Supplement of Atmos. Chem. Phys., 25, 9583–9600, 2025
https://doi.org/10.5194/acp-25-9583-2025-supplement

Atmospheric Chemistry and Physics
Open Access

Supplement of

Impact of topographic wind conditions on dust particle size distribution: insights from a regional dust reanalysis dataset

Xinyue Huang et al.

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

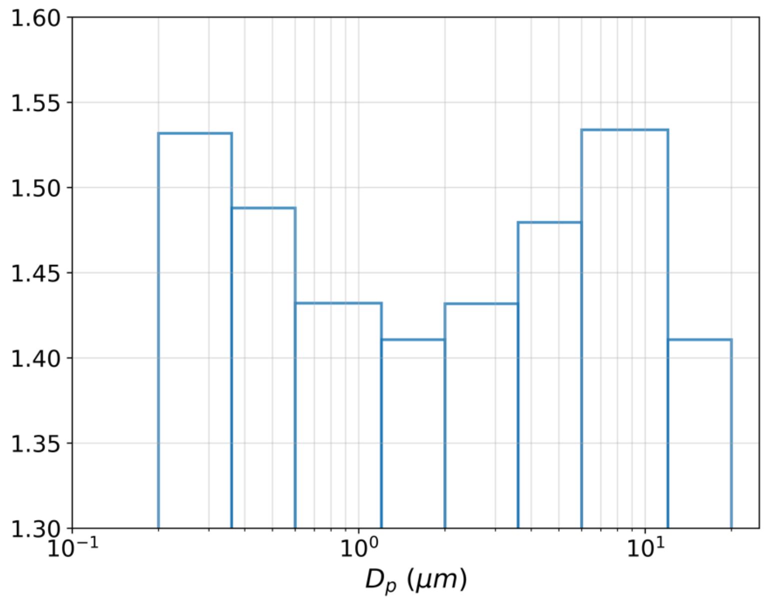

Figure S1. Ratio of first-guess dust concentration to its reanalysis across eight size bins. The average of dust concentrations in grids that contain any portion of selected Fennec segments were used as an example (see Section 2.2 for more details).

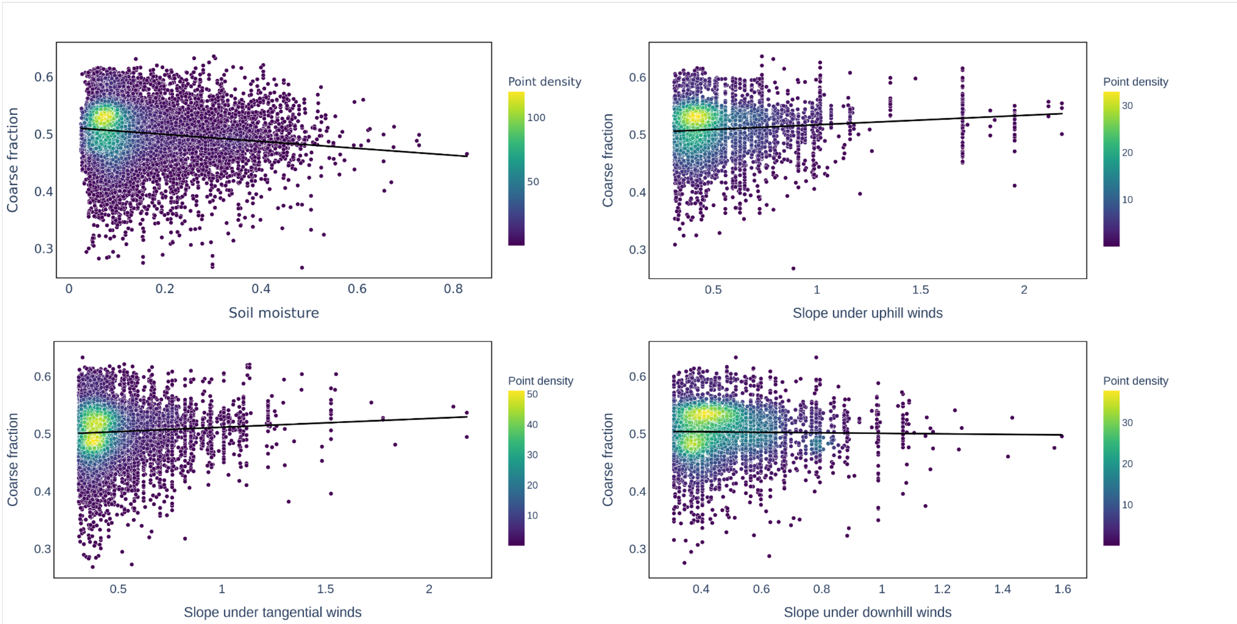

Figure S2. Scatter plots and linear trend lines of relationships between the coarse fraction of surface dust concentration and soil moisture and slope under three different wind directions. The color-codes present the number of overlapping data points.

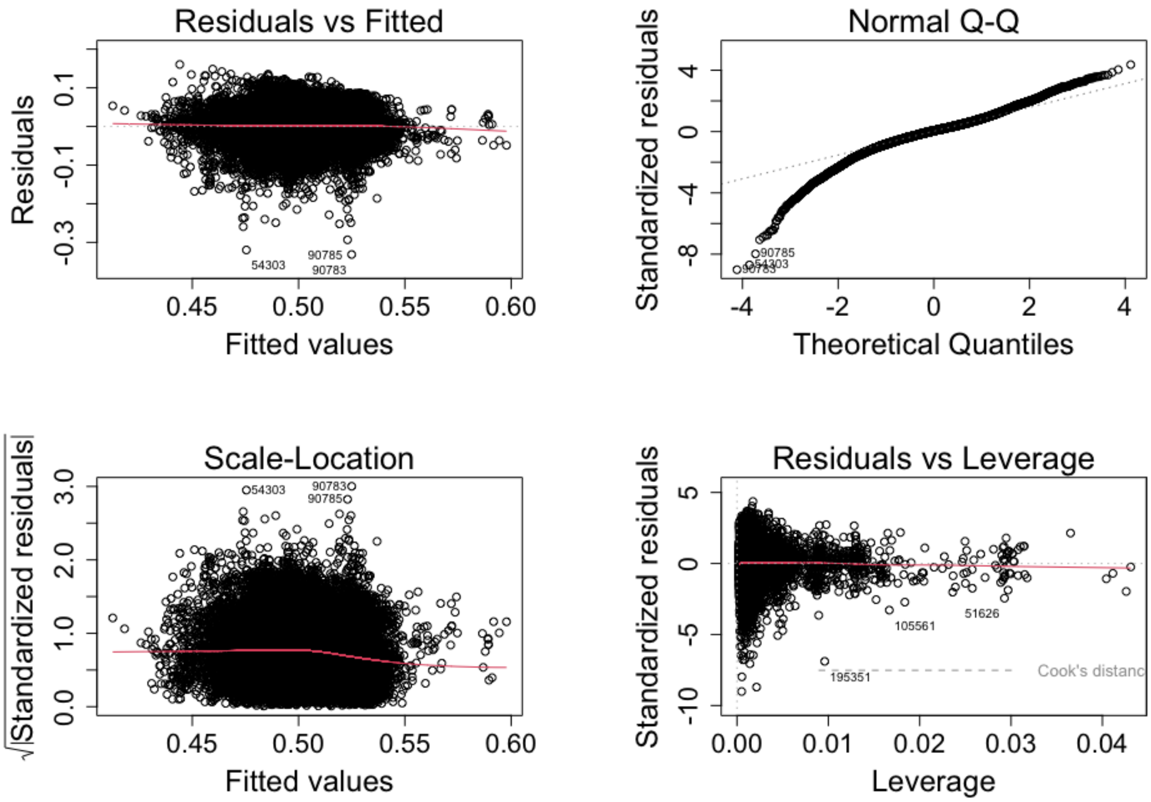

Figure S3. Diagnostic plots for the residuals of the linear model with only significant interactions (Eq. (4)). The "Residual vs Fitted" and "Scale-Location" panels indicate that the residuals exhibit uneven variances. The "Normal Q-Q" panel shows the deviation of data points from the dotted line, indicating that the residuals do not conform to a normal distribution.

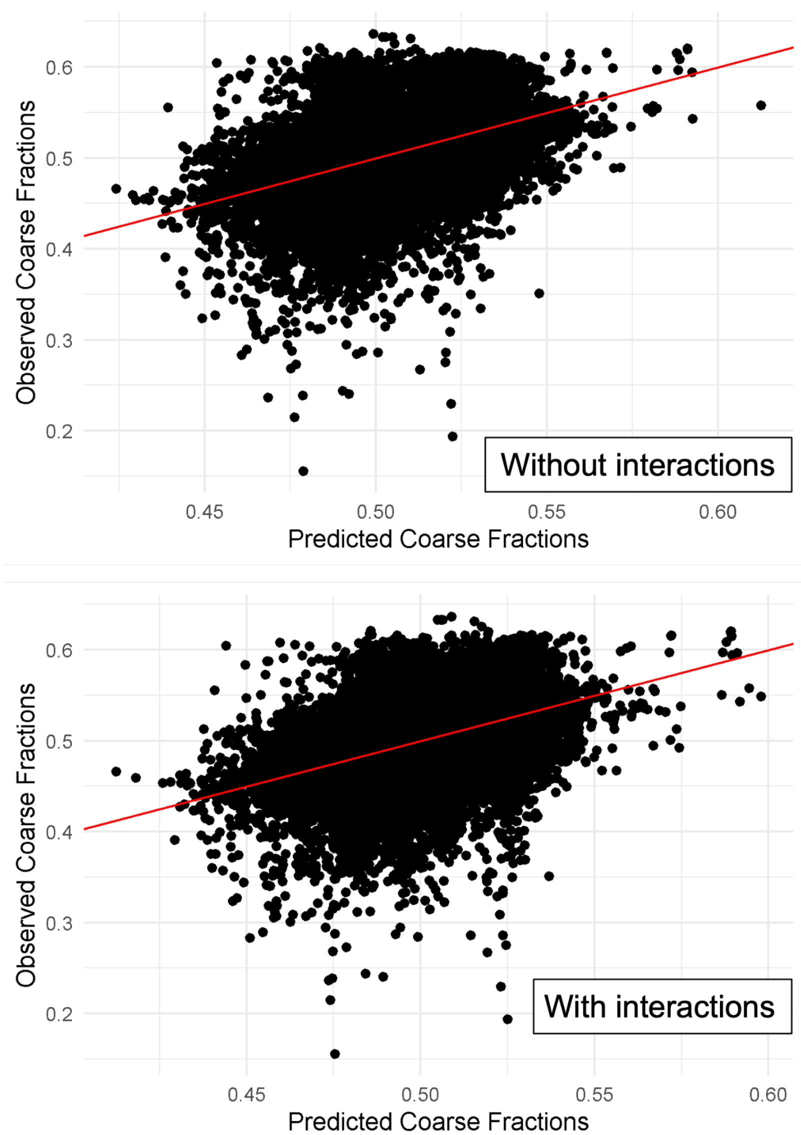

Figure S4. Predicted coarse fractions from the linear model without (Eq. (3)) and with interactions (Eq. (4)) compared with the observed coarse fractions. Despite some deviations for individual points, data points cluster around the red one-to-one line where the predictions equal the observations.

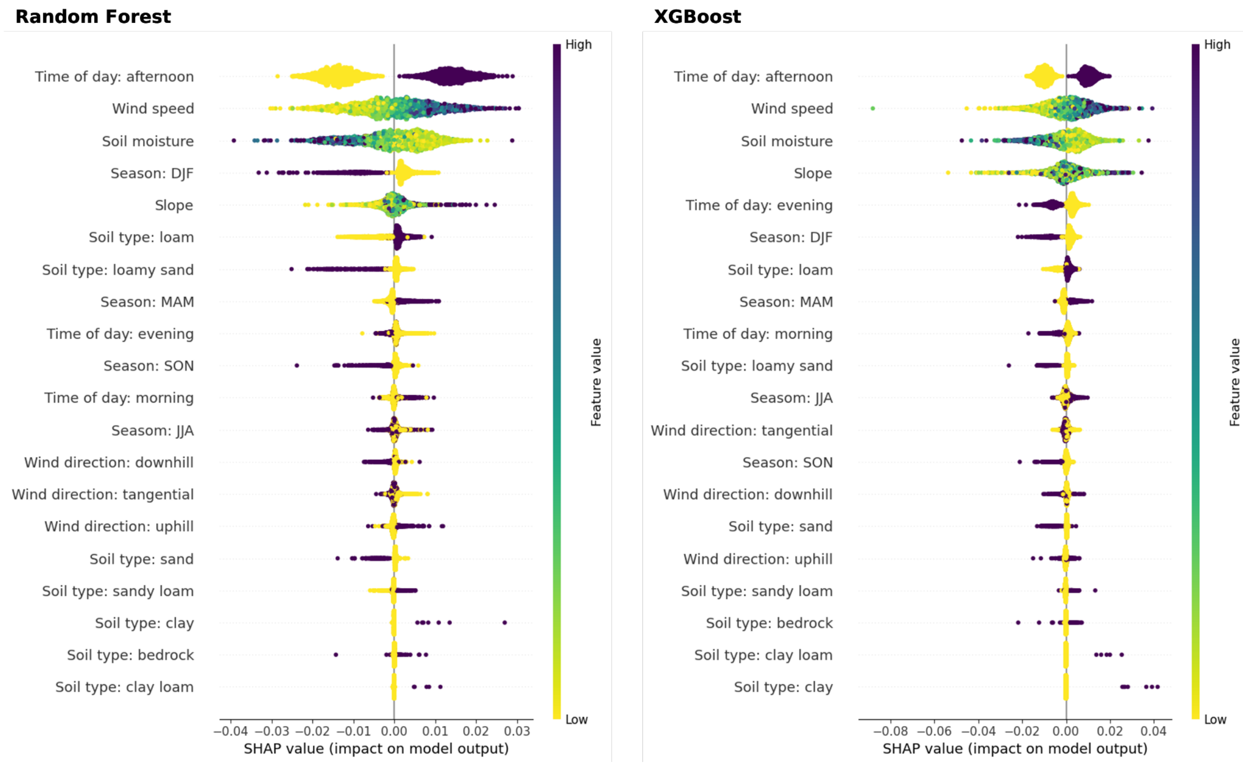

Figure S5. The SHapley Additive exPlanations (SHAP) summary plots (or beeswarm plots) for the optimized Random Forest and Extreme Gradient Boosting (XGBoost) models. Predictors are ranked by descending importance from top to bottom. The SHAP values on the x-axis represent the sign and magnitude of the impact on coarse fraction by each data point. Scatter points are colored by values of each predictor. Thus, dark purple scatter points dominating the right side and light-yellow points dominating the left side of the vertical centerline suggests that the corresponding predictor has an overall positive correlation with the coarse dust fraction, and vice versa.

Table S1. Hyperparameters used to fine-tune the Random Forest and XGBoost models. Bolded hyperparameters were selected for the optimized models.

Models	Hyperparameters	Values			
Random Forest	Number of trees	100	**300**	500	800
	Maximum depth of each tree	10	20	**30**	unconfined
	Minimum number of samples required to split a node	**2**	5	10	
	Minimum number of samples required at a leaf node	**1**	2	4	10
	Features considered for the best split	**square root**	log2	None	
	Whether bootstrap samples are used	**True**	False		
XGBoost	Number of boosting rounds	100	300	500	**800**
	Step size shrinkage to prevent overfitting	0.01	0.05	**0.1**	0.2
	Maximum depth of a tree	3	6	**10**	
	Fraction of samples used for training	0.6	0.8	**1**	
	Fraction of features used per tree	**0.6**	0.8	1	
	Minimum loss reduction required to make a split	**0**	0.1	0.2	0.5
	L1 regularization term (LASSO)	**0**	0.1	1	5
	L2 regularization term (Ridge)	1	**2**	5	10

Table S2. Generalized variance inflation factors (GVIFs) for all predictors in the model without interactions (Eq. (3)). The GIF adjusted for the degree of freedom (Df) values ($GVIF^{1/(2 \cdot Df)}$) being 1 (the minimum) indicates no collinearity and values smaller than 5 typically suggest low and acceptable collinearity.

Variables	**GVIF**	**Df**	$\boldsymbol{GVIF^{1/(2 \cdot Df)}}$
wind speed	1.139	1	1.067
slope with uphill winds	1.998	1	1.413
slope with tangential winds	2.654	1	1.629
slope with downhill winds	2.187	1	1.479
time of day	1.248	2	1.057
season	1.453	3	1.064
year	1.004	1	1.002
soil moisture	1.313	1	1.146
soil texture	1.651	8	1.032

Table S3. Estimates, standard errors, and p-values of all 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). Statistically significant (at 0.05 significance level) coefficients are bolded and their p-values are marked with "*", among which the negative coefficients are italic.

Coefficients for Variables	**Estimates**	**Standard errors**	**p-values**
Intercept (β_0)	0.0753	0.1574	0.6323
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***
time of day ("afternoon" as reference) (β_5)			
evening	***-0.0339***	***0.0006***	***<0.0001****
morning	***-0.0253***	***0.0006***	***<0.0001****
season ("DJF" as reference) (β_6)			
JJA	**0.0139**	**0.0008**	**<0.0001***
MAM	**0.0184**	**0.0007**	**<0.0001***
SON	**0.0124**	**0.0008**	**<0.0001***
year (β_7)	**0.0002**	**0.0001**	**0.0180***
soil moisture (β_8)	***-0.0742***	***0.0030***	***<0.0001****
soil texture ("sand" as reference) (β_9)			
loamy sand	***-0.0064***	***0.0014***	***<0.0001****
sandy loam	**0.0106**	**0.0013**	**<0.0001***
loam	**0.0151**	**0.0011**	**<0.0001***
sandy clay loam	**0.0204**	**0.0025**	**<0.0001***
clay loam	**0.0393**	**0.0036**	**<0.0001***
clay	**0.0841**	**0.0064**	**<0.0001***
organic materials	-0.0019	0.0042	0.6570
bedrock	**0.0179**	**0.0019**	**<0.0001***

Table S4. Coefficient estimates from linear models without interactions using different definitions for coarse fraction. Specifically, the ratios of cumulative dust mass in the coarsest one, two, or three bins to the total dust mass concentration (denoted as cf1- cf3) were applied in the models. Statistically significant (at 0.05 significance level) coefficients are bolded and marked with "*", among which the negative coefficients are italic.

Variables and coefficients	**Estimates (cf1)**	**Estimates (cf2)**	**Estimates (cf3)**
Intercept (β_0)	**0.8256***	0.0753	**0.2997***
wind speed (β_1)	***-0.0057****	**0.0075***	**0.0074***
slope with uphill winds (β_2)	0.0018	**0.0175***	**0.0159***
slope with tangential winds (β_3)	**0.0162***	**0.0081***	**0.0076***
slope with downhill winds (β_4)	**0.0086***	**0.0076***	**0.0069***
time of day ("afternoon" as reference) (β_5)			
evening	***-0.0032****	***-0.0339****	***-0.0290****
morning	**0.0170***	***-0.0253****	***-0.0240****
season ("DJF" as reference) (β_6)			
JJA	***-0.0075****	**0.0139***	**0.0140***
MAM	***-0.0091****	**0.0184***	**0.0180***
SON	***-0.0062****	**0.0124***	**0.0127***
year (β_7)	***-0.0003****	**0.0002***	**0.0002***
soil moisture (β_8)	**0.1954***	***-0.0742****	***-0.0933****
soil texture ("sand" as reference) (β_9)			
loamy sand	**0.0029***	***-0.0064****	***-0.0033****
sandy loam	***-0.0049****	**0.0106***	**0.0125***
loam	**0.0079***	**0.0151***	**0.0147***
sandy clay loam	***-0.0173****	**0.0204***	**0.0185***
clay loam	***-0.0108****	**0.0393***	**0.0358***
clay	***-0.0454****	**0.0841***	**0.0766***
organic materials	-0.0025	-0.0019	0.0012
bedrock	**0.0080***	**0.0179***	**0.0164***

Table S5. Estimates, standard errors, and p-values of all 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, as well as significant interaction terms between wind conditions and other independent variables. The interaction coefficients represent wind conditions (speed and direction) under various situations of time of day, season, and soil moisture. The symbols of coefficients are defined in Eq. (3) and (4). Statistically significant (at 0.05 significance level) coefficients are bolded and their p-values are marked with "*", among which the negative coefficients are italic.

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***
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****

	Estimates	Standard errors	p-values
MAM	***-0.0105***	***0.0032***	***0.0011****
SON	***-0.0090***	***0.0036***	***0.0116****
Other coefficients			
	Estimates	Standard errors	p-values
Intercept (β_0)	0.0703	0.1560	0.6522
time of day ("afternoon" as reference) (β_5)			
evening	***-0.1210***	***0.0044***	***<0.0001****
morning	***-0.0415***	***0.0042***	***<0.0001****
season ("DJF" as reference) (β_6)			
JJA	**0.0360**	**0.0047**	**<0.0001***
MAM	**0.0370**	**0.0045**	**<0.0001***
SON	**0.0166**	**0.0053**	**0.0018***
year (β_7)	**0.0002**	**0.0001**	**0.0149***
soil moisture (β_8)	0.0156	0.0208	0.4528
soil texture ("sand" as reference) (β_9)			
loamy sand	***-0.0061***	***0.0014***	***<0.0001****
sandy loam	**0.0108**	**0.0013**	**<0.0001***
loam	**0.0149**	**0.0011**	**<0.0001***
sandy clay loam	**0.0203**	**0.0025**	**<0.0001***
clay loam	**0.0391**	**0.0035**	**<0.0001***
clay	**0.0846**	**0.0063**	**<0.0001***
organic materials	-0.0024	0.0042	0.5634
bedrock	**0.0165**	**0.0019**	**<0.0001***