Supplementary Information for

"Importance of Biogenic Volatile Organic Compounds to Peroxyacyl Nitrates (PANs) Production in the Southeastern U.S. during SOAS 2013"

S. Toma et al.

Correspondence to: S. Bertman (steven.bertman@wmich.edu)

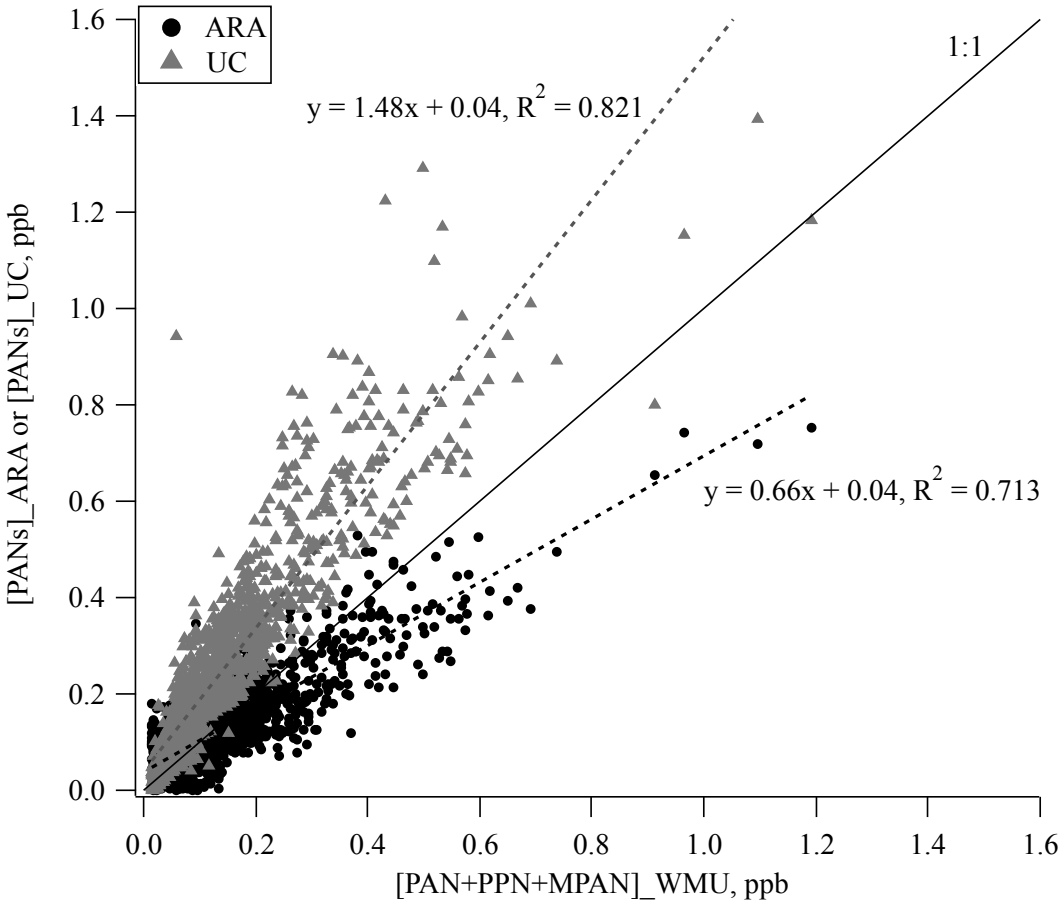

Figure S1. Relationship of total PANs with other research groups during SOAS 2013 campaign. The means with standard deviation of PANs from ARA, UC, and WMU were 0.129±0.092, 0.245±0.194, and 0.138±0.119 ppb respectively. The medians of PANs from ARA, UC, and WMU were 0.111, 0.204, and 0.103 ppb respectively.

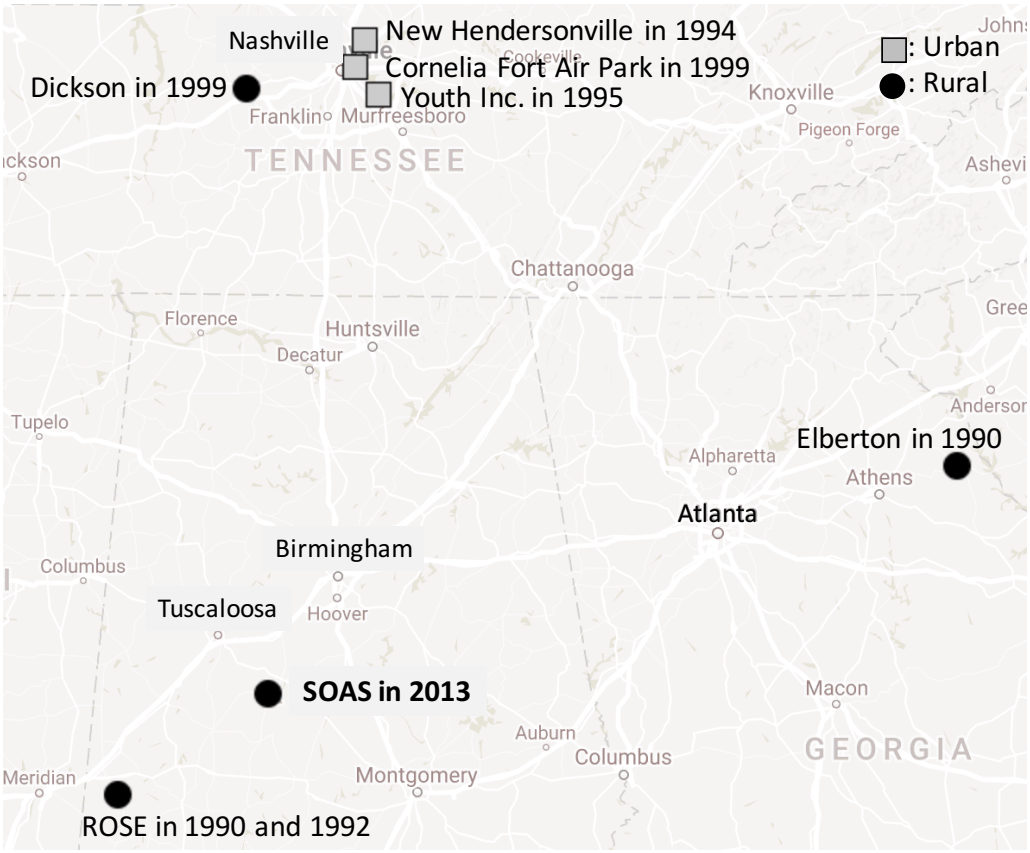

Figure S2. Various Locations of Other Measurements in the Southeastern U.S. since 1990. Sampling term: Elberton (June 24th – July 13 in 1990), ROSE 1990 (June 10th – July 20th in 1990), ROSE 1992 (June 19th – July 2nd in 1992), Henderson (June 22nd – July 19th in 1994), Youth Inc. (June 29th – July 26th in 1995), Dickson (June 13th – July 15th in 1999), Cornelia Fort Ground Site (June 14th – July 14th in 1999), and SOAS 2013 (June 1st – July 15th in 2013). (*Data map: Google map, 2016).*

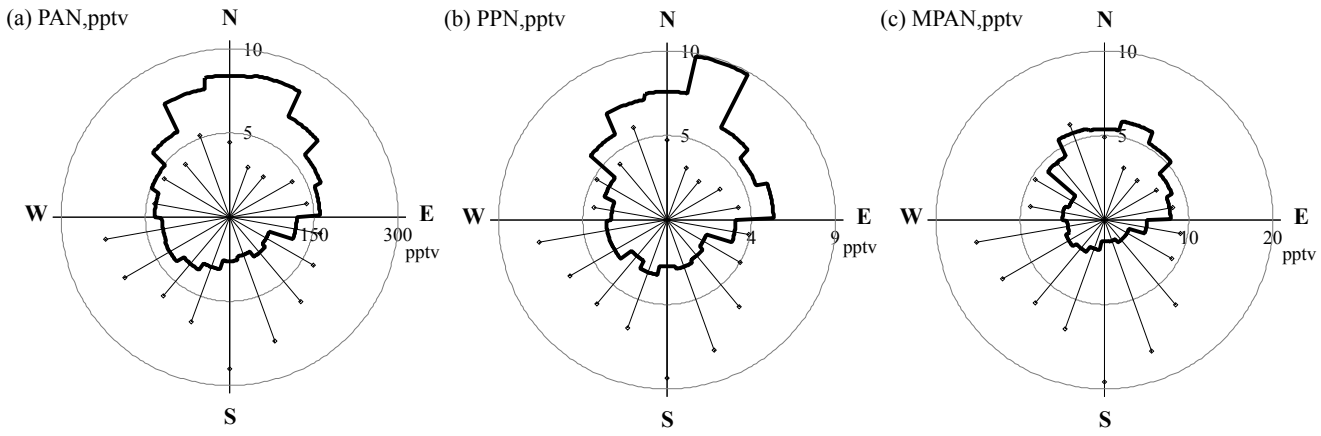

Figure S3. Polar plots of PANs concentrations as a function of wind direction in SOAS 2013 (excluded June 4th). The bold trace line in each plot indicates the average concentration of a PAN compound and the solid lines from the center are the frequency of wind direction.

Methods and Results of MLR analysis for PANs

$$[PAN] = A + B_1[MPAN] + B_2[PPN] \quad (1)$$

In a multiple linear regression (MLR) model as Equation (1), [PAN] is treated as a response variable and [MPAN] and [PPN] are used as independent predictor variables. B_1 and B_2 are partial regression coefficients on [MPAN] and [PPN]. The MLR statistical analysis conducted two steps of statistical testing. First, the *F*-test with ANOVA and R^2 investigated how well the model Eq. (1) fits the measurement data. However, *F*-test is impossible to directly find out which predictor variable is significantly useful. Therefore, in the next step, the significant utility of each partial regression coefficient was explored using the Student's *t*-test. The respective *t*-value was calculated from each partial regression coefficient divided by the standard error. When results of the *t*-test indicate presence of statistical significance for the partial regression coefficients, the magnitude of the standardized partial regression coefficient, β_i, allows us to compare the relative contribution of each independent predictor variable within the model.

As the notice to conduct MLR statistical analysis, high multicollinearity causes effects on the results of the analysis (e.g. Mendenhall et al., 2009). Although the assumption of the MLR statistical analysis on [PAN] takes a stance that each predictor variable is derived from different hydrocarbon precursor independently, the values of "tolerance" or "variance inflation factor (VIF)" were helpful to

assess the impact of the multicollinearity. The tolerance is calculated as 1- $R^2_{MPAN-PPN}$, where $R^2_{MPAN-PPN}$ is the coefficient of determination between MPAN and PPN and VIF is 1/tolerance. Large VIF value indicates strong multicollinearity of predictor variables. According to Stevens (2012), if the value of VIF is greater than 10, it indicates effective multicollinearity.

The statistical analysis was conducted using SPSS statistics software (versions 16, IBM). Results of *F*-test and R^2 on the MLR model for SOAS 2013 during the daytime are summarized in Table S1. Similar PANs data collected from Dickson, TN during the SOS experiment in 1999 is used as a comparable reference. The small *p*-value (*P* in Table S1) of *F*-test indicated that the overall fit of the model Eq. (1) is statistically significant in both the SOAS 2013 and Dickson 1999, and at least one independent predictor variable was significantly useful.

Table S1. Summary of the *F*-test and R^2.

Year	Number of data	*P* of *F*-value	*R*	R^2
Dickson, TN in 1999	486	<0.001	0.876	0.766
SOAS 2013	498	<0.001	0.775	0.601

A summary of coefficients of MPAN and PPN in both SOAS 2013 and Dickson 1999 is shown in Table S2. Since all VIF values were less than 10, there was no impact of multicollinearity in the MLR statistical analysis in both SOAS 2013 and Dickson 1999. The small *p*-value (*P* in Table S2) of the *t*-test of both MPAN and PPN in SOAS 2013 and Dickson 1999 indicates both predictor variables were useful to predict PAN. Therefore, respective partial regression coefficient values were available to estimate PAN in SOAS 2013 and Dickson 1999.

Table S2. Summary of coefficients on each independent predictor variable in *t*- test.

	Dickson, TN in 1999		SOAS 2013	
	MPAN	PPN	MPAN	PPN
Partial regression coefficient	B_1 5.098	B_2 5.762	B_1 7.596	B_2 6.910
Std. error of coefficient	0.305	0.178	0.469	0.725
P of *t*-test	<0.001	<0.001	<0.001	<0.001
VIF	1.036	1.036	1.427	1.427
β_i	0.374	0.725	0.549	0.323
r_i	0.509	0.795	0.726	0.624
Partial $R^2 = \beta_i r_i$	0.190	0.576	0.399	0.202

Std. error of coefficient means standard error of partial regression coefficient. *P* is statistical significant level. β_i is standardized partial regression coefficient. r_i is zero-order correlation. All dataset was during the daytime, 10 am – 4 pm.