

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Publications, Department of Statistics

Statistics, Department of

---

2020

## Statistical Downscaling with Spatial Misalignment: Application to Wildland Fire PM<sub>2.5</sub> Concentration Forecasting

Suman Majumder

Yawen Guan

Brian J. Reich

Susan O'Neill

Ana G. Rappold

Follow this and additional works at: <https://digitalcommons.unl.edu/statisticsfacpub>



Part of the [Other Statistics and Probability Commons](#)

---

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



# Statistical Downscaling with Spatial Misalignment: Application to Wildland Fire PM<sub>2.5</sub> Concentration Forecasting

Suman MAJUMDER , Yawen GUAN, Brian J. REICH, Susan O'NEILL, and Ana G. RAPPOLD

Fine particulate matter, PM<sub>2.5</sub>, has been documented to have adverse health effects, and wildland fires are a major contributor to PM<sub>2.5</sub> air pollution in the USA. Forecasters use numerical models to predict PM<sub>2.5</sub> concentrations to warn the public of impending health risk. Statistical methods are needed to calibrate the numerical model forecast using monitor data to reduce bias and quantify uncertainty. Typical model calibration techniques do not allow for errors due to misalignment of geographic locations. We propose a spatiotemporal downscaling methodology that uses image registration techniques to identify the spatial misalignment and accounts for and corrects the bias produced by such warping. Our model is fitted in a Bayesian framework to provide uncertainty quantification of the misalignment and other sources of error. We apply this method to different simulated data sets and show enhanced performance of the method in presence of spatial misalignment. Finally, we apply the method to a large fire in Washington state and show that the proposed method provides more realistic uncertainty quantification than standard methods.

**Key Words:** Image registration; Public health; Smoothing; Warping.

---

The views expressed in this manuscript are those of the individual authors and do not necessarily reflect the views and policies of the US Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

---

S. Majumder (✉) · B. J. Reich, Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695, USA (E-mail: [smajumd2@ncsu.edu](mailto:smajumd2@ncsu.edu)) (E-mail: [bjreich@ncsu.edu](mailto:bjreich@ncsu.edu)). Y. Guan, Department of Statistics, University of Nebraska–Lincoln, 340 Hardin Hall North Wing, Lincoln, NE 68583, USA (E-mail: [yguan12@unl.edu](mailto:yguan12@unl.edu)). S. O'Neill, Pacific Northwest Research Station, United States Forest Service, 400 N 34th Street, Suite 201, Seattle, WA 98103, USA (E-mail: [susan\\_oneill@firenet.gov](mailto:susan_oneill@firenet.gov)). A. G. Rappold, United States Environmental protection Agency, 101 Manning Drive, Chapel Hill, NC 27514, USA (E-mail: [rappold.ana@epa.gov](mailto:rappold.ana@epa.gov)).

© 2020 International Biometric Society

*Journal of Agricultural, Biological, and Environmental Statistics*, Volume 26, Number 1, Pages 23–44  
<https://doi.org/10.1007/s13253-020-00420-4>

## 1. INTRODUCTION

Air pollution associated with wildland fire smoke is an increasingly pressing health concern (Dennekamp and Abramson 2011; Rappold et al. 2011; Johnston et al. 2012; Dennekamp et al. 2015; Haikerwal et al. 2015, 2016; Wettstein et al. 2018). Reliable short-term forecasts of fire-associated health risk using numerical models facilitate informed decision making for local populations. Numerical models produce forecasts on a coarse grid and are prone to bias. Assimilating point-level monitor data with numerical-model output can reduce bias and provide more realistic uncertainty quantification (e.g., Berrocal et al. 2010a,b; Kloog et al. 2011; Zhou et al. 2011, 2012; Berrocal et al. 2012; Reich et al. 2014; Chang et al. 2014). However, most downscaling methods only correct additive and scaling biases and fail to guard against spatial misalignment errors where a forecasted event occurs in a different spatial location than forecasted. Spatial misalignment error in this context implies errors corresponding to predicting the location of a feature, such as fire plume, are wrong. Not accounting for it is problematic for wildland fire smoke forecasting because a common source of error is in predicting the direction of the fire plume which cannot be accounted for by additive and scaling correction to the forecast. This motivates us to develop a statistical downscaling method that accounts for spatial misalignment errors.

Spatial misalignment correction can be achieved using standard image registration (or warping) techniques, ranging from simple affine and polynomial transformations to more sophisticated methods such as Fourier-based transforms (Kuglin 1975; De Castro and Morandi 1987), nonparametric approaches like elastic deformation (Burr 1981; Tang and Suen 1993; Barron et al. 1994) and thin-plate splines (Bookstein 1989; Mardia and Little 1994; Mardia et al. 1996). Such applications of warping in image processing, especially medical imaging, has allowed us to use information from multiple sources simultaneously to improve our understanding. Beside image processing and medical imaging, warping is also popular in speech processing (Sakoe and Chiba 1978), handwriting analysis (Burr 1983), determination of alignment of boundaries of ice floes (McConnell et al. 1991) where they are used to improve pattern recognition capabilities. More recently, warping allowed improvement of weather forecast analysis and verification (Hoffman et al. 1995; Alexander et al. 1999; Sampson and Guttorp 1999; Reilly et al. 2004; Gilleland et al. 2010).

Sampson and Guttorp (1992) used warping of spatial coordinates to model non-stationary and non-isotropic spatial covariance structures. Anderes and Stein (2008) and Anderes and Chatterjee (2009) developed methods for estimating deformation of isotropic Gaussian random fields. The first attempt at using warping for forecast verification in statistics, to our knowledge, was proposed by Aberg et al. (2005). Image warping in wind field modeling was proposed by Ailliot et al. (2006), and Fuentes et al. (2008) used warping to assimilate two different sources of rainfall data in a single model. Kleiber et al. (2014) used warping in the context of model emulation and calibration framework. They assume the observations lie on a grid, and the spatial features are completely observed so that standard image registration techniques such as landmark registration can be used for estimating the warping function. However, this approach does not apply to our downscaling problem because the monitoring stations are spatially sparse and the shape and direction of the fire plume are not observed. The

estimation of the warping function is challenging and further complicated by the dynamic environment, such as changes in the wind pattern.

We propose a new statistical downscaling method that optimizes the information from available forecasts and real-time monitoring data. We achieve this through (1) introducing a warping function to allow for flexible model discrepancy beyond the additive and multiplicative biases and (2) multi-resolution modeling to allow the data to determine the appropriate spatial resolution to inform prediction. We estimate the spatial misalignment between the forecast and the observed data using a penalized B-spline approach. We also use spectral smoothing (Reich et al. 2014) to capture important patterns more vividly and reduce noise simultaneously. By coalescing these two methods in a single model, we propose a novel downscaling model that accounts for spatial misalignment as well as the usual additive and scaling biases while smoothing out the forecast to improve prediction.

The remainder of the paper proceeds as follows. Section 2 introduces the motivating dataset and Sect. 3 describes the proposed method. The performance of the model and its component models are studied extensively using a simulation study in Sect. 4. The method is applied to forecasting air pollution during a major fire in Washington State in Sect. 5, where we show that accounting for spatial misalignment provides better assessment of uncertainty. We finish with some concluding remarks in Sect. 6.

## 2. $PM_{2.5}$ DATA FOR WASHINGTON STATE

We have two sources of  $PM_{2.5}$  data: numerical model forecasts on a grid and ground monitoring station scattered around the state. Both data sets give hourly  $PM_{2.5}$  measurements for the state of Washington from August 13, 2015 to September 16, 2015, a period with severe wildland fires.

The numerical forecasts were generated by the BlueSky modeling system on a 4 km  $\times$  4 km grid resulting in a 200  $\times$  95 grid covering Washington. The model is run daily at midnight and provides an hourly forecast for the next 84 h of which we use only the forecasts for the first 24 h for our analysis. The model only forecasts  $PM_{2.5}$  levels created by wildland fires and does not contain any information about  $PM_{2.5}$  generated from other sources such as traffic or industry. The forecasts also do not assimilate observed  $PM_{2.5}$  data, instead they are mostly driven by location and intensity of the fire.

The second source of data is from the ground monitoring stations which measure the total  $PM_{2.5}$  level at the corresponding locations. We have 55 monitor stations throughout the state of Washington. These monitors include both permanent monitors that are likely to be preferentially located near high population areas and temporary monitors that are placed near the areas impacted by the fire. Approximately 7% of the observations from these monitors are missing.

Figure 1 shows the concentration of  $\log(1 + PM_{2.5})$  (in  $\log \mu\text{g}/\text{m}^3$ , called  $\log PM_{2.5}$  henceforth) on 22 August, 2015 at 04:00 GMT. The circles indicate the locations of the stations, and their colors correspond to the observed  $\log PM_{2.5}$  values; missing observations are colored in gray. The background map shows the numerical model forecast. There is obvious difference in the spatial resolution of the two sources of data as well as the type of

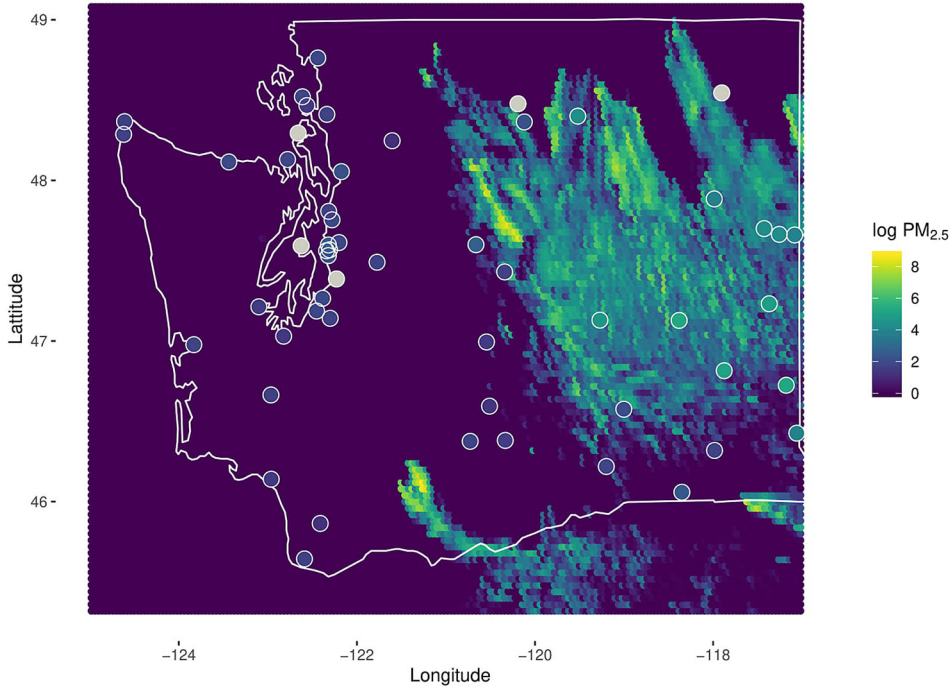


Figure 1.  $\log \text{PM}_{2.5}$  concentration in  $\log \mu\text{g}/\text{m}^3$  in Washington at 04:00 GMT on August 22, 2015. The background map shows the forecast from the numerical model, while the circle shows the location of the monitoring station. The color of the circle indicates the concentration level of the observed  $\log \text{PM}_{2.5}$  with missing value colored inlight gray.

the data. The numerical model forecast only gives information about wildland fire  $\text{PM}_{2.5}$  emissions, whereas the monitor station data includes both wildland fire  $\text{PM}_{2.5}$  emission and  $\text{PM}_{2.5}$  emission from other sources. This adds an additional level of difficulty to model and infer about the same phenomenon from the two sources of data. Naturally, excluding the other sources of  $\text{PM}_{2.5}$  from the forecast model probably results in the system ignoring some complex feedback between fire-related  $\text{PM}_{2.5}$  and  $\text{PM}_{2.5}$  from other sources. However, in a period with large fire, this effect is likely to be minimal and therefore can be overlooked.

### 3. STATISTICAL MODEL

Let  $Y_t(\mathbf{s})$  denote the measured  $\log \text{PM}_{2.5}$  from the monitor at spatial location  $\mathbf{s} = (s_1, s_2)^\top$  on day  $t$ , and  $X_t(\mathbf{s})$  be corresponding numerical forecast in log scale. Instead of directly relating these variables, we associate  $Y_t(\mathbf{s})$  to a smoothed and warped forecast to account for model discrepancies. Let  $w : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a warping function that maps  $\mathbf{s}$  to a new location  $w(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}))^\top$  to account for spatial misalignment (as discussed in Sect. 1) and  $\tilde{X}_t(\mathbf{s})$  to be smoothed forecast. The model is then

$$Y_t(\mathbf{s}) = \beta_0(\mathbf{s}) + \beta \tilde{X}_t(w(\mathbf{s})) + \epsilon_t(\mathbf{s}), \quad (3.1)$$

where  $\epsilon_t(\mathbf{s}) \stackrel{iid}{\sim} \text{Normal}(0, \sigma^2)$  is error. Since the smoothed and warped forecast is a product of an atmospheric dispersion model that already takes into account the spatiotemporal variability as well as the effects of meteorological components and other factors, we assume that the errors  $\epsilon_t(\mathbf{s})$  are independent over space and time. The slope parameter  $\beta$  is included to calibrate the difference in scale of the forecast and monitor data, perhaps due to the areal nature of the forecast and the point nature of the monitor.

### 3.1. MODEL FOR THE SPATIALLY VARYING INTERCEPT

A spatially varying intercept is employed to correct for possible additive bias. In our motivating example, additive bias in the monitor station observations come from other sources of PM<sub>2.5</sub>, such as traffic and industry that are not included in the numerical forecast. We model the spatially varying intercept using finite basis function expansion

$$\beta_0(\mathbf{s}) = b_0 + \sum_{j=1}^J \sum_{k=1}^K A_j^0(s_1) B_k^0(s_2) b_{jk}. \quad (3.2)$$

We use known basis functions for the two coordinates,  $A_j^0(s_1)$  and  $B_k^0(s_2)$ , and estimate the coefficients  $b_{jk}$  and  $b_0$ . Although other choices of basis functions are possible, we use an outer product of B-spline basis functions, that is,  $A_j^0(s_1)$  and  $B_k^0(s_2)$  are univariate B-spline basis functions with  $J$  and  $K$  knots, respectively. Cubic B-splines basis functions are a sensible choice as they can approximate any smooth function in a bounded domain.

A natural problem in finite-basis function expansion-based modeling is the choice of number of knots and their position. We select  $J$  and  $K$  to be large enough to capture the variability in the data with enough detail and use a penalized B-spline approach to prevent overfitting. Penalization is achieved by employing a Gaussian prior distribution on the coefficients  $\mathbf{b} = (b_{11}, b_{12}, \dots, b_{JK})^\top$  with mean  $\mathbf{0}$  and covariance  $\sigma_0^2 \boldsymbol{\Sigma}_0$ .  $\boldsymbol{\Sigma}_0$  has a conditional autoregressive (CAR) covariance structure, i.e.,  $\boldsymbol{\Sigma}_0 = (\mathbf{M}_0 - \rho_0 \mathbf{E}_0)^{-1}$ , where  $\mathbf{E}_0$  is the adjacency matrix for the coefficients in  $\mathbf{b}$  and  $\mathbf{M}_0$  is a diagonal matrix with the number of neighbors for each knot on the diagonal. The coefficients  $b_{jk}$  and  $b_{j'k'}$  are considered neighbors if  $|j - j'| + |k - k'| = 1$ .

### 3.2. MODEL FOR THE WARPING FUNCTION

We approximate the warping function using the finite basis function expansion

$$w_l(\mathbf{s}) = s_l + \sum_{j=1}^{J_1} \sum_{k=1}^{J_2} A_j(s_1) B_k(s_2) a_{jkl}, \quad l = 1, 2. \quad (3.3)$$

The warping function is defined by basis functions for the two coordinates  $A_j(s_1)$  and  $B_k(s_2)$  and the corresponding coefficients  $a_{jkl}$ . We use an outer product of B-spline basis functions for our model here as well, that is,  $A_j(s_1)$  and  $B_k(s_2)$  are univariate cubic B-spline basis functions with  $J_1$  and  $J_2$  knots, respectively. However, B-spline would not be a good choice

if the warped location is outside the bounded domain. This is tackled by forcing any point outside the grid to be remapped to its closest point on the grid boundary.

Other applications of warping in spatial statistics have used some restrictions on the form of the warping function. For example, [Sampson and Guttorp \(1992\)](#) restricted the class of warping functions to one-to-one functions and [Snelson et al. \(2004\)](#) restricted the warping functions to be monotone and have the entire real line as its range. Such restrictions are not necessary here since warping the space for covariates does not present problems of preserving measure-theoretic properties or positive definiteness of the covariance structure. Therefore, we can apply warping functions that map multiple locations to one point in the warped image. This may be unavoidable if the forecast is available only on a coarse spatial grid and multiple monitors reside in the same grid cell.

While insisting that the warping function is one-to-one is unnecessary and overly restrictive, we do impose a prior penalty to avoid overfitting. Our prior encourages the warping function to be smooth and centered around identity warp,  $w(\mathbf{s}) = \mathbf{s}$ . We consider identical priors for the coefficients  $\mathbf{a}_l = (a_{11l}, \dots, a_{J_1 J_2 l})$  for each  $l = 1, 2$  and that  $\mathbf{a}_1$  and  $\mathbf{a}_2$  are independent. To ensure a smooth warping function, we use a spatial prior for  $\mathbf{a}_l$  defined as a neighboring scheme based on the indices that involves the rook neighbors for each index when viewed to be placed on a two-dimensional integer grid. That is,  $a_{jkl}$  and  $a_{j'k'l}$  are neighbors if  $|j - j'| + |k - k'| = 1$ . A correlation structure for such a neighboring scheme is created by assigning a CAR covariance structure  $\Sigma_w = (\mathbf{M}_1 - \rho_w \mathbf{E}_1)^{-1}$  to the normally distributed coefficients, with  $\mathbf{E}_1$  being the adjacency matrix and  $\mathbf{M}_1$  being the diagonal matrix with  $i$ th diagonal entry equal to the number of neighbors of the  $i$ th point. This means that  $\mathbf{a}_l$  has a Gaussian distribution with mean  $\mathbf{0}$  and covariance  $\sigma_a^2 \Sigma_w$ . By setting  $E(\mathbf{a}) = \mathbf{0}$ , we shrink the warping function toward the identity function.

### 3.3. MODEL FOR THE SMOOTHING FUNCTION

Smoothing the forecast eliminates spurious small-scale variation and allows aligning large-scale features of the forecast such as smoke plumes with the monitor data. Since the forecast is on a regular grid, the smoothing can be achieved using the spectral downscalar proposed by [Reich et al. \(2014\)](#). The spectral representation of the forecast is

$$X_t(\mathbf{s}) = \int \exp(-i\boldsymbol{\omega}^\top \mathbf{s}) Z_t(\boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (3.4)$$

where  $\boldsymbol{\omega} \in \mathbb{R}^2$  is a frequency and

$$Z_t(\boldsymbol{\omega}) = \int \exp(i\boldsymbol{\omega}^\top \mathbf{s}) X_t(\mathbf{s}) d\mathbf{s} \quad (3.5)$$

is the inverse Fourier transform of the forecast. This decomposes the forecast's signals at different frequencies  $Z_t(\boldsymbol{\omega})$ . Processes that comprises of lower frequencies contain the information about the large-scale patterns, while processes corresponding to higher frequencies

holds local information. We capture the forecast features at  $L$  different resolutions using

$$\tilde{X}_{lt}(\mathbf{s}) = \int V_l(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^\top \mathbf{s}) Z_t(\boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (3.6)$$

where  $V_l(\boldsymbol{\omega})$  are known basis functions that serve as weights based on frequencies satisfying  $\int V_l(\boldsymbol{\omega}) d\boldsymbol{\omega} = 1$ ,  $\forall l = 1, \dots, L$  where  $L$  is the number of basis functions. A useful choice for the basis functions are Bernstein polynomials, as suggested by Reich et al. (2014) (see the Appendix-A for details). We then reconstruct the smoothed process by

$$\tilde{X}_t(\mathbf{s}) = \sum_{l=1}^L \alpha_l \tilde{X}_{lt}(\mathbf{s}). \quad (3.7)$$

Smoothing is achieved if  $\alpha_l \approx 0$  for terms with large  $\|\boldsymbol{\omega}\|$  as this essentially filters out high resolution features. On the other hand, if  $\alpha_l = 1$  for all  $l$ , the smoothed forecast reduces to the original forecast, i.e.,  $\tilde{X}_t(\mathbf{s}) = X_t(\mathbf{s})$ .

Constructing  $\tilde{X}_{lt}(\mathbf{s})$  requires computing the stochastic integrals in (3.5) and (3.6). For fast computing, these integrals are approximated using two dimensional discrete Fourier transform and inverse discrete Fourier transform as

$$Z_{pt} \approx \frac{1}{P} \sum_{q=1}^P \exp(i\boldsymbol{\omega}_p^\top \mathbf{s}_q) X_t(\mathbf{s}_q) \quad \text{and} \quad \tilde{X}_{lt}(\mathbf{s}) \approx \sum_{p=1}^P V_l(\boldsymbol{\omega}_p) \exp(-i\boldsymbol{\omega}_p^\top \mathbf{s}) Z_{pt}, \quad (3.8)$$

where the forecast is on a grid of  $P_1 \times P_2$  and  $P = P_1 P_2$ .

In (3.1), the scale of  $\beta$  and  $\alpha_1, \alpha_2, \dots, \alpha_L$  are not identified, so we reparametrize to  $\boldsymbol{\beta} = \beta(\alpha_1, \alpha_2, \dots, \alpha_L)^\top = (\beta_1, \beta_2, \dots, \beta_L)^\top$  and place a prior on  $\boldsymbol{\beta}$ . To prevent overfitting, we use the same penalized splines approach as before. We put another CAR covariance structure on  $\boldsymbol{\beta}$  with the neighboring scheme based on their indices, as before,

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma^2 \tau^2 \mathbf{D}_x),$$

where  $\mathbf{0}$  is the zero vector of length  $L$  and  $\mathbf{D}_x = (\mathbf{M}_2 - \rho_x \mathbf{E}_2)^{-1}$  is the corresponding CAR covariance structure,  $\mathbf{E}_2$  being the adjacency matrix with terms  $l$  and  $k$  considered neighbors if  $|l - k| = 1$  and  $\mathbf{M}_2$  being the corresponding diagonal matrix created similarly as before.

### 3.4. MODEL DETAILS

Since the forecast, and the spectral covariates, can only be computed on a grid and the monitoring sites are non-gridded points in  $\mathbb{R}^2$ , we use the nearest grid neighbor as a proxy for forecast at the station. That is, we use the model

$$Y_t(\mathbf{s}) = \beta_0(\mathbf{s}) + \sum_{l=1}^L \beta_l \tilde{X}_{lt}(\tilde{\mathbf{w}}(\mathbf{s})) + \epsilon_t(\mathbf{s}),$$



where  $\tilde{w}(\mathbf{s})$  is the location of the closest forecast grid cell to  $w(\mathbf{s})$ . Any point that goes outside the grid as a result of the warping is set at the nearest grid point, as mentioned earlier.

To complete the Bayesian model for jointly modeling the smoothing and warping, we specify the priors for the hyperparameters:  $\sigma_0^2 \sim IG(0.01, 0.01)$  and  $\sigma^2 \sim IG(0.01, 0.01)$ . We assume  $\sigma_a^2 \sim \text{Half-Normal}(0.15)$ . This sets the 99th percentile for the prior to be 1. This choice of prior strongly suggest the warping to be adequately smooth. We put a  $\text{Beta}(10,1)$  prior on the hyperparameters  $\rho_0$ ,  $\rho_a$  and  $\rho_x$ , suggesting a minimal level of spatial correlation being present. Instead of choosing a hyperprior for  $\tau^2$ , we set  $\tau^2 = 10$ . This helps avoid numerical instability in the computational process and yet provides enough prior uncertainty for the  $\beta$  parameter. We recommend choosing  $J$ ,  $K$ ,  $J_1$  and  $J_2$  to be large, e.g., so that the number of basis functions is roughly the same as the number of monitor stations, as the penalization should set the unnecessary coefficients to zero, thus reducing it to a simpler model. We use  $(J, K) = (J_1, J_2) = (6, 4), (10, 5)$  or  $(12, 8)$  for our simulation study scenarios with the corresponding number of monitor stations being 25, 50 or 100. For the data example, we use  $(J, K) = (J_1, J_2) = (11, 5)$ . We should chose  $L$  to be large as well since we added a penalization for that too. However, in this case we do not need to choose  $L$  to be as large as the number of monitor stations since we expect the smoothing operator to be a smooth function. For instance, we use  $L = 15$  throughout our studies and data example as we believe decomposing the information in  $X_t(\mathbf{s})$  into 15 partitions would allow us to filter out enough needless small-scale variations to achieve smoothing.

#### 4. SIMULATION STUDY

In this section, we conduct a simulation study to explore the performance of the proposed method in different scenarios. We consider four data generation processes and three sets of monitor station locations for each of these processes and create 30 datasets for each combination of these factors. We use the forecast from the dataset described in Sect. 2 for August 18, 2015 to August 22, 2015 as  $X_t(\mathbf{s})$ . The grid size for the simulation study was therefore the same as the forecast grid of the data,  $200 \times 95$ .

We consider five data generation processes. In the first case, data are generated by a simple linear regression (SLR) model with the forecast as the predictor, i.e.,

$$Y_t(\mathbf{s}) = \beta_0 + \beta_1 X_t(\mathbf{s}) + \epsilon_t(\mathbf{s}),$$

with  $\epsilon_t(\mathbf{s}) \stackrel{iid}{\sim} N(0, 1)$ ,  $\beta_0 = 1.5$  and  $\beta_1 = 0.25$ . Second, we use the smoothed forecast predictor

$$Y_t(\mathbf{s}) = \beta_0 + \sum_{l=1}^L \beta_l \tilde{X}_{lt}(\mathbf{s}) + \epsilon_t(\mathbf{s}),$$

where  $L = 10$ ,  $\epsilon_t(\mathbf{s}) \stackrel{iid}{\sim} N(0, 1)$ ,  $\beta_0 = 1.5$  and  $\beta_l$  were decreasingly ordered realizations of a  $N(0.25, 0.0625)$  random variate for  $l = 1, 2, \dots, 10$ . The descending order of the

coefficients ensures that the low-frequency terms have higher weights than high-frequency terms. The next two cases have a warped and smoothed forecast as predictor

$$Y_t(\mathbf{s}) = \beta_0 + \sum_{l=1}^L \beta_l \tilde{X}_{lt}(\tilde{w}(\mathbf{s})) + \epsilon_t(\mathbf{s})$$

with  $\tilde{w}(\cdot)$  being a warping function and the remaining components of the model being the same as in the previous scenario. These two cases are distinguished by their warping function. The first warping function is the translation warp

$$w(\mathbf{s}) = \mathbf{s} + \begin{pmatrix} 0.16 \\ 0.16 \end{pmatrix}.$$

The second warping function we used was diffeomorphism warp (Guan et al. 2019) that preserves the boundaries of the image. For  $0 \leq s_1, s_2 \leq 1$ ,

$$\begin{aligned} w_1(\mathbf{s}) &= s_1 - 2\theta_1 s_2 \sin s_1 \cos s_2 (\cos \pi s_1 + 1) (\cos \pi s_2 + 1) \\ w_2(\mathbf{s}) &= s_2 - 2\theta_2 s_1 \sin s_1 \sin s_2 \cos \frac{\pi s_1}{2} \cos \frac{3\pi s_2}{2}, \end{aligned}$$

where  $\theta_1$  and  $\theta_2$  are tuning parameters jointly deciding the location, direction and extent of the warp set equal to  $\theta_1 = 0.1$  and  $\theta_2 = 0.5$ . The fifth scenario we considered had a spatially varying intercept  $\beta_0(\mathbf{s}) = 0.5 + 1.25s_1 - 0.5s_2$  and we also set  $\beta = 1.2$  for this scenario (which increases the magnitude of the prediction errors for all methods). We use a spectral smoothing and translation warp as before to generate the data. To investigate the effect of number of monitor stations, we select 25, 50 or 100 monitor station observations randomly on the grid for each of the data generation processes.

A visualization of the data generation process for the fourth scenario can be seen in Fig. 2. The top left panel shows the original forecast which was smoothed using spectral smoothing. The smoothed forecast is shown in the top right panel. A diffeomorphism warp, shown in the bottom left panel, was then applied to this smoothed output. The resulting output in the bottom right panel shows a shrunken plume around the top middle part as a result.

For each of these scenarios, we fit a simple linear model to the data as well as three versions of the model proposed in Sect. 3 with or without the warping and smoothing components. For each method, a Markov Chain Monte Carlo (MCMC) chain was run for 20,000 iterations, of which the first 10,000 iterations were discarded as burn-in samples.

To compare models we use mean squared error (MSE) and mean absolute deviation (MAD) computed using the posterior mean as point forecast and pointwise coverage of 95% intervals and continuous ranked probability score (CRPS). The posterior predictive densities for the warped outputs can be skewed, heavy-tailed or even multi-modal and so metrics based on point predictions (MSE or MAD) may not capture the uncertainty properly. To evaluate the entire predictive distribution, CRPS (Gneiting and Raftery 2007) is therefore a more meaningful choice since it is a measure of integrated squared difference between the

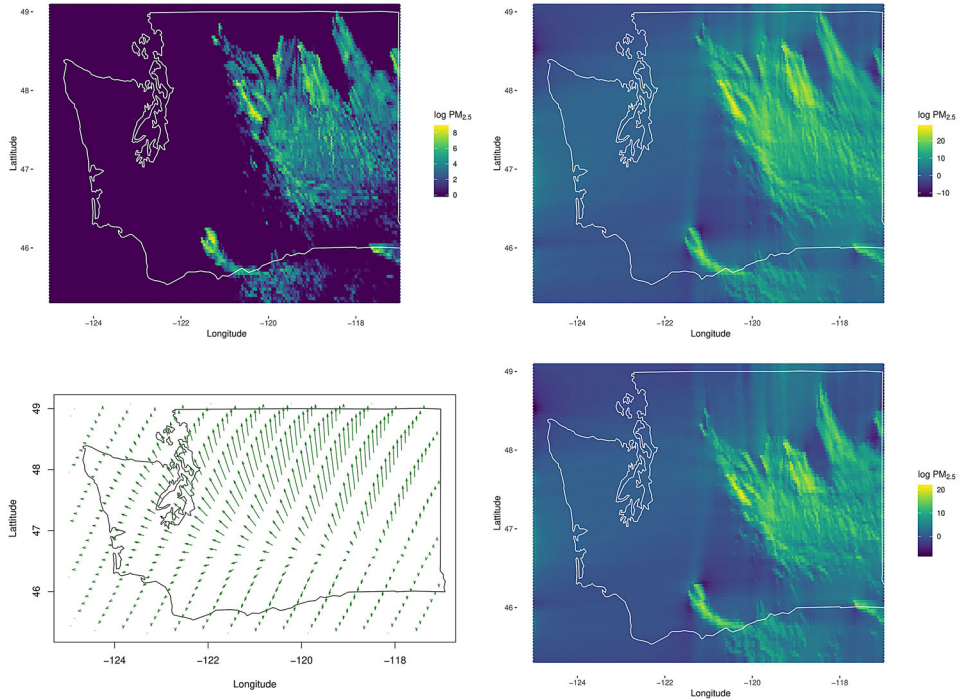


Figure 2. Original forecast  $X_T(\mathbf{s})$  in log-scale on August 22, 2015 at 4 : 00 am (top left); the smoothed forecast  $\tilde{X}_T(\mathbf{s})$  in log-scale (top right); the diffeomorphism warp  $w(\mathbf{s})$  (bottom left); the warped and smoothed forecast  $\tilde{X}_T(w(\mathbf{s}))$  in log-scale, created by using the diffeomorphism warp  $w(\mathbf{s})$  on the  $\tilde{X}_T(\mathbf{s})$ , used to generate the data. The log-concentrations are measured in  $\log \mu\text{g}/\text{m}^3$ .

cumulative distribution (CDF) function of the forecast and the corresponding CDF of the observations.

We compute the 3-day ahead forecast and compute the MSE, MAD, coverage and CRPS for the forecast of  $n$  monitor stations. For each of the 15 scenarios and for the corresponding 30 datasets in each scenario, MSE, MAD, coverage and CRPS for each of the four models are computed and averaged over space and time for all datasets. The MSE (MAD is similar) and CRPS for these cases are reported in Table 1. For all methods and cases, coverage is always between 94% and 100% and so it is not reported. Figure 3 presents a comparison of the true and estimated (posterior mean) of warping function  $w(\mathbf{s})$  for data sets with  $n = 25$  and  $n = 100$  from scenarios 3 and 4.

From Table 1, all methods perform similarly when data are generated from the SLR model. Therefore the added complexities of the full model do not result in overfitting in this case. The full model has smaller MSE and CRPS than the SLR model in the second case. However, the model with only smoothing component is somewhat better as that matches the true data generation model. In the later three cases, the full model provides the best results. The performance of all methods improves with increasing values of  $n$ . This is reflected in Fig. 3 which compares the true and estimated warp for both the warps used in this study. In both cases, the estimates are closer to the true value for  $n = 100$  than for  $n = 25$ . The

Table 1. MSE estimates (in  $\mu\text{g}^2/\text{m}^6$ ) and CRPS for the proposed model with both smoothing and warping components, only smoothing component and only warping component along with a SLR model for different scenarios

Warp	Smoothing	Spatially Varying Intercept	n	MSE			CRPS				
				SLR	Proposed model		SLR	Proposed model			
					Smooth	Warp	Both	Smooth	Warp	Both	
None	None	No	25	<b>1.01</b>	1.02	1.02	1.03	0.78	0.78	0.78	<b>0.77</b>
		No	50	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	1.01	0.78	<b>0.77</b>	<b>0.77</b>	<b>0.77</b>
		No	100	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.79	0.79	0.79	<b>0.78</b>
None	Spectral	No	25	1.27	<b>1.02</b>	1.35	1.03	0.87	<b>0.78</b>	0.85	<b>0.78</b>
		No	50	1.30	<b>1.00</b>	1.33	1.01	0.89	<b>0.77</b>	0.85	<b>0.77</b>
		No	100	1.31	<b>1.00</b>	1.35	<b>1.00</b>	0.90	<b>0.79</b>	0.86	<b>0.79</b>
Translation	Spectral	No	25	1.69	1.34	1.60	<b>1.19</b>	1.00	0.85	0.88	<b>0.78</b>
		No	50	1.62	1.25	1.45	<b>1.07</b>	0.99	0.82	0.84	<b>0.77</b>
		No	100	1.64	1.25	1.49	<b>1.09</b>	1.00	0.83	0.90	<b>0.78</b>
Diffeomorphism	Spectral	No	25	1.53	1.32	1.46	<b>1.17</b>	0.96	0.84	0.85	<b>0.76</b>
		No	50	1.50	1.23	1.39	<b>1.10</b>	0.95	0.82	0.83	<b>0.75</b>
		No	100	1.56	1.26	1.49	<b>1.12</b>	0.97	0.83	0.90	<b>0.77</b>
Translation	Spectral	Yes	25	32.56	8.76	14.50	<b>5.37</b>	4.56	1.85	2.59	<b>1.35</b>
		Yes	50	27.38	7.08	12.25	<b>2.80</b>	4.23	1.73	2.38	<b>1.00</b>
		Yes	100	28.69	7.04	25.53	<b>1.84</b>	4.37	1.74	3.76	<b>0.84</b>

Each horizontal block represents a true data generation scheme dictated by the first three columns and the last two sets of 4 columns each represent the performance of the 4 competing models in terms of MSE and CRPS. The lowest MSE and CRPS values in each case are in bold

estimation is more accurate for the translation warp, compared to the more complicated diffeomorphism warp.

## 5. APPLICATION TO PM<sub>2.5</sub> FORECASTING IN WASHINGTON STATE

In our simulation study, we fix the warping function to be constant over time. However, for the wildland fire application, the warp likely varies over time, following changes in the location of the fires and wind field. Therefore, we analyze the data separately by day with the first 18h of data as training and forecast on the next 6h for each of the 35 days. This strikes a balance between flexibility to capture dynamics of the warping function while still providing sufficient training data to estimate the warping function. The priors, models, computational details and metric of comparison are the same as the simulation study.

For each day, we compute predictive MSE, MAD, coverage and CRPS averaged over space and time. These metrics, averaged over days, are presented in Table 2. Day-by-day comparisons for MSE, MAD, coverage and CRPS are available in Appendix-A.

The models with smoothing, warping or both perform significantly better than SLR. Smoothing leads to the largest reduction in MSE and MAD while the full model leads to the largest reduction in CRPS. Therefore, smoothing appears to be sufficient if only a point estimate is required, but including the warping function provides a better fit to the full predictive distribution.

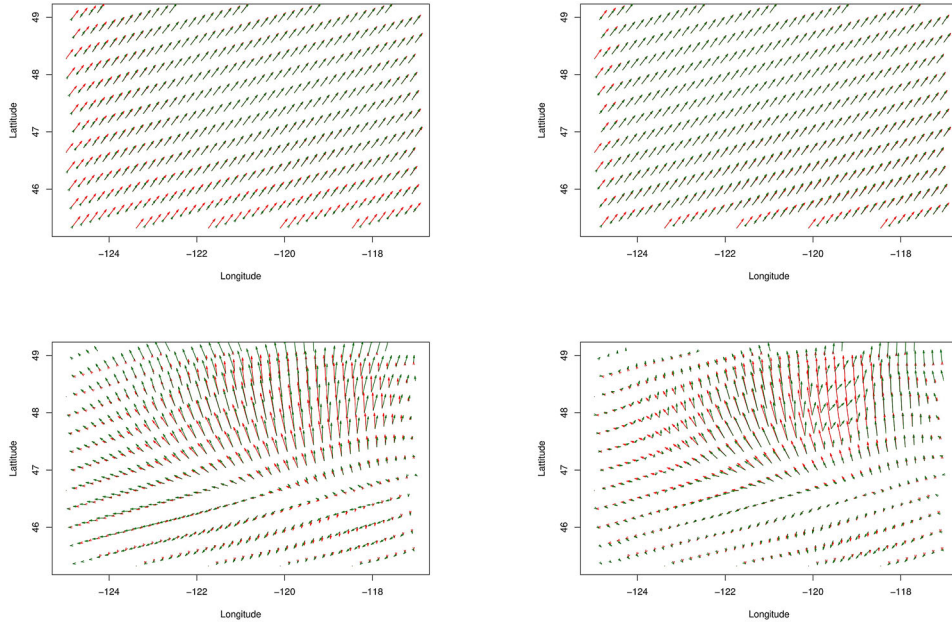


Figure 3. True (red) and estimated (green) warps for the translation warp (top row) and diffeomorphism warp (bottom row) for simulated data with  $n = 25$  (left) and  $n = 100$  (right).

Table 2. Values of MSE (in  $(\log \mu\text{g}/\text{m}^3)^2$ ), MAD (in  $\log \mu\text{g}/\text{m}^3$ ), coverage and CRPS for the data analysis for the different methods, averaged over days

	SLR	Smoothing	Warp	Full
MSE	0.4624	0.3357	0.3517	0.3675
MAD	0.4983	0.4081	0.4198	0.4219
Coverage	0.9237	0.9294	0.9126	0.9019
CRPS	0.4842	0.3688	0.3570	0.3500

To further illustrate how the warping models provide richer uncertainty quantification, we compute the posterior mean, standard deviation, skewness and kurtosis for each test set observation and present the distribution of these summary statistics as boxplots in Fig. 4. The mean values are similar for all models, but the warp-based models exhibit higher skewness and kurtosis.

Skewness and kurtosis often result from uncertainty in the warping function. For example, we compare the posterior predictive densities (PPD) of the four methods for a particular station located at the edge of the wildfire on August 22, 2015 at 7:00 pm in the left panel of Fig. 5. One would expect high uncertainty in estimation for such a location. The PPD from SLR method misses the true value (magenta) by quite some margin, while the three methods capture the true value within their respective PPDs. The PPD for the full model estimator has a heavier right tail and smaller peak, indicating high variance and kurtosis capturing the uncertainty of estimation in such a location. On the other hand, comparing the

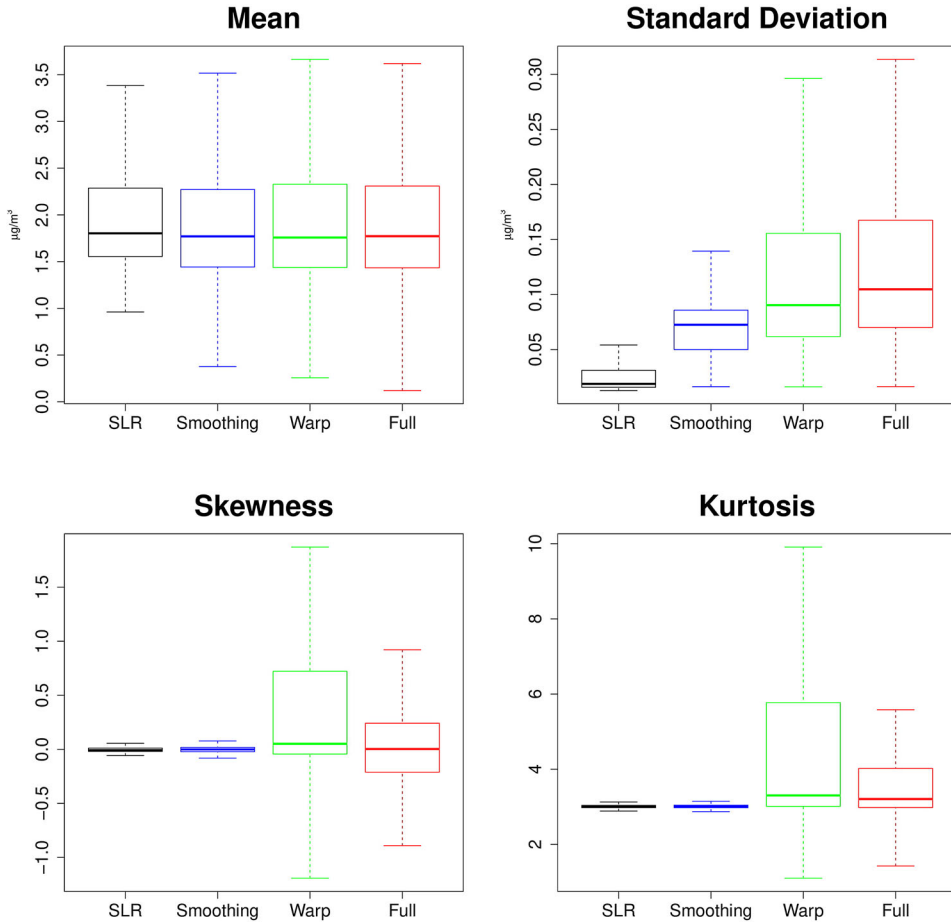


Figure 4. Boxplots of mean (in  $\log \mu\text{g}/\text{m}^3$ ), standard deviation (in  $\log \mu\text{g}/\text{m}^3$ ), skewness and kurtosis of the posterior predictive distributions for each model. The color schemes for each model is black (SLR), blue (Smoothing), green (Warp) and Red (Full) for each of the subfigures (Color figure online).

densities for a location that is in the middle of the wildfire for the same day and time shows that the PPDs behave similar to each other and have low skewness and thin tails, as can be seen in the right panel of Fig. 5. Figure 5 also shows the estimated warping function for the day. The red arrows imply a significant warp at the location at its base, while the green ones are non-significant (where warp at location  $s$  is significant if the 95% credible set of either component of  $w(s) - s$  excludes zero). The trace plots for the estimate of displacement due to the warping function ( $w(s) - s$ ) for the two locations marked in Fig. 5 are presented in Appendix-A showing adequate convergence for the MCMC procedure.



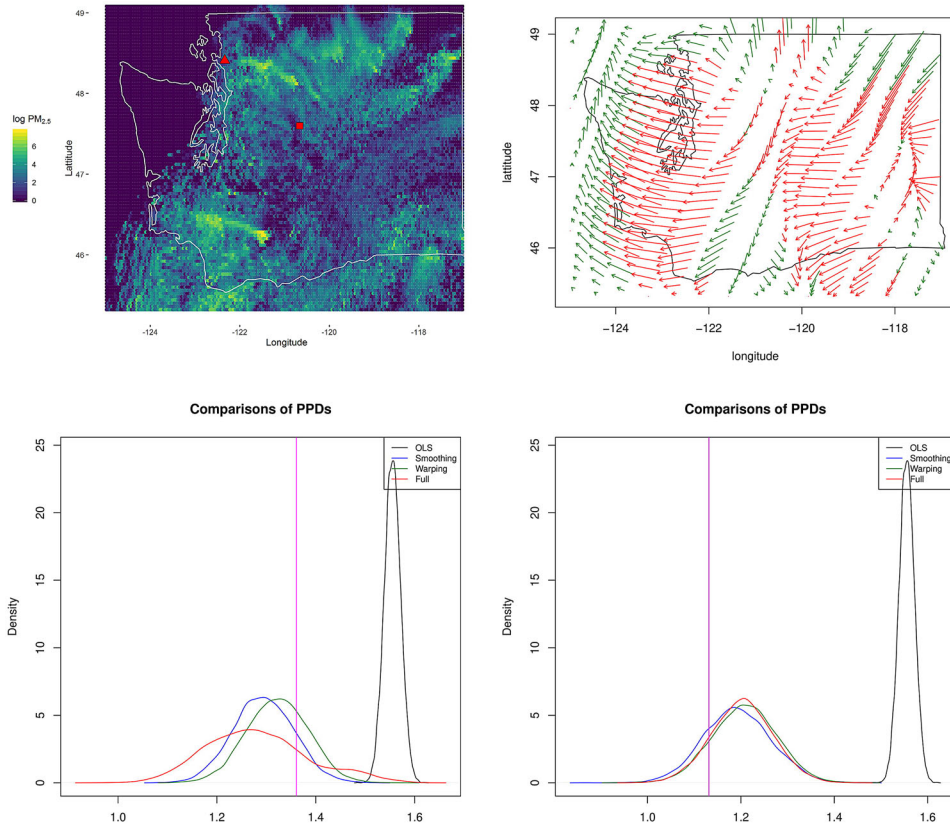


Figure 5. The numerical forecast of log-concentration of  $\text{PM}_{2.5}$  ( $\mu\text{g}/\text{m}^3$ ) on August 22, 2015 at 7:00 pm (top left) with two stations highlighted, one near the edge of fire (triangle) and one in the middle of fire (rectangle). The estimated warp for the day (top right) showing significant (red) and non-significant (green) warps. Comparisons of PPDs are made for the four competing models for the location marked as triangle (bottom left) and the location marked as rectangle (bottom right) (Color figure online).

## 6. CONCLUDING REMARKS

Motivated by a wildland fire application, we develop a new downscaling method that incorporates spectral smoothing and image warping techniques into a single downscaling method which is shown to improve forecast distributions for simulated and real data.

The proposed method could be extended in several ways. A simple extension would be to use forecasts that include background  $\text{PM}_{2.5}$  information as well. This will likely diminish the problem of having different quantities being measured by the two sources. Changing the slope in Eq. (3.1) to a spatially varying one could also be contemplated. Another extension could see the error distribution to have spatiotemporal dependence, especially in applications where the numerical model is unable to capture the important spatiotemporal trends observed in the data. While this would be straightforward to implement, it would add to the computational burden making it more challenging to apply in real time. Another possible extension is to allow for warping in both space and time. Another extension would be to use spatially varying the slope parameter  $\beta$  in Eq. 3.1. Warping the forecast in time would adjust

for timing errors, such as model misspecification of the rate of wildland fire expansion or the speed of a storm traveling through the spatial domain. We have not done this because we refit the model in fairly small spatiotemporal windows, but in other applications warping in time could be just as important as warping in space.

## ACKNOWLEDGEMENTS

We would like to thank the United States Forest Service (USFS) for providing the data. The authors were partially supported by NSF DMS-1638521, NIH ES027892, DOI 14-1-04-9 and KAUST 3800.2. We are grateful for this support.

*[Received February 2020. Accepted September 2020. Published Online October 2020.]*

## A. TECHNICAL DETAILS FOR THE MODEL

### A.1. CONSTRUCTION OF BASIS FUNCTIONS FOR SPECTRAL SMOOTHING

As mentioned in Sect. 3, we smooth our forecast using a spectral smoothing approach proposed by Reich et al. (2014). This process, using fast Fourier transform and inverse fast Fourier transform, breaks the original forecasts  $X_t(\mathbf{s})$  into several layers  $X_{lt}(\mathbf{s})$  by weighting them with basis functions  $V_l(\boldsymbol{\omega})$ . Each of the  $X_{lt}(\mathbf{s})$ s contains information about phenomena of different scales. We mentioned some restrictions on the basis functions to be used in Sect. 3.

A common choice for choosing this basis function is the Bernstein polynomial basis function, as suggested by Reich et al. (2014). This approach assumes that the dependence of frequency  $\boldsymbol{\omega}$  in constructing the basis functions is solely on the magnitude of the frequency  $\|\boldsymbol{\omega}\|$ . With this assumption, the basis functions can be written as

$$V_l(\boldsymbol{\omega}) = V_l(\|\boldsymbol{\omega}\|) = \binom{L-1}{l-1} \left( \frac{\|\boldsymbol{\omega}\|}{2\pi} \right)^{l-1} \left( 1 - \frac{\|\boldsymbol{\omega}\|}{2\pi} \right)^{L-l}, \quad (\text{A.1})$$

for  $l = 1, 2, \dots, L$ . This set up ensures that  $\int V_l(\boldsymbol{\omega}) d\boldsymbol{\omega} = 1, \forall l$ .

However, such representation of  $\tilde{X}_{lt}(\mathbf{s})$  may be subject to identifiability issues because of how the basis functions are defined in Eq. (A.1). To avoid such issues, we follow Reich et al. (2014) and define

$$\boldsymbol{\delta} = \begin{cases} \boldsymbol{\omega} & \text{if } \|\boldsymbol{\omega}\| \leq \|\bar{\boldsymbol{\omega}}\| \\ \bar{\boldsymbol{\omega}} & \text{if } \|\boldsymbol{\omega}\| > \|\bar{\boldsymbol{\omega}}\| \end{cases} \in [0, 2\pi), \quad (\text{A.2})$$

where  $\bar{\boldsymbol{\omega}} = [\mathbb{I}(\omega_1 > 0)(2\pi - \omega_1), \mathbb{I}(\omega_2 > 0)(2\pi - \omega_2)]^\top$ . After this, we define our basis functions as  $V_l(\boldsymbol{\omega}) = V_l(\|\boldsymbol{\delta}\|)$ . This ensures we avoid aliasing issues while retaining the other properties.



## A.2. COMPUTING

The warping function is not completely identifiable, that is to say that for two different warping functions  $w_1(\mathbf{s}) \neq w_2(\mathbf{s})$ , we may have the same warped output  $X_t(w_1(\mathbf{s})) = X_t(w_2(\mathbf{s}))$  for some  $\mathbf{s}$  and at some timepoint  $t$ . If the two warping functions differ only on how they map points with zero values to other points with zero values, then it is not possible to distinguish them. Assuming that the forecast would be non-constant over any region is unrealistic as it is bound to have regions with zero values, in general. This is not necessary for us to have the warping function identifiable, but it does create problems with convergence as parameters can fluctuate between two sets of values both of which give the same warped output.

Another concern for convergence is the large number of parameters in the model. The smoothing coefficients needed to be marginalized to achieve convergence in the full model. Convergence of component models (smoothing-only or warping-only) is much quickly achieved compared to the full model scenario and usually require no tricks such as marginalization, although we used marginalization for them as well. We used the simple Metropolis within Gibbs algorithm to run our MCMC chains throughout. Metropolis-adjusted Langevin algorithm (MALA) or Hamiltonian Monte Carlo (HMC) methods may provide quicker convergence but would add much complexity to each iteration.

Codes for generic-purpose use for these methods are available in the author's GitHub repository.

## B. SUPPLEMENTAL TABLES AND FIGURES

### B.1. ADDITIONAL TABLES FOR MAD AND COVERAGE ESTIMATES FROM THE SIMULATION STUDY

We present here additional tables obtained from the simulation study detailed in Sect. 4. These tables show the performance of the four models, the OLS model, the full model (see Sect. 3) and the two sub-models, smoothing only and warping only model (see Sect. 4), in four different data generation scenarios with three different values of  $n$  for each of the four cases. The results obtained here are similar to those in Sect. 4 (Tables 3, 4).

### B.2. ADDITIONAL FIGURES FROM DATA ANALYSIS

We present additional images from data analysis here. Figures 6 and 7 show the performance of the four models, as in Sect. 5, for every run (each run being based on each day) based on the metrics MAD and coverage. The inference is similar to that in Sect. 5. On most days with large fires, and thereby large plumes, the full model works better than the smoothing only model. All other models work better than the SLR model on almost any day.

Figure 8 shows the trace plots for the estimated displacement due to warp for the two locations in Fig. 5. The location in the middle of the fire (right panel) has values around zero, meaning a non-significant warp at the location. The estimate for the location at the

Table 3. MAD (standard error) estimates (in  $\mu\text{g}/\text{m}^3$ ) for the proposed model with both smoothing and warping components, only smoothing component and only warping component along with a SLR model for different scenarios

Warp	Smoothing	Spatially Varying Intercept	n	SLR	Proposed model		
					Smooth	Warp	Both
None	None	No	25	<b>0.80 (0.02)</b>	0.81 (0.02)	0.81 (0.02)	0.81 (0.02)
		No	50	<b>0.80 (0.01)</b>	<b>0.80 (0.01)</b>	<b>0.80 (0.01)</b>	<b>0.80 (0.01)</b>
		No	100	<b>0.80 (0.01)</b>	<b>0.80 (0.01)</b>	<b>0.80 (0.01)</b>	<b>0.80 (0.01)</b>
None	Spectral	No	25	0.90 (0.02)	<b>0.81 (0.02)</b>	0.93 (0.02)	<b>0.81 (0.02)</b>
		No	50	0.91 (0.01)	<b>0.80 (0.01)</b>	0.92 (0.01)	<b>0.80 (0.01)</b>
		No	100	0.91 (0.01)	<b>0.80 (0.01)</b>	0.93 (0.01)	<b>0.80 (0.01)</b>
Translation	Spectral	No	25	1.03 (0.02)	0.92 (0.02)	1.00 (0.03)	<b>0.87 (0.04)</b>
		No	50	1.01 (0.01)	0.89 (0.01)	0.95 (0.02)	<b>0.82 (0.03)</b>
		No	100	1.02 (0.01)	0.89 (0.01)	0.97 (0.01)	<b>0.83 (0.04)</b>
Diffeomorphism	Spectral	No	25	0.98 (0.03)	0.91 (0.02)	0.96 (0.03)	<b>0.86 (0.03)</b>
		No	50	0.97 (0.01)	0.88 (0.01)	0.93 (0.01)	<b>0.83 (0.02)</b>
		No	100	0.99 (0.01)	0.89 (0.01)	0.96 (0.01)	<b>0.84 (0.01)</b>
Translation	Spectral	Yes	25	4.69 (0.04)	2.11 (0.04)	2.98 (0.13)	<b>1.67 (0.04)</b>
		Yes	50	4.31 (0.03)	1.90 (0.02)	2.70 (0.07)	<b>1.17 (0.32)</b>
		Yes	100	4.43 (0.02)	1.91 (0.01)	4.06 (0.06)	<b>0.98 (0.23)</b>

The lowest MAD value in each case is in bold

Table 4. Coverage (standard error) estimates for the proposed model with only smoothing component, only warping component and the full model along with an SLR model for different scenarios

Warp	Smoothing	Spatially Varying Intercept	n	SLR	Proposed model		
					Smooth	Warp	Both
None	None	No	25	0.95 (0.01)	0.95 (0.01)	0.95 (0.01)	0.95 (0.01)
		No	50	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)
		No	100	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)
None	Spectral	No	25	0.95 (0.01)	0.95 (0.01)	0.95 (0.01)	0.95 (0.01)
		No	50	0.95 (0.01)	0.95 (0.00)	0.94 (0.01)	0.95 (0.00)
		No	100	0.95 (0.00)	0.95 (0.00)	0.94 (0.00)	0.95 (0.00)
Translation	Spectral	No	25	0.95 (0.01)	0.95 (0.01)	0.94 (0.01)	0.95 (0.01)
		No	50	0.95 (0.00)	0.95 (0.00)	0.95 (0.01)	0.95 (0.01)
		No	100	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)
Diffeomorphism	Spectral	No	25	0.95 (0.01)	0.94 (0.01)	0.94 (0.01)	0.95 (0.01)
		No	50	0.95 (0.00)	0.94 (0.00)	0.95 (0.00)	0.95 (0.00)
		No	100	0.95 (0.00)	0.94 (0.00)	0.95 (0.01)	0.95 (0.00)
Translation	Spectral	Yes	25	0.96 (0.00)	0.97 (0.00)	0.96 (0.01)	0.97 (0.00)
		Yes	50	0.95 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
		Yes	100	0.96 (0.00)	0.93 (0.00)	0.96 (0.00)	0.98 (0.01)

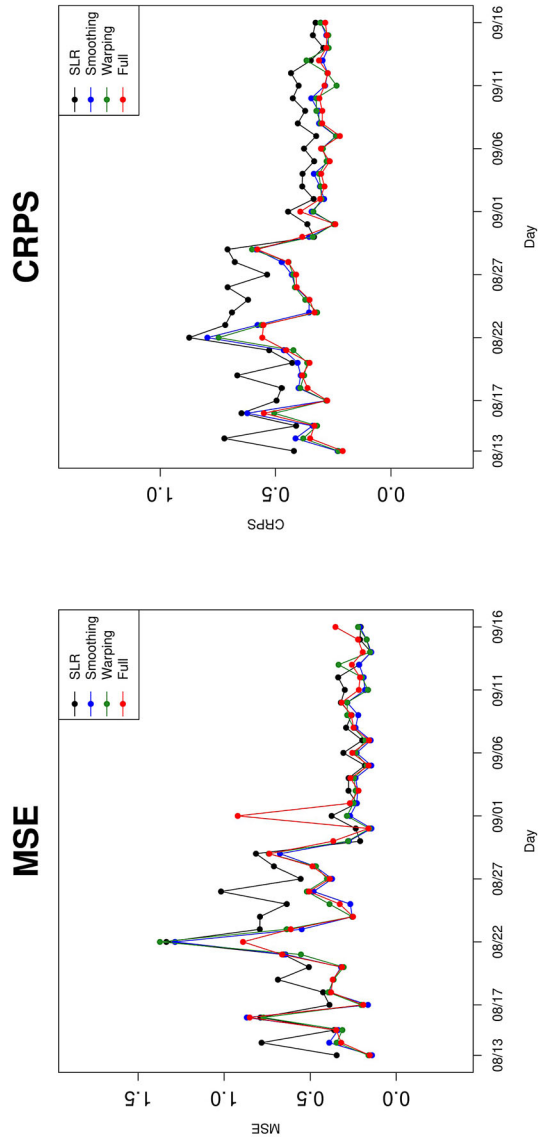


Figure 6. Daily prediction MSE (left panel) in  $\mu g^2/m^6$  and CRPS (right panel) for the four models.

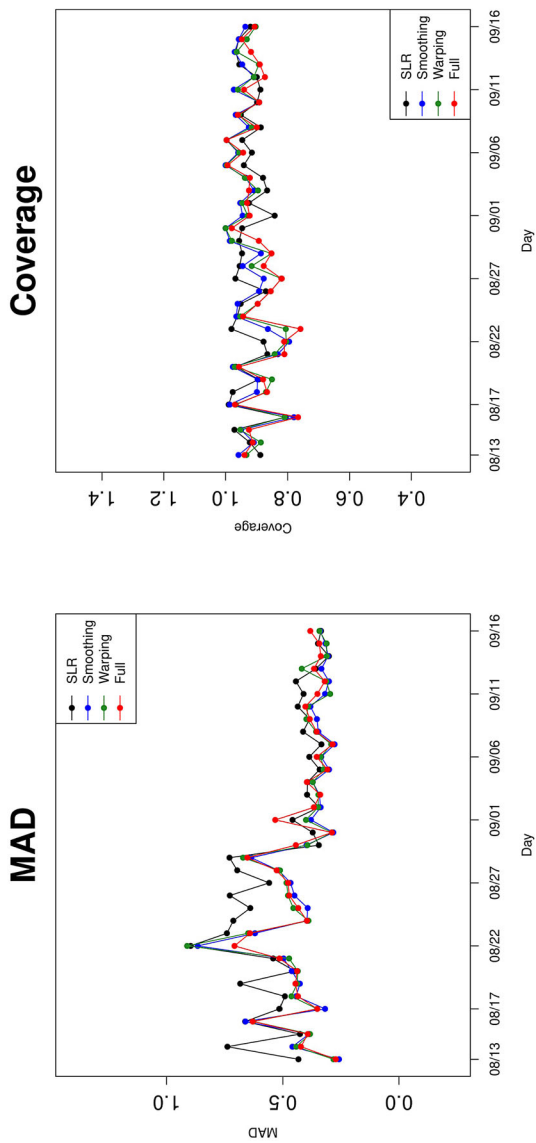


Figure 7. Daily prediction MAD (left panel) in  $\mu g / m^3$  and coverage (right panel) for the four models.

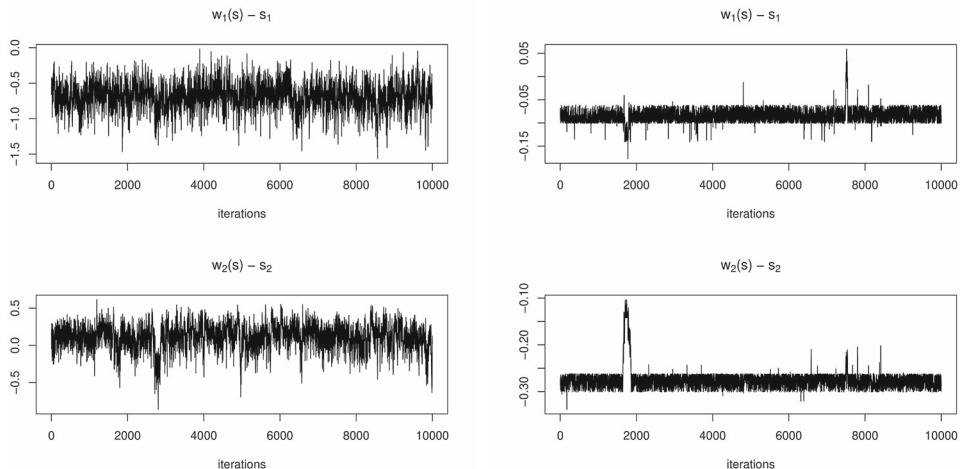


Figure 8. Trace plots for  $x$  (top) and  $y$  (bottom) coordinates of the estimated displacement due to warp ( $w(\mathbf{s}) - \mathbf{s}$ ) for the two locations flagging in Fig. 5. Left panel is for the location at the edge of fire and right panel for the location in the middle of it.

edge the fire is has a jagged trace with values away from zero, indicating a significant warp. In both cases, the MCMC algorithm mixes well.

## REFERENCES

- Aberg, S., Lindgren, F., Malmberg, A., Holst, J. and Holst, U. (2005) An image warping approach to spatio-temporal modelling. *Environmetrics*, **16**, 833–848.
- Ailliot, P., Monbet, V. and Prevosto, M. (2006) An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, **17**, 107–117.
- Alexander, G. D., Weinman, J. A., Karyampudi, V. M., Olson, W. S. and Lee, A. (1999) The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 superstorm. *Monthly Weather Review*, **127**, 1433–1457.
- Anderes, E. and Chatterjee, S. (2009) Consistent estimates of deformed isotropic gaussian random fields on the plane. *The Annals of Statistics*, **37**, 2324–2350.
- Anderes, E. B. and Stein, M. L. (2008) Estimating deformations of isotropic gaussian random fields on the plane. *The Annals of Statistics*, **36**, 719–741.
- Barron, J. L., Fleet, D. J. and Beauchemin, S. S. (1994) Performance of optical flow techniques. *International Journal of Computer Vision*, **12**, 43–77.
- Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2010a) A bivariate space-time downscaler under space and time misalignment. *The Annals of Applied Statistics*, **4**, 1942.
- Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2010b) A spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological, and Environmental Statistics*, **15**, 176–197.
- Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2012) Space-time data fusion under error in computer model output: an application to modeling air quality. *Biometrics*, **68**, 837–848.
- Bookstein, F. L. (1989) Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 567–585.
- Burr, D. J. (1981) A dynamic model for image registration. *Computer Graphics and Image Processing*, **15**, 102–112.

- Burr, D. J. (1983) Designing a handwriting reader. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 554–559.
- Chang, H. H., Hu, X. and Liu, Y. (2014) Calibrating modis aerosol optical depth for predicting daily pm 2.5 concentrations via statistical downscaling. *Journal of Exposure Science and Environmental Epidemiology*, **24**, 398.
- De Castro, E. and Morandi, C. (1987) Registration of translated and rotated images using finite fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 700–703.
- Dennekamp, M. and Abramson, M. J. (2011) The effects of bushfire smoke on respiratory health. *Respirology*, **16**, 198–209.
- Dennekamp, M., Straney, L. D., Erbas, B., Abramson, M. J., Keywood, M., Smith, K., Sim, M. R., Glass, D. C., Del Monaco, A., Haikerwal, A. and Tonkin, A. M. (2015) Forest fire smoke exposures and out-of-hospital cardiac arrests in Melbourne, Australia: A case-crossover study. *Environmental Health Perspectives*, **123**, 959–964.
- Fuentes, M., Reich, B., Lee, G. et al. (2008) Spatial–temporal mesoscale modeling of rainfall intensity using gage and radar data. *The Annals of Applied Statistics*, **2**, 1148–1169.
- Gilleland, E., Lindström, J. and Lindgren, F. (2010) Analyzing the image warp forecast verification method on precipitation fields from the icp. *Weather and Forecasting*, **25**, 1249–1262.
- Gneiting, T. and Raftery, A. E. (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.
- Guan, Y., Sampson, C., Tucker, J. D., Chang, W., Mondal, A., Haran, M. and Sulsky, D. (2019) Computer model calibration based on image warping metrics: an application for sea ice deformation. *Journal of Agricultural, Biological and Environmental Statistics*, **24**, 444–463.
- Haikerwal, A., Akram, M., Del Monaco, A., Smith, K., Sim, M. R., Meyer, M., Tonkin, A. M., Abramson, M. J. and Dennekamp, M. (2015) Impact of Fine Particulate Matter (PM<sub>2.5</sub>) Exposure During Wildfires on Cardiovascular Health Outcomes. *Journal of the American Heart Association*, **4**, e001653.
- Haikerwal, A., Akram, M., Sim, M. R., Meyer, M., Abramson, M. J. and Dennekamp, M. (2016) Fine particulate matter (PM<sub>2.5</sub>) exposure during a prolonged wildfire period and emergency department visits for asthma. *Respirology*, **21**, 88–94.
- Hoffman, R. N., Liu, Z., Louis, J.-F. and Grassoti, C. (1995) Distortion representation of forecast errors. *Monthly Weather Review*, **123**, 2758–2770.
- Johnston, F. H., Henderson, S. B., Chen, Y., Randerson, J. T., Marlier, M., DeFries, R. S., Kinney, P., Bowman, D. M. J. S. and Brauer, M. (2012) Estimated global mortality attributable to smoke from landscape fires. *Environmental Health Perspectives*, **120**, 695–701.
- Kleiber, W., Sain, S. R. and Wiltberger, M. J. (2014) Model calibration via deformation. *SIAM/ASA Journal on Uncertainty Quantification*, **2**, 545–563.
- Kloog, I., Koutrakis, P., Coull, B. A., Lee, H. J. and Schwartz, J. (2011) Assessing temporally and spatially resolved pm<sub>2.5</sub> exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment*, **45**, 6267–6275.
- Kuglin, C. (1975) The phase correlation image alignment method. In *Proceedings of the IEEE 1975 International Conference on Cybernetics and Society*.
- Mardia, K., Kent, J., Goodall, C. and Little, J. (1996) Kriging and splines with derivative information. *Biometrika*, **83**, 207–221.
- Mardia, K. V. and Little, J. A. (1994) Image warping using derivative information. In *Mathematical Methods in Medical Imaging III*, vol. 2299, 16–32. International Society for Optics and Photonics.
- McConnell, R., Kwok, R., Curlander, J. C., Kober, W. and Pang, S. S. (1991) psi-s correlation and dynamic time warping: two methods for tracking ice floes in sar images. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 1004–1012.
- Rappold, A. G., Stone, S. L., Cascio, W. E., Neas, L. M., Kilaru, V. J., Carraway, M. S., Szykman, J. J., Ising, A., Cleve, W. E., Meredith, J. T., Vaughan-Batten, H., Deyneka, L. and Devlin, R. B. (2011) Peat bog wildfire smoke exposure in rural North Carolina is associated with cardiopulmonary emergency department visits assessed through syndromic surveillance. *Environmental Health Perspectives*, **119**, 1415–1420.

- Reich, B. J., Chang, H. H. and Foley, K. M. (2014) A spectral method for spatial downscaling. *Biometrics*, **70**, 932–942.
- Reilly, C., Price, P., Gelman, A. and Sandgathe, S. A. (2004) Using image and curve registration for measuring the goodness of fit of spatial and temporal predictions. *Biometrics*, **60**, 954–964.
- Sakoe, H. and Chiba, S. (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **26**, 43–49.
- Sampson, P. D. and Guttorp, P. (1992) Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association*, **87**, 108–119.
- Sampson, P. D. and Guttorp, P. (1999) Operational evaluation of air quality models. *Environmental Statistics: Analysing Data for Environmental Policy*, **165**, 33–51.
- Snelson, E., Ghahramani, Z. and Rasmussen, C. E. (2004) Warped gaussian processes. In *Advances in Neural Information Processing Systems*, 337–344.
- Tang, Y. Y. and Suen, C. Y. (1993) Image transformation approach to nonlinear shape restoration. *IEEE Transactions on Systems, Man, and Cybernetics*, **23**, 155–172.
- Wettstein, Z. S., Hoshiko, S., Fahimi, J., Harrison, R. J., Cascio, W. E. and Rappold, A. G. (2018) Cardiovascular and Cerebrovascular Emergency Department Visits Associated With Wildfire Smoke Exposure in California in 2015. *Journal of the American Heart Association*, **7**, e007492.
- Zhou, J., Chang, H. H. and Fuentes, M. (2012) Estimating the health impact of climate change with calibrated climate model output. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 377–394.
- Zhou, J., Fuentes, M. and Davis, J. (2011) Calibration of numerical model output using nonparametric spatial density functions. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**, 531–553.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.