Bayesian complementary clustering, MCMC and Anglo-Saxon Placenames

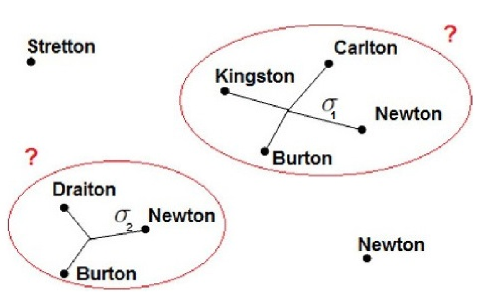
Giacomo Zanella

Department of Statistics, University of Warwick, UK

MOTIVATION AND MODELING

HISTORICAL HYPOTHESIS

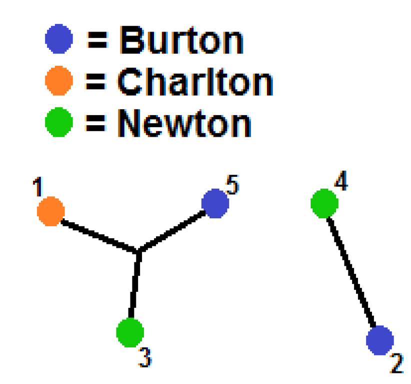
Anglo-Saxon settlements were grouped in clusters, each cluster collecting together settlements with distinctive administrative functions (hence names).



Questions:
Does the geographical distribution of placenames support this?
What is the typical intra-cluster dispersion σ?
Which subset of settlements is clustered?
Which placenames tend to cluster together?

MODEL REQUIREMENTS

k-type point process:
Marks (colors) represent placenames.
Complementary clustering:
Each color occurs at most once in each cluster.
Inferences on cluster partition:



$\mathbf{x} = (x_1, ..., x_{n(\mathbf{x})}) \leftrightarrow$ Given points

$\rho = \{C_1, ..., C_{N(\rho)}\} \leftrightarrow$ Partition of $\mathbf{x}$ $\quad \rho = \{\{1,3,5\},\{2,4\}\}$

$$\pi(\rho|\mathbf{x}) \quad \propto \quad \pi(\rho)\pi(\mathbf{x}|\rho)$$

A RANDOM PARTITION MODEL

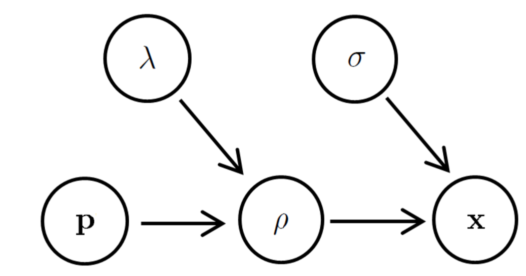
Data generation:

Cluster centers $\leftrightarrow$ Poisson Point Process driven by $\lambda\, g(\cdot)$
Cluster sizes $\leftrightarrow$ $|C_j| \sim (p_1, ..., p_k)$, $\quad j = 1, ..., N(\rho)$
Locations $\leftrightarrow$ Gaussian, mean at center, variance σ^2/π

Induced prior on ρ**:**

$$\pi(\rho|\mathbf{p}, \lambda) \quad \propto \quad \prod_{j=1}^{N(\rho)} \lambda\, p_{|C_j|}$$

with $\lambda \sim \text{Gamma}(k_\lambda, \theta_\lambda)$ and $\mathbf{p} = (p_1, ..., p_k) \sim \text{Dir}(1, ..., 1)$.



COMPUTATION: a Data Association problem

Intractability of the posterior

The full conditional posterior $\pi(\rho|rest)$ is intractable:

$$\prod_{j=1}^{N(\rho)} \left(\binom{k}{|C_j|}^{-1} \frac{p_{|C_j|}\, g(x_{C_j})}{|C_j|\,(2\pi\sigma^2)^{|C_j|-1}} \exp\left(-\frac{\sum_{i\in C_j}|x_i - x_{C_j}|^2}{2\sigma^2}\right)\right). \qquad (1)$$

	Finding MLE or MAP	Approximate Sampling
2 colors	$O(n^3)$ with the Hungarian Algorithm	polynomial-time MCMC* (Monomer-Dimer systems)
≥ 3 col colors	NP-hard Assignment Problem	Only heuristics (no guarantees of convergence)

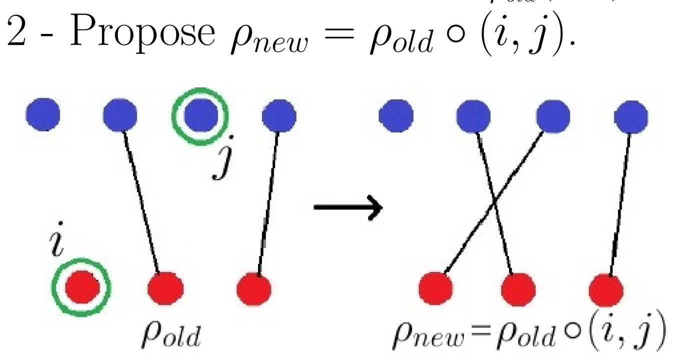
*Jerrum and Sinclair (1995), bound too large in practice.

MCMC for 2 colors

Sample space: matchings of a weighted bipartite graph.
Notation: ρ for matchings, (i,j) for edges and w_{ij} for weights.
Target distribution: $\pi(\rho) \propto \prod_{(i,j)\in\rho} w_{ij}$, with w_{ij} as in (1).
Algorithm: Metropolis-Hastings.

Proposal Distribution

1 - Given ρ_{old}, pick an edge (i,j) according to a distribution $q_{\rho_{old}}(i,j)$;
2 - Propose $\rho_{new} = \rho_{old} \circ (i,j)$.



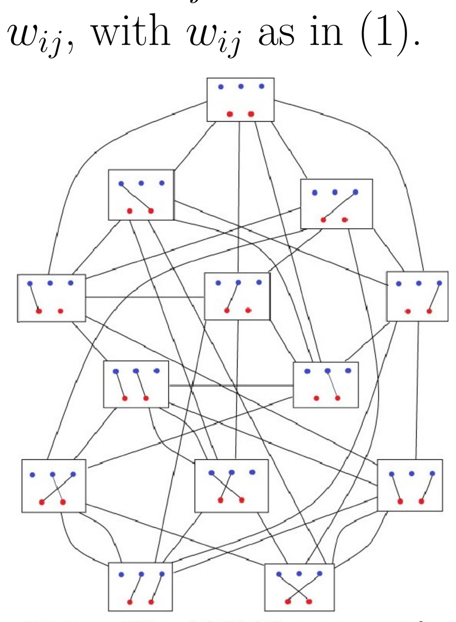


Figure: The MCMC as a graph.

What is the optimal choice for $q(i,j)$?

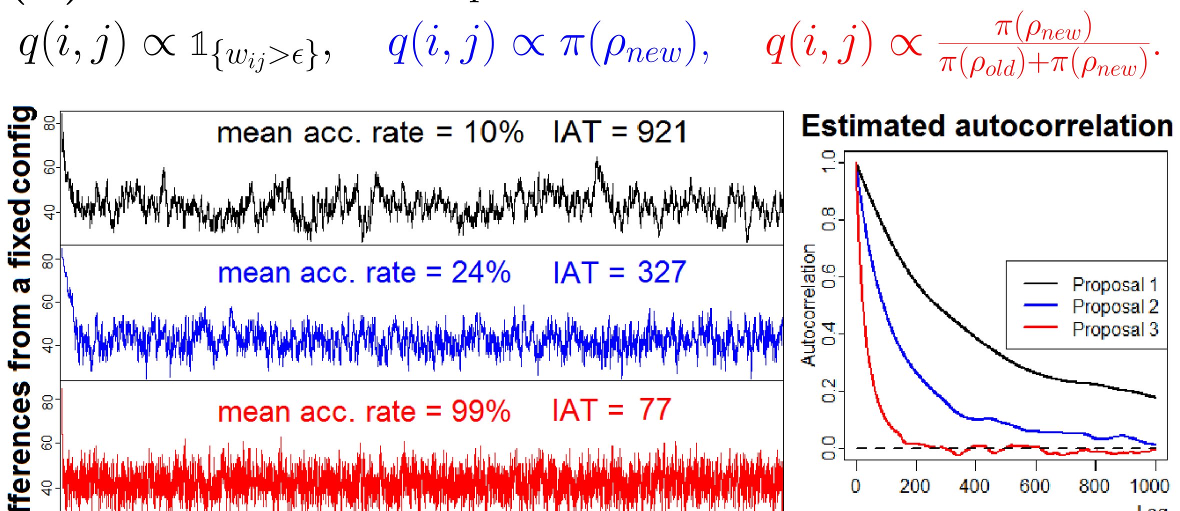
(A) Intuition: we seek $\frac{Q(\rho_{old},\rho_{new})}{Q(\rho_{new},\rho_{old})} \approx \frac{\pi(\rho_{new})}{\pi(\rho_{old})}$.
This is not obtained by $q(i,j) \propto \pi(\rho_{new})$. In fact

$$\frac{Q(\rho_{old},\rho_{new})}{Q(\rho_{new},\rho_{old})} = \frac{q_{\rho_{old}}(i,j)}{q_{\rho_{new}}(i,j)} = \frac{\pi(\rho_{new})/\sum_{i,j}\pi(\rho_{old}\circ(i,j))}{\pi(\rho_{old})/\sum_{i,j}\pi(\rho_{new}\circ(i,j))} =$$

$$= \frac{\pi(\rho_{new})\sum_{i,j}\pi(\rho_{new}\circ(i,j))}{\pi(\rho_{old})\sum_{i,j}\pi(\rho_{old}\circ(i,j))} \approx \frac{\pi(\rho_{new})}{\pi(\rho_{old})}\frac{\pi(\rho_{new})}{\pi(\rho_{old})} = \left(\frac{\pi(\rho_{new})}{\pi(\rho_{old})}\right)^2.$$

(B) Simulations: compare three choices

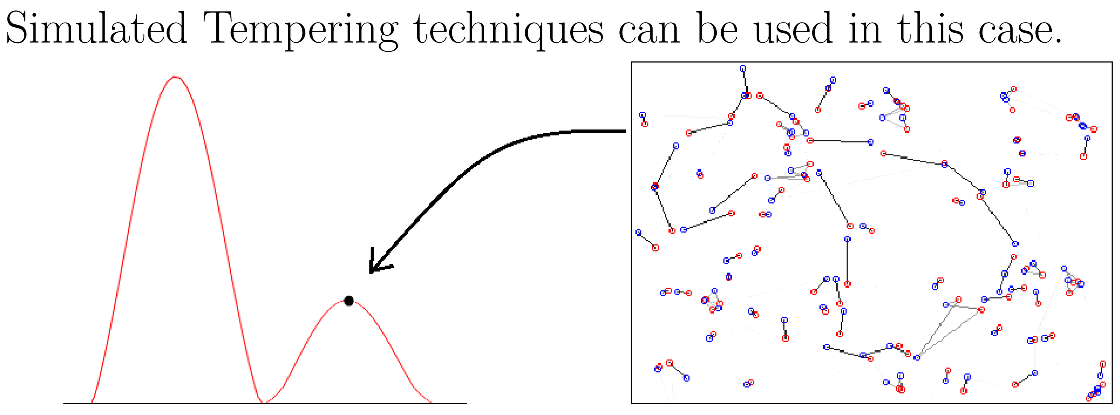
$q(i,j) \propto \mathbb{1}_{\{w_{ij}>\epsilon\}}$, $\quad q(i,j) \propto \pi(\rho_{new})$, $\quad q(i,j) \propto \frac{\pi(\rho_{new})}{\pi(\rho_{old})+\pi(\rho_{new})}$.



(C) Theoretical results? (work in progress)
1) Any $\tilde{q}(i,j)$ is (asymptotically) Peskun-dominated by a $q(i,j)$ satisfying asymptotic detailed balance, i.e. $\frac{Q(\rho_{old},\rho_{new})}{Q(\rho_{new},\rho_{old})} \rightarrow \frac{\pi(\rho_{new})}{\pi(\rho_{old})}$.
2) Many choices satisfies asymptotic detailed balance, e.g. $q(i,j) \propto \frac{\pi(\rho_{new})}{\pi(\rho_{old})+\pi(\rho_{new})}$, or $\sqrt{\pi(\rho_{new})}$, or $1 \wedge \frac{\pi(\rho_{new})}{\pi(\rho_{old})}$, or $1 \vee \frac{\pi(\rho_{new})}{\pi(\rho_{old})}$,...
None of those Peskun-dominates any other. How to choose?
3) In simple frameworks (e.g. hypercube with product target) it can be shown that $\frac{\pi(\rho_{new})}{\pi(\rho_{old})+\pi(\rho_{new})}$ minimizes the mixing time.

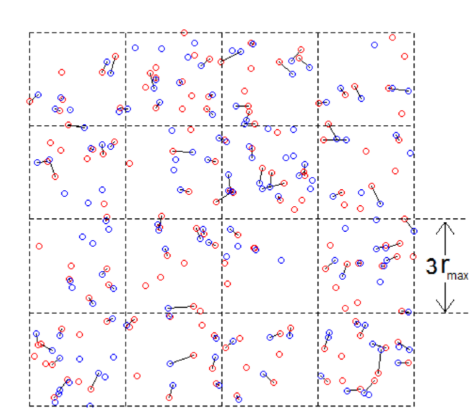
Multimodality and Simulated Tempering

For certain parameters values, the target measure exhibits multimodality (with local modes being cycle-like matchings). Simulated Tempering techniques can be used in this case.



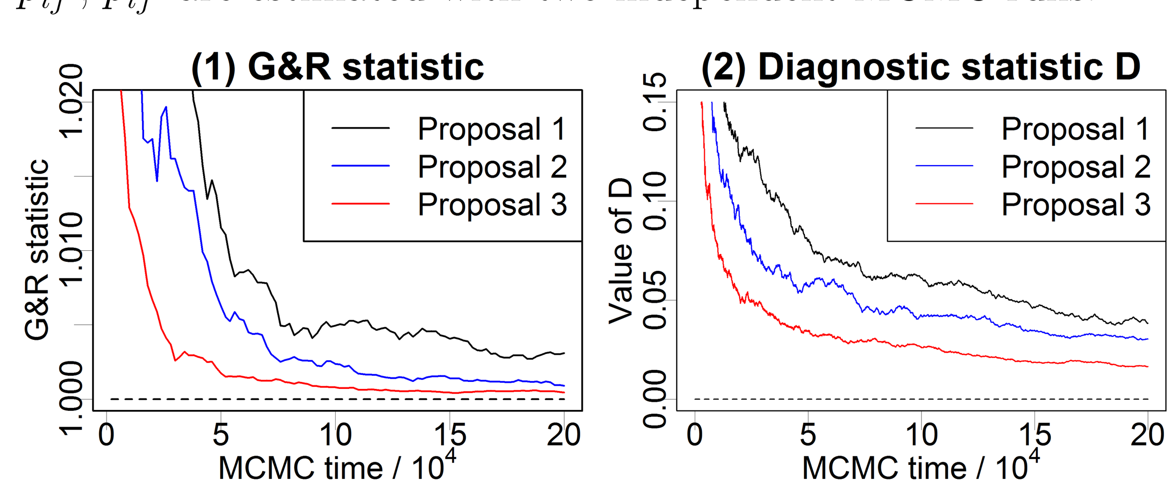
Multiple proposal scheme

Idea: propose a red-blue edge for each non-adjacent square of a grid.
Pros: simultaneous and independent proposal of many links (parallel computation).
Cons: truncation requirement (link length$<r_{max}$).



Two Convergence Diagnostic tools

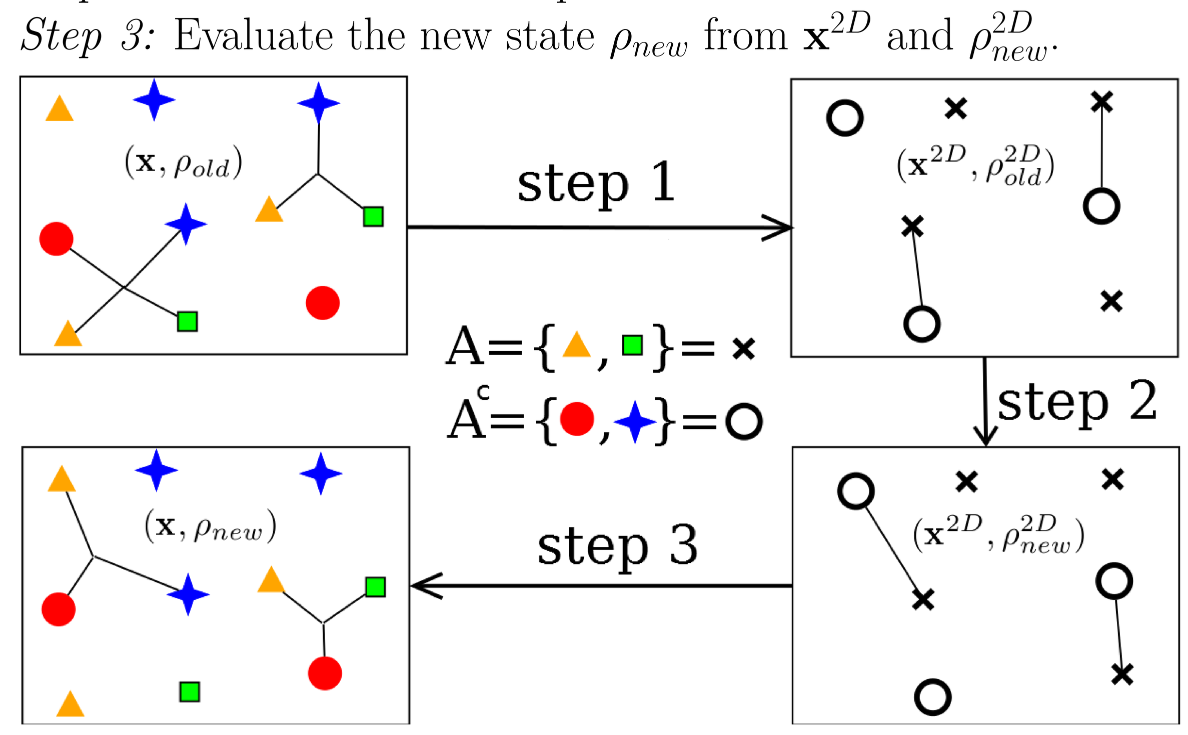
(1) Multi-dimensional Gelman and Rubin diagnostic.
(2) $D := \sup_{\{i,j\}} |\hat{p}_{ij}^{(1)} - \hat{p}_{ij}^{(2)}|$, where $p_{ij} = \mathbb{P}_\pi[(i,j) \in \rho]$ and $\hat{p}_{ij}^{(1)}$, $\hat{p}_{ij}^{(2)}$ are estimated with two independent MCMC runs.



MCMC for k colors

Sample space: matchings of a weighted k-partite hypergraph.
Notation: ρ for matchings, e for hyperedges, $w(e)$ for weights.
Target distribution: $\pi(\rho) \propto \prod_{e\in\rho} w(e)$, with $w(e)$ as in (1).
Algorithm: Mixture of Metropolis-Hastings.

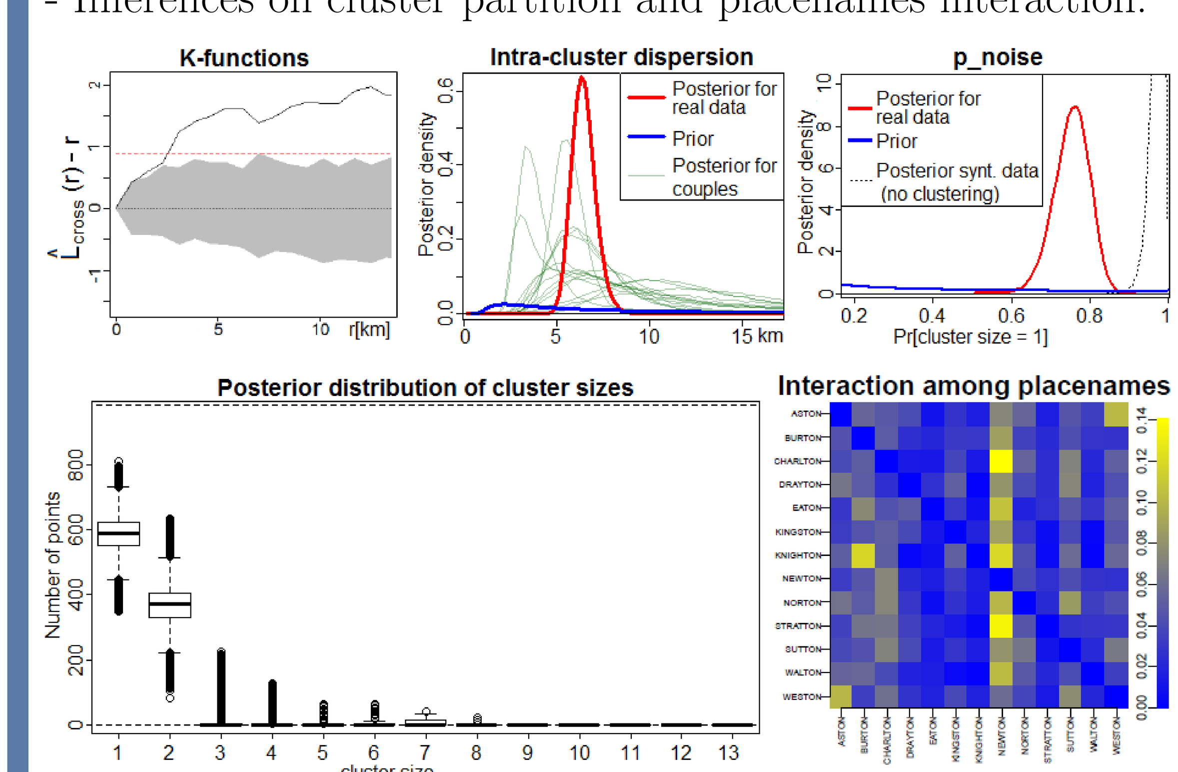
Proposal Distribution: given a subset of colors A
Step 1: Project (x, ρ_{old}) to a 2-color configuration $(\mathbf{x}^{2D}, \rho_{old}^{2D})$.
Step 2: Move in the 2-color space with the MCMC for 2-colors.
Step 3: Evaluate the new state ρ_{new} from $\mathbf{x}^{2D}$ and ρ_{new}^{2D}.

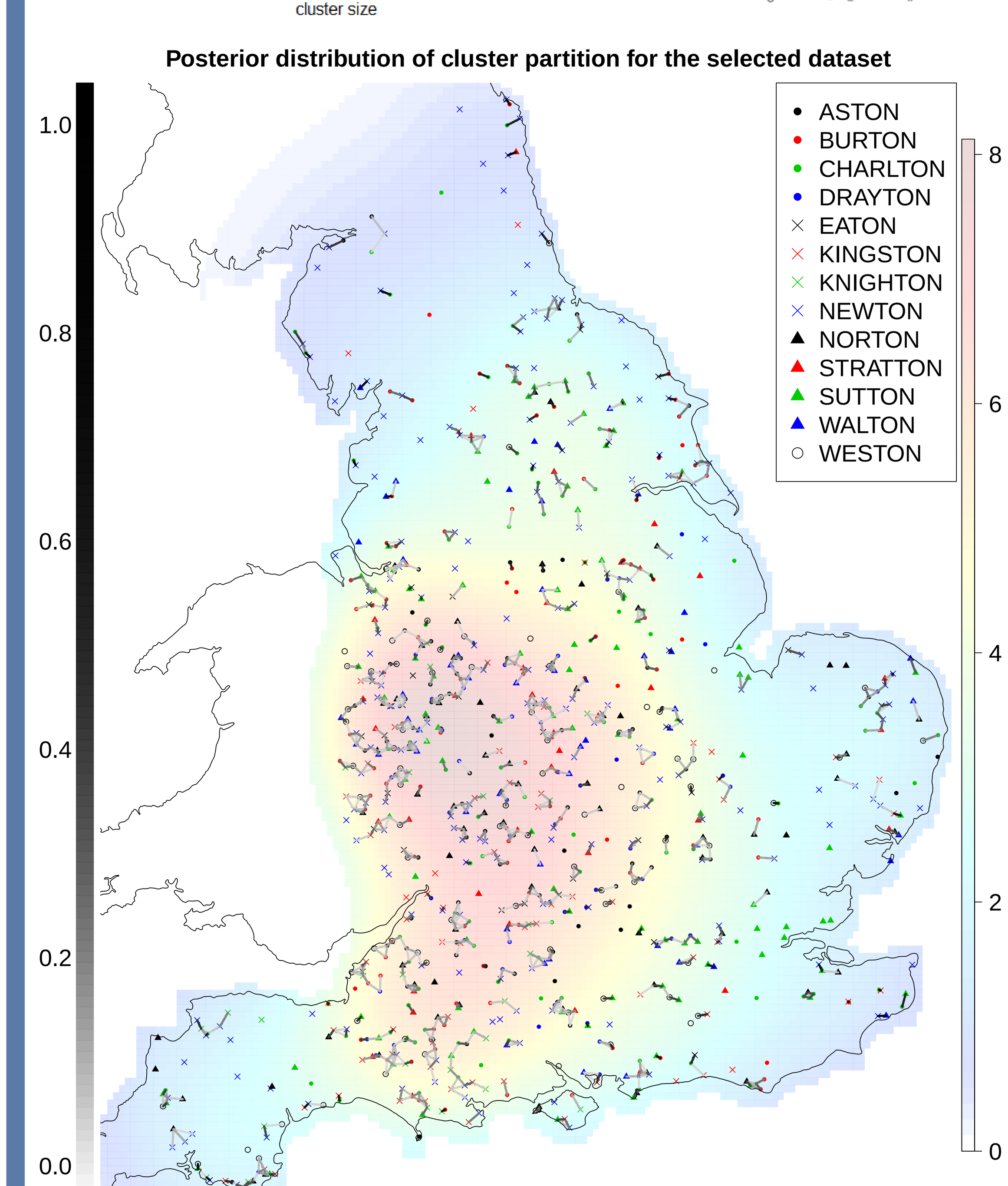


Remark: the projection of the target distribution π on the 2-color space (x^{2D}, ρ^{2D}) is the 2-color version of π itself!

DATA ANALYSIS

Dataset: 1021 settlements with 13 placenames.
Preliminary analysis: K-functions suggest interaction.
Random Partition Model analysis:
- Mild evidences in favor of interaction: $p_{noise} \neq 1$.
- With a flat prior, we estimate σ to lie in the range 5-8 km (support historians hypothesis).
- Inferences on cluster partition and placenames interaction.





Future Steps

- Include heterogeneity between high and low density region.
- Include other source of data (e.g. topology).
- Prove theoretical results related to the MH proposal.

Acknowledgements

Prof. Wilfrid Kendall for PhD supervision.
Prof. John Blair for collaboration and arranging supply of data.
CRiSM and EPSRC for funding.