Bayesian Nonparametric Scalar-on-Image Regression via Potts-Gibbs Random Partition Models

Mica Shu Xian Teo and Sara Wade

Abstract Scalar-on-image regression aims to investigate changes in a scalar response of interest based on high-dimensional imaging data. We propose a novel Bayesian nonparametric scalar-on-image regression model that utilises the spatial coordinates of the voxels to group voxels with similar effects on the response to have a common coefficient. We employ the Potts-Gibbs random partition model as the prior for the random partition in which the partition process is spatially dependent, thereby encouraging groups representing spatially contiguous regions. In addition, Bayesian shrinkage priors are utilised to identify the covariates and regions that are most relevant for the prediction. The proposed model is illustrated using the simulated data sets.

1 Introduction

Through advances in data acquisition, vast amounts of high-dimensional imaging data are collected to study phenomena in many fields. Such data are common in biomedical studies to understand a disease or condition of interest [2, 5, 39, 44], and in other fields such as psychology [3, 42], social sciences [7, 15, 17, 38], economics [12, 26, 27], climate sciences [30, 31], environmental sciences [4, 11, 22] and more. While extracting features from the images based on predefined regions of interest favours interpretation and eases computational and statistical issues, changes may occur in only part of a region or span multiple structures. In order to capture the

M. S. X. Teo · S. Wade (✉)
School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh James Clerk Maxwell Building, Edinburgh, UK
e-mail: sara.wade@ed.ac.uk

M. S. X. Teo
e-mail: mica.teo@ed.ac.uk

complex spatial pattern of changes and improve accuracy and understanding of the underlying phenomenon, sophisticated approaches are required that utilize the entire high-dimensional imaging data. However, the massive dimension of the images, which is often in the millions, combined with the relatively small sample size, which at best is usually in the hundreds, pose serious challenges.

In the statistical literature, this is framed as a scalar-on-image regression (SIR) problem [10, 14, 16, 19]. SIR belongs to the "large p, small n" paradigm; thus, many SIR models utilise shrinkage methods that additionally incorporate the spatial information in the image [10, 14, 16, 18, 19, 24, 37, 40, 46]. In the SIR problem, the covariates represent the image value at a single pixel/voxel, i.e. a very tiny region, and the effect on the response is most often weak, unreliable and difficult to interpret. Moreover, neighbouring pixels/voxels are highly correlated, making standard regression methods, even with shrinkage, problematic due to multicollinearity.

To overcome these difficulties, we develop a novel Bayesian nonparametric (BNP) SIR model that extracts interpretable and reliable features from the images by grouping voxels with similar effects on the response to have a common coefficient. Specifically, we employ the Potts-Gibbs model [21] as the prior of the random image partition to encourage spatially dependent clustering. In this case, features represent regions that are automatically defined to be the most discriminative. This not only improves the signal and eases interpretability, but also reduces the computational burden by drastically decreasing the image dimension and addressing the multicollinearity problem. Moreover, it allows sharp discontinuities in the coefficient image across regions, which may be relevant in medical applications to capture irregularities [46].

In this direction, [19] proposed the Ising-DP SIR model, which combines an Ising prior to incorporate the spatial information in the sparsity structure with a Dirichlet Process (DP) prior to group coefficients. Still, the spatial information is only incorporated in the sparsity structure and not in the BNP clustering model, which could result in regions that are dispersed throughout the image. Instead, we propose to incorporate the spatial information in the random partition model, encouraging spatially contiguous regions. Further advantages of the nonparametric model include a data-driven number of clusters, interpretable parameters, and efficient computations. Moreover, we combine this with heavy-tailed shrinkage priors [41] to identify relevant covariates and regions.

The remainder of this article is organized as follows. Section 2 outlines the development of the SIR model based on the Potts-Gibbs models. Section 3 derives the MCMC algorithm for posterior inference using the generalized Swendsen-Wang (GSW) [47] algorithm for efficient split-merge moves that take advantage of the spatial structure. Section 4 illustrates the methods through simulation studies. Section 5 concludes with a summary and future work.

2 Model Specification

We introduce the statistical models that form the basis of the proposed Potts-Gibbs SIR model: SIR, random image partition model and shrinkage prior.

2.1 Scalar-on-Image Regression

SIR is a statistical linear method used to study and analyse the relationship between a scalar outcome and two or three-dimensional predictor images under a single regression model [10, 14, 16, 19]. For each data point, $i = 1, \ldots, n$, we have

$$y_i = \mathbf{w}_i^T \boldsymbol{\mu} + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathrm{N}\left(0, \sigma^2\right), \tag{1}$$

where y_i is a scalar continuous outcome measure, $\mathbf{w}_i = (w_{i1}, \ldots, w_{iq})^T \in \mathbb{R}^q$ is a q-dimensional vector of covariates, and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T \in \mathbb{R}^p$ is a p-dimensional image predictor. Each x_{ij} indicates the value of the image at a single pixel with spatial location $\mathbf{s}_j = (s_{j1}, s_{j2})^T \in \mathbb{R}^2$ for $j = 1, \ldots, p$. We define $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_q)^T \in \mathbb{R}^q$ as a q-dimensional fixed effects vector and $\boldsymbol{\beta} = \left(\beta(\mathbf{s}_1), \ldots, \beta(\mathbf{s}_p)\right)^T$ (with $\beta_j := \beta(\mathbf{s}_j)$) as the spatially varying coefficient image described on the same lattice as $\mathbf{x}_i$. We model the high-dimensional $\boldsymbol{\beta}$ by spatially clustering the pixels into M regions and assuming common coefficients $\beta_1^*, \ldots, \beta_M^*$ within in each cluster, i.e. $\beta_j = \beta_m^*$ given the cluster label $z_j = m$. Thus, the prior on the coefficient image is decomposed into two parts: the random image partition model for spatially clustering the pixels and a shrinkage prior for the cluster-specific coefficients $\boldsymbol{\beta}^* = \left(\beta_1^*, \ldots, \beta_M^*\right)^T$. The SIR model in (1) can be extended for other types of responses through a generalized linear model framework (GLM) [23].

2.2 Random Image Partition Model

The image predictors are observed on a spatially structured coordinate system. Exchangeability is indeed no longer the proper assumption as the images contain covariate information, that we wish to leverage to improve model performance in this high-dimensional setting. To do so, we combine BNP random partition models, which avoid the need to prespecify the number of clusters, allowing it be determined and grow with the data, with a Potts-like spatial smoothness component [36]. Spatial random partition models in this direction are a growing research area, including Markov random field (MRF) with the product partition model (PPM) [32], with DP [29, 47], with Pitman-Yor process (PY)[21] and with mixture of finite mixtures (MFM) [13, 48]. Precisely, within the BNP framework, we focus on the class of Gibbs-type random partitions [1, 9, 20, 35], motivated by their comprise between tractable pre-

Table 1 Formulas of $V_p(M)$, $W_m(\phi)$ and terms of the predictive probability for assigning current cluster to either existing cluster or new cluster for DP, PY and MFM

	DP	PY	MFM
$V_p(M)$	$\frac{\Gamma(\alpha)\alpha^M}{\Gamma(\alpha+p)}$	$\frac{\Gamma(\alpha+1)\prod_{m=1}^{M-1}(\alpha+m\delta)}{\Gamma(\alpha+p)}$	$\sum_{l=1}^{\infty}\frac{\Gamma(\gamma l)l!}{\Gamma(\gamma l+p)(l-m)!}P_L(\cdot|\lambda)$
$W_m(\phi)$	$\Gamma(\mid C_m\mid)$	$\frac{\Gamma(\mid C_m\mid-\delta)}{\Gamma(1-\delta)}$	$\frac{\Gamma(\mid C_m\mid+\gamma)}{\Gamma(\gamma)}$
Existing cluster	$\frac{\Gamma(\mid C_m^{-A_o}\mid+\mid A_o\mid)}{\Gamma(\mid C_m^{-A_o}\mid)}$	$\frac{\Gamma(\mid C_m^{-A_o}\mid+\mid A_o\mid-\delta)}{\Gamma(\mid C_m^{-A_o}\mid-\delta)}$	$\frac{\Gamma(\mid C_m^{-A_o}\mid+\mid A_o\mid+\gamma)}{\Gamma(\mid C_m^{-A_o}\mid+\gamma)}$
New cluster	$\alpha\Gamma(\mid A_o\mid)$	$(\alpha+\delta M^{-A_o})\frac{\Gamma(\mid A_o\mid-\delta)}{\Gamma(1-\delta)}$	$\frac{V_p(M^{-A_o}+1)}{V_p(M^{-A_o})}\frac{\Gamma(\mid A_o\mid+\gamma)}{\Gamma(\gamma)}$

Note that the predictive probabilities are stated up to a proportionality constant

dictive rules and richness of the predictive structure, including important cases, such as the DP [6], PY [33, 34], and MFM [25]. The Potts-Gibbs models induce a distribution on the partition $\pi_p = \{C_1, \ldots, C_M\}$ of p pixels into M nonempty, mutually exclusive, and exhaustive subsets $C_1, \ldots, C_M$ such that $\cup_{C\in\pi_p} C = \{1, \ldots, p\}$. The model can be summarised as:

$$\mathrm{pr}(\pi_p) \propto \exp\underbrace{\left(\sum_{j\sim k,\, j<k}\upsilon_{jk}\mathbf{1}_{z_j=z_k}\right)}_{\text{Potts model}} \underbrace{\left(V_p(M)\prod_{m=1}^{M}W_m(\phi)\right)}_{\text{Gibbs-type random partition models}},$$

where $z_j \in \{1, \cdots, M\}$, $j \sim k$ means that j and k are neighbors, and $\mathbf{1}_{z_j=z_k}$ equals to 1 if j and k in the same cluster and 0 otherwise. In the following, we assume the spatial locations lie on a rectangular lattice with first-order neighbors and a common coupling parameter υ for all neighbor pairs; a higher value of υ encourages more spatial smoothness in the partition. We use the general notation ϕ to denote the parameters of the Gibbs-type partition models, and focus our study on three cases 1) DP with concentration parameter $\alpha > 0$; 2) PY with discount parameter $\delta \in [0, 1)$ and concentration parameter $\alpha > -\delta$; and 3) MFM with parameter $\gamma > 0$ (larger values encouraging more equally sized clusters) and a distribution $P_L(\cdot|\lambda)$ with parameter λ related to the prior on the number of clusters. The $\{V_p(M) : p \geq 1, 1 \leq M \leq p\}$ denotes the set of non-negative weights, which solves the backward recurrence relation $V_p(M) = (p - \delta M)V_{p+1}(M) + V_{p+1}(M + 1)$ with $V_1(1) = 1$. Table 1 describes the $V_p(M)$ and $W_m(\phi)$ for DP, PY and MFM models.

2.3 Shrinkage Prior

To identify relevant regions, we use heavy tailed priors for the unique values $(\beta_1^*, \ldots, \beta_M^*)$ of $(\beta(\mathbf{s}_1), \ldots, \beta(\mathbf{s}_p))$. Specifically, a t-shrinkage prior is used, motivated by its computational efficiency and nearly optimal contraction rate and selection consistency [41]:

$$\sigma^2 \sim \text{IG}\,(a_\sigma, b_\sigma)\,,$$
$$\left(\beta_m^*\right)|\sigma^2 \sim t_\nu(s\sigma), \quad \text{for all } m = 1, \ldots, M, \tag{2}$$

where $t_\nu(s\sigma)$ denotes t-distribution with degree of freedom ν and scale parameter $s\sigma$. For posterior inference, the t-distribution (2) is rewritten as a hierarchical inverse-gamma scaled Gaussian mixture,

$$\sigma^2 \sim \text{IG}\,(a_\sigma, b_\sigma)\,,$$
$$\eta_m^* \sim \text{IG}\left(a_\eta, b_\eta\right),$$
$$\left(\beta_m^*\right)|\sigma^2, \eta_m^* \sim N(0, \eta_m^*\sigma^2), \quad \text{for all } m = 1, \ldots, M,$$

where a_η and b_η are the shape and scaling parameter of the mixing distribution for each η_m^* respectively with $\nu = 2a_\eta$ and $s = \sqrt{b_\eta/a_\eta}$.

3 Inference

We aim to infer the posterior distribution of the parameters based on the proposed Potts-Gibbs SIR model:

$$y_i \mid \boldsymbol{\mu}, \boldsymbol{\beta}^*, \pi_p, \sigma^2 \sim N(\mathbf{w}_i^T\boldsymbol{\mu} + \mathbf{x}_i^{*T}\boldsymbol{\beta}^*, \sigma^2), \quad \text{for all } i = 1, \ldots, n,$$
$$\boldsymbol{\mu} \mid \sigma^2 \sim N(\mathbf{m}_\mu, \sigma^2\Sigma_\mu),$$
$$\boldsymbol{\beta}^* \mid \boldsymbol{\eta}^*, \sigma^2 \sim N(\mathbf{0}_M, \sigma^2\Sigma_{\beta^*}),$$
$$\sigma^2 \sim \text{IG}(a_\sigma, b_\sigma),$$
$$\eta_m^* \sim \text{IG}\left(a_\eta, b_\eta\right), \quad \text{for all } m = 1, \ldots, M,$$
$$\pi_p \sim \text{Potts-Gibbs}(\upsilon, \phi),$$

where $x_{im}^* = \sum_{j=1}^p x_{ij}\mathbf{1}(j \in C_m)/\sqrt{|C_m|}$ represents the total value, e.g. volume in the mth region of the image, $\mathbf{m}_\mu = (m_{\mu_1}, \ldots, m_{\mu_q})$, $\Sigma_\mu = \text{diag}(c_{\mu_1}, \ldots, c_{\mu_q})^T$, and $\Sigma_{\beta^*} = \text{diag}(\eta_1^*, \ldots, \eta_M^*)$. Note that when defining x_{im}^*, we rescale by the square root of cluster size , which is equivalent to rescaling the variance of β_m^* by the cluster size, encouraging more shrinkage for larger regions.

We develop a Gibbs sampler to simulate from the posterior with a generalized Swendsen-Wang (GSW) algorithm to draw samples from the Potts-Gibbs model. Poor mixing can be seen in single-site Gibbs sampling [8] due to the high correlation between the pixel labels. The SW algorithm [43] addresses this by forming nested clusters of neighbouring pixels, then updating all of the labels within a nested cluster to the same value. The generalisation of the technique for standard Potts models to generalised Potts-partition models is called GSW [47]. At each step of the algorithm, we proceed through the following steps:

1. Sample the image partition π_p given $\boldsymbol{\eta}^*$ and the data (with $\boldsymbol{\beta}^*$, $\boldsymbol{\mu}$, σ^2 marginalized). GSW is used to update simultaneously nested groups of pixels and hence improve the exploration of the posterior. The algorithm relies on the introduction of auxiliary binary bond variables, where $r_{jk} = 1$ if pixels j and k are bonded, otherwise 0. The bond variables define a partition of the pixels into nested clusters $A_1, \ldots, A_O$, where O denotes the number of nested clusters and each $A_o \subseteq C_m$ for some $m = 1, \ldots, M$. For each neighbor pair $j \sim k$ for $1 \leq j < k \leq p$, we sample the bond variables as follows, $r_{jk} \sim \text{Ber}\{1 - \exp(-\upsilon_{jk}\zeta_{jk}\mathbf{1}_{z_j=z_k})\}$, where we define $\zeta_{jk} = \kappa \exp\{-\tau d(\hat{\beta}_j, \hat{\beta}_k)\}$ with $\hat{\beta}_j$ denoting the estimated coefficient from univariate regression on the jth pixel and κ, τ are the tuning parameters of the GSW sampler. Notice that the algorithm reduces to single-site Gibbs when $\kappa = 0$, and recovers classical SW when $\kappa = 1$ and $\tau = 0$.

   As we are dealing with non-conjugate priors, we update the cluster assignment by extending Gibbs sampling with the addition of auxiliary parameters, which is widely known as Algorithm 8 [28]. We denote by A_o the current nested cluster; $C_1^{-A_o}, \ldots, C_M^{-A_o}$ the clusters without nested cluster A_o; M^{-A_o} the number of distinct clusters excluding A_o and h the number of temporary auxiliary variables. For each nested cluster A_o, it is assigned to an existing cluster $m = 1, \ldots, M^{-A_o}$ or a new cluster $m = M^{-A_o} + 1, \ldots, M^{-A_o} + h$ with probability as follows,

   $$\text{pr}(A_o \in C_m^{-A_o} \mid \cdots)$$
   $$\propto \begin{cases} \frac{\Gamma(|C_m^{-A_o}|+|A_o|-\delta)}{\Gamma(|C_m^{-A_o}|-\delta)}\text{pr}\left(\mathbf{y} \mid \pi_p^{A_o \to m}, \boldsymbol{\eta}^*\right) \\ \prod_{\{(j,k)|j \in A_o, k \in C_m^{-A_o}, r_{jk}=0\}} \exp\left\{\upsilon_{jk}(1-\zeta_{jk})\right\}, & \text{for } C_m^{-A_o} \in \pi_p^{-A_o}, \\ \\ \frac{1}{h}\frac{V_p(M^{-A_o}+1)}{V_p(M^{-A_o})}\frac{\Gamma(|A_o|-\delta)}{\Gamma(1-\delta)}\text{pr}\left(\mathbf{y} \mid \pi_p^{A_o \to M+1}, \boldsymbol{\eta}^*\right), & \text{for new } C_m^{-A_o}; \end{cases}$$

   where $\text{pr}\left(\mathbf{y} \mid \pi_p^{A_o \to m}, \boldsymbol{\eta}^*\right)$ and $\text{pr}\left(\mathbf{y} \mid \pi_p^{A_o \to M+1}, \boldsymbol{\eta}^*\right)$ denote the marginal likelihood of data obtained by moving A_o from its current cluster to existing clusters or newly created cluster respectively. Before updating the cluster assignments, we sample the nested clusters and compute the volume of each nested cluster for all images, with computational cost $\mathcal{O}(np)$. When updating the cluster assignments, the marginal likelihood dominates the computational cost, as it involves inversion and determinants of $(M + q) \times (M + q)$ matrices and updating the sufficient statistics for every nested cluster and every outer cluster allocation, i.e. the cost is $\mathcal{O}([[M + q]^3 + n[M + q]]OM)$.

2. Sample $\boldsymbol{\beta}^*$, $\boldsymbol{\mu}$, σ^2 jointly given the partition π_p, $\boldsymbol{\eta}^*$ and the data. Notationally, we reformulate $\tilde{\mathbf{x}}_i = (\mathbf{w}_i^T, \mathbf{x}_i^{*T})^T$ and $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\mu}^T, \boldsymbol{\beta}^{*T})^T$. We define $\tilde{\mathbf{X}}$ be the matrix with rows equal to $\tilde{\mathbf{x}}_i^T$. The corresponding full conditional for $\tilde{\boldsymbol{\beta}}$ and σ^2 is

   $$\sigma^2 \mid \cdots \sim \text{IG}(\hat{a}_\sigma, \hat{b}_\sigma),$$
   $$\tilde{\boldsymbol{\beta}} \mid \sigma^2, \cdots \sim \text{N}(\hat{\mathbf{m}}_{\tilde{\beta}}, \sigma^2 \hat{\Sigma}_{\tilde{\beta}}),$$

(a) Scenario 1: Truth   (b) Scenario 1: Ising-DP

(c) Scenario 1: Potts-DP   (d) Scenario 1: Potts-PY   (e) Scenario 1: Potts-MFM
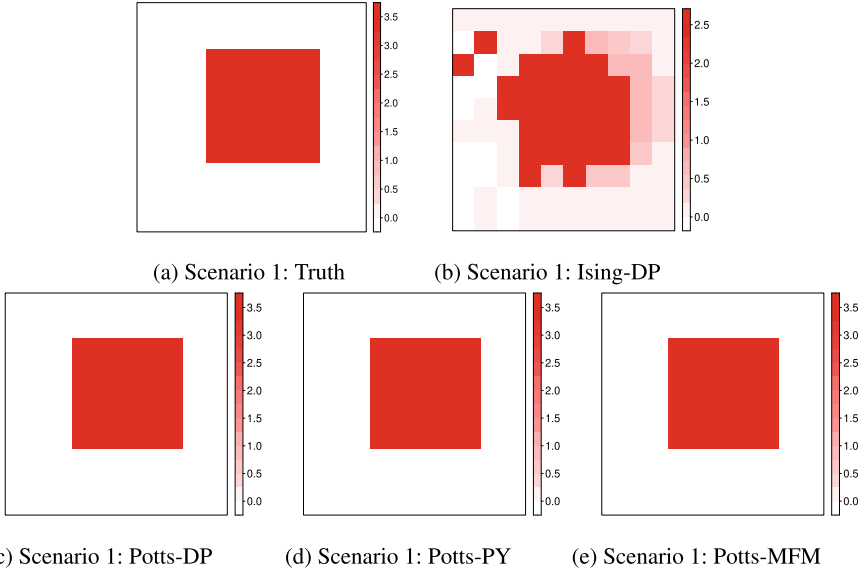
Fig. 1 Figures on the upper and bottom row showing the true and estimated coefficient matrix of the simulated data sets for scenario 1 under each model

where $\hat{\Sigma}_{\tilde{\beta}} = (\Sigma_{\tilde{\beta}}^{-1} + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$, $\hat{\mathbf{m}}_{\tilde{\beta}} = \hat{\Sigma}_{\tilde{\beta}}(\Sigma_{\tilde{\beta}}^{-1} \mathbf{m}_{\tilde{\beta}} + \tilde{\mathbf{X}}^T \mathbf{y})$, and $\mathrm{IG}(\hat{a}_\sigma, \hat{b}_\sigma)$ denotes the inverse-gamma distribution with updated shape $\hat{a}_\sigma = a_\sigma + n/2$ and scale $\hat{b}_\sigma = b_\sigma + [\mathbf{m}_{\tilde{\beta}}^T \Sigma_{\tilde{\beta}}^{-1} \mathbf{m}_{\tilde{\beta}} + \mathbf{y}^T \mathbf{y} - \hat{\mathbf{m}}_{\tilde{\beta}}^T \hat{\Sigma}_{\tilde{\beta}}^{-1} \hat{\mathbf{m}}_{\tilde{\beta}}]/2$.

3. Sample $\boldsymbol{\eta}^*$ given $\boldsymbol{\beta}^*$. The corresponding full conditional for each η_m^* is an inverse-gamma distribution with updated shape $\hat{a}_\eta = a_\eta + 1/2$ and scale $\hat{b}_\eta = b_\eta + (\beta_m^*)^2/(2\sigma^2)$:

$$\eta_m^* \mid \cdots \sim \mathrm{IG}(\hat{a}_\eta, \hat{b}_\eta), \quad \text{for } m = 1, \ldots, M.$$

4 Numerical Studies

We study through simulations the performance of the proposed model and compare it with Ising-DP [19]. We consider 2D images in this simulation. The $n = 300$ images are simulated on a two dimensional grid of size 10×10, with spatial locations $\mathbf{s}_j = (s_{j1}, s_{j2}) \in \mathbf{R}^2$ for $1 \le s_{j1}, s_{j2} \le 10$. For simplicity's sake, we include an intercept but do not consider others covariates, $\mathbf{w}_i$. We concentrate on the two simulation scenarios with true $M = 2$ and $M = 5$ as shown in Figs. 1 and 2. For each experiment, we summarise the posterior of the clustering structure of the data sets by minimising the posterior expected Variation of Information (VI) [45].
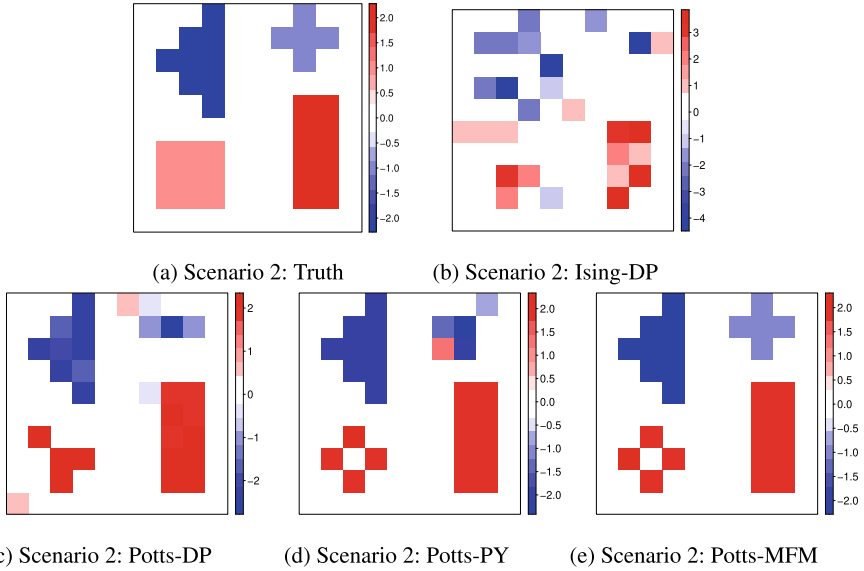
(a) Scenario 2: Truth   (b) Scenario 2: Ising-DP

(c) Scenario 2: Potts-DP   (d) Scenario 2: Potts-PY   (e) Scenario 2: Potts-MFM

Fig. 2 Figures on the upper and bottom row showing the true and estimated coefficient matrix of the simulated data sets for scenario 2 under each model

The Potts-Gibbs models can detect correctly the cluster structure under scenario 1 (Fig. 1). The Potts-Gibbs models are also capable of capturing and identifying the more complex cluster structure underlying the data for scenario 2 (Fig. 2) with the ARI 0.621–0.830 (Table 2). On the contrary, Ising-DP has failed terribly to recover the cluster structure for scenario 2, as illustrated in Fig. 2. It is observed that under the Potts-Gibbs models, most of the resultant clusters are spatially proximal, while under Ising-DP, the clusters are dispersed throughout the image. By taking into consideration spatial dependence in the random partition model via the Potts-Gibbs models, the proposed models produce spatially aware clustering and thus improve the predictions.

DP has a concentration parameter α, with larger values encouraging more new clusters and a rich-get-richer property that favours allocation to larger clusters. The PY has an additional discount parameter $\delta \in [0, 1)$ that helps to mitigate the rich-get-richer property and phase transition of the Potts model. The MFM has a parameter γ, with larger values encouraging more equal-sized clusters and helping to avoid phase transition of the Potts model, as well as additional parameters λ which are related to the prior on the number of clusters.

Table 2 Mean, standard deviation (in parentheses) and highest posterior density (HPD) interval of the posterior of adjusted Rand index (ARI), variation information (VI), mean squared error (MSE), mean squared prediction error (MSPE), and number of clusters for each scenario under each model

	Model	Scenario	Mean	HPD (95%)
ARI	Potts-DP	1	**1.0** (0.004)	(1.0, 1.0)
	Potts-PY		**1.0** (0.004)	(1.0, 1.0)
	Potts-MFM		**0.999** (0.007)	(1.0, 1.0)
	Ising-DP		0.307 (0.079)	(0.152, 0.464)
	Potts-DP	2	**0.621** (0.060)	(0.472, 0.684)
	Potts-PY		**0.713** (0.050)	(0.607, 0.818)
	Potts-MFM		**0.830**(0.036)	(0.756, 0.869)
	Ising-DP		0.038 (0.021)	(-0.001, 0.078)
VI	Potts-DP	1	**0.001** (0.010)	(2.22e-16, 2.22e-16)
	Potts-PY		**0.001** (0.009)	(2.220e-16, 2.220e-16)
	Potts-MFM		**0.001** (0.014)	(2.220e-16, 2.220e-16)
	Ising-DP		1.386 (0.154)	(1.083, 1.680)
	Potts-DP	2	**1.160** (0.211)	(0.902, 1.548)
	Potts-PY		**1.006** (0.147)	(0.640, 1.299)
	Potts-MFM		**0.599** (0.133)	(0.432, 0.866)
	Ising-DP		3.990 (0.159)	(3.691, 4.290)
MSE	Potts-DP	1	**1.33e-4** (5.59e-4)	(3.97e-9, 2.67e-4)
	Potts-PY		**1.03e-4** (8.73e-5)	(1.58e-7, 2.57e-4)
	Potts-MFM		**1.01e-4** (8.37e-5)	(4.21e-7, 2.66e-4)
	Ising-DP		0.807 (0.011)	(0.790, 0.828)
	Potts-DP	2	**0.246** (0.064)	(0.141, 0.374)
	Potts-PY		**0.157** (0.035)	(0.094, 0.224)
	Potts-MFM		**0.093** (0.014)	(0.079, 0.125)
	Ising-DP		0.980 (0.025)	(0.942, 1.020)
MSPE	Potts-DP	1	**4.215** (0.057)	(4.152, 4.317)
	Potts-PY		**4.213** (0.052)	(4.138, 4.311)
	Potts-MFM		**4.209** (0.052)	(4.136, 4.314)
	Ising-DP		145.912 (10.051)	(129.142, 165.950)
	Potts-DP	2	7.754 (2.653)	(3.175, 13.356)
	Potts-PY		**0.868**(0.168)	(0.669, 1.189)
	Potts-MFM		**0.850** (0.122)	(0.677, 1.108)
	Ising-DP		3.641 (0.526)	(2.766, 4.857)
Number of clusters	Potts-DP	1	2.019 (0.138)	(2.0, 2.0)
	Potts-PY		2.015 (0.122)	(2.0, 2.0)
	Potts-MFM		2.007 (0.081)	(2.0, 2.0)
	Ising-DP		4.575 (1.340)	(2.0, 7.0)
	Potts-DP	2	6.722 (0.901)	(5.0, 8.0)
	Potts-PY		6.882 (1.090)	(5.0, 9.0)
	Potts-MFM		5.232 (0.475)	(5.0, 6.0)
	Ising-DP		15.542 (1.554)	(13.0, 18.0)

5 Conclusion

We have developed novel Bayesian scalar-on-image regression models to extract interpretable features from the image by clustering and leveraging the spatial coordinates of the pixels/voxels. To encourage groups representing spatially contiguous regions, we incorporate the spatial information directly in the prior for the random partition through Potts-Gibbs random partition models. We have shown the potential of Potts-Gibbs models in detecting the correct cluster structure on simulated data sets. In our experiments, the hyperparameters of the Potts-Gibbs model were determined via a simple grid search on selected combinations of hyperparameters. However, future work will consist of investigating the influence of the various parameters inherent to the model and guidelines and tools to determine hyperparameters. The model will then be applied to real images, e.g. neuroimages. Motivated by examining and identifying brain regions of interest in Alzheimer's disease, we will use MRI images obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.adni-info.org). The proposed SIR model will be extended to classification problems through the GLM framework.

References

1. Cerquetti, A.: Generalized Chinese restaurant construction of exchangeable Gibbs partitions and related results. arXiv:0805.3853 (2008)
2. Craddock, R.C., Holtzheimer, P.E., III., Hu, X.P., Mayberg, H.S.: Disease state prediction from resting state functional connectivity. Magn. Reson. Med. **62**(6), 1619–1628 (2009)
3. Davatzikos, C., Shen, D., Gur, R.C., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B.I., Gur, R.E.: Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch. Gen. Psychiatry **62**(11), 1218–1227 (2005)
4. Debois, D., Ongena, M., Cawoy, H., De Pauw, E.: MALDI-FTICR MS imaging as a powerful tool to identify Paenibacillus antibiotics involved in the inhibition of plant pathogens. J. Am. Soc. Mass Spectrom. **24**(8), 1202–1213 (2013)
5. Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C.: Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. NeuroImage **41**(2), 277–285 (2008)
6. Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**, 209–230 (1973)
7. Ferwerda, B., Schedl, M., Tkalcic, M.: Using instagram picture features to predict users' personality. In: International Conference on Multimedia Modeling. Springer (2016)
8. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984)
9. Gnedin, A., Pitman, J.: Exchangeable Gibbs partitions and Stirling triangles. J. Math. Sci. **138**(3), 5674–5685 (2006)
10. Goldsmith, J., Huang, L., Crainiceanu, C.M.: Smooth scalar-on-image regression via spatial Bayesian variable selection. J. Comput. Graph Stat. **23**(1), 46–64 (2014)
11. Gundlach-Graham, A., Burger, M., Allner, S., Schwarz, G., Wang, H.A.O., Gyr, L., Grolimund, D., Hattendorf, B., Günther, D.: High-speed, high-resolution, multielemental laser ablation-inductively coupled plasma-time-of-flight mass spectrometry imaging: Part i. instrumentation and two-dimensional imaging of geological samples. Anal. Chem. **87**(16), 8250–8258 (2015)

12. Henderson, J.V., Storeygard, A., Weil, D.N.: Measuring economic growth from outer space. National Bureau of Economic Research, Cambridge, Mass (2009)
13. Hu, G., Geng, J., Xue, Y., Sang, H.: Bayesian spatial homogeneity pursuit of functional data: an application to the U.S. income distribution. arXiv:2002.06663 (2020)
14. Huang, L., Goldsmith, J., Reiss, P.T., Reich, D.S., Crainiceanu, C.M.: Bayesian scalar-on-image regression with application to association between intracranial DTI and cognitive outcomes. NeuroImage **83**, 210–223 (2013)
15. Hum, N.J., Chamberlin, P.E., Hambright, B.L., Portwood, A.C., Schat, A.C., Bevan, J.L.: A picture is worth a thousand words: a content analysis of Facebook profile photographs. Comput. Hum. Behav. **27**(5), 1828–1833 (2011)
16. Kang, J., Reich, B.J., Staicu, A.M.: Scalar-on-image regression via the soft-thresholded Gaussian process. Biometrika **105**(1), 165–184 (2018)
17. Kim, Y., Kim, J.H.: Using computer vision techniques on instagram to link users' personalities and genders to the features of their photos: an exploratory study. Inf. Process. Manage. **54**(6), 1101–1114 (2018)
18. Lee, K., Cao, X.: Bayesian group selection in logistic regression with application to MRI data analysis. Biometrics **77**(2), 391–400 (2021)
19. Li, F., Zhang, T., Wang, Q., Gonzalez, M.Z., Maresh, E.L., Coan, J.A.: Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. Ann. Appl. Stat. **9**(2), 687–713 (2015)
20. Lijoi, A., Prünster, I.: Models beyond the Dirichlet process. In: Bayesian Nonparametrics (2010)
21. Lü, H., Arbel, J., Forbes, F.: Bayesian nonparametric priors for hidden Markov random fields. Stat. Comput. **30**(4), 1015–1035 (2020)
22. Maloof, K.A., Reinders, A.N., Tucker, K.R.: Applications of mass spectrometry imaging in the environmental sciences. Curr. Opin. Environ. Sci. Health. **18**, 54–62 (2020)
23. McCullagh, P., Nelder, J.A.: Generalized linear models. Routledge (2019)
24. Mehrotra, S., Maity, A.: Simultaneous variable selection, clustering, and smoothing in function-on-scalar regression. Can. J, Stat (2021)
25. Miller, J.W., Harrison, M.T.: Mixture models with a prior on the number of components. J. Am. Stat. Assoc. **113**(521), 340–356 (2018)
26. Naik, N., Kominers, S.D., Raskar, R., Glaeser, E.L., Hidalgo, C.A.: Computer vision uncovers predictors of physical urban change. Proc. Natl. Acad. Sci. U.S.A. **114**(29), 7571–7576 (2017)
27. Naik, N., Raskar, R., Hidalgo, C.A.: Cities are physical too: using computer vision to measure the quality and impact of urban appearance. Am. Econ. Rev. **106**(5), 128–132 (2016)
28. Neal, R.M.: Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph Stat. **9**(2), 249–265 (2000)
29. Orbanz, P., Buhmann, J.M.: Nonparametric Bayesian image segmentation. Int. J. Comput. Vis. **77**(1–3), 25–45 (2007)
30. O'Neill, S.J.: Image matters: climate change imagery in US. UK and Australian newspapers. Geoforum **49**, 10–19 (2013)
31. O'Neill, S.J., Boykoff, M., Niemeyer, S., Day, S.A.: On the use of imagery for climate change engagement. Glob. Environ. Change **23**(2), 413–421 (2013)
32. Pan, T., Hu, G., Shen, W.: Identifying latent groups in spatial panel data using a Markov random field constrained product partition model. arXiv:2012.10541 (2020)
33. Perman, M., Pitman, J., Yor, M.: Size-biased sampling of Poisson point processes and excursions. Probab. Theory Relat. Fields **92**(1), 21–39 (1992)
34. Pitman, J.: Some developments of the Blackwell-Macqueen urn scheme. Lect. Notes-Monograph Ser. **30**, 245–267 (1996)
35. Pitman, J.: Lecture Notes in Mathematics. Springer (2006)
36. Potts, R.B., Domb, C.: Some generalized order-disorder transformations. Math. Proc. Cambridge Philos. Soc. **48**(1), 106 (1952). https://doi.org/10.1017/S0305004100027419
37. Reiss, P., Mennes, M., Petkova, E., Huang, L., Hoptman, M., Biswal, B., Colcombe, S., Zuo, X., Milham, M.: Extracting information from functional connectivity maps via function-on-scalar regression. NeuroImage **56**, 140–148 (2011)

38. Samany, N.N.: Automatic landmark extraction from geo-tagged social media photos using deep neural network. Cities **93**, 1–12 (2019)
39. Shi, J., Lepore, N., Gutman, B., Thompson, P., Baxter, L., Caselli, R., Wang, Y.: Genetic influence of apolipoprotein E4 genotype on hippocampal morphometry: an N = 725 surface-based Alzheimer's disease neuroimaging initiative study. Hum. Brain Mapp. **35**(8), 3903–3918 (2014)
40. Smith, M., Fahrmeir, L.: Spatial Bayesian variable selection with application to functional magnetic resonance imaging. J. Am. Stat. Assoc. **102**(478), 417–431 (2007)
41. Song, Q., Liang, F.: Nearly optimal Bayesian shrinkage for high dimensional regression. arXiv:1712.08964 (2017)
42. Sun, D., van Erp, T.G., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K.H., Toga, A.W., Cannon, T.D.: Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: Classification analysis using probabilistic brain atlas and machine learning algorithms. Biol. Psychiatry (1969) **66**(11), 1055–1060 (2009)
43. Swendsen, R.H., Wang, J.S.: Nonuniversal critical dynamics in Monte Carlo simulations. Phys. Rev. Lett. **58**(2), 86–88 (1987)
44. Van Walderveen, M., Kamphorst, W., Scheltens, P., Van Waesberghe, J., Ravid, R., Valk, J., Polman, C., Barkhof, F.: Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. Neurology **50**(5), 1282–1288 (1998)
45. Wade, S., Ghahramani, Z.: Bayesian cluster analysis: Point estimation and credible balls (with discussion). Bayesian Anal. **13**(2), 559–626 (2018)
46. Wang, X., Zhu, H., Initiative, A.D.N.: Generalized scalar-on-image regression models via total variation. J. Am. Stat. Assoc. **112**(519), 1156–1168 (2017)
47. Xu, R.Y.D., Caron, F., Doucet, A.: Bayesian nonparametric image segmentation using a generalized Swendsen-Wang algorithm. arXiv:1602.03048 (2016)
48. Zhao, P., Yang, H.C., Dey, D.K., Hu, G.: Bayesian spatial homogeneity pursuit regression for count value data. arXiv:2002.06678 (2020)