Title: Observed Impacts of Aerosol Regimes on Energy and Carbon Fluxes in the Amazon Forest

Response (blue color) to anonymous Referee #3 (black). The original manuscript was changed accordingly. The lines indicated in our answers correspond to the track version of the manuscript.

General comment

This comment was prepared as part of MSc course work at Wageningen University under supervision of Prof Wouter Peters. They were uploaded as a comment as they were regarded to be of good quality, and likely helpful to the authors and editor in the review process.

We would like to express our sincere gratitude to Professor Wouter Peters and his students for their interest in our manuscript. Their comments help us to improve our results and discussion.

This study examines how aerosol regimes affect energy and carbon fluxes in a pristine central Amazon forest. Using 2016–2022 meteorological and flux data from the Amazon Tall Tower Observatory (ATTO) and AOD (500 nm) from AERONET, it tests whether aerosol loading alters latent heat (LE), net radiation (Rn), and CO_2 fluxes (FCO_2). The study focuses on the dry season (August–November), when biomass burning elevates aerosol concentrations across the southern Amazon Basin. The authors define two aerosol regimes: clean (AOD < 0.13) and polluted (AOD > 0.40), consistent with previous studies such as Steiner et al. (2013) and Ross Herbert & Stier (2023). This threshold-based approach, derived from data percentiles, provides a simple yet robust framework for distinguishing contrasting aerosol loading conditions. Their analysis focuses on the 10:00–14:00 LT period to examine energy partitioning under contrasting aerosol regimes. Authors interestingly present, VPD vs Temperature (Figure 4), a combination of variables that I have not encountered in other studies reviewed during this process. It is particularly valuable, as it effectively illustrates—through the observed variables of VPD and temperature—the realistic delay caused by reduced shortwave incoming radiation (SWin) during polluted periods. They report a delay in the rise of temperature and VPD under polluted conditions, highlighting the moderating effect of aerosols. They conclude by confirming the well-documented finding that, paradoxically, CO_2 uptake increases under polluted conditions—by about 57.7% in this case—due to the diffuse radiation effect, where scattered sunlight enhances photosynthesis within shaded canopy layers. This result is in strong agreement with previous studies on Amazonian aerosol dynamics, particularly Rodrigues et al. (2024) and Cirino et al. (2014), which similarly observed elevated carbon uptake under high-AOD conditions. The study concludes by emphasizing the nonlinear and complex interactions between AOD and surface fluxes, demonstrated through MANCOVA and Random Forest Model analyses, underlining however the need for further investigation.

Remarks on several aspects:

(1) Midday Averaging

The authors assess the effects of aerosols on surface energy and carbon fluxes by averaging 30-minute flux measurements over the 10:00–14:00 LT period and then calculating percentage reductions between clean and polluted aerosol regimes. This time window is identified as representing the period of strongest radiative and convective activity (line 173). However, the diurnal cycle plots (Figs. 5 and 6) reveal uneven flux patterns, with noticeable uninvestigated areas (Figure 5 & 6 "white spaces", outside 10:00–14:00 LT window) within both the clean and polluted regimes. As a starting point, Figure 4 clearly shows a delay in the increase of temperature and vapor pressure deficit (VPD). Because natural processes evolve non-uniformly throughout the day, using a short and non-equidistant time subset which may bias the calculated percentage reductions and misrepresent the actual aerosol influence. The paper's methodology follows Steiner et al. (2013), who also analyzed fluxes over the 10:00–14:00 LT period and compared similar aerosol optical depth (AOD) regimes (AOD < 0.3 vs. > 0.5). However, within the text, fluxes reductions' comparisons are made with studies that employed different approaches to assess aerosol-load effects. For example, Rodrigues et al. (2024) and Cirino et al. (2014) estimated flux reductions under specific irradiance conditions, distinguishing Solar Zenith Angle (SZA) zones and thereby incorporating the time-of-day variability, rather than relying on a fixed midday average. A closer examination of Figures 5 and 6, which depict sensible, latent, and ground heat fluxes, reveals an interesting but unexamined pattern during i.e. the morning transition (06:00–10:00 LT). Both H and G occasionally exceed their respective values under polluted conditions, while the consistent dominance of the clean regime in LE appears to be underestimated. Early-morning CO_2 uptake (Figure 6) also exhibits a more dynamic behavior, with pronounced transitions between clean and polluted regimes. To better capture the full evolution of the phenomena and associated fluxes, the authors could integrate the area under the fluxes' curves over the 06:00–17:00 LT period and compare the resulting averages between the clean and polluted aerosol regimes. Alternatively, if there is sufficient data outside the window 10:00-14:00 LT the authors could consider reporting morning (06:00-10:00 LT) and afternoon sub-period (14:00-17:00 LT) averages separately to capture diurnal variability better. Analyzing relative irradiance would require substantially more methodological development and investigation by the authors; therefore, it is not recommended.

We thank the Referee #3 for this comment. We agree that surface fluxes exhibit variable features outside the 10:00–14:00 LT window. However, during these periods, flux variability may be influenced by boundary-layer dynamics and low solar elevation angles, which can affect H, LE, and FCO_2 and complicate the isolation of the radiative effects of aerosols. Moreover, radiometer uncertainties (typically within ~5%) are less significant when radiation levels are high. At low solar elevation angles (early morning and late afternoon), radiation magnitudes are smaller, which increases the relative importance of measurement errors and energy balance closure uncertainties. For these reasons, the 10:00–14:00 LT period provides more favorable conditions for isolating aerosol-induced radiative effects.

In the revised manuscript, we have added Figure 4 showing the full diurnal cycles of shortwave, longwave, and net radiation during the dry season (2016–2022). This figure demonstrates that peak net radiation consistently occurs between 10:00 and 14:00 LT, supporting our choice of this time window.

(2) Gaps Manipulation

The authors state that their initial dataset comprised 10,890 half-hourly observations (line 87), which, after several filtering steps, was reduced to 523 rows—of which only 370 belong to the dry season (lines 94–96). However, the paper does not clarify how these 10,890 records were originally obtained. Figure 2 further raises questions about data representativeness and statistical treatment: the monthly boxplots show means much higher than medians, indicating positive skewness, while the number of valid data points per month is not reported. The data filtering process is clearly described, resulting in 523 rows of 30-minute averaged meteorological, flux, and AOD values. However, the dataset distribution across years is highly uneven, as also noted by the authors (line 97: “The distribution…effects of aerosol”). Specifically, years contributing less than 5 % of the total dataset are treated equivalently to years such as 2020 and 2022 (42,4% and 29,2% data coverage respectively), despite potentially different atmospheric and surface conditions. This raises concerns regarding the robustness of the study’s conclusions. Evidentially, no quantitative assessment of data representativeness or uncertainty is provided. Similar studies (e.g., Schmitt et al., 2023) have explicitly visualized monthly data availability and included “fraction of missing data”. Moreover, the extremely low number of data rows for certain years warrants further examination, as such sparse temporal coverage could substantially affect the robustness of the Random Forest Model (RFM) used later in the statistical analysis. Limited data availability may lead to overfitting, biased feature importance when training and validation subsets are unevenly represented. It is recommended that the authors include the fraction of valid rows per month, which could be directly incorporated into Figure 2. Furthermore, the manuscript should clearly describe the origin of the initial 10,890 observations—specifying the time period covered, sampling frequency, and measured variables—to better contextualize the subsequent data filtering process.

We thank the Referee #3 for these important comments. They will certainly help to improve the methodology and discussion of the results.

We would like to begin our responses by stating that in the new version of the manuscript, we regrouped our data in a way that allowed us to include a greater number of runs (half-hour periods). In the previous version of the manuscript, in addition to excluding all periods when clouds were present, which is very common in the Amazon, we also excluded all data from a given day and time when a variable was missing. For example, if we did not have the reflected shortwave radiation measurement for a given time, we removed all other variables for that same time. This resulted in only 523 valid half-hour periods (370 dry season, 153 wet). In the new version of the Manuscript, we decided to regroup the variables so that they did not depend on each other. We first identified the periods in which we had the Clean and Polluted regimes and then identified how many runs of each variable were available for each regime. After this procedure, the number of runs increased substantially, as shown in Table R1, comparison between the dataset used in the first version of the manuscript (single database) and the dataset used for this new version (database by variable).

Table R1: Number of runs (half-hour periods) after all quality controls mentioned in section 2.2.

Variables	Single database					Database by variable				
	10:00 -14:00 LT		07:00 -17:00 LT		Total	10:00 -14:00 LT		07:00 -17:00 LT		Total
	No. Clean	No. Polluted	No. Clean	No. Polluted	No. Sample	No. Clean	No. Polluted	No. Clean	No. Polluted	No. Sample
SWin(Wm^{-2})	98	81	219	151	370	301	204	736	459	1195
SWout(Wm^{-2})	98	81	219	151	370	301	204	736	459	1195
LWatm(Wm^{-2})	98	81	219	151	370	301	200	733	453	1186
LWterr(Wm^{-2})	98	81	219	151	370	301	204	735	459	1194
Rn(Wm^{-2})	98	81	219	151	370	301	200	733	453	1186
H(Wm^{-2})	98	81	219	151	370	197	192	455	389	844
LE(Wm^{-2})	98	81	219	151	370	183	180	447	386	833
FCO_2($\mu mol m^{-2} s^{-1}$)	98	81	219	151	370	247	195	596	405	1001
G(Wm^{-2})	98	81	219	151	370	301	218	741	487	1228

The initial number of 10,890 observations does not represent the full raw eddy-covariance dataset, which contains 122,734 half-hourly records over 2016–2022. Instead, this number corresponds to the subset of 30-minute periods for which aerosol optical depth (AOD) data from AERONET (version 3, level 2) were available and could be matched with surface flux and radiation measurements. The text has been updated accordingly to improve clarity (Section 2.2 in the revised version of the manuscript).

L95-107: "*To eliminate cloud interference and investigate the role of aerosols in surface energy fluxes, the central objective of this study, we used data from the Aerosol Robotic Network (AERONET) at the ATTO site, specifically AOD (version 3, level 2). These data are free of cloud contamination due to pre and post-field calibration (Giles et al., 2019). Based on this, 30-minute averages were calculated between 2016 and*

2022 for which AOD data from AERONET were available, the initial combined dataset comprised 10,890 observations, including all variables listed in Table 1. This matched dataset served as the starting point for the subsequent quality control and filtering procedures. First, the turbulent fluxes underwent quality control based on Foken et al. (2004). Only data with flags "0" (best quality) and "1" (acceptable for general analysis) were used; data with flag "2" (poor quality) were discarded. Second, this study only considered the daytime period (from 7:00 to 17:00 LT) because the highest Rn values occur during this time. After filtering, the resulting dataset is summarized in Table S1 and S2."

As described in the previous comment, we regrouped our data in a way that allowed us to include a greater number of runs (half-hour periods). Based on this updated dataset, Figure R1 was revised and now includes the number of samples per month (*n*). The mean values are higher than the medians, particularly during the dry season, reflecting the influence of episodic high-AOD events (e.g., biomass burning, smoke intrusions) that shift the distributions toward positive skewness. We additionally verified that the main seasonal contrasts remain qualitatively unchanged when using median AOD instead of mean AOD.

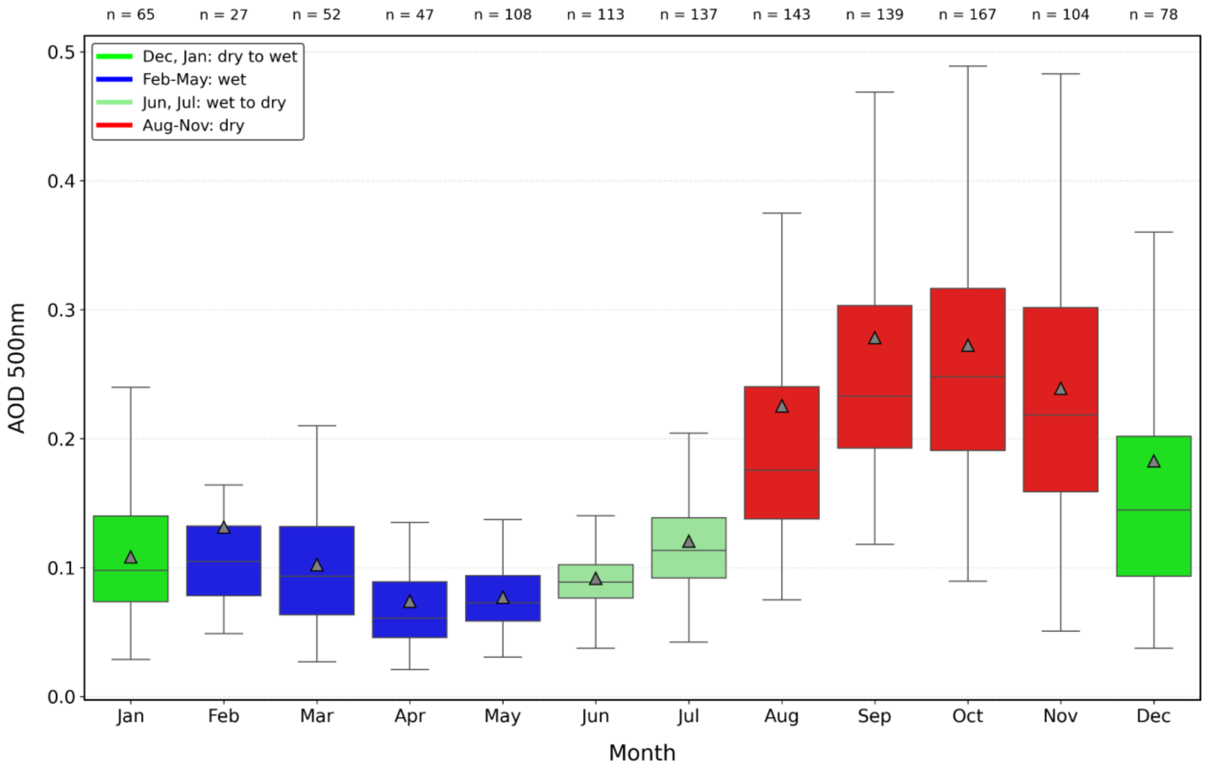

Figure R1. Box plot showing monthly AOD 500 nm values measured at the ATTO site between 2016 and 2022. The box represents the central 50% of the data, the whiskers represent the smallest and largest non-outlier values, while the means are indicated by the green triangles and the medians are the lines inside the box. Numbers above each month indicate the sample size (n) (Figure 2 in the revised version of the manuscript).

(3) Statistical Analysis

The study explores the relationship between aerosol optical depth (AOD) and surface fluxes (Rn, H, LE, FCO_2) implementing Spearman correlations, multivariate MANCOVA testing assessed by Pillai test and a Random Forest Model (RFM) to quantify nonlinear dependencies and variable importance. However, several methodological lack in processes or data-handling limitations seem to weaken the robustness of the conclusions. The manuscript provides a general introduction to the application of Pillai's test and outlines the advantages of using the Random Forest Model (RFM) to investigate nonlinear and complex interactions between variables and systems. However, it remains unclear to what extent these principles—particularly in the case of RFM—have been appropriately implemented and demonstrated in the study. In comparable RFM environmental works, such as Miao et al. (2018) and Zhang et al. (2023), linear correlation analyses were explicitly conducted to assess collinearity among key variables by providing comprehensive correlation matrices, providing direct linear insights. In contrast, Rocha et al. (2025) only briefly mention in line 272 that "the statistical relationships show low intensity or no statistical significance," without offering supporting analyses or graphical evidence. Furthermore, while Miao et al. (2018) thoroughly examined their multivariate equations and reported the statistical significance of their models and variables, Rocha et al. (2025) limit the discussion to the significance of Pillai's test (line 275), suggesting the absence of linear interactions without presenting sufficient analytical support or methodological transparency. Another major concern is data volume, as mentioned in major argument 2. Miao et al. (2018) utilized approximately 7,000 samples, and Zhang et al. (2023) worked with about 60,000 samples. In contrast, Rocha et al. (2025) rely on only 370 rows of data for the dry period, which raises serious concerns about potential overfitting of the RFM. Moreover, although the manuscript mentions a cross-validation approach in Table 3, it does not specify the technique used or report its results. Finally, the model assessment presented in Table 3 appears inadequate and leaves substantial uncertainty regarding the RFM's reliability. In the referenced studies, Miao et al. (2018) implemented multiple factor matrices, and Zhang et al. (2023) validated their models through scatter density plots and strong statistical metrics across training and testing datasets, including mean absolute error (MAE) and percentage variation analyses. Rocha et al. 2025 attempt to employ a RFM to capture the nonlinear influence of aerosols on surface fluxes. However, this approach lacks sufficient methodological justification and statistical robustness. The authors do not provide any evidence of cross-validation or other procedures to assess model generalisability. Furthermore, the dataset used for training—only 370 observations—is several orders of magnitude smaller than what is typically required for stable Random Forest performance, raising serious concerns about overfitting and the reliability of the reported metrics. Consequently, the predictive results presented in Table 3 should be interpreted with caution, as their statistical validity is uncertain. Given the limited dataset, the application of the Random Forest Model (RFM) in this study does not appear to add substantial value to the results or discussion. With such a small sample size, the model's capacity to generalise is minimal, and its predictive performance cannot be reliably validated. Moreover, the manuscript provides no detailed explanation of the model evaluation or validation procedures, which further undermines confidence in the reported outcomes. To strengthen the analysis, I suggest replacing or complementing the RFM with a correlation matrix to explicitly reveal potential collinearity among variables, particularly regarding the influence of AOD (as in Table 3). Additionally, presenting multivariate regression equations and reporting their levels of statistical significance would offer a clearer and more interpretable understanding of how other environmental factors interacts with AOD. Such an approach could also serve as a solid foundation for future studies investigating aerosol impacts on surface fluxes under polluted regimes.

We sincerely thank the MSc students at Wageningen University, under the supervision of Prof Wouter Peters, for their detailed and constructive feedback regarding our statistical methodology.

We agree with the referee and have removed the RFM analysis from the revised manuscript. We emphasize that the RFM was originally intended as a complementary exploratory tool, and its removal does not affect the main results or interpretations of the study. In the revised version, to assess whether clean and polluted regimes exhibit statistically significant differences in radiation and surface energy and CO_2 fluxes, we apply the Mann–Whitney U test, which is well suited for non-normally distributed data and unequal sample sizes. These revisions provide a clearer and more robust statistical framework to support our conclusions regarding aerosol impacts on surface–atmosphere interactions.

Minor arguments and typos:

Minor issue 1: Several sentences are poorly structured or ambiguous, leading to confusion or misinterpretation. Examples include lines 74–75, 97, 99–101, 112–113, 134, and 247–248, as well as the descriptions for Figures, especially 2 and 4, where I suggest rephrasing or clarifying.

L76: “*The climate is tropical humid and characterized by two seasons (wet and dry), driven by seasonal shifts of the Intertropical Convergence Zone over the Amazon Basin (Andreae et al., 2015).*”

L299: “*However, analysis of LE, which represents the fraction of available energy converted into evapotranspiration, shows a consistent decrease in the polluted regime compared to the clean regime (Figure 6), which contradicts this expectation.*”

Additionally, Section 2.3 (Analysis Methods) has been revised to address all the reviewer’s comments.

Minor issue 2: Several statements lack adequate justification or references, I suggest further elaboration on the statements:

Line 114: The use of a fourth-order polynomial is mentioned but not explained or visualized.

The polynomial fit shown in Figs. 4–6 was applied solely as a smoothing technique for visualization purposes. In the revised manuscript, we have included the 30-min observed data points in the figures to better illustrate data variability. This clarification has been incorporated into the manuscript as follows:

L139-141: "*To facilitate the visualization of the mean diurnal patterns, a 4th-order polynomial curve was applied exclusively as a smoothing technique to the observational data. This curve fitting was used solely for graphical purposes and does not represent a physical or predictive model. All analyses were based on the measured data.*"

Lines 135–136: Require citation or elaboration.

We appreciate your suggestion, but we have removed the RFM.

Lines 220–222: Could be expanded with a brief example of the described method.

The expansion was carried out as follows:

L261-267: "*They identified a correlation between relative irradiance, air temperature, and VPD. Meanwhile, Herbert and Stier (2023) and Palácios et al. (2024) reinforce the idea that AOD significantly influences temperature variations, particularly on a regional scale. For instance, Palácios et al. (2024) observed positive linear correlations between AOD and air temperature across distinct climatic phases, attributed to the absorption of solar radiation by biomass burning emissions resulting in atmospheric heating. Similarly, Herbert and Stier (2023) utilized reanalysis data to demonstrate that 2-meter air temperature increases as a function of AOD, consistent with localized heating of the smoke layer due to strong absorption of solar radiation.*"

Minor Issue 3: Some methodological descriptions (e.g., line 12 in the Abstract; lines 87–90 on data filtering; lines 137–144 on the RFM methodology) could be condensed, as they do not add substantial value to the manuscript.

We thank the referee for this comment. The RFM analysis has been removed from the revised manuscript, as detailed in lines 251-259 of this document.

Minor Issue 4: GPP is mentioned in the Abstract and Conclusion but is neither discussed nor analyzed in the main text.

The Abstract and Conclusion has been updated, and references to GPP have been removed in the revised manuscript.

Minor Issue 5: The manuscript refers to two towers at the ATTO site but does not specify which tower's data are used in the analyses and figures.

Thanks for this comment. The text has been revised to specify that the analyses are based on data from the Instant Tower (81 m).

L66: "*The data used in this study were collected as part of the ATTO project, a bilateral initiative between Brazil and Germany. Since 2012, ATTO has carried out continuous measurements, as described by Andreae et al. (2015), located in an area of pristine tropical forests in the central Amazon (Figure 1), which contains the Instant Tower of 81 meters (-2.1441°S, -58.9999°W).*"

P1, line 12: The last sentence of the Abstract adds no clear value to the manuscript and could be removed.

The text has been removed from the abstract in the new version of the manuscript.

P3, line 81: Change LiCor to LI-COR for correct company citation.

The text has been updated accordingly. Thanks.

P5, line 112: The text states that hourly averages are used, while figures show 30-minute values—this inconsistency should be corrected.

Section 2.3 (Analysis Methods) has been revised accordingly.

P14, Table 3 description: Typo — change FCO to FCO_2.

The table description has been updated accordingly. Thanks.

P15, line 312: Typo — change aerossol to aerosol.

The text has been updated accordingly. Thanks.

References:

Andreae, M. O., Acevedo, O. C., Araùjo, A., Artaxo, P., Barbosa, C. G. G., Barbosa, H. M. J., Brito, J., Carbone, S., Chi, X., Cintra, B. B. L., da Silva, N. F., Dias, N. L., Dias-Júnior, C. Q., Ditas, F., Ditz, R., Godoi, A. F. L., Godoi, R. H. M., Heimann, M., Hoffmann, T., Kesselmeier, J., Könemann, T., Krüger, M. L., Lavric, J. V., Manzi, A. O., Lopes, A. P., Martins, D. L., Mikhailov, E. F., Moran-Zuloaga, D., Nelson, B. W., Nölscher, A. C., Santos Nogueira, D., Piedade, M. T. F., Pöhlker, C., Pöschl, U., Quesada, C. A., Rizzo, L. V., Ro, C.-U., Ruckteschler, N., Sá, L. D. A., de Oliveira Sá, M., Sales, C. B., dos Santos, R. M. N., Saturno, J., Schöngart, J., Sörgel, M., de Souza, C. M., de Souza, R. A. F., Su, H., Targhetta, N., Tóta, J., Trebs, I., Trumbore, S., van Eijck, A., Walter, D., Wang, Z., Weber, B., Williams, J., Winderlich, J., Wittmann, F., Wolff, S., and Yáñez-Serrano, A. M.: The Amazon Tall Tower Observatory (ATTO): overview of pilot measurements on ecosystem ecology, meteorology, trace gases, and aerosols, Atmospheric Chemistry and Physics, 15, 10 723–10 776, https://doi.org/10.5194/acp-15-10723-2015, 2015.

Cirino, G. G., Souza, R. A. F., Adams, D. K., and Artaxo, P.: The effect of atmospheric aerosol particles and clouds on net ecosystem exchange in the Amazon, Atmospheric Chemistry and Physics, 14, 6523–6543, https://doi.org/10.5194/acp-14-6523-2014, 2014.

Foken, T., Göockede, M., Mauder, M., Mahrt, L., Amiro, B., and Munger, W.: Post-Field Data Quality Control, pp. 181–208, Kluwer Academic Publishers, ISBN 1402022646, https://doi.org/10.1007/1-4020-2265-4_9, 2004.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database– automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmospheric Measurement Techniques, 12, 169–209, https://doi.org/10.5194/amt-12-169-2019, 2019.

Miao, S., Zhang, X., Han, Y., Sun, W., Liu, C. and Yin, S.: Random Forest Algorithm for the Relationship between Negative Air Ions and Environmental Factors in an Urban Park, Atmosphere, vol. 9, no. 12, Art. no. 463, doi:10.3390/atmos9120463, 2018.

NASA JPL: NASADEM Merged DEM Global 1 arc second V001, [Data set], https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001, accessed 2020-12-30, 2020.

Palácios, R., Castagna, D., Barbosa, L., Souza, A. P., Imbiriba, B., Zolin, C. A., Nassarden, D., Duarte, L., Morais, F. G., Franco, M. A., Cirino, G., Kuhn, P., Sodré, G., Curado, L., Basso, J., Roberto de Paulo, S., and Rodrigues, T.: ENSO effects on the relationship between aerosols and evapotranspiration in the south of the Amazon biome, Environmental Research, 250, 118 516, https://doi.org/10.1016/j.envres.2024.118516, 2024.

RAISG: Geospatial Vetor Data of the biomes, limits and cities, [Data set], https://www.raisg.org/en/maps/, accessed: 2023-12-01, 2023.

Rodrigues, S., Cirino, G., Moreira, D., Pozzer, A., Palácios, R., Lee, S.-C., Imbiriba, B., Nogueira, J., Vitorino, M. I., and Vourlitis, G.:Enhanced net CO2 exchange of a semideciduous forest in the southern Amazon due to diffuse radiation from biomass burning, Biogeo-sciences, 21, 843–868, https://doi.org/10.5194/bg-21-843-2024, 2024.

Schmitt, A. U., Ament, F., de Araújo, A. C., Sá, M., and Teixeira, P.: Modeling atmosphere–land interactions at a rainforest site – a case study using Amazon Tall Tower Observatory (ATTO) measurements and reanalysis data, Atmospheric Chemistry and Physics, 23, 9323–9346, https://doi.org/10.5194/acp-23-9323-2023, 2023.

Steiner, A. L., Mermelstein, D., Cheng, S. J., Twine, T. E., and Oliphant, A.: Observed Impact of Atmospheric Aerosols on the Surface Energy Budget, Earth Interactions, 17, 1–22, https://doi.org/10.1175/2013ei000523.1, 2013.

Zhang, Y., Fu, R., Yu, H., Dickinson, R. E., Juarez, R. N., Chin, M., and Wang, H.: A regional climate model study of how biomass burning aerosol impacts land-atmosphere interactions over the Amazon, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007jd009449, 2008

1