General Comments from Reviewer #2:

This manuscript presents the Taiwan Reanalysis Downscaling Dataset (TReAD), a 2-km, 40-year d ynamically downscaled regional climate dataset produced with the WRF model. The dataset includ es six Essential Climate Variables and is evaluated against station observations, gridded datasets an d satellite-derived products. It addresses an important gap in long-term, high-resolution climate inf ormation for Taiwan, where complex topography limits the applicability of coarse-resolution reana lysis products.

The manuscript is generally well organised and presents a valuable dataset of broad interest to the climate, hydrology, ecology, disaster management and climate-impact research communities. The s tudy is well within the scope of ESSD. The manuscript would benefit from
clearer documentation and improved data accessibility, as well as a more transparent discussion of the dataset' s limitations and recommended uses.

I recommend publication after minor revision, provided that these points are adequately
addressed. Detailed comments are provided below.

Major comments

1. Although the model configuration is broadly described, the manuscript does not provide enough detail for a third party to reproduce the simulations. Please provide further clarification on:

- A complete inventory of the external forcing, configuration and preprocessing inputs.
- A more detailed explanation of the nesting, nudging and blending methods, how they were im plemented? and how the relevant parameters were specified?
- The preprocessing workflow used to generate the initial and lateral boundary conditions.

If configuration files, preprocessing scripts or workflow documentation can be made publicly a vailable, this would substantially improve reproducibility and long-term usability. I encourage the authors to provide such material.

*Thank you for this constructive suggestion regarding reproducibility. We have enhanced Section 2. 1 and 2.2 to provide details about the workflow, inputs, and configuration parameters. Below is the step-by-step summary of our simulation protocol and data preprocessing workflow:*

*1. External forcing and input data: ERA5 data inputs (6-hourly) include geopotential height, specifi c humidity, temperature, and uv wind components across 30 pressure levels, as well as surface fiel ds i.e., 10-m uv winds, 2-m temperature and dewpoint temperature, surface pressure, sea surface te mperature, skin temperature, and soil moisture and temperature across 4 soil layers.*

*2. Preprocessing and initialization workflow*

1. *Create intermediate meteorological files by metgrid package, which horizontally interpolat*

*es ERA5 data to WRF grid points.*

2. *Vertically interpolate intermediate meteorological files to sigma level of WRF by real package, which create initial condition and lateral boundary condition files.*
3. *Applied analysis blending scheme of Hsiao et al (2015), which blended the 3D wind, potential temperature, water vapor mixing ratio, and pressure in the initial condition file with those simulated from previous WRF run stored in the restart file for the outer domain. The inner domain is small and no blending process is needed. The cutoff length is 1200km, the same as what was used in Hsiao et al (2015).*
4. *Digital filter initialization, dfi_opt=3 and dfi_nfilter=7 (default Dolph filter), simulated backward for 30 minutes and then forward for 45 minutes were exercised for the outer domain to accelerate the spin up.*
5. *For the inner domain, soil temperature and moisture from previous WRF simulation stored in the restart file were kept to replace the ones in the initial condition file.*
6. *Run WRF for 6 hours to spin up the atmosphere, and the output won't be used.*
7. *Restart WRF runs for the rest of 6 hours and the output will be stored as TReAD data.*

*3. Nudging setup: Spectral nudging was exclusively applied during the initial one-month non-stop spin-up simulation at the very beginning to establish well-balanced soil temperature and moisture initial fields. It was not used during the main production integration cycles.*

*The above information has been incorporated in the revised Section 2.1 (Model configurations) & 2.2 (Reforecasting data and method).*

2. The Data Availability section should be expanded to provide clearer information for potential users. I recommend including a summary table of all variables, units, temporal and spatial resolutions, either in the main text or as a supplementary table. This would greatly improve the dataset's usability.

- The manuscript states that monthly products are available. Could the sub-daily outputs also be made available?

*Thank you for your practical recommendation to improve the usability and transparency of the TReAD dataset. We address your comments below:*

*1. Summary table of dataset variables: We have created a Summary Table listing key TReAD variables along with their corresponding physical units. To maintain manuscript readability while providing quick reference for potential users, this table has been included as Supplementary Table S3, and explicitly referenced in the updated "Data Availability" in the main text.*

*Table S3. Summary table of key dataset variables*

| *Variables* | *Unit* | *Temporal resolution* | *Spatial resolution* |
|---|---|---|---|
| *Mean temperature at 2 m height (T2MEAN)* | *℃* | *Monthly, daily, hourly* | *2km* |
| *Maximum temperature at 2 m height (T2MAX)* | *℃* | *Monthly, daily* | *2km* |
| *Minimum temperature at 2 m height (T2MIN)* | *℃* | *Monthly, daily* | *2km* |
| *Near-surface relative humidity (RH2)* | *%* | *Monthly, daily, hourly* | *2km* |
| *Mean precipitation (RAIN)* | *mm/day* | *Monthly, daily* | *2km* |
| | *mm* | *hourly* | |
| *Accumulated downwelling shortwave flux (ACSWDNB)* | *MJ/m2 per day* | *Monthly, daily* | *2km* |
| | *W/m2* | *hourly* | |
| *Surface pressure (PSFC)* | *hPa* | *Monthly, daily, hourly* | *2km* |
| *Wind speed at 10 m height (SPDUV10MEAN)* | *m/s* | *Monthly, daily, hourly* | *2km* |

*2. Availability of sub-daily outputs: Regarding the sub-daily products, we appreciate the interest in higher-frequency outputs. Sub-daily model outputs are not directly hosted on public cloud repositories primarily because higher temporal resolution data inherently carries higher subgrid noise and localized uncertainty, which require careful interpretation depending on the target application. To ensure proper use and data integrity, sub-daily outputs are made available upon reasonable request. Potential users are highly encouraged to contact the corresponding author to request access.*

*The "Data Availability" in the revised manuscript has been expanded to reflect these provisions:*
*"A detailed summary listing all available variables, units, and spatial and temporal resolutions is provided in Supplementary Table S3. In addition, daily and sub-daily simulation outputs are available upon reasonable request to the corresponding authors." …*

3. Several validation details require clarification

- How missing observational records were handled in the calculation of validation statistics? Were days with missing data excluded pairwise?

- Please clarify the elevation information used for the temperature adjustment and discuss the r emaining representativeness mismatch between point observations and TReAD grid-cell mean s. This issue is particularly relevant in complex terrain, where horizontal and vertical sub-grid variability may not be fully addressed by a simple lapse-rate correction.
- Is the 6h spin-up in each re-initialisation sufficient for the land surface model to reach equilibrium, particularly for variables with longer memory such as soil moisture? Please provide a justification for retaining only the second 6 hours of each 12-hour simulation. In particular, please discuss whether the 6-hour adjustment period is expected to be adequate for variables influenced by land-surface memory. If this was not explicitly tested, this should be a cknowledged as a potential limitation rather than necessarily requiring additional sensitivity ex periments.

*Thank you for these insightful questions regarding data validation protocols, topographic elevation adjustments, and land-surface spin-up considerations. We address each of your three points below:*

*1. Handling of missing observational records: Yes, days with missing observational data were excl uded pairwise. For all point-to-point and gridded statistical calculations (e.g., correlation and RMS D), model output days corresponding to missing observation records at a given station were omitte d prior to computing statistical metrics. This ensures that validation statistics are strictly paired and unskewed by missing observations.*

*2. Elevation adjustments and residual representativeness*
*Mismatch over complex terrain: To minimize the effect of grid-cell averaged terrain height of the WRF model, particularly on temperature, we adjusted temperatures by applying lapse rate with the differences between modelled raw heights of each grid and the actual heights of the closest station s. This adjustment was described in the second paragraph of Section 2.3. To better clarify the origi nal differences of terrain heights and the adjustment applied, we have revised the paragraph as belo w.*

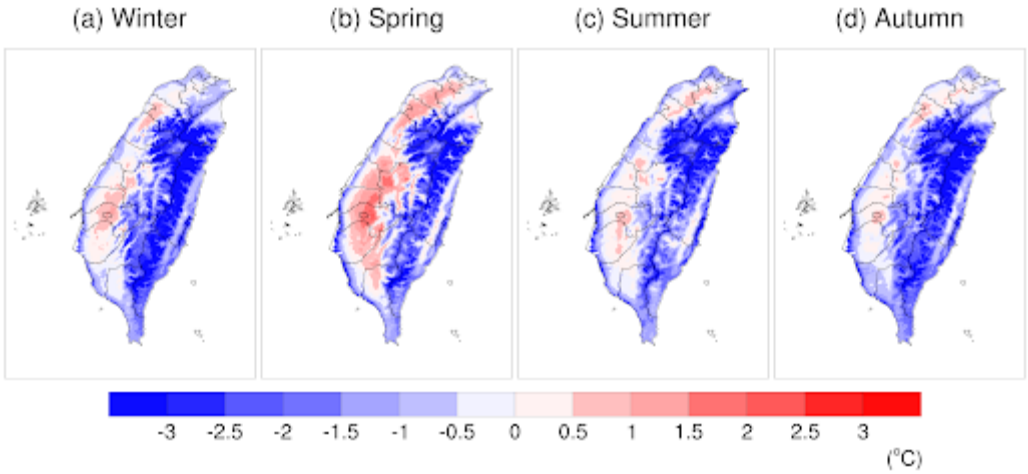
*"Taking into account the elevation difference between the point-based CWA observations and the grid-cell averages of TReAD (e.g., underestimated elevation at most mountainous locations), TRe AD temperatures were adapted using a constant lapse rate of 0.65°C per 100 m before comparison. Moreover, the TReAD grid cells nearest to the corresponding CWA station coordinates were chose n for comparison."*

*Remaining spatial mismatches: We have re-evaluated the temperature biases as follows:*

- *Lapse Rate Sensitivity: Replacing monthly-varying lapse rates with the standard 0.65°C/10 0m rate yielded negligible statistical differences, confirming that temperature biases stem f rom land-atmosphere boundary layer physics rather than vertical reduction rules.*
- *Physical Drivers of Spatial Biases: The western warm bias could be associated with an ove restimation of downward shortwave radiation, whereas the mountain cold bias is driven by overestimated rainfall and subsequent evaporative cooling.*
- *Spatial Scale & Local Convection: Model outputs (2-km) were spatially interpolated to 1-k*

*m for verification. The model exhibits limited capability of capturing the exact timing, spat ial distribution and intensity of localised convective systems, leading to maximum tempera ture differences in particular over high-mountain peaks like Yushan.*

- *Impact on Thermal Circulations: Long-term (40-yr) evaluation shows persistent underestim ation of daily Tmax and overestimation of Tmin. The dampened daytime thermal response could weaken the land-sea thermal gradient, thereby moderating the intensity of daytime o nshore sea breezes and upslope winds. While Tmax is underestimated on average across th e 22 stations, the gridded output shows no underestimation—and even a slight overestimati on—over lowland regions (see below the figure of biases of seasonal Tmax). Therefore, th e model can maintain a reasonable land-sea breeze circulation and should not be significan tly affected.*




*The discussion of diurnal local circulation and Tmax beyond the scope of this paper. We tend not t o add these discussion materials to the revised manuscript.*

*3. Spin-up period (6 hours) and Land Surface Model (LSM) memory constraints: We appreciate yo ur insightful point regarding land-surface memory. The 12-hour re-initialization cycle with a 6-hou r spin-up period (discarding the first 6 hours and retaining hours 7 – 12) was primarily designed to mitigate atmospheric boundary layer and precipitation spin-up spin-down artifacts while maintaini ng long-term atmospheric trajectory constraint from boundary conditions. However, we fully recog nize that a 6-hour spin-up duration is insufficient for long-memory land-surface state variables—p articularly deep soil moisture and soil temperature—to reach thermodynamic and hydrological equ ilibrium. Rather than conducting resource-prohibitive long-spin-up re-simulations, we have explicit ly acknowledged this short spin-up duration as a potential methodological uncertaintiesin Section 2. 2 and 4.2 (see below the paragraphs) in the revised manuscript. We note that while atmospheric fie lds (temperature, wind, precipitation) settle rapidly within 6 hours, deep soil moisture field continu ity may exhibit slight re-initialization shocks, which potential users evaluating soil-atmosphere fee dback should keep in mind.*

*"It should be noted that while a 6-hour spin-up period is adequate for atmospheric dynamical and thermodynamic variables to achieve physical equilibrium, it is not long enough for land-surface st*

*ate variables with longer memory, such as deep soil moisture and soil temperature." (in Section 2. 2)*

*"Model Parameterizations and LSM Memory Constraints: Physical parameterization schemes suc h as cloud microphysics, boundary layer mixing could introduce systemic biases. Furthermore, the 12-hour re-initialization cycle with a 6-hour spin-up, while effective for atmospheric variables, ma y be insufficient for some land-surface state variables (e.g., soil temperature and deep soil moistur e) to achieve full hydrological equilibrium." (in Section 4.2)*

4. The manuscript would benefit from a dedicated subsection summarising dataset limitations a nd sources of uncertainty relevant to potential users. Please discuss, where relevant:
- Limitations inherited from the ERA5 forcing.
- Uncertainty associated with model configuration and physical parameterisations.
- Potential limitations associated with the representation of land-use change.
- Uncertainty in the observational and gridded reference datasets used for validation, particular ly in mountainous regions.
- Limitations associated with comparisons between point observations and grid-cell averages.
I also recommend adding concise user guidance addressing:
- Applications and regions for which TReAD is particularly well suited.
- Regions or conditions where users should have particular caution.
- Variables for which systematic biases remain and users may wish to undertake application-sp ecific bias assessment or bias correction.
- Known limitations at high elevations and over small or remote islands.The purpose should be to help future users understand the appropriate and inappropriate uses of the dataset.

*Thank you for your recommendation. Adding a dedicated subsection on dataset limitations, uncerta inty, and user guidance significantly enhances the transparency and utility of the TReAD dataset fo r the scientific community. As recommended, we have added a new subsection titled "4.2 Data unc ertainties" in the revised manuscript. This section addresses the requested uncertainty dimensions (ERA5 boundary forcing, land-use representations, observational reference uncertainties over com plex terrain, and spatial scale mismatch) alongside a user guidance outlining appropriate/inappropri ate applications, caution zones, variable-specific bias alerts, and high-elevation/small-island caveat s.*

*Below is the complete text integrated into the revised manuscript.*

*TReAD is a regional downscaling data, not a product of regional reanalysis. This means there is no assimilation with observation data during data production and inevitable biases and uncertainty are expected. To facilitate informed usage of the TReAD dataset, potential users should consider the fo llowing inherited and internal sources of uncertainty:*

- *ERA5 boundary forcing inheritance: As a regional reanalysis downscaled from ERA5, TRe*

*AD inherits synoptic-scale biases and boundary forcing errors from ERA5, particularly reg arding the exact timing, trajectory, and intensity of synoptic systems (e.g., tropical cyclone s) over the Northwest Pacific.*

- *Static land-use representation: TReAD employs static land-use/land-cover (LULC) maps o f four years. Some transitions such as yearly changes in gradual urban expansion, agricult ural shifting, and canopy growth can not be presented by the LULC employed in this study and may introduce differences between simulation and observation. This contributes to the disparity in long-term terrestrial wind speed trends (roughness evolution) and localized sur face energy balance over expanding urban centers. Moreover, LULC is a product of satellit e retrievals and the retrieval process also can cause some uncertainty.*
- *Observational uncertainties in mountainous regions: The reference datasets (TCCIP gridde d observations) used for validation carry inherent uncertainties. Over high-altitude, rugged terrain (e.g., the Central Mountain Range), ground station density is extremely sparse. Inter polation artifacts in gridded observations over mountain slopes contribute to perceived mo del-observation discrepancies.*
- *Point-to-grid representativeness mismatch: Comparing point-based station measurements a gainst 2-km grid-cell averages is not a fair comparison and inherently introduces representa tiveness errors. A linear lapse-rate adjustment cannot fully capture sub-grid variability (rela ting to topography, radiative shading, etc.) particularly in narrow valleys, or high mountain summits like Yushan and Sunmoonlake.*
- *Typhoon performance generalizability: Although TReAD demonstrates high spatial fidelity during benchmark extreme events like Typhoon Morakot (2009), all typhoons across the en tire 40-year period have not been systematically evaluated . Users analyzing typhoon relate d hydrology are encouraged to conduct storm-specific evaluations when applying TReAD t o other typhoons.*

*To support appropriate scientific use of TReAD, we provide the following user guidelines. "TRe AD is well suited for regional climate assessments, mesoscale dynamics, and hydrological modelin g over agricultural plains, coastal plains and mid-elevation hills across Taiwan, where spatial patter ns and long-term trends are relatively well reproduced. However, users should be more cautious w hen applying TReAD to high-mountain regions, narrow valleys, and small or remote islands. Partic ularly, small islands would suffer the land-ocean blending characteristics, inducing modulated win d profile and diurnal surface condition. More importantly, precipitation and shortwave radiation ar e variables requiring bias correction depending on application. Localised convective rainfall displa cement over complex terrain leads to greater biases in intensity while overall spatial patterns are w ell represented. As for shortwave radiation against the satellite-derived product, TReAD reveals a g eneral overestimation of downward shortwave radiation, probably due to underestimation of simula ted cloud cover and cloud optical depth while faithfully depicting micro-scale terrain shading and g radients across mountainous regions that coarse reanalyses fail to capture."*

Minor Comments

1. The manuscript states that 17 stations on the plain and 5 mountainous stations were used, giving a total of 22 stations. However, Table S1 appears to contain only 20 stations. Please verify the station inventory and ensure that the manuscript, figures, tables, supplementary material and validation statistics consistently report the same number of stations.

*Thank you for carefully spotting this inconsistency between the main text and Supplementary Table S1. We appreciate the opportunity to rectify the station information across the manuscript components. We have thoroughly verified and updated the station inventory throughout the manuscript, figures, tables, and supplementary materials. Supplementary Table S1 has been fully updated to list all 22 CWA stations (17 plain stations and 5 high-mountain stations), complete with their station names, latitude and longitude, actual observation elevations, nearest TReAD model grid-cell raw elevations. Furthermore, we have cross-checked the revised manuscript to ensure that the validation statistics (e.g., correlation, RMSD), temporal/spatial figures, text descriptions, and supplementary tables correctly reflect the 22-station dataset.*

*Table S1. Names, elevations, longitudes and latitudes of the used CWA observation stations and the modelled elevation of the corresponding grids*

| No. | Name | Elevation (m) | Longitude | Latitude | Elevation_grid (m) |
|---|---|---|---|---|---|
| 1 | Tamsui | 19 | 121.4489 | 25.1649 | 16.2 |
| 2 | Keelung | 26.7 | 121.7405 | 25.1333 | 47.5 |
| 3 | Taipei | 6.3 | 121.5149 | 25.0377 | 14 |
| 4 | Hsinchu | 26.9 | 121.0142 | 24.8279 | 20.8 |
| 5 | Taichung | 84 | 120.6841 | 24.1457 | 95.2 |
| 6 | Penghu | 10.7 | 119.5631 | 23.5655 | 7.7 |
| 7 | Dongjidao | 43 | 119.6675 | 23.257 | 3.4 |
| 8 | Chiayi | 26.9 | 120.4329 | 23.4959 | 28.7 |
| 9 | Tainan | 40.8 | 120.2048 | 22.9932 | 15.7 |
| 10 | Kaohsiung | 2.3 | 120.3157 | 22.566 | 4.1 |
| 11 | Hengchun | 22.1 | 120.7463 | 22.0039 | 35.7 |
| 12 | Lanyu | 324 | 121.5583 | 22.037 | 128.4 |
| 13 | Dawu | 8.1 | 120.9038 | 22.3557 | 14 |
| 14 | Taitung | 9 | 121.1546 | 22.7522 | 9.7 |
| 15 | Chenggong | 33.5 | 121.3734 | 23.0975 | 33.2 |
| 16 | Hualien | 16 | 121.6133 | 23.9751 | 14.3 |
| 17 | Yilan | 7.2 | 121.7565 | 24.764 | 7.3 |
| 18 | Anbu | 825.8 | 121.5297 | 25.1826 | 797.4 |
| 19 | Zhuzihu | 607.1 | 121.5445 | 25.1621 | 744.7 |
| 20 | Sunmoonlake | 1017.5 | 120.9081 | 23.8813 | 721.8 |
| 21 | Alishan | 2413.4 | 120.8132 | 23.5082 | 2193.8 |
| 22 | Yushan | 3844.8 | 120.9595 | 23.4876 | 3328.1 |

2. The dataset is described as covering 1980-2019 and most of the validation uses this period. However, some analyses and figure captions refer to 1980-2018, including the comparison of a

nnual temperature time series and the spatial precipitation analysis (L255 and Fig.8). Please ex plain why 2019 is excluded from these/many analyses and ensure that the temporal coverage is stated consistently throughout the manuscript, figures, tables, captions and supplementary mat erial.

*Thank you for pointing out these discrepancies regarding the temporal coverage. We appreciate the opportunity to clarify and ensure complete consistency across all manuscript components. The TRe AD dataset covers the full 40-year period from 1980 to 2019, and all validation statistics, annual ti me series, and spatial analyses were computed accordingly. The "1980 – 2018" errors in temperatu re and precipitation sections (including the original Figure 8 for annual temperature and the origina l Figure 11 for spatial precipitation) have been corrected in the revised manuscript.*

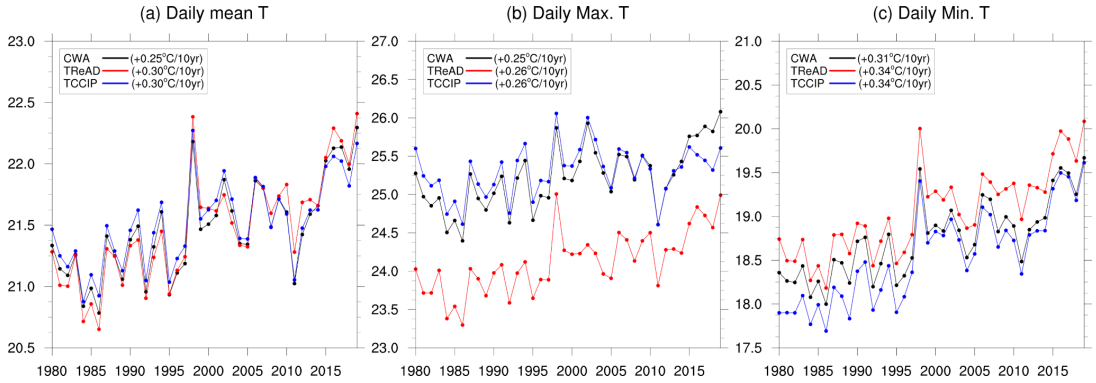



*Fig. Time series of daily mean, maximum and minimum temperature for the 1980-2019 period (th e original Figure 8)*

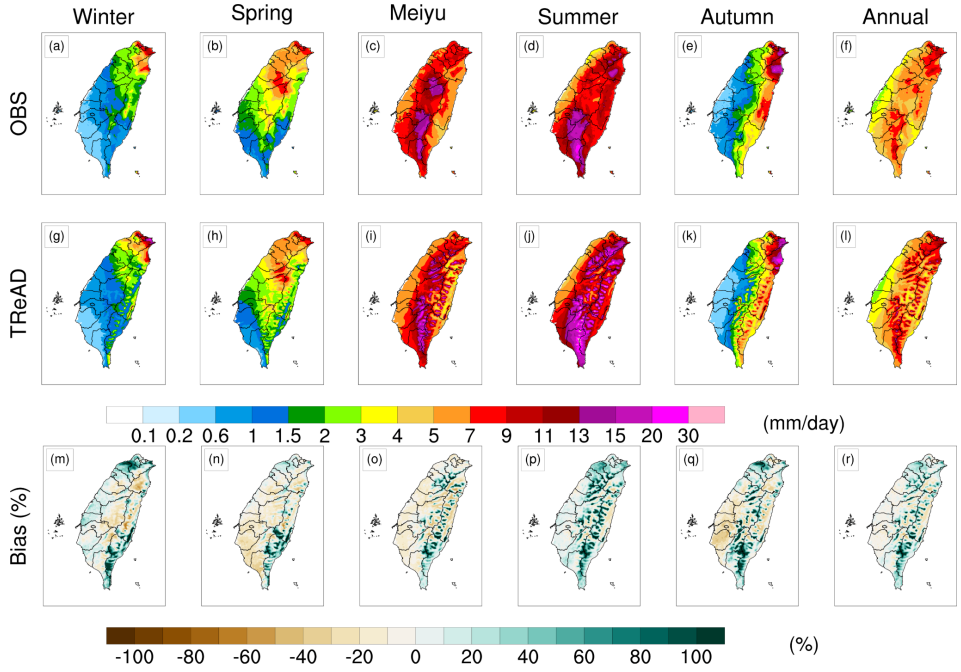



*Fig. Spatial comparison of long-term seasonal and annual mean of precipitation for the 1980-2019 period (the original Figure 11)*

3. The manuscript uses several datasets, including CWA observations, TCCIP data and Himawari-8-derived radiation data. Although the general sources are described, a concise summary table would make the validation framework considerably clearer. For each reference dataset, I suggest including: dataset name, variables used, spatial resolution, temporal resolution, period used in the present study, processing, interpolation, or elevation adjustment applied, purpose in the validation, relevant information on independence from ERA5/TReAD, where applicable.

*Thank you for this constructive suggestion to enhance the clarity of our validation framework. We address your comments below:*

*1. Summary table of reference datasets: We have created a summary table listing all reference observational datasets used for TReAD validation (CWA ground stations, TCCIP gridded observations, and Himawari-8 satellite-derived product). This table explicitly details dataset names, variables, spatial/temporal resolutions, time periods, processing/elevation adjustments, validation purposes, and independence from ERA5/TReAD. It has been incorporated into Supplementary Table S4 (see below).*

*2. Expanded descriptions in Section 2.3 (Observation Data for Evaluation): We have expanded discussion in the revised manuscript to incorporate required descriptions for both TCCIP gridded dataset and Himawari-8 satellite inferences:*

- *TCCIP Gridded Dataset: "Using the methodology developed by Weng and Yang (2018), TCCIP collected, digitized, and homogenized source-varying observational datasets before providing long-term, applicable, gridded daily temperature and precipitation datasets for the period from 1960 to 2023 (details of ground stations used refer to Hung, 2012; Tung et al., 2022; Kono et al., 2025). In addition to the aggregated 5-km horizontal resolution product, the 1-km high-resolution temperature and precipitation datasets were utilized in this study for spatial validation."*
- *Himawari-8 Satellite-Derived Shortwave Radiation: "Another key observational dataset is the shortwave radiation derived from the Himawari-8 geostationary satellite (Bessho et al., 2016), which features a spatial resolution of 0.01° (approx. 1-km). To evaluate simulated shortwave radiation in TReAD, both CWA station observations (available 1980 – 2019) and satellite-derived data (available 2011 – 2019) were utilized. Since shortwave radiation was not measured at certain stations during earlier years, only TReAD grid points corresponding to valid station observations were selected to ensure a fair comparison."*

*Table S4. Summary table of three reference datasets*

| *Dataset Name* | *CWA station Obs.* | *TCCIP gridded data* | *Himawari-8 satellite-derived product* |
|---|---|---|---|

| | | | |
|---|---|---|---|
| *Variables used* | *Temperature, Pressure, Relative Humidity, Wind, Rain, Shortwave radiation* | *Tmax, Tmin, Tmean, Rain* | *Shortwave radiation* |
| *Spatial Res.* | *Point* | *0.01°(1-km)* | *0.01°(1-km)* |
| *Temporal Res.* | *Daily* | *Daily* | *Daily* |
| *Study period* | *1980-2019* | *1980-2019* | *2011-2019* |
| *Processing / Adjustments* | *Pairwise exclusion; temperature lapse-rate (0.65 °C/100m); Hydrostatic pressure adj.* | *Homogenization & spatial interpolation (Weng & Yang 2018)* | *Clear-sky/cloud radiative retrieval algorithms (Bessho et al. 2016)* |
| *Validation Purpose* | *Baseline temporal trends, diurnal cycles & grid-to-point stats* | *Spatial pattern evaluation, regional precipitation biases* | *Fine-scale terrain shading & spatial solar irradiance mapping* |
| *ERA5/TReAD Independence* | *Fully Independent (Ground gauge network)* | *Fully Independent (Dense gauge interpolation)* | *Fully Independent (Top-of-atmosphere satellite)* |

4. The use of Typhoon Morakot as a precipitation case study is useful. However, please explain the rationale for selecting this event. For example, was it selected because of its severity, scientific importance, observational coverage or another reason? A broader multi-storm evaluation is not necessarily required. However, if TReAD has not been systematically evaluated for precipitation associated with typhoons having different tracks or intensities, this limitation should be acknowledged so that users do not overgeneralise the performance demonstrated for Morakot.

*Thank you for pointing out these important points regarding the case study selection rationale and the uncertainty range of typhoon precipitation performance. Typhoon Morakot (2009) was selected as the flagship precipitation case study primarily based on its extreme severity. Morakot was one of the most devastating typhoons in Taiwan's history, producing record-breaking accumulated rainfall over southern mountain regions within a three-day period. Testing TReAD against such an extreme event provides a stringent assessment of the dataset's capacity to reproduce ultra-heavy typhoon precipitation.*

*We agree with the Reviewer's caution about overgeneralizing TReAD's typhoon performance from*

*a single benchmark event. While Morakot demonstrates TReAD's robust capability during a slow-moving, topographically enhanced extreme rainfall system, TReAD precipitation performance has not been systematically evaluated across all tropical cyclones with varying tracks or intensities. Th erefore, we have acknowledged this as a methodological limitation in the newly added Section 4.2 (Data and Uncertainties, see below the added paragraph), advising users that typhoon rainfall accur acy may vary depending on storm speed, track, and synoptic environment. We have incorporated t he following paragraphs in the revised manuscript.*

*"To demonstrate its utility for climate impact and adaptation studies, we evaluated TReAD's abili ty to reproduce extreme precipitation during Typhoon Morakot in 2009 which is the one of the mo st devastating typhoons in Taiwan's history."*

*"Typhoon performance generalizability: Although TReAD demonstrates high spatial fidelity duri ng benchmark extreme events like Typhoon Morakot (2009), all typhoons across the entire 40-year period have not been systematically evaluated. Users analyzing typhoon related hydrology are enco uraged to conduct storm-specific evaluations when applying TReAD to other typhoons."*

Suggested but non-essential additions

1. The current validation is broad and already includes several metrics and comparisons. Never theless, the quantitative description of dataset performance could be further strengthened by on e or two additional diagnostics, such as:

- Mean Absolute Error
- Seasonal bias statistics, given the seasonal variations noted for temperature and radiation.
- Percentile-based error analyses, for example errors for extreme percentiles vs. moderate conditions.
- Probability distribution comparisons using quantile-quantile plots.

These are suggestions rather than mandatory requirements and I do not consider it necessary to include all of them.

*Thank you for this insightful suggestion to further strengthen the quantitative assessment of dataset performance. We highly appreciate your flexibility in allowing us to choose the most suitable addit ional diagnostics. To address your recommendation regarding Mean Absolute Error (MAE), percen tile-based distributions, and overall probability distribution (PDF) characteristics, we have added d edicated station-by-station boxplot analyses for each ECV using paired daily time series from TRe AD and CWA across the entire 1980 – 2019 period. Below are key values of the added boxplot fig ures.*

- *Percentile & PDF distribution alignment: Each box plot pairs a CWA ground station with i ts corresponding TReAD grid cell, visually depicting the median, interquartile range, and e*

*xtreme whiskers/outliers. This directly evaluates how well TReAD reproduces the observational PDF across both moderate and extreme regimes.*

- *Multi-station performance dispersion & MAE: The station-wise boxplots effectively showcase inter-station error dispersion and distributional spreads, providing a clear visual representation of magnitude-based error metrics such as MAE and interquartile biases across Taiwan's varied regions.*

*Here are examples of the renewed boxplot figures for temperature and precipitation.*

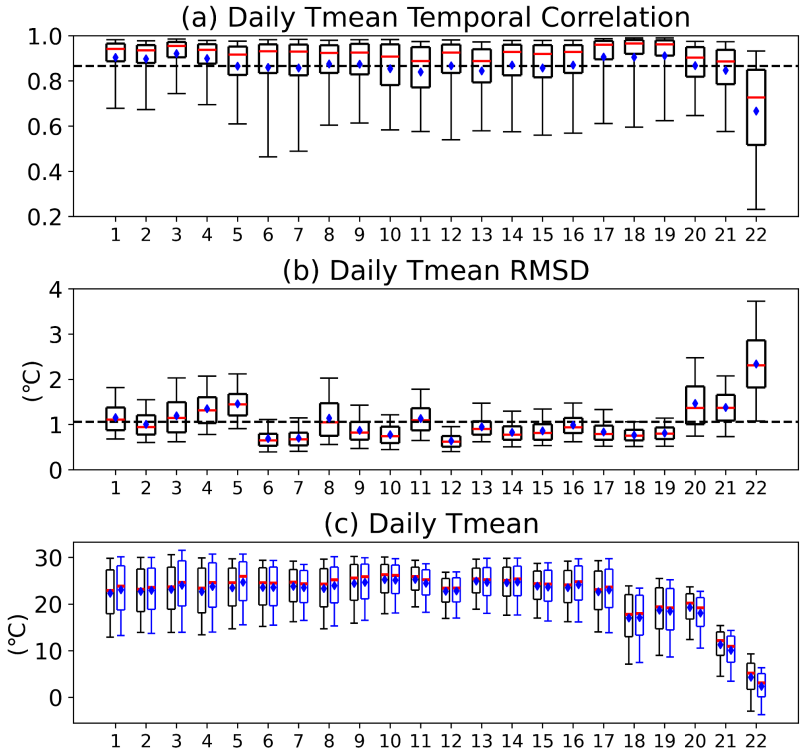



*Fig. (Renewed boxplot figure for temperature) The average (a) correlation coefficients and (b) RMSDs (℃) of daily temperatures between (c) the elevation-adjusted TReAD (blue boxes) and the station-based CWA data (black boxes) for the period from 1980 to 2019. The horizontal dashed lines represent the average of all stations and the blue points and red lines in each boxplot indicate means and medians, respectively.*

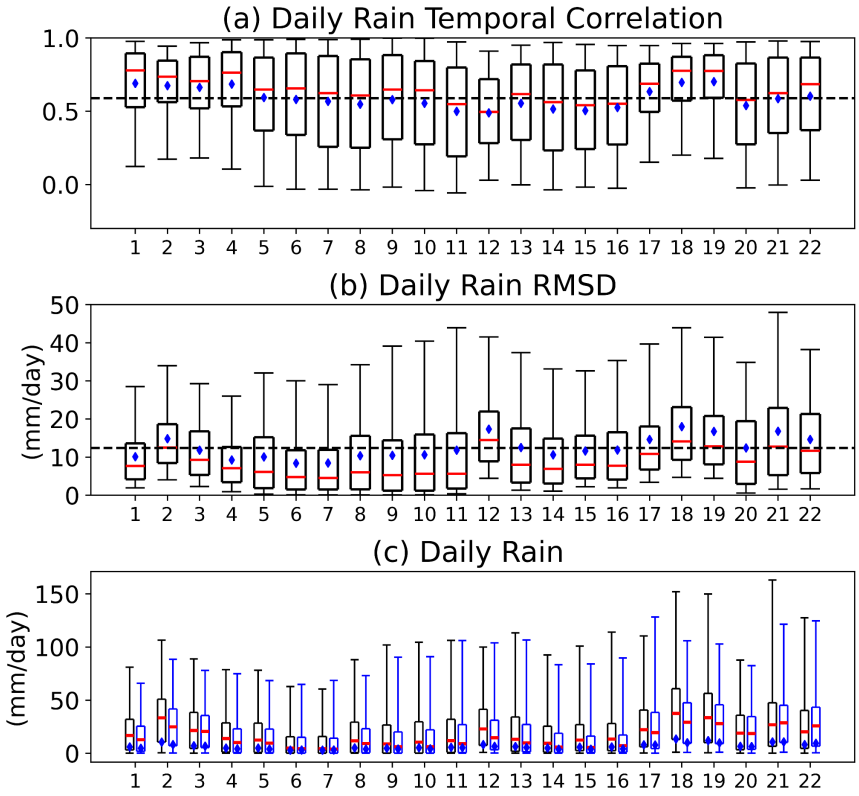



*Fig. (Renewed boxplot figure for precipitation) Same as the previous figure, but for daily precipita tion*

2. If available, a brief estimate of the computational resources and storage requirements associated with producing the dataset would be useful for researchers interested in reproducing or extending the production workflow.

*Thank you for this practical suggestion. We have provided in the supplementary material with an e stimate of the computational setup and resources used for the production workflow to assist researc hers interested in reproducing this dataset.*

*High-Performance Computing Infrastructure:*

- *System: Fujitsu HPC FX1000*
- *Processor: A64FX (48 cores @ 2.2 GHz per node)*
- *Resource Allocation: 8 compute nodes (totaling 384 CPU cores)*

*Computational Cost: It takes approximately 23.5 hours (nearly 1 full day) of continuous wall-clock time to complete a 1-month high-resolution simulation domain.*

Editorial Comments

- The manuscript requires careful English-language revision. Several sentences are unnecessarily long or awkwardly phrased. A thorough language revision is recommended, particularly in the Introduction and Conclusion.

*Thank you for this valuable feedback regarding the manuscript's readability and language quality. The entire manuscript has undergone a thorough language review to enhance the clarity of our rese arch.*

- Wind speed/velocity: Please clarify whether the manuscript refers to wind speed or wind velocity, as these terms have different scientific meanings. Use the appropriate term consistently throughout the manuscript, particularly in the Abstract and Section 3.5.

*Thank you for pointing out this important difference in physical terminology. Since our evaluation s, used metrics, time series, and spatial plots assess the scalar magnitude of wind movement, we ha ve systematically updated the manuscript to use "wind speed" throughout.*

- Please ensure that all figures include descriptive titles and appropriate units for RMSE and ot her relevant metrics. Use consistent font sizes across all panels and remove redundant phrasing from the figure captions.
- Fig. 1: Increase the font size of the legends and axis labels for better readability.
- Fig. 4: Improve readability by standardizing the font size to match that used in Fig. 3.
- Figs. 6, 10, 16, 20, 24: Increase the line thickness in these figures for better visual readabilit y.
- Fig. 7: Add units to the colorbar and include a descriptive title for the figure.
- Fig. 8: Verify and correct the time range shown (currently noted as 1980-2016) to align with the rest of the manuscript.
- Fig. 22: Remove the informal abbreviation "ie" from the figure or its label.
- Fig. 26: Change the y-axis label from "standardized deviations (Normalized)" to "Normalised standard deviation".

*Thank you for these detailed and constructive suggestions to enhance visual presentation which is e ssential for effective communication. In response to your comments, 23 out of 26 figures in the ma nuscript have been regenerated to ensure high graphical resolution, concise and standardized layou t design, and uniform styling across all panels.*

- *All regenerated figures contain titles (of panels), metric units (such as RMSD and anomali es) in labels/legends and captions.*
- *Font sizes for axis titles, axis labels, colorbars have been considerably increased across all panels to ensure effortless readability.*

- *Line width of the line plots across (merged) time series graphics (specifically the original Fig. 6, 10, 16, 20, and 24) has been increased to improve contrast and visual distinction.*
- *Colorbars of spatial bias and comparison maps have been refined for better distinction with physical units added.*
- *The time axis for the original Fig. 8 has been updated to consistently reflect the full period from 1980-2019.*
- *Fig. 22 and Fig. 26 have been updated per your suggestion.*

*Below is the updated Taylor diagram (the original Fig. 26).*

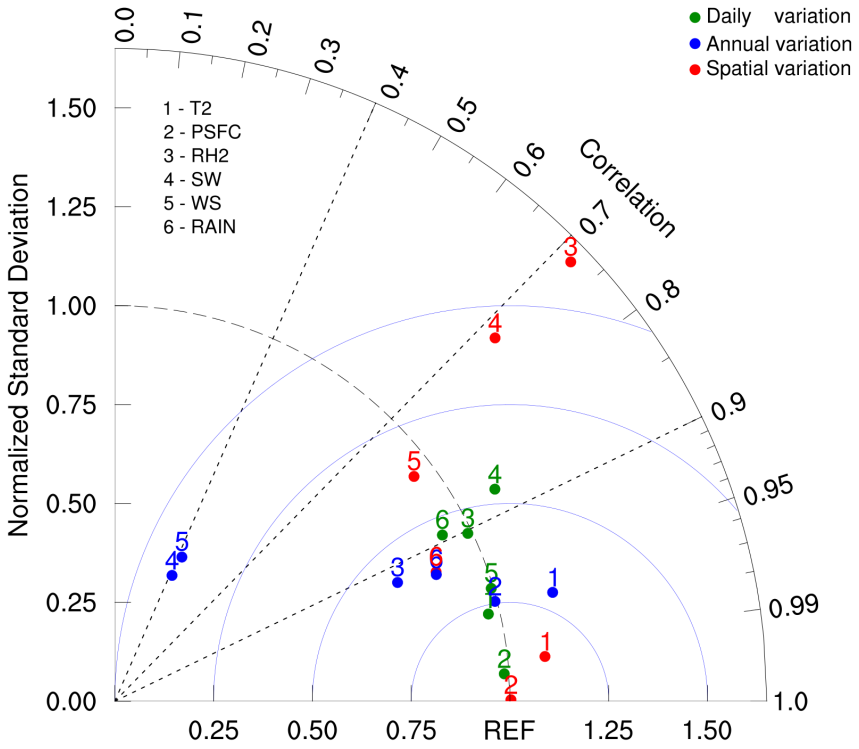



*Figure 26. Taylor diagram of the six ECVs namely 2-m air temperature (T2), atmospheric pressure (PSFC), RH (RH2), shortwave radiation (SW), wind speed (WS) and precipitation (RAIN) against CWA dataset. Red, blue, and green dots respectively indicate of spatial variation of climatology values, annual variation of 22-station averages, and daily variation of 22-station averages*

- L45: "Taiwan where its complexity of topography is high" for improved clarity. Rephrase, maybe to "Taiwan, where the topography is highly complex"

*Thank you for your comment. The sentence has been revised per your suggestion.*

- L66-67: Shorten the sentence "Such topography makes regional-scale climate analysis and application studies more difficult" for conciseness.

*Thank you for your comment. Please see below the original and refined sentence.*

*Original sentence:* *"Such topography makes regional-scale climate analysis and application studies more difficult."*
*Revised sentence:* *"Such topography complicates regional climate analysis and application."*

- L73, L74, L77, L78: Avoid the repeated use of "dynamically downscaling" in these lines.
- L81: Clarify the meaning of "aforementioned logics"; consider replacing it with a more precise term such as "methods" or "approaches".

*Thank you for your comments. Please see below the original and refined paragraphs.*

*Original paragraph:* *"The objective of this study is to reconstruct high-resolution historical climate data over Taiwan and the surrounding ocean via dynamically downscaling 40-year ERA5. Kanamitsu and Kanamaru (2007) produced a 10-km resolution dataset by dynamically downscaling the 57-year NCEP/NCAR reanalysis 1 using a regional climate model. Moreover, the results showed that the reconstructed data is of high quality and comparable to the regional analysis, NARR (Kanamaru and Kanamitsu, 2007). Similarly, Kayaba et al. (2016) used the 55-year JRA-55 to conduct dynamic downscaling and recreated a 5-km resolution dataset. As such, the dynamic downscaling has been proven to effectively produce historical data of climatic variables like temperature, rainfall, even extreme weather events for the regions with complicated terrain in particular. Therefore, we have followed the aforementioned logics and reconstructed a set of historical climate data over Taiwan at 2-km resolution for the period from 1980 to 2019"*

*Revised paragraph:* *"This study aims to reconstruct high-resolution historical climate data over Taiwan and adjacent oceans by dynamically downscaling 40 years of ERA5 reanalysis. Previous regional climate modeling by Kanamitsu and Kanamaru (2007) successfully refined the 57-year NCEP/NCAR reanalysis 1 into a 10-km dataset, yielding quality comparable to the North American Regional Reanalysis (NARR; Kanamaru and Kanamitsu, 2007). Similarly, Kayaba et al. (2016) regionalized the 55-year JRA-55 to generate a 5-km climate product. These efforts demonstrate that high-resolution regional downscaling effectively captures temperature, precipitation, and extreme events, particularly across complex terrain. Building on this framework, we produced a 2-km historical climate dataset for Taiwan spanning 1980 to 2019."*

- L91: Specify the location of the two nested domains within Figure 2, or ensure they are clearly identifiable in the figure itself.

*Thank you for your comment regarding domain configuration. Fig. 2 and its caption have been completely revised. Explicit latitude and longitude coordinate axes have been added. In addition, the*

*boundary frame of the inner 2-km domain has been remarkably thickened and distinctly labeled. S ee below the updated Figure 2 and its revised caption.*

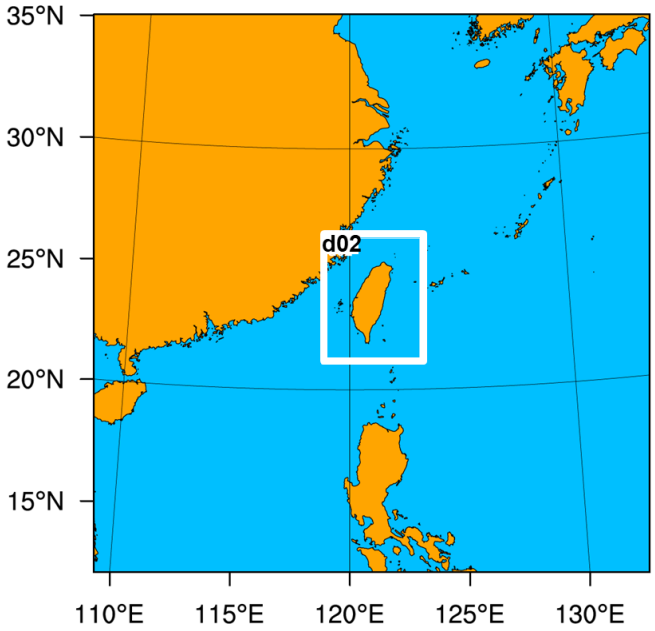



*Figure 2. Spatial configuration of the two nested WRF model domains used in TReAD downscalin g, including latitude and longitude coordinates. Domain 1 (10-km resolution, outer box) covers the regional large-scale circulation, while the inner Domain 2 (2-km resolution, highlighted by the bol d boundary frame) covers Taiwan and adjacent maritime areas.*

- L120 and L356: Define the abbreviation "UHI" at its first mention in L120 (where "heat islan d effect" appears), rather than introducing it later in L356.

*Thank you for your comment. Per your suggestion, the abbreviation "UHI" has been defined at its first mention and UHI is then used in the subsequent sections.*

- L164-165: "(see Table S1 for the details of the 22 stations including longitude, latitude and elevations)" : this parenthetical is redundant because Table S1 already provides those details. Consider shortening it to "(Table S1)" or "(see Table S1)".
- L166: "It is worthy mentioning" change to "It is worth mentioning".
- L224-225: Correct the figure reference, the "mean biases" discussed here refer to Fig. 7, not Fig. 5.

*Thank you for your comments. The above suggestions have been incorporated in the revised manu script.*

- L231-232: “when evaluating … separately (not shown here)”. Could the authors include these results for completeness? Alternatively, this statement could be removed if the results are not essential to the discussion.
- L249: Remove the qualifier "relatively weak" when describing the overestimation, as this characterization is not supported by the data.

*Thank you for your comments. The statement and “relatively weak” have been removed.*

- L313-314:” to be better … in 2009”: Reformulate, make it shorter ans clearer.

*Thank you for your comments. Please see below the original and refined sentences.*

***Original sentence:*** *“To be better applied to climate change impact and adaptation fields, the capability of reproducing extreme precipitation events is important and has been evaluated by assessing the typhoon Morakot that resulted in severe flooding events across Taiwan in 2009.”*
***Revised sentence:*** *“To demonstrate its utility for climate impact and adaptation studies, we evaluated TReAD's ability to reproduce extreme precipitation during Typhoon Morakot in 2009 which is the one of the most devastating typhoons in Taiwan’s history.”*

- L396-397: Rephrase the statement “since radiation … weather systems” and add appropriate supporting references.

*Thank you for your comments. Please see below the original and refined sentences with a supporting reference added.*

***Original sentence:*** *“Since radiation is significantly affected by the cloud-related weather systems (e.g., convective precipitation or typhoons), the simulated daily cumulative radiation has been further assessed by comparing it with CWA data sets (both station-based and Himawari-8 derived) for four observation stations (i.e., Taipei, Taichung, Dawu, and Kaohsiung) as shown in Figure 18.”*
***Revised sentence:*** *“As cloud processes and convective systems drive surface radiation variability (Stephens, 2005), the simulated daily cumulative radiation was further evaluated against CWA ground station and Himawari-8 satellite datasets across four key stations (Taipei, Taichung, Dawu, and Kaohsiung; Figure 18).”*

*Here is the newly added reference:*
*Stephens, G. L. (2005). Cloud feedbacks in the climate system: A critical review. Journal of climate, 18(2), 237-273. https://doi.org/10.1175/JCLI-3243.1*

- L470-473: Rephrase the entire passage for improved clarity and conciseness.

*Thank you for your comments. Please see below the original and refined sentences.*

***Original sentence:*** *"Relative humidity is another important variable for climate change impact st udy. In addition to the aforementioned UHI field, the research on the impact of climate change on animal husbandry also highlights the importance of relative humidity e.g. the Temperature Humidit y Index is frequently used as a hazard indicator for cows' industry (Liu et al., 2018)."*
***Revised sentence:*** *"Relative humidity is essential for climate impact assessments. In addition to urban heat studies, livestock research uses humidity-dependent metrics, such as the Temperature-H umidity Index (THI), to evaluate heat stress risks in dairy production (Liu et al., 2018)."*

- L480: "show varying results": change to "show varying performance".
- L582: "with cautions": change to "with caution".

*Thank you for your comments. The above suggestions have been incorporated in the revised manu script.*