Web Appendix for Statistical downscaling with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting

1 Technical Details for the Model

1.1 Construction of Basis Functions for Spectral Smoothing

As mentioned in Section 3 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting", 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 phenomenons of different scales. We mentioned some restrictions on the basis functions to be used in Section 3 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$

concentration forecasting".

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

$$V_l(\boldsymbol{\omega}) = V_l(||\boldsymbol{\omega}||) = \binom{L-1}{l-1} \left(\frac{||\boldsymbol{\omega}||}{2\pi}\right)^{l-1} \left(1 - \frac{||\boldsymbol{\omega}||}{2\pi}\right)^{L-l}, \tag{1.1}$$

for $l = 1, 2, \ldots, L$. 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 Equation (1.1). To avoid such issues, we follow Reich et al. (2014) and define

$$\boldsymbol{\delta} = \begin{cases} \boldsymbol{\omega} \text{ if } ||\boldsymbol{\omega}|| \leq ||\bar{\boldsymbol{\omega}}|| \\ \bar{\boldsymbol{\omega}} \text{ if } ||\boldsymbol{\omega}|| > ||\bar{\boldsymbol{\omega}}|| \end{cases} \in [0, 2\pi), \tag{1.2}$$

where $\bar{\boldsymbol{\omega}} = [\mathbb{I}(\omega_1 > 0)(2\pi - \omega_1), \mathbb{I}(\omega_2 > 0)(2\pi - \omega_2)]^{\mathsf{T}}$. 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.

1.2 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 function 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.

2 Supplemental Tables and Figures

2.1 Additional Tables for MAD and Coverage Estimates from the Simulation Study

We present here additional tables obtained from the simulation study detailed in Section 4 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting". These tables show the performance of the four models, the OLS model, the full model (see Section 3 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting") and the two sub-models, smoothing only and warping only model (see Section 4 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting"), 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 Section 4 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting".

2.2 Additional Figures from Data Analysis

We present additional images from data analysis here. The Figures 1 and 2 shows the performance of the four models, as in Section 5 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting", for every run (each run being based on each day)

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

Table 1: MAD (standard error) estimates (in $\mu g/m^3$) 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. The lowest MAD value in each case is in bold.

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)

Table 2: 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.

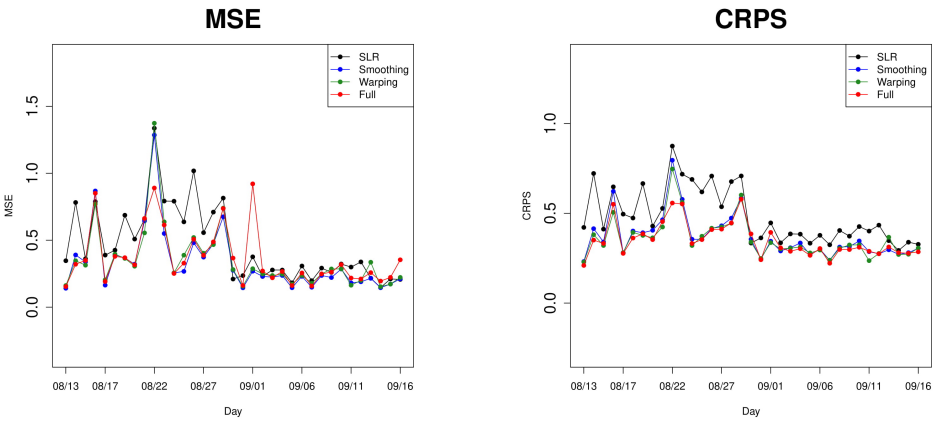

Figure 1: Daily prediction MSE (left panel) in $\mu g^2/m^6$ and CRPS (right panel) for the four models

based on the metrics MAD and coverage. The inference is similar to that in Section 5 of "Statistical downscaing with spatial misalignment: Application to wildland fire $PM_{2.5}$ concentration forecasting". 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 3 shows the trace plots for for the estimated displacement due to warp for the two locations in Figure **??**. 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 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.

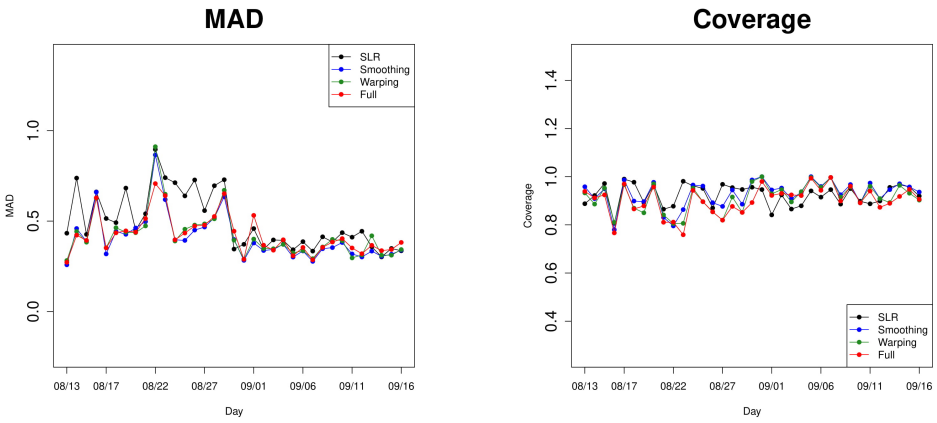

Figure 2: Daily prediction MAD (left panel) in $\mu g/m^3$ and coverage (right panel) for the four models

References

Reich, B. J., Chang, H. H., and Foley, K. M. (2014). A spectral method for spatial downscaling. *Biometrics* **70,** 932–942.

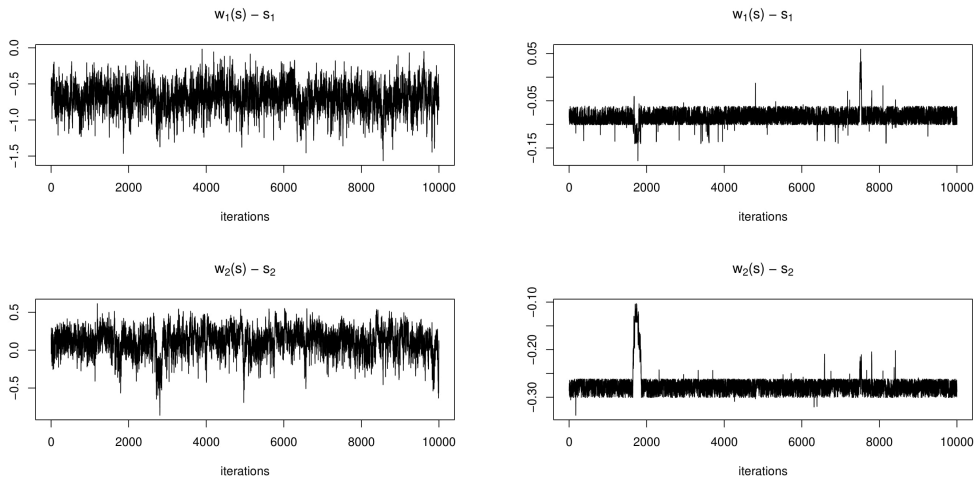

Figure 3: Trace plots for x (top) and y (bottom) coordinates of the estimated displacement due to warp ($w(\mathbf{s}) - \mathbf{s}$) for the two locations flagging in Figure **??**. Left panel is for the location at the edge of fire and right panel for the location in the middle of it.