Spatial Prediction of Sea Level Trends

Candace Berrett*[1], William F. Christensen†[2], Stephen R. Sain‡[3], Nathan Sandholtz§[4], David W. Coats¶[5], Claudia Tebaldi‖[6], and Hedibert F. Lopes**[7]

[1]*Department of Statistics, Brigham Young University, Provo, UT, USA*

[2]*Department of Statistics, Brigham Young University, Provo, UT, USA*

[3]*Jupiter Intelligence, Denver, CO, USA*

[4]*Department of Statistics, Simon Fraser University, Vancouver, BC, Canada*

[5]*Northrup Grumman Corporation, Ogden, UT, USA*

[6]*Climate Central, Boulder, CO, USA*

[7]*Department of Statistics and Econometrics, INSPER, Sao Paulo, Brazil*

September 25, 2018

Abstract

One of the major concerns engendered by a warming climate are changing sea levels and their lasting effects on coastal populations. Sea levels are now monitored by satellites, but long term records are only available at discrete locations along the coasts. Sea levels and long-term sea level trends must be better understood at the local level for local populations to adapt to sea level changes. We develop a model to facilitate spatial-prediction of known sea level trends, such as sea level rise and seasonal trends, and of remaining unknown trends in sea levels along the coast. By combining a spatially-varying coefficient modeling approach with spatio-temporal factor analysis methods in a Bayesian framework, we present a model that captures and predicts

*cberrett@stat.byu.edu
†william@stat.byu.edu
‡sainsr2@gmail.com
§nathan.sandholtz@gmail.com
¶davidwilliamcoats@gmail.com
‖ctebaldi@climatecentral.org
**hedibertfl@insper.edu.br

sea level trends and values at unmonitored locations along the coast. We show the value of the proposed model using thirty years of sea level data from thirty-eight stations along the Atlantic (east) coast of the United States. Specifically, one trend our approach provides a method for estimating is the rate of sea level rise, which ranges from roughly 1 mm/year in the northern and southern regions of the Atlantic coast to 5.4 mm/year in the middle region.

Keywords: spatio-temporal modeling, multivariate autoregressive model, spatial basis functions, confirmatory factor analysis, MCMC

1 Introduction

Changing sea levels can have significant global scale impacts on physical, social, and economic systems, but these impacts are largely felt at the local level. For example, Prime et al. (2015) describes the devastating impacts of rising sea levels combined with storm surges on catastrophic flooding events such as Hurricane Katrina and Hurricane Sandy in the United States which resulted in thousands of fatalities, billions of dollars in losses, and years of rebuilding important infrastructures. While it is estimated that *global* sea levels have increased by an average of between 1.5 to 1.9 mm/year between 1901 and 2010 and between 2.8 and 3.6 mm/year between 1993 and 2010 (IPCC, 2014), *local* sea level changes vary from one location to another. In order to mitigate impacts of such local sea level changes in the most cost-effective and locally-appropriate way (for examples of common mitigation approaches, see IPCC, 1997), sea levels must be understood at the local level. (For more examples on the impacts of changing local and global sea levels, see, for example: Mimura, 2013; Williams, 2013; Cozannet et al., 2017; and Wahl et al., 2017.)

Understanding sea levels at the local level presents many challenges. As is typical with environmental variables, there exist many sources of variability that must be accounted for and/or understood. For example, sea levels change constantly with the changing tides, sea levels exhibit seasonal trends such as natural rising in the spring and fall and dips in the winter and summer, and sea levels are strongly dependent and are impacted by regional or global trends due to weather events such as El Niño or other naturally occurring warmer and cooler years. Additionally, there are many physical

processes that contribute to local sea levels, including shifting surface winds, melting ice, tectonic movement, ocean temperature, ocean salinity, and more (see, for example, Church et al., 2013). Accounting for all these sources of variation becomes crucial when looking forward to projected changes in local sea level, since alternative futures will have dramatically different impact on the relative importance of these various sources (see for example Kopp et al., 2014). Here, as we focus on modelling past and current sea levels trends at the local level, we adopt a modeling approach that explains the ultimate outcome, without decomposing the sources explicitly.

Sea levels are not monitored at all locations along the coast. However, each location's response to changing sea levels must be based on local sea level changes and local economic and physical needs. Thus, our goal is to understand the spatial nature of sea levels in order to predict trends – including both known and unknown trends – at unmonitored locations. For example, predicting a linear sea level change (rise or fall) at unmonitored locations while accounting for other known (e.g., seasonal) and unknown trends can help local governments better prepare for their local sea level changes. To accomplish this goal, we propose a model that combines two spatio-temporal modeling tools: spatially-varying coefficient models and spatio-temporal factor analysis.

Bayesian spatially-varying coefficient models (Gelfand et al., 2003) have been used in many environmental and public health analyses (for examples, see, Waller and Gotway, 2004; Banerjee et al., 2014). These models were developed out of recognition that regression relationships are generally not constant over a region. In our case, we know that linear sea level changes and other trends in time are not constant along the coast and for our purposes it is desirable to model and learn about such changes in relationships in space. Thus, we use a spatially-varying coefficient model to directly model known sea level trends. This portion of the model is described in detail in Section 3.1.

To capture and model unknown sea level trends, we make use of Bayesian spatio-temporal factor analyses (Lopes et al., 2008, 2011). Bayesian spatio-temporal factor analysis methods were built on confirmatory factor analysis methods to capture the flexibility of factor analyses but to better identify spatial and temporal trends among the latent factors. We make use of this type of analysis

to capture remaining spatio-temporal trends after accounting for known trends. This portion of the model is described in detail in Section 3.2.

Previous analyses of sea levels focus on examining global and local sea level rise, or a linear change in time. Peltier (1996) examines historical carbon data to evaluate global sea level rise, while Church et al. (2004) uses global satellite altimeter data combined with local tide gauge data to identify regional sea level changes. Ezer (2013) examines global sea level estimates from different data sources. Hong et al. (1999), Müller et al. (2008), Ezer et al. (2013), and Yin and Goddard (2013) examine changing sea levels due to different ocean dynamics. In contrast, our analysis seeks to not only identify local sea level rise (or fall), but also to identify other local trends and to predict all trends at unmonitored locations.

Because of the impact that changing sea levels can have on storm surges and other extreme weather events, many statistical analyses of sea levels focus on learning about such storm surges. Zhang et al. (1999) uses hourly data from tide gauges to examine the frequency and intensity of storm surges while Sweet and Zervas (2011) identify a relationship between cool season storm surges and El Niño. Tebaldi et al. (2012) directly models the relationship between sea level rise and frequency of storm surges, predicting the frequency of similar storm surges in the future. In our analysis, we use monthly-averaged data and therefore do not examine short-term trends such as storm surges. However, the proposed model could be adapted to spatially predict strength and intensity of storm surges at unmonitored locations and thus be used by local officials to better prepare for such extreme weather events.

These analyses use a variety of methods to analyze sea levels such as models based on physical processes (Hong et al., 1999; Müller et al., 2008; Tebaldi et al., 2012; Yin and Goddard, 2013), frequency decomposition (Zhang et al., 1999), empirical mode decomposition (EMD; Ezer, 2013; Ezer et al., 2013), and empirical orthogonal functions (EOF's; Church et al., 2004). Methods used in our analysis are similar (for example, instead of EOF's to model spatial dependence, we use other basis functions), but are designed specifically to capture and predict sea level trends.

In the following sections, we first describe the data in Section 2. In Section 3 we motivate and develop the model. Section 4.1 examines the ability of our model to interpolate long-term trends along the coast and we also compare the model to modified versions to examine how well they capture trends that are not directly accounted for. We also examine the ability of the model to predict the actual sea level values at unmonitored locations in Section 4.2. Finally, we conclude with a discussion of the implications of our analysis in Section 5.

2 Data

The data were obtained from NOAA's tides and currents website https://tidesandcurrents.noaa.gov/ and contain hourly data from January 1, 1979 to December 31, 2008 of 93 tide and flood gauges along the United States coast. This work focuses on the Atlantic coast (east coast) of the United States, which includes 38 gauges. We are interested in longer-range temporal trends rather than very short-term temporal trends (i.e., hourly), therefore we averaged the data across lunar months (new moon to new moon), resulting in 371 lunar months during the 30-year time period. We used lunar months because of the relationship between the moon and sea levels/tides. The data were reported in meters relative to each location's mean high water (MHW); rather than working with the MHW data, we zero-centered the data from each location so that the mean at each location is 0.

Most locations have some amount of missing observations. After averaging, the amount of missing data at each location ranges from 0% to 73% missing, with most of the locations (approximately 3/4) having less than 10% missing. In the analysis, we do not eliminate sites with missing data and estimate each missing value using the model; indeed, this is one of the features of the model.

Figure 1 shows the locations of the stations. To more easily explore and plot the spatial relationships of trends, we make use of one-dimensional space by essentially stringing out the coast into one long line. Where two-dimensional (Euclidean) spatial distance is often described as "as the crow flies," we use the analogy "as the coast-guard walks" to describe this one-dimensional ("coastal") spatial distance. The coastal locations were computed by calculating the distances between latitude-

longitude points along a fine grid of the coast and matching the tide gauge locations to that grid. Because of this, the units of the one-dimensional space are essentially meaningless, but range from 0.41 to 73.70. The connection between the two spaces is shown in Figure 1.

Figure 1 also shows five locations that we used as test data to determine how well the model can predict trends and sea level values at unobserved locations. The test locations are location numbers (going from north to south) 4, 7, 16, 28, and 31 (station names: Seavey Island, ME; Newport, RI; Philadelphia, PA; Gloucester Point, VA; and Beaufort, Duke Marine Lab, NC) and are marked in red.

3 Model

The model we propose was motivated to estimate and predict spatial and temporal trends, but is also able to predict sea level values at unobserved locations. We examine model performance for both in Section 4, but describe the model in this section using the motivating trends.

To capture known temporal trends (i.e., sea-level rise and seasonal trends), we use a spatially-varying coefficient model. After accounting for the known trends, strong but unidentifiable trends remained across the residuals of the locations. Therefore, to capture these unknown trends, we use confirmatory spatio-temporal factor analysis. The spatially-varying coefficient portion of the model is described in Section 3.1 and the factor analysis portion of the model is described in Section 3.2.

3.1 Spatially-Varying Coefficient Model

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be a $T \times 1$ vector of sea level observations for location i observed at latitude-longitude location $\mathbf{s}_i$, for $i = 1, \ldots, n$. Because we use two types of spatial locations, we also note here that each two-dimensional $\mathbf{s}_i$ corresponds to a coastal location, denoted $\tilde{\mathbf{s}}_i$. We model the sea level at station i and time t by letting

$$y_{it} = \mu_i + (t - \bar{t})\beta_i + \sum_{j=1}^{J+1} b_j(t^*)\xi_{ij} + z_{it}, \tag{1}$$

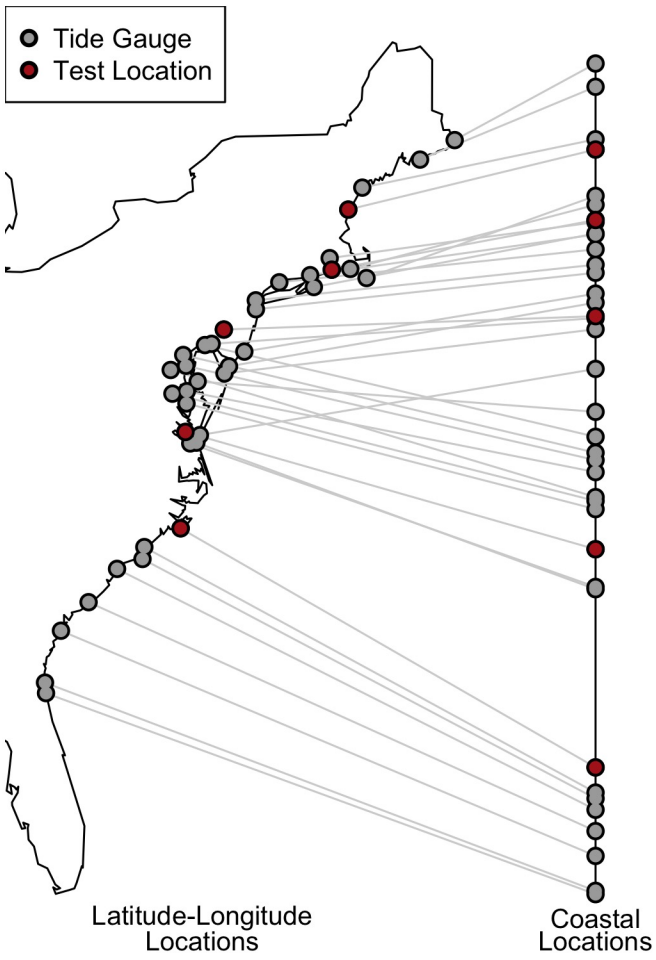

Figure 1: Map of the tide gauge locations. The left shows the latitude-longitude locations on the east coast of the United States while the right shows the coastal locations; the gray lines connect the two types of locations. Data from red locations are used as test data when applicable.

where μ_i is a location-specific mean, $(t - \bar{t})$ is the centered time index, β_i is the location-specific "sea-level change" or the coefficient of the linear change in time, $b_j(t^*)$ is the jth circular B-spline basis function, t^* is the day of the year for the time index t, and ξ_{ij} is the corresponding coefficient. z_{it} is the unexplained error. To capture additional spatio-temporal structure, we will model z_{it} using a spatio-temporal factor analysis, described in Section 3.2.

Each piece of the model in Equation (1) captures an interpretable long-range trend. μ_i captures the overall mean of location i. $(t-\bar{t})\beta_i$ captures the linearly increasing or decreasing trend in time. If the linear trend is increasing, this is what we think of as "sea-level rise." Finally, $\sum_{j=1}^{J+1} b_j(t^*)\xi_{ij}$ captures the seasonal annual trend. The splines are defined on the day of the year (we account for leap years by including 366 as the maximum day of year for each year). We used *circular* B-splines (Wood, 2017) so that the end of one year connects smoothly to the beginning of the following year. Thus, this piece captures the natural rise in the spring and fall and drops in the summer and winter exhibited by sea levels on the east coast (see Figure 4).

In exploratory analyses of the data (see, Coats, 2015), we saw strong evidence of spatial dependence in the known trends. For this reason, we use a spatially-varying coefficient model (e.g., Gelfand et al., 2003) to capture this dependence. By using a spatial model on the coefficients, we can predict the known trends at unobserved locations while accounting for the different relationships at each location. Specifically, for $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, let

$$\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{X}_\mu \boldsymbol{\alpha}_\mu, \boldsymbol{\Sigma}_\mu(\boldsymbol{\theta}_\mu)), \tag{2}$$

where $\mathbf{X}_\mu$ is an $n \times p$ matrix of covariates, $\boldsymbol{\alpha}_\mu$ is a $p \times 1$ vector of coefficients, and $\boldsymbol{\Sigma}_\mu(\boldsymbol{\theta}_\mu)$ is a spatial covariance matrix parameterized by $\boldsymbol{\theta}_\mu$. We use similar models for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$ and $\boldsymbol{\xi}_j = (\xi_{1j}, \ldots, \xi_{nj})'$:

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{X}_\beta \boldsymbol{\alpha}_\beta, \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_\beta)), \tag{3}$$

$$\boldsymbol{\xi}_j \sim \mathcal{N}(\mathbf{X}_{\xi_j} \boldsymbol{\alpha}_{\xi_j}, \boldsymbol{\Sigma}_{\xi_j}(\boldsymbol{\theta}_{\xi_j})), \tag{4}$$

where $\mathbf{X}_\beta$ and $\mathbf{X}_{\xi_j}$ are the covariate matrices, $\boldsymbol{\alpha}_\beta$ and $\boldsymbol{\alpha}_{\xi j}$ are the corresponding coefficients, and $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_\beta)$ and $\boldsymbol{\Sigma}_{\xi_j}(\boldsymbol{\theta}_{\xi_j})$ are the spatial covariance matrices parameterized by $\boldsymbol{\theta}_\beta$ and $\boldsymbol{\theta}_{\xi_j}$, respectively. Although not necessary, we let $\mathbf{X}_\mu = \mathbf{X}_\beta = \mathbf{X}_{\xi_j} \equiv \mathbf{X}$ so that all coefficients use the same covariates. Additionally, the spatial covariances do not need to be of the same form for all coefficients. For example, the spatial covariance of $\boldsymbol{\beta}$ could be from a Gaussian covariance function, while the spatial covariance of $\boldsymbol{\xi}_1$ may be from an exponential covariance function. In our analysis, we account for spatial dependence only in the $\mathbf{X}$ matrix by using bi-square spatial bases; then, for a given coefficient vector $\boldsymbol{\delta}$ (where $\boldsymbol{\delta}$ is any of $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, or $\boldsymbol{\xi}_j$), set $\boldsymbol{\Sigma}_\delta(\boldsymbol{\theta}_\delta) = \tau_\delta^2\,\mathbf{I}$, where τ_δ^2 is a covariance parameter and $\mathbf{I}$ is the identity matrix. We define the specific form of $\mathbf{X}$ used in our analysis in more detail in Section 3.3.

3.2 Spatio-Temporal Factor Analysis

The parameters $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, and $\boldsymbol{\xi}$ capture the known long-range trends, but the residuals z_{it} in Equation (1) still exhibit strong spatio-temporal dependence. An exploratory factor analysis of these residuals (Coats, 2015; Sandholtz, 2016) showed evidence of three factors and they each demonstrated spatial dependence with what we term "northern," "central," and "southern" factors, where, for example, the northern locations load most strongly on the northern factor. For this reason, we further model z_{it} with a spatio-temporal factor analysis model.

Let $\mathbf{z}_t = (z_{1t}, \ldots, z_{nt})'$ be the $n \times 1$ vector of residuals for time t, and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_T \end{bmatrix}$$

is the corresponding $n \times T$ matrix. Then, as in standard factor analysis,

$$\mathbf{Z} = \boldsymbol{\Lambda}\mathbf{F} + \mathbf{E}, \tag{5}$$

where $\boldsymbol{\Lambda}$ is an $n \times L$ matrix of factor loadings, $\mathbf{F}$ is an $L \times T$ matrix of L factors, and $\mathbf{E}$ is an $n \times T$ matrix of residuals.

When we combine the two portions of the model (Equations 1 and 5), we must make sure that the factors do not capture the trends meant to be modeled by the spatially-varying coefficient portion of the model. For example, nothing about this model prevents the factors from capturing portions of the seasonal annual trend or the linear trend, which would make the interpretation of those corresponding coefficients much less meaningful. To account for this, we apply ideas used to prevent spatial confounding as introduced by Hodges and Reich (2010) and Hughes and Haran (2013). Specifically, for all times at a given location we can write the covariate matrix from (1) as

$$\mathbf{U}_i = \begin{bmatrix} 1 & (1-(T+1)/2) & b_1(1^*) & \cdots & b_{J+1}(1^*) \\ 1 & (2-(T+1)/2) & b_1(2^*) & \cdots & b_{J+1}(2^*) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (T-(T+1)/2) & b_1(T^*) & \cdots & b_{J+1}(T^*) \end{bmatrix}. \tag{6}$$

In our analysis, we set $\mathbf{U}_i = \mathbf{U}$ for all i. Then for any location i, the orthogonal projection matrices $\mathbf{P} = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'$ and $\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P}$ can be used to divide the factors into a portion correlated with $\mathbf{U}$ and a portion orthogonal to $\mathbf{U}$. Specifically, for a given location i, we can write

$$\mathbf{Z}_i = (z_{i1}, \ldots, z_{iT})' = \mathbf{F}'\mathbf{\Lambda}_i + \mathbf{E}_i = \mathbf{P}\mathbf{F}'\mathbf{\Lambda}_i + \mathbf{P}^{\perp}\mathbf{F}'\mathbf{\Lambda}_i + \mathbf{E}_i,$$

where $\mathbf{\Lambda}_i$ is the $L \times 1$ vector of factor loadings corresponding to location i and $\mathbf{E}_i$ is the corresponding $T \times 1$ vector of residuals for the ith location. In this case, $\mathbf{P}\mathbf{F}'$ will be collinear with $\mathbf{U}$; thus, we remove this term from the model. $\mathbf{P}^{\perp}\mathbf{F}'$ will be orthogonal to $\mathbf{U}$; thus, we let $\tilde{\mathbf{F}} = \mathbf{P}^{\perp}\mathbf{F}$ be the factors of interest in our model so that

$$\mathbf{Z}_i = (z_{i1}, \ldots, z_{iT})' = \mathbf{P}^{\perp}\mathbf{F}'\mathbf{\Lambda}_i + \mathbf{E}_i = \tilde{\mathbf{F}}\mathbf{\Lambda}_i + \mathbf{E}_i.$$

Note that Hodges and Reich (2010) and Hughes and Haran (2013) use spectral decompositions of $\mathbf{P}$ and $\mathbf{P}^{\perp}$ to reduce the dimension of the partially-confounded spatial random effect. Spectral decompositions could also be used in this context.

To include spatio-temporal dependence in the factor analysis, we follow the approach of Lopes et al. (2008) by adding spatial dependence to the factor loadings, $\mathbf{\Lambda}$, and temporal dependence to the factors, $\mathbf{F}$. Consider the ℓth factor loading, $\boldsymbol{\lambda}_\ell = (\lambda_{1\ell}, \ldots, \lambda_{n\ell})'$, where $\lambda_{i\ell}$ is the loading for the ith location on the ℓth factor. *A priori*,

$$\boldsymbol{\lambda}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\lambda_\ell}(\boldsymbol{\theta}_{\lambda_\ell})), \tag{7}$$

where $\mathbf{0}$ is the $n \times 1$ vector of zeros and $\mathbf{\Sigma}_{\lambda_\ell}(\boldsymbol{\theta}_{\lambda_\ell})$ is the spatial covariance matrix parameterized by $\boldsymbol{\theta}_{\lambda_\ell}$. As with the spatially-varying coefficients, this spatial covariance can be any form and does not need to be the same for each $\boldsymbol{\lambda}_\ell$.

As in classical confirmatory factor analysis, for the model to be identifiable, we must constrain some of the parameters. This is typically done by fixing the loadings for L locations using a rank-L matrix of constants and we follow this approach. Let $\mathbf{\Lambda}_{fix}$ be the sub-matrix of $\mathbf{\Lambda}$ for the chosen fixed loadings. We set

$$\mathbf{\Lambda}_{fix} = \mathbf{I}_L, \tag{8}$$

where $\mathbf{I}_L$ is the $L \times L$ identity matrix. The distribution of the remaining $n-L$ random values for each $\boldsymbol{\lambda}_\ell$ is easily derived by combining Equations (7) and (8) and the conditional normal distribution.

We model the factors, $\mathbf{F}$, using a lag-1 multivariate autoregressive model. Note that because $\mathbf{P}^\perp$ is fixed, we model $\mathbf{F}$ rather than the factors of interest $\tilde{\mathbf{F}} = \mathbf{P}^\perp\mathbf{F}$. Let $\mathbf{F}_t = (F_{t1}, \ldots, F_{tL})'$ be the $L \times 1$ vector of factors at time t. Then, for $t = 2, \ldots, T$,

$$\mathbf{F}_t \sim N(\mathbf{\Omega}\mathbf{F}_{t-1}, \mathbf{\Sigma}_F), \tag{9}$$

where $\mathbf{\Omega}$ is an unknown autocorrelation matrix (not necessarily symmetric) and $\mathbf{\Sigma}_F$ is the covariance

matrix. For $t = 1$, we let

$$\mathbf{F}_1 \sim N(\mathbf{0}, \mathbf{\Sigma}_F),$$

where $\mathbf{0}$ is the $L \times 1$ vector of zeros and $\mathbf{\Sigma}_F$ is the same covariance matrix in (9).

3.3 Complete Model and Prior Distributions

Finally, we combine the spatially-varying coefficient model and the spatio-temporal factor model to define the full model for location i. We also define additional modeling choices we made in this analysis such as covariate matrices and prior distributions.

For location i, let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$. Then,

$$\mathbf{y}_i = \mathbf{U}\boldsymbol{\eta}_i + \tilde{\mathbf{F}}'\mathbf{\Lambda}_i + \mathbf{E}_i, \tag{10}$$

where $\mathbf{U}$ is defined in Equation (6), $\boldsymbol{\eta}_i = (\mu_i, \beta_i, \xi_{i1}, \ldots, \xi_{i(J+1)})'$, $\tilde{\mathbf{F}}$ is the $L \times T$ matrix of factors of interest, $\mathbf{\Lambda}_i$ is the $L \times 1$ factor loadings for location i, and $\mathbf{E}_i$ is a $T \times 1$ zero-mean normally-distributed error term such that $E_{it} \sim N(0, \sigma^2)$. Our data are centered at 0 for each location, therefore, we let $\mu_i = 0$ for all i so that $\boldsymbol{\eta}_i = (\beta_i, \xi_{i1}, \ldots, \xi_{i(J+1)})'$ and the column of 1's in $\mathbf{U}$ is removed. In fact, we can adjust $\mathbf{U}$ to add (or remove) other known trends of interest (or disinterest). Indeed, we consider different structures of $\mathbf{U}$ in Section 4 to better understand how the different portions of the model capture the sea-level trends across the space.

As stated in Section 3.1, we use $\mathbf{X}$ to capture the spatial dependence among the coefficients of the known trends by using spatial bases functions. While any basis function could be used (Cressie and Wikle, 2011), we use multiresolution bisquare bases (Cressie and Johannesson, 2008; Katzfuss and Cressie, 2012) using coastal distance. By using the coastal distance we can examine the spatial relationship of the coefficients across the locations along the coast with an easy-to-read plot. Multiresolution bisquare bases use R_q knots across the spatial domain at q resolutions. As defined in Cressie and Johannesson (2008) and adjusted here for the coastal locations, the rth basis

for the ith location at the qth resolution is defined to be,

$$c_{ir(q)} = \begin{cases} (1 - (|\tilde{s}_i - \tilde{v}_{r(q)}|/w_q)^2)^2, & |\tilde{s}_i - \tilde{v}_{r(q)}| \leq w_q \\ 0, & \text{otherwise} \end{cases}$$

where $\tilde{v}_{r(q)}$ is the rth knot of the qth resolution and w_q is 1.5 times the shortest distance between the knots of the qth resolution. In this analysis, we let $q = 2$, $R_1 = 3$, and $R_2 = 2$; thus,

$$\mathbf{X} = \begin{bmatrix} c_{11(1)} & c_{12(1)} & c_{13(1)} & c_{11(2)} & c_{12(2)} \\ c_{21(1)} & c_{22(1)} & c_{23(1)} & c_{21(2)} & c_{22(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n1(1)} & c_{n2(1)} & c_{n3(1)} & c_{n1(2)} & c_{n2(2)} \end{bmatrix}.$$

Because we use spatial bases for covariates, we let the covariance of the coefficients be independent with a common variance: for a given vector of coefficients $\boldsymbol{\delta}$, the covariance is defined to be $\tau_\delta^2\mathbf{I}$. We use the same prior distribution for all precision parameters, namely, $1/\tau_\delta^2 \sim \text{Gamma}(\gamma_{\tau^2}, \phi_{\tau^2})$, where γ_{τ^2} is the shape and ϕ_{τ^2} is the rate of the gamma distribution. We let $\gamma_{\tau^2} = 2$ and $\phi_{\tau^2} = 10^{-10}$. We use relatively diffuse prior distributions for the coefficients of $\mathbf{X}$ (Equations 3 and 4), namely for a given vector of spatially-varying coefficients, $\boldsymbol{\delta}$, $\boldsymbol{\alpha}_\delta \sim \mathcal{N}(\mathbf{0}, 100000\mathbf{I})$.

For the remaining parameters of the factor analysis, as proposed by Sandholtz (2016), we let $\boldsymbol{\Omega}_{ii} \sim \mathcal{N}(1, 1)$, where $\boldsymbol{\Omega}_{ii}$ is the ith diagonal of $\boldsymbol{\Omega}$, and $\boldsymbol{\Omega}_{ij} \sim \mathcal{N}(0, 2)$, where $\boldsymbol{\Omega}_{ij}$ is the (i, j) element of $\boldsymbol{\Omega}$ when $i \neq j$. This prior structure does not require $\boldsymbol{\Omega}$ to be symmetric but does appear to enforce $\boldsymbol{\Omega}$ to be positive definite. For the covariance of the factors, $\boldsymbol{\Sigma}_F \sim \text{InvWish}(S_F, \nu_F)$, where $\mathbf{S}_F$ is the scale matrix that we set equal to the identity matrix, and ν_F is the degrees of freedom that we set equal to L, the number of factors.

We fix the values of the factor loadings for locations 1, 22, and 37 (station names: Eastport, ME; Baltimore, Fort McHenry, Patapsco River, MD; and Fernandina Beach, FL) as defined in Equation (8). The prior covariance on the factor loadings, $\boldsymbol{\lambda}_\ell$, is a spatial dependence matrix. Because other

portions of the model capture the dependence across coastal distances, we use standard covariance functions on the Euclidean distances in this portion of the model. For each factor, we use an exponential covariance function to define the spatial dependence between locations. For the ℓth factor, the (i, j) element of $\mathbf{\Sigma}_{\lambda_\ell}(\boldsymbol{\theta}_{\lambda_\ell})$ in (7) is

$$\mathbf{\Sigma}_{\lambda_\ell}(\boldsymbol{\theta}_{\lambda_\ell})_{ij} = \tau^2_{\lambda_\ell} \exp(-||\mathbf{s}_i - \mathbf{s}_j||/\omega_{\lambda_\ell}),$$

where $\tau^2_{\lambda_\ell} \equiv \tau^2_\ell$ is the common variance among the ℓth factor loadings and $\omega_{\lambda_\ell} \equiv \omega_\ell$ is the spatial range parameter. We let τ^2_ℓ and ω_ℓ have the same prior distributions for all ℓ, namely, $1/\tau^2_\ell \sim$ Gamma$(\gamma_{\tau^2}, \phi_{\tau^2})$, where $\gamma_{\tau^2} = 2$ and $\phi_{\tau^2} = 10^{-10}$ for all ℓ, and $\omega_\ell \sim$ Gamma$(\gamma_\omega, \phi_\omega)$ and set $\gamma_\omega = 3$ and $\phi_\omega = 0.5$ for all ℓ.

Finally, we let the precision of the residuals, $1/\sigma^2$, follow a gamma distribution with prior parameters γ^2_σ and ϕ^2_σ which we set equal to 2 and 10^{-5}, respectively.

4 Model Comparison

We compare a selection of models to examine how different versions of the model capture trends and predict sea levels. Table 1 gives an overview of the models we compare. Specifically, we fit the complete model: a linear trend in time, a seasonal annual trend, and a spatio-temporal factor analysis with three factors. As mentioned in Section 3.3, we do not include the mean as the data have all been centered to 0 at each location. We then fit the model by leaving out various parts of the model, as defined in Table 1. Note that when the model only includes the spatio-temporal factor analysis portion (Model 4), $\mathbf{P}^\perp$ is not defined and thus is not used in the model. We fit each model twice: once using all of the observed data and once holding out five locations of observations as test data. The five locations were selected because they had varying amounts of missing observations and spanned the coast. Figure 1 shows the locations of the test data.

The goal for selecting these models is to examine the trends identified by each of the different models. For example, does the model with only the spatio-temporal factor analysis capture the

known trends in one (or more) of the factors? Additionally, we are interested in improvement (or worsening) of sea level predictions for additional model parameters. For example, does separating the known trends from the factor analysis improve model fit or is it simply useful for intuitive interpretation of these trends? There are many other models we could consider; for example, a model with a larger (or smaller) number of factors or a model with a different number of bases to capture the seasonal annual trend. However, our goal is not to delve into a comprehensive model comparison but rather to explore using a subset of models how the different portions of the model capture different aspects of the spatial and temporal trends.

We fit the models using a Metropolis-within-Gibbs MCMC sampler in `R` (R Core Team, 2018). Some full conditional distributions are known ($\boldsymbol{\beta}$, τ^2_β, $\boldsymbol{\alpha}_\beta$, $\boldsymbol{\xi}_j$, $\tau^2_{\xi_j}$, $\boldsymbol{\alpha}_{\xi_j}$, $\mathbf{F}_t$, $\boldsymbol{\Lambda}$, τ^2_ℓ, $\boldsymbol{\Sigma}_F$, σ^2, and missing values of y_{it}); for the remaining parameters with unknown full conditional distributions ($\boldsymbol{\Omega}$, ω_ℓ), we use adaptive Metropolis random walk algorithms using SCAM (Haario et al., 2005). We sampled 50000 values of each parameter and used the first 10000 as burn-in. Effective sample sizes for the remaining posterior draws for all parameters across all models range from 627 to 40000 and Monte Carlo standard errors (Jones et al., 2006; Flegal et al., 2008) range from 2×10^{-11} to 0.0085.

4.1 Spatial Prediction of Sea Level Trends

We first examine the competency of the complete model (Model 1 in Table 1) to capture the trends by examining the predicted trends fit with the test data held out compared to the predicted trends fit with the full data. We then compare these results to the remaining three models.

Model 1

First we examine the linear trend in time, or the "sea level rise." Figure 2 shows the posterior mean estimates of $\boldsymbol{\beta}$ versus the coastal locations, where the left side of the plot is the south end of the coast and the right side of the plot is the north end of the coast (Figure 1 shows how this relates to the two-dimensional coast by rotating it 90 degrees clockwise). The solid black dots are the estimates of $\boldsymbol{\beta}$ when the model is fit with the full data while the solid gray dots are the estimates

of $\boldsymbol{\beta}$ when the model is fit with the test data removed. For the most part, the posterior mean estimates are the same (the gray dots cover the black dots). As expected, however, the five test locations do not have the same posterior mean values. Additionally, as expected, the posterior 95% posterior intervals for $\boldsymbol{\beta}$ are much wider for test locations when their data was not included in the model fit. What we easily see from this plot is that there is a spatial trend in these estimates of the sea level rise at each location. Specifically, the sea level rise (for the observed time period) is larger for locations in the central part of the coast and lower for locations in the northern and southern parts of the coast. The predicted trend across the entire coastal space is shown in the thick solid black line (posterior mean of $\mathbf{X}_\beta \boldsymbol{\alpha}_\beta$ in Equation 3 estimated with the full data) and the gray dotted line (posterior mean of $\mathbf{X}_\beta \boldsymbol{\alpha}_\beta$ estimated when the test data is removed). These coefficients have a simple interpretation of the change in sea level over the thirty year period. The coefficients here have units of meters per lunar month which can be translated to mm per year and the posterior mean estimates indicate that the sea level is rising along the east coast depending on the location by between 0.99 and 5.44 mm per year.

Spatial trends for the coefficients $\boldsymbol{\xi}_j$ (Equations 1 and 4) are presented in Figure 3. As in Figure 2, the black represents the estimates and intervals from the model fit with the full data while the gray represents the estimates and intervals from the model fit with the test locations held out. As with the coefficients for the linear trend, the spatial trends of the coefficients are similar when the model is fit with all of the data and when it is fit with the test data held out. Each coefficient follows a different spatial trend. However, while some of the posterior mean estimates for the test locations of the coefficients are the same or close, some can be quite different, although the posterior intervals do overlap. For example, the posterior mean of the northernmost test location in $\boldsymbol{\xi}_5$ is higher when fit using the full data than when that location is held out, but the posterior intervals do overlap.

Recall that $\boldsymbol{\xi}_j$ are the coefficients for the seasonal trends. While these coefficients themselves do not have a simple interpretation, when combined we can examine how the differences in the coefficient estimates impact the predictions of the seasonal annual trend. Figure 4 shows the seasonal trends for the five test locations. Location numbers are ordered from north to south so that Location 4

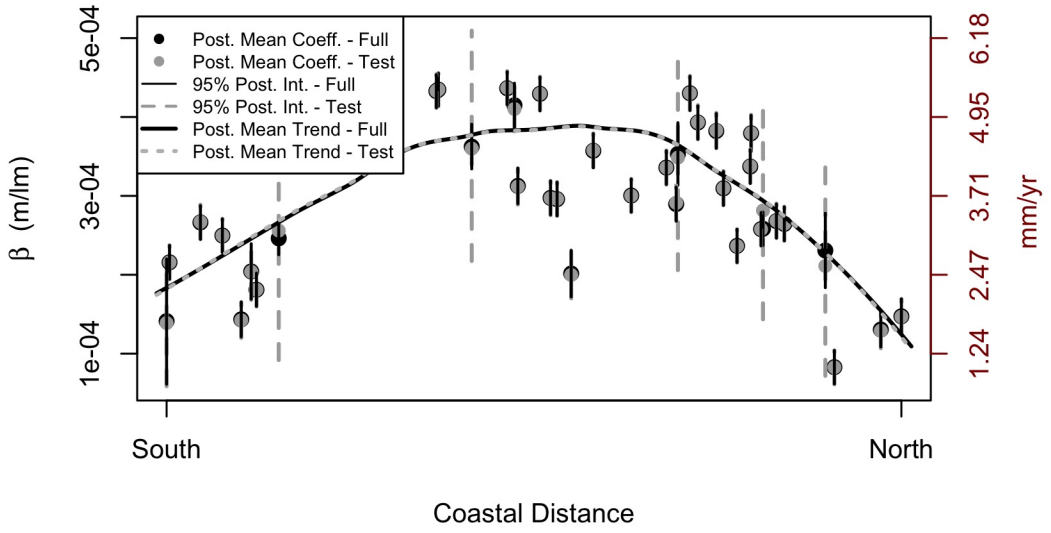

Figure 2: Posterior estimates and prediction intervals of the coefficients of the linear trend ("sea level rise") using Model 1. Black indicates that the entire/full data set was used to fit the model, while gray indicates that five test locations were held out to fit the model. The "trend line" is the posterior mean of $\mathbf{X}_\beta \boldsymbol{\alpha}_\beta$ in Equation (3). The coefficients represent the change in sea level over the thirty year period in meters per lunar month. The right (red) axis shows the estimates in mm per year.

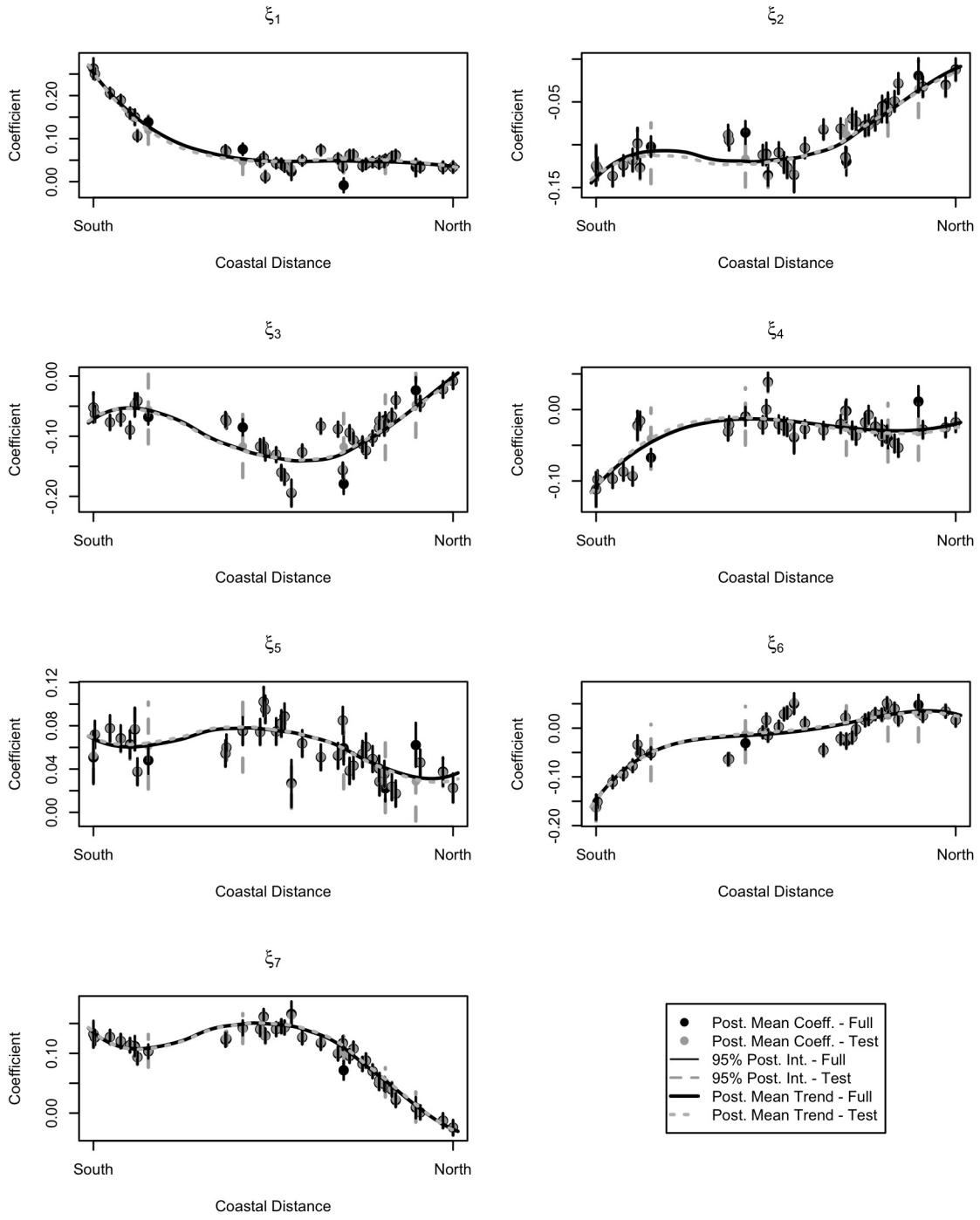

Figure 3: Posterior estimates and prediction intervals of the coefficients of the seasonal annual trend using Model 1. Black indicates that the entire/full data set was used to fit the model, while gray indicates that five test locations were held out to fit the model. The "trend line" is the posterior mean of $\mathbf{X}_\xi \boldsymbol{\alpha}_\xi$ in Equation (4).

is the northernmost test location in Figure 1 and Location 31 is the southernmost test location. The solid black line is the "true" seasonal annual trend, or the posterior mean when Model 1 is fit with the full data. The band around that line is the 95% credible interval; when there is a lot of data at that location, the credible interval is very narrow (as expected) and difficult to see in the figure. The dotted black line is the predicted seasonal annual trend, or the posterior mean when Model 1 is fit without the data from the test locations. While the trends do not match perfectly – and we would not expect them to – they do follow very closely with one another (and match the trend of the data points, marked as open circles) and the credible bands tend to overlap with one another. Thus, the model is able to predict seasonal annual trends at unmonitored locations anywhere along the coast. This figure also shows the spatial relationship of the seasonal trend along the coast; specifically, that northern locations have less pronounced seasons while southern locations have periods of higher sea levels during early summer (June) and fall (October).

We are also interested in examining the unknown trends captured by the spatio-temporal factor analysis portion. The posterior means and credible intervals of the factors when the model is fit with and without the test data are very similar, making it difficult to see a difference in a plot. Therefore, we provide a plot in the Supplementary Document and merely summarize quantitative differences of the factors here. The posterior distribution of the factors when estimated with or without the test data were virtually identical (MSE of the difference in posterior means across the factors was less than 0.00001, with posterior mean values of the factors ranging from -0.21 to 0.34). 95% credible intervals for the factors are similar, with the widths tending to be slightly larger (by an average of 0.0022 and maximum of 0.0061) when the model is fit without the test data. For the loadings, Figure 5 shows the posterior mean estimates with 95% credible intervals. Black indicates the estimates when Model 1 is fit with the full data while gray indicates the estimates when Model 1 is fit with the data from the test locations held out. The posterior mean estimates of the loadings exhibit slight differences in the non-test locations (average absolute difference at non-test locations was less than 0.01, with a maximum absolute difference of 0.06), and varied more among the test locations (average absolute difference at test locations was 0.05, with a maximum absolute difference of 0.23). Figure 5 clearly shows that the spatial pattern of the loadings indicate

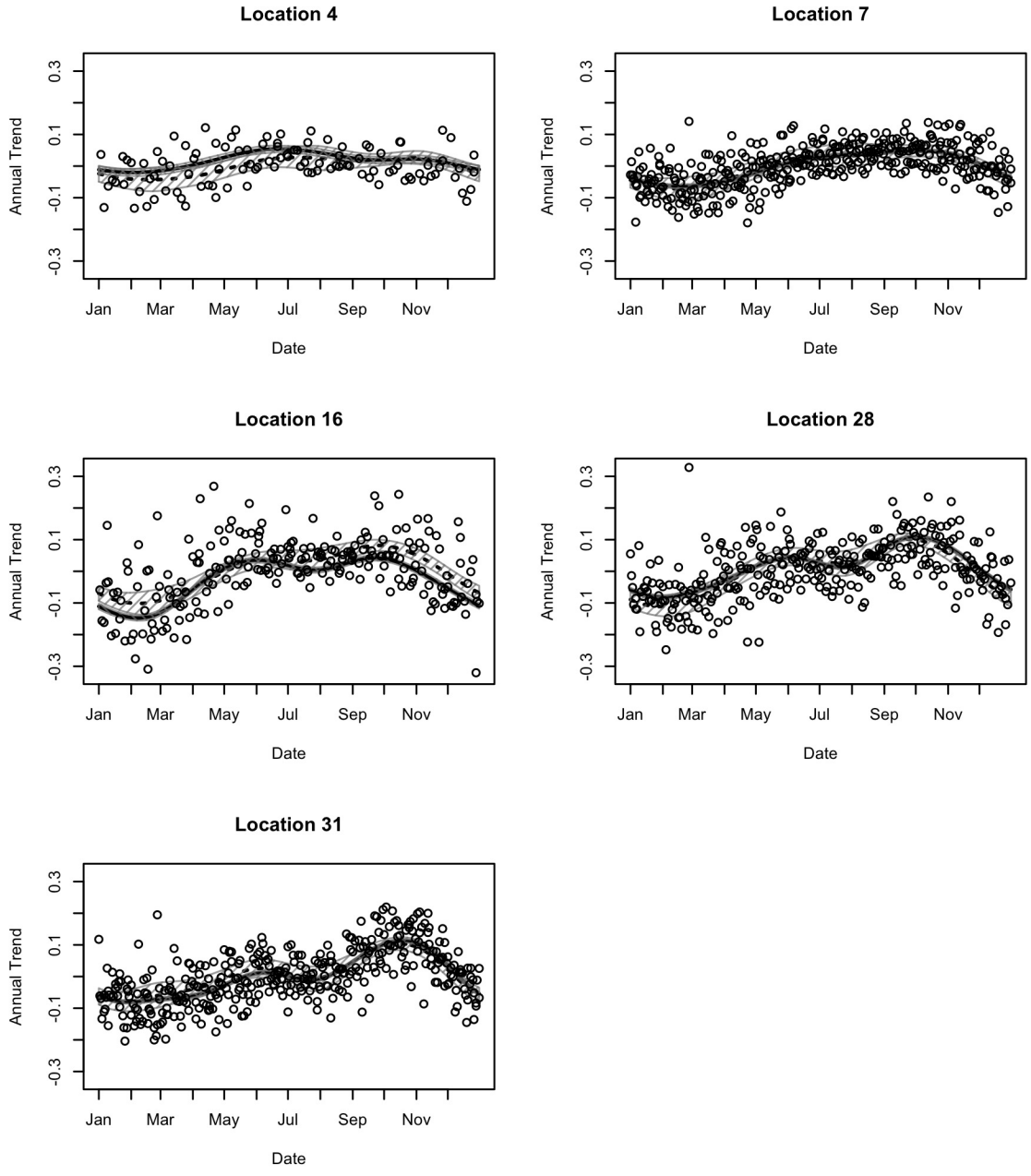

Figure 4: Seasonal annual trends for the five test locations when fit with Model 1. The solid black line is the posterior mean of the smoothed annual trend when the model is fit with all the data; the dotted black line is the posterior mean when the test locations are held out. The bands around each line are the respective 95% credible intervals for the annual trend. The points represent the observed data at each location.

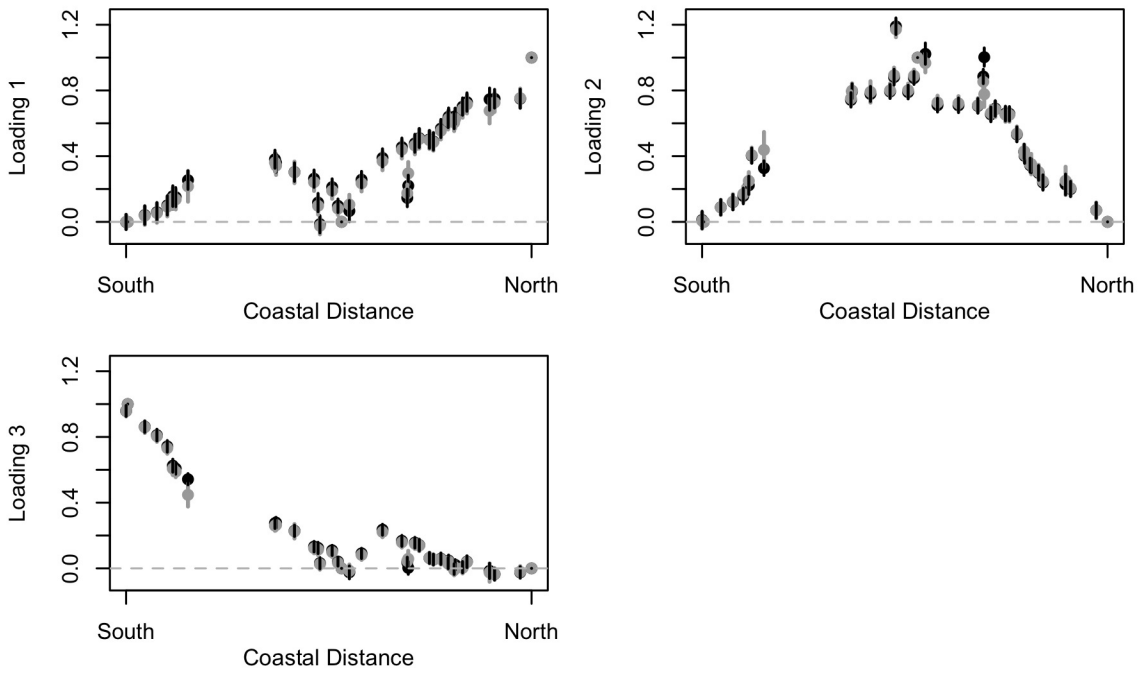

Figure 5: The points represent the posterior means of the loadings for Model 1 when fit with the full data (black) and fit with the test locations held out (gray). The vertical lines represent the 95% credible intervals for the corresponding points.

that Factor 1 loads most heavily on northern locations, Factor 2 loads most heavily on central locations, and Factor 3 loads most heavily on southern locations. Because of this, we call the factors "northern," "central," and "southern" factors, respectively.

In general, Model 1 – or the complete model – is able to capture and estimate trends at unobserved locations quite well. As expected, estimates are not exactly the same as the "true" trend and posterior intervals are larger, but they do follow very similar trends. In Section 4.2 we examine how well the models capture the actual sea level values, but first we continue our examination of the trends by looking at the trends present in other models.

Model 1 vs. Model 4

We developed the complete model as a way to capture, predict, and interpret known trends at unobserved locations. As observed above, Model 1 is able to do this well. However, if we don't directly model the known trends, the factors should be able to capture these trends either in a single

factor or a combination of factors. We proposed the other three models in Table 1 to examine this. To simplify and shorten the comparison, we only compare Models 1 and 4 here, but similar plots are provided in the Supplementary Document comparing the other models.

Figure 6 shows the posterior mean estimates of the loadings for Models 1 and 4. They follow similar patterns, but can differ a fair amount (by values of up to 0.25) from one another. This indicates that the factors from the two models capture slightly different trends. While there are still northern, central, and southern factors in Model 4, the loadings are slightly different and so, for example, the northern factor from Model 4 may capture additional or different trends than the northern factor from Model 1. This is expected since the factors in Model 4 should also be capturing the linear and seasonal annual trends.

While there are many ways to compare the ability of Model 4 (the factors-only model) to capture the known linear and seasonal trends, we choose to do so by comparing similar factors. As with Model 1, the loadings indicate that Model 4 identifies northern, central, and southern factors. We expect that if we subtract the northern factor estimated from Model 1 from the northern factor of Model 4, that the difference will exhibit the linear and/or seasonal annual trends that were not directly modeled in Model 4. Similarly for the central and southern factors. Figure 7 shows these differences using the posterior means when the models are fit with the full data. Immediately we see that each of these differences is estimating a positive linear trend and a cyclical seasonal annual trend, although different from one another. Recall that Factor 1 is the northern factor and that the northern locations have flatter seasonal annual trends (see Figure 4) and smaller linear trends (see Figure 2). The differences in Factor 1 show these properties. Factor 2 is the central factor and has a seasonal annual trend with approximately equal peaks (see Figure 4) and the largest linear trends (see Figure 2). Again, the differences in Factor 2 show these properties. Finally, Factor 3 is the southern factor and the trends seen in the differences of Factor 3 match the larger fall peaks of the seasonal annual trend of southern locations (see Figure 4) and the smaller linear trend (see Figure 2). Trends not directly accounted for in Model 4 are still captured. Likewise, Model 1 may be capturing other trends in the factors, although they are not as easily identified. Without

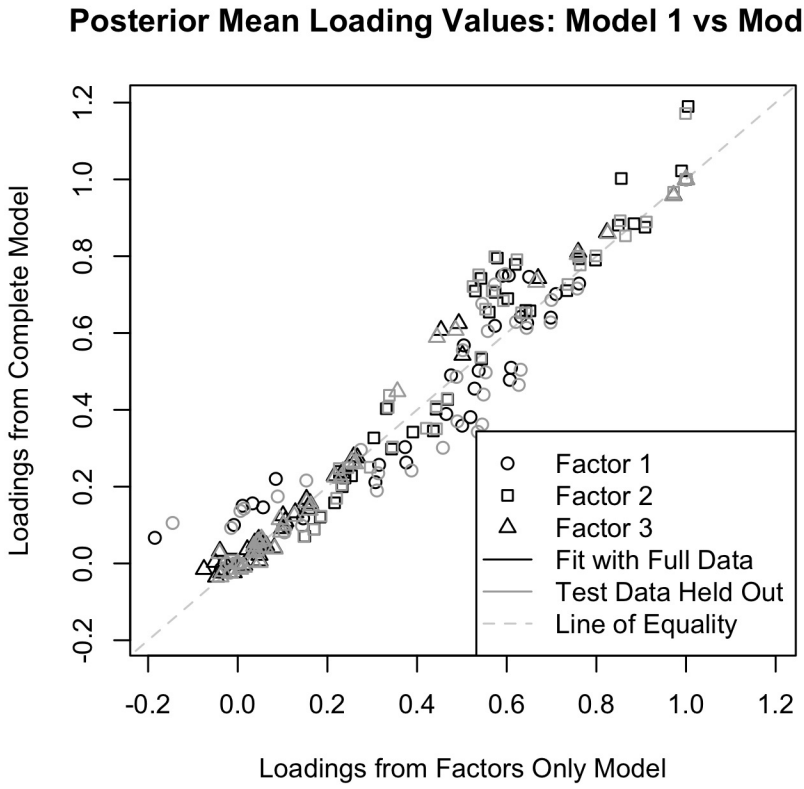

Figure 6: Posterior mean estimates of the loadings for Model 1 (y-axis) and Model 4 (x-axis) when fit with the full data (black) and when fit with the test data held out (gray). Each factor is denoted by a different symbol (circle for Factor 1, square for Factor 2, and triangle for Factor 3). The line indicating equality – where the mean loading values from Models 1 and 4 are equal – is the light gray dashed line.

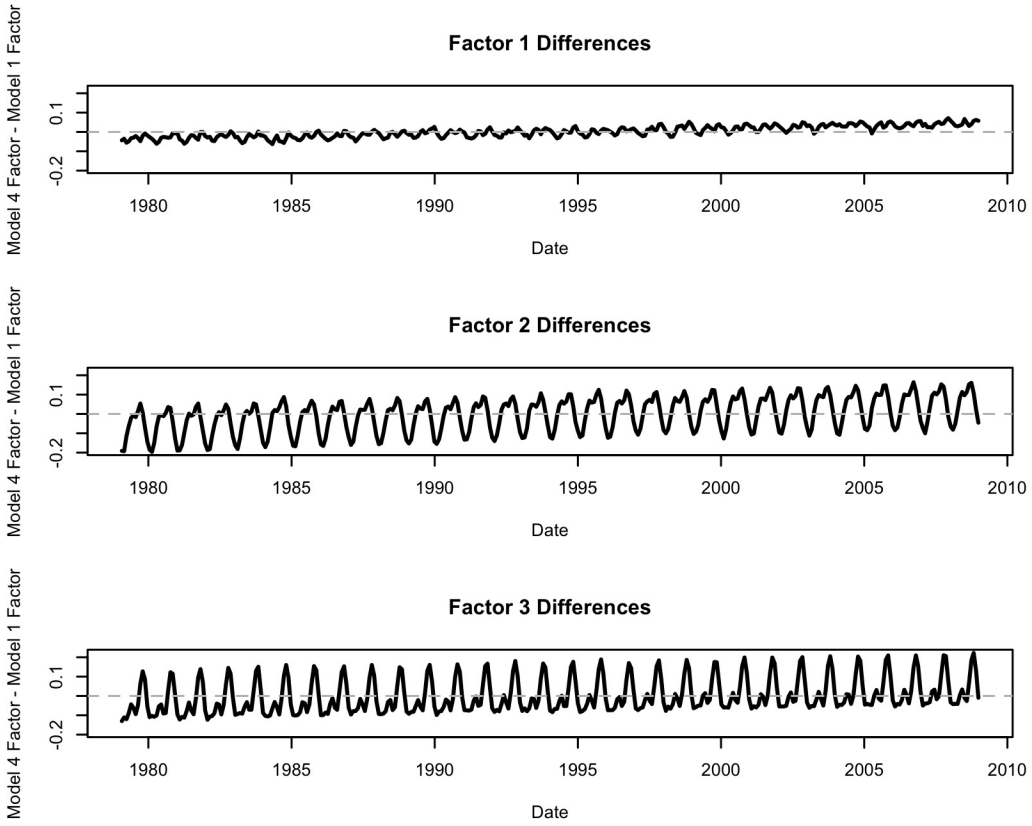

Figure 7: The solid black lines show the differences in the posterior mean estimates between Models 1 and 4 of each factor when the models are estimated with the full data. The dashed line is a horizontal line at 0 so that the linear trend is more easily visible.

subtracting the factors from Model 1, the factors from Model 4 show a visible linear trend, but not a visible seasonal annual trend (see Figure 9 in the Supplementary Document).

4.2 Spatial Prediction of Sea Levels

Here we compare the sea level predictions of the four models using root mean squared error (RMSE) and continuous rank probability scores (CRPS; Gneiting and Raftery, 2007) to compare how well each model predicts the test data. RMSE provides a summary of how well the posterior means match the observations while CRPS accounts for the uncertainty in the posterior distributions of the predictions. In the last section we compared trends using the model fit twice: once using the full data and once with the test data held out. We did this because we wanted to compare true trends

to predicted trends at unobserved locations. However, here we are only interested in predicting the actual sea level values; thus, we only use the models fit with the test data held out. To compute CRPS, we use the approximation method that makes use of MCMC posterior draws proposed by Krüger et al. (2016) and report the absolute values so that smaller CRPS values indicate improved performance.

Table 2 provides the RMSE and CRPS values of each of the models across each test location individually and across all test locations together. This table also provides the percent of missing observations from the data (noted in parentheses next to the location). The values are similar across all models, indicating that when it comes to predicting actual sea level values, directly modeling the trends does not provide a benefit. The bolded values are the smallest, or best, values across the rows; thus, when considering all held out observations (row labeled "All"), Models 1 and 3 perform the best according to both RMSE and CRPS.

Finally, we examine the observed values and the predicted values for the test locations. For brevity, we consider only two of the test locations – the location with the most accurate predictions (Location 7) and the location with the least accurate predictions (Location 16). Figure 8 shows the predictions from Model 1 (gray) and true sea level values (dotted black line) for these two locations. Because the RMSE was so small for Location 7, we are not surprised to see that the posterior mean predicted values fall right in line with the true sea level values. The RMSE was the largest for Location 16 and it's easy to see that the predicted values do not match the true values as well as they do for Location 7. However, we do see that the predicted values follow the pattern of the true sea level values and that in many places the predicted values just appear to be shifted slightly higher than the true values. This difference in predicted values may actually be due to imperfect data rather than imperfect predictions. Recall that we centered each location's sea level values to be 0. Because this location has about ten years of missing data at the beginning of the series and sea levels are rising, it makes sense that by zero-centering we may have shifted the data too far below 0. Had we had a more accurate centering value for these data, we may have seen that the prediction values were actually more accurate than they appear to be here.

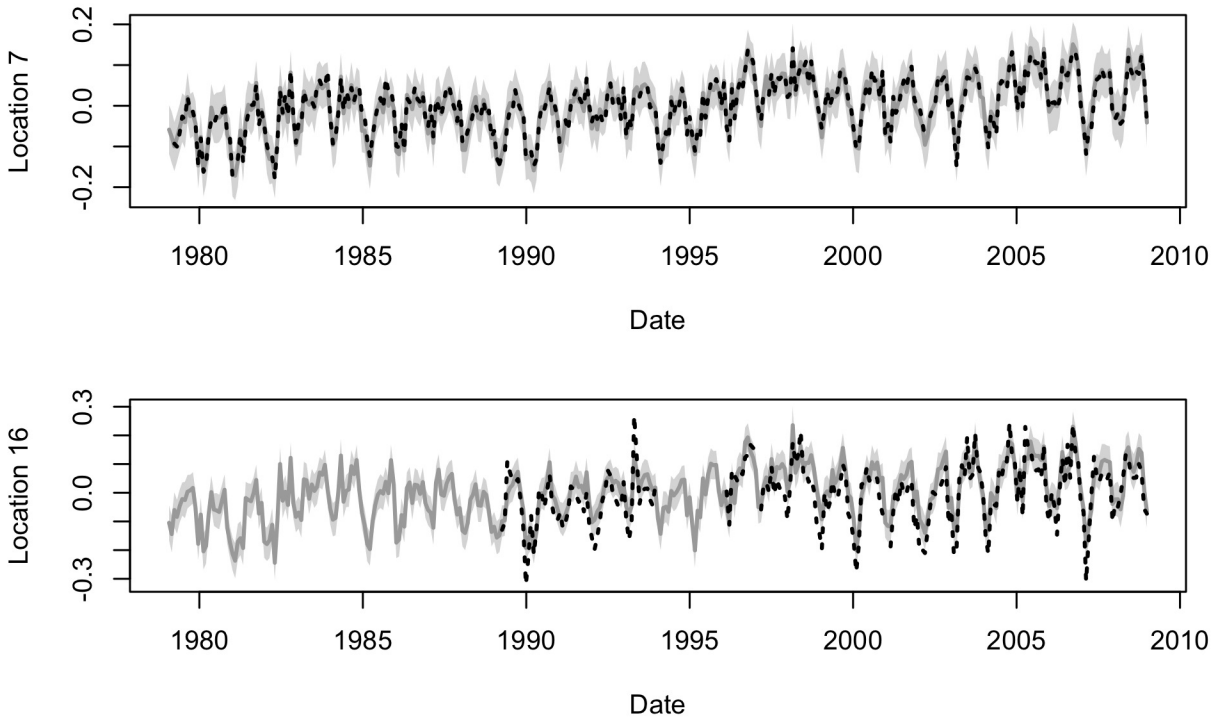

Figure 8: Posterior mean predictions (dark gray solid line) of two test locations, Locations 7 and 16, from Model 1. 95% prediction intervals are represented by the light gray bands. The true sea level values are marked by the dashed black line (times without the black line indicate there was no observed value at that time).

5 Discussion

We proposed a flexible model that is able to capture a variety of known and unknown trends in sea levels on the east coast of the United States and predict trends and sea levels at unmonitored locations. The model allows for nonseparable and nonstationariy space-time processes (Lopes et al., 2008), but with the simplicity and ease of separability and stationarity. Previous analyses have not focused on prediction of trends at local levels, although there is a need for understanding these patterns in order to prepare for future sea level changes and extreme weather events. Comparing different versions of the model, we saw that all versions predicted actual sea levels similarly, but that by directly modeling known trends, interpretable trends could be more easily predicted and understood at unmonitored locations. For example our model predicts the sea level rise ranges between 1 and 5.4 mm/year, depending on the location along the coast.

The spatial dependence of the loadings in the proposed model gives the factors meaning, but it was difficult to identify interpretable temporal trends from the factors. Lopes et al. (2008, 2011) provide a mechanism for enforcing more interpretability in the factors, for example, when defining factors, enforcing one factor to have a cyclical trend. Additionally, this model assumes homogeneous and independent errors. Christensen (2011) shows that for spatial prediction, it is beneficial to allow and account for correlation in the residuals. Allowing for spatial and temporal dependence in the residuals of our model may also be beneficial for more accurately predicting factors and loadings.

While forecasting future sea levels is possible with this model, we do not expect forecasts to be very accurate as the focus of this model was on infilling spatially and temporally unobserved or missing data. A different formulation of the model that makes use of more aspects of dynamic modeling (West and Harrison, 1999) may better be able to forecast future sea levels and offer insight into future sea level and extreme weather scenarios at the local level.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapmap & Hall/CRC, Boca Raton, FL, 2 edn.

Christensen, W. F. (2011), "Filtered kriging for spatial data with heterogeneous measurement error variances," *Biometrics*, 67, 947–957.

Church, J., Clark, P., Cazenave, A., Gregory, J., Jevrejeva, S., Levermann, A., Merrifield, M., Milne, G., Nerem, R., Nunn, P., Payne, A., Pfeffer, W., Stammer, D., and Unnikrishnan, A. (2013), "2013: Sea Level Change," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, B. V., and P. M. Midgley, pp. 1137–1216, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Church, J. A., White, N. J., Coleman, R., Lambeck, K., and Mitrovica, J. X. (2004), "Estimates of the Regional Distribution of Sea Level Rise over the 1950-2000 Period," *Journal of Climate*, 17, 2609–2605.

Coats, D. W. (2015), "Analyzing sea-level change on the east coast with spatiotemporally correlated data," Master's thesis, Brigham Young University, Provo, UT, USA.

Cozannet, G. L., Nicholls, R. J., Hinkel, J., Sweet, W. V., McInnes, K. L., Van de Wal, R. S. W., Slangen, A. B. A., Lowe, J. A., and White, K. D. (2017), "Sea Level Change and Coastal Climate Services: The Way Forward," *Journal of Marine Science and Engineering*, 5, doi:10.3390/jmse5040049.

Cressie, N. and Johannesson, G. (2008), "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society Series B*, 70, 209–226.

Cressie, N. and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Ezer, T. (2013), "Sea level rise, spatially uneven and temporally unsteady: Why the U.S. East Coast, the global tide gauge record, and the global altimeter data show different trends," *Geophyiscal Research Letters*, 40, 5439–5444, doi:10.1002/2013GL057952.

Ezer, T., Atkinson, L. P., Corlett, W. B., and Blanco, J. L. (2013), "Gulf Stream's induced sea level rise and variability along the U.S. mid-Atlantic coast," *Journal of Geophysical Research: Oceans*, 118, 685–697, doi:10.1002/jgrc.20091.

Flegal, J. M., Haran, M., and Jones, G. L. (2008), "Markov chain Monte Carlo: Can we trust the third significant figure?" *Statistical Science*, 23, 250–260.

Gelfand, A., Kim, H., Sirmans, C. F., and Banerjee, S. (2003), "Spatial Modeling with Spatially Varying Coefficient Processes," *Journal of the American Statistical Association*, 98, 387–396.

Gneiting, T. and Raftery, A. E. (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359–378.

Haario, H., Saksman, E., and Tamminen, J. (2005), "Componentwise adaptation for high dimensional MCMC," *Computational Statistics*, 20, 265–273, do: 10.1007/BF02789703.

Hodges, J. S. and Reich, B. J. (2010), "Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love," *The American Statistician*, 64, 325–334.

Hong, B. G., Sturges, W., and Clarke, A. J. (1999), "Sea Level on the U.S. East Coast: Decadal Variability Caused by Open Ocean Wind-Curl Forcing," *Journal of Physical Oceanography*, 30, 2088–2098.

Hughes, J. and Haran, M. (2013), "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models," *Journal of the Royal Statistical Society, Series B*, 75, 139–159.

IPCC (1997), *The Regional Impacts of Climate Change: An Assessment of Vulnerability [R.T.Watson, M.C.Zinyowera, R.H.Moss (eds.)]*, Cambridge University Press, United Kingdom.

IPCC (2014), *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team and Pachauri, R.K. and Meyer, L.A. (eds.)]*, IPCC, Geneva, Switzerland.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006), "Fixed-width output analysis for Markov chain Monte Carlo," *Journal of the American Statistical Association*, 101, 537–547.

Katzfuss, M. and Cressie, N. (2012), "Bayesian hierarchical spatio-temporal smoothing for very large datasets," *Environmetrics*, 23, 94–107, doi: 10.1002/env.1147.

Kopp, R. E., Horton, R. M., Little, C. M., Mitrovica, J. X., Oppenheimer, M., Rasmussen, D. J., Strauss, B. H., and Tebaldi, C. (2014), "Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites," *Earth's Future*, 2, 383–406.

Krüger, F., Lerch, S., Thorarinsdottir, T., and Gneiting, T. (2016), "Probabilistic Forecasting and Comparative Model Assessment Based on Markov Chain Monte Carlo Output," *ArXiv pre-prints*, arXiv:1608.06802.

Lopes, H. F., Salazar, E., and Gamerman, D. (2008), "Spatial Dynamic Factor Analysis," *Bayesian Analysis*, 3, 759–792, doi:10.1214/08-BA329.

Lopes, H. F., Gamerman, D., and Salazar, E. (2011), "Generalized Spatial Dynamic Factor Models," *Computational Statistics and Data Analysis*, 55, 1319–1330, doi:10.1016/j.csda.2010.09.020.

Mimura, N. (2013), "Sea-level rise caused by climate change and its implications for society," *Proceedings of the Japan Academy, Series B*, 89, 281–301, doi: 10.2183/pjab.89.281.

Müller, R. D., Sdrolias, M., Gaina, C., Steinberger, B., and Heine, C. (2008), "Long-Term Sea-Level Fluctuations Driven by Ocean Basin Dynamics," *Science*, 319, 1357–1362, doi: 10.1126/science.1151540.

Peltier, W. R. (1996), "Global sea level rise and glacial isostatic adjustment: An analysis of data from the east coast of North America," *Geophysical Research Letters*, 23, 717–720.

Prime, T., Brown, J. M., and Plater, A. J. (2015), "Physical and Economic Impacts of Sea-Level Rise and Low Probability Flooding Events on Coastal Communities," *PLoS ONE*, 10, DOI:10.1371/journal.pone.0117030.

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Sandholtz, N. G. (2016), "Bayesian Factor Analysis with Spatio-Temporal Dependence: An Application to Sea Level Rise on the East Coast of the United States," Master's thesis, Brigham Young University, Provo, UT, USA.

Sweet, W. V. and Zervas, C. (2011), "Cool-Season Sea Level Anomalies and Storm Surges along the U.S. East Coast: Climatology and Comparison with the 2009/10 El Niño," *Monthly Weather Review*, 139, 2290–2299.

Tebaldi, C., Strauss, B. H., and Zervas, C. E. (2012), "Modelling sea level rise impacts on storm surges along US coasts," *Environmental Research Letters*, 7, doi:10.1088/1748-9326/7/1/014032.

Wahl, T., Haigh, I. D., Nicholls, R. J., Arns, A., Dangendorf, S., Hinkel, J., and Slangen, A. B. A. (2017), "Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis," *Nature Communications*, 8, doi: 10.1038/ncomms16075.

Waller, L. A. and Gotway, C. A. (2004), *Applied spatial statistics for public health data*, vol. 368, John Wiley & Sons.

West, M. and Harrison, J. (1999), *Bayesian Forecasting and Dynamic Models*, Springer, New York, NY, 2 edn.

Williams, S. J. (2013), "Sea-Level Rise Implications for Coastal Regions," *Journal of Coastal Research*, 63, 184–196, doi:10.2112/SI63-015.1.

Wood, S. (2017), *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, 2 edn.

Yin, J. and Goddard, P. B. (2013), "Oceanic control of sea level rise patterns along the East Coast of the United States," *Geophysical Research Letters*, 40, 5514–5520, doi:10.1002/2013GL057992.

Zhang, K., Douglas, B. C., and Leatherman, S. P. (1999), "Twentieth-Century Storm Activity along the U.S. East Coast," *Journal of Climate*, 13, 1748–1761.

Table 1: Description of the four models used for comparison. "Y" indicates the piece is included in the model, while "N" indicates it is not. The final column indicates the number of factors used in the model.

Model	Linear Trend	Annual Trend	Factors
1	Y	Y	3
2	Y	N	3
3	N	Y	3
4	N	N	3

Table 2: Root mean squared errors (RMSE) and continuous rank probability scores (CRPS) for each model's prediction of the test location sea levels. The percents in parentheses to the right of the location indicate the amount of missing data in the observations at that location; the larger the amount of missing data, the smaller the number of observations used to compute the RMSE and CRPS values. Bolded values are the smallest ("best") values across the rows.

	Model	1	2	3	4
RMSE	Location 4 (71.2%)	0.0357	0.0364	**0.0356**	0.0379
	Location 7 (1.6%)	**0.0114**	0.0139	**0.0114**	0.0156
	Location 16 (41.8%)	0.0540	0.0522	0.0530	**0.0514**
	Location 28 (17.8%)	0.0279	**0.0271**	0.0283	0.0273
	Location 31 (0.0%)	**0.0240**	0.0263	0.0247	0.0271
	All	**0.0304**	0.0306	**0.0304**	0.0310
CRPS	Location 4	0.0200	0.0207	**0.0199**	0.0217
	Location 7	0.0084	0.0091	**0.0083**	0.0097
	Location 16	0.0342	0.0333	0.0332	**0.0328**
	Location 28	0.0153	**0.0148**	0.0155	0.0149
	Location 31	**0.0136**	0.0148	0.0140	0.0153
	All	**0.0163**	0.0167	**0.0163**	0.0170