JES

JOURNAL OF ENVIRONMENTAL SCIENCES

www.jesc.ac.cn

Origin of high particle number concentrations reaching the St. Louis, Midwest Supersite

Benjamin de Foy[1],*, James J. Schauer[2]

1. *Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO, USA. E–mail: bdefoy@slu.edu*
2. *Civil and Environmental Engineering, University of Wisconsin, Madison, WI, USA*

ARTICLE INFO

ABSTRACT

Ultrafine particles are associated with adverse health effects. Total Particle Number Concentration (TNC) of fine particles were measured during 2002 at the St. Louis — Midwest supersite. The time series showed overall low level with frequent large peaks. The time series was analyzed alongside criteria pollutant measurements and meteorological observations. Multiple regression analysis was used to identify further contributing factors and to determine the association of different pollutants with TNC levels. This showed the strong contribution of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) to high TNC levels. The analysis also suggested that increased dispersion resulting from faster winds and higher mixing heights led to higher TNC levels. Overall, the results show that there were intense particle nucleation events in a SO_2 rich plume reaching the site which contributed around 29% of TNC. A further 40% was associated with primary emissions from mobile sources. By separating the remaining TNC by time of day and clear sky conditions, we suggest that most likely 8% of TNC are due to regional nucleation events and 23% are associated with the general urban background.

© 2015 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences.
Published by Elsevier B.V.

Introduction

Due to the impacts of high particle number concentrations on human health and climate, there is a growing interest in understanding the sources and mechanism of the formation of ultrafine particles in the atmosphere (Oberdörster, 2001; Pierce and Adams, 2009). Much of the understanding of ultrafine particle number concentrations in urban areas has arisen from studies of the evaluation of emissions near roadways (Zhu et al., 2002) and regional nucleation events (Kulmala et al., 2004; Wang et al., 2013; Wu et al., 2007). However, several studies have shown that in urban areas, there are other sources that contribute to particle number concentrations such as recycling concrete (Kumar and

Morawska, 2014), industrial sources (Bycenkiene et al., 2014), ships (Donateo et al., 2014) wood burning and rail (Kuwayama et al., 2013). Due to the fact that ultrafine particles can be directly emitted from sources, can be formed in plumes in transport from source to receptor, and can result from regional nucleation events (Kumar et al., 2011, 2014), elucidating the sources of particle number concentrations in urban areas can be complex and is not expected to be generalizable across cities with very different air pollution sources, climatology, and weather. Nonetheless, there is a great need to better understand sources of particle number concentrations. Several studies have been conducted that have employed different methodologies to apportion particle number concentrations to sources including the use of PMF (Kuwayama et

* *Corresponding author*. E-mail: bdefoy@slu.edu (Benjamin de Foy).

al., 2013), particle size distribution fingerprints (Hussein et al., 2014), potential source contribution function (PSCF) (Bycenkiene et al., 2014), and the association with source tracers (Donateo et al., 2014).

In the current study, we examine particle number concentrations in East St. Louis, Missouri in the American Midwest. The time series was analyzed by Qian et al. (2007) who focused on regional nucleation events which occurred when sulfur dioxide and nitrogen oxides levels were low. We are interested in analyzing the rest of the time series, specifically the times when there were high levels of particle number concentrations associated with sulfur dioxide plumes and with elevated nitrogen oxides concentrations.

We focus on particle number concentrations in the absence of speciation data, which precludes the use of techniques such as Positive Matrix Factorization and Chemical Mass Balance for source apportionment. As an alternative, we use a multiple regression model to evaluate the dependence of the particle numbers on meteorological parameters and gas phase concentrations. Specifically, we examine the relationship of particle number concentrations with ozone which serves as an indicator of photochemical reactions, with sulfur dioxide which is present in large concentrations in an industrial plume and with nitrogen oxide which is a tracer for mobile source emissions impacting the supersite. We also examine the relationship with the following meteorological parameters: temperature, relative humidity, wind speed, solar radiation and mixing height.

1. Measurements and methods

1.1. Measurements

The measurement site is the St. Louis — Midwest Supersite (ESTL) which was funded by the United States Environmental Protection Agency (EPA). It is located in East St. Louis, approximately 3 km east of the Central Business District of St. Louis, on the other side of the Mississippi river in a low-density, mixed-use neighborhood impacted by industrial point sources nearby.

Measurements of particle number concentration are described in detail in Qian et al. (2007): particles with mobility diameters ranging from 3 to 40 nm were measured using a Nano-scanning mobility particle sizer (Nano-SMPS) consisting of a nano-DMA (TSI Model 3085) and an ultrafine condensation particle counter (TSI Model 3025). Particles with mobility diameters from 30 to 400 nm were measured with a SMPS consisting of a long column DMA and a TSI Model 3760 condensation particle counter. Particles from 0.1 to 2 μm were measured using a LASAIR Model 1002 single particle optical counter. Hourly Total Particle Number Concentrations (TNCs) were obtained for the range of 3 nm to 2 μm for the year 2002. The measurements were initially processed with 5 min intervals, which were averaged to 1 hr intervals for analysis.

Hourly measurements of nitrogen oxides (NO_x), sulfur dioxide (SO_2), ozone (O_3), and $PM_{2.5}$ were available from United States Environmental Protection Agency *via* the Air Quality System, and EC and OC were available as described in Bae et al. (2004). Hourly meteorological observations were obtained from St. Louis Downtown Airport (KCPS) in Cahokia, IL, 5 km south of the measurement site, from the Integrated Surface Hourly Data available from the National Climatic Data Center. Meteorological data from KCPS was used because it is in agreement with onsite observations but had fewer missing hours.

Estimates of solar radiation and planetary boundary layer heights were obtained at the supersite from numerical simulations using the Weather Research and Forecasting (WRF) model, version 3.5.1 (Skamarock et al., 2005), as described in detail in de Foy et al. (2014).

1.2. Data analysis and multiple regression model

In order to analyze different aspects of the TNC time series, we will use 3 different sets of groups. The first set of groups is a simple classification into 4 quadrants according to peak values of SO_2 and TNC. The second set of groups is designed to separate the effects of SO_2, NO_x and cloudiness. The final set of groups is used in the regression analysis to differentiate between night and day.

For the first part of the analysis we classify the TNC time series into 4 groups based on levels of SO_2 and TNC. There are different approaches to identifying peak levels in time series. For this case, both time series have an annual and a diurnal signal that reduces the effectiveness of using a single threshold to classify the data. We therefore use the regression model described below to obtain a smoothed time series of TNC and SO_2 that includes seasonal and diurnal variations. Multiple regression analysis was performed using as inputs sine and cosine functions with periods of 1, 1/2, 1/4, 1/8, and 1/16 of a year as well as 24, 12, 6, and 3 hr. Any hour with a TNC or SO_2 concentration that has a residual that is greater than one standard deviation of the residual of the data to the harmonic fit was deemed a peak value.

We perform a multiple regression analysis of the TNC time series at the supersite based on a combination of hourly measurements and model-produced variables. TNC is log-normally distributed, and so we carry out the model for log (TNC) which we scale to have a normal distribution with zero mean and unit standard deviation. Among the other input variables, the following are also approximately log-normally distributed: SO_2, NO_x, O_3, Planetary Boundary Layer Height (PBLH) and Solar Radiation (SR). For the variables with zero values, we use an offset to obtain the best fit to a normal distribution after taking the logarithm: the offsets used were 1 ppb for SO_2 and O_3, and 10 W/m^2 for solar radiation. We then took the log of the modified variables and rescaled the output to approximate a normal distribution with zero mean and unit standard deviation, as was done for TNC. We use a linear transformation for the remaining meteorological parameters: Temperature (T), Relative Humidity (RH) and Wind Speed (WS).

This means that the regression model obtains an expression for TNC of the following form:

$$TNC = Bkg \times (SO_2 + 1)^{\alpha 1} \times NOx^{\alpha 2} \times (O_3 + 1)^{\alpha 3} \times PBLH^{\alpha 4}$$
$$\times (SR + 10)^{\alpha 5} \times \exp(\alpha_6 T) \times \exp(\alpha_7 RH) \times \exp(\alpha_8 WS).$$
$$(1)$$

The regression is performed with a least squares inversion which determines the regression coefficients α_i and the background

term Bkg, PBLH is planetary boundary layer height. Because the least squares procedure is sensitive to outliers, we use Iteratively Reweighted Least Squares (IRLS) to increase the robustness of the results: after solving for the regression coefficients, we calculate the residual of the fit and exclude points with a residual greater than 2 times the standard deviation of the residual. This is performed iteratively to converge on a stable set of outliers. In some cases, it can be instructive to analyze the group of outliers separately as described below.

The standard error of the regression coefficients can be calculated from the least squares inversion. An alternative method of estimating the uncertainty of the coefficients is to use block-bootstrapping. If we assume that measurements, errors are randomly distributed, then selecting multiple samples at random from the existing data will sample times with different errors. Performing separate regressions with each random sample will therefore provide an estimate of the impact of measurement errors on the model results.

Because pollution events may be correlated in time (they typically last on the order of a couple of hours up to a day), we use block-bootstrapping with a block length of 24 hr: days to include in the analysis were selected at random with replacement for each realization of the model. By carrying out 100 realizations of the model, we were able to obtain the standard deviation of the uncertainty in the coefficients, as well as the covariation of the coefficients.

2. Results and discussion

2.1. Time series analysis

The time series of TNC and SO_2 are dominated by large peaks lasting no more than an hour or two, as can be seen in Figs. 1

and 2. Fig. 1 shows the data for the whole of 2002 and Fig. 2 shows a zoom over Jan 1-16 and July 16-30. The data points were classified into four groups according to the levels of SO_2 and TNC. The cutoff used for the classification of each hour varied by day of year and time of day using the results from the multiple linear regression using the harmonic functions described in Section 1.2 above. The TNC function varied smoothly from a minimum of 10,000 #/cm^3 in the summer to a maximum of 60,000 #/cm^3 in the winter. The diurnal range was 10,000 #/cm^3 in the summer and 20,000 #/cm^3 in the winter. For SO_2, the time series varied smoothly from a minimum of 1 ppb in the summer to a maximum of 6.5 ppb in the winter, and the diurnal range varied from 1.5 ppb in the summer to 2.5 ppb in the winter. High data points were the ones with concentration residuals of more than one standard deviation above the smooth temporal profiles: TNC levels more than 44,000 #/cm^3 above, and SO_2 concentrations more than 13.5 ppb above the smooth time series. In this way we were able to identify peaks during the day and/or the summer that would have been lost if we had been using a single threshold for the entire year. We obtain the following four non-overlapping groups: low SO_2/low TNC (4757 hr), high TNC/low SO_2 (462 hr), high SO_2/low TNC (564 hr), high SO_2/high TNC (289 hr).

The time series in Figs. 1 and 2 show the data points colored by category, which highlights the fact that many of the peaks in SO_2 and TNC match each other, and that SO_2 and TNC have higher peak values in the winter months. There are some peaks in SO_2 that do not have high TNC levels which tend to be at night, as can be seen for Jan 3–4 in Fig. 2. In July, the SO_2 peaks have much lower concentrations, and the TNC peaks are not associated with SO_2 to the same extent as in the winter.

Fig. 3 shows the mean TNC levels by SO_2 and NO_x concentration averaged over the entire measurement period. This
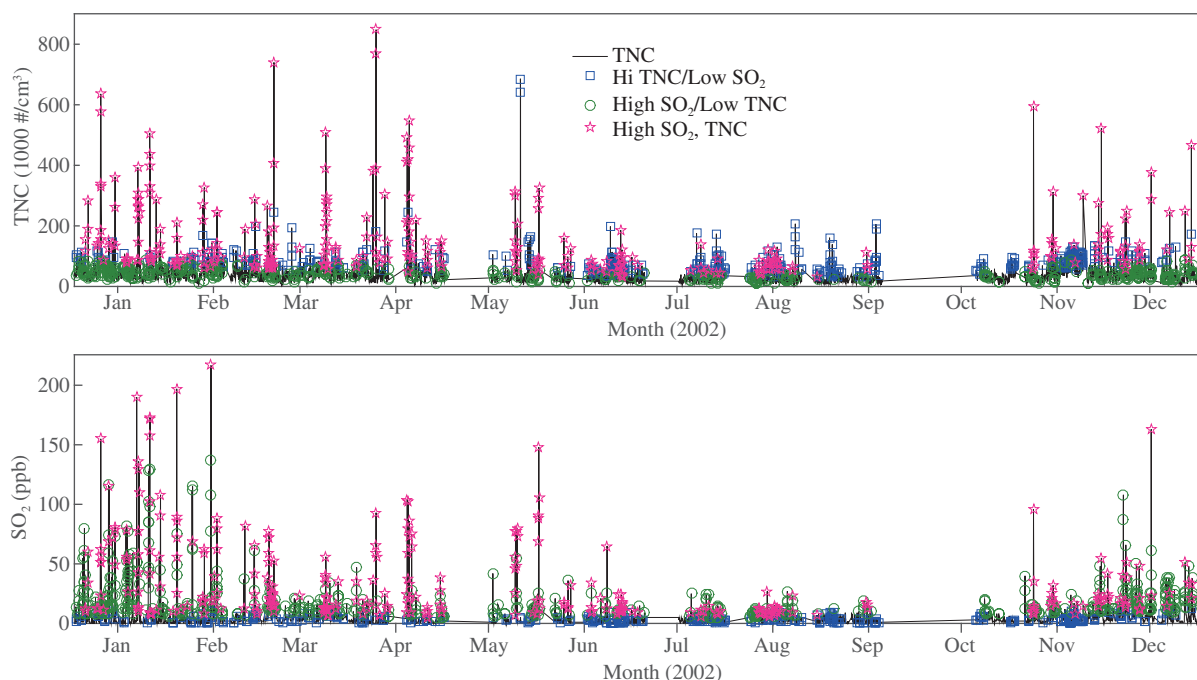

Fig. 1 – Time series of TNC (total particle number concentration) and sulfur dioxide (SO_2) with colored symbols showing three of the four groups used in the analysis.
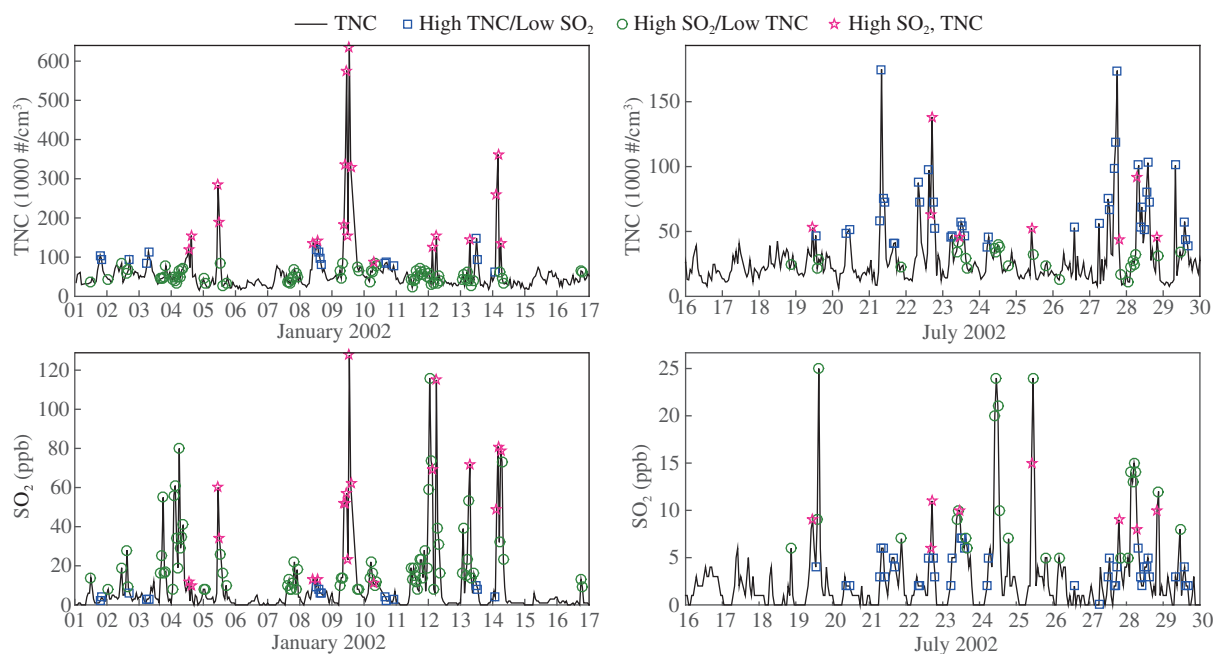
Fig. 2 – Zoom of Fig. 1 for 1–16 January and 16–30 July 2002 showing the time series of TNC and SO_2 for the four groups used in the analysis. Shading shows night time.

isopleth clearly shows that peak TNC levels are associated with high SO_2 plumes that are not high in NO_x. Nevertheless, high NO_x levels are associated with higher TNC levels compared to low NO_x and low SO_2 conditions. Levels of TNC decrease with both decreasing SO_2 and decreasing NO_x. This suggests that the greater fraction of particles are due either to formation in fresh SO_2 plumes, or to primary emissions alongside NO_x sources.

Fig. 4 shows histograms of the TNC and SO_2 levels in each group, with the variables transformed as described in the Measurements and methods section. For SO_2, this illustrates that there are low background levels with high peaks, but that

Fig. 3 – Isopleths of mean total particle number concentration (TNC) by sulfur dioxide (SO_2) and nitrogen oxides (NO_x) concentrations using all available measurements in 2002. Contour lines shown for 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500, 600 thousand #/cm³.

there is not a clear distinction between high SO_2/low TNC and high SO_2/high TNC. For TNC, the figure shows that peak levels are much higher than background levels, but now the histograms are different for the groups with high and low SO_2. The high SO_2/high TNC group has higher TNC levels than the high TNC/low SO_2 group. Likewise, the high SO_2/low TNC group has higher TNC than the low SO_2/low TNC group. Fig. 4 also shows the histograms of surface NO_x concentrations for the four groups. These are very similar to each other although the high TNC/low SO_2 and high SO_2/low TNC groups are associated with higher NO_x levels. The bottom panel shows the histogram by time of day which reveals that high SO_2/high TNC events occur most frequently in the late morning whereas high TNC/low SO_2 events occur most frequently during the morning rush-hour. The high SO_2/low TNC events occur throughout the day but do have an increased incidence between midnight and sunrise.

Table 1 shows summary statistics for these two datasets by group in order to quantify the contribution of peak concentrations observed in Figs. 1 and 2. The average concentration of SO_2 is 6 ppb, and the average TNC is 42,000 #/cm³. This hides very large variations between the groups. The Low SO_2/ Low TNC group occurs 78% of the time and has a mean SO_2 concentration of 2.7 ppb and TNC of 31,000 #/cm³, thereby accounting for 35% of SO_2 and 58% of particles at the measurement sites. The high TNC/low SO_2 group occurs 7.6% of the time and accounts for 15% of TNC but only 5% of SO_2. The high SO_2/low TNC group occurs 9% of the time and accounts for 31% of SO_2 and 9% of TNC, and the high SO_2/high TNC group occurs 5% of the time and contributes 29% of the SO_2 and 18% of TNC.

To refine the analysis of TNC levels, we split the entire time series into a new set of groups that includes the impact of NO_x concentrations and time of day. As in the previous set of High/
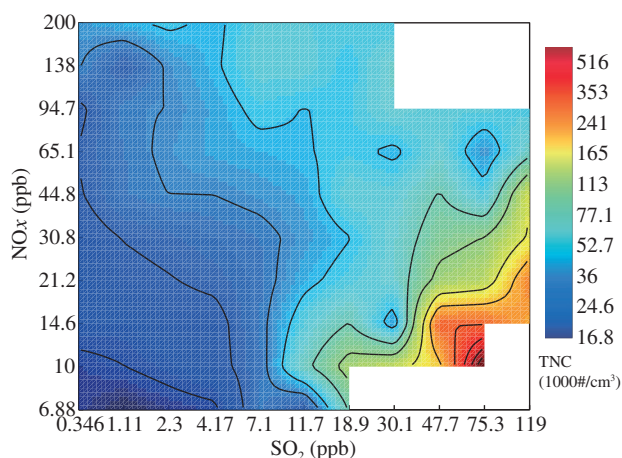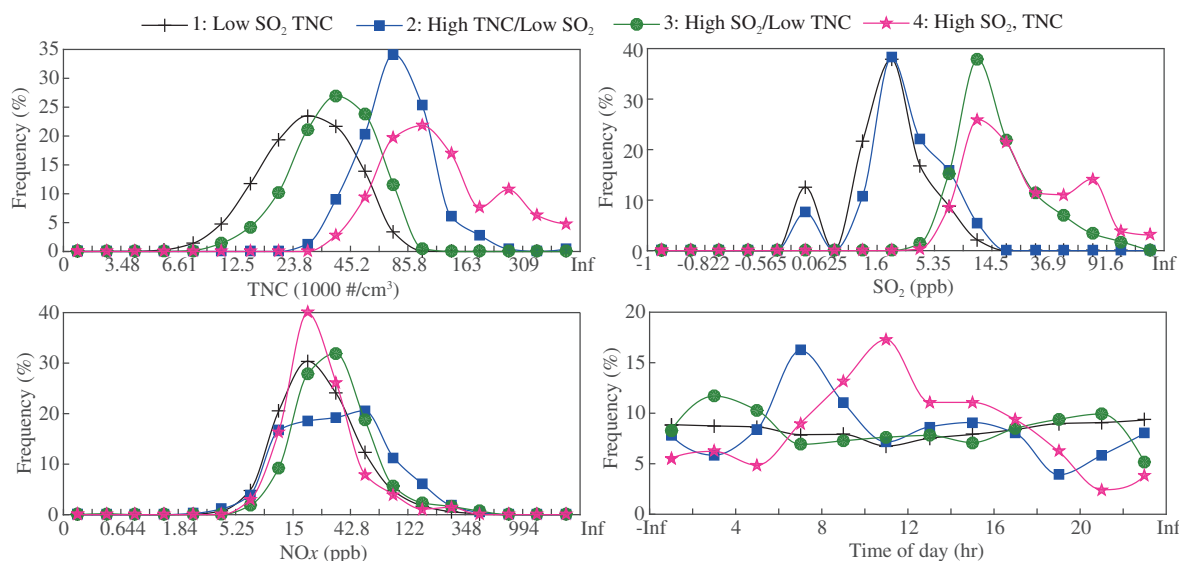
Fig. 4 – Histograms of total particle number concentration (TNC), sulfur dioxide (SO₂), nitrogen oxides (NOₓ) and time of day with colored symbols showing the four groups used in the analysis.

Table 1 – Number of data points, average concentrations and fraction of total contribution of total particle number concentration (TNC) and sulfur dioxide (SO₂) in each of the four groups discussed in the text.

Group	Data points (number)	Data fraction (%)	Mean TNC ($\#/cm^3$)	Mean SO₂ (ppb)	Fraction TNC (%)	Fraction SO₂ (%)
Low SO₂/low TNC	4757	78.3	30,952	2.7	57.8	34.9
HighTNC/low SO₂	462	7.6	82,616	3.6	15.0	4.6
High SO₂/low TNC	564	9.3	40,427	20.3	9.0	31.4
High SO₂/high TNC	289	4.8	160,775	36.6	18.2	29.1
All	6072	100.0	41,942	6.0	100.0	100.0

Low SO₂/TNC groups, these new groups do not overlap and together include all the data points. The four groups are: 1. High SO₂ (merger of high SO₂/low TNC and high SO₂/high TNC), 2. High NOₓ containing the times with non-peak SO₂ and higher than average NOₓ (concentrations in the top 50%, above 23 ppb), 3. Low SO₂, low NOₓ and clear skies (i.e., non-peak SO₂, NOₓ in the bottom 50% and clear skies at KCPS), and 4. Low SO2, low NOₓ and cloudy skies (1/8 to 8/8 cloud cover observed at KCPS). Fig. 5 shows the fraction of TNC contributed by each of the 4 groups for each hour of the day. While high SO₂ levels are always associated with higher than average TNC levels (see Fig. 4 and Table 1), the figure shows that the fraction of TNC associated with high SO₂ is higher during the day when there is solar radiation. The high NOₓ conditions contribute particle numbers mainly at rush hour and during the night. The cloudy group has a fairly uniform contribution throughout the day. This is in contrast to the clear sky group where the contribution increases during the day, which suggests that there is nucleation during times with high solar radiation.

We estimate the contribution of each group (high SO₂; high NOₓ; low SO₂, low NOₓ, clear skies; low SO₂, low NOₓ, cloudy skies) to the total TNC levels at the site. We calculate the mean TNC level for each group for each hour of the day. We then split the clear sky group into a background and a daytime excess by using the average TNC level from 9 pm to 9 am as a background, which came to 26,000 #/cm³. The daytime excess can be taken to represent the extra TNC levels due to
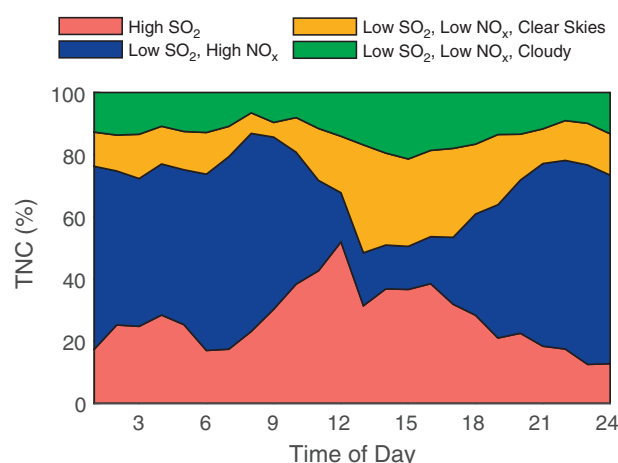

Fig. 5 – Contribution to total particle number concentration (TNC) at each hour of the day for four different conditions: peak sulfur dioxide (SO₂); non-peak SO₂ with high nitrogen oxides (NOₓ); non-peak SO₂ with low NOₓ and clear skies; and non-peak SO₂ with low NOₓ and cloudy skies. The extra TNC associated with SO₂ can be clearly seen during the day, the TNC associated with primary emissions can be seen at rush hour and at night. The regional nucleation events are most likely the ones contributing to the daytime bulge in TNC in the third group (low SO₂, low NOₓ, clear skies).

nucleation. Overall, this suggests that nucleation in the SO_2 plume accounted for 29% of TNC levels, emissions of fine particles associated with NO_x emissions, and hence with mobile sources, accounted for 40% of TNC levels, daytime nucleation contributed 8% of TNC and the remaining 23% could be considered as urban background.

2.2. Wind rose analysis

The analysis will now turn to wind roses to obtain more clues about the origin of SO_2 and TNC at the measurement site. The same technique was used to identify possible sources of nickel, vanadium and black carbon in Milwaukee (de Foy et al., 2012). Fig. 6 shows the wind roses colored by time of day for the four groups discussed above. The low SO_2/low TNC group (which occurred 78% of the time) has the fairly typical wind rose for this location as discussed in de Foy et al. (2014). There is wind transport from all directions, with a predominant southeasterly component and a more spread out northwesterly component. Calm winds occur 17% of the time. The high TNC/low SO_2 group is not that different, with higher TNC values associated with transport from the south and from the northwest. The striking difference is that high SO_2/low TNC and high SO_2/high TNC both have very strong signatures from the southwest. In the case of high SO_2/high TNC, these are exclusively from the southwest with a lower than average number of calms. The time of day difference can also be seen in these diagrams, with the high SO_2/low TNC group occurring
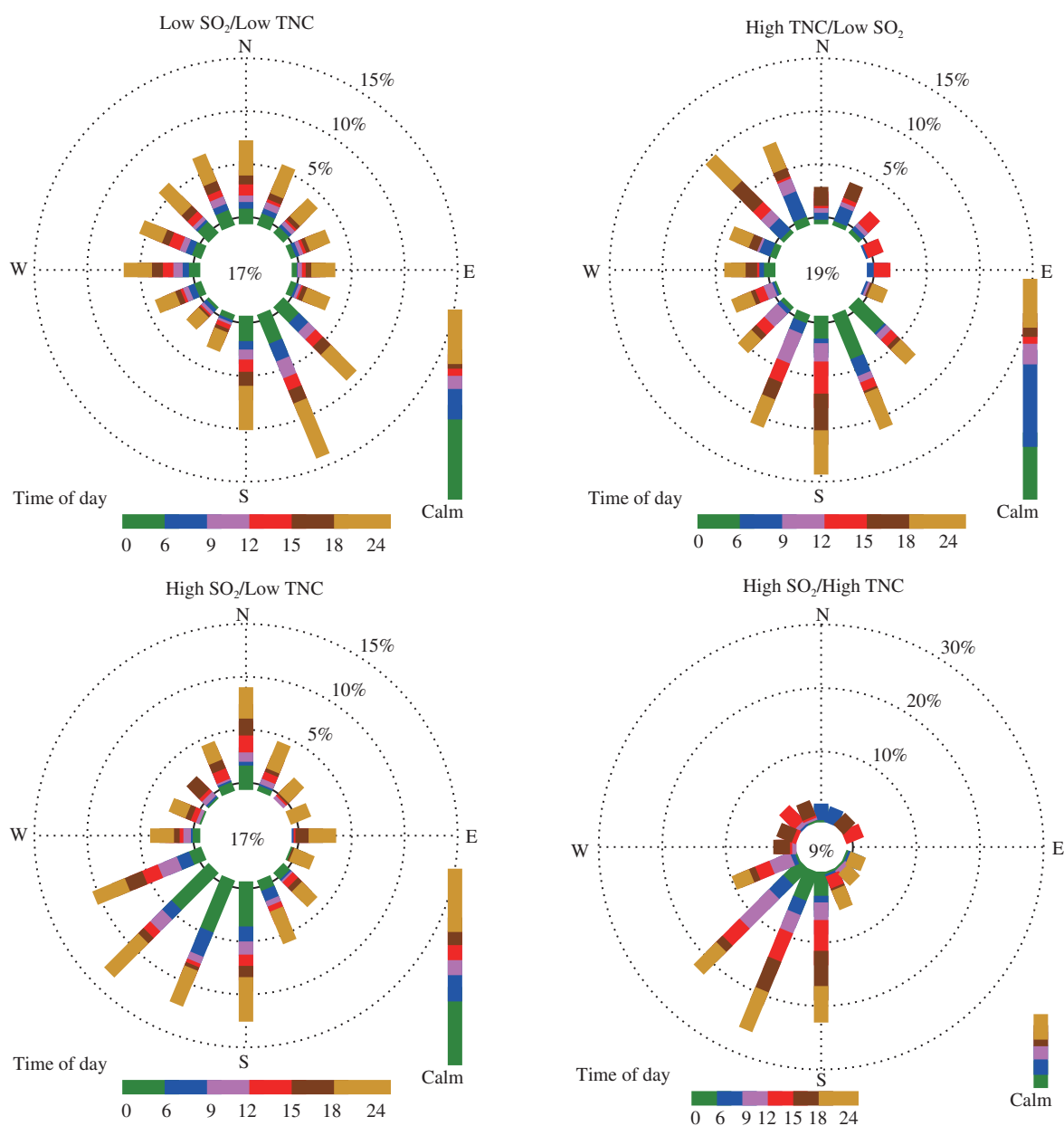

Fig. 6 – Wind roses showing wind direction and time of day for the four groups used in the analysis. Each bar is colored by time of day according to the legend at the bottom. The bar on the right shows the time of day when there are calms (wind speeds below 1.5 m/sec), the total percentage of calms is listed in the center of each rose.

predominantly at night and the high SO2/high TNC events occurring mostly during the day.

The SO_2 peaks at the supersite are from a point source that is 1.3 km away at a bearing of 215°. This is further corroborated by the long term time series of SO_2 that spans the closure of the source and by measurements from a separate site 2.3 km south of the supersite (see Supplementary information Figs. S1 and S2). Because this source is well characterized and identified, the results show with great confidence that there are specific SO_2 plumes that impact the site at all hours of the day, and that during the daytime there is particle nucleation in the plumes leading to elevated TNC levels.

2.3. Multiple regression model

To refine the analysis, a multiple regression model was used to estimate the impact of SO_2, NO_x, O_3, wind speed, temperature, relative humidity, boundary layer height (PBLH) and solar radiation (SR) on TNC levels. As described in Section 1.2, for TNC, SO_2, NO_x, O_3, PBLH and SR, we used the log of the values to obtain approximately normal distributions of the transformed variables. For WS, T, RH we used a linear rescaling of the values. In addition to performing the regression analysis for the

complete time series (the "All" group), we perform it for night time only ("Night", defined as clear sky radiation of 0 W/m^2) and for day time hours ("Day", defined as clear sky radiation greater than or equal to 100 W/m^2). We further carry out the regression for the outliers in the Day group that were rejected by the IRLS procedure ("Day-High"). Note that these are non-overlapping groups that account for most hours in the dataset except for the twilight hours (between 0 and 100 W/m^2). Finally, we also do the regression for the four groups based on SO_2 and TNC levels described in Section 2.1.

Fig. 7 shows the probability density plots for the group with All data points. Fig. 7a, b shows the distribution of the variables in transformed coordinates. Fig. 7c, d shows the distribution of the TNC measurements and model fit as well as the distribution of the residuals for the points retained in the regression model (*i.e.*, excluding the points that were removed by the IRLS). TNC levels follow closely the normal distribution in transformed coordinates. The model fit does too, but has a narrower distribution which indicates that it fails to simulate the levels at the extremes, both for high and low values. The residuals are normally distributed and in addition were uncorrelated with the input variables. Fig. 8 shows a scattergram of the TNC levels and the model output
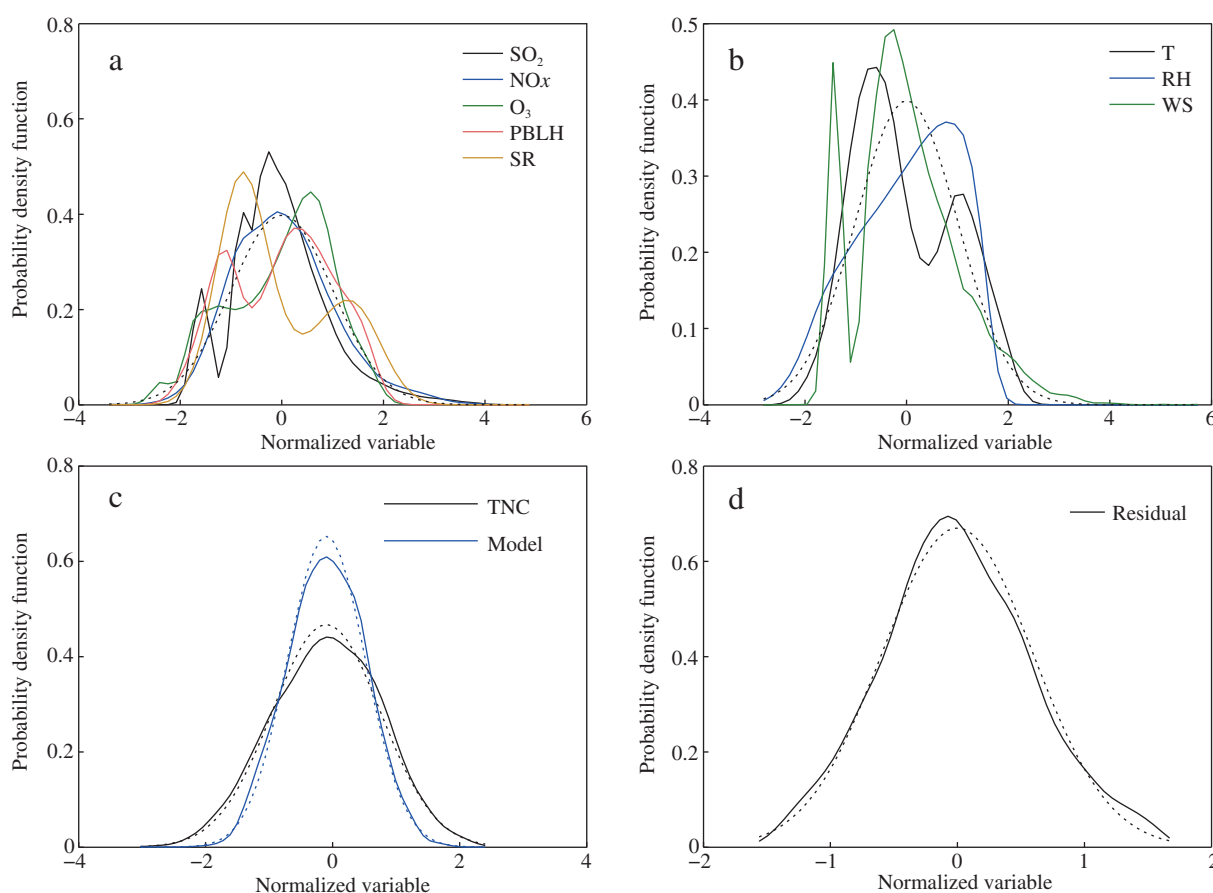

Fig. 7 – Probability density distributions for the regression model using all available data ("All group"). Top left shows the log transformed variables. (a) Log transformed variables; (b) linearly transformed variables; (c) distribution of the TNC measurements and of the regression model simulations in transformed space; (d) distribution of the residual. Dotted lines in (a) and (b) show the normal distribution with zero mean and standard deviation of one, dotted lines in (c) and (d) show the best-fit normal distributions for the corresponding variables.
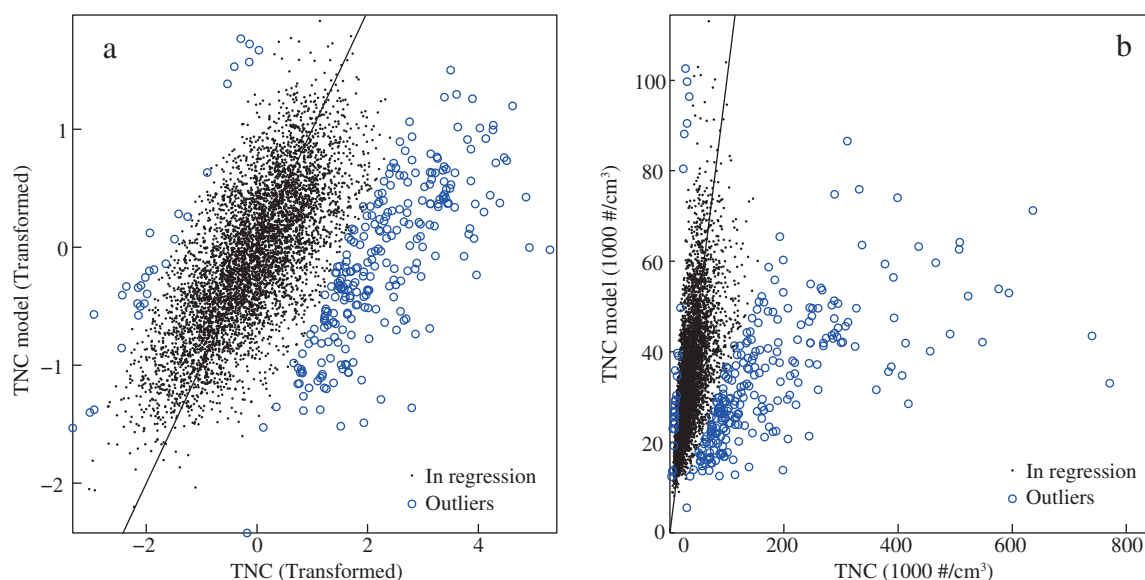
Fig. 8 – Scatter plot of regression model *versus* total particle number concentration (TNC) measurements in transformed space (a) and in original units of #/cm³ (b) for the All group. Black points show data used in the regression, blue circles show data points rejected by the Iteratively Reweighted Least Squares procedure, the lines show where the model equals the measurements.

in transformed space as well as in the original units (#/cm³). Table 2 presents statistics and details of the multiple regression fit: there are 5351 data points included in the fit and 283 outliers removed by the IRLS procedure, Pearson's correlation coefficient squared (r^2) was 0.51 without the outliers and 0.38 including the outliers. Average TNC levels were 36,004 #/cm³ in the data and 33,583 #/cm³ in the model fit. The Root Mean Square Error of the fit was 44,000 #/cm³ with the outliers and 16,000 #/cm³ without them, and the coefficient of variation of the fit was 0.4 compared with 0.6 for the data (excluding outliers). The outliers accounted for 5% of the data points and 19% of the total particle count.

To evaluate the uncertainty of the results we performed 100 realizations of the inversion using block-bootstrapping with a block length of 24 hr. There were 271 days with data. We therefore select 271 days at random with replacement for inclusion in the analysis. By randomly selecting days we test the sensitivity of the results to measurement errors as well as to the selection of pollution events. Fig. 9 shows the scatterplot and histograms of the regression coefficients produced by the 100 realizations of the least squares inversion. The histograms give an indication of the uncertainty in the coefficients, and the scatterplots show that there is limited covariance between the input variables. r^2 values

Table 2 – Results from the multiple regression model for TNC for the All group showing the regression coefficients, least squares standard error, uncertainty estimate based on 100 block-bootstrapped realizations and the influence relation of each parameter in the regression. Also shown are statistical metrics of the fit.

Variable	Regression coefficient	Standard error	STD[a] bootstrap	Influence relation
NO_x ppb	0.439	0.0122	0.031	NO_x^0.44
Solar radiation (W/m²)	0.051	0.0039	0.026	(SR + 10)^0.051
SO_2 (ppb)	0.093	0.0068	0.022	(SO_2 + 1)^0.093
PBLH[b](m)	0.042	0.0068	0.011	PBLH^0.042
O_3 (ppb)	−0.026	0.0095	0.021	(O_3 + 1)^−0.026
Wind speed (m/sec)	0.044	0.0032	0.008	exp(0.044*WS)
Temperature (°C)	−0.011	0.0006	0.002	exp(−0.011*T)
Relative humidity (%)	−0.008	0.0004	0.002	exp(−0.008*RH)
Statistics				
r^2	0.51		Mean	Bootstrap STD
No. Obs (#/cm³)	5351	Data	36004	971
No. outliers (#/cm³)	283	Model	33583	895

[a] STD: standard deviation.
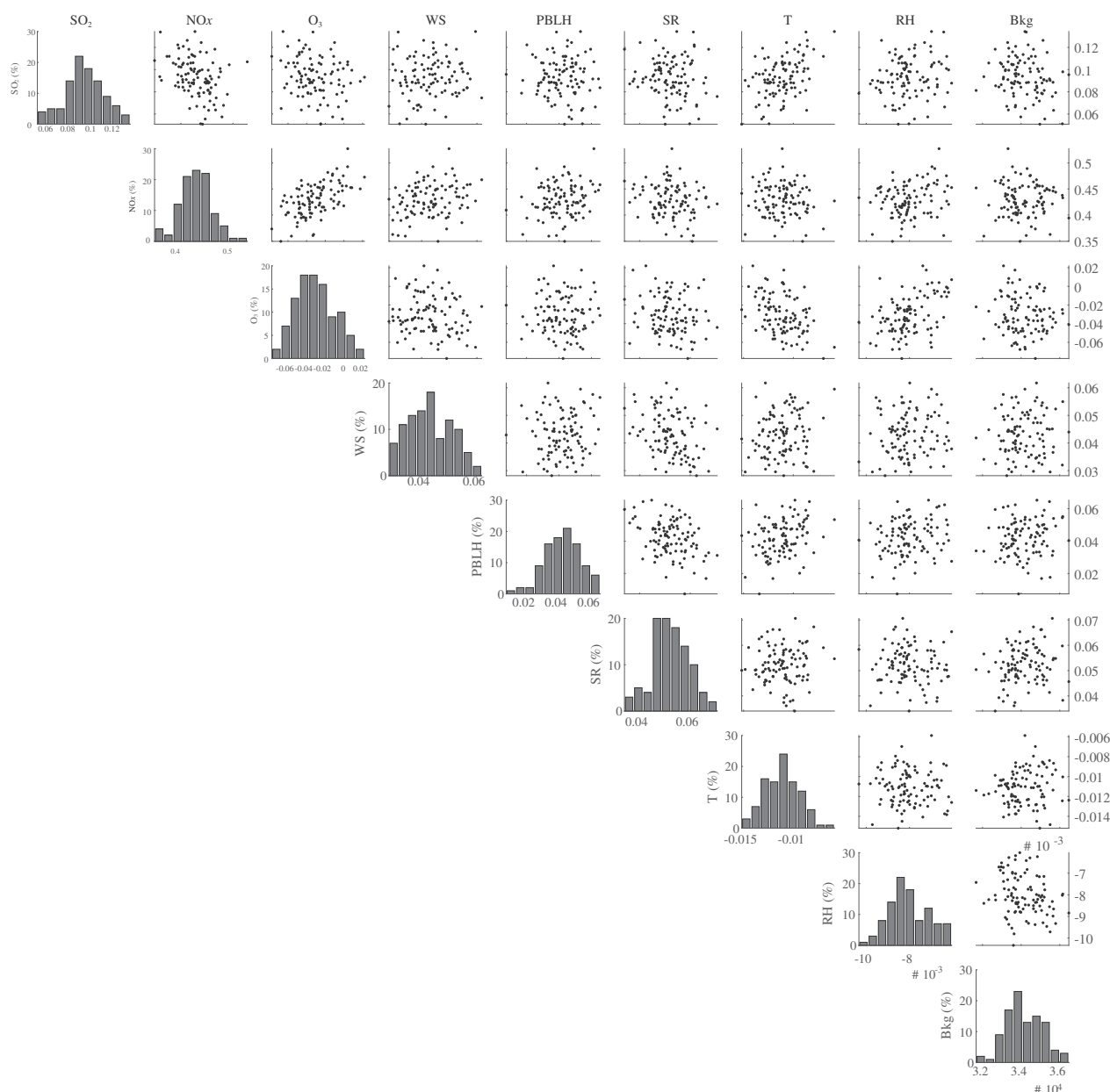[b] PBLH: planetary boundary layer height.

Fig. 9 – Bootstrapped estimates of uncertainties in the coefficients of the input variables for the regression model of TNC based on 100 realizations of the regression. Histograms show the uncertainties in each coefficient. Scatterplots show the covariance between the coefficients for different variables.

are lower than 0.05 for most pairs of coefficients. The largest exception is the correlation of O_3 with NO_x ($r^2 = 0.33$), T (0.18) and RH (0.18). The uncertainty in the regression for O_3 is the largest of all the input parameters, and this would be the first variable to be removed from the analysis. Sensitivity tests were performed by excluding O_3 and separately by excluding NO_x. As can be seen by comparing Figs. S5 and S6 with Fig. S4, keeping both O_3 and NO_x in the regression does not change the results of the analysis. Furthermore, O_3 is an indicator of photochemical activity and hence possibly of secondary production. We therefore decided to keep it in the final model.

2.4. Multiple regression model coefficients

For the variables that were log transformed, the regression coefficients correspond to the exponent of the variable in the fit as shown in Eq. (1). For the linearly transformed variables, the impact of the variable on the TNC model is the exponential of the parameter multiplied by the variable (see also Eq. (1)). These values are shown in Table 2 along with the standard error calculated by the least squares procedure and the standard deviation of the 100 bootstrapped realizations of the model. From these it can be seen that all coefficients are statistically highly significant. Note also that the uncertainties

estimated by the bootstrapping are much larger than the ones based on the standard error of the least squares inversion. We believe that the least square standard errors are underestimated and that using block-bootstrapping gives a more robust estimate of the uncertainties in the regression model.

The regression coefficients in Table 2 can be represented graphically by plotting the entire term in the table *versus* the values of the input variable, as shown in Fig. 10. For each of the 8 groups, the figure shows the impact of changing levels of the input variable (SO_2, NO_x *etc.*) to the TNC levels in the regression model. This yields an influence relation for each parameter. Note that the influence relations were rescaled to have an average value of one. For SO_2, for the All group, this figure shows that the highest SO_2 levels (above 300 ppb) correspond to TNC values that are around 1.4 times the average. Considering just the Day and high SO_2/high TNC groups, the high SO_2 events lead to TNC values that are double the average. The relationship between SO_2 and TNC is lowest for the low SO_2/low TNC group and for the high SO_2/low TNC group.

Because of the importance of SO_2 for particle nucleation, we expect that higher SO_2 concentrations should be associated with higher TNC levels. This is clearly seen for the high SO_2/high TNC, the Day and the Day-High groups which confirms that SO_2 during the day leads to high particle numbers. The lower regression coefficients, and hence weaker relationship between SO_2 and TNC, at Night and for the high SO_2/low TNC group confirm that solar radiation is needed for particle nucleation to take place and that the SO_2

plumes occurring at night do not have elevated TNC levels (see also Fig. 4).

The regression coefficients of NO_x in the model fit are larger than those for SO_2 and also more variable by group. High NO_x levels lead to a factor of 3 higher TNC levels for the Night and for the low SO_2/low TNC groups. For the Day and for the high SO_2/low TNC groups, the impact of NO_x on TNC levels is reduced to 1.5 to 2. The relationship changes sign for the high TNC/Low SO_2 group, with higher NO_x leading to lower TNC values. NO_x levels at the supersite are indicative of primary mobile source emissions (de Foy et al., 2014). This suggests that the large factors at night especially are indicative of emissions of fine particles from the same mobile sources. The positive relationship for the high SO_2 events suggests that NO_x in the atmosphere may be enhancing particle nucleation in the plumes. The negative relationship for the high TNC/low SO_2 group was more surprising as this would have been a good candidate for increased TNC levels with higher NO_x concentrations given that a significant fraction of TNC levels are associated with mobile source emissions. In addition to experiencing reduced TNC with higher NO_x, the high TNC/low SO_2 group is also the only group for which TNC levels decrease as wind speed increases. An additional element to consider is the role of scavenging in reducing TNC levels in polluted air masses. It is possible that for this group there are higher particle numbers for relatively cleaner and more stagnant air, which would be in agreement with regional nucleation events of the type described in Qian et al. (2007).

There is a very clear signal of decreasing levels of particle number with increasing temperature and relative humidity in
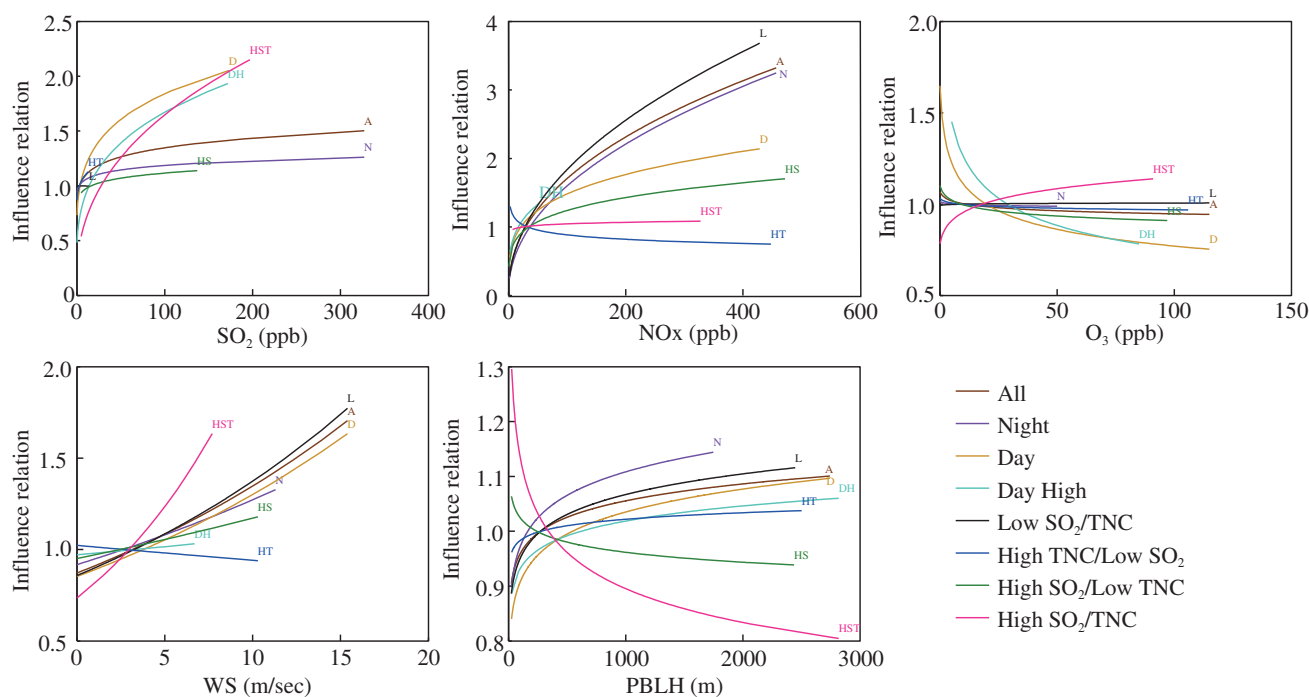

Fig. 10 – Influence relation of variables included in the multiple regression analysis for total particle number concentration (TNC) for the different groups outlined in the text. See Table 1 and Fig. 1 for group descriptions, see Fig. S3 for the influence relation of temperature, relative humidity and solar radiation.

the regression coefficients in Table 2 and for all the groups from the extra panels in Fig. S3 of the Supplementary information. In the model, TNC levels are always lower when the temperature is warmer, with a factor of two differences between the minimum and maximum temperatures. The impact of relative humidity is of the same magnitude as that of temperature, with no clear pattern in the different coefficients in each group. More mechanistic studies would be required to explain the drivers of these relationships.

The relationship between shortwave downward radiation at the surface was explored using results from the WRF model because measurements were not readily available. The influence relation is shown in Fig. S3 of the Supplementary information. As expected, higher solar radiation is associated with higher TNC levels for most groups, although the strength of the effect is variable. The groups with high SO_2 (high SO_2/low TNC and high SO_2/high TNC) are particularly susceptible to elevated TNC with more solar radiation. The daytime relationship is weak however. This could be related to the fact that TNC is more strongly related to NO_x than to SO_2 for these groups, which adds evidence to the suggestion that the NO_x relationship is indicative of primary emissions whereas the SO_2 relationship is indicative of nucleation in the plume.

The relationship of TNC to O_3 levels is one of the least clear-cut of the variables retained in the analysis. The scaling factors are close to one for most groups. The high SO_2/high TNC group has increased TNC levels with increased ozone, which could be expected from ozone as an indicator of photochemical activity. For the Day and Day-High groups however the relationship is reversed, with lower TNC linked to higher ozone concentrations. This runs counter to the idea that more particles would be associated with the more polluted and more photochemically reactive conditions signaled by higher O_3 concentrations. We performed a sensitivity test without NO_x in the regression, and found that higher O_3 concentrations are still associated with reduced TNC levels. This could be related to scavenging in polluted plumes or to multi-variable interactions which we cannot resolve at this time.

There is a very clear positive relationship between higher wind speeds and higher TNC levels for most of the groups. This is particularly strong for the High SO_2/High TNC group which consists of the large SO_2 and TNC peaks from the point source close-by. This suggests that higher TNC levels are associated with rapid, direct transport from the source rather than from longer transport times or from stagnant conditions. More research would be needed to identify the mechanism leading to this relationship. There is also the possibility that faster winds bring cleaner air masses which would have reduced condensation sinks and hence the potential for elevated TNC. Future research would need to consider the condensation sink in detail, as noted in Qian et al. (2007). In contrast, the Day-High and high TNC/low SO_2 groups have a very weak relation between TNC and WS suggesting that for these cases stagnation does enhance particle numbers.

Finally the relationship of TNC to the planetary boundary layer height, taken from the WRF model, is highly variable. For the high SO_2/low TNC and the high SO_2/high TNC groups (which are predominantly during the day), higher boundary layer heights lead to reduced TNC levels. The regression model already has the strongest positive relationship with solar radiation for these two groups, suggesting that aside from the impact of solar radiation when there is more mixing, there is more plume dilution and hence lower TNC. For all the other groups, higher PBLH is associated with higher TNC. This is especially true at night and for the low SO_2/low TNC group. Further research would be needed to determine the reasons for this, for example whether it is related to changes in the level of condensation sinks.

On a last note, the high SO_2/high TNC and Day-High groups contain some of the largest peaks and consequently could be expected to have similar coefficients in the regression model. This is true for SO_2 but not for O_3 where the coefficients have opposite signs, and not for solar radiation and PBLH which have a minor impact on the Day-High group but a large impact on the high SO_2/high TNC group. Although these two groups share some members, the Day-High group contains the times with the largest residuals in the regression model for the Day group. Because some of the largest TNC peaks are related to SO_2 peaks, these points do not have large residuals. Instead, the Day-High group has members in common with the high TNC/low SO_2 group which has TNC peaks that are not easily accounted for by SO_2. This further reinforces the conclusion that high SO_2/high TNC events are related to direct SO_2 plume hits from the point source, whereas Day-High events are due to a different source.

2.5. Multiple regression model for other species

To further explore the influence relationships, the regression analysis was carried out for different pollutants: $PM_{2.5}$, EC, OC and NO_x. Fig. 11 shows the influence for the All group for the 4 species along with TNC (see Fig. S4 for the complete set of species). This figure shows that TNC is the species most strongly associated with SO_2, but that higher SO_2 also leads to higher NO_x, OC and $PM_{2.5}$ concentrations whereas EC concentrations are not linked to SO_2. In contrast, NO_x is nearly the opposite. The strongest association is between NO_x and EC, a further indication that they share common sources. The relationship is progressively weaker for OC, TNC and $PM_{2.5}$.

One of the most surprising findings from Fig. 10 was the increase in TNC associated with both higher wind speeds and higher mixing heights. The regression model for the other species in Fig. 11 shows that these have lower concentrations associated with higher wind speeds and higher mixing, as could be expected from a simple box model view of dispersion: more dilution leads to lower concentrations. This clearly shows that there is a different dynamic taking place for TNC which should be explored in future studies, for example by measuring the condensation sinks.

The influence relationship shows that higher O_3 is strongly associated with lower NO_x, which could be a confounding factor in the relationship of O_3 and TNC discussed above. We see that TNC had the strongest negative relationship with temperature and relative humidity, and that these are the opposite of the relationships for $PM_{2.5}$ and OC, both of which have higher concentrations with higher temperature and relative humidity.
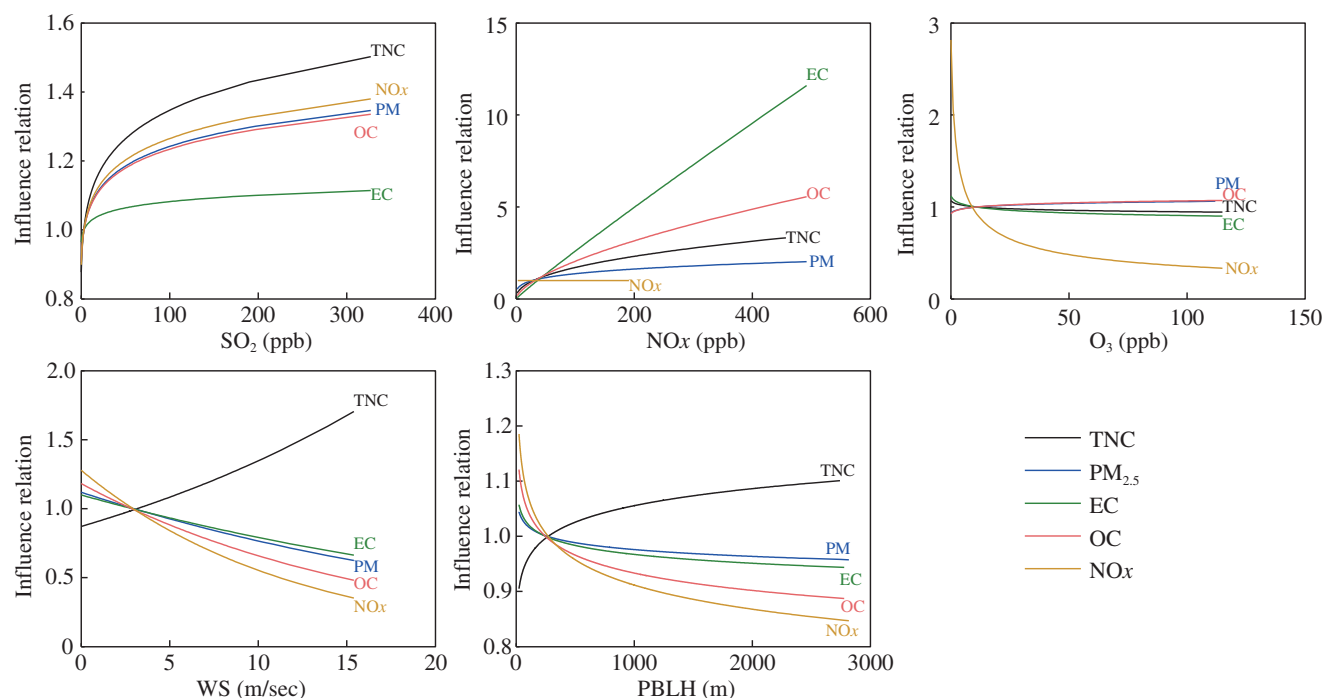
Fig. 11 – Influence relation of variables included in the multiple regression analysis for the All group for TNC, PM$_{2.5}$, elemental carbon (EC), organic carbon (OC) and NO$_x$. See Fig. S4 for the influence relation of temperature, relative humidity and solar radiation.

3. Conclusions and implications

Data analysis and multiple regression models were used to study the origin of particle numbers at the St. Louis, Midwest supersite based on a year-long time series of hourly measurements made in 2002. This showed that at that time the site was impacted by an industrial SO$_2$ plume accounting for a quarter to a third of particle numbers at the site. Between a third and a half of particles are likely associated with primary emissions from mobile sources. Of the remainder, up to a tenth is suggested to result from regional nucleation events leaving around a quarter of the particles due to the general urban background. Although the particular SO$_2$ source is no longer in operation and hence the conditions at the supersite have changed, the analysis emphasizes that local conditions can have significant impacts on particle number concentrations.

The tools presented in this study can be applied to future studies looking at nucleation events and high particle number concentrations, and would be valuable in helping to better understand the emissions and the meteorological conditions that lead to high particle number concentrations. Such understanding is needed to develop mitigation strategies to reduce particle number concentrations and elevated adverse impacts on human health and climate change due to high particle number concentrations. We would not expect that the results of the current study can be generalizable to other locations but the methods presented in this study could be readily applied to other study locations to develop a better understanding of sources and dynamics of high particle number concentrations and nucleation events.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jes.2014.12.026.

REFERENCES

Bae, M.S., Schauer, J.J., DeMinter, J.T., Turner, J.R., Smith, D., Cary, R.A., 2004. Validation of a semi-continuous instrument for elemental carbon and organic carbon using a thermal-optical method. Atmos. Environ. 38 (18), 2885–2893.

Bycenkiene, S., Plauskaite, K., Dudoitis, V., Ulevicius, V., 2014. Urban background levels of particle number

concentration and sources in Vilnius, Lithuania. Atmos. Res. 143, 279–292.

de Foy, B., Smyth, A.M., Thompson, S.L., Gross, D.S., Olson, M.R., Sager, N., et al., 2012. Sources of nickel, vanadium and black carbon in aerosols in Milwaukee. Atmos. Environ. 59, 294–301.

de Foy, B., Cui, Y.Y., Schauer, J.J., Janssen, M., Turner, J.R., Wiedinmyer, C., 2014. Estimating sources of elemental and organic carbon and their temporal emission patterns using a least squares inverse model and hourly measurements from the St. Louis—Midwest Supersite. Atmos. Chem. Phys. Discuss. 14, 12019–12070.

Donateo, A., Gregoris, E., Gambaro, A., Merico, E., Giua, R., Nocioni, A., et al., 2014. Contribution of harbour activities and ship traffic to PM2.5, particle number concentrations and PAHs in a port city of the Mediterranean Sea (Italy). Environ. Sci. Pollut. Res. 21 (15), 9415–9429.

Hussein, T., Mølgaard, B., Hannuniemi, H., Martikainen, J., Järvi, L., Wegner, T., et al., 2014. Fingerprints of the urban particle number size distribution in Helsinki, Finland: local versus regional characteristics. Boreal Environ. Res. 19, 1–20.

Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V.M., et al., 2004. Formation and growth rates of ultrafine atmospheric particles: a review of observations. J. Aerosol Sci. 35 (2), 143–176.

Kumar, P., Morawska, L., 2014. Recycling concrete: an undiscovered source of ultrafine particles. Atmos. Environ. 90, 51–58.

Kumar, P., Robins, A., Vardoulakis, S., Quincey, P., 2011. Technical challenges in tackling regulatory concerns for urban atmospheric nanoparticles. Particuology 9 (6), 566–571.

Kumar, P., Morawska, L., Birmili, W., Paasonen, P., Hu, M., Kulmala, M., et al., 2014. Ultrafine particles in cities. Environ. Int. 66, 1–10.

Kuwayama, T., Ruehl, C.R., Kleeman, M.J., 2013. Daily trends and source apportionment of ultrafine particulate mass (PM0.1) over an annual cycle in a typical California City. Environ. Sci. Technol. 47 (24), 13957–13966.

Oberdörster, G., 2001. Pulmonary effects of inhaled ultrafine particles. Int. Arch. Occup. Environ. Health 74 (1), 1–8.

Pierce, J.R., Adams, P.J., 2009. Uncertainty in global CCN concentrations from uncertain aerosol nucleation and primary emission rates. Atmos. Chem. Phys. 9, 1339–1356.

Qian, S., Sakurai, H., McMurry, P.H., 2007. Characteristics of regional nucleation events in urban East St. Louis. Atmos. Environ. 41 (19), 4119–4127.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., et al., 2005. A Description of the Advanced Research WRF Version 2. NCAR.

Wang, Z.B., Hu, M., Wu, Z.J., Yue, D.L., 2013. Research on the formation mechanisms of new particles in the atmosphere. Acta Chim. Sin. 71 (4), 519–527.

Wu, Z.J., Hu, M., Liu, S., Wehner, B., Bauer, S., ßling, A.M., et al., 2007. New particle formation in Beijing, China: statistical analysis of a 1-year data set. J. Geophys. Res.-Atmos. 112 (D9). http://dx.doi.org/10.1029/2006JD007406.

Zhu, Y.F., Hinds, W.C., Kim, S., Sioutas, C., 2002. Concentration and size distribution of ultrafine particles near a major highway. J. Air Waste Manag. Assoc. 52 (9), 1032–1042.