Anonymous Referee #2

This paper presents a statistical analysis of various climate fields to relate the tropospheric air quality in the Southern African continent, atmospheric circulation and SST.

The general idea is interesting and the paper is well structured (although I appreciate when figures appear in the text, where they are cited, rather than at the end of the manuscript, which makes reading rather tedious on a computer).

We thank the Reviewer for her/is interest in our work and the time s/he spent in making constructive comments, which helped us in substantially improving the quality of the manuscript. We apologise for the manuscript layout, but we have many figures with many panels and this is not easily manageable by MS Word. We provide below a point-by-point response to all the comments/requests. Reviewer's comments are in black, our responses are in blue.

My major comment is on the application of the statistical methodology. The authors seem to use ~15 years of geopotential data from the CAMS reanalysis. The rationale is that AOD data are only available on that period. But the authors use SST data that cover more than one century (and use only a small subset). I think it would be more appropriate to apply the k-means algorithm on a longer period of time (e.g. with ERA-I, ERA5, or NCEP reanalyses) to compute weather regimes in a statistically robust way, and then classify CAMS data onto such weather regimes. This would reduce the uncertainty on the computation of WRs.

We agree with the reviewer that an assessment of the robustness of the WR classification is needed. However we believe that the synoptic characterisation of the aerosol transport should be performed by using the CAMS product, which assure coherence between atmospheric circulation and aerosol data. Therefore we decided to keep using the CAMS classification to characterise the aerosol transport, and assess its robustness as you (and Referee #1) suggest. We tested: the sensitivity of the WR classification in CAMS to changes in the geographical domain and the number of retained PCs. Moreover the WR classification defined in 2003-2017 is compared with a WR classification defined in 1981-2020 by using ERA5 data. Section S2 has been added in the Supplement to discuss the sensitivity of the CAMS classification to changes in the geographical domain and the number of retained PCs, WR centroids and frequencies from the ERA5 classification have been added to Fig. 4 and 7, and a long paragraph has been added to Section 3.1 to discuss the classification in ERA5 and summarise the sensitivity tests. We can conclude that:

"The sensitivity tests performed on the WR classification of the CAMS data show a high degree of robustness with respect to changes in the time period, and a good degree of robustness with respect to changes in the geographical domain and the retained variance, highlighting that the classification well represents the main features of the synoptic circulation in the region".

The ERA5 classification is used in the revised version of the manuscript for the analysis of the intraseasonal and interannual variability of the WRs in the period 1981-2020 (see Fig. 7, 8, 12 and 13).

My second methodological suggestion is to use a cross-validation approach to E2C and C2E, by "learning" the associations between WR and AOD on a decade, and "testing/validating" this association on the remaining 5 years. This would give credence to the alleged predicting power of the statistical approach.

Thank you for this suggestion. We do agree that the predictive potential of the WR classification is worth to be investigated. However, the shortness of the AERONET time series prevent the application of the suggested cross validation approach. Please note that only in Ascension Island observations cover the whole studied period (2003-2017) with only 612 data points (40 per year on average), while the continental stations only cover 2-4 years, with at best less than 300 data points (see Table 1). However, we have tested the C2E approach with different time leads, namely 1, 2, and 4 days (see Fig. R1, R2 and R3). Results show little improvement for 1 and 2 day lead times, reflecting the 2-3 day persistence of the WRs, and ambiguous results for 4 day lead time. Predictive skill assessment of the WR classification should be done on data with larger time-space coverage (e.g. satellite products), but it is beyond the scope of this paper.

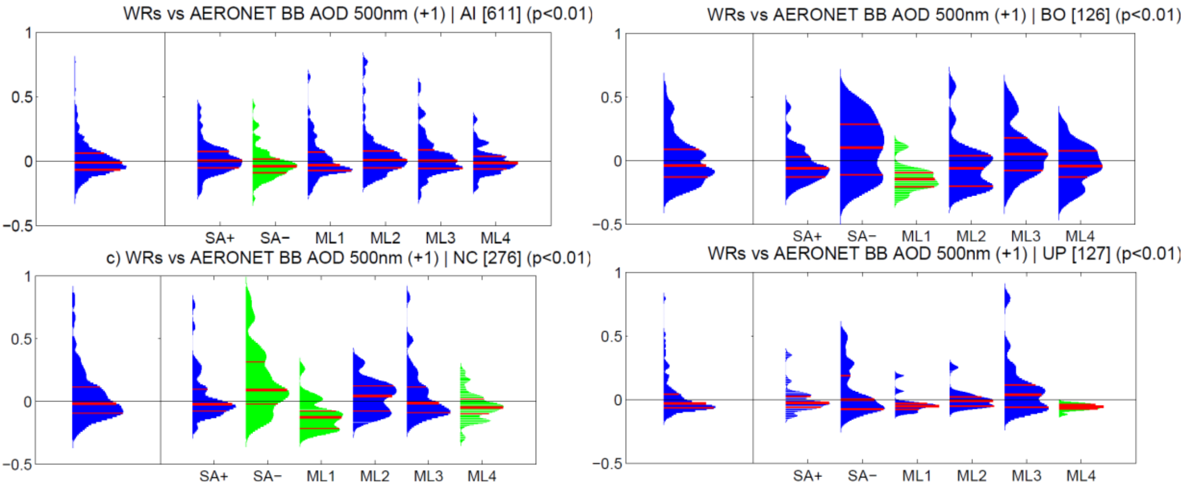

Figure R1. Circulation to environment characterisation: for each AERONET station (Ascension Island, Bonanza, Namibian Stations, Upington), (left panel) distribution of the AOD anomalies at 500 nm, and (right panel) for each CAMS WR, with 1 day lead time. Anomaly distributions significantly different from the climatological sample are displayed in green (p-values of the Kolmogorov-Smirnov test used to assess the significance of the differences are reported in Table S5). In titles, the number of available daily observations and the p-value of the ANOVA used to assess the WR characterisation are reported in brackets.

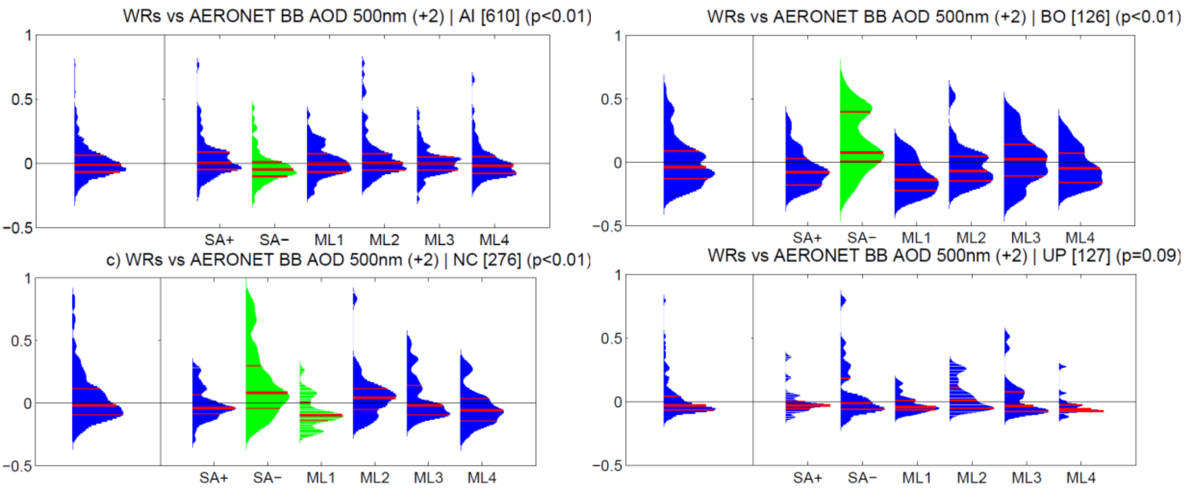

Figure R2. Same as Fig. R1, but for 2 day lead time.

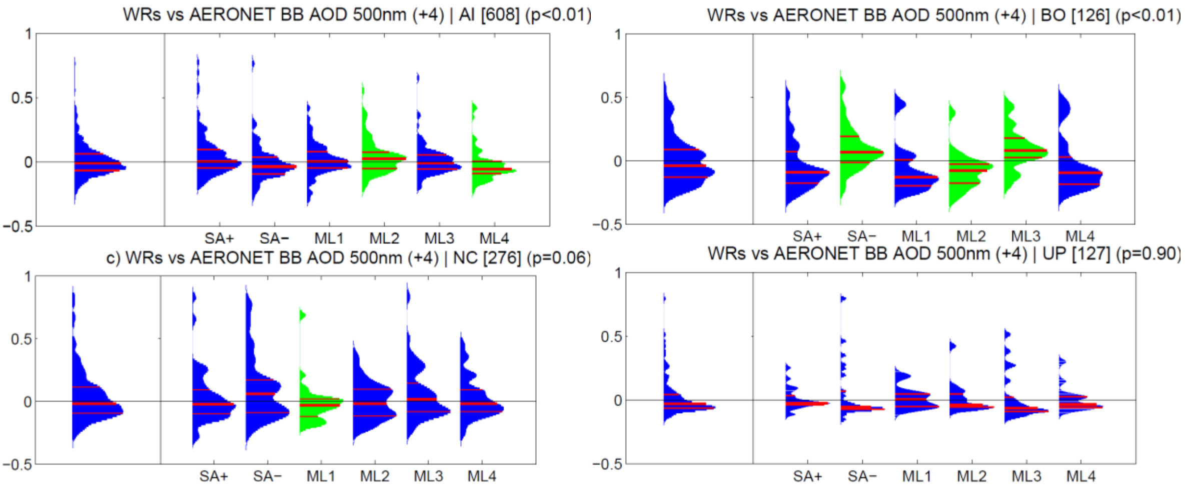

Figure R3. Same as Fig. R1, but for 4 day lead time.

Minor comments

The first paragraph of the introduction states that aerosols modify the radiative properties of the atmosphere. Fine. "As a consequence, they can influence on the atmospheric synoptic and large-scale dynamics" seems strange, as the radiative properties of aerosols are rather local, which contradicts large-scale atmospheric motion, where radiation is not so crucial. Please explain.

Thank you for this comment, the sentence is actually misleading. We referred to the role of aerosols as climate forcing, and we made the equivalence between climate and large scale dynamics, which is, we do agree, not appropriate. The sentence now reads:

"As a consequence, aerosols can influence the atmospheric and climate dynamics".

The end of the introduction lacks a paragraph that states the scientific question that the manuscript is dealing with. At present, the introduction states rather general questions, then states what the authors intend to do. How this endeavor corresponds to the many general questions seems to be left to the imagination of the reader.

The specific scientific questions we aim to respond are formulated explicitly in the Introduction:

"The scope of this paper is to fill the gaps in the understanding of atmospheric and aerosol dynamics during austral winter in the southern Africa/South Atlantic sector, by providing a characterisation of the synoptic variability of the atmospheric circulation, and determining the circulation patterns controlling the transport of BBA from the tropics to the extratropics. To this aim, an objective weather regime (WR) classification…"

When the authors compute the correlation between SST and WR frequency (Figure 9), they could do this on a much longer period, as WR can be determined from longer reanalyses. This would provide a more robust assessment of interannual relations.

Thank you for this comment. After building the WR classification in ERA5, which results to be consistent with the analysis of CAMS data, in the revised version of the manuscript we use this longer time series for analysing the interannual variability across 40 years (Fig. 12). New results show no large differences among the WRs, with no significant trends and comparable variability (STD between 3% ad 4%). We then use the interannual frequencies to investigate possible teleconnections. As expected, the correlation between SST anomalies in the tropical Pacific and SA+/SA- is limited when a longer time span is considered. However, the analysis of the associated atmospheric pattern shows a possible teleconnection (see Fig. 13). We highlight that the WR classification show that synoptic variability is dominated by transient disturbances, and the role of interannual teleconnections is limited.

Could the authors compare their results with computations of particle trajectories, for well-chosen events?

We do agree that computing particle trajectories for selected events would provide an interesting comparison with our results. However, the scope of the paper is to provide a comprehensive picture of the synoptic variability of circulation and aerosols, rather than investigating single events. Therefore, we applied a lead-lag correlation analysis to the CAMS AOD daily anomalies to highlight the development of the BBA transport events in the South Atlantic and southern Africa in the context of a wave pattern dominating the synoptic variability in the region (see Fig. 9 in the revised version). Results show that the AOD anomalies are modulated at the same pace of the WR lifetimes (3 days for the SA regimes, 4-5 days for the ML regimes), confirming the WR control on the BBA transport.

The paper does not present any discussion of comparisons with already existing results. I am not an expert on the subject, but I would have expected that the results reported by the authors could be placed in a context of existing literature.

Thank you for this comment, we agree that our results need to be placed in the context of existing research. As pointed out in the introduction, in our knowledge, no long term characterisation of the BBA transport in the South Atlantic/southern Africa has been presented in the literature to date. For this reason, comparison is only possible with papers analysing short time periods or case-studies. In the revised version of the manuscript, we discuss our findings in comparison with results from the SAFARI-92 and SAFARI 2000 campaigns. A new paragraph has been added in the conclusions:

"The analysis of the regional circulation patterns controlling the BBA transport the South Atlantic/southern Africa sector is reported in literature mainly as a complement in the discussion of field campaign results. During the SAFARI-92 field experiment, Lindesay et al. (1996) reported pronounced BBA transport across southern Africa towards the Indian Ocean, in association with El Niño conditions and intensified continental high. Conversely, during the SAFARI 2000 campaign (Swap et al., 2003), Stein et al. (2003) found an association between the occurrence of rivers of smoke heading towards the Indian Ocean and increased westerly waves and weaker continental high, concomitant with La Niña conditions (see also Garstang et al., 1996). These contrasting conclusions likely originate from to the limited robustness of the analysis due to the shortness of the observation periods. Based on a longer dataset, the WR characterisation suggest a key role of the westerly waves in controlling the rivers of smoke, supporting the hypothesis of Garstang et al. (1996), although it remains inconclusive concerning the role of ENSO phases".

Table 1. AERONET station used in this study: locations and data availability (Version 3 Direct Sun algorithm, level2).

Station	Country	Latitude	Longitude	Observations (coverage)
Ascension Island (AI)	UK Overseas Territory	8.0°S	14.4°W	612 (2003-2017)
Bonanza (BO)	Namibia	21.8°S	19.6°E	126 (2016-2017)
Namibian Stations (NS)	Namibia			276 (2013-2017)
Gobabeb (GO)	Namibia	23.6°S	15.0°E	219 (2015-2017)
Henties Bay (HB)	Namibia	22.1°S	14.3°E	139 (2013-2017)
HESS (HE)	Namibia	23.3°S	16.5°E	158 (2016-2017)
Simon's Town IMT (ST)	South Africa	34.2°S	18.4°E	127 (2016-2017)
Upington (UP)	South Africa	28.4°S	21.2°E	111 (2015-2016)

Table 2. WR transition rates in the CAMS classification, computed as the percentage of transitions from a WR (rows) towards the others (columns). By definition, the diagonal represents persistence. Transition rates above 1/6, i.e. the threshold for non-random transitions, are reported in bold.

WR	SA+	SA-	ML1	ML2	ML3	ML4
SA+	**61**	6	6	11	6	8
SA-	10	**59**	5	10	8	7
ML1	8	14	**39**	**28**	8	3
ML2	12	5	3	**46**	**31**	2
ML3	12	9	4	6	**45**	**24**
ML4	12	13	**29**	3	3	**40**

Table 3. WR transition rates in the ERA5 classification, computed as the percentage of transitions from a WR (rows) towards the others (columns). By definition, the diagonal represents persistence. Transition rates above 1/6, i.e. the threshold for non-random transitions, are reported in bold.

WR	SA+	SA-	ML1	ML2	ML3	ML4
SA+	**58**	8	4	13	6	10
SA-	7	**60**	9	10	8	6
ML1	7	18	**42**	**21**	8	4
ML2	11	5	3	**50**	**27**	3
ML3	4	12	6	4	**45**	**28**
ML4	14	7	**23**	5	5	**46**

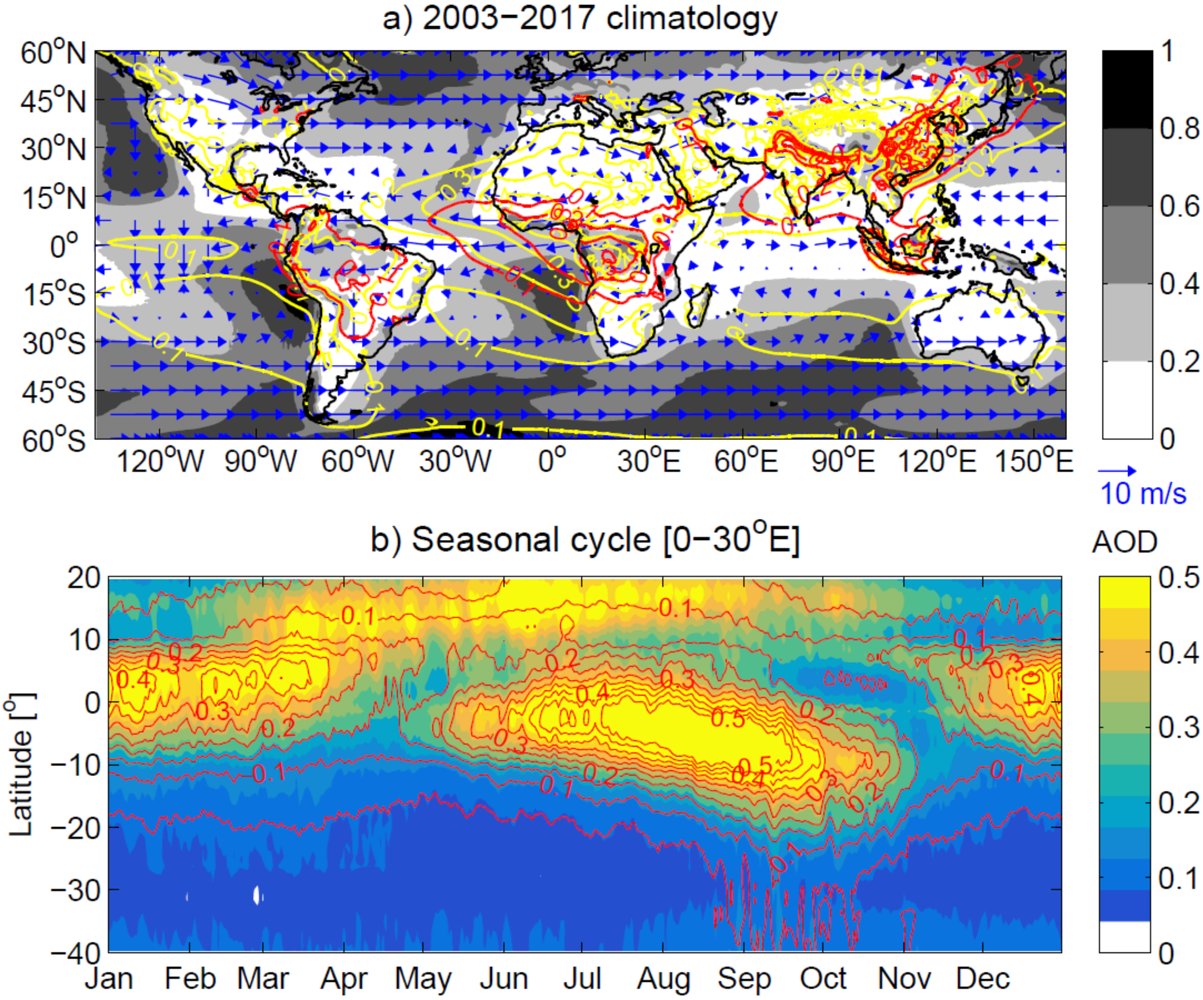

Figure 1. 2003-2017 climatology derived from CAMS reanalysis: (a) annual mean of total (yellow contours) and organic matter (red contours) aerosol optical depth (AOD) at 550 nm, low cloud cover fraction (shadings) and wind at 700 hPa (arrows); (b) annual cycle of total (shadings) and organic matter (red contours) AOD at 550 nm, averaged over Africa and South Atlantic [0-30°E].

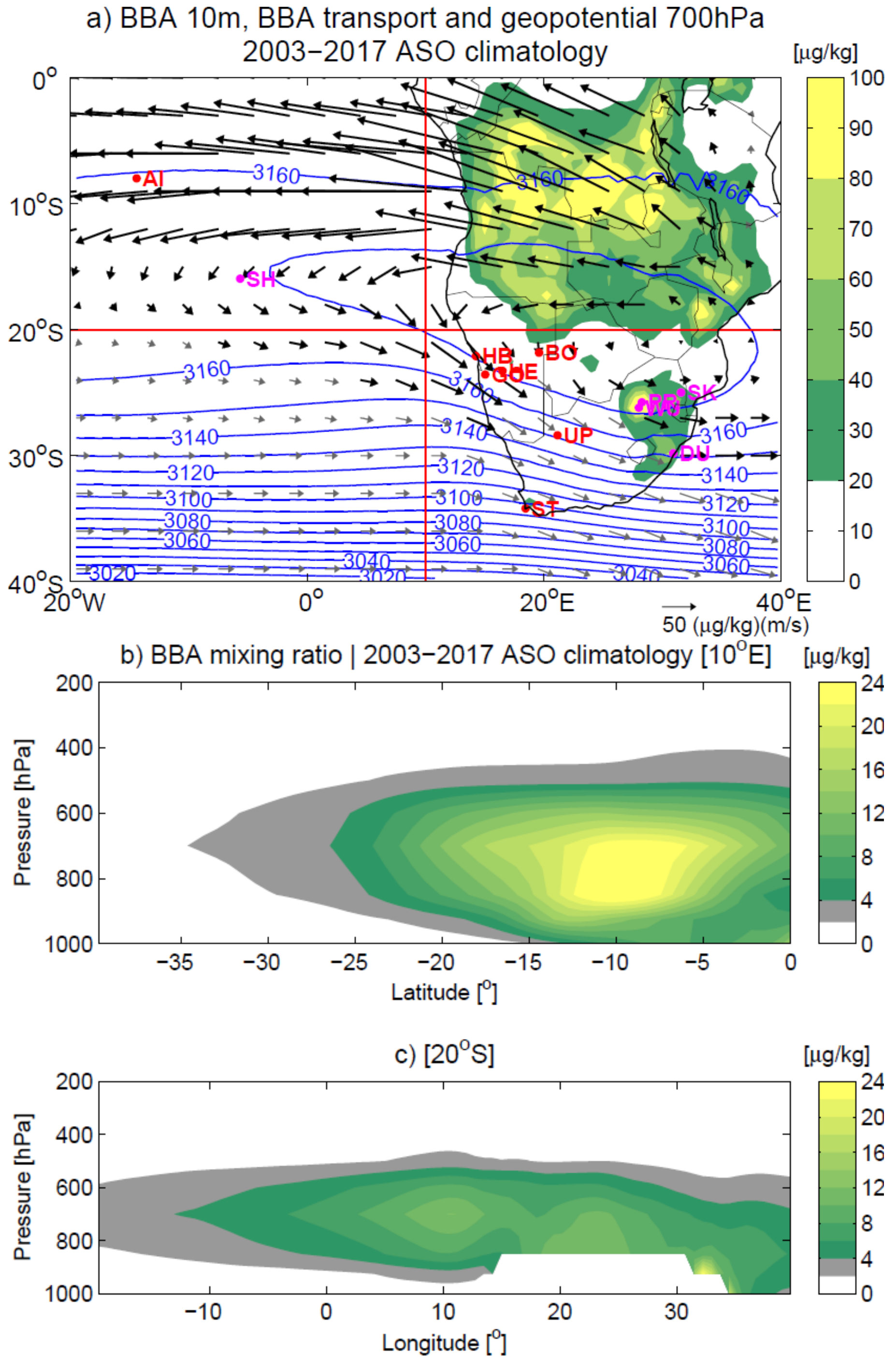

Figure 2. 2003-2017 ASO climatology derived from CAMS reanalysis: (a) organic matter mixing ratio at 10m (µg/kg, shadings), organic matter transport at 700 hPa (µg/kg m/s, arrows) and geopotential height at 700 hPa (m, contours); vertical cross-sections of the organic matter mixing ratio (µg/kg) at (b) 0°E and (c) 25°S. In (a), thick arrows highlight organic matter transport corresponding to organic matter mixing ratio greater than 4 µg/kg; red lines indicate where organic matter mixing ratio cross-sections are computed; red dots indicate the locations of the AERONET stations used in this study (see Table 1 for details); magenta dots indicate the locations of available stations not used in this study (see Section 2.2 for details).

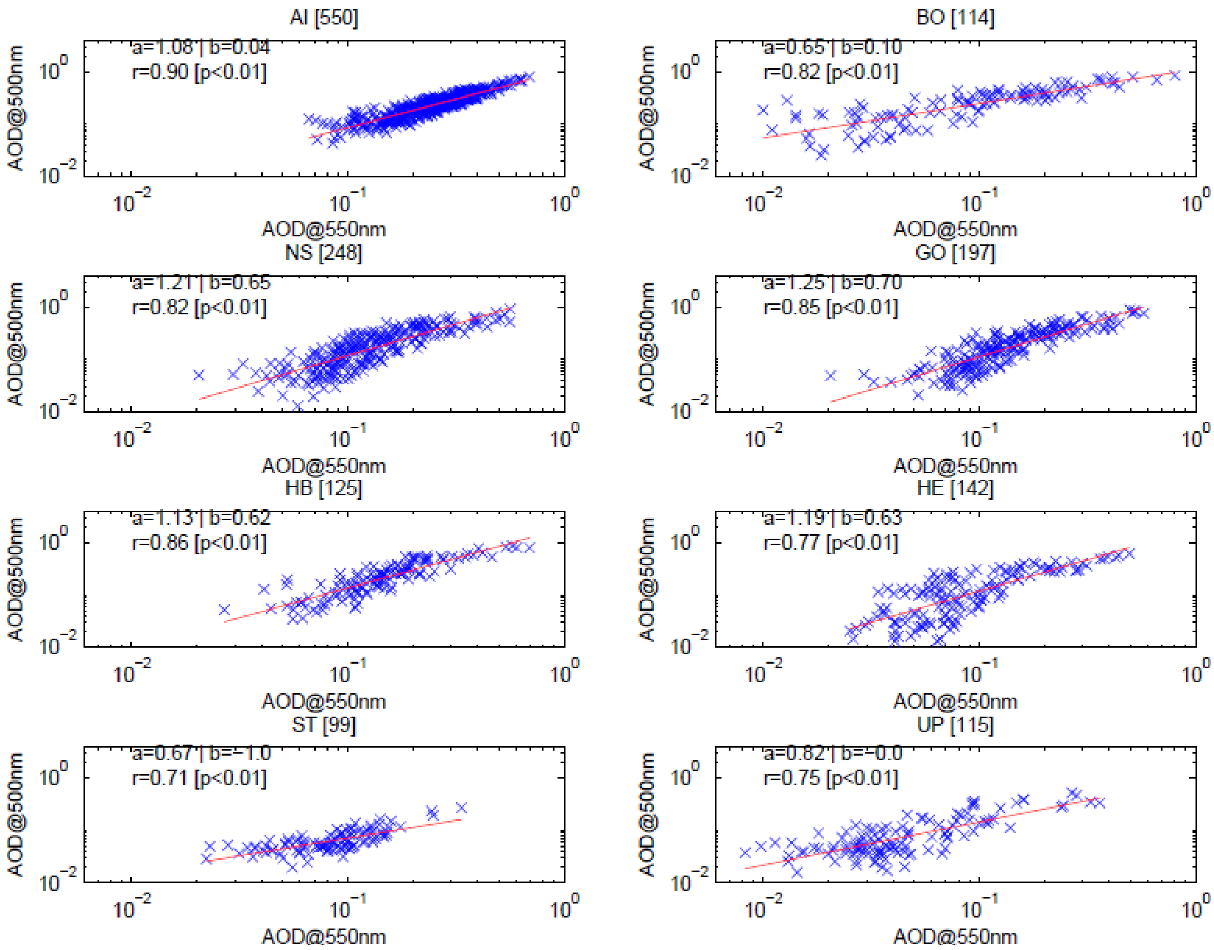

Figure 3. Daily data comparison for ASO 2003-2017: CAMS reanalysis AOD at 550 nm vs AERONET observed AOD at 500 nm. CAMS data are extracted at grid points the closest to the station coordinates (see Table 1). Red lines display the linear regression between CAMS and AERONET data, and the coefficients of the regression models are also reported in the plots, along with the correlation coefficient and the p-value. In titles, the size of the sample used in the linear regression model is reported in brackets (see Section 2.2 for details).

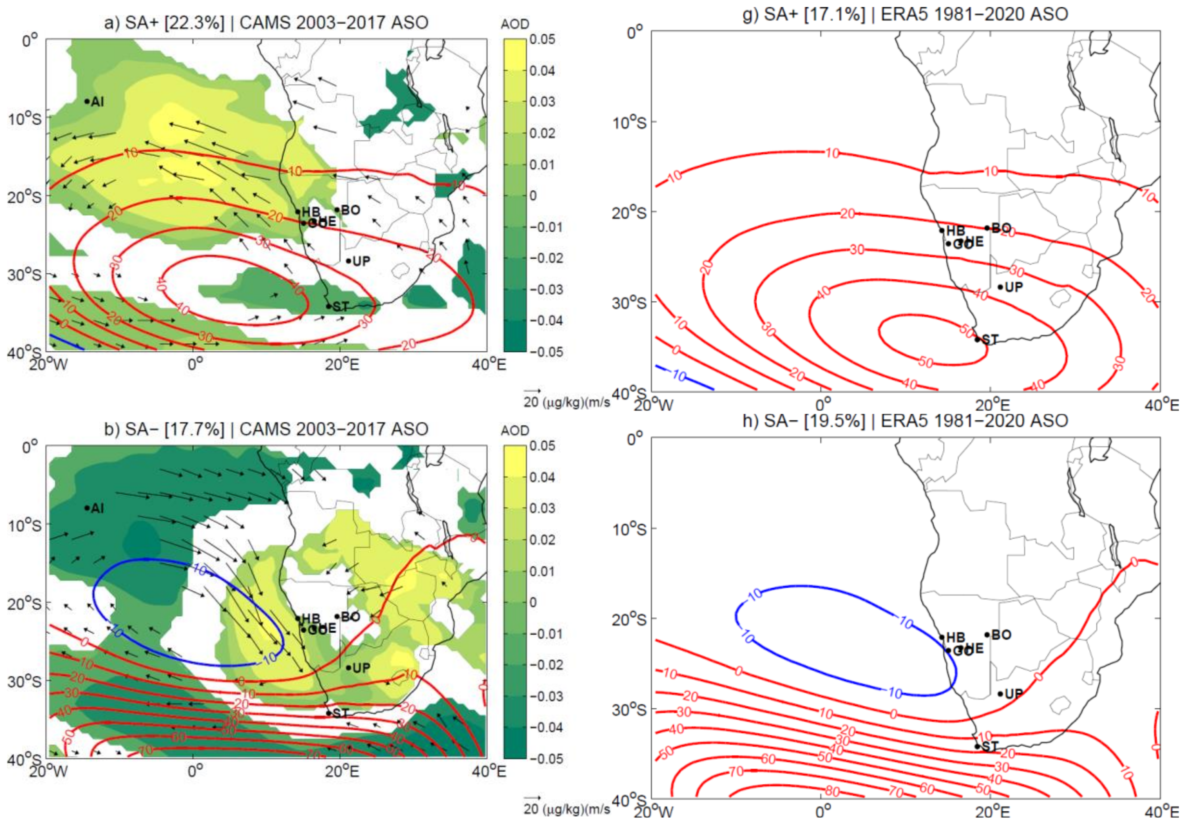

Figure 4. Left panels: anomaly patterns of CAMS geopotential height (m, contours), AOD at 550 nm (shadings) and organic matter transport ((µg/kg)(m/s), arrows) at 700 hPa associated with the WRs classified from CAMS geopotential height at 700 hPa in ASO 2003-2017. Right panels: anomaly patterns of ERA5 geopotential height (m, contours) at 700 hPa associated with the WRs classified from ERA5 geopotential height at 700 hPa in ASO 1981-2020. Dots indicate the locations of the AERONET stations used in this study. Frequency of the WRs is indicated in brackets. For AOD and organic matter transport, only values significant at 95% level of confidence after a Student's t-test are displayed.

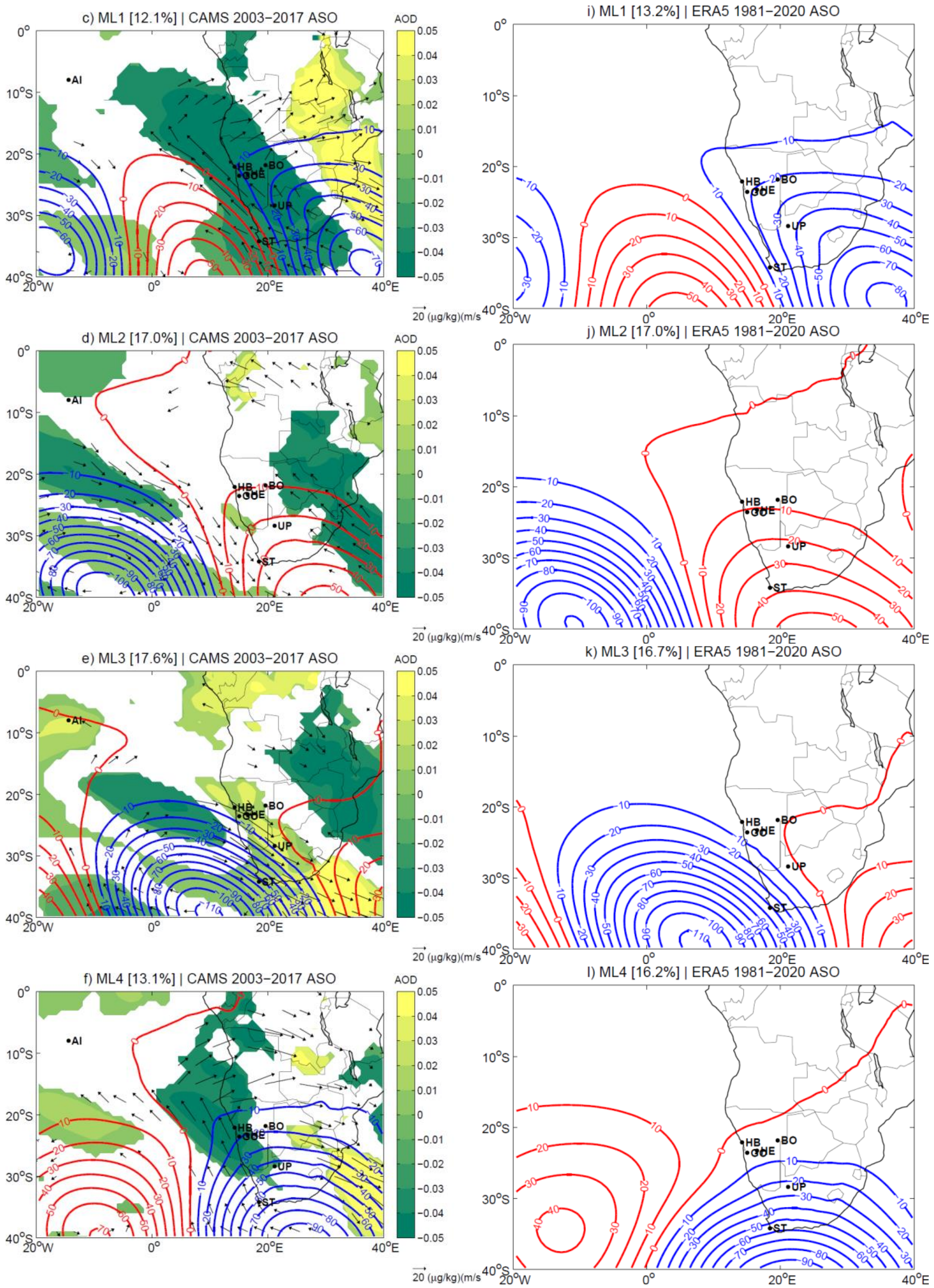

Figure 4. Continued.

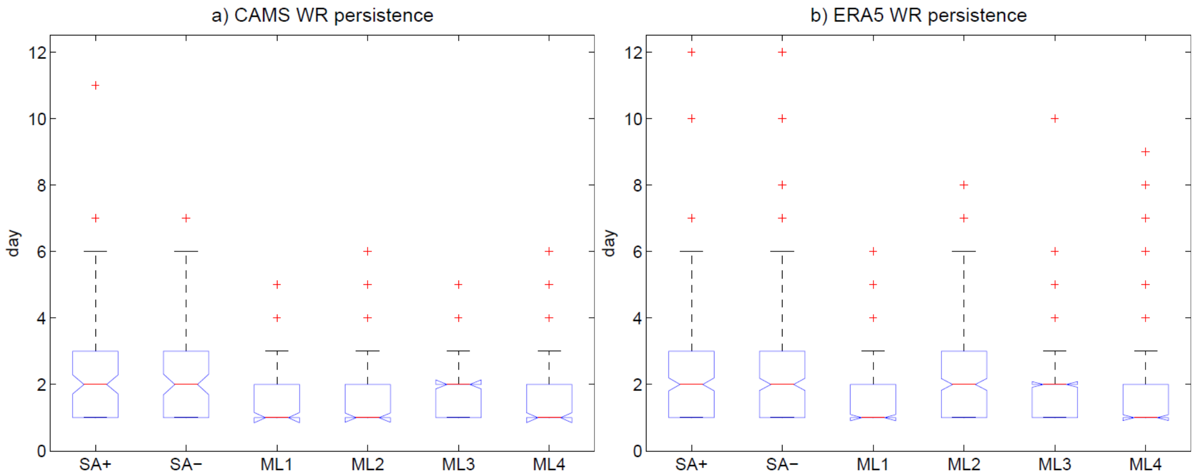

Figure 5. WR persistence in (a) CAMS and (b) ERA5 classifications, displayed as the distribution of the WR sequences. Red lines represent the median, boxes represent the interquartile range, whiskers extend up to the 1.5 of the interquartile range, outliers are displayed as red crosses.

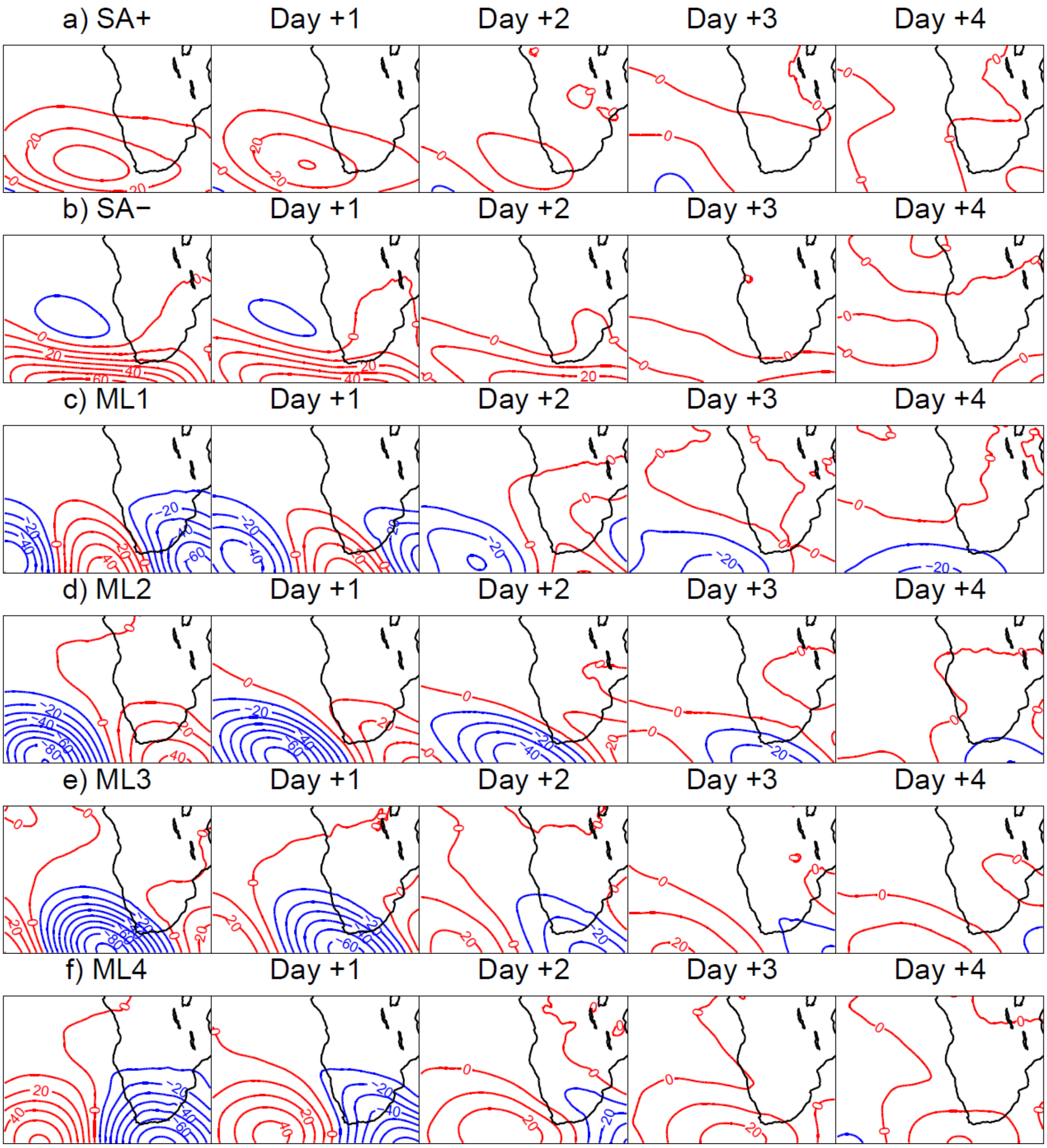

Figure 6. Daily evolution of the 700 hPa geopotential height anomalies (m, contours) associated with the CAMS WR classification, computed as composites from the WR occurrence (day 0) to day +4.

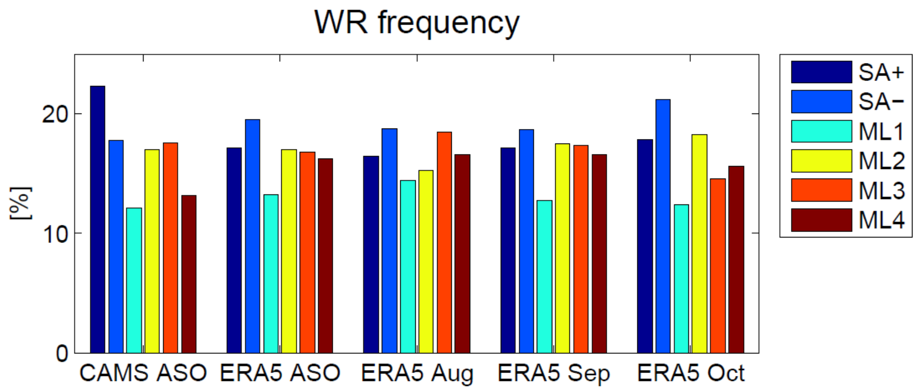

Figure 7. WR frequency in CAMS (computed in the period 2003-2017) and ERA5 (computed in the period 1981-2020) classifications.

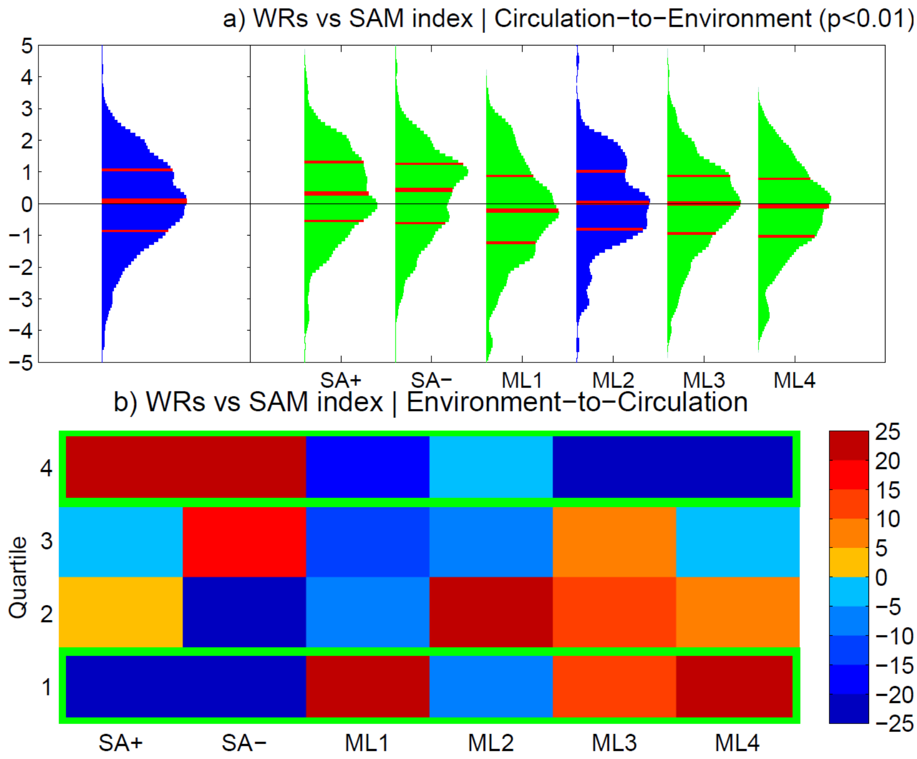

Figure 8. (a) Circulation to environment characterisation (C2E): (left panel) daily Southern Annular Model (SAM) index distribution, and (right panel) for each ERA5 WR. Probability density functions are estimated by using a normal kernel density; red lines represent 25th, 50th and 75th percentiles. Anomaly distributions significantly different from the climatological sample are displayed in green (significance is assessed by a Kolmogorov-Smirnov test at 95% level of confidence). The p-value of the ANOVA used to assess the C2E characterisation is reported in brackets. (b) Environment to circulation characterisation (E2C): ERA5 WR frequency anomaly for each quartile of the daily SAM index. Values represent percentage changes relative to climatological frequencies. Green boxes highlight significant changes in the WR frequencies, i.e. frequency anomalies exceeding the critical threshold (11.07) for the chi-squared statistics with 5 degrees of freedom at 95% level of confidence (see Section 2.4 for details).

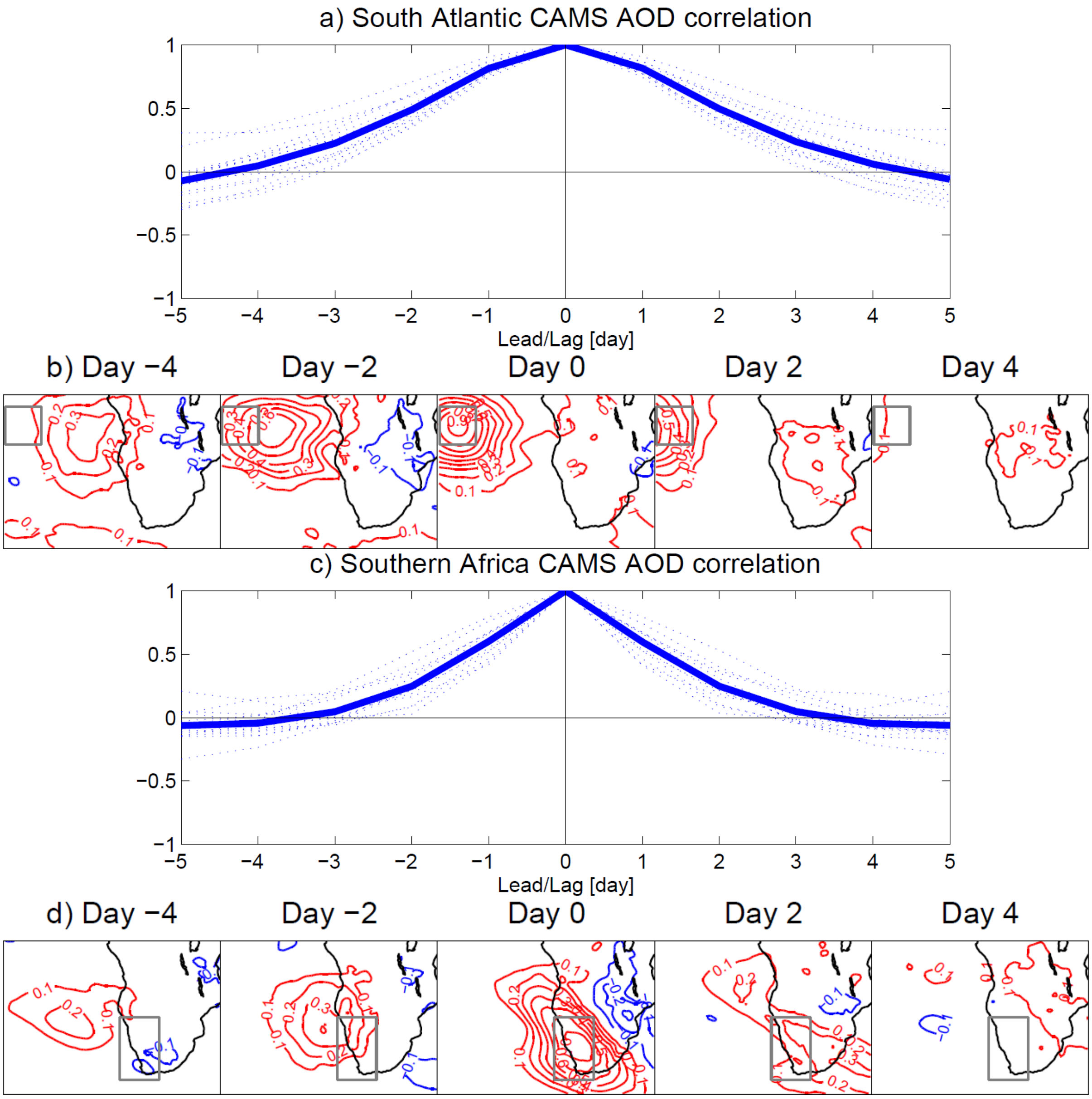

Figure 9. Daily lead-lag correlation analysis of CAMS AOD anomalies. AOD anomalies averaged in the (a) South Atlantic and (c) southern Africa are correlated with themselves: dotted lines display individual year correlations; South Atlantic and southern Africa domains are displayed as boxes in (b) and (d), respectively. Correlation maps: AOD anomalies averaged in the (b) South Atlantic and (d) southern Africa are correlated with the AOD anomalies in the South Atlantic/southern Africa domain. Correlations are computed year-by-year and averaged.

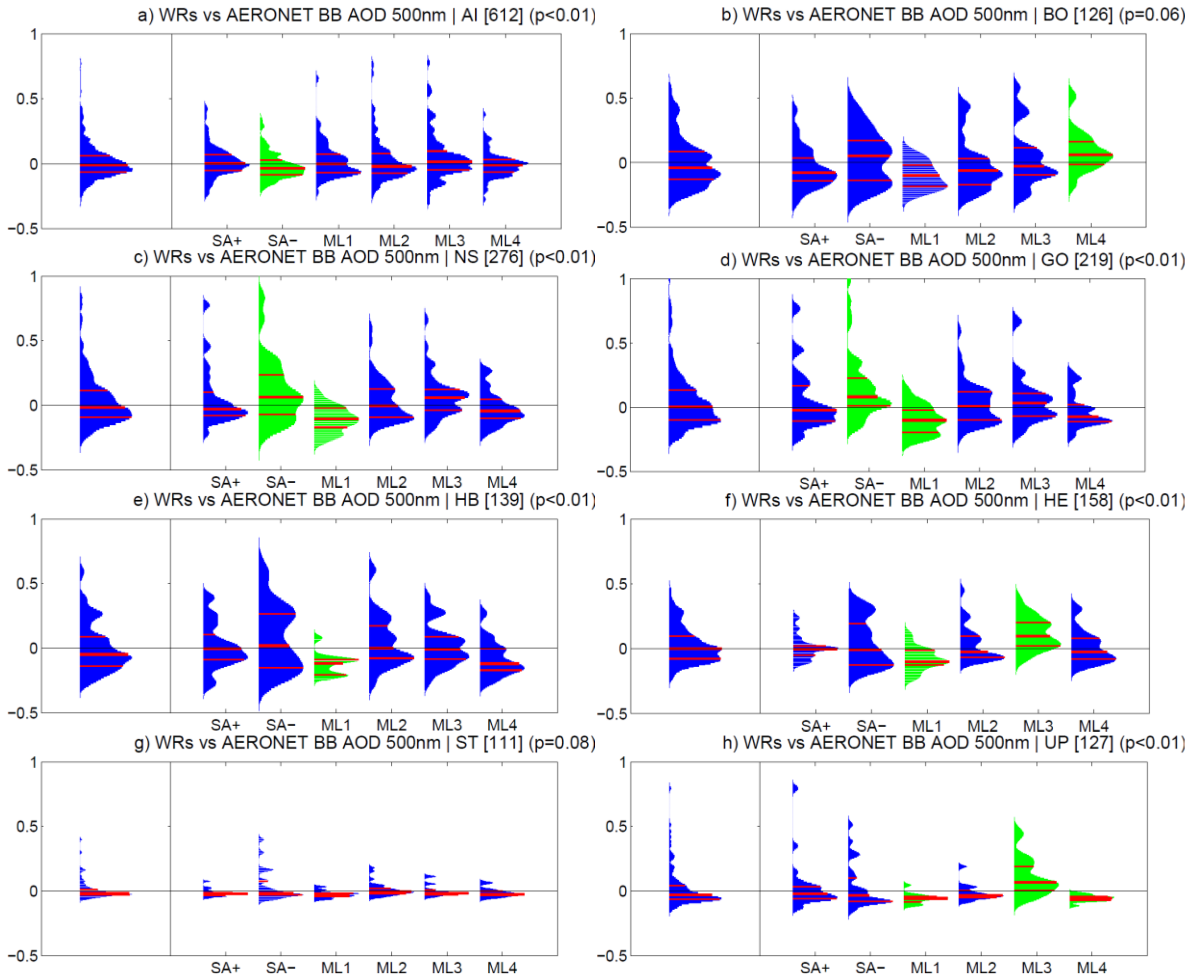

Figure 10. Circulation to environment characterisation: for each AERONET station, (left panel) distribution of the AOD anomalies at 500 nm, and (right panel) for each CAMS WR. Probability density functions are estimated by using a normal kernel density; red lines represent 25th, 50th and 75th percentiles. Anomaly distributions significantly different from the climatological sample are displayed in green (p-values of the Kolmogorov-Smirnov test used to assess the significance of the differences are reported in Table S5). In titles, the number of available daily observations and the p-value of the ANOVA used to assess the WR characterisation are reported in brackets.

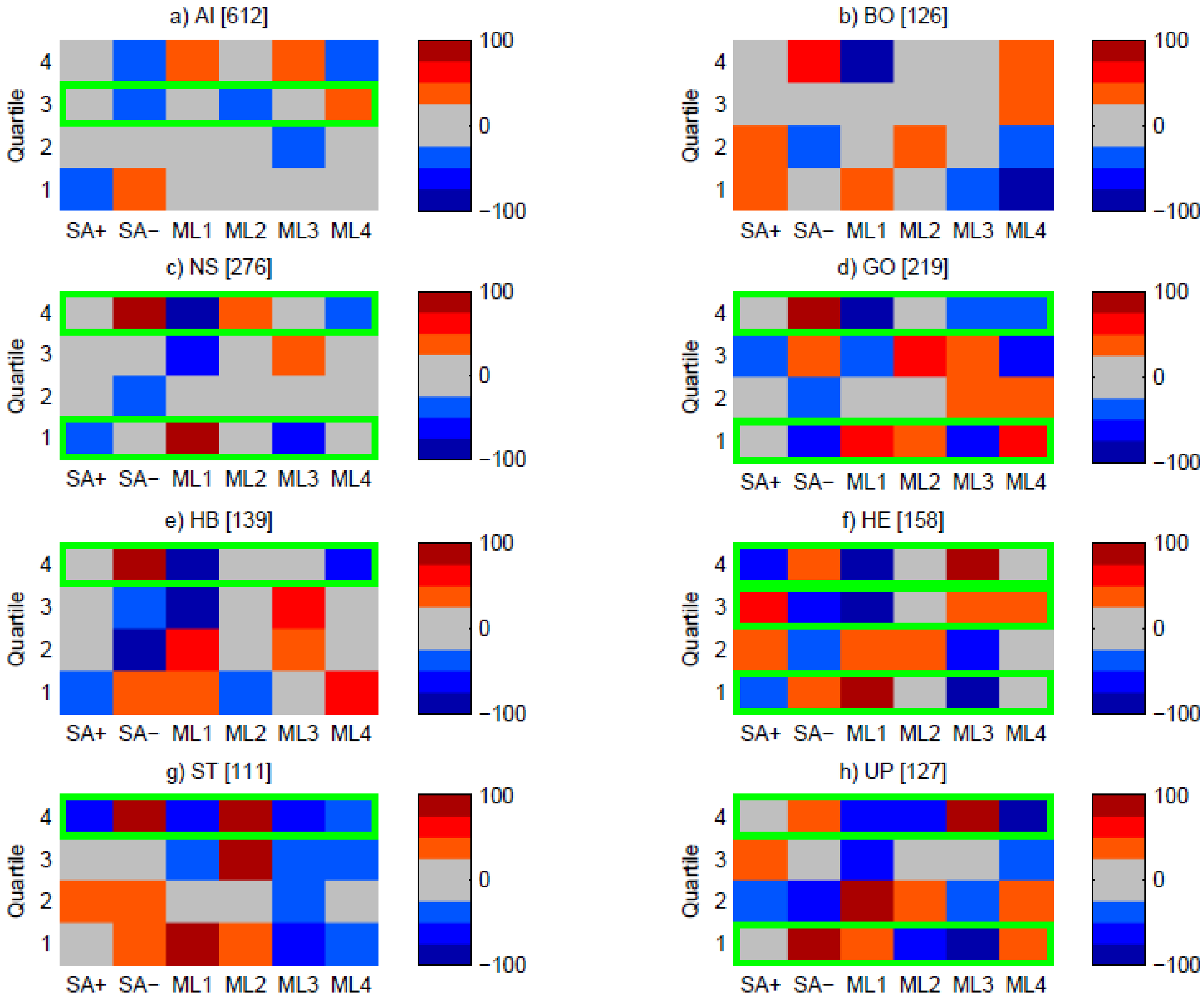

Figure 11. Environment to circulation characterisation: CAMS WR frequency anomaly for each quartile of the AOD anomalies at 500 nm at the AERONET stations. Values represent percentage changes relative to climatological frequencies. Green boxes highlight significant changes in the WR frequencies, i.e. frequency anomalies exceeding the critical threshold (11.07) for the chi-squared statistics with 5 degrees of freedom at 95% level of confidence (chi-squared statistics are reported in Table S6). In brackets, the number of available daily observations are indicated.

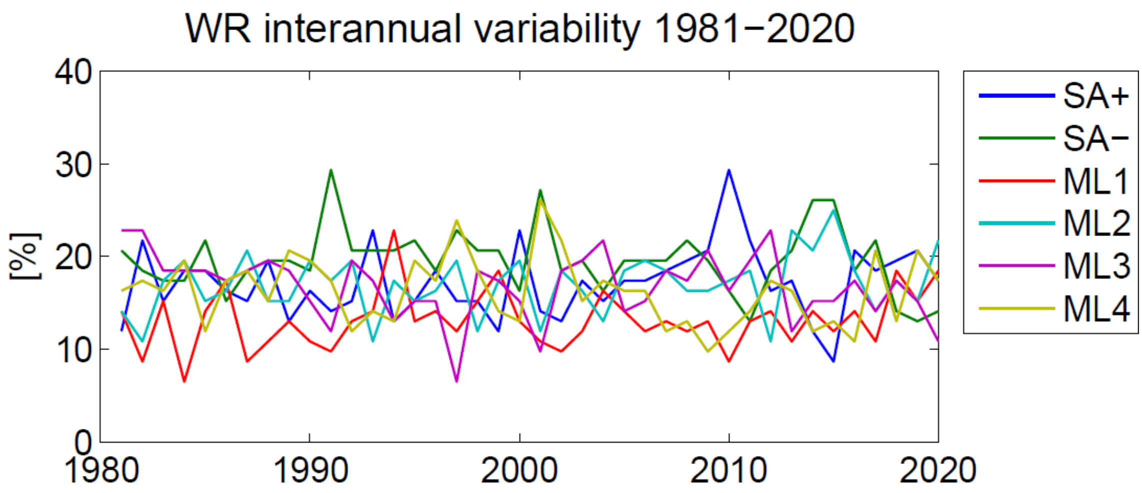

Figure 12. ERA5 WR frequency 1981-2020: interannual variability.

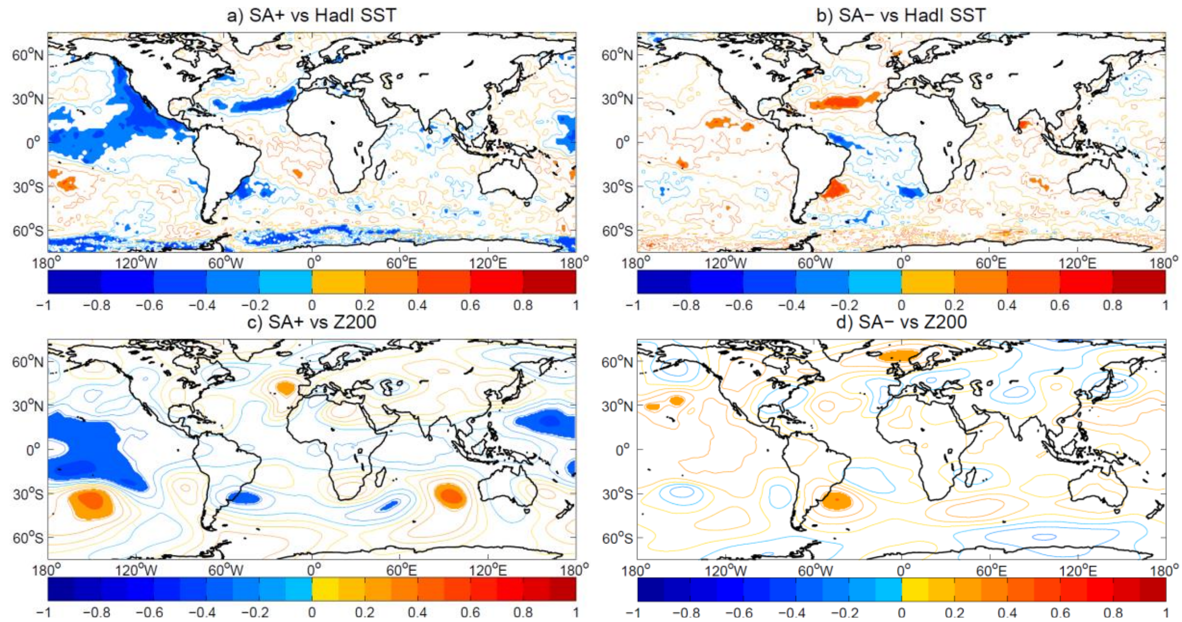

Figure 13. Interannual correlation over the period 1981-2020: SA+ and SA- frequency from the ERA5 classification vs (a, b) HadI sea surface temperature and (c, d) ERA5 geopotential height at 200 hPa in ASO. Shadings display significant correlations at 95% level of confidence. Time series are detrended and standardised.