Accounting for Location Measurement Error in Atomic Resolution Images of Crystalline Materials

Matthew J. Miller[a], Matthew J. Cabral[b],Elizabeth C. Dickey[c],
James M. LeBeau[d], and Brian J. Reich[a]

[a] Department of Statistics, North Carolina State University, Raleigh, NC

[b] The University Of Sydney, New South Wales, Australia

[c] Department of Materials Science and Engineering, North Carolina State University Raleigh, NC

[d] Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA

Abstract

Scanning transmission electron microscopy can directly image the atomic structure of materials. To resolve this structure, the material must be aligned along a direction such that columns of atoms are projected onto the image. The local relationships between the intensities and distances of these projected atom columns can inform our understanding of structure–property relationships to ultimately further improve the materials. Measurement error in the atom column locations can, however, introduce bias into parameter estimates. Here, we create a spatial Bayesian hierarchical model that treats the locations as parameters to account for measurement error, and lower the computational burden by approximating the likelihood using a non-contiguous block design around the atom columns. We conduct a simulation study and analyze real data to compare our model to standard spatial and non-spatial models. The results show our method corrects the bias in the parameter of interest, drastically improving upon the standard models.

1. Introduction

From the analysis of atomic resolution STEM images, researchers can determine atom column locations and intensities that reveal a material's local atomic structure and chemical composition, which can sensitively govern material properties. Engineering and controlling material behavior requires the accurate and precise characterization of chemical and structural correlations (Keen and Goodwin, 2015).

Our analysis tests hypotheses about the relationship between the positions of neighboring atom columns shown in Figure 1. However, error from these location measurements can alter this analysis. Therefore, it is important that we account for this measurement error (ME) in our statistical model to make correct conclusions. ME in covariates in linear regression settings results in biased parameter estimates that attenuate towards zero (Carroll et al., 2006). There are a variety of methods to correct for this bias in models with independent error terms, including regression calibration (Carroll and Stefanski, 1990; Gleser, 1990), simulation extrapolation (SIMEX) (Cook and Stefanski, 1994), and Bayesian hierarchical modeling with informative priors on the ME variance based on expert knowledge or repeated measures. Muff et al. (2015) provide a review of Bayesian ME models with several applications and use integrated nested Laplace approximations to carry out their analysis.

The STEM data in Figure 1 exhibit spatial dependence, and so we are interested in ME methods for spatial settings. ME methods for spatial statistics have particularly been developed for spatially misaligned data where covariates are observed at locations different from where the response is observed (Szpiro et al., 2011; Gryparis et al., 2008). Li et al. (2009) create a spatial linear mixed models ME framework and show that regression coefficient attenuation and variance inflation occur with naive estimates in spatial settings as well. Alexeeff et al. (2016) introduce SIMEX for spatial settings where either the data is misaligned or the model is misspecified, and Huque et al. (2016) present a spatial analogue to regression calibration. Recently, Tadayon and Torabi (2018) and Tadayon and Rasekh (2019) have

developed ME models for non-Gaussian settings by incorporating the ME variance into the spatial covariance. These methods all require knowledge of the ME variance, the ability to estimate it, or assumptions about the ME to make the model identifiable.

Spatial statistical models incorporate observation locations into the model design via covariates and covariance functions. In image analysis, algorithms that detect locations of interest may do so with error. Though we may not have prior information on the variance of this ME, we do have the information provided by the entire image. Instead of including informative priors on this variance, we can expand the model into a hierarchical setting that incorporates every pixel and treats the locations as parameters of the model. The data layer of the hierarchy treats pixel intensities as responses and weights each pixel's contribution to locations of interest by its distance from the location.

In STEM images, because atom column locations provide information about material chemistry and structure, ME in these locations could bias our understanding of local effects in these materials. Therefore, these images are natural candidates for the described hierarchical framework. Spatial correlation between pixels, however, creates computational issues, as the large size of the image results in an enormous covariance structure and a likelihood that is impossible to compute. Thus, we must approximate the likelihood or the covariance matrix (or both) in order to implement a computationally tractable Bayesian hierarchical model that accounts for ME in the atom column locations.

Heaton et al. (2019) compare the performance of various low rank and sparse covariance/precision approximations for large data sets. Low rank approximations are popular, but Stein (2014) showed that contiguous independent block likelihood approximations outperform low rank models when the nugget variance is small and the observations are dense. He points out that the independent contiguous block assumption is troubling, and suggests using the Vecchia likelihood method instead (Vecchia, 1988; Katzfuss and Guinness, 2017). These STEM images, however, are the ideal candidates for independent blocks. We see in Figure 1 that atom columns appear as bright circles in the images with dark, low-information

areas around them. We put boxes around the atom columns and treat the observations in one box as independent from the observations in another. We discard the observations outside of the boxes since they contain little information about atom positions. Thus, we have a collection of non-contiguous boxes that we can reasonably assume are independent.

This approach differs from other methods because it uses the data in the image to account for the ME, instead of estimating it or assuming something about the underlying process. Additionally, the computational time scales linearly with the number of atoms, making it feasible to use for very large data sets. Furthermore, while Den Dekker et al. (2005) and Van Aert et al. (2005) characterize structural parameters in atom columns using frequentist methods, they treat residuals as uncorrelated. We incorporate spatial correlation between pixels and atom columns into our model, use Bayesian methods to quantify uncertainty in our parameters, and take advantage of a hierarchical framework to perform inference on parameters that characterize physical and chemical relationships between atoms.

The rest of this article proceeds as follows. In Section 2, we explain how we collected the data. We describe the hierarchical model and approximate likelihood of the data layer in Section 3. In Section 4 we discuss the Markov chain Monte Carlo (MCMC) setup. We compare the hierarchical model with standard spatial and simple linear regression models in Section 5 via a simulation study. We apply and compare these methods on collected STEM image data in Section 6, and conclude in Section 7.

2.   Description of the STEM Image Data

$Pb(Mg_{1/3}Nb_{2/3})O_3$ (PMN) is a relaxor ferroelectric material with perovskite structure. Perovskite crystals have two main types of atom sites, generically called A- and B-sites. In PMN, the A-sites are exclusively lead, while one third of the B-sites contain magnesium and two thirds contain niobium on average. High angle annular dark-field (HAADF) STEM images allows us to directly view and identify columns of A- and B-sites based on intensity.

Figure 1 shows a 551×551 pixel image with $19^2 = 361$ identified B-sites (blue boxes) and $18^2 = 324$ identified A-sites (red boxes). The Appendix describes the atom column identification process. The proportion of niobium and magnesium in a B-site column determines the intensity. We are interested in the relationship between the intensity of the B-site columns and the displacement of the neighboring A-site column from its expected location. Atomic arrangement in relaxor ferroelectrics such as PMN drive their unique material properties. Understanding how individual atoms influence their surrounding structure is important for understanding the origin of material properties, and in turn, how to engineer them for even greater properties (Keen and Goodwin, 2015).

3.   Model Description

We use the Bayesian hierarchical framework for our statistical model. The data layer encapsulates the relationship between the intensities associated with each pixel and the atom column locations and intensities. The process layer models the association between the displacement of the A-site locations from their expected position and the intensities of neighboring B-sites using a Gaussian kernel. We compare the hierarchical model to the described spatial and simple linear regression models with fixed atom column locations.

3.1.   Data Layer

The model for the observed intensities is

$$Y(\mathbf{p}) = \beta_0 + \sum_{i \in \{A,B\}} \sum_{j=1}^{N_i} \beta_{ij} \exp\left(-\frac{\|\mathbf{p} - \mathbf{s}_{ij}\|^2}{2\psi_i^2} \right) + \epsilon(\mathbf{p}), \tag{1}$$

where $Y(\mathbf{p})$ is the intensity at pixel $\mathbf{p}$, N_i is the number of atoms of type i, $\mathbf{s}_{ij}$ is the coordinate vector of the jth atom column of type i, , and ψ_i is the bandwidth parameter for type-i atom columns. The expected intensity decays from the atom column location

Figure 1: *Left*: Scanning Transmission Electron Microscope (STEM) image of Lead Magnesium Niobate (PMN), with red boxes placed around the A-sites and blue boxes around the B-sites. *Bottom Right*: Zoomed in view of atom columns with plotted centers of the columns found from nonlinear least squares. *Top Right*: Rendering of the crystal structure of PMN, showing the A-sites as columns of lead (gray) and the B-sites as columns of niobium (green) and magnesium (yellow).
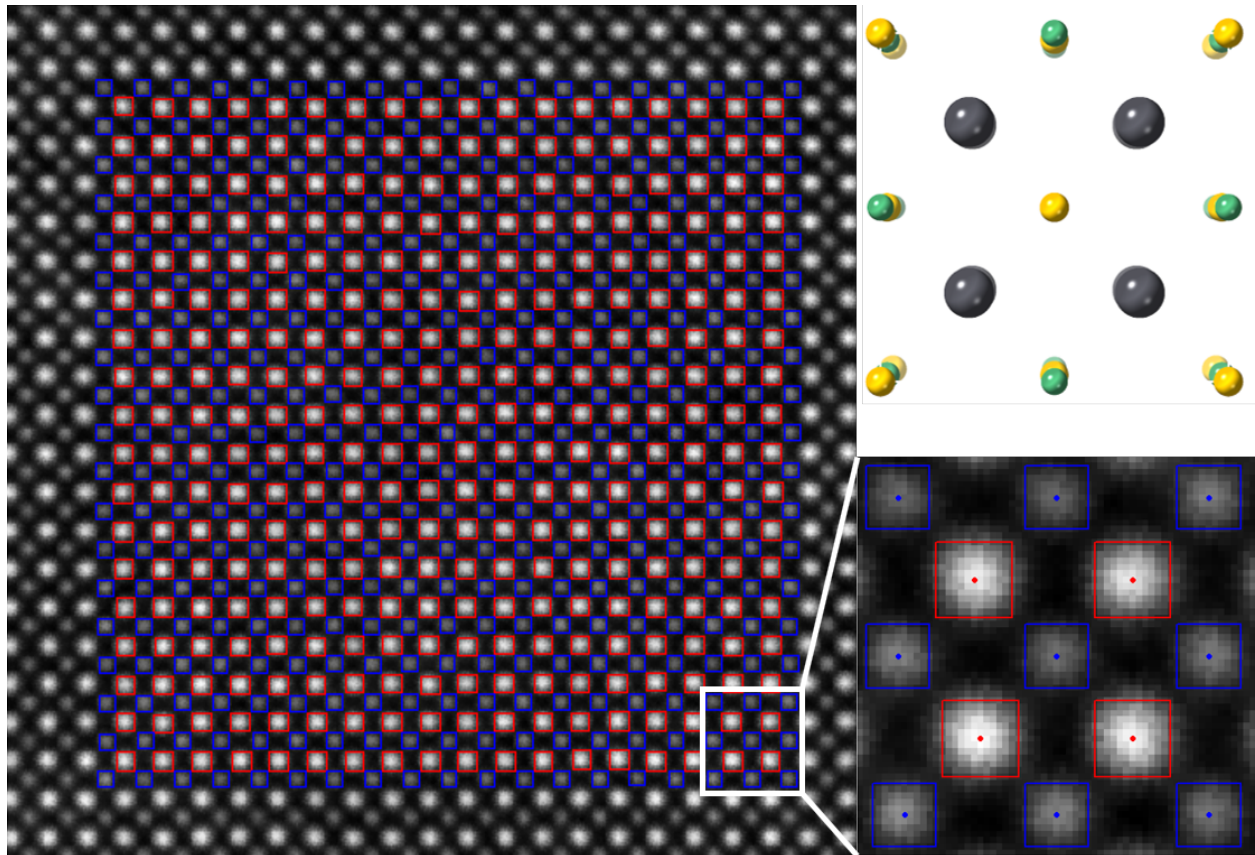
following a Gaussian kernel to place higher intensity value on pixels closer to the center of the nearest atom column. The background intensity β_0 and true atom column intensity β_{ij} control the mean intensity.

The residuals ϵ follow a Gaussian process denoted $GP(\sigma^2, r_{\text{pix}}, \rho_{\text{pix}})$ with mean 0 and exponential covariance function

$$C\big(\epsilon(\mathbf{p}), \epsilon(\mathbf{p'})\big) = \sigma^2 \Big[(1 - r_{\text{pix}}) I(\mathbf{p} = \mathbf{p'}) + r_{\text{pix}} \exp \big(- \frac{\|\mathbf{p} - \mathbf{p'}\|}{\rho_{\text{pix}}} \big) \Big], \qquad (2)$$

where σ^2 is the variance, r_{pix} is the proportion of variance that is spatial, ρ_{pix} is the spatial range and $I(\cdot)$ is the indicator function. The exponential covariance function is a part of the desirable Matérn class of covariance functions where the smoothness parameter is $\frac{1}{2}$ (Gelfand et al., 2010).

We approximate our model as independent across windows surrounding the atoms, as shown in Figure 1. We only consider pixels within square windows $\mathbf{W}_{ij}$ around column $\mathbf{s}_{ij}$, thus moving from the contiguous blocks described by Stein (2014) to non-contiguous blocks of equal sizes for each atom column type. Since the atom columns outside of the window are far from the pixels inside the window, we treat their contributions as negligible. We then approximate our model from (1) as

$$Y(\mathbf{p}_{ijk}) = \beta_0 + \beta_{ij} \exp \Big(- \frac{\|\mathbf{p}_{ijk} - \mathbf{s}_{ij}\|^2}{2\psi_i^2} \Big) + \epsilon(\mathbf{p}_{ijk}), \qquad (3)$$

where $Y(\mathbf{p}_{ijk})$ is the intensity of the kth pixel in $\mathbf{W}_{ij}$ and $\mathbf{p}_{ijk} \in \mathbf{W}_{ij}$ is the 2×1 coordinate vector of that pixel. The covariance for pixels within $\mathbf{W}_{ij}$ follows (2), and is 0 for pixels that are not in the same window.

3.2. Process Layer

The objective of our study is to test whether the displacement of the A-sites from the unweighted center of the neighboring B-sites relates to the intensity of the B-sites. For the

B neighboring B-sites of the jth A-site, the unweighted center is

$$\mathbf{u}_{Aj} = \frac{1}{B}\sum_{k\sim j}\mathbf{s}_{Bk}, \tag{4}$$

where $k\sim j$ denotes the kth neighbor of the jth site. The β_{Bk} parameters in the data layer are the intensities of the B-sites, so the weighted center is

$$\mathbf{w}_{Aj} = \frac{\sum_{j\sim k}\beta_{Bk}\mathbf{s}_{Bk}}{\sum_{k\sim j}\beta_{Bk}}. \tag{5}$$

Figure 2 displays the weighted and unweighted centers. The process layer models the coordinates of the A-site columns, conditioned on the locations of all B-sites $\mathbf{s}_B = \{\mathbf{s}_{Bk}\text{ for all }k\}$:

$$s_{Ajk}|\mathbf{s}_B, \boldsymbol{\beta}, \alpha_0, \alpha_1, \sigma_A^2 = u_{Ajk} + \alpha_0 + \alpha_1(w_{Ajk} - u_{Ajk}) + \varepsilon(s_{Ajk}), \tag{6}$$

where $k\in\{x,y\}$ and s_{Ajk}, u_{Ajk}, and w_{Ajk} are the kth coordinates of $\mathbf{s}_{Aj}, \mathbf{u}_{Aj}$, and $\mathbf{w}_{Aj}$, respectively. The residuals ε are independent between x- and y- coordinates and follow a mean-zero Gaussian process $GP(\sigma_A^2, r, \rho)$ with the exponential covariance structure defined in (2). This model frames the study's objective as a test of whether $\alpha_1 = 0$.

The 2×1 vector $\mathbf{s}_{Aj} - \mathbf{u}_{Aj}$ is the $x-$ and $y-$ displacement of the A-site from the central position, and the displacement resembles simple linear regression with covariate $\mathbf{w}_{Aj} - \mathbf{u}_{Aj}$. The slope parameter α_1 models the linear relationship between displacement of the A-site and the difference between the weighted and unweighted averages of its neighboring B-sites. In other words, a relatively high-intensity B-site is associated with greater A-site displacement. The intercept parameter is α_0 and σ_A^2 is the A-site variance.

We model the B-site locations as

$$\mathbf{s}_{Bj}|\widetilde{\mathbf{s}}_{Bj}, \sigma_B^2 \overset{\text{ind.}}{\sim} \mathbf{N}(\widetilde{\mathbf{s}}_{Bj}, \sigma_B^2\mathbf{I}_2), \tag{7}$$

where $\widetilde{\mathbf{s}}_{Bj}$ is the expected location of the B-site based on the symmetric properties of the

Figure 2: Diagram of negative A-site displacement in response to the difference in intensity-weighted and unweighted averages of neighboring B-sites. $\mathbf{s}_{Bj}$ are the B-sites neighboring A-site $\mathbf{s}_{A1}$, $\mathbf{W}_{A1}$ is the intensity-weighted average of the locations of the B-sites and $\mathbf{U}_{A1}$ is the unweighted average. The black B-site is more intense than the three gray B-sites.
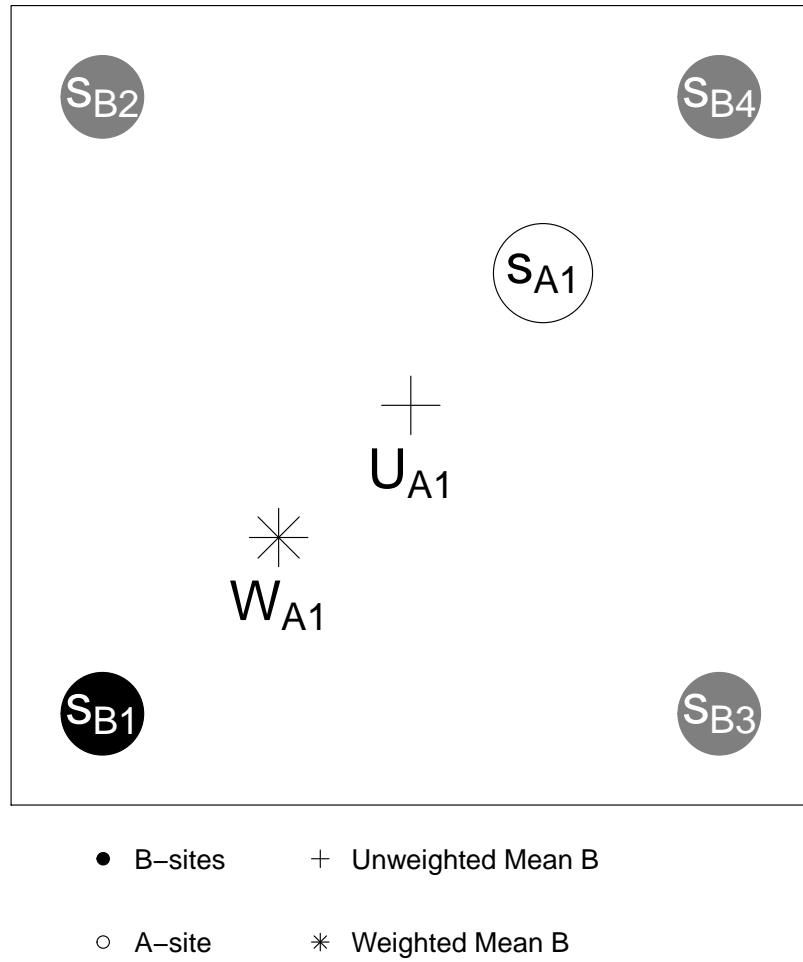
crystal structure of the material. σ_B^2 controls B-site displacement from the crystal structure. We treat the B-sites as uncorrelated because we expect the deviation of the sites from their expected location on the crystal lattice to be small.

3.3. Prior Layer

In general, we choose weakly informative priors for our parameters. The means for the B-sites are on an equispaced grid $\widetilde{\mu}_B$ calculated from the corner sites, which corresponds to the perovskite structure of PMN. We use OLS estimates of the β_{ij} to ground the hyperparameters $\sigma_{\beta_i}^2$ at reasonable values. In particular, we set the mean for $\sigma_{\beta_i}^2$ at the sample standard deviation of the OLS estimates of β_{ij} and the variance of $\sigma_{\beta_i}^2$ at 25^2. We let the priors for σ_i^2 follow Inverse Gamma distributions, so we solve for the shape and rate parameters based on the mean and variance settings. We provide a description of the parameters along with their prior distributions in Table 1.

4. Computing

We use Gibbs and Metropolis sampling in an MCMC framework to sample from the joint posterior distribution of the parameters. The description of the prior distributions of the hierarchical model parameters is in Table 1. For the non-hierarchical models, the regression coefficients α_0 and α_1 have conjugate $N(0, 1000^2)$ priors. We also use a Gibbs sampler for variance σ_A^2, with an InverseGamma$(0.01, 0.01)$ conjugate prior. In the spatial linear regression model, we use Metropolis samplers for the correlation parameters r and ρ with Uniform$(0, 1)$ and LogNormal$(0, 10)$ priors, respectively.

The hierarchical model contains $3(N_A + N_B) + 16$ parameters, where N_i is the number of type-i columns. As such, the number of parameters scale linearly with the number of atom columns. To mitigate the large computational burden we break the image into independent blocks, placing boxes around each column as described in Section 2. The boxes must not

Table 1: Description of parameters, hyperparameters and associated prior distributions for the hierarchical model. $\mathbf{s}_{Aj}$ and $\mathbf{s}_{Bj}$ are considered parameters in the data layer of the model, but responses in the process layer. μ_{Aj} is the mean defined in equation 6 and $\widetilde{\mu}_{Bj}$ is the grid location described in Section 3.1.3. The hyperparameters for the variance of the β_{ij} come from setting the mean to be the sample variance of the OLS estimates and the variance to be 25^2, with $a_i = \frac{\hat{\sigma}^2_{\beta_i}}{25^2} + 2$ and $b_i = \hat{\sigma}^2_{\beta_i}(\frac{\hat{\sigma}^2_{\beta_i}}{25^2} + 1)$.

Parameter(s)	Description	Prior
β_0	Intercept for pixel intensity	$\text{Normal}(0, 1000^2)$
β_{ij}	Slope associated with pixel intensity for atom j of type i	$\text{Normal}(\mu_{\beta_i}, \sigma^2_{\beta})$
$\mu_{\beta_A}, \mu_{\beta_B}$	Hyperparameters; means of the A- and B-site β's	$\text{Normal}(0, 1000^2)$
$\sigma^2_{\beta_A}, \sigma^2_{\beta_B}$	Hyperparameter; variance of β's	$\text{InvGamma}(a_i, b_i)$
ψ_A, ψ_B	Bandwidth for A- and B- site intensities	$\text{LogNormal}(0, 100)$
σ^2	Pixel intensity variance	$\text{InvGamma}(.01, .01)$
r, r_{pix}	Proportion of variance that is spatial for atoms and pixels, respectively	$\text{Uniform}(0, 1)$
ρ, ρ_{pix}	Spatial range for atoms and pixels, respectively	$\text{LogNormal}(0, 100)$
$\mathbf{s}_{Aj}$	Coordinates of the jth A-site	$\text{Normal}(\mu_{Aj}, \sigma^2_A)$
$\mathbf{s}_{Bj}$	Coordinates of the jth B-site	$\text{Normal}(\widetilde{\mu}_{Bj}, \sigma^2_B)$
α_0, α_1	Intercept and slope for A-site displacement vs difference of weighted and unweighted B-site averages	$\text{Normal}(0, 1000^2)$
σ^2_A, σ^2_B	A-site and B-site variance	$\text{InvGamma}(.01, .01)$

11

overlap, or we will count pixels more than once in our analysis and have an invalid model. Therefore, the size of the boxes is important, as they must contain the atom column while not overlapping with the other boxes. It is helpful to orient the image so that it is not necessary to rotate the boxes to be in line with the rows of atom columns.

After selecting box half-widths of h_A and h_B for the A- and B-sites, respectively, we create the boxes by rounding the estimates for the atom column locations to the nearest pixel, then adding and subtracting the half-widths from the x- and y-coordinates to get the pixels inside of the box. Thus we have square boxes of width $2h_i + 1$ around each atom column of interest, where $i \in \{A, B\}$. The approximate likelihood is then

$$p(\mathbf{Y}|\boldsymbol{\Theta}) \approx \prod_{i \in \{A,B\}} \prod_{j=1}^{N_i} p^*(\mathbf{Y}_{ij}|\mathbf{s}_{ij}, \boldsymbol{\Theta}), \tag{8}$$

where $p^*(\cdot)$ is the density from the approximate model in (3), $\mathbf{Y}_{ij}$ is the vector of pixels in window $\mathbf{W}_{ij}$, and Θ is the vector of parameters other than the location of the jth atom column of type i.

Because these boxes are the same size for each atom type, we need only to compute the two pixel-pixel distance matrices (one for A-sites and one for B-sites) for the covariance matrices in the likelihood, making likelihood calculations very efficient. See the Supplementary Materials for the derivations of the sampler updating steps. The code for our MCMC algorithm, simulations, and figures is available at https://github.com/reich-group/HierarchicalSTEM.

5. Simulation Study

We simulate 100 data sets for each model setting, drawing 10,000 posterior samples for each data set after a 10,000 iteration burn-in period. We compare the hierarchical model against the spatial and simple linear regression models with fixed atom column locations described in Section 5.2. The window half-widths are 6 pixels for the A-sites and 5 pixels for the

B-sites.

5.1. Data Generation

We generate data to have similar properties to the real data plotted in Figure 1. We also consider simulations with slightly different true parameters to understand the operating characteristics of the proposed method.

5.1.1. Atom Column Locations

We first draw 19^2 B-sites from a normal distribution where the mean is a grid of points 40 pixels apart and the standard deviation $\sigma_B = 0.25$. To simulate the locations of the corresponding 18^2 A-sites, we first need to generate the β's. We set $\beta_0 = 87$, and independently draw $\beta_{ij} \sim N(\mu_{\beta_i}, \sigma^2_{\beta_i})$, where $\mu_{\beta_A} = 3060$, $\mu_{\beta_B} = 1425$, and $\sigma^2_{\beta_A} = \sigma^2_{\beta_B} = 150$. Letting the A-site distance matrix d be defined by the unweighted means of neighboring B-sites in the mean grid, with $\alpha_0 = -0.08$, $\alpha_1 = -0.15$, $\sigma_A = 0.4, r = 0.73$, and $\rho = 100$, we draw the A-sites from the distribution defined in (6).

5.1.2. Pixel Intensities

We examine five model settings by fixing correlation parameter $r_{pix} = 0.53$ and varying intensity standard deviation σ between 140, 220, and 300 for the first three settings. For the last two, we fix $\sigma = 140$ and change r_{pix} to be 0.7 and 0.9. We set the bandwidth parameters $\psi_A = 4.3$ and $\psi_B = 3.7$ and pixel spatial range $\rho_{pix} = 5.5$. We draw the pixel intensity values based on (1) for pixels within a $2(h_i + 2) + 1$ width box around each atom of type i. The purpose of this is to ensure that the boxes with half-widths h_i drawn around the estimated atom locations contain pixels that follow the proper distribution. The remaining pixel intensities come from an i.i.d $N(\beta_0, 25)$ distribution.

5.1.3. Initial Atom Column Locations

The algorithm described in the Appendix chooses the initial atom column locations by first finding the intensity-weighted average of the nearby pixels and then using nonlinear least squares to refine this estimate. Because we already know the general location of each atom based on the boxes, we skip the Normalized Cross-Correlation (NCC) step, using the pixels inside each corresponding box. In most cases, the nonlinear least squares fit and the intensity-weighted average produce the same location.

5.2. Non-Hierarchical Models

Bayesian spatial and simple linear regression using fixed atom column locations provide faster and more straightforward analyses at the cost of bias and variance inflation from naive parameter estimates. We estimate the atom locations using the non-linear least squares method described in the Appendix, and assume them to be known for the remainder of the analysis. We modify the models from Cabral (2018) by combining the x- and y-displacements into one vector. The new models are of the form

$$\delta(\mathbf{s}_{Aj}) = \alpha_0 + \alpha_1 \mathbf{\Psi}(\mathbf{s}_{Aj}) + \epsilon(\mathbf{s}_{Aj}), \tag{9}$$

where $\delta(\mathbf{s}_{Aj}) = \mathbf{s}_{Aj} - \mathbf{u}_{Aj}$, $\mathbf{\Psi}(\mathbf{s}_{Aj}) = \mathbf{w}_{Aj}^* - \mathbf{u}_{Aj}$, and $\text{Cov}(\epsilon(\mathbf{s}_{Aj})) = \sigma_A^2 \mathbf{I}_2$. $\mathbf{u}_{Aj}$ is defined in (4) and

$$\mathbf{w}_{Aj}^* = \frac{\sum_{j \sim k} \hat{I}(\hat{\mathbf{s}}_{\mathbf{Bk}}) \hat{\mathbf{s}}_{Bk}}{\sum_{j \sim k} \hat{I}(\hat{\mathbf{s}}_{Bk})}, \tag{10}$$

which is the analogue for (5) when every pixel is not in the model. $I(\mathbf{s}_{Bk})$ is the intensity found from the nonlinear least squares fit in the Appendix. The covariance structure for the spatial linear regression model is the same as for the hierarchical model and the residuals for the simple linear regression model are i.i.d. normal with mean 0 and variance σ_A^2.

5.3. Results

We are primarily interested in the slope parameter α_1. Table 2 displays the bias of the posterior means, mean posterior standard deviation, coverage and estimated Mean Squared Error $(\widehat{MSE})$ in all model settings. The hierarchical model has the highest coverage and lowest $\widehat{MSE}$ for α_1 compared to simple and spatial linear regression for every setting. The hierarchical model captures the true regression coefficient, while the posterior mean estimator of α_1 in the spatial and simple linear regression models attenuates towards zero, as expected from the measurement error literature. The attenuation contributes to poor coverage in the naive models, whereas the hierarchical models perform well until the intensity standard deviation σ increases drastically. We see the model performance start to decline when $\sigma = 220$, and perform much worse when $\sigma = 300$.

Table 3 displays the results of more parameters from the initial model setting. For the parameters common between the three models, the hierarchical model has the best coverage, though the spatial linear regression model has tighter posteriors for the correlation parameters, which is reflected in MSE estimates. The data layer parameters have less than 95% coverage, but the bias and means of the posterior standard deviations show that they are close to the truth for the most part. The low coverage may be explained by the pixels inside the windows not capturing all of the information in the model. However, the parameters of interest are in the process layer, not the data layer, and this model sees better performance in the process layer parameters than the spatial and simple linear regression models.

6. STEM Image Analysis

The PMN image in Figure 1 contains 19^2 A-sites and 18^2 B-sites for analysis. We run the MCMC for each model for 90,000 iterations after a 10,000 iteration burn-in and check convergence visually via trace plots. We compare the hierarchical model with half-width 6 for the A-sites and 5 for the B-sites against the spatial and simple linear regression models

Table 2: Summary of simulation study performance for estimating $\alpha_1 = -0.15$ under various parameter settings for simple linear regression (SimpLR), spatial linear regression (SpatLR), and our new Bayesian hierarchical model (Hierarch). We simulated 100 data sets for each parameter setting. Monte Carlo standard errors are in parentheses.

Statistics	Model	$r_{pix} = 0.53$ $\sigma = 140$	$r_{pix} = 0.53,$ $\sigma = 220$	$r_{pix} = 0.53,$ $\sigma = 300$	$r_{pix} = 0.7,$ $\sigma = 140$	$r_{pix} = 0.9$ $\sigma = 140$
	SimpLR	0.037 (0.0013)	0.087 (0.0013)	0.130 (0.0150)	0.042 (0.0013)	0.052 (0.0013)
Bias	SpatLR	0.037 (0.0013)	0.086 (0.0013)	0.121 (0.0097)	0.042 (0.0012)	0.052 (0.0012)
	Hierarch	-0.002 (0.0016)	0.027 (0.0022)	0.081 (0.0114)	-0.004 (0.0018)	-0.004 (0.0020)
	SimpLR	0.016 (0.0001)	0.014 (0.0001)	0.015 (0.0006)	0.016 (0.0001)	0.015 (0.0001)
Mean Post. SD	SpatLR	0.012 (0.0001)	0.012 (0.0001)	0.013 (0.0006)	0.012 (0.0001)	0.012 (0.0001)
	Hierarch	0.017 (0.0002)	0.026 (0.0003)	0.026 (0.0003)	0.018 (0.0002)	0.020 (0.0002)
	SimpLR	37 (4.8)	0 (0)	0 (0)	21 (4.1)	3 (1.7)
% Coverage	SpatLR	22 (4.1)	0 (0)	0 (0)	7 (2.7)	0 (0)
	Hierarch	95 (2.2)	85 (3.6)	25 (4.3)	97 (2.0)	97 (1.7)
	SimpLR	0.15 (0.011)	0.77 (0.022)	3.94 (2.35)	0.20 (0.011)	0.29 (0.014)
$\widehat{MSE} \times 100$	SpatLR	0.15 (0.010)	0.76 (0.022)	2.39 (1.08)	0.19 (0.010)	0.28 (0.013)
	Hierarch	0.03 (0.004)	0.12 (0.014)	1.94 (1.40)	0.23 (0.005)	0.04 (0.005)

Table 3: Simulation study results for 100 simulated data sets with $\alpha_1 = -0.15$ for simple linear regression (SimpLR), spatial linear regression (SpatLR), and our new Bayesian hierarchical model (Hierarch). Coverage is the percent of 95% highest posterior density credible intervals that contain the parameter.

Parameter	Model	Truth	Bias	(SE)	Mean Post. SD	(SE)	Coverage (%)	(SE)	$\widehat{MSE}$	(SE)
α_0	SimpLR	-0.08	0.019	(0.008)	0.018	(0.0001)	40	(4.9)	0.0064	(0.0001)
	SpatLR		-0.018	(0.007)	0.070	(0.0017)	92	(2.7)	0.0055	(0.0008)
	Hierarch		-0.034	(0.007)	0.096	(0.0046)	95	(2.2)	0.0061	(0.009)
σ_A	SimpLR	0.4	0.053	(0.002)	0.013	(0.0001)	6	(2.4)	0.0032	(0.0002)
	SpatLR		0.064	(0.002)	0.027	(0.0001)	13	(3.4)	0.0045	(0.0003)
	Hierarch		0.034	(0.005)	0.049	(0.0002)	94	(2.4)	0.0034	(0.0008)
r	SpatLR	0.73	-0.048	(0.008)	0.082	(0.0015)	88	(3.2)	0.0091	(0.0016)
	Hierarch		-0.114	(0.007)	0.085	(0.0015)	80	(4.0)	0.0174	(0.0020)
ρ	SpatLR	100	-5.5	(3.2)	31.0	(1.30)	86	(3.5)	1053	(19.3)
	Hierarch		64.5	(9.9)	98.4	(10.11)	97	(1.7)	1385	(386.4)
β_0	Hierarch	87	7.59	(0.45)	4.78	(0.014)	67	(4.7)	77.4	(7.67)
β_{A100}	Hierarch	3006.21	63.04	(13.6)	79.4	(0.017)	69	(4.6)	22204	(3571)
σ	Hierarch	140	-2.22	(0.13)	1.18	(0.001)	47	(5.0)	6.62	(0.675)
ψ_A	Hierarch	4.3	-0.01	(0.0007)	0.008	(0.0000)	79	(4.1)	0.0001	(0.00002)
r_{pix}	Hierarch	0.57	-0.01	(0.0008)	0.008	(0.0000)	48	(5.0)	0.0003	(0.00003)
ρ_{pix}	Hierarch	5.5	-0.33	(0.024)	0.217	(0.0021)	63	(4.8)	0.17	(0.019)

Table 4: Posterior mean and highest posterior density 95% credible intervals for the 5 common parameters among the hierarchical, spatial linear regression, and spatial linear regression models.

| | Hierarchical Model | | Spatial LR | | Simple LR | |
	Mean	Credible Int.	Mean	Credible Int.	Mean	Credible Int.
α_0	-0.06	(-0.34, 0.21)	-0.09	(-0.32, 0.15)	-0.09	(-0.12, -0.06)
α_1	-0.29	(-0.36, -0.23)	-0.19	(-0.22, -0.16)	-0.19	(-0.24, -0.13)
σ_A	0.38	(0.28, 0.53)	0.42	(0.34, 0.52)	0.40	(0.37, 0.42)
r	0.72	(0.56, 0.87)	0.83	(0.73, 0.91)	–	–
ρ	205	(58, 440)	122	(54, 225)	–	–

described in Section 5.2.

The results of the analysis are as expected. Table 4 shows that the posterior means for α_1 in the simple and spatial linear regression models are much closer to zero than in the hierarchical model, and the variance is inflated, as we expect because of measurement error. The estimated effect is 53% larger for the full model than for the standard models. We visualize these results in the density plots in Figure 3. We also see the posterior intervals and means for the atom column locations in Table 4. The spatial linear regression model puts a wider interval on the intercept term α_0 than the simple linear regression model, which allows for a narrower interval around the regression coefficient of interest α_1. The spatial linear regression credible interval for α_1 does not overlap with the interval from the hierarchical model, providing strong evidence of attenuation.

All three models indicate strong evidence of a negative relationship between A-site column displacement and B-site intensity through the parameter α_1. In other words, A-site column locations tend to be further from B-sites with higher proportions of magnesium. These findings are consistent with observations made with x-ray diffraction that propose the distribution of magnesium and niobium directly influences the bonding between lead and oxygen (Chen et al., 1996; Jeong et al., 2005).

Figure 3: Posterior densities for the five common parameters between the hierarchical, spatial linear regression, and simple linear regression models. The regression parameter α_1 attenuates towards zero in the simple and spatial linear regression models.
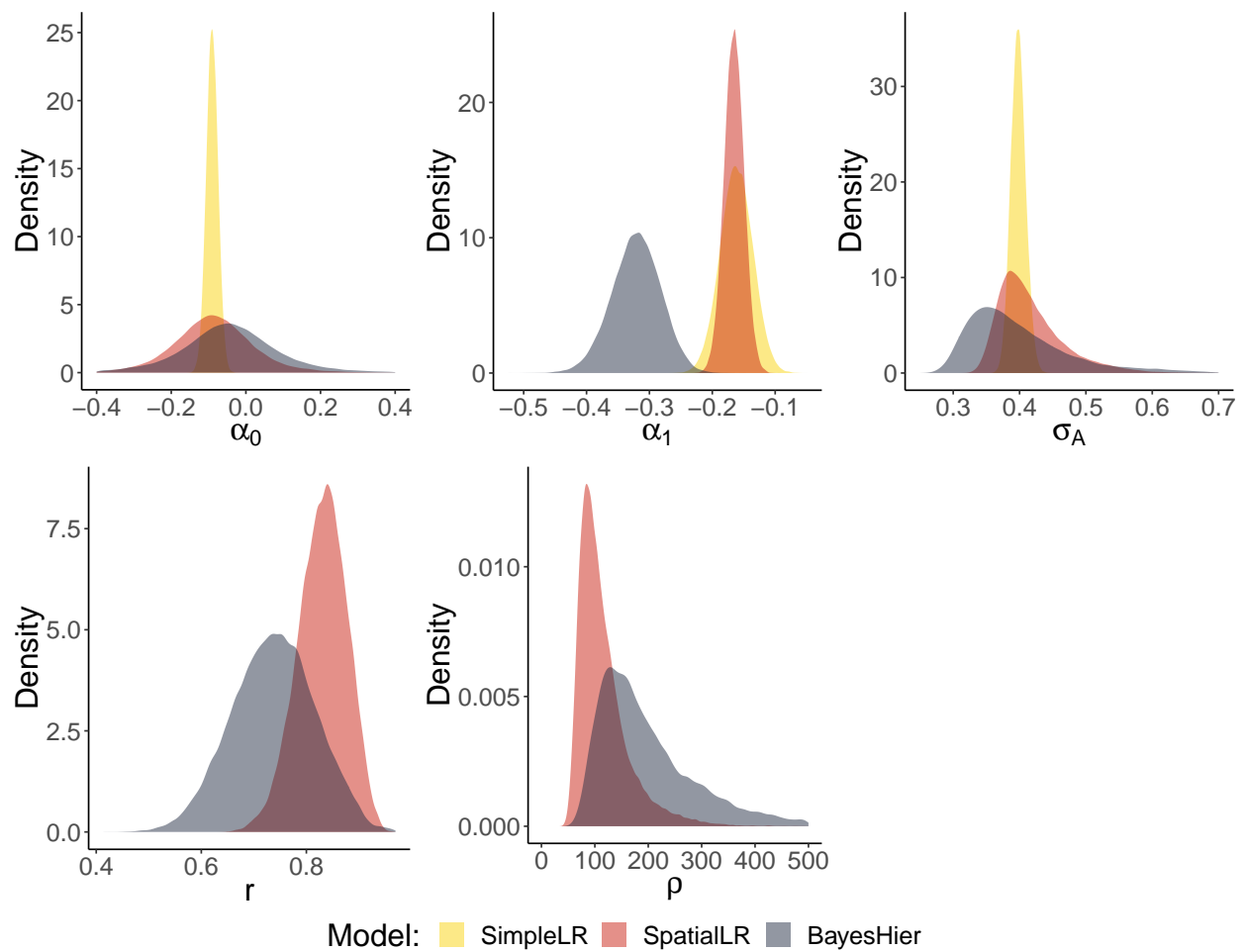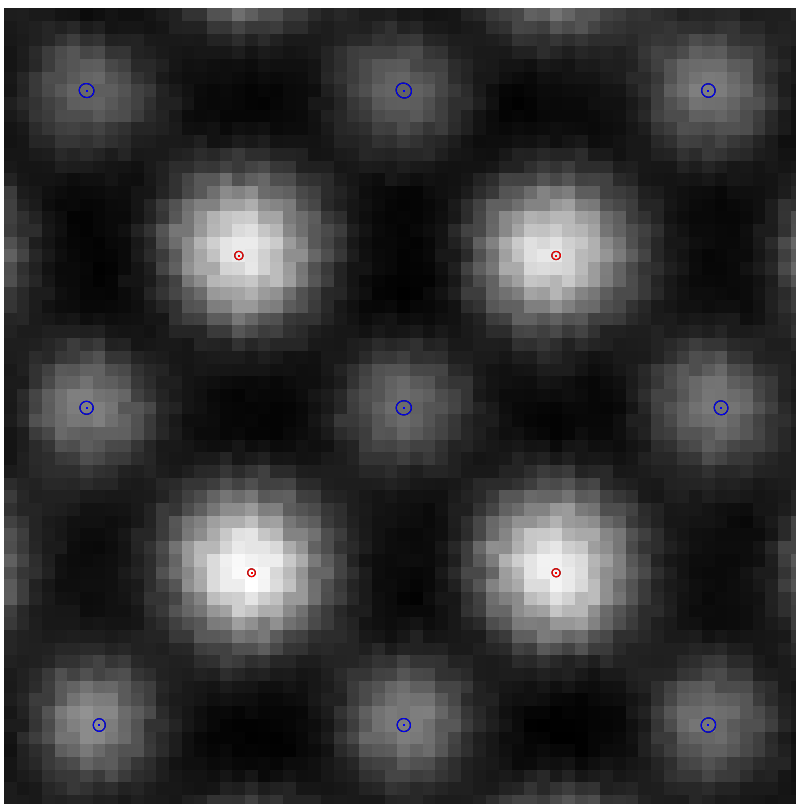
Figure 4: 95% posterior regions (circles) and means (points) for atom column locations from the inset image in Figure 1.

7. Discussion

Electron microscopy imaging techniques will continue to improve and provide us with an ever clearer picture of how local physical and chemical differences contribute to global material properties. This article describes a spatial Bayesian hierarchical model that accounts for measurement error in locations for atomic-scale images of crystalline materials. Our new method is a dramatic improvement over the standard analysis techniques, and as such we hope it will become an impactful tool for materials scientists. We apply this model to real and simulated STEM images of PMN, and show that it outperforms spatial and simple linear regression where the estimated locations are treated as the truth. This method is computationally intensive compared to the naive models, as the number of parameters scale with the number of atom columns and the data layer uses intensities at each pixel as responses. However, using independent non-contiguous blocks around the atom columns allows the time to scale linearly with the number of columns.

For the type of data explored in our application, the non-contiguous block method is limited by the maximum size of the windows around the atom columns. The blocks cannot overlap, because the information in the overlapping region would be counted twice. Rotating the image so that the angle of the rows of atom columns aligns with the blocks will help maximize the block size. We can also modify this model to apply it to different types of crystal structures and zone axes.

Appendix

A. Finding Initial Atom Column Locations

We adopt the methods described by Sang and LeBeau (2014) to find the initial estimates of the atom column locations, first by using the NCC to find the region for each column and then using the intensity-weighted average of the pixels as an initial location estimate. We use nonlinear least squares to fit the equation

$$g(x,y) = A \exp \left\{ - \left[\frac{\left((x - x_0)\cos\theta + (y - y_0)\sin\theta\right)^2}{\sigma_1^2} \right. \right.$$
$$\left. \left. - \frac{\left((x - x_0)\sin\theta - (y - y_0)\cos\theta\right)^2}{\sigma_2^2} \right] \right\} + Z, \quad (11)$$

where $g(x,y)$ is the atom column intensity, Z is the background intensity at pixel (x,y), A is the peak intensity with background removed, θ is the rotation angle, and σ_1 and σ_2 are the standard deviations along the axes of the ellipse. The initial value for background intensity Z_0 is the difference in the median and standard deviation of the column intensity, and the initial amplitude estimate A_0 is the difference between Z_0 and the maximum intensity for the column. The coordinates (x_0, y_0) are the true atom column position. We use non-linear least squares to minimize (over $\{A, \theta, \sigma_1^2, \sigma_2^2, z, x_0, y_0\}$) the average squared error between $g(x,y)$ and the fitted model and obtain estimates of x_0 and y_0. The spatial and simple linear regression models described in Section 3 used these fitted values as the known atom column locations and intensities, and the hierarchical model uses them as initial values for

the MCMC algorithm used to sample from the posterior.

B.   Full Conditional Derivations

Here we provide the derivations of the full conditional derivations used for MCMC. Let N_i be the number of atoms of type i and M_{ij} be the the number of pixels in window $\mathbf{W}_{ij}$.

B.1.   Data Layer

B.1.1.   Background Intensity Parameter β_0

Let the prior distribution for β_0 be $N(0, \sigma_{\beta_0}^2)$, and $\mathbf{Q}_i$ be the pixel-pixel precision matrix for a box around atom columns of type i. Let $\mathbf{X}_{ij} = [\exp(-\frac{\|\mathbf{p}_{ij1} - \mathbf{s}_ij\|^2}{2\psi_i^2}), \ldots, \exp(-\frac{\|\mathbf{p}_{ijN_i} - \mathbf{s}_ij\|^2}{2\psi_i^2})]^T$

and $\mathbf{u}_{ij} = \beta_0 \mathbf{1}_{M_i} + \beta_{ij} \mathbf{X}_{ij}$. Then, using the approximate likelihood from equation 8,

$$
\begin{aligned}
p(\beta_0 | \mathbf{Y}, \mathbf{\Theta}_{[-\beta_0]}) &\propto p(\mathbf{Y}|\mathbf{\Theta})p(\beta_0) \\
&\approx \prod_{i \in \{A,B\}} \prod_{j=1}^{N_i} p(\mathbf{Y}_{ij}|\mathbf{s}_{ij}, \mathbf{\Theta}) \exp\left(-\frac{\beta_0^2}{2\sigma_{\beta_0}^2}\right) \\
&\propto \exp\left(-\frac{\beta_0^2}{2\sigma_{\beta_0}^2}\right) \prod_{i \in \{A,B\}} \prod_{j=1}^{N_i} \exp\left[-\frac{1}{2}(\mathbf{Y}_{ij} - \mu_{ij})^T \mathbf{Q}_i (\mathbf{Y}_{ij} - \mu_{ij})\right] \\
&\propto \exp\left(-\frac{\beta_0^2}{2\sigma_{\beta_0}^2}\right) \\
&\quad \times \prod_{i \in \{A,B\}} \prod_{j=1}^{N_i} \exp\left\{-\frac{1}{2}\left[\beta_0^2 \mathbf{1}_{M_i}^T \mathbf{Q}_i \mathbf{1}_{M_i} - 2\beta_0 (\mathbf{Y}_{ij} - \beta_{ij}\mathbf{X}_{ij})^T \mathbf{Q}_i \mathbf{1}_{M_i}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\beta_0^2\left(\frac{1}{\sigma_{\beta_0}^2} + \sum_{i \in \{A,B\}} N_i(\mathbf{1}_{M_i}^T \mathbf{Q}_i \mathbf{1}_{M_i})\right)\right.\right. \\
&\quad \left.\left. - 2\beta_0 \sum_{i \in \{A,B\}} \sum_{j=1}^{N_i} (\mathbf{Y}_{ij} - \beta_{ij}\mathbf{X}_{ij})^T \mathbf{Q}_i \mathbf{1}_{M_i}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_{\beta_0}^2} + \sum_{i \in \{A,B\}} N_i(\mathbf{1}_{M_i}^T \mathbf{Q}_i \mathbf{1}_{M_i})\right]\right. \\
&\quad \times \left[\beta_0^2 - 2\beta_0\left(\frac{1}{\sigma_{\beta_0}^2} + \sum_{i \in \{A,B\}} N_i(\mathbf{1}_{M_i}^T \mathbf{Q}_i \mathbf{1}_{M_i})\right)^{-1}\right. \\
&\quad \left.\left. \times \sum_{j=1}^{N_i} (\mathbf{Y}_{ij} - \beta_{ij}\mathbf{X}_{ij})^T \mathbf{Q}_i \mathbf{1}_{M_i}\right]\right\}.
\end{aligned}
$$

If we let $V_{\beta_0} = \frac{1}{\sigma_{\beta_0}^2} + \sum_{i \in \{A,B\}} N_i(\mathbf{1}_{M_i}^T \mathbf{Q}_i \mathbf{1}_{M_i})$ and $M_{\beta_0} = \sum_{j=1}^{N_i} (\mathbf{Y}_{ij} - \beta_{ij}\mathbf{X}_{ij})^T \mathbf{Q}_i \mathbf{1}_{M_i}$, then

$$
\beta_0 | \mathbf{Y}, \mathbf{\Theta}_{[-\beta_0]} \sim N\left(\frac{M_{\beta_0}}{V_{\beta_0}}, \frac{1}{V_{\beta_0}}\right). \tag{12}
$$

B.1.2. A-site Intensity Parameters β_{Aj}

Let the prior distribution for β_{Aj} be $N(\mu_{\beta_A}, \sigma_\beta^2)$, and let μ_{Aj}, $\mathbf{X}_{Aj}$ and $\mathbf{Q}_A$ be defined as in the previous subsection. Then,

$$p(\beta_{Aj}|\mathbf{Y}, \boldsymbol{\Theta}_{[-\beta_{Aj}]}) \propto p(\mathbf{Y}|\boldsymbol{\Theta})p(\beta_{Aj})$$

$$\approx \prod_{i \in \{A,B\}} \prod_{j=1}^{N_i} p(\mathbf{Y}_{ij}|\mathbf{s}_{ij}, \boldsymbol{\Theta}) \exp(-\frac{(\beta_{Aj} - \mu_{\beta_A})^2}{2\sigma_{\beta_A}^2})$$

$$\propto p(\mathbf{Y}_{Aj}|s_{Aj}, \boldsymbol{\Theta}) \exp(-\frac{(\beta_{Aj} - \mu_{\beta_A})^2}{2\sigma_{\beta_A}^2})$$

$$\propto \exp\left\{ -\frac{1}{2} \left[(\mathbf{Y}_{Aj} - \mu_{Aj})^T \mathbf{Q}_A (\mathbf{Y}_{Aj} - \mu_{Aj}) + (\beta_{Aj} - \mu_{\beta_A})^2\right] \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left[\beta_{Aj}^2 (\mathbf{X}_{Aj}^T \mathbf{Q}_A \mathbf{X}_{Aj} + \frac{1}{\sigma_{\beta_A}^2}) \right.\right.$$

$$\left.\left. - 2\beta_{Aj}((\mathbf{Y}_{Aj} - \beta_0 \mathbf{1}_{M_A})^T \mathbf{Q}_A (\mathbf{Y}_{Aj} - \beta_0 \mathbf{1}_{M_A}) + \frac{\mu_{\beta_A}}{\sigma_\beta^2})\right] \right\}$$

Let $M_{\beta_{Aj}} = (\mathbf{Y}_{Aj} - \beta_0 \mathbf{1}_{M_A})^T \mathbf{Q}_A (\mathbf{Y}_{Aj} - \beta_0 \mathbf{1}_{M_A}) + \frac{\mu_{\beta_A}}{\sigma_\beta^2}$ and $V_{\beta_{Aj}} = \mathbf{X}_{Aj}^T \mathbf{Q}_A \mathbf{X}_{Aj} + \frac{1}{\sigma_{\beta_A}^2}$. After factoring out $V_{\beta_{Aj}}$ and completing the square, we see

$$\beta_{Aj}|\mathbf{Y}, \boldsymbol{\Theta}_{[-\beta_{Aj}]} \sim N(\frac{M_{\beta_{Aj}}}{V_{\beta_{Aj}}}, \frac{1}{V_{\beta_{Aj}}}) \tag{13}$$

B.1.3. B-site Intensity Parameters β_{Bj}

Because the B-sites are used in determining the A-site locations, we cannot derive a full conditional, and must use Metropolis sampling instead. However, we draw from a $N(\frac{M_{\beta_{Bj}}}{V_{\beta_{Bj}}}, \frac{1}{V_{\beta_{Bj}}})$ to get our candidate instead of using the standard method. $M_{\beta_{Bj}}$ and $M_{\beta_{Bj}}$ are calculated the same way as $M_{\beta_{Aj}}$ and $M_{\beta_{Aj}}$, but with replacing the A's for B's.

B.1.4. Variance Parameter σ^2

Let the prior distribution for σ^2 be $\text{InvGamma}(c, d)$. Let μ_{ij} be defined as before and $\mathbf{Q}_i^* = \sigma^2 \mathbf{Q}_i$. Then,

$$
\begin{aligned}
p(\sigma^2 | \mathbf{Y}, \mathbf{\Theta}_{[-\sigma^2]}) &\propto p(\mathbf{Y} | \mathbf{\Theta}) p(\sigma^2) \\
&\approx \prod_{i \in \{A, B\}} \prod_{j=1}^{N_i} p(\mathbf{Y}_{ij} | \mathbf{s}_{ij}, \mathbf{\Theta})(\sigma^2)^{-c-1} \exp(-\frac{d}{\sigma^2}) \\
&\propto (\sigma^2)^{-(\frac{N_a + N_b}{2} + c) - 1} \exp \Big\{ -\frac{1}{2\sigma^2} \Big[2d \\
&\quad + \sum_{i \in \{A, B\}} \sum_{j=1}^{N_i} (\mathbf{Y}_{ij} - \mu_{ij})^T \mathbf{Q}_i^* (\mathbf{Y}_{ij} - \mu_{ij}) \Big] \Big\},
\end{aligned}
$$

which is the kernel of an inverse gamma distribution. So,

$$
\sigma^2 | \mathbf{Y}, \mathbf{\Theta}_{[-\sigma^2]} \sim \text{IG}\Big(\frac{N_a + N_b}{2} + c, d + \frac{1}{2} \sum_{i \in \{A, B\}} \sum_{j=1}^{N_i} (\mathbf{Y}_{ij} - \mu_{ij})^T \mathbf{Q}_i^* (\mathbf{Y}_{ij} - \mu_{ij}) \Big) \tag{14}
$$

B.2. Process Layer

Let $\delta_x = (\delta_{x1}, \delta_{x2}, \cdots, \delta_{xa})^T$, where $\delta_{xj} = s_{A_j x} - U_{A_j x}$, the difference of the x-coordinates of the jth atom column location and the 2×1 vector defined in in equation 4. Let $\mathbf{\Psi}_x = (\mathbf{\Psi}_{x1}, \mathbf{\Psi}_{x2}, \cdots, \mathbf{\Psi}_{xa})^T$, where $\mathbf{\Psi}_{xj} = W_{A_j x} - U_{A_j x}$, the difference of the x-coordinates defined in equations 5 and 4, respectively. Define δ_y and $\mathbf{\Psi}_y$ similarly. Then, the distributions of their likelihoods are:

$$
\delta_i | \mathbf{\Theta} \overset{\text{ind}}{\sim} N(\alpha_0 \mathbf{1}_a + \alpha_1 \mathbf{\Psi}_i, \mathbf{V}), \tag{15}
$$

where

$$
\mathbf{V} = \sigma_A^2 \Big[(1 - r) \mathbf{I}_a + r \exp\big(-\frac{1}{\rho} \mathbf{D}_A \big) \Big]. \tag{16}
$$

$\mathbf{\Theta} = (\alpha_0, \alpha_1, \sigma_a, r, \rho, \mathbf{S}_B, \beta_{B1}, \ldots, \beta_{Bb})^T$, where $\mathbf{S}_B$ is the $b \times 2$ matrix of B-site locations and b is the number of B-sites. $\mathbf{1}_a$ is an $a \times 1$ vector of 1's, $\mathbf{D}_A$ is the Euclidean distance matrix

between the A-sites and $\mathbf{I}_a$ is the $a \times a$ identity matrix, where a is the number of A-sites. Finally, let $\boldsymbol{\Theta}_{[-p]}$ be the vector Θ with parameter p removed.

B.2.1. Intercept Parameter α_0

Recall from 1 the prior distribution for α_0 is $N(0, 1000^2)$. We will generalize here and let $\alpha_0 \sim N(0, \sigma_{\alpha_0}^2)$. Let $\mu_i = \delta_i - \alpha_0 \mathbf{1}_a - \alpha_1 \boldsymbol{\Psi}_i$ Then, we derive the full conditional distribution:

$$p(\alpha_0 | \delta_x, \delta_y, \boldsymbol{\Theta}_{[-\alpha_0]}) \propto p(\delta_x | \boldsymbol{\Theta}) p(\delta_y | \boldsymbol{\Theta}) p(\alpha_0)$$

$$\propto \exp\left\{ -\frac{1}{2}\Big[\sum_{i \in \{x,y\}} (\mu_i^T \mathbf{V}^{-1} \mu_i) + \frac{\alpha_0^2}{\sigma_{\alpha_0}^2} \Big] \right\}$$

$$\propto \exp\left\{ -\frac{1}{2}\Big[\sum_{i \in \{x,y\}} (\alpha_0^2 \mathbf{1}_a^T \mathbf{V}^{-1} \mathbf{1}_a \right.$$

$$\left. - 2\alpha_0 (\delta_i - \alpha_1 \boldsymbol{\Psi}_i)^T \mathbf{V}^{-1} \mathbf{1}_a) + \frac{\alpha_0^2}{\sigma_{\alpha_0}^2} \Big] \right\}$$

$$\propto \exp\left\{ -\frac{1}{2}\Big[\alpha_0^2 (2(\mathbf{1}_a^T \mathbf{V}^{-1} \mathbf{1}_a) + \frac{1}{\sigma_{\alpha_0}^2}) \right.$$

$$\left. - 2\alpha_0 \sum_{i \in \{x,y\}} (\delta_i - \alpha_1 \boldsymbol{\Psi}_i) \mathbf{V}^{-1} \mathbf{1}_a \Big] \right\}$$

$$\propto \exp\left\{ -\frac{2(\mathbf{1}_a^T \mathbf{V}^{-1} \mathbf{1}_a) + \frac{1}{\sigma_{\alpha_0}^2}}{2} \Big[\alpha_0^2 \right.$$

$$\left. - 2\alpha_0 \frac{\sum_{i \in \{x,y\}} (\delta_i - \alpha_1 \boldsymbol{\Psi}_i) \mathbf{V}^{-1} \mathbf{1}_a}{2(\mathbf{1}_a^T \mathbf{V}^{-1} \mathbf{1}_a) + \frac{1}{\sigma_{\alpha_0}^2}} \Big] \right\}.$$

If we let $V_{\alpha_0} = 2(\mathbf{1}_a^T \mathbf{V}^{-1} \mathbf{1}_a) + \frac{1}{\sigma_{\alpha_0}^2}$ and $M_{\alpha_0} = \sum_{i \in \{x,y\}} (\delta_i - \alpha_1 \boldsymbol{\Psi}_i) \mathbf{V}^{-1} \mathbf{1}_a$, we see after completing the square that

$$\alpha_0 | \delta_x, \delta_y, \boldsymbol{\Theta}_{[-\alpha_0]} \sim N\left(\frac{M_{\alpha_0}}{V_{\alpha_0}}, \frac{1}{V_{\alpha_0}}\right). \tag{17}$$

B.2.2. Slope Parameter α_1

As we did with α_0, we will generalize the prior for α_1 and let $\alpha_1 \sim N(0, \sigma_{\alpha_1}^2)$. Again let μ_i be defined as above. Then,

$$p(\alpha_1|\delta_x, \delta_y, \boldsymbol{\Theta}_{[-\alpha_1]}) \propto p(\delta_x|\boldsymbol{\Theta})p(\delta_y|\boldsymbol{\Theta})p(\alpha_1)$$

$$\propto \exp\left\{-\frac{1}{2}\Big[\sum_{i\in\{x,y\}}(\mu_i^T\mathbf{V}^{-1}\mu_i) + \frac{\alpha_1^2}{\sigma_{\alpha_1}^2}\Big]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\Big[\sum_{i\in\{x,y\}}(\alpha_1^2\boldsymbol{\Psi}_i\mathbf{V}^{-1}\boldsymbol{\Psi}_i - 2\alpha_1(\delta_i - \alpha_0\mathbf{1}_a)) + \frac{\alpha_1^2}{\sigma_{\alpha_1}}\Big]\right\}$$

$$\propto \exp\left\{-\frac{\sum_{i\in\{x,y\}}\boldsymbol{\Psi}_i\mathbf{V}^{-1}\boldsymbol{\Psi}_i + \frac{1}{\sigma_{\alpha_1}^2}}{2}\Big[\alpha_1^2 \right.$$

$$\left. -2\alpha_1\frac{\sum_{i\in\{x,y\}}(\delta_i - \alpha_0\mathbf{1}_a)}{\sum_{i\in\{x,y\}}\boldsymbol{\Psi}_i\mathbf{V}^{-1}\boldsymbol{\Psi}_i + \frac{1}{\sigma_{\alpha_1}^2}}\Big]\right\}$$

Letting $V_{\alpha_1} = \sum_{i\in\{x,y\}}\boldsymbol{\Psi}_i\mathbf{V}^{-1}\boldsymbol{\Psi}_i + \frac{1}{\sigma_{\alpha_1}^2}$ and $M_{\alpha_1} = \sum_{i\in\{x,y\}}(\delta_i - \alpha_0\mathbf{1}_a)$, after completing the square we have

$$\alpha_1|\delta_x, \delta_y, \boldsymbol{\Theta}_{[-\alpha_1]} \sim N\Big(\frac{M_{\alpha_1}}{V_{\alpha_1}}, \frac{1}{M_{\alpha_1}}\Big). \tag{18}$$

B.2.3. Variance Parameter σ_A^2

Let the prior distribution for σ_A^2 be InvGamma(f, g). Let $\mathbf{V}^* = \frac{1}{\sigma_A^2}\mathbf{V}$, and μ_i be defined the same as in the previous subsections. Then,

$$p(\sigma_A^2|\delta_x, \delta_y, \boldsymbol{\Theta}_{[-\sigma_A^2]}) \propto p(\delta_x|\boldsymbol{\Theta})p(\delta_y|\boldsymbol{\Theta})p(\sigma_A^2)$$

$$\propto |\sigma_A^2\mathbf{V}^*|^{-1}\exp\Big[-\frac{1}{2\sigma_A^2}\sum_{i\in\{x,y\}}\mu_i^T(\mathbf{V}^*)^{-1}\mu_i\Big]$$

$$\times (\sigma_A^2)^{-f-1}\exp\Big(-\frac{g}{\sigma_A^2}\Big)$$

$$\propto (\sigma_A^2)^{-(f+N_A)-1}\exp\left\{-\frac{1}{\sigma_A^2}\Big(g + \frac{\sum_{i\in\{x,y\}}\mu_i^T(\mathbf{V}^*)^{-1}\mu_i}{2}\Big)\right\},$$

27

which is the kernel of an inverse gamma distribution, so

$$\sigma_A^2|\delta_x, \delta_y, \boldsymbol{\Theta}_{[-\sigma_A^2]} \sim \text{InvGamma}(f + N_A, g + \frac{\sum_{i \in \{x,y\}} \mu_i^T (\mathbf{V}^*)^{-1} \mu_i}{2}). \tag{19}$$

B.3. Prior Layer

B.3.1. Mean and Variance for β_A and β_B

The mean and variance parameters for β_A and β_B have priors determined by the OLS estimates of the $\beta'_{ij}s$, as described in Table 1. Let $\hat{\mu_{\beta_i}}$ be the mean of the OLS estimates for the β_{ij}'s, and a_i and b_i be defined as in the Table 1 caption. Then we have conjugate posteriors, with

$$\mu_{\beta_i}|\beta_A, \sigma_{\beta_i}^2, \hat{\mu_{\beta_i}} \sim N\left(\frac{\frac{\hat{\mu_{\beta_i}}}{1000^2} + \sum_{j=1}^{N_i} \beta_{ij}}{\frac{1}{1000^2} + \frac{N_i}{\sigma_{\beta_i}^2}}, \frac{1}{\frac{1}{1000^2} + \frac{N_i}{\sigma_{\beta_i}^2}}\right) \tag{20}$$

and

$$\sigma_{\beta_i}^2|\mu_{\beta_i}, \beta_A, a_i, b_i \sim \text{InvGamma}\left(a_i + N_i/2, b_i + \frac{1}{2}\sum_{j=1}^{N_i}(\beta_{ij} - \mu_{\beta_i})^2\right) \tag{21}$$

B.3.2. B-site variance σ_B^2

Because the B-site locations are modeled as independent, the conjugate full conditional for σ_B^2 is standard. Let $\sigma_B^2 \sim \text{InvGamma}(l, m)$. Let μ_B be the matrix of grid locations described in Section 3.1.3 and $\mathbf{s}_B$ the matrix of B-site locations. Then,

$$\sigma_B^2|\widetilde{\mu}_B, \mathbf{s}_B \sim \text{InvGamma}\left(N_B + l, m + \frac{1}{2}\sum_{j=1}^{N_B}\|\mathbf{s}_{Bj} - \mu_{Bj}\|^2\right). \tag{22}$$

References

Alexeeff, S. E., Carroll, R. J., and Coull, B. (2016), "Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures," *Biostatistics*, 17, 377–389.

Cabral, M. C. (2018), "Quantifying Short-Range Chemical and Structural Order in Complex Oxides via Scanning Transmission Electron Microscopy," Ph.D. thesis, North Carolina State University.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman and Hall/CRC.

Carroll, R. J. and Stefanski, L. A. (1990), "Approximate quasi-likelihood estimation in models with surrogate predictors," *Journal of the American Statistical Association*, 85, 652–663.

Chen, I. W., Li, P., and Wang, Y. (1996), "Structural origin of relaxor perovskites," *Journal of Physics and Chemistry of Solids*, 57, 1525–1536.

Cook, J. R. and Stefanski, L. A. (1994), "Simulation-extrapolation estimation in parametric measurement error models," *Journal of the American Statistical Association*, 89, 1314–1328.

Den Dekker, A., Van Aert, S., Van den Bos, A., and Van Dyck, D. (2005), "Maximum likelihood estimation of structure parameters from high resolution electron microscopy images. Part I: A theoretical framework," *Ultramicroscopy*, 104, 83–106.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010), *Handbook of Spatial Statistics*, CRC press, chap. 2, pp. 24–26.

Gleser, L. (1990), "Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models," *Contemp. Math*, 112, 99–114.

Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J., and Coull, B. A. (2008), "Measurement error caused by spatial misalignment in environmental epidemiology," *Biostatistics*, 10, 258–274.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., et al. (2019), "A case study competition among methods for analyzing large spatial data," *Journal of Agricultural, Biological and Environmental Statistics*, 1–28.

Huque, M. H., Bondell, H. D., Carroll, R. J., and Ryan, L. M. (2016), "Spatial regression with covariate measurement error: A semiparametric approach," *Biometrics*, 72, 678–686.

Jeong, I.-K., Darling, T. W., Lee, J. K., Proffen, T., Heffner, R. H., Park, J. S., Hong, K. S., Dmowski, W., and Egami, T. (2005), "Direct observation of the formation of polar

nanoregions in Pb(Mg1/3Nb2/3)O3 using neutron pair distribution function analysis," *Physical Review Letters*, 94, 147602.

Katzfuss, M. and Guinness, J. (2017), "A general framework for Vecchia approximations of Gaussian processes," *arXiv preprint arXiv:1708.06302*.

Keen, D. A. and Goodwin, A. L. (2015), "The crystallography of correlated disorder," *Nature*, 521, 303–309.

Li, Y., Tang, H., and Lin, X. (2009), "Spatial linear mixed models with covariate measurement errors," *Statistica Sinica*, 19, 1077.

Muff, S., Riebler, A., Held, L., Rue, H., and Saner, P. (2015), "Bayesian analysis of measurement error models using integrated nested Laplace approximations," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64, 231–252.

Sang, Xiahan, O. A. A. and LeBeau, J. M. (2014), "Atom column indexing: atomic resolution image analysis through a matrix representation," *Microscopy and Microanalysis*, 20, 1764–1771.

Stein, M. L. (2014), "Limitations on low rank approximations for covariance matrices of spatial data," *Spatial Statistics*, 8, 1–19.

Szpiro, A. A., Sheppard, L., and Lumley, T. (2011), "Efficient measurement error correction with spatially misaligned data," *Biostatistics*, 12, 610–623.

Tadayon, V. and Rasekh, A. (2019), "Non-Gaussian covariate-dependent spatial measurement error model for analyzing big spatial data," *Journal of Agricultural, Biological and Environmental Statistics*, 24, 49–72.

Tadayon, V. and Torabi, M. (2018), "Spatial models for non-Gaussian data with covariate measurement error," *Environmetrics*, e2545.

Van Aert, S., Den Dekker, A., Van Den Bos, A., Van Dyck, D., and Chen, J. (2005), "Maximum likelihood estimation of structure parameters from high resolution electron microscopy images. Part II: a practical example," *Ultramicroscopy*, 104, 107–125.

Vecchia, A. V. (1988), "Estimation and model identification for continuous spatial processes," *Journal of the Royal Statistical Society: Series B (Methodological)*, 50, 297–312.