



HAL
open science

Bayesian nonparametric spatial prior for traffic crash risk mapping: a case study of Victoria, Australia

Jean-Baptiste Durand, Florence Forbes, Cong Duc Phan, Long Truong, Hien D Nguyen, Fatoumata Dama

► **To cite this version:**

Jean-Baptiste Durand, Florence Forbes, Cong Duc Phan, Long Truong, Hien D Nguyen, et al.. Bayesian nonparametric spatial prior for traffic crash risk mapping: a case study of Victoria, Australia. 2021. hal-03138803v1

HAL Id: hal-03138803

<https://inria.hal.science/hal-03138803v1>

Preprint submitted on 11 Feb 2021 (v1), last revised 7 Sep 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian nonparametric spatial prior for traffic crash risk mapping: a case study of Victoria, Australia

J.-B. Durand^{1*}, F. Forbes¹, C.D. Phan², L. Truong², H.D. Nguyen², and F. Dama¹

Inria, LJK, Statisfy team and La Trobe University

Summary

We investigate the use of Bayesian nonparametric (BNP) models coupled with Markov random fields (MRF) in a risk mapping context, to build partitions of the risk into homogeneous spatial regions. In contrast to most existing methods, the proposed approach does not require an arbitrary commitment to a specified number of risk classes and determines their risk levels automatically. We consider settings in which the relevant information are counts and propose a so called BNP Hidden MRF (BNP-HMRF) model that is able to handle such data. The model inference is carried out using a variational Bayes Expectation–Maximisation algorithm and the approach is illustrated on traffic crash data in the state of Victoria, Australia. The obtained results corroborate well with the traffic safety literature. More generally, the model presented here for risk mapping offers an effective, convenient and fast way to conduct partition of spatially localised count data.

Key words: Road safety; Traffic crashes; Risk mapping; Bayesian nonparametrics; Markov random field; Variational Bayes Expectation–Maximisation algorithm

1. Introduction

Traffic-related injuries and deaths are major problems in contemporary societies. Social economic losses from traffic crashes, in particular from motor vehicle crashes, are enormous. This makes road and traffic safety a major concern, worldwide. The nondecreasing relationship between crash casualties and population suggests that safety improvements could be gained from a better prediction of crash occurrences. Traffic crashes are complex events involving the interactions of various factors. In particular, since road transport involves distances by nature, most studies call for spatial analysis to account for geographical locations and environments in which crashes occur. The goal is often to accurately predict the risks at

*Author to whom correspondence should be addressed.

¹ Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Inria Grenoble Rhone-Alpes, 655 av. de l'Europe, 38335 Montbonnot, France

² School of Engineering and Mathematical Sciences, La Trobe University, Bundoora, Australia

Email: Jean-Baptiste.Durand@inria.fr

Acknowledgment. The authors are partly supported by the Inria project Lander.

Opinions and attitudes expressed in this document, which are not explicitly designated as Journal policy, are those of the author and are *not* necessarily endorsed by the Journal, its editorial board, its publisher Wiley or by the Australian Statistical Publishing Association Inc.

16 different locations (Lord & Mannering 2010) and to link these risk values to other variables
17 for interpretability, or to assess the impact of several risk factors (Theofilatos & Yannis 2014;
18 Papadimitriou et al. 2019) and road safety measures (Elvik et al. 2009). The hope is to
19 identify the potential causal sources of crashes, and to apply appropriate control procedures
20 and protection measures; see e.g., Truong & Currie (2019).

21 In this work, our goal is not primarily to predict the number of crashes or the related risk
22 as per se, although predictions are an output of our model. We primarily aim at highlighting
23 areas with different risk levels, with respect to various covariates, such as population density,
24 traffic density, signalisation density, etc. The interest of such partitioning is to highlight spatial
25 heterogeneity, to locate high risk areas (so called risk hot spots), and to determine whether
26 they exhibit some structure in space that could be analysed or directly interpreted. Moreover,
27 aggregation of connected regions with low numbers of crashes and the same risk level can
28 also be used to increase the effective sample size dedicated to estimating that risk level.

29 Such partitions can be obtained by applying risk mapping models. Standard risk
30 mapping models usually produce a continuous estimation of the risk that requires a post-
31 processing classification step to obtain clearly delimited risk zones, usually based on the
32 arbitrary choice of risk levels or of risk intervals. Sophisticated ways to set these thresholds
33 are illustrated in the Crash Risk Mapping Technical Specifications report of the European
34 Road Assessment Programme: <https://eurorap.org/crash-rate-mapping/>.

35 Most statistical methods for risk mapping of aggregated data (e.g., Mollié 1999;
36 Richardson et al. 1995; Pascutto et al. 2000; Lawson et al. 2000) are based on a Poisson log-
37 linear mixed model and follow the so called BYM model of Besag, York & Mollié (1991),
38 and extended by Clayton & Bernadinelli (1992), which is called the convolution model by
39 Mollié (1996). This model is based on a Hidden Markov Random Field (HMRF), where the
40 latent intrinsic risk field is modeled by a Markov field with continuous state space, namely a
41 Gaussian Conditionally Auto-Regressive (CAR) model. Other developments in this context
42 concern spatio-temporal mapping (Knorr-Held & Richardson 2003; Robertson et al. 2010;
43 Lawson & Song 2010) and multivariate risk mapping (Knorr-Held, Rasser & Becker 2002;
44 MacNab 2010).

45 For all of these procedures, the model inference results in a real-valued estimation of the
46 risk at each location and one of the main reported limitations is that local discontinuities
47 in the risk field are not modelled (see e.g., Green & Richardson 2002), leading to
48 potentially oversmoothed risk maps. Also, in some cases, coarser representations where
49 areas with similar risk values are grouped are desirable (e.g., Abrial et al. 2005). Grouped
50 representations have the advantage of providing clearly delimited areas in which more
51 focused studies could be conducted to better understand the crashes determinants. These areas
52 at risk can be viewed as clusters as in Knorr-Held & Rasser (2000), but we prefer to interpret

53 them as risk classes, as in the seminal work of Schlattmann & Böhning (1993) and Böhning,
54 Dietz & Schlattmann (2000), and with additional spatial constraints by Green & Richardson
55 (2002) and Alfo, Nieddu & Vicari (2009).

56 Indeed, geographically separated areas representing different clusters can have similar
57 risks and be grouped in the same class. Consequently, the classes can be less numerous than
58 the clusters, and their interpretation is made easier for decision-makers. Using the BYM
59 model, it is possible to derive such a grouping from the output, using either fixed risk ranges
60 (usually difficult to choose in practice), or more automated clustering techniques (see e.g.,
61 Fraley & Raftery 2007). In any case, this post-processing step is likely to be sub-optimal.
62 For this reason, there have been several attempts to design procedures that can directly model
63 such a risk classification.

64 Green & Richardson (2002) propose replacing the continuous risk field by a partition
65 model, involving the introduction of a finite number of risk levels and allocation variables to
66 assign each area under study to one of these levels. Spatial dependencies are then taken into
67 account by modeling the allocation variables as a latent discrete state-space Markov field,
68 such as the Ising (two classes or states) or Potts (more than two classes) model (Chandler
69 1987; Stoehr 2017). In the same spirit, in the work by Fernandez & Green (2002), the
70 spatial dependence is pushed one level higher resulting in a more flexible model, but more
71 difficult parameter estimation problem. These various attempts are based on discrete HMRF
72 modelling and all use MCMC techniques for inference, which can seriously limit and even
73 prevent their application to large data sets in a reasonable time.

74 Beyond inference, one additional concern is the automatic selection of the proper
75 number of classes in the data, or equivalently the number of states in the HMRF. In the
76 independent data case, several criteria exist to select this number automatically based on
77 penalised likelihoods (e.g., the Akaike information criterion, Bayesian information criterion,
78 integrated classification likelihood, etc.) and have been extended in the HMRF framework
79 using variational approximations (Forbes & Peyrard 2003). They require running several
80 models with different class numbers so as to choose the best one, with a potential waste
81 of computational effort as all but the best model are usually discarded. Other techniques
82 use a fully Bayesian setting, including a prior on the number of components, with the
83 most celebrated method being the reversible jump Markov chain Monte Carlo (Green 1995).
84 Although simplifications in inference have been proposed by Miller & Harrison (2018), the
85 computational cost of reversible jump techniques remains considerably high.

86 In this work, to handle discontinuities in the spatial structure of the risk, without having
87 to arbitrary choose their number, we propose to operate in the framework of Bayesian
88 nonparametric (BNP) methods (Ghosal & Van der Vaart 2017). More specifically, we build
89 on methods recently proposed for the modelling of continuous observations by Lu, Arbel &

90 Forbes (2020). We extend the approach, referred to as BNP-HMRF, to the modelling of count
 91 data. We derive the corresponding variational Bayes Expectation–Maximisation (VBEM)
 92 algorithm for the model estimation. The approach is then illustrated on traffic crash data
 93 in the state of Victoria, Australia. The analysis provides risk zones and risk levels that are
 94 globally coherent with other findings in the literature.

95 The proposed BNP-HMRF model is explained in Section 2. The model implementation
 96 using variational approximation is detailed in Section 3. The application to crash risk mapping
 97 of Victoria is detailed in Section 4. A conclusion and perspectives are provided in Section 5.

98 2. BNP-HMRF model for count data

99 The study aims at providing some risk mapping of traffic crashes, based on data
 100 regarding geographical zones. Since, on average, the number of traffic crashes increases with
 101 respect to other variables characterising the traffic importance (e.g., population size, traffic
 102 intensity, length of road network), the numbers of crashes have to be normalised with respect
 103 to at least one of these variables. The obtained ratio provide what we interpret as risks. One
 104 objective is then to account for some spatial heterogeneity regarding the observed risks. The
 105 model described in the following lines aims at clustering regions with close risks to provide
 106 a labelled map, where each label is associated with some risk level.

Considering J regions, where we let y_j represent the number of crashes occurring in
 region $j \in \{1, \dots, J\}$, characterised by a normalisation variable N_j ; e.g., the population size
 of region j . For the sake of clarity, we first assume that there is a finite set of K risk levels
 $\mathbf{\Lambda} = \{\lambda_0, \dots, \lambda_{K-1}\}$, that are ordered so that λ_k is the $(k + 1)$ th smallest level. Since the
 risk level associated to region j is not known in advance, a variable $z_j \in \{0, \dots, K - 1\}$ is
 introduced to indicate the assigned risk level, i.e. $z_j = k$, when region j is at risk level λ_k .
 When region j is at risk level λ_k , the number of traffic crashes y_j is then assumed to be
 Poisson distributed with mean $\lambda_k N_j$. That is, y_j conditioned on $z_j = k$ has probability mass
 function

$$p(y_j | z_j = k; \mathbf{\Lambda}, N_j) = \mathcal{P}(y_j; \lambda_k N_j), \quad (1)$$

107 where $\mathcal{P}(\cdot; \lambda_k N_j)$ denotes the probability mass function of the Poisson distribution with
 108 parameter $\lambda_k N_j$. From a generic point of view, $p(y_j | z_j = k; \mathbf{\Lambda}, N_j)$ is referred to as an
 109 emission distribution. As a consequence, the mean number of crashes is a linear function
 110 of N_j : $E[y_j | z_j = k; \mathbf{\Lambda}, N_j] = \lambda_k N_j$.

111 The goal is then to estimate the risk levels $\mathbf{\Lambda} = \{\lambda_0, \dots, \lambda_{K-1}\}$ and the most likely risk
 112 mapping through the most likely values of the z_j s. In practice, risk levels are likely to vary

113 smoothly across regions. It is more likely that neighbouring regions have the same risk level
 114 with possible abrupt changes from a region to another if they have contrasting characteristics.
 115 Thus, for a better estimation of risk levels, a Markov random field (MRF) model is used
 116 for the set of labels $\mathbf{z} = \{z_j, j \in J\}$, to account for spatial dependencies between connected
 117 regions.

118 Formally, the regions are seen as the vertices of a graph G . They are connected by an
 119 edge in the graph whenever they share a boundary, although other types of connections could
 120 be considered (e.g., they either share a boundary or have a common neighbour, etc.). The
 121 probability for neighbouring regions having either a similar or different label is controlled
 122 by some scalar positive parameter denoted by β . The higher the value of β , the more likely
 123 neighbouring regions are at the same risk level.

124 The number K of risk levels is not usually known in advance and has to be chosen
 125 adaptively by users. To avoid this commitment to a fixed number K , we propose an extension
 126 of the model that does not restrict the levels to a finite number K . This extension is based
 127 on so called Dirichlet Process Mixtures (Lu, Arbel & Forbes 2020) and is referred to as a
 128 Bayesian Non-Parametric Hidden Markov Random Field (or more concisely, BNP-HMRF).

129 The BNP-HMRF model is defined as follows. The set of J regions under consideration is
 130 associated to a graph structure $G = (J, E)$, where each $j \in J$ corresponds to a node of G and
 131 the set of edges E represents all pairs of regions with a common boundary. The likelihood
 132 part of the model is given by (1). The observations are counts $\mathbf{y} = \{y_j, j \in J\}$ distributed
 133 independently given $\mathbf{\Lambda}$ and \mathbf{z} with for every j ,

$$p(y_j | z_j = k; \mathbf{\Lambda}, N_j) = \mathcal{P}(y_j; \lambda_k N_j).$$

The risk class labels $\mathbf{z} = \{z_j, j \in J\}$ are assumed to be distributed as a Markov random field
 on G with the following distribution:

$$\begin{aligned} p(\mathbf{z} | \beta, \boldsymbol{\pi}) &\propto \exp \left(\sum_{j=1}^J \ln \pi_{z_j} + \beta \sum_{\{i,j\} \in E} \mathbf{1}_{(z_i=z_j)} \right) \\ &= \left(\prod_{j=1}^J \pi_{z_j} \right) \exp \left(\beta \sum_{\{i,j\} \in E} \mathbf{1}_{(z_i=z_j)} \right), \end{aligned} \quad (2)$$

134 where $\mathbf{1}_{(z_i=z_j)}$ is the indicator function equal to 1 when $z_i = z_j$ and 0 otherwise, $\{i, j\} \in E$
 135 indicates that $\{i, j\}$ is an edge in G , β is some unknown scalar parameter, and the π_k s are

136 weights defined for every $k \geq 0$ as

$$\pi_k(\boldsymbol{\tau}) = \tau_k \prod_{l < k} (1 - \tau_l),$$

where $\boldsymbol{\tau}^\top = (\tau_0, \tau_1, \dots)$ is a sequence of independent, identically distributed (i.i.d.) random variables with distribution $\text{Beta}(1, \alpha)$. The parameter α is an hyperparameter which follows a gamma distribution:

$$\alpha | s_1, s_2 \sim \mathcal{G}(s_1, s_2),$$

while each parameter λ_k in (1) is also distributed according to a gamma distribution:

$$\lambda_k | a_k, b_k \sim \mathcal{G}(a_k, b_k).$$

The construction of the π_k s corresponds to a stick-breaking construction (see Lemma 3.4 in Ghosal & Van der Vaart 2017 and Sethuraman 1994, for details) and guarantees that $\sum_{k=0}^{\infty} \pi_k = 1$, which in turns ensures that the distribution in (2) is a valid Markov field (see Lu, Arbel & Forbes 2020, for details). The complete hierarchical model can thus be stated, for $\mathbf{z} = \{z_1, \dots, z_J\}$ and $k = 0, 1, \dots$, as follows:

$$\begin{aligned} \lambda_k | a_k, b_k &\sim \mathcal{G}(a_k, b_k), \\ \alpha | s_1, s_2 &\sim \mathcal{G}(s_1, s_2), \\ \tau_k | \alpha &\sim \mathcal{B}(1, \alpha), \\ \pi_k(\boldsymbol{\tau}) &= \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), \\ p(\mathbf{z} | \boldsymbol{\tau}, \beta) &\propto \prod_{j \in J} \pi_{z_j}(\boldsymbol{\tau}) \exp(\beta \sum_{\{i,j\} \in E} \mathbf{1}_{(z_i=z_j)}), \\ y_j | z_j; \boldsymbol{\Lambda}, N_j &\sim \mathcal{P}(y_j; \lambda_{z_j} N_j), \quad \text{for each } j \in J. \end{aligned}$$

137

3. Inference using variational approximation

138 In what follows, we propose an adaptation of the VBEM procedure from Lu, Arbel &
139 Forbes (2020) to Poisson emission probabilities.

140 The observed counts are denoted by $\mathbf{y} = \{y_j, j \in J\}$ and the normalising variables
141 are denoted by $\mathbf{N}_J = \{N_j, j \in J\}$. The set of parameters to estimate divides into
142 two subsets, $\boldsymbol{\Phi}^\top = (\beta, s_1, s_2, \mathbf{a}^\top)$ with $\mathbf{a} = \{a_k, b_k, k = 1, \dots\}$, which are unknown but
143 fixed parameters, and $\boldsymbol{\Theta}^\top = (\boldsymbol{\alpha}^\top, \boldsymbol{\Lambda}^\top)$, which are random parameters. The hierarchical
144 representation of the model above induces additional latent variables $(\mathbf{z}^\top, \boldsymbol{\tau}^\top)$. There is no

145 formal difference in the treatment of random parameters and latent variables, but it is standard
 146 to distinguish them. The term latent usually refers to variables whose number increases with
 147 the number of observations, while parameters are usually of fixed dimension. Parameters Φ
 148 are determined by an empirical Bayes principle:

$$\hat{\Phi} = \arg \max_{\Phi} p(\mathbf{y}|\Phi) = \arg \max_{\Phi} \int p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \Theta|\Phi) dz d\boldsymbol{\tau} d\Theta.$$

149 The random parameters and latent variables are handled via their posterior density
 150 $p(\mathbf{z}, \boldsymbol{\tau}, \Theta|\mathbf{y}, N_J, \hat{\Phi})$. This posterior is not available in close-form due to an intractable
 151 normalising constant. In this work, we use a variational approximation principle to provide
 152 an approximation of the true posterior. More specifically, the posterior is approximated using
 153 a member of a family of distributions q that factorise and that are truncated appropriately to
 154 exhibit only a finite number of terms. The infinite state space for each z_i is dealt with by
 155 choosing a truncation of the state space to a maximum label K (Blei & Jordan 2006).

156 In practice, this consists of assuming that the variational distribution $q(\mathbf{z}) = \prod_{j \in J} q(z_j)$
 157 and that the q_{z_j} s satisfy $q_{z_j}(k) = 0$, for $k \geq K$, and that the variational distribution on $\boldsymbol{\tau}$ also
 158 factorises as $q_{\boldsymbol{\tau}}(\boldsymbol{\tau}) = \prod_{k=0}^{K-2} q_{\tau_k}(\tau_k)$, with the additional condition that $\tau_{K-1} = 1$. We thus
 159 have the variation approximation, below:

$$q(\mathbf{z}, \boldsymbol{\tau}, \Theta) = q_{\alpha}(\alpha) \prod_{j=1}^J q_{z_j}(z_j) \prod_{k=0}^{K-2} q_{\tau_k}(\tau_k) \prod_{k=0}^{K-1} q_{\lambda_k}(\lambda_k). \quad (3)$$

160 In (3), K does not represent the number of classes that is actually assumed to exist in the data
 161 but an upper bound of it. In practice, the exact value of K is not critical, K has to be fixed to
 162 a value large enough so as to be higher than the maximum expected number of classes.

163 The variational approximation procedure is justified by the now standard statement (see
 164 e.g., Lu, Arbel & Forbes 2020) that for every function $q(\mathbf{z}, \boldsymbol{\tau}, \Theta)$, the marginal likelihood
 165 is lower-bounded, where the lower bound is the Kullback-Leibler divergence between
 166 $p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \Theta|N_J, \Phi)$ and $q(\mathbf{z}, \boldsymbol{\tau}, \Theta)$,

$$\ln p(\mathbf{y}|\Phi) \geq E_{q(\mathbf{z}, \boldsymbol{\tau}, \Theta)} \left[\ln \frac{p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \Theta|N_J, \Phi)}{q(\mathbf{z}, \boldsymbol{\tau}, \Theta)} \right]. \quad (4)$$

167 The derived variational algorithm is then an alternate maximisation of this bound with
 168 respect to each factor in $q(\mathbf{z}, \boldsymbol{\tau}, \Theta)$ and Φ . Each maximisation step has some explicit
 169 functional expression although from the computational point of view, some steps may require
 170 further approximations. Their descriptions at a coarser level in terms of blocks of parameters
 171 is as follows. The iteration index is denoted by (r) in the successive update formulas:

172 • VE- \mathbf{z} :

$$q_z^{(r)}(\mathbf{z}) \propto \exp \left(E_{q_{\theta, \tau}^{(r-1)}} [\ln p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \boldsymbol{\Theta} | \mathbf{N}_J, \boldsymbol{\Phi}^{(r-1)})] \right); \quad (5)$$

173 • VE- $\boldsymbol{\Theta}, \boldsymbol{\tau}$:

$$q_{\theta, \tau}^{(r)}(\boldsymbol{\Theta}, \boldsymbol{\tau}) \propto \exp \left(E_{q_z^{(r)}} [\ln p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \boldsymbol{\Theta} | \mathbf{N}_J, \boldsymbol{\Phi}^{(r-1)})] \right); \quad (6)$$

174 • VM- $\boldsymbol{\phi}$:

$$\boldsymbol{\Phi}^{(r)} = \arg \max_{\boldsymbol{\phi}} E_{q_z^{(r)} q_{\theta, \tau}^{(r)}} [\ln p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \boldsymbol{\Theta} | \mathbf{N}_J, \boldsymbol{\Phi})]. \quad (7)$$

175 Some of these steps are not specific to the Poisson emissions model and have similar
176 forms as the ones derived by Lu, Arbel & Forbes (2020) for Gaussian emissions. These steps
177 are briefly recalled below, omitting the iteration index for simplicity.

VE- α step

We have

$$q_\alpha(\alpha) \propto p(\alpha | s_1, s_2) \exp \left(\sum_{k=0}^{K-2} E_{q_{\tau_k}} [\ln p(\tau_k | \alpha)] \right) = \mathcal{G}(\alpha; \hat{s}_1, \hat{s}_2),$$

with

$$\begin{aligned} \hat{s}_1 &= s_1 + K - 1, \\ \hat{s}_2 &= s_2 - \sum_{k=0}^{K-2} E_{q_{\tau_k}} [\ln(1 - \tau_k)], \end{aligned}$$

178 and

$$E_{q_{\tau_k}} [\ln(1 - \tau_k)] = \psi(\gamma_{k,2}) - \psi(\gamma_{k,1} + \gamma_{k,2}), \quad (8)$$

179 where $\psi(\cdot)$ is the digamma function and $(\gamma_{k,1}, \gamma_{k,2})$ are the parameters defining q_{τ_k} (see
180 below).

VE- τ_k step

For every $0 \leq k < K$,

$$q_{\tau_k}(\tau_k) = \mathcal{B}(\tau_k; \hat{\gamma}_{k,1}, \hat{\gamma}_{k,2})$$

181 with

$$\hat{\gamma}_{k,1} = 1 + \sum_{j=1}^J q_{z_j}(k) = 1 + n_k,$$

182 and

$$\hat{\gamma}_{k,2} = \frac{\hat{s}_1}{\hat{s}_2} + \sum_{l=k}^{K-1} n_l,$$

183 where we have used the notation $n_k = \sum_{j=1}^J q_{z_j}(k)$. Note also that $\sum_{k=0}^{K-1} n_k = J$.

184 **VM- β step**

185 The handling of parameter β is particularly nuanced. Standard variational approximations
 186 focus on solving the posterior intractability resulting from the combination of prior
 187 distributions and likelihoods, both usually tractable. Unfortunately, when dealing with
 188 Markov random field priors, the issue does not only arise from the prior/likelihood
 189 combination but already from the intractable normalising constant of the Markov prior. This
 190 constant cannot be discarded as it depends on β . Thus, the VM- β step has no explicit solution
 191 and requires further approximation. The additional approximation we propose consists of
 192 approximating the gradient, with respect to β , of the lower bound (4):

$$\sum_{k=0}^{K-1} \sum_{i \sim j} q_{z_j}^{(r)}(k) q_{z_i}^{(r)}(k) - \sum_{k=0}^{K-1} \sum_{i \sim j} \tilde{q}_{z_j}^{(r)}(k|\beta) \tilde{q}_{z_i}^{(r)}(k|\beta),$$

193 with $\tilde{q}_{z_j}^{(r)}(z_j|\beta)$, defined by:

$$\text{for all } k = 0 \dots K-1, \quad \tilde{q}_{z_j}^{(r)}(z_j = k|\beta) = \frac{\exp(\ln \pi_k(\tilde{\tau}) + \beta \sum_{i \in N(j)} q_{z_i}^{(r)}(k))}{\sum_{l=0}^{\infty} \exp(\ln \pi_l(\tilde{\tau}) + \beta \sum_{i \in N(j)} q_{z_i}^{(r)}(l))} \quad (9)$$

194 where $N(j)$ is the set of neighbours of j in G and $\tilde{\tau} = E_{q_{\tilde{\tau}}}[\tau]$. Note that in the BNP-HMRF
 195 setting, the use of $\tilde{\tau}$ is also an additional necessary approximation; details can be found in
 196 Lu, Arbel & Forbes (2020). Regarding β , the approximation can be interpreted as the transfer
 197 to the Potts prior of the variational approximation used for the Markov posterior.

198 The remaining steps are specific to Poisson emission distributions.

VE- λ_k step

For every $0 \leq k < K$, from the general result in (6) it follows that

$$\begin{aligned}
 q_{\lambda_k}(\lambda_k) &\propto \exp\left(\mathbb{E}_{q_z} \left[\ln \left(p(\lambda_k | a_k, b_k) \prod_{j=1}^J p(y_j | z_j, \lambda_{z_j}, N_j) \right) \right]\right) \\
 &\propto p(\lambda_k | a_k, b_k) \exp\left(\sum_{j=1}^J q_{z_j}(k) \mathbb{E}_{q_{z_j}} \left[\ln p(y_j | \lambda_k, N_j) \right]\right) \\
 &\propto p(\lambda_k | a_k, b_k) \exp\left(\sum_{j=1}^J q_{z_j}(k) \left(-N_j \lambda_k + y_j \ln(N_j \lambda_k) - \ln(y_j!) \right)\right) \\
 &\propto p(\lambda_k | a_k, b_k) \exp\left(\sum_{j=1}^J q_{z_j}(k) \left(-N_j \lambda_k + y_j \ln(\lambda_k) \right)\right).
 \end{aligned}$$

199 Thus q_{λ_k} is a gamma density $\mathcal{G}(\lambda_k | \hat{a}_k, \hat{b}_k)$, where

$$\hat{a}_k = a_k + \sum_{j=1}^J y_j q_{z_j}(k) \quad \text{and} \quad \hat{b}_k = b_k + \sum_{j=1}^J N_j q_{z_j}(k).$$

200 This is a consequence of the conjugacy property of gamma priors for Poisson likelihood
 201 functions.

202 VE- Z_j step

203 For $j \in J$, the general result in (5) leads to

$$q_{z_j}(z_j) \propto \exp\left(\mathbb{E}_{q_{\lambda_{z_j}}} \left[\ln p(y_j | \lambda_{z_j}, N_j) \right] + \mathbb{E}_{q_\tau} \left[\ln \pi_{z_j}(\tau) \right] + \beta \sum_{i \in N(j)} q_{z_i}(z_j) \right), \quad (10)$$

where $N(j)$ is the set of neighbours of j in G , and for $z_j = k$,

$$\mathbb{E}_{q_\tau} \left[\ln \pi_k(\tau) \right] = \mathbb{E}_{q_{\tau_k}} \left[\ln \tau_k \right] + \sum_{l=0}^{k-1} \mathbb{E}_{q_{\tau_l}} \left[\ln(1 - \tau_l) \right],$$

with

$$\mathbb{E}_{q_{\tau_k}} \left[\ln(\tau_k) \right] = \psi(\gamma_{k,1}) - \psi(\gamma_{k,1} + \gamma_{k,2}),$$

204 where $\mathbb{E}_{q_{\tau_l}} \left[\ln(1 - \tau_l) \right]$ is given by (8).

The term $\mathbb{E}_{q_{\lambda_k}} \left[\ln p(y_j | \lambda_k, N_j) \right]$ is obtained as follows:

$$\begin{aligned} \mathbb{E}_{q_{\lambda_k}} \left[\ln p(y_j | \lambda_k, N_j) \right] &= \mathbb{E}_{q_{\lambda_k}} \left[-N_j \lambda_k + y_j \ln(N_j \lambda_k) - \ln(y_j!) \right] \\ &= -N_j \mathbb{E}_{q_{\lambda_k}} \left[\lambda_k \right] + y_j \ln(N_j) + y_j \mathbb{E}_{q_{\lambda_k}} \left[\ln(\lambda_k) \right] - \ln(y_j!), \end{aligned}$$

205 where due to the fact that q_{λ_k} is a gamma density $\mathcal{G}(\lambda_k | \hat{a}_k, \hat{b}_k)$ (see above), $\mathbb{E}_{q_{\lambda_k}} \left[\lambda_k \right] =$
 206 \hat{a}_k / \hat{b}_k and $\mathbb{E}_{q_{\lambda_k}} \left[\ln(\lambda_k) \right] = \psi(\hat{a}_k) - \ln(\hat{b}_k)$.

207 **VM- (a_k, b_k) step**

208 For every $0 \leq k < K$, the VM- (a_k, b_k) step is conducted by maximising

$$\mathbb{E}_{q_{\lambda_k}} \left[\ln \frac{p(\lambda_k | a_k, b_k)}{q_{\lambda_k}(\lambda_k)} \right],$$

209 which is equivalent to minimising the Kullback-Leibler divergence between $p(\lambda_k | a_k, b_k)$ and
 210 q_{λ_k} . Since both densities are gamma, the minimum is obtained whenever $p(\lambda_k | a_k, b_k) = q_{\lambda_k}$
 211 and thus $(a_k, b_k) = (\hat{a}_k, \hat{b}_k)$.

212 **VM- (s_1, s_2) step**

213 The VM- (s_1, s_2) step is conducted by maximising

$$\mathbb{E}_{q_\alpha} \left[\ln \frac{p(\alpha | s_1, s_2)}{q_\alpha(\alpha)} \right].$$

214 By a similar argument as for the VM- (a_k, b_k) step, $(s_1, s_2) = (\hat{s}_1, \hat{s}_2)$, since both densities
 215 are in the gamma family.

216 Finally, to start the iterative procedure, initial values of the parameters have to be
 217 provided and the choice generally has an effect on the quality of the final estimates. Here
 218 we resort to several runs of the k -Means algorithm with random initial labels and using the
 219 ratio Y_j / N_j as input data. For each run, the final clustering yields a value \hat{z}_j for each z_j , and
 220 $q_{z_j}(k) = \mathbf{1}_{(k=\hat{z}_j)}$. Parameters $\hat{\mathbf{A}}_k$ are initialised as the sample means of observations within
 221 initial class k . Initial values of parameters a_k and b_k are defined so that $\mathbb{E}_{q_{\lambda_k}} \left[\lambda_k \right] = \hat{\mathbf{A}}_k$ and
 222 $\text{var}_{q_{\lambda_k}} \lambda_k = \min_{\ell, \hat{\mathbf{A}}_\ell > 0} \hat{\mathbf{A}}_\ell$. For each of these possible initial variational parameters from the
 223 multiple runs, we keep the value that yields the highest free energy (4). The computation
 224 of the free energy is detailed in Appendix I. In this step, initial values are also required for
 225 (s_1, s_2, β) , which are arbitrarily set to $(1.4, 1, 0)$.

226

4. Application to traffic crash risk mapping

227

The results presented here comprise of data from Victoria, Australia. Crash data between 2014 and 2018 were obtained from Victoria's open data directory (see, Truong & Currie 2019). The number of traffic crashes were aggregated at statistical areas level 2 (SA2s), which are medium-sized functional areas within the Australian Statistical Geography Standard (ASGS). The total number of SA2s in Victoria is 458, excluding several SA2s that have no population according to the Australian Bureau of Statistics (ABS) 2016 census. The SA2 scale has the advantage of offering a good compromise between the number of regions and the spatial resolution, thus ensuring both tractability of algorithms and the interpretability of results.

236

Since an absolute validation of the risk mapping is difficult, we resort to covariates to assess whether risk level classes encode relevant and contrasted characteristics between different classes. In practice, the set of covariates is the same as the set of possible normalising variables. The primary set of covariates is given in the upper part of Table 1. Additional covariates are derived from the primary set of covariates, whenever they appear to make sense as normalising variables. These derivatives are listed in the lower part of Table 1. The road density variable (RnDens) roughly corresponds to the squared ratio of the road length on the edge length of a square region. The VtrFAR19 variable represents the traffic density as opposed to the absolute traffic load VktFAR19. The VtpopFAR19 variable represents the traffic load per inhabitant.

4.1. Method

247

We distinguish between two kinds of analyses: exploratory analyses and full runs of the model. Exploratory analyses aim at quickly assessing the potential of several variables as normalising factors for the risk. In this context we resort to reduced numbers of random initialisations and iterations (1,000 and 300, respectively), while full runs may require more initial values and iterations for some variables (see hereinafter). The methodology regarding exploratory analyses is the following:

253

1. A principal component analysis (PCA) is performed on the variables listed in Table 1.

254

If a group of variables appears to be strongly correlated to the same axis, we keep only one variable in the group and discard all others.

255

256

2. For each possible normalising variable N_j , the ratio $(y_j/N_j)_{1 \leq j \leq J}$ are computed,

257

quantised into seven bins and represented on a map. Boundaries between bins are

258

defined by sample quantiles. These bins are anticipated to be close to the expected

259

initial partition.

Table 1. Available covariates: primary set (upper part) and additional derived ones (lower part) with their description sometimes truncated.

Code	Description
AREASQ16	Square area in km ²
pop16	Population in 2016 (resident)
VktFAR19	Vehicle-Kilometres Travelled on Freeways and Arterial roads
RnNoItcNS	Number of non-signalised intersection (excl. roundabout)
RnNoItcS	Number of signalised intersection
RnNoItcR	Number of roundabout intersection
RnNoItc	Total number of intersection (sum of 3 types)
RnLen	Total length of all road network (km)
RnSZ7080	Length of posted speed 70 - 80 km/h road (km)
RnSZ90100	Length of posted speed 90 - 100 km/h road (km)
RnSZo100	Length of posted speed over 100 km/h road (km)
RnDens	road density ($RnLen^2/AREASQ16$)
PopDens	population density ($pop16/AREASQ16$)
VtrFAR19	VktFAR19/RnLen, Vehicle-Kilometres Travelled on Freeways etc. per km of road
VtpopFAR19	VktFAR19/pop16, Vehicle-Kilometres Travelled on Freeways etc. per inhabitant
PropSign	proportion of signalised intersections over total number of intersections

- 260 3. The numbers of crashes are plotted against the normalising variable. If the model was
261 adequately characterising the observed data, for each risk level, the average number of
262 crashes should be a linear function of N_j , so that the 2D plot should exhibit sets of
263 points clustered around a number of lines that pass through the origin. In particular,
264 if there was only one risk level, the whole data set would exhibit a linear structure.
265 Therefore, as a first evaluation, a linear regression model is estimated and the regression
266 line is presented. If needed, non-linear transforms of N_j are considered (exponential,
267 logarithm, monomials) until the figure exhibits linear structures. This step is purely
268 exploratory.
- 269 4. The BNP-HMRF is estimated with $K = 10$ and provides for each region its probability
270 to be at each risk level. A segmentation or partition of the graph into K segments is
271 then obtained by assigning each region/node to the level with the highest probability.
272 The segmentation is represented as a K -color map.
- 273 5. The 2D plot described in Step 3 is drawn again, now using different colours, where
274 one colour is used for each label. To assess label separation, lines passing through the
275 origin are drawn using slopes $E_{q\lambda_k} \left[\lambda_k \right]$. These are compared with linear regression
276 models estimated separately for each class by least squares. Label separation is also
277 quantified by label marginal entropies represented on a map in grey scale.

278 6. If the number of estimated risk classes in the data is less than K , the model is
279 considered as potentially relevant.

280 In the latter case, VBEM is run again with more initial values and iterations if necessary (up
281 to 5,000 and 600, respectively), again with $K = 10$. If the relative growth of free energy falls
282 under 10^{-5} before the specified maximum number of iterations is reached, the algorithm is
283 considered to have converged and is stopped. Then analyses of variance are performed on
284 each covariate to interpret the risk classes and to examine how their values segregate within
285 regions. Tests are performed at significance level 0.01 and p -values are provided, as well as
286 per-label boxplots. If the number of risk classes is K , it may indicate that the BNP-HMRF
287 model has trouble finding well-separated classes and that the number of classes could be
288 arbitrarily high when the number of regions increases. In that case, a model with continuous
289 risk levels could be more appropriate.

290 The data, code, python notebook and system environments (conda, docker) required
291 to reproduce these analyses and results are available at [https://gitlab.inria.fr/
292 satisfy_public/anzj-crash/-/tree/master/notebooks](https://gitlab.inria.fr/statisfy_public/anzj-crash/-/tree/master/notebooks).

293 4.2. Results

294 4.2.1. Normalising variable selection

295 The first plane in the variable space of the PCA is depicted in Figure 1. The following
296 variables: RnNoItc, RnLen, RnSZ7080, RnDens, and RnSZ90100 were highly correlated
297 with AREASQ16 and thus were discarded. The variables RnSZo100 and VtpopFAR19 have
298 low variability over the regions and were thus not considered as informative. Among the
299 remaining variables, we illustrate here our procedure by considering the population size
300 (variable pop16) and the traffic density (variable VtrFAR19) as normalising variables in turn.
301 Results with other normalising variables are provided in Appendix II.

302 4.2.2. Risk with respect to the population size

303 We now consider the population size of a region as N_j , i.e., risks are clustered according
304 to the impact of population size on the number of crashes. For illustrative purposes, the map
305 obtained from Step 2 of the exploratory analysis is included in Appendix II as Figure 11. The
306 model led to seven clusters, among which four have negligible frequencies (see Figures 2
307 and 4). The frequency corresponds to the size of the cluster divided by the total number
308 of regions. Here we do not provide any detailed interpretation for 8 regions in four clusters
309 (clusters labelled as 5 to 9) that essentially are outliers (the regions correspond to zones with 6
310 to 184 inhabitants, where 9 to 209 crashes occur); see Figures 4 and 5. The estimate $\hat{\beta} \approx 0.34$

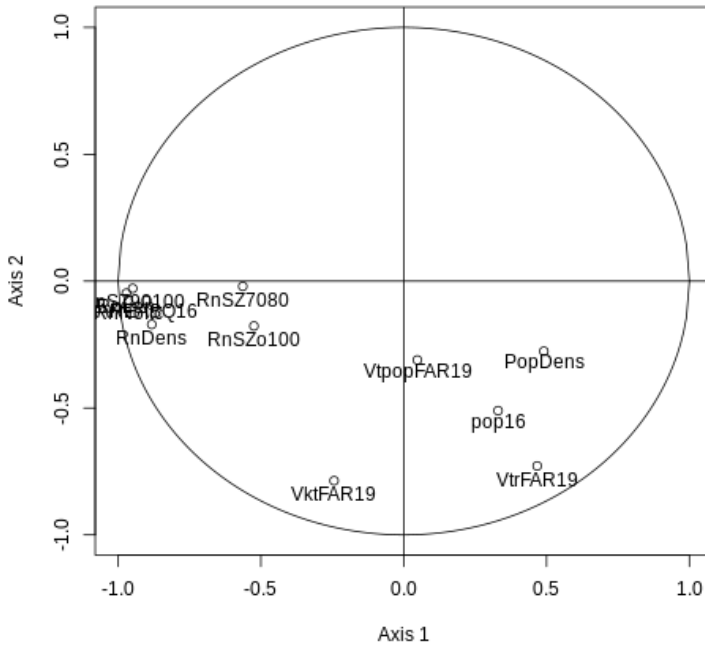


Figure 1. Principal component analysis on covariates: first plane in the variable space.

311 indicates rather low spatial aggregation of clusters. The cumulative marginal entropy (i.e.
 312 the sum over regions of the assignment probability entropies) is 20.7, indicating moderate
 313 uncertainty regarding the labels in some regions (compared to the other models considered,
 314 hereafter), although the marginal state entropy is below 0.2 in the majority of the regions (see
 315 Figure 3). Note that due to the model eliminating some states k during the estimation process
 316 (if and only if $\max_j q_{z_j}(k) < 0.5$), the remaining labels may not have consecutive indices
 317 since they are identified with their initial indexing in $\{0, \dots, K - 1\}$.

318 We ordered the labels by increasing values of risk levels. It can be seen from Figure 4
 319 that the fitted linear regression lines are in accordance with slopes induced by expected risk
 320 levels, indicating well-separated classes. There is however some larger discrepancy between
 321 the two lines in risk level 4, where the regression line has a larger slope than the expected risk
 322 level due to possible confusions between levels 4 and 2 in some regions.

323 Risk level 0 is related to peripheral regions that are close to the capital of Melbourne,
 324 and enclaves, which are often regional towns or rural centres with substantial residential

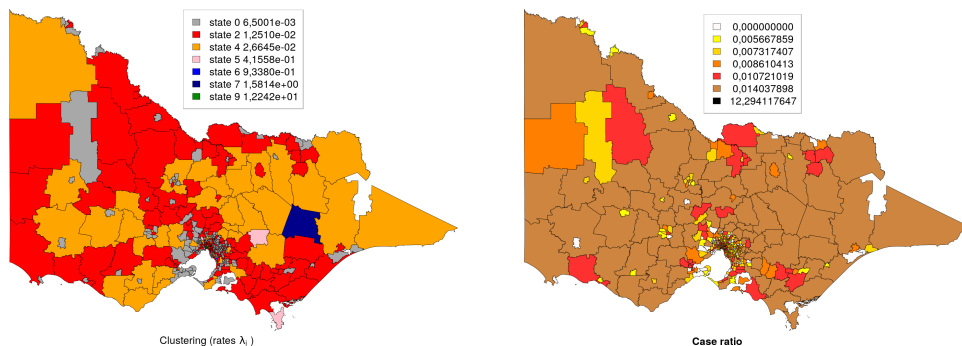


Figure 2. Risk mapping with respect to population size (variable pop16). Right-hand part: segmentation using quantiles on ratio.

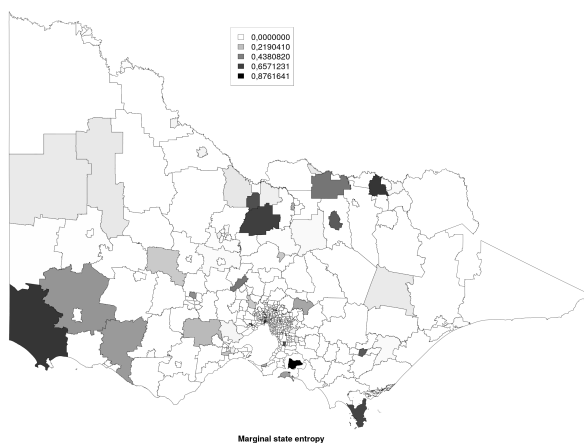


Figure 3. Marginal state entropy for each region regarding risk levels with respect to population size.

325 developments (see Figure 2[†]). These are small regions (Figure 5 b) with high population sizes
 326 (Figure 5 a), high absolute traffic (Figure 5 c) and high traffic densities (Figure 5 d). Risk level
 327 2 is related to peripheral and central regions. These are medium-sized zones with medium
 328 population sizes, absolute traffic and traffic densities. Risk level 4 is related to far peripheral
 329 and hypercentral regions (relative to the capital). These are sparsely populated, have varying
 330 sizes, with high absolute traffic and traffic densities. The four variables considered in Figure 5
 331 are well discriminated by the risk levels, with ANOVA p -values between 10^{-10} and 10^{-15}
 332 regarding the effects of the classes.

333 Since regions with higher population sizes have lower risks, it is possible that population
 334 has some non-linear effect, which is an avenue for further investigation.

[†]High-resolution maps are available as supporting information available in online publication

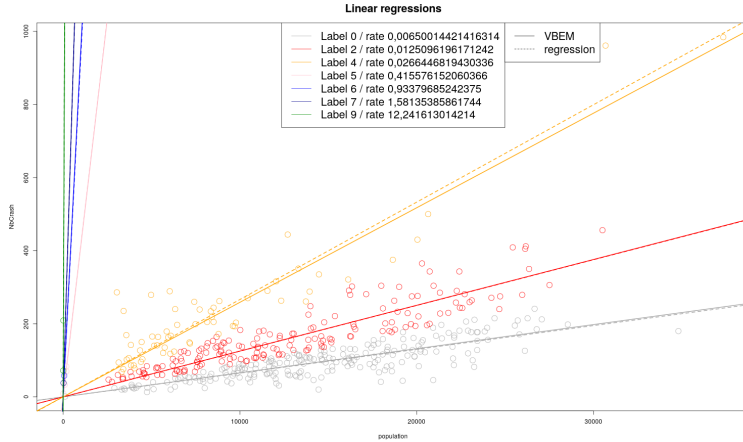


Figure 4. Cluster-wise regressions of crash numbers on population size.

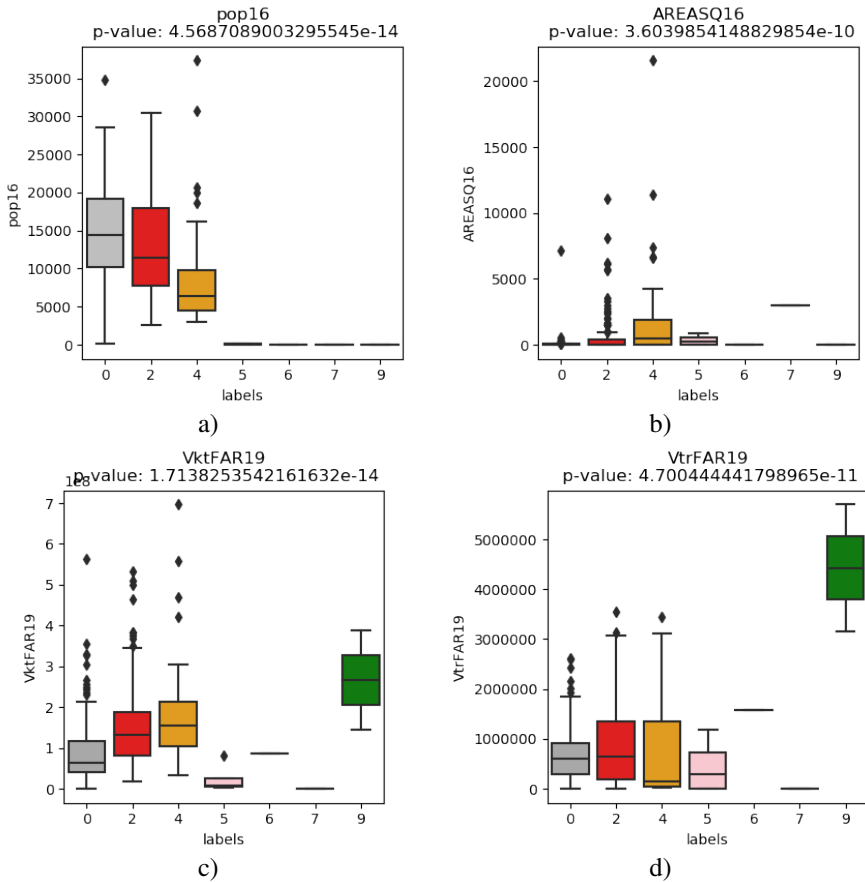


Figure 5. ANOVA of: a) pop16 (population size), b) AREASQ16 (region size), c) VktFAR19 (absolute traffic) and d) VtrFAR19 (traffic density) on risk levels with respect to population size.

335 **4.2.3. Risk with respect to the traffic density: VtrFAR19 variable**

336 We now set N_j to be the ratio VtrFAR19 of vehicle-kilometers travelled on freeways
 337 and arterial roads divided by the total road length in the region, i.e., risks are clustered
 338 according to the impact of VtrFAR19 on the number of crashes. N_j can be null in some
 339 regions, which leads to a degenerate model, so we set the minimal value to one in VktFAR19
 340 when computing the ratio (this happens in the zone of French Island, only). We obtain nine
 341 clusters, among which two have negligible frequencies (see Figures 6 and 8). The estimate
 342 $\hat{\beta} \approx 0.54$ indicates good spatial aggregation of regions. The cumulative marginal entropy is
 343 15.8, indicating mostly low uncertainty regarding risk levels. From Figure 7, the entropy is
 344 very close to 0 in a large majority of regions and close to 0.693 (i.e., $\ln 2$) in a small number of
 345 regions, which is a value that represents equiprobability between two risk levels. It can be seen
 346 from Figure 8 that linear regression lines are in accordance with slopes induced by expected
 347 risk levels, indicating well-separated classes. There is however some larger discrepancies
 348 between the two values in risk levels 0 and 5, where the regression lines have a larger slopes
 349 than the expected risk levels, due to possible confusions between levels 0 and 2, and levels 5
 350 and 6, in some regions.

351 Risk level 0 is related to regions in the close periphery of the centre (see Figure 6).
 352 These are small regions with high traffic densities, medium population sizes and population
 353 densities (see Figure 9). Risk level 2 is related to regions in the close periphery of the
 354 capital city centre as well as hypercentral regions, and central regions and enclaves. These are
 355 small regions with intermediate traffic densities, high population sizes and medium to high
 356 population densities. Risk level 3 is related to regions in the close periphery of the centre as
 357 well as hypercentral regions, central regions and enclaves. These are small regions with low
 358 traffic densities, high population sizes and low population densities. Risk level 5 is related to
 359 peripheral regions. These are medium-sized regions with very low traffic densities, population
 360 sizes and population densities. Risk level 6 is related to far peripheral regions. These are large
 361 regions with very low traffic densities, population sizes and populatoion densities. The four
 362 variables considered in Figure 9 are well discriminated by the risk levels, with ANOVA p -
 363 values between 10^{-21} and 10^{-69} regarding the effects of classes.

364 Since regions with higher VtrFAR19 have lower risks, it is possible that VtrFAR19 has
 365 some non-linear effect, which we again leave to further investigations.

366 **4.2.4. Combining classes issued from two variables**

367 The analyses performed in Subsections 4.2.2 and 4.2.3 yield different mappings. It
 368 would be possible to define new classes as pairs of classes issued from each model. However
 369 this would lead to more complex interpretations. Indeed, there are strong dependencies

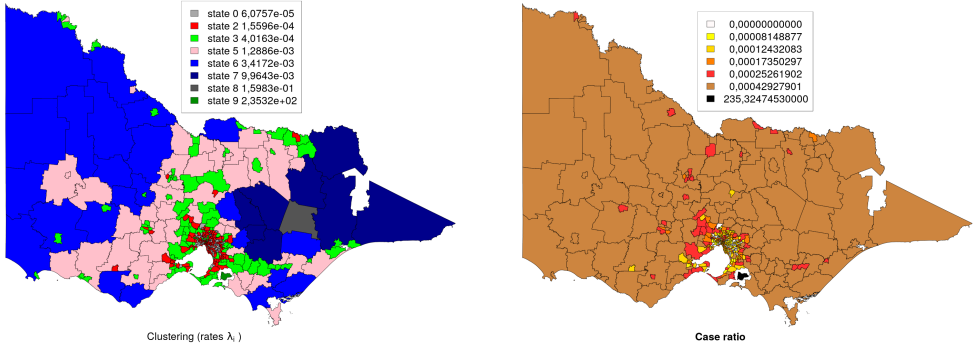


Figure 6. Risk mapping with respect to the traffic density (variable VtrFAR19). Left-hand part: traffic per region. Right-hand part: segmentation using quantiles on ratio.

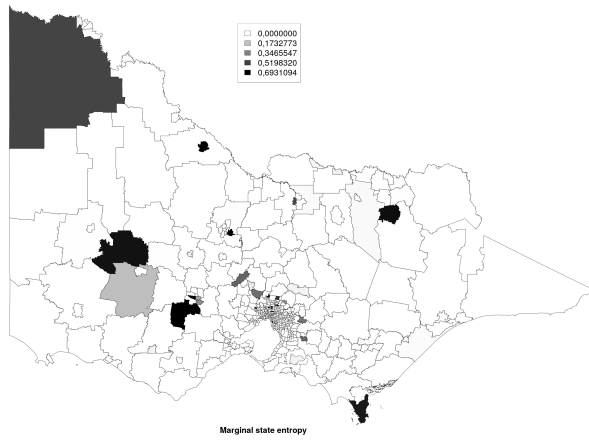


Figure 7. Marginal state entropy for each region regarding risk levels with respect to traffic density VtrFAR19.

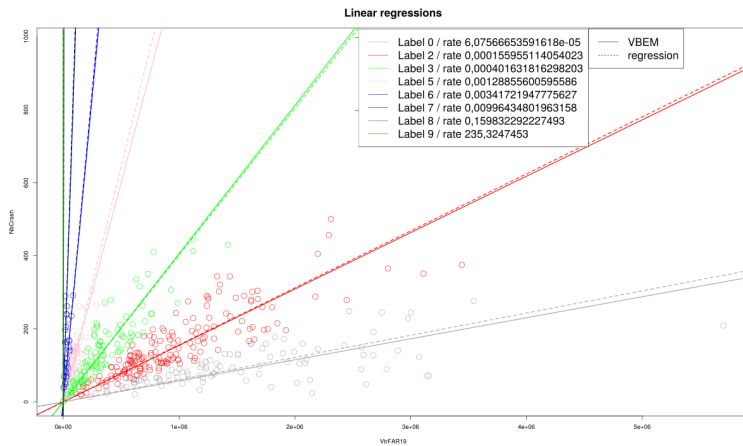


Figure 8. Cluster-wise regressions of crash numbers on traffic density VtrFAR19

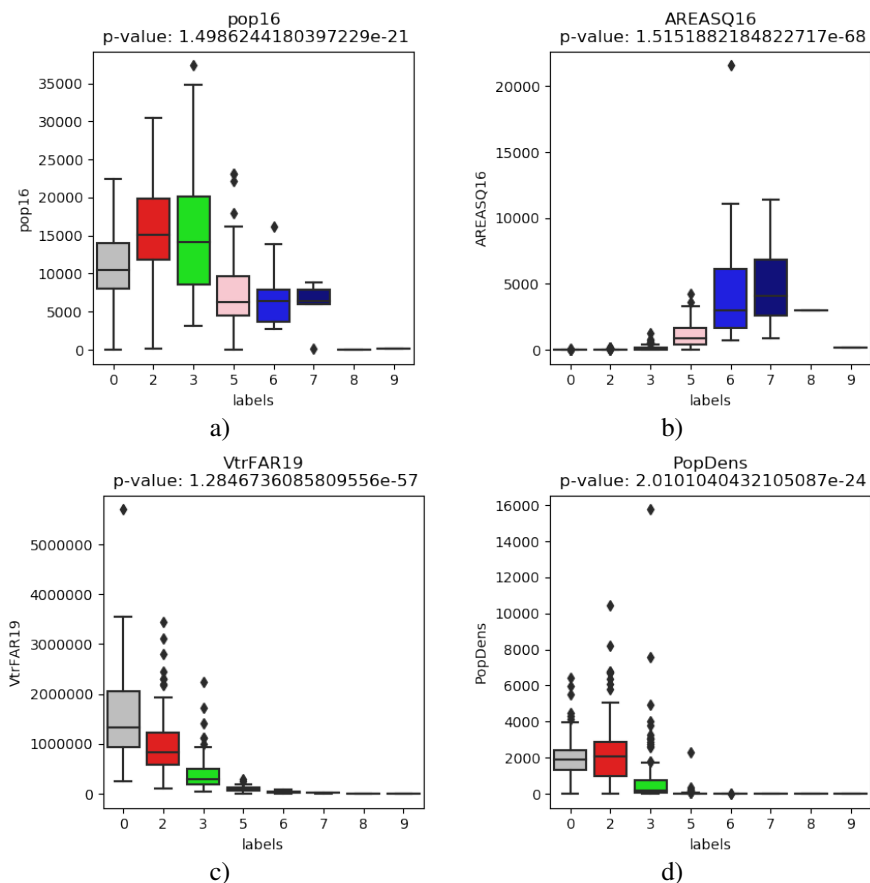


Figure 9. ANOVA of: a) pop16 (population size), b) AREASQ16 (region size), c) VtrFAR19 (traffic density), d) population density on risk levels with respect to traffic density VtrFAR19.

370 between both classes. A χ^2 independence test rejects the assumption of independence with
 371 p -value of 10^{-14} . To visualise associations between classes, a correspondence analysis (CA)
 372 was performed. We ignored outlier classes and also class VtrFar19_7 (meaning class 7 in
 373 the model obtained with normalising risk VtrFar19), which perfectly corresponds to Pop16_4
 374 (higher risk levels in both models). The first CA plane is represented in Figure 10. This
 375 shows that labels VtrFar19_2 and VtrFar19_3 are mainly associated into Pop16_0, while
 376 Pop16_4 and VtrFar19_5 are strongly associated, VtrFar19_6 is associated with Pop16_2 and
 377 VtrFar19_0 is split between Pop16_0 and Pop16_2. As a conclusion, the orders or risk levels
 378 in each model are preserved, except for level 6 in VtrFar19.

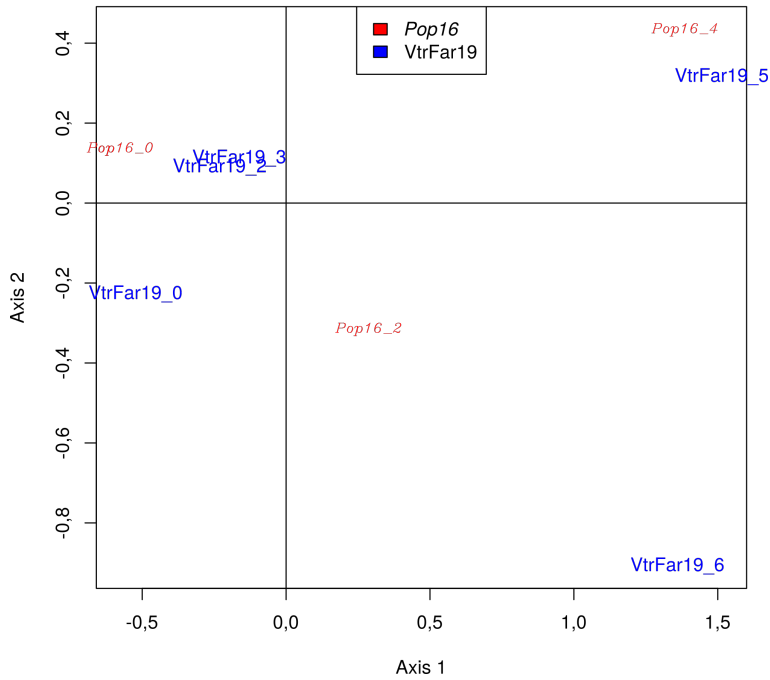
CA Biplot: Classes for Pop16 and VtrFar19

Figure 10. Correspondence analysis between the classes associated with the Pop16 and VtrFAR19 models: first principal plane.

379 **4.2.5. Risk with respect to other variables**

380 Among the different choices of N_j considered for risk normalisation, the following
 381 ones yielded some reduced numbers of classes: VktFAR19, PropSign and PopDens. The
 382 associated results are presented in Appendix II. In contrast, RnNoItcNS, RnLen and RnDens
 383 had increasing numbers of classes whenever K increased further beyond 10, indicating
 384 some failure in modelling meaningful classes of risks when applying our approach to those
 385 variables.

386

5. Conclusion and perspectives

387 The BNP-HMRF model presented here for risk mapping offers a convenient and fast
 388 approach for conducting segmentation of count data regressions indexed by graphs. For
 389 example running the 300 iterations required to obtain the results from Subsection 4.2.2 takes
 390 between 35 seconds and 40 seconds on a Laptop with an Intel Core 8th i7 8665U CPU
 391 with 4 cores, HT, 1.9Ghz, 4.8Ghz Turbo, 8Mo/UHD 620, using hyperthreading. This running

392 time includes selection of the number of classes, which is performed implicitly with state
 393 probabilities vanishing when states are not relevant.

394 The workload required to add new models to our software is rather low, since the specific
 395 modifications due to the Poisson assumption can all be embedded into a new class, which is
 396 added to the package and by using class inheritance principles. This holds if the added model
 397 remains in the exponential family, using conjugate priors.

398 This models would also offer the possibility to handle missing y_j s by introducing new
 399 variational factors q_{y_j} in our VBEM approximation. Besides handling incomplete data sets,
 400 this would also allow modellers to perform model selection by cross-validation for example.

401 The proposed model was effective in identifying clusters with distinct risk levels.
 402 Detailed analyses of these clusters showed that regions with higher traffic densities tend
 403 to have lower traffic density-based crash risk levels, while regions with higher population
 404 sizes tend to have lower population-based crash risk levels. These findings corroborate well
 405 with the traffic safety literature. It is well-established that crash risks tend to decrease with
 406 increasing exposure, such as population or the number of road users (safety-in-number
 407 effect, Elvik 2014). Regarding further application to traffic crashes, the model could be
 408 extended to consider multiple crash exposure variables. However, the possibility to use
 409 several risk-normalising variables leads to multiplying classes and makes their interpretation
 410 more difficult. To solve this issue, we could build meta-labels by considering for each region a
 411 risk signature. This signature would be a vector of risk levels, each component corresponding
 412 to a specific variable N_j . However, this would greatly increase the dimension of the state
 413 space to be considered. We could in addition consider either spatial clustering on those
 414 signatures (possibly equipped with some metrics) or continuous latent variables obtained
 415 with multiple correspondence analysis, introducing continuous latent variables in our model.
 416 Moreover, our approach could be compared with clusterings obtained from CAR models,
 417 which would require some ranges for grouping CAR random effects as discussed in Section
 418 1.

419 Ideally, all possibly relevant variables $N_{i,j}$ should be included into a unique multiple
 420 regression model such as

$$p(y_j | \lambda_k, N_{1,j}, \dots, N_{I,j}) = \mathcal{P}(y_j | \lambda_0 + \sum_{i=1}^I \lambda_{k,i} N_{i,j})$$

421 or

$$p(y_j | \lambda_k, N_{1,j}, \dots, N_{I,j}) = \mathcal{P} \left(y_j | \exp \left[\lambda_0 + \sum_{i=1}^I \lambda_{k,i} N_{i,j} \right] \right), \quad (11)$$

422 and

$$p(y_j | \lambda_k, N_{1,j}, \dots, N_{I,j}) = \mathcal{P} \left(y_j | \exp \left[u_j + \lambda_0 + \sum_{i=1}^I \lambda_{k,i} N_{i,j} \right] \right), \quad u_j \sim \mathcal{N}(0, \tau^2), \quad (12)$$

423 where (11) and (12) have the respective advantages of enabling non-positive linear predictors
 424 and modelling over-dispersion (see also Waller & Carlin 2010). However, such models do
 425 not yield explicit VE-steps in the VBEM algorithm and further approximations would be
 426 required. Although zero-inflated count data did not seem to invalidate our analyses and
 427 results, they may have to be considered in other applications, either at the level of Dirichlet
 428 Processes (as discussed in Canale et al. 2017) or via alternative emission densities.

429 To model centrifugal/centripetal effects through a radius r or more generally, the effect
 430 of continuous latent variables z_j , we could define the risk Λ as a stochastic, monotonic
 431 function of z_j . Another possibility would be the extension of a scalar β by a function
 432 parameterized by the difference in z_j values between two regions.

433 BNP-HMRF models could also be extended to handle multivariate count data,
 434 particularly to address modelling problems in ecology where counts correspond to the
 435 number of observed species. The emission densities could thus be replaced by Join Species
 436 Distribution Models (JSDMs), in which spatial dependencies and heterogeneity sources are
 437 usually modelled with univariate CARs (see, e.g., Saas & Gosselin 2014) but are ignored in
 438 the multivariate count data.

439 Appendix I

440 Free energy computation

Here, we present how to compute the free energy (4). This is mainly used to define a
 stopping criterion in VBEM, which is why we ignore terms that do not depend on hyper-
 parameters or variational parameters. The free energy can be decomposed into two terms:

$$\begin{aligned} \mathbb{E}_{q(z, \tau, \Theta)} \left[\ln \frac{p(\mathbf{y}, \mathbf{z}, \tau, \Theta | N_J, \Phi)}{q(z, \tau, \Theta)} \right] &= \mathbb{E}_{q(z, \tau, \Theta)} [\ln p(\mathbf{y}, \mathbf{z}, \tau, \Theta | N_J, \Phi)] \\ &\quad - \mathbb{E}_{q_z} [\ln q_z] - \mathbb{E}_{q_\tau} [\ln q_\tau] - \mathbb{E}_{q_\Theta} [\ln q_\Theta], \end{aligned}$$

441 where the last three terms are entropies and the first term has already been already calculated
 442 in the E-step.

443 **Entropy terms**

444 In this section, we provide the expressions for $H[q_z] = -E_{q_z}[\ln q_z(\mathbf{z})]$, $H[q_\tau] =$
 445 $-E_{q_\tau}[\ln q_\tau]$ and $H[q_\theta] = -E_{q_\theta}[\ln q_\theta(\Theta)]$. Firstly,

$$H[q_z] = \sum_{j=1}^J H[q_{z_j}] = - \sum_{j=1}^J \sum_{k=0}^{K-1} q_{z_j}(k) \ln q_{z_j}(k),$$

446 where $q_{z_j}(k)$ is given in (10). Next,

$$H[q_\theta] = H[q_\alpha] + \sum_{k=0}^{K-1} H[q_{\lambda_k}]. \quad (13)$$

447 The first term above can be derived directly from the known entropy expression for a gamma
 448 density:

$$H[q_\alpha] = \hat{s}_1 - \ln \hat{s}_2 + \ln \Gamma(\hat{s}_1) + (1 - \hat{s}_1)\psi(\hat{s}_1).$$

449 The other terms are obtained as follows:

$$-H[q_{\lambda_k}] = E_{q_{\lambda_k}} \left[\ln(\lambda_k | \hat{a}_k, \hat{b}_k) \right] = \psi(\hat{a}_k) - \ln(\hat{b}_k). \quad (14)$$

450 Moreover,

$$H[q_\tau] = \sum_{k=0}^{K-1} H[q_{\tau_k}],$$

451 where for every k ,

$$\begin{aligned} H[q_{\tau_k}] &= \ln B(\gamma_{k,1}, \gamma_{k,2}) - (\gamma_{k,1} - 1)[\psi(\gamma_{k,1}) - \psi(\gamma_{k,1} + \gamma_{k,2})] \\ &\quad - (\gamma_{k,2} - 1)[\psi(\gamma_{k,2}) - \psi(\gamma_{k,1} + \gamma_{k,2})] \end{aligned}$$

452 and $B(\gamma_{k1}, \gamma_{k2}) = \Gamma(\gamma_{k1})\Gamma(\gamma_{k2})/\Gamma(\gamma_{k1} + \gamma_{k2})$.

453 **E-step terms**

The $E_{q(\mathbf{z}, \tau, \Theta)}[\ln p(\mathbf{y}, \mathbf{z}, \tau, \Theta | \mathbf{N}_J, \Phi)]$ term decomposes into 5 terms:

$$\begin{aligned} E_{q(\mathbf{z}, \tau, \Theta)}[\ln p(\mathbf{y}, \mathbf{z}, \tau, \Theta | \mathbf{N}_J, \Phi)] &= \sum_{j=1}^J E_{q_{z_j} q_\Lambda}[\ln p(y_j | z_j, \Lambda)] + E_{q_\alpha}[\ln p(\alpha | s_1, s_2)] \\ &+ \sum_{k=0}^{K-1} E_{q_{\tau_k} q_\alpha}[\ln p(\tau_k | \alpha)] + \sum_{k=0}^{K-1} E_{q_{\lambda_k}}[\ln p(\lambda_k | a_k, b_k)] + E_{q_z q_\tau}[\ln p(\mathbf{z} | \tau, \beta)]. \end{aligned}$$

454 If the free energy is computed at the end of each VBEM iteration, as per Section 3 (VM-
 455 (s_1, s_2) and VM- (a_k, b_k) steps), we have $(s_1, s_2) = (\hat{s}_1, \hat{s}_2)$ and $(a_k, b_k) = (\hat{a}_k, \hat{b}_k)$. Thus, in
 456 the free energy $E_{q_\alpha}[\ln p(\alpha|s_1, s_2)]$ cancels out with $H[q_\alpha]$. Similarly, $E_{q_{\lambda_k}}[\ln p(\lambda_k|a_k, b_k)]$
 457 cancels out with $H[q_{\lambda_k}]$ in (13).

458 $\sum_{j=1}^J E_{q_{z_j, q_\Lambda}}[\ln p(y_j|z_j, \Lambda)]$ **term.**

459 For each $1 \leq j \leq J$,

$$E_{q_{z_j, q_\Lambda}}[\ln p(y_j|z_j, \Lambda)] = \sum_{k=0}^{K-1} q_{z_j}(k) E_{q_{\lambda_k^*}}[\ln p(y_j|\lambda_k^*)].$$

In the case of Poisson emission densities,

$$\begin{aligned} E_{q_{z_j, q_\Lambda}}[\ln p(y_j|z_j, \Lambda)] &= \sum_{k=0}^{K-1} q_{z_j}(k) \left[y_j E_{q_{\lambda_k}}[\ln(\lambda_k)] - N_j \frac{\hat{a}_k}{\hat{b}_k} + y_j \ln(N_j) - \ln(y_j!) \right] \\ &= y_j \sum_{k=0}^{K-1} q_{z_j}(k) E_{q_{\lambda_k}}[\ln(\lambda_k)] - N_j \sum_{k=0}^{K-1} q_{z_j}(k) \frac{\hat{a}_k}{\hat{b}_k} - y_j \ln(N_j) - \ln(y_j!), \end{aligned} \quad (15)$$

460 with $E_{q_{\lambda_k}}[\ln(\lambda_k)|\hat{a}_k, \hat{b}_k]$ given in (14). Since the last two terms in (15) depend on the data
 461 only, they do not need to be computed to monitor the convergence of the VBEM algorithm.

462 $E_{q_{\tau_k, q_\alpha}}[\ln p(\tau_k|\alpha)]$ **term.**

463 Using the expression of a Beta distribution cross-entropy, it follows that

$$E_{q_{\tau_k, q_\alpha}}[\ln p(\tau_k|\alpha)] = \psi(\hat{s}_1) - \ln \hat{s}_2 + \left(\frac{\hat{s}_1}{\hat{s}_2} - 1 \right) [\psi(\gamma_{k2}) - \psi(\gamma_{k1} + \gamma_{k2})].$$

464 $E_{q_{z, q_\tau}}[\ln p(z|\tau, \beta)]$ **term.**

465 This term cannot be computed exactly due to the intractable normalising constant \mathcal{K} in
 466 the Potts model expression:

$$\ln p(z|\tau, \beta) = \sum_{j=1}^J \ln \pi_{z_j}(\tau) + \beta \sum_{i \sim j} \mathbf{1}_{(z_i = z_j)} - \ln \mathcal{K}(\beta, \tau).$$

467 The first two terms can be computed easily but the last one requires approximation. As for
 468 the estimation of β , we can use a mean-field like approximation and approximate at each

469 iteration the true value of $\mathcal{K}(\beta, \boldsymbol{\tau})$ by $\tilde{\mathcal{K}}(\beta, \boldsymbol{\tau})$, the normalising constant of \tilde{q}_z , defined in (9):

$$\tilde{\mathcal{K}}(\beta, \boldsymbol{\tau}) = \prod_{j=1}^J \left(\sum_{l=1}^{\infty} \exp(\ln \pi_l(\boldsymbol{\tau}) + \beta \sum_{i \in N(j)} q_{z_i}^{(r)}(l)) \right),$$

470 where as in Section 3, $\boldsymbol{\tau}$ in the above expression could be replaced by $\mathbb{E}_{q_{\boldsymbol{\tau}}}[\boldsymbol{\tau}]$ to avoid the
471 dependence in a random $\boldsymbol{\tau}$. However, this would correspond to a zeroth order approximation
472 and we can do better with a first order approximation as follows:

$$\mathcal{K}(\beta, \boldsymbol{\tau}) \approx \tilde{\mathcal{K}}(\beta, \boldsymbol{\tau}) \exp(\mathbb{E}_{\tilde{q}_z}[V(\mathbf{z}; \boldsymbol{\tau}, \beta) - \tilde{V}(\mathbf{z}; \boldsymbol{\tau}, \beta)]), \quad (16)$$

473 where $V(\mathbf{z}; \boldsymbol{\pi}, \beta)$ is defined as

$$V(\mathbf{z}; \boldsymbol{\pi}, \beta) = \sum_{j=1}^N \log \pi_{z_j} + \beta \sum_{i \sim j} \mathbf{1}_{(z_i = z_j)}$$

and similarly,

$$\tilde{V}(\mathbf{z}; \boldsymbol{\tau}, \beta) = \sum_{j=1}^N \left(\ln \pi_{z_j}(\boldsymbol{\tau}) + \beta \sum_{i \in N(j)} q_{z_i}^{(r)}(z_j) \right).$$

Note that all terms are tractable:

$$\mathbb{E}_{\tilde{q}_z}[V(\mathbf{z}; \boldsymbol{\tau}, \beta) - \tilde{V}(\mathbf{z}; \boldsymbol{\tau}, \beta)] = \beta \sum_{j=1}^N \sum_{k=0}^{K-1} \tilde{q}_{z_j}(k | \boldsymbol{\tau}, \beta) \left(\frac{\sum_{i \in N(j)} \tilde{q}_{z_i}(k | \boldsymbol{\tau}, \beta)}{2} - \sum_{i \in N(j)} q_{z_i}^{(r)}(k) \right).$$

Then,

$$\begin{aligned} \mathbb{E}_{q_{\boldsymbol{\tau}}}[\ln \mathcal{K}(\beta, \boldsymbol{\tau})] &\approx \mathbb{E}_{q_{\boldsymbol{\tau}}}[\ln \tilde{\mathcal{K}}(\beta, \boldsymbol{\tau})] + \mathbb{E}_{q_{\boldsymbol{\tau}}}[\mathbb{E}_{\tilde{q}_z}[V(\mathbf{z}; \boldsymbol{\tau}, \beta) - \tilde{V}(\mathbf{z}; \boldsymbol{\tau}, \beta)]] \\ &\approx \sum_{j=1}^N \mathbb{E}_{q_{\boldsymbol{\tau}}}[\ln(\sum_{l=1}^{\infty} \exp(\ln \pi_l(\boldsymbol{\tau}) + \beta \sum_{i \in N(j)} q_{z_i}^{(r)}(l)))] \\ &\quad + \beta \sum_{j=1}^N \sum_{k=0}^{K-1} \mathbb{E}_{q_{\boldsymbol{\tau}}} \left[\tilde{q}_{z_j}(k | \boldsymbol{\tau}, \beta) \left(\frac{\sum_{i \in N(j)} \tilde{q}_{z_i}(k | \boldsymbol{\tau}, \beta)}{2} - \sum_{i \in N(j)} q_{z_i}^{(r)}(k) \right) \right]. \end{aligned}$$

474

Appendix II

475

Additional figures and results on crash data analysis

476

Here, we present further figures and analyses obtained by using different covariates in our model.

478

Figure 11 depicts the map obtained from Step 2 of the exploratory analysis: population size and number of crashes per region together with the segmentation obtained by using 7 quantiles on the ratio of the ‘number of crashes on population size’.

479

480

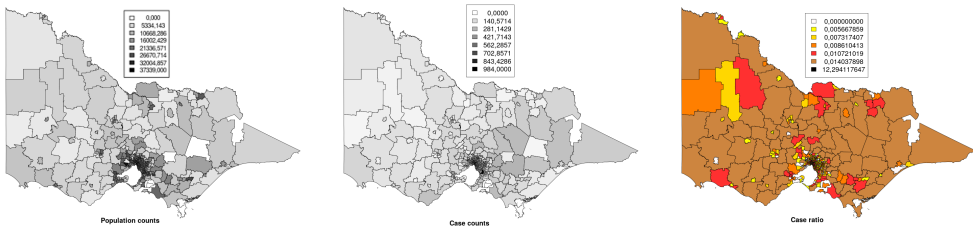


Figure 11. Maps of data and risk ratio use in exploratory using population size (variable pop16). Left: population size per region. Middle: number of crashes per region. Right: segmentation using quantiles on ratio.

481

The following subsections reproduce similar analyses as those that appear in Subsections 4.2.2 and 4.2.3 but using other variables to normalise risks of crashes.

482

483 Risk with respect to the traffic: VktFAR19 variable

484

We now set N_j as the vehicle-kilometres travelled on freeways and arterial roads, i.e., risks are clustered according to the impact of vehicle-kilometres travelled on the number of crashes. The risk appears as somewhat spatially homogeneous and applying the model directly does not yield several well-separated clusters. We consider the transform $p(y_j|z_j = k; \mathbf{\Lambda}, N_j) = \mathcal{P}(y_j; \lambda_k \sqrt{N_j})$. Here, N_j can be null in some regions, which leads to a degenerate model, so we set the minimal value to one. We obtain three clusters, all of which are well represented (see Figure 12). The estimate $\hat{\beta} \approx 0.42$ indicates a moderate spatial aggregation of regions. The cumulative marginal entropy is 32.6, indicating some rather high uncertainty regarding risk levels in some regions. It can be seen from Figure 13 that linear regression lines are in accordance with slopes induced by expected risks levels, indicating well-separated classes. There is however some larger discrepancies between the two quantities in risk level 8, where the regression line has a lower slope than the expected risk level, due to possible confusions between levels 8 and 4 in some regions.

491

Risk level 1 is related to peripheral far west and northern regions (see Figure 12).

492

These are regions with low traffic densities, population sizes and population densities (see

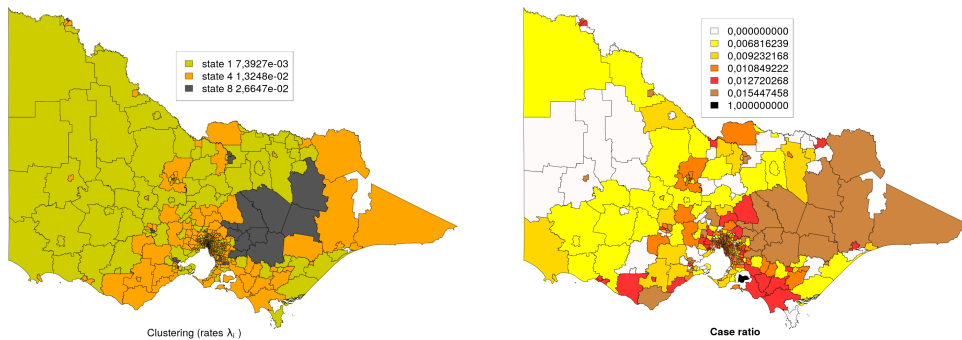


Figure 12. Risk mapping with respect to the traffic (variable VktFAR19). Left-hand part: traffic per region. Right-hand part: segmentation using quantiles on ratio.

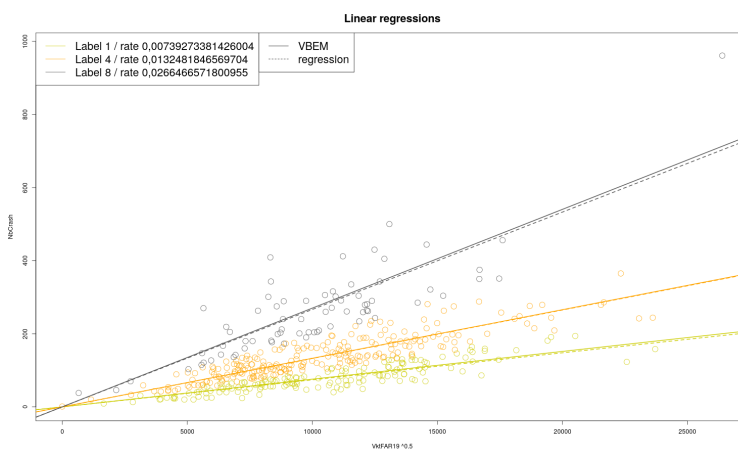


Figure 13. Cluster-wise regressions of crash numbers on VktFAR19.

499 Figure 14). Risk level 4 is related to regions in the close periphery of the capital city centre
 500 and far east regions. These have intermediate traffic densities, population sizes and population
 501 densities. Risk level 8 is related hypercentral regions and those at the close east periphery.
 502 These have high traffic densities, population sizes and population densities.

503 We remark that VktFAR19 has no effect on the labels (p -value 0.69), which seems to
 504 confirm the linear relationship between VktFAR19 and crash numbers (cluster-wise).

505 **Risk with respect to the proportion of signalised intersections: PropSign variable**

506 The number of signalised intersections (including roundabouts) was generally quite low
 507 and the number of non-signalised intersections was highly correlated with the number of
 508 intersections (Pearson correlation: 0.99988). However, the ratio r of the number of signalised
 509 intersections on total number of intersections what somehow variable, so we could use it as

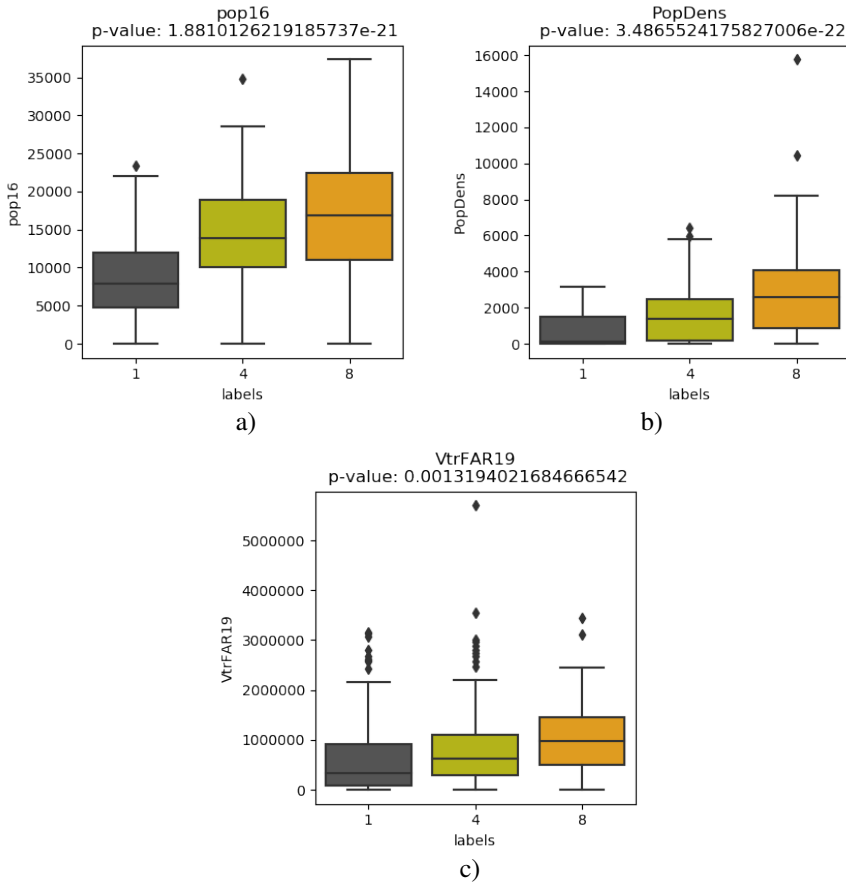


Figure 14. ANOVA of: a) pop16 (population size), b) PopDens (population density) and c) VtrFAR19 (traffic density) on risk levels with respect to absolute traffic intensity VktFAR19.

510 the normalising variable. We subtracted the minimal value (rounded to the closest multiple
 511 of 0.1, here 0.8) and considered $N_j = (r - 0.8)^{1.5}$. In this scenario, we consider that risks
 512 are clustered according to the impact of relative increase of the proportion of signalised
 513 intersections with respect to the minimal ratio on the number of crashes. We obtained
 514 nine clusters, among which four have negligible frequencies (see Figure 8). The estimate
 515 $\hat{\beta} \approx 0.27$ indicates low spatial aggregation of regions. The cumulative marginal entropy is
 516 15.8, indicating mostly low uncertainty regarding risk level. Labels are mostly explained
 517 by centripetal gradients. It can be seen from Figure 16 that the accordance between linear
 518 regression lines and slopes induced by expected risks levels is moderate, indicating possible
 519 confusions between classes 3 versus 4, 4 versus 5 and 5 versus 6.

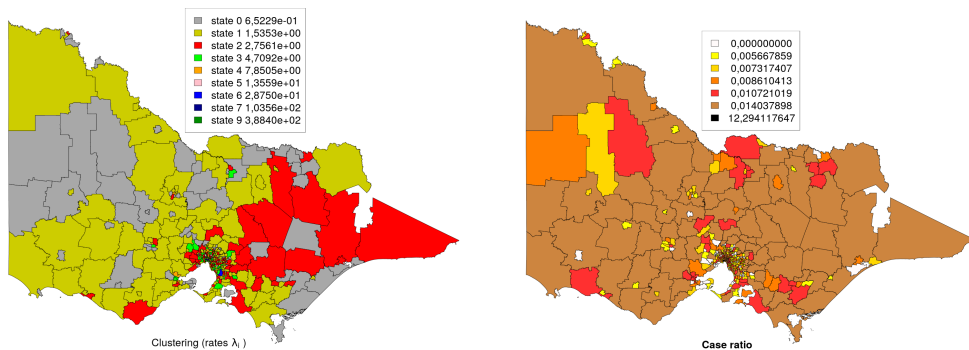


Figure 15. Risk mapping with respect to the ratio of signalised on total number of intersections (variable PropSign). Left-hand part: ratio per region. Right-hand part: segmentation using quantiles on ratio.

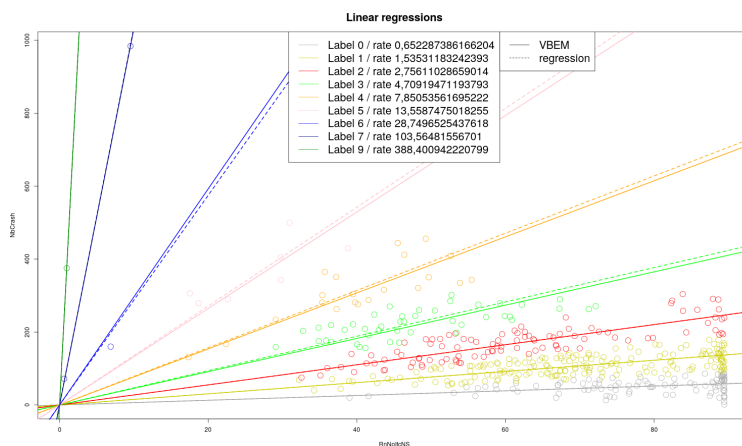


Figure 16. Cluster-wise regressions of crash numbers on the ratio of signalised on total number of intersections.

520 Risk levels 5, 6, 7 and 9 apply to a very small numbers of regions (with small values of
 521 N_j , see Figure 15) and can be neglected. Risk level 0, 1 and 2 apply to far peripheral regions
 522 with low absolute traffics, traffic densities, population sizes and population densities (with
 523 an increasing risk regarding those four variables). Risk level 3 is less peripheral than 0, 1, 2
 524 with higher absolute traffics, traffic densities, population sizes and population densities. Risk
 525 level 4 applies to more central regions than 3, with slightly higher population densities but
 526 somewhat lower population sizes and traffic.

527 **Risk with respect to the population density: PopDens**

528 We now set N_j as the population density PopDens, i.e., risks are clustered according
 529 to the impact of PopDens on the number of crashes. We consider the transform $p(y_j|z_j =$

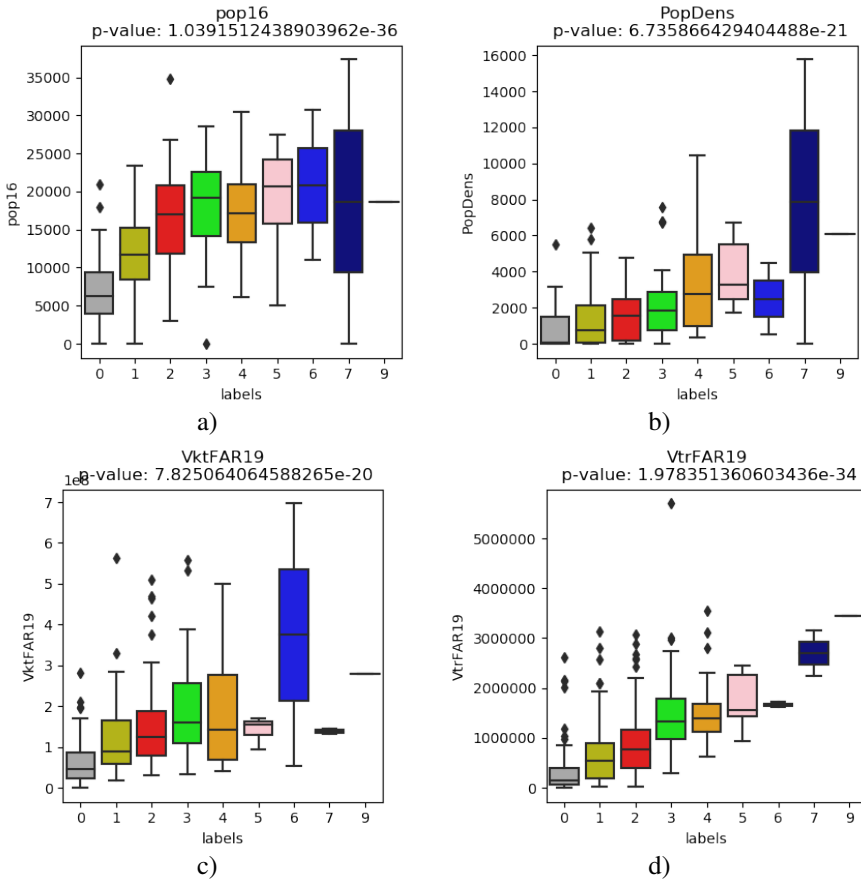


Figure 17. ANOVA of: a) pop16 (population size), b) PopDens (population density), c) VktFAR19 (absolute traffic intensity) and d) VtrFAR19 (traffic density) on risk levels with respect to the proportion of signalised intersections.

530 $k; \mathbf{\Lambda}, N_j) = \mathcal{P}(y_j; \mathbf{\Lambda}, \sqrt{N_j})$. We obtain nine clusters, among which three have negligible
 531 frequencies (see Figure 19). The estimate $\hat{\beta} \approx 0.45$ indicates moderate spatial aggregation of
 532 regions. The cumulative marginal entropy is 17.7, indicating moderate uncertainty regarding
 533 risk level. It can be seen from Figure 19 that the accordance between linear regression
 534 lines and slope induced by expected risks levels is moderate, indicating possible confusions
 535 between classes 2 versus 3, 3 versus 4 (mainly), 4 versus 5 and 5 versus 6.

536 Risk levels 7 to 9 apply to a very small numbers of regions (hypercentral or
 537 hyperperipheral, with small N_j , see Figure 18) and can be neglected. Risk levels 0 to 6 are
 538 mostly explained by centrifugal gradients: from small central regions, with high population
 539 densities, low absolute traffics but high traffic densities to large peripheral regions, with low

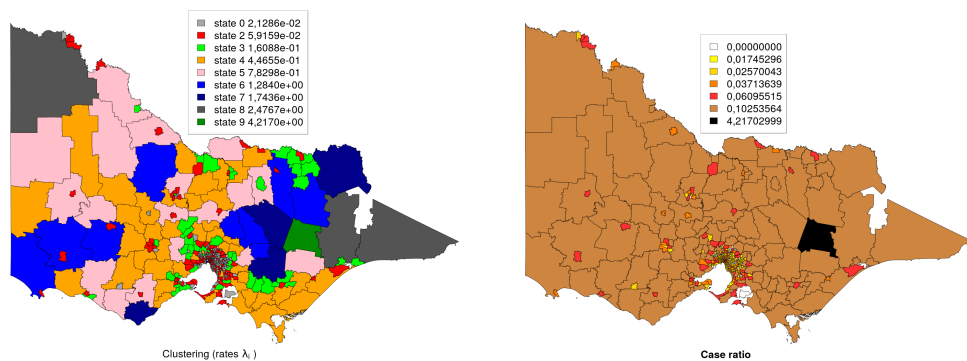


Figure 18. Risk mapping for variable PopDens. Left-hand part: population density per region. Middle part: assigned risk level entropy. Right-hand part: segmentation into a finite number of risk levels.

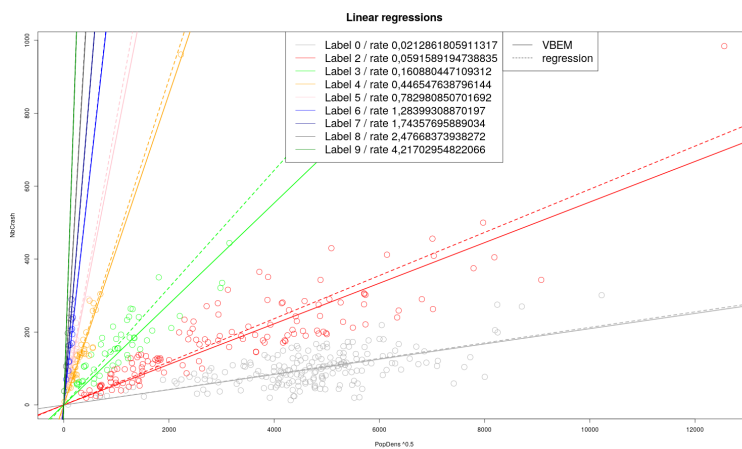


Figure 19. Cluster-wise regressions of crash numbers on PopDens.

540 population densities, high absolute traffics but low traffic densities (see Figure 20). Enclaves
 541 mostly have risk levels 2 or 0, while the risks of their surrounding regions are higher.

542 We remark that if a region has low $VktFAR_{19}$ and high $VtrFAR_{19}$, it means that it has
 543 low absolute traffic but dense traffic, when considering the total road network length.

544 Since regions with higher PopDens have lower risks, it is possible that PopDens has
 545 some non-linear effect, which is a point to be further investigated.

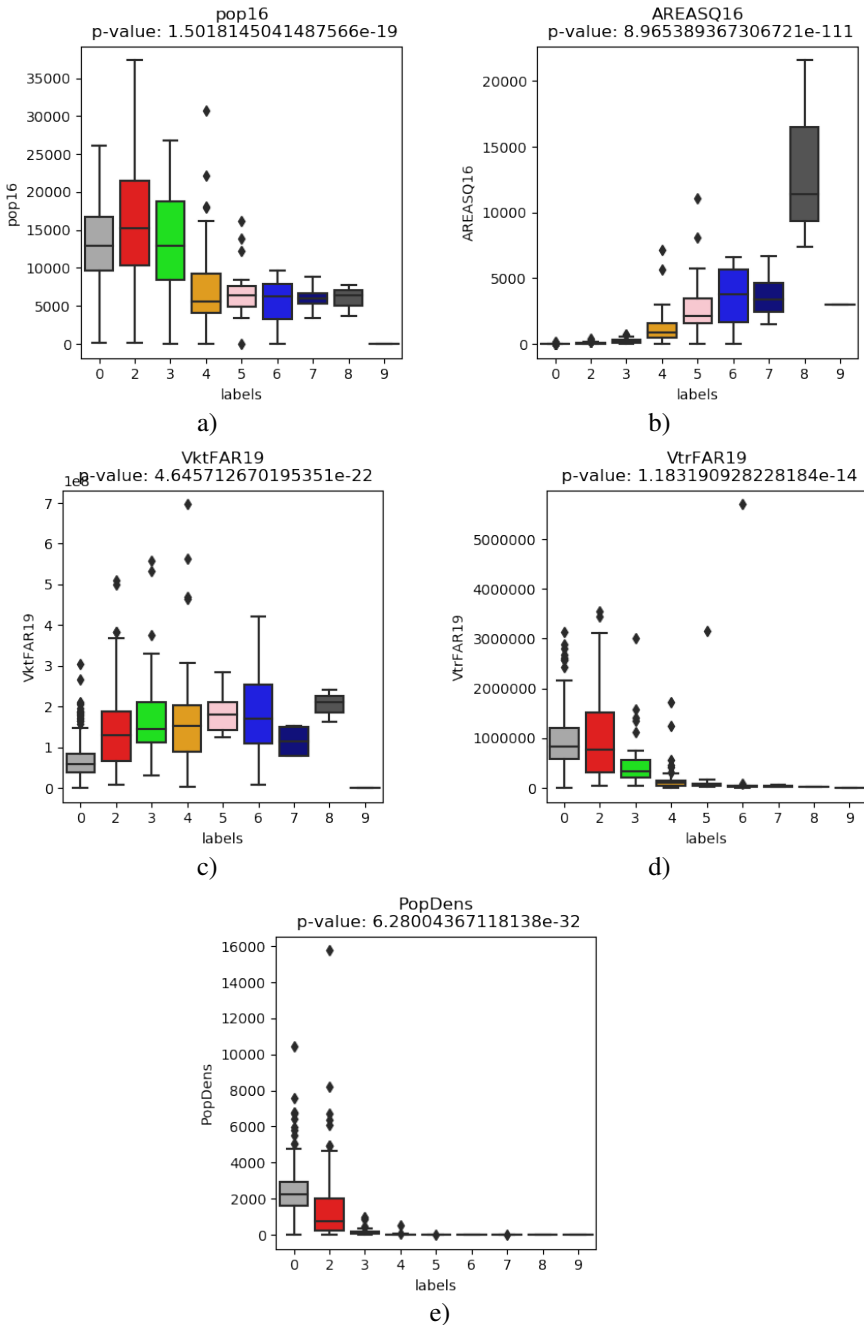


Figure 20. ANOVA of: a) pop16 (population size), b) AREASQ16 (region size), c) VktFAR19 (absolute traffic intensity), d) VtrFAR19 (traffic density) and e) PopDens (population density) on risk levels with respect to population density.

546

References

- 547 ABRIAL, D., CALAVAS, D., JARRIGE, N. & DUCROT, C. (2005). Spatial heterogeneity of the risk of BSE in
548 France following the ban of meat and bone meal in cattle feed. *Preventive veterinary medicine* **67**, 69–
549 82. doi:10.1016/j.prevetmed.2004.10.004. URL [http://www.ncbi.nlm.nih.gov/pubmed/
550 15698909](http://www.ncbi.nlm.nih.gov/pubmed/15698909).
- 551 ALFO, M., NIEDDU, L. & VICARI, D. (2009). Finite mixture models for mapping spatially dependent
552 disease counts. *Biometrical Journal* **51**, 84–97.
- 553 BESAG, J., YORK, J. & MOLLIE, A. (1991). Bayesian image restoration, with two applications in spatial
554 statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–59.
- 555 BLEI, D.M. & JORDAN, M.I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Anal.*
556 **1**, 121–143.
- 557 BÖHNING, D., DIETZ, E. & SCHLATTMANN, P. (2000). Space-time mixture modelling of public health
558 data. *Statistics in Medicine* **19**, 2333–2344.
- 559 CANALE, A., LIJOI, A., NIPOTI, B. & PRÜNSTER, I. (2017). On the pitman-yor process with spike and
560 slab base measure. *Biometrika* **104**, 681–697.
- 561 CHANDLER, D. (1987). *Introduction to modern statistical mechanics*. New York, Oxford: Oxford University
562 Press. URL <http://opac.inria.fr/record=b1081336>.
- 563 CLAYTON, D. & BERNADINELLI, L. (1992). Bayesian methods for mapping disease risk. *Geographical
564 and Environment Epidemiology: Methods for Small Area Studies*, eds P.Elliot, J.Cuzik, D.English, and
565 R.Stern, Oxford, UK:Oxford University Press , 205–220.
- 566 ELVIK, R. (2014). *Towards a general theory of the relationship between exposure and risk*. Institute of
567 Transport Economics, Oslo, Norway.
- 568 ELVIK, R., VAA, T., HOYE, A. & SORENSEN, M. (2009). *The handbook of road safety measures*. Emerald
569 Group Publishing.
- 570 FERNANDEZ, C. & GREEN, P. (2002). Modelling spatially correlated data via mixtures: a Bayesian approach.
571 *Journal of the Royal Statistical Society: Series B (Methodological)* **64**, 805–826.
- 572 FORBES, F. & PEYRARD, N. (2003). Hidden Markov model selection based on mean field like
573 approximations. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**, 1089–1101.
- 574 FRALEY, C. & RAFTERY, A. (2007). Bayesian regularization for normal mixture estimation and model-based
575 clustering. *Journal of Classification* **24**, 155–181.
- 576 GHOSAL, S. & VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference*, vol. 44.
577 Cambridge University Press.
- 578 GREEN, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model
579 determination. *Biometrika* **82**, 711–732.
- 580 GREEN, P.J. & RICHARDSON, S. (2002). Hidden Markov models and disease mapping. *Journal of the
581 American Statistical Association* **97**, 1–16.
- 582 KNORR-HELD, L. & RASSER, G. (2000). Bayesian detection of clusters and discontinuities in disease maps.
583 *Biometrics* **56**, 13–21.
- 584 KNORR-HELD, L., RASSER, G. & BECKER, N. (2002). Disease mapping of stage-specific cancer incidence
585 data. *Biometrics* **58**, 492–501.
- 586 KNORR-HELD, L. & RICHARDSON, S. (2003). A hierarchical model for space-time surveillance data on
587 meningococcal disease incidence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*
588 **52**, 169–183.
- 589 LAWSON, A., BIGGERI, A., BOEHNING, D., LESAFFRE, E., VIEL, J., CLARK, A., SCHLATTMANN, P.
590 & DIVINO, F. (2000). Disease mapping models: an empirical evaluation. *Statistics in Medicine* **19**,
591 2217–2241.
- 592 LAWSON, A. & SONG, H. (2010). Bayesian hierarchical modeling of the dynamics of spatio-temporal
593 influenza season outbreaks. *Spatial and Spatio-temporal Epidemiology* **1**, 187–195.

- 594 LORD, D. & MANNERING, F. (2010). The statistical analysis of crash-frequency data: A review and
595 assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* **44**,
596 291–305. doi:<https://doi.org/10.1016/j.tra.2010.02.001>. URL <http://www.sciencedirect.com/science/article/pii/S0965856410000376>.
- 598 LU, H., ARBEL, J. & FORBES, F. (2020). Bayesian nonparametric priors for hidden Markov random fields.
599 *Statistics and Computing* URL <https://hal.archives-ouvertes.fr/hal-02163046>.
- 600 MACNAB, Y. (2010). On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods*
601 *in Medical Research* **20**, 49–68.
- 602 MILLER, J.W. & HARRISON, M.T. (2018). Mixture models with a prior on the number of components.
603 *Journal of the American Statistical Association* **113**, 340–356.
- 604 MOLLIE, A. (1996). Bayesian mapping of disease. *Markov Chain Monte Carlo in Practice*, eds. W. Gilks, S.
605 Richardson, and D. J. Spiegelhalter, London: Chapman and Hall , 359–379.
- 606 MOLLIE, A. (1999). Bayesian and empirical bayes approaches to disease mapping. In *Disease mapping and*
607 *risk assessment for public health*, eds. A. Lawson, A. Biggeri & D. Bohning. Wiley, pp. 15–29.
- 608 PAPADIMITRIOU, E., FILTNESS, A., THEOFILATOS, A., ZIAKOPOULOS, A., QUIGLEY, C. & YANNIS, G.
609 (2019). Review and ranking of crash risk factors related to the road infrastructure. *Accident Analysis*
610 *& Prevention* **125**, 85 – 97.
- 611 PASCUTTO, C., WAKEFIELD, J., BEST, N., RICHARDSON, S., BERNARDINELLI, L., STAINES, A. &
612 ELLIOTT, P. (2000). Statistical issues in the analysis of disease mapping data. *Statistics in Medicine*
613 **19**, 2493–2519.
- 614 RICHARDSON, S., MONFORT, C., GREEN, M., DRAPER, G. & MUIRHEAD, C. (1995). Spatial variation of
615 natural radiation and childhood leukaemia incidence in great britain. *Statistics in Medicine* **14**, 2487–
616 2501.
- 617 ROBERTSON, C., NELSON, T., MACNAB, Y. & LAWSON, A. (2010). Review of methods for space-time
618 disease surveillance. *Spatial and Spatio-temporal Epidemiology* **1**, 105–116.
- 619 SAAS, Y. & GOSSELIN, F. (2014). Comparison of regression methods for spatially-autocorrelated count data
620 on regularly- and irregularly-spaced locations. *Ecography* **37**, 476–489.
- 621 SCHLATTMANN, P. & BÖHNING, D. (1993). Mixture models and disease mapping. *Statistics in Medicine*
622 **12**, 1943–1950.
- 623 SETHURAMAN, J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica* **4**, 639–650.
- 624 STOHR, J. (2017). A review on statistical inference methods for discrete Markov random fields. *arXiv*
625 *e-prints* , arXiv:1704.03331v1[math.CO].
- 626 THEOFILATOS, A. & YANNIS, G. (2014). A review of the effect of traffic and weather characteristics on
627 road safety. *Accident Analysis & Prevention* **72**, 244 – 256.
- 628 TRUONG, L. & CURRIE, G. (2019). Macroscopic road safety impacts of public transport: A case study of
629 Melbourne, Australia. *Accident Analysis & Prevention* **132**, 105270.
- 630 WALLER, L. & CARLIN, B. (2010). Disease mapping. In *Handbook of spatial statistics*, eds. A. Gelfand,
631 P. Diggle, P. Guttorp & M. Fuentes, vol. 2010, chap. 14. Handbook of Modern Statistical Methods. Boca
632 Raton: Chapman & Hall, CRC Press, pp. 217–244.