

Authors' responses to Referees' comments

Journal: Geoscientific Model Development

Manuscript Number: gmd-2024-78

Title: NAQPMS-PDAF v2.0: A Novel Hybrid Nonlinear Data Assimilation System for Improved Simulation of PM_{2.5} Chemical Components

Authors: Hongyi Li, Ting Yang, et al.

Note:

Comment (12-point black italicized font).

Reply (indented, 12-point blue normal font).

“Revised text as it appears in the text (in quotes, 12-point blue italicized font)”.

Anonymous Referee #1

1 General comments:

*This manuscript integrates an ensemble Kalman filter-based non-linear data assimilation method with an atmospheric chemistry transport model (CTM). The primary advancement is the coupling of the hybrid Kalman-Nonlinear Ensemble Transform Filter (KNETF) with an adaptive forgetting factor to the CTM model. The method was tested using a real-world dataset, with experiments varying ensemble sizes and evaluated against multiple metrics. **The presentation quality is good, though some minor issues need to be addressed.***

Authors' response:

We thank the reviewer for the positive assessment and constructive suggestions of our manuscript.

2 Detailed Comments:

1) Line 15: Replace "difficulty" with "challenge".

Authors' response:

We thank the reviewer for the suggestion, and we agree that "challenge" is more accurate. The revised text is shown below.

Abstract, Line 15: “*However, accurately describing spatiotemporal variations of PM_{2.5}*”

chemical components remains a challenge.”

2) Line 147: The term "level-2" is not adequately introduced. Consider moving the reference from Lines 152-153 to the beginning of the paragraph for clarity.

Authors' response:

We thank the reviewer for the suggestion and agree that the reference from Lines 152-153 should be moved to the beginning of the paragraph for clarity. Since the term “level-2” is detailed in our previous work (Wang et al. 2022), this manuscript dose not adequately introduce.

For NAQPMS-PDAF v2.0, we designed the level-2 parallel computational framework as in NAQPMS-PDAF v1.0 by using the Message Passing Interface (MPI) standard to ensure high computational efficiency in ensemble data assimilation. The level-2 parallel computational framework can simultaneously perform the parallel computation within the atmospheric chemistry transport model NAQPMS and the parallel computation of the ensemble tasks. For example, running 20 ensembles means executing 20 model tasks, each of which requires integral computation in a large model grid. With enough computational resources, 20 model tasks can be executed simultaneously at different computational nodes, and each model task can perform parallel computation in the large model grid with multiple processors. For practice, NAQPMS-PDAF v2.0 initializes the main communicator (MPI_COMM_WORLD) into three sub-communicators (Fig. R1), namely COMM_model, which is responsible for model integration, COMM_filter, which is responsible for the filter analysis, and COMM_couple, which is responsible for the information transfer between the model and the filter. Taking the example of calling 12 processors, assuming that the ensemble size is 3, NAQPMS-PDAF v2.0 will initialize 3 COMM_models for 3 model tasks, and each model task can be assigned 4 processors. Then, each model grid can be cut into 4 sub-grids for parallel computation. COMM_couple combines the communicators for different model tasks. COMM_filter occupies the same number of processors as the first model task (COMM_model 1) with 4 processors.

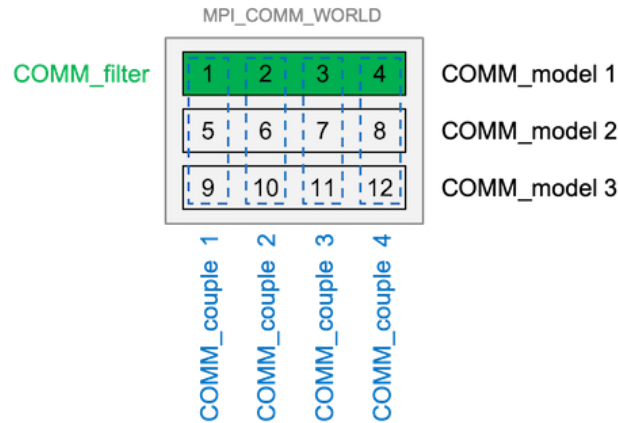


Figure R1: Example of a typical configuration of the communicators using a parallelized model (quoted from <https://pdaf.awi.de/trac/wiki>)

According to the reviewer’s suggestion, we moved the reference to the beginning of the paragraph for clarity and added the description to the term “level-2”. The revised text is shown below.

Section 2.3.1, Line 146-153: “NAQPMS-PDAF v2.0 implements an online coupling between NAQPMS and PDAF v2.1 with OMI, utilizing a level-2 parallel computational framework. The description of level-2 parallel implementation was detailed in our previous work (Wang et al., 2022). The online coupling ensures the continuous operation of model forecasts and assimilation analysis at each time step, achieved by directly integrating PDAF routines into the prototype code of NAQPMS (the right portion of Fig. 1, the blue represents NAQPMS main routines, while the yellow represents PDAF main routines). The level-2 parallel computational framework, which utilizes the Message Passing Interface standard (MPI), facilitates concurrent processing and data exchange among multiple ensemble members and parallel computation among model state matrixes within each ensemble member, enhancing the efficiency of ensemble analysis and numerical model computations. For instance, the operation of twenty ensemble members necessitates the execution of twenty model tasks, each of which performs integral calculations on a large model grid. Twenty model tasks can be executed simultaneously at twenty computational nodes with sufficient computational resources. Each model task can then perform parallel computation with multiple processors by splitting the large model grid into multiple sub-grids.”

Reference

Wang, H., Yang, T., Wang, Z., Li, J., Chai, W., Tang, G., Kong, L., and Chen, X.: An aerosol vertical data assimilation system (NAQPMS-PDAF v1.0): development and application, *Geosci. Model Dev.*, 15, 3555-3585, <https://doi.org/10.5194/gmd-15-3555-2022>, 2022.

3) Line 215: *Properly cite online resources instead of directly inserting hyperlinks.*

Authors' response:

We thank the reviewer for the reminder and apologize for the inappropriate citation. The revised text is shown below.

Section 2.3.3, Line 214-215: “...which are subsequently transformed into non-Gaussian distribution matrixes through non-Gaussian process generation v1.2 (Cheynet, 2024).”

Reference

Cheyne, E.: Non-Gaussian process generation, https://github.com/ECheyne/Gaussian_to_nonGaussian/releases/tag/v1.2, GitHub. Retrieved July 7, 2024.

4) Figure 3: *The coloring of Domain 3 is difficult to distinguish.*

Authors' response:

We thank the reviewer for the suggestion, and we revised the Fig. 3 in the original manuscript by changing the line color of domain 3. The revised version is shown below.

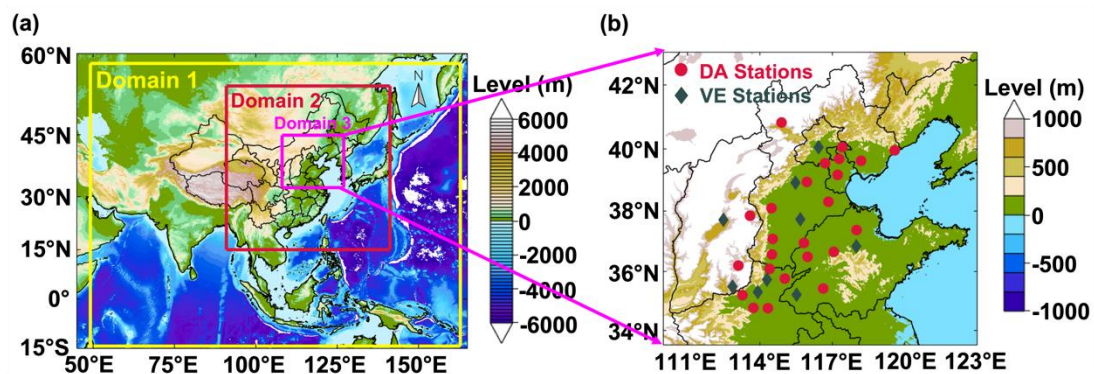


Figure 3: *The model domains in WRF simulation (a) and the location of observations*

(b). Domain 3 in (a) is the target area of this study. Twenty-four red sites in (b) represent the sites for data assimilation, and nine green sites in (b) represent the sites for spatial independent validation. The topographic dataset is from the ETOPO1 1 arc-minute Global Relief Model, taken from the National Geophysical Data Center (Amante and Eakins, 2009).

Reference

Amante, C. and Eakins, B. W.: ETOPO1 arc-minute global relief model: procedures, data sources and analysis, 2009.

5) *Line 395: Remove the word "deeply".*

Authors' response:

We thank the reviewer for the suggestion, and we removed the word "deeply" in Line 395 of the original manuscript. The revised text is shown below.

Section 3.1, Line 395: *"Therefore, in the Discussion section, we discuss the uncertainties of ensemble perturbations."*

6) *Figure 4: It is unclear from the manuscript whether the experiments were run multiple times, particularly for plot d). The stochastic nature of this method may introduce variation in running time. Indicate whether the presented values are the mean of multiple runs or include the mean and uncertainty band for multiple runs.*

Authors' response:

We thank the reviewer for the suggestion. To test the dependence on ensemble size, we controlled a fixed assimilation frequency of 1 hour and changed the ensemble size to 2, 5, 10, 15, 20, 30, 40, and 50. Then we assimilated the hourly observations of five PM_{2.5} chemical components from all sites with 48 timesteps from 00:00 (LST) on February 2nd to 23:00 (LST) on February 3rd, 2022. This means that the Model Integration process and Assimilation process iteratively looped 48 times, which is equivalent to performing 48 experiments. Therefore, the presented values in Figure 4 are statistical results over 48 timesteps. For Fig. 4d, the elapsed time of the system processes are the statistical averages over 48 timesteps. According to the Reviewer's

suggestions, we indicated the presented values in Fig. 4 in the Line 381 of the original manuscript. The revised text is shown below.

Section 3.1, Line 381: *“Figure 4 shows the mean CRPS, RMSE and CORR values and the statistical averages of the elapsed time over 48 timesteps with the ensemble sizes of 2, 5, 10, 15, 20, 30, 40, and 50.”*