

APPLICATION PAPER

Modeling the relationship between vertical temperature profiles and acute surface-level ozone events in the US Southwest by spatially smoothing a functional quantile regression estimator

Brook T. Russell^{1,*}  and William C. Porter²

¹School of Mathematical and Statistical Sciences, Clemson University, Clemson, South Carolina 29634, USA

²Department of Environmental Sciences, University of California, Riverside, Riverside, California 92521, USA

*Corresponding author. E-mail: brookr@clemson.edu

Received: 21 September 2022; Accepted: 28 October 2022

Keywords: Air pollution; atmospheric profile variables; functional covariates; functional data analysis; gridded data products

Abstract

With the proliferation of gridded data products, modeling the *relationship* between a scalar response and a functional covariate over the points on a regular lattice is becoming increasingly important. In this work, our overall aim is to better understand the relationship between high quantiles of surface-level ozone and the vertical temperature profile (VTP), a functional covariate, over the US Southwest in the summer. We develop our penalized functional quantile regression based approach within the framework provided by functional data analysis. As we assume that coefficient functions at points on the lattice exhibit spatial similarity, we obtain improved estimates by penalizing dissimilarity between nearby coefficient function estimates. In order to better account for the high degree of diversity of this region, we more strongly weight differences between coefficient function estimates at cells that exhibit a higher degree of geographical and climatological similarity. Our analysis suggests that the VTP is associated with acute surface-level ozone events during the summertime over this region, and the nature of this relationship differs spatially.

Impact Statement

Acute surface-level ozone events have a multitude of negative health consequences. Meteorology plays a major role in the development and intensification of these events; therefore, better understanding the complex interplay between meteorology and surface-level ozone is paramount. For this purpose, in this work, we develop a methodology to model the relationship between vertical temperature profiles and high quantiles of surface-level ozone. We find that acute surface-level ozone events are associated with air temperature inversions over much of the US Southwest in the summer; however, hot weather conditions look to be the primary driver over the Sierra Nevada Mountains and in far Northern California.

1. Introduction

Surface-level ozone (O_3) has negative health consequences that may be further magnified when it reaches its highest levels. The elderly and those with complicating health conditions, such as asthma, may be even more vulnerable to the impacts of surface-level O_3 . O_3 is formed via a chemical reaction between NO_x and volatile organic compounds (VOCs) in the presence of sunlight. Therefore, emissions play a role in

characterizing O_3 levels on the surface. However, it is not uncommon for 2 days with identical NO_x and VOCs emissions to have drastically different surface-level O_3 amounts. It is well established that local and regional meteorological conditions account for a large portion of this discrepancy (Jacob and Winner, 2009; Tai et al., 2010; Liu and Cui, 2014; Porter et al., 2015). Broadly speaking, these works conclude that meteorological variables such as surface-level air temperature, circulation, and stagnation are related to surface-level air pollution; however, there are often several approaches for characterizing these meteorological conditions.

One way to capture these types of meteorological conditions is by considering atmospheric profile variables (APVs). An APV is considered to be a function mapping the altitude above a location on Earth's surface to a scalar measurement. The vertical temperature profile (VTP), which reports the temperature of the air at increasing altitude is an important APV in climate sciences. Modeling surface-level air pollution as a function of APVs has been considered previously (Du et al., 2013; Rendón et al., 2014; Wolf et al., 2014; Russell et al., 2017; Russell and Porter, 2021), but has received much less attention in the literature in comparison to approaches that consider surface-level (scalar) covariates exclusively. In this work, we investigate the association between the VTP and high levels of surface-level O_3 through the use of methods developed within the framework of functional data analysis (FDA). In FDA, some or all observations are taken to be functions, as opposed to scalars and/or vectors. To this end, in this work, we consider the VTP to be a *functional* covariate.

Modeling a scalar response variable as a function of a functional covariate is often done via the use of the functional linear model Horváth and Kokoszka (2012). As in a standard linear regression model, a functional linear model (with a scalar response) estimates the conditional mean response given the functional covariate. If one is interested in higher quantiles of the response distribution, a functional quantile regression model may be of use. Along these lines, Russell and Dyer (2017) suggest a penalized functional quantile regression model to investigate the impact of APVs on surface-level air pollution at locations in South Carolina and Florida. These authors conclude that surface-level air pollution is related to APVs at different pressure levels, and that these associations differ by location and season. As acute levels of surface-level O_3 may have a disproportionately higher impact on human health outcomes, we focus on modeling higher conditional quantiles of the O_3 response in this work.

The last 10 or 15 years have seen the proliferation of gridded data products, which offer climate and air pollution data over large spatial domains with excellent temporal coverage and no missing values. By its nature, the grid utilized in a gridded data product could be thought of as a regular spatial lattice. Others have considered approaches for analysis of data in these contexts (Huang et al., 2010; Zhu et al., 2010; Zheng and Zhu, 2012; Reyes et al., 2015), although our research objectives and methods differ from these works. We emphasize that our primary goal is to better understand the relationship between the VTP and high surface-level O_3 in the summer in the US Southwest based on output from a gridded data product. That is, our objective is to model the *relationship* between a functional covariate and conditional quantiles of a scalar response at points on a regular lattice.

A simple analysis approach would be to fit independent functional linear regression models at each point on the lattice. Unfortunately, such an approach would likely be difficult to interpret due to the presence of excess noise. For our data, we believe that it is reasonable to assume that the functional regression relationships at adjacent points on the lattice are similar. For this reason, we consider an approach that institutes a penalty for the dissimilarity between estimated coefficient functions at adjacent locations on the lattice. This approach helps us with our objective of better understanding the relationship between VTP and key quantiles of surface-level O_3 at all points on the lattice.

2. Pointwise Functional Quantile Regression Modeling

For the random process $\{Y, X\}$ with iid copies $\{Y_t, X_t\}_{t \in \mathbb{N}}$, assume the response variable of interest is $Y_t \in \mathbb{R}$, and $X_t : [0, H] \rightarrow \mathbb{R}$ is a functional covariate. Further, assume that $\{X_t\}_{t \in \mathbb{N}}$ are in \mathcal{L}^2 , the separable Hilbert space determined by the set of measurable real-valued square-integrable functions on $[0, H]$. Given

the value of the functional covariate, we directly model $Q_\tau(Y_t|X_t)$, the conditional τ th quantile of the scalar response variable for some $\tau \in (0, 1)$, via

$$Q_\tau(Y_t|X_t) = \alpha_\tau + \int_0^H X_t^c(s)\beta_\tau(s)ds. \tag{1}$$

We note that the intercept $\alpha_\tau \in \mathbb{R}$, and we call $X_t^c(s) = X_t(s) - E(X(s))$ the centered covariate function. Additionally, the twice differentiable function β_τ is called the *coefficient function* for the τ th quantile. In a typical analysis, estimation and inference regarding β_τ in equation (1) is a primary aim, as portions of $[0, H]$ over which the coefficient function is positive (negative), imply a positive (negative) relationship with that conditional quantile of the response. Additionally, we assume that $\beta_\tau'' \in \mathcal{L}^2$, and that β_τ and β_τ' are both absolutely continuous.

As this is commonly done in FDA, we assume that β_τ can be approximated by a linear combination of a finite set of known basis functions, $\{\phi_k\}_{k=1,\dots,K}$. This implies $\beta_\tau(s) \approx \sum_{k=1}^K b_k \phi_k(s) = \phi^T(s)\mathbf{b}_\tau$. In reality, functional covariates $\{x_i\}_{i=1,\dots,n}$ are infinite dimensional and therefore not fully be observed. Instead, the analyst observes $\{x_i(h)\}_{i=1,\dots,n}$ for $h \in \{h_1, \dots, h_U\}$ such that $0 < h_1 < \dots < h_U \leq H$. We further assume that we are able to identify $\mathbf{a}_i \in \mathbb{R}^M$ such that $x_i(s) \approx \sum_{m=1}^M a_{i,m} \psi_m(s) = \psi^T(h)\mathbf{a}_i$, where $\{\psi_m\}_{m=1,\dots,M}$ is another set of known basis functions. Therefore, the conditional quantile in equation (1) can be expressed via

$$Q_\tau(Y_t|X_t) \approx \alpha_\tau + \int_0^H (\mathbf{a}_i^T \psi(h)) (\phi^T(h)\mathbf{b}_\tau) dh = \alpha_\tau + \mathbf{a}_i^T \mathbf{J} \mathbf{b}_\tau. \tag{2}$$

Here, the $M \times K$ dimensional matrix \mathbf{J} is constructed such that the entry in the m th row and k th column is $\int_0^H \psi_m(h)\phi_k(h)dh$. We estimate \mathbf{b}_τ and α_τ via

$$\left(\widehat{\alpha}_\tau, \widehat{\mathbf{b}}_\tau\right)^T = \underset{\mathbf{b} \in \mathbb{R}^K, \alpha \in \mathbb{R}}{\operatorname{argmin}} \sum_{t=1}^n \rho_\tau(y_t - \alpha - \mathbf{a}_i^T \mathbf{J} \mathbf{b}), \tag{3}$$

where ρ_τ is the *check-loss* function commonly used in quantile regression (Koenker and Bassett Jr., 1978). The resulting optimum from equation (3) leads to the coefficient function estimator via the relationship $\widehat{\beta}_\tau(h) = \phi^T(h)\widehat{\mathbf{b}}_\tau$.

3. Functional Quantile Regression Modeling on a Grid

Climate scientists often make use of gridded data products, which can be thought of as producing output over the points on a regular spatial lattice. Assume that \mathcal{D} is a spatial domain of interest, and that we observe data over the set of points on a regular lattice $\mathcal{D}' \subset \mathbb{Z}^2$, such that $\mathcal{D}' \subset \mathcal{D}$. Despite the fact that the primary objective is to describe the association between the functional covariate and a key quantile of the scalar response everywhere in \mathcal{D} , we assume that modeling this relationship $\forall \mathbf{s} \in \mathcal{D}'$ will be useful. Additionally, we assume that the true coefficient functions at nearby locations are similar, which makes the approach of borrowing information from nearby cells a potentially promising way to improve coefficient function estimates.

Assume that at time $t \in \{1, \dots, n\}$, we observe realizations of a functional covariate and scalar response at locations $\{s_l\}_{l=1,\dots,L}$ such that $s_l \in \mathcal{D}' \forall l \in \{1, \dots, L\}$, at a sequence of pressure levels $0 < h_1 < \dots < h_U \leq H$. We propose a procedure to simultaneously estimate all L coefficient functions $\{\beta_{\tau,l}\}_{l=1,\dots,L}$, under the assumptions described above. We again assume that the true coefficient function at s_l is well approximated by a finite linear combination of known basis functions, giving $\beta_{\tau,l}(h) \approx \sum_{k=1}^K b_{l,k,\tau} \phi_k(h) = \phi^T(h)\mathbf{b}_{l,\tau}$. Observations from the functional covariate are denoted by $\{x_{l,t}(h)\}_{l \in \{1, \dots, L\}, t \in \{1, \dots, n\}}$, and assume further that there exists $\mathbf{a}_{l,t} \in \mathbb{R}^M$ such that

$\mathbf{x}_{l,t}(h) \approx \sum_{m=1}^M a_{l,t,m} \psi_m(h) = \boldsymbol{\psi}^T(h) \mathbf{a}_{l,t}$. We then approximate the conditional τ th quantile of the response at time t and location s_l by

$$Q_\tau(Y_{l,t} | \mathbf{X}_{l,t}) \approx \alpha_{l,\tau} + \int_0^H (\mathbf{a}_{l,t}^T \boldsymbol{\psi}(h)) (\boldsymbol{\phi}^T(h) \mathbf{b}_{l,\tau}) dh = \alpha_{l,\tau} + \mathbf{a}_{l,t}^T \mathbf{J} \mathbf{b}_{l,\tau}. \tag{4}$$

One could consider the estimator

$$\left(\tilde{\mathbf{A}}_\tau^T, \tilde{\mathbf{B}}_\tau^T \right)^T = \underset{\mathbf{B} \in \mathbb{R}^{KL}, \mathbf{A} \in \mathbb{R}^L}{\operatorname{argmin}} \Lambda_\tau(\mathbf{A}_\tau, \mathbf{B}_\tau), \tag{5}$$

for $\mathbf{A}_\tau = (\alpha_{1,\tau}, \dots, \alpha_{L,\tau})^T$ and $\mathbf{B}_\tau = (\mathbf{b}_{1,\tau}^T, \dots, \mathbf{b}_{L,\tau}^T)^T$, where

$$\Lambda_\tau(\mathbf{A}_\tau, \mathbf{B}_\tau) = \sum_{l=1}^L \left(\sum_{t=1}^n \rho_\tau(y_{lt} - \alpha_{l,\tau} - \mathbf{a}_{l,t}^T \mathbf{J} \mathbf{b}_{l,\tau}) \right). \tag{6}$$

The estimator in equation (5) does not incorporate any sort of penalty for dissimilarity between neighboring estimated coefficient functions. As we believe that nearby coefficient functions exhibit similarity in our data application, we supplement the loss function in equation (5) by an additional penalty term.

3.1. Penalizing spatial dissimilarity

Our spatial region of interest is composed of lush forests, arid deserts, coastal regions, and high mountain terrain. Because of this geographical heterogeneity, some pairs of neighboring cells on the lattice may exhibit a higher (or lower) degree of spatial similarity in terms of their true underlying coefficient functions. For this reason, for neighboring locations s_l and $s_{l'}$ ($l \neq l'$), we propose the weighting function

$$w_{\theta,\gamma}(s_l, s_{l'}) = \theta \exp(-\gamma \delta(s_l, s_{l'})), \tag{7}$$

where $\theta > 0$ and $\gamma \geq 0$ are unknown parameters and the function δ models geographic and or climatological dissimilarity between two locations. We propose the penalized estimator

$$\left(\hat{\mathbf{A}}_\tau^T, \hat{\mathbf{B}}_\tau^T \right)^T = \underset{\mathbf{B} \in \mathbb{R}^{KL}, \mathbf{A} \in \mathbb{R}^L}{\operatorname{argmin}} \Lambda_\tau(\mathbf{A}, \mathbf{B}) + \Lambda_{\text{spat}}(\mathbf{A}, \mathbf{B}, \theta, \gamma, \tau), \tag{8}$$

where Λ_τ is defined in equation (6) and

$$\Lambda_{\text{spat}}(\mathbf{A}, \mathbf{B}, \theta, \gamma, \tau) = \sum_{1 \leq l < l' \leq L} w_{\theta,\gamma}(s_l, s_{l'}) \int_0^H |\boldsymbol{\phi}^T(h) \mathbf{b}_{l,\tau} - \boldsymbol{\phi}^T(h) \mathbf{b}_{l',\tau}| dh \mathbb{I}\{s_l \sim s_{l'}\}. \tag{9}$$

We note that equation (9) gives the sum of the spatial dissimilarity penalties over all pairs of adjacent cells in \mathcal{D}' . Here, $\mathbb{I}\{\cdot\}$ denotes the indicator function, and $s_l \sim s_{l'}$ implies that s_l and $s_{l'}$ are adjacent. The definition of adjacency is flexible and can be selected in a way that makes sense for a specific data application. Importantly, the $\hat{\mathbf{B}}_\tau = \left(\hat{\mathbf{b}}_{s_{1,\tau}}^T, \dots, \hat{\mathbf{b}}_{s_{L,\tau}}^T \right)^T$ that minimizes (8) leads to the coefficient functions estimates $\left\{ \hat{\boldsymbol{\beta}}_{\tau,l} \right\}_{l=1,\dots,L}$. The absolute loss penalty in equation (9) is used, because the absolute value function can be expressed in terms of the check-loss function, as seen in equation (10). This makes it straightforward to implement the penalty by augmenting the quantile regression design matrix.

4. Analysis of Surface-Level Ozone in the US Southwest

The primary research objective in this work is to gain a better level of understanding regarding the effects of VTP on acute daily surface-level O_3 events over the US Southwest. The O_3 and VTP data come from the GEOS-Chem model version 12.2.1 (Bey et al., 2001), run for the period of late summer (July/August)

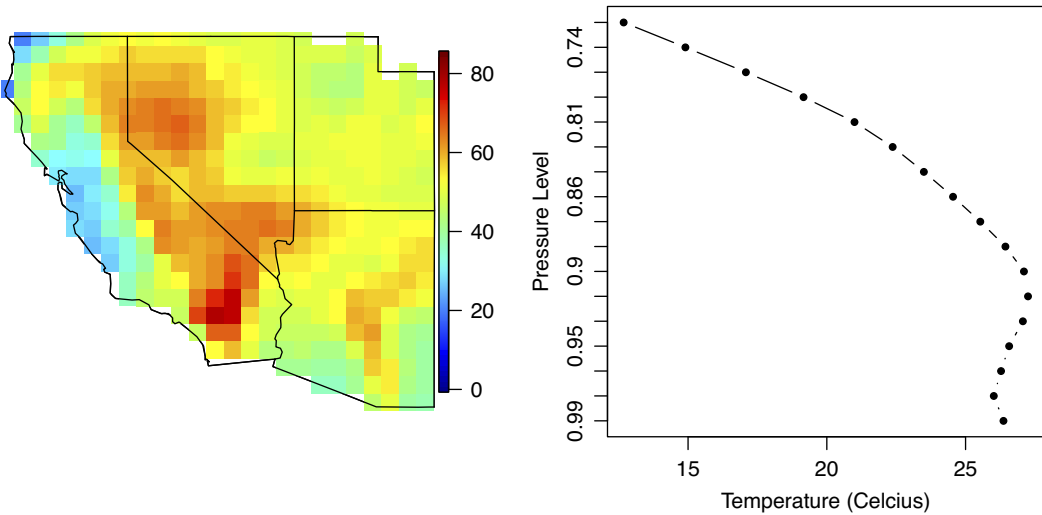


Figure 1. We plot surface-level O_3 (in ppb) at all US Southwest grid cells on July 30, 2016 (left), and the VTP for the grid cell containing Riverside, CA on this day (right). Riverside, CA is plotted on the map in the left panel.

2016. Importantly, GEOS-Chem obtains meteorological variables from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (Gelaro et al., 2017). For illustrative purposes, we plot surface-level O_3 levels (in ppb) for all US Southwest grid cells on July 30, 2016, in the left panel of Figure 1. On this day, surface-level O_3 is fairly high over much of the region. The right panel of Figure 1 plots the VTP for the grid cell containing Riverside, CA on this day. For reference, Riverside, CA is plotted on the map in the left panel. Typically, the temperature of the air warms at lower altitudes; however, we see that this is not the case on this day. This plot of the VTP indicates the presence of a *temperature inversion*, where cooler air at the surface is trapped under warmer air. The high levels of surface-level O_3 in Riverside may be due in part to the presence of this inversion.

In order to more effectively penalize the dissimilarity between adjacent grid cells, we seek to identify a data product that is able to assess the degree to which two adjacent grid cells are similar from a geographical and climatological perspective. For this purpose, we use the PRISM data product (Daly et al., 1997) as it explicitly incorporates information about rainfall and implicitly incorporates information about geographical features such as elevation. Denote the 30-year average annual precipitation totals for PRISM cell i ($i = 1, \dots, N_{\text{Prism}}$) via $\text{Pr}_{30}(i)$. We define $\delta(s_l, s_r) = |\eta(s_l) - \eta(s_r)|$, where $\eta(s_l) = \frac{\sum_{i=1}^{N_{\text{Prism}}} (\text{Pr}_{30}(i) \mathbb{I}\{i \in l\})}{\sum_{i=1}^{N_{\text{Prism}}} \mathbb{I}\{i \in l\}}$, and $i \in l$ is the event that the midpoint of PRISM cell i is contained in cell l . Essentially, $\eta(s_l)$ is the average of all PRISM 30-year annual precipitation totals for all PRISM locations in the grid cell l . The PRISM data product takes both geographical and climatological information into account, and therefore we find it useful to leverage it for this purpose. We implement our analysis procedure using B-spline basis functions to represent both the coefficient function as well as the functional covariates. We utilize 7 B-spline basis functions of order 4 with equally spaced interior knots to represent the functional covariates and estimated coefficient functions. The parameters θ and γ are estimated via a two-dimensional grid search using Bayesian information criterion (BIC) as the model comparison criterion. In keeping with our objective of modeling acute surface-level O_3 , we take $\tau = 0.90$.

Figure 2 plots the resulting coefficient function estimates at all locations in the spatial domain for a sequence of eight equally spaced pressure levels: 825, 850, ..., 975, 1,000. Recall that a pressure level of approximately 1,000 corresponds with the Earth’s surface. At higher pressure levels (above 925 or 950), warmer temperatures are associated with larger high O_3 events. Around the pressure level 950, we begin to see this relationship disappear over large portions of this region. Instead, as we go lower in the

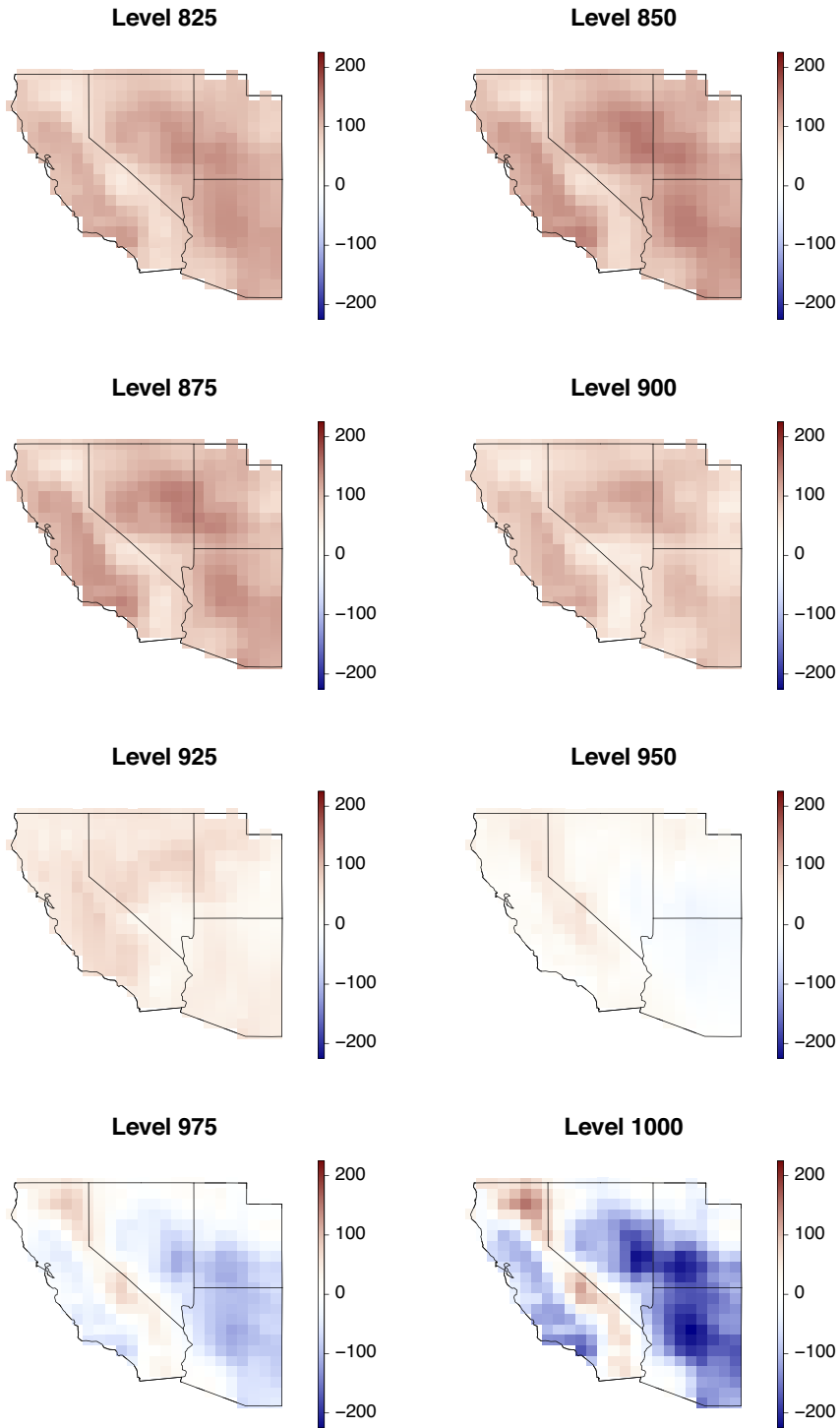


Figure 2. For a sequence of eight pressure levels, we plot the resulting coefficient function estimates for all cells on our spatial grid.

atmosphere, larger high O₃ events are associated with cooler air over Nevada, Southern Utah, and Arizona, as well as coastal California. Interestingly, we see the opposite over the high mountains of California and in the far northern part of the state. This implies that air temperature inversions are associated with acute surface-level O₃ events over the majority of the US Southwest; in contrast, *warmer* air at the surface is linked to these types of air pollution events over the Sierra Nevada Mountains and in Northern California.

5. Discussion

In this work, at each location on a spatial lattice, we estimate the relationship between the VTP and high daily surface-level O₃ events during the summer in the US Southwest states of California, Nevada, Arizona, and Utah. We take a functional quantile regression approach, and as we believe that nearby coefficient functions are similar, we estimate at all points on the lattice simultaneously and penalize the spatial dissimilarity. In this work, we model acute surface-level O₃ events by estimating the conditional 0.90 quantile of the response distribution. Because of the geographic heterogeneity of this diverse region, we weight differences between coefficient function estimates at neighboring cells more if they are similar geographically and/or climatologically. Our analysis concludes that temperature inversions are a primary driver of acute O₃ events over Arizona, Utah, Nevada, and coastal California. In contrast, higher temperatures are associated with acute O₃ events over far Northern California and the Sierra Nevada Mountains.

We determine two cells on the grid to be adjacent if they are connected to the N, S, E, or W. In the future, we hope to expand this definition to further outlying cells. Doing so would complicate estimation, but may improve estimates. Also, the model developed in this work does not incorporate temporal dependence. Although we believe that the results presented here would not change greatly, we hope to refine the modeling approach to account for temporal dependence in the future. We would also like to further refine our methodology to simultaneously consider other quantiles of interest, and to investigate the impact of alternative weighting functions. In the future, we believe that it would be interesting to consider combinations of other regions and seasons, to compare and contrast the results presented here. Similarly, it would be interesting to consider models other than GEOS-Chem model version 12.2.1. Other extensions to our modeling procedure include developing modeling procedures that are able to incorporate both gridded model output and *in situ* observations (O₃ and VTP). We hope to consider these ideas in future work.

We note that equation (9), with its absolute value penalty terms written in terms of finite differences, can also be expressed via the check-loss function, relying on the fact that $\forall \tau \in (0,1)$ and $u \in \mathbb{R}$:

$$\begin{aligned} |u| &= |u|\mathbb{I}\{u > 0\}\tau + |u|\mathbb{I}\{u > 0\}(1 - \tau) \\ &\quad + |u|\mathbb{I}\{u < 0\}\tau + |u|\mathbb{I}\{u < 0\}(1 - \tau) \\ &= |0 - -u|\mathbb{I}\{0 - -u > 0\}\tau + |0 - -u|\mathbb{I}\{0 - -u < 0\}(1 - \tau) \\ &\quad + |0 - u|\mathbb{I}\{0 - u > 0\}\tau + |0 - -u|\mathbb{I}\{0 - -u < 0\}(1 - \tau) \\ &= \rho_\tau(0 - u) + \rho_\tau(0 - -u). \end{aligned} \tag{10}$$

Therefore, in practice, the optimization in equation (8) can be performed in a computationally efficient manner by constructing the design matrix determined by the loss function in equation (6), and augmenting it with the design matrix determined by the penalty Λ_{spat} in equation (9), relying on the relationship in equation (10)). For this reason, functions designed for quantile regression with scalar covariates, such as R's *quantreg* package, may be utilized for optimization. With a large number of adjacent cells, the resulting design matrix will be very large, but also very sparse. This sparsity can be leveraged to make estimation reasonable.

Acknowledgments. We acknowledge two reviewers, whose comments and suggestions helped to improve this manuscript.

Author Contributions. Conceptualization: both authors; Data curation: W.P.; Formal analysis: B.R.; Methodology: B.R.; Software: B.R.; Writing—original draft: B.R.; Writing—review and editing: both authors. Both authors approved the final submitted draft.

Competing Interests. The authors declare none.

Data Availability Statement. The ozone data and atmospheric profile variables data are available upon reasonable request from the corresponding author. The PRISM 30-year precipitation averages are available at <https://prism.oregonstate.edu/normals/>.

Funding Statement. This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

Provenance. This article is part of the Climate Informatics 2022 proceedings and was accepted in *Environmental Data Science* on the basis of the Climate Informatics peer review process.

References

- Bey I, Jacob DJ, Yantosca RM, Logan JA, Field BD, Fiore AM, Li Q, Liu HY, Mickley LJ and Schultz MG** (2001) Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research: Atmospheres* 106(D19), 23073–23095.
- Daly C, Taylor GH and Gibson WP** (1997) The PRISM approach to mapping precipitation and temperature. In *Proceedings of 10th AMS Conference on Applied Climatology*, Reno, NV, October 20–23. *American Meteorological Society*. 10–12.
- Du C, Liu S, Yu X, Li X, Chen C, Peng Y, Dong Y, Dong Z and Wang F** (2013) Urban boundary layer height characteristics and relationship with particulate matter mass concentrations in Xi'an, central China. *Aerosol and Air Quality Research* 13, 1598–1607.
- Gelaro R, McCarty W, Suarez MJ, Todling R, Molod A, Takacs L, Randles CA, Darmenov A, Bosilovich MG, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva AM, Gu W, Kim G-K, Koster R, Lucchesi R, Merkova D, Nielsen JE, Partyka G, Pawson S, Putman W, Rienecker M, Schubert SD, Sienkiewicz M and Zhao B** (2017) The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate* 30 (14), 5419–5454.
- Horváth L and Kokoszka P** (2012) *Inference for Functional Data with Applications*. New York: Springer Science & Business Media.
- Huang H-C, Hsu N-J, Theobald DM and Breidt FJ** (2010) Spatial Lasso with applications to GIS model selection. *Journal of Computational and Graphical Statistics* 19(4), 963–983.
- Jacob DJ and Winner DA** (2009) Effect of climate change on air quality. *Atmospheric Environment* 43(1), 51–63.
- Koenker R and Bassett Jr G** (1978) Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1), 33–50. <https://doi.org/10.2307/1913643>
- Liu J and Cui S** (2014) Meteorological influences on seasonal variation of fine particulate matter in cities over southern Ontario, Canada. *Advances in Meteorology* 2014. 15 <https://doi.org/10.1155/2014/169476>
- Porter W, Heald C, Cooley D and Russell B** (2015) Investigating the observed sensitivities of air quality extremes to meteorological drivers via quantile regression. *Atmospheric Chemistry and Physics Discussions* 15(10), 14075–14109.
- Rendón AM, Salazar JF, Palacio CA, Wirth V and Brötz B** (2014) Effects of urbanization on the temperature inversion breakup in a mountain valley with implications for air quality. *Journal of Applied Meteorology and Climatology* 53(4), 840–858.
- Reyes A, Giraldo R and Mateu J** (2015) Residual kriging for functional spatial prediction of salinity curves. *Communications in Statistics-Theory and Methods* 44(4), 798–809.
- Russell BT, Wang D and McMahan CS** (2017) Spatially modeling the effects of meteorological drivers of PM_{2.5} in the Eastern United States via a local linear penalized quantile regression estimator. *Environmetrics*. 28:e2448. <https://doi.org/10.1002/env.2448>
- Russell BT and Dyer JL** (2017) Investigating the link between PM_{2.5} and atmospheric profile variables via penalized functional quantile regression. *Environmental and Ecological Statistics* 24, 363–384.
- Russell B.T. and Porter W.C.** (2021) Using spatial smoothing to model a functional regression estimator to points on a lattice with application to surface-level ozone in the eastern United States. *Environmental and Ecological Statistics*, 29, 127–147. <https://doi.org/10.1007/s10651-021-00505-4>
- Tai AP, Mickley LJ and Jacob DJ** (2010) Correlations between fine particulate matter (PM_{2.5}) and meteorological variables in the United States: Implications for the sensitivity of PM_{2.5} to climate change. *Atmospheric Environment* 44(32), 3976–3984.
- Wolf T., Esau I. and Reuder J.** (2014) Analysis of the vertical temperature structure in the Bergen valley, Norway, and its connection to pollution episodes. *Journal of Geophysical Research: Atmospheres* 119(18), 10,645–10,662, doi:10.1002/2014JD022085.
- Zheng Y and Zhu J** (2012) On the asymptotics of maximum likelihood estimation for spatial linear models on a lattice. *Sankhya A* 74(1), 29–56.
- Zhu J, Huang H-C and Reyes PE** (2010) On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(3), 389–402.

Cite this article: Russell BT and Porter WC (2022). Modeling the relationship between vertical temperature profiles and acute surface-level ozone events in the US Southwest by spatially smoothing a functional quantile regression estimator. *Environmental Data Science*, 1: e21. doi:10.1017/eds.2022.25