SUPPLEMENT C TO "BAYESIAN COMPLEMENTARY CLUSTERING, MCMC AND ANGLO-SAXON PLACENAMES": MULTIPLE PROPOSAL SCHEME

By Giacomo Zanella*

University of Warwick

We define the multiple proposal scheme mentioned in Section 4.1.1 of Zanella (2014) and we prove that it targets the correct distribution. Then we demonstrate its performances on a synthetic sample and we give a rough prediction of the speedup that can be obtained for the analysis of the Anglo-Saxon placename location dataset studied in Zanella (2014) when using parallel programming.

1. Defining the square grid. Suppose that we observe a bivariate point pattern $\mathbf{x}$ in a square window $W = [0, a] \times [0, a]$, with $a > 0$ (otherwise consider a square containing the observed window). In the spirit of Besag's coding method (Besag, 1974), we first divide the window according to a grid of squares of side of length (at least) $2r_{max}$ like in Figure 1 (left). Then we divide the squares in 4 groups, in order to have no adjacent (nor corner adjacent) squares in the same group. Say for simplicity that we have l squares for each group from 1 to 4. We denote the squares as $\left\{S_s^g\right\}_{s=1,\dots,l}^{g=1,2,3,4}$, where the superscripts denotes the group and the subscript the square in the group (see Figure 1, left).

2. Defining the transition step. Each step of the multiple proposal scheme works as follows:

1. Choose an index g uniformly at random from $\{1, \dots, 4\}$;
2. For s running from 1 to l:
   (a) Define R_s^g as the set of all the red points inside S_s^g;
   (b) Define B_s^g as the set of blue points inside S_s^g or at distance at most r_{max} from S_s^g that are *not* linked to any red point in a square different from S_s^g (see the right-hand side of Figure 1);
   (c) Choose a red-blue couple (i, j) uniformly at random from $R_s^g \times B_s^g$;
   (d) Propose to move to $\rho_{new} = \rho_{old} \circ (i, j)$ and accept the move with probability $1 \wedge \frac{\hat{\pi}(\rho_{new})}{\hat{\pi}(\rho_{old})}$.

*Supported by EPSRC through a PhD position under the CRiSM grant EP/D002060/1

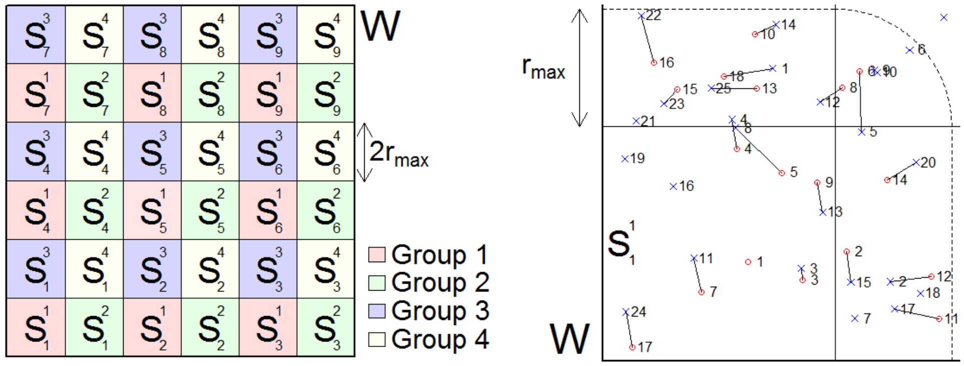

FIG 1. *Left: the observed window W divided into squares. Right: in this case* $R_1^1 = \{1, 3, 4, 5, 7, 9, 17\}$ *and* $B_1^1 = \{3, 4, 6, 7, 8, 9, 10, 11, 13, 16, 18, 19, 21, 24\}$.

Here $\hat{\pi}$ is the truncated version of the full conditional posterior distribution of ρ for the model described in Section 3 of Zanella (2014) with $k = 2$. Note that, since we are using truncation, only points closer than r_{max} can be linked.

For simplicity, in step $2(c)$ we considered (i, j) to be chosen uniformly at random from $R_s^g \times B_s^g$. The extension to a general proposal $q_{\rho_{old}}(i, j)$ like in Section 4.1 of Zanella (2014) is straightforward: one simply needs to take into account for the proposal in the MH acceptance probability in step $2(d)$.

Note that, when the target measure $\hat{\pi}$ factorizes as for the Poisson model (see equation (3.3) of Zanella, 2014), the l different MH steps in step 2 of such multiple proposal scheme can be implemented in an embarrassingly parallel fashion, meaning that they can be performed without the need of any communication between them. In fact it is easy to see that such l moves involve separate subgraphs of the original bipartite graph and, since $\hat{\pi}$ factorizes, the l acceptance-rejection steps are independent.

3. Showing the correctness of the induced MCMC. Before testing such scheme on a synthetic sample we need to show that the induced Markov chain is ergodic with stationary distribution $\hat{\pi}$. This is not obvious and indeed a careful choice of the sets R_s^g and B_s^g (like the one in step $2(a)-(b)$) is crucial in order for all the moves to be reversible and for the proposal distribution to be symmetric.

LEMMA 1. *The Markov Chain induced by the multiple proposal scheme is an ergodic Markov chain with stationary distribution $\hat{\pi}$.*

PROOF. The transition kernel P induced by the multiple proposal scheme can be seen as a mixture of 4 cycles of l transition kernels: $P = \frac{1}{4}\big(P_1^1 \cdots P_l^1 + P_1^2 \cdots P_l^2 + P_1^3 \cdots P_l^3 + P_1^4 \cdots P_l^4\big)$, where P_s^g is the MH transition kernel with proposal Q_s^g given by steps $2(a) - (d)$ for fixed $s \in \{1, \dots, l\}$ and $g \in \{1, 2, 3, 4\}$. If each P_s^g satisfies detailed balance conditions with respect to $\hat{\pi}$, then it follows that P satisfies them too. We need to show that, for any s and g, $Q_s^g(\rho_{old}, \rho_{new}) = Q_s^g(\rho_{new}, \rho_{old})$ for any couple of matchings (ρ_{old}, ρ_{new}), from which it follows that $1 \wedge \frac{\hat{\pi}(\rho_{new})}{\hat{\pi}(\rho_{old})}$ is the correct MH acceptance probability.

The probability of choosing a certain couple $(i, j) \in R_s^g \times B_s^g$ in step $2(c)$ is $\frac{1}{|R_s^g||B_s^g|}$. Note that the set R_s^g does not depend on ρ_{old}. On the other hand the set B_s^g does depend on ρ_{old}, but it does not change for $\rho_{new} = \rho_{old} \circ (i, j)$ with $(i, j) \in R_s^g \times B_s^g$. Therefore, when the current matching becomes ρ_{new} the probability of choosing (i, j) remains $\frac{1}{|R_s^g||B_s^g|}$ with the same R_s^g and B_s^g. Let $\rho_{new} = \rho_{old} \circ (i, j)$, with $(i, j) \in R_s^g \times B_s^g$ (otherwise $Q_s^g(\rho_{old}, \rho_{new})$ is clearly 0). Let us first consider the case where $\rho_{old} \circ (i, j)$ is and addition or deletion move (see equation (4.1) of Zanella, 2014). In this case the only way to propose to move from ρ_{old} to ρ_{new} (and back from ρ_{new} to ρ_{old}) with Q_s^g is by choosing the red-blue couple (i, j) in step 2(c). Therefore $Q_s^g(\rho_{old}, \rho_{new}) = Q_s^g(\rho_{new}, \rho_{old}) = \frac{1}{|R_s^g||B_s^g|}$. If $\rho_{old} \circ (i, j)$ is a switch move then the only way to propose to move from ρ_{old} to ρ_{new} is by choosing the red-blue couple (i, j), while the only way to propose to move back from ρ_{new} to ρ_{old} is by choosing either the couple (i', j) or the couple (i, j'), depending on whether $\rho_{old} \circ (i, j)$ equals $\rho_{old} - (i', j) + (i, j)$ or $\rho_{old} - (i, j') + (i, j)$ respectively. In the first case since $(i', j) \in \rho_{old}$ and $j \in B_s^g$ then, by definition of B_s^g, $i \in R_s^g$. In the second case, since $(i, j') \in \rho_{old}$ then the j'-th blue point has a distance smaller than r_{max} from S_s^g and therefore, again since $(i, j') \in \rho_{old}$, $j' \in B_s^g$. Therefore, since $(i', j) \in R_s^g \times B_s^g$, or $(i, j') \in R_s^g \times B_s^g$ respectively, $Q_s^g(\rho_{old}, \rho_{new}) = Q_s^g(\rho_{new}, \rho_{old}) = \frac{1}{|R_s^g||B_s^g|}$. Finally, if $\rho_{old} \circ (i, j)$ is a double-switch move then there are respectively two ways to propose to move from ρ_{old} to ρ_{new}, i.e. choosing (i, j) and (i', j') in step 2(c), and two ways to propose to move back from ρ_{new} to ρ_{old}, i.e. choosing (i, j') and (i', j). Analogously to the switch move one can show that $i' \in R_s^g$ and $j' \in B_s^g$. Therefore $Q_s^g(\rho_{old}, \rho_{new}) = Q_s^g(\rho_{new}, \rho_{old}) = \frac{2}{|R_s^g||B_s^g|}$.

The desired ergodicity follows from the fact that P is an aperiodic and irreducible Markov transition kernel on a finite state space satisfying detailed balance conditions with respect to $\hat{\pi}$. □

4. Demonstration of performances on a synthetic sample. We test the multiple proposal scheme on the posterior $\pi(\rho|\sigma, \mathbf{p}^{(c)}, \lambda, \mathbf{x})$ given by the Poisson model where $k = 2$, $\sigma = 0.3$, $p_1^{(c)} = p_2^{(c)} = 0.5$, $\lambda = 400$ and the center intensity $g(\cdot)$ is uniform over a window $W = [0, 20] \times [0, 20]$. Here $\mathbf{x}$ is a synthetic sample generated according to the Poisson model with such parameters. The sample $\mathbf{x}$ happened to be made of 310 red and 316 blue points (see Figure 2). We consider three cases, $l = 1, 4, 9$, in order to show that the mixing of the MH Markov chain improves roughly at rate equal to l. We used the convergence diagnostic techniques presented in Section 4.1.2 of Zanella (2014). The results are shown in Figure 2 and Table 1.

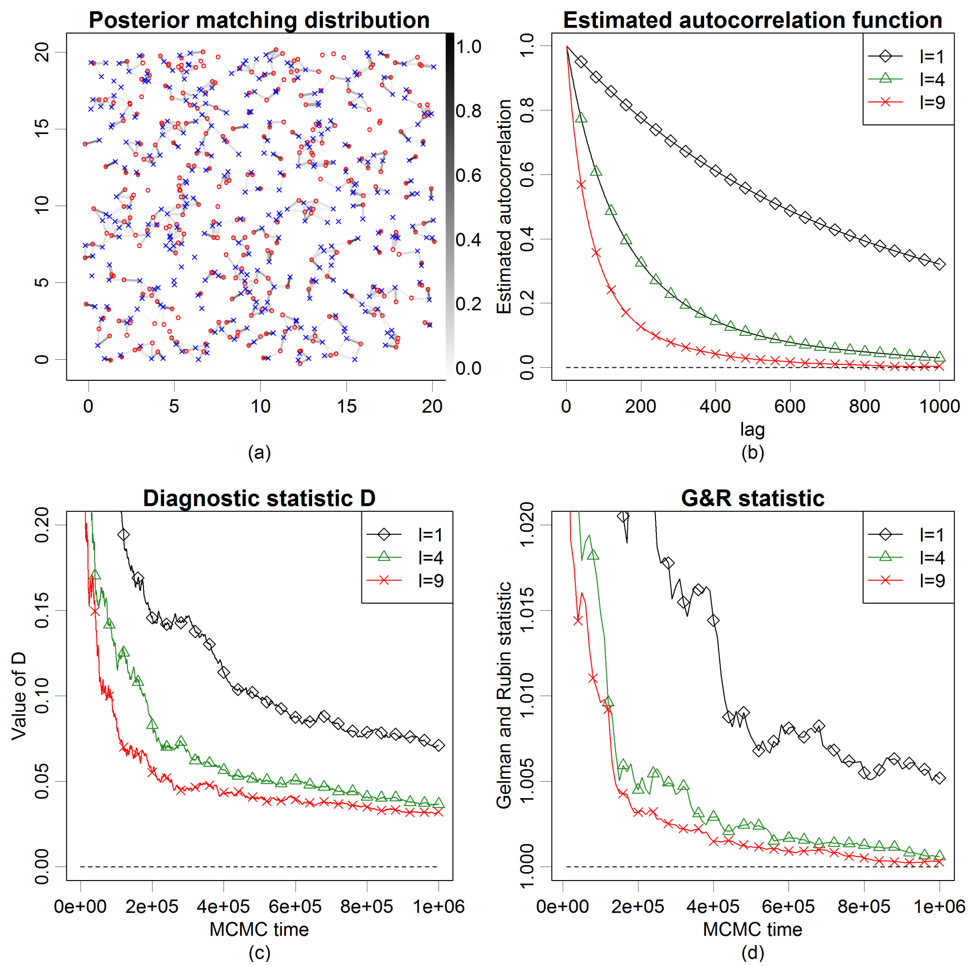

FIG 2. *Comparison of the multiple proposal scheme for $l = 1, 4, 9$ using the convergence diagnostic techniques of Section 4.1.2 of Zanella (2014)*

5. An approximate prediction of the speed-up for the real data. Given the historical interpretation of the clusters we are looking for (groups

	Estimated IAT	ESS for 10^4 steps	steps to reach $D < .1$	steps to reach reach $GR < .01$
$l = 1$	2030	32	4.9e05	4.3e05
$l = 4$	428	161	1.7e05	1.2e05
$l = 9$	193	262	7e04	1e05

Table 1

Performances of the multiple proposal scheme for $l = 1, 4, 9$ on configuration in Figure 2 (a) averaged over 5 independent runs for each value of l. GR denotes the multivariate Gelman and Rubin statistic (potential scale reduction factor).

of settlements interacting from the administrative and political point of view) two settlements in the same administrative cluster should be close enough to allow inhabitants to walk from one settlement to the other, spend time transacting business, and then returning all in a single day (thus, 3 hours to go and 3 hours to come back). In fact the historians involved in the project consider it to be implausible for two settlements in the same cluster to be separated a distance greater than 15 km. Therefore, when analyzing the settlements dataset, a reasonable choice of r_{max} could be, for example, 20 km (increasing on the upper bound given by the historians in order to have additional confidence of not imposing conditions that are too restrictive).
The area where we observe settlements is roughly 53 000 km. Therefore if we were to divide that area in squares of side $2r_{max} = 40$ km we would obtain approximately $\frac{53\,000}{40^2} \approx 33$ different squares. Therefore we would have approximately 8 squares for each group, i.e. $l = 8$. Therefore, given the results above, it is reasonable to expect a parallel implementation of such scheme to yield approximately an 8-times speed-up of the MH Markov Chain.

References.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 192–236.

Zanella, G. (2014). Bayesian Complementary Clustering, MCMC and Anglo-Saxon placenames. *arXiv preprint arXiv:1409.6994*.

Department of Statistics
University of Warwick
Coventry, CV4 7AL
United Kingdom
E-mail: g.zanella@warwick.ac.uk