D.4. *Simulation 3 - Cluster Overspecification Results*

D.4.1. $L = 3, \sigma^2 = 2$

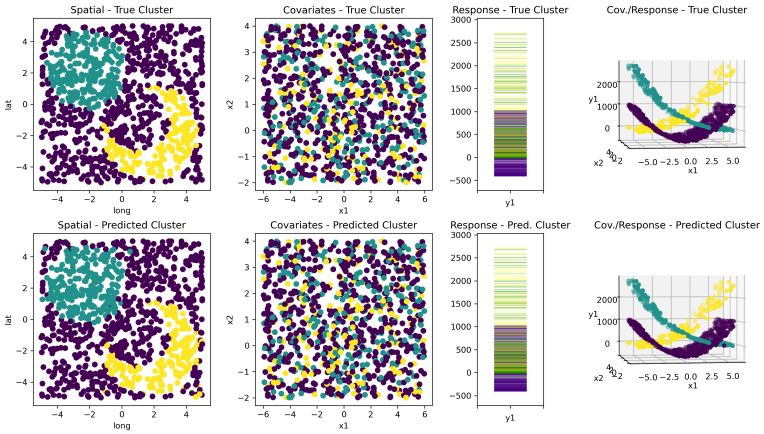
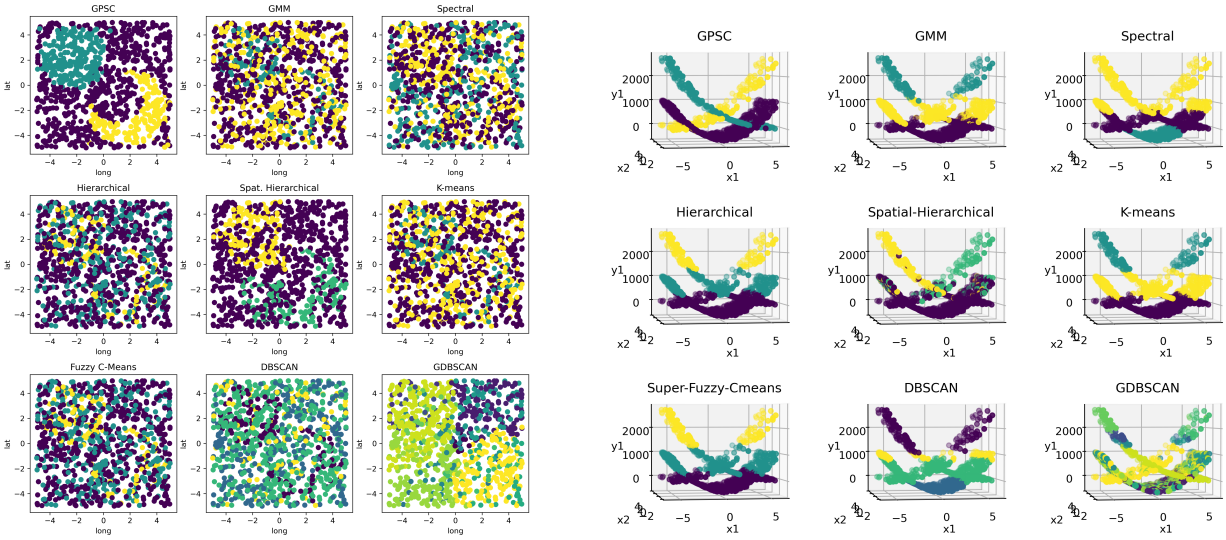


Fig. 19: GPSC results for for Simulation 3, $L = 3, \sigma^2 = 2$, colored by cluster and separated by data domain as in previous simulation. The first row indicates ground truth with results from GPSC in the second.



(a) Spatial Domain (b) XY Domain

Fig. 20: Comparisons to spatial and supervised clustering algorithms for Simulation 3, $L = 3, \sigma^2 = 2$.

Table 7: Adjusted Rand index and adjusted mutual information of different methods against the true labels for for Simulation 3, $L = 3, \sigma^2 = 2$, replicated over 50 random seeds reported as mean ± standard deviation.

METHOD	ARI	AMI	METHOD	ARI	AMI
GPSC	$\mathbf{0.72 \pm 0.27}$	$\mathbf{0.70 \pm 0.24}$	GMM	0.16 ± 0.02	0.14 ± 0.02
K-MEANS	0.17 ± 0.02	0.13 ± 0.01	C-MEANS	0.16 ± 0.02	0.13 ± 0.01
HIER.	0.17 ± 0.03	0.13 ± 0.03	SPAT. HIER.	0.16 ± 0.11	0.17 ± 0.08
DBSCAN	0.22 ± 0.04	0.15 ± 0.03	GDBSCAN	0.10 ± 0.02	0.24 ± 0.04
SPECTRAL	0.08 ± 0.02	0.16 ± 0.02			

D.4.2. $L = 4, \sigma^2 = 2$

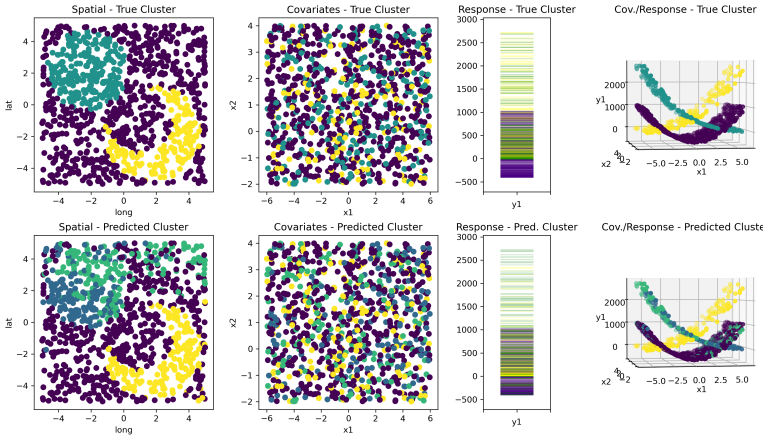
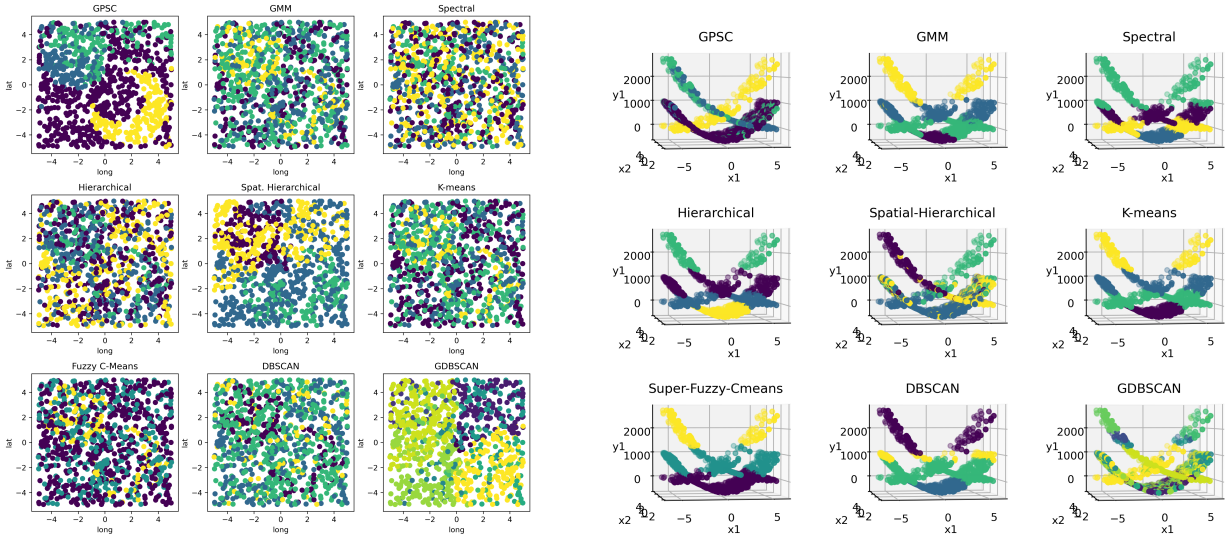


Fig. 21: GPSC results for for Simulation 3, $L = 4, \sigma^2 = 2$, colored by cluster and separated by data domain as in previous simulation. The first row indicates ground truth with results from GPSC in the second.



(a) Spatial Domain (b) XY Domain

Fig. 22: GPSC and comparisons to spatial clustering and supervised clustering algorithms, $L = 4, \sigma^2 = 2$.

Table 8: Adjusted Rand index and adjusted mutual information of different methods against the true labels for for Simulation 3, $L = 4, \sigma^2 = 2$, replicated over 50 random seeds reported as mean ± standard deviation.

METHOD	ARI	AMI	METHOD	ARI	AMI
GPSC	$\mathbf{0.58 \pm 0.14}$	$\mathbf{0.69 \pm 0.11}$	GMM	0.08 ± 0.01	0.18 ± 0.02
K-MEANS	0.15 ± 0.02	0.22 ± 0.02	C-MEANS	0.16 ± 0.02	0.13 ± 0.01
HIER.	0.14 ± 0.03	0.19 ± 0.04	SPAT. HIER.	0.13 ± 0.10	0.20 ± 0.07
DBSCAN	0.22 ± 0.04	0.15 ± 0.03	GDBSCAN	0.10 ± 0.02	0.24 ± 0.04
SPECTRAL	0.09 ± 0.01	0.14 ± 0.15			

D.4.3. $L = 5, \sigma^2 = 2$

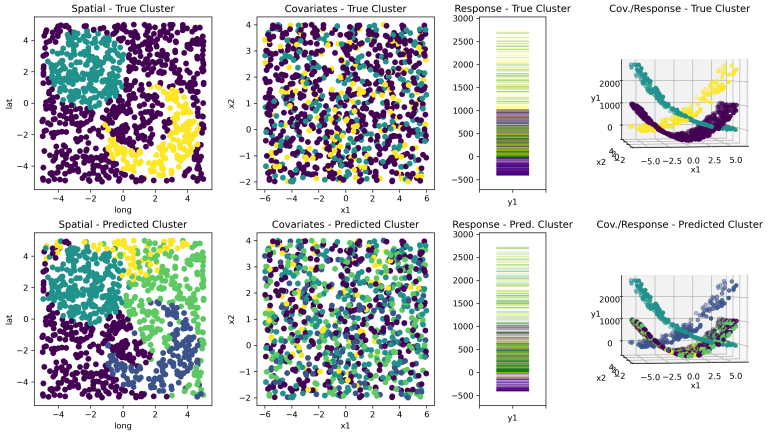
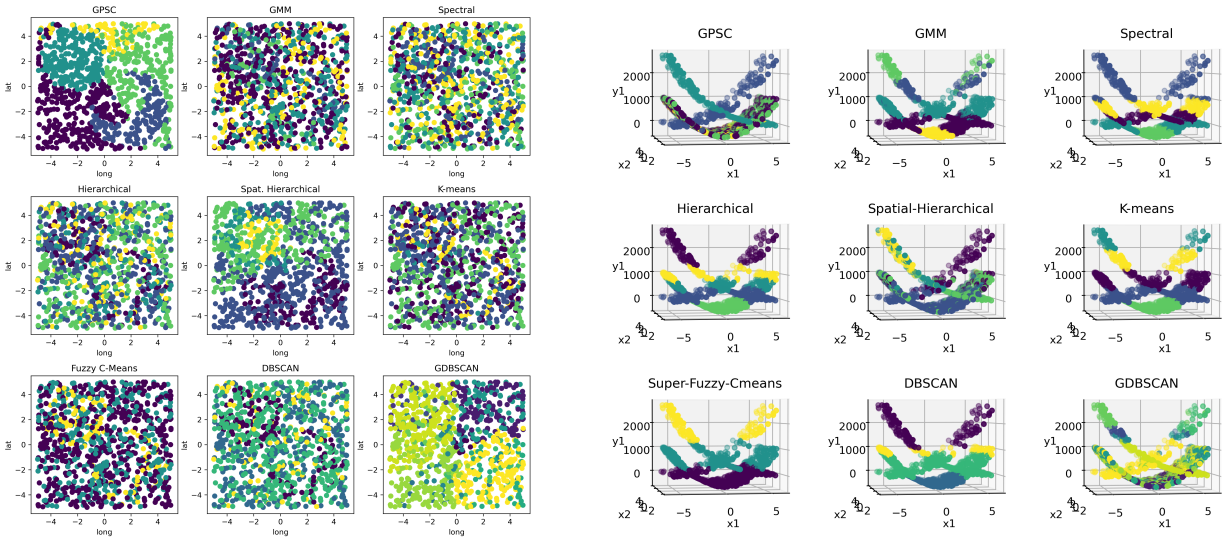


Fig. 23: GPSC results for for Simulation 3, $L = 5, \sigma^2 = 2$, colored by cluster and separated by data domain as in previous simulation. The first row indicates ground truth with results from GPSC in the second.



(a) Spatial Domain (b) XY Domain

Fig. 24: GPSC and comparisons to spatial clustering and supervised clustering algorithms, $L = 5, \sigma^2 = 2$.

Table 9: Adjusted Rand index and adjusted mutual information of different methods against the true labels for for Simulation 3, $L = 5, \sigma^2 = 2$, replicated over 50 random seeds reported as mean ± standard deviation.

METHOD	ARI	AMI	METHOD	ARI	AMI
GPSC	$\mathbf{0.47 \pm 0.08}$	$\mathbf{0.65 \pm 0.09}$	GMM	0.10 ± 0.04	0.20 ± 0.06
K-MEANS	0.15 ± 0.02	0.20 ± 0.02	C-MEANS	0.16 ± 0.02	0.13 ± 0.01
HIER.	0.14 ± 0.03	0.20 ± 0.03	SPAT. HIER.	0.13 ± 0.09	0.22 ± 0.06
DBSCAN	0.22 ± 0.04	0.15 ± 0.03	GDBSCAN	0.10 ± 0.02	0.24 ± 0.04
SPECTRAL	0.09 ± 0.01	0.14 ± 0.02			

D.4.4. $L = 6, \sigma^2 = 2$

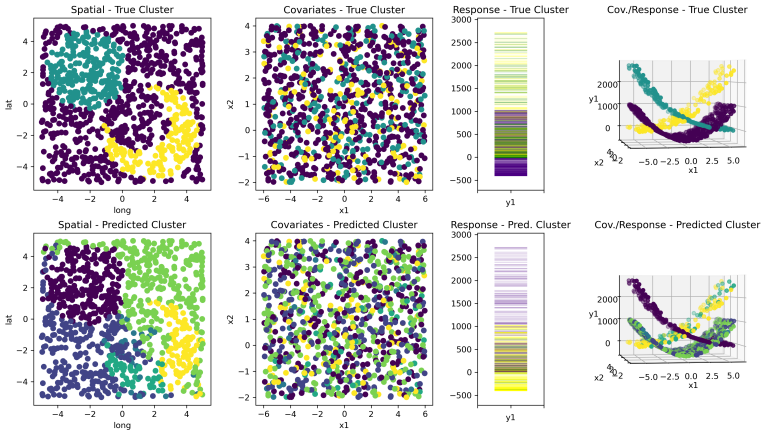
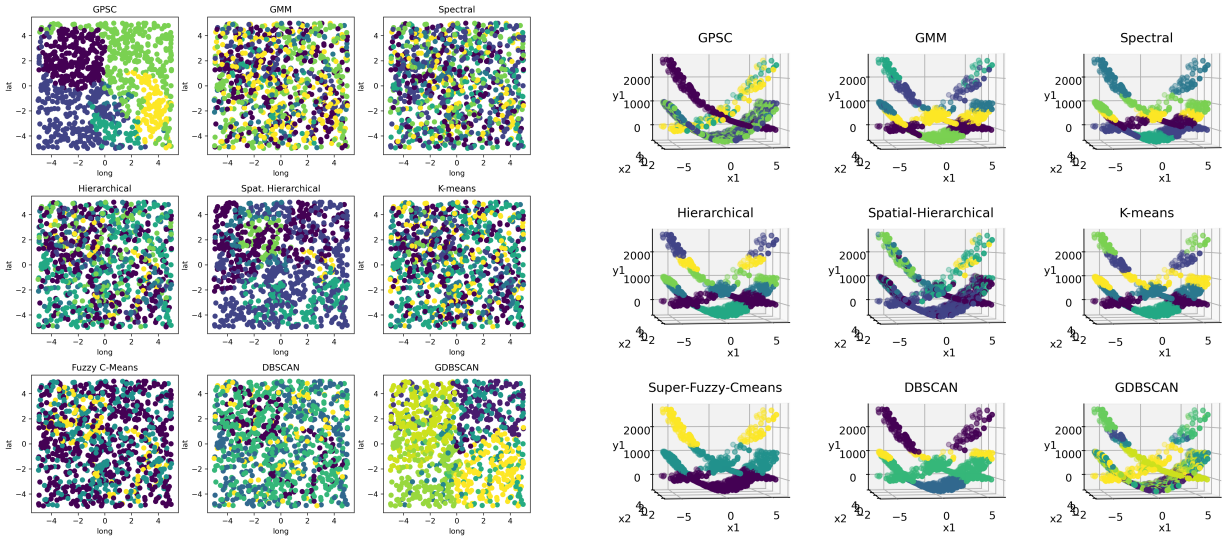


Fig. 25: GPSC results for for Simulation 3, $L = 6, \sigma^2 = 2$, colored by cluster and separated by data domain as in previous simulation. The first row indicates ground truth with results from GPSC in the second.



(a) Spatial Domain (b) XY Domain

Fig. 26: GPSC and comparisons to spatial clustering and supervised clustering algorithms for Simulation 3, $L = 6, \sigma^2 = 2$.

Table 10: Adjusted Rand index and adjusted mutual information of different methods against the true labels for for Simulation 3, $L = 6, \sigma^2 = 2$, replicated over 50 random seeds reported as mean ± standard deviation.

METHOD	ARI	AMI	METHOD	ARI	AMI
GPSC	$\mathbf{0.43 \pm 0.05}$	$\mathbf{0.63 \pm 0.04}$	GMM	0.10 ± 0.03	0.24 ± 0.04
K-MEANS	0.14 ± 0.02	0.19 ± 0.01	C-MEANS	0.16 ± 0.16	0.13 ± 0.01
HIER.	0.14 ± 0.04	0.19 ± 0.02	SPAT. HIER.	0.13 ± 0.08	0.23 ± 0.06
DBSCAN	0.22 ± 0.04	0.15 ± 0.03	GDBSCAN	0.10 ± 0.02	0.24 ± 0.04
SPECTRAL	0.08 ± 0.01	0.18 ± 0.02			

D.5. *Simulation 3 - Functions of s and x*

The set up here is exactly as in Simulation 3, however, the functions are now functions of both the spatial domain and the covariate domain. It can be see that GPSC is still able to recover the true clusters under these conditions. Exact implementation and final parameters can be found in the submitted code.

For each cluster, y is generated as a function of just x_i with independent Gaussian distributed noise $\epsilon \sim N(0, 2)$. For cluster 1, the true nonlinear function is:

$$y = 10(s_1 + s_2)^2 + 40(x_1)^2 + (x_2)^2 - 500 + \epsilon.$$

For cluster 2, the true nonlinear function is:

$$y = 10(s_2)^2 - (x_1 - 8)^3 + (x_2)^3 + \epsilon.$$

For cluster 3, the true nonlinear function is:

$$y = -10(s_1)^2 + (x_1 + 8)^3 + (x_2)^3 - 20 + \epsilon.$$

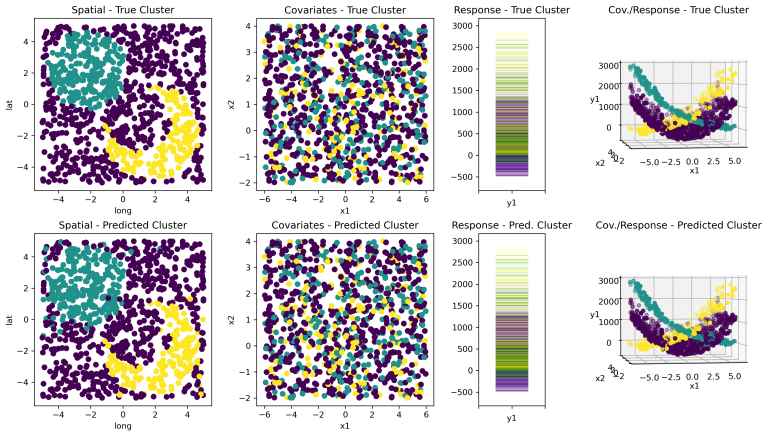
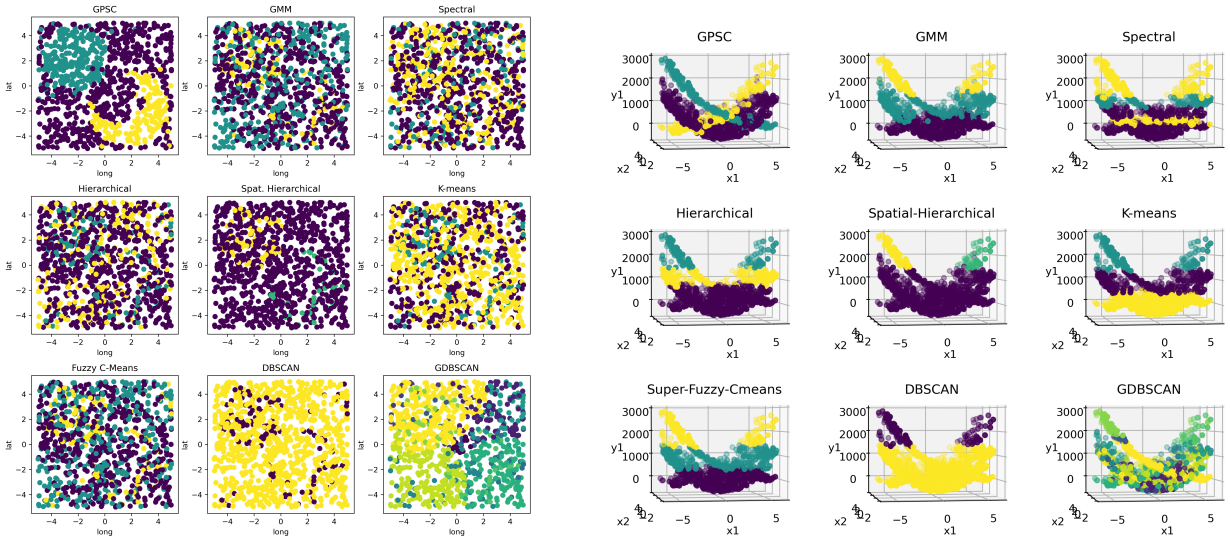


Fig. 27: GPSC results for Simulation 3 analog with functions of both s and x, colored by cluster and separated by data domain as in previous simulation. The first row indicates ground truth with results from GPSC in the second.

As can be seen, GPSC performs well regardless of whether the functional relationships are based on s, x or s and x. Regardless of the which case the true functional relationship is in, the full vector (s, x, y) is used as input and GPSC is able to accurately recover the clusters.



(a) Spatial Domain (b) XY Domain

Fig. 28: GPSC and comparisons to spatial clustering and supervised clustering algorithms for Simulation 3 analog with functions of both s and x.

Table 11: Adjusted Rand index and adjusted mutual information of different methods against the true labels for Simulation 3 analog with functions of both s and x, replicated over 50 random seeds reported as mean ± standard deviation.

METHOD	ARI	AMI	METHOD	ARI	AMI
GPSC	$\mathbf{0.69 \pm 0.28}$	$\mathbf{0.65 \pm 0.25}$	GMM	0.09 ± 0.02	0.09 ± 0.02
K-MEANS	0.08 ± 0.01	0.05 ± 0.01	C-MEANS	0.08 ± 0.01	0.05 ± 0.01
HIER.	0.08 ± 0.02	0.05 ± 0.02	SPAT. HIER.	0.10 ± 0.06	0.10 ± 0.07
DBSCAN	0.15 ± 0.02	0.09 ± 0.02	GDBSCAN	0.09 ± 0.02	0.21 ± 0.03
SPECTRAL	0.01 ± 0.05	0.03 ± 0.02			