

Dr. Sergey Gromov
Handling Topic Editor
Geoscientific Model Development

January 20, 2025

Dear Dr. Sergey Gromov,

We extend our gratitude for providing us with the topic editor report. We also wish to express our appreciation for your suggestions. The manuscript has undergone revisions to adhere to the archive standards of GMD.

Sincerely,
Prof. Patrick Rinke and Hilda Sandström

Topic editor report

Dear Hilda Sandström and Prof. Patrick Rinke,

After the internal discussion with the Executive Editors of the GMD, we have concluded that the manuscript needs a minor revision before it can be published. The reason is still the compliance with the GMD policies regarding the data used – we cannot accept references to pre-existing datasets in their current form (see the details below). Unfortunately, this implies some additional work on your side; fortunately, this will improve the publication quality and ensure reproducibility of the results.

Currently, the “Code and data availability” section of the MS reads: “The datasets used for this analysis are all freely available from original publications or the database website (see Table 1).” Unfortunately, referenced datasets are not available in compliance with our policy. I quote the decision of Juan A. Añel, the GMD Executive Editor who reviewed this case: “For this work the authors use seven pre-existing datasets (they list them in Table 1) that they do not link or publish in one of the repositories that we can accept. In some cases (e.g. Wang) they only cite a paper, not a repository. These datasets are: Wang, Gecko, Quinones, QM9, nablaDFT, MassBank Europe, MassBank of North America. The authors should store in a permanent repository the data they have used from these datasets, or ask the owners or publishers of such datasets to do it. I understand that it is not their data, but if they can use them, probably they can republish them, at least the part that they have used. Therefore, given that the authors have submitted a new version of their work for a new round of reviews, before continuing with it, please, send it back to the authors to address the issues with the datasets.”

I therefore would like to ask you to resubmit the manuscript with the following amendments introduced: - Either provide the reference to the data repository (containing archived data in exact form and version you have used) or store in a permanent repository the data you have used from these datasets, or ask the owners or publishers of such datasets to do it. The requirements (specifically the archive standards) are listed on GMD website at: https://www.geoscientific-model-development.net/policies/code_and_data_policy.html and should be satisfied for every dataset. The updated references may be accommodated in Table 1; - I encourage making a copy of every used dataset, if such is attainable; - Please combine the contents of the code and workflow datasets you provide (currently there are two separate Zenodo publications) and add a sufficient description of their contents (listing the title of the manuscript, I guess you can reuse some of the contents of your ‘readme.md’ files). The dataset at <https://zenodo.org/records/14007835> provides mere classes but not the main workflow/code to obtain the results you present.

I am looking forward to receiving the revised manuscript,
With best regards,
S. Gromov

1. *I therefore would like to ask you to resubmit the manuscript with the following amendments introduced: - Either provide the reference to the data repository (containing archived data in exact form and version you have used) or store in a permanent repository the data you have used from these datasets, or ask the owners or publishers of such datasets to do it. The requirements (specifically the archive standards) are listed on GMD website at: https://www.geoscientific-model-development.net/policies/code_and_data_policy.html and should be satisfied for every dataset. The updated references may be accommodated in Table 1; - I encourage making a copy of every used dataset, if such is attainable;*

We have now created a new repository with copies of the dataset versions used in our similarity analysis. The datasets are now provided along with their source and licensing information. These resources are now available in the following Zenodo repository: <https://doi.org/10.5281/zenodo.14671496>, DOI:10.5281/zenodo.14671496.

We have also updated the code and data availability statement to read: "The current version of model is available from the project website: https://github.com/hilsan/atmospheric_compound_similarity_analysis under the GNU General Public License v3.0. The exact version of the model used to produce the results used in this paper is archived on Zenodo (<https://doi.org/10.5281/zenodo.14671496>), as are the input data and scripts to run the model and produce the plots for all the simulations presented in this paper (<https://doi.org/10.5281/zenodo.14671496>). The datasets used in this study are sourced from publicly available repositories, each under a specific license. Detailed licensing information is provided in the Zenodo repository accompanying this manuscript."

2. *Please combine the contents of the code and workflow datasets you provide (currently there are two separate Zenodo publications) and add a sufficient description of their contents (listing the title of the manuscript, I guess you can reuse some of the contents of your 'readme.md' files). The dataset at <https://zenodo.org/records/14007835> provides mere classes but not the main workflow/code to obtain the results you present.*

We have now created a new repository combining the versions of the datasets used for analysis, the previous two code and workflow repositories to align with archive standards. We have also extended the repository description and README.md accordingly. We have also added the manuscript title to the repository description. These resources are now available in the following Zenodo repository: <https://doi.org/10.5281/zenodo.14671496>, DOI:10.5281/zenodo.14671496.