



Published in final edited form as:

Technometrics. 2019 ; 61(4): 494–506. doi:10.1080/00401706.2018.1546622.

Spatial Signal Detection Using Continuous Shrinkage Priors

An-Ting Jhuang¹, Montserrat Fuentes², Jacob L. Jones³, Giovanni Esteves³, Chris M. Fancher⁴, Marschall Furman¹, Brian J. Reich¹

¹Department of Statistics, North Carolina State University, Raleigh, NC 27695

²College of Humanities and Sciences, Virginia Commonwealth University, Richmond, VA 23284

³Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC 27695

⁴Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Abstract

Motivated by the problem of detecting changes in two-dimensional X-ray diffraction data, we propose a Bayesian spatial model for sparse signal detection in image data. Our model places considerable mass near zero and has heavy tails to reflect the prior belief that the image signal is zero for most pixels and large for an important subset. We show that the spatial prior places mass on nearby locations simultaneously being zero, and also allows for nearby locations to simultaneously be large signals. The form of the prior also facilitates efficient computing for large images. We conduct a simulation study to evaluate the properties of the proposed prior and show that it outperforms other spatial models. We apply our method in the analysis of X-ray diffraction data from a two-dimensional area detector to detect changes in the pattern when the material is exposed to an electric field.

Keywords

Bayesian variable selection; High-dimensional data; Image analysis; X-ray diffraction

1 Introduction

X-ray diffraction (XRD) is a powerful technique to characterize the atomic structure of a material. There are nearly 300,000 structures cataloged in the International Center for Diffraction Data (Thomas, 2012; Editorial, 2014). For example, we analyze the XRD pattern of lead zirconate titanate (PZT). PZT is a useful functional material known as a piezoelectric. As a piezoelectric, PZT changes shape when an electric field is applied, making it useful for actuation, positional control and energy conversion. The material we investigate in the present study is a commercially available PZT polycrystalline ceramic material known by the tradename K350 (Piezo Technologies). This material has been studied previously under varying temperature, electric field, pressure and deformation, making it an established material (e.g., Katrusiak, 2008; Dutta and Singh, 2011; Gorfman et al., 2011; Esteves et al., 2015) that is suitable for the present investigation. The data in the present experiment are acquired using a two-dimensional area detector that measures diffracted intensity; this method can be referred to as 2-D XRD. The data are shown in

Figure 1 and described in detail in Section 2. In a 2-D XRD image, each pixel represents the z-score under an electric field relative to several baseline images. There are few z-scores absolutely greater than two, and these pixels tend to cluster in rings indicative of atomic structure. Therefore, we analyze these data using sparse spatial signal detection methods. Our objective is to develop a more powerful statistical method for detecting and quantifying structural changes in 2-D XRD patterns under different experimental conditions.

The statistical problem of identifying spatial regions affected by an experimental factor has many applications, including epidemiology, neuroscience and materials science. As a result, there is a rich literature on the topic. The spatial scan statistic (Naus, 1965; Kulldorff, 1997; Costa and Kulldorff, 2009; Liao et al., 2017) searches for regions with different means than the background, but does not estimate the signal and is not well suited for large images with many significant subregions. The two-dimensional fused lasso (Friedman et al., 2007), the graphical lasso (Friedman et al., 2008) and the smooth-sparse decomposition method (Yan et al., 2017) are penalized regression methods that account for spatial structure in the signal using penalties to encourage spatial smoothness. Yan et al. (2017) incorporates denoising and signal detection into one step for images with a smooth background and facilitates to keep sharp boundaries which is not achieved in traditional two-step procedures (Qiu and Yandell, 1997; Bradley and Roth, 2007; Sollie, 2013). Spatial wavelet shrinkage methods impose a threshold on coefficients in the wavelet domain to recover a sparse signal (Donoho and Johnstone, 1994; Taswell, 2000; Jansen, 2001; Yadav et al., 2014; He and Wang, 2017). These regularization methods can be applied to high-dimensional data, but require presetting the tuning parameters via cross validation (Mallick and Yi, 2013) and fail to account for all sources of uncertainty.

Our approach builds on Bayesian variable selection methods. This allows us to fully account for uncertainty in the posterior distribution and incorporate known atomic structure in the prior. An intuitive sparse prior is a two-component mixture (spike and slab) with one component concentrated near zero for the unimportant features and the other diffuse for the signals (Mitchell and Beauchamp, 1988; George and McCulloch, 1993; Yuan and Lin, 2005; Geweke, 1996; George and McCulloch, 1997; O'hara and Sillapää, 2009; Ro ková and George, 2016). Two-component mixture priors can be extended to the spatial setting to identify subregions of interest (Goldsmith et al., 2014; Boehm Vock et al., 2015; Li et al., 2015; Kang et al., 2018). However, the two-component construction is computationally challenging because posterior sampling requires a search over a considerable space of complex models and is plagued by slow convergence and poor mixing (Carvalho et al., 2010; Johnson and Rossell, 2012; Mallick and Yi, 2013).

The computational difficulties of spike-and-slab priors are abated by continuous shrinkage priors (Carvalho et al., 2010; Griffin and Brown, 2010; Armagan et al., 2013; Bhattacharya et al., 2014; Bhadra et al., 2016b; Piironen and Vehtari, 2016). Rather than a discrete mixture over two components, these priors continuously mix over shrinkage parameters. For example, Carvalho et al. (2010) proposed the horseshoe prior that assumes a normal prior with mean zero and standard deviation that follows a half-Cauchy prior. The horseshoe prior has high concentration around zero for sparsity, heavy tails to avoid excessive shrinkage of

signals and attractive theoretical properties (Datta and Ghosh, 2013; van der Pas et al., 2014; Bhadra et al., 2016a; van der Pas et al., 2017; Zhang et al., 2017).

To our knowledge, we propose the first continuous shrinkage prior for spatial data. We extend the horseshoe prior to account for spatial dependence in the signal at nearby observations. We prove that the proposed spatial horseshoe prior has the univariate horseshoe prior marginally at each location, and that the induced joint distribution for pairs of nearby sites has higher concentration around zero and heavier tails compared to independent horseshoe priors. The form of the continuous shrinkage prior permits simple expressions for the gradients of the posterior, and thus we use the Hamiltonian Monte Carlo (HMC) algorithm (Neal, 1994) for efficient sampling in high-dimensions. A simulation study demonstrates that the proposed method is effective at identifying spatial signals. When applied to the 2-D XRD data, we find improved cross-validation performance compared to other methods.

The paper is organized as follows. We describe the data in Section 2. In Sections 3 and 4, we introduce the proposed shrinkage prior and show its theoretical properties. In Section 5, we evaluate the proposed prior through the simulation study. In Section 6, we apply our method to the 2-D XRD data and compare with other methods. We conclude with remarks and comments in Section 7.

2 Description of the 2-D XRD data

The 2-D XRD data were acquired using high-energy X-rays generated at a synchrotron source (Advanced Photon Source, Argonne National Laboratory). The beamline 11-ID-C was used with an energy of 105 keV. The sample was located in the X-ray beam and a silicon-based detector was placed in the transmitted direction, approximately 2 meters from the sample position. Electrical connections were made from a high voltage power supply to the top conductive electrode of the PZT sample. The bottom electrode was grounded. 2-D XRD patterns were measured at sequentially increasing and then decreasing electric field amplitudes from -2 kV/mm to 2 kV/mm. The objective is to examine if introducing electric field changes the structure of the PZT sample and to capture the diffraction pattern. The detector records data over time which yields images that are 2048×2048 pixel matrices. The intensity of the XRD measured in counts per second (cps) indicates the relative abundance at each pixel. In addition, we transform the image of size 2048×2048 into size 1023×1023 by removing the first and last rows consisting of all zeros due to experimental setup and extracting all odd rows and columns. We surmise that this step is reasonable in that the intensity is close to the values nearby, and the diffraction pattern is still apparent. (See Supplementary Materials A for the data before and after this reduction.)

We compare the diffraction pattern of PZT with electric field to that without electric field and investigate the locations where changes occur. There are 100 baseline images recorded without electric field. We define the response $Y(s)$ as the standardized intensity,

$$Y(s) = \frac{X(s) - \hat{\mu}(s)}{\hat{\sigma}(s)},$$

where $X(s)$ is the intensity at pixel s , $\hat{\mu}(s)$ is sample mean of the 100

baseline images at pixel s , and $\hat{\sigma}(s)$ is sample standard deviation of the 100 baseline images at pixel s . The yellow rings in Figure 1 indicate the larger change of diffraction pattern at $t =$

20 seconds. We choose the image at time $t = 20$ seconds since this is when the maximum electric field is applied. The main features in Figure 1 are the rings centered on the middle of the image. We project the spatial shrinkage process into the polar coordinates similar to a Radon transform (Radon, 1986) to accommodate these annular features in Sections 3 and 6. The Radon transformation has been used successfully for image analysis problems related to motion detection (Carretero-Moya et al., 2009; Xu et al., 2011), deblurring (Cho et al., 2011) and classification (Kinoshita et al., 2008; Deepak and Sivaswamy, 2011; Acharya et al., 2016).

3 Model description

3.1 The univariate horseshoe prior

Define Y as the response variable and β as the signal variable. The likelihood for Y with horseshoe prior for β is

$$Y|\beta \sim N(\beta, \sigma^2), \beta|\lambda \sim N(0, \tau^2 \lambda^2), \lambda \sim C^+(0, 1), \tag{1}$$

where σ^2 and τ^2 are the error variances, and λ follows the standard half-Cauchy distribution on the positive reals. If $\sigma^2 = 1$, the posterior mean is

$$E(\beta|Y) = \int_0^1 (1 - \kappa) Y P(\kappa|Y) d\kappa = [1 - E(\kappa|Y)] Y, \tag{2}$$

where $\kappa = \frac{1}{1 + \lambda^2}$ is the shrinkage coefficient. $E(\kappa|Y)$ determines the amount of shrinkage

towards zero. The name of the prior comes from the horseshoe-shaped Beta $(\frac{1}{2}, \frac{1}{2})$ distribution on the shrinkage coefficient κ that is induced by the half-Cauchy distribution on λ . The shape of the shrinkage density implies that the shrinkage coefficient is close to either zero or one with high probability which shrinks null signals towards zero and avoids shrinking the true signals. This property facilitates separating signals from the noise. Although marginally over λ this prior for β does not have a closed form of the probability density function, it enjoys nice theoretical properties including tight bounds, high concentration near zero and heavy tails (Carvalho et al., 2010).

3.2 The spatial horseshoe prior (SHP)

To extend the horseshoe prior to the spatial setting, define $Y(s)$ as the real-valued response at a spatial location $s = (s_1, s_2) \in \mathbb{R}^2$ and $\beta(s)$ as the true signal at s . For data observed at locations s_1, \dots, s_n , define $\mathbf{Y} = [Y(s_1), \dots, Y(s_n)]^T$ as the $n \times 1$ response vector and $\boldsymbol{\beta} = [\beta(s_1), \dots, \beta(s_n)]^T$ as the signal vector. Extending (1), let

$$Y|\boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \sigma^2 I_n), \boldsymbol{\beta}|\boldsymbol{\lambda} \sim N(0, \boldsymbol{\Lambda} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \boldsymbol{\Lambda}), \tag{3}$$

where σ^2 is the error variance, I_n is the identity matrix of size n , $\boldsymbol{\Lambda}$ is diagonal with the diagonal elements $\boldsymbol{\lambda} = [\lambda(s_1), \dots, \lambda(s_n)]^T$, and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is a spatial covariance matrix.

Independence of the response \mathbf{Y} given the signals $\boldsymbol{\beta}$ is a strong assumption, but justified for

the 2-D XRD data where the distribution of photons should theoretically follow a Poisson process with independent counts across pixels. Physics dictates that the photons should follow a Poisson process so that the number of photons that land in disjoint regions are independent of each other. The photon detectors are identical and operate independently and so this assumption should carry through to the observed data.

Spatial shrinkage is induced by the spatial process model for $\lambda(\mathbf{s})$. We propose a Gaussian copula model (Nelsen, 2006) that preserves the marginal half-Cauchy distribution and captures spatial dependence,

$$\lambda(\mathbf{s}) = \lambda_0 I[\theta(\mathbf{s})], \tag{4}$$

where λ_0 is a global scale parameter, $\theta(\mathbf{s})$ is a latent spatial Gaussian process with mean zero and variance one for all \mathbf{s} , $I(\cdot) = F_{C^+}^{-1}[\Phi(\cdot)]$ is the half-Cauchy link function, $F_{C^+}^{-1}(\cdot)$ is the inverse cumulative density function of the half-Cauchy distribution, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. The half-Cauchy link function provides a marginal half-Cauchy distribution for $I[\theta(\mathbf{s})]$ if $E[\theta(\mathbf{s})] = 0$ and $\text{Var}[\theta(\mathbf{s})] = 1$. We interpret $I[\theta(\mathbf{s})]$ as a local shrinkage parameter and λ_0 as a global shrinkage parameter, respectively.

The latent shrinkage process $\theta(\mathbf{s})$ could follow another Gaussian process prior with mean zero, variance one and spatial correlation. We also consider a low-rank representation of the latent shrinkage process

$$\theta(\mathbf{s}) = \sum_{j=1}^J \frac{x_j(\mathbf{s})}{\sqrt{\sum_{j=1}^J x_j^2(\mathbf{s})}} b_j, \tag{5}$$

where $x_j(\mathbf{s})$ is the j^{th} basis function for location \mathbf{s} and $b_j \stackrel{iid}{\sim} N(0, 1)$. The spatial model gives the same marginal distribution as (1), but with spatial dependence both in the distribution for $\beta | \lambda$ and the distribution for λ . We will elaborate on the properties of this model in Section 4.

4 Theoretical properties

4.1 Properties of the spatial horseshoe prior

In this section we study the properties of the SHP for a pair of spatial locations as a function of the distance between the locations. We compare the SHP with the simple model with the independent β and λ , namely, $\Sigma_\beta = I_2$ and $\lambda_1, \lambda_2 \stackrel{iid}{\sim} C^+(0, 1)$. The joint density of β in this special case is (Carvalho et al., 2010)

$$P_0(\beta_1, \beta_2) = \frac{1}{2\pi^3} e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right), \tag{6}$$

where $E_1(\cdot)$ is the exponential integral function. Next, we consider the joint shrinkage parameter $\lambda = \lambda(\mathbf{s}_1) = \lambda(\mathbf{s}_2) \sim C^+(0, 1)$ and the covariance Σ_β with the diagonal elements 1

and the spatial correlation ρ_{β} which stands in for the distance between the two locations. Then we obtain (Supplementary Materials B.1) the joint density

$$P_1(\beta_1, \beta_2) = 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (1 - \rho_{\beta}^2)^{\frac{1}{2}} (\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2)^{-\frac{1}{2}} - (2\pi)^{-1} (1 - \rho_{\beta}^2)^{\frac{1}{2}} e^{\frac{1}{2}(\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2)} \operatorname{erfc}\left(\sqrt{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2}}\right), \tag{7}$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function.

Figure 2 plots $\log[P_0(\beta_1, \beta_2)]$ and $\log[P_1(\beta_1, \beta_2)]$ for $\rho_{\beta} = 0, 0.5, 0.9$. When $\rho_{\beta} = 0$, the log density with a common shrinkage parameter has higher spike around zero and heavier tail than the log density with independent shrinkage parameters. Increasing the ρ_{β} gives the higher concentration on the 45-degree line of the log density.

4.1.1 Concentration around zero—The densities $P_0(\beta_1, \beta_2)$ and $P_1(\beta_1, \beta_2)$ both have high concentration towards zero. It can be shown that $\lim_{\beta_1, \beta_2 \rightarrow 0} P_0(\beta_1, \beta_2) = \infty$ and

$\lim_{\beta_1, \beta_2 \rightarrow 0} P_1(\beta_1, \beta_2) = \infty$. Moreover, we show that the density with joint shrinkage

concentrates towards zero more than the density with independent shrinkage by proving that the ratio of $P_1(\beta_1, \beta_2)$ to $P_0(\beta_1, \beta_2)$ diverges as β_1 and β_2 approach zero. Let R be the ratio,

$$R = \frac{P_1(\beta_1, \beta_2)}{P_0(\beta_1, \beta_2)} = \pi^{\frac{3}{2}} e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2)} \left(\frac{\beta_1^2 + \beta_2^2}{2}\right)^{-\frac{1}{2}} \left[E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right)\right]^{-1} - \pi^2 \frac{\operatorname{erfc}\left(\sqrt{\frac{\beta_1^2 + \beta_2^2}{2}}\right)}{E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right)}. \tag{8}$$

We show that $\lim_{\beta_1, \beta_2 \rightarrow 0} R = \infty$ with the divergence rate $(\sqrt{\beta_1^2 + \beta_2^2} \log \beta_1 \log \beta_2)^{-1}$

(Supplementary Materials B.2). This indicates that joint sparsity is achieved more rapidly in $P_1(\beta_1, \beta_2)$ than in $P_0(\beta_1, \beta_2)$.

4.1.2 Tail behavior—We also investigate the capability to retain signals of large values through examining the tail densities with joint and independent shrinkage. In Figure 2, we observe that the tail density is lower in $P_0(\beta_1, \beta_2)$ than $P_1(\beta_1, \beta_2)$ and becomes larger when the spatial dependence is stronger. The higher tail density implies a better capability to avoid excessive shrinkage. Theoretically, it is shown that $\lim_{\beta_1, \beta_2 \rightarrow \infty} P_0(\beta_1, \beta_2) = \infty$ with the

divergence rate $\frac{\beta_1^2 \beta_2^2}{\sqrt{\beta_1^2 + \beta_2^2}}$ (Supplementary Materials B.3). This result suggests the higher tail density with joint shrinkage and o its stronger power to incorporate large signals.

4.2 Properties of the spatial horseshoe posterior

Let the error variance σ^2 be 1, then the posterior mean can be expressed as $E(\beta | Y, \lambda) = (\mathbf{I}_2 - \mathbf{K})\mathbf{Y}$, where the shrinkage matrix $K = I_2 - (I_2 + \Omega_\beta^{-1})^{-1}$ and $\Omega_\beta = \Lambda \Sigma_\beta \Lambda$. To examine the joint shrinkage induced by the SHP, we consider the one-number complexity summary (Efron, 2004)

$$df = tr(I - K) = \frac{2(1 - \rho_\beta^2)\lambda_1^2\lambda_2^2 + \lambda_1^2 + \lambda_2^2}{(1 - \rho_\beta^2)\lambda_1^2\lambda_2^2 + \lambda_1^2 + \lambda_2^2 + 1}. \tag{9}$$

If $\lambda_1 = \lambda_2 = \lambda$, then

$$df = \frac{2(1 - \rho_\beta^2)\lambda^4 + 2\lambda^2}{(1 - \rho_\beta^2)\lambda^4 + 2\lambda^2 + 1}. \tag{10}$$

In this special case with $\rho_\beta = 0$, $df = \frac{2\lambda^2}{\lambda^2 + 1}$, and thus its prior distribution is under the SHP resembles the horseshoe $Beta\left(\frac{1}{2}, \frac{1}{2}\right)$ density scaled by 2. Figure 3 plots the density of df . The SHP with common shrinkage has spikes at zero and two corresponding to joint shrinkage. This property encourages noises to be shrunk to zero and signals to be retained at the original values. The density of df with independent λ has a spike around one which corresponds to shrinking only one of β_1 and β_2 , instead of shrinkage or non-shrinkage for both variables. This characteristic even holds when $\rho_\beta = 0.9$ which is contradictory to the expectation that a high correlation between two variables leads to similar shrinkage behaviors.

5 Simulation Study

We investigate how the SHP performs in detecting sparse signals compared with other methods using a simulation study.

5.1 Data generation

We consider the two true surfaces:

$$\begin{aligned}
 &1. \beta_0(s) = 6I[(s_1 - 10)^2 + (s_2 - 12)^2 \leq 1] \\
 &+ 5I[(s_1 - 10)^2 + (s_2 - 12)^2 > 1 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 4] \\
 &+ 4I[(s_1 - 10)^2 + (s_2 - 12)^2 > 4 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 16] \\
 &+ 3I[(s_1 - 10)^2 + (s_2 - 12)^2 > 16 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 36]
 \end{aligned}$$

and

$$\begin{aligned}
 &2. \beta_0(s) = 6I[(s_1 - 10)^2 + (s_2 - 12)^2 \leq 1] \\
 &+ 5I[(s_1 - 10)^2 + (s_2 - 12)^2 > 1 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 4] \\
 &+ 4I[(s_1 - 10)^2 + (s_2 - 12)^2 > 4 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 16] \\
 &+ 3I[(s_1 - 10)^2 + (s_2 - 12)^2 > 16 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 36] \\
 &+ 4I[(s_1 - 28)^2 + (s_2 - 30)^2 \leq 4] \\
 &+ 3I[(s_1 - 28)^2 + (s_2 - 30)^2 > 4 \text{ and } (s_1 - 28)^2 + (s_2 - 30)^2 \leq 16].
 \end{aligned}$$

The observations are generated with $Y(s) | \beta_0(s) \stackrel{indep}{\sim} N(\beta_0(s), \sigma^2)$ on a 40×40 grid of $n = 1,600$ locations. The proportion of signals with nonzero $\beta_0(s)$ is 7.06% and 10.13% in the two surfaces. In addition, we vary the error variance $\sigma^2 = 0.5^2, 1^2, 2^2$. For each combination of true surface and σ^2 , we generate $N = 100$ data sets. Figure 4 shows the two surfaces and a representative simulated data set for each surface.

5.2 Models

Although the SHP can accommodate other covariance structures, we use the conditionally autoregressive prior (CAR, Carlin and Banerjee, 2003) for Σ_{β} . The CAR covariance is $\Sigma_{\beta} = (M - \rho_{\beta}A)^{-1}$, where M is the diagonal matrix with the elements m_1, \dots, m_n indicating the number of neighbors for locations s_1, \dots, s_n , ρ_{β} is the spatial dependence parameter, and A is the adjacency matrix with $A_{ij} = 1$ if locations s_i and s_j are neighbors and $A_{ij} = 0$ otherwise. Note that while $\rho_{\beta} \in (0, 1)$ determines the strength of spatial dependence, it is not a correlation parameter as in Section 4. The CAR prior is a natural choice for data on a discrete grid as discussed in Cressie (2013) and Gelfand (2010). This is an intuitive prior specification and has the advantage of a sparse inverse covariance matrix which leads to a dramatic computational savings. For our analysis of over a million pixels, a prior that did not give a sparse inverse covariance would require approximation and/or prohibitively long run times. Computational details are given in Supplementary Materials C.

We fit five models for $\beta(s)$ to each simulated dataset. The first *four* are versions of the spatial horseshoe differentiated by their flexibility in modeling the shrinkage process $\lambda(s)$. The fifth method is soft-thresholded Gaussian process (STGP) model of Kang et al. (2018) that places prior mass exactly at zero.

Gaussian: $\lambda(s)$ as a constant across space, $\lambda(s) = \lambda_0$ which leads to

$$\beta | \lambda_0 \sim N\left(0, \lambda_0^2 (M - \rho_{\beta}A)^{-1}\right)$$

The error variance σ^2 and the scale parameter λ_0^2 follow

uninformative inverse gamma prior $IG(0.1, 0.1)$, and the spatial dependence parameter $\rho\beta$ follows a beta prior $Beta(10,1)$. This is the usual Gaussian CAR model.

SHS_quad: Low-rank representation using the half-Cauchy link function with quadratic basis expansions, $\lambda(s_i) = \lambda_0 l[\theta(s_i)]$, where $l[\cdot]$ is the half-Cauchy link function,

$$\theta(s) = \sum_{j=1}^J \frac{x_j(s)}{\sqrt{\sum_{j=1}^J x_j^2(s)}} b_j, \text{ priors } b_j \stackrel{iid}{\sim} N(0, 1),$$

$$x_1(s_i) = 1, x_2(s_i) = s_{i1}, x_3(s_i) = s_{i2}, x_4(s_i) = s_{i1}^2, x_5(s_i) = s_{i2}^2 \text{ and } x_6(s_i) = \mathbf{s}_{i1}\mathbf{s}_{i2} \text{ for } i = 1, \dots, n.$$

SHS_B-spline 1: Low-rank representation using the half-Cauchy link function with the B-spline basis expansions with five degrees of freedom for each coordinate, $\lambda(\mathbf{s}_i) = \lambda_0 l[\theta(\mathbf{s}_i)]$, where $l[\cdot]$ is the half-Cauchy link function, $\theta(\mathbf{s}_i) = \sum_{j=1}^{100} x_j(\mathbf{s}_i) b_j$, priors $b_j \stackrel{iid}{\sim} N(0, 1)$ and $x_j(\mathbf{s}_i)$ is the j^{th} product of B-spline bases at location \mathbf{s}_i .

SHS_B-spline 2: Same structure as the model SHS_B-spline 1, but with ten degrees of freedom for each coordinate.

STGP: Soft-thresholded Gaussian process prior for $\beta(\mathbf{s})$. Let $\tilde{\beta}(s)$ follow a multivariate normal distribution with zero mean and CAR covariance with first-order neighborhood. Then $\beta(s) = g_\kappa[\tilde{\beta}(s)]$, where g_κ is the soft-thresholding function to map $\tilde{\beta}(s)$ near zero to exact zero and thus gives a sparse prior,

$$g_\kappa(x) = \begin{cases} 0 & \text{if } |x| \leq \kappa \\ \text{sign}(x)(|x| - \kappa) & \text{if } |x| > \kappa \end{cases}$$

The thresholding parameter κ follows a uniform prior $U(0,10)$ and controls the degree of sparsity.

5.3 Evaluation metrics

We compare methods in terms of root mean squared error (RMSE), coverage probability, Type I error and power. For simulated data set k , let $\hat{\beta}_k(s)$ be the posterior mean and $[l_k(s), u_k(s)]$ be the 95% credible interval of $\beta(s)$. We list the formulae of the measures as follows:

$$RMSE = \sqrt{\frac{1}{Nn} \sum_{k=1}^N \sum_{i=1}^n [\hat{\beta}_k(s_i) - \beta_0(s_i)]^2}$$

$$\text{Coverage} = \frac{1}{Nn} \sum_{k=1}^N \sum_{i=1}^n I\{\beta_0(s_i) \in [l_k(s_i), u_k(s_i)]\}$$

$$\text{Type I error} = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^n I\{P[\hat{\beta}_k(s_i) > 0] > 0.95\} \cdot I[\beta_0(s_i) = 0]}{\sum_{i=1}^n I[\beta_0(s_i) = 0]}$$

$$\text{Power} = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^n I\{P[\hat{\beta}_k(s_i) > 0] > 0.95\} \cdot I[\beta_0(s_i) \neq 0]}{\sum_{i=1}^n I[\beta_0(s_i) \neq 0]}.$$

5.4 Results

The results are given in Tables 1 and 2. RMSE decreases as the number of basis functions in the spatial shrinkage process $\lambda(\mathbf{s})$ increases since a larger number of basis functions enhances the flexibility in $\lambda(\mathbf{s})$. RMSE in the SHS_B-spline 2 model drops 34.06% compared with RMSE in the Gaussian model under the low-noise case in Signal 2. SHS_B-spline models have the smallest RMSE in the low-noise and mid-noise cases while STGP outperforms other models in RMSE under the other scenario. Coverage and type I error are at or near the nominal level for all methods except for the Gaussian model in the low-noise cases. All models have strong power when the error variance is small. However, large error variance distinguishes the models. STGP model has about 66% the power of SHS models. The loss of power in STGP model results from excessive shrinkage from the soft-thresholding formulation, especially in the large-noise scenario.

Figures 5 and 6 illustrate the simulation results under Signal 2 with $\sigma = 0.5$ and 2, respectively. Increasing the flexibility of the shrinkage process improves signal identification and leads to a smoother signal surface. The Gaussian model tends to have more false positives, especially when the error variance is small. This explains why it has strong power but poor RMSE. The simulation plots under other scenarios including data generated with spatially-correlated errors are shown in Supplementary Materials D. For the computing time, analyzing a simulated data set takes 0.33, 0.87, 1.23, 3.40 and 10.29 minutes for Gaussian, SHS_quad, SHS_B-spline 1, SHS_B-spline 2 and STGP models, respectively.

6 Analysis of the 2-D XRD data

6.1 Model comparisons

In this section we apply the proposed model in Section 3 to the 2-D XRD data. We tailor the SHP model to the 2-D XRD data to capture the ring-shaped pattern visible in Figure 1. Rather than using the basis expansions of square coordinates in the simulation study, we consider a basis expansion of the radius, $r \geq 0$, and the angle, $a \in [0, 2\pi)$, from the central point. In the simulation study the true features were defined on disks and thus we used radial basis functions for the smoothing process, $\lambda(\mathbf{s})$. As is apparent in Figure 1, the features in the XRD data are annuli, and we thus choose basis functions define in polar coordinates to capture the shape of these features. We use Fourier and B-spline basis expansions for angle and radius, respectively. $A_1(a), \dots, A_{k_a}(a)$ be the Fourier basis functions and $B_1(r), \dots, B_{k_r}(r)$ be the B-spline basis functions, where k_a and k_r are the numbers of basis functions in the angle

and the radius, respectively. The Fourier basis functions are $A_1(a) = \sin(a), A_2(a) = \cos(a), A_3(a) = \sin(2a)$, and $A_4(a) = \cos(2a)$, etc. At pixel s , the basis functions consist of $J = k_a \times k_r$ products of Fourier and B-spline basis expansions, i.e.

$$x(s) = [x_1(s), \dots, x_J(s)]^T = \left\{ A_1[a(s)] \times B_1[r(s)], \dots, A_{k_a}[a(s)] \times B_{k_r}[r(s)] \right\}^T.$$

The priors and computing details remain the same as in Section 5.

We implement five-fold cross-validation to evaluate the prediction in models with varying flexibility of shrinkage across space. We consider $(k_a, k_r) = (4,10), (4, 25), (4,50), (4,100), (8, 5)$ and $(8, 50)$. We do not include STGP in the model comparisons because of the heavy computation. Table 3 presents the results as RMSE-1, since for the z-scores with independent standard normal error, one is the lowest achievable prediction RMSE. Generally, predicted RMSE declines as flexibility of shrinkage grows, similar to the simulation results shown in Section 5. However, a larger number of basis functions of angle tends to have a smaller prediction error when the total number of basis expansions is the same, e.g. SHS1 versus SHS5 and SHS3 versus SHS6. According to Wilcoxon signed rank test, RMSE of the models SHS1-SHS6 is significantly less than RMSE of the Gaussian model. Coverage is close to the nominal level 95% for all models.

6.2 Summary of fitted models

In this section we fit the Gaussian and SHS models as above to the full dataset. Because the number of pixels is large, we implement the Bayesian spatial false discovery rate (BSFDR) procedure (Sun et al., 2015) with rate 0.05 to control for multiple testing. We consider the one-sided null and alternative hypotheses $H_0 : \beta(s) \leq 0$ and $H_1 : \beta(s) > 0$, respectively. The BSFDR procedure provides a critical probability that the null hypothesis is rejected if the posterior probability is greater than the critical probability. The critical probabilities are 92.93% for all models, except for 91.92% for the SHS4 model. The number of pixels for which H_{0s} is rejected are also similar for the Gaussian and SHS models, from 1.42% to 2.02%.

Figure 7 displays the fitted results for the Gaussian and SHS6 (with the lowest RMSE) models. The posterior mean of the shrinkage parameter $\lambda(s)$ (Figure 7, bottom left) is large for only a few radii and for only a subset of the angles on the rings. These areas also have the largest posterior mean of $\beta(s)$ and posterior probability that $\beta(s)$ is positive. The map of difference in posterior probability (Figure 7, top right) exhibits that the model SHS6 encourages shrinkage along the rings. Comparing models, we find more significant pixels and stronger spatial clustering of the signal in the SHS model compared to the Gaussian model. In addition, Figure 8 illustrates that the density of posterior mean $\beta(s)$ from the SHP model (SHS6) has higher concentration around zero and heavier tails than the density of $\beta(s)$ from the Gaussian model.

Figure 7 is an analysis of one time point ($t = 20$). We investigate the temporal trend of signals by fitting the model SHS6 for the standardized intensity $Y(s)$ at all time points. We use the sum of squared signals, $D = \sum_{i=1}^n [\beta(s_i)]^2$, as a measure of strength of structural change of the PZT sample. Figure 9 displays the boxplots of the posterior of D and the

electric field by time. Generally, the structural change becomes larger when the magnitude of electric field increases. The measure D is smaller in the first half of experimental period under a positive charge compared to the second half of period under a negative charge. The reason for the rapid increase in D between 54 and 58 seconds is a material phenomenon known as ferroelectric switching; the electric field amplitude at this time exceeds that required for substantial reorientation of material elements, leading to a significant change in the diffraction pattern. The example of this method on ferroelectrics demonstrates the versatility of the method at efficiently assessing 2-D XRD images for the purposes of materials characterization.

7 Conclusion

In this paper, we propose a new method for sparse signal deflection with application to the 2-D XRD imaging data. To our knowledge, the SHP is the first continuous shrinkage prior for spatial data. Theoretically, we prove that the SHP has the univariate horseshoe prior marginally at each location, and that the induced joint distribution for pairs of nearby sites has higher concentration around zero and heavier tails compared to independent horseshoe priors. Further, we facilitate the computation via the HMC algorithm (Neal, 1994) for high-dimensional data. The simulation and empirical results both show the improvement of estimation and prediction when we consider spatial dependence.

A potential limitation of our method is that the covariance structure of data is assumed to be independent given the mean vector β . This is sufficient for the 2-D XRD imaging data because the distribution of photons ought to follow a Poisson process with independent counts across pixels, but it may not hold for different data. Our simulation results (Supplemental Materials D) show that for data with moderate spatial correlation our working independence model maintained reasonable error rates but had lower power than our previous simulation with independent errors. For simulated data with strong residual spatial correlation, our model with working independent assumption did not work well. The shortcomings of the working independence model led us to consider a richer model that includes spatial correlation in the residuals. A second restriction is the assumption of CAR covariance structure with the first order neighbors. The covariance structure can be modified to meet needs in other situations. For instance, the Matérn covariance (Stein, 1999) is a covariance function depending on the distance between two given locations, a smooth parameter and a range parameter. The general class of covariance functions includes the exponential covariance and the squared exponential covariance. The flexibility may be preferred and computationally feasible for smaller datasets.

Deep learning via convolutional neural networks (CNN) have emerged as immensely powerful tools for extracting information from images (Goodfellow et al., 2016; Rawat and Wang, 2017; Yamashita et al., 2018). The standard CNN is a supervised learning algorithm that uses an image predictor. However, in our setting we do not have known labels (e.g., images with locations that are known to have a change) to train the model. Also, while CNNs are strong for prediction, they are weak for formal statistical inference which is our primary objective.

In the future, the proposed method can be extended to Bayesian variable selection in spatial linear regression. In spatial regression, the covariates effect can vary by site (Gelfand et al., 2003), and the SHP for the spatially-varying effects would encourage sparsity in these effects. We can also extend the SHP to a spatio-temporal horseshoe model for longitudinal data. We can use the extended model to test for spatio-temporal anomalies in a surveillance study (e.g. Li et al., 2012) or process monitoring (e.g. Yan et al., 2018).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/oe-public-access-plan>). The authors thank the editor, associate editor, and two anonymous referees for insightful and constructive comments on an earlier version of the manuscript.

References

- Acharya UR, Fujita H, Sudarshan VK, Mookiah MRK, Koh JEW, Tan JH, Hagiwara Y, Chua CK, Junnarkar SP, Vijayanathan A, and Ng KH (2016). An integrated index for identification of fatty liver disease using radon transform and discrete cosine transform features in ultrasound images. *Information Fusion*, 31:43–53.
- Armagan A, Dunson DB, and Lee J (2013). Generalized double Pareto shrinkage. *Statistica Sinica*, 23(1):119–143. [PubMed: 24478567]
- Bhadra A, Datta J, Li Y, Polson NG, and Willard B (2016a). Prediction risk for the horseshoe regression. *ArXiv e-prints*.
- Bhadra A, Datta J, Polson NG, and Willard B (2016b). Default Bayesian analysis with global-local shrinkage priors. *Biometrika*, 103(4):955–969.
- Bhattacharya A, Pati D, Pillai NS, and Dunson DB (2014). Dirichlet-laplace priors for optimal shrinkage. *ArXiv e-prints*.
- Boehm Vock LF, Reich BJ, Fuentes M, and Dominici F (2015). Spatial variable selection methods for investigating acute health effects of fine particulate matter components. *Biometrics*, 71(1):167–177. [PubMed: 25303336]
- Bradley D and Roth G (2007). Adaptive thresholding using the integral image. *Journal of Graphics Tools*, 12(2):13–21.
- Carlin PB and Banerjee S (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. *Bayesian Statistics*, 7:45–63.
- Carretero-Moya J, Gismero-Menoyo J, Asensio-López A, and Blanco-del Campo A (2009). Application of the radon transform to detect small-targets in sea clutter. *IET Radar, Sonar & Navigation*, 3(2):155–166.
- Carvalho CM, Polson NG, and Scott JG (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.
- Cho TS, Paris S, Horn BKP, and Freeman WT (2011). Blur kernel estimation using the Radon transform. In *CVPR 2011*, pages 241–248.
- Costa MA and Kulldorff M (2009). Applications of spatial scan statistics: a review In Glaz J, Pozdnyakov V, and Wallenstein S, editors, *Scan Statistics: Methods and Applications*, pages 129–152. Birkhäuser Boston, Boston, MA.

- Cressie NAC (2013). *Statistics for Spatial Data*. Wiley, New York, NY.
- Datta J and Ghosh JK (2013). Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Analysis*, 8(1):111–132.
- Deepak KS and Sivaswamy J (2011). Automatic assessment of macular edema from color retinal images. *IEEE Transactions on Medical Imaging*, 31(3):766–776. [PubMed: 22167598]
- Donoho DL and Johnstone JM (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.
- Dutta I and Singh RN (2011). Dynamic in situ x-ray diffraction study of antiferroelectric-ferroelectric phase transition in strontium-modified lead zirconate titanate ceramics. *Integrated Ferroelectrics*, 131(1):153–172.
- Editorial (2014). Crystallography matters. *Nature Materials*, 13.
- Efron B (2004). The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association*, 99(467):619–632.
- Esteves G, Fancher CM, and Jones JL (2015). In situ characterization of polycrystalline ferroelectrics using x-ray and neutron diffraction. *Journal of Materials Research*, 30(3):340–356.
- Friedman J, Hastie T, Höfling H, and Tibshirani R (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332.
- Friedman J, Hastie T, and Tibshirani R (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441. [PubMed: 18079126]
- Gelfand AE, Kim H-J, Sirmans CF, and Banerjee S (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.
- Gelfand AE, D. P. G. P. F. M. (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL.
- George EI and McCulloch RE (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George EI and McCulloch RE (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7(2):339–374.
- Geweke J (1996). Variable selection and model comparison in regression. *Bayesian Statistics*, 5:609–620.
- Goldsmith J, Huang L, and Crainiceanu CM (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *Journal of Computational and Graphical Statistics*, 23(1):46–64. [PubMed: 24729670]
- Goodfellow I, Bengio Y, and Courville A (2016). *Deep Learning*. MIT Press <http://www.deeplearningbook.org>.
- Gorfman S, Keeble DS, Glazer AM, Long X, Xie Y, Ye Z-G, Collins S, and Thomas PA (2011). High-resolution x-ray diffraction study of single crystals of lead zirconate titanate. *Physical Review B*, 84:020102.
- Griffin JE and Brown PJ (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.
- He L and Wang Y (2017). Wavelet frame based image restoration using sparsity, nonlocal and support prior of frame coefficients. *The Visual Computer*.
- Jansen M (2001). *Noise Reduction by Wavelet Thresholding*. Springer, New York, NY.
- Johnson VE and Rossell D (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107(498):649–660.
- Kang J, Reich BJ, and Staicu A-M (2018). Scalar-on-image regression via the soft- thresholded Gaussian process. *Biometrika*, 105(1):165–184. [PubMed: 30686828]
- Katrusiak A (2008). High-pressure crystallography. *Acta Crystallographica Section A*, 64(1):135–148.
- Kinoshita SK, Azevedo-Marques PM, Pereira RR Jr, Rodrigues AH, and Ranganathan RM (2008). Radon-domain detection of the nipple and the pectoral muscle in mammograms. *Journal of Digital Imaging*, 21(1):37–49. [PubMed: 17436047]
- Kulldorff M (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, 26(6):1481–1496.

- Li F, Zhang T, Wang Q, Gonzalez MZ, Maresh EL, and Coan JA (2015). Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *The Annals of Applied Statistics*, 9(2):687–713.
- Li G, Best N, Hansell AL, Ahmed I, and Richardson S (2012). BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics*, 13(4):695–710. [PubMed: 22452805]
- Liao Y, Li X, and Wang J (2017). The study on modified spatial scan statistic In Yang W, editor, *Early Warning for Infectious Disease Outbreak*, pages 329–342. Academic Press, Cambridge, MA.
- Mallick H and Yi N (2013). Bayesian methods for high dimensional linear models. *Journal of Biometrics and Biostatistics*,
- Mitchell TJ and Beauchamp JJ (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.
- Naus JI (1965). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*, 60(310):532–538.
- Neal RM (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, 111(1):194–203.
- Nelsen RB (2006). *An Introduction to Copulas*. Springer, New York, NY.
- O’hara RB and Sillapää MJ (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1):85–118.
- Piironen J and Vehtari A (2016). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. *ArXiv e-prints*.
- Qiu P and Yandell B (1997). Jump detection in regression surfaces. *Journal of Computational and Graphical Statistics*, 6(3):332–354.
- Radon J (1986). On the determination of functions from their integral values along certain manifolds (Parks PC Trans.). *IEEE-Transactions on Medical Imaging*, MI-5(4):170–176.
- Rawat W and Wang Z (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9):2352–2449. [PubMed: 28599112]
- Ro ková V and George EI (2018). The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444.
- Sollie P (2013). *Morphological Image Analysis: Principles and Applications*. Springer Science & Business Media, Berlin, Germany.
- Stein ML (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, NY.
- Sun W, Reich BJ, Tony Cai T, Guindani M, and Schwartzman A (2015). False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):59–83. [PubMed: 25642138]
- Taswell C (2000). The what, how, and why of wavelet shrinkage denoising. *Computing in Science Engineering*, 2(3):12–19.
- Thomas JM (2012). The birth of x-ray crystallography. *Nature*, 491:186–187. [PubMed: 23135450]
- van der Pas S, Szabó B, and van der Vaart A (2017). Uncertainty quantification for the horseshoe. *Bayesian Analysis*, 12(4):1221–1274.
- van der Pas SL, Kleijn BJK, and van der Vaart AW (2014). The horseshoe estimator: posterior concentration around nearly black vectors. *Electronic Journal of Statistics*, 8(2):2585–2618.
- Xu J, Yu J, Peng Y, and Xia X (2011). Radon-Fourier transform for radar target detection, I: Generalized Doppler filter bank. *IEEE Transactions on Aerospace and Electronic Systems*, 47(2): 1186–1202.
- Yadav M, Yadav S, and Sharma D (2014). Image denoising using orthonormal wavelet transform with stein unbiased risk estimator. In *Electrical, Electronics and Computer Science (SCEECS), 2014 IEEE Students’ Conference*, pages 1–4, Bhopal, India.
- Yamashita R, Nishio M, Do RKG, and Togashi K (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4):611–629. [PubMed: 29934920]
- Yan H, Paynabar K, and Shi J (2017). Anomaly detection in images with smooth background via smooth-sparse decomposition. *Technometrics*, 59(1):102–114.

- Yan H, Paynabar K, and Shi J (2018). Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics*, 60(2):181–197.
- Yuan M and Lin Y (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100(472):1215–1225.
- Zhang Y, Reich BJ, and Bondell H (2017). High dimensional linear regression via the R2-D2 shrinkage prior. ArXiv e-prints.

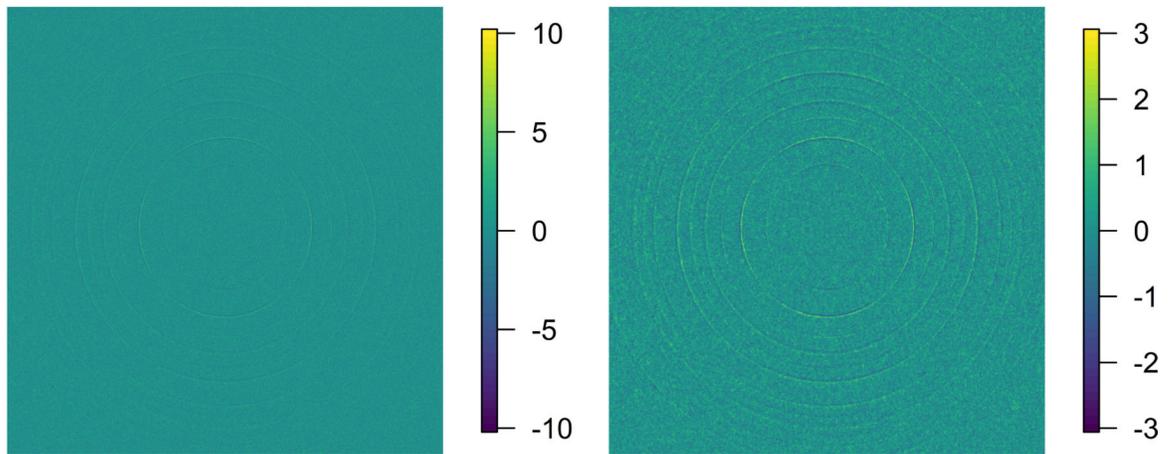


Figure 1:

The map of the standardized intensity $Y(s)$ (left) with the maximum electric field 2 kV/mm at time $t = 20$ seconds using the sample mean $\hat{\mu}(s)$ and the sample standard deviation $\hat{\sigma}(s)$ of the 100 baseline images without electric field and the smoothed $Y(s)$ (right) using a Gaussian kernel with the bandwidth 0.75 pixels.

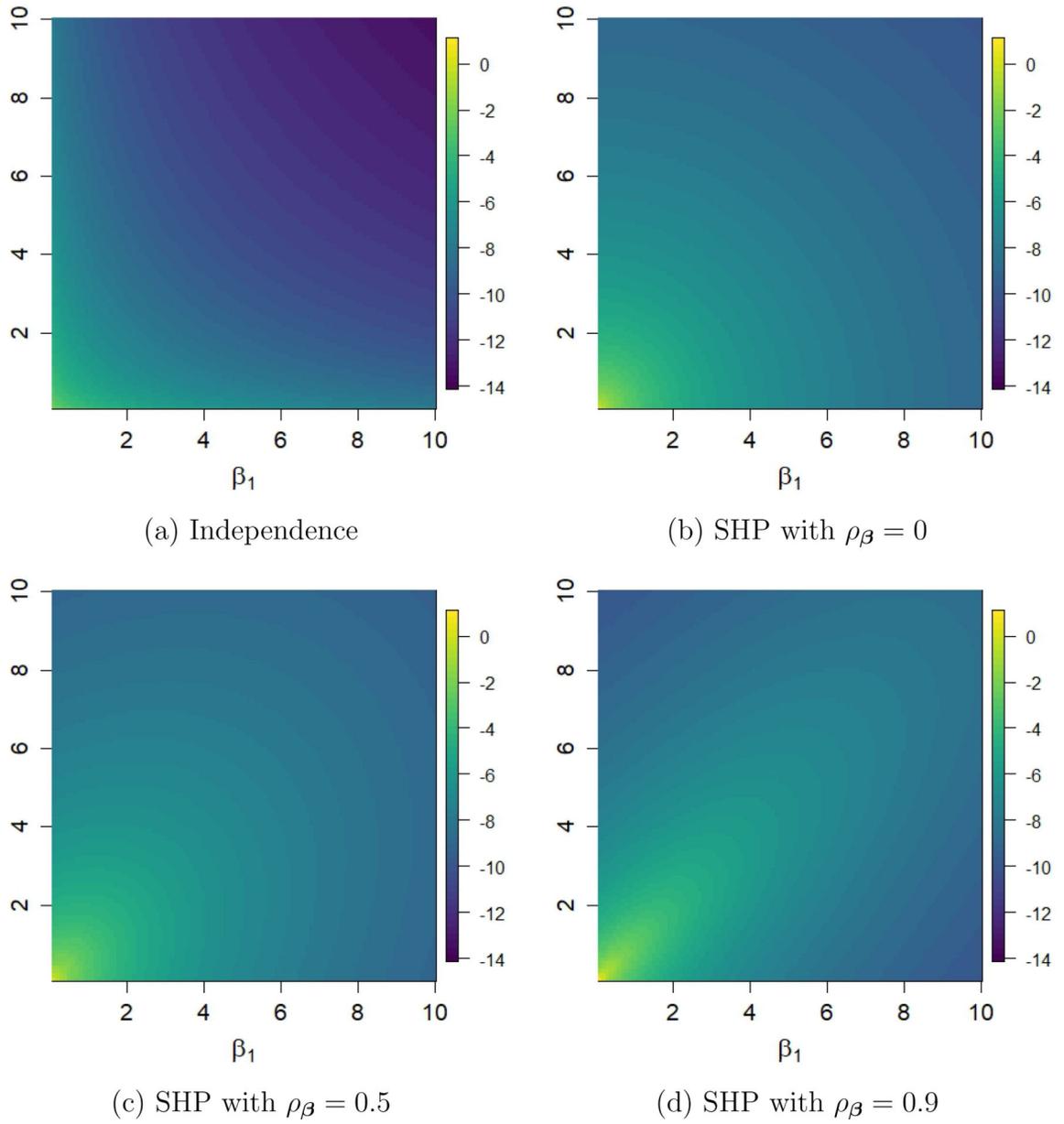


Figure 2: Log prior density of β_1 and β_2 in the first quadrant for (a) independent horseshoe priors ($\log P_0$) and (b)-(d) the spatial horseshoe prior ($\log P_1$) with $\rho_{\beta} \in \{0, 0.5, 0.9\}$.

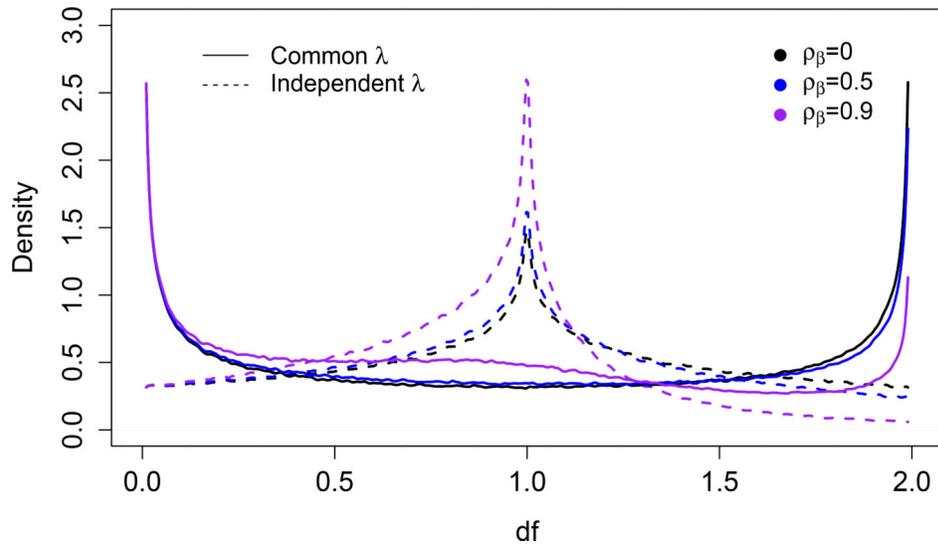


Figure 3: Induced prior density plot of the complexity measure $df = \text{tr}(\mathbf{I} - \mathbf{K})$ with independent and common shrinkage with $\rho_\beta \in \{0, 0.5, 0.9\}$.

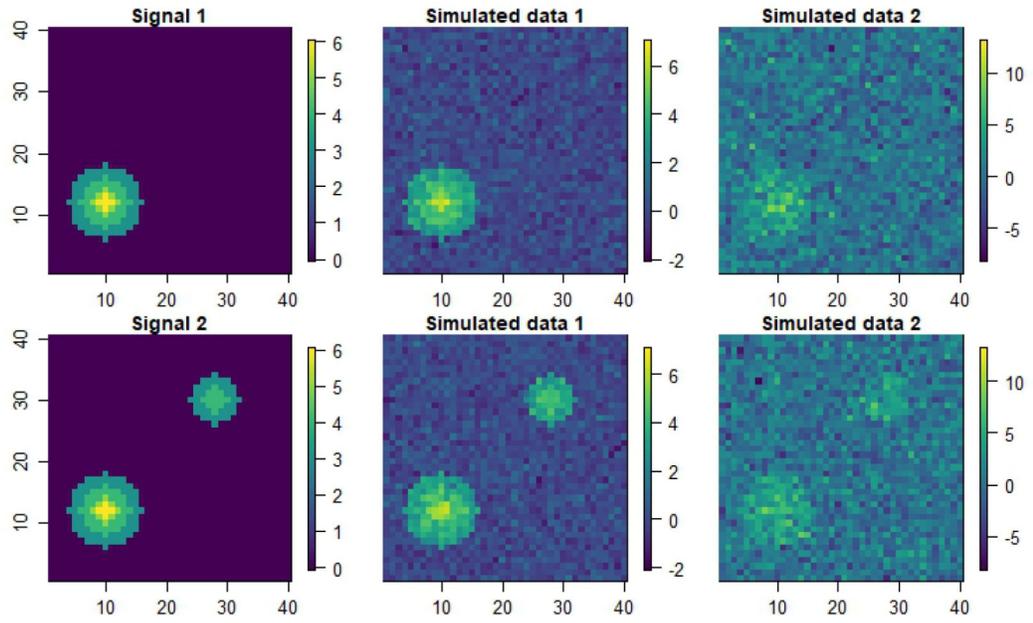


Figure 4: True surface $\beta_0(\mathbf{s})$ (left) and a simulated data set $Y(\mathbf{s})$ with error standard deviation $\sigma = 0.5$ (middle) and $\sigma = 2$ (right).

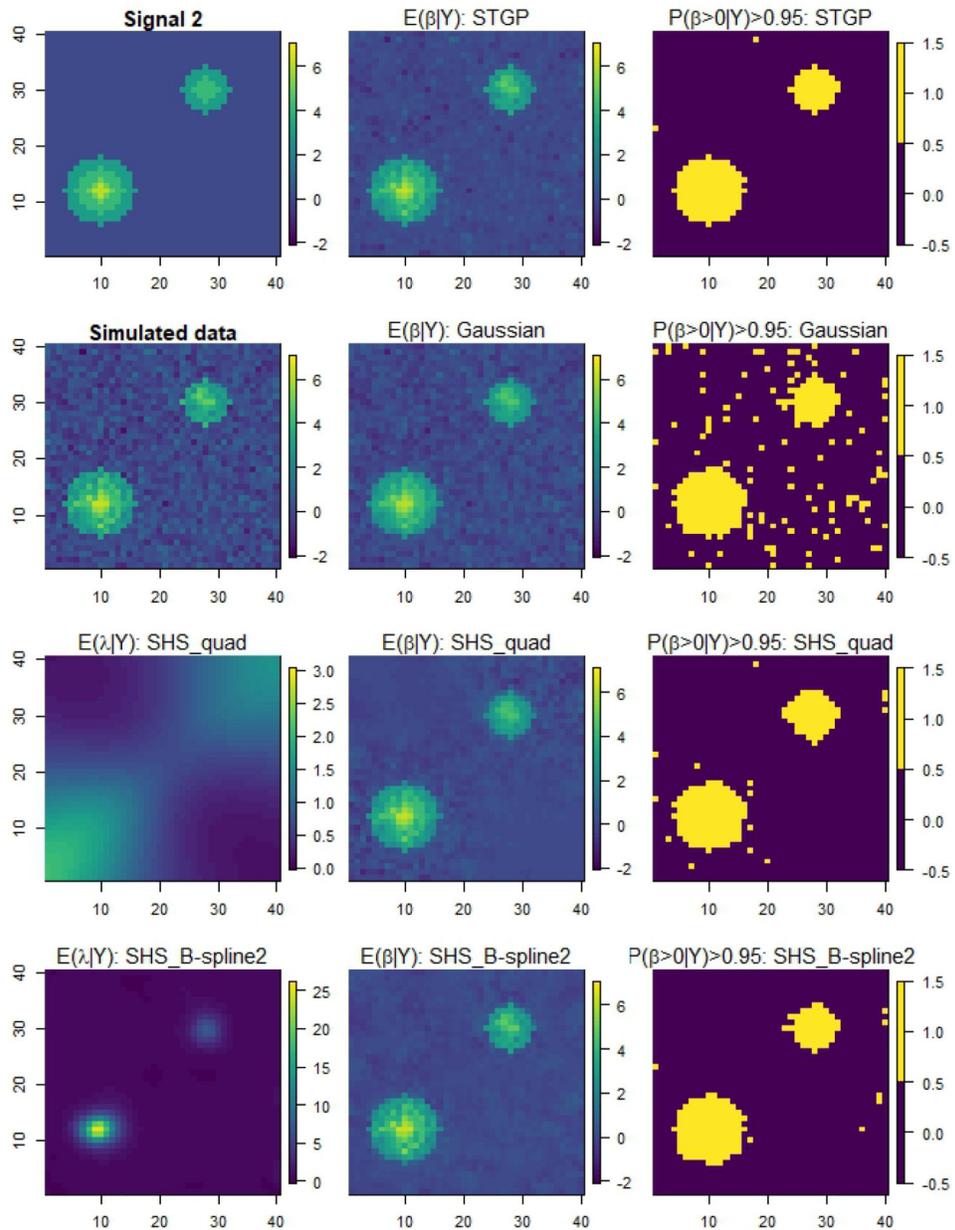


Figure 5: Simulation results for Signal 2 with $\sigma = 0.5$ for the four models (rows): the first row shows the true signal $\beta_0(s)$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(s)$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

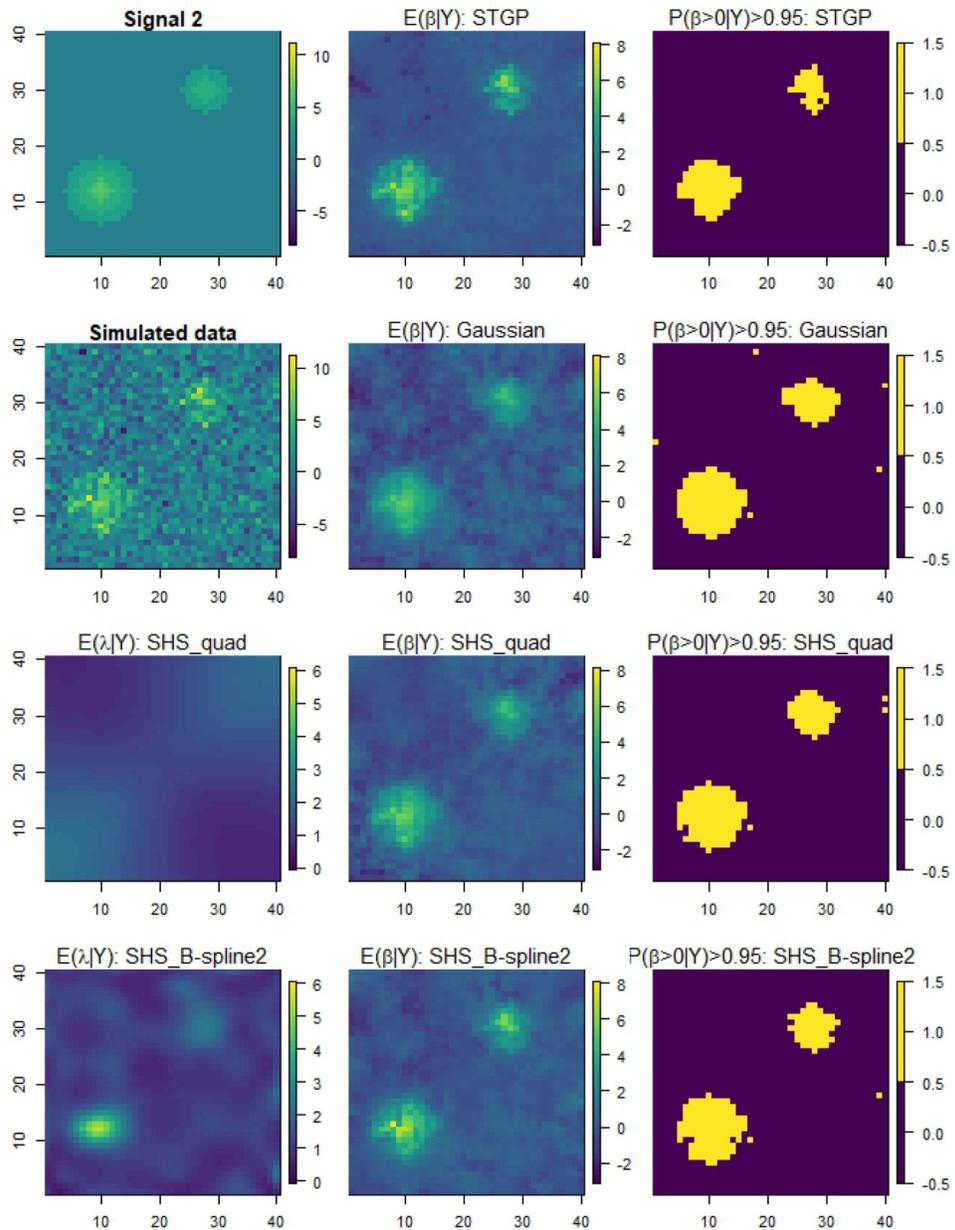


Figure 6: Simulation results for Signal 2 with $\sigma = 2$ for the four models (rows): the first row shows the true signal $\beta_0(s)$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(s)$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

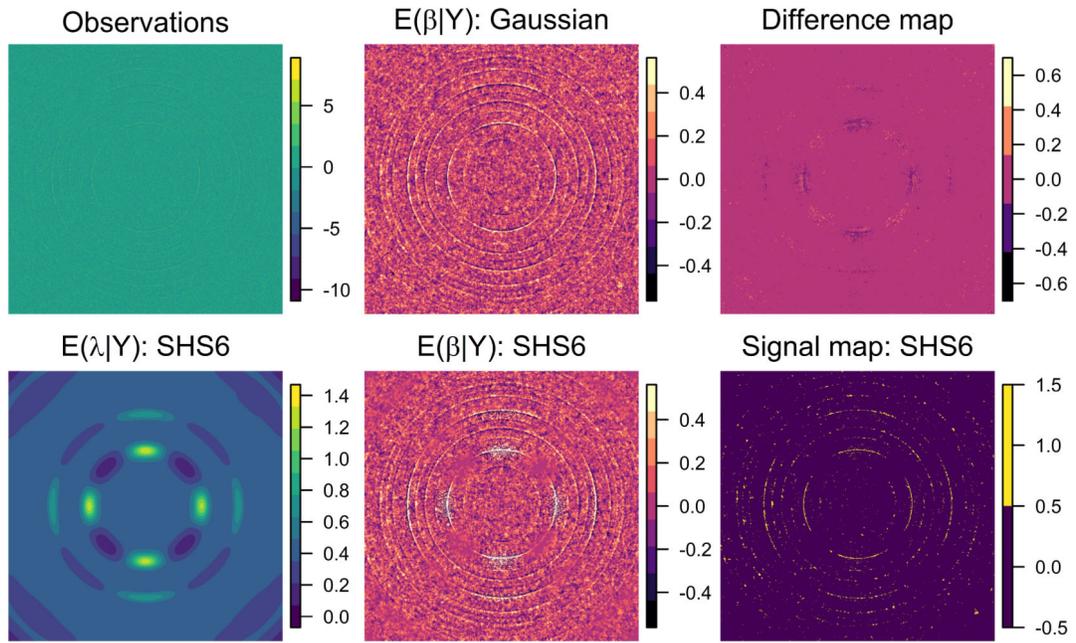


Figure 7: Results of the Gaussian model (top row), and the best prediction performance model SHS6 (bottom row) and columns in different statistics; the first column plots the observations and posterior mean of shrinkage process $\lambda(\mathbf{s})$ from the model SHS6, the second column plots the truncated (at ± 0.5) posterior mean of signal $\beta(\mathbf{s})$ to illustrate contrast between the two models, and the third column plots the difference in posterior probability that $\beta(\mathbf{s})$ is positive, i.e. $P(\beta(\mathbf{s}) > 0 \mid \mathbf{Y})_{\text{SHS6}} - P(\beta(\mathbf{s}) > 0 \mid \mathbf{Y})_{\text{Gaussian}}$, and the signal map for the model SHS6 with the Bayesian spatial false discovery rate controlled at 0.05.

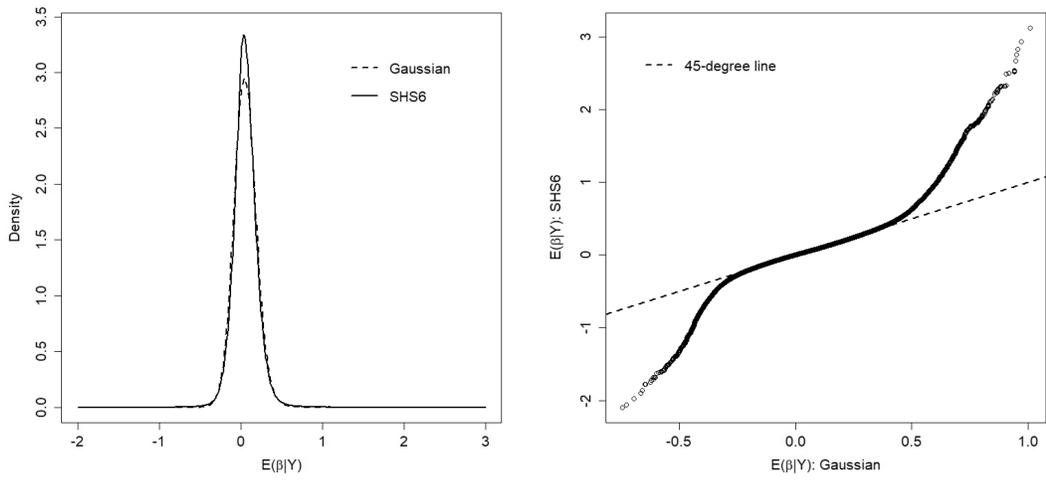


Figure 8: Density plot (left) and quantile-quantile plot (right) of the posterior mean of signal $\beta(s)$ from the Gaussian and SHS6 models.

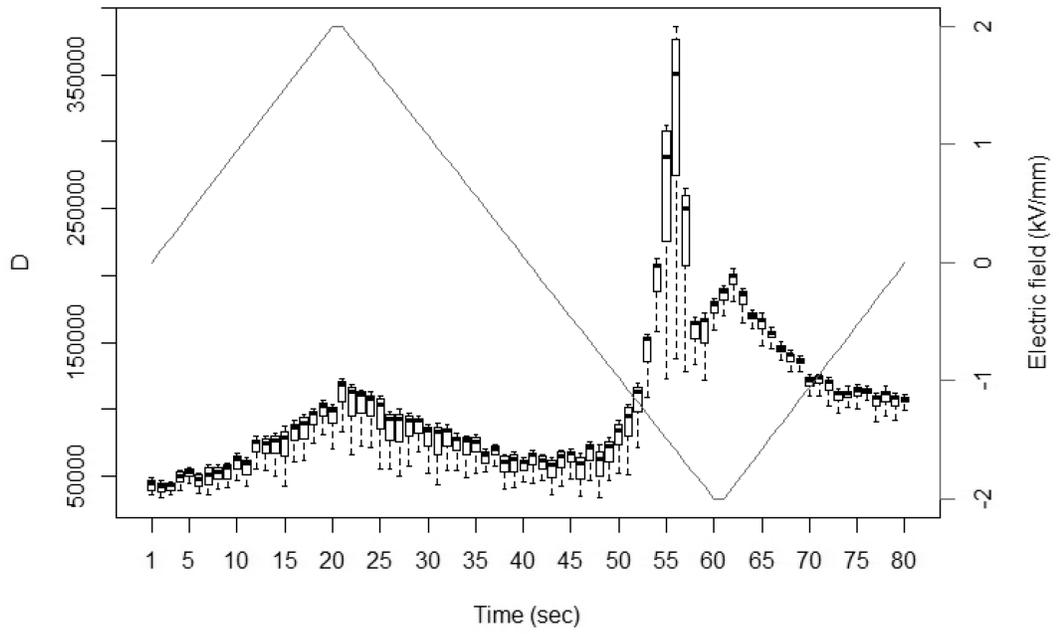


Figure 9:
 Time series plot of the posterior of $D = \sum_{i=1}^n [\beta(s_i)]^2$ (boxplots) from the model SHS6 applied separately for a bipolar electric field from -2 kV/mm to 2 kV/mm (line).

Table 1:

Summary of the simulation study under Signal 1 by error variance σ^2 for the Gaussian, spatial horseshoe with quadratic (SHS_quad), spline 1 (SHS_B-spline 1) and spline 2 (SHS_B-spline 2) shrinkage models and the soft-thresholded Gaussian process (STGP) model.

Statistics	Model	$\sigma^2 = 0.5^2$		$\sigma^2 = 1^2$		$\sigma^2 = 2^2$	
		Estimate	SE	Estimate	SE	Estimate	SE
100×RMSE	Gaussian	33.08	0.11	42.34	0.17	52.21	0.26
	SHS_quad	22.32	0.07	31.92	0.14	43.84	0.28
	SHS_B-spline 1	19.67	0.13	28.97	0.21	42.32	0.39
	SHS_B-spline 2	21.31	0.12	31.19	0.14	44.31	0.34
	STGP	24.05	0.22	30.75	0.45	39.76	0.33
Coverage (%)	Gaussian	91.67	0.17	97.46	0.05	98.66	0.05
	SHS_quad	98.09	0.04	98.68	0.03	99.28	0.03
	SHS_B-spline 1	97.91	0.51	97.78	0.68	98.61	0.60
	SHS_B-spline 2	97.61	0.36	97.31	0.08	98.22	0.13
	STGP	98.15	0.07	99.48	0.03	99.62	0.02
Type I error (%)	Gaussian	7.25	0.14	2.60	0.06	1.18	0.05
	SHS_quad	1.79	0.05	1.27	0.04	0.79	0.04
	SHS_B-spline 1	1.73	0.40	1.91	0.55	1.06	0.19
	SHS_B-spline 2	1.96	0.23	2.27	0.09	1.44	0.15
	STGP	0.56	0.03	0.15	0.02	0.01	0.00
Power (%)	Gaussian	100.00	0.00	99.97	0.02	98.00	0.14
	SHS_quad	100.00	0.00	99.86	0.03	95.36	0.25
	SHS_B-spline 1	100.00	0.00	99.86	0.03	95.65	0.25
	SHS_B-spline 2	100.00	0.00	99.94	0.02	96.60	0.27
	STGP	100.00	0.00	98.48	0.15	67.85	0.62

Table 2:

Summary of the simulation study under Signal 2 by error variance σ^2 for the Gaussian, spatial horseshoe with quadratic (SHS_quad), spline 1 (SHS_B-spline 1) and spline 2 (SHS_B-spline 2) shrinkage models and the soft-thresholded Gaussian process (STGP) model.

Statistics	Model	$\sigma^2 = 0.5^2$		$\sigma^2 = 1^2$		$\sigma^2 = 2^2$	
		Estimate	SE	Estimate	SE	Estimate	SE
100×RMSE	Gaussian	37.04	0.12	49.21	0.18	60.98	0.27
	SHS_quad	27.48	0.08	39.66	0.15	53.13	0.27
	SHS_B-spline 1	24.87	0.09	36.78	0.22	52.65	0.42
	SHS_B-spline 2	24.41	0.10	36.66	0.13	52.53	0.36
	STGP	27.03	0.20	43.07	0.46	49.55	0.37
Coverage (%)	Gaussian	88.09	0.24	96.47	0.07	98.59	0.05
	SHS_quad	96.74	0.08	97.74	0.06	98.76	0.05
	SHS_B-spline 1	97.33	0.11	97.75	0.22	98.29	0.40
	SHS_B-spline 2	97.40	0.12	96.87	0.07	98.13	0.08
	STGP	98.15	0.07	98.42	0.07	99.31	0.03
Type I error (%)	Gaussian	9.79	0.17	3.67	0.08	1.39	0.05
	SHS_quad	2.87	0.08	2.07	0.07	1.08	0.05
	SHS_B-spline 1	2.40	0.09	1.85	0.14	1.45	0.35
	SHS_B-spline 2	2.19	0.09	2.53	0.09	1.36	0.08
	STGP	0.56	0.03	0.65	0.05	0.03	0.01
Power (%)	Gaussian	100.00	0.00	99.90	0.02	94.93	0.21
	SHS_quad	100.00	0.00	99.79	0.04	92.69	0.27
	SHS_B-spline 1	100.00	0.00	99.66	0.05	92.43	0.32
	SHS_B-spline 2	100.00	0.00	99.79	0.03	92.65	0.33
	STGP	100.00	0.00	98.91	0.09	64.36	0.56

Table 3:

Comparison of prediction on the 2-D XRD data using the Gaussian and spatial horseshoe (SHS) models with k_a basis functions in the angle and k_r basis functions in the radius (SHS1-SHS6). Methods are compared using $100 \times (\text{root mean squared error} - 1)$ as RMSE, coverage (%) and computing time per 100 iterations in minutes.

Model	k_a	k_r	RMSE		Coverage (%)		Computing time (mins)
			Estimate	SE	Estimate	SE	
Gaussian	-	-	3.64	0.08	94.72	0.02	2.78
SHS1	4	10	3.59	0.08	94.78	0.02	6.57
SHS2	4	25	3.57	0.08	94.79	0.02	9.91
SHS3	4	50	3.55	0.08	94.80	0.03	17.45
SHS4	4	100	3.54	0.08	94.80	0.02	41.61
SHS5	8	5	3.56	0.08	94.80	0.02	7.38
SHS6	8	25	3.50	0.09	94.84	0.02	8.62