Authors' response regarding "The radiative forcing of $PM_{2.5}$ heavy pollution, its influencing factors and importance to precipitation during 2014-2023 in the Bohai Rim, China" by Zhu et al.

We appreciate the reviewer's comments and suggestions. The manuscript has been revised accordingly. Our point-by-point responses to the comments are presented below. The comments are in black, followed by responses in blue and revised manuscript in blue marked by underline.

Response to Referee Comment #1

Authors present research on an interesting topic: the influence of meteorological parameters and aerosol concentrations measured at ground level on radiative forcing. They use up-to-date methods and data sources; however, in my opinion, the manuscript in its current form is not suitable for publication, and some points should be explained, changed, or even require additional data analysis.

Response: We are grateful to you for your time, insightful comments, and constructive suggestions, which have been invaluable in helping us to significantly improve the quality and clarity of our work. We have carefully considered each point raised and have made revisions to the manuscript accordingly.

1. Moreover, the way the authors present their findings is chaotic and sometimes hard to follow. The figures are sometimes unreadable, captions are too small, etc. There is probably too much data presented in the manuscript; maybe moving some data to an appendix would help and allow readers to focus on the main scope of the manuscript.

Response: ①To present our work more clearly, we added a Methodology section in Section 2 as "2.4 Methodology", incorporating "2.4.1 k-means clustering method" and "2.4.2 Importance estimation based on Random Forest algorithms" in this section. And a new subsection entitled "2.4.3 The data usage workflow" has also been added after that. This section specifically addresses ambiguities and unclear points regarding methods in the original text, and includes an additional figure: Figure 1. The data usage workflow and framework in this work. The figure clearly illustrates the data, methods, and objectives employed in this work, along with the data and methods used for each specific objective. Besides, we added the introduction of the data usage workflow and framework as "Figure 2 exhibited the data usage workflow and framework in this work. The CMEE $PM_{2.5}$ data at sites and CERES net radiation flux data were used to show the characteristics of $PM_{2.5}$ and RF on regional heavy pollution days (RHPD) in the Bohai Rim.….

Subsequently, CMEE, CERES, and ERA5 data were combined to analyze the impact of meteorological profile types (temperature and winds) on surface $PM_{2.5}$ and RF on regional heavy pollution days by using the k-means clustering method at the stations.….

To explore the relative importance of factors to pollution RF and precipitation on regional heavy pollution days in the Bohai regions, the gridded data of CERES, ERA-5, and TAP were jointly applied to the random forest algorithm.…"

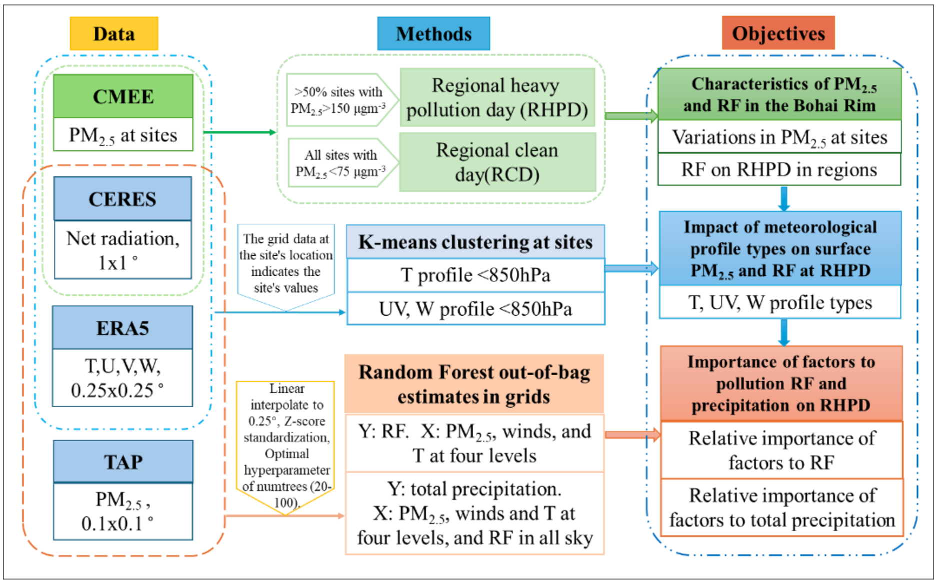

Figure 1. The data usage workflow and framework in this work. (Figure 2 in revised manuscript).

②Regarding your concerns about image clarity and excessive data presentation, we have revised Figures 5 and 6 in the original manuscript. We have enlarged the text within the figures and condensed the original Figures 5 and 6 to include only wind classification, $PM_{2.5}$, and surface RF subplots (as shown below: Figures 2 and 3). The remaining sections of the original text (percentage statistics at stations, and radiative forcing (RF) at the top and in the atmosphere) have been relocated to supplementary file (Figure S6 and Figure S7). This approach enables readers to focus on the core content and objectives of the manuscript.

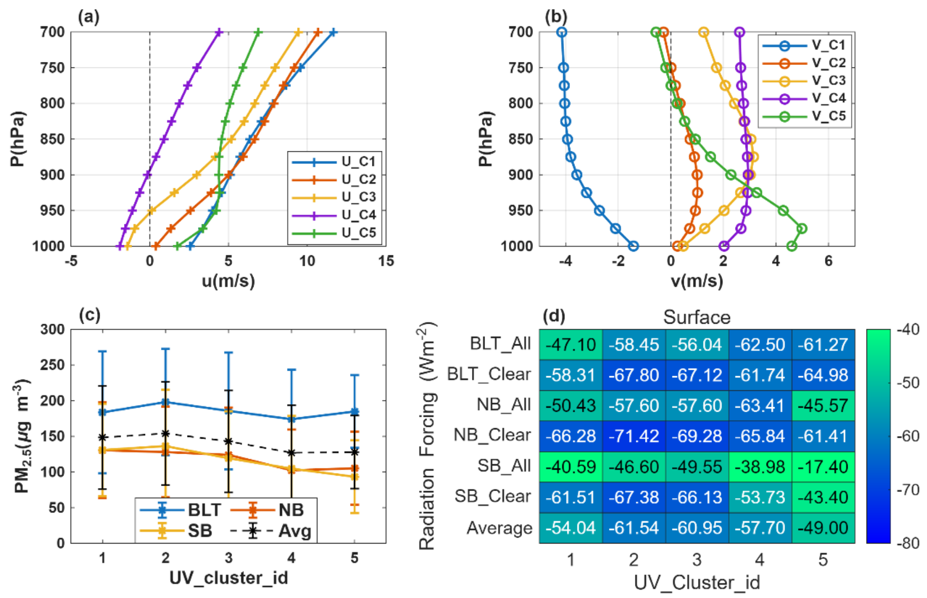

Figure 2. The five horizontal wind profiles under 850 hPa by k-means clustering on the regional heavy pollution days: (a) u wind, (b) v wind, (c) the $PM_{2.5}$ concentrations, and (d) the radiative forcing in the three regions in clear- and all-sky at the surface at the five clusters. (the Figure 5 in origin version and Figure 6 in revised version).

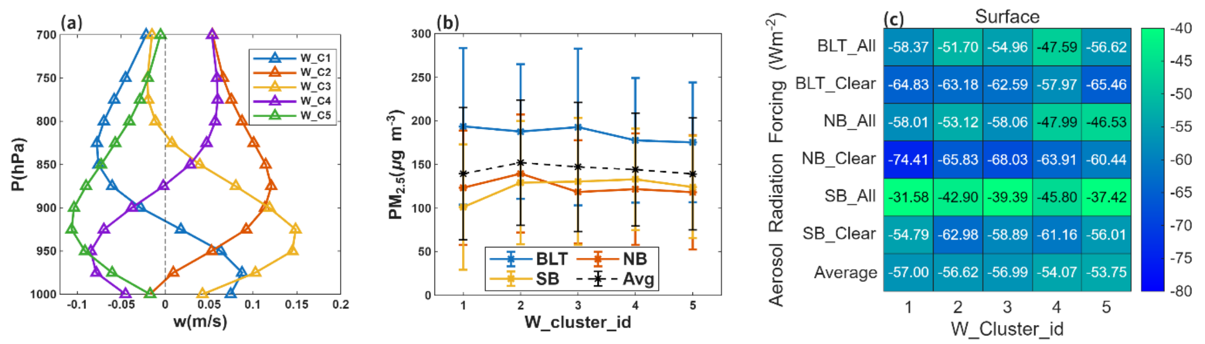

Figure 3. The five vertical wind profiles under 850 hPa by k-means clustering on the regional heavy pollution days: (a) the w wind, (b) the $PM_{2.5}$ concentrations, and (c) the radiative forcing in the three regions in clear- and all-sky at the surface. (the Figure 6 in origin version and Figure 7 in revised version).

2. The main issue of the manuscript, in my opinion, is how the authors "compare" in-situ data with satellite measurements. They use multisource data, which is fine; however, the relationship between them and the transition from in-situ measurement data to gridded data is unclear.

Response: This comment helped us recognize that the original manuscript lacked sufficient explanation regarding the comparison between ground-based observational data and satellite measurement data. To address this, we have expanded "2.4.3 the data usage workflow" section within Section 2, and explicitly describe the comparison between in-situ and grid data as "…CMEE, CERES, and ERA5 data were combined to analyze the impact of meteorological profile types (temperature and winds) on surface $PM_{2.5}$ and RF on regional heavy pollution days by using the k-means clustering method at the stations. The RF and meteorological parameters at stations were extracted from grid data of CERES and ERA5, respectively. The extracting method is that the grid data at the site's location indicates the site's values….". For Section 3.2, site-based $PM_{2.5}$ data (derived from in situ observations) were employed. Radiative forcing data were calculated using satellite CERES data from the grid cell containing the site. Vertical meteorological profiles were sourced from ERA5 data corresponding to the grid cell at the site location. For Section 3.3, all data employed are grid-based. A random forest importance analysis was conducted using grid-based TAP, ERA5, and CERES data; this section did not utilize site-based observational data.

3. For instance, they use a monitoring network to identify high-concentration episodes, which is acceptable. Then, some gridded data are used in statistical analysis. What happens between in-situ and gridded data? How does TAP work? A brief description is needed in the manuscript. By the way, why do the authors use ERA-5 data while the TAP website claims that meteorological data are combined with aerosol data? If ERA-5 data are better, then what is the quality of TAP aerosol data? The authors use PM2.5 data as a predictor in the Random Forest model. They claim it is a proxy for anthropogenic sources; however, elsewhere they discuss the influence of transport on local concentrations, and in another place, they state that columnar optical properties determine radiative forcing. So, is local PM2.5 a predictor of radiative forcing or not?

Maybe it would be better to use emission inventories—TAP claims that one is incorporated—to estimate anthropogenic sources instead of PM2.5 concentrations? Moreover, in section 3.3, the authors again claim that PM2.5 is related to emissions while meteorological profiles reflect diffusion. I cannot agree: emissions together with diffusion factors influence PM2.5 concentrations. So, PM2.5 is not an independent variable, contrary to what is stated in the conclusion.

Response: ①Regarding the insufficient description of data usage and the handling of in-situ and grid data, we have already outlined our revisions in the first two responses and will not elaborate further here.

②For the TAP data, Figure 4 shows the operational process of the near real-time $PM_{2.5}$ data generated from TAP (Geng et al., 2021). In the revised version, additional descriptions have been included to demonstrate the quality of the data, as follows: "The TAP $PM_{2.5}$ is estimated based on a two-stage machine learning model coupled with the synthetic minority oversampling technique and a tree-based gap-filling method, which improves the $PM_{2.5}$ estimations on highly polluted days. The TAP $PM_{2.5}$ showed a higher regression slope (0.97) when evaluated against ground measurements (Geng et al., 2021).".

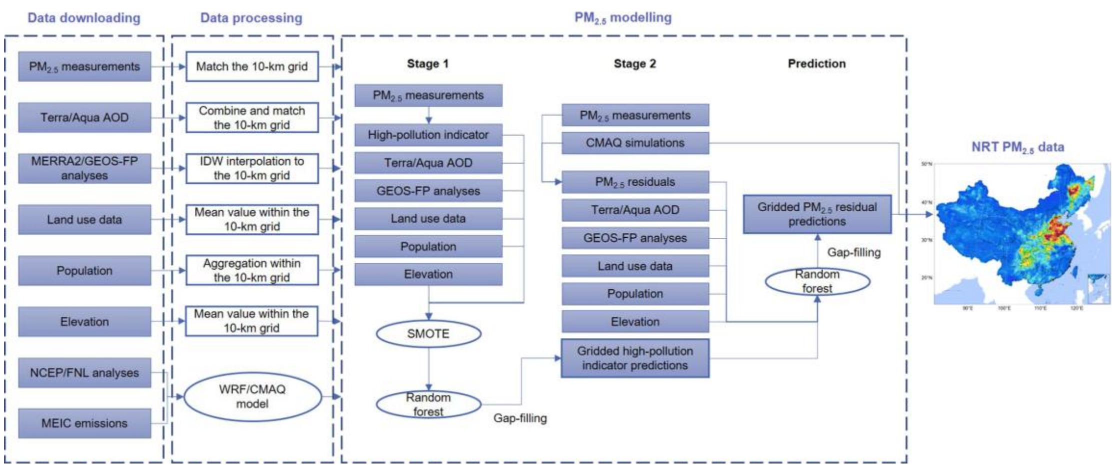

Figure 4. Operational process of the near real-time $PM_{2.5}$ data generated from TAP (Geng et al., 2021).

③Regarding the use of TAP $PM_{2.5}$ as a predictive variable, we are most grateful for your identification of this critical issue. We recognize that the original manuscript contained inconsistencies and imprecision in its discussion of $PM_{2.5}$. Our original intent was to examine the contribution of surface $PM_{2.5}$ to RF by the Random Forest model, rather than treating it as an independent emission source. Consequently, characterizing it as an approximate emission source in the original text was indeed imprecise. In the revised manuscript, we have modified this section's wording, removed the statement that it represents anthropogenic emission sources and replaced it with "surface aerosol concentrations" (our work focuses on regional heavy pollution days with $PM_{2.5}$>150 $\mu g m^{-3}$). The reason we did not employ emission inventories is that current inventories typically lack sufficient spatial-temporal resolution. In contrast, TAP $PM_{2.5}$ offers high spatiotemporal resolution (near real-time daily full-coverage $PM_{2.5}$ data at a spatial resolution of 10 km). Besides, it performs well under conditions of high pollution (a higher regression slope of 0.97 when evaluated against ground measurements in Beijing-Tianjin-Hebei (Geng et al., 2021)), and thus better reflects the actual amount of radiation-related surface aerosols formed. Furthermore, we explicitly recognize the limitations of TAP $PM_{2.5}$ being influenced by atmospheric diffusion transport in the discussion section of the paper as "TAP $PM_{2.5}$ data contains the meteorological information, and some other factors have not been taken into account in the machine learning such as terrain elevation.".

Reference: Geng, Guannan, Qingyang Xiao, Shigan Liu, et al. 2021. 'Tracking Air Pollution in China: Near Real-Time PM2.5 Retrievals from Multisource Data Fusion'. *Environmental Science & Technology* 55 (17): 12106–15. https://doi.org/10.1021/acs.est.1c01863.

4. Another example concerns the "mean profiles" of meteorological parameters. It is written that the authors used profiles at over 11 stations. How are they representative for the 0.25 × 0.25 grid used in the Random Forest analysis? Is there any local orography favoring aerosol transport or accumulation in valleys? What about sea-land differences? I can understand that clustering is performed over land (land stations). It seems that clustering and Random Forest are independent. So, it should be explained somewhere why the authors perform such investigations.

Response: This opinion has provided us with inspiration and deeper insight.

①As previously addressed in our earlier response, clustering and random forests are distinct and independent methodologies. Cluster analysis focuses on station-based investigations, primarily examining how variations in temperature and wind profiles within the boundary layer (below 850 hPa) influence $PM_{2.5}$ concentrations and RF, all conducted at land-based observation stations. Random forests, however, represent a regional analysis, aiming to investigate the relative importance of $PM_{2.5}$ and temperature and wind at various altitudes (1000, 850, 700, and 500 hPa) for RF. To address this, we have incorporated relevant clarifications in the revised manuscript, such as the data utilization workflow in the Methods section in the above response, and added the rationale for this analysis in the first paragraph of Section 3.3 as "…. Additionally, the above impact was based on stational statistics. Therefore, our analysis in this section was conducted at all the grids within the Bohai regions….".

②Regarding the influence of topography and land-ocean differences, it is certain that these factors affect both $PM_{2.5}$ and radiative forcing. Regarding topography, Figure 1 in the manuscript illustrates the terrain characteristics within the study area. The northwestern in study area is mountainous, where southerly or easterly winds favors pollution accumulation in land, while westerly winds facilitate pollutant transport from land to the ocean. In the regional variation analysis, we added the discussion of spatial heterogeneity related to driving factors such as aerosol concentrations, RF values, local topography and geographical location in section 3.2.1 as "The importance of influencing factors to pollution RF showed heterogeneity across different regions. This may be related to variations in aerosol concentrations and pollution RF. In clear sky, the BLT region featured high $PM_{2.5}$ concentrations (Figure 3) coupled with high RF values (Figure 4a), whereas the SB region exhibited low $PM_{2.5}$ concentrations alongside low RF, illustrating the higher importance of $PM_{2.5}$ impacts on RF at the surface in these two regions (Figure 8a). Conversely, the NB region, with low $PM_{2.5}$ concentrations but high RF, showed less importance of $PM_{2.5}$ than in the BLT and SB. Differences in the importance of meteorological parameters across regions may be related to local topography and geographical location. Figure 1 indicates that the northwestern parts of the BLT and NB regions border mountainous terrain, where wind direction and speed significantly influence pollution dispersion and transport. Consequently, wind (V-wind in the BLT region, U- and V-winds in the NB region) exhibited greater importance for RF in these regions compared to the SB region. Conversely, the SB region's lower latitude and absence of nearby high mountain ranges resulted in temperature factors being more important than in the BLT and NB regions.”.

③ Regarding the land-sea differences, we demonstrated the variations in influencing factors between land and ocean at the end of Section 3.3.1. In the revised manuscript, we have expanded the analysis of influencing factors related to land-ocean differences, as “The relative importance of influencing factors between land and ocean also exhibited certain variations. Regarding the surface and atmosphere RF, the significance of $PM_{2.5}$ over land surpassed that over the ocean, while the importance of U and V over the ocean exceeded that over land areas (U500 in clear-sky, V1000 in all-sky). Given that land surfaces exhibited higher aerosol concentrations than oceans, the prominence of $PM_{2.5}$ over land was entirely justified. The heightened significance of U and V over the ocean was primarily linked to transport from land to the ocean. Based on the land-ocean distribution in Figure 1, it is evident that zonal winds (particularly westerlies) facilitate the transport of land-based pollutants towards the ocean, while meridional winds also promote this transport (both southerly and northerly winds). Conversely, transport from the ocean to land also occurred, though its impact was comparatively minor relative to the influence of high-concentration aerosols over the land.”.

5. Another question is: what are the profiles during the rest of the analyzed time, not only during high-concentration episodes? For example, temperature profile clusters 2 to 4 exhibit inversion from around 950 to 850 hPa. The frequency of occurrence during the investigated episodes is around 30–40%. What happens during the rest of the investigated period (autumn, winter)? Another issue is how “wind clusters” are presented. Figure 5a is completely unreadable. Maybe the authors will find another way to present changes with the altitude of wind speed and direction?

Response: ①As for this question, we compared the average meteorological profiles during the study period between all days, regional clean days, and regional heavy pollution days across the region (see Figure 5 below). It can be observed that the average profiles showed subtle differences between heavy polluted and clean days, failing to reflect the presence of temperature inversions. Consequently, we clustered temperature profiles from all stations during autumn and winter into also five categories (decreasing, decreasing with isotherms, low-level inversion, mid-level inversion, and upper-level inversion). We then calculated the proportion of each profile type across all days, regional clean days, and regional heavy pollution days, as shown in Figure 6. The comparison clearly demonstrates that the proportion of inversion conditions on regional heavy pollution days is significantly higher than that on regional clean days, and also exceeds the proportion across all days, indicating the crucial role of inversions during heavy pollution episodes. This section has been placed in the supplementary file (Figure S5) and we added the comparison in section 3.2.1 as "The comparison of proportions of different temperature profile types on all days, regional clean days, and regional heavy pollution days in the study period (Figure S5) showed clearly that the proportion of temperature inversions occurring on regional heavy pollution days was significantly higher than that on regional clean days, and also exceeded the proportion across all days, indicating the significant role of temperature inversions on regional heavy pollution days." .

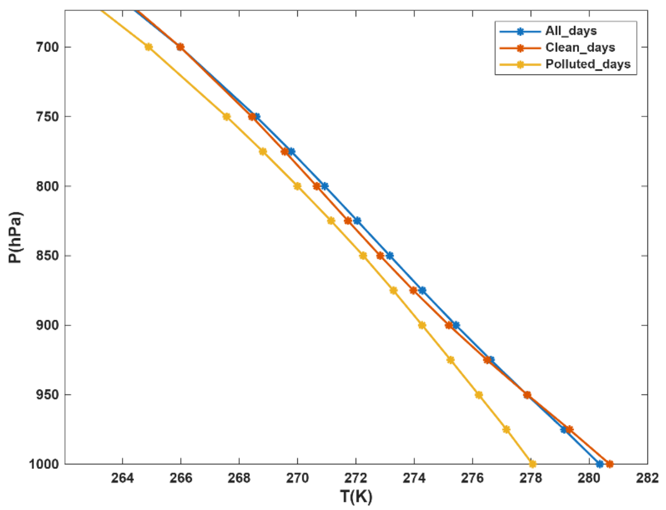

Figure 5. Average temperature profiles for all days, regional clean days, and regional heavy pollution days during autumn and winter at monitoring stations within the study region.

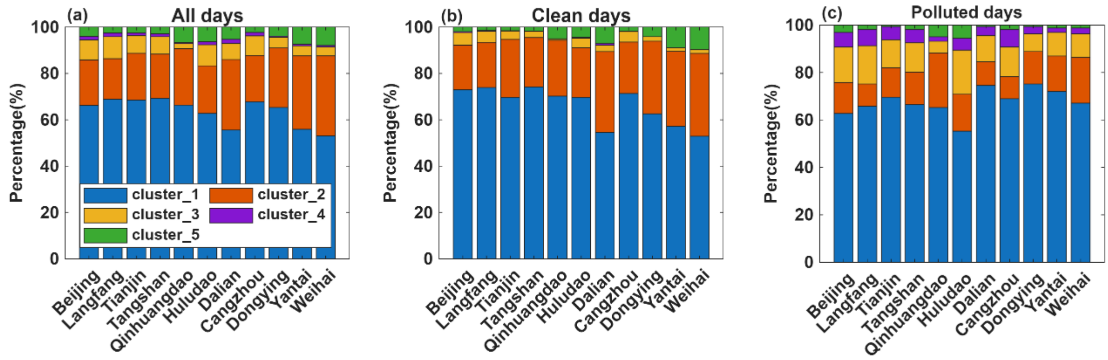

Figure 6. Proportions of different temperature profile types (C1: decreasing, C2: decreasing with isotherms, C3: low-altitude inversion, C4: mid-altitude inversion, and C5: upper-altitude inversion) on all days (a), regional clean days (b), and regional heavy polluted days (c) at each station during autumn and winter 2014-2023.

②To address the presentation of wind clustering, we have redrawn the figures, separating the u and v components in Figure 5a to enhance clarity (Figure 2 in this response document). The frequency statistics and RF_TOA and RF_Atmos subplots have been placed in the supplementary file (Figure S6 in the new supplementary file).

6. Regarding mean clustered profiles, it would be interesting to connect temperature profiles with wind profiles. I suggest performing multidimensional cluster analysis to find meteorological situations favoring large PM2.5 concentrations—for example, inversion and low wind speed near the ground.

Response: We agree with your proposal. The k-means clustering method we employed is a multidimensional clustering technique characterized by high computational efficiency. Following your recommendation, we applied k-means clustering to the combined temperature and wind field profiles, utilizing the elbow method to determine the number of cluster (2 clusters) (results shown in Figure 7) or the same 5 clusters as in the original manuscript (results shown in Figure 8). We can find that these classification approaches failed to capture variations in the height of temperature inversion layers (our attempts with clusters 2–10 all proved incapable of reflecting inversion layer height changes). It also cannot distinguish which types of conditions favors increased $PM_{2.5}$ concentrations. Furthermore, this combined clustering fails to reflect the relative importance of temperature and wind profile variations on $PM_{2.5}$ and

RF. Therefore, although clustering temperature and wind profiles together is meaningful, this combined approach struggles to identify meteorological conditions conducive to elevated $PM_{2.5}$ concentrations (such as temperature inversions and low near-surface wind speeds). Consequently, our work employed separate clustering methods for temperature and wind profiles. The results of this joint clustering analysis have been added to the supplementary file and were described in the Methods section 2.4.3 of the article, as "For the k-means clustering, we clustered the T, horizontal, and vertical winds separately to capture variations in the height of temperature inversion layers and to reveal the relative importance of temperature and wind profile variations below 850 hPa on $PM_{2.5}$ and RF. The clustering of combined T and winds by the k-means method failed to capture the temperature inversion layers, and could not reveal the relative importance (see Figure S1 and S2).".

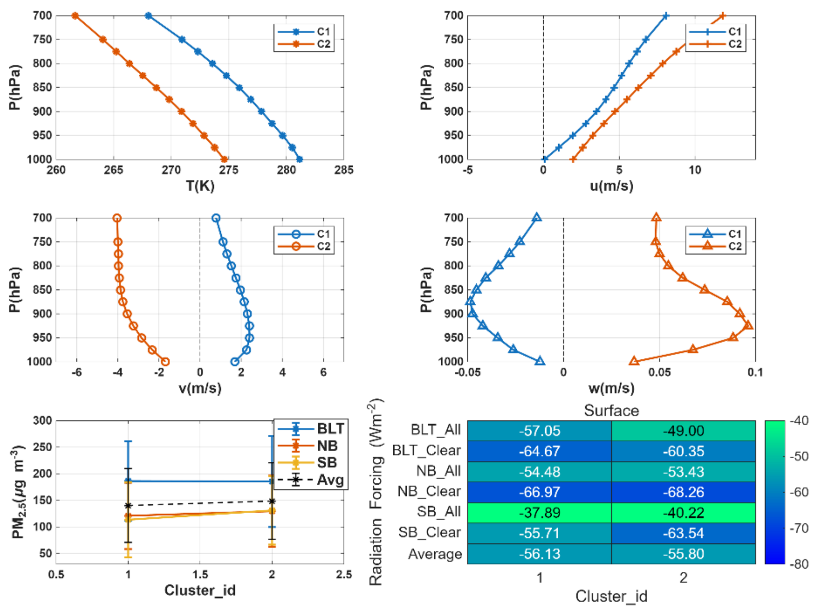

Figure 7. The results of clustering of combined T and winds by the k-means algorithm using the elbow method to determine the number of clusters of 2 (this clustering of 2 is distinct for temperature and winds, but failed to capture the temperature inversion layers). (Figure S1 in the supplementary file)

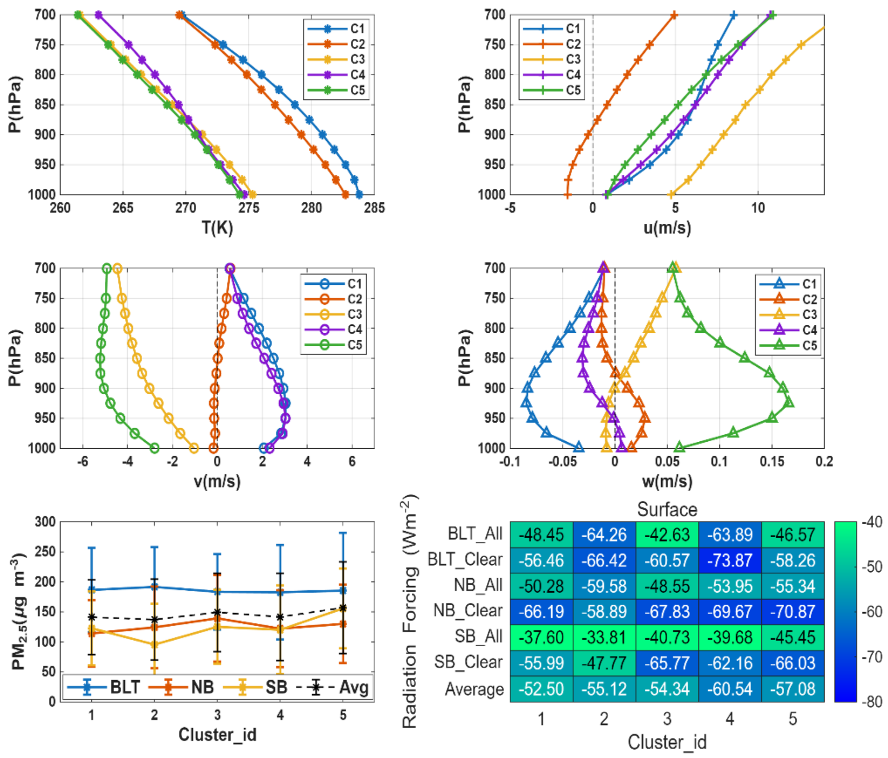

Figure 8. The results of clustering of combined T and winds by the k-means algorithm using the same number of clusters in the manuscript of 5 (this clustering of 5 also failed to capture the temperature inversion layers). (Figure S2 in the supplementary file)

7. One major weakness of the manuscript is the insufficient discussion of the Random Forest analysis. The authors should elaborate on why individual parameters influence radiative forcing, separately for the clear-sky and all-sky cases. What are the potential mechanisms? What is the influence of clouds? Furthermore, the land-sea aspect requires a more in-depth analysis, particularly regarding aerosol transport between sea and land—for example, the influence of wind speed and direction.

Response: In accordance with the proposed suggestions, we have expanded the analysis and discussion concerning the random forest results in section 3.3.1. The additional content includes:

① The mechanisms underlying differences in influencing factors between clear-sky and all-sky conditions. In clear-sky, the direct radiative effect of high-concentration surface aerosols dominates the radiative forcing at the surface and in the atmosphere. In all-sky, cloud radiative effects become significant, with cloud-radiation interactions influenced jointly by aerosols and meteorological conditions, where meteorological factors may play a more prominent role. The added content is as follows: "The importance of factors to the pollution RF in clear-sky and all-sky showed a distinct difference: Surface $PM_{2.5}$ was the most important factor to surface and atmosphere RF in clear-sky, but V wind in high level (500 hPa) in all-sky. In clear-sky, high-concentration surface aerosols directly interact with shortwave radiation (through scattering and absorption), becoming the most important factors of pollution RF at the surface and in the atmosphere. Other meteorological factors may indirectly influence RF by altering aerosol distribution (including horizontal and vertical distribution) and aerosol chemistry formation to affect the RF indirectly, but their effect is far less pronounced than the direct interaction between high-concentration surface aerosols and radiation. In the all-sky, however, the pollution RF is primarily influenced by cloud-radiation interactions. Cloud radiative forcing is affected by cloud characteristics (cloud cover, cloud height, cloud type, etc.), which in turn are influenced by both aerosols and meteorological conditions, with the latter potentially playing a more significant role. The critical role of upper-level (500 hPa) V-wind may be attributed to the strong association between weather systems over the Bohai Rim region and upper-level zonal winds.".

② Analysis of differences in influencing factors between land and sea. These differences primarily stem from the high importance of $PM_{2.5}$ over land and the relatively high importance of u and v factors over the ocean. We have linked these differences in influencing factors to pollutant transport between land and sea. The added content is as follows: "The relative importance of influencing factors between land and ocean also exhibited certain variations. Regarding the surface and atmosphere RF, the significance of $PM_{2.5}$ over land surpassed that over the ocean, while the importance of U and V over the ocean exceeded that over land areas (U500 in clear-sky, V1000 in all-sky). Given that land surfaces exhibited higher aerosol concentrations than oceans, the prominence of $PM_{2.5}$ over land was entirely justified. The heightened significance of U and V over the ocean was primarily linked to transport from land to the ocean. Based on the land-ocean distribution in Figure 1, it is evident that zonal winds (particularly westerlies) facilitate the transport of land-based pollutants towards the ocean, while meridional winds also promote this transport (both southerly and northerly winds). Conversely, transport from the ocean to land also occurred, though its impact was comparatively minor relative to the influence of high-concentration aerosols over the land.”.

Response to Referee Comment #2

This study integrates ground observations, satellite remote sensing, reanalysis data, and machine learning techniques to systematically analyze the radiative forcing (RF) characteristics, influencing factors, and their relative importance to precipitation during severe autumn and winter PM2.5 heavy pollution events in the Bohai Sea region (2014–2023). While the work is meaningful, the following issues need to be addressed:

Response: We sincerely appreciate your valuable comments and thoughtful suggestions for providing us with important guidance for improving the manuscript. We have carefully addressed each of the issues you raised below with the following amendments.

1. Page 6, Line 135: The term "clean days" is used without a clear definition. Please specify the exact PM2.5 concentration threshold or criteria used to classify a day as "clean".

Response: Thank you for your suggestion. We have clarified the regional clean day in the section of data and method in the modified version: "The regional clean day is defined as the day with all stations $PM_{2.5}$ < 75 μgm^{-3} within the study area".

2. Although machine learning algorithms are central to this study, the manuscript lacks quantitative performance evaluation. Please provide essential statistical metrics, such as the coefficient of determination (R^2), RMSE, and MAE, for both the radiative forcing and precipitation prediction models to validate their accuracy.

Response: We have incorporated your recommendation by adding an explanation of the statistical metrics for the machine learning models concerning radiative forcing and total daily precipitation in the section "2.4.2 Importance estimation based on Random Forest algorithms" in the revised manuscript as "The coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the model performance, which was shown in Table S2.". Additionally, we have included the following Table 1 (Table S2 in the new supplementary file) in the supplementary materials:

Table 1. The essential statistical metrics of the evaluation of Random Forest training models.

Statistical metrics	RF surface clear	RF surface all	RF TOA clear	RF TOA all	RF Atmos clear	RF Atmos all	Total pre
R^2	0.6721	0.6722	0.6852	0.6512	0.6557	0.6742	0.5728
RMSE	0.5697	0.5697	0.5562	0.5881	0.5837	0.5683	0.6454
MAE	0.4277	0.4291	0.3921	0.4381	0.4442	0.4416	0.2410

It should be emphasized that this paper focuses solely on $PM_{2.5}$, temperature, and winds as factors. Consequently, the model's key statistical metrics are not particularly high. However, our objective is to compare the relative importance of $PM_{2.5}$, temperature, and winds, rather than the accuracy of predictive outcomes. Therefore, these model evaluation metrics have been placed in the supplementary file. Nevertheless, we recognize that the importance factors for radiation and precipitation derived through machine learning methods may be influenced by other unaccounted variables, such as terrain elevation and land surface conditions. Consequently, we added the machine learning model accuracy limitations and the shortcomings of excluding factors in the discussion, as "TAP $PM_{2.5}$ data contains the meteorological information and some other factors have not been taken into account in the machine learning such as terrain elevation. The mechanism and degree of the impact of pollution RF on precipitation are not unequivocal.".

3. Page 20, Line 455: The discussion regarding regional heterogeneity is currently superficial. The authors should elaborate on the underlying physical mechanisms driving these differences. Specifically, the analysis should link the results to sub-regional variations in surface types, emission intensities, and topographical features.

Response: According to your suggestion, we have added the discussion regarding regional heterogeneity in section 3.3.1 and section 4. In section 3.3.1, we discussed the reason for regional heterogeneity in depth. And in section 4, we also briefly explain the regional heterogeneity. The adding content in Section 3.3.1 is as followed: "The importance of influencing factors to pollution RF showed heterogeneity across different regions. This may be related to variations in aerosol concentrations and pollution RF. In clear sky, the BLT region featured high $PM_{2.5}$ concentrations (Figure 3) coupled with high RF values (Figure 4a), whereas the SB region exhibited low $PM_{2.5}$ concentrations alongside low RF, illustrating the higher importance of $PM_{2.5}$ impacts on RF at the surface in these two regions (Figure 8a). Conversely, the NB region, with low $PM_{2.5}$ concentrations but high RF, showed less importance of $PM_{2.5}$ than in the BLT and SB. Differences in the importance of meteorological parameters across regions may be related to local topography and geographical location. Figure 1 indicates that the northwestern parts of the BLT and NB regions border mountainous terrain, where wind direction and speed significantly influence pollution dispersion and transport. Consequently, wind (V-wind in the BLT region, U- and V-winds in the NB region) exhibited greater importance for RF in these regions compared to the SB region. Conversely, the SB region's lower latitude and absence of nearby high mountain ranges resulted in temperature factors being more important than in the BLT and NB regions." And in the Section 4 with "The regional heterogeneity for the important factors in the Bohai Rim regions may relate to aerosol concentrations, RF values, local topography, and geographical location."

4. Please clarify whether data standardization was performed prior to K-means clustering. Given the disparate units and scales of the input variables, omitting standardization could significantly bias the clustering results towards variables with larger magnitudes. If standardization was not applied, a rigorous justification is required.

Response: For k-means clustering, we did not perform data standardization. This is because our clustering is conducted separately for meteorological parameters, where there is no issue of inconsistent units. Furthermore, we selected parameters below 850 hPa for clustering, meaning the data used for clustering exhibited minimal variation in magnitude. Concurrently, in response to your comments, we performed standardization prior to clustering. The results were largely consistent with those obtained without standardization, as illustrated in Figure 9 below. Regarding this standardization issue, we have included the figure in the supplementary file and provided an explanation in the section "2.4.3 The data usage workflow**"** as "Since we cluster based on separate parameters in the boundary, the standardization before clustering becomes unnecessary, as standardized and unstandardized results are essentially consistent (see Figure S3)”.

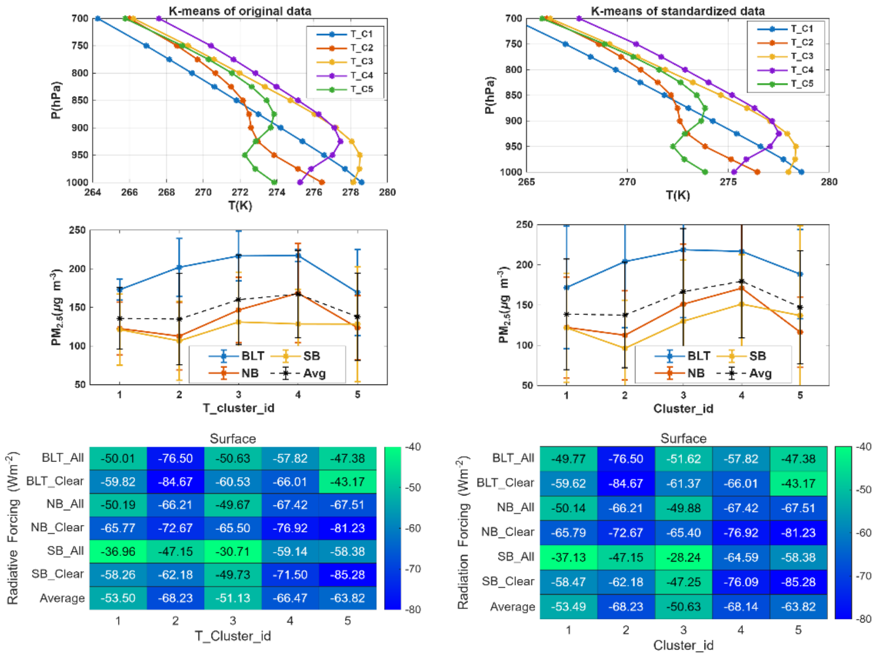

Figure 9. The comparison of k-means clustering by using original data and standardized data (Z-score of T). (Figure S3 in new Supplementary file).

5. Data Quality and Pre-processing: The validity of applying machine learning to linearly interpolated CERES radiation data requires justification. Linear interpolation may introduce artifacts or smooth out extreme values, potentially biasing the training process of the machine learning models. The authors should discuss: The extent of missing data prior to interpolation. Whether the spatial resolution of CERES data is sufficient to capture the local variability of pollution events in the study area. How these uncertainties affect the generalizability of the conclusions.

Response: We agree with your viewpoint.

①Regarding the extent of missing data prior to interpolation, we used the CERES-SYN1deg product; “CERES-SYN1deg provides hourly gridded observed top of atmosphere (TOA) fluxes and computed surface fluxes from the Fu-Liou radiative transfer model, which is suitable for regional diurnal and process studies. The CERES-SYN1deg products have been validated by other measurements (Doelling et al., 2016; Fillmore et al., 2022; Rutan et al., 2015)”. Our work utilizes the daily numerical data, thus eliminating the issue of missing data prior to interpolation.

②Regarding whether the spatial resolution of the CERES data can capture local variability in pollution events within the study area, it can be assessed from Figure 10 (Figure S4 in the new supplemental file). The figure demonstrates that the raw 1x1° data exhibits differences between the three study regions (BLT, SB, and NB), with some local variability also present within each region. Naturally, the CERES SYNIdeg 1°× 1° spatial resolution data cannot resolve minute gradients within urban areas. However, this study focuses on regional radiative forcing and its influencing factors, with the temporal focus being regional heavy pollution days (regional pollution events). Regional heavy pollution days typically exhibit strong spatial consistency across larger scales. Therefore, this dataset can still capture regional variations in pollution events to a certain extent in study regions on the regional heavy pollution days.

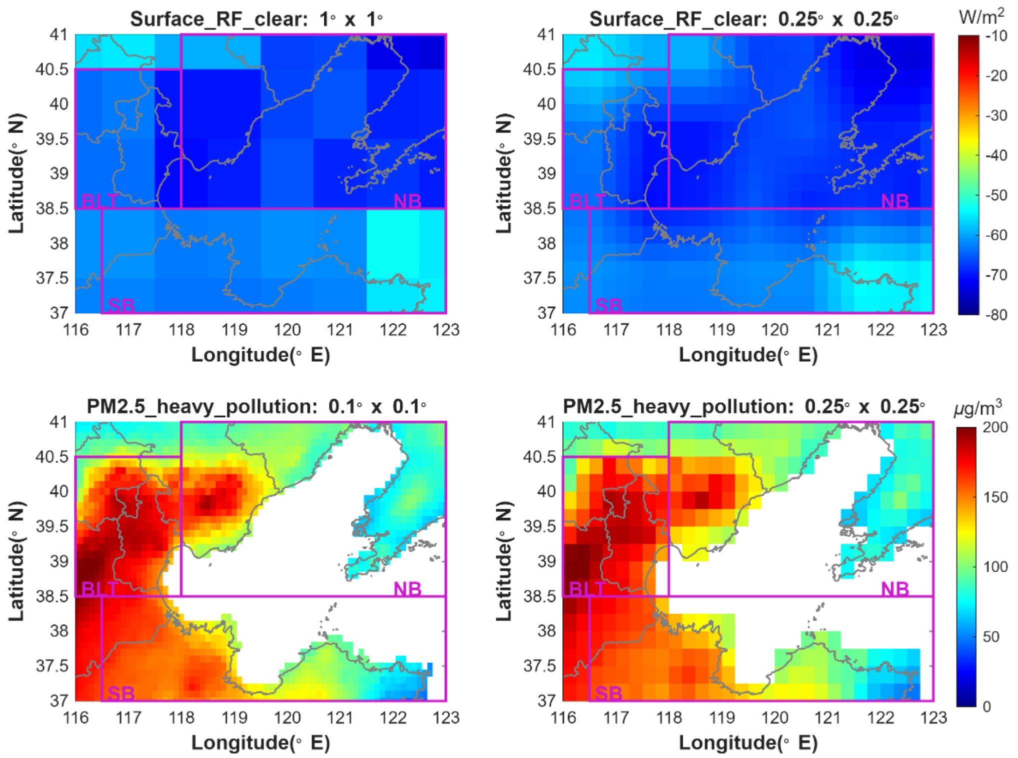

Figure 10. The comparison between the original data (first column) and the interpolated data (second column) in the study area. The first row is the radiative forcing (RF) from CERES at the surface in clear sky, and the second row is the interpolation for the $PM_{2.5}$ from TAP. The purple rectangle indicates the three regions.

③ Concerning the bias issues arising from machine learning following interpolation, we can find in Figure 10 that "The spatial distributions of the interpolated and original datasets are generally consistent, with minor discrepancies observed only at a few grid points exhibiting abrupt value changes. Since the results of this study primarily focus on regional averages, the errors introduced at a limited number of grid points have less impact on the regional mean outcomes." Besides, we have provided an explanation in the discussion section of the paper. Due to issues with data resolution, higher-precision observational data will be required in the future, as "The interpolation of datasets with different spatial resolutions used for training of the machine learning algorithm may cause some uncertainty. Further studies are required to quantify the impacts of these factors on RF and precipitation and to explore the underlying physical mechanisms or connections through additional observations in diverse regions and in higher spatial resolution, and numerical simulations."

6. In Section 2.4 ： The current description of K-means and Random Forest is too generic. This section should be condensed to focus on the specific implementation details specific to this study, such as hyperparameter settings, input feature selection, and the cross-validation strategy employed.

Response: Thanks for this suggestion. We have condensed the description of the section about K-means and Random Forest and added some specific implementation details to this study including hyperparameter settings, input feature selection, and the cross-validation strategy employed. Besides, we added a section of "2.4.3 The data usage workflow" and Figure 1 (this response file) to show more implementation details.

For k-means, the description is as "We classified the T and wind components profiles in the boundary layer (below 850 hPa) adopting the k-means clustering method (Lloyd, 1982). K-means clustering is an unsupervised machine learning algorithm used to partition a dataset into distinct groups or clusters, and popular in a wide variety of applications due to its simplicity, efficiency and effectiveness. Through calculation, there were 161 days of regional $PM_{2.5}$ heavy pollution days in the study regions, which is shown in the next section. Then, the profiles at the 11 stations in Bohai Rim on the 161 regional heavy polluted days are used to cluster (the number of samples is 11*161).

We use the k-means ++ algorithm for cluster center initialization (Arthur and Vassilvitskii 2006) and the squared Euclidean distance to measure the similarity to the centroid. We selected the numbers of clusters (2–8) for classification of T, and then combined the elbow method (the corner of the Sum Square Error) and the representativity of T profiles to determine the last number of clusters (=5 in this study). The numbers of horizontal and vertical wind component clusters (also 5 clusters) were selected along the T clustering."

For Random Forest, the description is "We used the Random Forest algorithm to compare and rank the importance of various factors to pollution RF and daily total precipitation. The variable factors concerned in this study were $PM_{2.5}$, T, and 3 wind components at four levels (500, 700, 850, and 1000 hPa). The Random Forest method is a popular ensemble learning technique that combines multiple decision trees to improve prediction accuracy and reduce overfitting (Breiman, 2001). In addition, it performs excellently for evaluating the independent variables' importance (Cutler et al., 2007). This study mainly used the "out-of-bag" observations method (Archer and Kimes, 2008) in the Random Forest regression model to calculate the importances of variables. Out-of-bag predictor importance estimates by permutation measure how influential the model's predictor variables are at predicting the response. Thus, the larger the calculated value, the greater its importance. For the random forest model training, this study employed a widely used 10-fold cross-validation (CV) method. Through repeated tests, we obtained the optimal hyperparameter of the number of trees from 20 to 100 and the number of leaf used the default of 5. The coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the model performance, which was shown in Table S2."