

Brazilian Atmospheric Inventories - BRAIN: A comprehensive database of air quality in Brazil

MS No.: essd-2023-305

Dear reviewers,

We greatly appreciate the comments and suggestions you provided, which are very constructive and have contributed to enhance the content of the revised manuscript. Please find below the reply to all the reviewer's comments.

Best regards

Leonardo Hoinaski and coauthors

Reply to comments by Reviewer #1:

Reply to general comments

General comment:

The manuscript shows the development and analysis of an air quality management system to help air quality policies and management in Brazil. The paper brings a very interesting topic, specifically due to: i) the current lack of numerical simulated dataset needed to work with air quality in Brazil; ii) Not everyone that works with meteorological and/or air quality in knows how to produce and work with atmospheric models. Thus, having the possibility to use a platform, as proposed by the authors for a specific region, is an indubitable advantage. I would recommend the paper for publishing after addressing all the following comments.

My major concern is on creating numerical simulation for a country with such a large size as Brazil. For that, a massive statistic evaluation is mandatory and not only in cities with large amount of pollutants emissions but also in background regions, such as the Northwest of the Amazonas state (near to the Amazon Tall Tower Observatory -ATTO). In cases where a monitoring network is not available, the authors should use remote sensing to evaluate the numerical experiments. There is a good amount of satellite products to use in this case: MODIS to evaluate AOD; TROPOMI to evaluate NO₂ and HCHO; IASI+GOME2 to evaluate O₃ (on the 20km grid). If the authors are proposing to make available numerical simulations for the entire country, it requires evaluation for the whole domain or at least an evaluation scale where it shows a confiability level of the data. In order to assume the premises of good modeling results for regions without validation, the authors need to build an underlying analysis by showing how the model performs on key scenarios, such as:

i) Background regions.

ii) Under anthropogenic emission effects (also analyze the model's abilities with plume transport).

ii) Long range transport (specially in the Amazon region during the wet season with (between January and May), when the Intertropical Convergence Zone (ITCZ) is more intense in the south, allowing the long transport of BC (biomass burning emissions) and dust from Africa (Sahara desert) (Artaxo et al., 2013; Martin et al., 2016; Pöhlker et al., 2018, 2019). With the availability of several years of BC background measurements at the ATTO tower, the authors could separate African episodic events from the rather constant regional BC concentrations that are relevant when comparing with modeled values not under anthropogenic influence.

Once it is shown that the model is capable of representing key scenarios, the idea of using modeling data for regions not fully evaluated is more reasonable.

Reply: We appreciate the positive feedback and all points raised by the reviewer, which will improve our manuscript and give new directions for the continuity of our work. The BRAIN air quality dataset evaluation uses observations extracted from 244 air quality monitoring stations in Brazil in nine Brazilian states. We agree that BRAIN must be tested in background regions and other unmonitored areas. A new comparison with gridded data from MERRA-2 and TROPOMI has been prepared for inclusion in the final version of the manuscript, showing the model's performance in background regions. We will also demonstrate that BRAIN can represent key scenarios capturing observed temporal patterns in background regions. The influence of long-range transport will be addressed in a future version of the database by implementing boundary contributions from GEOSCHEM and other tools. We will be happy to include these considerations in the manuscript.

Reply to specific comments

I would strongly recommend the authors to put more focus on the states with the most extensive networks of AQS such as SP, RJ and MG, in addition of course to PR, SC and RS (I am suspicious it was done that way because the authors are based in SC).

Reply: We provide a detailed model evaluation by monitoring stations in nine Brazilian states, including SP, RJ, and MG. We used statistical metrics to measure the model's performance in each station. We also present visual statistical analyses through scatter and line plots by AQS in SM 12-14. We have planned to provide datasets with higher spatial resolution in Southeast Brazil in future versions of our database. New sentences will be included in the manuscript to clarify that we will provide further evaluation of BRAIN and more refined datasets for Southeast Brazil.

The authors should have a modeling strategy component focused on the States with AQS to allow the increase of the spatial resolution, and thus, get a more robust statistical evaluation for medium and small-sized cities that, otherwise, would not be properly covered by coarse resolution model simulations (10 km or more). I also recommend the authors to do a full evaluation by selecting a region with AQS available and combine observational data from satellites and field campaigns (if available). The idea here is to try to have a case where the model can be massive evaluated vertically and horizontally.

Reply: We agree that the coarser domain does not properly cover small municipalities. It will be addressed in future versions by providing datasets with higher spatial resolutions. We envision making available more detailed datasets for regions other than southern Brazil. It is worth emphasizing that BRAIN will be a long-term project. These considerations will be included in the manuscript.

Besides the evaluation using 244 AQS, we prepared a spatial comparison of BRAIN with the MERRA-2 and TROPOMI datasets to evaluate the database in background areas. A Full statistical evaluation with AQS has been presented in SM 11-14, showing that BRAIN performs adequately even with 20x20 km of resolution. It will be a good starting point for understanding the air pollution process in Brazil.

If there is no observations to compare with in States other than those with AQS, I do not think a country-scale simulation is really worthwhile here. Maybe the authors should provide a strategy/approach for validade/evaluate regions without AQS (remote sensing?).

Reply: For permitting and licensing purposes, the United States Environmental Protection Agency recommends the inclusion of background concentrations as part of a cumulative impact analysis. According to the Title 40 of the USA Code of Federal Regulations (appendix-W part 51), "*in those cases where adequately representative monitoring data to characterize background concentrations are not available, it may be appropriate to use results from a regional-scale photochemical grid model, or other representative model application*". In this context, BRAIN is truly worthwhile and could be a step forward for a good procedure for licensing new sources of air pollution. We understand that BRAIN requires a deeper evaluation. We believe in an ongoing process that will improve the database. We emphasize that BRAIN has limitations and requires further evaluation.

We prepared a spatial comparison of BRAIN with the MERRA-2 and TROPOMI datasets to evaluate the database in unmonitored areas. It will be included in the manuscript as well as the considerations about the purpose of the database and its limitations.

As most of the air quality stations used for model evaluation are placed in southeastern Brazil, why not focusing on the largest metropolitan areas of the country, such as, the metropolitan areas of São Paulo, Rio de Janeiro, Belo Horizonte, etc?. Or maybe use passed campaigns such as The Green Ocean Amazon experiment (GoAmazon2014/5) or The Regional Carbon Balance in Amazonia (BARCA) to evaluate the chemistry component of the model in different vertical levels and with high spatial resolution.

Reply: Our goal with BRAIN is to provide air quality and emission datasets to overcome data gaps and the lack of spatial representativeness of AQS. We aim to fulfill the requirements of dose-response studies and permitting procedures for new sources in Brazilian municipalities. There is extensive literature addressing the air quality in great metropolitan areas. We have decided to use a wider approach to somehow contribute to the literature. Future analysis with BRAIN could evaluate the chemistry and physics of air pollution in targeted areas. We will clarify the focus/goals of our database in the manuscript. Also, we will include considerations about BRAIN's limitations and future perspectives.

The authors claimed that, currently available initiatives including reanalysis and satellite products are still not providing datasets with large spatial and temporal resolutions for developing air pollution policies in Brazil. In the biomass emission perspective, have the authors checked the fire products from Visible Infrared Imaging Radiometer Suite (VIIRS) on board the Suomi NPP satellite - VIIRS (375 m resolution for fire activities). The paper Ferrada et al., 2022 (Introducing the VIIRS-based Fire Emission Inventory version 0 (VFEIv0)) shows a new open biomass burning inventory that relies on the fire radiative power (FRP) data from VIIRS.

Reply: We appreciate the recommendation. Indeed, VIIRS brings more detailed information and could potentially reproduce estimations with smaller uncertainties. We have prepared BRAIN since 2020. All codes/scripts were developed for data available in 2020. We have implemented the outdated FINNv1.5 in this first version. However, new data sources could be included in our emissions inventory in future work, which will optimize the database. We just finished a new version of BRAIN for 2020 using the new version of FINN and BRAVES. Therefore, the issue with outdated FINN has been already addressed. We will clarify that the database has the potential to incorporate new sources of data in future versions.

BRAVES uses activity data from field campaigns conducted in the metropolitan area of São Paulo. It really does not make any sense take this region out of a very high-resolution simulation design.

Reply: BRAVES has been developed by our group since 2018 to provide multispecies and high-spatiotemporal-resolution data of vehicular emissions covering the entire country. BRAVES totally suits this work. VEIN model (<https://gmd.copernicus.org/articles/11/2209/2018/>) is the one that uses activity data from field campaigns conducted in the metropolitan area of São Paulo. We will include more detailed information of activity data used in BRAVES modelling.

NCAR (FINN) version 1.5 (Wiedinmyer et al., 2011). This version is pretty outdated. An updated version is now available <https://egusphere.copernicus.org/preprints/2023/egusphere-2023-124/egusphere-2023-124.pdf>.

Reply: Indeed, FINNv1.5 is pretty much outdated. We just finished a new version of BRAIN for 2020 using the new version of FINN fixing this issue. We plan to compare different sources of emission data in future versions of our database. We appreciate the recommendation, and we believe that it will strengthen BRAIN.

The authors mentioned that the WRF model demonstrated the ability to reproduce diurnal and seasonal variability of winds in the Brazilian North-East region (Souza et al., 2022a). Although, this resolution is slightly lower than the one used in this work for the parent domain. It wouldn't be better to just focus on the largest metropolitan areas of Brazil?. I am assuming that you have set up the model simulations at 20 km resolution in an attempt to avoid out-of-memory and space of storage issues, but if that is the case, why not just focusing on high densely areas?

Reply: There is a huge gap in air quality data (surface/ground level concentration) in Brazil, where most municipalities could not derive their own air quality management system, which motivated the BRAIN development. We focused on providing data on a national scale as a starting point for licensing new sources. As we did for Southern Brazil, we will reproduce datasets with better resolution for other Brazilian regions. We do have storage and processing limitations. Even though we have struggled with our small budget and team, we believe that BRAIN contributes to our field of knowledge. We will reinforce the BRAIN focus and where it could potentially be used.

The authors mentioned that the lack of data quality assurance may compromise the credibility of the available air quality observations in Brazil. It is true, and consequently, this could potentially compromise any analysis conducted on these data sets. In my view, a modeling study that centers on the States with the most extensive networks of AQS would have had a more effective simulation strategy, as the one mentioned in the previous comments. This point also brings the importance of using data from previous campaigns with high quality assurance.

Reply: Indeed, modeling study that centers on the States with the most extensive networks of AQS can be better evaluated. However, we have focused on providing data where AQS is not available. We believe that our database brings essential information for licensing new sources and deriving air quality management systems in Brazil. Further evaluation will bring robustness to our database. We will clarify the future perspectives of BRAIN in the manuscript.

Reply to comments by Reviewer #2:

Reply to general comments

The current paper aims to develop an emission database for the entire country of Brazil based on local emissions inventories for industries, vehicular emissions estimations, MEGAN for biogenic emissions, and FINN for fire emissions. Additionally, the authors aim to make this dataset available to the community.

The initiative is really good not only for Brazil but also for South America. The countries inside South America have their connection and impact of long-range transport of pollutants. Thus, an emissions inventory for the largest country is appreciated. Primarily when the government and environmental protection agencies do not provide the official data for the community

The authors have put much effort into the work, gathering all data (meteorology, emissions, temporal profile, chemical profile, different models, different databases), processing the models, and performing the air quality modeling.

The database is interesting and valuable for Brazil and South America and will be very appreciated by the community. As mentioned by the authors, the lack of this information holds the improvement in the air quality across Brazil. I would recommend the paper for publishing after addressing the following comments.

[Reply: We deeply appreciate the positive feedback.](#)

Reply to specific comments

1. Introduction

The introduction shows the situation of the region of study regarding air quality problems, fire emissions, countryside problems with air pollution, and vehicle emissions across the country, among others. Additionally, the authors exposed the lack of air quality monitoring stations and emissions inventory. This situation is a concern for any country since the lack of this information is critical for air quality management.

The authors stated: “Currently, available initiatives, including reanalysis and satellite products, are still not providing datasets with large spatial and temporal resolutions for developing air pollution policies in Brazil.” I would like some publications to enforce this statement. There are ERA reanalysis data with one hour of time resolution and less than 10km of spatial resolution. Additionally, there are MODIS data for air quality, probably not hourly, but there is a frequency that may help.

[Reply: We will include these initiatives in the introduction chapter, contrasting with BRAIN features.](#)

2. BRAIN Database

Error in the Figure 1 b): “Annual Industrial”

[Reply: OK! We fixed this issue](#)

2.1.1 Vehicular emissions

The BRAVES model was used to estimate the vehicle emissions. The database was validated in another paper, but I would like to see some explanation on how the authors estimated the activity data, especially over countryside cities. This information is crucial for understanding vehicular emissions.

Reply: BRAVES uses fuel consumption by the municipality to represent the activity data. We will include a more detailed description of BRAVES in the manuscript, explaining how the activity data has been estimated.

2.1.2 Industrial emissions

It used the emissions inventory of Espírito Santo, Minas Gerais, and Santa Catarina, which are local emissions inventory provided by the Environmental Protection Agency. Additionally, the emissions inventory developed by Kawashima et al. (2020) was also used. These emissions inventories are not from the same year and do not cover the entire state. Did the authors scale the emissions for the current modeling year, which is 2019? Please explain this in the text.

Additionally, have the authors considered applying EDGAR emissions across the regions without emissions inventory, especially for the industry sector?

Reply: We will clarify this point by including more details of the industrial inventory. We did not convert the emissions inventory to the current modeling year. We agree that it is a drawback in our database that must be addressed in further detail. Future efforts will be expended to improve the industrial inventory, scaling to the current year and including not inventoried sources.

2.1.3 Biomass burning emissions

FINN is a good database for biomass burning, primarily for providing daily Biomass burning emissions. I noticed the authors used the FINN 1.5, but the FINN version 2.5 is available. I recommend the authors to use the updated version.

Reply: Indeed, FINNv1.5 is outdated. This issue has been fixed in a new version of BRAIN which is already available.

2.1.5 Sea spray aerosol emissions

The Sea spray aerosol is handled inline in CMAQ. Is there a way to extract these values to use in another photochemical model, such as WRF-CHEM or CAMx?

Reply: We are not sure if one can extract sea spray emissions from CMAQ. We provide the OCEAN file (https://github.com/USEPA/CMAQ/blob/main/DOCS/Users_Guide/Tutorials/CMAQ_UG_tutorial_oceanfile.md) in our database. You can derive the sea spray emissions using this file. You can also derive your own OCEAN file using this Python script: <https://github.com/leohoinaski/CMAQrunner/blob/master/hoinaskiSURFZONEv2.py>

2.3 Air Quality & 2.3.1 Models' performance

The air quality was developed using CMAQ v5.3.2. The boundary conditions for the Brazilian domain are based on the standard profile. I would recommend that the authors provide Boundary conditions from a global model, such as GEOS-Chem, as mentioned in the paper, in this version or future work. This will account for the long-range transport of air pollutants across the country.

Regarding the model performance. The lack of air quality monitoring stations across Brazil was mentioned. However, there are some important areas in which the model should be validated for at least a few days across the modeling scenarios. This helps to understand if the emissions inventory should increase or decrease the total emissions. A validation with satellite data will improve this analysis.

Additionally, there is only the validation and plot for the 20x20km. How did the 4x4km modeling domain perform? I recommend adding this information to the text and supplementary material.

Reply: We agree that we should have included boundary conditions from a global model, such as GEOS-Chem. It will be completely addressed in future versions of our database. We have included a spatial comparison between the BRAIN dataset and MERRA-2 + TROPOMI to evaluate the database in unmonitored areas. We will provide a modeling evaluation for the 4x4km domain in the revised manuscript.