A Bayesian space-time model for clustering areal units based on their disease trends
Supplementary Materials

GARY NAPIER*, DUNCAN LEE

School of Mathematics and Statistics, University of Glasgow, G12 8SQ, UK
Gary.Napier@glasgow.ac.uk

CHRIS ROBERTSON

Department of Mathematics and Statistics, University of Strathclyde, G1 1XH, UK

ANDREW LAWSON

Department of Public Health Sciences, Medical University of South Carolina, USA

INTRODUCTION

This supplementary material has the following sections. Appendix A presents additional data description for the two case studies, while Appendix B describes the (MC)^3 estimation algorithm used to fit the model. Appendix C summarises the computational demand of fitting our model to data of different sizes, while Appendix D presents the data generation and additional results from the simulation study. Appendix E presents additional sensitivity analyses for the two case studies.

*To whom correspondence should be addressed.
APPENDIX A - ADDITIONAL DATA DESCRIPTION FOR THE TWO CASE STUDIES

A.1 Measles susceptibility case study

The study region is Scotland, and a map with the main towns and cities is presented in Figure 1. Scotland has a population of 5.4 million people, and for this study has been partitioned into $K = 1235$ non-overlapping intermediate zones (IZ), which are a key geography designed by the Scottish Government for distributing small-area statistics. The spatial pattern in the proportion susceptible (un-vaccinated) $\{\hat{\theta}_{kt}\}$ in 2004 is displayed in Figure 2, which shows that the less densely populated regions in the north and west of Scotland appear to have higher susceptibility rates compared to the more densely populated regions, such as Glasgow and Edinburgh. One of the reasons for this could be that general practitioners (doctors) and health centres are not as accessible in these remote rural regions, and so in combination with the negative press towards the MMR vaccine, parents in those areas may have been less likely to have their children vaccinated.

The temporal trend in $\{\hat{\theta}_{kt}\}$ for all areas is displayed in panel (a) of Figure 3, which shows a clear change point in 2004 which corresponds to the year of the retraction of the discredited article by Wakefield.

A.2 Respiratory hospitalisation case study

The study region for the respiratory hospitalisation case study is the Greater Glasgow and Clyde Health Board, which is displayed as the shaded region in Figure 4. The region has been partitioned into $K = 271$ non-overlapping intermediate zones, and the spatial pattern in the SMR, $\hat{\theta}_{kt}$, in 2005 is displayed in Figure 5. The spatial pattern shows elevated SMR values in the east of the city and just south of the river Clyde (the white line running south-east), which are some of the poorest parts of the study region. The temporal trends are displayed in panel (b) of Figure 3, and show no clear trends over the 10 year period.
Bayesian space-time model for clustering areas based on their disease trends

Fig. 1: Map of Scotland showing the main towns and cities.
Fig. 2: Spatial pattern in the estimated proportions $\hat{\theta}_{kt}$ of pre-school children susceptible to measles in Scotland in 2004. The panels in the top-left corner highlight the cities Glasgow and Edinburgh in the centre of Scotland.
Bayesian space-time model for clustering areas based on their disease trends

Fig. 3: Temporal trends in: (a) susceptibility to measles $\hat{\theta}_{kt}$; and (b) risk of hospitalisation due to respiratory disease.
Fig. 4: The Study region of the Greater Glasgow and Clyde Health Board (shaded region) overlaid on a Google map.
Bayesian space-time model for clustering areas based on their disease trends

Fig. 5: Spatial pattern in the SMR \( \hat{\theta}_{kt} \) for respiratory hospitalisation in 2005.
Appendix B - (MC)$^3$ estimation algorithm

For our (MC)$^3$ algorithm $M$ Markov chains $C = (C_1, \ldots, C_M)$ are run in parallel with temperatures $T_1 = 1 > T_2 > \ldots > T_M > 0$. The first chain with $T_1 = 1$ is known as the cold chain, and is the one upon which final posterior inference is based. Denote the collection of model parameters for the $i$-th iteration of the $m$-th chain as $\Psi_{C_{mi}} = \{\beta_{C_{mi}}, \phi_{C_{mi}}, \omega_{C_{mi}}, \lambda_{C_{mi}}, \rho_{C_{mi}}, \gamma_{C_{mi}}, \gamma_{C_{mi}}^2\}$, where $\gamma_{C_{mi}}$ is the complete vector of parameters comprising the $S$ trend functions. Then the (MC)$^3$ algorithm is defined as follows.

**((MC)$^3$ Algorithm)**

1. Specify initial values, $\Psi_{C_{m0}}$ for each parameter in each Markov chain for $m = 1, \ldots, M$.

2. Repeat the following steps for $i = 1, \ldots, n$ iterations of the sampling scheme.

   (a) At iteration $i$ repeat the following steps for each Markov chain for $m = 1, \ldots, M$, and each model parameter $\psi_{C_{mi}} \in \Psi_{C_{mi}}$.

      i. Sample a proposal value $\tilde{\psi}_{C_{mi}}$ from a proposal distribution $q(\tilde{\psi}_{C_{mi}}|\psi_{C_{mi}})$.

      ii. Accept $\tilde{\psi}_{C_{mi}}$ as the next value in the chain, that is set $\psi_{C_{mi+1}} = \tilde{\psi}_{C_{mi}}$, with probability $\min(1, \alpha_1)$, where

      \[
      \alpha_1 = \frac{p(\tilde{\psi}_{C_{mi}})^T q(\tilde{\psi}_{C_{mi}}|\psi_{C_{mi}})}{p(\psi_{C_{mi}})^T q(\psi_{C_{mi}}|\psi_{C_{mi}})}
      \]

      and $p(\cdot)$ denotes the full conditional distribution of $\tilde{\psi}_{C_{mi}}$ or $\psi_{C_{mi}}$.

   (b) To couple the chains randomly select two of the chains, say $\tilde{\Psi}_{C_{ji}}$ and $\tilde{\Psi}_{C_{ki}}$.

      i. The proposal values for a Metropolis type move are $\tilde{\Psi}_{C_{ji}} = \Psi_{C_{ki}}$ and $\tilde{\Psi}_{C_{ki}} = \Psi_{C_{ji}}$.

      ii. Swap the chains with probability $\min(1, \alpha_2)$, where

      \[
      \alpha_2 = \frac{p(\tilde{\Psi}_{C_{ki}})^T p(\tilde{\Psi}_{C_{ji}})^T}{p(\tilde{\Psi}_{C_{ki}})^T p(\tilde{\Psi}_{C_{ji}})^T}.
      \]
Bayesian space-time model for clustering areas based on their disease trends

Note that for parameters that can be Gibbs sampled the temperatures do not apply, and the acceptance rate is always equal to one. From the sampling scheme it is clear that increasing the temperature level $T_m$ increases the acceptance probability $\alpha_1$. The acceptance rates for accepting a new proposed value, $\tilde{\psi}_{C,m}$, are controlled by updating the standard deviation of the proposal distribution $q(\cdot)$, such that an acceptance rate of 40-50% is achieved. Through simulation testing we found that swapping chains with an acceptance rate of 20-30% was sufficient to remove the multimodality issue, which is similar to the results of Atchadé and others (2011). In order to obtain such an acceptance rate geometric spacing of the temperatures was used, that is $T_{m+1} = c \cdot T_m$, where $T_1 = 1$ and $0 < c < 1$. The constant $c$ is updated during the algorithm to ensure the spacing between the temperatures of the chains provides an acceptance rate of 20-30%, such that, if the acceptance rate is greater than 30%, the difference between the temperatures will widen, and hence shrink if the acceptance rate of swapping chains is less than 20%. Choosing the number of chains $M$ is an extra choice to be made in this algorithm and can be application specific (Altekar and others, 2004), and here $M = 4$ appeared to work well following exploratory runs of the algorithm for both the simulated and real data sets. Tuning of the proposal distributions to have the acceptance rates discussed above is automatically undertaken in the algorithm, meaning the user does not need to manually tune the algorithm.

Appendix C - Computation time and scalability

Table 1 displays the computation times for fitting the mixture model proposed in the main paper to data of various sizes, which illustrates the scalability of our approach. Run times are provided for both the MCMC and the (MC)$^3$ algorithms, and in all cases the models are run with $S = 4$ temporal trends on regular spatio-temporal grids. Inference in all cases was based on 100,000 iterations, and the (MC)$^3$ algorithm had $M = 4$ parallel chains (as used in the simulation and real data studies) resulting in 400,000 iterations in total. The timings were carried out using
an Apple iMac computer with a 3.2 GHz Intel Core i5 processor and 16 GB MHz DDR3 memory. The table displays run times for data sizes between 1,000 (10 × 10 spatial grid with 10 time periods) and 50,000 (50 × 50 spatial grid with 20 time periods) data points, and the maximum run time for 100,000 iterations is under 18 hours. The Table shows that run times increase with increasing numbers of data points as expected, and that the (MC)³ algorithm is around 3 times slower than the MCMC algorithm despite generating 4 times as many iterations (it is based on 4 parallel chains).

Table 1: Summary of the time taken to run the mixture model in seconds (in hours, minutes and second in brackets) on a regular grid with different square grid sizes (K), number of time periods (N) and total number of data points (N.all).

<table>
<thead>
<tr>
<th>K</th>
<th>N</th>
<th>N.all</th>
<th>Timing in seconds (h:m:s) (MC)³</th>
<th>MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>10,000</td>
<td></td>
<td>1,523 (25.23)</td>
<td>492 (8.12)</td>
</tr>
<tr>
<td>100</td>
<td>20,000</td>
<td></td>
<td>2,508 (41.48)</td>
<td>803 (13.23)</td>
</tr>
<tr>
<td>400</td>
<td>10,000</td>
<td></td>
<td>5,003 (1:23.23)</td>
<td>1,637 (27.17)</td>
</tr>
<tr>
<td>400</td>
<td>20,000</td>
<td></td>
<td>8,000 (2:46.43)</td>
<td>3,063 (51.3)</td>
</tr>
<tr>
<td>900</td>
<td>18,000</td>
<td></td>
<td>22,396 (6:13.16)</td>
<td>6,908 (1:55.8)</td>
</tr>
<tr>
<td>900</td>
<td>27,000</td>
<td>48,000</td>
<td>33,730 (9:22.10)</td>
<td>10,065 (2:47.45)</td>
</tr>
<tr>
<td>1600</td>
<td>48,000</td>
<td>62,270</td>
<td>62,270 (17:17.50)</td>
<td>17,562 (4:52.42)</td>
</tr>
<tr>
<td>2500</td>
<td>50,000</td>
<td>64,010</td>
<td>22,396 (6:13.16)</td>
<td>18,648 (5:10.48)</td>
</tr>
</tbody>
</table>

Appendix D - Data generation and additional results for the simulation study

D.1 - Data generation

The data generated in the simulation study come from the following Poisson log-linear variant of the general model (3.1) in the main paper.

\[ Y_{kt} \sim \text{Poisson}(E_{kt}R_{kt}), \quad k = 1, \ldots, K = 271, \quad t = 1, \ldots, N = 9, \]

\[ \ln(R_{kt}) = \beta_0 + \phi_k + \sum_{s=1}^{4} \omega_{ks} f_s(t|\gamma_s), \]

where \(E_{kt} = 100\ \forall \ k \text{ and } t\). The random effects \(\phi\) are generated from the Leroux CAR model outlined by (3.2) in the main paper, with \(\rho = 0.99\) and \(\tau^2 = 0.001\). The temporal trend functions
$f_s(t|\gamma_s)$ used to generate the data are presented in Figure 2 in the main paper, and show constant, linearly increasing, linearly decreasing and change point forms.

**D.2 - Accuracy of the trend estimation**

The trends $f_s(t|\gamma_S)$ estimated in the simulation study are represented by:

1. Constant trend - $\beta_1$.
2. Linear increasing trend - $\beta_1 + \gamma_1 t$, with $\gamma_1 > 0$.
3. Linear decreasing trend - $\beta_1 + \gamma_2 t$, with $\gamma_2 < 0$.
4. Change point trend - $\beta_1 + \gamma_3 t + \gamma_4 (t - 5)_+$, with $\gamma_3 > 0$ and $\gamma_4 < 0$.

Tables 2 (scenarios (i) and (ii)) and 3 (scenarios (iii) and (iv)) display the bias, root mean square error (RMSE), and coverage probabilities of the 95% credible intervals for the parameters $(\beta_1, \gamma_1, \ldots, \gamma_4)$ associated with the 4 temporal trends, from using both the (MC)$^3$ and MCMC algorithms. Note, that in allocation mechanism (B) the only trend present in the data is the linearly decreasing one, and hence $(\gamma_1, \gamma_3, \gamma_4)$ do not have true values for the computation of the metrics described above. Unsurprisingly, the results mirror those observed in the classification percentages presented in the main paper, and have the following general trends.

For allocation mechanism (B) where all areas are assigned to a decreasing trend the (MC)$^3$ algorithm performs well in all scenarios, with all 95% credible intervals exhibiting close to the nominal coverage level and all estimates having negligible biases. Similar results are observed for the MCMC algorithm, with the exception being scenario (i) where the coverage probabilities drop to around 80% and the RMSE values are larger compared with the (MC)$^3$ algorithm. For allocation mechanisms (A) and (C) the (MC)$^3$ algorithm consistently outperforms the MCMC algorithm in scenarios (i) to (iii), with lower RMSE values in 29 out of 30 cases. This trend is
reversed in scenario (iv) where the MCMC algorithm gives the lowest RMSE values in 9 out of 10 cases, although in this case the magnitudes in the differences are generally not large. Finally, the (MC)$^3$ algorithm has coverage probabilities close to the nominal 95% levels in most cases, where as for the MCMC algorithm this is only true for scenario (iv). For scenarios (i) to (iii) the MCMC algorithm produces coverage probabilities between 11% and 89%, which is due to the relatively poor parameter estimation as measured by RMSE.

**Table 2:** Bias, root mean square error (RMSE) and coverage probabilities of the 95% credible intervals from the parameters of the temporal trends and the intercept term for scenarios (i) and (ii) and allocation mechanisms (A) to (C) using the (MC)$^3$ (left) and MCMC (right) algorithms.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Parameter</th>
<th>(MC)$^3$ algorithm</th>
<th>MCMC algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Coverage</td>
</tr>
<tr>
<td>(i) A</td>
<td>$\beta_1$</td>
<td>0.0041</td>
<td>0.0192</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>0.0084</td>
<td>0.0131</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>-0.0045</td>
<td>0.0221</td>
</tr>
<tr>
<td></td>
<td>$\gamma_3$</td>
<td>-0.0054</td>
<td>0.0087</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4$</td>
<td>0.0170</td>
<td>0.0523</td>
</tr>
<tr>
<td>(i) B</td>
<td>$\beta_1$</td>
<td>-0.0075</td>
<td>0.0506</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>0.0011</td>
<td>0.0114</td>
</tr>
<tr>
<td>(i) C</td>
<td>$\beta_1$</td>
<td>-0.0075</td>
<td>0.0506</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>0.0011</td>
<td>0.0114</td>
</tr>
<tr>
<td>(ii) A</td>
<td>$\beta_1$</td>
<td>-0.0004</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>0.0002</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>-0.0005</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>$\gamma_3$</td>
<td>-0.0005</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4$</td>
<td>0.0003</td>
<td>0.0076</td>
</tr>
<tr>
<td>(ii) B</td>
<td>$\beta_1$</td>
<td>-0.0001</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>-0.0001</td>
<td>0.0009</td>
</tr>
<tr>
<td>(ii) C</td>
<td>$\beta_1$</td>
<td>-0.0002</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>-0.0002</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>0.0000</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>$\gamma_3$</td>
<td>-0.0002</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4$</td>
<td>0.0003</td>
<td>0.0230</td>
</tr>
</tbody>
</table>

**D.3 - Results from not including a true trend in the model**

The simulation study presented above has fitted model (3.1) with all temporal trends that generated the data included as possible trends in the model. However, this does not illustrate
Table 3: Bias, root mean square error (RMSE) and coverage probabilities of the 95% credible intervals from the parameters of the temporal trends and the intercept term for scenarios (iii) and (iv) and allocation mechanisms (A) to (C) using the \( (MC)^3 \) (left) and MCMC (right) algorithms.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Parameter</th>
<th>(MC)(^3) algorithm</th>
<th>Standard algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>(iii) A</td>
<td>( \beta_1 )</td>
<td>0.0002</td>
<td>0.0044</td>
</tr>
<tr>
<td></td>
<td>( \gamma_1 )</td>
<td>0.0000</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>( \gamma_2 )</td>
<td>-0.0007</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>( \gamma_3 )</td>
<td>-0.0006</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>( \gamma_4 )</td>
<td>0.0033</td>
<td>0.0148</td>
</tr>
<tr>
<td>(iii) B</td>
<td>( \beta_1 )</td>
<td>-0.0004</td>
<td>0.0050</td>
</tr>
<tr>
<td></td>
<td>( \gamma_2 )</td>
<td>0.0000</td>
<td>0.0009</td>
</tr>
<tr>
<td>(iii) C</td>
<td>( \beta_1 )</td>
<td>-0.0001</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td>( \gamma_1 )</td>
<td>0.0001</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>( \gamma_2 )</td>
<td>-0.0005</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>( \gamma_3 )</td>
<td>-0.0001</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>( \gamma_4 )</td>
<td>0.0001</td>
<td>0.0095</td>
</tr>
<tr>
<td>(iv) A</td>
<td>( \beta_1 )</td>
<td>-0.0003</td>
<td>0.0043</td>
</tr>
<tr>
<td></td>
<td>( \gamma_1 )</td>
<td>0.0005</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>( \gamma_2 )</td>
<td>-0.0015</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>( \gamma_3 )</td>
<td>-0.0010</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>( \gamma_4 )</td>
<td>0.0022</td>
<td>0.0085</td>
</tr>
<tr>
<td>(iv) B</td>
<td>( \beta_1 )</td>
<td>-0.0008</td>
<td>0.0050</td>
</tr>
<tr>
<td></td>
<td>( \gamma_1 )</td>
<td>-0.0001</td>
<td>0.0009</td>
</tr>
<tr>
<td>(iv) C</td>
<td>( \beta_1 )</td>
<td>0.0003</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td>( \gamma_1 )</td>
<td>0.0009</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>( \gamma_2 )</td>
<td>-0.0010</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td>( \gamma_3 )</td>
<td>-0.0017</td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td>( \gamma_4 )</td>
<td>0.0067</td>
<td>0.0157</td>
</tr>
</tbody>
</table>

what happens if an important trend that is in the data is excluded in the model fitting. Therefore we repeated the simulation study by fitting model (3.1) without the linearly decreasing trend using the \( (MC)^3 \) estimation algorithm.

The results are displayed in Table 4 in two parts; (a) the percentages of those areas that were generated as having a decreasing trend (not included in the model) that were classified to the trends included in the model; and (b) for areas whose true trend was included in the model the percentage correct classification. The results from part (a) suggest that for all scenarios and allocation combinations, areas that have a decreasing trend are allocated to similar trends as expected, which is either the constant or change point trends. The latter can to some extent mimic an overall decreasing trend by the increasing part of the trend having a much smaller slope than the decreasing part of the trend. Unsurprisingly, none of these areas are allocated to
the increasing trend, and these results provides a sense check that the model performs as expected if an important trend is excluded from the model.

The results from part (b) show correct classification probabilities between 76% and 94%, which, while still quite high, are lower than those presented in the main study when all true trends are included in the model. This reduction in performance is not surprising, because not including the linearly decreasing trend as an option in the model causes those areas with a linearly decreasing trend to be misclassified to one of the other trends (constant or change point). This incorrect classification leads to incorrect parameter estimates for \((\beta, \gamma)\), leading to areas being misclassified as the estimated trends do not have the correct magnitude. Note, that for allocation mechanism (B) all areas have linearly decreasing trends, and hence there are no areas to consider in part (b) of the table.

**Table 4**: The results from omitting the linearly decreasing trend when fitting the model. Section (a) summarises which trends the areas with a true decreasing trend were assigned to. Section (b) summarises the overall correct classification percentages for areas whose true trend was included in the model.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>(a) Decreasing trend areas classified to</th>
<th>(b) Correct classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>constant</td>
<td>Increasing</td>
</tr>
<tr>
<td>(i) A</td>
<td>76.59%</td>
<td>0%</td>
</tr>
<tr>
<td>(i) B</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>(i) C</td>
<td>91.61%</td>
<td>0%</td>
</tr>
<tr>
<td>(ii) A</td>
<td>81.35%</td>
<td>0%</td>
</tr>
<tr>
<td>(ii) B</td>
<td>0.03%</td>
<td>0%</td>
</tr>
<tr>
<td>(ii) C</td>
<td>94.15%</td>
<td>0%</td>
</tr>
<tr>
<td>(iii) A</td>
<td>69.46%</td>
<td>0%</td>
</tr>
<tr>
<td>(iii) B</td>
<td>0.08%</td>
<td>0%</td>
</tr>
<tr>
<td>(iii) C</td>
<td>93.96%</td>
<td>0%</td>
</tr>
<tr>
<td>(iv) A</td>
<td>79.89%</td>
<td>0%</td>
</tr>
<tr>
<td>(iv) B</td>
<td>0.10%</td>
<td>0%</td>
</tr>
<tr>
<td>(iv) C</td>
<td>97.2%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Table 5 displays the trend allocation for each of the 1235 intermediate zones in Scotland for models (A) to (C), which have change points in 2002, 2004 and 2006 respectively. There is very little difference in these trend allocations, with almost all of the areas being classified to the change point trend. Furthermore, we refitted the model with hyperparameter values of $a = b = 0.001$ in the inverse-gamma prior for $\tau^2$, and the results again showed little change.

**Table 5:** Sensitivity analysis of the trend allocations from changing both the year of the change point and the hyperparameters in the inverse-gamma prior for $\tau^2$.

<table>
<thead>
<tr>
<th>Hyperparameters</th>
<th>Changepoint</th>
<th>Constant</th>
<th>Decreasing</th>
<th>Increasing</th>
<th>Change Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = 1, b = 0.1$</td>
<td>2002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1235</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1233</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1230</td>
</tr>
<tr>
<td>$a = b = 0.001$</td>
<td>2002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1235</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1233</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1230</td>
</tr>
</tbody>
</table>

**E.2 - Respiratory hospitalisation case study**

Table 6 displays the trend allocations for each of the intermediate zones in the Greater Glasgow and Clyde health board for the linear (A) and monotonic ((B) - 1 knot and (C) - 2 knots) increasing and decreasing trend models. The table shows very similar numbers of IZs allocated to each trend from each model, which shows the results are robust to this choice. Finally, the results from the area specific linear time trends model similar to that proposed by Bernardinelli and others (1995), and described in Section 3.4, is compared against our clustering model (3.1) in Figure 6. The figure displays boxplots of the posterior medians of the area specific temporal slope parameters $\{\delta_k\}$ grouped by which trend (decreasing, constant, increasing) model (3.1) assigned the area to. The figure shows that the two models are largely in agreement, in the sense that areas allocated to the decreasing trend have negative $\delta_k$ estimates, where as areas allocated to
Fig. 6: Boxplots displaying the estimated $\{\delta_k\}$ linear slope estimates grouped by which trend (decreasing, constant, increasing) model (3.1) assigned the area to.

the increasing trend have positive $\delta_k$ estimates.

Table 6: Sensitivity analysis of the trend allocations from using linear and non-linear temporal trends and hyperparameters in the inverse-gamma prior for $\tau^2$.

<table>
<thead>
<tr>
<th>Hyperparameters</th>
<th>Model</th>
<th>Temporal trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant</td>
<td>Decreasing</td>
</tr>
<tr>
<td>$a = 1, b = 0.1$</td>
<td>Linear</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Monotonic (1)</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Monotonic (2)</td>
<td>127</td>
</tr>
<tr>
<td>$a = b = 0.001$</td>
<td>Linear</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>Monotonic (1)</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Monotonic (2)</td>
<td>128</td>
</tr>
</tbody>
</table>

REFERENCES

REFERENCES


[Received August 1, 2010; revised October 1, 2010; accepted for publication November 1, 2010]