Response to Reviewer #3 for Geoscientific Model Development: Manuscript gmd-2024-148 By Liu et al.

We sincerely thank Reviewer #3 for thoughtful and constructive feedback. We have carefully considered each comment and made every effort to implement all the suggested changes. The notes below address each comment in detail. Please note that Reviewer's comments are shown in bold type and our responses in plain type.

Reviewer #3

This work develops a coupled chemistry-meteorology data assimilation (DA) system for the China Meteorological Administration (CMA). It proposes a coupled chemistry meteorology model and adds black carbon (BC) mass concentration to the operational CMA-GFS 4D-Var system as a control variable. This work (part 1) focuses on system description, validation of tangent linear and adjoint models, and investigation of the analysis increment.

Response: We sincerely appreciate the reviewer's thorough summary and insightful comments on our study. The reviewer has correctly captured the key aspects of our work. As stated, this work (Part 1) primarily focuses on the system description, validation of the tangent linear and adjoint models, and investigation of the analysis increment. In subsequent work, we plan to present a more comprehensive analysis of the impact of BC assimilation on BC and meteorological fields in Part II. Thanks again for the valuable feedback.

General comments:

- **The authors generally clearly describe the DA system. However, some important information is discussed in the result section, and some reordering may be needed (see specific comments below).**

  Response: We thank the reviewer for the constructive feedback. We appreciate the suggestion regarding the reordering of the content. In response, we have moved the relevant information from the results section to the section describing the DA system, as recommended. For details, please refer to the response to the issue regarding Lines 464-476. This restructuring allows for a clearer presentation of the methodology and ensures that the information is logically organized. We believe this improves the flow of the manuscript and enhances the clarity of the DA system description.

- **The authors compare the differences in analysis increments between uncoupled, weakly coupled and strongly coupled DA experiments in Section 4.4. The reviewer has some questions about the results, particularly on Figure 10 (see below).**

  Response: Thanks for the comments and interest in our results. We acknowledge that the original expression in Section 4.4 may not have been sufficiently clear, and we appreciate the opportunity to improve it. After carefully considering the reviewer's feedback, along with comments from the other two reviewers, we have completely rewritten this section, which is now presented as Section 5.3 in the revised manuscript.

  In the updated version, we have clearly introduced the objective of the four experiments, which is to investigate the impact of different BC assimilation strategies on both BC and atmospheric variables. We have renamed the four experiments as DA_BC, DA_MET, DA_MET_then_BC, and DA_MET_BC_simult. The revised Table 3 now provides a clear description of the four experiments. We have also compared the BC analysis increments obtained from the DA_BC, DA_MET_then_BC, and DA_MET_BC_simult experiments, noting that the BC analysis increments from the DA_MET experiment are very small. Additionally, we compare the atmospheric analysis increments caused by BC assimilation in DA_BC, DA_MET_then_BC (DA_MET_then_BC - DA_MET), and DA_MET_BC_simult (DA_MET_BC_simult - DA_MET).

Our main conclusions from this analysis are as follows: The preliminary results obtained from this set of four experiments indicate that different BC assimilation strategies have little impact on BC analysis increments but significantly affect the analysis increments of atmospheric variables. When only BC observations are assimilated, the influence of BC on atmospheric variables is more pronounced, whereas the simultaneous assimilation of meteorological observations moderates this influence. This suggests that in BC assimilation, meteorological observations can help constrain the uncertainty introduced by BC observations on atmospheric variables, thereby improving the reliability of the assimilation results. Moreover, these results demonstrate the successful implementation of the newly developed CMA-GFS-AERO 4D-Var system and highlight it as an effective approach for investigating the feedback of BC data assimilation on meteorological forecasts.

In the future, we will conduct batch experiments using CMA-GFS-AERO 4D-Var to gain deeper insights into the role of BC assimilation in numerical weather prediction and further refine the system for broader applications.

Additionally, in response to another reviewer's suggestion, we have adjusted the radius of influence for BC observations to 2 km, 10 km, and 20 km for urban, rural, and remote stations, respectively, according to Elbern et al. (2007). Consequently, all experiments in Section 5.3 have been redone using the updated radii, and the corresponding figures and text have been revised accordingly to reflect the new results.

For more details on the analysis, especially the discussion related to Figure 10, please refer to Section 5.3 of the revised manuscript. We once again appreciate the reviewer's valuable suggestions.

Specific comments:

- **Line 137: The reviewer is not sure whether "inverse integration" is a proper expression for the adjoint model.**

  Response: Thanks for pointing this out. We have revised the sentence to use "backward integration", which better reflects the role of the adjoint model in integrating backward in time.

- **Line 230-235: The authors provide a pragmatic way to convert the BC mass concentration (C_{bc}) to the mass mixing ratio of BC (ψ_{bc}). The reviewer has two questions regarding this:**
  - **Are the weights (ω^n) calculated locally?**

Response: Thanks for the insightful comment. The distribution weights (ω^n) are not calculated locally; instead, they are determined based on the entire three-dimensional background field. This approach ensures that the weight distribution reflects the global characteristics, rather than being influenced by local variations. By doing so, a global weighting factor is provided, which allows for a reasonable allocation of analysis increments.

Additionally, we have added the following explanation in the revised manuscript: "…Firstly, the distribution weights (ω^n) of each size bin of ψ_{bc}^n in the background field are calculated based on the entire three-dimensional domain, following the equation $\omega^n = \frac{\sum_1^N {\psi_{bc}}^n}{\sum_{n=1}^6 \left(\sum_1^N {\psi_{bc}}^n\right)}$, where N represents the number of three-dimensional grid points…"

- **In reality, could the change in C_{bc} be mainly caused by ψ_{bc} for one size bin (one of the six variables)? For example, the observation of the BC surface concentration is larger than the background, and it is mainly due to ψ_{bc} of one size bin. However, the DA system will assign the analysis increments to ψ_{bc} of all size bins.**

  Response: We appreciate the reviewer's concern. In reality, the change in C_{bc} may indeed be primarily driven by the variation of ψ_{bc}^n in a specific size bin. While the use of distribution weights (ω^n) ensures that the analysis increment is proportionally assigned across all size bins. For instance, if the change in C_{bc} is mainly due to the second size bin of ψ_{bc}^n, then the weight ω^n for this bin will be relatively larger, leading to a greater proportion of the analysis increment being assigned to it. This approach maintains consistency while still reflecting the dominant contributors to C_{bc} variations.

- **Line 253: The forward model was used to refer to the forecast model. It should be clear here whether this is an observation forward model (observation operator) or a forecast forward model.**

  Response: Thanks for the insightful comment. To clarify, the "forward model" in this context refers specifically to the forecast forward model. Following the recommendation, we have revised it as follows:

  "…Representativeness errors reflect the inaccuracies in the forecast forward model…"

- **Line 298: It would be a good idea to add some references to show the use of the SOAR function in operational data assimilation, e.g., Ballard et al., (2016). https://doi.org/10.1002/qj.2665**

  Response: Thanks for the valuable suggestion. We have added a reference to Ballard et al. (2016) to highlight the use of the SOAR function in operational data assimilation systems. We appreciate the recommendation, as it strengthens the context and support for this statement. The revised sentence now reads as follows:

  "…The horizontal correlation of the background error for the control variable BC is calculated by the second-order auto-regressive (SOAR) correlation function, which is commonly used in operational data assimilation systems (Ballard et al., 2016), expressed as…"

- **Line 312-313: Information is described repeatedly.**
  Response: Thanks for pointing this out. We have simplified the description in Line 312-313 to avoid duplicating the detailed information already provided in Section 2.2. The revised sentence now reads as follows:
  "…Referring to the running scheme of the CMA-GFS 4D-Var system described in Section 2.2, the CMA-GFS-AERO 4D-Var system adopts the same 6-h cycling schedule and assimilation windows…"

- **Line 316: How about the iteration number in the outer loop?**
  Response: Thanks for pointing this out. In our configuration, the outer loop is performed only once. This is consistent with the operational setup of the CMA-GFS 4D-Var system and has been found sufficient to achieve convergence within the experimental framework of this study. We have updated the text to include this information for clarity in the revised manuscript as follows:
  "…The maximum minimization iteration number in the inner loop was set to 50, while the outer loop was performed only once. This setting is consistent with the operational configuration of the CMA-GFS 4D-Var system and has been found sufficient for achieving convergence in our experiments…"

- **Line 464-476: The authors discuss the coupling of the DA system in the result section (Section 4.3). The reviewer suggests providing this information in the section that describes the DA system.**
  Response: We sincerely appreciate the reviewer's valuable suggestion. Following the recommendation, we have provided this information in Section 3 that describes the DA system, as follows:
  "3.3.5 Flow-dependent background error covariance in CMA-GFS-AERO 4D-Var
  In the strongly coupled aerosol-meteorology assimilation system, interactions between the atmospheric variables and BC allow BC observations to influence the analysis increment of atmospheric variables and vice versa. The incremental 4D-Var algorithm implicitly evolves the background error covariances (**B**) throughout the assimilation window according to the TL model dynamics. This process modifies prior background error variance estimates and induces non-zero correlations between model variables (Smith et al., 2015). By utilizing the fully coupled TLM and ADM in the inner loops of the strongly coupled assimilation system, cross-covariance information between BC and atmospheric variables is generated. This enables observations of one variable to produce analysis increments in the other, leading to more consistent analyses.
  Specifically, if the BC observation is assumed to take place at the initial of the assimilation window, the 4D-Var assimilation is equivalent to the 3D-Var assimilation. Since the BC variable is assumed to be uncorrelated with the atmospheric variables in the static **B**, and there is no direct relationship between the BC observation operator and the atmospheric variables, the BC observation does not lead to the generation of the analysis

increments of atmospheric variables. In this case, the merits of a coupled data assimilation system cannot be fully manifested by only assimilating a BC observation at the beginning of the window. If the BC observation is assumed to take place at the middle and the end of the assimilation window, $\mathbf{B}$ evolves within the assimilation time window through the TLM $\mathbf{M}_{0\rightarrow i}$, obtaining the implicit background error covariance matrix $\mathbf{M}_{0\rightarrow i}\mathbf{B}\mathbf{M}_{0\rightarrow i}^{T}$ that evolves with time. $\mathbf{M}_{0\rightarrow i}\mathbf{B}\mathbf{M}_{0\rightarrow i}^{T}$ includes the cross-covariances information of BC and atmospheric variables, and can realize the feedback of the BC observation to the atmospheric variables through the CMA-GFS-AERO ADM $\mathbf{M}_{0\rightarrow i}^{T}$, further producing analysis increments of atmospheric variables."

Additionally, we have revised the description of Figure 7 in the manuscript to make it more concise:

"Figure 7 depicts the analysis increments of temperature at the first model level at the initial time of the assimilation time window (0300 UTC), with the BC observation placed at 0600 and 0900 UTC, respectively. It can be seen that when the BC observation is placed at 0600 and 0900 UTC, positive analysis increments of temperature are generated, with the value of about 0.02 K near the observation location. The mechanism behind the generation of these temperature increments is detailed in Section 3.3.5. This indicates that the temperature of the analysis field will increase due to the assimilation of the BC observation."

- **Line 540: In EXP3, if the atmospheric and the BC variable were minimized separately (as stated in line 503), would the assimilation of the BC variable affect the atmospheric variable? If so, is it due to a change in the background when the model is coupled? Does this also mean that the results of EXP2 and EXP1 should be similar (the authors did not show the results of EXP2)?**

  Response: We sincerely appreciate the reviewer's insightful comments. We once again apologize for the confusion caused by the unclear description in Section 4.4 of the original manuscript. After carefully considering the reviewer's comments, we have completely rewritten this section, which is now presented as Section 5.3 in the revised manuscript. In this revision, we have clarified the distinction between different assimilation approaches and refined the explanation of how BC assimilation interacts with atmospheric variables, ensuring a more accurate and intuitive presentation of the results. These updates also address the reviewer's concerns about the impact of BC assimilation on atmospheric variables and the relationship between different experiments.

- **Figure 10: How many BC observations are used in this plot?**

  Response: Thanks for the insightful comment. The assimilation window of CMA-GFS-AERO 4D-Var is 6 hours. We assimilated all available BC observations at an hourly frequency within this window, excluding any missing data, resulting in 135 observations in total.

- **Line 549: 10f is not similar to 10d and 10e.**

  Response: Thanks for pointing this out. As we mentioned in our previous response, after carefully considering the reviewer's comments, we have completely rewritten the original Section 4.4, which is now presented as Section 5.3 in the revised manuscript. In the updated version, the statement that "10f is similar to 10d and 10e" has been removed. We once again appreciate the reviewer's valuable feedback.

- **Line 553-556: This sentence seems to compare the analysis increments in EXP4 with those in EXP1 and EXP3. However, the right panels of Figure 10 show the differences in analysis increments between EXP4 and EXP2, not the analysis increments in EXP4.**
  Response: We sincerely appreciate the reviewer's insightful comments. As we mentioned in our previous response, after carefully considering the reviewer's comments, we have completely rewritten the original Section 4.4, which is now presented as Section 5.3 in the revised manuscript. In the updated version, we have explicitly clarified that the atmospheric analysis increments shown in Figure 10 are solely generated by BC assimilation and do not include contributions from atmospheric observations assimilation. Throughout the analysis, we have ensured that all related descriptions are precise and clearly articulated. We once again thank the reviewer's valuable comments.

- **Line 557-558 and 606-608: To reveal the difference in analysis increments between weak and strong coupling, a comparison between EXP4-EXP2 and EXP3-EXP2 is needed.**
  Response: Thanks for the valuable comment. As we mentioned in our previous response, our expression in Section 4.4 in the original manuscript was not sufficiently clear, which may have caused confusion. We apologize for any misunderstanding. In Section 5.3 of the revised manuscript, we have compared the atmospheric analysis increments caused by BC assimilation in DA_MET_then_BC (DA_MET_then_BC - DA_MET) and DA_MET_BC_simult (DA_MET_BC_simult - DA_MET). Additionally, the statement in Lines 606-608 of the Conclusion section has been removed accordingly. We once again appreciate the reviewer's valuable suggestions.

- **Tables 4, 5 and 6: Are the computation times in these tables the average of multiple realisations?**
  Response: Thanks for the insightful comment. Yes, the computation times in Tables 4, 5, and 6 are the averages of multiple realizations. We conducted several repeated experiments to ensure the robustness and representativeness of the results. The values shown in the tables are the mean computation times from all the repeated runs.

Technical corrections:

- **Line 111-112: "… (Canadian Aerosol Module; (Gong et al., 2003)) …)"**
  Response: Thanks for pointing this out. We have revised it as follows:
  "…CAM (Canadian Aerosol Module; Gong et al., 2003) ..."

- **Equation 12: R is used to define two things in the same equation.**
  Response: Thanks for pointing this out. We revised the equation to avoid ambiguity by replacing R with R_d in $k_z = \frac{g^2}{(R_d T_0)^2} k_p$. The revised text now reads as follows:

"… $k_z = \frac{g^2}{(R_d T_0)^2} k_p$, g denotes the gravitational acceleration, R_d represents the gas constant for dry atmospheric air …"

- **Line 552: Extra space between "values" and "in".**
  Response: Thanks for pointing out this formatting issue. We have corrected the extra space between "values" and "in".