Responses to Review Comments

Dear Editor,

We thank you and both reviewers very much for their careful review and valuable comments. We have addressed each of their review comments carefully and revised our manuscript accordingly. Below please find our responses to all comments. Please note that the reviewer's remarks are in black, our responses are highlighted in blue, and extracts from the manuscript are in red, with **new texts** that have been added/edited marked in bold. We hope that you and both reviewers find our responses and revised manuscript satisfactory. Thank you very much.

Kind regards,

Ming Luo, on behalf of all co-authors

Reviewer #1

This manuscript presents a daily 1km dataset of six near-surface atmospheric moisture indicators over China for 2003–2020 by integrating a variety of data sources, such as meteorological stations, remote sensing, climate reanalysis, and population data. The topic is highly relevant to the journal, and the data itself will benefit broad research applications. The manuscript is clearly organized and well written. I have some comments/suggestions for improving the quality of this work.

***Response*:** We gratefully appreciate for your positive comment and valuable suggestions. We have accordingly revised our manuscript and prepared a list of point-to-point responses for you further assessment.

Major comments.

Model validation data. The authors split daily samples randomly into 80% training and 20% validation subsets. Because daily observations from the same station, nearby stations, and adjacent dates are strongly autocorrelated, a sample-level random split is likely to place observations from the same stations and closely related dates in both subsets. Consequently, the reported R^2, MAE, and RMSE values may characterize interpolation among familiar stations and dates rather than performance at other unseen locations. Have the authors tried adding station-held-out or spatially blocked validation, in which all dates from a few selected stations are excluded from training? Additionally, the authors should explain how errors and systematic biases in LST, ERA5 and other inputs may influence the model.

Response:

Thank you for this comment. To address the potential dependence between training and validation samples, we have added both spatial and temporal hold-out validations. For the spatial hold-out validation, 20% of the stations were excluded from model training, and all observations from these stations were used only for validation; for the temporal hold-out validation, the model was trained using data from 2003 to 2018 and evaluated using data from 2019 to 2020 (Lines 358–366): "**We perform spatial hold-**

out validation by holding out 20% of stations from training, and temporal hold-out validation by training the model on data from 2003 to 2018 and validating on 2019 to 2020. The model maintains strong performance under both strategies, with modest decreases in R^2 compared to the random split (e.g., AVP R^2=0.989, 0.976, and 0.977 under random, spatial hold-out, and temporal hold-out, respectively; Tables S1). ME, MAE, and RMSE exhibit comparable performance across the three validation strategies (Tables S1). These results confirm that our model extrapolates well across both space and time, demonstrating the robustness of the reported accuracy by ruling out validation leakage."

Additionally, we have expanded the discussion in Section 4.5 to clarify how input uncertainties may influence the model (Lines 660–673): "For instance, ERA5-Land 2-m SAT and DPT are subject to biases inherited from precipitation forcing and underestimated uncertainties from the land surface model (Muñoz-Sabater et al., 2021)**, which may introduce biases into the background fields**. These uncertainties may propagate through the machine-learning framework and affect the final estimates of atmospheric moisture indicators. Similarly, **LST is derived from clear-sky observations and may be less representative under cloudy conditions, potentially affecting the characterization of surface energy balance and subsequent moisture estimation (Zhang et al., 2022). Vapor pressure is available at a relatively coarse resolution and is resampled to 1 km, which may smooth fine-scale spatial variability and limit the model's capacity to represent local moisture gradients. As no explicit elevation correction is applied between station and grid level elevations, residual errors may remain in mountainous regions where elevation differences are substantial. Consequently,** uncertainties in remote sensing products, topographic data, and ancillary covariates may affect local predictions."

Reference

Zhang, T., Zhou, Y., Zhu, Z., Li, X., and Asrar, G. R.: A global seamless 1 km resolution daily land surface temperature dataset (2003–2020), Earth System Science Data, 14, 651–664, https://doi.org/10.5194/essd-14-651-2022, 2022.

Table S1. Comparison of R^2, ME, MAE, and RMSE values for six moisture indicators under different validation strategies.

Metrics	Split strategy	RH (%)	AVP (hPa)	VPD (hPa)	DPT (°C)	MR (g kg^{-1})	SH (g kg^{-1})
R^2	Random	0.877	0.989	0.906	0.985	0.988	0.988
	Spatial hold-out	0.825	0.976	0.840	0.978	0.974	0.975
	Temporal hold-out	0.824	0.977	0.854	0.974	0.975	0.975
ME	Random	0.029	0.002	0.010	0.023	0.003	0.005
	Spatial hold-out	0.097	0.017	0.006	0.044	0.009	0.010
	Temporal hold-out	-0.262	0.039	0.002	0.028	0.032	0.022
MAE	Random	4.849	0.677	0.983	1.120	0.459	0.451
	Spatial hold-out	5.956	0.936	1.347	1.317	0.639	0.626
	Temporal hold-out	6.201	1.009	1.335	1.454	0.685	0.671
RMSE	Random	6.520	0.933	1.411	1.608	0.633	0.620
	Spatial hold-out	7.784	1.363	1.952	1.884	0.922	0.902
	Temporal hold-out	8.060	1.378	1.896	1.991	0.933	0.912

Baseline comparison. I suggest the authors do some baseline comparison against the input datasets. For example, compare native ERA5-Land values with ground stations; native ERA5 with resampled ERA5 values; resampled ERA5 with ground stations. These will help examine whether the main feature-importance results that show ERA-derived moisture variables are the dominant predictors are partly biased by the close mathematical and statistical relationships between these highly correlated variables.

Response:

Thank you for your suggestion. We have added a new subsection "4.2 Baseline comparison of input datasets against station observations" in the Discussion of revised manuscript (Lines 523–534):

"4.2 Baseline comparison of ERA5-Land against station observations

To assess the baseline performance of ERA5-Land and the effect of spatial resampling, we compare native ERA5-Land, 1-km resampled ERA5-Land, and station observations (Fig. 8). Native and resampled ERA5-Land products show nearly identical performance against stations (R^2=0.648 and 0.651, MAE=8.225% and 8.186%, RMSE=10.906% and 10.852%, respectively), with strong agreement between them (R^2=0.999, Fig. 8c, f), confirming that artificial distortion is not introduced in the resampling process. Some biases are found for both datasets, with relatively large positive biases over the Sichuan Basin and Tibetan Plateau and negative biases over parts of Xinjiang (Fig. 8a, b). These results show that ERA5-Land-derived variables carry substantial physical information about moisture variability, and their high feature importance is not an artifact of inter-variable correlations."

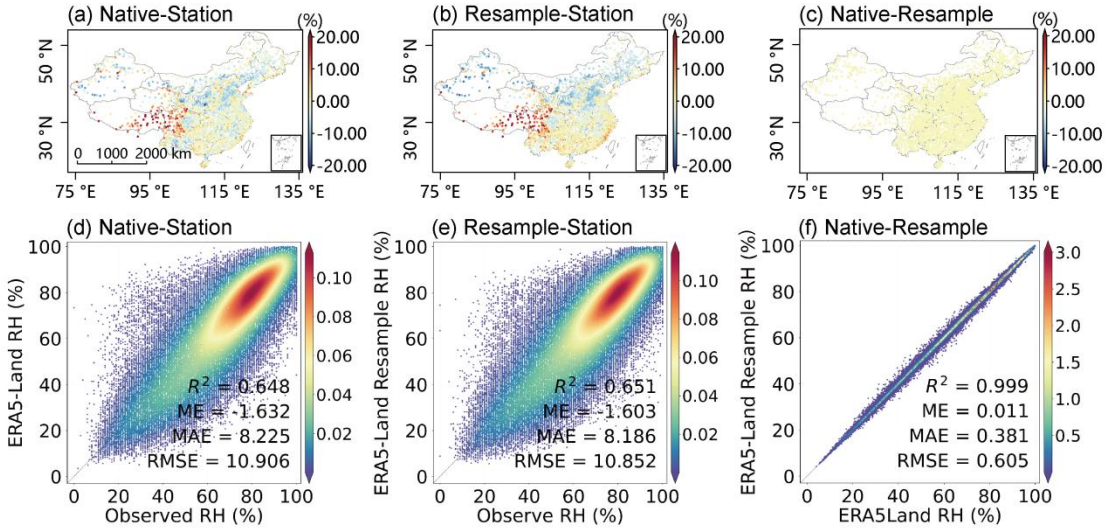

Figure 8. Comparison of native and resampled ERA5-Land against station observations for RH over China during 2003–2020: (a) Spatial distribution of ME at individual stations for native ERA5-Land, (b) Same as (a) but for resampled ERA5-Land, (c) Spatial distribution of ME at individual stations for native and resampled ERA5-Land comparison, (d) Scatter plots of predicted and observed RH for native ERA5-Land, (e) Same as (d) but for resampled ERA5-Land, (f) Scatter plots of predicted and observed RH for native and resampled ERA5-Land comparison.

More details are needed regarding harmonization of heterogeneous input datasets for

reproducibility. For example, what are the exact spatial resampling methods for each variable; how hourly data were converted to daily values; how missing MODIS observations were handled; how station and grid-cell elevation differences were treated.

Response:

Thank you for your suggestion. We have expanded the methods section to provide more detailed preprocessing procedures for each dataset.

The conversion from hourly to daily station observations has been clarified in Lines 194–196: "The station dataset includes daily mean near-surface air temperature (SAT), RH, and surface pressure (PRS). **Daily values are calculated as the arithmetic mean of all available hourly observations for each station.**"

The treatment of elevation differences between station and grid cells has been clarified in Lines 238–242: "Elevation data are thus fetched from the Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT DEM; Yamazaki et al. (2017)), from which slope was subsequently derived. **We include MERIT DEM directly as a covariate in the model to account for topographic influences, rather than applying explicit elevation correction between station and grid cells.**"

The harmonization of covariate datasets has been detailed in lines 251–261: "They are then harmonized to a consistent spatial extent, coordinate reference system, and spatial resolution using mosaicking, reprojection, resampling, clipping, and aggregation. **Specifically, the MODIS products (i.e., LST, land cover, and column water vapor) are first mosaicked from individual tiles to generate seamless coverages. Missing values in MODIS data are retained as fill values (i.e., -9999) and are not used for interpolation. All covariates are then reprojected to the WGS84 coordinate system and resampled to 1 km. Land cover is resampled using the majority rule to preserve class integrity (Kim et al., 2024). Elevation, slope, and population density are aggregated to 1 km using the mean value. Vapor pressure, 2-m dewpoint temperature, and 2-m temperature are resampled to 1 km using bilinear interpolation. All covariates are then clipped to the study domain.**"

Reference

Kim, D.-H., Johnson, J. M., Clarke, K. C., and McMillan, H. K.: Untangling the impacts of land cover representation and resampling in distributed hydrological model predictions, Environmental Modelling & Software, 172, https://doi.org/10.1016/j.envsoft.2023.105893, 2024.

Minor comments.

The authors should report quantitative comparisons in the text. The authors state in several places that "closest agreement", "larger deviations", etc. The exact values (especially for MAE and RMSE) and significance tests are needed when comparing different products.

Response: Thank you for your suggestion. We have added quantitative comparisons in this revised version:

Lines 570–574: "CMFD and CDMet show larger deviations from the station measurements **(CMFD MAE/RMSE: 6.779%/9.326%; CDMet: 5.750%/7.450%)**, with CMFD underestimating RH across most regions. In contrast, HiMIC-Daily shows closer agreement with observations **(MAE=3.980%, RMSE=5.311%)**".

Lines 589–590: "HiMIC-Daily shows the closest agreement with the station observations **(HiMIC-Daily MAE/RMSE: 4.640%/6.194%; CMFD: 7.194%/9.712%; CDMet: 7.712%/11.604%)**"

Lines 591–597: "CMFD captures the seasonal evolution of RH but shows larger deviations in the magnitude **(spring MAE/RMSE: 7.408%/9.872%; summer: 6.759%/9.066%; autumn: 7.220%/9.790%; winter: 7.202%/9.804%)**, whereas CDMet exhibits clear seasonal biases, with overestimation in spring **(MAE=10.172%, RMSE=15.378%)** and underestimation in summer **(MAE=7.440%, RMSE=11.232%)** and autumn **(MAE=6.825%, RMSE=9.690%)**."

Lines 599–603: "HiMIC-Daily maintains the best agreement with observed RH **(MAE=4.094%, RMSE=5.493%)**, while CMFD closely follows the observed temporal variations but with noticeable biases **(MAE=6.750%, RMSE=8.704%)**.

CDMet shows larger seasonal deviations again **(spring MAE/RMSE: 9.196%/14.868%; summer: 5.972%/9.017%; autumn: 6.948%/9.764%; winter: 5.251%/7.139%)**."

The manuscript should describe HiMIC-Daily more explicitly as a machine learning downscaling/bias-correction product, to avoid claiming that 1-km information is independently observed.

Response: Per your suggestion, we have revised the statement as a machine learning downscaling product.

Lines 24–26: "Here, we present HiMIC-Daily, a seamless daily 1-km-resolution **dataset for** near-surface atmospheric moisture **in** China, 2003–2020 based on **machine learning downscaling**."

Lines 141–142: "To address this gap, we develop HiMIC-Daily—a daily 1 km **downscaling dataset for** multi-indicator atmospheric moisture **in** China covering 2003–2020."

Lines 686–690: "The high-resolution (daily and 1 km) multi-indicator atmospheric moisture dataset over China during 2003–2020 (HiMIC-Daily) is **a machine learning downscaling product** publicly available through the National Tibetan Plateau Data Center (TPDC) of China at https://doi.org/10.11888/Atmos.tpdc.303449 (last access: May 2026; **Su et al. (2026b)**)."

Reference:

Su, Z., Zhang, H., Wu, S., and Luo, M.: HiMIC-Daily: A high-resolution (daily and 1 km) atmospheric moisture collection over China during 2003 – 2020 (V1.0) [dataset], https://doi.org/10.11888/Atmos.tpdc.303449, 2026.

The full dataset is not shared on Zenodo. Now it only includes some sample data.

Response: Thank you. We have now updated the Zenodo deposit with the full dataset (version V1.1). The data are accessible via the same DOI

https://doi.org/10.5281/zenodo.20124067.

L21: remove “extreme”. It is repetitive with “extremes”

Response: Removed.

L39-40: claims of higher accuracy and more realistic temporal variability should be supported by quantitative comparisons.

Response: Thank you for your suggestion. We have revised the manuscript to include the quantitative comparisons (Lines 40–44): “Compared with two existing products **(e.g., CMFD and CDMet)**, HiMIC-Daily provides finer spatial detail, higher accuracy, and more realistic temporal variability across different climatic regions **(HiMIC-Daily MAE/RMSE for RH in 2013: 4.640%/6.194%; CMFD: 7.194%/9.712%; CDMet: 7.712%/11.604%)**.”

L64: “has” to “have”

Response: Amended.

L111: “of assessing” to “for assessing”

Response: Amended.

L180: “daily estimates” to “estimates”

Response: Amended.

Table 2 and L232: version of MODIS doesn’t seem consistent.

Response:

Thank you for pointing out this. We confirm that the land cover data used in our study are derived from the MCD12Q1.006 product, and we have revised the version of MODIS (Table 2): "https://doi.org/10.5067/MODIS/MCD12Q1.**006**"

Table 4: is cited several times in the text but absent.

Response:

We have removed the original Table 4 as we have integrated its statistical metrics (R^2, ME, MAE, RMSE) directly into Figure 3. The original citations to Table 4 have been replaced with references to Figure 3.

Lines 342–343: "All six indicators achieve high accuracy, with R^2 values exceeding 0.877 (**Fig. 3**)."

Lines 350–355: "MAE and RMSE for each indicator are within acceptable ranges, with RH recording an MAE of 4.849% and an RMSE of 6.520%, followed by AVP with an MAE of 0.677 hPa and an RMSE of 0.933 hPa, VPD with an MAE of 0.983 hPa and an RMSE of 1.411 hPa, DPT with an MAE of 1.120°C and an RMSE of 1.608°C, MR with an MAE of 0.459 g kg^{-1} and an RMSE of 0.633 g kg^{-1}, and SH with an MAE of 0.451 g kg^{-1} and an RMSE of 0.620 g kg^{-1} (**Fig. 3**)."

Figure 5 caption: "Deeper red indicates larger values, reflecting better model performance" is true only for R^2. Larger MAE and RMSE indicate worse performance.

Response:

Thank you for pointing this out. We have checked Figure 5 and confirm that the colorbar for MAE and RMSE was reversed relative to that for R^2. Accordingly, we have revised the Figure 5 caption (Lines 458–460): "Deeper red indicates **better model performance for R^2, MAE, and RMSE, corresponding to higher R^2 values but lower MAE and RMSE values. For ME, deeper red and deeper blue indicate larger positive and negative biases.**"

Reviewer #2

The authors of the manuscript titled "HiMIC-Daily: A high-resolution (daily and 1 km) multi-indicator atmospheric moisture collection over China, 2003–2020" present a machine learning approach which utilises the LightGBM model to develop a 1 km daily moisture dataset for China for the years spanning 2003–2020, using meteorological stations, reanalysis data, remote sensing and population density. Additionally, the dataset presented here is useful for the community and fits well with the journal's scope. I have however a few concerns/comments before publishing the paper:

Response: Thank you for your positive comments on our work. We greatly appreciate your recognition of the importance of our high-resolution atmospheric moisture dataset and your thoughtful suggestions for improving our work. Following your suggestion, we have addressed all comments point by point to further improve the clarity and quality of our manuscript.

The "Data and Methods" section needs to be expanded to include all the methodological details of the data pipeline. The authors state that "all gridded datasets are unified at a 1 km spatial resolution and consistent coordinate reference system" and later that "They are then harmonized to a consistent spatial extend, coordinate reference system (WGS84), and spatial resolution (1 km) using mosaicking, reprojection, resampling, clipping, and aggregation". More details for each of the presented datasets is needed in the methods section and which method was applied to each. In addition to this, in lines 292-293, which hyperparameters of the LightGBM model were tuned in the grid search procedure and was the grid search exhaustive or random sampling?

Response:

Thank you for your suggestion. In this revision, we have expanded the Data and Methods sectionto provide more detailed preprocessing information for each covariate dataset, including mosaicking, reprojection, resampling strategies for each variable type.

The harmonization of covariate datasets has been detailed in Lines 251–261: "They are then harmonized to a consistent spatial extent, coordinate reference system, and spatial resolution using mosaicking, reprojection, resampling, clipping, and aggregation. **Specifically, the MODIS products (i.e., LST, land cover, and column water vapor) are first mosaicked from individual tiles to generate seamless coverages. Missing values in MODIS data are retained as fill values (i.e., -9999) and are not used for interpolation. All covariates are then reprojected to the WGS84 coordinate system and resampled to 1 km. Land cover is resampled using the majority rule to preserve class integrity (Kim et al., 2024). Elevation, slope, and population density are aggregated to 1 km using the mean value. Vapor pressure, 2-m dewpoint temperature, and 2-m temperature are resampled to 1 km using bilinear interpolation. All covariates are then clipped to the study domain.**"

In addition, we have clarified the hyperparameter tuning procedure in Lines 307–313: "Hyperparameter tuning is performed using **an exhaustive** grid search combined with 5-fold cross-validation within the training subset, and the final hyperparameter combination is selected by minimizing RMSE (Sarker, 2021). **The tuned hyperparameters include the number of estimators, maximum tree depth, maximum number of leaves, minimum number of samples per leaf, minimum sum of instance weights in a leaf, bagging fraction, feature fraction, and learning rate (Ke et al., 2017).**"

Reference

Kim, D.-H., Johnson, J. M., Clarke, K. C., and McMillan, H. K.: Untangling the impacts of land cover representation and resampling in distributed hydrological model predictions, Environmental Modelling & Software, 172, https://doi.org/10.1016/j.envsoft.2023.105893, 2024.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: Lightgbm: A highly efficient gradient boosting decision tree, Advances in Neural Information Processing Systems, **https://api.semanticscholar.org/CorpusID:3815895 (last access: 20 March 2026)**, 2017.

The split into training and validation sets was implemented by following a random 80/20 split of the pooled data from all the stations. While this is also informative, it's not adequate in this context as there is significant validation leakage risk that can artificially inflate the performance metrics. Nearby days from the same station or data from nearby stations for the same day can likely appear in both sets. A temporal (split by months or even years) and spatial (leave a set of stations out) cross-validation should be added to the evaluation procedures.

Response:

Thank you for this comment. To address the potential dependence between training and validation samples, we have added both spatial and temporal hold-out validations. For the spatial hold-out validation, 20% of the stations were excluded from model training, and all observations from these stations were used only for validation; for the temporal hold-out validation, the model was trained using data from 2003 to 2018 and evaluated using data from 2019 to 2020 (Lines 358–366): "**We perform spatial hold-out validation by holding out 20% of stations from training, and temporal hold-out validation by training the model on data from 2003 to 2018 and validating on 2019 to 2020. The model maintains strong performance under both strategies, with modest decreases in R^2 compared to the random split (e.g., AVP R^2=0.989, 0.976, and 0.977 under random, spatial hold-out, and temporal hold-out, respectively; Tables S1). ME, MAE, and RMSE exhibit comparable performance across the three validation strategies (Tables S1). These results confirm that our model extrapolates well across both space and time, demonstrating the robustness of the reported accuracy by ruling out validation leakage.**"

The authors made the choice to train a separate model for each year to account for possible year-to-year variations in the relationships between atmosphere moisture and its predictors. This has the potential to introduce artificial year-to-year discontinuities and complicate trend analysis. Have the authors made any comparison with a single model for all the years and this approach to validate this? The possibility of month-

to-month variation in this relationship to be larger than year-to-year also exists (due to seasonal variations), so another option would be to train 12 individual models for each month or 4 seasonal ones (pool the data from all the years for all Januaries for example).

Response:

Thank you very much for this suggestion. We have compared the annual modeling strategy with a single all-year model, 12 monthly models, and 4 seasonal models. The results show comparable performance across all strategies, indicating that the model accuracy is not strongly dependent on how the training data are grouped in time. For example, the R^2 values for AVP are 0.989, 0.971, 0.977, and 0.977 for the annual, single, monthly, and seasonal modeling strategies, respectively (Tables S2). We have revised the manuscript to clarify that this comparison evaluates the sensitivity of model performance to the temporal modeling strategy.

We have added this discussion to Section 3.1.2 in Lines 414–421: "**To evaluate the annual modeling strategy, we compare it with a single all-year model, 12 monthly models, and 4 seasonal models (Tables S2). The results show comparable performance across all four strategies (e.g., AVP R^2: 0.989 for annual, 0.971 for single, 0.977 for monthly, and 0.977 for seasonal). These results indicate that the overall predictive performance is insensitive to the temporal grouping of the training data. We retain the annual model strategy because it allows the model to accommodate interannual changes in response relationships and facilitates operational updates as new years of data become available.**"

Table S2. Comparison of R^2, ME, MAE, and RMSE values for six moisture indicators under different modeling strategies.

Metrics	Model strategy	RH (%)	AVP (hPa)	VPD (hPa)	DPT (°C)	MR (g kg^{-1})	SH (g kg^{-1})
R^2	Single model	0.797	0.971	0.823	0.968	0.971	0.969
	Monthly models	0.812	0.977	0.840	0.974	0.975	0.975

	Seasonal models	**0.822**	**0.977**	**0.845**	**0.974**	**0.976**	**0.975**
ME	**Single model**	**0.006**	**0.001**	**0.002**	**-0.001**	**-0.001**	**0.001**
	Monthly models	**0.055**	**0.013**	**-0.011**	**0.017**	**0.006**	**0.009**
	Seasonal models	**-0.003**	**0.002**	**0.001**	**-0.003**	**0.001**	**0.001**
MAE	**Single model**	**6.332**	**1.100**	**1.375**	**1.672**	**0.721**	**0.732**
	Monthly models	**5.982**	**0.983**	**1.316**	**1.477**	**0.659**	**0.651**
	Seasonal models	**5.846**	**0.986**	**1.294**	**1.483**	**0.652**	**0.652**
RMSE	**Single model**	**8.284**	**1.501**	**1.986**	**2.260**	**0.987**	**0.995**
	Monthly models	**7.940**	**1.353**	**1.917**	**2.055**	**0.910**	**0.897**
	Seasonal models	**7.706**	**1.349**	**1.883**	**2.033**	**0.896**	**0.891**

Line 303: The Mean Error should also be included in the evaluation metrics to assess for systematic bias.

Response:

Thanks for this suggestion, we have incorporated mean error into the evaluation metrics and updated the relevant sections throughout the manuscript. In abstract section, we have added Mean Error (ME) values to demonstrate that the model performs well across all indicators in Lines 36–40: "The strongest performance is obtained for AVP, DPT, MR, and SH, with R^2 values exceeding 0.985, **mean error values consistently near zero (e.g., 0.005 hPa for AVP)**, and error metrics remain within acceptable ranges for all indicators (e.g., mean absolute error **(MAE)** of 0.677 hPa and a root mean square error (RMSE) of 0.933 hPa for AVP)."

In "Overall accuracy" section, we have reported ME values for all six indicators in lines 347–350: "**ME are consistently close to zero across all indicators, demonstrating marginal systematic bias. Specifically, RH exhibits an ME of 0.036%, while AVP, VPD, DPT, MR, and SH record ME values of 0.005 hPa, 0.006 hPa, 0.020°C, 0.004 g kg^{-1}, and 0.006 g kg^{-1}, respectively.**"

In "Validation on yearly, monthly, and daily scales" section, we have included

mean error results across temporal scales:

Lines 381–384: "**The annual ME values are consistently close to zero across all indicators and years, ranging from approximately -0.180 to 0.420 for RH and within ±0.060 for the remaining indicators (Fig. 4d).** The annual MAE and RMSE values for all indicators remained relatively stable throughout the study period (Fig. 4**g**, **j**)".

Lines 394–396: "**Monthly ME values remain consistently near zero across all indicators, with RH shifting from positive in winter to negative in summer and autumn, while the remaining indicators exhibit no clear seasonal pattern.**"

Lines 406–409: "RH reaches its highest accuracy around day 30 and its lowest around days 26–27, **with corresponding larger ME around days 26–27 and near-zero ME around day 30**. VPD performs best in early-month to mid-month, **while the remaining indicators show daily ME values within ±0.050 hPa**."

In "Spatial variation in prediction accuracy" section, we have also added a comprehensive analysis of the spatial distribution of ME for all indicators in lines 438–444: "**For ME, all indicators show small biases at most stations, indicating limited systematic bias across most regions (Fig. 5g–l). AVP, DPT, MR, and SH exhibit similar spatial patterns, with negative biases predominantly in southern China and positive biases in northern China. In contrast, RH and VPD display more randomly distributed biases without clear regional patterns.** In terms of error metrics, MAE and RMSE exhibit consistent spatial distributions within each indicator (Fig. 5**m**–**x**)."

Accordingly, we have added ME to Figure3, Figure4 and Figure 5.

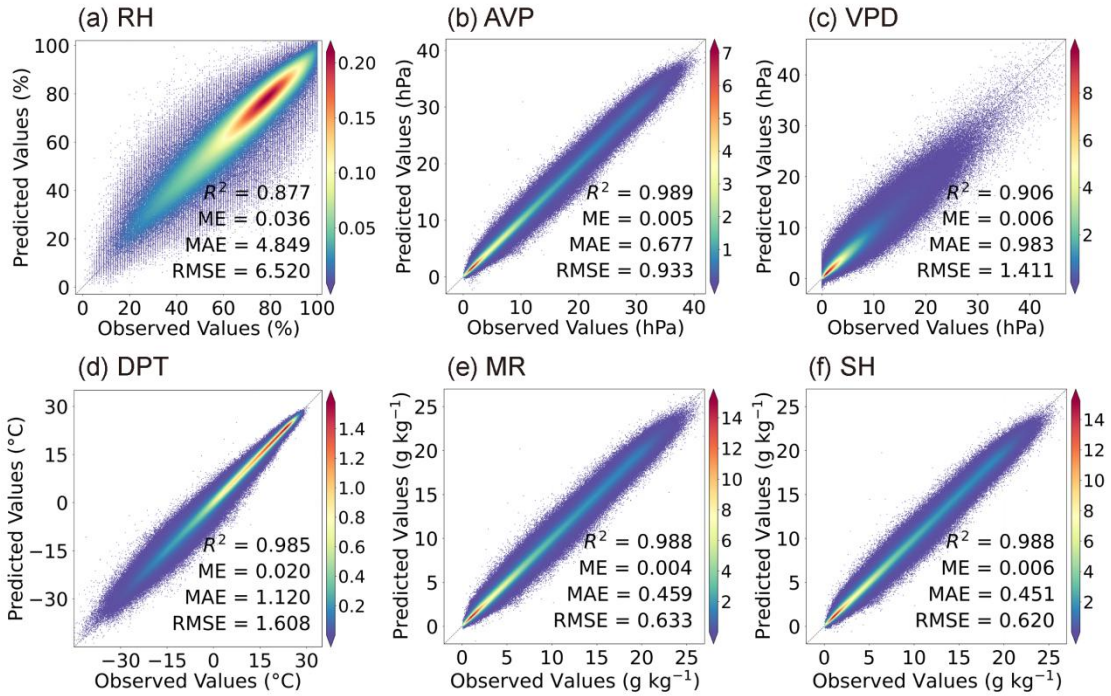

Figure 3. Scatter plots of predicted and observed atmospheric moisture indicators over China during 2003–2020: (a) RH, (b) AVP, (c) VPD, (d) DPT, (e) MR, and (f) SH. Colors indicate the density of data points, with red representing higher density and blue representing lower density. The black dashed line is the 1:1 line.

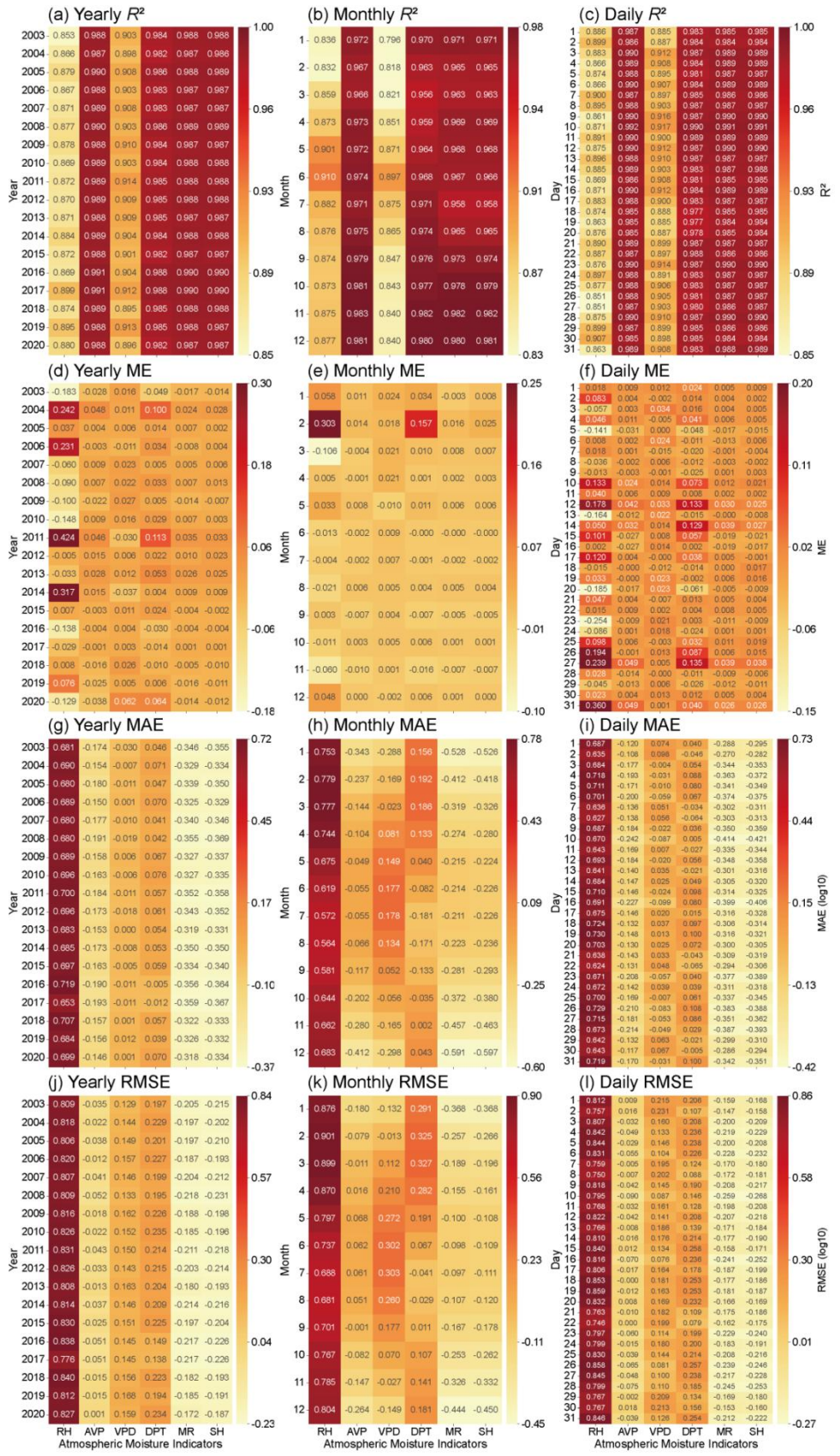

Figure 4. Annual, monthly, and daily R^2, **ME,** MAE, and RMSE of six atmospheric moisture indicators over China during 2003–2020. Colors represent values, with red for higher values and yellow for lower. MAE and RMSE values are shown on a logarithmic scale, i.e., log (10).

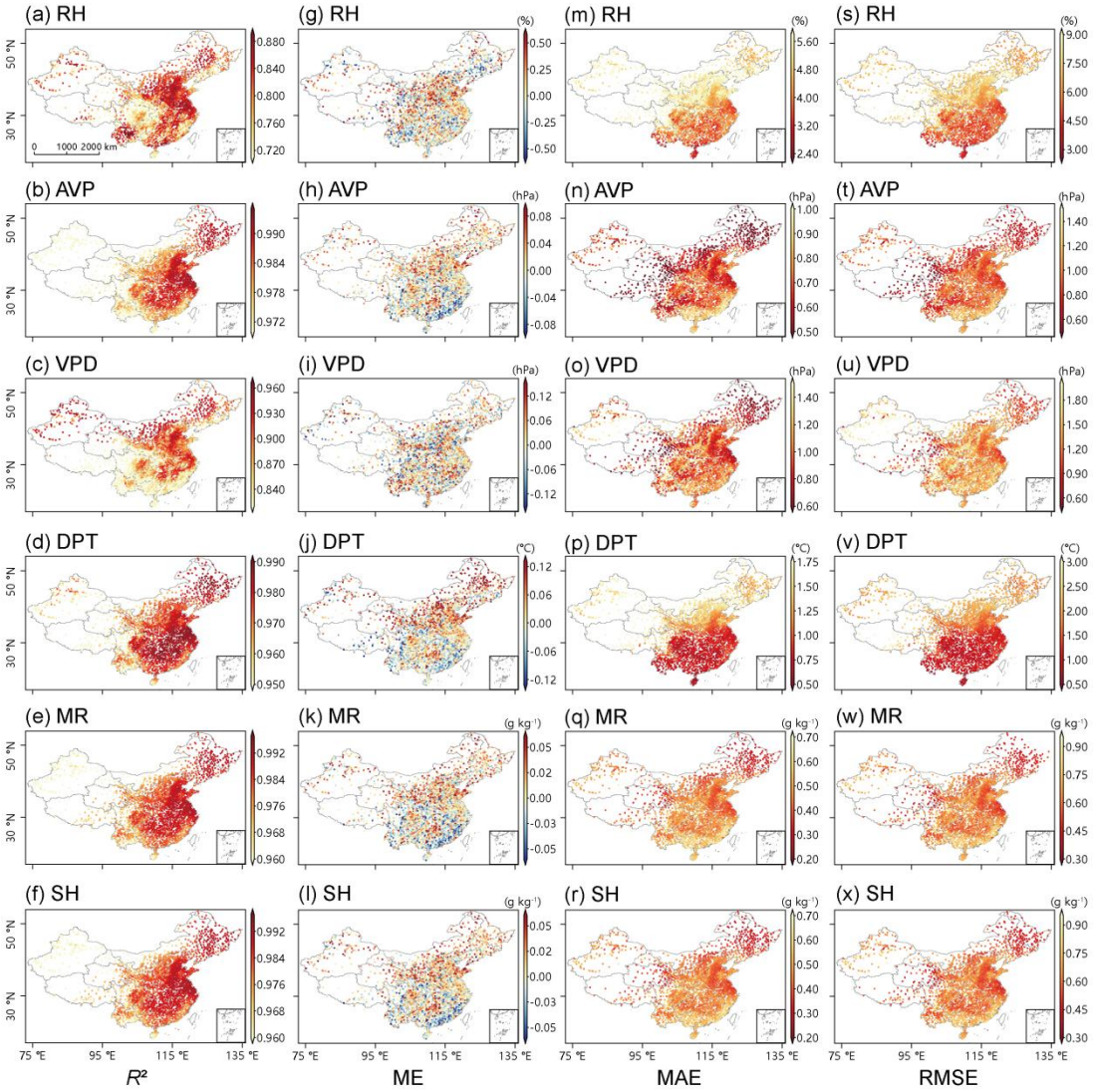

Figure 5. Spatial distribution of R^2, **ME,** MAE, and RMSE values for six atmospheric moisture indicators at meteorological stations over China during 2003–2020: Each row corresponds to one moisture indicator (from top to bottom: RH, AVP, VPD, DPT, MR, and SH), while **(a-f)**, **(g-l)**, **(m-r)**, and **(s-x)** correspond to R^2, **ME,** MAE, and RMSE, respectively. Deeper red indicates **better model performance for ,R^2, MAE, and RMSE, corresponding to higher R^2 values but lower MAE and RMSE values. For ME, deeper red and deeper blue indicate larger positive and negative biases.**

Figure 5: The colormap used makes comparisons very difficult. Consider changing (e.g. viridis or magma).

Response: We have updated the colormaps in Figure 5 (see above) to improve visual clarity. Specifically, we now use "YlOrRd" for R^2, "YlOrRd_r" for MAE and RMSE,

and “RdYlBu_r” for ME.

The ERA5-Land dataset in coastal areas has a lot of missing data over land (when interpolating to a 1x1 km grid). How did the authors treat this? Please clarify whether any coastal stations were excluded or interpolated and if prediction accuracy was checked separately for coastal vs inland stations.

Response:

Thanks for this question. ERA5-Land variables were resampled to 1 km using nearest-neighbor interpolation, and values at station locations were extracted directly. Stations with completely missing values were excluded (36 stations) (Lines 262–264): “**Station values are then extracted directly from the resampled grids, without additional spatial interpolation. Stations without ERA5-Land values were excluded (36 stations, Fig. S1a).**”

Coastal stations (174 stations) are defined as those within 30 km of the coastline (Fan et al., 2025; Chen et al., 2018). We perform separate accuracy assessments for coastal and inland stations (Lines 448–452): “**We further evaluate model performance separately for coastal and inland stations. Coastal stations (174 stations) are defined as those within 30 km of the coastline (Fan et al., 2025; Chen et al., 2018). The model performed consistently well for both coastal and inland areas (R^2>0.79), with only marginal differences in error metrics, demonstrating its robustness across geographic settings (Fig. S1b).**”

Reference:

Fan, T., Wang, D., Chan, P. W., Zeng, Z., Huang, L., and Chen, Y.: Comparison of diurnal variations of pre-summer extreme hourly rainfall between inland and coastal regions over the Greater Bay Area, Urban Climate, 61, https://doi.org/10.1016/j.uclim.2025.102427, 2025.

Chen, G., Lan, R., Zeng, W., Pan, H., and Li, W.: Diurnal Variations of Rainfall in Surface and Satellite Observations at the Monsoon Coast (South China),

Journal of Climate, 31, 1703–1724, https://doi.org/10.1175/jcli-d-17-0373.1, 2018.

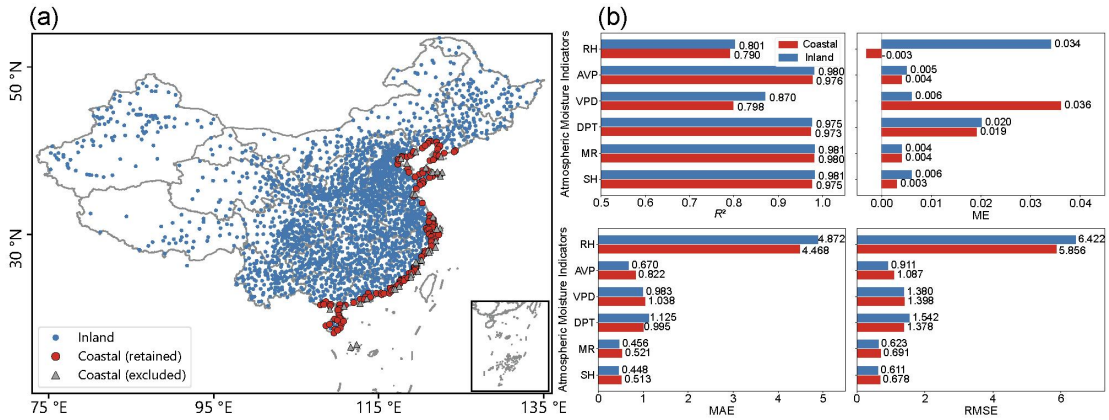

Figure S1. Comparison of prediction accuracy between coastal and inland stations: (a) Spatial distribution of meteorological stations. Inland, retained coastal, and excluded coastal stations are shown as blue circles, red circles, and gray triangles, respectively, (b) Performance comparison (R^2, ME, MAE, RMSE) between coastal and inland stations for six moisture indicators. Red and blue bars denote coastal and inland stations, respectively.

ERA5-Land-derived indicators dominate feature importance for all six target variables. This raises the question of whether HiMIC-Daily performs genuine sub-grid downscaling or is primarily a bias-corrected sharpening of the ERA5-Land field, with other covariates adding only secondary local adjustments. I suggest for a sample of ERA5-Land grid cells across contrasting terrain, compute the variance of the corresponding 1 km HiMIC-Daily predictions within each cell. Low within-cell variance relative to between-cell variance would indicate the product largely reproduces the coarse ERA5-Land field.

Response:

Per your suggestion, we have computed the within-grid variance of HiMIC-Daily predictions across contrasting terrain regions on July 28, 2013 (Lines 535–543): "**To further investigate whether HiMIC-Daily performs genuine downscaling, we compute the within-grid variance of 1-km predictions across contrasting terrains**

(Fig. S2). The mean within-grid standard deviation is 2.91% in the North China Plain (Song and Deng, 2015) and 3.90% in the mountainous Chongqing region (Liu et al., 2017) (Fig. S2c, d). The larger within-grid variability indicates that HiMIC-Daily captures sub-grid spatial details beyond the coarse ERA5-Land input. This is further supported by its improved accuracy over ERA5-Land on July 28, 2013 (HiMIC-Daily: R^2=0.849, RMSE=5.311%; ERA5-Land: R^2=0.578, RMSE=8.880%), and the additional spatial detail captured by HiMIC-Daily (Fig. S2a, b)."

Reference

Liu, Y., Yue, W., Fan, P., Zhang, Z., and Huang, J.: Assessing the urban environmental quality of mountainous cities: A case study in Chongqing, China, Ecological Indicators, 81, 132–145, https://doi.org/10.1016/j.ecolind.2017.05.048, 2017.

Song, W. and Deng, X.: Effects of Urbanization-Induced Cultivated Land Loss on Ecosystem Services in the North China Plain, Energies, 8, 5678–5693, https://doi.org/10.3390/en8065678, 2015.

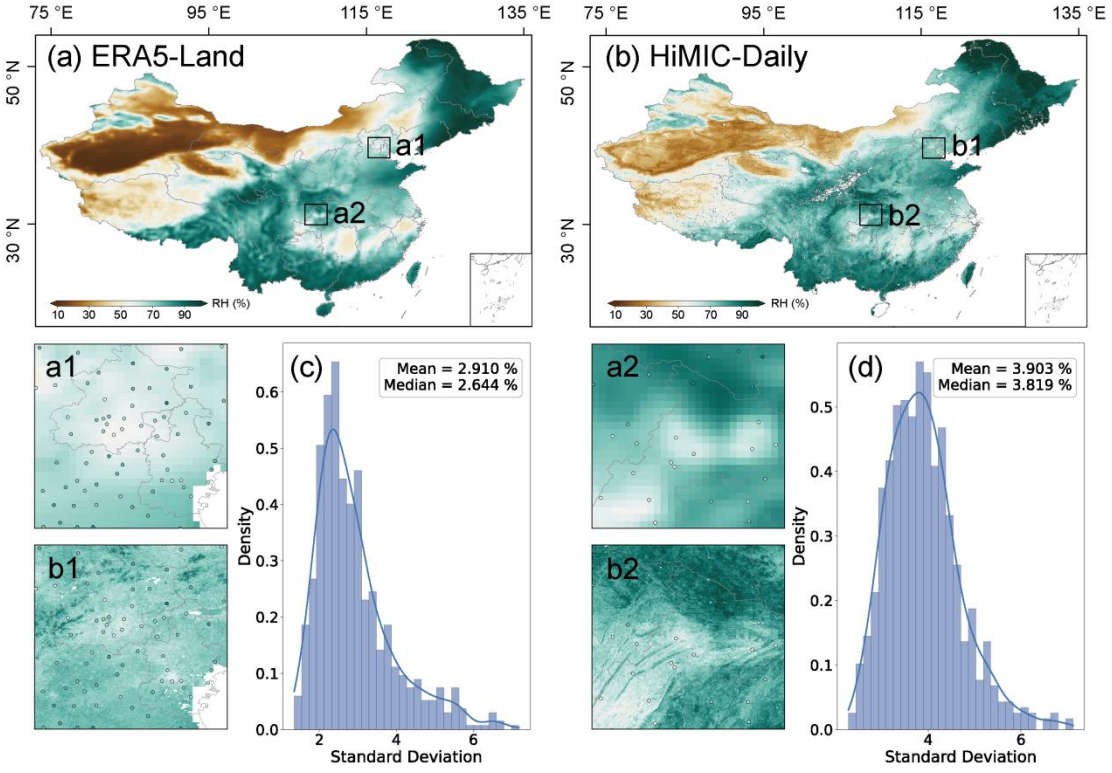

Figure S2. Comparison of spatial patterns between ERA5-Land and HiMIC-Daily for RH over the mainland of China and across contrasting terrain regions on July 28, 2013 (1: North

China Plain, 2: Chongqing mountainous region): (a) ERA5-Land and (b) HiMIC-Daily RH predictions over the mainland of China. (c) Within-grid standard deviation of HiMIC-Daily predictions across ERA5-Land grid cells in the North China Plain. (d) Same as (c) but for the Chongqing mountainous region.

Additionally, while the manuscript is well written and has a good structure and flow there are some minor mistakes that need to be fixed:

Line 21: Extreme is repeated

Response: Amended.

Line 481: fine-scale informations (should be information)

Response: Amended.

Table 4 could not be located in the manuscript as provided (mentioned twice in text).

Response:

Thank you for pointing this out. We have removed the original Table 4 as we have integrated its statistical metrics (R^2, ME, MAE, RMSE) directly into Figure 3. The original citations to Table 4 have been replaced with references to Figure 3.

Lines 342–343: "All six indicators achieve high accuracy, with R^2 values exceeding 0.877 (**Fig. 3**)."

Lines 350–355: "MAE and RMSE for each indicator are within acceptable ranges, with RH recording an MAE of 4.849% and an RMSE of 6.520%, followed by AVP with an MAE of 0.677 hPa and an RMSE of 0.933 hPa, VPD with an MAE of 0.983 hPa and an RMSE of 1.411 hPa, DPT with an MAE of 1.120°C and an RMSE of 1.608°C, MR with an MAE of 0.459 g kg^{-1} and an RMSE of 0.633 g kg^{-1}, and SH with an MAE of 0.451 g kg^{-1} and an RMSE of 0.620 g kg^{-1} (**Fig. 3**)."

We thank both reviewers again for their positive comments and helpful suggestions, which have improved the clarity and robustness of our work.