Supplement of Atmos. Chem. Phys., 25, 9295–9314, 2025
https://doi.org/10.5194/acp-25-9295-2025-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Aerosol impacts on isolated deep convection: findings from TRACER

Dié Wang et al.

Correspondence to: Dié Wang (diewang@bnl.gov)

Text S1. TINT algorithm

Tracking convective rainfall cores initiated during DCC days in the anticyclonic regime is accomplished using the TINT algorithm, developed by Raut et al. (2021). TINT, an open-source tracking tool, automatically estimates the trajectories of moving cloud objects in sequential images, applicable to various two-dimensional datasets, including remote sensing measurements and model simulations (Fridlind et al. 2019). For a detailed explanation of the TINT algorithm, refer to Raut et al. (2021).

The inputs for TINT are derived from the 2-km Z values derived from gridded NEXRAD KHGX-Houston radar data (1 km horizontal resolution). The selection of the 2 km level is made to accurately sample precipitation areas, following the method presented in Oue et al. (2022). The position of each cell is determined as the difference-weighted center of the region with Z exceeding a certain threshold value. Tracking is conducted for all selected DCC days, starting from 0500 LT and continuing until 0459 LT the following day. To ensure the optimal performance of TINT, several thresholds are applied:

- Definition of contiguous object: To mitigate signal noise contamination, a single object is characterized as a contiguous area including a minimum of eight radar grid points (1 km × 1 km) with a minimum Z threshold of 10 dBZ, signifying the presence of precipitation echoes.
- Identification of convective cells: An object is labeled as a convective cell if its highest 2-km Z exceeds 30 dBZ. This threshold is widely used by radar systems for detecting convective clouds, as demonstrated in studies such as Petersen et al. (1996), Kumar et al. (2016), and Gupta et al. (2023).
- Definition of deep convective event: A deep convective event is defined as a track with a 2-km Z exceeding 40 dBZ and a 30-dBZ echo top height (ETH) higher than 5 km at any point during its lifespan, following a similar definition to Dixon and Wiener (1993). Convective cells during these events are called convective rainfall cores which are used in our study.
- Domain exclusion: Convective rainfall cores that move into or out of the radar domain (400 km × 400 km) during their life cycle are excluded from the statistical analysis.
- Lifetime duration: To minimize potential misidentification due to uncertainties in radar data or the tracking method, the analysis primarily focuses on deep convection with rainfall cores lasting longer than 40 minutes. This criterion ensures that the convective rainfall cores are detected in at least 7 or 8 consecutive radar scans.

Text S2. Definitions of key DCC parameters

Several key parameters are derived and analyzed for these tracked convective rainfall cores, including the maximum 2-km *Z*, rainfall core area, 30-dBZ ETH/15-dBZ ETH, core lifetime, and core propagation speed. Here is a summary of these key parameters:

1. Maximum 2-km *Z*: This parameter represents the highest intensity of radar reflectivity observed within the convective cells. It indicates the strength of the precipitation and provides information about the intensity of the convective activity.
2. Core lifetime: The lifetime of a convective rainfall core is determined by the duration between its first identification time and its last identification time. This parameter provides insights into how long the precipitating cells persist in the study region on average.
3. Core area: The core area is calculated by multiplying the number of continuous grid points with 2-km *Z* greater than 30 dBZ by the grid resolution. It provides an estimate of the spatial extent covered by the convective rainfall cores and indicates the size of the area experiencing moderate precipitation.
4. 30-dBZ ETH/15-dBZ ETH: The ETH is defined as the maximum height at which the radar echo reaches a Z of 30 dBZ/15 dBZ. It serves as a proxy for the strength of the convective updraft or upward mass flux, providing information about the vertical extent and intensity of the convective activity (Heymsfield et al., 2010).
5. Core propagation speed: The propagation speed of a convective rainfall core is determined as the mean rate at which a cell advances from its first radar detected location to its final radar detected position.
6. Initiation timing and location: These parameters indicate when and where the rainfall cores first develop. They provide insights into the preferred timing and spatial distribution of rainfall initiation in the radar region.

It is important to note that the TINT algorithm used in our study is not designed to efficiently identify the splitting and merging of convective cores, which can be more prevalent in organized convective systems. As a result, this limitation may have some impact on the estimation of convective properties in those clouds. However, our study specifically focuses on relatively isolated DCCs, which is a common type of DCC in southeastern Texas during the summer months in the anticyclonic regime (Fridlind et al., 2019). In these isolated cases, the occurrence of mergers and splitters is relatively less frequent compared to organized DCCs. Therefore, we anticipate that the impact of this limitation of TINT on our statistical analysis of convective rainfall core life cycle characteristics will be minimal.

Text S3. Sensitivity tests of tracking thresholds

We conducted sensitivity tests to evaluate the robustness of the criteria used in defining and tracking convective rainfall cores in the prior subsection. The tests were performed using data from June to August 2022 (regardless of SOM regimes), and the results are summarized in Table S3. Here are the details of the sensitivity tests and their impact on DCC properties:

1. Single object definition: In this test, a single object was defined as a contiguous area of at least 10 radar grid points with a minimum 2-km *Z* threshold of 10 dBZ. This resulted in a 5% reduction in the total number of cells, and a slight increase of 5 km² in the median core area compared to the original criteria.
2. Tracking using 3-km *Z*: In this test, the tracking of cells was performed using a 3-km *Z* threshold. No significant changes were found in terms of the convective properties of interest.
3. Decreased lifetime threshold: The DCC lifetime limit was decreased from 40 minutes to 30 minutes in this test. As expected, the number of cores significantly increased by 32%, and the median rainfall core lifetime was shortened by 8 minutes.
4. Calibration offset: To account for the possible NEXRAD *Z* offset mentioned in previous studies (Gourley et al., 2003; Ryzhkov et al., 2005), an offset of 3 dBZ was applied to the data. This test aimed to investigate the potential impact of the calibration issue on derived convective properties. The results showed that this calibration offset had only a slight impact on the number of cores and core properties of interest.

Overall, the mean state of core properties was somewhat sensitive to the thresholds used in the tracking model, but most of the changes were insignificant according to results from a *t*-test. The most significant changes in cell core area occurred when the lifetime threshold was loosened. However, no significant differences are found in terms of 30-dBZ ETH when using different thresholds.

Text S4. Key assumptions of the MLR models

We evaluate the MLR models by assessing and validating the key assumptions. These assumptions include linearity, homoscedasticity, normality, independence, and multicollinearity. In the following paragraphs, we provide a detailed analysis and explanation of these tests and plots. Overall, the results indicate that the data meet the essential assumptions of the MLR model. This fulfillment of assumptions strengthens the reliability of the following estimated causal effects of aerosols.

To test for linearity, we examine whether there is a linear relationship between the predictor variable and the outcome variable by plotting the model residuals as a function of the fitted variables. The residuals are the differences between the observed and predicted values of the outcome variable after fitting a linear model to the data. The fitted values, on the other hand, are the predicted values of the outcome variable based on the MLR model. As illustrated in Figure S5, when using DCC cases identified within a radius of 50 km from the M1 site, the residuals are evenly spread around a horizontal line with no distinct patterns when considering different exposures, indicating the absence of non-linear relationships. Similar patterns are shown for other scenarios considered in the study. This provides some evidence for linearity between the outcome and predictor variables in our data.

To validate the assumption of homoscedasticity in the MLR model, we create a Spread-Location or Scale-Location plot as shown in Figure S6. The plot is generated by plotting the square root of the absolute value of the standardized residuals on the y-axis against the fitted values on the x-axis. The standardized residuals were calculated by dividing the residuals by their estimated standard deviation, resulting in a mean of zero and a standard deviation of one. When the assumptions of linearity and constant variance are met, the Spread-Location plot will show a roughly horizontal line with no discernible pattern, indicating that the residuals have a constant variance across the range of fitted values. In Figure S6, the red line is close to being horizontal, and the points are equally and randomly spread, indicating good homoscedasticity. We find similar results for other scenarios considered in the study.

For validating the assumption of normality in the MLR model, we employ a quantile-quantile (QQ) plot of standardized residuals plotted against theoretical quantiles. The theoretical quantiles are the expected values if the data followed a normal distribution. If the standardized residuals are normally distributed, the plot will show a linear relationship between the two sets of values, as depicted in Figure S7. Based on the plot, we observe that the data closely follows a straight line, indicating that the residuals are normally distributed. Thus, we can conclude that the assumption of normality has been met for our MLR regression model.

Multicollinearity refers to the situation where two or more independent variables in a MLR model are highly correlated with each other. This can lead to unstable regression coefficients, inflated standard errors, and inaccurate predictions. To check for the presence of little or no multicollinearity in this study, we used the Variance Inflation Factor (VIF). The VIF measures the extent to which the variance of the estimated regression coefficient is increased due to multicollinearity. A VIF of 1 indicates no multicollinearity, while a VIF greater than 1 suggests the presence of multicollinearity. Typically, a VIF greater than 5 or 10 is a cause for concern.

When using the most-unstable parcel and the first data pairing method, the VIF values for three covariates in the model for predicting 30-dBZ ETH are all around 1, which suggest limited multicollinearity. Note that standardizing the covariates significantly reduces the VIF value for all predictors, and we, therefore, used the standardized covariates for subsequent calculations of g-computation. When using different parcels as the surface or mixed-layer parcel, the VIF values only change slightly, but are all below 2.

The independence of observations in the MLR model is assessed using the Durbin-Watson test. This test is a statistical method used to test for the presence of autocorrelation in the residuals. Autocorrelation refers to the situation where the residuals are correlated with each other over time, which violates the assumption of independence. The Durbin-Watson test calculates a test statistic that ranges from 0 to 4. A value of 2 indicates that there is no autocorrelation, while a value below 2 indicates positive autocorrelation, and a value above 2 indicates negative autocorrelation. Typically, a value between 1.5 and 2.5 is considered acceptable. The Durbin-Watson values for all scenarios are around 2, which indicates that our observations are independent of each other. Therefore, we can conclude that there is no evidence of autocorrelation in the residuals of our MLR models.

Assessing the influence of individual observations on the results of a regression analysis is crucial in identifying and eliminating any strong influencers. To accomplish this, a leverage plot is conducted to show the leverage values of each case against the standardized residuals. Leverage values represent the degree to which an observational point can affect the slope of the regression line. Observations with high leverage values and large standardized residuals have the potential to strongly impact the slope, coefficients, or other aspects of the model, such as the statistical significance of the predictors. These observations may be outliers or may have extreme values of the predictor variables, meaning that they do not follow the trend observed in most cases. Cook's distance, which measures the influence of an observation on the regression coefficients, is used in conjunction with the leverage plot to identify influential observations. Typically, a cutoff value of 1 is used, which is plotted as a red dashed line in Figure S8. This means that any observation with a Cook's distance greater than 1 is considered influential. In our dataset, all cases are below the Cook’s distance cutoff

values, indicating that they have low Cook's distance scores and do not have a large impact on the estimated regression results. Similar results are shown for other scenarios as well.

Table S1. Intense operation days during the TRACER Intense Operation Period in 2022.

Date (MMDD)
0601, 0602, 0604, 0617, 0621, 0622, 0626, 0629, 0706, 0710, 0711, 0712, 0713, 0727, 0729, 0730, 0801, 0802, 0804, 0807, 0808, 0809, 0810, 0821, 0822, 0826, 0827, 0828, 0831, 0901, 0902, 0906, 0907, 0916, 0917, 0918, 0919, 0925

Table S2. Correlation coefficients between meteorological variables and ETH when the surface parcel or mix-layer parcel is used in the calculations of the meteorological variables.

Distance to M1	LNB	CAPE	LCL	LFC	ELR_3	ELR_6	LWS	RH
30-dBZ ETH, Surface Parcel								
20 km	x	x	x	x	x	x	x	x
30 km	x	x	0.3	x	0.3	x	x	x
40 km	x	0.3	0.3	x	0.3	x	x	x
50 km	0.3	0.2	0.4	x	0.4	x	x	x
30-dBZ ETH, Mixed-layer Parcel								
20 km	x	x	x	-0.3	x	x	x	x
30 km	x	x	0.4	x	0.3	x	x	x
40 km	x	x	0.3	x	0.3	x	x	x
50 km	0.2	x	0.4	x	0.4	x	x	x
15-dBZ ETH, Surface Parcel								
20 km	x	x	x	x	x	x	-0.3	x
30 km	x	x	x	x	x	x	x	x
40 km	x	x	x	x	x	x	x	x
50 km	x	x	x	x	x	x	x	x
15-dBZ ETH, Mixed-layer Parcel								
20 km	x	x	x	x	x	x	-0.3	x
30 km	x	x	x	x	x	x	x	x
40 km	x	x	x	x	x	x	x	x
50 km	0.3	0.3	x	x	x	x	x	x

Table S3. DCC properties from several sensitivity tests performed on three-month data from June to August in 2022.

Z threshold [dBZ]	Min. no. of grid	Min. lifetime [min]	Level of Z for tracking [km]	No. of samples	Median max. Z [dBZ]	Median core area [km2]	Median 30-dBZ ETH [km]	Median lifetime [min]
30	8	40	2	9568	46.4	47	6.5	80
30	10	40	2	9149	46.8	52	7	80
30	8	40	3	9689	46.4	47	7	80
30	8	30	2	12635	46.0	42	6.5	72
33	8	40	2	8434	47.5	45	6.5	78

Table S4. The Adjusted R^2 values for the fitted MLR models and the 95% confidence intervals for the independent variables. The outcome variable is the 30 dBZ ETH and the CAPE is calculated when assuming the most unstable parcel would rise.

Exposure, Distance to the ARM site	Adjusted R^2	95% confidence intervals for the exposure variables	95% confidence intervals for CAPE	95% confidence intervals for ELR_3
All cases				
N_{cn}, 40 km	0.1	[0.01, 1.58]	[-0.14, 0.91]	[-0.32, 0.82]
N_{cn}, 50 km	0.3	[-0.01, 1.41]	[-0.16, 0.79]	[-0.53, 0.55]
N_{ufp}, 20 km	0.1	[0.60, 3.79]	[-1.69, 0.37]	[-1.20, 0.70]
N_{ufp}, 30 km	0.2	[0.11, 1.98]	[-0.49, 0.71]	[-0.32, 0.94]
N_{ufp}, 40 km	0.2	[0.12, 1.68]	[-0.36, 0.69]	[-0.31, 0.78]
N_{ufp}, 50 km	0.4	[0.40, 1.72]	[-0.30, 0.58]	[-0.38, 0.59]
Sea breeze cases only				
N_{ccn1}, 30 km	0.3	[0.13, 4.12]	[-0.88, 1.97]	[-1.87, 0.86]
N_{ccn1}, 40 km	0.3	[-0.34, 2.73]	[-0.21, 1.66]	[-1.08, 1.03]
N_{ccn1}, 50 km	0.4	[-0.52, 1.66]	[0.43, 1.80]	[-0.75, 0.76]
N_{ccn08}, 30 km	0.3	[0.13, 4.12]	[-0.88, 1.97]	[-1.87, 0.86]
N_{ccn08}, 40 km	0.4	[-0.22, 2.98]	[-0.01, 1.91]	[-0.91, 1.10]
N_{ccn08}, 50 km	0.4	[-0.43, 2.23]	[0.14, 1.79]	[-0.86, 0.87]
N_{ccn06}, 30 km	0.3	[0.13, 4.12]	[-0.88, 1.97]	[-1.87, 0.86]
N_{ccn06}, 40 km	0.4	[0.22, 3.05]	[-0.01, 1.69]	[-1.17, 0.64]
N_{ccn06}, 50 km	0.5	[0.61, 2.79]	[-0.03, 1.29]	[-1.07, 0.36]
N_{ccn04}, 30 km	0.4	[0.80, 5.09]	[-1.07, 2.13]	[-1.71, 0.72]
N_{ccn04}, 40 km	0.4	[-0.22, 2.98]	[-0.01, 1.91]	[-0.91, 1.10]
N_{ccn04}, 50 km	0.4	[0.33, 2.80]	[-0.06, 1.45]	[-1.01, 0.60]
N_{ccn02}, 20 km	0.5	[-0.75, 6.75]	[-2.47, 4.81]	[-2.15, 1.22]
N_{ccn02}, 30 km	0.5	[0.18, 2.93]	[0.18, 2.93]	[-1.46, 1.20]

N_{ccn02}, 50 km	0.3	[-0.82, 1.60]	[0.16, 1.91]	[-0.62, 1.09]
N_{ccn01}, 30 km	0.5	[0.38, 3.76]	[0.18, 2.93]	[-1.46, 1.20]
N_{cn}, 40 km	0.2	[-0.36, 2.56]	[-0.40, 1.49]	[-0.52, 1.24]
N_{cn}, 50 km	0.5	[0.44, 2.29]	[-0.48, 0.82]	[-0.59, 0.72]
N_{ufp}, 50 km	0.5	[0.19, 2.13]	[-0.45, 0.85]	[-0.61, 0.75]

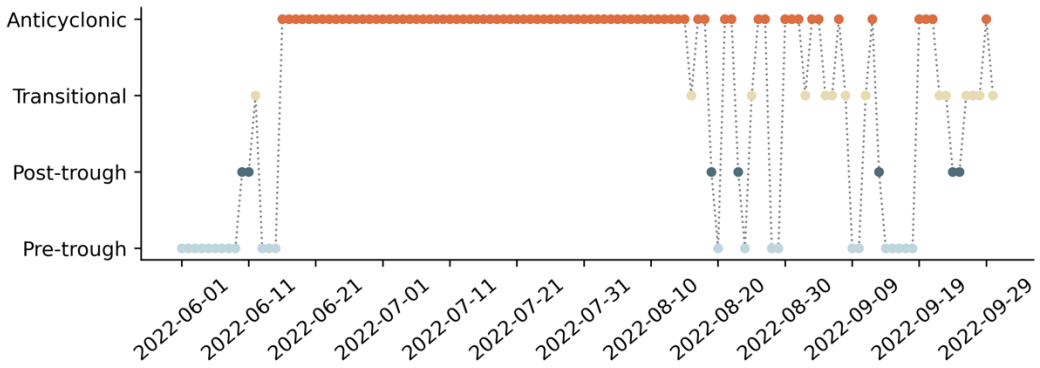

Figure S1. Synoptic regimes as a function of day for 2022, classified using the Self-Organizing Map.

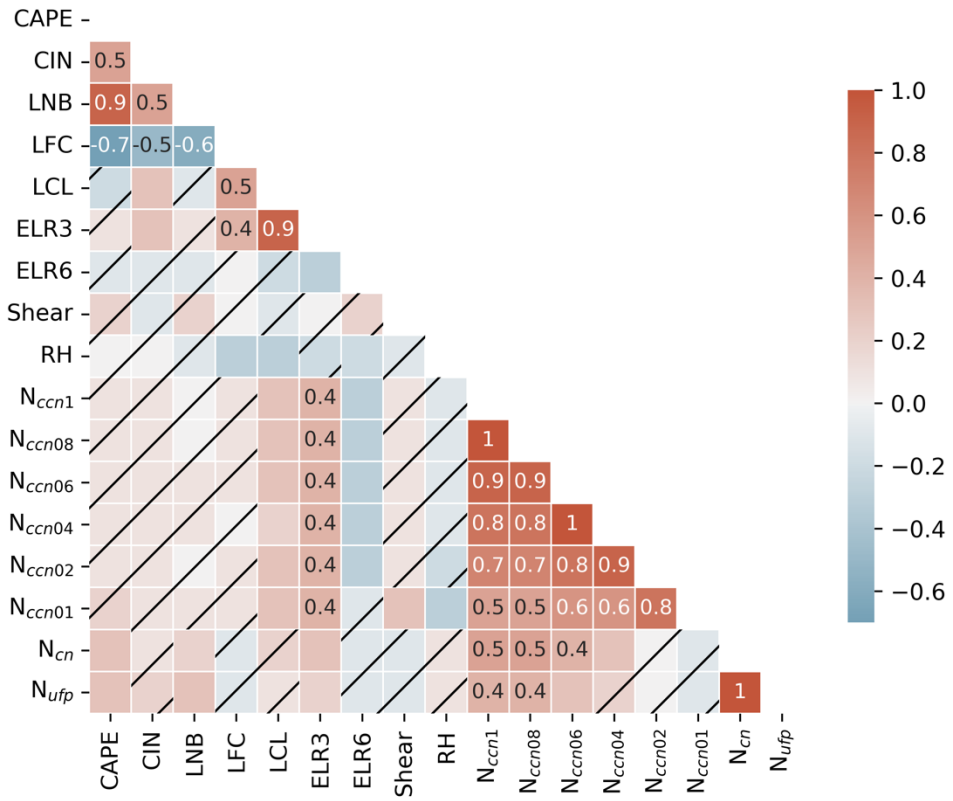

Figure S2. Correlation matrix between the meteorological variables and aerosol number concentrations for DCC cases identified within a radius of 50 km from the M1 site. The meteorological variables are calculated using ARM soundings when assuming the surface parcel would rise to form a convection.

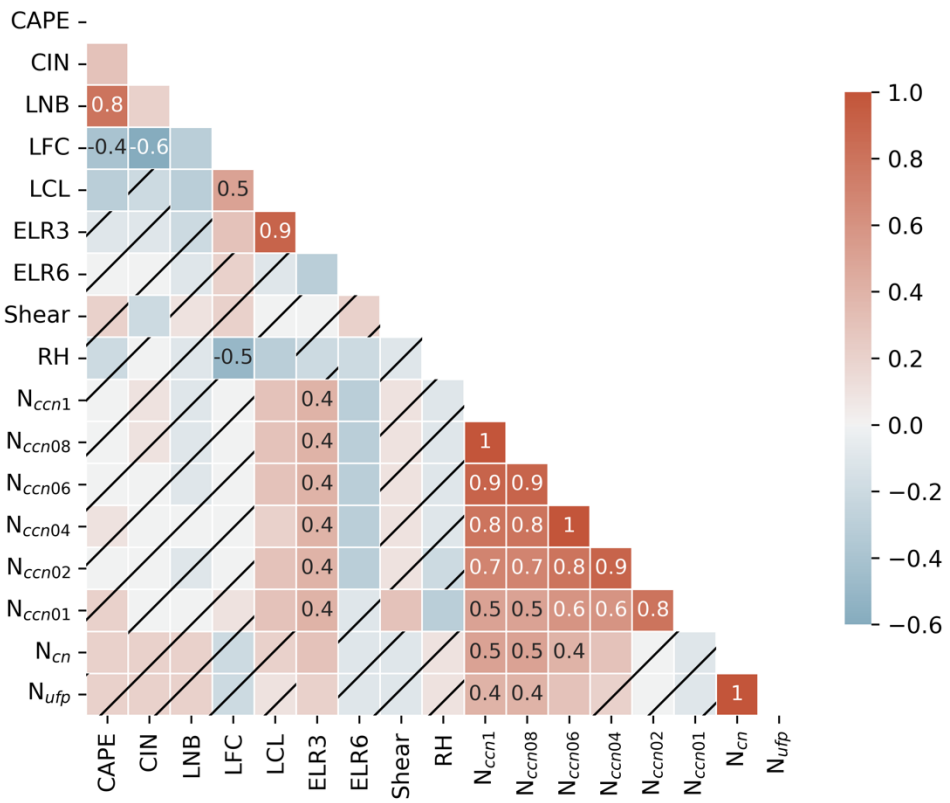

Figure S3. Same as Figure S3 but using a mixed-layer parcel.

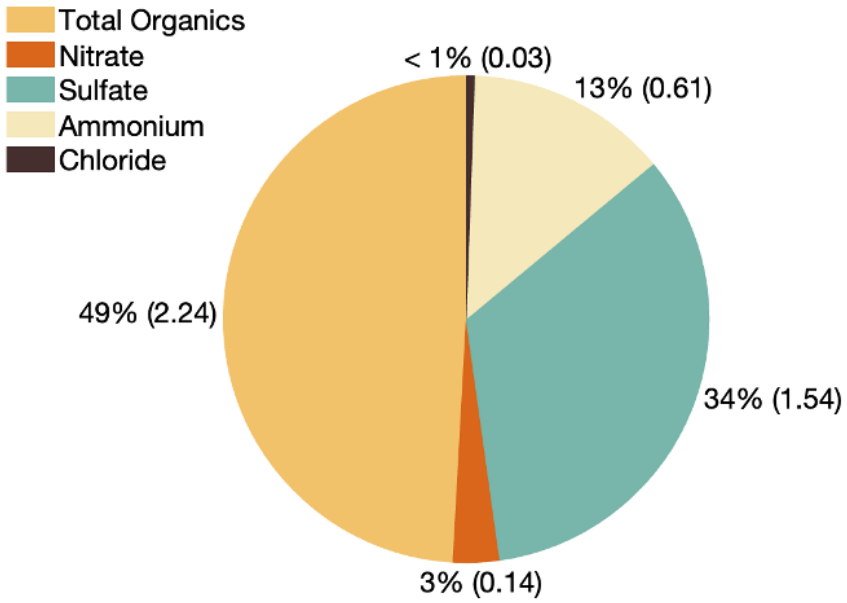

Figure S4. Averaged percentage contribution of aerosol mass concentrations measured during a one-hour period following sounding launches on selected DCC days.

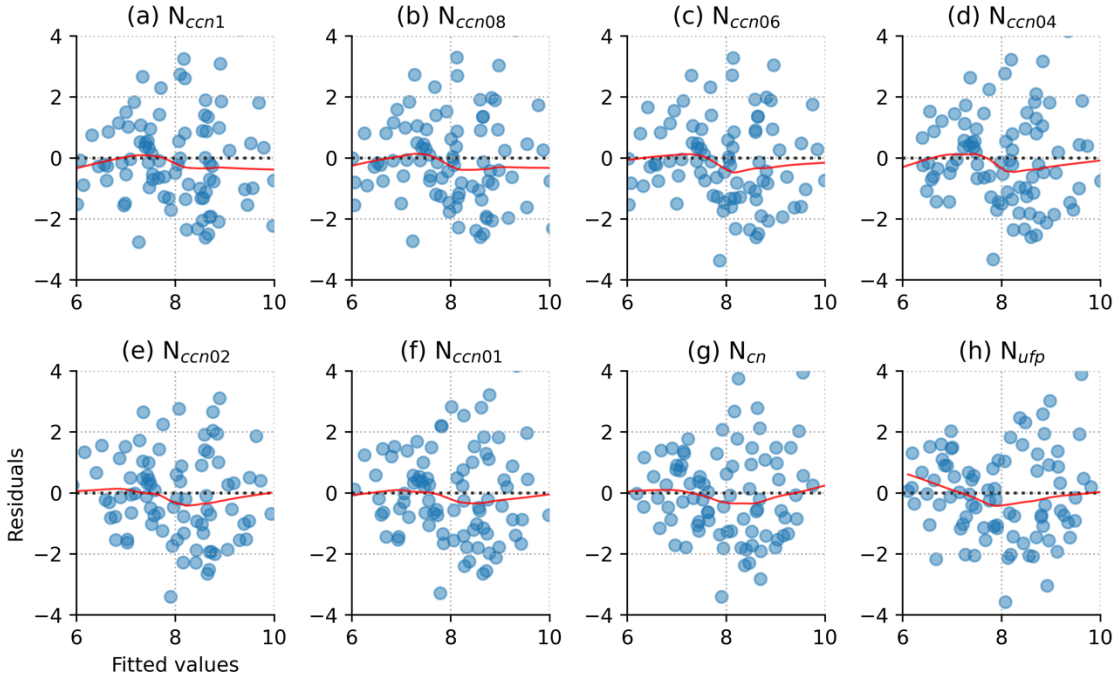

Figure S5. MLR model residuals as a function of fitted values for DCCs formed within a radius of 50 km from the M1 site using the post-sounding data averaging method. The most unstable parcel is used for the calculations of meteorological variables. The outcome variable is 30-dBZ ETH.

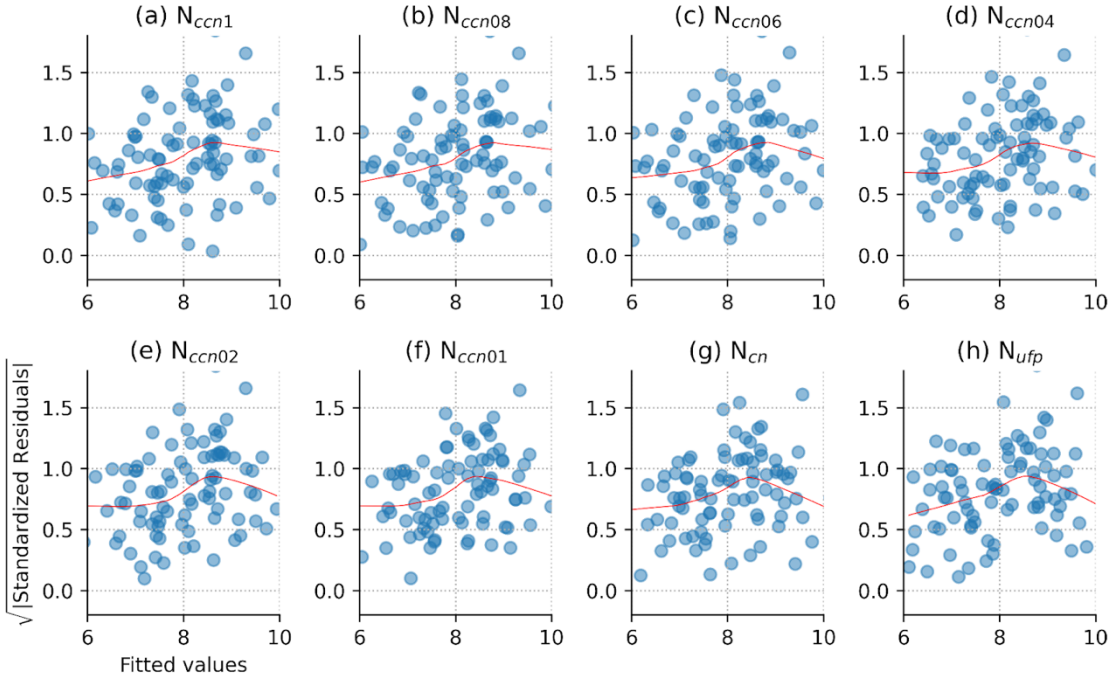

Figure S6. Standardized residuals as a function of fitted values for DCCs formed within a radius of 50 km from the M1 site using the post-sounding data averaging method. The most unstable parcel is used for the calculations of meteorological variables. The outcome variable is 30-dBZ ETH.

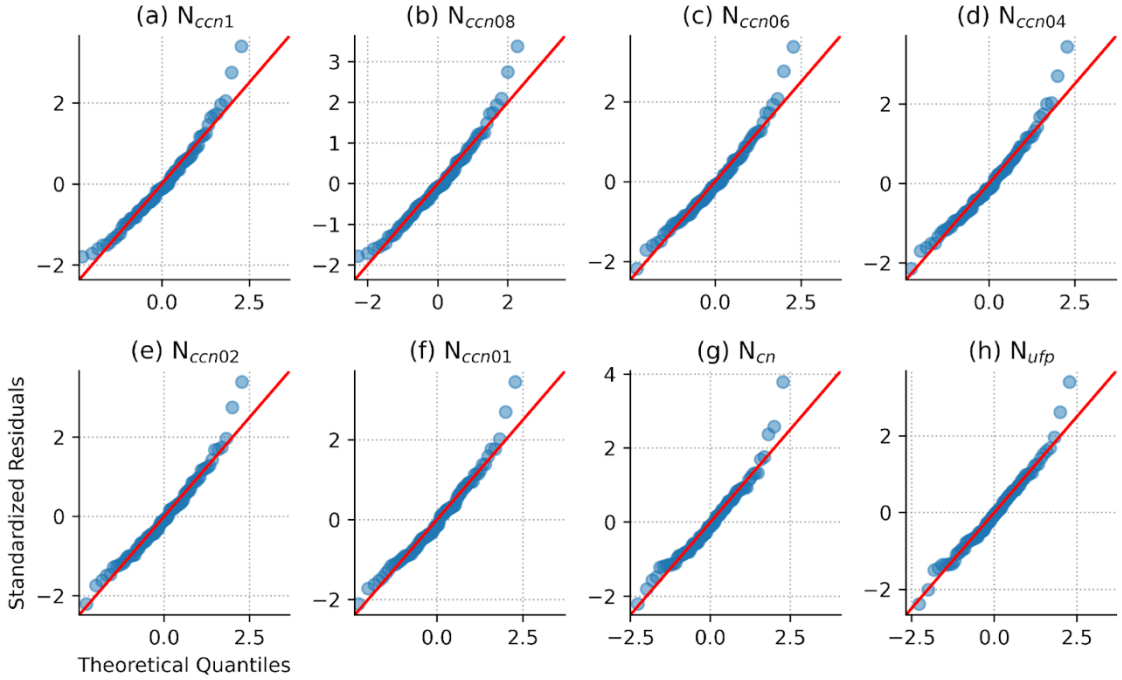

Figure S7. Quantile-quantile plot for DCCs formed within a radius of 50 km from the M1 site using the post-sounding data averaging method. The most unstable parcel is used for the calculations of meteorological variables. The outcome variable is 30-dBZ ETH.

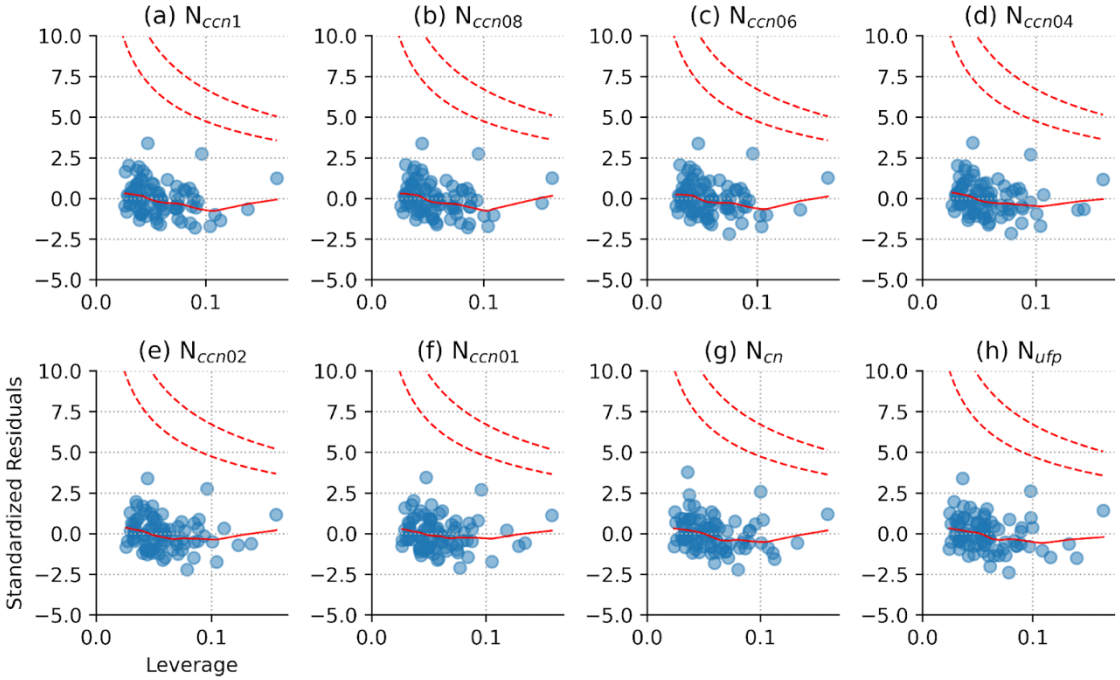

Figure S8. leverage plot for DCCs initiated within a radius of 50 km from the M1 site using the post-sounding data averaging method. The most unstable parcel is used for the calculations of meteorological variables. The outcome variable is 30-dBZ ETH.