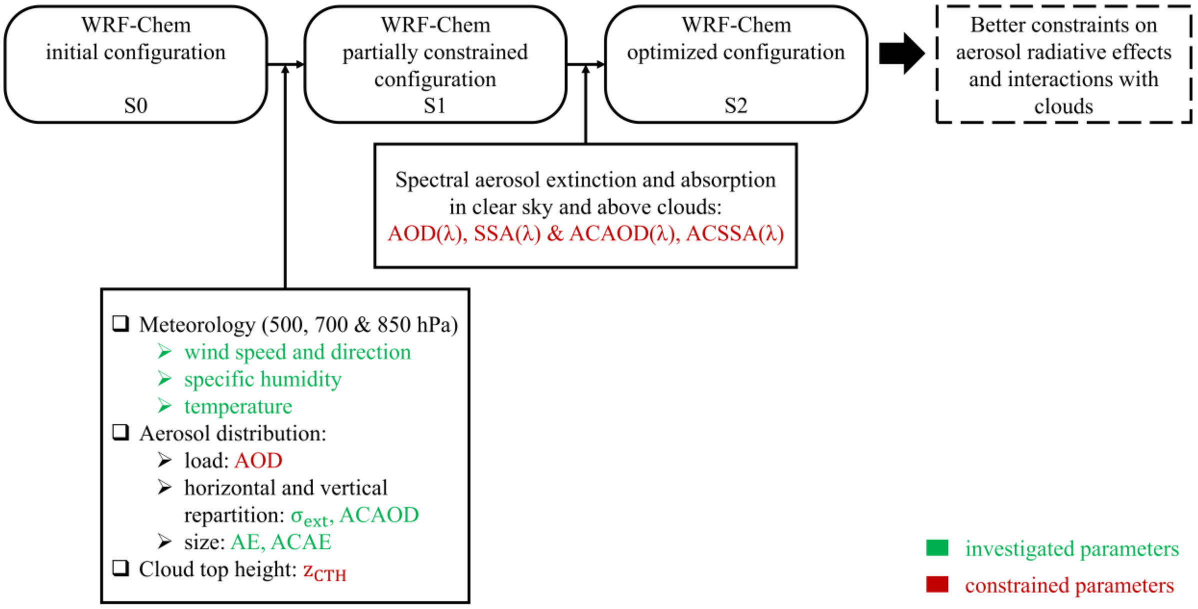
RC1: 1. In this review, I'll reveal myself as a remote sensing specialist with less expertise in models such as WRF-Chem. With that in mind, I'd like to question the two step method of constraining WRF-Chem with POLDER and other data. Is this common practice? What is the benefit compared to constraining in one step, in other words adjusting WRF-Chem all at once based on POLDER and met data? I could understand differentiating between constraints that come from the POLDER retrievals and other data. But that is not the case – AOD from POLDER is used in step 1, microphysical properties in step 2. The problem is that the POLDER algorithms work by simultaneously retrieving all of those parameters, which I would think should be considered when using those data to adjust WRF-Chem. For example, if one were to impose an incorrect AOD in a POLDER retrieval, there would potentially be an incorrect retrieval of aerosol absorption (and vice versa). In a similar fashion, by initially modifying emissions to match AOD, and then later constraining aerosol optical properties might mean AOD no longer matches. I wonder if the POLDER/WRF-Chem match for AOD in Figure 16 and the ACAOD in Figure 17 could be further improved by constraining emissions a second time at this point. Which begs the question: why not just do it all at the same time. Am I missing something here?

AC: We agree with the reviewer that both AOD and SSA are linked and some changes on one parameter can simultaneously have an impact on the other. In our approach, we decided to constrain WRF-Chem in two steps to investigate separately the major factors impacting the AOD and then the SSA. The idea was to constrain first the AOD as this optical property is an extensive parameter, i.e., integrating all aerosols regardless of the origin of the source. Once this parameter well simulated, we could look further into the aerosol distribution by studying the sensitivity of the SSA to the chemical composition.

Thanks to this approach, we first evaluated the general performance of WRF-Chem in terms of aerosol emissions and transport modelling, i.e., the dynamics. We concluded that the WRF-Chem partially constrained configuration correctly simulated the general pattern of the transport of aerosols both in terms of location and intensity. Once the AOD constrained after adjusting the strength of biomass burning and desert dust aerosol emissions sources, we then studied the main microphysical properties influencing the SSA, i.e., the size, the chemical composition and the level of absorption of aerosols. We were thus able to determine an optimized configuration of WRF-Chem reproducing well the AOD and SSA both in clear and cloudy atmospheres for the studied period. Although it might be surprising, from a modelling point of view, checking and adjusting aerosol emissions is often an unavoidable preliminary step, because emissions inventories and schemes in models remain too uncertain, as highlighted in many publications (e.g., Turquety et al. 2014; Flaounas et al. 2016).

RC1: 1b. As an aside, I find the nomenclature for S1 "reference configuration" to be confusing. "Reference," to me, implies some external (to this effort) source of data, while instead S1 represents the first stage of constrained WRF-Chem results. Alternatives that may be better: "partially constrained configuration" or "emissions adjusted configuration" or something else more informative. Furthermore, Figure 3 could be made more useful by also listing what is constrained in WRF-Chem at each step.

AC: We appreciate your propositions. We replaced "reference configuration" by "partially constrained configuration" in the article. We modified Figure 3 by adding the investigated (in green color) and constrained (in red color) parameters in WRF-Chem (see below).



We modified its caption like “Figure 3: Schematic diagram of the general approach applied to constrain WRF-Chem simulations of aerosols using A-Train satellite coincident retrievals of the distributions of aerosols and clouds. AOD, ACAOD, SSA, ACSSA, σ_{ext}, AE, ACAE and z_{CTH} stand for aerosol optical depth, above-cloud aerosol optical depth, single scattering albedo, above-cloud single scattering albedo, aerosol extinction coefficient, Angstrom exponent, above-cloud Angstrom exponent and cloud top height.”.

RC1: 2. Continuing the discussion on constraint methodology: why is it so approximate? In section 3.2.1 you adjust the APIFLAMEv1 BB emissions inventory by a factor of 1.5, while the ratio of WRF-Chem to POLDER-3/GRASP AOD(565) is 0.49/1.10 = 2.24. I imagine one reason could be that AOD does not directly relate to emissions estimates, but a discussion of how you go from that ratio value to 1.5 is needed. The same could be said for the choice of the Williams et al. (2007) BC refractive index, BCx2, OC/2.5 and 2.5% BrOC: is this the best match you can make? Could you do some more fine tuning to get a better constraint, or are the computational needs too much?

AC: Indeed, the AOD does not only relate to emissions estimates, but also on all the physico-chemical processes influencing the aerosol population. We first applied a multiplying factor of 2 to the APIFLAMEv1 BB emissions inventory and the result was not concluding. Over land (12°E-30°E, 10°S-0°S), simulated AOD values ($AOD_{mod,565}$ = 0.98) were in much better agreement with those retrieved by POLDER-3/GRASP ($AOD_{obs,565}$ = 1.10), with a moderate underestimation of 0.12 on a geographical average for July 2008. Nevertheless, above the ocean, simulated AOD values ($AOD_{mod,565}$ = 0.86) were overestimated compared to POLDER-3/GRASP retrievals ($AOD_{obs,565}$ = 0.61), with a positive bias of 0.25 on a geographical average for July 2008. This could be explained by a higher emitted and then transported aerosol load. This could also be interpreted as a numerical diffusion problem of a dense biomass burning aerosol (BBA) plume, which causes a spread of the load between the continent and the ocean due to the width of the mesh of WRF-Chem (as a reminder, a grid of 30x30 km^2). Finally, the adjustment factor of 1.5 was chosen as being a good compromise, even if small biases were still present. This point is now clarified in the revised manuscript.

To our knowledge, the mass absorption cross section (MAC) associated with the Williams et al. (2007) BC refractive index, $\mathrm{MAC_{BC,fresh}} = 7.2\ \mathrm{m^2.g^{-1}}$, is the closest to the reference value of $\mathrm{MAC_{BC,fresh}} = 7.5 \pm 1.2\ \mathrm{m^2.g^{-1}}$ (Bond and Bergstrom, 2006) from the literature. Therefore, it seemed interesting to include this refractive index in our sensitivity study.

Our sensitivity study, including 96 WRF-Chem experiments, took several months to run and to analyze results. The final result (BCx2, OC/2.5 and 2.5% BrOC) might not be the finest result but represents our best match.

As explained in the revised manuscript, a score is computed for each WRF-chem numerical simulation. The score is equal to the number of times the result of the simulation is included within the range of the retrieval uncertainties. This score is computed over all the pixels for the two selected areas (with and without clouds) and over all available spectral aerosol parameters (i.e., AOD, SSA, 4 spectral bands for AERO-AC and 6 for GRASP). As a result, with the BC refractive index of Bond and Bergstrom (2006), we were unable to find a realistic solution of WRF-Chem within the uncertainties of APIFLAMEv1 (Turquety et al. 2014). With the BC refractive index of Williams et al. (2007), this optimized configuration of WRF-Chem was obtained.

2b. By the way, I think table 4 would be more informative if the 'score' you used in section 3.3 were also included. And regarding how that score is calculated: why not use statistically appropriate parameters such as the Mean Absolute Error (MAE) or Root-Mean Squared Error (RMSE)?

AC: This is an interesting suggestion but we kept Table 4 as it for clarity and conciseness. Indeed, it would require to add three more lines per numerical experiment, the scores being different for each proportion of brown carbon. The scores that we calculated allowed us to establish a ranking of the numerical simulations in descending order, no numerical simulation having obtained exactly the same score. In addition, the higher the score, the more the bias is reduced between the WRF-Chem numerical simulation and the POLDER-3 retrievals. This method of calculation being sufficient, we were able to do the ranking without using these statistical parameters (MAE and RMSE).

RC1: 3. Another issue relates to how you visualize and assess the differences between POLDER and the various versions of WRF-Chem. I like the images in, for example, Fig. 7, 8, 11, etc. since they give a good intuitive understanding of the spatial context in comparing the two. However, it is qualitative. Since you've aggregated the POLDER data to the coarser WRF-Chem model grid, you have the ability to do a pixel by pixel comparison – in which case you can calculate and visualize more statistics. In most cases, you present a mean difference between POLDER and WRF-Chem, when instead you could show a histogram of differences, and make a mean-bias ("Bland-Altman") or some other sort of plot. Please just don't use the (unfortunately common) scatterplot and calculate correlation coefficients, as that would be statistically inappropriate (see for example, Altman and Bland, 1983, https://doi.org/10.2307/2987937). One of the reasons this matters is because AOD (and presumably ACAOD) is not normally distributed (Sayer and Knobelspiesse, 2019, https://doi.org/10.5194/acp-19-15023-2019).

AC: Thank you for this interesting suggestion. As suggested, we included a statistical part with histograms to Fig. 8, for example, in the revised manuscript.

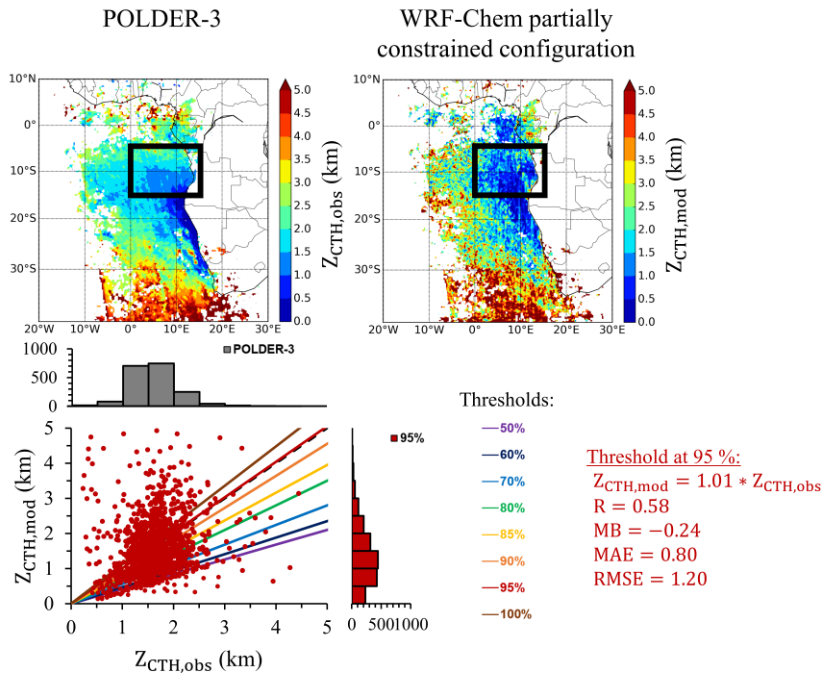


Figure 8: Monthly averaged cloud top heights retrieved by POLDER-3 (top, left) and simulated with the WRF-Chem partially constrained configuration with the threshold of 95 % (top, right) for July 2008, with an associated statistical analysis of the threshold values of the cloud top height tested in the model (bottom). The black frames (0°E-15°E, 15°S-5°S) represent the study area.

This revised Fig. 8 now better helps to understand the choice of the threshold of 95 % in the WRF-Chem partially constrained configuration.

RC1: 4. Essentially, you're only assessing half of one month of one year. I realize that many of your choices are driven by computational resources, but how do you expect these results to fit within the annual cycle of BB in southern Africa, and how relevant is it considering inter-annual variability? I was expecting some discussion on this. These results need to be expressed in terms of their general usefulness. Perhaps a few words on computation resources needed for this analysis would be useful too, should somebody want to scale up to a full year or years.

AC: We extended the study period until September 30, 2008 with the WRF-Chem optimized configuration to test its ability to represent the absorption cycle observed during the fire season of 2008. Figure A shows the temporal evolutions of the ACAOD (left) and ACSSA (right) retrieved by POLDER-3/AERO-AC (solid line curves) and simulated with the WRF-Chem optimized configuration (dotted curves) above clouds at 490, 550 and 865 nm (blue, green and red curves, respectively), during the July-August-September 2008 quarter.

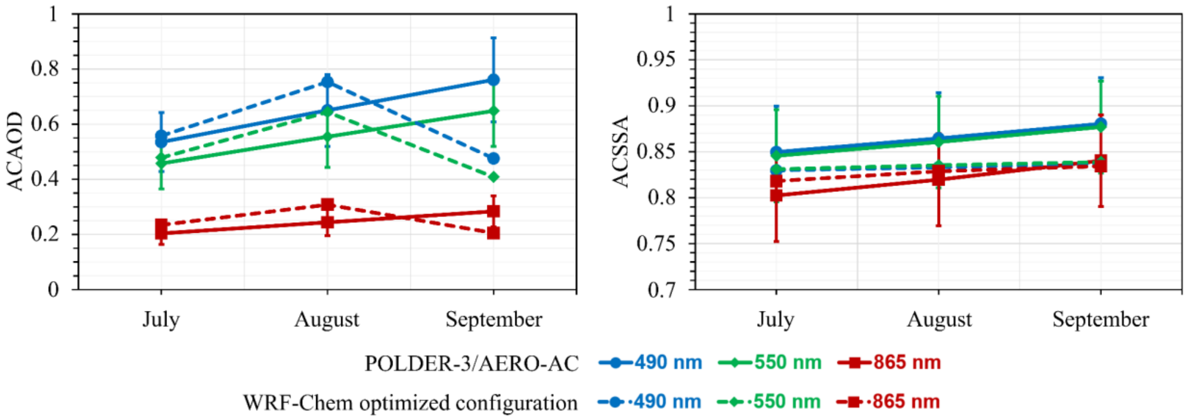


Figure A: Temporal evolutions of the ACAOD (left) and the ACSSA (right) retrieved by POLDER-3/AERO-AC (solid line curves) and simulated with the WRF-Chem optimized configuration (BCx2, OC/2.5 with 2.5% of BrOC, dotted curves) above clouds over the Southeast Atlantic Ocean at 490, 550 and 865 nm (blue, green and red curves, respectively). The optical properties are averaged geographically within the coordinate study area of (0°E-15°E, 15°S-5°S) and monthly over the July-September 2008 period. The vertical error bars correspond to the uncertainties associated with POLDER-3 data with an accuracy of ±20 % (per pixel estimate) on the aerosol optical depth and ±0.05 (per pixel estimate) on the single scattering albedo over the entire spectral domain.

On the one hand, Figure A shows that the WRF-Chem optimized configuration reproduces in part at least the increase in ACAOD retrieved by POLDER-3/AERO-AC between July and August, but does not reproduce the variability observed between August and September 2008. For September 2008, the WRF-Chem optimized configuration fails to adequately simulate the aerosol concentration peak which is typically observed at this time of year over the Southeast Atlantic region. Indeed, the WRF-Chem optimized configuration underestimates the amounts of aerosols transported over the marine stratocumulus clouds with an average bias of -0.24 for ACAOD at 550 nm compared to the POLDER-3/AERO-AC retrievals in September 2008. This underestimation could be related to uncertainties about the evolution of biomass burning emissions during the season in the inventory used (APIFLAMEv1, Turquety et al. 2014). Recently, a numerical sensitivity study was conducted by Pan et al. (2019) testing six global biomass burning emissions inventories (excluding APIFLAMEv1) in the NASA's GEOS-Chem chemistry-transport model. In particular, this modelling study focused on the South African continent in September 2008. The results of the study showed an underestimation of the simulated AODs over land for these six inventories, with a maximum bias of 0.23 compared to MODIS-Aqua data at 550 nm and at least 50% compared to data from the AERONET site in Mongu, Zambia. This discrepancy between modelling and observations could not be explained by different synoptic conditions because the meteorology and aerosol mechanisms (transport and removal) were identical for these six numerical experiments. The authors of this study therefore suggest that estimates of the amounts of aerosols emitted by biomass burning should be revised high, in particular by updating the emission factors according to the season and the conditions of activity of the fires.

On the other hand, Figure A shows that the WRF-Chem optimized configuration struggles to represent very satisfyingly the seasonal cycle of ACSSA retrieved by POLDER-3/AERO-AC, which is characterized by a decrease in aerosol absorption (increase in ACSSA values) during the advance of the fire season. This result suggests that the chemical composition of the BBA plumes, especially the carbonaceous aerosols, changes during the dry season, likely in relation to a change in the type of fuel burned and in combustion conditions (Zuidema et al. 2016b; Pan et al. 2019). It is therefore necessary to study the aerosols physico-chemical and optical properties on a month-by-month basis in order to better understand and constrain the BBA absorption cycle over the South-East Atlantic region in WRF-Chem.

This point is now discussed in more details in the conclusion section of the revised manuscript. We added the following sentences at line 127, page 32: "First results show that the WRF-Chem optimized configuration struggles to represent very satisfyingly the seasonal cycle of ACSSA retrieved by POLDER-3/AERO-AC, which is characterized by a decreased aerosol absorption (increase in ACSSA values) during the advance of the fire season. This suggests that the chemical composition of the BBA plumes, especially the carbonaceous aerosols, changes during the dry season, likely in relation to a change in the type of fuel burned and in combustion conditions (Zuidema et al. 2016b; Zuidema et al. 2018; Pan et al. 2019; Pistone et al. 2019).".

RC1: 5. Is the RETRO anthropogenic emissions inventory, which doesn't cover the year of your analysis, the best source for that information?

AC: We chose the RETRO global anthropogenic emissions inventory for its higher temporal resolution (monthly) than other global anthropogenic emissions inventories (generally yearly) such as EDGARv4.2 (Janssens-Maenhout et al. 2010). We agree with the reviewer that RETRO does not cover the specific year of our analysis, but this long-term climatology still remains representative of the anthropogenic influence during the African dry season.

RC1: 6. When considering POLDER measurement uncertainty, please recall that this often refers to per pixel estimates. In that sense, are the error bars in Figure 14 appropriate, since they are applied to geographically averaged values?

AC: The reviewer is right that uncertainties are for each pixel. In our approach, only the data with the lowest uncertainties are kept and these uncertainties (±15% for AOD and ±0.05 for SSA) are the same for each pixel. Hence the averaging procedure does not change our results. This point is now clarified in the revised manuscript.

RC1: More minor points:

RC1: 1. Why are the reference wavelengths for SSA and AOD different for the POLDER/GRASP and above cloud products different? I agree that they're close enough to not matter, but it should be explained.

AC: The POLDER-3/AERO-AC product does not use the wavelength at 565 nm because it is not a polarized channel. For the POLDER-3/AERO-AC retrievals, the 550 nm channel was added as a request of modelers who typically use this wavelength in the visible range for the evaluation of simulated aerosol parameters. Considering that the absorption of aerosols does not depend on the wavelength between 490 and 865 nm in POLDER-3/AERO-AC, the aerosol optical properties can be calculated at 550 nm with the Mie's theory, knowing the size of the particles, their absorption and their optical thickness (at 865 nm).

RC1: 2. Page 2, line 45: Clarify if "BBA direct radiative forcing" refers or global or SEAO.

AC: “BBA direct radiative forcing” refers to the SEAO. We added “Over the SEAO” at the beginning of the sentence, line 45, page 2.

RC1: 3. Page 14, line 125: I think you mean “specialty” not “specificity”.

AC: Indeed, we mean “specialty”.

RC1: 4. Table 2 and discussion in text. I’m a little confused what you mean by “range of uncertainties” here. Does this really mean range of the values reported in the literature?

AC: You are right. This really means range of the values reported in the literature. We replaced “range of uncertainties” by “range of values” in the article.

RC1: 5. Page 21, lines 507-8: Is the really the primary source of the differences, or could it also be due to the nature of your WRF-Chem constraint as well?

AC: The SSA of aerosols depends on both their size and chemical composition. In section 3.2.3 entitled “Particle size”, we checked that the size of aerosols simulated by WRF-Chem is realistic in comparison with the POLDER-3 retrievals, without any modification of the physico-chemical processes in the model. Furthermore, the relative humidity is another important factor that may impact the SSA. In section 3.1 entitled “Meteorology in the WRF-Chem initial configuration (S0)”, we ensured that the humidity is correctly simulated by WRF-Chem in comparison with the ECMWF reanalysis, with spatial correlation coefficients higher than 0.93 and very low mean errors, mean absolute errors and root mean square errors values. Therefore, we made the assumption that the chemical composition of aerosols is the primary source of the differences on the SSA between WRF-Chem and the POLDER-3 retrievals. Finally, one important key parameter included in the aerosol chemical composition is the aerosol mixing state. In our study, we only considered the homogeneous internal mixing state of particles in our WRF-Chem numerical simulations as in the POLDER-3 satellite inversion algorithms. The core-shell configuration, which appears to be important for very old BBA plumes transported off the coast of Southern Africa (Zuidema et al. 2018; Denjean et al. 2020; Taylor et al. 2020; Wu et al. 2020), could also have an influence on BBA optical properties and it would deserve to be tested.

RC1: 6. Page 25, line 577: Is the score calculated for AOD, SSA, AND the top of atmosphere spectral observations? If so, the latter are used to derive the former, so the logic seems a bit circular.

AC: The score is calculated for the AOD(λ) and the SSA(λ) over the clear-sky area (see Fig. 2, left) based on the POLDER-3/GRASP retrievals, and for the ACAOD(λ) and the ACSSA(λ) over the cloudy area (see Fig. 2, right) based on the POLDER-3/AERO-AC retrievals.

RC1: 7. Page 29, line 657: You mention here and elsewhere a “spatial correlation coefficient” but give no description of what you mean by this or how it is calculated. Similarly on Page 30, line 687. Are those error metrics what I’m calling for in 3? Regardless, they need to be explained.

AC: Thank you for this suggestion. This point is now clarified in the revised manuscript. Those errors metrics are the statistical parameters you are calling for in question 3. We added an appendix A to the revised manuscript describing the statistical parameters that we used in our study. We also added the following sentence at line 331, page 12: "The performance of the WRF-Chem numerical simulations will be assessed using commonly used statistical parameters (see Appendix A) (Thunis et al. 2011; Lingard et al. 2013): the Pearson correlation coefficient (R), the mean bias (MB), the mean absolute error (MAE) and the root mean square error (RMSE)."

Appendix A: Description of the statistical parameters

The Pearson correlation coefficient (R) is calculated as follows:

$$R = \frac{\sum_{i=1}^{N}(M_i - \overline{M})\,(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N}(M_i - \overline{M})^2}\,\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}}$$

M_i, O_i, $\overline{M}$, $\overline{O}$ et N are respectively the simulated value, the observed value, the average of the simulated and observed values, and the total number of pixels. R is between -1 and +1, with a value tending towards zero indicating no linear correlation.

The mean bias (MB) is the average of the difference between a simulated and an observed value over an area or over a specified period:

$$MB = \frac{\sum_{i=1}^{N}(M_i - O_i)}{N}$$

If the difference between the simulated value and the observed value is non-zero then the simulated value is said to be biased, positively in the case of overestimation or negatively otherwise.

The mean absolute error (MAE) is calculated from the absolute value of the difference between a simulated and an observed value:

$$MAE = \frac{\sum_{i=1}^{N}|M_i - O_i|}{N}$$

This quantity is therefore always positive and tends towards zero when the simulated values are close to those observed. This statistical parameter is more restrictive than the average bias because it avoids possible error compensations.

The root mean square error (RMSE) is calculated from the square root of the root mean square difference between the simulated and observed values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(M_i - O_i)^2}{N}}$$

Point but large biases can produce a high RMSE value. This statistical parameter is therefore commonly used as a measure of the overall performance of a model.

AC: References

Bond, Tami C., et Robert W. Bergstrom. 2006. « Light Absorption by Carbonaceous Particles: An Investigative Review ». Aerosol Science and Technology 40 (1): 27-67. https://doi.org/10.1080/02786820500421521.

Denjean, Cyrielle, Thierry Bourrianne, Frederic Burnet, Marc Mallet, Nicolas Maury, Aurélie Colomb, Pamela Dominutti, et al. 2020. 'Overview of Aerosol Optical Properties over Southern West Africa from DACCIWA Aircraft Measurements'. Atmospheric Chemistry and Physics 20 (8): 4735–56. https://doi.org/10.5194/acp-20-4735-2020.

Flaounas, Emmanouil, Vassiliki Kotroni, Konstantinos Lagouvardos, Martina Klose, Cyrille Flamant, and Theodore M. Giannaros. 2016. 'Assessing Atmospheric Dust Modelling Performance of WRF-Chem over the Semi-Arid and Arid Regions around the Mediterranean'. Atmospheric Chemistry and Physics Discussions, May, 1–28. https://doi.org/10.5194/acp-2016-307.

G. Janssens-Maenhout, A. M.R. Petrescu, M. Muntean & V. Blujdea (2011) Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, Greenhouse Gas Measurement and Management, 1:2, 132-133, DOI: 10.1080/20430779.2011.579358

Lingard, Justin, Lorenzo Labrador, Daniel Brookes, et Andrea Fraser. 2013. « Statistical Evaluation of the Input Meteorological Data Used for the UK Air Quality Forecast (UK-AQF) », no 1: 34.

Pan, Xiaohua, Charles Ichoku, Mian Chin, Huisheng Bian, Anton Darmenov, Peter Colarco, Luke Ellison, et al. 2019. 'Six 1100 Global Biomass Burning Emission Datasets: Inter-Comparison and Application in One Global Aerosol Model'. Preprint. Aerosols/Atmospheric Modelling/Troposphere/Physics (physical properties and processes). https://doi.org/10.5194/acp-2019-475.

Taylor, Jonathan W., Huihui Wu, Kate Szpek, Keith Bower, Ian Crawford, Michael J. Flynn, Paul I. Williams, et al. 2020. 'Absorption Closure in Highly Aged Biomass Burning Smoke'. Atmospheric Chemistry and Physics 20 (19): 11201–21. https://doi.org/10.5194/acp-20-11201-2020.

Thunis, P., Emilia Georgieva, et Stefano Galmarini. 2011. A procedure for air quality models benchmarking, Version 2.

Turquety, S, Laurent Menut, B Bessagnet, Alessandro Anav, Nicolas Viovy, Fabienne Maignan, and M Wooster. 2014. APIFLAME v1.0: High-Resolution Fire Emission Model and Application to the Euro-Mediterranean Region. Vol. 7. https://doi.org/10.5194/gmd-7-587-2014.

Williams, T. C., C. R. Shaddix, K. A. Jensen, et J. M. Suo-Anttila. 2007. « Measurement of the dimensionless extinction coefficient of soot within laminar diffusion flames ». International Journal of Heat and Mass Transfer 50 (7): 1616-30. https://doi.org/10.1016/j.ijheatmasstransfer.2006.08.024.

Wu, Huihui, Jonathan W. Taylor, Kate Szpek, Justin M. Langridge, Paul I. Williams, Michael Flynn, James D. Allan, et al. 2020. 'Vertical Variability of the Properties of Highly Aged Biomass Burning Aerosol Transported over the Southeast Atlantic during CLARIFY-2017'. Atmospheric Chemistry and Physics 20 (21): 12697–719. https://doi.org/10.5194/acp-20-12697-2020.

Zuidema, Paquita, Jens Redemann, James Haywood, Robert Wood, Stuart Piketh, Martin Hipondoka, and Paola Formenti. 2016b. 'Smoke and Clouds above the Southeast Atlantic: Upcoming Field Campaigns Probe Absorbing Aerosol's Impact on Climate'. Bulletin of the American Meteorological Society 97 (7): 1131–35. https://doi.org/10.1175/BAMS-D-15-00082.1.

Zuidema, Paquita, Arthur J. Sedlacek, Connor Flynn, Stephen Springston, Rodrigo Delgadillo, Jianhao Zhang, Allison C. Aiken, Annette Koontz, and Paytsar Muradyan. 2018. 'The Ascension Island Boundary Layer in the Remote Southeast Atlantic Is Often Smoky'. Geophysical Research Letters 45 (9): 4456–65. https://doi.org/10.1002/2017GL076926.