I thank the reviewer for taking the time to review this manuscript. I will try to clarify some of the confusions revealed by this review. Please see my responses below.

1. The paper is potentially publishable, but their are many issues in both the approach and the presentation that need to be improved before it can be published.1. The author is not careful about describing the quality of the data analyzed in the paper in the methods section. All 'observational' and model data should include some information about the quality from the literature.
A. All of these records are very qualitative, of course. Are there any studies which show that visibility can be used as a proxy? please cite and examine how much you can conclude from qualitative data, or don't use the station data.

Response: The manuscript has lengthy descriptions (Section 2) of the data source used, including present weather and visibility from weather stations, as well as three external dust optical depth datasets from MODIS (by Voss and Amato, 2020) and CAMS and MERRA2 reanalyses. Evaluation of the station data quality is fully described in Section 3. It has been mentioned in a number of places in the manuscript that weather stations have long been used for dust monitoring (e.g., two previous global studies by Shao et al. 2013 and Mahowald et al., 2007), and that the visibility data, when constrained by concurrent dust event reports, have been previously used to estimate ambient dust concentrations around the world (see Section 1 last paragraph and Section 2.1, 2.2).

B. MERRA model output is presented as if it were observations. "CAMS and MERRA2 represent recent advances in developing atmospheric composition reanalysis using global model systems with capabilities to assimilate satellite observations of atmospheric aerosols and gaseous species (Gelaro et al., 2017; Inness et al., 2019)." The way you describe this, it is as if you think these are assimilations of aerosols, but really very little aerosol data is assimilated and they tend not to do very well against the observations, so please discuss the discrepancies and use these model results NOT as observations but as model results. For example: https://acp.copernicus.org/articles/20/10047/2020/.

Response: Section 2.3 clearly describes that CAMS and MERRA2 are aerosol reanalysis generated from global models, which have the capability of assimilating satellite observations from AVHRR, MODIS, MISR, AERONET, and AATSR, with or without bias corrections. The sheer volume of these data is anything but "very little". Using observation-constrained model simulations, rather than free runs, is expected to capture some of the observed dust variability and fill the data gaps caused by limited satellite overpass and clouds, etc. The discrepancy between the surface stations, satellite (MODIS), and model reanalysis is discussed in Section 4.2, Table 3 & 4, and Fig. 7 & 8.

2. I also do not think there is value in presenting all the codes in the first few graphs. Do they really have different correlations or trends? If you only showed 2 of these codes, would we learn less? Please think carefully what your take-away messages are and if we would actually learn more if you presented less.

Response: The qualitative dust weather reports (ww=06-09, 30-35, 98) are used to (1) derive the dust event frequency (FR, in percentage), and (2) combine with visibility to derive a more quantitative measure of the dust burden than FR. Thus it is necessary to carefully examine the data quality (especially the temporal consistency) of ww reports in the following aspects: (1) is the present weather observed in uniform schedule or frequency over time (addressed by Fig. 1); (2) are the dust codes being used consistently over time according to the WMO definition? (addressed by Fig. 2); and (3) do the dust codes represent the same level of dust burden consistently over time (addressed by Fig. 3 and Table 2).

3. Pretty much all the tables and figures have NO description of what is in the table and figures in the figure captions. Please describe what is actually plotted clearly. I probably misunderstood most of the plots and would need to re-review once the plots are explained.

Response: see my responses to your specific comments below.

4. The author wants to indicate that drought caused some of the dust, but doesn't show statistical studies of this relationship. Please add in any mechanisms that you want to evaluate and make sure your results are statistically significant.

Response: Section 4.2 (last 2 paragraphs) includes the analysis on the dust-drought relationship, including the correlation and significance test results between dustiness and drought index, as well as a brief discussion on the linkage of drought, vegetation, and dust emission.

"Figure 1. Analysis of the continuity of present weather (ww) reports in ISD. (a) Global monthly number of stations (Nstn), number of ww reports (Nww), and number of dustevent reports (Ndu). For clarity Ndu is multiplied by 30; (b) Nww from different report types."

"Therefore, the SYNOP data between 1986 and 2019 are most suitable for global-scale dust analysis." I assume this means you will only present the SYNOP data in this paper. This figure can go into the supplement, and say this in the methods instead. But wait, you are using the other times periods, after you say that you shouldn't use them? Seems odd??? Please justify their use, and then discuss how that changes your results whenever you use the data that isn't very consistent.

Response: See my response to #2 above. SYNOP data between 1986 and 2019 are considered most suitable for several reasons: (1) the number of ww reports are continuous (see Fig. 1); (2) the usage of most dust codes (except 09) is continuous (with no artefact from the alteration of code definition), especially the most frequent ww=06, 07 and the codes of dust storm events (ww=30-35) (see Fig. 2); and (3) the quantitative meaning of dust codes are nearly consistent over time (see Fig. 3 and Table 2). Generally speaking, to use any long-term measurements for variability and trend analysis of a physical phenomenon, the data quality must be carefully examined, especially in terms of temporal consistency and continuity.

Table 2 is really unreadable because we can't understand what these codes are. I think in Table 1 you should come up with some short acronym that describes each of these codes, and use it for

the rest of the paper, so that we know which code is which and what they mean. Or just exclude this whole section, as not really very important or interesting compared to the others.

Response: See my responses to #2 and #4 above.

"Figure 2. Analysis of the continuity of dust weather code usage. From top to bottom: monthly number of reports of all and individual dust codes in the global SYNOP data. Horizontal lines are all-time averages. Dust weather codes are described in Table 1". This figure caption does not tell us what is plotted. Please describe the variable plotted. Is this the number of observations? Why do you present so many? Are they actually different? The same? If they are the same, then show one. If they are different, tell us how they are different and why.

Response: See my responses to #2 and #4 above.

"Figure 3. Temporal consistency of the harmonic mean visibility associated with dust weather codes. Dash lines are all-time averages. Dust weather codes are described in Table 1." What variable is presented in figure3? It should be described in the figure caption. Is this the number of observations? Or the correlation? Do you really need to present all the variables, or could you just show one and the offset between them?

"Figure 3 shows that while there are significant year-to-year fluctuations in the harmonic mean visibility associated with dust codes, they generally fall into three clusters: ww =06–09 (3.7 km), ww = 30–32, 98 (1.5 km), and ww = 33–35 (0.7 km)." please tell use why this is important. Does this make sense? Etc. right now there is no context for this statement.

Response: It means that if the dust burden represented by the qualitative reports changes from year to year, and it's important to check if there are significant trends which will distort the use of dust weather reports in tracking long-term dust changes. The dust codes fall into three clusters, indicating that they contain some quantitative information of the dust burden, albeit being qualitative reports.

"Figure 4. Decadal mean dust event frequency (FR). Gray dots are stations with FR< 1%." Please indicate in the figure caption exactly which data you use to define this, since there are many ways to do this. Repeat information in the methods to be clear. Your figure captions are so brief as to make reading your paper much more difficult than it needs to be.

Response: FR is defined in Eq. 1. The Fig.4 caption now reads "Decadal mean dust event frequency (f, %), defined as the percentage of dust events out of the manual present weather reports. Gray dots are stations with f < 1%. The number of stations (including those with f < 1%) is shown."

"Using Eq. 1, the decadal mean FR is calculated for stations with at least 5 years'data in each decade, as shown in Fig. 4." Please don't write sentences like this that are difficult to understand. Tell us again what equation 1 uses to calculate the frequency of occurrence in English. (for example: Using dust frequency as calcualted from the visibilty data (equation 1), we evaluate the decadal values...)

Response: Again, FR is defined in Eq. 1. It is derived from the present weather report, not visibility. The Fig. 4 caption will be revised to clarify that (see the response above).

These two paragraphs are very hand wavey, but on important, and easily plottable points: "The distinctive changes of decadal mean FR can be linked to multidecadal climate variations, especially the occurrence of mega-drought events lasting several years or even decades. Elevated dust activity can be observed in areas affected by persistent drought, where the reduction of soil moisture and vegetation leaves the ex-posed, dry soil prone to wind erosion. For example, a striking feature in the 1950s is the widespread, frequent dust events in the U.S. Southwest and Midwest, with several stations reporting FR > 20% in the High Plains of Texas and Colorado. The heightened dust activity was fueled by a 11-year-long (1946–1956) drought that afflicted a massive area centered in the Southwest U.S. (Fye et al., 2003). The 1950s drought was characterized by a prolonged lack of precipitation and excessive warm temperatures, which caused crop failure and livestock feed shortage (Goudie and Middleton, 1992). As the drought came to an end in the spring of 1957, FR started to decline and has since remained low in the last 50 years. Similarly, North Africa experienced progressively drier conditions during the 1970–80s in the Sahel, a semiarid dryland belt at the southern border of Sahara Desert (Giannini et al., 2008). The Sahelian drought was triggered by anomalous sea surface temperature (SST) in the tropic Atlantic and Indian Ocean (Dai, 2011). The Sahelian dust frequency during drier-than-normal years, especially in the 1980s when drought was most severe, is significantly higher compared to the pre- or post-drought periods. The drought-induced dust enhancement is also evident from the frequent dust weather observed downstream, including the Caribbean, Gulf of Mexico, and Iberian Peninsula. This is consistent with the long-term in situ dust measurements in Barbados and Miami, Florida, indicating a positive cor-relation between the Sahel dry anomaly and African dust outflow across the tropical North Atlantic (Prospero and Lamb, 2003; Zuidema et al., 2019). With the amelioration of Sahelian drought in the 2000s, FR experienced significant decreases at the source and downwind, consistent with ground and satellite observations (Hsu et al., 2012; Li etal., 2014). In the past decade, increased dust activity can be observed in West Africa and the Middle East, which will be discussed later." Please show this is true statistically significantly, in a clear way. perhaps show the 1-sigma and 2-sigma bounds and highlight the time periods above or below, or do a correlatoin.

Response: The discussions presented here are simply based on the visual examination of the decadal mean FR and VI from the 1950s to the 2010s. The discussions are connected to and supported by a number of previous studies cited in the paper. Please be specific about the 'hand waving'.

"if weather stations provide a consistent view of global dust variations, FR and VI are compared with the datasets described in Sect. 2.3." You can compare the dust variations to satellite data and AERONET data and evaluate the dust variables, but please do not pretend that MERRA output is more than model output.

Response: No pretension is made here. Section 2 clearly describes CAMS and MERRA2 as global model reanalysis. Although they are model output with uncertainties, the correlation

analysis in Table 3 shows good agreement between surface stations, MODIS, and model reanalysis in terms of the long-term variability.

"Figure 6. Global monthly (gray) and annual (black dotted) FR and VI . Horizontal lines are all-time averages." Please describe what you are plotting in complete detail, including repeating which variables are used, etc, so that your figures are self-standing. How are you averaging over the globe when you only have spotty data? Needs to be described in the methods section, and make sure you are doing this in a manner that is consistent with the observations and models you are comparing again (for example, pick each point from the dust station data, and match to the model output at the same grid box, so are weighting similarly).
Response: The global mean FR and VI are derived by aggregating all global stations together. The station map is given in Fig. S1, S2, and is also shown in Fig. 4 and 5. The stations have fairly good coverage of the major dust source areas. Similarly, the global mean DOD is calculated from MODIS (Voss and Amato, 2020), CAMS, and MERRA2 without taking into account coincidence in space and time, as one normally would do in using these data in analyzing the global mean. Collocation would be needed for a critical comparison of MODIS vs. reanalysis, which is outside the scope of this study.

"Trends are further calculated from the monthly anomalies using the pyMannKendallpackage developed by Hussain and Mahmud (2019), which consists of multiple Mann-Kendall test options to accommodate the seasonality and serial correlation in the data. The Mann-Kendall test is a non-parametric test of the presence of monotonic trend in the data, and has advantage over parametric methods (e.g., t test) for its in-sensitivity to outliers, missing values, and the statistical distribution of the data. TheMann-Kendall test is designed for serially independent data and thus can be influenced by the presence of autocorrelation in the data, which either increases the uncertainty of estimated trends or prolongs the length of time period required to detect a given trend (Weatherhead et al., 1998)." All methods should be in the methods section, or in figure captions, not in the results section.

Response: My preference is to briefly describe the trend detection method here, rather than using a separate subsection.

"The decadal mean FR and VI (Fig. 4 and 5) indicate mega-drought events are associated with extremely active dust periods in the 20th century." To make this statement you need to have compared against precipitation or P-E data and show a statistically significant change/relationship

Response: This statement repeats the finding from the discussions in Section 4.1, which is also supported by a number of previous studies cited in the manuscript, e.g., Goudie and Middleton, 1992; Prospero and Lamb, 2003.

"Dust variation in North Africa from 1986 to 2019. From top to bottom: monthly Nwwand Ndu, and anomalies of FR, VI, MERRA2 DOD, wind speed, soil moisture, and theJones North Atlantic Oscillation (NAO) index. DOD and soil moisture are averaged overregion 10◦N–20◦N, 20◦W–20◦E; Dotted curves are annual averages. Dashed lines areTheil-Sen linear regression of

FR, VI and DOD, with the annual trends and confidence intervals (CI) shown." What is Nww? Ndu? Should be explained in the figure caption, as well as FR, VI.

Response: The figure caption now reads "Dust variation in North Africa from 1986 to 2019. From top to bottom: number of present weather reports (Nww), number of dust event reports (Ndu, multiplied by 3 for clarity), and the anomalies of dust event frequency (f), dust extinction coefficient (β), MERRA2 dust optical depth (DOD), wind speed, soil moisture, and the Jones North Atlantic Oscillation (NAO) index. DOD and soil moisture are averaged over region 10°N–20°N, 20°W–20°E. Gray curves are monthly data, while dotted black curves are annual averages. Dashed lines are Theil-Sen linear regression of f, β and DOD, with the annual trends and confidence intervals (CI) shown.".

"Consistent with previous studies, Fig. 11 shows that African dust was at a historic high during the late 1980s due to abnormally strong winds and low soil moisture, both favorable for dust production (e.g., Prospero and Lamb, 2003; Chiapello et al., 2005; Evan et al., 2016)." Is this statistically significantly true? Please check or put your statistical significance on the plot.

Response: As shown in Fig. 11, the monthly FR and VI anomalies are well above the zero mean and at a maximum in the 1980s and early 1990s, in agreement with the past studies cited in the manuscript.

"Indeed, significant correlations are found between NINO 3.4 and scPDSI ($r = 0.27, p < 0.001$), as well as between PDO and scPDSI ($r = 0.53$, $p < 0.001$). PDO also appears to be more correlated ($r = -0.36$, $p < 0.001$) with MERRA2 DOD than NINO3.4 ($r = -0.15$, $p < 0.005$). As seen in Fig. 13, PDO has transitioned into positive and weak negative phases since 2015, which has led to the amelioration of drought and consequently, a decrease of dust activity in the Middle East." Is this based on the time series in the plots? Please specify.

Response: It is based on the data shown in Fig. 13. NINO 3.4 covaries with PDO and is not shown in Fig. 13, though.

"The dust decline was driven by the weakening of surface winds, as shown in Fig. 15 and suggested by past studies, which reported widespread decrease of surface winds in China during the 1970s– 1990s, with greatest reduction during spring and in the gusty wind segment (Guo et al., 2011; Lin et al., 2013)." Again, do not show your results without showing that they are statistically significant. Please show a correlation coefficient and that it is significant before you make such important, but unclear statements.

Response: These statements are backed by both previous studies and statistical analysis (with significance test results) presented in section 5.3 (see line 448-476).