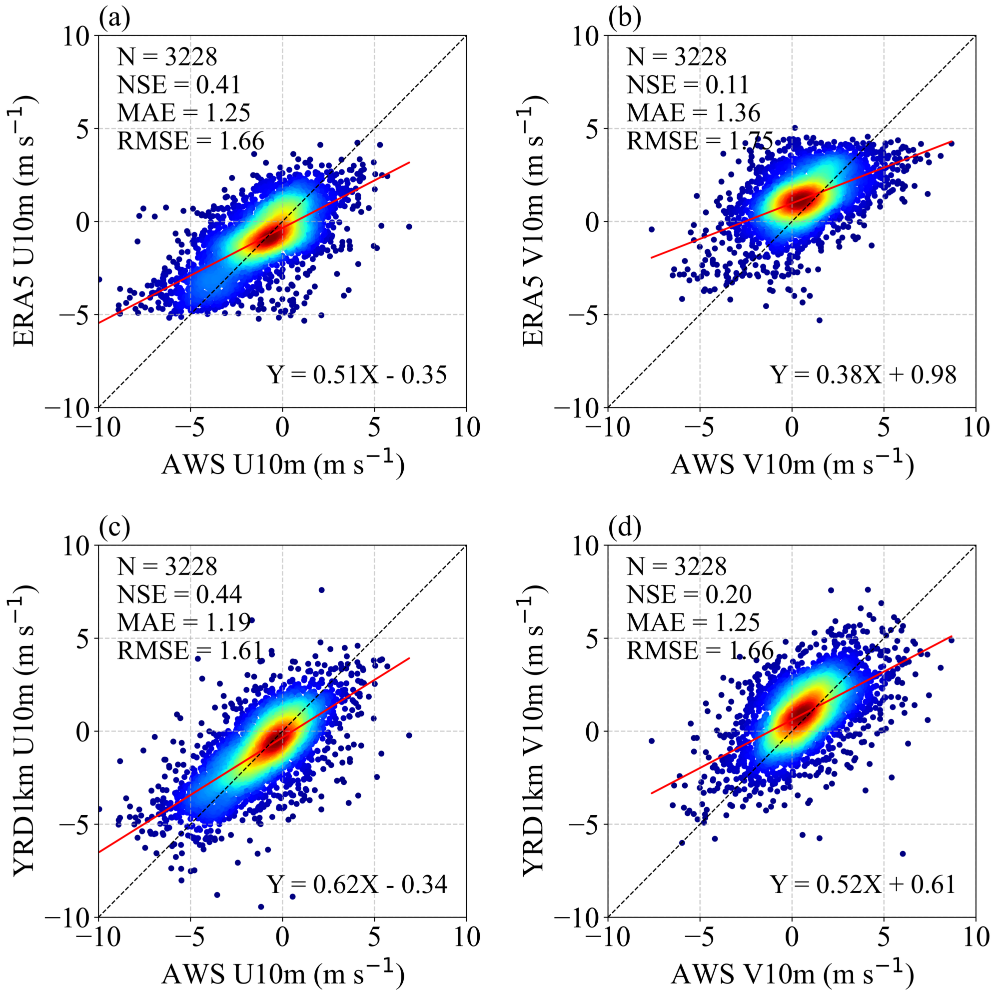

Figure 4. Independent validation scatterplots of 10-m wind components over the YRD region: (a) ERA5 U10m, (b) ERA5 V10m, (c) YRD1km U10m, and (d) YRD1km V10m.

- Are there not other studies/datasets with high-resolution downscaling over the YRD, that could be compared against? Purely from an interpolation/scale standpoint, of course the YRD1km dataset would be better able to resolve the winds than ERA5, especially if ERA5 grid points were interpolated to the same observation points for verification (i.e. multiple obs stations within the same ERA5 grid box would have the same ERA5 wind interpolated to them, but would have different WRF winds from different WRF grid points).

This ties into the utility of the dataset and the claims of such utility within the article. Statements like "The scale-dependent improvements emphasize the application value of YRD1km for both short-term weather monitoring and long-term climate analyses in the YRD region" are not backed up within the article. There just isn't enough verification conducted, over long enough time periods

+ seasons, against other competitive datasets, to make such claims. In particular, YRD1km isn't useful for "short-term weather monitoring" because it is dynamically downscaled from a reanalysis dataset (i.e. not useful as a forecast or nowcast dataset, only hindcasting), nor is it long enough yet for long-term climate analyses. I would recommend the authors be more specific about what the YRD1km can be used for, e.g. as a future training dataset for AI-based downscaling (with a longer dataset), or as input into high-resolution air-quality dispersion modelling or even sub-kilometer scale downscaling (down to LES scales), for specific applications (these were stated in the conclusion, but could be made more explicit and specific in other parts of the article).

Response: Thank you for this important and constructive comment regarding the role of spatial interpolation in the reported skill improvements and the scope of the dataset's stated applications.

At present, there is a lack of publicly available, kilometer-scale wind field datasets over the YRD. This data gap was a primary motivation for conducting this study. We agree that, from a purely spatial interpolation perspective, a higher-resolution product will generally yield lower errors when evaluated against dense surface observations, particularly when multiple stations fall within a single coarse-resolution ERA5 grid cell. However, the improvements demonstrated by YRD1km cannot be attributed solely to spatial interpolation of ERA5 fields. Unlike single-variable interpolation, YRD1km is produced through a fully dynamical framework that jointly constrains multiple atmospheric variables under physical laws, incorporates high-resolution terrain and land-surface information, and resolves mesoscale and vertical structures that cannot be recovered through interpolation alone. This distinction is particularly evident in the three-dimensional wind field analyses presented in the manuscript. For example, the vertical wind structures associated with the representative convective cases show clear convergence bands and wind speed enhancement zones that are absent in ERA5 and cannot be reproduced through spatial interpolation of reanalysis data. These features reflect physically coherent mesoscale dynamics rather than pointwise interpolation effects.

In particular, we acknowledge that YRD1km, being dynamically downscaled from a reanalysis product, is not intended for real-time forecasting or nowcasting and should therefore not be described as a dataset for "short-term weather monitoring." We have revised the manuscript to remove or rephrase these statements and to more accurately reflect the hindcast nature and temporal limitations of the dataset.

Following your suggestion, we have revised the manuscript to make the application scope of the YRD1km dataset more explicit, precise, and aligned with the presented validation. The revised text now emphasizes its current suitability for high-resolution diagnostic analyses of three-dimensional wind structures, event-based investigations of high-impact weather, and detailed characterization of urban and complex-terrain wind fields. In addition, YRD1km is framed as a physically consistent background dataset for applications that require fine-scale wind information, such as air-quality dispersion modeling and wind energy resource assessment. Potential uses in data-driven or AI-based downscaling studies are now clearly presented as future applications, contingent on further temporal extension of the dataset.

These revisions ensure that the claimed utility of YRD1km is fully supported by the presented validation and aligns with the dataset's design, coverage, and intended use.

Minor Comments:

- Should the citation/URL of the dataset be listed directly in the abstract? (not sure on this, please double check)

Response: Thank you for this comment. We have carefully checked the ESSD author guidelines and confirmed that including the dataset citation and DOI directly in the abstract is appropriate and consistent with ESSD formatting requirements.

- Statements like "high spatiotemporal resolution is fundamental to modern meteorological services, wind energy development, and the safe operation of low-altitude economy..." should be backed with citations, even if seemingly self-evident

Response: Thank you for the suggestion. Relevant citations have now been added to support the statement.

- MODIS2001 ---> did you mean MODIS2010?

Response: Thank you for raising this point. The land-use dataset used in the WRF configuration is indeed the default MODIS 2001 land-use classification provided within the WRF modeling system.

- Inconsistencies with abbreviations, e.g. LU-ESA2020 in the text but LUT-ESA2020 in the table

Response: Thank you for noting this inconsistency. The abbreviations have now been standardized as LU-EWC2020 throughout the manuscript.

- "Figures 5 presents" --> line 360

Response: Thank you for pointing this out. The grammatical error has been corrected to “Figure 5 presents”.