

Reply to the reviewer #2

We sincerely thank the reviewer for taking the time to evaluate our manuscript and for providing insightful and constructive comments. We have carefully considered all suggestions and revised the manuscript accordingly. Below, we provide point-by-point responses to the reviewer's comments (responses marked in blue and sections included in manuscript in violet). We fully agree with the reviewer's observations, and all recommended modifications have been incorporated into the revised version of the manuscript.

1. The study adopts LightGBM for downscaling but lacks comparison with other mainstream downscaling models (e.g., random forest, XGBoost, regression models), so the superiority of LightGBM cannot be fully verified.

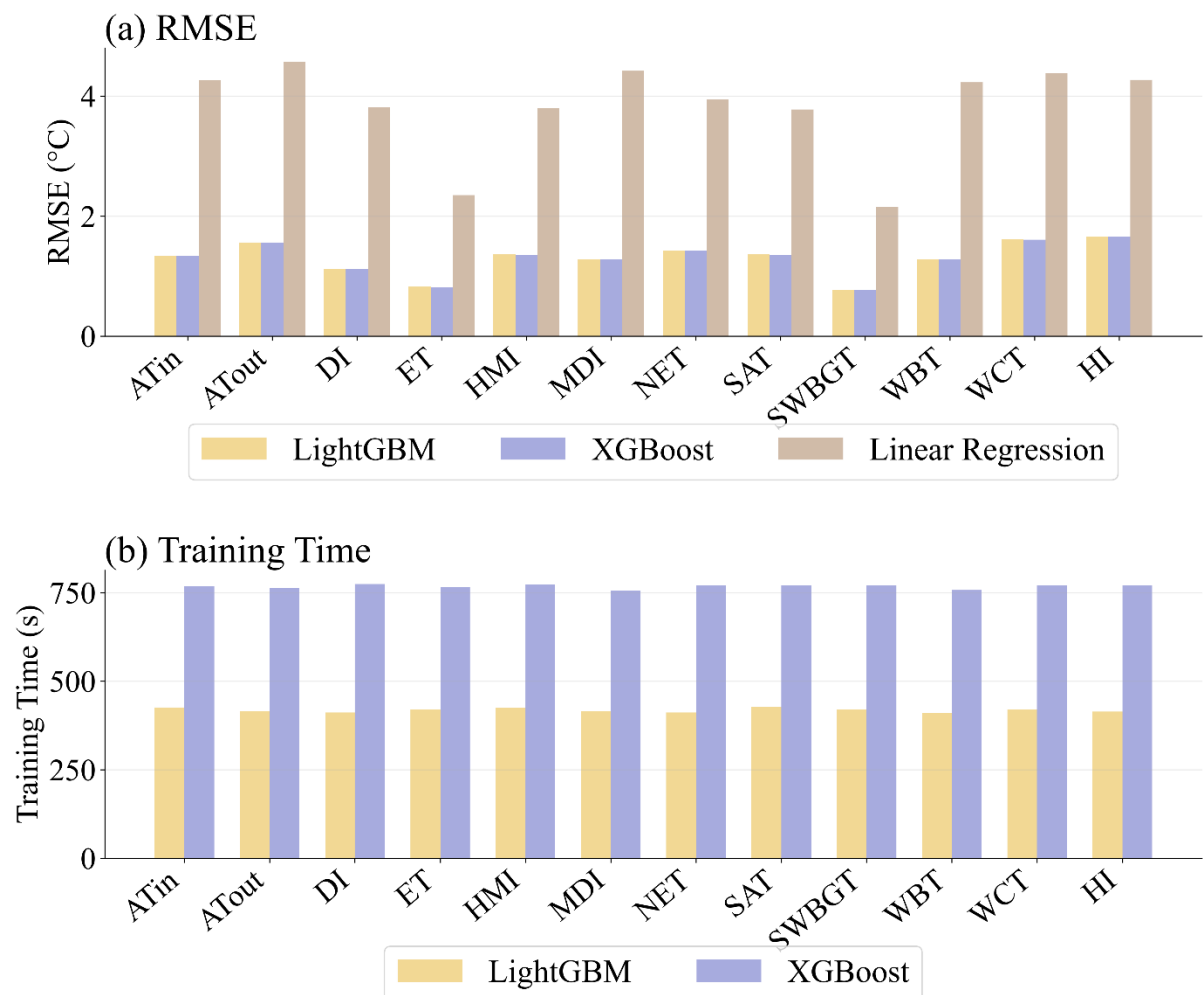
Response:

We thank the reviewer for this valuable suggestion. To justify the selection of LightGBM, we compared its performance with XGBoost and Linear Regression using the same training dataset and identical train–test split for all 12 HPT indices. Random Forest was also evaluated but was not computationally feasible at this data scale because of its substantially higher memory requirements and longer training time. LightGBM achieved predictive performance comparable to XGBoost while reducing training time by approximately 1.5–2 times, whereas Linear Regression showed considerably lower accuracy, highlighting the importance of nonlinear modelling for HPT estimation. We further note that, because LOYO cross-validation was performed for all 12 HPT indices, using XGBoost would have required substantially greater computational time with negligible improvement in predictive performance. These findings are consistent with previous studies (Ke et al., 2017; Candido et al., 2021; Su et al., 2021). Accordingly, we have incorporated this comparison and its interpretation into the Discussion section (lines 523-538) of the revised manuscript, with the corresponding results presented in Supplementary Figure 12.

Discussion:

“To assess the suitability of LightGBM for HiTIC-India, we compared it with XGBoost and Linear Regression using the same training dataset (23 million samples) and identical random train-test split for all 12 HPTs. Random Forest was also tested but was not feasible at this scale due to excessive memory requirements and longer training times. LightGBM and XGBoost achieved nearly identical accuracies (ATin test $R^2 = 0.9846$ for both; WCT test $R^2 = 0.9805$ and 0.9806 , supplementary figure 12), whereas Linear Regression showed substantially lower performance ($R^2 = 0.77$ – 0.87), indicating the importance

of nonlinear relationships in thermal stress estimation. LightGBM, however, reduced training time by approximately 1.5–2 times compared with XGBoost while maintaining comparable accuracy and minimal train–test differences (supplementary figure 12) for each HPT index. Furthermore, since LOYO cross-validation was conducted for all 12 HPT indices, XGBoost would require greater computation time than LightGBM to complete the analysis, despite offering negligible gains in predictive performance. This is consistent with previous comparative studies reporting that LightGBM's histogram-based split-finding and leaf-wise growth strategy substantially reduce computational cost relative to XGBoost's level-wise growth, without compromising accuracy (Candido et al., 2021; Ke et al., 2017; Su et al., 2021). Therefore, LightGBM was selected as the most suitable approach for generating daily 1 km thermal stress estimates across India.”



“**Supplementary Figure 12.** Comparison of LightGBM against XGBoost and Linear Regression across the 12 HPT indices. (a) Test-set RMSE (°C) for LightGBM, XGBoost, and Linear Regression.

LightGBM and XGBoost achieve nearly identical RMSE across all indices, both substantially outperforming Linear Regression. (b) Training time (s) for LightGBM and XGBoost. Despite comparable predictive accuracy, LightGBM trains roughly twice as fast as XGBoost across all 12 indices, reflecting its greater computational efficiency.”

2. Covariate selection is not discussed for multicollinearity among LST, PWV, population density and topographic factors, which may interfere with model stability.

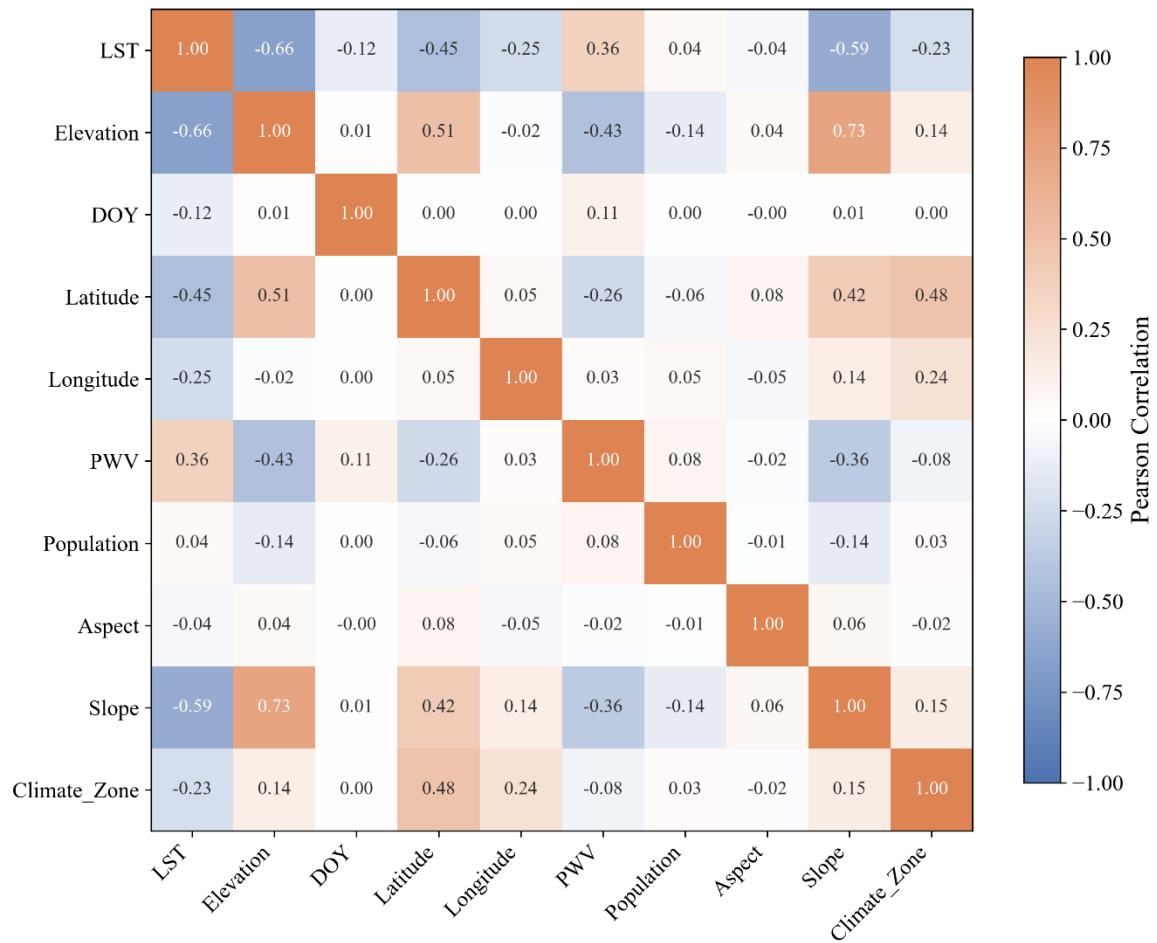
Response:

We thank the reviewer for this valuable suggestion. To assess potential multicollinearity among the selected covariates, we performed a Variance Inflation Factor (VIF) analysis (Kutner et al., 1996; Thompson et al., 2017). All VIF values were below 3.2, well under the commonly accepted threshold of 5, indicating that LST, PWV, population density, and the topographic variables provide complementary and largely independent information. These results have been incorporated into the Methods (lines 156-159) and Discussion section (lines 442-448) of the revised manuscript, and the corresponding VIF analysis is presented in Figure 15.

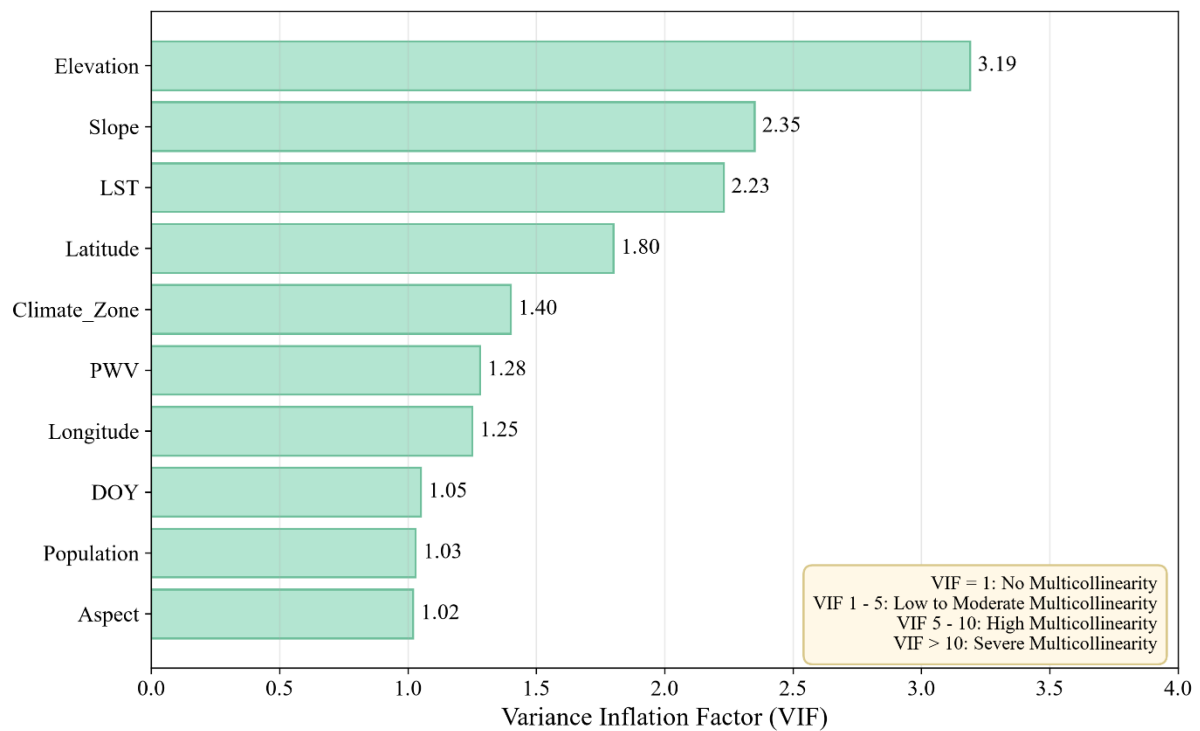
Discussion:

“The Variance Inflation Factor (VIF) analysis further demonstrated that the covariates set did not exhibit multicollinearity, with all VIF values remaining below 3.2 (figure 15), and well under the commonly used threshold of 5 (Kutner, 1996; Thompson et al., 2017). The highest value was associated with Elevation, which is expected because of its natural correlation with slope and temperature-related variables (LST). These results indicate that LST, pwv, population density, and the topographic covariates provide complementary and largely independent information, supporting the physical consistency and robustness of our downscaling framework.”

(a) Predictor Correlation Matrix



(b) VIF by Predictor (Training Set)



“Figure 15. Multicollinearity assessment of the 10 predictor variables used in the LightGBM models. (a) Pearson correlation matrix among predictors. (b) Variance inflation factor (VIF) for each predictor, computed on the training set. All predictors fall well below the conventional multicollinearity threshold of 5. The absence of severe multicollinearity ($VIF > 10$) indicates that LST, PWV, population, and the topographic covariates each contribute largely independent, complementary information to the model.”

3. The study uses ERA5 reanalysis data as the benchmark, but no validation against in-situ meteorological station observations across India is conducted; the reliability of model outputs in real terrain remains unclear.

Response:

We thank the reviewer for this valuable suggestion. We agree that validation against independent ground observations is important. Accordingly, we incorporated an independent validation using observations from six India Meteorological Department (IMD) stations, where daily maximum air temperature, minimum air temperature, and wind speed data were available. Consequently, independent validation could be performed for SAT and WCT, the two HPT indices that can be directly derived from the available station observations. The validation methodology has been described in the Methods section (line 145-148), and the corresponding results have been added to the Results section (lines 355-366, Figure 10). We also clarify that this comparison provides an independent evaluation of the ERA5-downscaled products against available ground observations, rather than direct validation of all HPT indices.

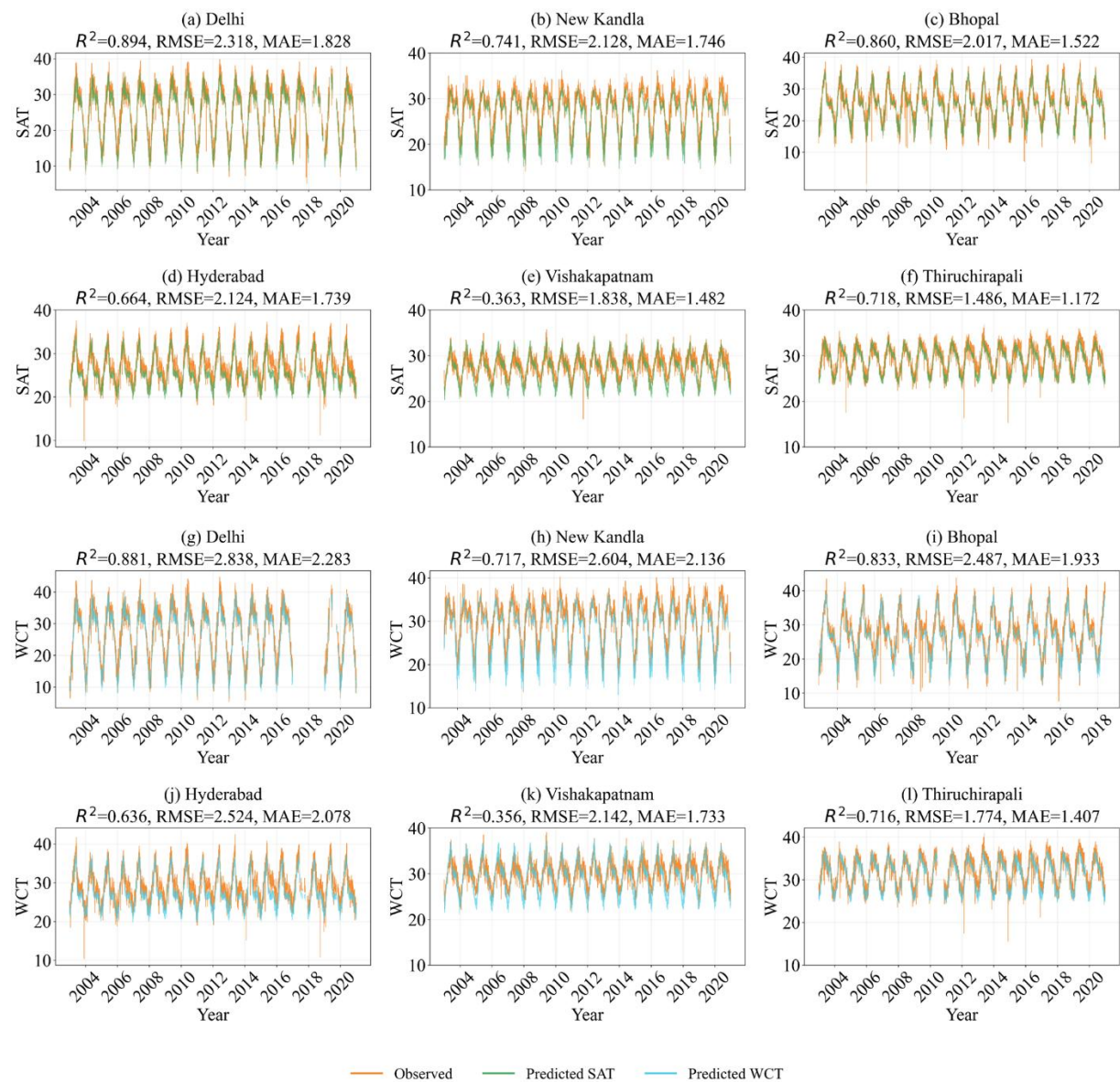
Method:

“Independent validation of the downscaled products was performed using ground-based observations from six India Meteorological Department (IMD) stations for SAT and WCT, the two HPT indices that could be directly evaluated using the available station measurements.”

Results:

“Model predictions for SAT and WCT are validated against independent IMD station observations at six locations. Independent validation was limited to these two HPT indices because only daily maximum air temperature (T_{max}), minimum air temperature (T_{min}), and wind speed observations were available at the selected IMD stations. Station records are screened for noise by removing values beyond

five standard deviations of the station mean, with comparisons restricted to days where both station and predicted values are available. R^2 ranged from 0.363 to 0.894 for SAT (figure 10(a)-10(f)) and 0.356 to 0.881 for WCT across stations (figure 10(g)-10(l)), with the strongest agreement at Delhi and Bhopal and the lowest at Vishakhapatnam. However, at Vishakhapatnam, our predictions closely track the observed seasonal cycle and annual variability, indicating that the model captures the site's overall thermal regime despite greater day-to-day variability. Overall, the comparison demonstrates good agreement between the ERA5-downscaled predictions and the available IMD observations, providing an independent evaluation of the model performance.”



“Figure 10. Time series comparison of predicted and observed (a–f) surface air temperature (SAT) and (g–l) wind chill temperature (WCT) at six India Meteorological Department (IMD) stations (Delhi, New Kandla, Bhopal, Hyderabad, Visakhapatnam, and Thiruchirapalli) during 2003–2020. Observed SAT was derived as the daily mean of the maximum and minimum air temperatures, whereas observed WCT was calculated from the daily mean air temperature and wind speed using the same equation applied to the HiTIC-India dataset. Orange lines denote the in-situ observations, while green and cyan lines denote the predicted SAT and WCT, respectively.”

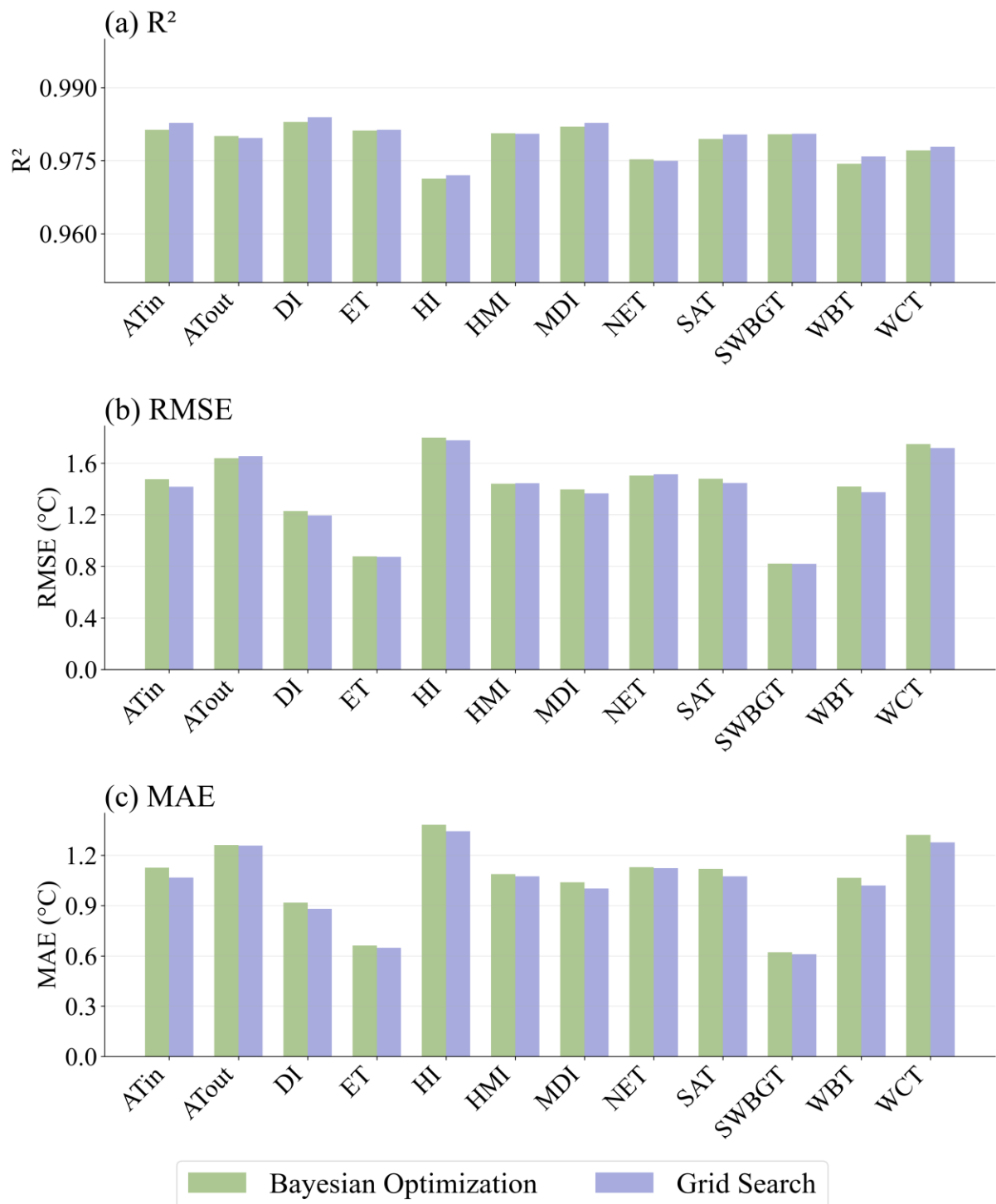
4. Hyperparameter tuning only applies grid search and 5-fold cross-validation, without testing more advanced strategies (e.g., Bayesian optimization), potentially limiting model performance optimization.

Response:

We thank the reviewer for this valuable suggestion. To evaluate the robustness of the hyperparameter optimization strategy, we compared the grid search approach used in this study with Bayesian optimization using the same five-fold cross-validation framework. Owing to the substantially higher computational time of Bayesian optimization for our large dataset, the comparison was performed using 25 optimization trials. Both approaches achieved nearly identical predictive performance (R^2 differences ≤ 0.002 ; RMSE differences ≤ 0.056 °C), indicating that additional Bayesian optimization trials would likely provide only marginal improvements in model performance while substantially increasing computational time. Accordingly, we retained the grid search approach for hyperparameter optimization in the final models. This comparison has been incorporated into the Discussion section (lines 539-547) of the revised manuscript, with the corresponding results presented in Supplementary Figure 13.

“We further assessed the robustness of the hyperparameter tuning strategy by comparing grid search with Bayesian optimization (Wu et al., 2019) using the same five-fold cross-validation framework. Both approaches achieved nearly identical predictive performance (R^2 differences ≤ 0.002 ; RMSE differences ≤ 0.056 °C), with grid search performing equally or slightly better for 9 of the 12 HPT indices (supplementary figure 13). Bayesian optimization was limited to 25 trials because of its higher computational time associated with iterative surrogate-model construction and updating, and under this

conditions yielded negligible improvements in its prediction performance. These findings indicate that the adopted grid-search strategy was sufficient to identify near-optimal hyperparameter combinations while providing a computationally efficient solution for generating the HiTIC-India dataset.”



“Supplementary Figure 13. Comparison of test-set model performance between Bayesian optimization (orange) and staged grid search (blue) for hyperparameter tuning of the 12 HPT indices: ATin, ATout, DI, ET, HI, HMI, MDI, NET, SAT, SWBGT, WBT, and WCT. (a) Coefficient of determination (R^2),

(b) root mean square error (RMSE, °C), and (c) mean absolute error (MAE, °C). The two tuning strategies yield nearly indistinguishable performance across all 12 indices, with R^2 differences ≤ 0.002 and RMSE differences $\leq 0.056^\circ\text{C}$.”

5. Model evaluation mainly relies on overall statistical metrics (R^2 , RMSE, MAE), while spatial heterogeneity of errors across different climatic/topographic zones of India is not analysed.

Response:

We thank the reviewer for this valuable suggestion. To evaluate the spatial heterogeneity of model performance across India's diverse climatic and topographic conditions, we conducted a comprehensive climate-zone-specific assessment. The original 18 fine-scale Koppen–Geiger climate classes were first aggregated into six broad thermal zones, and the classification is now provided in Supplementary Table 2. We then evaluated the performance of the LightGBM model trained using the complete Indian dataset separately within each climate zone using R^2 , RMSE, and MAE. In addition, we compared its performance with six independently trained climate-zone-specific LightGBM models using the same 80/20 train–test split. The results showed consistently high predictive performance across all climate zones and comparable accuracies between the India-wide and climate-zone-specific models, demonstrating that the proposed framework generalizes well across India's diverse climatic and topographic conditions. These additions have been incorporated into the Methods (lines 135-140), Results (lines 305-312), and Discussion sections (lines 467-483) of the revised manuscript, with the corresponding results presented in Figures 4–6, Supplementary Figure 5, and the climate-zone classification summarized in Supplementary Table 2.

Method:

“For each of the 12 HPT indices, a LightGBM model was trained using the complete dataset of India, which included samples from all six Koppen–Geiger climate zones. The 18 fine-scale Koppen-Geiger climate classes in India were aggregated into 6 broad thermal zones based on dominant thermal regime, as summarized in supplementary table 2. Then, the performance of this model was subsequently evaluated and compared with six independently trained LightGBM climate-zone-specific models to assess spatial heterogeneity and model generalizability in each HPTs.”

Result:

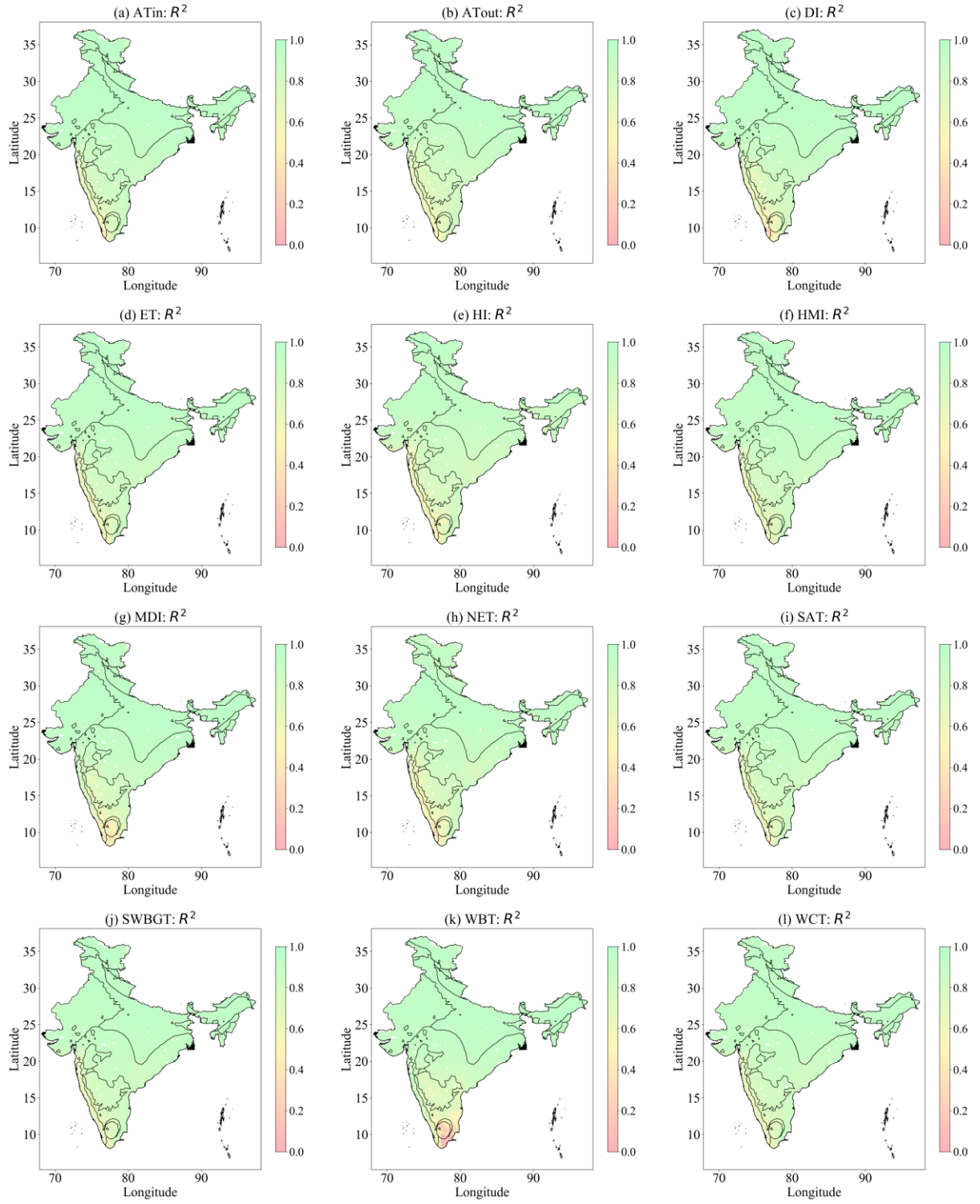
“The performance of the LightGBM model trained using the complete Indian dataset was further evaluated for R^2 , RMSE, and MAE (figure 4-6). The model consistently achieved high accuracy across India, with R^2 ranging from 0.86 to 0.99 for the 12 HPT indices. Spatially, R^2 remained uniformly high across most of India, while RMSE and MAE showed generally low and homogeneous error distributions. Slightly higher errors were observed over parts of the Cold Continental and Himalayan region for several HPT indices, although the overall differences among climate zones were small. These results indicate that the model maintains consistent performance across India's diverse climatic and topographic regions.”

Discussion:

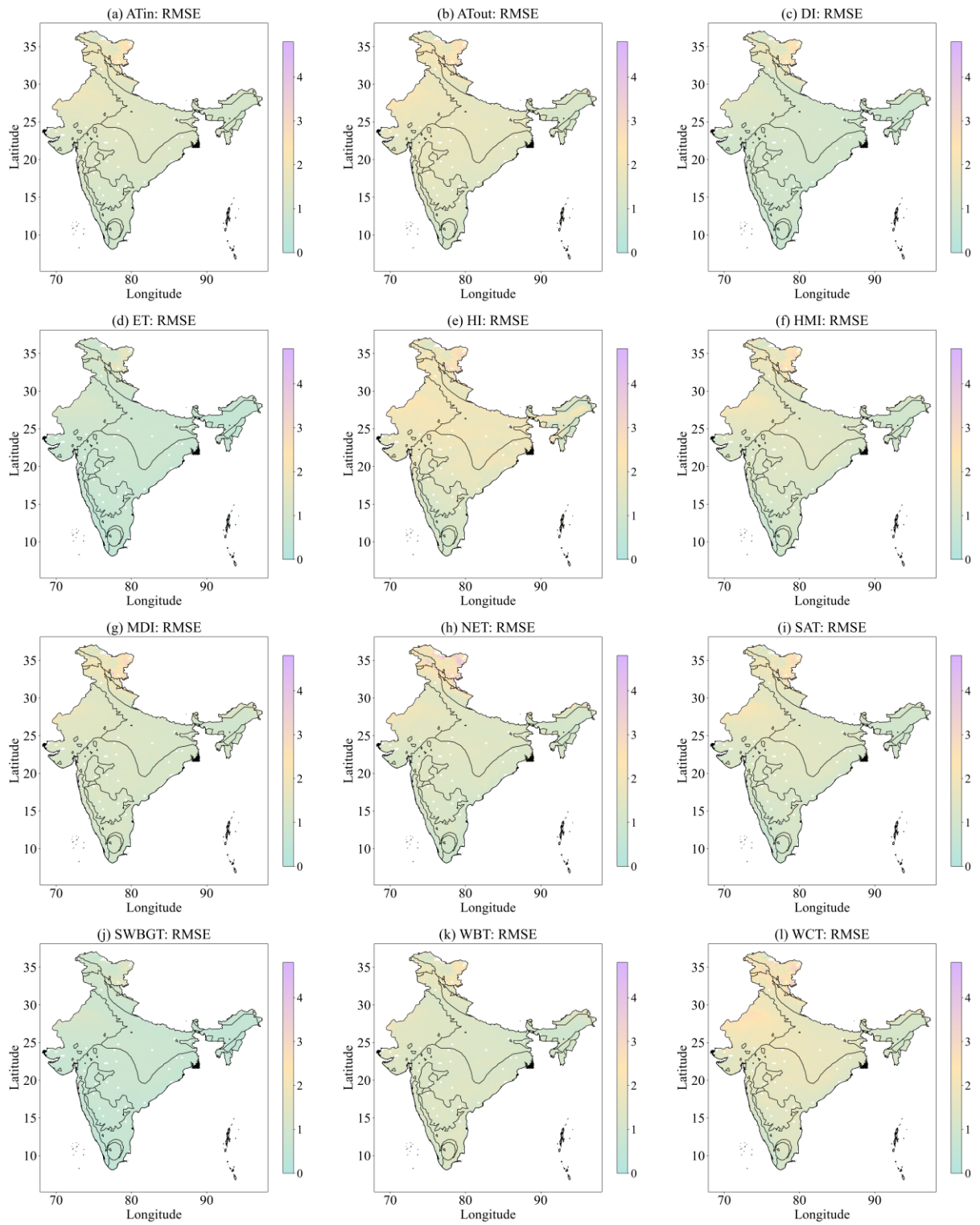
“Further evidence for the robustness of the modelling framework was obtained by comparing a LightGBM model trained using the complete Indian dataset, encompassing all six Köppen–Geiger climate zones, with six independently trained climate-zone-specific models. For all models, the available samples were randomly divided into 80% for training and 20% for independent testing. The India-wide model was trained using data from all climate zones, whereas each climate-zone-specific model was trained exclusively using data from its respective climate zone. Both approaches achieved comparable performance (supplementary figure 5), with R^2 generally exceeding 0.95, RMSE mostly within 0.7–2.0 °C, and MAE within 0.5–1.5 °C. The close agreement between the India-wide and climate-zone-specific models indicates that the learned predictor–target relationships remains stable. Slightly higher errors for SAT and WCT in the Cold Continental and Himalayan zone. These results demonstrate that a single model trained using data from all Indian climate zones provides reliable predictions across diverse climatic regions without substantial loss of accuracy compared with climate-zone-specific models. Li et al. (2023) similarly trained a single model using region-wide covariates across the North China Plain without climate-zone-specific stratification and reported comparable or better accuracies. However, the North China Plain is considerably smaller and more climatically homogeneous than India. Therefore, the comparable performance of the India-wide model across all six Köppen-Geiger climate zones provides strong evidence of its spatial generalizability.”

Supplementary Table 2. Mapping of fine-scale Köppen-Geiger classes to broad thermal zones.

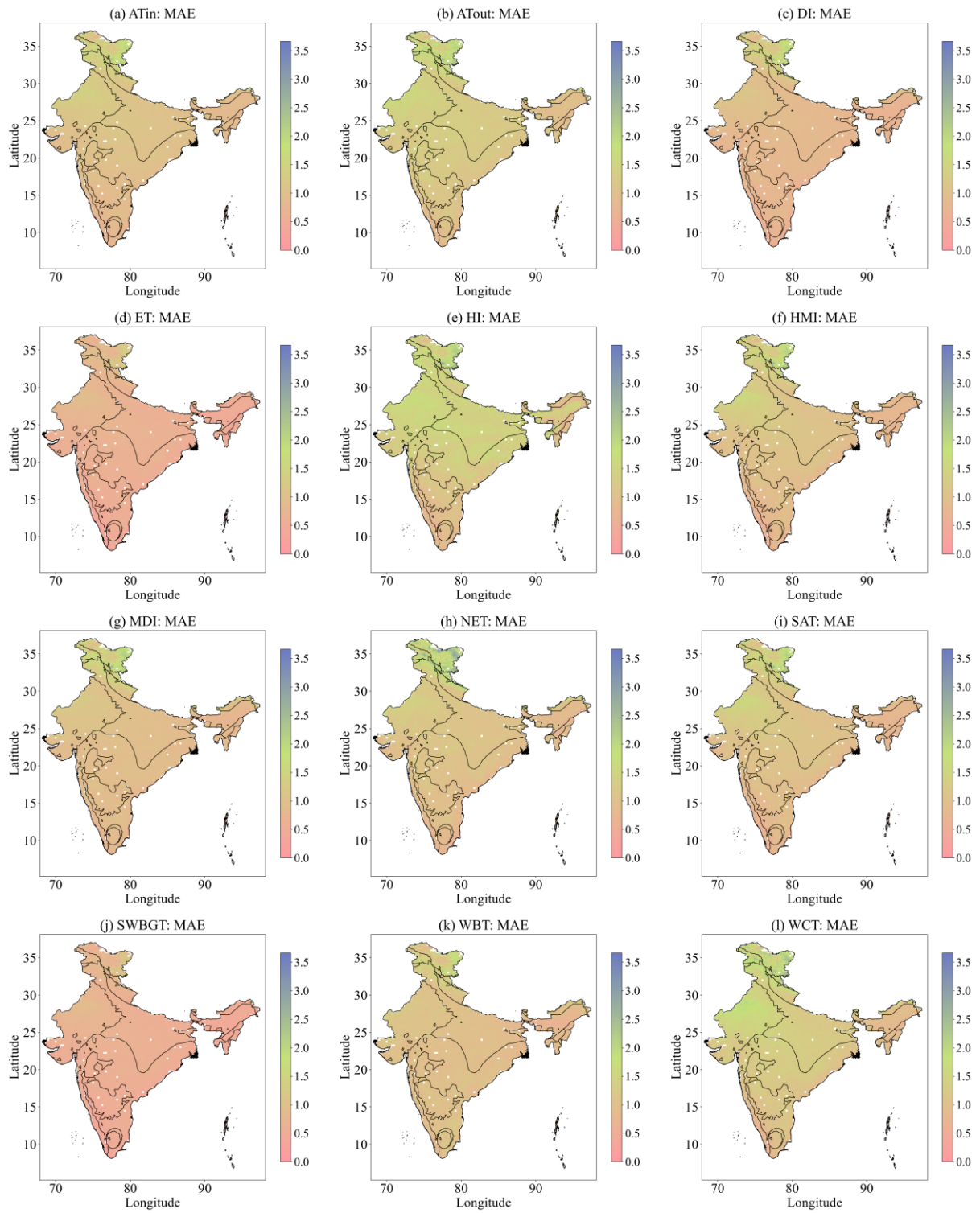
Broad Zone	Constituent Köppen Codes	Original Fine Classes	No. of Classes Merged
Tropical Monsoon	Am	Tropical Monsoon	1
Tropical Savanna	Aw	Tropical Savannah	1
Arid and Semi-Arid	BWh, BWk, BSh, BSk	Arid Desert Hot, Arid Desert Cold, Arid Steppe Hot, Arid Steppe Cold	4
Humid Subtropical	Cwa, Cfa	Temperate Dry Winter Hot Summer, Temperate Without Dry Season Hot Summer	2
Mountain and Temperate	Csa, Csb, Csc, Cwc, Cfc	Temperate Dry Summer Hot Summer, Temperate Dry Summer Warm Summer, Temperate Dry Summer Cold Summer, Temperate Dry Winter Cold Summer, Temperate Without Dry Season Cold Summer	5
Cold Continental and Himalayan	Dsa, Dwa, Dwb, Dwd, Dfa	Cold Dry Summer Hot Summer, Cold Dry Winter Hot Summer, Cold Dry Winter Warm Summer, Cold Dry Winter Very Cold Summer, Cold Without Dry Season Hot Summer	5



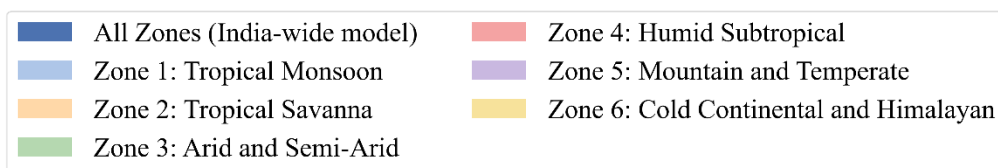
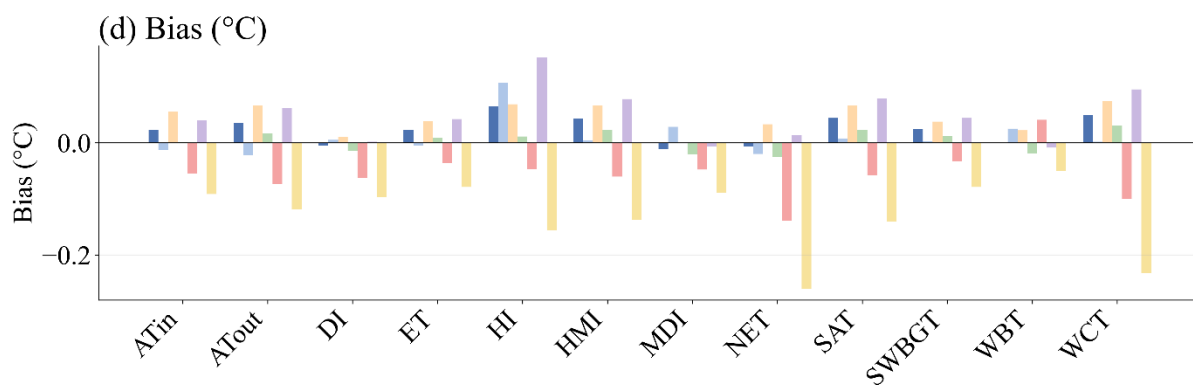
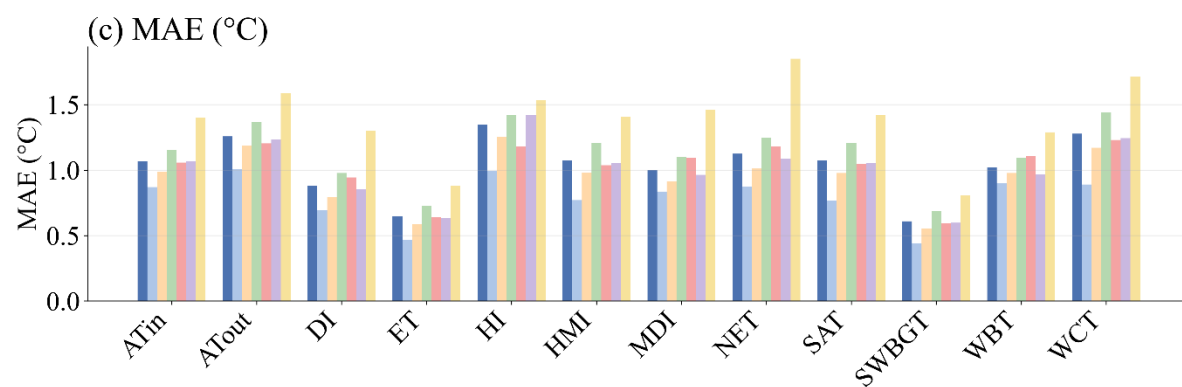
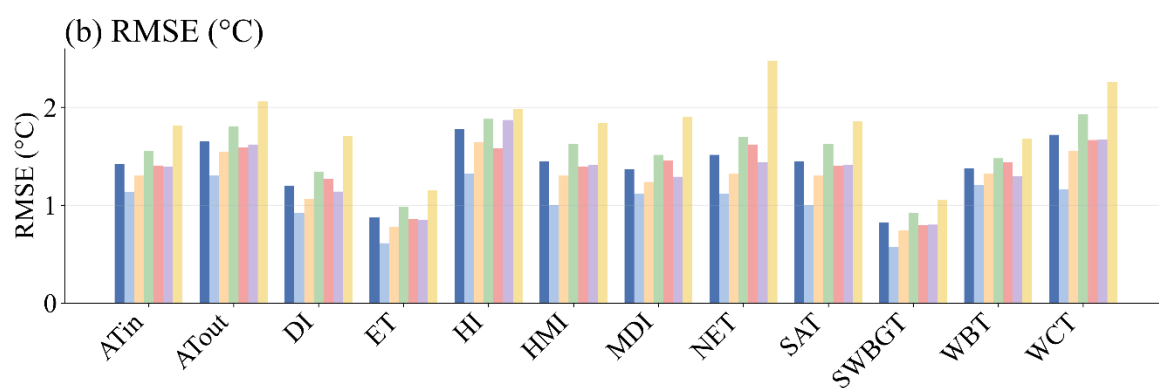
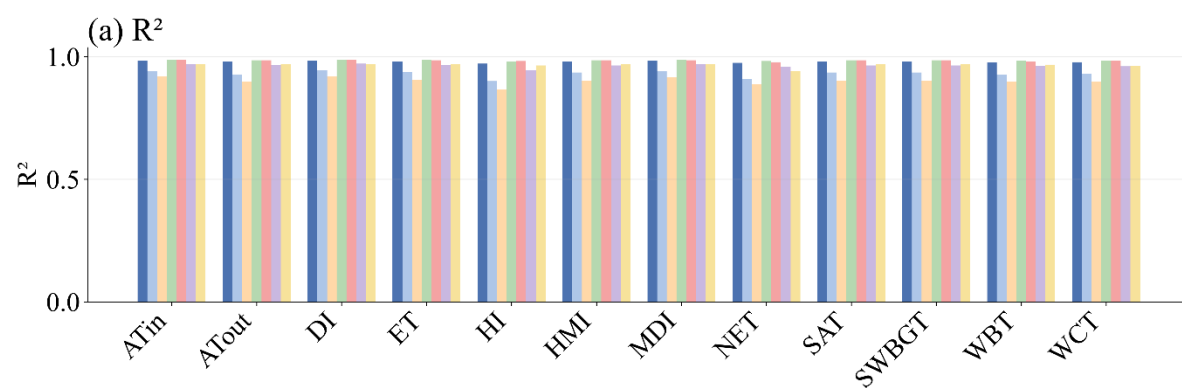
“**Figure 4.** Spatial distribution of the coefficient of determination (R^2) for the 12 human-perceived temperature (HPT) indices across India. The model maintains high R^2 values (>0.90) over most regions, with only slightly lower values in some parts of the Tropical Monsoon and Tropical Savanna zones.”



“Figure 5. Spatial distribution of the root mean square error (RMSE, °C) for the 12 human-perceived temperature (HPT) indices across India. RMSE values are generally low and spatially uniform, indicating consistent model performance across the country. Slightly higher RMSE values are observed over parts of the Cold Continental and Himalayan region for several HPT indices.”



“Figure 6. Spatial distribution of the mean absolute error (MAE, °C) for the 12 human-perceived temperature (HPT) indices across India. MAE values are generally low and spatially uniform, with only slightly higher values observed over parts of the Cold Continental and Himalayan region for several HPT indices.”



“Supplementary Figure 5. Comparison of zone-wise test-set performance metrics (R^2 , RMSE, MAE) between the All zones (India-wide model) model and zone-specific models for the 12 HPT indices. Close agreement between the two approaches across all zones indicates that predictor–target relationships learned from the full domain transfer well to individual climate zones.”

6. Error variation under extreme weather (heatwaves, cold waves) is not assessed; the model’s performance for extreme thermal stress events is unknown.

Response:

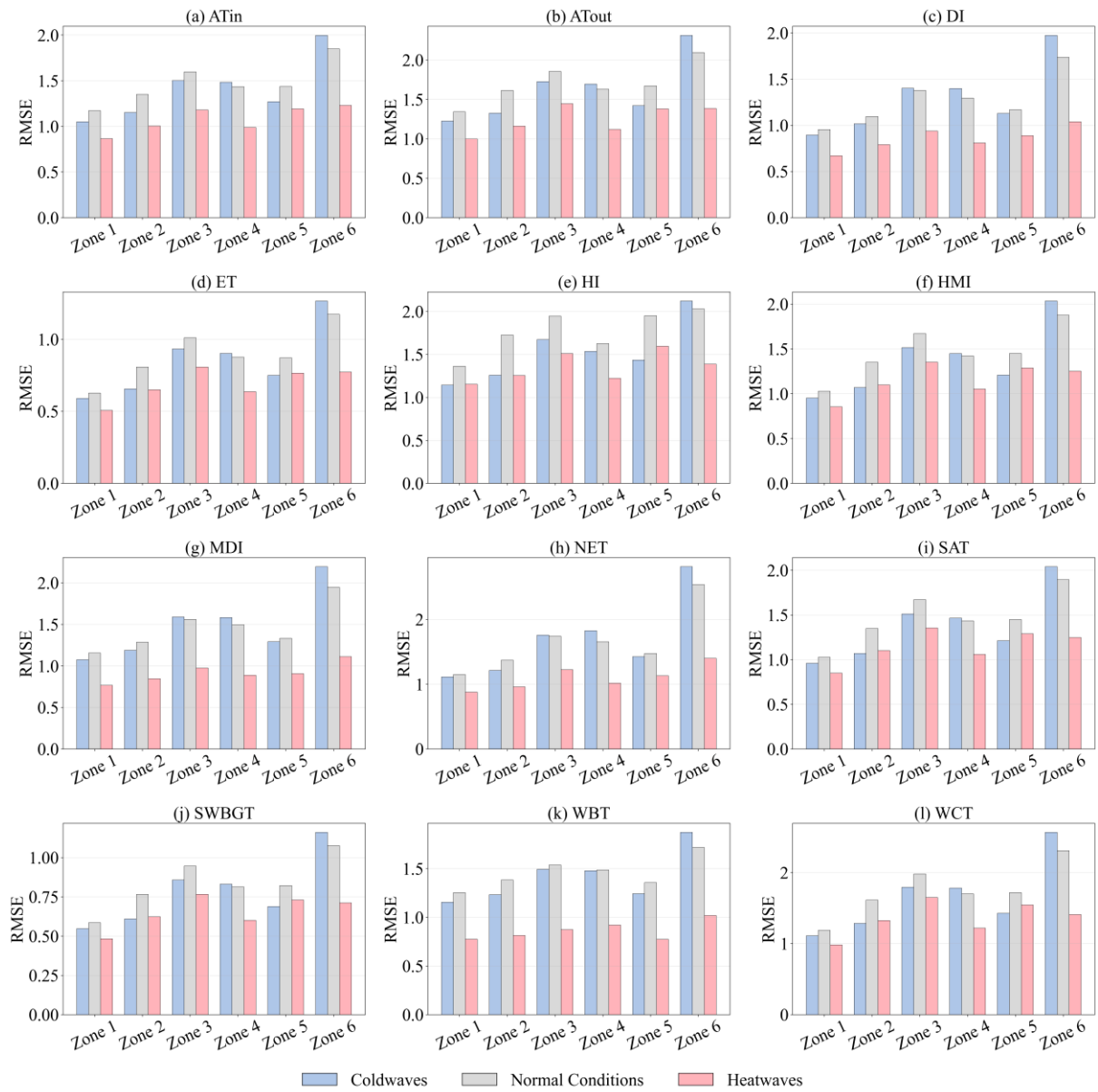
We thank the reviewer for this valuable suggestion. To evaluate the model performance under extreme thermal stress conditions, we conducted an additional analysis by stratifying the predictions into coldwave (<10th percentile), normal (10th–90th percentile), and heatwave (>90th percentile) regimes within each of the six Koppen–Geiger climate zones. For each thermal regime, we computed RMSE, MAE, and bias for all 12 HPT indices. The results showed a consistent error pattern across the different thermal regimes and climate zones. Model bias remained small, with a slight positive bias during coldwave conditions and a slight negative bias during heatwave conditions, while remaining close to zero under normal conditions. Error magnitudes during coldwave and normal conditions were generally comparable, whereas the lowest errors were observed during heatwave conditions for nearly all HPT indices. Although the Cold Continental and Himalayan zone exhibited comparatively higher errors, reflecting its greater climatic variability and complex terrain, the overall model performance remained robust across all thermal regimes. These additional analyses have been incorporated into the Methods (lines 140-143) and Results sections (lines 328-340) of the revised manuscript, with the corresponding results presented in Figures 7–9.

Methods:

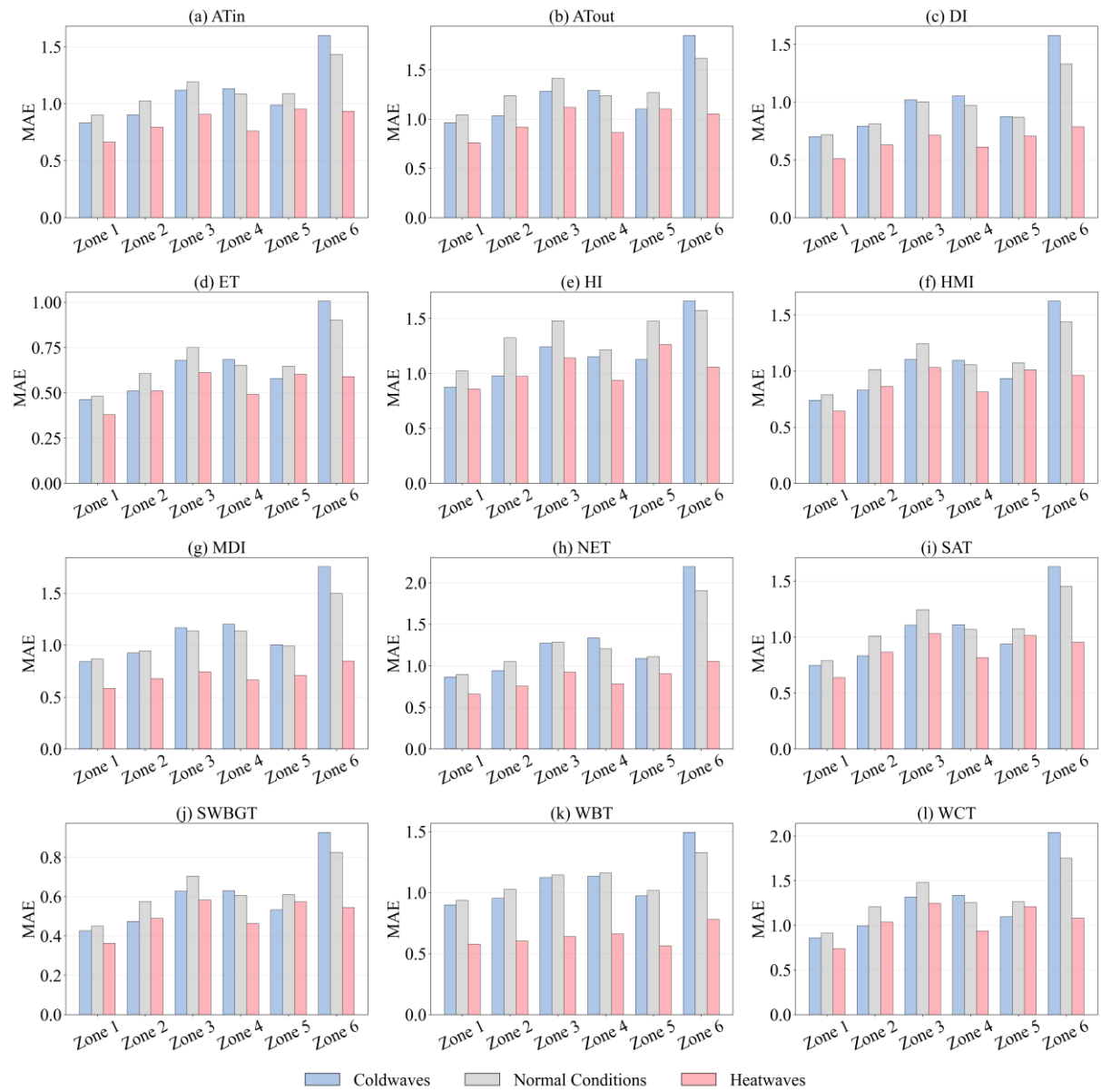
“Our model performance under extreme thermal conditions is further evaluated by classifying daily values into coldwaves (<10th percentile), normal conditions (10th–90th percentile), and heatwaves (>90th percentile) regimes within each zone, with Bias, RMSE, and MAE computed for each climatic zone.”

Results:

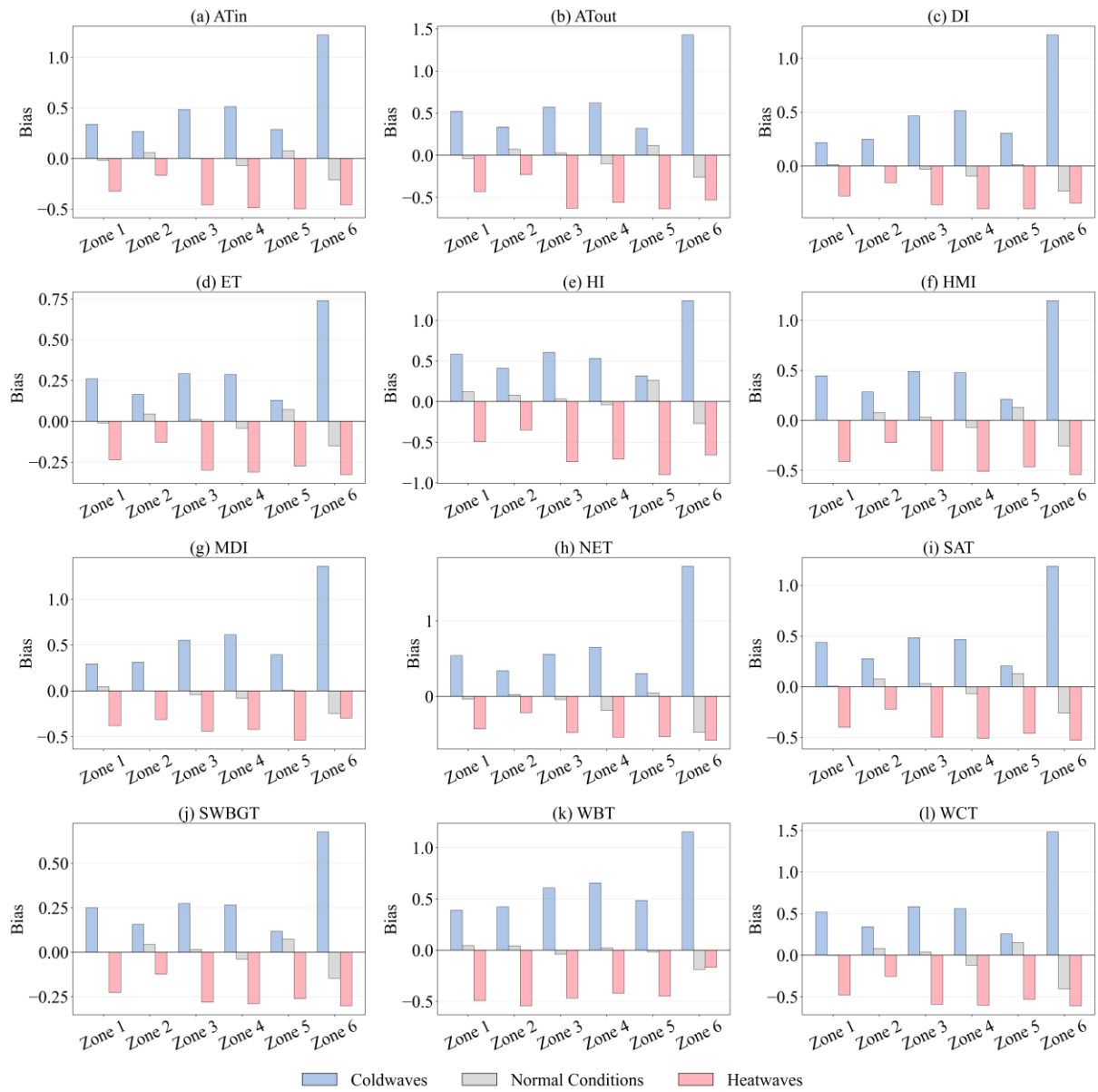
“To assess model performance under different thermal regimes, RMSE, MAE, and bias (figure 7–9) are computed separately for coldwave (<10th percentile), normal (10th–90th percentile), and heatwave (>90th percentile) conditions across the climate zones. A consistent error pattern is observed across all the HPT indices and climate zones. During coldwave conditions, the models generally exhibited a slight positive bias, indicating a tendency to overestimate the HPT indices by approximately 0.2–0.6 °C in most climate zones and by 1.2–1.8 °C in the Cold Continental and Himalayan zone. In contrast, during heatwave conditions, the models showed a slight negative bias (–0.4 to –0.9 °C), indicating a tendency to underestimate the HPT indices, whereas bias remained close to zero under normal conditions. RMSE and MAE were generally comparable during coldwave and normal conditions, while the lowest errors were observed during heatwave conditions for most HPT indices. For instance, the MAE of WCT ranges from 0.86–2.03°C during coldwaves, 0.91–1.75 °C under normal conditions, and 0.74–1.62 °C during heatwaves. Across all three thermal regimes, the Cold Continental and Himalayan zone consistently exhibited slightly higher bias, RMSE, and MAE than the other climate zones.”



“**Figure 7.** Spatial distribution of RMSE (°C) for the 12 HPT indices across climate zones of India (a)–(l) in coldwave (<10th percentile), normal (10th–90th percentile), and heatwave (>90th percentile) conditions.”



“Figure 8. Spatial distribution of MAE (°C) for the 12 HPT indices across climate zones of India (a)–(l) in coldwave (<10th percentile), normal (10th–90th percentile), and heatwave (>90th percentile) conditions.”



“Figure 9. Spatial distribution of bias (°C) for the 12 HPT indices across climate zones of India (a)–(l) in coldwave (<10th percentile), normal (10th–90th percentile), and heatwave (>90th percentile) conditions.”

7. The 12 HPT indices show differentiated accuracy, but the reasons for performance gaps among indices (e.g., wind-related indices with higher errors) are not explained.

Response:

We thank the reviewer for this valuable suggestion. We agree that differences in predictive performance among the 12 HPT indices merit further discussion. Accordingly, we have expanded the Future scope and limitations section to explain that the slightly higher errors observed for the wind-dependent indices (NET and WCT) are primarily attributable to the high spatial and temporal variability of wind speed,

which is strongly influenced by local topography, surface roughness, and land-cover heterogeneity that cannot be fully resolved at the 1 km spatial resolution used in this study. We also compare our findings with previous studies (Yan et al., 2021; Zhang et al., 2023; Li et al., 2023), which similarly reported lower predictive accuracy for wind-dependent thermal indices. Finally, we discuss that incorporating finer-resolution wind fields, improved terrain descriptors, and dynamic surface roughness information could further reduce these uncertainties and improve the performance of wind-related HPT indices in future versions of the dataset. These additions have been incorporated into the Future scope and limitations section (lines 573-584) of the revised manuscript.

Future scope and limitations:

“Among the 12 HPT indices, the wind-dependent indices NET and WCT consistently showed slightly higher RMSE and MAE across climate zones, seasons, and extreme thermal conditions. This is likely due to the strong spatial and temporal variability of wind speed, which is influenced by local topography, surface roughness, and land-cover heterogeneity that are not fully resolved at 1 km resolution. Similar behaviour has been reported in previous HPT downscaling studies. Zhang et al. (2023) and Li et al. (2023) found that the inclusion of wind speed increased prediction errors for NET and WCT, while (Yan et al., 2021) reported the lowest accuracy for wind-dependent indices in areas with complex terrain where local wind fields vary sharply over short distances. The agreement of these findings across independent studies suggests that wind representation remains a key source of uncertainty in high-resolution thermal downscaling. Future improvements may therefore be achieved with finer-resolution wind fields, enhanced terrain descriptors, and dynamic surface roughness information.”

8. Temporal validation only presents annual overall performance, lacking assessment of seasonal error differences between hot and cold seasons.

Response:

We thank the reviewer for this valuable suggestion. To evaluate potential seasonal differences in model performance, we conducted an additional seasonal analysis by separately assessing the hot (April–June) and cold (October–December) seasons for all 12 HPT indices over the study period (2003–2020). We computed R^2 , RMSE, and MAE independently for each season and found consistently high predictive

performance under both seasonal conditions. Although the cold season exhibited slightly lower errors and higher R^2 than the hot season, the differences were small, indicating that the proposed downscaling framework remains robust throughout the year. We further discuss that the slightly higher errors during the hot season are physically consistent with the greater variability in temperature, humidity, and wind conditions during the pre-monsoon period, particularly for wind-dependent and moisture-sensitive HPT indices. These additional analyses have been incorporated into the Methods (lines 143-145) and Discussion sections (lines 496-504) of the revised manuscript, with the corresponding results presented in Supplementary Figures 7–9.

Methods:

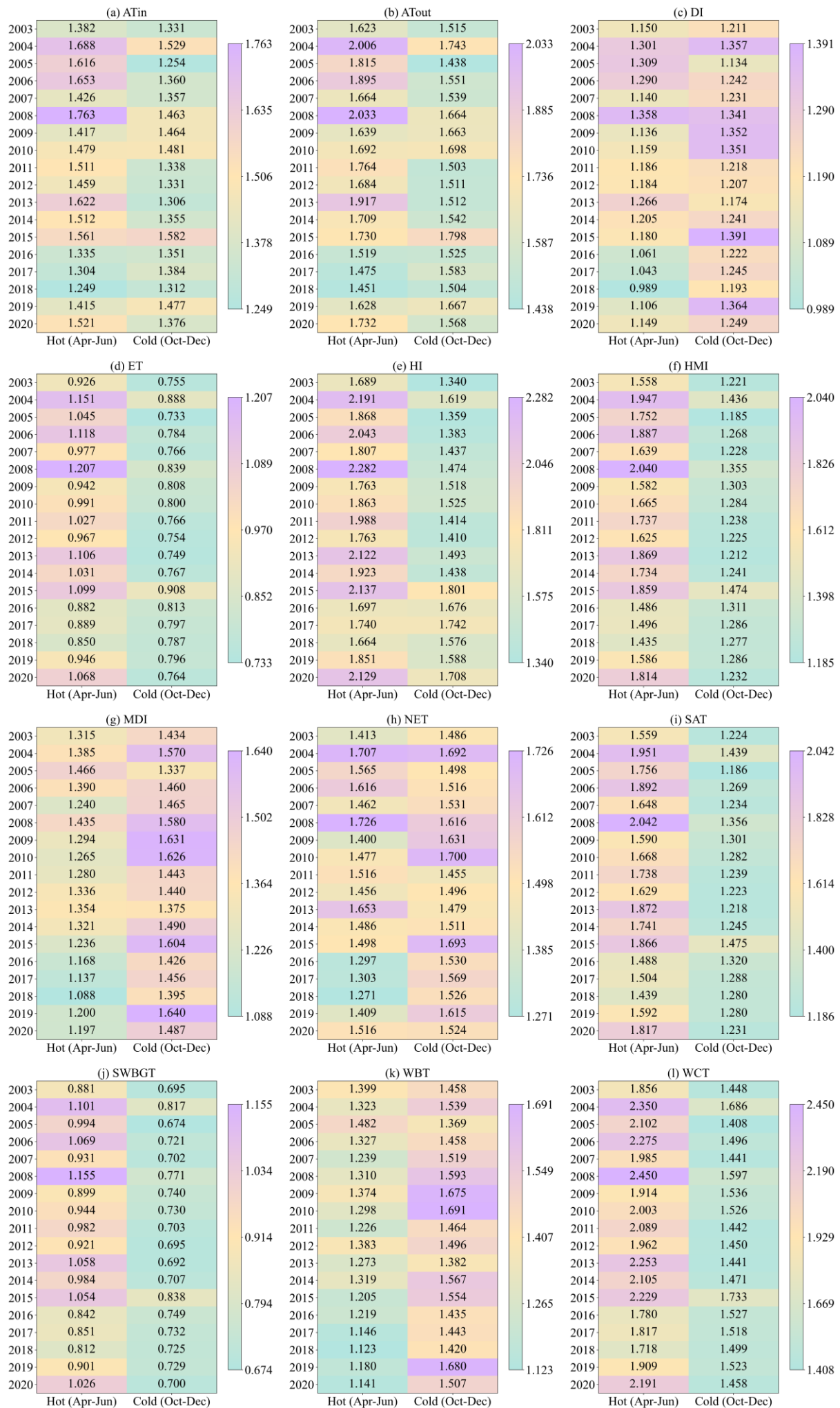
“Additionally, the seasonal error variations are assessed by evaluating these metrics separately for the hot season (April–June) and cold season (October–December) across all 18 years (2003–2020).”

Discussion:

“Seasonal evaluation showed consistently high model performance across all 12 HPTs and all years (2003–2020), with R^2 ranging from 0.94 to 0.99 in both the hot (April–June) and cold (October–December) seasons (supplementary figures 7–9). The cold season consistently exhibited slightly higher R^2 and lower RMSE and MAE than the hot season, with the largest differences observed in the wind-dependent indices (ATout, WCT) and moisture-sensitive indices (HI, MDI). For instance, the hot-season RMSE of WCT ranged from 1.86 to 2.45 °C, compared with 1.41–1.73 °C in the cold season. This pattern is physically consistent with the stronger atmospheric variability during the pre-monsoon period, when temperature, humidity, and wind conditions are more variable than during the relatively stable post-monsoon season.”



“Supplementary Figure 7. Year-wise R^2 for hot-season (April–June) and cold-season (October–December) predictions across the 12 HPT indices, 2003–2020. R^2 remains consistently high in both seasons (0.94–0.99) across all years and indices, with the cold season generally showing marginally higher values than the hot season.”



“Supplementary Figure 8: RMSE is consistently higher in the hot season than the cold season across nearly all years and indices, most notably for the wind-dependent (ATout, WCT) and moisture-sensitive (HI, MDI) indices; 2008 shows the highest hot-season RMSE for most indices.”



“Supplementary Figure 9: The seasonal pattern of MAE closely mirrors that of RMSE, with hot-season errors consistently exceeding cold-season errors across all 12 indices.”

9. The downscaling framework ignores diurnal variations of thermal factors; the daily mean dataset cannot reflect intra-day human thermal stress changes.

Response:

We thank the reviewer for this valuable comment. We agree that the current HiTIC-India dataset is limited to daily mean HPT values and therefore does not capture intra-day (diurnal) variations in human thermal stress. We have explicitly acknowledged this limitation in the Future scope and limitations section (lines 549-555) of the revised manuscript. Specifically, we clarify that, although the dataset is suitable for regional- and national-scale assessments of daily thermal conditions, it may not fully represent short-duration thermal extremes occurring at different times of the day, such as morning, midday, evening, or nighttime. We further note that the development of an hourly or diurnal-resolved HiTIC dataset using sub-daily meteorological covariates represents an important direction for future research to improve the characterization of human thermal stress and broaden the potential applications of the dataset.

“The HiTIC-India dataset represents daily mean values for the HPT indices and therefore does not capture intra-day (diurnal) variations in thermal stress. Human thermal exposure can differ substantially between morning, midday, evening, and nighttime periods, and daily means may underestimate short-duration extreme conditions relevant to applications such as heat advisories, outdoor labour safety, and nighttime thermal stress assessments. Accordingly, the dataset should be interpreted as a daily-mean thermal product, while the development of an hourly or diurnal-resolved dataset using sub-daily meteorological covariates is identified as an important direction for future work.”

10. The contribution degree of each covariate to HPT simulation is not quantified via feature importance analysis, so the driving mechanism of surface/topographic factors is unclear.

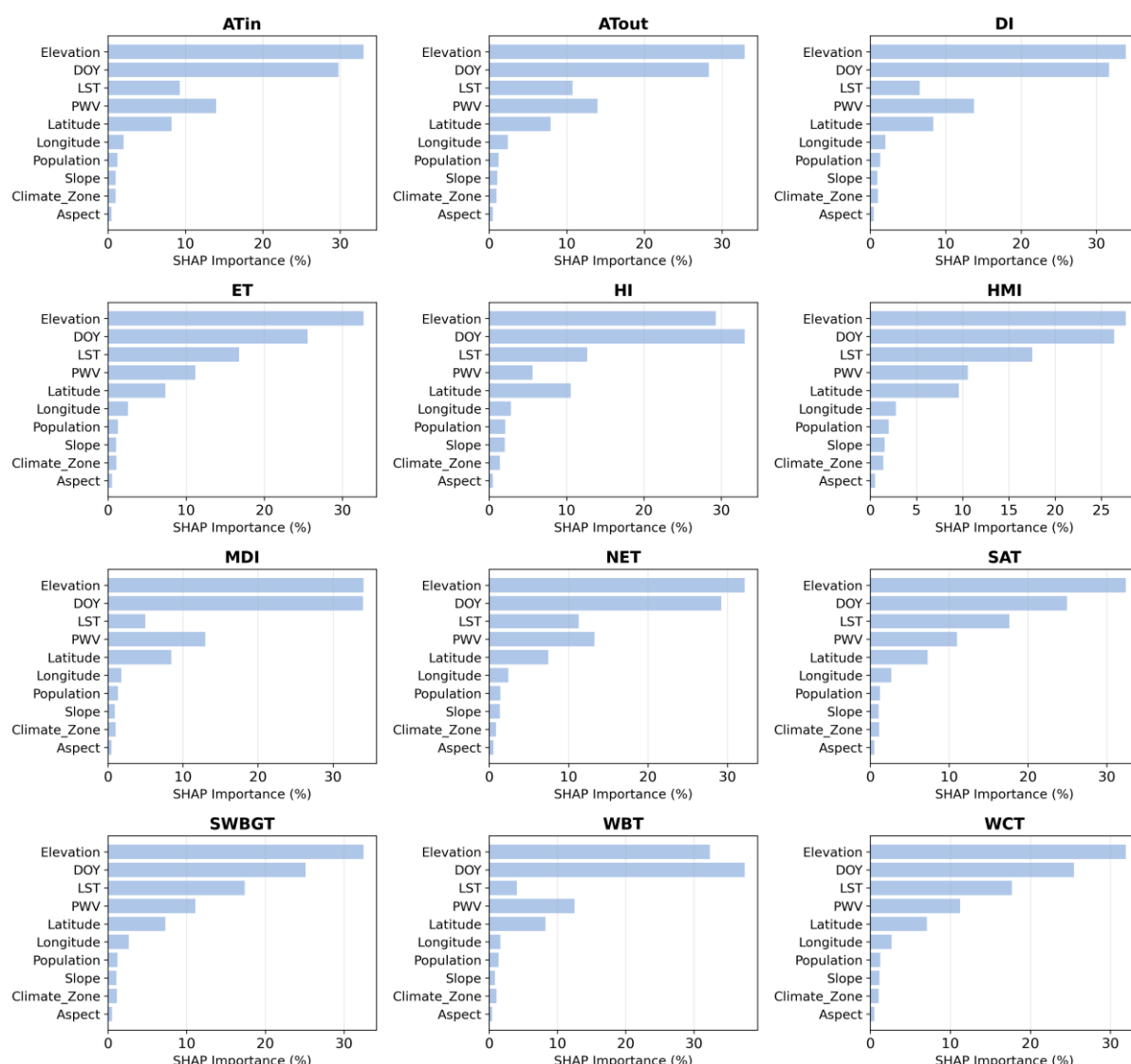
Response:

We thank the reviewer for this valuable suggestion. To quantify the contribution of individual covariates to the simulation of the 12 HPT indices and improve the interpretability of the modelling framework,

we performed a SHAP (shapley additive explanations) feature importance analysis (Lundberg and Lee, 2017). The analysis showed that elevation and day of year (DOY) were consistently the two most influential predictors across all 12 HPT indices, together accounting for approximately 55–65% of the total feature importance, highlighting the dominant roles of topography and seasonal variability in controlling thermal conditions across India. In contrast, LST contributed only 5–18%, while PWV, latitude, and population density provided smaller but meaningful contributions. Aspect, slope, and climate zone exhibited consistently low importance across all indices. These findings demonstrate that the proposed downscaling framework is driven by a combination of climatic, topographic, and surface characteristics rather than being dominated by a single remotely sensed temperature variable. We have incorporated the SHAP feature importance analysis and its interpretation into the Methods (lines 156-159) and Discussion section (lines 430-439) of the revised manuscript, with the corresponding results presented in Figure 14.

Discussion:

“Understanding the relative influence of predictor variables is essential for interpreting the physical basis of machine learning models used for thermal downscaling. SHAP analysis (Lundberg and Lee, 2017) showed that elevation and doy were consistently the two most influential predictors across all 12 HPTs, together accounting for approximately 55–65% of the total feature importance (figure 14). This indicates that topographic and doy (seasonal controls) are the primary drivers of thermal variability across India. While lst contributed only about 5–18%, pwv, latitude, and population density provided smaller but still meaningful contributions (2–14%). Aspect, slope, and climatic zones showed consistently low importance across the indices (figure 14). Overall, the SHAP results indicate that the downscaling framework captures the complex interactions among topographic, climatic, and surface-related variables that govern human thermal conditions.”



“Figure 14. SHAP feature importance (%) for the 10 covariates used in the LightGBM models for each of the 12 HPT indices.”

11. The dataset covers 2003–2020, but trend analysis of long-term thermal stress changes based on the new dataset is missing in the results.

Response:

We thank the reviewer for this valuable suggestion. To demonstrate the applicability of the HiTIC-India dataset for assessing 18 year changes in human thermal conditions, we have incorporated a pixel-wise linear trend analysis for all 12 HPT indices over the period 2003–2020 into the Results section (lines 411-424) of the revised manuscript and the updated the process in methods section (154-156). The robustness of the identified trends was further evaluated through pixel-wise statistical significance

testing ($p < 0.05$), with statistically significant trends shown in Supplementary Figure 1. The analysis revealed spatially coherent warming trends across the Cold Continental and Himalayan, north-eastern hill, and Tropical Savanna regions for most HPT indices, while several indices exhibited significant decreasing trends over the Mountain and Temperate zone. These additions demonstrate the capability of the HiTIC-India dataset to characterize long-term spatial patterns of human thermal stress across India. At the same time, we acknowledge that the current trend analysis is limited to the 2003–2020 period because of the temporal availability of the input datasets. Accordingly, we have explicitly discussed this limitation in the Future scope and limitations section (lines 556-561) of the revised manuscript, noting that the reported trends should be interpreted with appropriate caution, as they do not span the 30-year climatological baseline (Gent, 2022). We further identify extending the input datasets to longer time periods as an important direction for future research to enable more robust assessments of long-term changes in human thermal stress.

Methods:

“Long-term trends in the gridded predictions were estimated using pixel-wise linear regression, and their statistical significance was assessed using a two-tailed test ($p < 0.05$) (Gibbons and Pratt, 1975).”

Results:

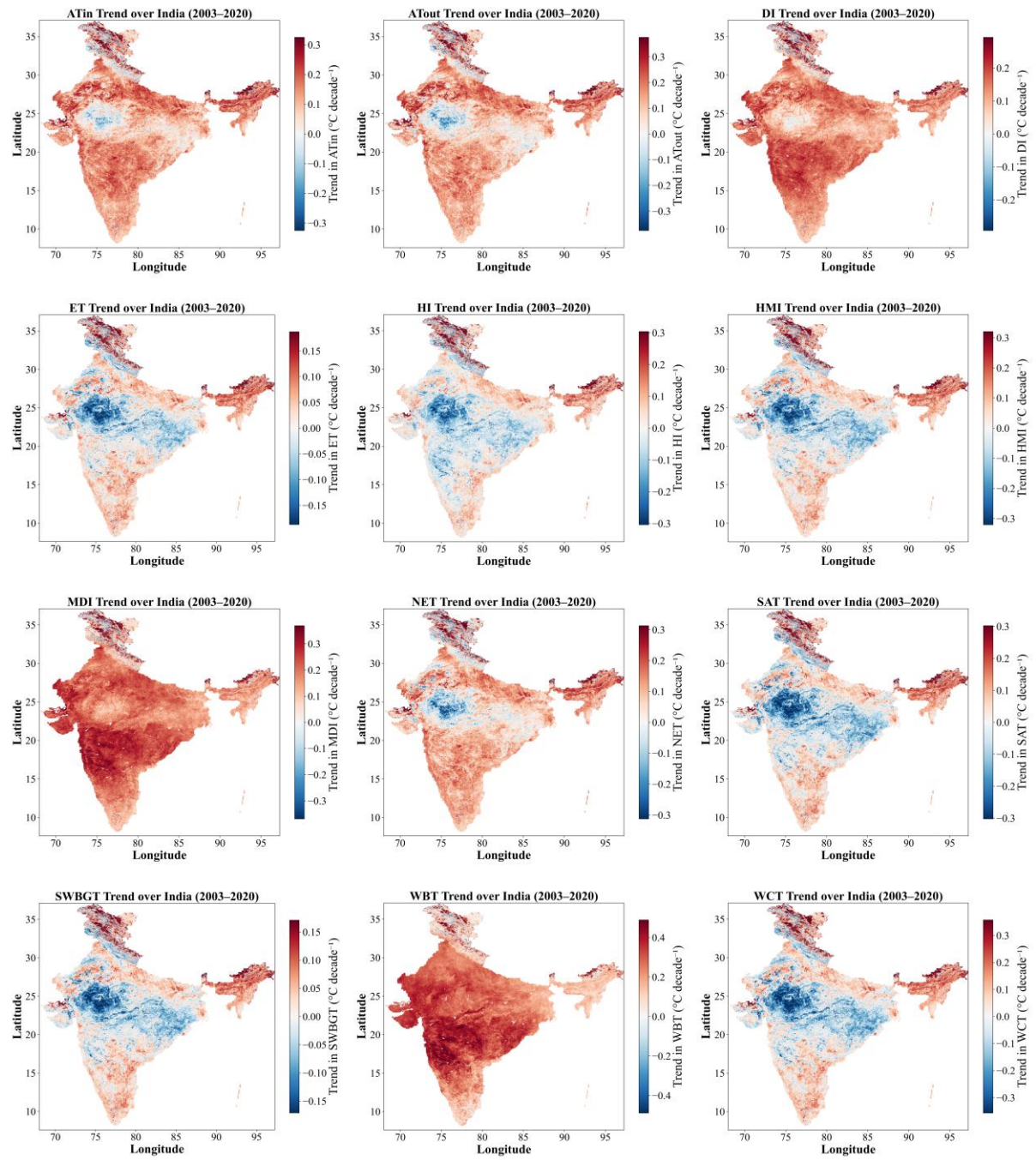
“Pixel-wise linear trends for all 12 human thermal indices (HPTs) during 2003–2020 are depicted in figure 13. Most indices show widespread warming over the northern Cold Continental and Himalayan zone, the north-eastern hill states, and the southern Tropical Savanna zone. Among the 12 indices, AT_{in}, AT_{out}, DI, MDI, NET, and WBT show warming trends across almost all climate zones, with MDI and WBT exhibiting the most spatially extensive and consistent increasing trends. In contrast, ET, HI, HMI, SAT, SWBGT, and WCT show a decreasing trend over the Mountain and Temperate zones of central India.

To evaluate the robustness of these trends, statistical significance was tested at each pixel with $p < 0.05$. The pixels with significant trends are shown in supplementary figure 1. The results confirm that the warming observed over the northern Himalayan region, the north-eastern hill states, and the southern Tropical Savanna zone is statistically significant and spatially coherent across most indices. Similarly, the decreasing trend in the Mountain and Temperate zone is statistically significant for ET, HI, HMI,

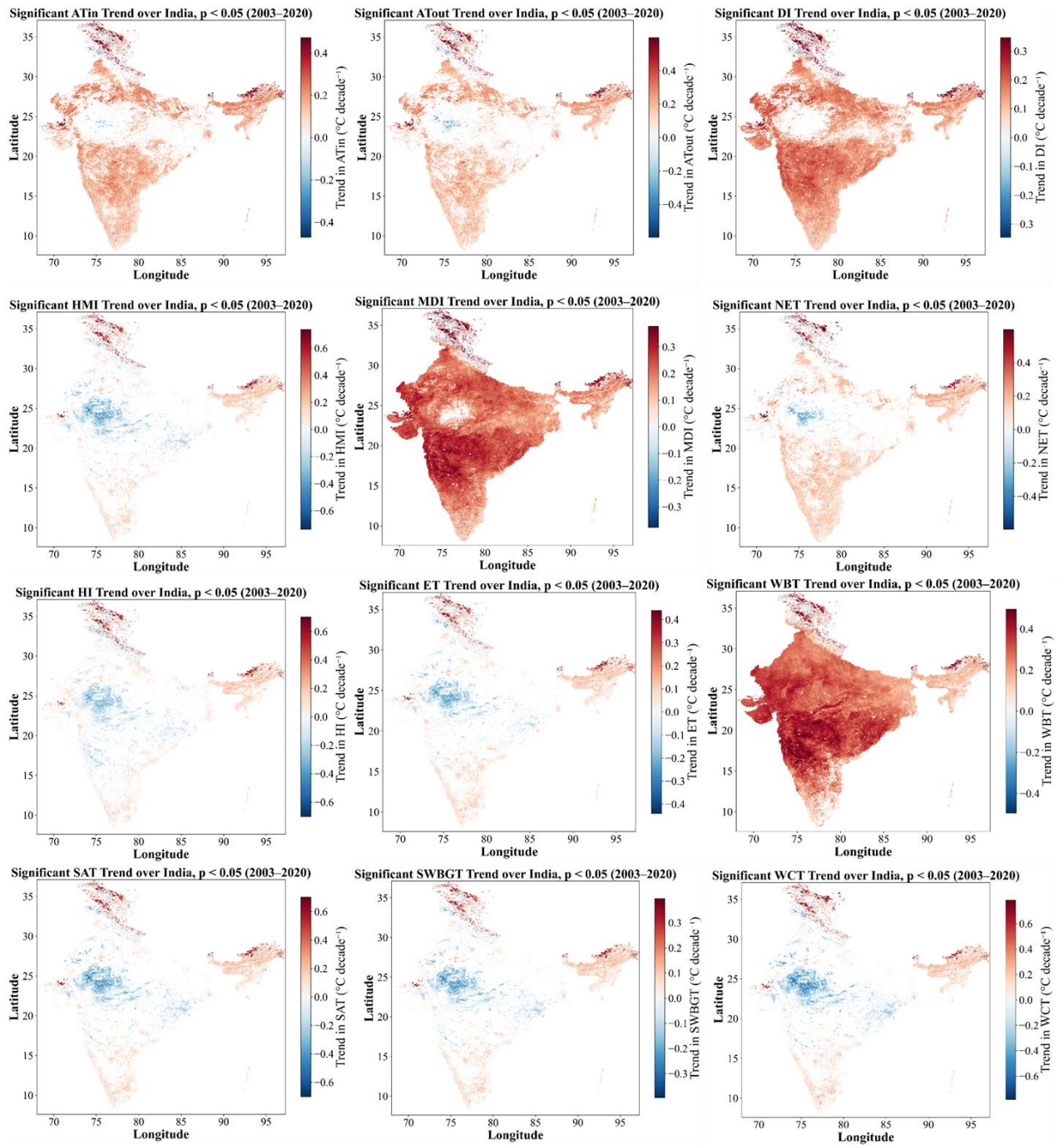
SAT, SWBGT, and WCT, indicating that this pattern represents a consistent regional signal rather than random variability.”

Future scope and limitation:

“Climate trend analysis is conventionally performed over a minimum 30-year baseline (Gent, 2022) to reliably separate long-term trends from natural inter-annual variability. The trends presented here span only 2003–2020 (18 years), limited by the availability of the underlying input datasets, and should therefore be interpreted with caution as they may be more sensitive to short-term climate variability than a full 30-year baseline would allow. Future work extending the input datasets over a longer period would enable more robust trend estimation.”



“Figure 13. Pixel-wise linear trend (°C decade⁻¹) in predicted daily values of the 12 HPT indices over India, 2003–2020.”



“Supplementary Figure 1. Statistically significant ($p < 0.05$) pixel-wise linear trends ($^{\circ}\text{C decade}^{-1}$) in predicted daily values of the 12 HPT indices over India, 2003–2020.: ATin, ATout, DI, HMI, MDI, NET, HI, ET, WBT, SAT, SWBGT, and WCT. Only pixels with a significant trend are shown.”

12. The study demonstrates seasonal spatial patterns of HPT but does not discuss urban-rural differences in thermal stress, despite incorporating population density as a covariate.

Response:

We thank the reviewer for this valuable suggestion. To address this concern, we incorporated an additional analysis to evaluate urban–rural differences in thermal stress using a boundary-based

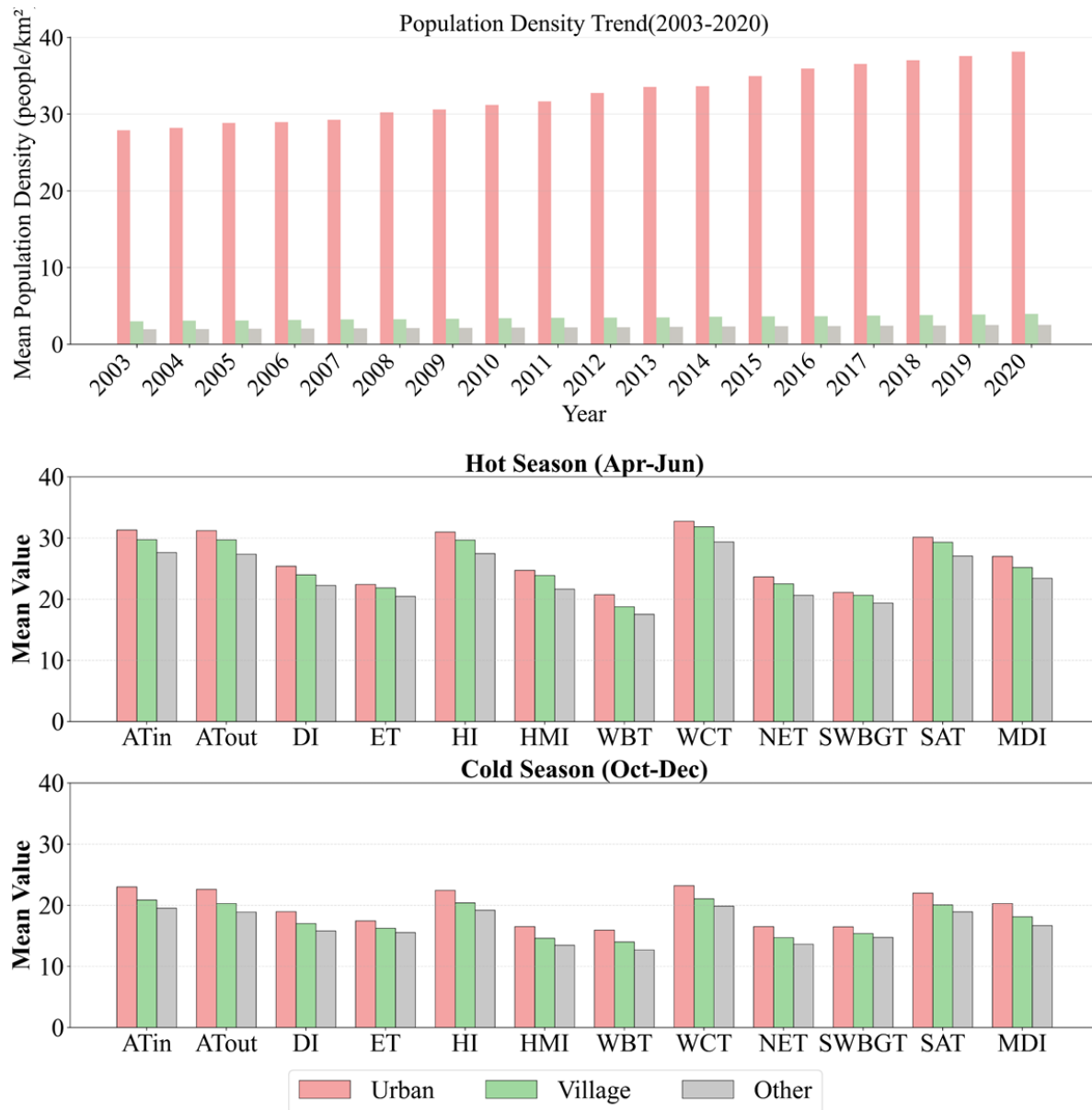
classification derived from the Global Human Settlement Urban Centre Database (urban areas) and Census 2011 village boundaries (Survey of India). We then compared the mean HPT values across urban, rural, and other land-use classes during the hot and cold seasons. The analysis revealed a consistent thermal gradient across all 12 HPT indices, with urban areas exhibiting the highest thermal stress, followed by rural and other areas. Furthermore, the observed increase in population density in urban areas during the study period (Jaysawal and Saha, 2014; Tumbe, 2016; Sarkar and Lakshmana, 2024) corresponded with elevated thermal stress, indicating that population density is an effective covariate for representing anthropogenic thermal influences within the downscaling framework. These additions have been incorporated into the Methods (lines 159-162) and Discussion sections (lines 486-495) of the revised manuscript, with the corresponding results presented in Supplementary Figure 6.

Methods:

“Further, to assess urban-rural differences in thermal stress, grid cells are classified into Urban, rural, and other land-use categories using a boundary-based classification derived from the Global Human Settlement Urban Centre Database (urban areas) and Census 2011 village boundaries (survey of India), and mean HPT values are compared across these classes for the hot and cold seasons.”

Discussion:

“Increasing urbanization across India during 2003–2020 (Jaysawal and Saha, 2014; Tumbe, 2016) provides an opportunity to assess whether population density captures anthropogenic thermal effects within the downscaling framework. Additionally, the population density increased consistently over the study period in urban areas (Ijas et al., 2026; Sarkar and Lakshmana, 2024), with a smaller rise in rural areas, whereas other areas remained relatively stable (supplementary figure 6, upper panel). Correspondingly, a consistent thermal gradient was observed across all 12 HPTs in both hot and cold seasons, with urban areas showing the highest values, followed by village and other areas (supplementary figure 6, bottom panel). This correspondence between rising population density and elevated thermal stress in the same areas suggests that population density is an effective covariate for capturing anthropogenic thermal influence across India.”



“**Supplementary Figure 6.** Temporal changes in population density and comparison of HTP values across land-use classes. The top panel shows the increasing trend in mean population density (people km⁻²) from 2003 to 2020 for Urban, Village, and Other areas. The middle and bottom panels present the mean values of 12 HTP indices during the hot season (April–June) and cold season (October–December), respectively. Across both seasons, Urban areas consistently exhibit higher thermal index values than Village and Other areas, indicating stronger thermal stress in densely populated regions. The persistent urban–rural thermal gradient highlights the contribution of population density in representing spatial variations associated with anthropogenic thermal influences.”

13. The conclusion overstates dataset applicability; it fails to clearly state the limitations and applicable scope of HiTIC-India in practical scenarios.

Response:

We thank the reviewer for this valuable suggestion. Accordingly, we have revised the Conclusion section (lines 586-602) to provide a more balanced summary of the HiTIC-India dataset. We now explicitly define its scope by clarifying that the dataset provides daily mean HPT values and is therefore intended for applications such as climate–health studies, long-term thermal exposure assessment, and adaptation planning. We also acknowledge the higher uncertainties associated with the wind-dependent indices (NET and WCT), particularly over mountainous regions. These revisions provide a clearer description of the dataset's applicability and limitations.

Conclusion

“This study presents HiTIC-India, a daily gridded dataset of 12 HPT indices at 1-km resolution across India for 2003–2020, generated by downscaling ERA5 reanalysis data using a LightGBM model. The dataset shows high and stable accuracy (RMSE = 1.38°C, R^2 = 0.97, MAE = 1.03°C) across climate zones, seasons, and extreme heatwave and coldwave conditions. Further, the reliability of the datasets is supported through several independent checks, including LOYO cross-validation, comparison between zone-specific models and a global dataset model, in-situ validation against IMD weather station records, and comparison with independent ERA5 and MERRA-2 datasets. Additionally, HiTIC-India provides daily-mean rather than hourly values, which is primarily intended for studies related to 18 long-term patterns, such as climate–health research, mapping chronic heat and cold exposure, and adaptation planning, rather than for real-time heat-warning applications that require finer time resolution. The two wind-based indices (NET and WCT) show slightly higher errors, particularly in mountainous regions; hence, results for these indices should be interpreted with some caution. Within these limits, the HiTIC-India datasets provide a high-resolution, spatially consistent representation that urban areas experience greater thermal stress than rural areas, while several regions of India, including the Himalayas and parts of the northeast and south, are undergoing significant 18 years long warming. Hence, it offers a valuable, high-resolution, openly accessible resource for research and decision-making related to human thermal stress, climate adaptation, and public health.”

14. No uncertainty analysis is performed for the final 1 km gridded dataset, which affects its credibility for climate-health and decision-making applications.

Response:

We thank the reviewer for this valuable suggestion. We agree that assessing the reliability of the final 1 km HiTIC-India dataset is important for its intended applications. However, direct pixel-level uncertainty quantification was not feasible because dense, independent 1 km observational datasets covering India are not available. To address this limitation, we incorporated an alternative uncertainty assessment by spatially aggregating the 1 km predictions to the native resolutions of ERA5 (0.25°) and the independent MERRA-2 (0.5° × 0.625°) reanalysis datasets and comparing the resulting 18-year mean fields with the corresponding coarse-resolution products using R², RMSE, MAE, bias, and spatial correlation. While ERA5 served as the primary training dataset, comparison with the independent MERRA-2 product provided an external evaluation of whether the downscaling framework preserved genuine climatological signals rather than source-specific artefacts. The results demonstrated strong agreement with both reference datasets, indicating that the proposed downscaling framework reliably preserves the large-scale thermal patterns of the driving reanalyses. We have incorporated the uncertainty assessment methodology into the Methods section (lines 164-174) and the corresponding results and discussion into the Discussion section (lines 506-521) of the revised manuscript, with the comparisons presented in Supplementary Figures 10 and 11.

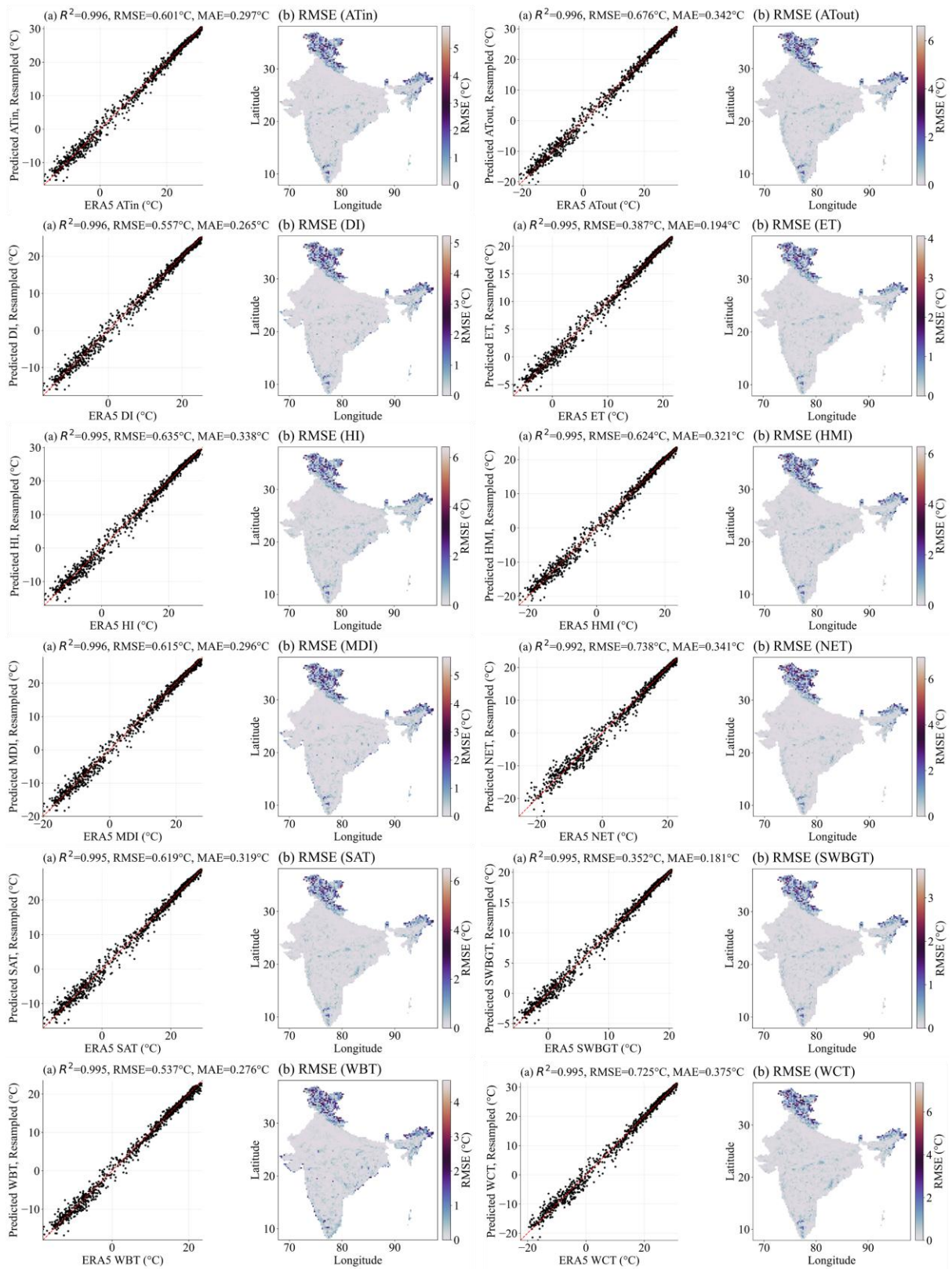
Method:

“Additionally, as dense, independent 1 km observations were not available across India, direct pixel-level uncertainty quantification was not feasible. So, the 1 km daily predictions were spatially aggregated back to the native resolution of two independent reference products ERA5 reanalysis (0.25°) and MERRA-2 (0.5° × 0.625°) and the resulting 18-year mean fields were compared against the original coarse-resolution data using R², RMSE, MAE, and bias. ERA5 served as the primary training source, while MERRA-2, an independent product not used for training, assessed whether the downscaling preserved genuine climatological signals rather than source-specific artifacts. As MERRA-2 does not natively provide relative humidity or vapour pressure, these were derived from its air temperature and dew-point temperature using the Tetens equation (saturation vapour pressure from air temperature, actual vapour pressure from dew-point temperature, and relative humidity as their ratio) (Tetens, 1930),

and used together with MERRA-2 air temperature and wind speed to compute the twelve HPT indices following the same formulations applied to the primary dataset.”

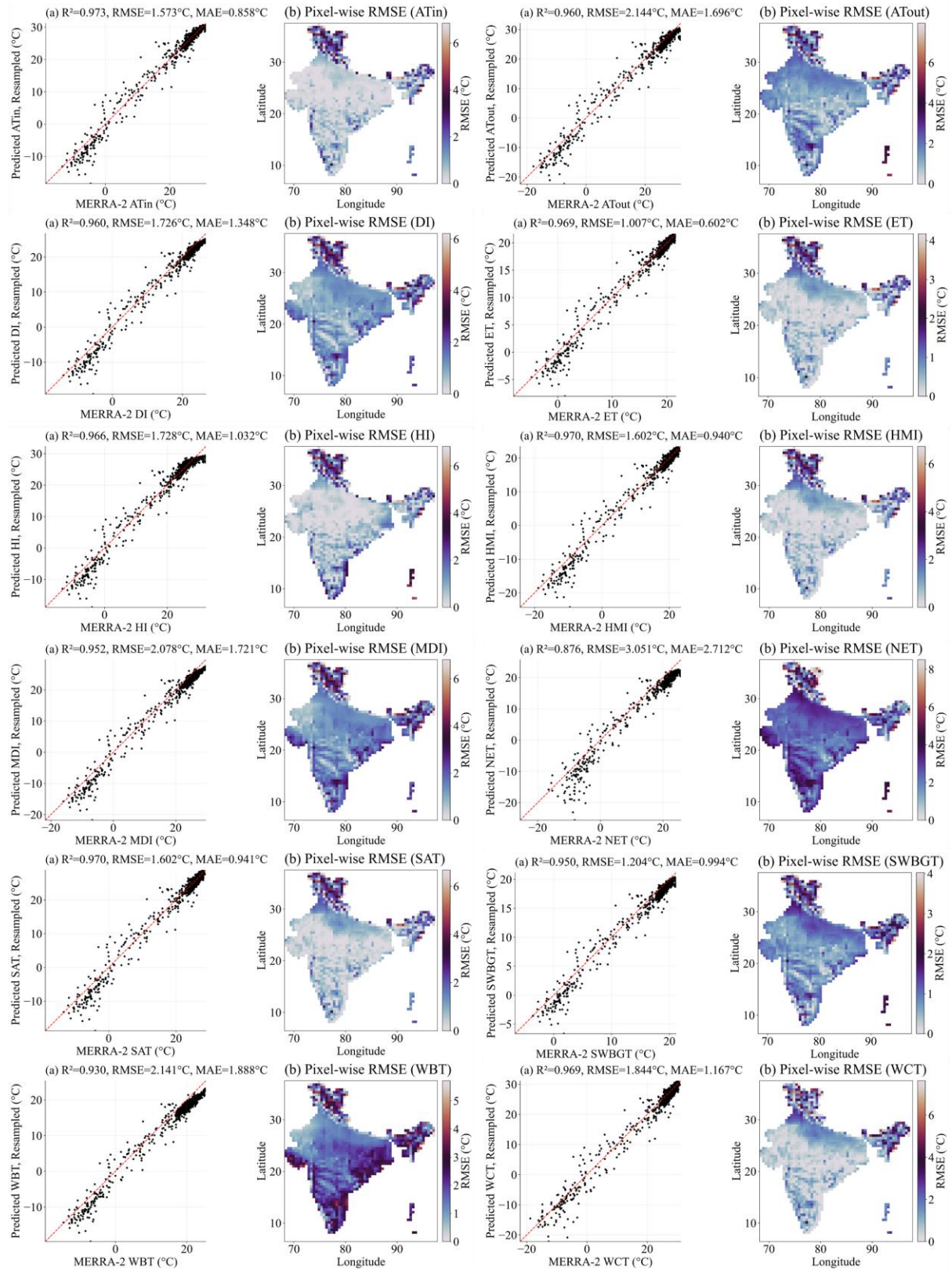
Discussion:

“Direct pixel-level uncertainty quantification was not feasible given the absence of dense, independent 1 km observational stations across India. As an alternative, we incorporated prediction reliability by resampling the 1 km daily predictions to the native resolution of two independent reanalysis products ERA-5 (0.25°) and MERRA-2 (0.5°) and comparing the resulting 18-year mean fields against the corresponding coarse-resolution source data for all 12 HPTs (supplementary figure 10). Against ERA-5, resampled predictions showed very high consistency ($R^2 = 0.992\text{--}0.996$, $\text{RMSE} = 0.35\text{--}0.74^\circ\text{C}$, bias within $\pm 0.08^\circ\text{C}$), which is expected because ERA5 was the primary training source. Agreement with MERRA-2, an independent product not used during model training, remained strong overall ($R^2 = 0.876\text{--}0.973$, $\text{RMSE} = 1.0\text{--}3.1^\circ\text{C}$), although slightly lower than for ERA5 (supplementary figure 11). The weakest agreement was observed for NET ($R^2 = 0.876$, $\text{RMSE} = 3.05^\circ\text{C}$), consistent with its greater sensitivity to humidity and wind velocity inputs and with the additional uncertainty introduced by deriving some MERRA-2 humidity variables through Tetens equation (Tetens, 1930). The consistently strong agreement across both independent references indicates that the downscaling framework preserves the large-scale climatological signal of the driving reanalysis data, while the comparatively lower MERRA-2 agreement reflects genuine cross-product differences between the source reanalyses rather than a deficiency specific to the downscaling model.”



“Supplementary Figure 10. Comparison of the resampled HiTIC-India predictions with the original ERA5 reanalysis for the 12 human-perceived temperature (HPT) indices based on the 18-year (2003–2020) mean fields. For each HPT index, panel (a) shows the scatter plot between the resampled HiTIC-

India predictions and the corresponding ERA5 values, together with the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Panel (b) shows the spatial distribution of the RMSE across India.”



“Supplementary Figure 11. Comparison of the resampled HiTIC-India predictions with the independent MERRA-2 reanalysis for the 12 human-perceived temperature (HPT) indices based on the 18-year (2003–2020) mean fields. For each HPT index, panel (a) shows the scatter plot between the resampled HiTIC-India predictions and the corresponding MERRA-2 values, together with the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Panel (b) shows the spatial distribution of the RMSE across India.”

15. The study does not discuss the influence of data gaps and interpolation errors of original MODIS LST and PWV products on HPT simulation results.

Response:

We thank the reviewer for highlighting this important point. We have now clarified the influence of the MODIS input datasets on the HiTIC-India product in the Future scope and limitations section (lines 562-572). Specifically, we explain that the LST covariate was derived from the seamless gap-filled dataset of Zhang et al. (2022b), and therefore any interpolation uncertainty originates from that validated product. In contrast, the MODIS PWV product retains a small number of short-duration data gaps. Rather than interpolating these missing values, HPT predictions were generated only for pixel-days where valid observations were available for all input variables, thereby avoiding additional uncertainty associated with PWV gap-filling at the expense of a slight reduction in spatiotemporal completeness.

Future scope and limitations:

“The MODIS-derived covariates used in this study differ in their data completeness. The LST covariate was obtained from the seamless global 1 km dataset of Zhang et al. (2022b), in which missing values in the original MODIS LST product were reconstructed using a validated spatiotemporal gap-filling framework. Consequently, the LST input is spatially and temporally complete, although some uncertainty associated with the gap-filling process may propagate into the HPT estimates. In contrast, the MODIS PWV product retains its original data gaps caused primarily by cloud and aerosol contamination. These gaps are generally limited to short-duration periods and constitute only a small fraction of the overall time series. Rather than interpolating the missing values, HPT predictions were generated only for pixel-days where valid observations were available for all input variables. This

approach avoids introducing additional uncertainty from PWV gap-filling at the expense of a small reduction in spatiotemporal completeness.”