REDUCED-RANK SPATIO-TEMPORAL MODELING OF AIR POLLUTION CONCENTRATIONS IN THE MULTI-ETHNIC STUDY OF ATHEROSCLEROSIS AND AIR POLLUTION[1]

BY CASEY OLIVES*, LIANNE SHEPPARD*, JOHAN LINDSTRÖM†,
PAUL D. SAMPSON*, JOEL D. KAUFMAN* AND ADAM A. SZPIRO*

University of Washington and Lund University†*

There is growing evidence in the epidemiologic literature of the relationship between air pollution and adverse health outcomes. Prediction of individual air pollution exposure in the Environmental Protection Agency (EPA) funded Multi-Ethnic Study of Atheroscelerosis and Air Pollution (MESA Air) study relies on a flexible spatio-temporal prediction model that integrates land-use regression with kriging to account for spatial dependence in pollutant concentrations. Temporal variability is captured using temporal trends estimated via modified singular value decomposition and temporally varying spatial residuals. This model utilizes monitoring data from existing regulatory networks and supplementary MESA Air monitoring data to predict concentrations for individual cohort members.

In general, spatio-temporal models are limited in their efficacy for large data sets due to computational intractability. We develop reduced-rank versions of the MESA Air spatio-temporal model. To do so, we apply low-rank kriging to account for spatial variation in the mean process and discuss the limitations of this approach. As an alternative, we represent spatial variation using thin plate regression splines. We compare the performance of the outlined models using EPA and MESA Air monitoring data for predicting concentrations of oxides of nitrogen (NO_x)—a pollutant of primary interest in MESA Air—in the Los Angeles metropolitan area via cross-validated R^2.

Our findings suggest that use of reduced-rank models can improve computational efficiency in certain cases. Low-rank kriging and thin plate regression splines were competitive across the formulations considered, although TPRS appeared to be more robust in some settings.

1. Introduction. There is growing evidence in the epidemiologic literature of the relationship between air pollution and adverse health outcomes. Early findings were based on somewhat crude regional, and possibly temporally specific, assignment of exposures [Dockery et al. (1993), Pope et al. (2002), Samet et al. (2000)].

Yet, methods for assigning individual exposure to cohort study participants have become much more sophisticated. Recent studies have assigned individual exposure using the value measured at the nearest monitoring location [Miller et al. (2007), Ritz, Wilhelm and Zhao (2006)]; using "land use regression" estimates based on spatially distributed or Geographic Information Systems (GIS) based covariates [Brauer et al. (2003), Hoek et al. (2008), Jerrett et al. (2005a)]; and by interpolation with geostatistical methods such as kriging and semi-parametric smoothing [Jerrett et al. (2005b), Künzli et al. (2005), Paciorek et al. (2009)].

Motivated by the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) study [Kaufman et al. (2012)], Szpiro et al. (2010), Sampson et al. (2011) and Lindström et al. (2013) developed a flexible spatio-temporal prediction model based on monitoring data from existing regulatory networks as well as supplementary MESA Air monitoring data to predict concentrations for individual MESA cohort members. This work integrates land-use regression with kriging to account for spatial dependence in pollutant concentrations. Temporal variability is captured using temporal trends estimated via sparse singular value decomposition and temporally varying spatial residuals [Fuentes, Guttorp and Sampson (2006), Sampson et al. (2011), Szpiro et al. (2010)].

In general, spatio-temporal models are limited in their efficacy for large data sets due to computational intractability. For example, in the purely spatial setting, computation typically is of the order $\mathcal{O}(n^3)$, where n is the number of spatial locations. The computational effort for log-likelihood evaluation of the MESA Air spatio-temporal model typically grows at least as fast, but slower than $\mathcal{O}(N^3)$, where N is the total number of spatio-temporal observations [Lindström et al. (2013)]. Methods for reducing the computational burden in spatio-temporal models are becoming more common in the spatial statistics literature. Several authors have proposed dynamic frameworks for modeling residual spatial and temporal dependence, although these approaches continue to suffer from computational intractability [Gelfand, Banerjee and Gamerman (2005), Stroud, Müller and Sansó (2001)]. In the large spatial data context, approximate likelihood and sampling-based approaches have been proposed to reduce computational burden [Fuentes (2007), Pace and LeSage (2009)]. An alternative to approximate methods involves reducing the spatial process to a K-dimensional subspace ($K \ll n$) in order to increase computational efficiency [Banerjee et al. (2008), Crainiceanu, Diggle and Rowlingson (2008), Kammann and Wand (2003), Nychka and Saltzman (1998), Stein (2007, 2008)]. These so-called "low-rank" or "reduced-rank" approaches can reduce computation to $\mathcal{O}(K^3)$.

In the current work, we develop reduced-rank versions of the spatio-temporal model outlined in Lindström et al. (2013), Szpiro et al. (2010). Specifically, we apply the approach proposed by Kammann and Wand (2003) to achieve low-rank kriging to account for spatial variation in the mean process and spatially varying temporal trends. We discuss the limitations of this approach and, as an alternative, represent spatial variation using thin plate regression splines [Wood (2003)]. We

compare the performance of the outlined models using Environmental Protection Agency (EPA) and MESA Air monitoring data for predicting oxides of nitrogen (NO_x) concentrations in the Los Angeles metropolitan area.

2. Description of data.

2.1. *Air Quality System* (*AQS*).   The national AQS network of regulatory monitors, managed by the EPA, reports concentrations of a wide variety of air pollutant concentrations on an ongoing basis, most typically hourly averages. For this study, we include NO_x measurements from 21 AQS monitors in the Los Angeles area, one of six metropolitan areas where MESA Air cohort members live. Monitor locations are shown in Figure 1 (left). As MESA Air supplementary monitoring is done at the 2-week average scale, we aggregate AQS monitoring data to 2-week averages. Due to skew in the data, all 2-week averages are log transformed.

2.2. *MESA Air.*   As part of the MESA Air project goals to provide high quality individual exposure prediction, additional monitoring data were collected in each of the study's six geographic regions, including Los Angeles. The goal of the supplementary monitoring was to provide geographically complementary data to the AQS monitoring data and to systematically span the design space based on proximity to traffic. Additionally, supplementary monitoring data included measurements collected at a subset of cohort participant homes. The sampling strategy is described in more detail by Cohen et al. (2009).
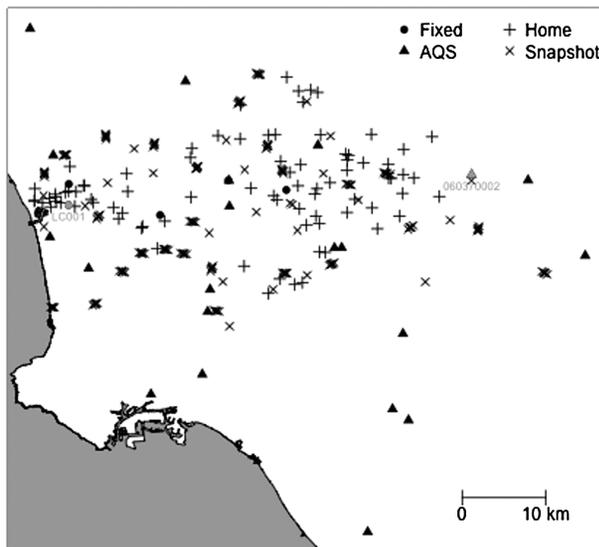

FIG. 1.   *Map of AQS and MESA Air monitoring locations in Los Angeles*, *California. "Home outdoor" monitors have been jittered for participant confidentiality.*

The MESA Air supplementary data is comprised of three classes of monitors, which we refer to as "fixed site," "home outdoor" and "community snapshot." There are a total of five "fixed sites" included in this study in the Los Angeles area. These "fixed-sites" began measuring 2-week average concentrations in November of 2005, for a total of 426 measurements by June 1, 2009. A total of 84 "home outdoor" locations were included in this study. These sites were sampled during 2-week periods starting in May of 2006 and ending in February of 2008, for a total of 155 measurements. The sampling plan calls on each home to be measured two times during different seasons. Last, the "community snapshot" sub-campaign consists of 177 sites measured in three rounds of spatially rich sampling during single 2-week periods from July 5, 2006 to January 1, 2007, for a total of 449 measurements. In each round of the "community snapshot" monitoring, most monitors were clustered in groups of six, with three on each side of a major roadway at distances of about 50, 100 and 300 meters, and locations were chosen to span the domain of various land-use categories and to cover a wide geographic region. All MESA Air monitoring locations as of June 1, 2009 are displayed in Figure 1. Likewise, temporal coverage and sampling frequency during the study period for each monitoring location and type is depicted in Figure 2. Table 1 provides summary statistics on the native and log-scales for both EPA and MESA Air data.
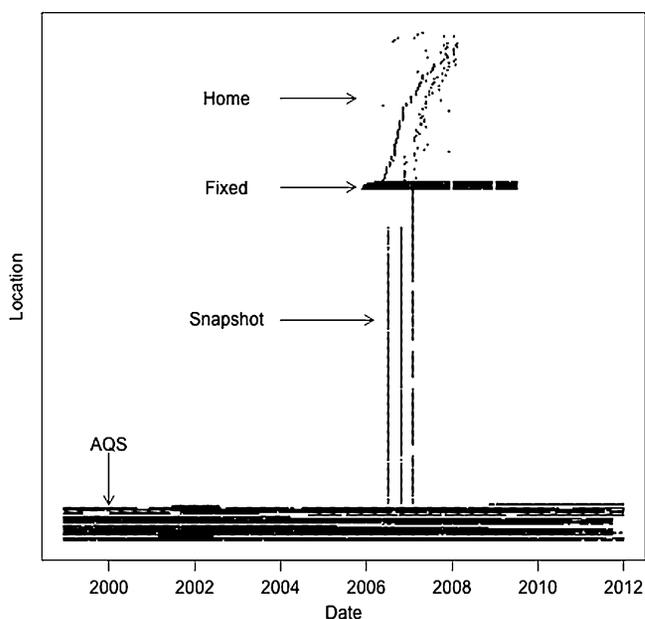

FIG. 2. *Schematic of sampling schedule for AQS and MESA Air monitors between* 1999 *and* 2012. *Each point represents a two-week sampling period.*

TABLE 1
Summary of statistics of NO_x monitoring data at EPA AQS and MESA Air supplementary monitoring sites

Type of site	NO_x ppb		$\log(NO_x$ ppb)	
	Mean	**SD**	**Mean**	**SD**
AQS/fixed site				
2-wk	53.30	40.10	3.72	0.75
LTA	45.35	17.27	3.74	0.39
Community snapshot				
2006-07-05 (summer)	34.24	11.49	3.47	0.39
2006-10-25 (fall)	75.09	23.47	4.27	0.32
2007-01-31 (winter)	95.29	26.99	4.51	0.30
Home outdoor	45.65	28.30	3.63	0.64

2.3. *GIS.* In addition to the monitoring data, spatial prediction at locations where there are no measurements rely heavily on GIS-based covariates and so-called "land-use regression" techniques [Jerrett et al. (2005a)]. In this paper, we considered a limited set of geographic covariates: (i) log distance to A1, A2 or A3 roadway [TeleAtlas (2000)], (ii) log Caline3QHCR point predictions averaged over 9 kilometer buffer [Eckhoff and Braverman (1995)], (iii) distance to nearest coast [TeleAtlas (2000)], (iv) distance to city hall [TeleAtlas (2000)], (v) normalized difference vegetation index averaged over 250 meter buffer [Carroll et al. (2004)], (vi) log elevation, and (vii) percent impervious surface in 50 meter buffer [Fry et al. (2011)].

3. Methods.

3.1. *Review of full-rank spatio-temporal model.* The existing spatio-temporal model as initially described by Szpiro [Szpiro et al. (2010)] takes the form

$$y(s, t) = \mu(s, t) + \nu(s, t),$$

where $y(s, t)$ is the log two-week average of pollutant measurements at location s and time t, $\mu(s, t)$ is the mean field and $\nu(s, t)$ is the residual field. The mean field, μ, is defined as a linear combination of temporal basis functions with spatially varying coefficients. The spatially varying coefficients are comprised of a land-use regression component in addition to spatially structured random fields. These coefficients capture spatial heterogeneity in the amplitude of the temporal basis functions. As such, the mean field is written as

$$\mu(s, t) = \sum_{j=1}^{m} \{ \mathbf{X}_j \boldsymbol{\alpha}_j + \beta_j(s) + \psi_j(s) \} f_j(t),$$

where the $\mathbf{X}_j$ are design matrices containing GIS/land-use covariates of dimension $n \times (p_j + 1)$, where n is the total number of observed sites and $\boldsymbol{\alpha}_j$ is a vector of regression land-use regression coefficients of dimension $p_j + 1 \times 1$. The $\beta_j(\mathbf{s})$ where $\mathbf{s} = (s_1, \ldots, s_n)$ are Gaussian spatial random fields distributed as

$$\beta_j(\mathbf{s}) \sim N(\mathbf{0}, \Sigma_{\beta_j}(\boldsymbol{\theta}_j)).$$

Here, $\Sigma_{\beta_j}(\boldsymbol{\theta}_j)$ is the covariance matrix of dimension $n \times n$ indexed by the vector of parameters $\boldsymbol{\theta}_j$. Generally, we assume a spatial exponential decay model with range ϕ_j and partial sill τ_j^2. The $\psi_j(\mathbf{s})$ are i.i.d. random effects distributed as

$$\psi_j(\mathbf{s}) \sim N(\mathbf{0}, \sigma_j^2 \mathbf{I}).$$

Note $\psi_j(\mathbf{s})$ can equivalently be thought of as the nugget for the $\beta_j(\mathbf{s})$-field. The original formulation of this model did not include a provision for a nugget [Szpiro et al. (2010)], although more recent work allowed for but did not utilize this parameter [Lindström et al. (2013)]. We later discuss the implications of excluding the nugget for computation and predictive performance.

The $f_j(t)$ are temporal basis functions with $f_1(t) \equiv 1$ for all t (typically m is small, ≤ 3) estimated by modified singular value decomposition. See Fuentes, Guttorp and Sampson (2006), Sampson et al. (2011), Szpiro et al. (2010) for a more thorough discussion of trend estimation. Figure 3 depicts these smooth temporal basis functions and their fit to the EPA and MESA Air NO$_x$ monitoring data at two sites.

Last, we specify the model for the residual field, $\nu(s, t)$. Consistent with Lindström et al. (2013), Sampson et al. (2011), Szpiro et al. (2010), we assume that the mean model accounts for the mean structure and all temporal correlation. Thus, the spatio-temporal residuals are assumed to have zero mean and to be independent in time, so that

$$\nu(\mathbf{s}, t) \sim N(\mathbf{0}, \Sigma_\nu^t(\boldsymbol{\theta}_\nu)),$$

where $\Sigma_\nu^t(\boldsymbol{\theta}_\nu)$ is a covariance matrix of dimension $n_t \times n_t$ and n_t is the number of sites observed at time t with $\sum_t n_t = N$, the total number of observations. Once again, we assume that the ν field follows a spatial exponential decay model with range ϕ, partial sill τ^2 and (possibly) nugget σ^2.

A concise representation of this model is given as

(1) $$\mathbf{Y} = \mathbf{F}\mathbf{X}\boldsymbol{\alpha} + \mathbf{F}\mathbf{B} + \mathbf{F}\mathbf{P} + \mathbf{V},$$

where $\mathbf{Y}$ is an $N \times 1$ vector of stacked responses $y(s, t)$ (first varying s then t), $\mathbf{F} = (f_{st,is'})$ is an $N \times mn$ matrix that has elements

$$f_{st,is'} = \begin{cases} f_i(t), & \text{if } s = s', \\ 0, & \text{else}, \end{cases}$$

$\mathbf{X}$ is a block diagonal matrix with diagonal blocks $\{\mathbf{X}_j\}_{j=1}^m$, $\boldsymbol{\alpha}$ is an $\sum_{j=1}^m \{p_j + 1\} \times 1$ stacked vector of the $\boldsymbol{\alpha}_j$, $\mathbf{B}$ is an $mn \times 1$ vector of the stacked β_j, $\mathbf{P}$ is an
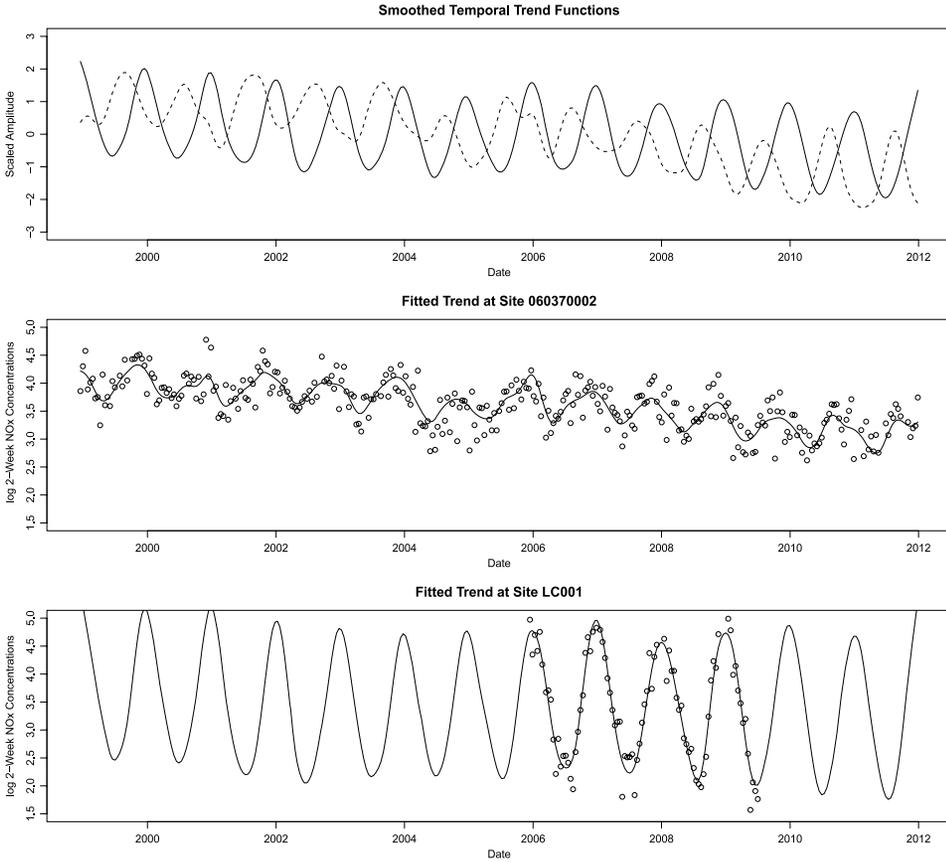
FIG. 3. (*Top*) *Two temporal basis functions estimated by modified singular value decomposition from Los Angeles monitoring data*; (*middle and bottom*) *raw log-transformed data and fits to the two temporal basis functions at sites near* (*LC*001) *and far* (06037002) *from the coastline.*

$mn \times 1$ vector of the stacked nuggets, ψ_j, and $\mathbf{V}$ is an $N \times 1$ vector of the stacked ν (first varying s then t). This model is thus indexed by the land use regression coefficients, $\boldsymbol{\alpha}$, and the covariance parameters

$$\boldsymbol{\theta}_B = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m), \qquad \boldsymbol{\theta}_j = (\phi_j, \tau_j^2), \qquad j = 1, \ldots, m,$$

$$\boldsymbol{\theta}_P = (\sigma_1^2, \ldots, \sigma_m^2),$$

$$\boldsymbol{\theta}_V = (\phi_\nu, \tau_\nu^2, \sigma^2).$$

To simplify notation, we collect the covariance parameters into the vector $\boldsymbol{\Xi} = (\boldsymbol{\theta}_B, \boldsymbol{\theta}_P, \boldsymbol{\theta}_V)$. In the remainder of the manuscript, for the sake of brevity we suppress the dependence of covariance matrices on their respective parameters, except where an explicit dependence is illustrative.

Model (1) is typically fit using profile maximum likelihood methods, although full maximum likelihood and restricted maximum likelihood approaches are also possible [Lindström et al. (2013)]. Sampson used a multi-stage "pragmatic" approach to fitting (1) and generating predictions [Sampson et al. (2011)]. Lindström adapted the model to allow for time-varying covariates, although this extension is not presented here [Lindström et al. (2013)]. This model is implemented in the R-package, SpatioTemporal, available at http://cran.r-project.org/package=SpatioTemporal.

3.2. *Motivation for reduced-rank spatial smoothing.* Although the above formulation of the model has been successful for predicting air pollution concentrations, we note two limitations of this formulation, particularly with respect to the β-fields. First, we note that it is not natural to interpret the β-fields as random effects since it is difficult to imagine the data generating mechanism that might give rise to such fields [Hodges (2013), Hodges and Clayton (2011)]. Second, the range parameters in the β-fields tend to be challenging to estimate in practice. Moreover, Zhang showed that in the case of spatial generalized linear mixed models, this quantity is not consistently estimable [Zhang (2004)].

As such, we consider a spline-based representation of the β-fields in the mean model. To motivate, we note that the Gaussian spatial β-fields, as defined above, can be represented as spatial splines as follows. Let

$$\Sigma_{\beta_j} = \tau_j^2 \mathbf{\Omega},$$

where $\mathbf{\Omega}$ is a matrix such that

$$\mathbf{\Omega} = \left\{ C\left(\|\mathbf{s}_i - \mathbf{s}_j\| \right) \right\}_{i,j \in \mathcal{S}}$$

and $\mathcal{S}$ is the set of observed spatial locations. For the exponential model, $C(r) = \exp\{-|r|/\phi\}$. It follows that the β-fields can be expressed as

$$\beta_j(\mathbf{s}) = \mathbf{\Omega}^{1/2} \boldsymbol{\delta}_j,$$

where $\boldsymbol{\delta}_j \sim \text{MVN}(\mathbf{0}, \tau_j^2 \mathbf{I})$. The n columns of the matrix $\mathbf{\Omega}^{1/2}$ represent n spatial basis functions indexed by the parameter ϕ_j. Written as such, the β-fields can be viewed as random linear combinations of spatial basis functions. Exploiting the connection between linear mixed models and penalized splines, we can view the β_j-fields as penalized spatial splines with smoothing parameters σ^2/τ_j^2 [Ruppert, Wand and Carroll (2003)]. Having represented the β-fields as penalized splines, it is natural to consider penalized reduced-rank splines instead as a means of improving model performance and computational efficiency.

We note that an analogous argument can be made for the residual field. However, it is also the case that the ν-field is well understood within the traditional framework of random effects models. That is, the ν-field captures extra random spatial variation that arises from time point to time point. Furthermore, the range

parameter in the ν-field tends to be more stably estimated in practice due to the repeated measurements over time.

In the following sections, we describe reduced-rank representations of the β-fields using low-rank kriging and thin plate regression splines.

3.3. *Low-rank kriging.* We follow the approach outlined by Kammann and Wand (2003) and Ruppert, Wand and Carroll (2003) for low-rank kriging (LRK) of the β-fields. Specifically, LRK is achieved by replacing $\boldsymbol{\Omega}$ with $\mathbf{Z}\tilde{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^{\top}$, where

$$\mathbf{Z} = \{C(\|\mathbf{s}_i - \boldsymbol{\kappa}_j\|)\}_{i \in \mathcal{S}, j \in \mathcal{K}}, \qquad \tilde{\boldsymbol{\Omega}} = \{C(\|\boldsymbol{\kappa}_i - \boldsymbol{\kappa}_j\|)\}_{i,j \in \mathcal{K}},$$

and $\mathcal{K}$ is the set of spatial knot locations, $\boldsymbol{\kappa}$, of cardinality $K \ll n$. It follows that we can approximate $\beta_j(\mathbf{s})$ by $\mathbf{Z}\tilde{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{\delta}_j$, where $\boldsymbol{\delta}_j$ is now a K-vector distributed as $\text{MVN}(0, \Sigma_{\delta_j} = \tau_j^2\mathbf{I})$. We note that this approach bears strong resemblance to the predictive processes presented by Banerjee [Banerjee et al. (2008)]. In fact, Banerjee noted that LRK is a re-projection of his predictive process. As such, these approaches are computationally identical despite the fact that the predictive process is derived formally from a full-rank parent process.

Letting $\mathbf{Z}_B = \{\mathbf{Z}\tilde{\boldsymbol{\Omega}}^{-1/2}\}_{j=1}^m$ and $\tilde{\mathbf{B}}$ be the stacked vector of $\boldsymbol{\delta}_j$s, we can express the spatio-temporal model as

$$(2) \qquad\qquad Y = \mathbf{FX}\boldsymbol{\alpha} + \mathbf{FZ}_B\tilde{\mathbf{B}} + \mathbf{FP} + \mathbf{V}.$$

Model (2) can be re-expressed as

$$\mathbf{Y} \sim \text{MVN}(\mathbf{FX}\boldsymbol{\alpha}, \tilde{\Sigma}),$$

where

$$\tilde{\Sigma} = \mathbf{FZ}_B \Sigma_{\tilde{B}} \mathbf{Z}_B^{\top}\mathbf{F}^{\top} + \mathbf{F}\Sigma_P\mathbf{F}^{\top} + \Sigma_V,$$

$\Sigma_{\tilde{B}}$ is a block-diagonal matrix with diagonal elements $\{\Sigma_{\delta_j}\}_{j=1}^m$, Σ_P is a block-diagonal matrix with diagonal elements $\{\sigma_j^2\mathbf{I}_n\}_{j=1}^m$, and Σ_V is a block-diagonal matrix with diagonal elements $\{\Sigma_\nu^t\}_{t=1}^T$. The log-likelihood is given by

$$l(\boldsymbol{\alpha}, \boldsymbol{\Xi} \mid \mathbf{Y}) \propto -\log|\tilde{\Sigma}| - (\mathbf{Y} - \mathbf{FX}\boldsymbol{\alpha})^{\top}\tilde{\Sigma}^{-1}(\mathbf{Y} - \mathbf{FX}\boldsymbol{\alpha}).$$

Consistent with Lindström et al. (2013), Szpiro et al. (2010), we estimate regression coefficients $\boldsymbol{\alpha}$ using the profile maximum likelihood. It is easy to show that

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}^{\top}\mathbf{F}^{\top}\tilde{\Sigma}^{-1}\mathbf{FX})^{-1}\mathbf{X}^{\top}\mathbf{F}^{\top}\tilde{\Sigma}^{-1}\mathbf{Y},$$

so that the profile log likelihood is simplified to

$$(3) \qquad l_p(\boldsymbol{\Xi} \mid \mathbf{Y}) \propto -\log|\tilde{\Sigma}| - (\mathbf{Y} - \mathbf{FX}\hat{\boldsymbol{\alpha}})^{\top}\tilde{\Sigma}^{-1}(\mathbf{Y} - \mathbf{FX}\hat{\boldsymbol{\alpha}}),$$

and the remaining parameters are estimated as those quantities that maximize (3). We estimate all parameters using the L-BFGS-B algorithm as implemented in the

`optim` function in `stats` package in R. This is an iterative method that allows for box constraints on all parameters [Byrd et al. (1995)].

Prediction is achieved by assuming a joint distribution between observed data $\mathbf{Y}$ and unobserved data $\mathbf{Y}^*$,

$$
\begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}^* \end{pmatrix} \sim \left(\begin{pmatrix} \mathbf{FX} \\ \mathbf{F}^*\mathbf{X}^* \end{pmatrix} \boldsymbol{\alpha}, \begin{bmatrix} \tilde{\Sigma} & \tilde{\Sigma}_{.*} \\ \tilde{\Sigma}_{*.} & \tilde{\Sigma}_{**} \end{bmatrix} \right),
$$

where $\tilde{\Sigma}_{**}$ is the covariance of $\mathbf{Y}^*$ and $\tilde{\Sigma}_{.*}$ is the cross covariance of $\mathbf{Y}$ and $\mathbf{Y}^*$. Predictions are based on the conditional expectation $E[\mathbf{Y}^*|\mathbf{Y}]$ with MLEs plugged in, namely,

$$
\hat{\mathbf{Y}}^* = \mathbf{F}^*\mathbf{X}^*\hat{\boldsymbol{\alpha}} + \hat{\tilde{\Sigma}}_{.*}\hat{\tilde{\Sigma}}^{-1}(\mathbf{Y} - \mathbf{FX}\hat{\boldsymbol{\alpha}})
$$

with conditional prediction variance

$$
V(\mathbf{Y}^* \mid \mathbf{Y}, \hat{\Xi}, \hat{\boldsymbol{\alpha}}) = \hat{\tilde{\Sigma}}_{**} - \hat{\tilde{\Sigma}}_{*.}^{\top}\hat{\tilde{\Sigma}}^{-1}\hat{\tilde{\Sigma}}_{.*}.
$$

A drawback of LRK is the dependence of the basis functions on the range parameters ϕ_j, $j = 1, \ldots, m$. Kammann and Wand, for purely spatial data, address this issue by fixing the value of this parameter at the maximum spatial distance observed in the data [Kammann and Wand (2003)]. Although it is attractive to condition on fixed spatial basis functions, arbitrary selection of these parameters could lead to worse predictive performance. The range parameters can be estimated from the data, albeit at the expense of more challenging numerical optimization and with the caveat that they may not be consistently estimable [Zhang (2004)].

An alternative approach which sidesteps these issues and leverages the spatial spline formulation calls for the use of alternative spline bases. Thin plate regression splines are a popular alternative, and we explore their application in the current problem below.

3.4. *Summary of thin plate regression splines.* Thin plate regression splines (TPRS) present an alternative to the LRK approach and mitigate the issue of estimating the range parameter(s) [Wood (2003)]. Although these models are widely used (implementation is available in the R package `mgcv`, e.g.), we briefly summarize the approach with the goal of describing parallels between TPRS and LRK.

Assume that we wish to estimate the function f based on (purely spatial) observations $\mathbf{Y}$ at locations $\mathbf{s} = (s_1, s_2)$ such that

$$
Y_i = f(s_i) + \varepsilon_i
$$

by minimizing this penalized objective function

$$
\|\mathbf{Y} - f(\mathbf{s})\| + \lambda \int_{s_1}\int_{s_2} \left(\frac{\partial^2 f}{\partial s_1^2} + \frac{\partial^2 f}{\partial s_2^2} + \frac{\partial^2 f}{\partial s_1 \partial s_2} \right)^2 ds_1\, ds_2.
$$

It can be shown that the solution is given by

$$(4) \qquad f(s) = \sum_{i=1}^{n} \zeta_i \eta(\|s - s_i\|) + \sum_{j=1}^{3} \gamma_j \iota_j(s),$$

where the ι_j are linearly independent polynomials spanning the space of polynomials in $\mathcal{R}^2$ (of degree less than 2) and $\eta(r) = 2^{-3}\pi^{-1}r^2 \log(r)$. Further, $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$ are fixed unknown coefficients subject to the constraint $\mathbf{T}^\top \boldsymbol{\zeta} = 0$ with $T_{ij} = \iota_j(s_i)$ [Green and Silverman (1994)].

Let $\mathbf{E}$ be a matrix so that $E_{ij} = \eta(\|s_i - s_j\|)$. Wood presents a reduced-rank approximation of this problem, which is the solution to the unconstrained optimization problem

$$\text{minimize}\|\mathbf{Y} - \mathbf{U}_K \mathbf{D}_K \mathbf{W}_K \boldsymbol{\zeta}^* - \mathbf{T}\boldsymbol{\gamma}\| + \lambda \boldsymbol{\zeta}^{*\top} \mathbf{W}_K^\top \mathbf{D}_K \mathbf{W}_K \boldsymbol{\zeta}^*,$$

where $\boldsymbol{\zeta}^*$ is a $K - 3 \times 1$ vector of fixed unknown coefficients, $\mathbf{U}\mathbf{D}\mathbf{U}^\top$ is the eigendecomposition of $\mathbf{E}$ so that the n columns of $\mathbf{U}$ are equal to the eigenvectors of $\mathbf{E}$ ordered by their associated eigenvalues from largest to smallest, $\mathbf{D}$ is a diagonal matrix of these eigenvalues, $\mathbf{U}_K$ is a matrix of the first K columns of $\mathbf{U}$, and $\mathbf{D}_K$ is a matrix of the first K rows and columns of $\mathbf{D}$. Last, $\mathbf{W}_K$ is a $K \times K - 3$ orthogonal column basis such that $\mathbf{T}^\top \mathbf{U}_K \mathbf{W}_K = \mathbf{0}$ (to account for the constraint) [Wood (2003)].

It is easy to see that this unconstrained optimization is equivalent to fitting the linear mixed model

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\gamma} + \mathbf{U}_K \mathbf{D}_K \mathbf{W}_K \boldsymbol{\zeta}^* + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\zeta}^* \sim \text{MVN}(\mathbf{0}, \sigma_\zeta^2 (\mathbf{W}_K^\top \mathbf{D}_K \mathbf{W}_K)^{-1})$, $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$, and $\lambda = \sigma_\varepsilon^2 / \sigma_\zeta^2$. Equivalently, let $\boldsymbol{\zeta}^* = (\mathbf{W}_K^\top \mathbf{D}_K \mathbf{W}_K)^{-1/2} \boldsymbol{\delta}^*$, where $\boldsymbol{\delta}^* \sim \text{MVN}(0, \sigma_\zeta^2 \mathbf{I})$, then the above equation becomes the following:

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\gamma} + \mathbf{U}_K \mathbf{D}_K \mathbf{W}_K (\mathbf{W}_K^\top \mathbf{D}_K \mathbf{W}_K)^{-1/2} \boldsymbol{\delta}^* + \boldsymbol{\varepsilon}.$$

3.5. *Formulation of β-fields as thin plate regression splines.* We consider modeling the β-fields as TPRS using the relationship between penalized splines and mixed models [Ruppert, Wand and Carroll (2003)]. Following the above formulation, we can approximate $\beta_j(\mathbf{s})$ in (1) as $\mathbf{T}\boldsymbol{\gamma}_j + \mathbf{Z}^* \boldsymbol{\delta}_j^*$, where $\mathbf{T}$ contains the spatial coordinates of the monitoring locations, $\boldsymbol{\gamma}_j$ is a 2×1 vector of fixed unknown coefficients, $\mathbf{Z}^* = \mathbf{U}_K \mathbf{D}_K \mathbf{W}_K (\mathbf{W}_K^\top \mathbf{D}_K \mathbf{W}_K)^{-1/2}$, and $\boldsymbol{\delta}_j^*$ is a $K - 3 \times 1$ vector distributed as $\text{MVN}(\mathbf{0}, \tau_j^2 \mathbf{I})$.

We can succinctly incorporate this approximation into our modeling framework as follows. First, augment the design matrices $\mathbf{X}_j$ by appending the matrix $\mathbf{T}$ so that $\mathbf{X}_j^* = (\mathbf{X}_j \ \mathbf{T})$ for $j = 1, \ldots, m$ (if $\mathbf{X}_j$ already contains the spatial coordinates as predictors, then this step is unnecessary). Additionally, append the vector $\boldsymbol{\gamma}_j$

to the $\boldsymbol{\alpha}_j$ so that $\boldsymbol{\alpha}_j^{*\top} = (\boldsymbol{\alpha}_j^\top \ \boldsymbol{\gamma}_j^\top)$ for $j = 1, \ldots, m$. Last, letting $\mathbf{Z}_B^*$ be a block-diagonal matrix with diagonal elements $\{\mathbf{Z}^*\}_{j=1}^m$ and $\boldsymbol{\alpha}^*$ and $\tilde{\mathbf{B}}^*$ be the stacked vectors of $\boldsymbol{\alpha}_j^*$ and $\boldsymbol{\delta}_j^*$ for $j = 1, \ldots, m$, respectively, we formulate the TPRS version of the spatio-temporal model as a linear mixed model, as follows:

$$(5) \qquad \mathbf{Y} = \mathbf{F}\mathbf{X}^*\boldsymbol{\alpha}^* + \mathbf{F}\mathbf{Z}_B^*\tilde{\mathbf{B}}^* + \mathbf{F}\mathbf{P} + \mathbf{V}.$$

We note the similarities between equations (2) and (5). In fact, Nychka showed that thin plate splines are equivalent to kriging using a generalized covariance function [Nychka (2000)]. It is clear that the difference between LRK and TPRS has to do primarily with the choice of basis functions. However, we also emphasize that the TPRS bases are not dependent on any additional (e.g., range) parameters. Estimation of model parameters and prediction follows as described in Section 3.3.

4. Computational considerations.
Evaluation of (3) directly is computationally intensive, with the number of computations growing as $\mathcal{O}(N^3)$. However, the computational burden can be eased considerably by taking advantage of the block-diagonal nature of the Σ_B and Σ_V. Namely, Lindström showed that reformulation of (3) can reduce the computational burden to $\mathcal{O}(m^3 n^3)$ [Lindström et al. (2013)]. Typically, low-rank models boast a computational advantage over their full-rank counterparts. Yet, reducing the computational burden in spatio-temporal data is nuanced. In the following, we discuss how the formulation of the β-fields using either LRK or TPRS impacts computation. We illustrate the computational burden of calculating (3) by considering the determinant term $|\tilde{\Sigma}|$, employing a similar reformulation to that employed in Lindström et al. (2013) to exploit the block diagonal nature of Σ_B and Σ_V. Proofs of the following results and the corresponding reformulation of the full likelihood in (3) are provided in the Online Supplement [Olives et al. (2014)].

By application of known identities, it can be shown that

$$|\tilde{\Sigma}| = |\mathbf{F}\mathbf{Z}_B \Sigma_{\tilde{B}} \mathbf{Z}_B^\top \mathbf{F}^\top + \mathbf{F}\Sigma_P \mathbf{F}^\top + \Sigma_V|$$

$$(6) \qquad = |\Sigma_{\tilde{B}}||\Sigma_P||\Sigma_V||\Sigma_P^{-1} + \mathbf{F}^\top \Sigma_V^{-1}\mathbf{F}|$$

$$\times |\Sigma_{\tilde{B}}^{-1} + \mathbf{Z}_B^\top \mathbf{F}^\top (\Sigma_V^{-1} - \Sigma_V^{-1}\mathbf{F}(\mathbf{F}^\top \Sigma_P^{-1}\mathbf{F} + \Sigma_V^{-1})^{-1}\mathbf{F}^\top \Sigma_V^{-1})\mathbf{F}\mathbf{Z}_B|.$$

For highly unbalanced data like that which we typically encounter in MESA Air, (6) is dominated by the calculation of $|\Sigma_P^{-1} + \mathbf{F}^\top \Sigma_V^{-1}\mathbf{F}|$. Computation of this component grows at $\mathcal{O}(m^3 n^3)$, the same rate as the full-rank model.

As mentioned, the full-rank spatio-temporal model originally published by Szpiro did not include the nugget, $\mathbf{P}$, in the β-fields [Szpiro et al. (2010)]. When the nugget is not present, the determinant $|\tilde{\Sigma}|$ reduces to

$$(7) \qquad \begin{aligned} |\tilde{\Sigma}| &= |\mathbf{F}\mathbf{Z}_B \Sigma_{\tilde{B}} \mathbf{Z}_B^\top \mathbf{F}^\top + \Sigma_V| \\ &= |\Sigma_{\tilde{B}}||\Sigma_V||\Sigma_{\tilde{B}}^{-1} + \mathbf{Z}_B^\top \mathbf{F}^\top \Sigma_V^{-1}\mathbf{F}\mathbf{Z}_B|. \end{aligned}$$

Interestingly, in (7), computation will generally be dominated by calculation of $|\tilde{\Sigma}_B^{-1} + \mathbf{Z}_B^\top \mathbf{F}^\top \Sigma_V^{-1} \mathbf{F} \mathbf{Z}_B|$, which grows at $\mathcal{O}(m^3 K^3)$. This makes it clear that, when the nugget is not present, reducing the rank of the β-fields can lead to some improvement in terms of computation. We note that in the case where the data are more balanced, it is possible that computation of $|\Sigma_V|$ (or, equivalently, Σ_V^{-1}), which grows at $\mathcal{O}(\sum_t n_t^3)$, will dominate computation in both cases.

In Figure 4, we plot the CPU time required for optimized log-likelihood evaluation in full-rank and reduced-rank models with $K = 25$ with and without the nugget present for both LRK and TPRS. We see that for LRK and TPRS, as the number of sites increases, full-rank models take large steps in computation time required, whereas reduced-rank models grow much more slowly when a nugget is
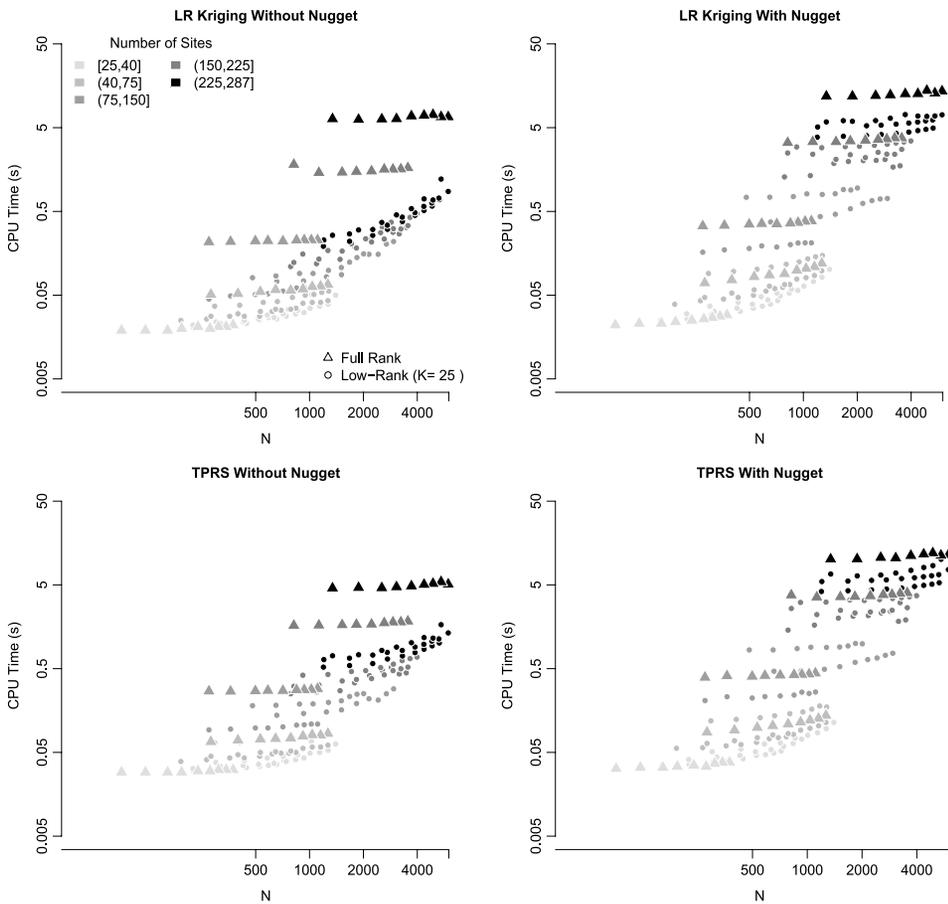

FIG. 4. *CPU time required for a single log-likelihood evaluation of LRK and TPRS models for the EPA AQS and MESA Air NO_x monitoring data in Los Angeles, California. Triangles indicate models where the rank of the spatial smooth is equal to the number of sites and circles indicate models where the rank of the smooth is equal to 25 in various depleted MESA Air data sets.*

not present. However, there is very little difference in computational growth between full- and reduced-rank models as the number of sites increases when the nugget is present.

5. Application to NO$_x$ monitoring data in Los Angeles.   We apply the proposed reduced-rank spatio-temporal models to NO$_x$ data collected in the Los Angeles area as part of the MESA Air monitoring campaign and via the EPA regulatory network.

5.1. *Models considered.*   We fit a variety of models to the data which vary in three aspects: (1) the choice of spline basis, (2) the rank of β-field smooth, and (3) the inclusion of the nugget. In all models considered, we employ two time trends ($m = 2$) as depicted in Figure 3. Likewise, the residual ν-field is always specified as exponentially distributed with a nugget. And, last, all of the GIS covariates are present in each of the $\mathbf{X}_j$ matrices.

5.1.1. *Choice of spline basis.*   We have outlined two possible classes of spline bases, exponential (used in LRK) and thin plate splines. As previously indicated, the use of exponential basis functions requires handling of the range parameters in each of the β-fields by either fixing its value at some *ad hoc* data-derived value or through full optimization. To investigate the trade-off between optimization of an additional range parameter and fixing this parameter at an arbitrary conservative value, we assume the range parameters in the β-fields are both fixed, and in separate models that they are estimated. To assess the sensitivity to the fixed value, we set the range parameters in all fields equal to the maximum, one half, one quarter and one eighth of the observed maximum spatial range in the data (80.7 km). Additionally, to assess the sensitivity of model performance to the choice of spline basis, we fit TPRS smooths to the β-fields.

5.1.2. *Rank of smooth.*   As a general rule of thumb, Ruppert, Wand and Carroll suggest that the number of knots, K, be chosen as $\max(20, \min\{150, n/4\})$ [Ruppert, Wand and Carroll (2003)]. In the case of our MESA Air and EPA data, this would result in $K = 71$. Although this rule of thumb is convenient, it is unclear how the number of knots in the spatial component of the mean model will influence spatio-temporal prediction. For our purposes, we explore a variety of different ranks on spatio-temporal prediction, $K = 287, 100, 50$ and 25. We note that the models with $K = 287$ correspond to full-rank models.

Knot location can also play an important role in LRK. Kammann and Wand choose knot locations using efficient space-filling algorithms [as implemented by the `cover.design()` function in the R package `fields`] [Kammann and Wand (2003)]. In our primary investigations, we choose knot locations using space-filling of monitoring sites within the study area (see Figure 5). Although space-filling of observed locations is a convenient approach to choosing the knot
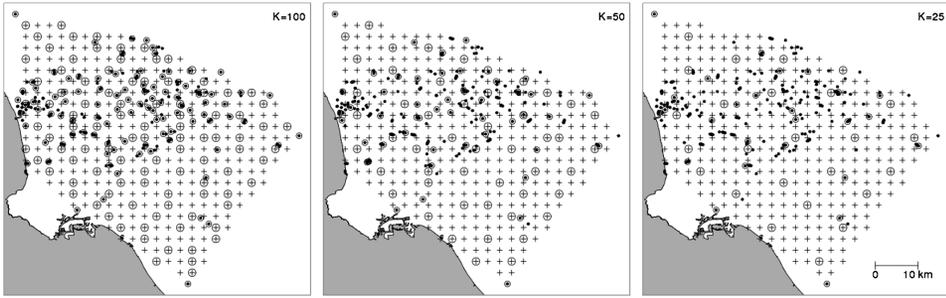
FIG. 5. *Map of knot locations chosen by efficient space-filling of monitoring locations (small open circle) and of regular grid locations (large open circles). Small black dots represent participant locations and crosses represent grid locations. "Home outdoor" monitors have been jittered for participant confidentiality.*

locations in our analysis, it is natural to consider knots chosen at alternative locations. For example, an attractive option could be to specify knot locations on a regular grid over the study area. To investigate, in addition to the primary analysis, we also fit models where knot locations are chosen using space-filling of a regular grid of the convex hull of the study region, where each grid cell is approximately 2.5 kilometers on each side (see Figure 5).

5.1.3. *Nugget effect.*   Given the analytical findings suggesting that reduced-rank modeling leads to a computational advantage in the case when the nugget is not present, we fit models both with and without the nugget. However, we note that while analytically feasible, models which exclude the nugget from the β-fields are less conceptually defensible. Namely, exclusion of the nugget from the β-fields makes it difficult for the model to capture fine-scale variability in the mean process. Moreover, preliminary investigations showed that very low-rank smooths in models without a nugget in the β-fields were unstable. As such, we present a limited set of results for reduced-rank models where the nugget is not present.

5.2. *Model validation.*   We employ cross-validation to assess model predictive performance. Our primary interest is in prediction of long-term averages of NO_x concentrations. Unfortunately, in this data set there are only 26 AQS and/or MESA "fixed sites" that provide adequately long time-series for long-term average validation. These sites tend to be more homogeneous in their geographic covariate distribution and have larger spatial spread when compared to MESA participant locations, which could potentially limit our ability to adequately assess predictive performance.

As such, in addition to cross-validation of AQS and MESA "fixed sites," we also consider cross-validation of MESA "community snapshot" and "home outdoor" sites. We apply tenfold cross-validation to each type of monitor. In each of these three scenarios, all remaining data are used to estimate model parameters and to predict at left-out locations.

Due to the varying nature of sampling at sites in each of the three scenarios, Lindström suggests calculating RMSE and R^2 slightly differently in each case [Lindström et al. (2013)]. At "fixed"/AQS sites, we calculate RMSE and R^2 metrics on both the 2-week and long-term average scales. Long-term averages at left-out sites are computed only over times where data are observed, so that

$$c(s) = \sum_{\tau \,:\, \exists y(s, \tau)} \frac{\exp\{y(s, \tau)\}}{|\{t : \exists y(s, t)\}|}.$$

The cross-validated R^2 on the long-term average scale is given by [Szpiro, Sheppard and Lumley (2011)]

$$(8) \qquad R^2 = \min\left\{0, 1 - \frac{\text{RMSE}(\hat{c}(s))^2}{\text{Var}(c(s))}\right\}.$$

For the second scenario, we perform cross-validation of the "community snapshot" locations. We cross-validate all three sampling periods/seasons simultaneously and calculate cross-validated RMSE and R^2 by season. Doing so allows us to assess the spatial predictive ability of the model across multiple seasons. Likewise, as each of the "community snapshot" locations were sampled during the same two-week periods, we can view the resulting metrics as representative of the pure spatial predictive capacity of the model.

Last, we also consider cross-validation of "home outdoor" sites. As the "home outdoor" sites are repeatedly sampled over time and typically at different time points, much of the R^2 is likely to reflect a temporal signal, which is strong in these data. As such, in addition to the raw cross-validated R^2, we also consider a de-trended version of the R^2 where the variance, $\text{Var}(c(s))$, in (8) is replaced by the variance of observations after removing the predictions from a reference model that accounts for (some) temporal variability. Here, we use a reference model based on the spatial average of measurements at AQS/"fixed sites" at each time point. Thus, the de-trended R^2 represents the improvement in performance of our models compared to central site predictions commonly used in air pollution epidemiology studies [Pope et al. (1995)].

5.3. *Comparison with other reduced-rank spatio-temporal models.* Although the current model was developed specifically to address the complexities arising in the context of MESA Air, a number of other methods for reduced-rank spatial and spatio-temporal modeling have been published, including fixed-rank filtering, Gaussian Markov random field approximations, covariance tapering, predictive processes and generalized additive models. Unfortunately, fixed-rank filtering is not available in an off-the-shelf package, and implementing this model for these data is a project unto itself. We further note that the application of Gaussian Markov random field approximations and covariance tapering in this setting is nuanced and may not result in any computational savings for these data. See the

Online Supplement [Olives et al. (2014)] for further discussion of the application of these two approaches in the current modeling framework.

As mentioned previously, there appears to be an explicit correspondence between predictive processes and LRK, as noted in Banerjee et al. (2008). As such, formally modeling the β-fields in (1) as reduced-rank predictive processes would not provide any additional insight into this work. That being said, one version of a predictive process *spatio-temporal* model is implemented in the spBayes package in R. Namely, the function spDynLM fits the following model:

$$y(s, t) = \mathbf{X}_t(\mathbf{s})\beta_t + u_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \qquad t = 1, 2, \ldots, T,$$
$$\varepsilon_t(\mathbf{s}) \sim N(0, \tau_t^2),$$
$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \Sigma_\eta),$$
$$\beta_0 \sim N(m_0, \Sigma_0),$$
$$u_t(\mathbf{s}) = u_{t-1}(\mathbf{s}) + w_t(\mathbf{s}), \qquad w_t(\mathbf{s}) \sim \mathrm{GP}(0, C_t(\cdot, \theta_t)),$$
$$u_0(s) = 0.$$

The spatial process w_t, here assumed to be exponential, can be replaced with a predictive process of reduced rank to reduce computational burden. This model significantly deviates from our own and may not perform well in the context of such highly imbalanced data as that which we analyze here. Nevertheless, we apply it to our data in an effort to make a fair comparison between published approaches to reduced-rank spatio-temporal modeling and our method. Specifically, we fit two models:

1. full-rank model ($K = 287$) for all w_t fields, and
2. reduced-rank ($K = 50$) for all w_t with knots chosen on a grid.

We note that the spDynLM function requires that knots be chosen on a grid when utilizing the reduced-rank predictive process machinery. In both cases, we fit the models assuming the following priors for the θ_t, Σ_η, τ_t^2:

$$1/\phi_t \sim \mathrm{Unif}\big(1/(0.9 \times \max \text{ distance}), 3/(0.05 \times \max \text{ distance})\big),$$
$$\sigma_t^2 \sim \mathrm{InvGamma}(2, 10),$$
$$\tau_t^2 \sim \mathrm{InvGamma}(2, 5),$$
$$\Sigma_\eta^2 \sim \mathrm{InvWish}(2, 0.001\mathbf{I}_p).$$

These priors are largely based on the example code available in the spDynLM documentation, with some small changes to reflect the data. Model predictions were the median of 500 posterior draws, after a burn-in period of 1500. We cross-validated these models for "fixed sites" using the same cross-validation groups as before.

Last, for an additional comparison with methods available in off-the-shelf soft-
ware, we considered a generalized additive model that reformulates the mean pro-
cess $\mu(s, t)$ without resorting to a dynamic model. Namely, we replaced $\mu(s, t)$
with the following:

$$\mathbf{X}(\mathbf{s})\boldsymbol{\alpha} + \eta_t + g(\mathbf{s}) + h(\mathbf{s}, t).$$

Here both g and h are modeled using TPRS. For investigating models with spatial
rank of K, we set the degrees of freedom for g equal to K and the degrees of
freedom for h equal to $K \times 14$ (e.g., when $K = 50$, h has 700 df), where 14 is
the number of years represented in the data. Note that both g and h can be viewed
as penalized regression splines with structure similar to what we outline in the
paper. But for h, we are now assuming a nonseparable model for space and time
which differs from the tensor product approach used in our model. Moreover, we
do not rely on predefined temporal basis functions to model time. The η_t are i.i.d.
Gaussian random effects that capture nonsmooth temporal variation. Note that the
v-field remains the same as outlined in the paper. We fit this model using the `gamm`
function in the `mgcv` package in `R`.

6. Results.

6.1. *Performance of proposed reduced-rank models in LA.* Table 2 shows
the results of the cross-validation at "fixed sites" for models when the nugget is

TABLE 2
*Cross-validated RMSE and R^2 for "fixed sites" when nugget is present. R^2 have been multiplied
by 100 for presentation*

Basis/K	RMSE					R^2				
	287	100	50	25	0	287	100	50	25	0
2-wk										
LRK ($\phi = $ est)	15.43	15.52	15.72	16.35	17.82	85	85	85	83	80
LRK ($\phi = $ max)	15.64	15.64	15.83	15.87	17.82	85	85	84	84	80
LRK ($\phi = $ max/2)	15.52	15.56	15.24	16.14	17.82	85	85	86	84	80
LRK ($\phi = $ max/4)	15.31	15.32	15.33	15.74	17.82	85	85	85	85	80
LRK ($\phi = $ max/8)	15.04	15.08	15.13	15.59	17.82	86	86	86	85	80
TPRS	16.38	15.29	15.11	16.01	17.82	83	85	86	84	80
LTA										
LRK ($\phi = $ est)	10.43	10.56	11.11	11.41	12.41	68	67	64	62	55
LRK ($\phi = $ max)	10.48	10.46	10.53	10.84	12.41	68	68	67	65	55
LRK ($\phi = $ max/2)	10.40	10.39	10.08	10.72	12.41	68	68	70	66	55
LRK ($\phi = $ max/4)	10.30	10.31	10.33	11.04	12.41	69	69	69	64	55
LRK ($\phi = $ max/8)	10.26	10.28	10.36	10.85	12.41	69	69	68	65	55
TPRS	10.83	9.99	9.88	10.60	12.41	65	71	71	67	55

present. For LRK models, the choice of range does not appear to be a strong determinant of the predictive performance, with fully optimized models performing nearly as well as those models with the range parameter fixed at various values. Likewise, TPRS models exhibit highly competitive predictive performance with a slight edge over LRK models at lower ranks for long-term averages. Cross-validated R^2 values stay relatively consistent across ranks until $K = 25$, at which point both 2-week and long-term average predictive scores drop off. In all cases, models with some spatial smoothing ($K > 0$) perform better than models without any smoothing ($K = 0$).

Table 3 show the results of cross-validation at "community snapshot" sites. We typically see the best performance in the Winter as compared with the Fall and Spring seasons. Once again, there appears to be little difference in model performance as the choice of range parameters varies. TPRS models continue to compete strongly with LRK models. The rank of the β-field smooth does not tend to influence performance heavily, although again spatial smoothing at any rank does tend to improve predictive performance.

TABLE 3

Cross-validated RMSE and R^2 for "community snapshot" locations when nugget is present. R^2 have been multiplied by 100 for presentation

Basis/K	RMSE					R^2				
	287	100	50	25	0	287	100	50	25	0
Summer										
LRK ($\phi =$ est)	6.74	6.71	6.73	6.98	6.97	66	66	66	63	63
LRK ($\phi =$ max)	6.68	6.67	7.03	6.70	6.97	66	66	63	66	63
LRK ($\phi =$ max/2)	6.69	6.68	6.66	6.96	6.97	66	66	66	63	63
LRK ($\phi =$ max/4)	6.71	6.72	6.66	6.76	6.97	66	66	66	65	63
LRK ($\phi =$ max/8)	6.76	6.74	6.83	6.82	6.97	65	66	65	65	63
TPRS	6.62	6.71	6.70	6.72	6.97	67	66	66	66	63
Fall										
LRK ($\phi =$ est)	11.64	11.61	11.99	11.55	11.78	75	76	74	76	75
LRK ($\phi =$ max)	11.66	11.59	11.75	11.83	11.78	75	76	75	75	75
LRK ($\phi =$ max/2)	11.66	11.64	11.61	11.97	11.78	75	75	76	74	75
LRK ($\phi =$ max/4)	11.65	11.63	11.53	11.35	11.78	75	75	76	77	75
LRK ($\phi =$ max/8)	11.66	11.89	11.95	11.86	11.78	75	74	74	74	75
TPRS	11.92	12.10	12.14	12.11	11.78	74	73	73	73	75
Winter										
LRK ($\phi =$ est)	13.01	12.92	12.98	12.99	15.32	77	77	77	77	68
LRK ($\phi =$ max)	13.04	12.94	13.11	14.00	15.32	77	77	76	73	68
LRK ($\phi =$ max/2)	13.03	12.99	12.59	13.63	15.32	77	77	78	75	68
LRK ($\phi =$ max/4)	13.02	12.98	12.63	13.8	15.32	77	77	78	74	68
LRK ($\phi =$ max/8)	13.05	13.51	12.65	13.95	15.32	77	75	78	73	68
TPRS	13.27	13.04	13.08	14.19	15.32	76	77	77	72	68

TABLE 4

Cross-validated RMSE and R^2 for "home outdoor" locations when nugget is present. R^2 have been multiplied by 100 *for presentation*

Basis/K	RMSE					R^2				
	287	100	50	25	0	287	100	50	25	0
Raw										
LRK ($\phi =$ est)	5.51	5.52	5.88	5.89	7.03	93	93	92	92	88
LRK ($\phi =$ max)	5.54	5.54	5.71	6.41	7.03	93	93	92	90	88
LRK ($\phi =$ max/2)	5.53	5.54	5.28	6.01	7.03	93	93	93	92	88
LRK ($\phi =$ max/4)	5.53	5.58	5.52	6.14	7.03	93	93	93	91	88
LRK ($\phi =$ max/8)	5.52	5.49	5.48	6.26	7.03	93	93	93	91	88
TPRS	5.52	5.50	5.53	6.00	7.03	93	93	93	92	88
Detrended										
LRK ($\phi =$ est)						84	84	82	82	75
LRK ($\phi =$ max)						84	84	83	79	75
LRK ($\phi =$ max/2)						84	84	86	81	75
LRK ($\phi =$ max/4)						84	84	84	81	75
LRK ($\phi =$ max/8)						84	84	85	80	75
TPRS						84	84	84	81	75

Table 4 shows the results of the cross-validation study at "home outdoor" locations. Here, the choice of range parameter model appears to have even less of important role locations than it did at "fixed sites." Namely, cross-validated RMSE increases only slightly, resulting in a minimal decrease in R^2, as the rank decreases in the raw home predictions. Detrended R^2 did show some decay as the rank decreased, but still remained relatively high. TPRS models performed as well as LRK models across ranks. Again, models with some spatial smoothing outperformed those models with no smoothing.

Figure 6 compares the cross-validated R^2 for a set of models of rank $K =$ 287, 100, 50 and 25 with and without the nugget present in the β-fields at AQS/"fixed sites," "community snapshot" and "home outdoor" locations. Note, for LRK results, the range parameter has been estimated from the data. The figure suggests that while full rank models ($K = 287$) are comparable across these two specifications, predictive performance of models without the nugget in the β-fields tend to drop off rapidly as the rank of the smooth decreases, particularly in the case of LRK, where in select cases the R^2 decreases to zero when $K = 25$. TPRS models tend to be more robust, although the decrease in R^2 in TPRS models without a nugget tends to be greater than in TPRS models with a nugget.

Figure 7 compares the results of fitting full and LRK models to the MESA Air data when the knots were chosen using space-filling of either monitoring locations or a regularly spaced grid of locations. Generally speaking, models where knots were chosen at monitoring sites performed better than those where knots were chosen at grid locations.
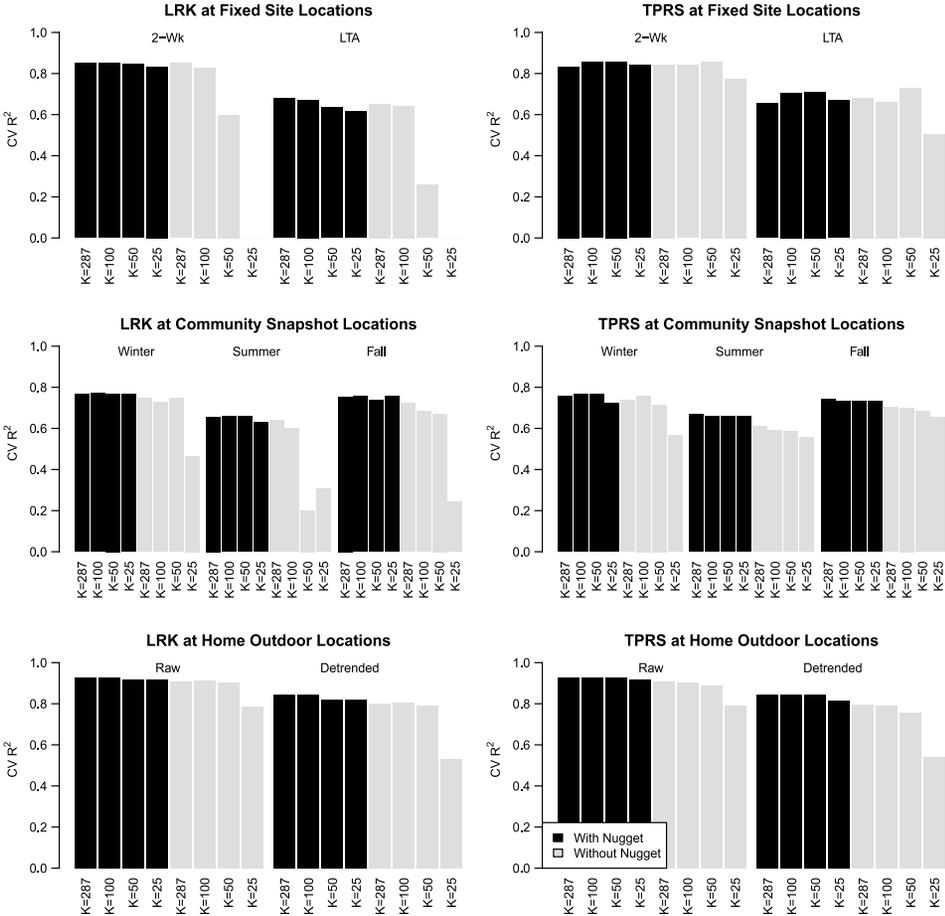
FIG. 6. *Comparison of cross-validated R^2 at "fixed site," "community snapshot," and "home outdoor" locations using low-rank kriging and TPRS.*

6.2. *Performance of other reduced-rank spatio-temporal modeling methods.* We found that the `spDynLM` implementation did not work well for our data, possibly due to the large imbalance across space and time. In both models ($K = 50, 287$), the time-varying range parameter was not well identified and varied significantly, thus resulting in poor characterization of the rate of spatial decay. The temporal sparsity of the data may also have contributed to the poor performance due to the dynamic nature of the model's temporal trend. While the in-sample fits for these models are quite good, the out-of-sample predictions are highly variable, resulting in cross-validated R^2 equal to zero for all scenarios considered. In Figure 8 we show the scatter plots of observed and predicted values in fitted models. The histograms in this same figure represent the distribution of predictions at unobserved times and locations. We note that these are on the log-scale, so
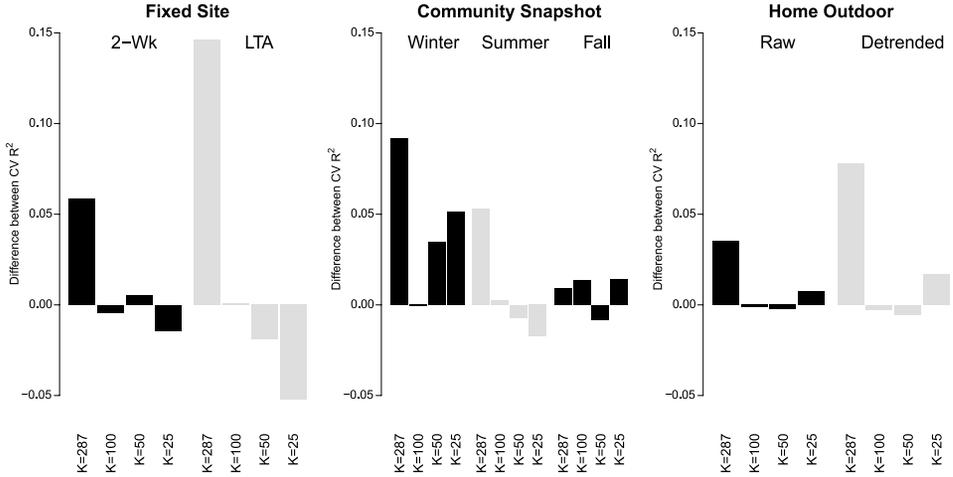
FIG. 7. *Differences between cross-validated R^2 in LRK models with knots chosen at monitoring locations and on a regular grid by rank. Models assume that the nugget, **P**, is present in all β-fields and the range parameters are estimated by maximum likelihood.*
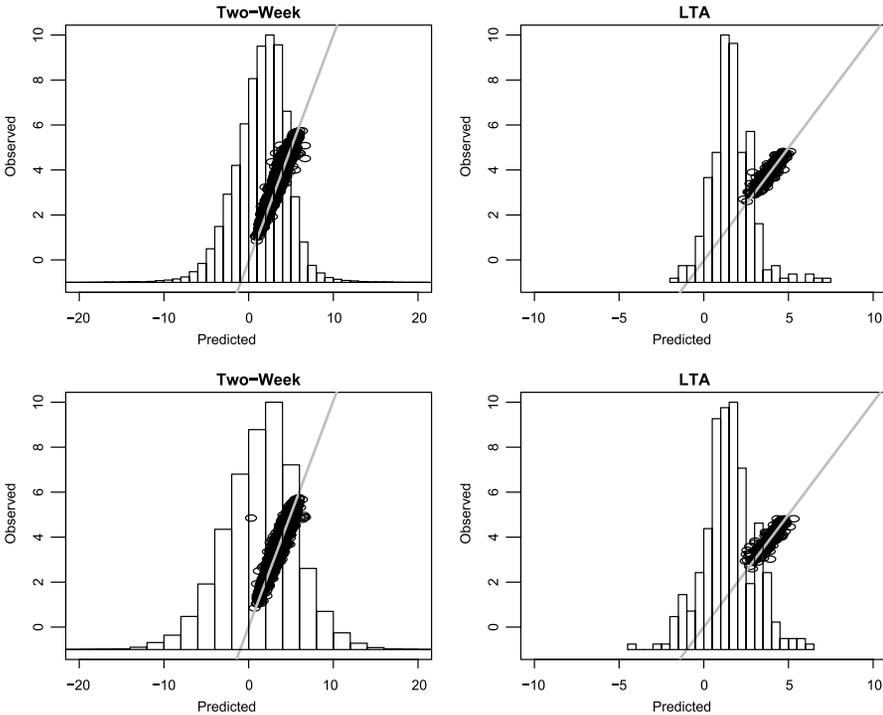
FIG. 8. (*Top row*) In-sample fits *for full-rank models fit in* spBayes. (*Bottom row*) In-sample fits *for reduced-rank models* ($K = 50$) *fit in* spBayes. *Histograms represent the distribution of posterior predictions at time points/locations without observed data.*
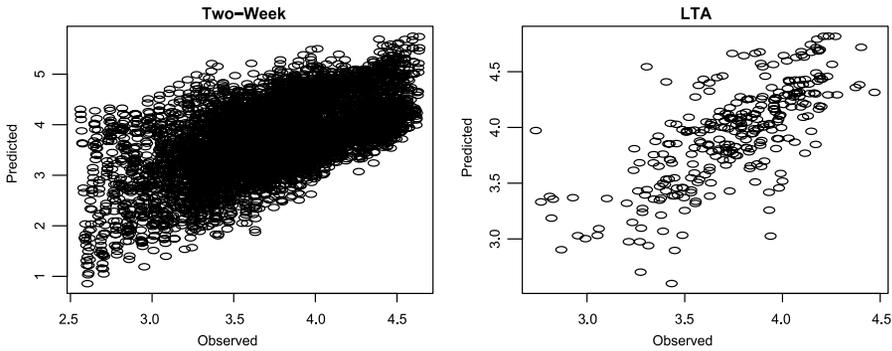
FIG. 9.   In-sample fits *for reduced-rank models fit in* mgcv *with $K = 50$ on the two-week scale* (*left*) *and long-term average scale* (*right*).

that when exponentiated to the native scale, many predicted values at unobserved times/locations are extremely large.

The results of our gamm implementation were only marginally better. While the in-sample fits of this approach were more promising (see Figure 9), the predictions were nowhere near the caliber of those achieved using our model. Inspection of the residuals suggests that there remains significant temporal correlation that is unaccounted for by the mean model. We found that the cross-validated R^2 was equal to 0 on both the two-week and long-term average scale using this approach. This low R^2 was driven by the presence of outlying predictions for a handful of sites in two different cross-validation groups. Additionally, the model failed to converge for a single cross-validation group.

7. Discussion.   This paper focuses on presentation of LRK and TPRS representations of the mean process in the spatio-temporal model proposed by Szpiro et al. (2010), Sampson et al. (2011), and Lindström et al. (2013). Our approach allows for a reduced-rank representation of the β-fields in the mean process of the original model, which tends to be the most time-consuming piece to evaluation in likelihood optimization. In certain cases, we have shown that such reduced-rank representations of the β-fields can lead to a computational advantage over the full rank specification. Namely, when the nugget of the β-field is not present, we have shown that our low-rank approach leads to slower growth in the CPU time required for likelihood evaluation.

The formulation of the β-fields in the mean process of the model as spatial splines is attractive for a number of other reasons. For example, oftentimes predictions of air pollution concentrations are used as inputs into health models to estimate health effects. Typically, the predictions are based on spatially misaligned data and ignoring this fact can lead to biased results and overly optimistic standard errors [Szpiro, Sheppard and Lumley (2011)]. The expression of the β-fields as splines places GIS covariates and spatial smoothing on more equal footing.

Namely, in this form we can think of the GIS covariates and the spatial basis functions as unpenalized and penalized spatial covariates, respectively. This interpretation leads to a more coherent approach to measurement error correction for spatially misaligned data [Szpiro and Paciorek (2013)]. It is also important to note that the computational advantage gained in log-likelihood evaluation extends analogously to prediction, thus reducing computation time needed to predict at potentially many new locations.

For LRK models, we explored the choice of range parameters of prediction, ranging from the case where the range was fully estimated from the data to the case where it was fixed at an arbitrary conservative value indicated by the data. In the scenario when the range parameter is fixed, we showed that the original specification of the full-rank model can also be interpreted as a standard penalized spatial spline.

Likewise, we discussed the parallels between kriging and TPRS. We emphasize that a limitation of the kriging basis functions is the reliance on the range parameter and that TPRS is not subject to the same limitation. That being said, we note that there is an equivalence between thin plate splines and kriging using a Matern-covariance with infinite range [Kimeldorf and Wahba (1970), Nychka (2000), Wahba (1981)]. As such, one might view the use of TPRS as making an implicit assumption about the range parameter. The fact that TPRS and LRK were competitive in our results indicates that TPRS is a valid and attractive option for spatial smoothing in these models. To further this argument, we performed additional analyses (results included in the Online Supplement [Olives et al. (2014)]) comparing out-of-sample prediction variances and AIC as a means of model selection. These analyses indicated that TPRS models tended to result in more stable prediction variances across rank specification when compared to LRK models. However, there was little notable difference between AIC values in LRK and TPRS models. Rather, AIC values indicated full-rank LRK models were preferable to reduced-rank ones in all cases. TPRS models with $K = 100$ had the lowest AIC.

Our approach to model assessment relies on cross-validation. As we are primarily interested in prediction of long-term averages, the cross-validation approach outlined isolates the spatial predictive capacity of the models. We applied our approach to ambient MESA Air and EPA NO_x data collected in the Los Angeles area as well as traditional road covariates and Caline point predictions models. We found that generally speaking, the choice of the range parameter in the LRK exponential spatial basis functions had little impact on the model performance. In fact, reducing the rank of the model tended to also have little impact in most cross-validation scenarios for ranks of moderate size ($K = 50, 100$). We note that the recommendation of Ruppert, Wand and Carroll ($K = 71$ for the MESA Air data) falls squarely in this range [Ruppert, Wand and Carroll (2003)]. However, we found that reduction of the rank of the β-fields below $K = 50$ tended to noticeably impact model predictions. This impact was further exacerbated by exclusion

of the nugget in the β-fields. This finding is not a surprise, as exclusion of the nugget in the β-fields amounts to attributing all extra variation in the mean beyond what is explained by the GIS covariates to the spatial β-fields. Reduction of the rank of the smooth of these random fields results in a spatial smooth that is unlikely to be able to capture spatial heterogeneity.

This unfortunate finding is at odds with the goal of reducing the computational burden of full-rank spatio-temporal likelihood evaluations. Although the original specification published by Szpiro et al. did not include a nugget in the β-field, it is our feeling that such models are less defensible than those that include a nugget, since it is unlikely that the GIS covariates in the model account for all nonsmooth spatial variation.

That being said, the results herein described are based on a single data setting. Indeed, there almost surely exists other data sets where inclusion of a nugget in the β-fields is contraindicated. In these cases, use of a moderate rank smooth could lead to both a computational and predictive advantage.

Last, we examined a number of other approaches and specification to modeling NO_x concentrations in the current data set and found poor performance for two off-the-shelf packages. Our findings confirm that the long history of methodological development of the model under study in the context of modeling air pollution exposures for MESA Air was indeed well guided and that current off-the-shelf packages are not ideal for analyzing these data. Future research should, however, include investigations into the extension of the current model using covariance tapering of either the β-fields covariance or even of the overall covariance matrix $\tilde{\Sigma}$.

The LA NO_x data application is meant to exemplify the current methods. However, we note that this model is being applied more broadly to four separate pollutants in six major cities in the United States as part of MESA Air [Keller et al. (2014)]. Furthermore, a rigorous approach to model selection, that varies the number of trends, covariates and β-field models, is also being applied to choose the best performing predictive models. Taking into account cross-validation, this effort includes the fitting of hundreds of models, representing a significant investment of time on the part of MESA Air investigators. To further emphasize the impact of the current methods, we performed a separate set of analyses replicating a large subset of the cross-validation scenarios for NO_x data in Los Angeles considered by MESA Air investigators in their development of exposure models for use in primary MESA Air health analyses. We found that TPRS models achieved highly competitive results in roughly half the time, suggesting that had these methods been available during model development, potentially hundreds of computer hours could have been saved during the model development process. As we move toward incorporating the current methods into the highly optimized `SpatioTemporal` package, and further optimize the reduced-rank model fitting procedures, we expect that the gains in computational time will increase in orders of magnitude, to roughly 5 times faster. As such, we believe that the current work will continue to have tangible implications for MESA Air investigators and their collaborators who

continue to use the MESA Air spatio-temporal model as the basis for exposure assessment in air pollution cohort studies.

SUPPLEMENTARY MATERIAL

Supplement to "Reduced-rank spatio-temporal modeling of air pollution concentrations in the Multi-Ethnic Study of Atherosclerosis and Air Pollution" (DOI: 10.1214/14-AOAS786SUPP; .pdf). We provide a detailed derivation of the optimized likelihood, comparisons of the prediction variances, discussion model selection by AIC for the paper "Reduced-rank spatio-temporal modeling of air pollution concentrations in the Multi-Ethnic Study of Atherosclerosis and Air Pollution" by Casey Olives, Lianne Sheppard, Johan Lindström, Paul D. Sampson, Joel D. Kaufman and Adam A. Szpiro.

REFERENCES

BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 825–848. MR2523906

BRAUER, M., HOEK, G., VAN VLIET, P., MELIEFSTE, K., FISCHER, P., GEHRING, U., HEINRICH, J., CYRYS, J., BELLANDER, T., LEWNE, M. and BRUNEKREEF, B. (2003). Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems. *Epidemiology* **14** 228–239.

BYRD, R. H., LU, P., NOCEDAL, J. and ZHU, C. Y. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16** 1190–1208. MR1346301

CARROLL, M. L., DIMICELI, C. M., SOHLBERG, R. A. and TOWNSHEND, J. R. G. (2004). 250m MODIS Normalized Difference Vegetation Index, 250ndvi28920033435, Collection 4. Univ. Maryland, College Park, Maryland, Day 289, 2003.

COHEN, M. A., ADAR, S. D., ALLEN, R. W., AVOL, E., CURL, C. L., GOULD, T., HARDIE, D., HO, A., KINNEY, P., LARSON, T. V., SAMPSON, P., SHEPPARD, L., STUKOVSKY, K. D., SWAN, S. S., LIU, L. J. S. and KAUFMAN, J. D. (2009). Approach to estimating participant pollutant exposures in the multi-ethnic study of atherosclerosis and air pollution (MESA air). *Environmental Science & Technology* **43** 4687–4693.

CRAINICEANU, C. M., DIGGLE, P. J. and ROWLINGSON, B. (2008). Bivariate binomial spatial modeling of *Loa loa* prevalence in tropical Africa. *J. Amer. Statist. Assoc.* **103** 21–37. MR2420211

DOCKERY, D. W., POPE, C. A. 3RD, XU, X., SPENGLER, J. D., WARE, J. H., FAY, M. E., FERRIS, B. G. JR and SPEIZER, F. E. (1993). An association between air pollution and mortality in six U.S. cities. *N. Engl. J. Med.* **329** 1753–1759.

ECKHOFF, P. A. and BRAVERMAN, T. N. (1995). Addendum to the user's guide to CAL3QHC version 2.0 (CAL3QHCR user's guide). Technical Support Division, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

FRY, J., XIAN, G., JIN, S., DEWITZ, J., HOMER, C., YANG, L., BARNES, C., HEROLD, N. and WICKHAM, J. (2011). Completion of the 2006 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering & Remote Sensing* **77** 858–864.

FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.* **102** 321–331. MR2345545

FUENTES, M., GUTTORP, P. and SAMPSON, P. D. (2006). Using transforms to analyze space-time processes. *Monogr. Statist. Appl. Probab.* **107** 77.

GELFAND, A. E., BANERJEE, S. and GAMERMAN, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* **16** 465–479. MR2147537

GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*: *A Roughness Penalty Approach. Monographs on Statistics and Applied Probability* **58**. Chapman & Hall, London. MR1270012

HODGES, J. S. (2013). *Richly Parameterized Linear Models*. Chapman & Hall, Boca Raton.

HODGES, J. and CLAYTON, M. K. (2011). Random effects old and new. Technical report, Univ. Minnesota, Minneapolis, MN.

HOEK, G., BEELEN, R., DE HOOGH, K., VIENNEAU, D., GULLIVER, J., FISCHER, P. and BRIGGS, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* **42** 7561–7578.

JERRETT, M., ARAIN, A., KANAROGLOU, P., BECKERMAN, B., POTOGLOU, D., SAHSUVAROGLU, T., MORRISON, J. and GIOVIS, C. (2005a). A review and evaluation of intraurban air pollution exposure models. *J. Expo. Anal. Environ. Epidemiol.* **15** 185–204.

JERRETT, M., BURNETT, R. T., MA, R., POPE, C. A. 3RD, KREWSKI, D., NEWBOLD, K. B., THURSTON, G., SHI, Y., FINKELSTEIN, N., CALLE, E. E. and THUN, M. J. (2005b). Spatial analysis of air pollution and mortality in los angeles. *Epidemiology* **16** 727–736.

KAMMANN, E. E. and WAND, M. P. (2003). Geoadditive models. *J. Roy. Statist. Soc. Ser. C* **52** 1–18. MR1963210

KAUFMAN, J. K., ADAR, S. D., ALLEN, R. W., BARR, R. G., BUDOFF, M. J., BURKE, G. L., CASILLAS, A. M., COHEN, M. A., CURL, C. L., DAVIGLUS, M. L., DIEZ ROUX, A. V., JACOBS, D. R. JR, KRONMAL, R. A., LARSON, T. V., LIU, S. L., LUMLEY, T., NAVAS-ACIEN, A., O'LEARY, D. H., ROTTER, J. I., SAMPSON, P. D., SHEPPARD, L., SISCOVICK, D. S., STEIN, J. H., SZPIRO, A. A. and TRACY, R. P. (2012). Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease the multi-ethnic study of atherosclerosis and air pollution (MESA air). *American Journal of Epidemiology* **176** 825–837.

KELLER, J. P., OLIVES, C., KIM, S.-Y., SHEPPARD, L., SAMPSON, P. D., SZPIRO, A. A., ORON, A. P., LINDSTRÖM, J., VEDAL, S. and KAUFMAN, J. D. (2014). A unified spatiotemporal modeling approach for prediction of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environ. Health Perspect.* To appear.

KIMELDORF, G. S. and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502. MR0254999

KÜNZLI, N., JERRETT, M., MACK, W. J., BECKERMAN, B., LABREE, L., GILLILAND, F., THOMAS, D., PETERS, J. and HODIS, H. N. (2005). Ambient air pollution and atherosclerosis in Los Angeles. *Environ. Health Perspect.* **113** 201–206.

LINDSTRÖM, J., SZPIRO, A. A., SAMPSON, P. D., ORON, A., RICHARDS, M., LARSON, T. and SHEPPARD, L. (2013). A flexible spatio-temporal model for air pollution with spatial and spatio-temporal covariates. *Environ. Ecol. Stat.* 1–23.

MILLER, K. A., SISCOVICK, D. S., SHEPPARD, L., SHEPHERD, K., SULLIVAN, J. H., ANDERSON, G. L. and KAUFMAN, J. D. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *N. Engl. J. Med.* **356** 447–458.

NYCHKA, D. W. (2000). Spatial-process estimates as smoothers. In *Smoothing and Regression: Approaches, Computation, and Application* 393–424. Wiley, New York.

NYCHKA, D. and SALTZMAN, N. (1998). Design of air quality networks. In *Case Studies in Environmental Statistics* (D. Nychka, L. Cox and W. Piegorsch, eds.). *Lecture Notes in Statistics* **132** 51–76. Springer, New York.

OLIVES, C., SHEPPARD, L., LINDSTRÖM, J., SAMPSON, P. D., KAUFMAN, J. D. and SZPIRO, A. A. (2014). Supplement to "Reduced-rank spatio-temporal modeling of air pollution concentrations in the Multi-Ethnic Study of Atherosclerosis and Air Pollution." DOI:10.1214/14-AOAS786SUPP.

PACE, R. and LESAGE, J. (2009). A sampling approach to estimate the log determinant used in spatial likelihood problems. *Journal of Geographical Systems* **11** 209–225.

PACIOREK, C. J., YANOSKY, J. D., PUETT, R. C., LADEN, F. and SUH, H. H. (2009). Practical large-scale spatio-temporal modeling of particulate matter concentrations. *Ann. Appl. Stat.* **3** 370–397. MR2668712

POPE, C. A. 3RD, THUN, M. J., NAMBOODIRI, M. M., DOCKERY, D. W., EVANS, J. S., SPEIZER, F. E. and HEATH, C. W. JR (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am. J. Respir. Crit. Care Med.* **151** 669–674.

POPE, C. A. 3RD, BURNETT, R. T., THUN, M. J., CALLE, E. E., KREWSKI, D., ITO, K. and THURSTON, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association* **287** 1132–1141.

RITZ, B., WILHELM, M. and ZHAO, Y. (2006). Air pollution and infant death in southern California, 1989–2000. *Pediatrics* **118** 493–502.

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720

SAMET, J. M., DOMINICI, F., CURRIERO, F. C., COURSAC, I. and ZEGER, S. L. (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *N. Engl. J. Med.* **343** 1742–1749.

SAMPSON, P. D., SZPIRO, A. A., SHEPPARD, L., LINDSTRÖM, J. and KAUFMAN, J. D. (2011). Pragmatic estimation of spatio-temporal air quality model with irregular monitoring data. *Atmospheric Evnironment* **45** 6593–6606.

STEIN, M. L. (2007). Spatial variation of total column ozone on a global scale. *Ann. Appl. Stat.* **1** 191–210. MR2393847

STEIN, M. L. (2008). A modeling approach for large spatial datasets. *J. Korean Statist. Soc.* **37** 3–10. MR2420389

STROUD, J. R., MÜLLER, P. and SANSÓ, B. (2001). Dynamic models for spatiotemporal data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 673–689. MR1872059

SZPIRO, A. A. and PACIOREK, C. J. (2013). Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics* **24** 501–517. MR3161971

SZPIRO, A. A., SHEPPARD, L. and LUMLEY, T. (2011). Efficient measurement error correction with spatially misaligned data. *Biostatistics* **12** 610–623.

SZPIRO, A. A., SAMPSON, P. D., SHEPPARD, L., LUMLEY, T., ADAR, S. D. and KAUFMAN, J. D. (2010). Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics* **21** 606–631. MR2842271

TELEATLAS (2000). TeleAtlas Dynamap 2000. [CD_ROM], TeleAtlas, Lebanon, NH.

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16. MR0618629

WOOD, S. N. (2003). Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 95–114. MR1959095

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc*. **99** 250–261. MR2054303

C. OLIVES
J. D. KAUFMAN
DEPARTMENT OF ENVIRONMENTAL
  AND OCCUPATIONAL HEALTH SCIENCES
UNIVERSITY OF WASHINGTON
4225 ROOSEVELT WAY NE
SEATTLE, WASHINGTON 98105
USA
E-MAIL: colives@uw.edu
        joelk@uw.edu

J. LINDSTRÖM
MATHEMATICAL STATISTICS
CENTER FOR MATHEMATICAL SCIENCES
LUND UNIVERSITY
BOX 118, SE-221 00 LUND
SWEDEN
E-MAIL: johanl@maths.lth.se

L. SHEPPARD
A. A. SZPIRO
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF WASHINGTON
BOX 357232
HEALTH SCIENCES BUILDING ROOM F600
1705 NE PACIFIC
SEATTLE, WASHINGTON 98195-7232
USA
E-MAIL: sheppard@uw.edu
        aszpiro@uw.edu

P. D. SAMPSON
DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195-4322
E-MAIL: pds@stat.washington.edu