POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

# Bayesian distance-based clustering of areal data

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

**Author:** Filippo Galli

**Advisor:** Prof. Alessandra Guglielmi

**Co-advisor:** Matteo Gianella

**Academic year:** 2025-2026

## 1. Introduction

The growing availability of large-scale spatial data across government, environmental, and commercial domains has created a pressing need for statistical methods that are both theoretically principled and computationally feasible. In the Bayesian framework, inference via Markov chain Monte Carlo (MCMC) methods often becomes prohibitively expensive as data sizes increase, with computational complexity frequently scaling super-linearly with sample size. This limitation restricts the applicability of state-of-the-art approaches to real-world problems where timely decision-making is essential and datasets may encompass millions of observations across thousands of spatial units.

The specific challenge addressed in this thesis is the clustering of areal data—spatial data aggregated over geographic regions such as census tracts, municipalities, or administrative districts. Unlike point-referenced data where observations occur at precise locations, areal data represent summaries over regions, introducing unique methodological challenges regarding spatial dependence, boundary detection, and the analysis of distributional outcomes rather than scalar summaries. The motivation arises from a concrete policy application: identifying spatial patterns in socioeconomic data, specifically income distributions across Public Use Microdata Areas (PUMAs) and Italian municipalities, to inform geographically targeted interventions. Understanding these patterns enables policymakers to identify regions with similar economic profiles, detect spatial inequalities, and design resource allocation strategies that respect geographic coherence.

The computational barrier becomes evident when examining current Bayesian methodology for spatial areal data. The SPMIX framework of Gianella, Beraha, et al. (2023), which represents a current standard for spatial boundary detection, employs a sophisticated finite Gaussian mixture model with logistic multivariate conditionally autoregressive (CAR) priors on mixture weights to induce spatial dependencies. While statistically rigorous, this approach requires approximately ten hours of computation for 80,000 individual income observations aggregated across 93 PUMAs in the Los Angeles area on a modern desktop machine equipped with an Intel i7-1255U processor and 32GB RAM. Such computational demands preclude repeated analyses under different model specifications, comprehensive sensitivity studies, or scaling to larger geographic regions such as statewide or national analyses. These constraints effectively

prevent applied researchers from rapidly exploring alternative scenarios or validating findings across multiple datasets.

This thesis proposes a scalable alternative that bridges statistical methodology and practical feasibility. The core innovation is to transform the inferential problem from individual-level observations to area-specific distributional representations. Rather than modeling 80,000 individual records, we compute kernel density estimations (KDE) for each areal unit, reducing the problem to 93 density objects and their pairwise distances. This aggregation preserves essential distributional information—capturing shape, skewness, and tail behavior of regional income distributions—while dramatically reducing dimensionality. Clustering is then performed on these distances using a Bayesian nonparametric framework that combines the distance-based cohesion-repulsion likelihood of Natarajan et al. (2024) with a partition prior induced by the Normalized Generalized Gamma (NGG) process, extended with spatial adjacency and covariate information. The computational gains are substantial: inference completes in under two seconds compared to ten hours for existing approaches, enabling interactive exploratory analysis and scaling to thousands of areal units. An efficient MCMC scheme synthesizing several recent algorithmic advances—including the Sequential Allocated Merge-Split sampler, distance-based Locality-Sensitive Sampling, and Smart-Dumb-Dumb-Smart move-type selection—ensures rapid convergence and effective exploration of the partition space. All code is publicly available as a modular, high-performance C++ library with R bindings at github.com/Filippo-Galli/BNPClust, facilitating reproducibility and extension by other researchers.

## 2. Model specification

The proposed framework is an areal Product Partition Model with Covariates (aPPMx) where posterior inference on the partition of areal units is obtained from pairwise distances between area-specific kernel density estimates of income distributions. This section details the model components: the distance-based likelihood capturing cohesion and repulsion, the NGG-induced partition prior enabling data-driven inference on the number of clusters, extensions incorporating spatial and covariate information, and the efficient MCMC sampling scheme.

### 2.1. Distance-based likelihood

The likelihood draws upon the cohesion–repulsion framework of Natarajan et al. (2024), which extends the partial likelihood paradigm of Duan et al. (2021) with an inter-cluster repulsion term that penalizes small distances between observations assigned to different clusters. This dual structure addresses a fundamental identifiability limitation in cohesion-only approaches: without repulsion, the likelihood provides limited information for determining the number of clusters, as adding clusters always improves within-cluster cohesion.

Formally, the likelihood is specified as:

$$\mathcal{L}(D \mid \boldsymbol{\theta}, \boldsymbol{\lambda}, \rho_n) = \prod_{k=1}^{K} \prod_{\substack{i,j \in C_k \\ i<j}} f(D_{ij} \mid \lambda_k) \cdot \prod_{(k,t)\in A} \prod_{\substack{i \in C_k \\ j \in C_t}} g(D_{ij} \mid \theta_{kt}), \tag{1}$$

where $D = \{D_{ij}\}$ is the $n \times n$ pairwise distance matrix derived from the $n$ areal kernel density estimates, $\rho_n = \{C_1, \ldots, C_K\}$ is a partition of $n$ areal units into $K$ clusters, and $A = \{(k,t) : 1 \le k < t \le K\}$ enumerates all distinct cluster pairs. The first product constitutes the Likelihood Cohesion Factor (LCF), promoting compact clusters by favoring small intra-cluster distances. The second product is the Likelihood Repulsion Factor (LRF), enforcing separation by penalizing small inter-cluster distances.

For computational convenience and interpretability, we specify $f$ and $g$ as gamma distributions:

$$f(d \mid \lambda_k) = \text{Gamma}(\delta_1, \lambda_k), \tag{2}$$

$$g(d \mid \theta_{kt}) = \text{Gamma}(\delta_2, \theta_{kt}), \tag{3}$$

where $\delta_1, \delta_2 > 0$ are fixed shape parameters encoding the clustering objective. Setting $\delta_1 < 1$ produces a decreasing density that encourages small intra-cluster distances, while $\delta_2 > 1$ produces an increasing density favoring larger inter-cluster distances. This asymmetry naturally encodes the clustering objective without requiring hard constraints.

Conjugate gamma priors on the rate parameters, $\lambda_k \sim \text{Gamma}(\alpha, \beta)$ and $\theta_{kt} \sim \text{Gamma}(\zeta, \gamma)$, enable analytical marginalization over these parameters. This yields an integrated likelihood depending only on the distance matrix $D$ and the partition $\rho_n$, substantially simplifying posterior computation and avoiding expensive parameter sampling during MCMC iterations.

## 2.2. NGGP-induced partition prior

To enable clustering without pre-specifying the number of components, we employ a normalized generalized gamma process (NGGP) prior on the random partition. The NGGP generalizes the Dirichlet Process through a discount parameter $\sigma \in (0, 1)$, offering greater flexibility in controlling the variance of the cluster count and the "rich-get-richer" behavior where large clusters are increasingly favored to grow.
The exchangeable partition probability function (EPPF) induced by an NGGP sample is:

$$P[\rho_n, U \in du] = \frac{a^{|\rho_n|} u^{n-1}}{\Gamma(n)(u+\tau)^{n-\sigma|\rho_n|}} \cdot \mathrm{e}^{-(a/\sigma)[(u+\tau)^\sigma - \tau^\sigma]}\, du \cdot \prod_{c \in \rho_n} \frac{\Gamma(|c| - \sigma)}{\Gamma(1-\sigma)}, \quad (4)$$

where $a > 0$ is a mass parameter controlling the expected number of clusters, $|\rho_n|$ is the number of clusters, $\sigma \in (0, 1)$ governs the variance of the cluster count and moderates the rich-get-richer effect, $\tau > 0$ is a rate parameter, and $|c|$ is the number of areas in the cluster $c$. The auxiliary variable $U$, which arises from the posterior characterization of normalized random measures (James et al. 2009), acts as a random concentration parameter facilitating tractable inference. The Dirichlet Process emerges as a special case when $\sigma \to 0$ and $\tau \to 1$, establishing the NGGP as a flexible generalization.

## 2.3. Spatial and covariate extensions

In areal data applications, neighboring regions frequently exhibit similar socioeconomic characteristics due to shared infrastructure, labor markets, and historical development patterns. We incorporate this domain knowledge through prior modifications that encourage spatially coherent clustering without imposing hard constraints.

Spatial adjacency is incorporated through a prior modification following Cremaschi et al. (2023) and Gianella, Quintana, et al. (2026):

$$\pi(\rho_n \mid W) \propto \pi_0(\rho_n) \cdot \exp\left(\lambda_s \sum_{h=1}^{K} \|W_h\|\right), \quad (5)$$

where $W$ is the binary spatial adjacency matrix, $\|W_h\|$ counts the number of adjacencies (shared boundaries) within cluster $h$, and $\lambda_s > 0$ is a tuning parameter controlling spatial similarity strength. Larger values of $\lambda_s$ assign higher prior probability to partitions where geographically adjacent areas are grouped together, reflecting the empirical observation that spatial proximity correlates with outcome homogeneity.
Auxiliary covariates are incorporated via a PPMx similarity function (Müller et al. 2011):

$$\pi(\rho_n \mid \underline{x}) \propto \pi(\rho_n \mid W) \cdot \prod_{h=1}^{K} g(x_h^*), \quad (6)$$

where $x_h^* = \{x_i : i \in C_h\}$ collects covariates for observations in cluster $h$, and $g(x_h^*) = \int \prod_{i \in C_h} q(X_i \mid \zeta_h)\, q(\zeta_h)\, d\zeta_h$ is the marginal likelihood under an auxiliary probability model. This formulation treats covariates as informative for clustering without assuming they are random at the observation level, upweighting partitions where observations within clusters exhibit coherent covariate patterns.

## 2.4. MCMC sampling scheme

The conjugacy structure enables marginalization of component parameters $\lambda$ and $\theta$, allowing the MCMC to target the posterior of the partition $\rho_n$ directly. This collapsed representation improves mixing by eliminating nuisance parameters and reducing the state space dimension.
The sampler integrates several methodological advances: (i) the Sequential Allocated Merge-Split (SAMS) sampler (Dahl et al. 2022) for partition proposals, which employs sequential allocation during split moves for improved computational efficiency; (ii) a distance-based adaptation of Locality-Sensitive Sampling (LSS) (Luo et al. 2018) for observation selection, which assigns probability weights based on distance-based proximity to improve proposal quality; (iii) the Smart-Dumb-Dumb-Smart (SDDS) move-type selection scheme (Wang et al. 2015), which probabilistically selects between "smart" moves (us-

ing SAMS) and "dumb" moves (uniform allocation) based on observation properties; (iv) periodic Gibbs scans via Algorithm 3 of (Neal 2000) for local refinement; and (v) shuffle moves (Martínez et al. 2014) for fine-grained reassignment within existing clusters. The auxiliary variable $U$ in (4) is updated via adaptive Random Walk Metropolis-Hastings with proposal variance tuned via the Robbins–Monro process (Garthwaite et al. 2016).

## 3. Simulation study

Before applying the framework to real data, we validate its ability to recover known cluster structures under controlled conditions using three simulated datasets from Natarajan et al. (2024). Each dataset contains $n = 100$ observations drawn from a mixture of $K = 10$ multivariate Gaussian clusters with centers at the vertices of the standard $d$-dimensional simplex and identity covariance matrices scaled by $\sigma^2 I_d$. We evaluate three scenarios of increasing difficulty: $(\sigma, d) = (0.18, 10)$ representing well-separated clusters; $(0.25, 10)$ with moderate overlap; and $(0.2, 50)$ with high-dimensional overlap. Higher values of $\sigma$ increase within-cluster dispersion, while higher $d$ increases the overlap between intra- and inter-cluster distance distributions. To evaluate the spatial extension, we construct an artificial adjacency matrix using eight-nearest neighbors in Euclidean space.

| Configuration | $(0.18, 10)$ | $(0.25, 10)$ | $(0.2, 50)$ |
|---|---|---|---|
| NGG only | 1.00 | 0.90 | 0.65 |
| NGG + Spatial | 1.00 | 0.90 | 0.82 |
| Natarajan et al. | 1.00 | 0.85 | 0.90 |

Table 1: Clustering accuracy measured by Adjusted Rand Index (ARI) for simulated datasets under three difficulty levels.

The NGGP prior achieves perfect recovery in the easiest scenario and maintains competitive performance in the moderate case. Incorporating spatial adjacency provides a substantial boost in the high-dimensional scenario, improving ARI from 0.65 to 0.82, demonstrating the value of spatial structure even when artificially imposed. The reference model of Natarajan et al. (2024) achieves the best performance in the most challenging setting (ARI = 0.90), likely due to its specialized geometry for simplex-centered clusters. These results establish that our framework achieves competitive clustering accuracy while offering greater flexibility through nonparametric priors and spatial extensions.

## 4. California census income dataset

The primary application considers 93 PUMAs in the Los Angeles metropolitan area, encompassing Los Angeles, Ventura, and Orange counties, with approximately 80,000 individual income observations from the 2020 American Community Survey. Personal income values are log-transformed to account for right-skewness, and kernel density estimates are computed for each PUMA. Pairwise Jeffrey divergences serve as the distance metric, selected for superior effective sample size (ESS) properties compared to the 1-Wasserstein distance, which exhibited slower mixing in preliminary analyses.

Three model variants are evaluated: the baseline NGGP with NGG process induced prior; NGGPW incorporating spatial adjacency; and NGGPWx incorporating demographic covariates (standardized mean age and modal sex). All models use hyperparameters $a = 1$, $\tau = 1$, $\sigma = 0.1$, with spatial strength $\lambda_s = 1$ where applicable.

### 4.1. Clustering results and interpretation

The baseline identifies five clusters including one singleton PUMA, revealing distinct income distribution profiles: a high-income coastal cluster (Cluster 3, mean \$114,342), moderate-income suburban clusters (Clusters 1 and 2, means \$70,290 and \$62,723), a lower-income central cluster (Cluster 4, mean \$28,929), and a singleton low-income outlier (Cluster 5, mean \$29,624). Both NGGPW and NGGPWx yield a six-cluster solution where the previous low-income cluster splits into two more homogeneous subgroups—one with higher mean but lower variance (\$29,970, 2 PUMAs) and one with lower mean but higher variance (\$28,148, 3 PUMAs)—suggesting that spatial regularization helps identify distinct socioeconomic zones within previously aggregated areas.

The three NGGP-based models produce highly concordant partitions (ARI $\geq 0.99$), indicating

robustness to the inclusion of spatial and covariate information. By contrast, the cohesion-only LCF model (omitting the repulsion term) identifies a substantially different three-cluster partition with dispersed, less cohesive geographic patterns (ARI $\approx$ 0.23 with NGGP variants), confirming the essential role of repulsion for well-separated clustering. A frequentist baseline using the REDCAP algorithm (Guo 2008) with $K = 6$ produces markedly different partitions emphasizing spatial contiguity over distributional similarity (ARI $\approx$ 0.23 with NGGPWx), highlighting that our probabilistic approach captures fundamentally different structure than optimization-based regionalization.
Covariate analysis reveals meaningful demographic patterns: the high-income coastal cluster has the highest mean age (51.1 years), while the lowest-income clusters exhibit younger populations (43.6–45.8 years). Sex distribution is relatively balanced across clusters, with only the singleton cluster showing female predominance.

### 4.2. Computational performance

| Model | ESS | Time (sec) | ESS/s |
|---|---|---|---|
| LCF | 831 | 0.64 | 1298 |
| NGGP | 646 | 0.92 | 703 |
| NGGPW | 1105 | 1.02 | 1087 |
| NGGPWx | 944 | 1.14 | 826 |

Table 2: Computational efficiency for the California dataset (20,000 iterations, 5,000 burn-in). ESS denotes effective sample size for the number of clusters.

All model variants complete 20,000 MCMC iterations in approximately one second, representing a reduction from approximately ten hours (SPMIX) to under two seconds. The NGGPW model achieves the optimal trade-off between mixing quality and computational cost, with ESS/s improving 54.6% over baseline. The spatial prior enhances mixing by 71% in ESS at only 11% additional computational cost, demonstrating that incorporating domain knowledge improves both statistical and computational efficiency. The LCF model, while fastest (ESS/s = 1298), produces different clustering as evidenced by its divergence from NGGP-based solutions.

## 5. Italian municipalities dataset

To test scalability to larger spatial systems, we apply the framework to 7,903 Italian municipalities from the ISTAT 2020 census. Unlike the California case where individual-level data are available, the Italian data provide aggregated income histograms with seven brackets per municipality. This necessitates computing the discrete Jeffrey divergence directly on histogram representations rather than estimated densities, demonstrating the framework's flexibility to different data modalities.
The substantially larger number of areal units tests the $\mathcal{O}(n^2)$ scaling of the full likelihood: the distance matrix contains approximately 31 million entries, and each likelihood evaluation requires processing all intra- and inter-cluster distances.

### 5.1. Clustering results and geographic patterns

The NGGP and NGGPW models both identify 23 clusters with nine singletons. The geographic distribution shows limited spatial coherence without covariate information, with clusters fragmented across regions.
The NGGPWx model, incorporating the percentage of females as a demographic covariate, yields a dramatically more parsimonious six-cluster solution. This consolidation demonstrates that auxiliary covariates provide additional structure that enables the model to identify meaningful socioeconomic groupings without requiring numerous small clusters. The six clusters exhibit distinct geographic and economic profiles. Cluster 1 is the dominant large cluster, covering 4,773 municipalities with a moderate-high mean income of €23,598, while Cluster 2 represents a secondary large group of 2,737 municipalities characterised by a lower mean income of €17,101 and a slightly higher standard deviation in female percentage. Cluster 3, comprising 200 municipalities, stands out for its very low mean income of €12,507, in contrast to Cluster 4, which groups 130 municipalities with the highest mean income of €29,048. Cluster 5 contains 57 municipalities with a moderate-high mean income of €22,360 and the highest standard deviation in the percentage of females across all clusters. Finally, Cluster 6 is a tiny group of just 6 municipal-

ities representing outliers with unique income-demographic profiles.
The geographic map reveals that NGGPWx produces spatially coherent macro-regions: Cluster 1 dominates the northwest and center, Cluster 2 occupies the northeast and Adriatic regions, while Clusters 3–5 identify specific urban or suburban zones. This coherence emerges organically from the combination of income distribution similarity and demographic covariates, without hard spatial constraints.

### 5.2. Scalability and computational performance

| Model | ESS | Time (sec) | ESS/s |
|---|---|---|---|
| LCF | 843 | 1665 | 0.507 |
| NGGP | 373 | 4800 | 0.078 |
| NGGPW | 506 | 4735 | 0.107 |
| NGGPWx | 521 | 5153 | 0.101 |

Table 3: Computational efficiency for the Italian municipalities dataset (20,000 iterations).

The $\mathcal{O}(n^2)$ scaling becomes evident: inference requires approximately 80 minutes for the full likelihood models versus 28 minutes for the $\mathcal{O}(n)$ LCF variant, representing a 67% runtime reduction. The NGGPW model improves ESS by 36% over the baseline at marginal additional cost, maintaining favorable efficiency. The high ARI between NGGP and NGGPW ($\geq 0.99$) confirms stable clustering structure, while the lower agreement with NGGPWx (ARI $\approx 0.91$) reflects the meaningful parsimony induced by covariate integration.
Comparison with the frequentist PAM algorithm (6 clusters) reveals moderate agreement at best (ARI $\approx$ 0.28–0.61), with PAM producing fragmented, geographically incoherent clusters lacking the spatial-socioeconomic structure identified by our Bayesian approach.

## 6. Conclusions

This thesis developed and empirically validated a scalable Bayesian nonparametric framework for clustering areal data based on pairwise distances between kernel density estimates. The approach addresses a critical computational bottleneck in spatial Bayesian inference, reducing computational demands from hours to seconds for moderate-scale problems (93 PUMAs, ∼80,000 observations) while maintaining scalability to thousands of areal units (7,903 Italian municipalities).
The main methodological contributions include: (i) adaptation of the cohesion–repulsion distance-based likelihood to areal density objects, enabling probabilistic clustering on distributional representations; (ii) an NGG-induced partition prior extended with spatial adjacency and PPMx covariate terms, providing flexible, data-driven inference on cluster structure; and (iii) an efficient MCMC scheme synthesizing SAMS split-merge moves, distance-based LSS observation selection, SDDS move-type selection, and adaptive auxiliary variable updates. Empirical validation demonstrates that the spatial prior substantially improves posterior mixing (up to 71% ESS increase) and that the repulsion term is essential for well-separated, interpretable clustering.
Three limitations suggest directions for future research. First, the $\mathcal{O}(n^2)$ likelihood scaling due to the repulsion term becomes a computational bottleneck for large $n$, as evident in the Italian analysis where full models require 80+ minutes. Sparse approximations exploiting spatial structure—such as including only distances between neighboring regions or adaptive subsampling—are natural remedies that could restore linear scaling while preserving repulsion effects. Second, the framework involves several hyperparameters ($\sigma$, $\lambda_s$, $\delta_1$, $\delta_2$) whose current settings follow Natarajan et al. (2024); data-driven calibration via empirical Bayes or hierarchical priors would strengthen applicability across domains. Third, the choice of distance metric meaningfully affects mixing, with Jeffrey divergence outperforming Wasserstein in our experiments, but optimal selection requires domain expertise; systematic comparison of metrics and their interaction with likelihood specification warrants investigation.
Future extensions include hierarchical modeling for nested administrative structures (regions containing provinces containing municipalities), spatiotemporal panels for monitoring evolving regional inequalities, and integration with causal inference frameworks to evaluate policy interventions targeting identified clusters. The public release of the software implementation facili-

tates these extensions and enables application to diverse policy contexts including public health, environmental monitoring, and urban planning.

## References


Cremaschi, A. et al. (2023). "A Change-Point Random Partition Model for Large Spatio-Temporal Datasets". In: *arXiv preprint arXiv:2312.12396*. DOI: 10.48550/arXiv.2312.12396.

Dahl, D. B. and S. Newcomb (2022). "Sequentially allocated merge-split samplers for conjugate Bayesian nonparametric models". In: *Journal of Statistical Computation and Simulation* 92.7, pp. 1487–1511. DOI: 10.1080/00949655.2021.1998502.

Duan, L. L. and D. B. Dunson (2021). "Bayesian Distance Clustering". In: *Journal of Machine Learning Research* 22.189, pp. 1–27.

Garthwaite, P. H., Y. Fan, and S. A. Sisson (2016). "Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process". In: *Communications in Statistics - Theory and Methods* 45.17, pp. 5098–5111. DOI: 10.1080/03610926.2014.936562.

Gianella, M., M. Beraha, and A. Guglielmi (2023). "Bayesian nonparametric boundary detection for income areal data". In: *arXiv preprint arXiv:2312.13992*. DOI: 10.48550/arXiv.2312.13992.

Gianella, M., F. A. Quintana, and A. Guglielmi (2026). "Consensus Monte Carlo for Large Spatial Dataset". Manuscript in preparation.

Guo, D. (2008). "Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP)". In: *International Journal of Geographical Information Science* 22.7, pp. 801–823. DOI: 10.1080/13658810701674970.

James, L. F., A. Lijoi, and I. Prünster (2009). "Posterior Analysis for Normalized Random Measures with Independent Increments". In: *Scandinavian Journal of Statistics* 36.1, pp. 76–97. DOI: 10.1111/j.1467-9469.2008.00609.x.

Luo, C. and A. Shrivastava (2018). "Scaling-up Split-Merge MCMC with Locality Sensitive Sampling (LSS)". In: *arXiv preprint arXiv:1802.07444*. Proceedings of the 35th International Conference on Machine Learning (ICML).

Martínez, A. F. and R. H. Mena (2014). "On a Nonparametric Change Point Detection Model in Markovian Regimes". In: *Bayesian Analysis* 9.4, pp. 823–858. DOI: 10.1214/14-BA911.

Müller, P., F. Quintana, and G. L. Rosner (2011). "A Product Partition Model With Regression on Covariates". In: *Journal of Computational and Graphical Statistics* 20.1, pp. 150–169. DOI: 10.1198/jcgs.2011.09066.

Natarajan, A. et al. (2024). "Cohesion and Repulsion in Bayesian Distance Clustering". In: *Journal of the American Statistical Association* 119.546, pp. 1374–1384. DOI: 10.1080/01621459.2023.2191821.

Neal, R. M. (2000). "Markov chain sampling methods for Dirichlet process mixture models". In: *Journal of Computational and Graphical Statistics* 9.2, pp. 249–265.

Wang, W. and S. Russell (2015). "A Smart-Dumb/Dumb-Smart Algorithm for Efficient Split-Merge MCMC". In: *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pp. 846–855.