Impact of Topographic Wind Conditions on Dust Particle Size Distribution: Insights from a Regional Dust Reanalysis Dataset

Xinyue Huang[1], Wenyu Gao[2], Hosein Foroutan[1]

[1]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, 24060, USA
[2] Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC, 28223~~North Carolina~~, USA

Correspondence to: Hosein Foroutan (hosein@vt.edu)

Abstract. The size of windblown dust particles plays a critical role in determining their geochemical and climate impacts. This study investigates the relationship between topographic wind conditions (speed and direction relative to land slope) and the particle size distribution of dust emissions on a regional scale. We used the Multiscale Online Nonhydrostatic Atmosphere Chemistry (MONARCH) dust reanalysis dataset, which assimilates satellite data on coarse-mode dust optical depth (DOD_{coarse}). Validation against flight measurements from the 2011 Fennec campaign confirms the reanalysis's effectiveness in capturing coarse to super coarse dust. A 10-year dust reanalysis underwent selective screening to identify events with fresh emissions and the fraction of coarse dust concentrations was calculated as a surrogate for size distribution. The coarse fractions and a~~A~~ssociated meteorological and land characteristics obtained from various datasets were incorporated into multiple linear regression and machine learning models. Results indicate that dust particle size increases with wind speed, likely due to a higher fraction of fresh emissions and reduced deposition of coarse dust under stronger winds. A positive correlation between dust size and uphill slope suggests that enhanced vertical transport of dust by topography outweighs the impact of shifting emission microphysics over veering winds. Both positive correlations weaken in the afternoons and summer, likely due to the turbulence during haboob storms, which can suspend coarse dust from aged emissions, competing with the effect of uphill slopes. These findings on dust size dependency on topographic winds could improve dust cycle representation and its impacts.

1 Introduction

Windblown dust particles emitted from arid and semi-arid areas are the largest terrestrial contributor to global aerosols (Brasseur and Jacob, 2017). Understanding the size of the airborne dust particles is crucial for assessing their impacts on global climate and biogeochemistry. Size plays an important role in determining the emission and deposition of dust particles and thus the redistribution of soil nutrients around the globe (Knippertz, 2017; Duce and Tindale, 1991; Wang et al., 2023). Additionally, particle size can affect the likelihood of microorganisms' attachment to dust aerosols (Polymenakou et al., 2008; Yamaguchi et al., 2012) and the distance these associated microorganisms can travel (Prospero et al., 2005; Kellogg and Griffin, 2006). Particle size, along with other factors such as mineralogy, chemical composition, and shape, controls the climate impacts of dust particles which can vary drastically from warming to cooling (Adebiyi and Kok, 2020; Kok et al., 2023; Mahowald et al., 2014). Realistic representation of dust particle size distribution (PSD) in the atmosphere requires an understanding of the dependencies of dust size at emission (Kok, 2011; Rosenberg et al., 2014).

Relationships between dust PSD and various environmental factors have been extensively studied. Soil moisture has been widely reported to increase the proportion of coarse dust in emissions by enhancing bonding forces among fine particles within soil aggregates (Dupont, 2022; González-Flórez et al., 2023; Shao et al., 2020), but one recent study also argues that the effect is non-

monotonic (Ma et al., 2023). Another surface characteristic of interest is soil texture. Emitted dust PSD during wind tunnel experiments was reported to be significantly influenced by the fully dispersed soil size distribution (Wang et al., 2021), and proportion of emitted submicron particles and PM_{10} increased after tillage practices which broke down soil aggregates (Fernandes et al., 2023; Katra, 2020). The effect of atmospheric stability on dust size is inconsistent across studies. Some reported that an unstable atmospheric boundary is associated with richer submicron particles (Khalfallah et al., 2020; Shao et al., 2020), whereas others found no effect (Dupont, 2022; González-Flórez et al., 2023). Deposition impacts are universally acknowledged and were evaluated using characteristics of dust events or dust measurements, including fetch length (González-Flórez et al., 2023), dust age (Dupont et al., 2015; Ryder et al., 2013), and dust measurement height (Khalfallah et al., 2020; Shao et al., 2020). Nevertheless, the effects of multiple factors often intertwine during dust events, making the overall impact on dust size obscure.

Wind speed, or the resulting friction velocity (u_*) exerted on the ground surface, as the driving force of dust emissions, is one of the most essential parameters for dust emissions. The relationship between wind speed and PSD of dust emissions has been widely studied, yet consensus is lacking. Saltation-bombardment and aggregate disintegration are usually considered the primary mechanism for dust emissions (Kok et al., 2012). Parametrization of saltation-bombardment proposed that higher u_* leads to higher energy in saltating particles and thus enhances the breaking down of soil aggregates and ejection of fine particles (Shao, 2001). This theory is supported by multiple wind tunnel experiments (Alfaro et al., 2022; Wang et al., 2021) and field measurements (Chkhetiani et al., 2021; Dupont, 2022; Khalfallah et al., 2020). The brittle fragmentation theory, on the other hand, postulates that PSD of vertical dust flux is independent of u_*, backed by compiled data from multiple wind tunnel and field measurements (Kok, 2011). Some also proposed that detachment of submicron particles from the surface of soil aggregates is more common when kinetic energy of impacting particles is low and the ejection of coaster particle from fragmentation becomes increasingly dominant with higher impaction intensity (Malinovskaya et al., 2021). These discrepancies are partially due to the interplay of other factors, such as inconsistencies in dust emission measurements (Khalfallah et al., 2020; Shao et al., 2020), soil moisture (Ishizuka et al., 2008; Shao et al., 2020; Webb et al., 2021), and whether steady state saltation is reached (Mahowald et al., 2014). Selection of the dust emission properties (e.g., dust flux or dust concentration; Shao et al., 2020) and the height of dust measurements (Khalfallah et al., 2020) can alter the dust PSD. The effects of soil moisture on shifting dust PSD at emission can get entangled with the potential effects of u_*. For instance, the fine fraction in dust emissions counterintuitively increasing with decreasing u_* after light rain was caused by drying of the weakly crusted soils over time (Shao et al., 2020). With interferences of various factors, predicting the general dependency of PSD of dust emission on u_* at regional scales over longer time becomes complex.

The role of topography in altering size of dust emission is of emerging interest. The orographic ~~channeling~~channelling of winds can affect the dust emission or transport (Caton Harrison et al., 2021; Rosenberg et al., 2014). Uphill winds can enhance the vertical transport of dust particles through flow separation, especially increase the proportion of coarse particles in the elevated dust based on computational simulations (Heisel et al., 2021). Moreover, the veering angle between wind vectors and the surface inclination can affect~~ed~~ the emitted dust PSD. A study over a local field ~~(Malinovskaya et al., 2021)~~observed that ~~C~~compared to winds that blew more parallel to the ridges of the slopes (i.e., tangential winds), frontal uphill winds generated a higher fraction of fine particles (0.2-2 μm) because of more sputtering of fine particles on the windward slope due to resistance from the secondary aeolian structures, as well as less generation of coarse particles (2-5 μm) on the leeward slope with the recirculation zone (Malinovskaya et al., 2021). Other potential topographic impacts include the generation of erodible material by certain orographic winds (Washington et al., 2006) and the triggering of convective storms by mountains (Knippertz et al., 2007). However, their associations with dust PSD are understudied. Overall, it remains unclear whether the observed effects of wind over local topography on the PSD of dust emissions is detectable at regional scales and what the overall relationship between topography and dust PSD will be.

79   Understanding the relationship between topographic wind conditions and PSD of dust emission on a regional scale is important
80   for simulating dust activities and impacts in atmospheric or climate models. Complementing the accumulating field data on PSD
81   of dust emissions (Shao et al., 2020; González-Flórez et al., 2023; Fernandes et al., 2023), this study aims to explore the impacts
82   of topographic wind conditions on dust PSD on a regional scale through data analysis. The regional scale means that the dust
83   emission of concern inevitably includes near-source transport and deposition. Here, we selected "fresh" dust emission events from
84   the Multiscale Online Nonhydrostatic AtmospheRe Chemistry model (MONARCH) dust reanalysis data (Di Tomaso et al., 2022)
85   and constructed models to investigate the correlations between PSD of surface dust concentrations in fresh emissions and wind
86   conditions over slopes, while taking into account other relevant meteorological and surface conditions. Our methodology,
87   sensitivity analysis, and evaluations are described in Section 2, and the discussion of our findings followed by the main takeaways
88   are presented in Sections 3 and 4.

89

2 Data and Methods

### 2.1 Datasets and variables

92   The study domain (12-38° N, 18W-36° E) encompasses the Sahara Desert, the largest dust source on Earth, which contributes to
93   around 60% of the global dust loading (Tanaka and Chiba, 2006). Various monitoring or reanalysis datasets are available for this
94   region, providing information on African dust sizes and the associated environmental conditions needed for this study.

95   The Multiscale Online Nonhydrostatic AtmoshpheRe Chemistry model (MONARCH) dust reanalysis (Di Tomaso et al., 2022)
96   dataset provides size-resolved dust information from 2007 to 2016 with a 3-hour intervals. The dataset covers North Africa, the
97   Middle East, and Europe using a rotated-pole projection with a spatial resolution of 0.1°. The assimilation data of reanalysis
98   assimilated the coarse-mode dust optical depth (DOD_{coarse}) were derived from the Moderate Resolution Imaging Spectroradiometer
99   (MODIS)-Aqua Deep Blue level 2 aerosol products (Collection 6), including aerosol optical depth (AOD), and the Ångström
100 exponent, and the single scattering albedo at different wavelengths (Ginoux et al., 2012; Pu and Ginoux, 2016). MONARCH's
101 first-guess dust size distribution follows the brittle-fragmentation theory of Kok (2011) with perturbations across 12 ensemble
102 members. By applying a local ensemble transform Kalman filter with four-dimensional extension (4D-LETKF) at each 24-hour
103 assimilation window, The DOD_{coarse} observations were incorporated with reanalysis increments were added to the model ensemble
104 simulations (first-guess) to match the DOD_{coarse} observations during each 24-hour assimilation window. Specifically, the dust state
105 vector of the total coarse dust mixing ratio (distributed across five coarser bins from 1.2 to 20 μm) was updated, then the increments
106 for the finer three bins were determined proportionally to their total relative mass. Therefore, although MONARCH reanalysis
107 does not directly assimilate fine-mode DOD, corrections in the coarse bins propagated to the entire PSD through the assimilation
108 state vector and physical parameterizations, aligning the PSD more closely with dust-specific observations. Consequently, if the
109 prior PSD is biased—for instance by placing too much mass in the largest bin or not enough in a medium bin—that bias may
110 persist to some extent after assimilation. Despite the limitation, validation against AERONET data indicates that fine dust is still
111 captured satisfactorily (Di Tomaso et al., 2021; Mytilinaios et al., 2023), supporting the reliability of the dataset to investigate dust
112 PSD. We hypothesize that even though the dust PSDs at emission are fixed in the dataset for the first-guess, the dust concentration
113 reanalysis captures the potential regional effects of topographic wind conditions on the dust PSD via assimilation of the satellite
114 DOD_{coarse} observations, nudging the size-resolved concentrations to generate expected optical properties. We showcase the The
115 adjustments in dust concentration PSD during data assimilation were showcased by investigating the uneven ratio of the first-guess

dust concentration to its reanalysis across eight size bins (see section 2.2 and Fig. S1). The MONARCH model was run in 40 hybrid pressure-sigma model layers and the dust concentration in the lowest layer is saved as the surface dust concentration. The surface dust concentration best represents near-source dust emissions and transport, and was chosen for this study. The concentration is available in eight size bins (i.e., 0.2–0.36, 0.36–0.6, 0.6–1.2, 1.2–2.0, 2.0–3.6, 3.6–6.0, 6.0–12.0, and 12.0–20.0 µm; Klose et al., 2021)). For easier comparison, these size-resolved concentrations were condensed into a single index called the "coarse fraction", defined as the sum of mass concentrations of the coarsest two bins (6-12 µm) divided by the total mass concentration of all eight bins (referred to as "cf2"). The delineation of fine and coarse particles is somewhat arbitrary and case-specific across studies. In general, a cutoff diameter from submicron to above 10 µm was used (Dupont, 2022; Fernandes et al., 2023; Panebianco et al., 2023) and the most common range is roughly 2-5 µm (Ishizuka et al., 2008; Ryder et al., 2019; Shao et al., 2020; Webb et al., 2021), which generally aligns with the lower boundary of 6 µm used in this study. Additionally, we tested alternative definitions of coarse fraction (namely "cf1" and "cf3", where the coarsest one or three bins are assigned as coarse dust) in the subsequent statistical analysis and results suggest that cf2 is a representative surrogate for dust particle size. More details on the comparison are included in Table S4 and Section 3.3.

Because wind conditions associated with dust concentrations are not available from the MONARCH dust reanalysis dataset, we sourced the information from the Modern-Era Retrospective analysis for Research and Applications (MERRA-2) data (Gelaro et al., 2017). MONARCH ensemble simulations applied meteorological inputs from two reanalysis datasets, i.e., MERRA-2 and ERA-Interim. Given that wind from both reanalyses are highly constrained by observations, and there is a substantial overlap in the assimilated data used by the two (Fujiwara et al., 2024; Rienecker et al, 2008; Dee et al., 2011; Gelaro et al., 2017), it is reasonable to use MERRA-2 wind vectors to inform the wind conditions of MONARCH dust reanalysis. The available wind components nearest to the surface and most relevant to dust emissions are at 2 meters above ground, provided as hourly average with a spatial resolution of 0.5° latitude × 0.625° longitude in the product M2I1NXASM. Wind speed and wind direction are subsequently calculated.
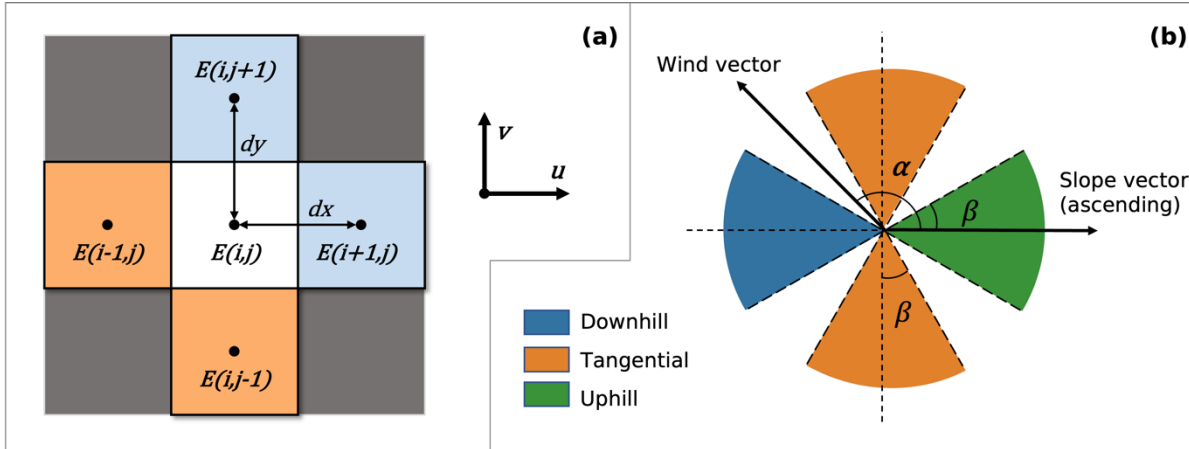
Figure 1. (a) Variables used in the calculation of upwind slope. The elevations in the target grid, E(i,j) and the upwind grids (in orange if the wind components were positive; in light blue if the wind components were negative) were used to calculate the upwind slope components in x and y directions. (b) Methods for assigning the type of wind direction over topography. The angle between the upwind slope and the wind vector, α, and the predefined cut-off angle, β, together determine whether the wind from an event is typical for any of the three categories of relative wind direction, namely downhill, tangential, and uphill winds.

146 The wind direction types relative to topography (hereafter "relative wind direction type") were determined based on the angle, α

147 $(0 \leq \alpha < 360°)$ between the wind vector ($u(i,j)$ and $v(i,j)$) and the slope vector ($S_x(i,j)$ and $S_y(i,j)$). Elevation data for calculating the

148 slope were retrieved from the NASA Shuttle Radar Topography Mission Global 3 arc-second (SRTM GL3) dataset (Farr et al.,

149 2007). The SRTM elevation data are expected to be compatible with the MERRA-2 wind reanalysis because 1) both the SRTM

150 and GTOPO30, which is used in the Goddard Earth Observing System (GEOS-5) model (MERRA-2's first-guess), have much

151 finer than MERRA-2's grid, making their spatially averages comparable, and 2) MERRA-2 wind reanalysis are highly constrained

152 by assimilated observations, reducing its dependency on topographic input. Dust concentration is most relevant to emissions in

153 upwind grids, therefore, slope vector components in each grid (i,j) were determined as gradients of elevation between the target

154 grid and the two neighboring grids in the upwind directions (Fig. 1(a) & Eq. (1)). To the best of the authors' knowledge, this study

155 presents the first derivation of the upwind slope over North Africa.

156 The wind direction over topography was categorized into tangential, uphill, and downhill winds depending on the angle α between

157 the slope and wind vectors as well as a predefined cut-off angle, β $(0 < \beta \leq 45°)$ (Fig. 1(b) & Eq. (2)). A smaller cut-off angle

158 results in a more selective process for assigning the relative wind direction types.

159

160
$$S_x(i,j) = \begin{cases} \frac{E(i,j)-E(i-1,j)}{dx}, & u(i,j) > 0 \\ \frac{E(i,j)-E(i+1,j)}{dx}, & u(i,j) < 0 \end{cases}, S_y(i,j) = \begin{cases} \frac{E(i,j)-E(i,j-1)}{dy}, & v(i,j) > 0 \\ \frac{E(i,j)-E(i,j+1)}{dy}, & v(i,j) < 0 \end{cases}, \tag{1}$$

161

162 where (i,j) denotes the location of a grid cell, $E(i,j)$ represents the elevation, and $S_x(i,j)$ and $S_y(i,j)$ represent the slope components

163 in x and y directions, respectively. The $u(i,j)$ and $v(i,j)$ are horizontal wind components.

164
$$\text{wind type} = \begin{cases} \text{uphill,} & 0 < \alpha < \beta \text{ or } (360° - \beta) < \alpha < 360° \\ \text{tangential,} & (90° - \beta) < \alpha < (90° + \beta) \text{ or } (270° - \beta) < \alpha < (270° + \beta), \\ \text{downhill,} & (180° - \beta) < \alpha < (180° + \beta) \end{cases} \tag{2}$$

165

166 The land characteristics of soil texture and soil moisture that are expected to cast impacts on PSD of dust emissions were

167 considered. Spatial distribution of soil texture was adopted from the map used in the Global Land Data Assimilation System version

168 2 (GLDAS2) Noah land surface model (Rodell et al., 2004), derived from the global soil dataset by Reynolds et al. (2000). The

169 texture of the top layer of soil was categorized into the 16 classes developed by the Food and Agricultural Organization (FAO),

170 varying in sand, silt, and clay fractions (Jahn et al., 2006). Soil moisture data were retrieved from the MERRA-2 product

171 M2T1NXLND, providing average water content in the top 5-centimeter layer of soil hourly with a spatial resolution of 0.5° latitude

172 × 0.625° longitude.

173 All datasets were co-registered onto a universal 0.1° Plate Carrée coordinate. MERRA-2 and MONARCH data were regridded

174 using the Python xesmf package version 0.7.1 (Zhuang et al., 2023). The 2 m wind vectors and the soil moisture from the MERRA-

175 2 reanalysis underwent upsampling using the "nearest source to destination" algorithm to match MONARCH's finer resolution.

176 This algorithm does not bring in artificial variations so is the safest choice for regridding. Coarser spatial resolution of the MERRA-

177 2 data means some neighboring grids will inevitably share the same wind vector and soil moisture, diminishing the potential effects

178 of wind conditions on dust PSD. The SRTM elevation data with a much higher original resolution were downsampled using the

179 "average" method provided by the Python package of geowombat version 2.1.6 (Graesser, 2023). Sub-grid information on

topography was lost, but handling topography information at the same scale as the wind data is reasonable because the terrain variations at finer resolution are deemed smooth when scaling up. Soil texture data at 0.25° lat-lon coordinates were also projected to the 0.1° coordinates. To match the instantaneous 3-hourly timesteps of the MONARCH reanalysis, we picked the average wind components and soil moisture from the precedent hour, relevant to the initial dust emissions that can remain airborne during that period. Maps showing the average wind speed, the average upwind slope, the most common wind direction types, and the average surface dust concentration at 0.1° resolution over the 10 years are presented in Fig. 2. While we acknowledge the inherent resolution limitations of reanalysis datasets, the focus of this study is on the broader-scale modulation of dust emission by wind conditions, and data assimilation combined with upsampling techiniques ensure that our conclusion remain interpretable in this context.
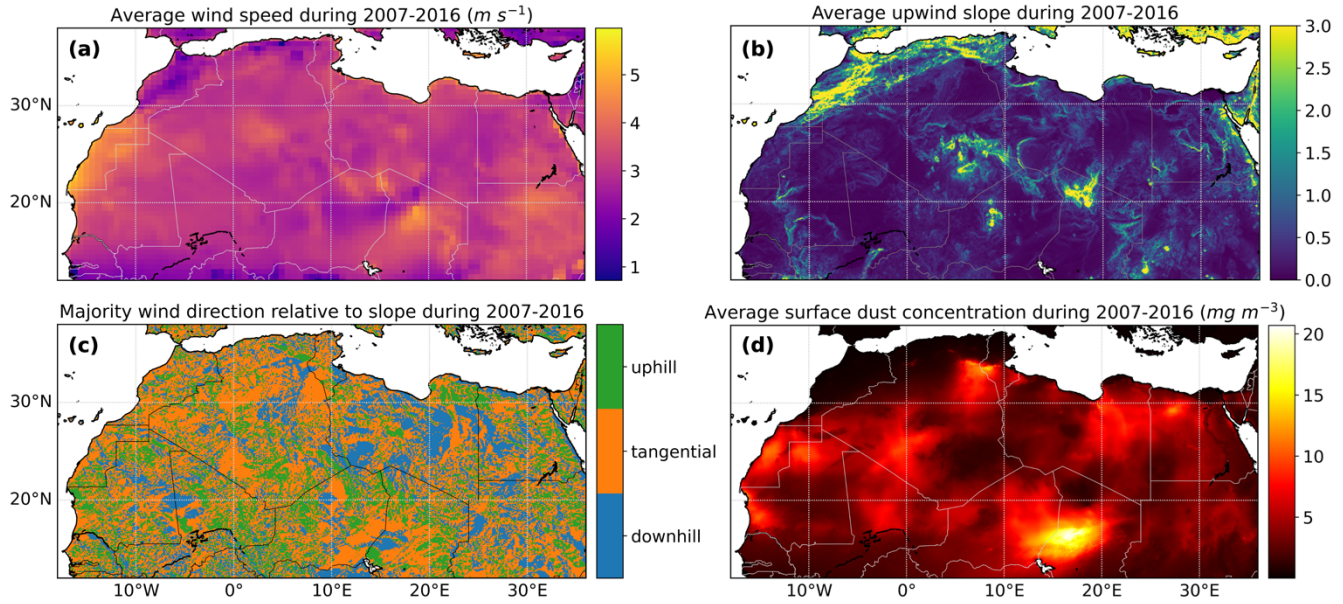
Figure 2. General spatial patterns of (a) 2-m wind speed from the MERRA-2 reanalysis, (b) calculated upwind slope, (c) derived wind direction type relative to slope using cut-off angle of 10°, and (d) MONARCH total dust concentration at ground surface from 2007 to 2016. The average values are shown for wind speed, upwind slope, and dust surface concentration, and the most frequent types are shown for the relative wind direction.

2.2 Validation against the Fennec measurements

The MONARCH dust reanalysis dataset was previously evaluated against observations from the Aerosol Robotic Network (AERONET) retrievals (Di Tomaso et al., 2021; Mytillinaios et al., 2023). Comparison between the~~The~~ 3-hourly MONARCH reanalysis of DOD$_{coarse}$ at 550 nm ~~was evaluated against~~ and the coarse-mode AOD at 500 nm from ~~the Aerosol Robotic Network (AERONET)~~ retrievals over Sahara generate a~~. The~~ Pearson correlation of~~was~~ 0.81 with a root mean square error of 0.15 (Di Tomaso et al., 2022). Here, we present an additional case study to particularly evaluate the performance of dust reanalysis on capturing fresh dust emissions. Observational data were obtained from ~~We expanded the evaluation of MONARCH dust concentrations using data of~~ the 2011 Fennec campaign, where size-resolved dust emissions over western Africa were intensively sampled using wing-mounted instruments (Ryder et al., 2013). Segments of three flights (b600–602), each lasted 10 minutes, over northern Mali on 17–18 June 2011, were identified to be associated with fresh dust uplifts by low-level jets (Ryder et al., 2013; Ryder et al., 2015). Measured dust number concentration during these flight segments was converted to volumetric concentration for easier comparison with the MONARCH reanalysis data. The MONARCH reanalysis grids containing any portion of these

flight trajectories were identified, and the associated dust mass concentrations were retrieved. These concentrations were weighted averaged by flight duration in each grid cell to yield an overall binned dust concentration. The MONARCH dust mass concentrations were also converted into volumetric concentrations using ~~an average~~ the dust particle density of 2500 kg m^{-3} for the finer four bins and 2,650 kg m^{-3}, for the coarser four bins (Klose et al., 2021) ~~which had previously been used to convert number concentration of dust particles to mass concentration for Fennec measurements (Ryder et al., 2019)~~. As shown in Fig. 3, the trends of two dust PSDs generally agree well across all the eight size bins of the MONARCH dataset. Most notably, MONARCH reanalysis is effective at capturing the coarse to super coarse modes (defined as dust with diameter greater than 10 µm; Meng et al., 2022) represented by the last three bins (3.6–20 µm), outperforming several recent dust simulations that lack the data assimilation (Adebiyi and Kok, 2020; Meng et al., 2022). Additional investigations (see Fig. S1) revealed that the reanalysis dust concentration changes non-monotonically from its first-guess, leading to the conclusion that not only the total dust concentration but also its PSD has changed due to assimilation. Overall, ~~the comparison~~ the predicted concentration for fresh emitted dust lies in an acceptable range, and the data assimilation process improved concentration across all bins as well as adjusted the dust size distribution, suggesting that the MONARCH reanalysis reasonably represents fresh dust emissions ~~confirms the reliability of the MONARCH dust reanalysis in representing dust size distribution~~.
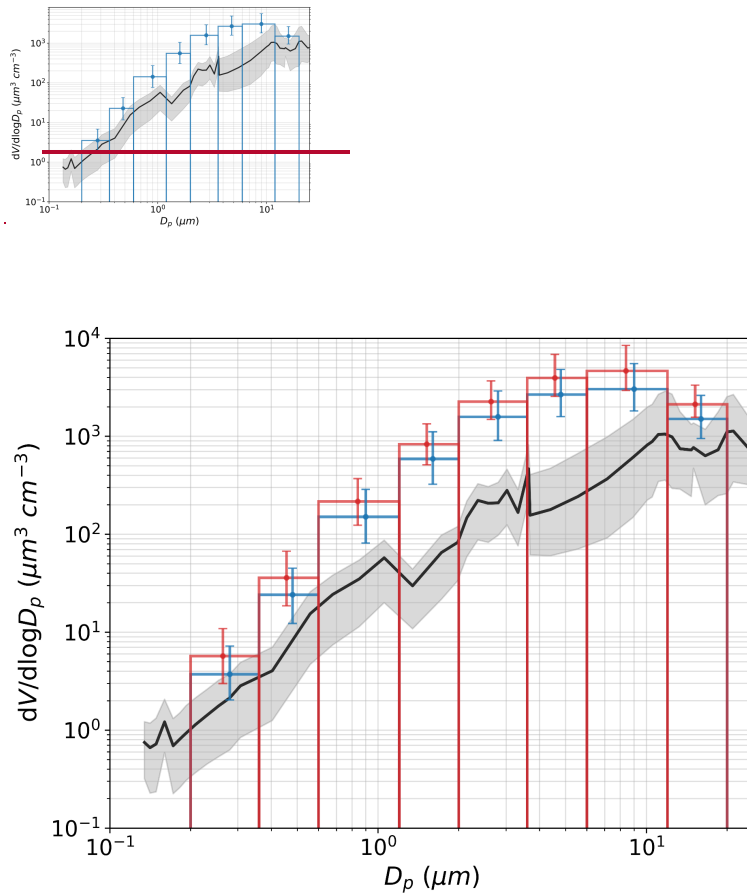

Figure 3. The black line shows the average volumetric concentration (µm^3 cm^{-3}) of dust sampled during three Fennec flights (6 flight segments) and the grey shaded area denotes the range of values. The blue bars show the volumetric concentration (µm^3 cm^{-3}) of dust in corresponding grids from the MONARCH reanalysis calculated from the weighted average mass concentration, with a particle density of 2,500 kg m^{-3} for the finer four bins and 2,650 kg m^{-3} for the coarser four bins ~~assuming a particle density of 2650 kg m^{-3}~~. The error bars denote the range of values.

2.3 Event selection and sensitivity tests

Addressing the specific goals of this study requires the selection of the most relevant dust events from a decade of data over northern Africa. Two goals guided the screening process: 1) excluding the aged dust, and 2) focusing on wind conditions over distinctive terrain variations. Several screening criteria were accordingly developed that must be met simultaneously for dust events to be eligible for further analysis. Specifically, the selected dust events must occur within dust sources and over terrain with prominent slopes, be concurrent with high wind speeds and typical wind directions over slope, and have notable increases in dust surface concentrations. The procedures for screening these events are illustrated in Fig. 4 and described below in detail. This highly selective approach was made possible by the abundance of data from the MONARCH reanalysis.

To start with, dust events were confined to dust sources to exclude long-range transported and likely aged dust far from dust-source regions. The 10-year average of surface dust concentrations was calculated for the entire study domain, and pixels with values above the threshold percentile were designated as dust sources. For example, dust sources selected using the 80th percentile of the 10-year average dust concentration as the threshold are presented in Fig. 4(a). The map of dust sources features the Bodélé Depression, Great Sand Sea, Tanezrouft, and the Atlantic Coastal Desert, consistent with the dust sources identified in other studies (Formenti et al., 2011; Di Tomaso et al., 2022). Further screening steps were performed independently within these dust sources.
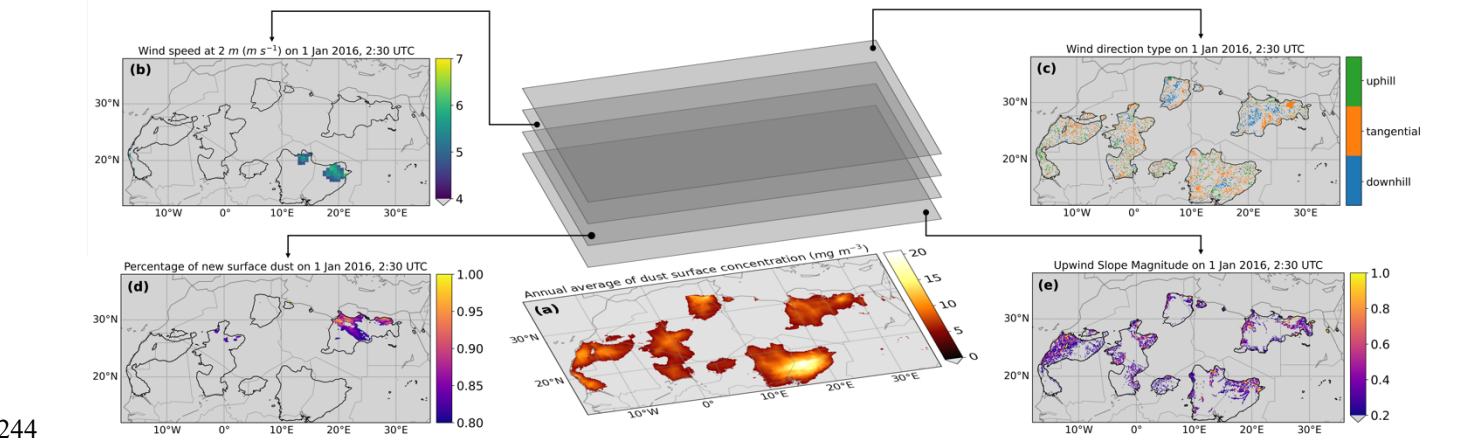
Figure 4. Schematics showing the procedures for screening the 10-year dust reanalysis data for fresh dust emissions. First, dust sources were assigned to regions where the annual average of dust surface concentration is among the top percentile (i.e., 20% as shown in (a)) of all values across the study domain of North Africa. Subsequently, additional screening criteria were applied simultaneously to events that occurred within these dust sources. These criteria include (b) high wind speed, (c) wind direction over topography, (d) high increase in dust surface concentration, and (e) steep slope. Maps (b–e) are examples of filters used in the "initial" run (see in Table 1) for 1 January 2016, at 2:30 UTC.

To maximize the likelihood of capturing predominantly fresh dust emissions, high wind speed was a necessary criterion, as low wind speeds are unlikely to generate sufficient fresh dust. A single cut-off value for high wind speed was chosen as a top percentile of the 10-year average wind speed over the whole domain. Similarly, a single threshold slope was used to select the events over prominent terrain variations. Dust events occurring over relative flat surfaces were excluded to magnify the potential signals of shifted PSD due to terrain variation. Wind direction over slopes was categorized following procedures described in Section 2.1 and a cut-off angle smaller than 45° was used to exclusively select typical uphill, tangential, or downhill winds over slopes. Another criterion for increasing the probability of capturing dominantly fresh emissions was identifying sharp increases in surface concentrations. This approach favored the initiation of significant dust emissions with relative clean background dust levels. Examples of selected dust events based on each of the above criteria at 2:30 UTC on 1 January 2016, are provided in Fig. 43(b–

260 e), using the configurations for the "initial" run (i.e., a cut-off angle of 10° for categorization of wind direction, the 80th percentile

261 as threshold for high wind speeds and steep slope, and 80% as the threshold ratio of the increased surface dust concentration to the

262 total dust concentration as shown in Table 1).

263 We performed several sensitivity tests to account for the uncertainties associated with the criteria used for event selection. We

264 perturbed each of the threshold percentiles or the cut-off values used in the "initial" run in a pair of sensitivity runs. Specifically,

265 we added and subtracted 10% from the threshold percentiles of 80% used for screening dust sources, steep slope, high wind speed,

266 and high fraction of fresh emissions. Cut-off angles, β of 5° and 20° were tested compared to the initial 10° for typical wind

267 direction. Configurations of all the sensitivity runs are shown in Table 1. A combination of all the stricter criteria was used in the

268 "final" run.

269

270 **Table 1.** Thresholds or cut-off values used in all runs. Locations where the 10-year average surface dust concentration is above the threshold
271 percentile value of all locations were designated as dust sources. Events over dust sources with slope and wind speed above the threshold
272 percentiles of temporally averaged values over the whole domain, and increases in dust surface concentration above the threshold values were
273 selected. The wind direction relative to topography was assigned using the cut-off angle. Values marked in bold represent the different criteria
274 compared to the "initial" run.

Test name	Dust sources threshold percentile	Wind direction cut-off angle	Slope threshold percentile	Wind speed threshold percentile	Increase in dust surface concentration threshold
Initial	80	10°	80	80	80%
90%source	**90**	10°	80	80	80%
70%source	**70**	10°	80	80	80%
5cutoff	80	**5°**	80	80	80%
20cutoff	80	**20°**	80	80	80%
90%slope	80	10°	**90**	80	80%
70%slope	80	10°	**70**	80	80%
90%wdsp	80	10°	80	**90**	80%
70%wdsp	80	10°	80	**70**	80%
90%sconc	80	10°	80	80	**90%**
70%sconc	80	10°	80	80	**70%**
Final	**90**	**5°**	**90**	**90**	**90%**

275

276 **2.4 Statistical analysis**

277 Based on the selected events for "fresh" dust emissions, we explored the relationships between wind conditions and the PSD of

278 emitted dust, taking into consideration the effects of various meteorological and landscape factors. Specifically, the dependent

279 variable was the coarse fraction of surface dust, and the independent variables of focus were the wind speed and the ~~three variables~~

280 ~~of~~ slope associated with uphill, tangential, and downhill winds. ~~We separated the slope by wind direction types to provide a more~~

281 ~~holistic representation of the effects of veering wind over topography.~~ The five additional independent variables are environmental

282 variables that can potentially affect dust PSD as well as the relationships between dust PSD and wind conditions, including the

283 continuous variables of year and soil moisture, and the categorical variables of time of day, season, and soil texture type. Dust

284 events with missing values in soil texture class or wind direction type were excluded.

285 Exploratory data analysis was first conducted to assess data quality, identify intrinsic patterns, and guide the selection of

286 appropriate statistical models. ~~Exploratory data analysis was conducted to assess the quality and to identify the patterns of data, as~~

287 ~~well as to inform the selection of statistical~~ regression ~~models.~~ We selected and modified our statistical models based on their

288 adherence to model assumptions, ability to explain the variability in the coarse fraction, and overall complexity. An initial ~~One~~

~~potential~~ choice was~~is~~ the multiple linear regression model~~,~~ which has the advantage of high explainability. We separated the slope by wind direction types to provide a more holistic representation of the effects of veering wind over topography. Significant coefficients for continuous variables, such as wind speed, represent the change in the coarse fraction of dust concentrations associated with a one-unit change in that corresponding independent variable.~~Significant coefficients for continuous variables such as wind speed indicate the amount of change in coarse fraction of dust concentrations per unit change in the independent variables.~~ Categorical variables (e.g., time of day) are encoded as binary dummy variables, each representing a distinct category. The coefficients of these variables reflect the change in coarse fraction relative to the reference category chosen during the encoding process.~~Categorical variables (e.g., time of day) are converted into dummy binary variables, each representing a category, and their coefficients indicate the changes in coarse fraction when switching from the randomly decided reference category.~~ Interaction terms can also be added to linear models in order to ~~To~~ reflect the interplay between wind conditions and other factors~~, interaction terms are considered~~. An interaction can be expressed as $x_m \times x_p$ ~~$\beta_{mp}x_mx_p$~~, where x_m is one of the four wind condition variables and x_p is one of the two additional continuous variables (i.e., year and soil moisture) or a dummy variable representing ~~for the~~ one of the three additional categorical variables (i.e., time of day, season, and soil texture). The adjustment~~s~~ in the coefficient for x_m due to x_p would be represented by $\beta_{mp}x_p$. A valid linear model requires linear relationships, normality of errors, constant variance (homoscedasticity)~~normality, and constant variance on the errors (homoscedasticity)~~, and low correlations among independent variables. Collinearity or multicollinearity among predictors can inflate standard errors and reduce the statistical significance of regression coefficients. To assess this, we calculated the Generalized Variance Inflation Factors (GVIFs) for all predictors in the linear models using the VIF function in R (R. Core Team, 2023). For categorical variables, the GVIFs were adjusted by the degrees of freedom (Df), expressed as $GVIF^{1/(2 \cdot Df)}$. An adjusted GVIF of 1 (the smallest value) indicates no collinearity, while values below 5 generally suggest low and acceptable collinearity.

When assumptions of normality and constant variances were violated, we attempted standardization and transformations,~~,~~ ~~standardization, and transformation,~~ including~~such as~~ the Box-Cox transformation (Box and Cox, 1964) and the logit transformation (Berkson, 1944), ~~were applied on the variables~~ to tackle the~~se~~ issues. We also tested weighted linear regression, also known as weighted least squares (WLS) (Kiers, 1997), which can handle non-constant variance (heteroscedasticity) by assigning weights to observations.~~.Weighted linear regression, also known as weighted least squares (WLS) (Kiers, 1997)), was also tested to handle the non-constant variance (heteroscedasticity).~~

In addition, Beta regression model~~s~~ ~~was considered~~were implemented, which ~~is~~are particularly useful for fractional variables that range between 0 and 1, such as the coarse fraction in this study (Douma and Weedon, 2019). ~~It~~Beta regression has been previously ~~used for~~applied to ~~variables such as~~ air quality-related health metrics within the standard unit interval (Lu et al., 2021) and to particle size data with skewed distributions (Peleg, 2019).

Several evaluation metrics were used to compare model performance, including the root mean square error (RMSE), the mean absolute error (MAE), the adjusted coefficient of determination (adjusted R^2, which evaluates the amount of variability in the coarse fraction explained by the model while penalizing model complexity). Higher adjusted R^2, and lower RMSE or MAE values suggest that more variations of the data are captured by models. ~~The final model was chosen based on its compliance with model assumptions, ability to explain the variability of the coarse fraction, and model complexity. The adjusted R^2 was used for comparison among linear models to simultaneously evaluate the amount of variability in coarse fraction explained by the model, with penalty on complexed models. For more complex models, selection criteria based on the model likelihood values, such as the Akaike information criterion (AIC; Bozdogan, 1987), Bayesian information criterion (BIC; Neath and Cavanaugh, 2011), and the likelihood ratio test (Lewis et al., 2010), were considered.~~ We also calculated the prediction interval accuracy at the 95% confidence

level, defined as the proportion of observations covered by prediction intervals. The performance metrics were also averaged using the 10-fold cross-validation (CV), where the dataset was randomly divided into 10 subsets, and models were trained on 9 subsets and tested on the remaining subset in each iteration, with a total of 10 iterations. All~~The~~ statistical analyses were performed using R ~~(R. Core Team, 2023)~~ and the relevant packages (Zeileis and Hothorn, 2002; Grün et al., 2012; Cribari-Neto and Zeileis, 2010; Fox and Weisberg, 2019; Venables and Ripley, 2002).

Moreover, we constructed machine learning models to account for the large dataset and potential non-linear relationships. Categorical variables (wind direction, soil texture type, season, and time of day) were converted into dummy variables, resulting in a total of 22 predictors when combined with continuous predictors (wind speed, slope, and year). We built Random Forest and Extreme Gradient Boosting (XGBoost) models, both are widely used regressors (Bacanin et al., 2024; Brokamp et al., 2017; Keller & Evans, 2019; Zhang et al., 2022). Though both models rely on decision trees, Random Forest aggregates multiple trees trained on randomly sampled subset of data, whereas the XGBoost sequentially refines decision trees through iterative training. Model hyperparameters were fine-tuned to maixmize the predictive performance, with the search grids determined based on sample size, predictor count, and computational efficiency. The search grids and the optimal hyperparameter combinations are listed in Table S1. As with the linear models, we assessed the accuracy of prediction interval coverage at 95% confidence level through 10-fold CV. Machine learning models are known to have high prediction accuracy but can be challenging to interpret. One technique to assess the contribution of individual predictors to the coarse fraction based on decision trees is the SHapley Additive exPlanations (SHAP) analysis, which was performed on the optimized models. These analyses were conducted using Python packages including scikit-learn (version 1.2.2; Pedregosa et al., 2011), xgboost (version 1.7.6; Chen and Guestrin, 2016), and shap (version 0.44.0; Lundberg and Lee, 2017).

Eventually, the more complex models did not outperform the multiple linear models to a large extent. Given their high explainability, ability to incorporate interactions between predictors, and competitive performance, we utimately selected linear models for further analysis. SHAP results from the machine learning models were also included for cross-validating key findings. More details on the model performance and results are described in Section 3.3.

3 Results and Discussion

3.1 Sensitivity tests show minor variations

As a result of the screening processes, a total number of 461,183 dust events were identified in the "initial" run and 25,884 events were identified in the "final" run from around 3.5 billion records of dust concentrations at specific locations and times. Figure 5 shows the percentage changes in median coarse fraction by wind direction types from all sensitivity runs. In general, nearly all perturbations of any single screening criterion result in around ±1% of changes in average or median coarse fraction grouped by wind direction. The exception is restricting wind speed to the top 10% percentile (the "90%wdsp" run), which leads to a roughly 2% increase in the median values for each wind direction type. Coarse fraction of dust emissions with downhill winds are usually more sensitive to the threshold values used for screening than emissions with the other two wind directions. When all criteria were restricted simultaneously in the "final" run, the median coarse fraction of dust emission increases by less than 2% under tangential or uphill winds and around 3.5% under downhill winds. Considering the minor variations in coarse fraction among sensitivity runs, we decided to focus on one run for more detailed analysis. The "final" run was chosen because, in theory, the selected events are most representative for fresh dust emissions.

365
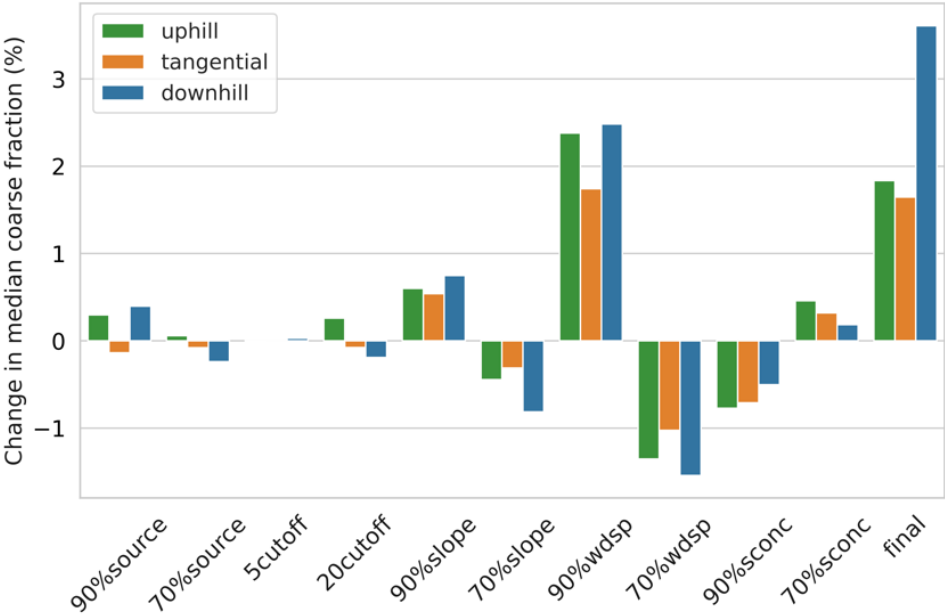
Figure 5. Percentage change in the median coarse fraction by wind direction type for all sensitivity runs as compared to the "initial" run.

368

3.2 Exploratory analysis

To gain an impression of the general distribution of the data, we started with~~by~~ plotting the dust coarse fraction against four variables of wind conditions in Fig. 6. Scatter points in the panel for wind speed are color-coded by number of overlapping observations, and data points in other three panels for slopes are color-coded by wind speed. Across all four panels, a pattern of heteroscedasticity is revealed, that is, the variance of coarse fraction is greater for dust events associated with lower wind speed or slope than for those with high wind speed or slope. The vertically aligned scatter points with varying colors at a slope of around 1.7 in the panel for "slope under uphill winds" represent 93 dust events that occurred at a same location under different wind speeds near the northern border of Western Sahara during 2007–2016, illustrating how a larger number of dust events can lead to higher variance. No obvious non-linear relationships between the four wind condition variables and the coarse fraction are observed. The trend lines based on simple linear regression models of the coarse fraction against each wind condition variable indicate general trends in the dust PSD with varying wind conditions, but the significance of these relationships is not assured. Additional plots for the soil moisture and slope under three wind conditions with color-codes representing the density of overlapping data points is presented in Fig. S2.
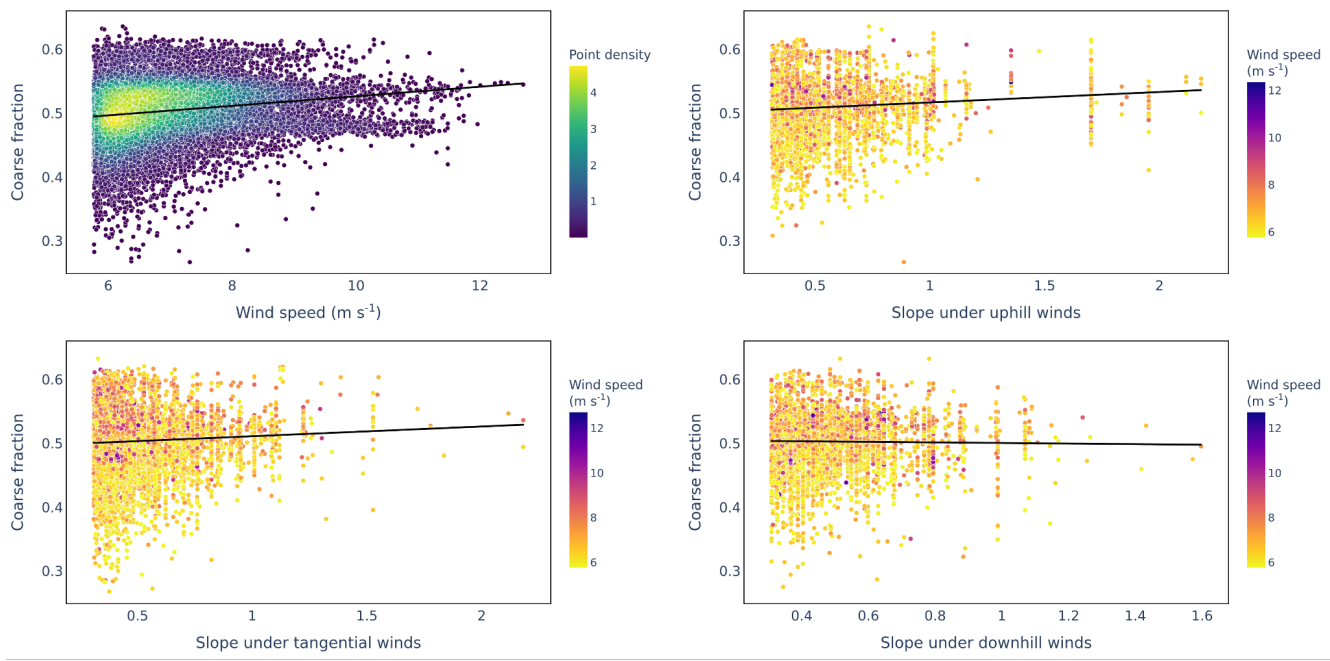
382
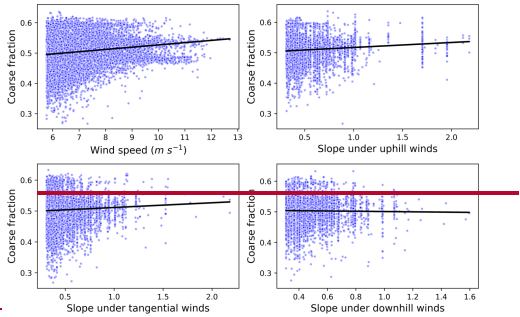
383 .

Figure 6. Scatter plots and linear trend lines of relationships between the coarse fraction of surface dust concentration and the wind speed and the slope under three different wind directions. In the upper left panel for wind speed, the color-codes present the number of overlapping data points. In the other three panels for slope, the color of each scatter point represents the associated wind speed for the dust event.

3.3 General associations between wind conditions and dust coarse fraction

~~Due to the~~ The absence of obvious non-linear relationships between coarse fraction and wind conditions from the exploratory analysis~~,~~ further motivated us to initially ~~we first~~ construct~~ed~~ linear regression models, in addition to it being a common starting point. The linear model ~~that~~ includ~~ing~~es all independent variables but no interactions has an adjusted R^2 value of 0.224, RMSE with 10-fold CV of 0.070, and MAE with 10-fold CV of 0.059. Additionally, we added all possible interactions into the model and fine-tuned it by removing ~~then removed~~ the insignificant interaction terms~~,~~. The resulting model with interactions has similar performance, with ~~an~~ the adjusted R^2 of 0.239, RMSE with 10-fold CV of 0.070, and MAE with 10-fold CV of 0.058. The somewhat weak correlations may be related to factors that are not included in the model, such as deposition, variability within the same categories of soil texture, etc. ~~Additional~~Residual analyses ~~suggest~~indicate ~~that the residuals~~ violate~~ions of~~ the assumptions of evenly distributed variance and normality (Fig. S3~~2~~). ~~We attempted variations of linear models, including standardization, Box-Cox transformation, logit transformation, and WLS but they all failed to resolve the issues. Beta regression models were also tested, and the best link functions were the log-log link function for the mean and the identity link function for the dispersion. However, the residuals still exhibit heteroscedasticity against the wind condition variables.~~ We attempted to address these issues and improve model performance through various linear model adaptations but only obtained indefinite or marginal improvements—standardizing the coarse fraction and wind conditions variables yields an adjusted R^2 of 0.241, RMSE with 10-fold CV of 1.678, and MAE with 10-fold CV of 1.420; and logit transformation on the coarse fraction generates an adjusted R^2 of 0.237, RMSE with 10-fold CV of 1.420, and MAE with 10-fold CV of 0.487. Furthermore, the Box-Cox transformation, weighted least squares (WLS), or Beta regression models with the best-performing configuration (with a log-log link function for the mean and an identity link function for the dispersion) all failed to resolve heteroscedasticity in the residuals with respect to wind condition variables. For the machine learning models, the optimized Random Forest and XGBoost models achieved R^2 values of 0.407 and 0.474, respectively, which dropped to 0.259 and 0.273 after 10-fold CV, indicating potential overfitting. The coverage rates of

410 prediction intervals with the 10-fold CV are 92.5% for Random Forest and 54.4% for XGBoost, both lower than the 94.0%

411 coverage by the linear models both with and without interactions. Linear models outperforming machine learning models on

412 prediction accuracy strongly encourages the selection of linear models. ~~Following all these tests and considering model complexity~~

413 ~~and interpretability~~Considering their comparative satisfactory performance, simplicity, and directly interpretable coefficients, we

414 decided to proceed with the linear regression models among all models. ~~The results can serve as initial guidance for our study. To~~

415 ~~better understand the general effects of wind conditions, we first present the model without interaction terms.~~We first assessed the

416 influence of individual predictors based on the linear regression model without interactions, with cross comparison with results

417 from the machine learning models (Fig. S5). Subsequently, we used the linear model with interaction terms to investigate the

418 effects of interactions among predictors.

419 Linear model without interaction terms are used to infer the general effects of wind conditions on coarse dust fraction. These linear

420 models are not intended to imply strictly linear relationships between dust PSD and wind conditions, but rather to provide initial

421 guidance on the directionality of these relationships. Although individual data points present deviations, our models effectively

422 predict the overall trend, as suggested by the response vs. fitted value plots (Fig. S4), where the predicted values align closely with

423 observed values and cluster around the one-to-one red line. Adjusted GVIF values for the model without regression were

424 consistently below 2 with most values close to 1 (Table S2), indicating that multicollinearity among continuous or categorical

425 predictors does not significantly affect the regression model. The multiple linear regression model for the dust coarse fraction

426 includes four independent variables for wind conditions (speed and three options for slope) and additional factors that may affect

427 the PSD of dust emissions~~.~~. ~~It~~allow~~s~~ing us to investigate the effects of topographic wind conditions while controlling interferences

428 from other environmental factors. The model can be expressed as:

430 $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \epsilon, \tag{3}$$

432 where, y represents the coarse fraction of dust emissions, x_1 represents wind speed, and x_2, x_3, and x_4 represent slope under uphill,

433 tangential, and downhill winds, respectively~~:~~. ~~The~~ x_5 is the categorical variable of time of day, including three levels of morning

434 (6:00–12:00 local time), afternoon (12:00–18:00 local time), and evening (18:00–6:00 local time)~~:~~. ~~The~~ x_6 is the categorical

435 variable of season, comprising DJF (winter months of December, January, and February), MAM (spring months of March, April,

436 and May), JJA (summer months of June, July, and August), and SON (autumn months of September, October, and November)~~:~~.

437 ~~The~~ x_7 and x_8 represent the continuous variables of year and soil moisture; and~~. The~~ x_9 is the categorical variable of soil texture

438 class, which contains eight levels of the FAO soil texture classes (Jahn et al., 2006). The ~~C~~coefficients β_i ~~are the slopes of the~~

439 ~~linear relationships between variable~~ x_i ~~and~~ y ~~for continous variables~~represent the expected changes in the response variable y per

440 unit increase in the continuous predictor x_i, and ~~are the changes in the intercept when switching from the reference level to another~~

441 ~~level for categorial variables.~~ the difference in y relative to the reference category for categorical predictor x_i, while holding all

442 other variables constant. The ϵ represents the residuals of the model. The default coarse fraction is cf2 (defined in Section 2.1),

443 and t

444 ~~The~~ corresponding estimated values, standard errors, and significance of the wind condition coefficients are presented in Table 2,

445 ~~and the complete information for~~ with full details of all coefficients ~~can be found~~ in Table S3~~1~~. As noted in Section 2.1, we also

446 tested two alternative definitions of coarse fraction (cf1 and cf3) and compared the estimated coefficients with their statistical

447 significance in Table S4. The coefficient estimates based on cf2 and cf3 are largely consistent, whereas those based on cf1 show

some distinct patterns. Given that dust in the top bin (12-20 μm) falls into the "super coarse" dust category (> 10 μm; Meng et al., 2022), these results suggest that super coarse dust responds differently to varying wind conditions compared to other coarse dust. Therefore, the cf2 definition serves as a robust representation of coarse dust particles.

The coefficient for wind speed from the linear model is statistically significant and positive (Table 2), suggesting that controlling~~excluding~~ the effects of other factors, the coarse fraction of dust emissions increases with the wind speed. This result is consistent with the SHAP analyses on both machine learning models as showcased in the SHAP summary plot (Fig. S5), where purple data points with high wind speed are concentrated on the right side of the vertical centerline and yellow points with lower wind speed dominates the left side, indicating a positive correlation between wind speed and coarse fraction. The observed trend contradicts implications of the theory based on the saltation-bombardment emission mechanism, which ~~states~~ predicts that ~~the~~ higher kinetic energy of impact particles ~~under higher~~ from greater wind speed can intensify the disintegration of soil aggregates and thus the release of finer particles (Shao, 2001; Alfaro et al., 1997). Conversely, our result aligns with the observed shift in the dominant emission mechanism from "shaking-off" of submicron particles to the generation of coarser microparticles from fragmentation as the velocity of saltating particles increases (Malinovskaya et al., 2021) ~~agrees with the observed trend~~. An alternative ~~possibility~~ explanation for the observed positive ~~correlation~~ effect of wind speed on dust size at emission is related to soil conditions (Ishizuka et al., 2008; Panebianco et al., 2023). Emissions of super coarse dust (> 10 μm) increased with wind speed while the emissions of ultrafine dust (< 1 μm) remained nearly invariant over sandy soil (Panebianco et al., 2023); in addition, the fraction of fine dust (< 2 μm) decreased with friction velocity on slightly crusted surfaces (Ishizuka et al., 2008). ~~Both~~ These phenomena were likely due to weaker cohesive forces and thus easier emission for coarse particles than fine particles. Though our sensitivity test using cf1 rejects the increased emission of super coarse dust with wind speed (Table S4), unmeasured changes in the fine dust emissions leading to an overall higher coarse fraction remain one possibility. ~~Although s~~Soil texture and soil moisture were ~~controlled~~ included in the model, ~~but~~ subtle discrepancies ~~among~~across events with~~in~~ the same soil class or soil moisture are not eliminated. Last but not least, since the fresh dust emissions at regional scale inevitably include transported dust, ~~y~~Yet another potential explanation is ~~that~~ unrelated to the emission, but to the transport process—as wind speed increases, more fresh emissions are generated, which undergo less deposition and contain a higher fraction of coarse particles than the aged background dust (González-Flórez et al., 2023).

Table 2. Estimates, standard errors, and p-values of wind condition coefficients for the multiple linear model of dust coarse fraction. The model includes the independent variables of wind conditions (i.e., wind speed and slope under three wind direction types), time of day, season, year, soil moisture, and soil texture. The symbols of coefficients are defined in Eq. (3).

Coefficients for Variables~~VariablesTerms and coefficients~~	Estimates	Standard errors	p-values
wind speed (β_1)	0.0075	0.0002	<0.0001
slope with uphill winds (β_2)	0.0175	0.0013	<0.0001
slope with tangential winds (β_3)	0.0081	0.0015	<0.0001
slope with downhill winds (β_4)	0.0076	0.0016	<0.0001

The coefficients for slopes under all three wind direction~~s~~ ~~types~~ from the linear regression are ~~all~~ significant and positive (Table 2), suggesting that ~~while holding the conditions of other factors constant,~~ the coarse fraction of dust emissions increases with the slope regardless of the relative wind direction. The largest coefficient for uphill winds ~~Between the~~ among all three slopes~~,~~ ~~however, uphill winds~~ indicates that ~~show the~~it has the strongest effect on dust coarse fraction. Similarly, the SHAP analysis on the optimized Random Forest model (Fig. S5) suggests a positive relationship between slope and coarse fraction, with uphill winds

further accentuate the positive effects. The XGBoost model indicates mixed effects of wind directions but the results are less reliable due to its lower prediction accuracy. Overall, the linear model and the Random Forest model agree on the positive correlation between slope, espacially under uphill winds, and dust size. The strong increase in coarse fraction with uphill slope aligns with previous findings using large eddy simulations, and which was explained by the enhancement in vertical transport of dust particles , which isbeing more prominent for coarser particles (Heisel et al., 2021). On the contraryIn contrast, the microphysics of dust emission proposed that compared to tangential winds, uphill winds against the slope resulted in more detachment of fine particles from the surface of soil aggregates on the windward slope due to the secondary aeolian structures, and meanwhile less ejection of coarse particles from the fragmentation of soil aggregates upon hitting the leeward slope (Malinovskaya et al., 2021). The overall trend of coarse fraction with slope under uphill windsOur results suggests that at the regional scale of our analysis, the effect of near-source transport of emitted dust at scales of hundreds to thousands of meters is more dominant than dominates over the impact of microphysics of dust emission related to secondary dune structure at scales of centimeter to meters. The overall elevated coarse fraction with slopes might be attributed to the orographic wind channelling (Rosenberg et al., 2014), increased availability of coarser particles on hills (Samuel-Rosa et al., 2013; Washington et al., 2006) and their greater mobility under gravity. Effects of slope under tangential and downhill winds are less pronounced, with linear model suggesting their weaker positive relationships with coarse fraction compared to uphill winds (Table 2), and the Random Forest model indicating negative impacts of tangential and downhill wind directions on coarse fraction even though the effect of slope is positive (Fig. S5). These weaker effects can possibly be explained by the lack of effective enhancement in vertical transport of coarse particles on the windward side of slopes. On downhill slopes, increasing coarse fraction of dust emission with slope may be related to the acceleration of saltating particles under gravity and strong impaction effective in break down soil aggregates into coarse particles. The positive correlation between slope under tangential winds and coarse fraction may be explained by the increasing fraction of sand with slope (Samuel-Rosa et al., 2013) or the increased mobility of coarse particles by gravity on hills.

3.4 Associations between wind conditions and dust coarse fraction under varying environmental conditions

Interaction terms in the model allow us to investigate how the relationships between wind conditions and dust coarse fraction may vary depending on the additional variables for time and surface characteristics. The model including all significant interactions is shown in Eq. (4). Results for coefficients related to the interactions are shown in Table 3 and the complete results are included in Table S52.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{15} x_1 x_5 + \beta_{16} x_1 x_6 + \beta_{18} x_1 x_8 + \beta_{25} x_2 x_5 +$$
$$\beta_{26} x_2 x_6 + \beta_{28} x_2 x_8 + \beta_{35} x_3 x_5 + \beta_{38} x_3 x_8 + \beta_{46} x_4 x_6 + \epsilon, \tag{4}$$

where, $x_1 - x_9$, y, and $\beta_1 - \beta_9$ are the same as in Eq. (3). The β_{ij} are coefficients for interactions between x_i ($1 \leq i \leq 4$) and x_j ($5 \leq j \leq 9$) and their interpretations are described in Section 2.4.

With interactions in the model, the coefficient β1 indicates the slope of linear correlation between wind speed and the coarse fraction when the variables that have interactions with wind speed are at the reference levels for categorical variables (i.e., "afternoon" for time of day and "DJF" for season) or at zero for continuous variable (i.e., soil moisture). Adjustments of the correlation under other conditions are indicated by the coefficients for interactions with wind speed (β15, β16, and β18). The overall coefficient for wind speed stays positive with varying time of day and season, except for the rare cases when soil moisture is high

520 (> 50%), which agrees with the results from the model without interactions (Table 2). As suggested by the adjustments of
521 coefficients, the positive correlation between wind speed and coarse fraction is weakened during events that happen in the
522 afternoon, in summer, or are associated with higher soil moisture. All these conditions are typical for haboob dust storms which
523 are capable of generating intense dust emissions (Heinold et al., 2013; Knippertz, 2017). Therefore, a potential explanation for the
524 observed patterns is that the dust PSD dependency on wind speed is reduced during convective conditions associated with haboobs.
525 Reasons behind the weakened correlation could be related to turbulent atmospheric conditions. Based on the above assumption
526 that the coarse fraction increases with wind speed due to the associated higher proportion of fresh emissions, during convective
527 events, the role of turbulent flows in keeping dust particles suspended regardless of the magnitude of wind speed may blur the
528 effects of wind speed. Moreover, the positive relationship between wind speed and coarse fraction diminishes with soil moisture.
529 Higher soil moisture may inhibit fresh dust emissions, which, as discussed in Section 3.3, potentially contribute to the positive
530 correlation between wind speed and coarse dust fraction, thereby weakenings the correlation.

531

532 Table 3. Estimates, standard errors, and p-values of the interaction coefficients for the multiple linear model of dust coarse fraction.
533 The model includes the independent variables of wind conditions (i.e., wind speed and slope under three wind direction types),
534 time of day, season, year, soil moisture, and soil texture, as well as significant interaction terms between wind conditions and other
535 independent variables. The interaction coefficients represent wind conditions (speed and direction) under various situations of time
536 of day, season, and soil moisture. The symbols of coefficients are defined in Eq. (3) and (4). Statistically significant (at 0.05
537 significance level) coefficients are bolded and their p-values are marked with "*", among which the negative coefficients are
538 italicized. The significantly positive, significantly negative, and insignificant coefficients at a confidence level of 95% (p value >
539 0.05) are marked with orange, blue, and grey backgrounds, respectively.

Multiple linear model coefficients for wind speed under various conditions		
	Estimates	Standard errors	p-values
Afternoon, DJF, and soil moisture of 0 (reference levels; β_1)	**0.0076**	**0.0007**	**<0.0001***
Adjustments with time of day (β_{15})			
evening	**0.0122**	**0.0006**	**<0.0001***
morning	**0.0016**	**0.0006**	**0.0058***
Adjustments with season (β_{16})			
JJA	*-0.0028*	*0.0007*	*<0.0001**
MAM	*-0.0023*	*0.0006*	*0.0003**
SON	-0.0003	0.0007	0.6500
Adjustments with soil moisture (β_{18})	*-0.0154*	*0.0029*	*<0.0001**
Multiple linear model coefficients for slope with uphill winds under various conditions		
	Estimates	Standard errors	p-values
Afternoon, DJF, and soil moisture of 0 (reference levels; β_2)	**0.0135**	**0.0030**	**<0.0001***
Adjustments with time of day (β_{25})			
evening	**0.0061**	**0.0024**	**0.0118***
morning	**0.0159**	**0.0026**	**<0.0001***
Adjustments with season (β_{26})			
JJA	*-0.0098*	*0.0028*	*0.0005**
MAM	*-0.0138*	*0.0029*	*<0.0001**
SON	-0.0056	0.0031	0.0672
Adjustments with soil moisture (β_{28})	0.0521	0.0107	<0.0001
Multiple linear model coefficients for slope with tangential winds under various conditions		
	Estimates	Standard errors	p-values
Afternoon, soil moisture of 0 (reference levels; β_3)	-0.0038	0.0025	0.1261
Adjustments with time of day (β_{35})			
evening	**0.0134**	**0.0024**	**<0.0001***
morning	**0.0110**	**0.0027**	**<0.0001***

17

	Estimates	Standard errors	p-values
Adjustments with soil moisture (β_{38})	**0.0351**	**0.0115**	**0.0022***
Multiple linear model coefficients for slope with downhill winds under various conditions			
	Estimates	Standard errors	p-values
DJF (reference level; β_4)	**0.0148**	**0.0026**	**<0.0001***
Adjustments with season (β_{46})			
JJA	*-0.0101*	*0.0031*	*0.0011**
MAM	*-0.0105*	*0.0032*	*0.0011**
SON	*-0.0090*	*0.0036*	*0.0116**

540

541 The adjustments of the relationship between slope and coarse fraction with other environmental variables are generally consistent
542 across three wind direction types. Overall, the coefficients for slope with three wind directions stay positive under most
543 circumstances, aligning with results from the model without interactions (Table 2). Notably, the effect of the uphill slope is
544 strongest in the morning. Unlike Harmattan surges, which can induce dust emission throughout the day, or haboob storms, which
545 mostly occur in the afternoons, dust uplift due to the breakdown of night-time low-level jets (NLLJs) is limited to the period around
546 sunrise to midday (Fiedler et al., 2015; Heinold et al., 2013). Therefore, the result indicates that the role of uphill slope in facilitating
547 transport of coarse dust is particularly relevant during emissions related to NLLJs. Moreover, the effect of slope in increasing
548 coarse fraction is weakest during afternoon events under both uphill and tangential winds. With both uphill and downhill winds,
549 the positive correlation between slope and coarse fraction is the strongest in winter, and the weakest in spring and summer. The
550 reduced correlation of dust PSD with uphill slope in both afternoon and summer suggests a diminished effect of slope during
551 haboob dust storms. This can be explained by the stronger turbulence associated with convective storms, which readily stirs up the
552 air and facilitate particles transport, thereby weakening the additional effect of uphill slope by elevating coarse particles through
553 flow separation.

554 The effect of slope on increasing coarse fraction of dust also becomes more apparent with increasing soil moisture under uphill
555 and tangential winds. Low soil moisture might be associated with low water vapor content in lower-Saharan Air Layer, which can
556 lead to continued vertical motions of the atmosphere into the night due to increased atmospheric longwave heating (Ryder, 2021).
557 Conversely, the air is more stable with higher relative humidity, making the enhancement by topography more critical for the
558 transport of coarse dust.

559

4 Conclusion

561 This study aims to explore the relationship between topographic wind conditions and particle size distribution (PSD) of dust
562 emissions on a regional scale through data analysis. The Multiscale Online Nonhydrostatic AtmospheRe Chemistry model
563 (MONARCH) dust concentrations were first evaluated against flight measurements of fresh dust emissions from the 2011 Fennec
564 campaign and were proven to be effective in capturing concentrations of coarse to super coarse dusts in fresh dust emissions. For
565 our analysis, size-resolved surface dust mass concentrations from the MONARCH dust reanalysis over the Sahara during 2007–
566 2016 were condensed into an index of coarse fraction (the ratio of the sum of concentrations in the top two bins (6–20 μm) to the
567 total concentration in eight bins (0.2–20 μm)), serving as the proxy for size distribution. Information on wind vectors and soil
568 moisture, elevation, and soil texture was obtained from the Modern-Era Retrospective analysis for Research and Applications
569 (MERRA-2) reanalysis data, the NASA Shuttle Radar Topography Mission Global 3 arc-second (SRTM GL3) dataset, and the
570 inputs to the Global Land Data Assimilation System version 2 (GLDAS2) Noah land surface model, respectively. Several highly

selective criteria were applied to maximize the probability of selecting fresh dust emissions with typical wind conditions over topography. Scatter plots of coarse fraction against four wind conditions variables (i.e., wind speed, slope under uphill, tangential, and downhill winds) reveal unevenly distributed variance without obvious non-linear trends. We ~~finally chose~~ ultimately selected the multiple linear models after testing several model variations to quantify and explain the trends in data, with key findings cross-validated using machine learning models.

The linear model without and with significant interaction terms can explain 22% and 24% of the variability of coarse fraction, respectively. The model, however, fails to fulfil~~l~~ the assumptions on homoscedasticity (constant variance) and normality, and this issue could not be resolved by other parametric modelling approaches including linear regression models with transformations of variables, weighted least square, or Beta regression models with several options for link functions. The strong intrinsic pattern of non-constant variance and the abundance of data points require more advanced models, which is beyond the scope of our current work. Other uncertainties arose~~came~~ from the varying original resolution of datasets and the resampling process. Moreover, even though we applied multiple criteria to exclusively pick fresh dust emissions, we cannot totally exclude the portion of transported dust. The analysis focuses on the~~is~~ general trend for North Africa, and more detailed insights rely on analysis for smaller geographic domains.

Despite ~~the~~ some limitations, the multiple linear models achieved high predictive accuracy—over 94% under 10-fold cross-validation (CV)—demonstrating their capability to provide meaningful insights for interpretation. The optimized Random Forest model, which attained 92% predictive accuracy with 10-fold CV despite potential overfitting, adding insights to the influence of each predictor by applying the SHapley Additive exPlanations (SHAP) analysis. Both the linear and machine learning models consistently reveal positive ~~correlations~~ associations between coarse dust fraction and both wind speed and slope~~and slope under all relative wind directions, while controlling for other factors~~. The impact of wind speed ~~The increase in coarse fraction with wind speed~~ may possibly be attributed to a higher proportion of fresh emissions, or ~~the~~ a greater resistance of surface soil~~'s resistance~~ to disintegration into fine dust particles. The persistent influence of slope on increasing coarse fraction might be related to meteorological (e.g., the orographic wind channelling) and soil conditions (e.g., increased availability and gravatational mobility of coarser particles on hills). The effect of slope is most pronounced under uphill winds, highlighting the ~~The observed effect of uphill slope aligns with~~ transport enhancement by topography over distances of hundreds to thousands of meters, which overrides the potential impact of local emission mechanisms at scales of centimeters to meters. Including interaction terms in the model allows us to investigate shifts in the effects of wind conditions on dust size under different environmental conditions. The positive correlation between wind speed and coarse fraction diminishes during afternoon and summer events and under high soil moisture. This reduction is likely due to decreased differences in dust size distribution by deposition during haboob convective storms when turbulence is strong. The uphill slope exhibits the strongest effect on increasing dust size in the morning, suggesting that the enhanced vertical transport is particularly effective in uplifting coarse dust during emissions related to the breakdown of night-time low-level jets. The effect of uphill slope is weakened during summer and afternoons, indicating that turbulence during haboob dust storms has competing effects in sustaining airborne coarse dust.

This work provides insights into the controlling factors of dust PSD on a regional scale using a meta-analysis of a 10-year dust reanalysis dataset, which complements the accumulating knowledge from recent field measurements. The study highlights the overlapping effects and interactions among various environmental factors on size distribution of dust emissions and calls for more work focusing on different factors and their interactions at various scales. Meanwhile, the workflow for screening fresh dust events developed in this study will facilitate future work of utilizing datasets at different scales.

Data Availability

The open-access MONARCH dust reanalysis data, prepared by the Barcelona Supercomputing Center (BSC), are available at https://earth.bsc.es/thredds_dustclim/homepage/. The Modern-Era Retrospective analysis for Research and Applications (MERRA-2) data (Gelaro et al., 2017) are managed by the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) and can be accessed at https://disc.gsfc.nasa.gov/datasets?project=MERRA-2. The NASA Shuttle Radar Topography Mission Global 3 arc-second (SRTM GL3) dataset is available at https://lpdaac.usgs.gov/products/srtmgl3v003/. The soil texture map input to the Global Land Data Assimilation System version 2 (GLDAS2) Noah land surface model (Rodell et al., 2004) is described and provided by NASA at https://ldas.gsfc.nasa.gov/gldas/soils.

Competing Interests

The authors declare that they have no conflict of interests.

Author Contribution

HF secured funding for the research. HF and XH conceptualized the study, developed methodology for data collection and processing, and provided the scientific interpretation of the results. XH processed the data and created the data visualizations. WG conducted the statistical analysis of the processed data, drafted the corresponding methods (Section 2.4), and verified the initial interpretation based on the statistical results. XH wrote the initial draft of the manuscript, and all authors contributed to the final version.

References

Adebiyi, A. A. and Kok, J. F.: Climate models miss most of the coarse dust in the atmosphere, Science Advances, 6, doi:10.1126/sciadv.aaz9507, 2020.

Alfaro, S. C., Bouet, C., Khalfallah, B., Shao, Y., Ishizuka, M., Labiadh, M., Marticorena, B., Laurent, B., and Rajot, J. L.: Unraveling the Roles of Saltation Bombardment and Atmospheric Instability on Magnitude and Size Distribution of Dust Emission Fluxes: Lessons From the JADE and WIND-O-V Experiments, Journal of Geophysical Research: Atmospheres, 127, doi:10.1029/2021jd035983, 2022.

20

Bacanin, N., Perisic, M., Jovanovic, G., Damaševičius, R., Stanisic, S., Simic, V., Zivkovic, M., and Stojic, A.: The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere, Science of The Total Environment, 929, 172195, doi:10.1016/j.scitotenv.2024.172195, 2024.

Berkson, J.: Application of the Logistic Function to Bio-Assay, Journal of the American Statistical Association, 39, 357–365, doi:10.2307/2280041, 1944.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, Journal of the Royal Statistical Society Series B: Statistical Methodology, 26, 211–243, doi:10.1111/j.2517-6161.1964.tb00553.x, 1964.

~~Bozdogan, H.: Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, Psychometrika, 52, 345–370, doi:10.1007/BF02294361, 1987.~~

Brasseur, G. P. and Jacob, D. J.: Modeling of Atmospheric Chemistry, Cambridge University Press, Cambridge, doi:10.1017/9781316544754, 2017.

Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., and Ryan, P.: Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches, Atmospheric Environment, 151, 1–11, doi:10.1016/j.atmosenv.2016.11.066, 2017.

Caton Harrison, T., Washington, R., Engelstaedter, S., Jones, R. G., and Savage, N. H.: Influence of Orography Upon Summertime Low-Level Jet Dust Emission in the Central and Western Sahara, Journal of Geophysical Research: Atmospheres, 126, doi:10.1029/2021jd035025, 2021.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, doi:10.1145/2939672.2939785, 2016.

Chkhetiani, O. G., Vazaeva, N. V., Chernokulsky, A. V., Shukurov, K. A., Gubanova, D. P., Artamonova, M. S., Maksimenkov, L. O., Kozlov, F. A., and Kuderina, T. M.: Analysis of Mineral Aerosol in the Surface Layer over the Caspian Lowland Desert by the Data of 12 Summer Field Campaigns in 2002–2020, Atmosphere, 12, doi:10.3390/atmos12080985, 2021.

Cribari-Neto, F. and Zeileis, A.: Beta Regression in R, Journal of Statistical Software, 34, 1–24, doi:10.18637/jss.v034.i02, 2010.

Dee, D. P., Uppala, S., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., and Bauer, d. P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Di Tomaso, E., Escribano, J., Basart, S., Ginoux, P., Macchia, F., Barnaba, F., Benincasa, F., Bretonnière, P. A., Buñuel, A., Castrillo, M., Cuevas, E., Formenti, P., Gonçalves, M., Jorba, O., Klose, M., Mona, L., Montané Pinto, G., Mytilinaios, M., Obiso, V., Olid, M., Schutgens, N., Votsis, A., Werner, E., and Pérez García-Pando, C.: The MONARCH high-resolution reanalysis of desert dust aerosol over Northern Africa, the Middle East and Europe (2007–2016), Earth Syst. Sci. Data, 14, 2785–2816, doi:10.5194/essd-14-2785-2022, 2022.

Douma, J. C. and Weedon, J. T.: Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression, Methods in Ecology and Evolution, 10, 1412–1430, doi:10.1111/2041-210x.13234, 2019.

Duce, R. A. and Tindale, N. W.: Atmospheric transport of iron and its deposition in the ocean, Limnology and Oceanography, 36, 1715–1726, doi:10.4319/lo.1991.36.8.1715, 1991.

Dupont, S.: On the Influence of Thermal Stratification on Emitted Dust Flux, Journal of Geophysical Research: Atmospheres, 127, doi:10.1029/2022jd037364, 2022.

Dupont, S., Alfaro, S. C., Bergametti, G., and Marticorena, B.: Near-surface dust flux enrichment in small particles during erosion events, Geophysical Research Letters, 42, 1992–2000, doi:10.1002/2015gl063116, 2015.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Reviews of Geophysics, 45, doi:10.1029/2005rg000183, 2007.

Fernandes, R., Dupont, S., and Lamaud, E.: Experimental evidence of dust flux size distribution variation along two consecutive erosion seasons, Aeolian Research, 61, doi:10.1016/j.aeolia.2023.100863, 2023.

Fiedler, S., Kaplan, M. L., and Knippertz, P.: The importance of Harmattan surges for the emission of North African dust aerosol, Geophysical Research Letters, 42, 9495–9504, doi:10.1002/2015gl065925, 2015.

Formenti, P., Schütz, L., Balkanski, Y., Desboeufs, K., Ebert, M., Kandler, K., Petzold, A., Scheuvens, D., Weinbruch, S., and Zhang, D.: Recent progress in understanding physical and chemical properties of African and Asian mineral dust, Atmospheric Chemistry and Physics, 11, 8231–8256, doi:10.5194/acp-11-8231-2011, 2011.

Fox, J. and Weisberg, S.: An R Companion to Applied Regression, 2019.

Fujiwara, M., Martineau, P., Wright, J. S., Abalos, M., Šácha, P., Kawatani, Y., Davis, S. M., Birner, T., and Monge-Sanz, B. M.: Climatology of the terms and variables of transformed Eulerian-mean (TEM) equations from multiple reanalyses: MERRA-2, JRA-55, ERA-Interim, and CFSR, Atmos. Chem. Phys., 24, 7873–7898, doi:10.5194/acp-24-7873-2024, 2024.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, doi:10.1175/JCLI-D-16-0758.1, 2017.

Gesch, D., Williams, J., and Miller, W.: A comparison of US geological survey seamless elevation models with shuttle radar topography mission data, IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), 754–756, 2001.

Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, Journal of Geophysical Research: Atmospheres, 106, 20255 –20273, doi:10.1029/2000JD000053, 2001.

Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C., and Zhao, M: Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products, Rev. Geophys., 50, RG3005, doi:10.1029/2012RG000388, 2012.

González-Flórez, C., Klose, M., Alastuey, A., Dupont, S., Escribano, J., Etyemezian, V., Gonzalez-Romero, A., Huang, Y., Kandler, K., Nikolich, G., Panta, A., Querol, X., Reche, C., Yus-Díez, J., and Pérez García-Pando, C.: Insights into the size-resolved dust emission from field measurements in the Moroccan Sahara, Atmos. Chem. Phys., 23, 7177–7212, doi:10.5194/acp-23-7177-2023, 2023.

Graesser, J.: GeoWombat: Utilities for geospatial data [code], https://github.com/jgrss/geowombat, 2023.

Grün, B., Kosmidis, I., and Zeileis, A.: Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned, Journal of Statistical Software, 48, 1–25, doi:10.18637/jss.v048.i11, 2012.

Heinold, B., Knippertz, P., Marsham, J. H., Fiedler, S., Dixon, N. S., Schepanski, K., Laurent, B., and Tegen, I.: The role of deep convection and nocturnal low-level jets for dust emission in summertime West Africa: Estimates from convection-permitting simulations, J Geophys Res Atmos, 118, 4385–4400, doi:10.1002/jgrd.50402, 2013.

Heisel, M., Chen, B., Kok, J. F., and Chamecki, M.: Gentle Topography Increases Vertical Transport of Coarse Dust by Orders of Magnitude, Journal of Geophysical Research: Atmospheres, 126, doi:10.1029/2021JD034564, 2021.

Ishizuka, M., Mikami, M., Leys, J., Yamada, Y., Heidenreich, S., Shao, Y., and McTainsh, G. H.: Effects of soil moisture and dried raindroplet crust on saltation and dust emission, Journal of Geophysical Research: Atmospheres, 113, doi:10.1029/2008jd009955, 2008.

Jahn, R., Blume, H. P., Asio, V., Spaargaren, O., and Schád, P.: FAO Guidelines for Soil Description, 2006.

Katra, I.: Soil Erosion by Wind and Dust Emission in Semi-Arid Soils Due to Agricultural Activities, Agronomy, 10, doi:10.3390/agronomy10010089, 2020.

Keller, C. A. and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10, Geosci. Model Dev., 12, 1209–1225, 10.5194/gmd-12-1209-2019, 2019.

Kellogg, C. A. and Griffin, D. W.: Aerobiology and the global transport of desert dust, Trends Ecol Evol, 21, 638–644, doi:10.1016/j.tree.2006.07.004, 2006.

Khalfallah, B., Bouet, C., Labiadh, M. T., Alfaro, S. C., Bergametti, G., Marticorena, B., Lafon, S., Chevaillier, S., Féron, A., Hease, P., Henry des Tureaux, T., Sekrafi, S., Zapf, P., and Rajot, J. L.: Influence of Atmospheric Stability on the Size Distribution of the Vertical Dust Flux Measured in Eroding Conditions Over a Flat Bare Sandy Field, Journal of Geophysical Research: Atmospheres, 125, doi:10.1029/2019jd031185, 2020.

Kiers, H. A. L.: Weighted least squares fitting using ordinary least squares algorithms, Psychometrika, 62, 251–266, 1997.

Klose, M., Jorba, O., Gonçalves Ageitos, M., Escribano, J., Dawson, M. L., Obiso, V., Di Tomaso, E., Basart, S., Montané Pinto, G., Macchia, F., Ginoux, P., Guerschman, J., Prigent, C., Huang, Y., Kok, J. F., Miller, R. L., and Pérez García-Pando, C.: Mineral dust cycle in the Multiscale Online Nonhydrostatic AtmospheRe CHemistry model (MONARCH) Version 2.0, Geosci. Model Dev., 14, 6403–6444, doi:10.5194/gmd-14-6403-2021, 2021.

Knippertz, P.: Mineral Dust Generation across Northern Africa and Its Impacts, in: Oxford Research Encyclopedia of Climate Science, doi:10.1093/acrefore/9780190228620.013.516, 2017.

Knippertz, P., Deutscher, C., Kandler, K., Müller, T., Schulz, O., and Schütz, L.: Dust mobilization due to density currents in the Atlas region: Observations from the Saharan Mineral Dust Experiment 2006 field campaign, Journal of Geophysical Research: Atmospheres, 112, doi:10.1029/2007jd008774, 2007.

Kok, J. F.: Does the size distribution of mineral dust aerosols depend on the wind speed at emission?, Atmospheric Chemistry and Physics, 11, 10149–10156, doi:10.5194/acp-11-10149-2011, 2011.

Kok, J. F., Parteli, E. J. R., Michaels, T. I., and Karam, D. B.: The physics of wind-blown sand and dust, Reports on Progress in Physics, 75, 106901, doi:10.1088/0034-4885/75/10/106901, 2012.

Kok, J. F., Albani, S., Mahowald, N. M., and Ward, D. S.: An improved dust emission model – Part 2: Evaluation in the Community Earth System Model, with implications for the use of dust source functions, Atmos. Chem. Phys., 14, 13043 –13061, doi:10.5194/acp-14-13043-2014, 2014.

Kok, J. F., Storelvmo, T., Karydis, V. A., Adebiyi, A. A., Mahowald, N. M., Evan, A. T., He, C., and Leung, D. M.: Mineral dust aerosol impacts on global climate and climate change, Nature Reviews Earth & Environment, 4, 71–86, doi:10.1038/s43017-022-00379-5, 2023.

Lewis, F., Butler, A., and Gilbert, L.: A unified approach to model selection using the likelihood ratio test, Methods in Ecology and Evolution, 2, 155–162, doi:10.1111/j.2041-210X.2010.00063.x, 2010.

Lu, S., Symonds, P., Verschoor, N., Chalabi, Z., Taylor, J., and Davies, M.: Use of Beta Regression to investigate the link between home air infiltration rate and self-reported health, Journal of Physics: Conference Series, 2069, doi:10.1088/1742-6596/2069/1/012178, 2021.

Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017.

Ma, X., Gao, Q., Jiang, X., Chen, S., Gan, Y., Zhang, T., Lu, X., and Wang, X.: Direct Effects of Air Humidity on Dust Aerosol Production: Evidences for the Surprising Role of Electrostatic Forces, Geophysical Research Letters, 50, doi:10.1029/2023gl103639, 2023.

Mahowald, N., Albani, S., Kok, J. F., Engelstaeder, S., Scanza, R., Ward, D. S., and Flanner, M. G.: The size distribution of desert dust aerosols and its impact on the Earth system, Aeolian Research, 15, 53–71, doi:10.1016/j.aeolia.2013.09.002, 2014.

Malinovskaya, E. A., Chkhetiani, O. G., and Maksimenkov, L. O.: Influence of Wind Direction on the Size Distribution of Aeolian Microparticles, Izvestiya, Atmospheric and Oceanic Physics, 57, 472–485, doi:10.1134/S0001433821050108, 2021.

Marticorena, B. and Bergametti, G.: Modeling the atmospheric dust cycle: 1. Design of a soil-derived dust emission scheme, Journal of geophysical research: atmospheres, 100, 16415 –16430, 1995.

Meng, J., Huang, Y., Leung, D. M., Li, L., Adebiyi, A. A., Ryder, C. L., Mahowald, N. M., and Kok, J. F.: Improved Parameterization for the Size Distribution of Emitted Dust Aerosols Reduces Model Underestimation of Super Coarse Dust, Geophys Res Lett, 49, doi:10.1029/2021GL097287, 2022.

Mytilinaios, M., Basart, S., Ciamprone, S., Cuesta, J., Dema, C., Di Tomaso, E., Formenti, P., Gkikas, A., Jorba, O., Kahn, R., Pérez García-Pando, C., Trippetta, S., and Mona, L.: Comparison of dust optical depth from multi-sensor products and MONARCH (Multiscale Online Non-hydrostatic AtmospheRe CHemistry) dust reanalysis over North Africa, the Middle East, and Europe, Atmos. Chem. Phys., 23, 5487–5516, doi:10.5194/acp-23-5487-2023, 2023.

~~Neath, A. A. and Cavanaugh, J. E.: The Bayesian information criterion: background, derivation, and applications, WIREs Computational Statistics, 4, 199–203, doi:10.1002/wics.199, 2011.~~

Panebianco, J. E., Avecilla, F., Funk, R., and Buschiazzo, D. E.: Influence of friction velocity and saltation intensity on the particle-size distribution of dust (PM32) emitted from an agricultural soil, Earth Surface Processes and Landforms, 48, 1176–1187, doi:10.1002/esp.5541, 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Peleg, M.: Beta distributions for particle size having a finite range and predetermined mode, mean or median, Powder Technology, 356, 790–794, doi:10.1016/j.powtec.2019.09.015, 2019.

Polymenakou, P. N., Mandalakis, M., Stephanou, E. G., and Tselepides, A.: Particle size distribution of airborne microorganisms and pathogens during an Intense African dust event in the eastern Mediterranean, Environ Health Perspect, 116, 292–296, doi:10.1289/ehp.10684, 2008.

Prospero, J. M., Blades, E., Mathison, G., and Naidu, R.: Interhemispheric transport of viable fungi and bacteria from Africa to the Caribbean with soil dust, Aerobiologia, 21, 1–19, doi:10.1007/s10453-004-5872-7, 2005.

Pu, B., and Ginoux, P.: The impact of the Pacific Decadal Oscillation on springtime dust activity in Syria, Atmos. Chem. Phys., 16, 13431–13448, doi:10.5194/acp-16-13431-2016, 2016.

R. Core Team: R: A Language and Environment for Statistical Computing, 2023.

Reynolds, C. A., Jackson, T. J., and Rawls, W. J.: Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions, Water Resources Research, 36, 3653–3662, doi:10.1029/2000wr900130, 2000.

Rienecker, M. M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R., and Gelaro, R.: The GEOS-5 Data Assimilation System-Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0, 2008.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, doi:10.1175/BAMS-85-3-381, 2004.

Rosenberg, P. D., Parker, D. J., Ryder, C. L., Marsham, J. H., Garcia-Carreras, L., Dorsey, J. R., Brooks, I. M., Dean, A. R., Crosier, J., McQuaid, J. B., and Washington, R.: Quantifying particle size and turbulent scale dependence of dust

flux in the Sahara using aircraft measurements, Journal of Geophysical Research: Atmospheres, 119, 7577–7598, doi:10.1002/2013jd021255, 2014.

Ryder, C. L.: Radiative Effects of Increased Water Vapor in the Upper Saharan Air Layer Associated With Enhanced Dustiness, Journal of Geophysical Research: Atmospheres, 126, doi:10.1029/2021jd034696, 2021.

Ryder, C. L., Highwood, E. J., Walser, A., Seibert, P., Philipp, A., and Weinzierl, B.: Coarse and giant particles are ubiquitous in Saharan dust export regions and are radiatively significant over the Sahara, Atmospheric Chemistry and Physics, 19, 15353–15376, doi:10.5194/acp-19-15353-2019, 2019.

Ryder, C. L., Highwood, E. J., Rosenberg, P. D., Trembath, J., Brooke, J. K., Bart, M., Dean, A., Crosier, J., Dorsey, J., Brindley, H., Banks, J., Marsham, J. H., McQuaid, J. B., Sodemann, H., and Washington, R.: Optical properties of Saharan dust aerosol and contribution from the coarse mode as measured during the Fennec 2011 aircraft campaign, Atmospheric Chemistry and Physics, 13, 303–325, doi:10.5194/acp-13-303-2013, 2013.

Ryder, C. L., McQuaid, J. B., Flamant, C., Rosenberg, P. D., Washington, R., Brindley, H. E., Highwood, E. J., Marsham, J. H., Parker, D. J., Todd, M. C., Banks, J. R., Brooke, J. K., Engelstaedter, S., Estelles, V., Formenti, P., Garcia-Carreras, L., Kocha, C., Marenco, F., Sodemann, H., Allen, C. J. T., Bourdon, A., Bart, M., Cavazos-Guerra, C., Chevaillier, S., Crosier, J., Darbyshire, E., Dean, A. R., Dorsey, J. R., Kent, J., O'Sullivan, D., Schepanski, K., Szpek, K., Trembath, J., and Woolley, A.: Advances in understanding mineral dust and boundary layer processes over the Sahara from Fennec aircraft observations, Atmospheric Chemistry and Physics, 15, 8479–8520, doi:10.5194/acp-15-8479-2015, 2015.

Samuel-Rosa, A., Dalmolin, R. S. D., and Miguel, P.: Building predictive models of soil particle-size distribution, Revista Brasileira de Ciência do Solo, 37, 422–430, doi:10.1590/s0100-06832013000200013, 2013.

Santos, J., Sakagami, Y., Haas, R., Passos, J., Machuca, M., Radünz, W., Dias, E., and Lima, M.: Wind speed evaluation of MERRA-2, ERA-interim and ERA-5 reanalysis data at a wind farm located in brazil, Proceedings of the ISES Solar World Congress, 1 –10, 2019.

Shao, Y.: A model for mineral dust emission, Journal of Geophysical Research: Atmospheres, 106, 20239–20254, doi:10.1029/2001jd900171, 2001.

Shao, Y., Zhang, J., Ishizuka, M., Mikami, M., Leys, J., and Huang, N.: Dependency of particle size distribution at dust emission on friction velocity and atmospheric boundary-layer stability, Atmospheric Chemistry and Physics, 20, 12939–12953, doi:10.5194/acp-20-12939-2020, 2020.

Tanaka, T. Y. and Chiba, M.: A numerical study of the contributions of dust source regions to the global dust budget, Global and Planetary Change, 52, 88–104, doi:10.1016/j.gloplacha.2006.02.002, 2006.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, 2002.

Wang, R., Li, Q., Wang, R., Chang, C., Guo, Z., Li, J., and Zhou, N.: Influence of Wind Velocity and Soil Size Distribution on Emitted Dust Size Distribution: A Wind Tunnel Study, Journal of Geophysical Research: Atmospheres, 126, doi:10.1029/2020jd033768, 2021.

Wang, X., Wang, Q., Prass, M., Pöhlker, C., Moran-Zuloaga, D., Artaxo, P., Gu, J., Yang, N., Yang, X., Tao, J., Hong, J., Ma, N., Cheng, Y., Su, H., and Andreae, M. O.: The export of African mineral dust across the Atlantic and its impact

over the Amazon Basin, Atmospheric Chemistry and Physics, 23, 9993–10014, doi:10.5194/acp-23-9993-2023, 2023.

Washington, R., Todd, M. C., Lizcano, G., Tegen, I., Flamant, C., Koren, I., Ginoux, P., Engelstaedter, S., Bristow, C. S., Zender, C. S., Goudie, A. S., Warren, A., and Prospero, J. M.: Links between topography, wind, deflation, lakes and dust: The case of the Bodélé Depression, Chad, Geophysical Research Letters, 33, doi:10.1029/2006gl025827, 2006.

Webb, N. P., LeGrand, S. L., Cooper, B. F., Courtright, E. M., Edwards, B. L., Felt, C., Van Zee, J. W., and Ziegler, N. P.: Size Distribution of Mineral Dust Emissions From Sparsely Vegetated and Supply-Limited Dryland Soils, Journal of Geophysical Research: Atmospheres, 126, doi:10.1029/2021jd035478, 2021.

Yamaguchi, N., Ichijo, T., Sakotani, A., Baba, T., and Nasu, M.: Global dispersion of bacterial cells on Asian dust, Sci Rep, 2, 525, doi:10.1038/srep00525, 2012.

Zeileis, A. and Hothorn, T.: Diagnostic Checking in Regression Relationships, R News, 2, 7–10, 2002.

Zhang, B., Zhang, Y., and Jiang, X.: Feature selection for global tropospheric ozone prediction based on the BO-XGBoost-RFE algorithm, Scientific Reports, 12, 9244, doi:10.1038/s41598-022-13498-2, 2022.

Zhuang, J., Dussin, R., Huard, D., Bourgault, P., Banihirwe, A., Raynaud, S., Malevich, B., Schupfner, M., Filipe, Levang, S., Jüling, A., Almansi, M., RichardScottOz, RondeauG, Rasp, S., Smith, T. J., Stachelek, J., Plough, M., Pierre, Bell, R., and Li, X.: pangeo-data/xESMF: v0.7.1 (v0.7.1), Zenodo [code], doi:10.5281/zenodo.7800141, 2023.