

# Supplementary materials for the article titled “Spatial Product Partition Models”

Garritt L. Page\*

Departamento de Estadística

Pontificia Universidad Católica de Chile

page@mat.puc.cl

Fernando A. Quintana

Departamento de Estadística

Pontificia Universidad Católica de Chile

quintana@mat.uc.cl

August 20, 2015

## 1 Example of Partition Probabilities

To further explore differences among prior partition probabilities available for each cohesion we consider here a slightly more complex toy example than the one presented in the manuscript. The example is based on the regular spatial configuration found in Figure S.1 with four spatial locations  $\mathbf{s}_1 = (0, 0)$ ,  $\mathbf{s}_2 = (1, 0)$ ,  $\mathbf{s}_3 = (1, 0)$ , and  $\mathbf{s}_4 = (1, 1)$ . The remaining nine points found in Figure S.1 are possible cluster centroids and are provided to illustrate distances used in  $C_1$ . For example,  $\bar{\mathbf{s}}_{12} = (1/2, 0)$  denotes the centroid for the cluster  $\{\mathbf{s}_1, \mathbf{s}_2\}$ . In this toy example there are 15 possible ways to partition locations and the symmetry of locations produces symmetry in the prior probabilities. We use the same parameter values for  $C_3$  and  $C_4$  as those used previously. Additionally we set  $a = 1$  (the

---

\*Corresponding author

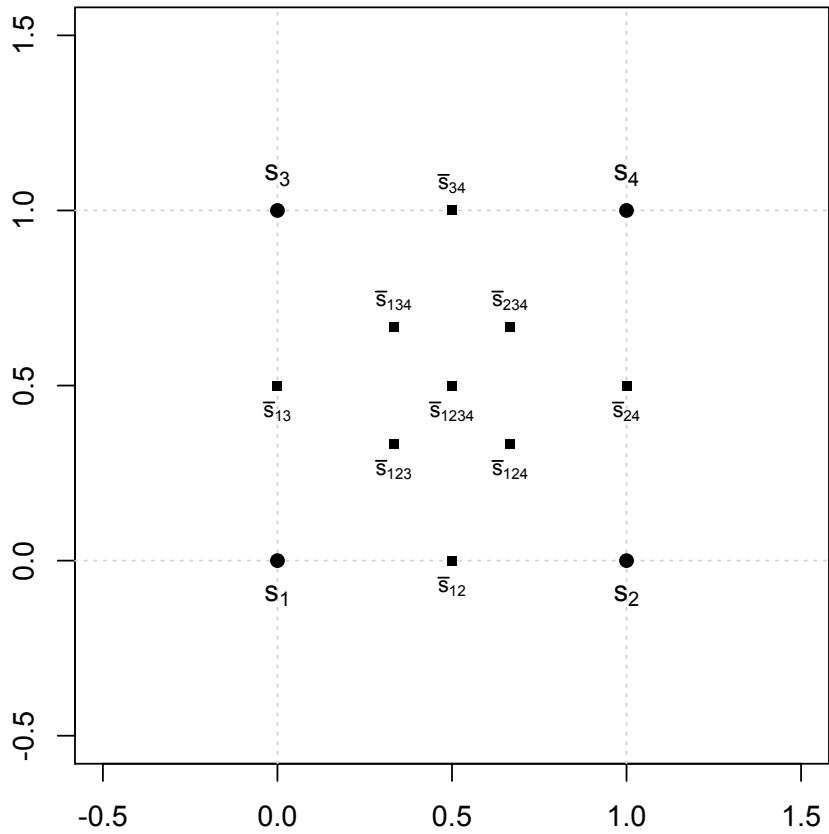


Figure S.1: Spatial Configuration of toy example. Circles represent locations considered. Squares correspond to centroids associated with different clusters. For example  $\bar{s}_{1234}$  is the centroid for cluster  $\{1, 2, 3, 4\}$  and is equal to  $(0.5, 0.5)$

median distance among all pairwise distances) and  $\alpha = 1$ . For each partition we list the probabilities for  $M = 1$  in Table S.1.

Notice that partitions that are not spatially connected  $\{\{1, 4\}, \{2, 3\}\}$ ,  $\{\{1, 4\}, \{2\}, \{3\}\}$ , and  $\{\{1\}, \{2, 3\}, \{4\}\}$  (rows 8, 11, and 12 of Table S.1) have the smallest probabilities for  $C_1$ ,  $C_4$ , and  $C_2$ . For  $C_3$  they are the smallest among similar partitions. Apart from  $C_2$  all cohesions place highest probability on partition  $\rho = \{1, 2, 3, 4\}$ . Generally speaking, for this set up, cohesions favor larger clusters that are spatially connected.

Table S.1: Partition probabilities associated with spatial configuration found in Figure S.1 and for  $M = 1$ .

$\rho$	$C_1(S_h, \mathbf{s}_h^*)$	$C_2(S_h, \mathbf{s}_h^*)$	$C_3(S_h, \mathbf{s}_h^*)$	$C_4(S_h, \mathbf{s}_h^*)$
	$Pr(\rho)$	$Pr(\rho)$	$Pr(\rho)$	$Pr(\rho)$
$\{1, 2, 3, 4\}$	0.085	0.000	0.419	0.188
$\{1, 2, 3\} \{4\}$	0.071	0.000	0.080	0.080
$\{1, 2, 4\} \{3\}$	0.071	0.000	0.080	0.080
$\{1, 3, 4\} \{2\}$	0.071	0.000	0.080	0.080
$\{1\} \{2, 3, 4\}$	0.071	0.000	0.080	0.080
$\{1, 2\} \{3, 4\}$	0.068	0.143	0.034	0.053
$\{1, 3\} \{2, 4\}$	0.068	0.143	0.044	0.079
$\{1, 4\} \{2, 3\}$	0.044	0.000	0.026	0.035
$\{1, 2\} \{3\} \{4\}$	0.068	0.143	0.026	0.054
$\{1, 3\} \{2\} \{4\}$	0.068	0.143	0.026	0.054
$\{1, 4\} \{2\} \{3\}$	0.055	0.000	0.020	0.036
$\{1\} \{2, 3\} \{4\}$	0.055	0.000	0.020	0.036
$\{1\} \{2, 4\} \{3\}$	0.068	0.143	0.026	0.054
$\{1\} \{2\} \{3, 4\}$	0.068	0.143	0.026	0.054
$\{1\} \{2\} \{3\} \{4\}$	0.068	0.143	0.015	0.037

## 2 Posterior Computation

### 2.1 MCMC Implementation

Fitting the models described in the main paper is a fairly straightforward MCMC exercise. The algorithm we employ is based on Neal (2000)'s algorithm number 8 in that it can be divided into two basic pieces. The first updates the partition  $\rho$  via the Polya urn scheme of Blackwell and MacQueen (1973) (and further developed by Quintana 2006) and the other updates the remaining model parameters using a Gibbs sampler Metropolis hybrid (Geman and Geman 1984; Gelfand and Smith 1990 and Metropolis et al. 1953).

To update the cluster membership of location  $i$ , cluster weights are created by comparing the unnormalized posterior for the  $h$ th cluster when location  $i$  is excluded to that when location  $i$  is included. In addition to weights for existing clusters, algorithm 8 of Neal (2000) requires calculating weights for  $p$  empty clusters whose cluster specific parameters are auxiliary variables generated from the prior. To make this more concrete, let  $S_h^{-i}$  denote the  $h$ th cluster and  $k_n^{-i}$  the number of clusters when subject  $i$  is not considered. Similarly  $s_h^{\star-i}$  will denote the vector of locations corresponding to cluster  $h$  when location  $i$  has been removed. Now letting  $f(\cdot|\boldsymbol{\theta}^*, \boldsymbol{\phi})$  denote a likelihood with global parameters  $\boldsymbol{\phi}$  and cluster specific parameters  $\boldsymbol{\theta}^*$ , the multinomial weights associated with the  $k_n^{-i}$  existing clusters and the  $p$  empty clusters are

$$Pr(c_i = h | -) \propto \begin{cases} f(y_i | \boldsymbol{\theta}_h^*, \boldsymbol{\phi}) \frac{C(S_h^{-i} \cup \{i\}, s^{\star-i} \cup s_i)}{C(S_h^{-i}, s^{\star-i})} & \text{for } h = 1, \dots, k_n^{-i} \\ f(y_i | \boldsymbol{\theta}_{new,h}, \boldsymbol{\phi}) C(\{i\}, s_i) & \text{for } h = k_n^{-i} + 1, \dots, k_n^{-i} + p. \end{cases} \quad (\text{S.1})$$

The  $\boldsymbol{\theta}_{new,h}$  are auxiliary variables drawn from their respective priors as required by algorithm 8. Care must be taken when location  $i$  belongs to a singleton cluster as removing the  $i$ th location produces an empty cluster. This in turn requires relabeling the existing cluster specific components to avoid gaps in the cluster labeling. Once  $\rho$  is updated, commonly used techniques in spatial modeling can be employed to update  $(\boldsymbol{\theta}^*, \boldsymbol{\phi})$ . More details are provided in the next section.

### 2.2 Posterior Prediction Distributions

Location dependent predictive distributions are readily available from the sPPM and draws from predictive distribution can be collected within the MCMC algorithm. The procedure

is to first assign location  $\mathbf{s}_{n+1}$  to an existing cluster or a new cluster using the following multinomial weights

$$Pr(c_{n+1} = h | -) \propto \begin{cases} \frac{C(S_h \cup \{n+1\}, \mathbf{s}_h^* \cup \mathbf{s}_{n+1})}{c(S_h, \mathbf{s}_h^*)} & \text{for } h = 1, \dots, k_n \\ C(\{n+1\}, \mathbf{s}_{n+1}) & \text{for } h = k_n + 1. \end{cases} \quad (\text{S.2})$$

Once location  $\mathbf{s}_{n+1}$  is assigned to say the  $\ell$ th cluster,  $y(\mathbf{s}_{n+1})$  can be generated from  $f(y(\mathbf{s}_{n+1}) | \mathbf{y}_\ell, c_{n+1} = \ell)$  (the posterior predictive distribution for cluster  $\ell$ ). For more details see Müller et al. (2011).

## 2.3 Full Conditionals

We considered three models in the main body of the paper to illustrate how seamlessly a variety of likelihoods can be paired with sPPM. As mentioned in the previous section the general algorithms employed to fit all three models are based on Neal (2000)'s algorithm number 8 which divides the MCMC updating scheme into two basic pieces. The first updates the partition  $\rho$ , and then, given  $\rho$ , the remaining model parameters  $(\boldsymbol{\theta}^*, \boldsymbol{\phi})$  are updated. The resulting algorithm is a hybrid of Metropolis and Gibbs sampler updating steps with parameters being updated on an individual basis via full conditionals or by way of a random walk Metropolis step with a Gaussian proposal density. For each of the three models we list full conditionals that are of recognizable form (for sake of clarity we also list the models in their entirety). In what follows  $[\theta | -]$  denotes the distribution of  $\theta$  given all other parameters and data. Also, to simplify notation we drop explicitly denoting objects as functions of  $\mathbf{s}_i$ .

### 2.3.1 Conditional Model with Prior Spatial Structure (CPS)

The conditional model in its entirety is

$$\begin{aligned} y_i | \boldsymbol{\mu}^*, c_i, \boldsymbol{\beta}, \mathbf{x}_i &\stackrel{ind}{\sim} N(\mu_{c_i}^* + \mathbf{x}'_i \boldsymbol{\beta}_1, \sigma^2) \text{ for } i = 1, \dots, n \text{ and } \sigma \sim UN(0, 10) \\ \mu_h^* &\stackrel{iid}{\sim} N(\mu_0, \sigma_0^2) \text{ with } \mu_0 \sim N(m, s^2) \text{ and } \sigma_0 \sim UN(0, 10) \\ \boldsymbol{\beta}_1 &\sim N(\mathbf{b}_0, s_\beta^2 \mathbf{I}) \\ Pr(\rho) &\propto \prod_{h=1}^{k_n} C(S_h, \mathbf{s}_h^*). \end{aligned}$$

Parameters with recognizable full conditionals are listed

$$\begin{aligned} [\beta| -] &\sim N([\sigma^{-2} \sum \mathbf{x}_i \mathbf{x}'_i + s_\beta^{-2} \mathbf{I}]^{-1} [\sigma^{-2} \sum \mathbf{x}_i (y_i - \mu_{c_i}^*) + s_\beta^{-2} \mathbf{b}_0], [\sigma^{-2} \sum \mathbf{x}_i \mathbf{x}'_i + s_\beta^{-2} \mathbf{I}]^{-1}) \\ [\mu_h^*| -] &\sim N([n_h/\sigma^2 + 1/\sigma_0^2]^{-1} [\sigma^{-2} \sum (y_i - \mathbf{x}'_i \boldsymbol{\beta}) + \sigma_0^{-2} \mu_0]), [n_h/\sigma^2 + 1/\sigma_0^2]^{-1}) \\ [\mu_0| -] &\sim N([k_n/\sigma_0^2 + 1/s^2]^{-1} [\sigma_0^{-2} \sum \mu_h^* + ms^{-2}], [k_n/\sigma_0^2 + 1/s^2]^{-1}). \end{aligned}$$

$\sigma^2$  and  $\sigma_0^2$  are updated using Metropolis steps. Updating  $\rho$  was done using (S.1).

### 2.3.2 Joint Model with Prior Spatial Structure (JPS)

The JPS model in hierarchical form is

$$\begin{aligned} \mathbf{y}_i | \boldsymbol{\mu}^*, c_i, \boldsymbol{\Sigma} &\stackrel{ind}{\sim} N_2(\boldsymbol{\mu}_{c_i}^*, \boldsymbol{\Sigma}) \text{ for } i = 1, \dots, n \text{ and } \boldsymbol{\Sigma} \sim IW(2, \mathbf{I}) \\ \mu_h^* | \boldsymbol{\mu}_0, \mathbf{T} &\stackrel{iid}{\sim} N_2(\boldsymbol{\mu}_0, \mathbf{T}) \text{ with } \mathbf{T} \sim IW(2, \mathbf{I}) \\ \boldsymbol{\mu}_0 &\sim N_2(\mathbf{0}, 10^2 \mathbf{I}) \\ Pr(\rho) &\propto \prod_{h=1}^{k_n} C(S_h, \mathbf{s}_h^*) \end{aligned}$$

Parameters with recognizable full conditionals are

$$\begin{aligned} [\boldsymbol{\mu}_h^*| -] &\sim N_2([n_h \boldsymbol{\Sigma}^{-1} + \mathbf{T}^{-1}]^{-1} [\boldsymbol{\Sigma}^{-1} \sum_{i:c_i=h} \mathbf{y}_i + \mathbf{T}^{-1} \boldsymbol{\mu}_0], [n_h \boldsymbol{\Sigma}^{-1} + \mathbf{T}^{-1}]) \\ [\boldsymbol{\mu}_0| -] &\sim N_2([k_n \mathbf{T}^{-1} + 10^{-2} \mathbf{I}]^{-1} [\mathbf{T}^{-1} \sum \boldsymbol{\mu}_h^*], [k_n \boldsymbol{\Sigma}^{-1} + 10^{-2} \mathbf{I}]) \\ [\boldsymbol{\Sigma}| -] &\sim IW(2 + n, \mathbf{I} + \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}_{c_i}^*)(\mathbf{y}_i - \boldsymbol{\mu}_{c_i}^*)') \\ [\mathbf{T}| -] &\sim IW(2 + k_n, \mathbf{I} + \sum_{h=1}^{k_n} (\boldsymbol{\mu}_h^* - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_h^* - \boldsymbol{\mu}_0)') \end{aligned}$$

$\rho$  is updated using (S.1).

### 2.3.3 Joint Model with Likelihood Spatial Structure (JLS)

The JLS model in hierarchical form is

$$\begin{aligned}
\mathbf{y}(\mathbf{s}_i) | \boldsymbol{\mu}^*, c_i, \boldsymbol{\theta}_i, \Sigma &\stackrel{ind}{\sim} N_2(\boldsymbol{\mu}_{c_i}^* + \boldsymbol{\theta}_i, \Sigma) \text{ with } \Sigma \sim IW(\nu_0, \mathbf{S}_0) \\
\begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix} &= \mathbf{A} \begin{pmatrix} \tilde{\theta}_{1i} \\ \tilde{\theta}_{2i} \end{pmatrix} \text{ where } \mathbf{A} = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} \text{ and } \gamma \sim UN(-1, 1) \\
[\tilde{\theta}_{j1}, \dots, \tilde{\theta}_{jn}] | \tau_j^2, \phi_j &\stackrel{ind}{\sim} N_n(\mathbf{0}, \mathbf{K}_j) \text{ with } (\mathbf{K}_j)_{\ell,m} = \tau_j^2 \exp\{-\phi_j ||\mathbf{s}_\ell - \mathbf{s}_m||\} \text{ for } j = 1, 2 \\
\boldsymbol{\mu}_h^* | \boldsymbol{\mu}_0, \mathbf{T} &\stackrel{iid}{\sim} N_2(\boldsymbol{\mu}_0, \mathbf{T}) \text{ for } h = 1, \dots, k_n \text{ and } \mathbf{T} \sim IW(2, \mathbf{I}) \\
\boldsymbol{\mu}_0 &\sim N(\mathbf{0}, 10^2 \mathbf{I}) \\
Pr(\rho) &\propto \prod_{h=1}^{k_n} C(S_h, \mathbf{s}_h^*).
\end{aligned}$$

The last prior specifications are  $\tau_j^2 \sim Gamma(1, 1)$ ,  $\phi_j \sim UN(0.5, 30)$ . We will use  $k_j^{i\ell}$  to denote the  $i, \ell$ th entry of  $K_j^{-1}$  for  $j = 1, 2$ . Now the following full conditionals are derived using straightforward arguments.

$$\begin{aligned}
[\Sigma | -] &\sim IW\left(2 + n, \mathbf{I} + \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}_{c_i}^* - \boldsymbol{\theta}_i)(\mathbf{y}_i - \boldsymbol{\mu}_{c_i}^* - \boldsymbol{\theta}_i)'\right) \\
[\boldsymbol{\mu}_h^* | -] &\sim N_2\left([n_h \Sigma^{-1} + \mathbf{T}^{-1}]^{-1} [\Sigma^{-1} \sum_{i:c_i=h} \{\mathbf{y}(\mathbf{s}_i) - \boldsymbol{\theta}_i\} + \mathbf{T}^{-1} \boldsymbol{\mu}_0], [n_h \Sigma^{-1} + \mathbf{T}^{-1}]^{-1}\right) \\
[\boldsymbol{\mu}_0 | -] &\sim N_2\left([k_n \mathbf{T}^{-1} + 10^{-2} \mathbf{I}]^{-1} [\mathbf{T}^{-1} \sum \boldsymbol{\mu}_h^*], [k_n \Sigma^{-1} + 10^{-2} \mathbf{I}]\right) \\
[\mathbf{T} | -] &\sim IW\left(2 + k_n, \mathbf{I} + \sum_{h=1}^{k_n} (\boldsymbol{\mu}_h^* - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_h^* - \boldsymbol{\mu}_0)'\right) \\
[\boldsymbol{\theta}(\mathbf{s}_i) | -] &\sim N_2\left([\mathbf{A}' \Sigma^{-1} \mathbf{A} + \text{diag}(k_1^{ii}, k_2^{ii})]^{-1} [\mathbf{A}' \Sigma^{-1} (\mathbf{y}(\mathbf{s}_i) - \boldsymbol{\mu}_{c_i}^*) - \mathbf{H}], [\mathbf{A}' \Sigma^{-1} \mathbf{A} + \text{diag}(k_1^{ii}, k_2^{ii})]^{-1}\right),
\end{aligned}$$

where  $\mathbf{H} = (\sum_{\ell \neq i} \theta_{1\ell} k_1^{i\ell}, \sum_{\ell \neq i} \theta_{2\ell} k_2^{i\ell})'$ . Updating  $\rho$  was done using (S.1) and a random walk metropolis step with Gaussian proposals was used to update  $\gamma$ , and  $\tau_j^2, \phi_j$  for  $j = 1, 2$ .

## 3 Selecting and Appropriate $M$

From the simulation study it is fairly evident that selecting an appropriate value for  $M$  in applied modeling is cohesion dependent. When conducting the simulation study found in Section 4 of the main paper,  $M$  values equal to  $10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$  were also

considered. For these values of  $M$  the same general trends found in Tables 3 and 4 of the main paper continue. The results can be found in Table S.2 for the data generating scenario that contains four clusters and Table S.3 for data that contain one cluster. We do not include values for SR as it is not influenced by  $M$ . Values for SSB were not included either as they remained essentially the same as those listed for  $M = 0.01$  in Tables 3 and 4 of the main paper.

The results suggest that  $C_3$  and  $C_4$  hit a lower bound in terms of  $M$ 's influence in partitions fairly quickly ( $M = 10^{-3}$ ). However, for  $C_1$  it appears that MSPE values continue improving until  $M = 10^{-5}$ . To better see this a graph of MSPE values is provided in Figure S.2. We only graph results from data scenario that contains four clusters. In Figure S.2 Data 1 refers to Gaussian error with square clusters, data 2 Gaussian error with irregular clusters, data 3 mixture error with square clusters and data 4 mixture error with irregular clusters. From the figure it is evident that  $C_2$ ,  $C_3$ , and  $C_4$  attain their lowest MSPE values for relatively large values of  $M$ , while  $C_1$  requires smaller  $M$  values. This information guided the  $M$  values used in the Chilean education data analysis.

In addition to taking into account information from the simulation study, to select an appropriate  $M$  value for each cohesion in the data analysis, a cross validation study was carried out. Of the 1215 schools, 600 were randomly selected as training observations and the remaining 615 observations were treated as testing data. For each cohesion the 600 training observations were fit to the data and out of sample MSPE was computed for  $M \in \{1 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, 5 \times 10^{-4}, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5\}$ . Using information gleaned from the simulation study in addition to the cross validation study resulting in setting  $M$  equal to  $5 \times 10^{-5}$ , 0.1, 1.0, and 0.5 for cohesions 1-4 respectively.

## 4 Exploratory Data Analysis of the Chilean Education Data

As an exploratory data analysis, a scatter plot of SIMCE vs. mother's education is provided in Figure S.3. There is clearly a positive, linear relation between the two variables. To explore the spatial dependence in the SIMCE scores, Figure S.3 also contains a variogram plot. The red line corresponds to an estimated variogram based on exponential covariance function. This resulted in a nugget estimate of 0.75, partial sill of 0.25, and asymptotic range of approximately 1.25 displaying the existence of spatial dependence. Here, the `geoR` package (Ribeiro Jr. and Diggle 2001) of R was employed to fit and graph the variogram.

Table S.2: Simulation study results when data are generated with four clusters.

Error	Cluster	Method	$M = 10^{-2}$			$M = 10^{-3}$			$M = 10^{-4}$		
			RAND	LPML	MSPE	RAND	LPML	MSPE	RAND	LPML	MSPE
Gaussian	Square	CPS $C_1$	0.16	-178.07	2.43	0.26	-180.02	2.34	0.37	-182.24	2.29
		CPS $C_2$	0.49	-182.78	2.32	0.50	-182.65	2.35	0.50	-183.31	2.32
		CPS $C_3$	0.43	-184.09	2.41	0.39	-185.88	2.46	0.40	-185.91	2.42
		CPS $C_4$	0.59	-181.58	2.27	0.55	-182.00	2.28	0.51	-183.41	2.28
	Irregular	CPS $C_1$	0.27	-176.76	2.28	0.41	-179.27	2.32	0.50	-181.39	2.33
		CPS $C_2$	0.52	-183.28	2.32	0.53	-185.25	2.44	0.54	-184.66	2.46
		CPS $C_3$	0.57	-182.55	2.30	0.55	-183.59	2.38	0.53	-185.91	2.51
		CPS $C_4$	0.74	-178.99	2.09	0.73	-180.13	2.21	0.72	-181.08	2.26
Mixture	Square	CPS $C_1$	0.16	-176.90	2.36	0.25	-180.71	2.36	0.37	-180.84	2.27
		CPS $C_2$	0.47	-182.74	2.28	0.48	-184.25	2.37	0.49	-183.36	2.34
		CPS $C_3$	0.43	-184.64	2.35	0.43	-185.55	2.46	0.40	-185.82	2.43
		CPS $C_4$	0.57	-181.92	2.21	0.56	-182.90	2.33	0.53	-183.02	2.30
	Irregular	CPS $C_1$	0.27	-176.37	2.27	0.40	-179.89	2.30	0.51	-181.26	2.26
		CPS $C_2$	0.54	-182.95	2.30	0.53	-184.42	2.37	0.54	-184.77	2.38
		CPS $C_3$	0.58	-182.77	2.27	0.53	-184.92	2.41	0.53	-185.10	2.37
		CPS $C_4$	0.77	-178.58	2.07	0.73	-180.15	2.21	0.74	-180.78	2.21
$M = 10^{-5}$			$M = 10^{-6}$			$M = 10^{-7}$					
Gaussian	Square	CPS $C_1$	0.44	-181.08	2.20	0.46	-181.86	2.31	0.45	-182.27	2.31
		CPS $C_2$	0.49	-183.23	2.32	0.49	-183.42	2.39	0.50	-183.90	2.38
		CPS $C_3$	0.38	-185.51	2.36	0.41	-185.37	2.46	0.41	-185.05	2.40
		CPS $C_4$	0.49	-183.03	2.24	0.51	-183.26	2.32	0.52	-182.76	2.29
	Irregular	CPS $C_1$	0.56	-180.31	2.32	0.57	-181.27	2.20	0.58	-179.05	2.27
		CPS $C_2$	0.56	-183.69	2.42	0.54	-185.34	2.36	0.55	-182.97	2.37
		CPS $C_3$	0.52	-184.19	2.44	0.54	-184.35	2.32	0.54	-182.24	2.38
		CPS $C_4$	0.69	-180.35	2.27	0.70	-181.10	2.15	0.71	-178.96	2.20
Mixture	Square	CPS $C_1$	0.45	-181.55	2.32	0.46	-182.10	2.30	0.48	-181.79	2.32
		CPS $C_2$	0.50	-184.79	2.44	0.49	-184.43	2.39	0.49	-183.51	2.43
		CPS $C_3$	0.43	-185.17	2.46	0.43	-186.15	2.39	0.40	-184.94	2.46
		CPS $C_4$	0.53	-182.99	2.35	0.52	-183.94	2.32	0.52	-182.84	2.36
	Irregular	CPS $C_1$	0.57	-181.22	2.26	0.57	-178.62	2.22	0.58	-179.36	2.24
		CPS $C_2$	0.54	-184.81	2.40	0.54	-182.07	2.33	0.54	-182.84	2.37
		CPS $C_3$	0.52	-184.94	2.40	0.52	-181.89	2.31	0.53	-182.33	2.35
		CPS $C_4$	0.71	-180.47	2.20	0.70	-178.20	2.15	0.72	-178.73	2.19

Table S.3: Simulation study results when data are generated with one cluster.

Error	Cluster	Method	$M = 10^{-2}$			$M = 10^{-3}$			$M = 10^{-4}$		
			RAND	LPML	MSPE	RAND	LPML	MSPE	RAND	LPML	MSPE
Gaussian	Square	CPS $C_1$	0.00	-178.16	2.03	0.00	-176.28	2.03	0.00	-177.69	2.04
		CPS $C_2$	0.00	-178.63	2.03	0.00	-176.80	2.03	0.00	-178.33	2.04
		CPS $C_3$	0.81	-180.67	2.12	0.83	-177.92	2.10	0.81	-179.95	2.12
		CPS $C_4$	0.07	-179.25	2.07	0.15	-177.39	2.07	0.04	-178.81	2.08
	Irregular	CPS $C_1$	0.00	-175.32	1.91	0.00	-174.69	1.97	0.00	-173.37	1.91
		CPS $C_2$	0.00	-175.85	1.92	0.00	-175.70	1.98	0.00	-174.57	1.93
		CPS $C_3$	0.05	-177.26	1.98	0.04	-176.94	2.04	0.01	-175.72	2.00
		CPS $C_4$	0.00	-176.24	1.94	0.00	-175.82	2.00	0.00	-174.18	1.94
Mixture	Square	CPS $C_1$	0.00	-178.42	2.02	0.00	-178.39	2.05	0.00	-176.85	2.05
		CPS $C_2$	0.00	-178.77	2.03	0.00	-178.57	2.06	0.00	-177.35	2.06
		CPS $C_3$	0.91	-180.53	2.12	0.83	-180.24	2.11	0.87	-178.78	2.11
		CPS $C_4$	0.09	-179.70	2.09	0.09	-178.88	2.08	0.06	-178.01	2.09
	Irregular	CPS $C_1$	0.00	-175.37	1.86	0.00	-174.74	1.94	0.00	-175.68	1.94
		CPS $C_2$	0.00	-176.63	1.89	0.00	-175.63	1.96	0.00	-176.90	1.96
		CPS $C_3$	0.01	-177.54	1.93	0.02	-176.65	2.02	0.02	-177.41	2.00
		CPS $C_4$	0.00	-176.08	1.88	0.00	-175.89	1.97	0.00	-176.76	1.97
$M = 10^{-5}$			$M = 10^{-6}$			$M = 10^{-7}$					
Gaussian	Square	CPS $C_1$	0.00	-176.98	2.02	0.00	-175.82	2.02	0.00	-174.82	2.08
		CPS $C_2$	0.00	-177.88	2.03	0.00	-177.46	2.02	0.00	-177.98	2.06
		CPS $C_3$	0.81	-179.16	2.09	0.81	-178.44	2.07	0.73	-179.18	2.15
		CPS $C_4$	0.08	-178.21	2.05	0.03	-177.52	2.03	0.00	-178.15	2.07
	Irregular	CPS $C_1$	0.00	-175.50	1.99	0.00	-173.31	1.96	0.00	-174.71	1.95
		CPS $C_2$	0.00	-176.61	2.03	0.00	-175.46	1.97	0.00	-176.82	1.95
		CPS $C_3$	0.02	-177.65	2.06	0.01	-176.61	2.02	0.01	-177.61	2.01
		CPS $C_4$	0.00	-176.63	2.02	0.00	-175.52	1.97	0.00	-176.12	1.95
Mixture	Square	CPS $C_1$	0.00	-176.96	2.05	0.00	-177.12	2.02	0.00	-176.83	2.01
		CPS $C_2$	0.00	-177.61	2.05	0.00	-177.96	2.02	0.00	-178.46	1.99
		CPS $C_3$	0.80	-179.19	2.13	0.76	-179.43	2.08	0.70	-179.95	2.06
		CPS $C_4$	0.08	-178.29	2.09	0.01	-178.57	2.04	0.00	-178.80	1.99
	Irregular	CPS $C_1$	0.00	-173.33	1.89	0.00	-173.05	2.02	0.00	-172.75	1.96
		CPS $C_2$	0.00	-174.33	1.90	0.00	-174.08	2.02	0.00	-175.64	1.98
		CPS $C_3$	0.01	-175.70	1.95	0.00	-175.19	2.10	0.01	-176.14	2.01
		CPS $C_4$	0.00	-174.56	1.91	0.00	-173.75	2.04	0.00	-175.09	1.97

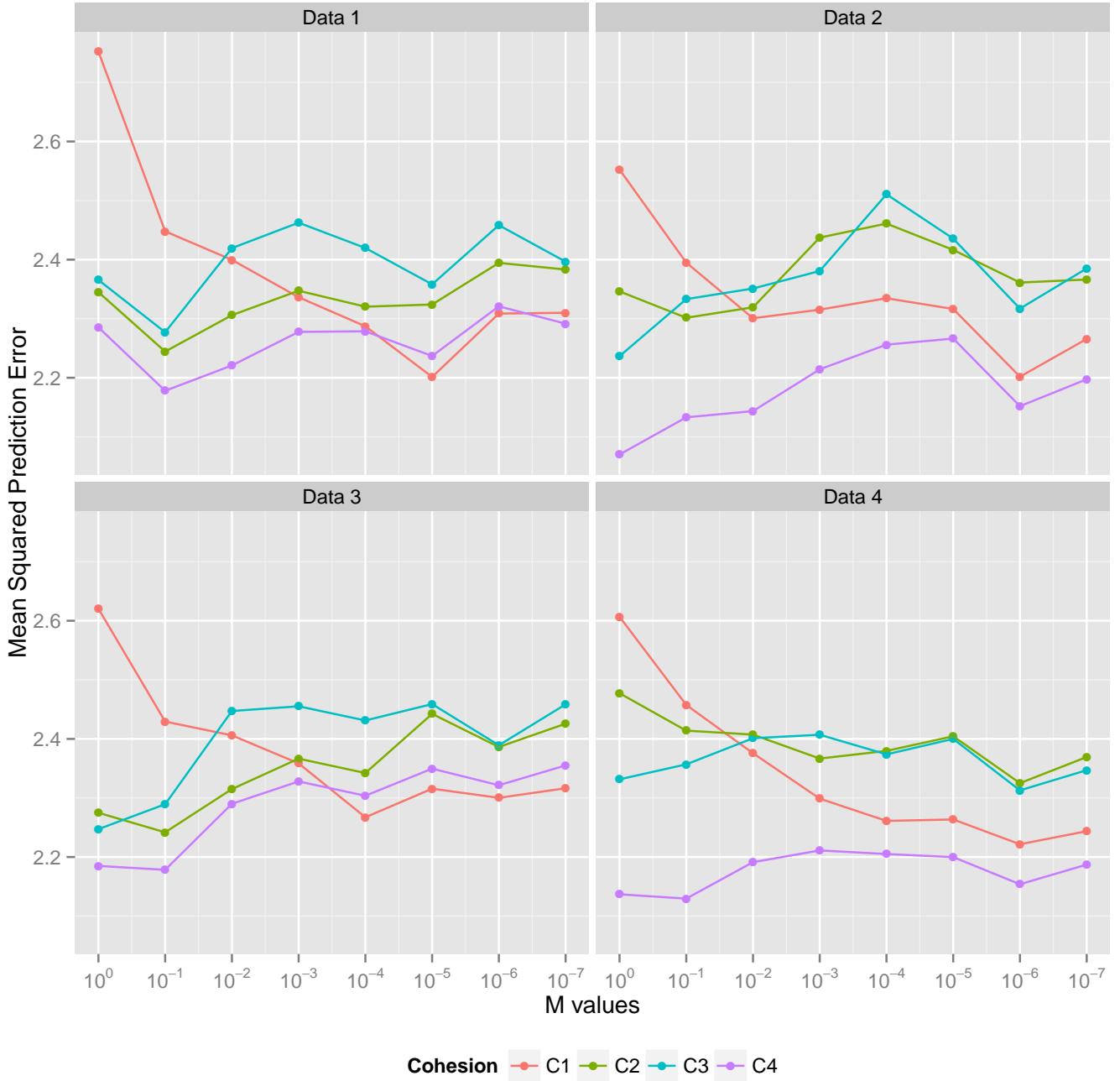


Figure S.2: MSPE values averaged over 100 datasets for a range of  $M$  values.

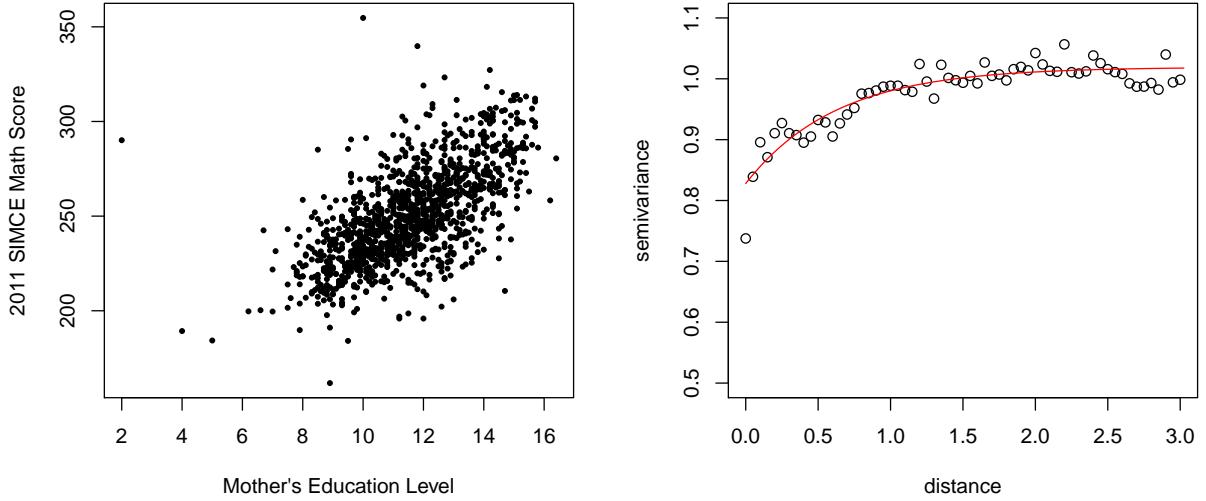


Figure S.3: The left figure corresponds to scatterplot between average 2011 Math SIMCE score vs. average mother’s education level. The right figure corresponds to an empirical exponential semivariogram fit for SIMCE math scores.

## 5 $\rho$ Estimates

An estimated partition of the 600 training observations is provided for each of the four cohesions and the three models in the body of the paper. To estimate  $\rho$  we employed the minimum least squares criteria of Dahl (2006). Figure S.4 corresponds to the CPS model, Figure S.6 to the JPS and Figure S.5 to the JLS model. For each of the three models the estimated number of clusters for each cohesion is relatively similar. Cohesions 1 and 2 produce about 40 clusters, cohesion 3 produces approximately 8 and the 4th cohesion produces 30 clusters. It appears that cohesion 3 produces a partition with fairly large clusters that have hard boundaries. Cohesion 1’s clusters are more local, but do not follow a hard boundary. All the estimated partitions display a spatial structure in the sense that clusters are all generally local.

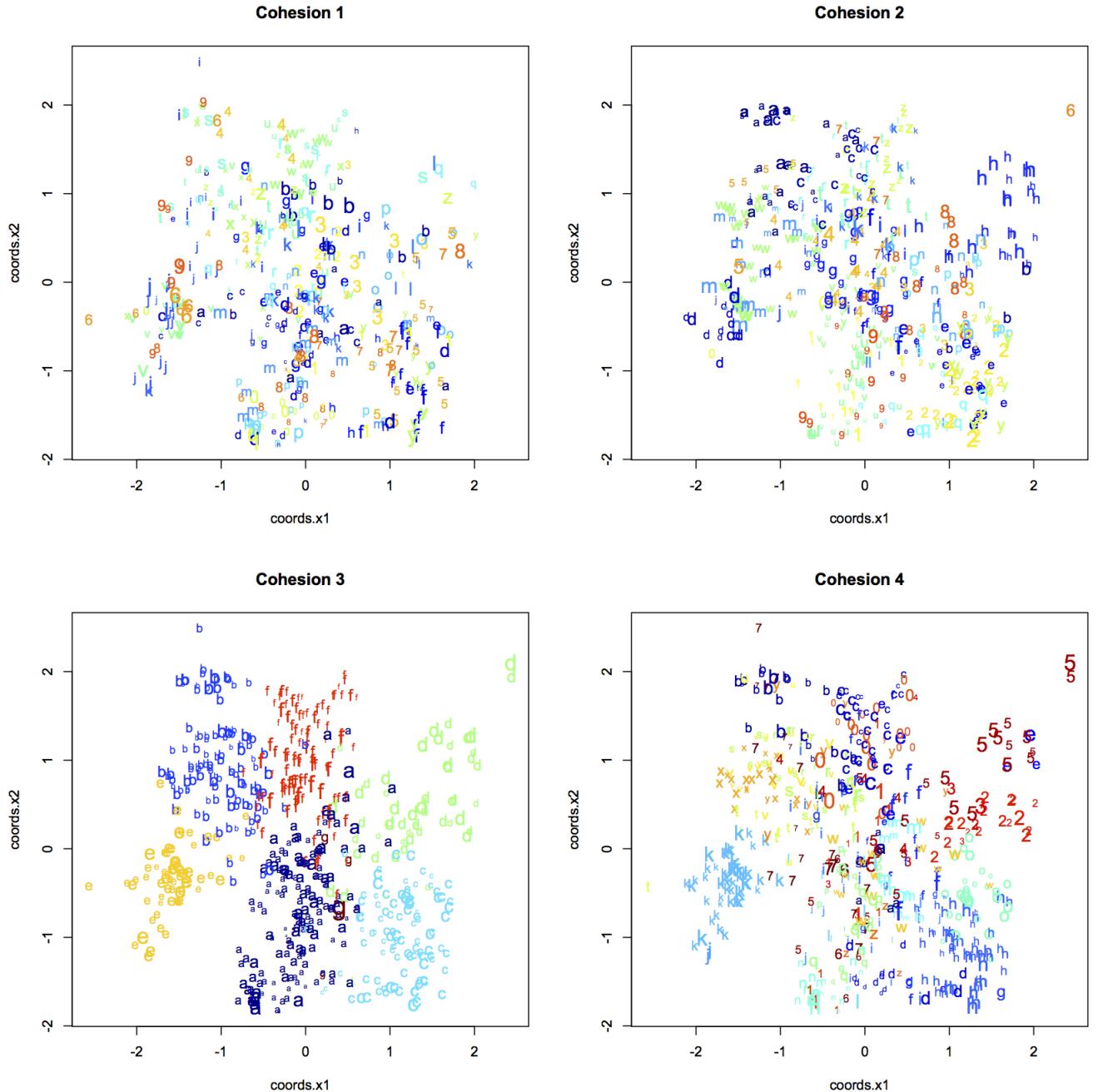


Figure S.4: Estimated partition using Dahl's method for CPS.

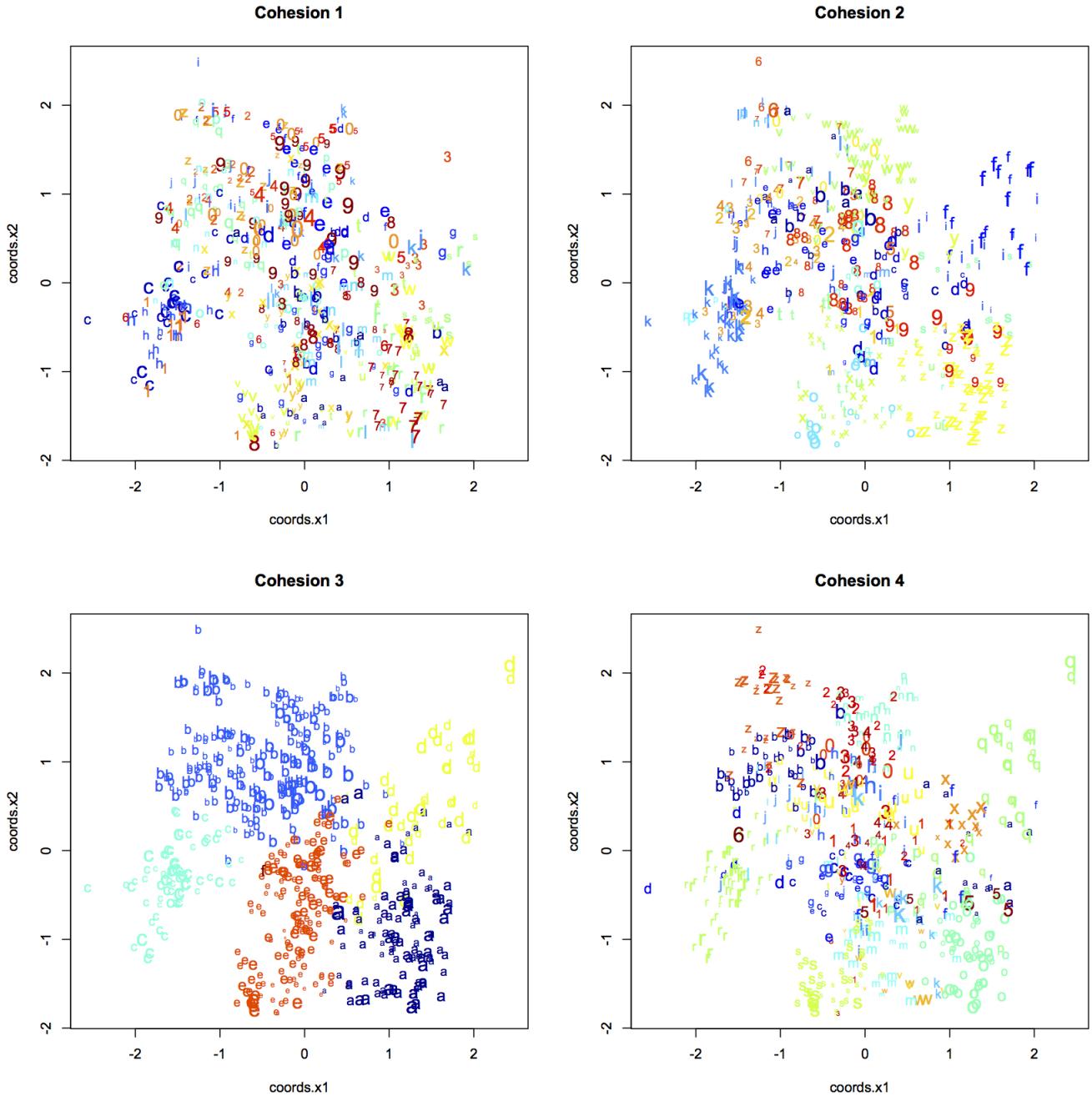


Figure S.5: Estimated partition using Dahl's method for JLS.

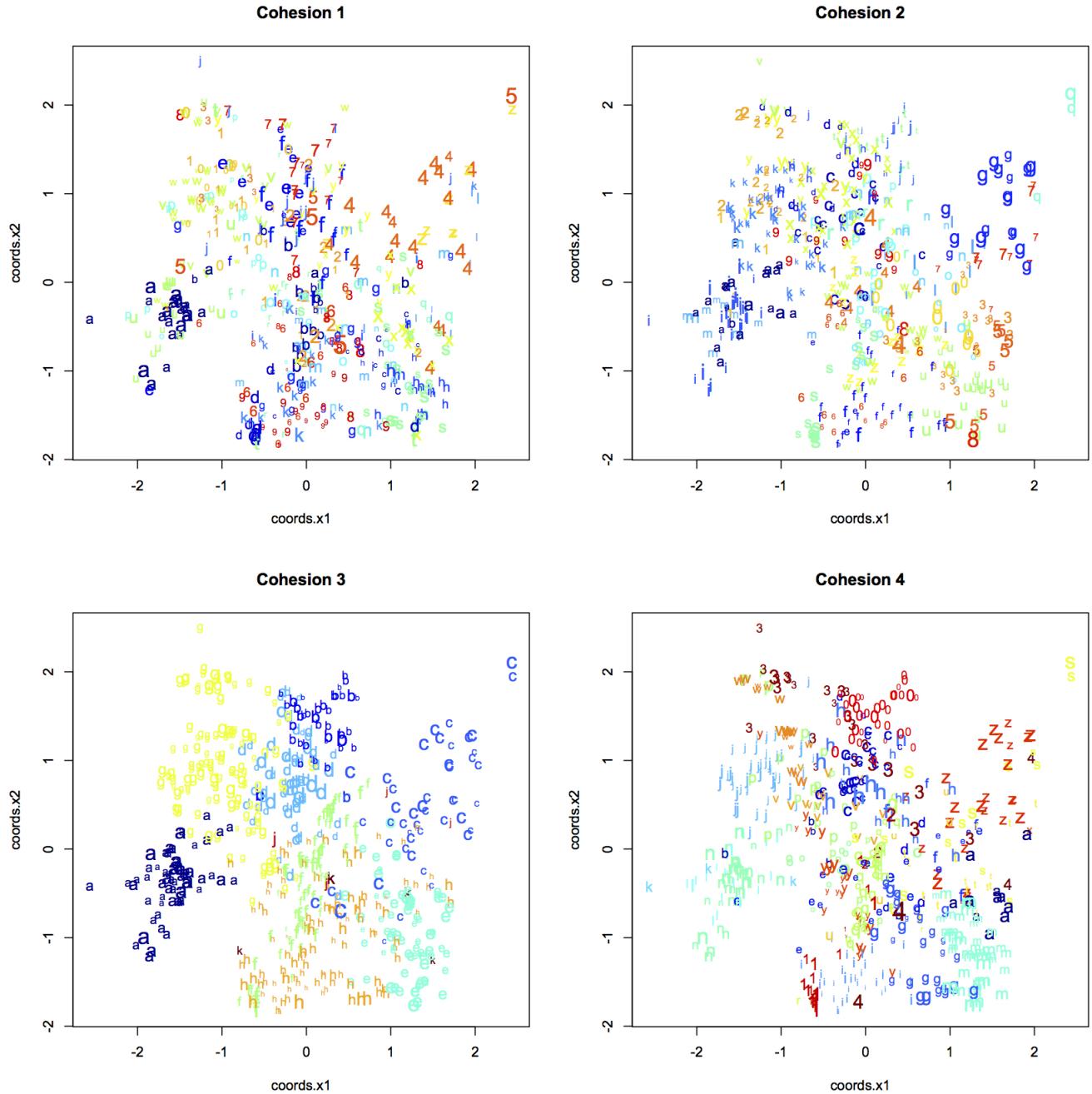


Figure S.6: Estimated partition using Dahl's method for JPS.

## References

- Blackwell, D. and MacQueen, J. B. (1973). “Ferguson Distributions via Pólya Urn Schemes.” *The Annals of Statistics*, 1: 353–355.
- Dahl, D. B. (2006). “Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model.” In Vannucci, M., Do, K. A., and Müller, P. (eds.), *Bayesian Inference for Gene Expression and Proteomics*, 201–218. Cambridge University Press.
- Gelfand, A. E. and Smith, A. F. M. (1990). “Sampling-Based Approaches to Calculating Marginal Densities.” *Journal of the American Statistical Association*, 85: 398–409.
- Geman, S. and Geman, D. (1984). “Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis of Machine Intelligence*, 6: 721–741.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). “Equations of State Calculations by Fast Computing Machines.” *Journal of Chemical Physics*, 21: 1087–1091.
- Müller, P., Quintana, F., and Rosner, G. L. (2011). “A Product Partition Model With Regression on Covariates.” *Journal of Computational and Graphical Statistics*, 20(1): 260–277.
- Neal, R. M. (2000). “Markov Chain Sampling Methods for Dirichlet Process Mixture Models.” *Journal of Computational and Graphical Statistics*, 9: 249–265.
- Quintana, F. A. (2006). “A Predictive View of Bayesian Clustering.” *Journal of Statistical Planning and Inference*, 136: 2407–2429.
- Ribeiro Jr., P. J. and Diggle, P. J. (2001). “geoR: a package for geostatistical analysis.” *R-NEWS*, 1(2): 14–18. ISSN 1609-3631.  
URL <http://CRAN.R-project.org/doc/Rnews/>