Journal of Data Science **11**(2013), 623-638

Nonparametric Assessment of Aftershock Clusters of the Maule Earthquake $M_w = 8.8$

Javier E. Contreras-Reyes
University of Valparaíso

Abstract: We study the spatial distribution of clusters associated to the aftershocks of the megathrust Maule earthquake M_W 8.8 of 27 February 2010. We used a recent clustering method which hinges on a nonparametric estimation of the underlying probability density function to detect subsets of points forming clusters associated with high density areas. In addition, we estimate the probability density function using a nonparametric kernel method for each of these clusters. This allows us to identify a set of regions where there is an association between frequency of events and coseismic slip. Our results suggest that high coseismic slip is spatially related to high aftershock frequency.

Key words: Aftershock distribution, kernel density estimation, Maule 2010 earthquake, nonparametric density-based clustering, slip model.

1. Introduction

Recent applications of clustering techniques range over an enormous set of disciplines, both in the natural sciences and in the social sciences. A standard account is the book of Kaufman and Rousseeuw (1990), which is focused mainly on the more classical methods, based of the notion of distance between objects. An alternative, relatively more recent approach is the model-based clustering formulation which regards the observed data as generated by a probability distribution of finite mixture of multivariate random variables having distribution belonging to some parametric family. In the implementation of this approach, the most common option is to adopt the multivariate Gaussian assumption for each of the density components, and estimate their parameters using an EM-type algorithm. An useful account to this approach is provided by Dasgupta and Raftery (1998). By applying the model-based clustering approach to earthquake data in the coastal area of central California, Dasgupta and Raftery (1998) have obtained six clusters, some of which are clearly linked to active faults. However one or two of their clusters do not correspond to some already identified area.

In a further approach to the clustering problem, the notion of an underlying density function $f(x)$ is retained, but the assumption that $f(x)$ as a finite mixture of components is removed, and so also the connected parametric assumption of the components. Therefore, this approach is based on the construction of a nonparametric estimate $\hat{f}(x)$ of $f(x)$, and the association of a cluster to each of the observed modes of the density. We concentrate on the kernel-type estimate formulation which is the most popular and conceptually very intuitive (Bowman and Azzalini, 1997). In the univariate case, the kernel estimate is defined by

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \varphi(x - x_i; h), \tag{1}$$

where $\varphi(z; h)$ denotes the normal density function with zero-mean and standard deviation h evaluated at point z. The normal density is adopted for simplicity and it could be replaced by another symmetric density without much effect on the outcome. A much more important role is played by h which in this context is called "smoothing parameter". In the d-dimensional case, $\varphi(z; h)$ is replaced by a multivariate density with a zero-mean vector; the simplest and most popular choice is the product of d such terms, with a different smoothing parameter for each of the d components. Clusters are then formed by the data points associated to the modes of $\hat{f}(x)$, and the clusters are separated by regions of low density of points. This logic procedure is referred as a nonparametric clustering (NPC).

The aim of the present work is to examine the spatial dependence between areas of high coseismic slip and the aftershock frequency (seismic clusters). To achieve this, we identify the distribution of clusters in the rupture area of the 2010 earthquake. The hypocenter data are taken from the SSN (Servicio Sismológico Nacional, Chile) and we use the coseismic model of Moreno *et al.* (2012), which includes all available geodetic data for the Maule earthquake. We adopted the NPC formulation of Azzalini and Torelli (2007), which has the advantage of not requiring specification of the number of existing clusters. We apply this methodology to the SSN aftershock catalogue data of the Maule earthquake and focus on the data of the area between Valparaíso and Tirúa [33-38.5°S], aiming the identification of seismic clusters and its spatial relationship with regions of coseismic slip.

2. Methodology

We briefly describe the NPC method proposed by Azzalini and Torelli (2007). This works by assuming that the available set of d-dimensional observations $S = \{x_1, \cdots, x_n\}$ represent a set of points drawn from a continuous multivariate random variable having an unknown probability density function $f(x)$, $x \in \mathbb{R}^d$.

Here, $x = (\text{latitude}, \text{longitude})$ denotes a geographical position ($d = 2$) and the data points $x_1, \cdots, x_n$ represent the positions of the observed seismic events.

For any given constant α such that $0 \leq \alpha \leq \max_{x \in \mathbb{R}^2} f(x)$, consider the high density region defined by

$$R_\alpha = \{x : \ x \in \mathbb{R}^2, \ f(x) \geq \alpha\}, \tag{2}$$

which has an associated probability $p(\alpha) = \int_{R_\alpha} f(x)dx$. The region R_α is in general, formed by a number m of connected sets, where $m \in \{1, 2, \cdots\}$.

Now we let α move along its range. This causes both m and $p = p(\alpha)$ to move accordingly, and we can regard m as a function of p since p is monotonic with respect to α, write $m(p)$. Note that $m(p)$ is instead a non-monotonic function. With the additional conventional settings $m(0) = m(1) = 0$, it can be shown that the total number of increments of the step function $m(p)$ is equal to the number M of modes of $f(x)$, hence to the number of clusters, in the sense defined earlier. As α varies along its range, and so does $p(\alpha)$, the corresponding connected components of $R(\alpha)$ form a hierarchical tree structure.

Translating the above idea into a working methodology requires some additional specifications and algorithmic work. We have sketched here only the main steps of the procedure. Full details of the method are given by Azzalini and Torelli (1997) and its implementation in R language is provided by Azzalini *et al.* (2011). This procedure will later referred to as the "pdfCluster method".

1. A nonparametric estimate $\hat{f}(x)$ of the density is obtained from the observed sample S.

2. For any α in the range $0 \leq \alpha \leq \max_i \hat{f}(x_i)$, it considers the sample analogue of (2) given by $S_\alpha = \{x_i : \ x_i \in S, \ \hat{f}(x_i) \geq \alpha\}$.

3. The above step is replicated for a grid of values spanning the admissible range of α, from 0 to $\max_i \hat{f}(x_i)$. This generates a mode function $\hat{m}(p)$ and a tree structure of the modes.

4. At the end of the earlier step we have obtained a tree structure of the modes of $\hat{f}(x)$. Moreover, for each of the M modes, we have allocated some elements of the sample S to the given mode. For each unallocated point x_0, we must select one of distributions $\hat{f}_1(x), \cdots, \hat{f}_M(x)$ which represent the estimated densities of the cluster cores. This allocation is most naturally based on the likelihood ratio, that is we allocate x_0 to the j-th cluster core such that the ratio $\hat{f}_j(x_0)/\max_{k \neq j} \hat{f}_k(x_0)$ is highest.

The final outcome of the clustering process is represented by the a partition of the sample S into a set of clusters, say $C_1, \cdots, C_M$.

2.1 Density-Based Silhouette Diagnostic

In classical cluster analysis, the term 'silhouette' refers to a diagnostic tool for the validation of the outcome of the clustering process (Rousseeuw, 1987). This technique provides a graphical representation of how appropriately the data has been clustered. The idea arises from the comparison of the small distance of each observation to the cluster where it has been allocated and a measure of separation from the closest alternative cluster.

An adaption of the silhouette idea to density-based clustering methods has been proposed by Menardi (2010). The method, called density-based silhouette (DBS) diagnostic, is based on the posterior probabilities

$$p_j(x_i) = \frac{\hat{\pi}_j \hat{f}_j(x_i)}{\sum_{j=1}^{M} \hat{\pi}_j \hat{f}_j(x_i)}, \qquad j = 1, \cdots, M,$$

where $\hat{\pi}_j$ plays the role of prior probability of cluster C_j; in practice it is taken to be the proportion of points allocated to C_j. The DBS index of observation x_i is

$$\mathrm{DBS}(x_i) = \frac{\log\left(\frac{p_{j_0}(x_i)}{p_{j_1}(x_i)}\right)}{\max_{k\in\{1,\cdots,n\}}\left|\log\left(\frac{p_{j_0}(x_k)}{p_{j_1}(x_k)}\right)\right|},$$

where j_0 denotes the cluster to which x_i has been allocated, and j_1 refers to the alternative cluster index for which p_j is maximum, $j \neq j_0$. In our case, after partitioning the SSN data with the pdfCluster method, we applied the DBS diagnostic to assess the quality of the outcome.

2.2 Temporal Analysis

We briefly describe four indexes to illustrate the consistency of the NPC method across the time. These indexes provide information on the accuracy of the adopted clustering in reconstructing the correct categories, compared to random choice. In Table 1, $N(F_i, O_j)$ denotes the number of forecasts in category i that had observations in category j, $N(F_i)$ denotes the total number of forecasts in category i, $N(O_j)$ denotes the total number of observations in category j, N is the total number of forecasts and $i, j = 1, \cdots, M$ are the indexes of M clusters. The NSS index corresponds to normal skill score with range $[0, 1]$, such that 0 indicates no skill and 1 indicates perfect score. The index HSS correspond to Heidke skill score (Brier and Allen, 1952) with range $(-\infty, 1]$, such that 0 indicates no skill and 1 indicates perfect score. The index HK correspond to Hanssen and Kuipers discriminant (Hanssen and Kuipers, 1965) with range

$[-1, 1]$, such that 0 indicates no skill, and 1 indicates perfect score. The NSS, HSS, and HK indexes are defined as

$$\text{NSS} = \frac{1}{N}\sum_{i=1}^{M} N(F_i, O_i),$$

$$\text{HSS} = \frac{NSS - \frac{1}{N^2}\sum_{i=1}^{M} N(F_i)N(O_i)}{1 - \frac{1}{N^2}\sum_{i=1}^{M} N(F_i)N(O_i)},$$

$$\text{HK} = \frac{N^2 - \sum_{i=1}^{M} N(F_i)N(O_i)}{N^2 - \sum_{i=1}^{M} N(O_i)^2}\text{HSS},$$

respectively.

Table 1: Multi-category Contingence table with M clusters

		Observed Category				
		1	2	...	M	Total
	1	$N(F_1, O_1)$	$N(F_1, O_2)$	$\cdots$	$N(F_1, O_M)$	$N(F_1)$
Forecast	2	$N(F_2, O_1)$	$N(F_2, O_2)$	$\cdots$	$N(F_2, O_M)$	$N(F_2)$
Category	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	M	$N(F_M, O_1)$	$N(F_M, O_2)$	$\cdots$	$N(F_M, O_M)$	$N(F_M)$
Total		$N(O_1)$	$N(O_2)$	...	$N(O_M)$	N

Hubert and Arabie (1985) noticed that the Rand index is not corrected for chances that are equal to zero and for random partitions having the same number of objects in each class. They introduced the corrected Rand index, whose expectation is equal to zero under random allocation. The adjusted Rand (AR) index is

$$\text{AR} = \frac{r - E(r)}{\max(r) - E(r)},$$

where

$$r = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} N(F_i, O_j)[N(F_i, O_j) - 1],$$
$$E(r) = \frac{1}{2N(N-1)}N^*(O)N^*(F),$$
$$\max(r) = \frac{1}{2}[N^*(O) + N^*(F)],$$

$$N^*(O) = \frac{1}{2}\sum_{i=1}^{M} N(F_i)[N(F_i) - 1],$$

$$N^*(F) = \frac{1}{2}\sum_{j=1}^{M} N(O_j)[N(O_j) - 1].$$

This maximum value is questionable since the number of common joined pairs is necessarily bounded by $\inf\{N^*(O), N^*(F)\}$, but $\max(r)$ ensures that the maximum value of AR is 1 when the two partitions are identical.

3. Results

Our numerical work is based on data extracted from the SSN catalogue, available at http://ssn.dgf.uchile.cl/. Specifically, we have considered 6,714 aftershocks in a map [32–40°S]×[69–75.5°E], for a period between 27 February 2010 and 13 July 2011 (see Figure 1) and for local magnitudes $M_l \geq 2.0$ (Contreras-Reyes and Arellano-Valle, 2012). All of these observations have been pre-processed with SEISAN 8.3 software starting from the information provided by 22 stations located in a map with coordinates $[-33.32, -39.80]$ latitude and $[-70.29, -73.24]$ longitude.

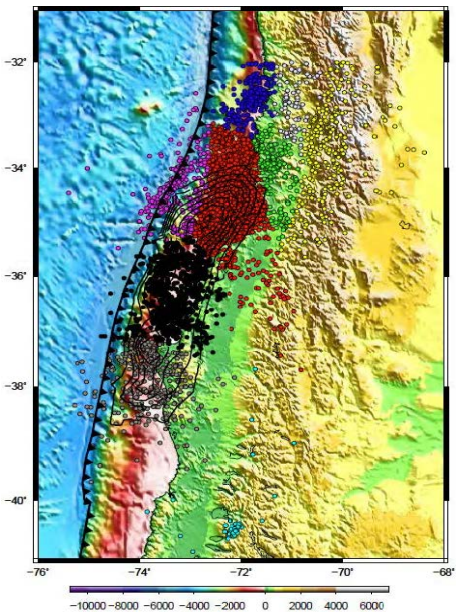

Figure 1: Map of the Chile region analyzed for slip (solid black line) and post-seismicity correlation with clustering events. The black-triangle line correspond to the trench

For the numerical processing, we used the R computing environment software (R Development Core Team, 2012). Most of the work was done using the R package `pdfCluster` by Azzalini *et al.* (2011); this package comprises the functions `pdfCluster` for the NPC method, `dbs` for DBS diagnostics and `kepdf` for kernel density estimation. In addition, R provides the `wilcox.test` function for individual test of means.

3.1 Clustering process

Figure 1 displays the geographical map of the area of interest with the points denoting the locations of the events. These points have been clustered using the pdfCluster method described in Section 2, leading to groups identified by different colors. Figure 2 shows the silhouette diagnostic and cluster tree. These indicate a lack of clear separation among clusters, which is not surprising, given the close geographical proximity of the clouds of points visible in the top panel of the figure.

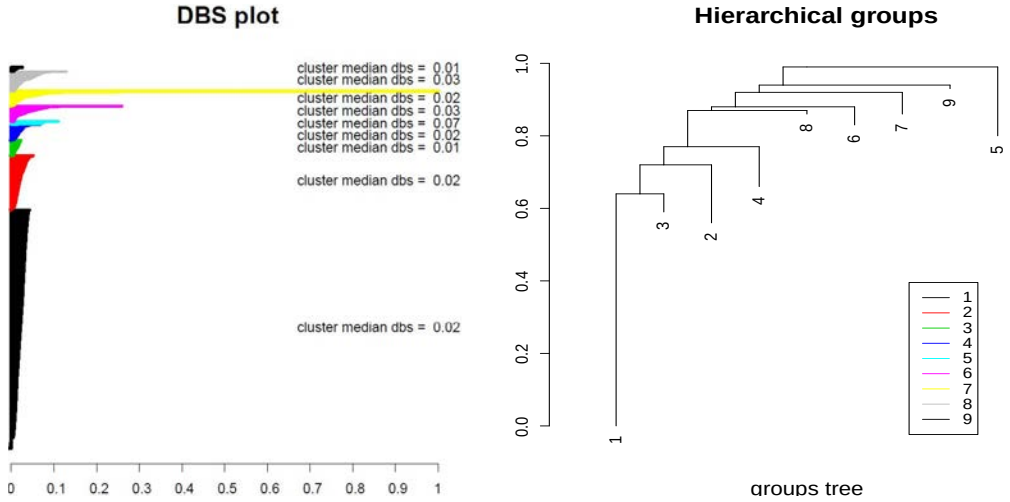

Figure 2: Plots of NPC method results

In the second stage of our numerical work, we have introduced two variants. One was to consider only the areas were slip took place in a non-negligible form, since our aim was exactly to examine the implications of the slip model. In addition, some numerical exploration has indicated that log-transformation of the quantities, slip and density function, lead to a more meaningful outcome. Since for a number of points the slip value is 0, we adopted the modification commonly used in similar cases of adding a small positive quantity, that is, working with $\log(k + \text{slip})$ for some small value k. Since slip is measured in meters, then we adopted $k = 0.01$ which represents a perturbation of only 1 cm of the original data. In the following, the term log-slip will be used for referring to $\log(0.01 + \text{slip})$.

Furthermore, association between slip and events density can be examined in two different ways. One is to chose a regular grid of points in the region of interest, and evaluate these variables or their log-transforms over this grid. The other option is to evaluate these variables at the observed points of the seismic events.

Figures 3, 4 and 5 refers to the first form, for three choices of the geographical area over which the grid of points is constructed. More precisely, the sets of points for which the computations have been performed have been obtained as the intersection of rectangular grids of sizes (192×214), (149×214) and (119×214) with the three regions shown on the left side of Figures 3, 4 and 5, respectively, of different geographical size. This process led then to consider three non-rectangular grids, comprising 19630, 10696 and 4812 points, respectively. The area covered by first grid includes the largest number of points associated of the seismic events of the data set, while the last one refers to the area with highest concentration of events. The right side of the figure displays the scatter plots of log-slip and log-density of the points, separately for each cluster. Only clusters labeled No. 1, 2, 3, 6, and 8 are considered here ($M = 5$, see Table 1); the other clusters have been dropped because the covered area of the slip model is included in the area of the retained clusters. Therefore only the selected clusters can be used for the analysis of the correlation of aftershock density with the slip model of Moreno *et al.* (2012). However, all clusters are considered in the 'dbs' plot and silhouette analysis of Figure 2 because the NPC analysis was run over the 6,714 aftershocks.

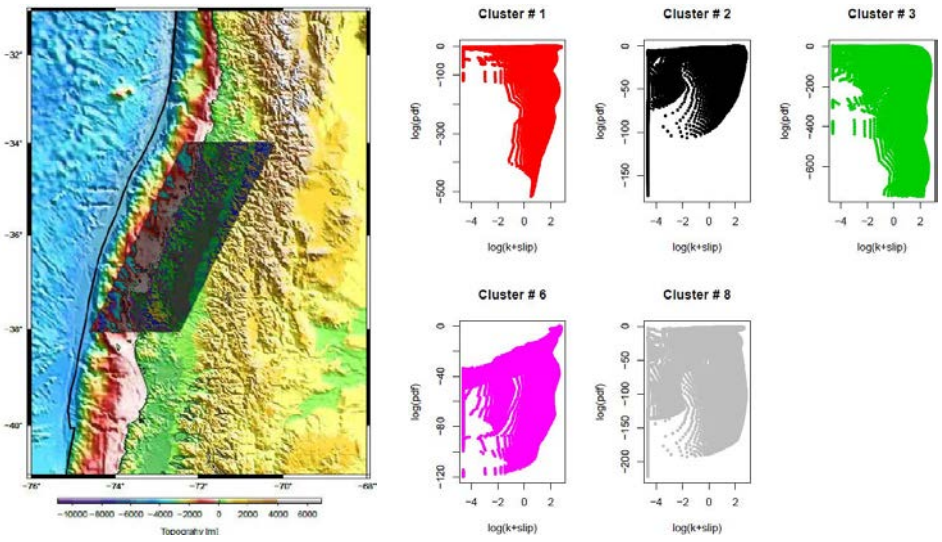

Figure 3: Left: Rectangular grid of size (192×214) points considered for the NPC method (blacked shadows). Right: Relationship between log-slip and post-seismicity log-frequency. The points with slip $= 0$ produce verticals strips of points at abscissa $\log(0.01) \approx -4.6$ in some of the top plots

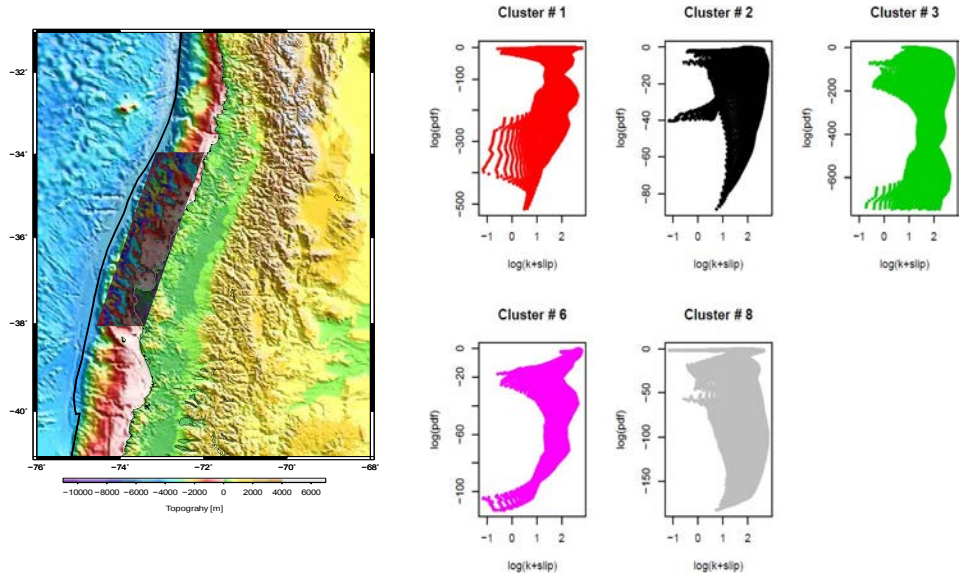

Figure 4: Left: Rectangular grid of size (149×214) points considered for the NPC method (blacked shadows). Right: Relationship between log-slip and post-seismicity log-frequency. The points with slip $= 0$ produce verticals strips of points at abscissa $\log(0.01) \approx -4.6$ in some of the top plots

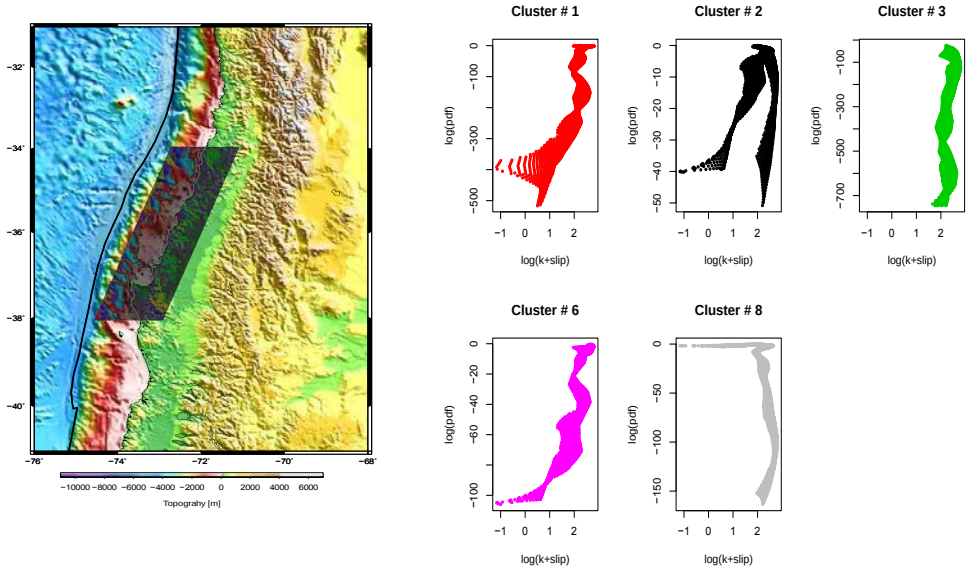

Figure 5: Left: Rectangular grid of size (119×214) points considered for the NPC method (blacked shadows). Right: Relationship between log-slip and post-seismicity log-frequency. The points with slip $= 0$ produce verticals strips of points at abscissa $\log(0.01) \approx -4.6$ in some of the top plots

Even if the selection of the three grids is somewhat subjective, the overall indication of Figures 3, 4 and 5 provides convincing evidence of the presence of association between the variables for the clusters under consideration, specifically those labeled with No. 1, 2, and 6. This association is becoming more and more marked as we move from the first to the last grid, that is, when we focus on the area with greater intensity of events. The type of association is definitely non-linear, and so admittedly it does not lend itself to simple interpretation, but it is clearly present, especially so in the bottom portion of the figure.

Figure 6 refers instead to the second form of comparison, where evaluation of log-slip and log-density is performed at the observed location of events instead of a regular grid of points. Also this figure exhibits some noteworthy features. One is that in the red, black, violet, and grey clusters there exists a clear positive association between log-pdf and log-slip. Hence, the maximum slip is associated with high frequency of events. The green cluster does not display any association, presumably so because several events matching with null slip zone where probably have not been involved with the main earthquake; however the slip produced in this zone is lower. Pichilemu city (34.38°S, 72.02°W) is located in the middle of the red cluster, approximately, which where the maximum slip is 16.6 meters. The sky-blue cluster correspond to seismic activity produced by Puyehue volcano (40.35°S, 72.50°W) eruption (June, 2011). Mocha Island (38.39°S, 73.87°W) is located at the bottom of the gray cluster, where the maximum slip is 11.9 meters (see Figure 1). In the black and gray clusters, we can see a positive association: the values of log-slip increment as log-pdf increases.

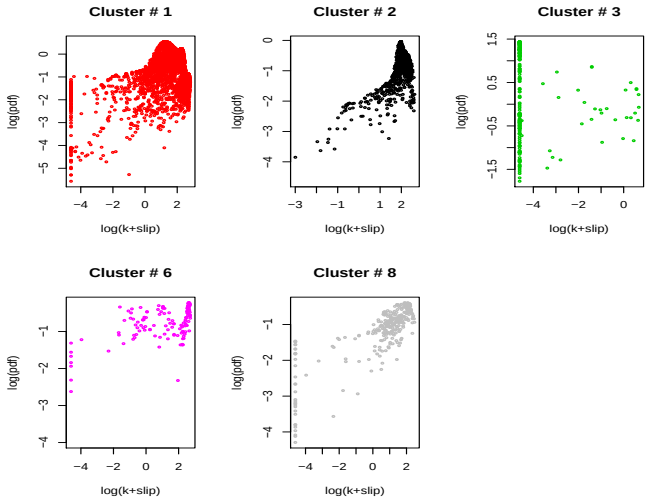

Figure 6: Relationship between log-slip and post-seismicity log-frequency for the clusters obtained for the NPC method of Figure 1. The points with slip $= 0$ produce verticals strips of points at abscissa $\log(0.01) \approx -4.6$ in some of the top plots

3.2 Clusters Comparisons

To examine where the differences among the groups are, we make use of the Wilcoxon test (Wilcoxon, 1945). In essence, this test compares the medians of two groups under the null hypothesis that the two medians are equal. The results for all possible pair comparisons are shown in Table 2. If each p-value is consider isolatedly, there is only on non-significant comparison at 5% level, but we must make an allowance for repeated testing; in this case, 10 testing procedures have been performed. The more classical form of allowance for repeated testing is via the Bonferroni correction, which here leads to consider the $0.05/10 = 0.005$ significance level. Therefore, also the value 0.0181 must be regarded as non-significant. We can see in Table 3 that the red cluster representing the Pichilemu zone has the higher maximum of slip in relation with red (Constitución zone) and violet (Pichilemu's offshore coast zone) clusters. The gray (Arauco zone) and green (Rancagua city zone) clusters display low values of slip; in practice, the green cluster does not exihibit slip effect.

Table 2: P-values for Wilcoxon test for clusters slip

	red	black	green	violet	gray
red	-	0	0	0.018	0
black	0	-	0	0.350	0
green	0	0	-	0	0
violet	0.018	0.350	0	-	0
gray	0	0	0	0	-

Table 3: Summary statistics of the slip variable by cluster and geodetic distance of clusters from the trench

	Statistics					Geodetic Distance		
Cluster	Mean	Min	Max	S.D.	N	Min	Max	Mean
red	6.20	0	16.57	4.33	4165	18.34	326.78	100.67
black	7.58	0.04	13.73	2.62	950	1.89	173.39	74.53
green	0.08	0	1.95	0.33	265	98.82	219.64	147.55
violet	7.57	0	15.04	5.75	149	1.31	168.93	27.56
gray	4.04	0	11.88	3.03	308	3.42	212.91	68.75

The clustering outcome can evolve with number of days after the mega earthquake and the observations involved in each day; so it may produce non-constant results. Hence, for $t > 1$, we compare the results at day t respect to day $t-1$ to compare two alternative partitions of the same set. In each comparison it is necessary to keep the same number of clusters at day t and $t+1$. Figure 7 shows the

consistency of NP method along 200 days since the moment of the great earthquake with values higher than 0.89 for AR case, 0.95 for NSS and HSS cases, and 0.7 for HK case. In the first 200 days, some compressions between the clusters estimation at day t versus day $t-1$ produce lower values of the indexes by the incorporation of one or more new groups related to the added observations.

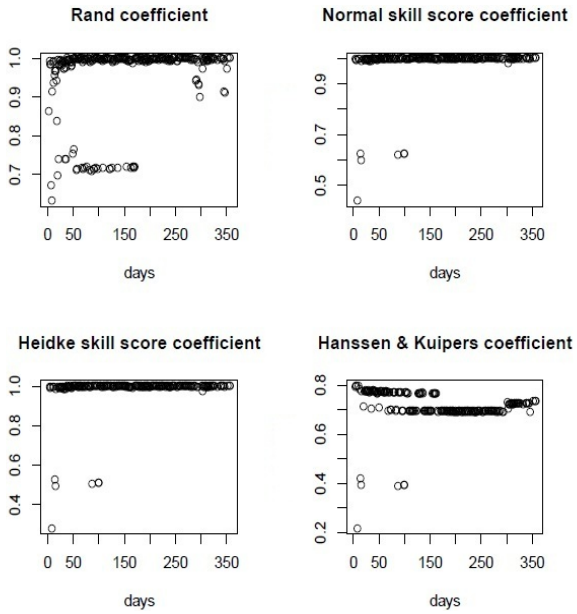

Figure 7: Plots of indexes of NPC results comparisons of data set at t day versus $t-1$ day

4. Discussion

The most recent major Chilean earthquake occurred on February 27th, 2010 ($M_w = 8.8$) filled a seismic gap (Ruegg *et al.*, 2009) that has experienced little seismic activity since 1835, when it broke with an estimated magnitude of $M_w \sim 8.5$ (Darwin, 1851, p. 768). Moreno *et al.* (2010) showed that the two regions of high coseismic slip of the Maule earthquake appeared to be highly locked before the earthquake. Subsequent geodetic studies have established that the main coseismic slip patch (>15 m) is located in the northern part of the rupture area, with a secondary concentration of slip to the south (5-12 m) (e.g., Lay *et al.*, 2010; Delouis *et al.*, 2010; Lorito *et al.*, 2011; Vigny *et al.*, 2011; Moreno *et al.*, 2012). Lorito *et al.* (2011) concluded that increased stress on the unbroken southern patch may have increased the probability of another great earthquake there in the near future, but his model has poor resolution on this area. In addition, Moreno *et al.* (2012) suggests that coseismic slip heterogeneity at the scale of single asperities appears to indicate seismic potential for future great earth-

quakes. These studies are not limited to the analysis of geodetic data; seismic and tsunami data have been used as well.

The pronounced crustal aftershock activity with mainly normal faulting mechanisms is found in the Pichilemu region (Farías *et al.*, 2011; Lange *et al.*, 2012, Rietbrock *et al.*, 2012). Lange *et al.* (2012) consider the processed events between 15 March and 30 September 2010 to estimate local magnitudes (M_l) in the Pichilemu region, where those magnitudes are comparable with the SSN magnitudes for large events. Specifically, a crustal aftershock activity is found in the region of Pichilemu ($\sim 34.5°$S) where the crustal events occur in a ~ 30 km wide region with sharp inclined boundaries and oriented oblique to the trench. On the other hand, the aftershock seismicity parallel to the trench is apparent at 50-120 km distance perpendicular to the trench (see Table 3). Near $\sim 35°$S and in the southern part of the rupture at $\sim 38°$S, significant aftershock activity occurs after the megathrust earthquake. This seismicity takes place in regions of high coseismic slip (see Table 3). Aftershocks and coseismic slip of the Maule 2012 earthquake terminate ~ 50 km south of the prolongation of the subducting Mocha Fracture zone around $\sim$(73.5°W, 38.5°S), near of the bottom of gray cluster (see Figure 1).

We have proposed an alternative way to clustering the aftershocks seismicity of the 2010 Maule earthquake M_W 8.8. The nonparametric clustering has shown to be consistent in the measure that the dairy aftershocks events are added in the analysis and we present the diagnostic tools to illustrate this feature. We employ a nonparametric kernel method to fit the high aftershock frequency, which were highly correlated with the used coseismic slip model. Our findings can be explored further by considering an extended data set, including the events with delayed effect, and modeling the relationship of high coseismic slip areas and aftershock clusters. Also, this catalogue should be considered to the study of the behavior of an aftershock sequence, to identify outliers and to classify sequences into groups exhibiting similar aftershock behavior (Schoenberg and Tranbarger, 2008). Finally, this analysis can be considered in attempting identification of an increasing risk of occurrence of another major earthquake.

Acknowledgements

The author would like to thank Giovanna Menardi and Adelchi Azzalini for useful comments and suggestions in connection with the pdfCluster methodology; to Reinaldo Arellano-Valle and Eduardo Contreras-Reyes for useful comments, suggestions and computational help; and to Marcos Moreno and Hector Massone for seismicity data base used in this article. The author also would like to thank referee and the editor for thoughtful comments which led to substantial improvement in presentation of the material.

References

Azzalini, A. and Torelli, N. (2007). Clustering via nonparametric density estimation. *Statistics and Computing* **17**, 71-80.

Azzalini, A., Menardi, G. and Rosolin, T. (2011). R package pdfCluster: cluster analysis via nonparametric density estimation (version 0.1-13). http://cran.r-project.org/package=pdfCluster.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*: *The Kernel Approach with S-Plus Illustrations.* Oxford University Press, Oxford.

Brier, G. W. and Allen, R. A. (1952). Verification of weather forecast. In *Compendium of Meteorology* (Edited by T. F. Malone), 841-848. American Meteorological Society, Boston.

Contreras-Reyes, J. E. and Arellano-Valle, R. B. (2012). Kullback-Leibler divergence measure for multivariate skew-normal distributions. *Entropy* **14**, 1606-1626.

Darwin, C. (1851). *Geological Observations on Coral Reefs, Volcanic Islands and on South America.* Smith, Elder and Company, London.

Dasgupta, A. and Raftery, A. E. (1998). Detecting features in spatial point processes with cluster via model-based clustering. *Journal of the American Statistical Association* **93**, 294-302.

Delouis, B., Nocquet, J. M. and Vallee, M. (2010). Slip distribution of the February 27, 2010 $M_w = 8.8$ Maule earthquake, central Chile, from static and high-rate GPS, InSAR, and broadband teleseismic data. *Geophysical Research Letters* **37**, L17305.

Farías, M., Comte, D., Roecker, S., Carrizo, D. and Pardo, M. (2011). Crustal extensional faulting triggered by the 2010 Chilean earthquake: the Pichilemu seismic sequence. *Tectonics* **30**, TC6010.

Hanssen, A. W. and Kuipers, W. H. A. (1965). On the relationship between the frequency of rain and various meteorological parameters. *Mededelingen en verhandelingen* **81**, 2-15.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification* **2**, 193-218.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data*: *An Introduction to Cluster Analysis.* Wiley, New York.

Lange, D., Tilmann, F., Barrientos, S. E., Contreras-Reyes, E., Methe, P., Moreno, M., Heit, B., Agurto, H., Bernard, P., Vilotte, J. P. and Beck, S. (2012). Aftershock seismicity of the 27 February 2010 M_w 8.8 Maule earthquake rupture zone. *Earth and Planetary Science Letters* **317-318**, 413-425.

Lay, T., Ammon, C., Kanamori, H., Koper, K. D., Sufri, O. and Hutko, A. R. (2010). Teleseismic inversion for rupture process of the 27 February 2010 Chile ($M_w = 8.8$) earthquake. *Geophysical Research Letters* **37**, L13301.

Lorito, S., Romano, F., Atzori, S., Tong, X., Avallone, A., McCloskey, J., Cocco, M., Boschi, E. and Piatanesi, A. (2011). Limited overlap between the seismic gap and coseismic slip of the great 2010 Chile earthquake. *Nature Geosciences* **4**, 173-177.

Menardi, G. (2010). Density-based silhouette diagnostics for clustering methods. *Statistics and Computing* **21**, 295-308.

Moreno, M., Rosenau, M. and Oncken, O. (2010). Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone. *Nature* **467**, 198-202.

Moreno, M., Melnick, D., Rosenau, M., Baez, J., Klotz, J., Oncken, O., Tassara, A., Chen, J., Bataille, K., Bevis, M., Socquet, A., Bolte, J., Vigny, C., Brooks, B., Rider, I., Grund, V., Smalley, B., Carrizo, D., Bartsch, M. and Hase, H. (2012). Toward understanding tectonic control on the M_w 8.8 2010 Maule Chile earthquake. *Earth and Planetary Science Letters* **321-322**, 152-165.

Pollitz, F. F., Brooks, B., Tong, X., Bevis, M. G., Foster, J. H., Burgmann, R., Smalley, R., Vigny, C., Socquet, A., Ruegg, J. C., Campos, J., Barrientos, S., Parra, H., Baez Soto, J. C., Cimbaro, S. and Blanco, M. (2011). Coseismic slip distribution of the February 27, 2010 M_w 8.8 Maule, Chile earthquake. *Geophysical Research Letters* **38**, L09309.

R Development Core Team (2012). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rietbrock, A., Ryder, I., Hayes, G., Haberland, C., Comte, D., Roecker, S. and Lyon-Caen, H. (2012). Aftershock seismicity of the 2010 Maule $M_w = 8.8$,

Chile, earthquake: correlation between co-seismic slip models and aftershock distribution? *Geophysical Research Letters* **39**, L08310.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53-65.

Ruegg, J. C., Rudloff, A., Vigny, C., Madariaga, R., de Chabalier, J. B., Campos, J., Kausel, E., Barrientos, S. and Dimitrov, D. (2009). Interseismic strain accumulation measured by GPS in the seismic gap between Constitución and Concepción in Chile. *Physics of the Earth and Planetary Interiors* **175**, 78-85.

Schoenberg, F. P. and Tranbarger, K. E. (2008). Description of earthquake aftershock sequences using prototype point patterns. *Environmetrics* **19**, 271-286.

Vigny, C., Socquet, A., Peyrat, S., Ruegg, J. C., Métois, M., Madariaga, R., Morvan, S., Lancieri, M., Lacassin, R., Campos, J., Carrizo, D., Bejar-Pizarro, M., Armijo, R., Aranda, C., Valderasa-Bermejo, M. C., Ortega, I., Bondoux, F., Baize, S., Lon-Caen, H., Pave, A., Vilotte, J. P., Bevis, M., Brooks, B., Smalley, Jr. R., Parra, H., Baez, J. C., Blanco, M., Cimbara, S. and Kendrick, E. (2011). The 2010 M_w 8.8 Maule mega-thrust earthquake of central Chile, monitored by GPS. *Science* **332**, 1417-1421.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics* **1**, 80-83.

Received January 9, 2013; accepted March 13, 2013.

Javier E. Contreras-Reyes
Department of Statistics
University of Valparaíso
Gran Bretaña 1111, Playa Ancha, 2360102 Valparaíso, Chile
jecontrr@uc.cl