Generalized Additive Modeling to Characterize $PM_{2.5}$ Behavior in California

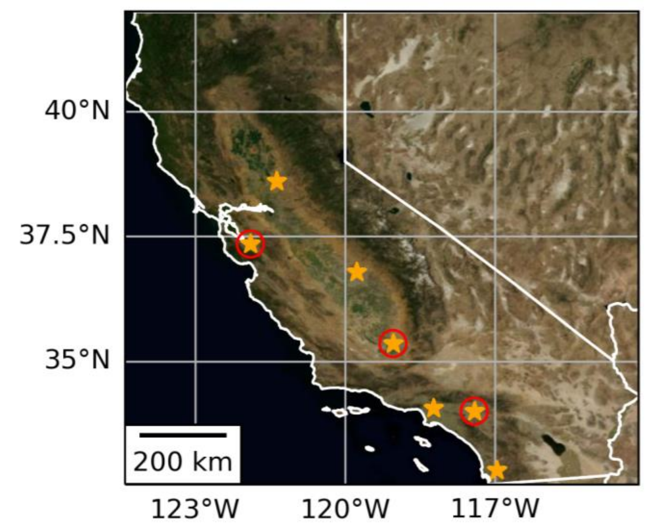
Duncan Quevedo[1], Ziqi Gao[2], & Cesunica Ivey[1]

[1]Department of Civil & Environmental Engineering, University of California, Berkeley; [2]School of Civil & Environmental Engineering, Georgia Institute of Technology

Objective

Despite declining emissions, much of California remains in nonattainment of $PM_{2.5}$ NAAQS. To understand why, we characterize historical $PM_{2.5}$ behavior by:

- Identifying driving variables of speciated $PM_{2.5}$ in California
- Modeling the response of speciated $PM_{2.5}$ to those driving variables



Methods

The general form for a generalized additive model (GAM) is

$$g(\mu) = \beta_0 + \sum_i f_i(x_i) + \varepsilon_i \quad (1)$$

Where g is the link function, μ is the expected value of the target variable, β_0 is the model intercept, f_i are the smooth functions to fit to the covariates x_i, and ε_i is the model error. For our model specification, we choose the log link and our covariates are detailed in Table 1.

Table 1. Summary of terms in (2).

Covariate	Fit	Description	Units
TMAX	Penalized cubic regression spline	Daily maximum temperature	°C
AWND	Penalized cubic regression spline	Daily average wind speed	m s^{-1}
RH_{sfc}	Penalized cubic regression spline	Daily maximum surface relative humidity	%
SR	Penalized cubic regression spline	Daily maximum solar radiation	W m^{-2}
U_{850mb}	Penalized cubic regression spline	Daily average 850 mb U component of wind (east-west)	m s^{-1}
V_{850mb}	Penalized cubic regression spline	Daily average 850 mb V component of wind (north-south)	m s^{-1}
RH_{850mb}	Penalized cubic regression spline	Daily average 850 mb relative humidity	%
eNO_X	Penalized cubic regression spline	Anthropogenic NO_X emissions	tons day^{-1}
eROG	Penalized cubic regression spline	Anthropogenic ROG emissions	tons day^{-1}
eSO_X	Penalized cubic regression spline	Anthropogenic SO_X emissions	tons day^{-1}
eNH_3	Penalized cubic regression spline	Anthropogenic NH_3 emissions	tons day^{-1}
$ePM_{2.5}$	Penalized cubic regression spline	Anthropogenic $PM_{2.5}$ emissions	tons day^{-1}
ONI	Penalized cubic regression spline	Oceanic Niño Index	°C
Season	Factor variable	Season	N/A

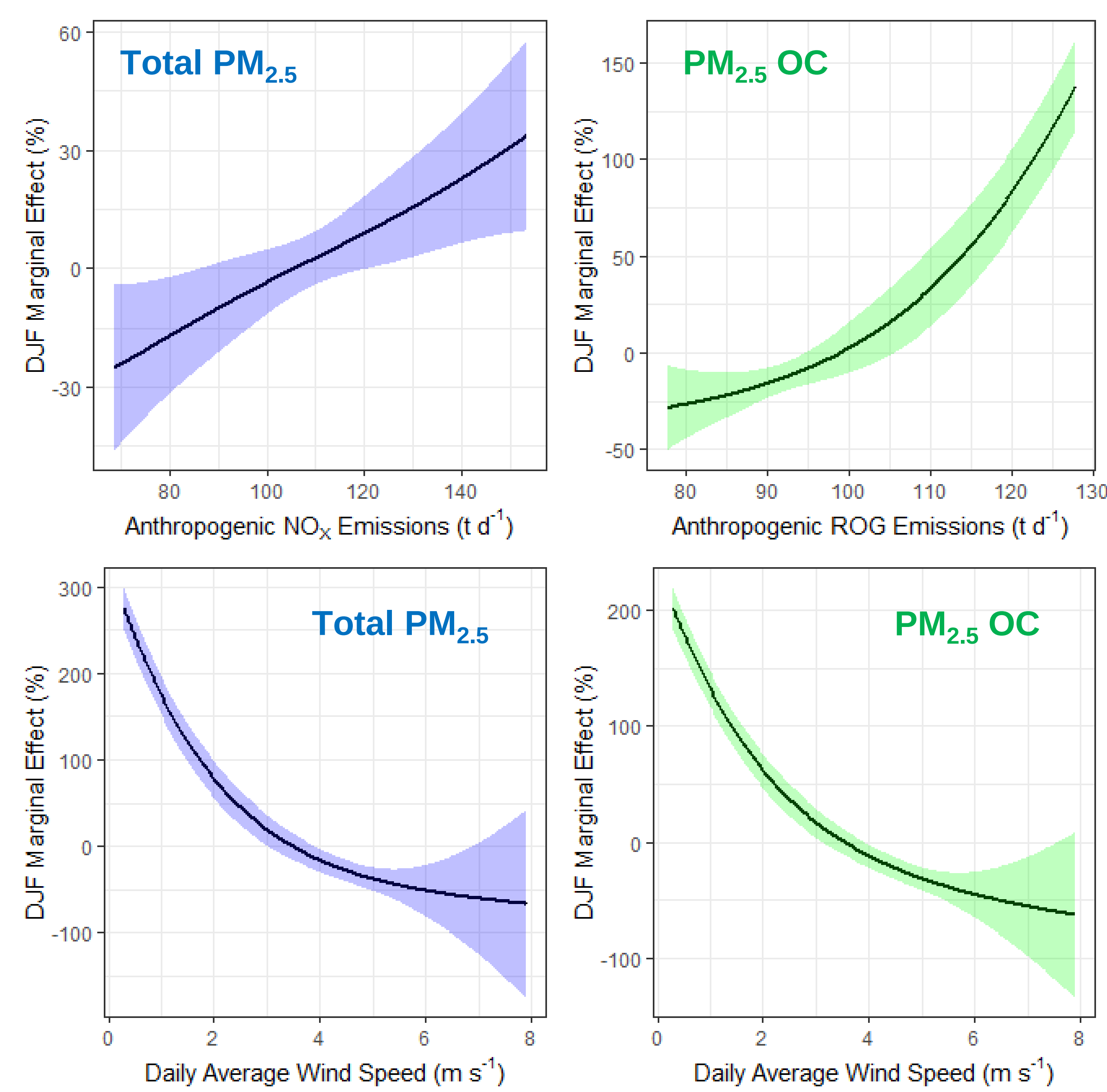
All models are fit with the same set of covariates. Spline fitting employs penalization for automated model selection. Marginal effects are calculated as

$$ME_i = 100\% \times [\exp\{s(x_i)\} - 1] \quad (2)$$

Results

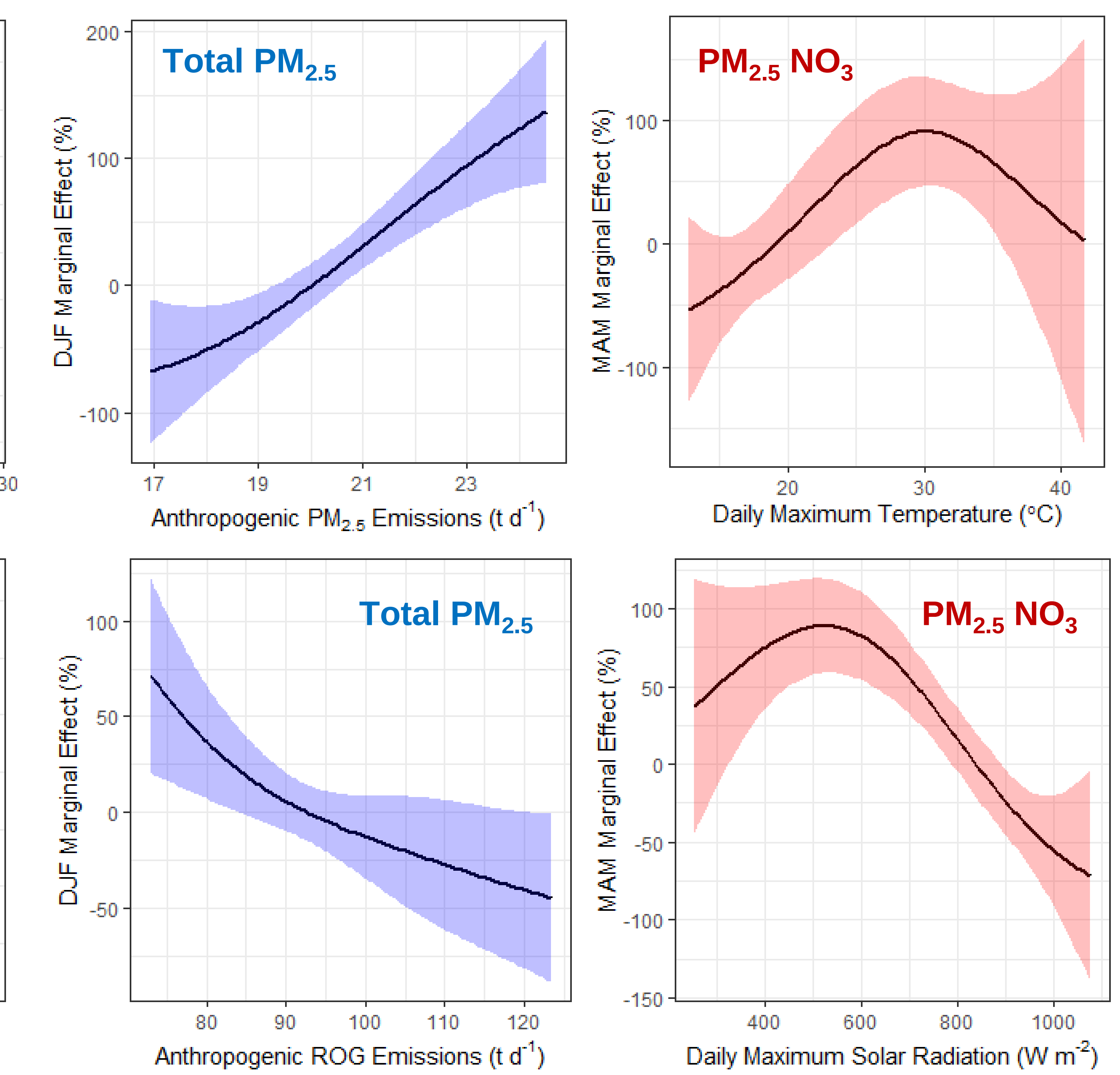
Site: San Jose – Jackson Street

Quantity	Total $PM_{2.5}$	OC	Units
Annual mean	10.773 (2.581)	3.937 (1.808)	µg m^{-3}
MB	-1.350 (0.667)	-0.059 (0.226)	µg m^{-3}
RMSE	2.654 (0.675)	0.940 (0.314)	µg m^{-3}
R^2	0.622 (0.235)	0.791 (0.172)	-



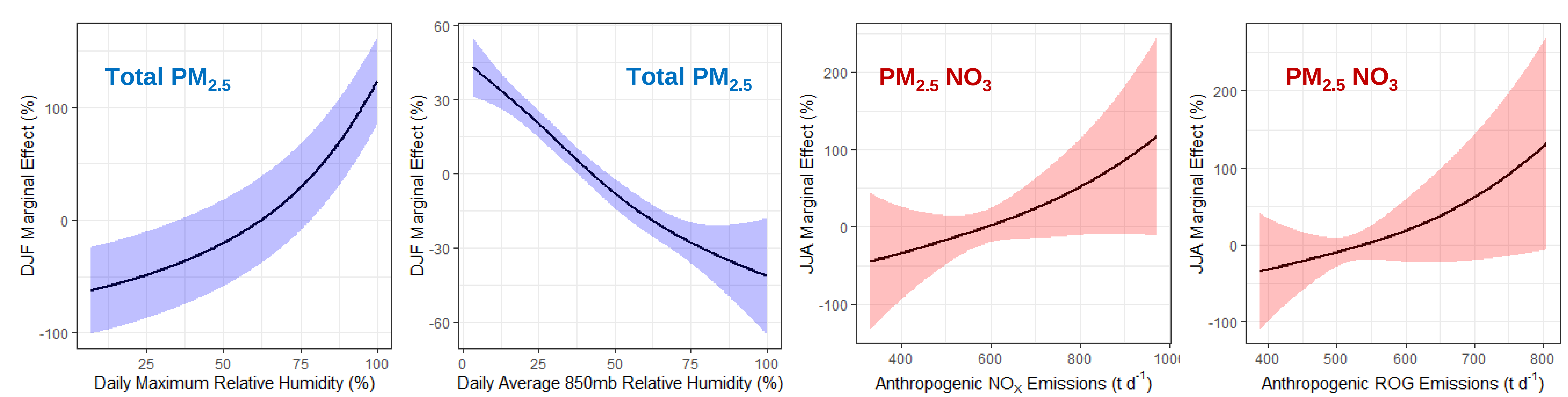
Site: Bakersfield – California Avenue

Quantity	Total $PM_{2.5}$	NO_3	Units
Annual mean	17.498 (3.757)	5.193 (1.885)	µg m^{-3}
MB	0.394 (1.844)	0.313 (1.705)	µg m^{-3}
RMSE	7.446 (3.175)	4.083 (4.860)	µg m^{-3}
R^2	0.489 (0.201)	0.537 (0.232)	-



Site: Riverside - Rubidoux

Quantity	Total $PM_{2.5}$	NO_3	Units
Annual mean	17.343 (5.714)	6.099 (3.019)	µg m^{-3}
MB	0.027 (0.758)	0.048 (0.523)	µg m^{-3}
RMSE	3.326 (0.716)	1.806 (1.132)	µg m^{-3}
R^2	0.781 (0.089)	0.793 (0.086)	-



We report annual performance metrics for our models of total $PM_{2.5}$ and of the speciated component that is the largest fraction of total $PM_{2.5}$ for selected Chemical Speciation Network (CSN) sites. We include marginal effects plots of the two most important drivers for the season with the greatest response. Performance metrics reported as mean (standard deviation) based on 5-times repeated 10-fold cross validation. Results are summarized in Table 2.

Table 2. Summary of results

Site	Species ($PM_{2.5}$, largest fraction)	Dominant season	Top 2 drivers
San Jose	$PM_{2.5}$	DJF	eNO_X, AWND
	OC	DJF	eROG, AWND
Bakersfield	$PM_{2.5}$	DJF	$ePM_{2.5}$, eROG
	NO_3	MAM	TMAX, SR
Riverside	$PM_{2.5}$	DJF	RH, RH_{850mb}
	NO_3	JJA	eNO_X, eROG

Conclusions

- Total and speciated $PM_{2.5}$ are modeled in a way that enables covariate-by-covariate analysis
- This covariate-by-covariate analysis facilitates the identification of driving variables, i.e., covariates that dominate the $PM_{2.5}$ response
- The models resolve seasonal interactions, allowing us to model marginal effects seasonally
- We can rank the importance of drivers using variable importance procedures on a seasonal basis
- This can help inform regulatory policy design by identifying what control targets will produce the greatest reductions in $PM_{2.5}$
- Our models provide a means of predicting changes in total and speciated $PM_{2.5}$ as individual covariates change, identifying what fraction of total change is apportionable to which covariate
- Altogether, these conclusions highlight our models' utility for informing $PM_{2.5}$ control strategies

Acknowledgements

- Khanh Do and Charles Blanchard for their insight and technical support
- The California Air Resources Board

The statements and conclusions on this poster are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

References

1) Camalier, L.; Cox, W.; Dolwick, P. The Effects of Meteorology on Ozone in Urban Areas and Their Use in Assessing Ozone Trends. *Atmospheric Environment* **2007**, *41* (33), 7127–7137. https://doi.org/10.1016/j.atmosenv.2007.04.061.
2) Blanchard, C. L.; Shaw, S. L.; Edgerton, E. S.; Schwab, J. J. Emission Influences on Air Pollutant Concentrations in New York State: I. Ozone. *Atmospheric Environment: X* **2019**, *3*, 100033. https://doi.org/10.1016/j.aeaoa.2019.100033.
3) Gao, Z.; Ivey, C. E.; Blanchard, C. L.; Do, K.; Lee, S.-M.; Russell, A. G. Separating Emissions and Meteorological Impacts on Peak Ozone Concentrations in Southern California Using Generalized Additive Modeling. *Environmental Pollution* **2022**, *307*, 119503. https://doi.org/10.1016/j.envpol.2022.119503.
4) Pearce, J. L.; Beringer, J.; Nicholls, N.; Hyndman, R. J.; Tapper, N. J. Quantifying the Influence of Local Meteorology on Air Quality Using Generalized Additive Models. *Atmospheric Environment* **2011**, *45* (6), 1328–1336. https://doi.org/10.1016/j.atmosenv.2010.11.051.