STATISTICAL CHARACTERIZATION OF SEISMIC PROCESS EXTREMES IN AN INTRAPLATE AREA – CASE STUDY: JOÃO CÂMARA – RN - BRAZIL.

SILVA, R. N. C., Universidade Federal do Acre - UFAC, Centro de Ciências Exatas e Tecnológicas - CCET, nonato@ufac.br
LUCIO, P. S., Universidade Federal do Rio Grande do Norte – Brasil e Centro de Geofísica de Évora – Portugal, pslucio@ccet.ufrn.br / pslucio@uevora.pt
do NASCIMENTO, A. F., Universidade Federal do Rio Grande do Norte, Departamento de Geofísica, UFRN, Brasil, aderson@geofisica.ufrn.br
MEDEIROS, W. E., Universidade Federal do Rio Grande do Norte, Departamento de Geofísica, UFRN, Brasil, walter@dfte.ufrn.br
FERREIRA, J. M., Universidade Federal do Rio Grande do Norte, Departamento de Geofísica, UFRN, Brasil, joaquim@dfte.ufrn.br

This paper was prepared for presentation during the 11th International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, August 24-28, 2009.

Abstract

The aim of this work is to make a brief discussion of methods to estimate the parameters of the Generalized Pareto distribution (GPD). In order to illustrate the methods some parameters were estimated for the GPD, considering a sequence of intraplate earthquakes which occurred in the city of João Câmara in the northeastern region of Brazil. This region was continuously monitored for two years (1987 and 1988). In order to estimate seismic hazard in this area we address the following techniques: Moments, Maximum Likelihood, Biased Probability Weighted Moments, Unbiased Probability Weighted Moments, Mean Power Density Divergence, Median, Pickands, Maximum Penalized Likelihood, Maximum Goodness-of-Fit and the Maximum Entropy, the focus of this manuscript. Based on the threshold of 1.5 m_b was estimated the seismic hazard for the city, and estimated the level of return to earthquakes of magnitudes 1.5, 2.0, 2.5, 3.0 m_b and a 5.2 m_b earthquake which occurred in November 1986.

Introduction

Earthquakes occurrences can cause impacts on society, therefore both the government and the researchers have tried to model this phenomenon so that they can be predicted with some certainty and efficiency, thus helping in preventive actions and in planning and implementation of public policies.

In many practical situations there is interest in modeling the tail of the distributions, as occurs, for example, in earthquake distribution. This modeling should be done through a sequence of earthquakes, where you must observe the distribution of the maximum of a sequence of random variables independent and identically distributed and assume that F_x is unknown, and look for families of approximate models $\left[F_x(X)\right]^n$ is to use the theory of extreme values, proposed by Fisher and Tippett (1928) or use an important theorem known as limit distributions above a threshold (POT), known as the theorem of Gnedenko-Pickands-Balkema-Haan (1941).

The purpose of this study is to model the seismic risk for the municipality João Câmara, NE do Brasil, the generalized Pareto distribution (GPD), through a threshold selected by the graph of mean residual life, then make predictions / forecasts based on the period and level of return of earthquakes above that threshold and finally estimate the return period of the earthquake in history occurred in the town 1986 which reached a magnitude of 5.2 m_b.

Methodology

Let $X_1,....,X_n$ be random variables independent and identically distributed, with distribution function F_X. Let also consider x_{F_x} the upper limit of the distribution of F_X, We call "excedances" those values X_i such that $X_i > u$ (a selected threshold). Denoting by N_u the number of excess of the threshold u, we define

$$N_u = \sum_{i=1}^{n} 1_{(X_i > u)}$$

where:

$$1_{(X_i>u)} = 1 \text{ if } X_i > u$$
$$1_{(X_i>u)} = 0\text{, otherwise.}$$

The excedances beyond a threshold *u*, denoted by $Y_1,....,Y_{nu}$ are the values $X_i - u \geq 0$. Figure 1 shows the observations $X_1,....,X_{12}$ and the excedances beyond the threshold *u*=4.

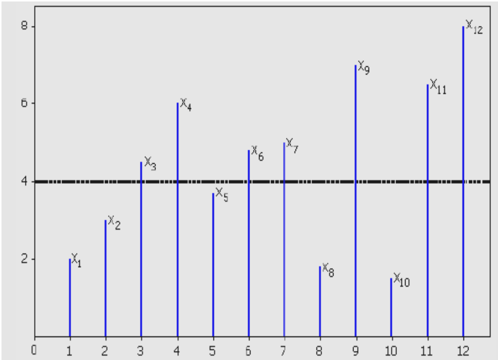

Figure 1: Illustration of the bar graph of the observations of a sequence of random variables $X_1,....,X_{12}$, which highlight the excesses above the threshold *u*=4.

Hence, given a threshold *u*, the distribution of values of x over u is given by:

$$P\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)},\ y > 0,$$

which represents the probability exceeds the value of x u by at most an amount y, where y=x-*u*. F is the distribution of a generalized extreme value, such that:

$$F(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \text{ for any } \mu, \sigma > 0$$

e $\xi \in \Re$. Then $n \ln F(x) \approx -\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}$, and that for high values of x should be an expansion to Taylor so that $\ln F(x) \approx -\{1-F(x)\}$, replacing and re-arranging for u, we have:

$$1 - F(u) \approx \frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}$$

and on a similar way for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1+\xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}.$$

Hence,

$$P\{X > u + y \mid X > u\} = \frac{1-F(u+y)}{1-F(u)} = \frac{\frac{1}{n}\left[1+\xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}}{\frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}} = \left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}$$

with $\tilde{\sigma} = \sigma + \xi(u-\mu)$. Thus, the distribution function of $(X-\mu)$ conditioned to $X > u$, is approximately:

$$H(y) = 1 - \left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}$$

defined in $\left\{y : y > 0\ e\left(1+\xi y/\tilde{\sigma}\right) > 0\right\}$, where $\tilde{\sigma} = \sigma + \xi(u-\mu)$.

Coles (2001) states that the family of distributions defined above is called the generalized Pareto family. The conditional distribution function is approximately the generalized Pareto distribution (GPD), representing the three distributions in one way, under the γ-parameterization: $W(x;\gamma) = 1-(1+\gamma x)^{-1/\gamma}$. Just as the generalized extreme value distributions (GEV) are the distributions limit to the maximum of the GPD are the type of parametric forms to limit distribution of excesses (The Balkema and de Haan Theorem). Generalized Pareto distributions are of the form of Exponential ($\gamma = 0$), Pareto type II ($\gamma > 0$) and Pareto or Beta ($\gamma < 0$).

The parameters of the generalized Pareto distribution to excesses that exceed thresholds (POT) are determined by those associated with the generalized extreme value distributions (GEV). In the limit of $F(x)$ when $\xi \to 0$ has the cumulative distribution of Gumbel:

$$F(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right],$$

and the distribution function of $(X-\mu)$, conditional to $X > u$, is approximately: $H(y) = 1 - \exp\left(-\frac{y}{\sigma}\right)$, com $y > 0$.

Figure 2 presents the graphs of the function of distribution of GPD to $\xi = -0{,}4$ (Pareto or Beta), $\xi \to 0$ (exponential) and $\xi = 0{,}4$ (Pareto type II), all with $\sigma = 2$. Note that as in GEV the parameter ξ determines the tails shape of the distribution.

Finally, the GPD and GEV distributions are related as follows:

$$G(x) = 1 + \ln(H(x)),\ \ln(H(x)) > -1.$$

This relationship explains why the densities of GPD have extreme tail asymptotically equivalent to a GEV.

Figures 3 and 4, illustrate this fact and shows the closeness of the GPD tails of some distributions with some GEV.

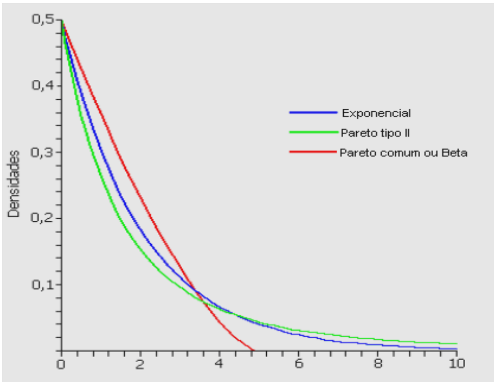

Figure 2: Illustration of probability density function of the three forms of the generalized Pareto distribution (GPD).

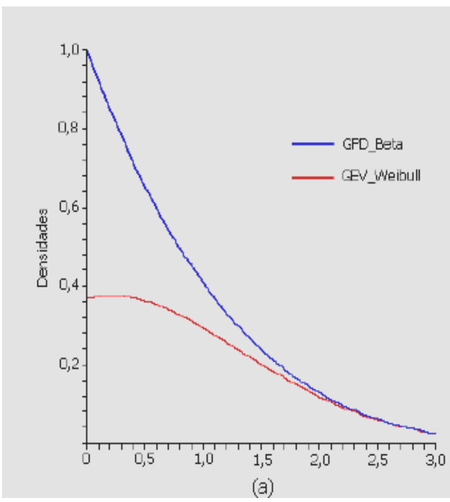

Figure 3: Densities of the GPD and GEV, common Pareto (Beta) and Weibull, both with $\xi = -0{,}2$.

Threshold Selection

The choice of threshold u in the face with some problems, because a value for u too "high" implies a small number of observations in the tail, which may result in increased variability of the estimators. However, a threshold is not high enough not satisfy the theoretical assumptions and estimates could result in distorted, so one idea is to track the extreme values as will be described

To determine the threshold used to the graphic analysis of the linearity n_u observations that exceed certain thresholds in the various samples. Thus the plot of mean residual life, used for visual determination of u is constructed as follows

$$:\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i - u)\right) : u < x_{\max}\right\},$$

here $x_1, x_2, \ldots, x_{n_u}$ consist of observations that exceed u and $x_{\max}$ is the higher of the observations. After the selection of the threshold the next step is the estimation of the parameter to be seen.

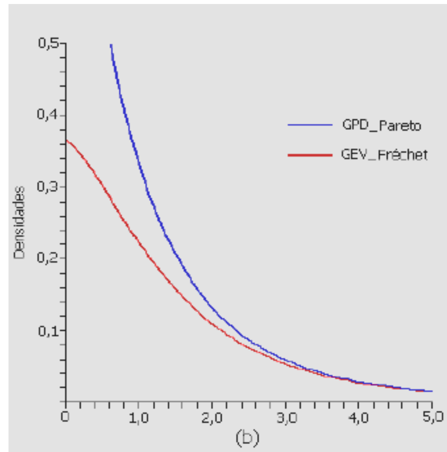

Figure 4: Densities of the GPD and GEV, Fréchet and Pareto type II, both with $\xi = 0{,}2$.

GPD Parameters Estimation

Several methods of estimating the parameters of the GPD have been proposed, which in recent years the method of maximum entropy (POME) has been used by several authors, generally Sing and Guo (1995), Oztekin (2004), where the POME when compared with other methods, obtained lower mean square error. So to estimate the parameters of the GPD by maximum entropy simply solve the following equations.

$$E\left\{\ln\left[1 - \frac{\xi(x-\mu)}{\sigma}\right]\right\} = -\xi$$

$$E\left\{\ln\left[\frac{1}{1 - \frac{\xi(x-\mu)}{\sigma}}\right]\right\} = -\frac{1}{1-\xi}$$

$$\operatorname{var}\left[\ln\left(1 - \frac{\xi(x-\mu)}{\sigma}\right)\right] = \xi^2$$

The solution of these equations must be done by numerical methods and simulations.

To check the accuracy of the estimator he was compared by mean square error with the following methods:

methods of moments (MOM), Pickands (Pickands), by the probability weighted moments: biases and non-bias (PWMB, PWMU), average density difference (MDPD), maximun goodnes of fit (MGF), median (MED), penalized maximum likelihood (MPLE) and the maximum likelihood (MLE), while there, depending on the value of not always in the terms of regularities are observed, however Brabson e Patutikof (2000), simulations using observed that $\xi \in (-0{,}5;0{,}5)$, then the maximum likelihood estimator of satisfies the conditions of regularity, so to find it simply solve the equations below by numerical methods, since the analytical solution is not possible.

$$\frac{\partial L}{\partial \xi} = \frac{\sum_{i=1}^{n} \ln\left(1 - \frac{\xi(x_i - \mu)}{\sigma}\right)}{\xi^2} + \frac{(1-\xi)\sum_{i=1}^{n} \frac{x_i - \mu}{\sigma\left(1 - \frac{\xi(x_i - \mu)}{\sigma}\right)}}{\xi} = 0,$$

$$\frac{\partial L}{\partial \xi} = \frac{n}{\sigma} - \frac{(1-\xi)\sum_{i=1}^{n} \frac{\xi(x_i - \mu)}{\sigma^2\left(1 - \frac{\xi(x_i - \mu)}{\sigma}\right)}}{\xi} = 0.$$

Discussions

Data analyses were done based on the R software (Language and Environment for Statistical Computing), version 2.8, through the package POT. The data used were from an earthquake sequence continuously monitored from 23/05/1987 until 07/07/1988 in the town of João Câmara, NE Brazil.

In Figure 5, we plot the mean residual life time in order to find the threshold of about 1.4m_b. The choice of the threshold using this procedure is not easy. However, in the same figure, we can see that some kind of linearity starts around 1.5m_b For this reason, we have chosen u = 1.4 m_b. In Figure 6 we see that the choice of the threshold follows the recommendations of Coles (2001). Since the threshold is very low, it will affect the asymptotic behavior.

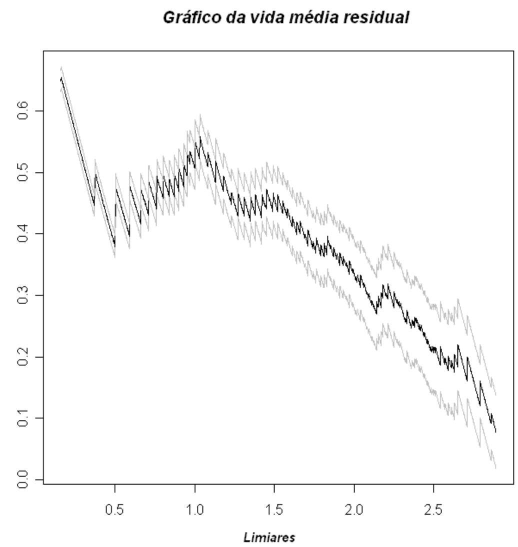

Figure 5: Graphical representation of the mean residual life time of the earthquakes – a tool for selecting the threshold of extreme values.

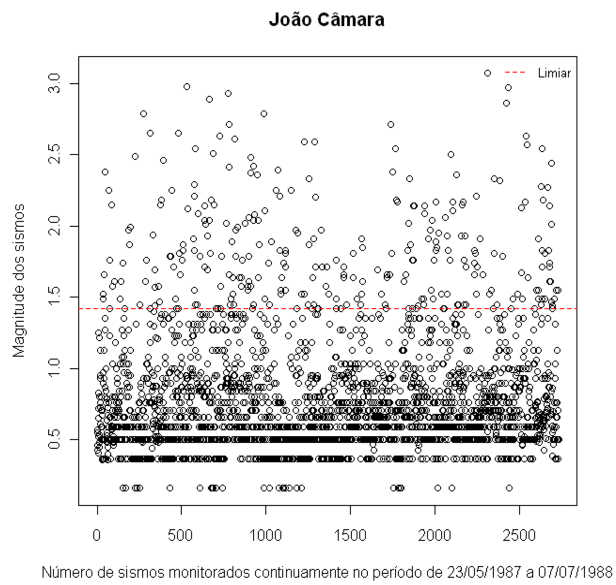

Figure 6: Graphical representation of the temporal dispersion of the earthquakes. The red line is the selected threshold.

After the selection of the threshold *u*, we will examine the parameters of the GPD, drawing greater attention to the shape parameter, because it defines the type of generalized Pareto distribution used to estimate the earthquake. In Table 1 we present not only the estimates for the shape parameter, but also for the scale of all proposed methods and the standard error associated to each one.

As the shape parameter defines the type of distribution we see in Table 1 that the proposed distribution to model the earthquake hazard is the Pareto or Beta, both for Maximum Entropy (POME) and Maximum Likelihood (MLE), since they have smallest standard errors.

Figures 7 and 8 show the fit diagnostics for GPD, considering the POME and MLE methods. We note that the graph of probabilities and quantiles the fittings are very good. It can also be observed from this diagnostics the goodness of the empirical density fitting and the information brought by the bottom right graph in Figure 7, related to the level and return period indicating that we expect a $1.5m_b$ earthquake at least once every ten days. However, the $5.2m_b$ earthquake is more unlikely to occur again, because it is expected to occur at least once every in about three hundred years by extrapolation, considering both the MLE and the POME.

Table 1: Estimated parameters of the generalized Pareto distribution, through the proposed methods of estimation and the standard error of the parameters of shape and scale.

Method	Estimate			Standard Error	
	$\hat{\xi}$	$\hat{\sigma}$	$\hat{u}$	$\hat{\xi}$	$\hat{\sigma}$
POME	-0,2998	0,4564	1,4340	0,0506	0,0455
MLE	-0,2892	0,5820	1,4340	0,0555	0,0466
PICKANDS	-0,4899	0,5496	1,8070	0,9124	0,8260
MOM	-0,2163	0,4427	1,8070	0,0864	0,0522
PWMB	-0,1737	0,4272	1,8070	0,1049	0,0554
PWMU	-0,1682	0,4252	1,8070	0,10455	0,0551
MDPD	-0,2766	0,4660	1,8070	0,3589	0,3245
MED	-0,2356	0,5127	1,8070	0,2583	,3015
MGF	-0,2163	0,4427	1,8070	0,0864	0,0522

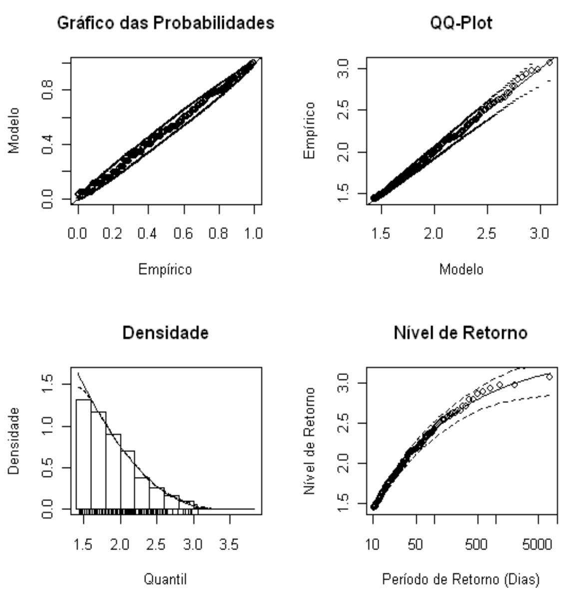

Figure 7: Earthquakes at João Câmara from GPD via the Maximum Entropy Methodology (POME).

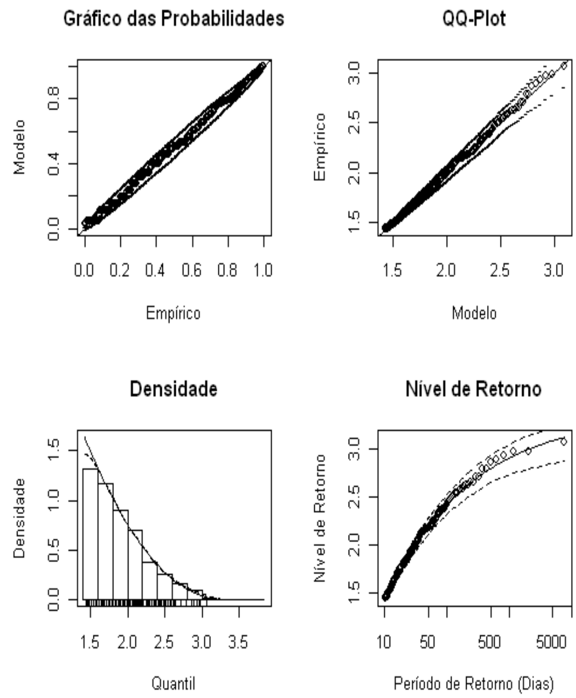

Figure 8: Earthquakes at João Câmara from GPD via the Maximum Likelihood Methodology (MLE).

Final Remarks

As the methodology for estimating the parameters of the Generalized Pareto Distribution (GPD) of the methods of maximum likelihood and maximum entropy (POME) were those that were more satisfactory.

The Generalized Pareto distribution (GPD), with $\xi = -0{,}2998$ by POME and $\xi = -0{,}2892$ by MLE, showing that the earthquakes in the city of João Câmara can be modeled in a satisfactory way by a GPD with negative shape parameters, so the tail should be modeled by a Pareto or Beta distribution.

We point out the fact that the return period for a 1.5 m_b earthquake is ten days. For an earthquake of magnitude $5.2m_b$ the return period of approximately 300 years.

Acknowledgements

The authors PSL, AFDN, WEM thank CNPq for their PQ grants.

Appendix: Observational Dataset

Our study area is located in the eastern part of the Potiguar basin, northeastern Brazil (Figure A1), which, within the present knowledge of historical and contemporary seismicity, is one of the most active areas in intraplate South America. Seismic activity in this area is known since 1808 and has been mainly characterized by earthquake swarms that can last four years and by events up to 5.2 body-wave magnitude (m_b). The earthquakes have occurred in the upper crust, which has not undergone any major tectonic event since the Pangea breakup in the Cretaceous (Bezerra et al., 2007).

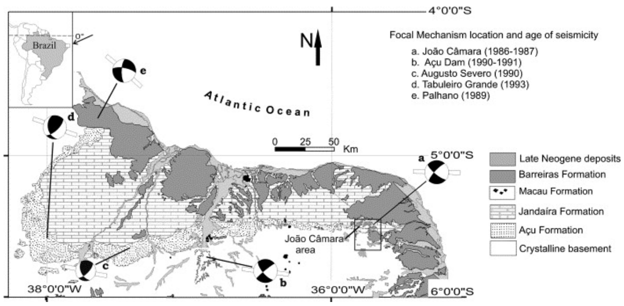

Figure A 1: Simplified geological map of the Potiguar basin and area studied in detail. Focal mechanisms indicate main epicentral areas. The white bars indicate direction of P (SHmax) axis. Inset: South American continent.

Figure A2 shows the epicenters used in this study. We identified two seismogenic faults in the João Câmara area. These epicenters occur both in the crystalline basement and in the Potiguar basin. Their related hypocenters, however, are located in the crystalline basement at a depth from ~ 1 to ~ 9 km (Bezerra et al., 2007).

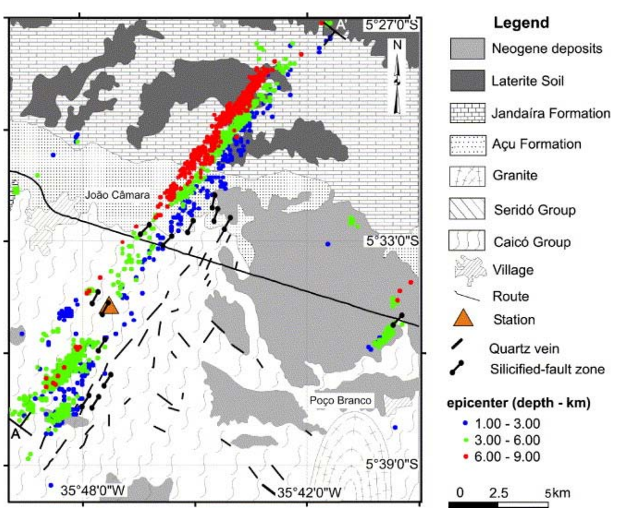

Figure A 2: Map of João Câmara (JC) epicentral area depicting geology and seismicity. Epicenters are from the selected telemetric-network dataset. Station JCAZ (triangle) is used as a reference. The orientation of patterns in the legend indicates foliation trend. Station JC01 quoted in text is ~ 7 km west of JCAZ and outside the map (from Bezerra et al., 2007)

References

[1] Brabson, B.B. and Palutikof, J.P. (2000). "Tests of the Generalized Pareto Distribution for predicting extreme wind speeds." Journal of Applied Meteorology, p.39.
[2] Coles S. (2001). Introduction to Statistical Modelling of Extreme Values, Springer.
[3] Fisher, R.A., Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society, p.180-190.
[4] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une serie aléatorie. Annales des Mathemátiques, p.423-453.
[5] Oztekin, T. (2004). Comparison of Parameter Estimation Methods for the Three- Parameter Generalized Pareto Distribution, University of Gaziosmanpaßa, Faculty of Agriculture, Agriculture Technology, Tokat – TURKEY.
[6] Patutikof, J.P., Brabson, B.B., Lister, D.H. and Adcock, S.T. (1999). "A review of methods to calculate extreme wind speeds." Meteorol. Appl. 6, p.119-132.
[7] Pickands III, J. (1975). Statistical inference using extreme order statistics. Ann. Statist. 3, p.119-131.
[8] Shannon C. E. (1998). "Communication in the presence of noise", Proc. Institute of Radio Engineers, vol. 37, no.1, pp. 10-21, Jan. 1949. Reprint as classic paper in: Proc. IEEE, Vol. 86, No. 2.
[9] Singh, V. P. and Rajagopal, A. K. (1986). A new method of parameter estimation for hydrologie frequency analysis. Hydrol. Sci. Technol. , p. 33-40.
[10] Singh, V. P. and H. Guo (1995). Parameter estimation for 3-parameter generalized pareto distribution by the principle of maximum entropy (POME), Department of Civil Engineering, Louisiana State University, Baton Rouge, Louisiana, USA.
[11] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
[12] Bezerra, F H R; Takeya, M K; Sousa, M O L; do Nascimento, A. F. Coseismic Reactivation of the Samambaia Fault in the Potiguar Basin, Brazil. Tectonophysics (Amsterdam), v. 430, p. 27-39, 2007