Thank you very much for the constructive comments. Please see my replies below.

<u>Review report acp-2020-813</u>

The current study deals with the analysis of dust long-term variability over the major deserts of the planet and the main downwind regions. To realize, observations from weather stations, derived by the NOAA Integrated Surface Database (ISD), have been analyzed. Overall, it is a very nice and comprehensive study which fits very well to the issues covered by the Atmospheric Chemistry and Physics journal. Moreover, the structure of the work is well defined and the manuscript it is very well written. Therefore, I recommend the submitted study to be published after taking into account the following minor comments and suggestions.

1. **Lines 50-55:** Here you have to add satellite sensors (SEVIRI, CATS, IASI, CALIPSO) which have been widely used in various studies focusing on dust aerosols.

   Response: this section is not intended to give a comprehensive review of satellite aerosol sensors, but briefly mentions a few with long-term observations of dust burden needed for the analysis of interannual to multidecadal dust variations (the objective of this study). Passive sensors with a wide swath and sufficiently long records are considered more suitable for such purpose. Among those, only a few provide aerosol information over deserts.

   It now reads "For contemporary dust, satellite remote sensing has greatly advanced the monitoring capability of large-scale dust events on the daily basis. Specifically, the long-term aerosol climatology derived from low- and moderate-resolution passive sensors in sun-synchronous orbits remain the most widely used datasets to track long-term dust trends and variations at global and regional scales. To name a few, the Advanced Very High Resolution Radiometer (AVHRR) sensors onboard NOAA's weather satellite series provide the longest over-ocean aerosol optical depth (AOD) dataset since the 1970s, available under the NOAA Climate Data Record program (Zhao et al., 2008). As a semi-quantitative measure of column aerosol burden, the absorbing aerosol index (AAI) derived from the UV bands of Total Ozone Mapping Spectrometer (TOMS) and similar instruments allows dust detection over desert surfaces, and the derivation of a global dust source map for improving dust model simulations (Ginoux et al., 2001; Prospero et al., 2002). Under the NASA Earth Observing System (EOS) program, newer instruments designed specifically for measuring the atmospheric composition with improved sensor characteristics, notably the twin Moderate Resolution Imaging Spectroradiometer (MODIS) sensors aboard Terra and Aqua, have further expanded the capability of large-scale dust characterization and source mapping from space (Ginoux et al., 2012). Many other instruments, including those from European satellites which are being reprocessed under the ESA Climate Change Initiative, offer a complementary view of the dust aerosol burden and properties since the 1990s, mostly over downwind areas where elevated dust layers can be effectively distinguished from the dark background of vegetation or ocean surfaces (Popp et al., 2019)".

2. **Lines 63-65:** It is missing a discussion about the advantages of passive satellite sensors (e.g. spatial coverage) as well as few sentences about the benefits and the drawbacks of active satellite sensors (e.g. CALIOP).
   Response: see my reply to #1 above. It now reads "While spaceborne sensors offer an unprecedented view of large-scale dust events, they can be limited by the difficulty of separating dust from the total aerosol signal, limited sampling frequency in sun-synchronous orbits (usually once per day), incapability of detecting dust under clouds, and sensor calibration instability. When it comes to the analysis of interannual to multidecadal dust variations governed by low frequency climate variability, satellite records are further limited by the relatively short length

and cross-sensor consistency and continuity, especially over dust source areas where quantitative information of the column aerosol burden is available only for the last two decades."

3. **Line 121:** Could you please explain better this sentence?
Response: In the manual report of present weather (ww), the code varies from 00 (lowest priority) to 99 (highest priority). During the data preprocessing, I follow this rule by keeping the highest value of the ww records in ISD; However, if there is a dust code in the ww values, the dust code is kept, instead. ISD has 7 place holders for ww reports at any given time, but for most of the time, only has 1 ww report, so the above rule does not apply. When multiple ww reports do occur, the above rule makes sure that the dust event is counted and not overwritten by higher-priority events. It is rare, however, that a dust event is reported at the same time with other higher-priority weather events, specifically fog and precipitation.

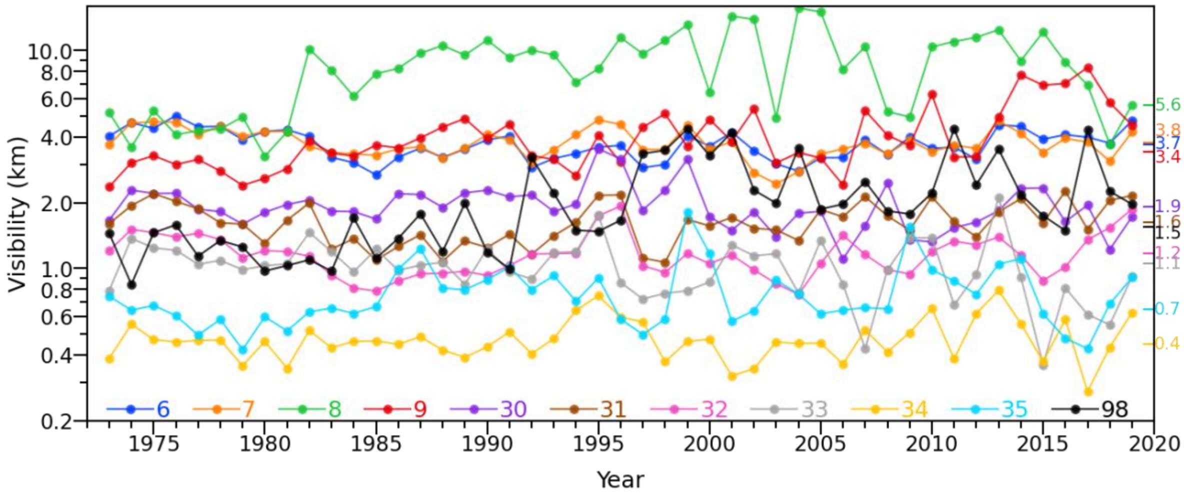
4. **Equation 1:** Please provide a definition for the FR abbreviation.
Response: FR is the dust event frequency, expressed in the percentage of dust event reports out of all ww reports. In the revised manuscript, FR will be abbreviated as "f", while VI will be replaced by the dust extinction coefficient, β, which is more intuitive and easier to interpret.

5. **Lines 167 – 173:** MERRA-2 winds have been also used for the sea-salt aerosols.
Response: It now reads "To separate the dust component over ocean, Voss and Evan (2020) followed Kaufman et al. (2005) by removing the contribution of fine-mode (anthropogenic and biomass burning) and sea salt aerosols (derived from reanalyzed winds) from the total AOD."

6. **Lines 177 – 180:** Few corrections must be implemented here. MODIS radiances, instead of AOD, are used and are transformed to bias-corrected AOD with respect to AERONET. MISR AODs, without bias correction, are assimilated in MERRA-2 only over bright surfaces whereas AERONET measurements are taken into account until 2014.
Response: Thanks for pointing it out. It now reads "MERRA2 is produced by the Goddard Earth Observing System version 5 (GEOS-5) which simulates dust emission and dispersion using the Goddard Chemistry, Aerosol, Radiation, and Transport model (GOCART), and assimilates multiple aerosol products, including bias-corrected AOD derived from the observed radiances by AVHRR (over ocean only, until 2002) and MODIS (over ocean and dark surfaces, from Terra since 2000 and Aqua since 2002), MISR AOD over bright land surfaces (surface albedo > 0.15, without bias correction), as well as ground-based AOD measurements from AERONET until 2014 (see details in Randles et al., 2017)."

7. **Line 187:** What do you mean "*even false recording*"? If it is not reliable, then why it is considered in the analysis?
Response: Removed.

8. **Figure 3:** It is hard to distinguish among the curves and the dashed lines. It is better to produce separate plots (for individual codes or groups) and give the all-time averages in a table or insert them in the plots.
Response: Fig. 3 has been remade. See below.



9. Figure 7: Which MODIS DOD are you using? From Voss and Evan (2020)? Have you tried to reproduce the same plot but considering only the coincident measurements among MODIS, CAMS and MERRA-2 for the common period? Please keep in mind that MODIS provides single measurements per day whereas the reanalyses datasets take into account the diurnal cycle. Can you comment on this? Moreover, it is needed an explanation of how the global means have been calculated. In Figure 5 in Levy et al. (2009), it is evident that the calculation of the domain averages is affected by the selected approach. This is quite critical for the satellite data in which there are gaps due to the inability of the applied algorithm to provide a retrieval.

Response: The MODIS dust optical depth (DOD) in Fig.7 is from Voss and Evan (2020). The figure caption is revised to reflect that. The global DOD mean is calculated without taking into account the data match in time or space. Collocation would be needed for a critical comparison of MODIS vs. reanalysis, which is outside the scope of this study. Rather, the DOD mean is calculated simply as the domain average of the monthly data of Voss and Evan (2020), CAMS, and MERRA2, as one would normally do when using them for global analysis. The manuscript does include discussions on the disparity of MODIS and model reanalyzed AODs: "The global average DOD is calculated from the MODIS DOD (550 nm) by Voss and Evan (2020) and the dust extinction optical depth (550 nm) from the CAMS and MERRA2 reanalyses, without taking into account the collocation in space and time. Figure 7 reveals systematic differences between the DOD datasets. For the overlapping period (2003–2018), the mean DOD is 0.042, 0.013 and 0.024 from MODIS, CAMS and MERRA2, respectively. CAMS also appears to be less variable than MODIS and MERRA2. The offsets in the mean DOD can be attributed to a number of factors. For example, MODIS DOD is based on once-per-day observations over both ocean and land, including deserts where large AOD values (up to 5) are allowed in the Deep Blue algorithm, whereas CAMS and MERRA2 are generated by global models which are capable of simulating the diurnal cycle of dust loading and assimilating satellite aerosol observations, including over deserts. Specifically, CAMS assimilates the MODIS Deep Blue AOD without bias correction (Inness et al., 2019). Unlike CAMS, MERRA2 does not use the Deep Blue products, but rather assimilates the MISR AOD without bias correction over desert surfaces (Randles et al., 2017). In addition, the models used to produce CAMS (ECMWF IFS) and MERRA2 (NASA GEOS-5) have a variety of differences in the model configuration, dust parameterization, dust optical properties, meteorological input, and AOD assimilation, all of which may contribute to the discrepancy between CAMS and MERRA2 DOD."

10. Figure 8: Please consider to reproduce the plot with the collocated data for the common period.
Response: See my response to #9. Fig. 8 is to compare the multi-year average annual dust cycle over the common period of 2003-2018, based on all available information from each

dataset.

11. **Tables 3 and 4:** It would be interesting to include also other dust optical depth databases such as MIDAS (Gkikas et al., 2020) and LIVAS (Amiridis et al., 2015).
    Response: Thanks for the suggestion. I will consider including these datasets in an upcoming manuscript as an extension of this study.

12. **Lines 305 – 306:** As I have already mentioned above, MERRA-2 assimilates MISR AODs above bright surfaces. Likewise, in MERRA-2 the anthropogenic dust sources are not considered.
    Response: Fixed. See my response to #9.

13. **Figure 9:** I would like to see the results at station level. More specifically, three global maps are needed with the stations colored based on the correlation coefficient of VI with scPDSI, soil moisture and wind.
    Response: I agree that station level analysis would provide information on the dust-climate relationship at the local scale. However, I think using global aggregated data is sufficient for studying the global mean behavior of dust in response to climate, as shown in the correlation reanalysis. The same is also true for the regional level analysis for North Africa, Middle East and East Asia. In addition, the dust-climate connection may not be properly manifested at station levels, because most stations are impacted by transported events from upwind sources and therefore, the climate variables (wind, soil wetness) observed at those stations likely do not properly represent the climate conditions driving the dust variations. On the other hand, the stations located near dust source areas are complemented by downwind observations to create a full long-term dust record, given that those stations have a great chance of interruption and discontinuity.

14. **Lines 324-325:** The vertical structure of the dust layers plays a key role when attempting to compare spaceborne retrievals with near-surface observations (see Section 4.4 in Gkikas et al. (2016)).
    Response: Thanks for pointing this out. It now reads "weather stations focus on dust conditions in the lower boundary layer, whereas satellites retrieve the total aerosol column and are unable to detect aerosols below clouds. As a result, the variable vertical profile of dust can affect the relationship between the ambient dust condition (f and β) and total column dust burden.".

15. **Line 338:** How the global means from the weather stations are calculated? Are you using any weight based on the data availability?
    Response: the global mean from weather stations is calculated by aggregating the weather reports from all eligible stations on the monthly basis (see Eq. 1). No weighting is applied.

16. **Figures 11, 13 and 15:** Same comment as for Figure 9. In addition, please use always red and blue color for the positive and negative phases, respectively, for the teleconnection patterns.
    Response: see my response to #9. For the color scheme of climate indices, blue represents wet conditions, and red represents dry conditions in the study region. I can make the change as suggested.