Spatial Correlation in Weather Forecast Accuracy: A Functional Time Series Approach

Phillip A. Jang* David S. Matteson†

Abstract

A functional time series approach is proposed for investigating spatial correlation in daily maximum temperature forecast errors for 111 cities spread across the U.S. The modelling of spatial correlation is most fruitful for longer forecast horizons, and becomes less relevant as the forecast horizon shrinks towards zero. For 6-day-ahead forecasts, the functional approach uncovers interpretable regional spatial effects, and captures the higher variance observed in inland cities vs. coastal cities, as well as the higher variance observed in mountain and midwest states. The functional approach also naturally handles missing data through modelling a continuum, and can be implemented efficiently by exploiting the sparsity induced by a B-spline basis.

The temporal dependence in the data is well-characterized by AR(1)-GARCH(1,1) processes with Student-t innovations, which capture the persistence of basis coefficients over time and the seasonal heteroskedasticity reflecting higher variance in winter. Through exploiting autocorrelation in the basis coefficients, the functional time series approach also suggests a method for improving weather forecasts and uncertainty quantification.

Key Words: Functional Time Series, Spatio-temporal, Random Effects, GARCH, Weather Forecast Data, Data Expo 2018

1. Introduction

A *functional time series* is a time-indexed sequence of stochastic processes $\{f_t(\tau)\}_{t=1}^{\infty}$ where each $f_t(\cdot)$ is a random function on the domain U. By unifying functional data analysis with time series analysis, it presents an approach to modelling randomness on curves, surfaces, and other phenomena varying over a spatial continuum where these functional data are observed regularly over time and exhibit serial dependence. Ramsay & Silverman (2005) and Tsay (2010) provide background for functional data analysis and time series analysis respectively, while Hörmann & Kokoszka (2012) provides background for functional time series. Aue et al. (2017) develops the theory for functional GARCH models.

In this application, the spatial domain U is a rectangle in $\mathbb{R}^2$ containing the range of longitudes and latitudes covering the lower 48 states. The data set is taken from the 2018 American Statistical Association Data Expo[1], consisting of daily maximum temperature forecasts from the National Weather Service for 111 cities spread across the US (excluding Alaska and Hawaii) over the period from July 2014 to September 2017. The locations of the cities are illustrated in Figure 1. Forecasts range from same-day to six-days-ahead and are compared to actual temperature recorded at city airports.

*Cornell University, Center for Applied Mathematics, 657 Frank H.T. Rhodes Hall, Ithaca, NY 14850

†Cornell University, Department of Statistical Science and ILR School, 1196 Comstock Hall, Ithaca, NY 14853

[1] http://ww2.amstat.org/sections/graphics/datasets/DataExpo2018.zip

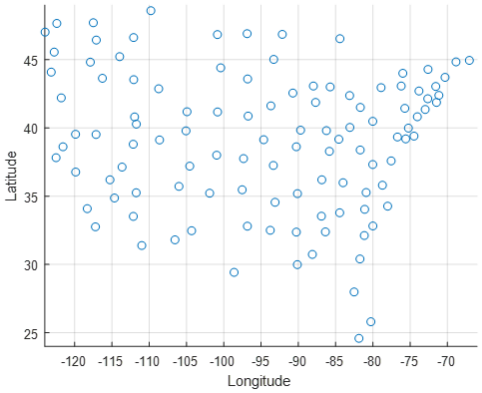

Figure 1: Locations of cities included in weather forecast data

A functional time series approach is applied to investigate and extract the structure of the spatial correlation in forecast errors. Through modelling the entire continuum instead of individual points, the proposed functional data approach also naturally handles missing data. This is a vital benefit, as data records are frequently incomplete and forecasts are not always available at every location.

2. Methodology

We first fix a forecast horizon h of interest, in this case between 0 to 6 days ahead. For the given forecast horizon, let $Y_t(\tau)$ be the forecast error of maximum temperature on day t for the city located at spatial coordinates $\tau \equiv (longitude, latitude)$. Spatial correlation is captured through the following spatio-temporal random effect model:

$$Y_t(\tau) = \mu(\tau) + \sum_{k=1}^{K} \beta_{kt}\varphi_k(\tau) + \varepsilon_t(\tau)$$

Here, $\mu(\tau)$ represents the mean forecast error and is assumed fixed over time. Estimation of the mean function is described in Section 2.1. The spatial basis functions $\varphi_1(\tau), \ldots, \varphi_K(\tau)$ describe the main modes of variation in the forecast errors, capturing spatial dependence across different regions of the US. The construction of the basis functions is described in Section 2.1. The random coefficients β_{kt} capture temporal correlation. They are modelled using independent AR(1)-GARCH(1,1) processes, as described in Section 2.2. Lastly, $\varepsilon_t(\tau)$ is a white noise process independent of the random coefficients, and is assumed i.i.d. $N(0, \sigma^2)$ for all t and τ. It is used to capture, including any measurement errors, the remaining variation not explained by the spatio-temporal random effect.

Writing $\Phi(\tau) = [\varphi_1(\tau) \cdots \varphi_K(\tau)]^T$ and $\boldsymbol{\beta}_t = [\beta_{1t} \cdots \beta_{Kt}]^T$, this model implies the following spatio-temporal covariance function:

$$Cov(Y_t(\tau), Y_{t'}(\tau')) = \Phi(\tau)^T \mathbb{E}[\boldsymbol{\beta}_t \boldsymbol{\beta}_{t'}^T] \Phi(\tau') + \sigma^2 \mathbb{1}_{\{t=t', \tau=\tau'\}}$$

2.1 Constructing Spatial Basis Functions

The following procedure is employed for constructing the fixed spatial basis functions φ_k. For simplicity, we consider the spatial domain as a subset of $\mathbb{R}^2$. Specifically, we define a 2-D cubic B-spline basis over the rectangle $[-124, -66] \times [24, 49]$, which contains the range of longitudes and latitudes covering the lower 48 states. The 2-D splines are built from the tensor product of 1-D cubic B-splines on longitude and latitude individually. Refer to Bartels et al. (1995) for further background.

For this particular dataset, knot sequences with 13 equally space interior knots captured salient features without over- or under-smoothing. This results in 17 cubic B-splines in each dimension, and 289 2-D splines in the resulting tensor product, denoted by $S_1(\tau), \ldots, S_{289}(\tau)$.

$$\mathcal{K}^{\text{Lon}} = [-124, -124, -124, -124, -119.86, -115.71, \ldots, -66, -66, -66, -66],$$
$$\mathcal{K}^{\text{Lat}} = [24, 24, 24, 24, 25.79, 27.57, \ldots, 49, 49, 49, 49], \text{ and}$$

$$\begin{bmatrix} S_1(lon, lat) \\ \vdots \\ S_{289}(lon, lat) \end{bmatrix} = \begin{bmatrix} B_{1,\mathcal{K}^{\text{Lon}}}(lon) \\ \vdots \\ B_{17,\mathcal{K}^{\text{Lon}}}(lon) \end{bmatrix} \otimes \begin{bmatrix} B_{1,\mathcal{K}^{\text{Lat}}}(lat) \\ \vdots \\ B_{17,\mathcal{K}^{\text{Lat}}}(lat) \end{bmatrix}.$$

For each day t, we fit splines such that the coefficients $\hat{c}_{t,1}, \ldots, \hat{c}_{t,289}$ solve the following optimization problem:

$$\min_{c_{t,1},\ldots,c_{t,289}} \sum_{i=1}^{n_t} \left[Y_t(\tau_{t,i}) - \sum_{j=1}^{289} c_{t,j} S_j(\tau_{t,i}) \right]^2,$$

where $\tau_{t,1}, \ldots, \tau_{t,n_t}$ are the observation locations available on day t. This handles (moderate amounts of) missing data naturally since missing observations $Y_t(\tau)$ are simply omitted from the objective function. Also, since B-splines have compact support, the resulting system is sparse and can be solved efficiently.

We denote the resulting coefficient matrix by $C = [\hat{c}_{t,j}]$. We use the column means $\bar{C} = \left[\bar{c}_1, \ldots, \bar{c}_{289}\right]$ to estimate the mean function $\mu(\tau)$ as:

$$\hat{\mu}(\tau) = \bar{\bar{Y}} + \sum_{j=1}^{289} \bar{c}_j S_j(\tau)$$

where $\bar{\bar{Y}} = \frac{1}{\sum_{t=1}^{T} n_t} \sum_{t=1}^{T} \sum_{i=1}^{n_t} Y_t(\tau_{t,i})$ is the sample mean of all observed forecast errors.

After mean-centering the columns of C, the singular value decomposition $C - \bar{C} = U\Sigma V^T$ provides principal component loadings as the columns of V, allowing for dimension reduction that is mean-square optimal in the coefficient space $\text{Col}(C - \bar{C})$. Assuming the SVD is written in descending order of singular values, the spatial basis functions $\varphi_k(\tau)$ are built using the first K columns of V, representing the K most important principal components.

$$\varphi_k(\tau) = \sum_{j=1}^{289} V_{jk} S_j(\tau) \text{ for } k = 1, \ldots, K \ll 289.$$

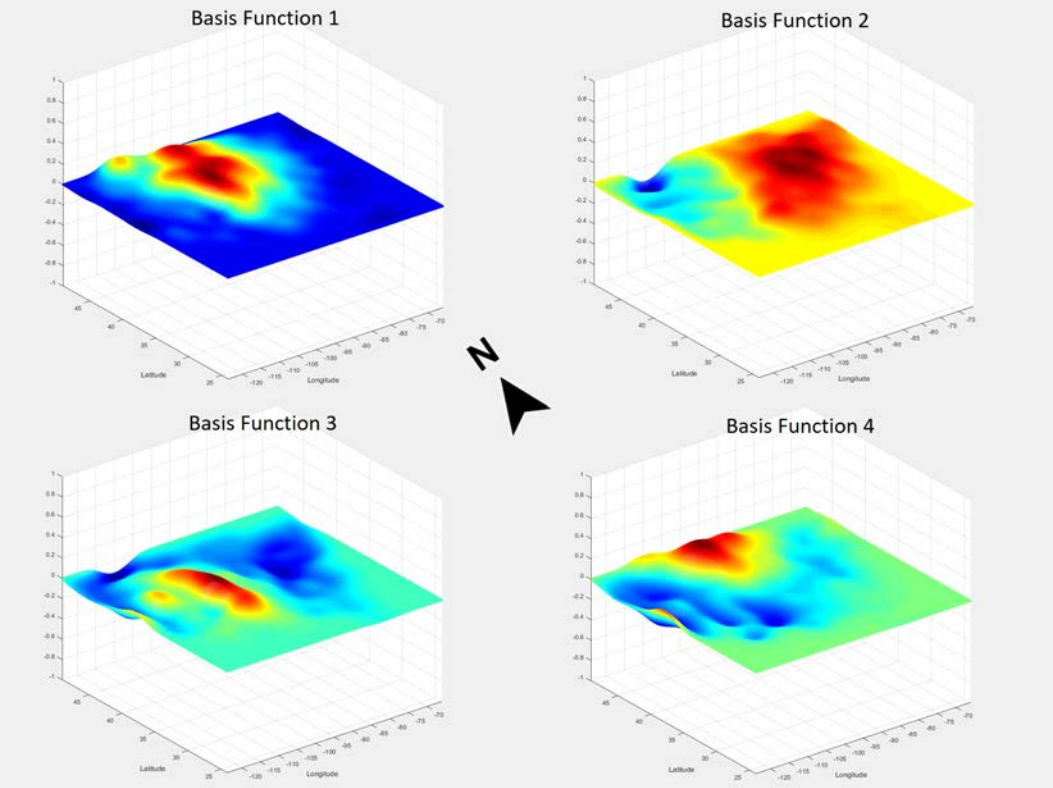

Figure 2: First 4 basis functions for 6-day-ahead maximum temperature forecasts

The first four basis functions are shown in Figure 2 for the 6-day-ahead forecast horizon. The first basis function represents an inland vs. coastal effect, as cities further inland had greater variance in their forecast errors compared to coastal cities. The second basis function represents an east vs. west effect, with the opposing signs of the regions allowing for a differentiation between the regions. The third and fourth basis functions represent mountain state and midwest state effects respectively, as these regions have the most unpredictable weather.

This structure of spatial correlation is most prevalent in 6-day-ahead forecasts, but vanishes as the forecast horizon shrinks to zero. For example, Figure 3 shows the first four basis functions for same-day forecasts, and these basis functions lack any coherent spatial structure.

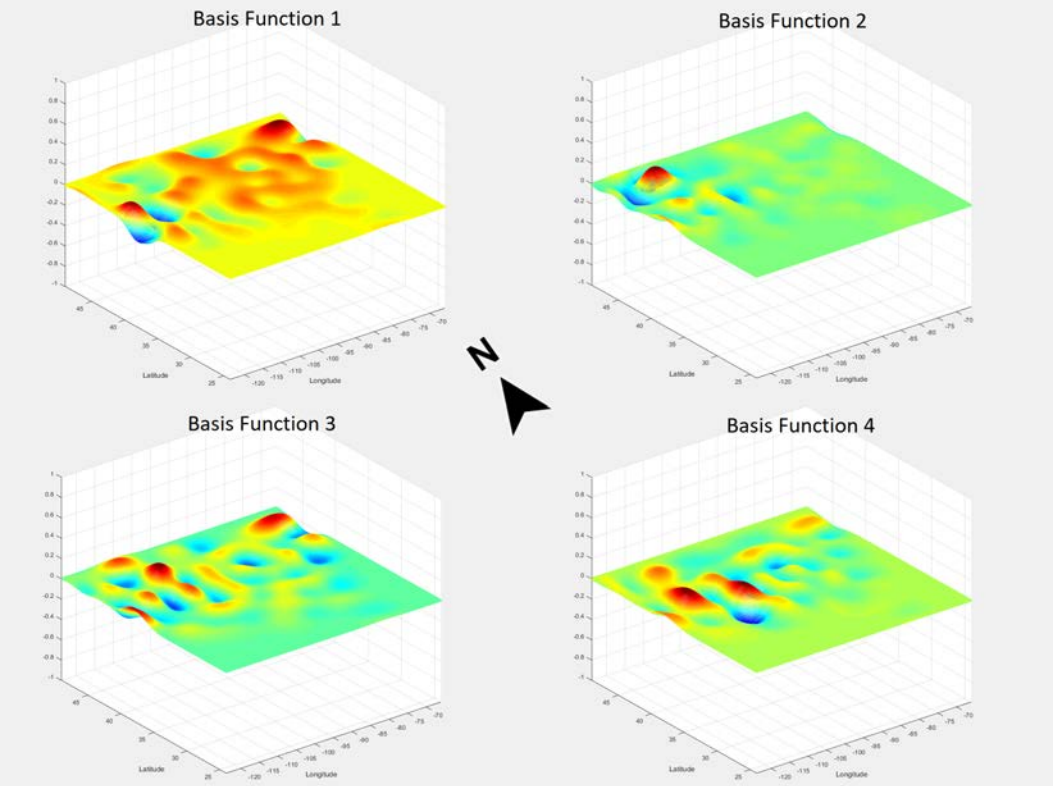

Figure 3: First 4 basis functions for same-day (0-day-ahead) maximum temperature forecasts

2.2 Modelling Random Coefficients

Once the spatial basis functions are defined as above, coefficients $\beta_{1t}, \ldots, \beta_{Kt}$ for each day t are estimated on this reduced K-dimensional basis as solutions to the following optimization problem:

$$\min_{\beta_{1t},\ldots,\beta_{Kt}} \sum_{i=1}^{n_t} \left[Y_t(\tau_{t,i}) - \hat{\mu}(\tau_{t,i}) - \sum_{k=1}^{K} \beta_{kt} \varphi_k(\tau_{t,i}) \right]^2 .$$

Temporal dependence in forecast errors is then modelled through these coefficients. Empirically, an AR(1)-GARCH(1,1) model with Student-t innovations was found to provide a good description of the observed coefficients. Specifically,

$$\begin{aligned} \beta_{kt} &= \psi_k \beta_{k,t-1} + u_{kt}, & u_{kt}|\mathcal{F}_{t-1} &\sim t_{\nu_k}(0, \eta_{kt}^2) \\ \eta_{kt}^2 &= \omega_k + \alpha_k u_{k,t-1}^2 + \gamma_k \eta_{k,t-1}^2 \end{aligned}$$

where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$ (the σ-field generated by $u_{k,t-1}, u_{k,t-2}, \ldots$ for all k). $u_{1t}, \ldots, u_{Kt}$ are assumed conditionally independent given $\mathcal{F}_{t-1}$.

The resulting innovations for the first basis function u_{1t} are shown in Figure 4 on the top left. Notably, the innovations exhibit a seasonal heteroskedasticity with winter weather being the most unpredictable. The GARCH process characterizes the heteroskedasticity

well, as the standardized innovations u_{1t}/η_{1t} exhibit approximately constant variance, and the squared standardized innovations show no significant autocorrelation.

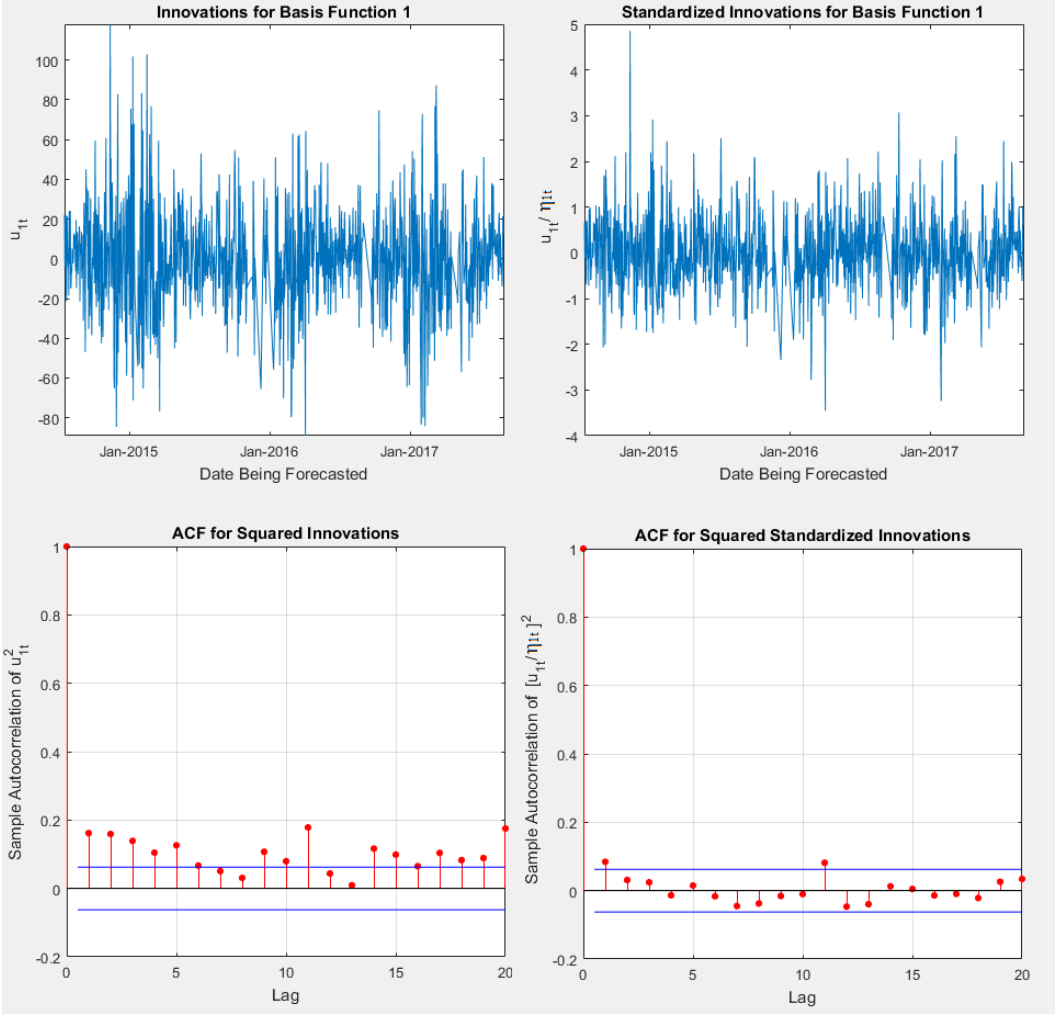

Figure 4: Estimated innovations u_{1t} (top left) and standardized innovations u_{1t}/η_{1t} (top right) for basis function 1 at forecast horizon 6-days-ahead, and their respective sample autocorrelation functions (bottom, left and right respectively). Approximate 95% pointwise confidence intervals are drawn as horizontal lines in the autocorrelation plots.

For the 6-day-ahead weather forecasts, the AR-GARCH parameter estimates for the first four basis functions are shown in Table 1 below. All fitted models were stationary with similar amounts of autocorrelation based on the similar values of ψ_k. Furthermore, all exhibit a high persistence in variance, indicated by the large values of γ_k. The earlier basis functions had conditional distributions with heavier tails, indicated by lower values of ν_k.

Table 1: AR(1)-GARCH(1,1) parameter estimates for first four basis functions for 6-day-ahead forecasts

Basis Function 1	ψ_1	ω_1	α_1	γ_1	ν_1
Estimate	0.65	13.50	0.09	0.89	8.33
Approx. Std. Error	0.02	6.91	0.02	0.02	2.05
t-ratio	26.91	1.95	4.25	37.14	4.06
p-value	<0.0001	0.0507	<0.0001	<0.0001	<0.0001

Basis Function 2	ψ_2	ω_2	α_2	γ_2	ν_2
Estimate	0.57	18.67	0.05	0.92	10.90
Approx. Std. Error	0.03	11.43	0.02	0.03	3.95
t-ratio	21.08	1.63	3.00	34.18	2.76
p-value	<0.0001	0.1024	0.0027	<0.0001	0.0057

Basis Function 3	ψ_3	ω_3	α_3	γ_3	ν_3
Estimate	0.53	7.11	0.03	0.96	13.31
Approx. Std. Error	0.03	6.46	0.01	0.02	4.87
t-ratio	19.95	1.10	2.32	47.14	2.73
p-value	<0.0001	0.2709	0.0203	<0.0001	0.0063

Basis Function 4	ψ_4	ω_4	α_4	γ_4	ν_4
Estimate	0.57	13.01	0.06	0.91	14.75
Approx. Std. Error	0.03	7.56	0.02	0.03	6.97
t-ratio	21.21	1.72	3.50	31.04	2.11
p-value	<0.0001	0.0853	0.0005	<0.0001	0.0344

3. Empirical Performance

For each pair of cities (i, j), sample correlations for the 6-day-ahead forecasts is computed before and after accounting for the spatial basis functions. More specifically, given a pair of cities located at τ_i and τ_j, the top of Figure 5 shows

$$\rho_{i,j}^{\text{before}} = \frac{\sum_t [Y_t(\tau_i) - \bar{Y}_\cdot(\tau_i)][Y_t(\tau_j) - \bar{Y}_\cdot(\tau_j)]}{\sqrt{\sum_t [Y_t(\tau_i) - \bar{Y}_\cdot(\tau_i)]^2 \sum_t [Y_t(\tau_j) - \bar{Y}_\cdot(\tau_j)]^2}},$$

and the bottom of Figure 5 shows

$$\rho_{i,j}^{\text{after}} = \frac{\sum_t [\hat{\varepsilon}_t(\tau_i) - \bar{\hat{\varepsilon}}_\cdot(\tau_i)][\hat{\varepsilon}_t(\tau_j) - \bar{\hat{\varepsilon}}_\cdot(\tau_j)]}{\sqrt{\sum_t [\hat{\varepsilon}_t(\tau_i) - \bar{\hat{\varepsilon}}_\cdot(\tau_i)]^2 \sum_t [\hat{\varepsilon}_t(\tau_j) - \bar{\hat{\varepsilon}}_\cdot(\tau_j)]^2}},$$

for the residuals $\hat{\varepsilon}_t(\tau) = Y_t(\tau) - \hat{\mu}(\tau) - \sum_{k=1}^{K} \beta_{kt} \varphi_k(\tau)$. Above, the sums are over the days t with no missing observations, and $\bar{Y}_\cdot(\tau_i)$ and $\bar{\hat{\varepsilon}}_\cdot(\tau_i)$ are sample averages of $Y_t(\tau_i)$ and $\hat{\varepsilon}_t(\tau)$ respectively over such t.

The cities are numbered from 1 to 111 (shown on the x- and y-axes), and cities are ordered from east to west, resulting in a concentration of high correlation along the main diagonal in the first figure. After accounting for $K = 20$ spatial basis functions, the second figure indicates the lack of spatial correlation in the residuals and provides evidence that the

proposed model provides an adequate approximation of the observed spatial correlation in forecast errors.

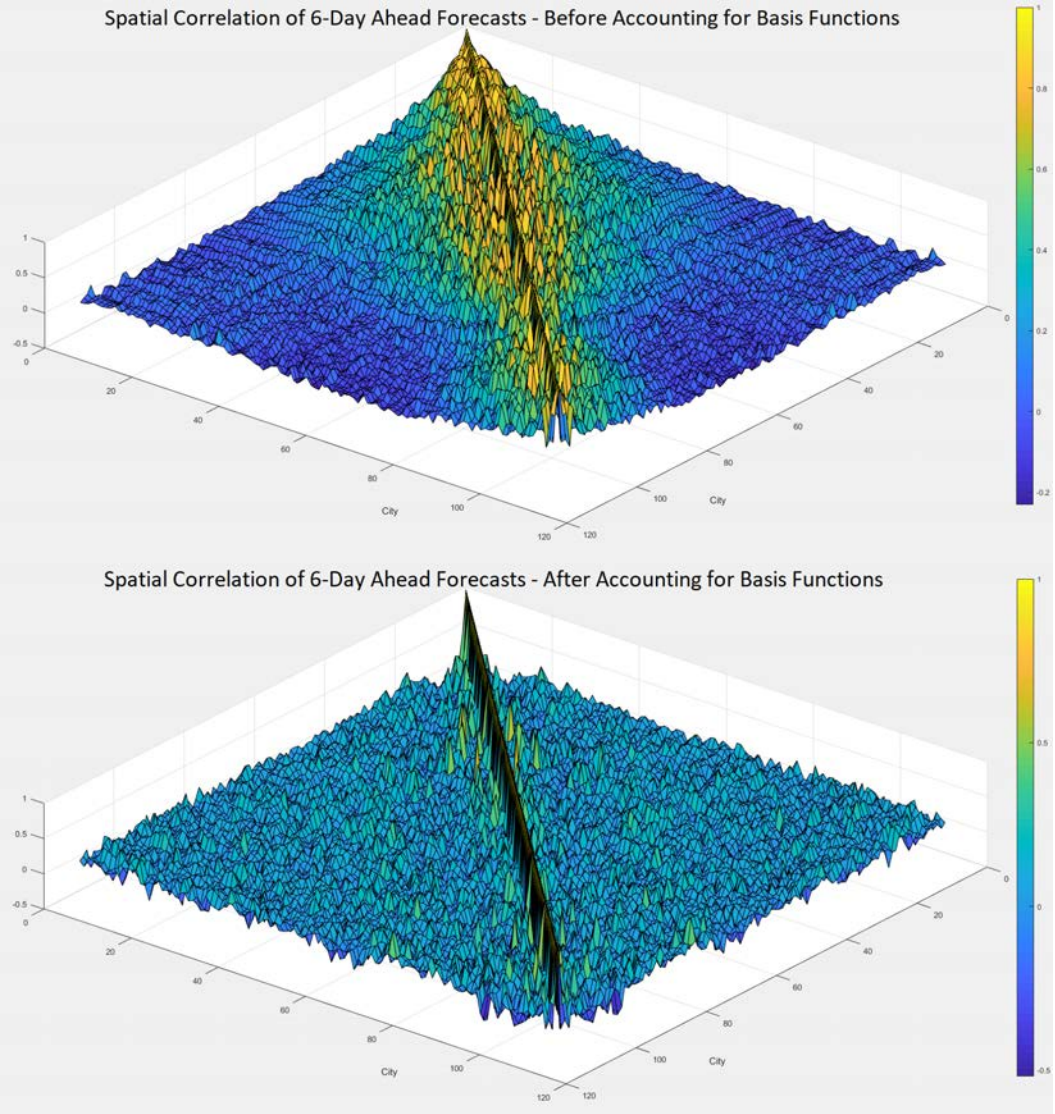

Figure 5: Spatial correlation in 6-day-ahead forecast errors for before (top) and after (bottom) accounting for basis functions

3.1 Predicting Forecast Errors to Improve Forecast Accuracy

The AR-GARCH parameter estimates can be used to predict the next day's basis coefficients, as

$$\beta_{k,T+1}|\mathcal{F}_T \sim t_{\nu_k}(\psi_k\beta_{kT}, \eta^2_{k,T+1}), \text{ and}$$
$$\eta^2_{k,T+1} = \omega_k + \alpha_k u^2_{kT} + \gamma_k \eta^2_{kT}.$$

Substituting the parameter estimates $\hat{\psi}_k$, $\hat{\omega}_k$, $\hat{\alpha}_k$, $\hat{\gamma}_k$, and $\hat{\nu}_k$ in place of the true parameters yields an approximate distribution which can be used for prediction and uncertainty quantification.

Setting the predicted coefficient to $\hat{\beta}_{k,T+1} = \hat{\psi}_k \beta_{kT}$, this can then be used to predict next day's weather forecast errors by setting

$$\hat{Y}_{T+1}(\tau) = \hat{\mu}(\tau) + \sum_{k=1}^{K} \hat{\beta}_{k,T+1} \varphi_k(\tau)$$

These predicted errors can then be used to adjust the next day's weather forecast (of the same horizon) accordingly.

4. Conclusions

We have introduced a functional time series approach to investigating spatial correlation in weather forecast accuracy. The modelling of spatial correlation is most fruitful for the longer forecast horizons, and becomes less relevant as the forecast horizon shrinks towards zero. For 6-day-ahead weather forecasts, the functional approach uncovers interpretable regional spatial effects, and captures the higher variance observed in inland cities vs. coastal cities, as well as the higher variance observed in mountain and midwest states. The functional approach also naturally handles missing data and can be implemented efficiently by exploiting the sparsity induced by using a B-spline basis.

The observed temporal dependence in the data is well-characterized by an AR(1)-GARCH(1,1) process with Student-t innovations, as the model captures the persistence of coefficients over time and the seasonal heteroskedasticity reflecting higher variance in winter. Autocorrelation in the basis coefficients can further be exploited to improve weather forecasts, especially at longer horizons.

5. Acknowledgements

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (PGS-D 502888).

REFERENCES

Aue, A., Horváth, L. & Pellatt, D. F. (2017). "Functional generalized autoregressive conditional heteroskedasticity", *Journal of Time Series Analysis* **38**(1), 3-21.

Bartels, R. H., Beatty, J. C., & Barsky, B. A. (1995). "An Introduction to Splines for Use in Computer Graphics and Geometric Modeling", *The Morgan Kaufmann Series in Computer Graphics*, Elsevier.

Hörmann, S. & Kokoszka, P. P. (2012). "Functional Time Series." in T. S. Rao (Ed.), *Time Series Analysis: Methods and Applications* (Vol. 30, pp. 157-186). (Handbook of Statistics; Vol. 30). Amsterdam: Elsevier B.V.. DOI: 10.1016/B978-0-444-53858-1.00007-7

Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd Edition. Springer, New York.

Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3rd Edition, John Wiley & Sons, Hoboken. DOI: 10.1002/9780470644560