Modelling seismic activity using a Bayesian non-parametric method

S. Hernández
Laboratorio de Procesamiento de Información Geoespacial
Universidad Católica del Maule.
Talca, Chile.
Email: shernandez@ucm.cl

P. Sallis
Geoinformatics Research Centre
Auckland University of Technology
Auckland, New Zealand.
Email: philip.sallis@aut.ac.nz

***Abstract*—Machine learning consists of a set of computational tools for performing large multi-dimensional data set analysis where standard statistical tests are not easily implemented. Many parametric approaches for machine learning consist of model selection and at least a two-step process. Using these techniques the underlying structure of the observed data may not be fully realised. On the other hand, Bayesian non-parametric methods perform inference operations over an infinitely greater number of parameters and because the inherent model uncertainty is also incorporated in the single-step approach, this can lead to a more robust estimation of resulting values. This paper applies this approach to the modelling geophysical events, which is a challenging spatio-temporal problem domain. This paper contributes to the ongoing investigation of optimal methods for geophysical event modelling by introducing a numerical computation solution using a Bayesian unsupervised learning algorithm with earthquake magnitude and location data from Central Chile following a recent 8.8 magnitude earthquake that destroyed many buildings and other property. It is envisaged that this method could be applied to other major earthquakes and further work is gathering data for analysis in this regard.**

I. Introduction

Earthquakes are random geophysical events that can have catastrophic dimensions and deeply affect the lives of people. The study of the statistical properties of earthquakes have a long tradition in physics,mathematical and applied statistics but because of their un-predictable nature, no solution has been found for alerting people when one is about to arrive. Instead, analyzing seismic activity data leads to explanations of what the nature of the event is and more importantly, what the probability is of a new earthquake given all recorded seismic events [10].

Conventional statistical modeling of geophysical data assumes a linear and Gaussian distribution of observations. The variogram is a widely used descriptor of spatial dependency for a group of observations and kriging is used to interpolate data from sparsely sampled observations [4]. Important to note here is that this standard spatial prediction method utilizes a stationary Gaussian process but many geophysical events are best described as point patterns instead [16].

Earthquakes locations and magnitudes are observed as randomly scattered events, and suitable statistical models are point processes [13]. Figure 1 shows the distribution of the number of earthquakes by magnitude in Richter scale. Point processes are stochastic models for random events happening in space and time, and the summary statistics of a point process is given by a function which is also known in geostatistics as the hazard function.

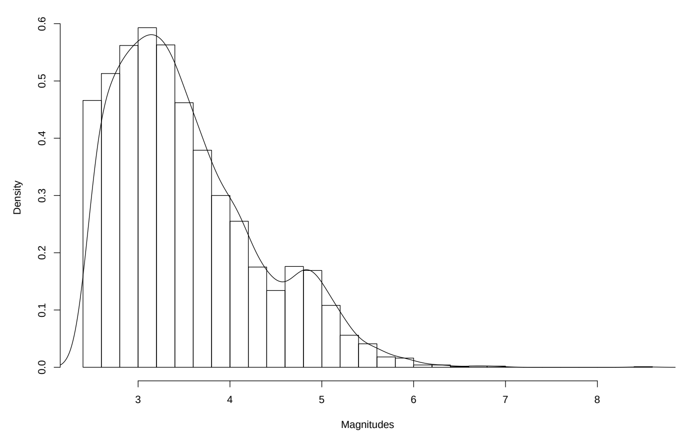

Fig. 1. Histogram of earthquake magnitudes in central Chile between years 2006 and 2010. Most of the seismic activity is concentrated in low magnitude events (less than 5 points in Richter scale), and few events with magnitude above 7 points. A single event of more than 8 points produces a long tail in the distribution.

Earthquakes can be clustered by using the main event (a single and largest magnitude) surrounded by several aftershocks, or alternatively by swarms of closely spaced events with similar magnitudes [7]. In order to perform spatial inference of the swarming behavior, we can fit a finite mixture model to represent the seismic activity represented by the hazard function of the underlying point process. Finite mixture models are probabilistic or model-based approaches for soft clustering and are characterized by the parametrization of mixing proportions, also having specific mixture component densities. Given that the neither the locations or the number of earthquakes is known *a priori*, the number of clusters or the number of mixture components have to be determined in a model selection step.

II. Related work

This paper has connections with other previous work where artificial intelligence and machine learning techniques have been used to model environmental systems [3]. In this context, supervised methods like neural networks have been preferred for building classifier systems. Also, unsupervised neural networks such as self-organizing maps have been proposed for modelling geophysical systems [11]. More closely related to this work, Ansari et.al. [1] proposed a fuzzy clustering algorithm for analyzing regionalized seismic hazard. The authors used a heuristic approach based on fuzzy hyper-volumes and partition density indexes for determining the number of clusters.

In this paper, we propose a Bayesian non-parametric approach for soft clustering earthquakes hazards. The Dirichlet process [6] is used to sample a distribution from an infinitely countable number of probability measures. In this approach, we integrate model selection (determining the number of clusters) and parameter estimation (determining the cluster centroids and mixing proportions) into a single inference step. Bayesian non-parametric methods for spatial mixture modeling were also formulated for rainfall measurements [8] and clustering cells in immunofluorescence histology [9].

III. Spatial mixture modelling

One of the main purposes of spatial modeling is prediction or estimating the realization of a random variable $y(x)$ in a spatial location x by means of a stationary Gaussian process $\theta(x)$. The residual is usually modelled as a zero-mean Gaussian process $\epsilon(x)$, and estimates at different locations yields a predictive surface of the process. Equation 1 represents the resulting Gaussian process process.

$$y(x) = \mu(x) + \theta(x) + \epsilon(x) \tag{1}$$

Now we concentrate in the Dirichlet process specifications, according to [8] we allow the stationary Gaussian process $\theta(x)$ to be a realization of a dependent Ditichlet process. A Dirichlet process is specified by a base distribution $G_o(x)$ and a concentration parameter α, so a distribution $G(x)$ is a sample from a DPM when:

$$G(x) \sim DP(\alpha, G_0(x)) \tag{2}$$

Extending finite mixture model to a non-parametric approach can be achieved by means of the Dirichlet Process Mixture (DPM) model [2]. More particularly, a Dirichlet process is used as to sample the conditional distribution of a finite mixture model with k components using a Dirichlet process prior. Given that many spatial models are neither Gaussian or stationary, parametric methods such as finite mixture models can be used to represent spatial dependency among a set of variables. More specifically, a Gaussian mixture model is a combination of a finite number of Gaussian densities, parametrized by their component parameters, such as the mean and covariance being written as $\phi_i = (\mu_i, \Sigma_i)$, but also having a mixing parameter π_i with $i = 1, \ldots, k$. The resulting density of a data point y can be written as:

$$p(y|\Theta) = \sum_{i=1}^{k} \pi_i \mathcal{N}(\phi_i) \tag{3}$$

Whose parameters are distributed according to:

$$y|c_i(x), \Phi \sim \mathcal{N}(\phi_i) \tag{4}$$
$$c_i(x)|\Pi \sim \mathcal{M}(\pi_1, \ldots, \pi_k) \tag{5}$$
$$\phi_i \sim G_0(x) \tag{6}$$
$$\pi_i \sim Dir(\alpha/k, \ldots, \alpha/k) \tag{7}$$

Where $c_i(x)$ represents a location-aware conditional latent variable that indicates the class where the data point y belongs, $\Pi = \{\pi_1, \ldots, \pi_k\}$ and $\Theta = \{\theta_1, \ldots, \theta_k\}$ represents the collections of all mixing and component parameters, and $\mathcal{M}(\cdot)$ and $Dir(\cdot)$ represents the multinomial and Dirichlet distributions respectively.

Taking n realizations of the spatial process $(y_1, \ldots, y_n)$ also yields a distribution of the indicator variables $c_i(x)$ given the mixing probabilities Π:

$$p(c_{1,i}(x), \ldots, c_{n,i}(x)|\Pi) = \prod_{i=1}^{k} \pi_i^{n_i} \tag{8}$$

with $n_i = \sum_{j=1}^{n} \delta_{j,i}$ being the number of observations belonging to class i. Assuming now the Dirichlet prior on π leaves the following conjugate form for the class consitional indicators:

$$p(c_{j,i}(x)|c_{\setminus j,i}(x), \alpha) = \frac{n_{\setminus j,i} + \alpha/k}{n - 1 + \alpha} \tag{9}$$

where $c_{\setminus j,i}$ represents all indicator variables for class j excepting the data point y_j, and $n_{\setminus j,i} = \sum_{l \neq j} \delta_{l,i}$

As the number of components k tends to infinity, from Neal [12] using a “Chinese Restaurant process” we represent the class conditional indicator variables as:

$$p(c_{j,i}(x)|c_{\setminus j,i}(x), \alpha) = \frac{n_{\setminus j,i}}{n - 1 + \alpha} \tag{10}$$
$$p(c_i \neq c_j \forall j < i\,|c_1, \ldots, c_{i-1}) = \frac{\alpha}{n - 1 + \alpha} \tag{11}$$

A. Spatial Hierarchical Dirichlet process mixture

Now we would like to concentrate on problems where a spatial DPM might not be able to successfully represent the diversity of a group of samples, so the spatial distribution also introduces a hierarchical structure by using a dependent Dirichlet process mixture (HDPM) [15]. The spatial HDPM extends the spatial DPM in a way that a new set of clusters is generated by each cluster of the base DPM. This setup allows to model spatial heterogeneity among a set of observations that shares a common feature.

In the case of earthquakes, events can be clustered around their magnitudes, but the spatial distribution does not have to be an stationary Gaussian random field. In this case, we allow the spatial distribution to be a DPM itself. Furthermore, the hierarchical extension is an straightforward extension of the DPM formulation, having now a base distribution H_0 specified by:

$$G(x) \sim DP(\gamma, H_0) \tag{12}$$
$$H_0 \sim DP(\alpha, G_0(x)) \tag{13}$$

B. Markov chain Monte Carlo implementation for the spatial DPM and HDPM

Markov chain Monte Carlo (MCMC) methods, and specially the Gibbs sampler plays a central role in Bayesian mixture modelling [5], where conjugate priors on the component parameters are used to for a hierarchichal sampling scheme. Gibbs sampling is an iterative MCMC scheme where each variable is updated in turn, using the its conditional distribution given all other variables.

$$p(y|c_i) \sim \mathcal{N}(\mu_i, \Sigma_i) \tag{14}$$

In the case of multivariate Gaussian mixtures, the prior for the mean μ_i is specified by a multivariate Gaussian distribution with hyper-parameters λ and r, so the prior can be written as:

$$p(\mu_i|\lambda, r) \sim \mathcal{N}(\lambda, r) \tag{15}$$

The hyper-parameters λ and r are conjugate priors, specified by:

$$p(\lambda) \sim \mathcal{N}(\mu_y, \Sigma_y) \tag{16}$$
$$p(r) \sim IW(1, \Sigma_y) \tag{17}$$

where μ_y and Σ_y are the mean and covariance of the data repectively, and IW represents the inverse-Wishart distribution.

Now, using the data likelihood from Equation 3, the posterior distribution of the means, conditioned on the prior and the indicator variables can be written as:

$$p(\mu_i|\, c, y, \Sigma_j, \lambda, r) \sim \mathcal{N}(\frac{\bar{y}_i n_i \Sigma_i + \lambda r}{n_i \Sigma_i + r}, \frac{1}{n_i \Sigma_i + r}) \tag{18}$$
$$\bar{y}_i = \frac{1}{n_j} \sum_{c_j = i} y_j \tag{19}$$

where n_i is the *occupation* number and $\bar{y}_j$ is the class conditional mean. Consequently, the posterior distribution of the hyper-parameters is given by:

$$p(\lambda|\mu_1, \ldots, \mu_k, r) \sim \mathcal{N}(\frac{\mu_y\, \Sigma_y^{-2} + r \sum_{j=1}^{k} \mu_j}{\Sigma_y^{-2} + kr}, \frac{1}{\Sigma_y^{-2} + kr}) \tag{20}$$

The component covariances Σ_j are also sampled from an inverse-Wishart distribution $p(\Sigma_j|\beta, w)$ with hyper-parameters β and w with the following distributions:

$$p(\beta) \sim IG(1, 1) \tag{21}$$
$$p(w) \sim IW(1, \Sigma_y) \tag{22}$$

The extension to the infinite limit has been [12] and [14], and consists of allocating data points to mixture components or creating new components using Equations 11. The extension to the hierarchical setup is performed by marginalizing the random effect variable, allocating data points to the resulting mixture model and creating a new DPM for each subset of the data.

IV. CASE STUDY : EARTHQUAKES MAGNITUDE IN CENTRAL CHILE

In order to exemplify the non-parametric approach, we analyze seismic activity in central Chile between the years 2006 and 2010. Chile is characterized by its continual seismic activity, but recently a devastating 8.8 magnitude earthquake hit the central part of the country. Figure 2 displays a summary of the recorded epicentres and magnitudes.

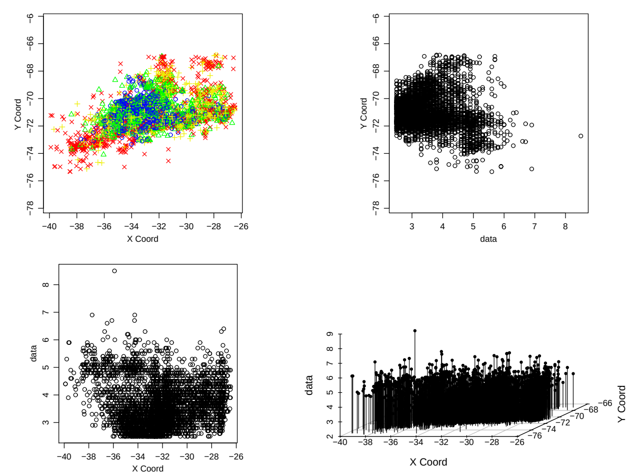

Fig. 2. Summary of the earthquakes dataset. The magnitude data is plotted against the latitude (X coord) and Longitude (Y coord).

The catastrophic dimensions of the earthquake leaded to several hundred human losses and more than US\$30 billion required to reconstruct the cities. Furthermore, several aftershocks with magnitude above 5 points in Richter scale continued to affect the country more than 3 months after the main earthquake. Figure 3 shows the locations of the main earthquakes in central Chile between the years 2006 and 2010.

Now we concentrate on the output of MCMC sampler for the spatial DPM. Figure 4 shows the number of mixture components used to fit the hazard function of a point process model of the data. Cluster centres are expected to be found in areas where there is more seismic activity, and aftershocks should be concentrated around those areas. Because neither

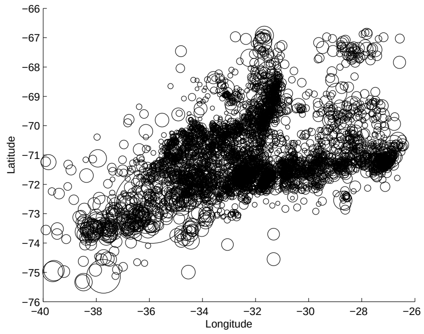

Fig. 3. Earthquakes locations in central Chile. Circles are used to represent the spatial location of an earthquake, and the radius is proportional to the magnitude. Several low magnitude earthquakes were recorded during the period of study, and the spatial distribution of bigger magnitude earthquakes shows no apparent sign of spatial clustering.

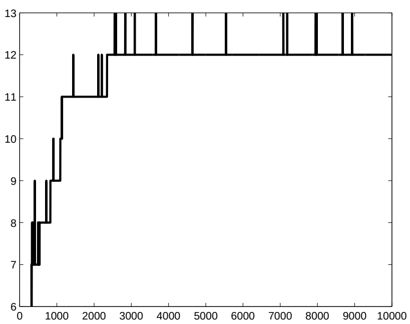

Fig. 4. Number of mixture components v/s MCMC iterations. The first 3000 iterations are taken as MCMC burn-in period, and the remaining of a total of 10000 iterations are used as the output of the algorithm. The MCMC sampler produces multiple configurations of the earthquakes hazard function being fitted with a finite mixture model.

the number of clusters or their locations is given, the DPM is able to sample multiple configurations of the hazard function.

Several clusters represent areas of low magnitude earthquakes (below 5 point of magnitude), which can be associated with background seismicity areas. It is worth noting that the “clustering” effect of the Chinese restaurant process prior defined in Equation 11 is in accordance with the distribution discussed for the magnitudes in Figure 1. Most earthquakes are below 5 points magnitude so they would enter into a cluster given the number of existing events associated to it.

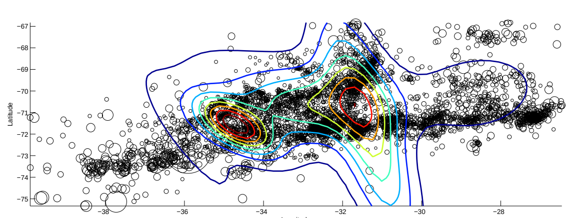

Fig. 5. Sample from a DPM after 10000 iterations. A 12 component mixture model is used to represent the hazard function of earthquakes magnitudes at different locations.

From Figure 5 we can see that earthquakes with big magnitudes are still not represented by a locally stationary Gaussian random field. Figure 6 represent the number of components obtained after 10000 iterations.

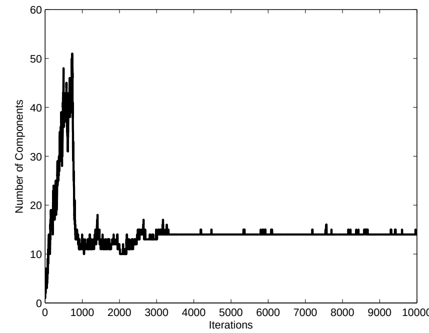

Fig. 6. Number of mixture components v/s MCMC iterations for the spatial HDPM. After 10000 iterations, a 14 component model is used to represent the magnitudes.

Now, we concentrate on the subset of the data that was under-represented by the spatial DPM. Taking the data points belonging to the higher magnitude events, we run a DPM for the spatial distribution (locations) of that subset of the data. Figure 7 shows a sample of the resulting HDPM after 10000 iterations.

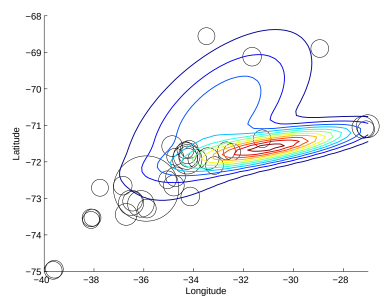

Fig. 7. Sample from a HDPM after 10000 iterations. A 2 component mixture model represents the spatial distribution of earthquakes above 5 degrees magnitude.

V. Conclusions

We proposed a Bayesian non-parametric approach to modeling geophysical events. The model is based on the spatial Dirichlet process mixture, and we have shown that the implementation is an straightforward extension of standard MCMC procedures for Dirichlet processes. This approach allowed us to overcome a complex model selection step, which is not easy to solve in many geophysical problems where the collected data might not be fully explanatory for the response variables.

Moreover, we have also highlighted the potential issues of modelling geo-referenced data as stationary Gaussian processes. In that sense, a finite mixture model enables to relax that assumption, providing a non-Gaussian representation to the posterior density. A Bayesian approach for mixture models is also taken, and the resulting hierarchical model is also sorted with the same MCMC algorithm.

Further work will consider the associated time differences of foreshocks and aftershocks, as well as the depth of the earthquakes. In that case a spatio-temporal Markovian model can be considered, so an extension of the hierarchical approach could be used.

References

[1] A. Ansari, A. Noorzad, and H. Zafarani. Clustering analysis of the seismic catalog of iran. *Computers & Geosciences*, 35(3):475 – 486, 2009.

[2] C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Institute of Mathematical Statistics*, 2(6):1152–1174, 1974.

[3] Serena H. Chen, Anthony J. Jakeman, and John P. Norton. Artificial intelligence techniques: An introduction to their use for modelling environmental systems. *Math. Comput. Simul.*, 78(2-3):379–400, 2008.

[4] N. Cressie. *Statistics for Spatial Data*. New York. John Wiley and Sons, 1993.

[5] J. Diebolt and C. P. Robert. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2):363–375, 1994.

[6] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, March 1973.

[7] V. I. Keilis-Borok G. I. Barenblatt and M. M. Vishik. Model of clustering of earthquakes. *Proceedings of the National Academy of Sciences of the United States of America,*, 78(9):5284–5287, Sept. 1981.

[8] A. E. Gelfand, A. Kottas, and S. N. MacEachern. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, 100(471):1021–1035, 2005.

[9] C. Ji, D. Merl, T. B. Kepler, and M. West. Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology. *Bayesian Analysis*, 4(2):191–412, 2009.

[10] Y. Kagan. Likelihood analysis of earthquake catalogues. *Geophysical journal international*, 106(1):135–148, 1991.

[11] A. Köhler, M. Ohrnbergera, and F. Scherbaum. Unsupervised feature selection and general pattern discovery using self-organizing maps for gaining insights into the nature of seismic wavefields. *Computers & Geosciences*, 35(9):1757–1767, 2009.

[12] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

[13] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[14] C. E. Rasmussen. The infinite Gaussian mixture model. *Advances in Neural Information Processing Systems*, (12):554–560, 2000.

[15] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[16] D. Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1):1–62, 1970.