

Smoothing seismicity techniques applied to seismic source characterization and probabilistic seismic hazard analysis in Brazil

Marlon Pirchiner^{1,2} and Marcelo Assumpção¹

Corresponding author: Marlon Pirchiner, Department of Geophysics, IAG, USP, São Paulo, Brazil. (marlon.pirchiner@iag.usp.br)

¹Department of Geophysics, Institute of Astronomy, Geophysics and Atmospherical Sciences (IAG), University of São Paulo (USP), São Paulo, Brazil.

²Applied Math School (EMAp), Getúlio Vargas Foundation (FGV), Rio de Janeiro, Brazil.

Intraplate seismicity is a hard problem to solve in tectonics. In Brazil, as other intraplate tectonic settings, mainly during the lack of observed seismicity, the seismic hazard assessment are also difficult and the methods applied are often close correlated to the geometries defined by the experts with valuable criteria.

Three smoothed seismicity techniques are reviewed and their application on construct models of Brazilian seismicity and its applications to long-term seismic hazard assessment in Brazil are also discussed. All these smoothing methods are based on a Kernel Density Estimation (KDE). The main difference between them is the bandwidth selection process.

The method proposed by *Frankel* [1995] uses a *fixed* bandwidth in space. *Woo* [1996] suggests a bandwidth as function of the *magnitude*. *Helmstetter and Werner* [2012] propose a *locally-adaptive* kernel bandwidths in space and time, chosen for each event as background (time-independent) model on their forecast experiments.

All these methods compute the seismicity rate surface in a binned space, time and magnitude, which able us to use this occurrence rate to define a grid of seismogenic point sources as the input source model to the hazard computation.

In addition, we used the information score and the error diagrams [*Kagan*, 2009] to evaluate the models, optimize the smoothing distance on the *Frankel* [1995] method, and finally give a weighted source models logic-tree to hazard computation.

All the hazard computations was performed on Openquake (OQ) [*Pagani et al.*, 2014] open source software from Global Earthquake Model (GEM) Foun-

dation. No focal mechanism information has been included or considered on these smoothed seismicity models. All hazard computations has been done using a single ground motion prediction equation (GMPE). A more detailed study on GMPEs selection should be developed further and it is beyond the scope of this work.

The results shown that all methods clearly most of the well known seismogenic regions and could be used in the Brazilian seismicity modeling. All methods improve the spatial resolution of the GSHAP hazard model, the only previously available for the region. These results should be considered on future reviews of Brazilian design building codes and procedures.

1. Introduction

1.1. Intraplate seismicity and previous studies in Brazil

Brazil occupies a large stable continental region with seismicity rate lower than Central and Eastern North America, Australia, India or Northern Europe [Giardini *et al.*, 1999; Johnston, 1996]. Spanning most of the central region of the South American plate, Brazil has earthquakes above magnitude 6 *mb* with a return period of 50 years or so. Historical records are very short with almost no data before the XX century. Historical evidence of magnitude a 7 has been reported in the Amazon [Berrocal *et al.*, 1984; Veloso, 2014]. However, as in other intraplate settings, seismic hazard is not negligible due to the low attenuation in the cold and thick lithosphere (magnitude 5 *mb* can be felt up to 500 km away [Assumpção *et al.*, 2014]) and special seismic hazard studies have been required for critical facilities such as nuclear power plants [Almeida *et al.*, 2013].

Few studies related to seismic hazard, such as Berrocal *et al.* [1996]; Santos *et al.* [2010] for SE and NE Brazil, have been carried out so far and the first Brazilian seismic building code [ABNT, 2006; Santos and Lima, 2008] was based on the GSHAP [Giardini *et al.*, 1999] results for South America [Tanner and Shepherd, 1997; Shedlock and Tanner, 1999]. Brazilian seismicity is concentrated in several seismic zones [Berrocal *et al.*, 1984; Assumpção *et al.*, 2014]. Except for a peak of activity in Northeast Brazil, both the GSHAP model and the ABNT [2006] zoning (even when the Brazilian recommended levels are in line with those of other nations with higher occurrence of earthquakes [Hampshire De C. Santos *et al.*, 2013]) have not taken into proper account some of the other highly active seismic zones, such as the Porto dos Gaúchos, Mato Grosso state, with recurrent activity since the occurrence of the largest recorded Brazilian event with magnitude 6.2 *mb* in 1955 [Barros *et al.*, 2009], or the Central Amazon region where concentration of events in the last 50 years is consistent with the location of the possible

historical earthquake of magnitude 7 [Veloso, 2014]. For this reason, it is important to review and update the hazard estimates for Brazil, especially since the number of Brazilian earthquake records has been increased in the last three decades.

Despite attempts to explain the origin of some of the seismic zones, such as stress concentrations from lithospheric thinning [Assumpção *et al.*, 2004, 2014; *de Azevedo et al.*, 2015] or flexural effects [Zoback and Richardson, 1996; Assumpção and Sacek, 2013], most of them have ill-defined boundaries. Attempts to produce seismic hazard maps [Dourado, 2014] based on seismogenic zones are heavily dependent on subjective expert opinion, even more so in low seismic areas such as Brazil. Like a contribution to some new seismic hazard model for Brazil, here we briefly review recent methods of smoothing seismicity rates with application to Brazilian earthquakes.

1.2. Smoothed seismicity models

Although the standard *Cornell* [1968]-*McGuire* [1976] methodology often used to assess seismic hazard makes use of expert opinions to predefined seismogenic zones, one goal of this paper is to review a few smoothed seismicity methods [Riznichenko, 1959] based on KDE which results in a grid or a smoothed seismicity surface without define any seismic zone polygon (also known as *zoneless* approaches).

The (smoothed) seismicity KDE techniques allow us to define another grid of seismic point sources given an earthquake catalog. The seismicity rate (*a*-value) associated to each source is defined by the smoothed seismicity methods.

To perform the KDE, both its shape and bandwidth needs to be defined. Most kernels are Gaussian, power-laws or one of their variants like Fisher kernels. The main distinction between the methods described here is the bandwidth selection process.

Frankel [1995] smoothing method applies a Gaussian *Nadaraya* [1964]-*Watson* [1964] kernel smooth technique with a *fixed smoothing-distance* bandwidth to estimate the seismicity rate by smoothing the 2D histogram from earthquakes counting. The actual version of the algorithm is already implemented on the Hazard Modeller's Toolkit (HMTK) [*Weatherill et al.*, 2012; *Weatherill*, 2014] with a simple variant of Frankel's method discussed by *Zechar and Jordan* [2010] which adds the contribution of each earthquake kernel on given cell.

Woo [1996] proposed a *magnitude-dependent* kernel bandwidth based on the regression for nearest neighbor mean distance between earthquakes binned by magnitude.

Helmstetter and Werner [2012] proposed a *locally-adaptive* bandwidth on both space and time dimensions for the background (time-independent) seismicity rate on their long-term forecasts experiments.

Using the long-term background seismic rate computed for forecast and assuming its time-independence, the same Probabilistic Seismic Hazard Analysis (PSHA) assumption, it is possible, preserving their purposes differences [*Marzocchi and Zechar*, 2011], to use these background seismicity rate values on seismic sources characterization [*Weatherill and Pagani*, 2014].

1.3. Information scores and error diagrams (EDs)

From these three smoothing methods, only *Helmstetter and Werner* [2012] since its formalization cross-validate their model parameters. All the other methods are concept to use the whole catalog and some fixed parameter (e.g. Frankel's smoothing distance).

To compare the performance of these three models, the information scores and error diagrams (EDs) [*Molchan and Kagan*, 1992; *Kagan*, 2007, 2009] were computed as al-

ternative to the tests suggested to the Regional Earthquake Likelihood Models (RELM) forecasts experiments [*Schorlemmer et al.*, 2007].

On the evaluation phase all models learned from earthquakes into 1970-2004 observation time and were tested against the target realizations occurred between 2004 and 2014-01-01.

Scores from some Frankel-based models (from different smoothing distances) was computed and the best scores gave us an reasonable choice for this parameter.

HYBRID DECISION APPROACH BIRD/JORDAN[2013?]

1.4. Hazard computation

To perform the hazard computation, the Openquake (OQ) suite [*Pagani et al.*, 2014] was chosen. The modeling tool was the HMTK also open available by Global Earthquake Model (GEM) scientific board. The Frankel method was already implemented on the toolkit and two others was rewritten in the context of this paper based on sources given by the authors and their papers.

The other seismic source parameters, like b -values, minimum and maximum magnitudes for the frequency-magnitude distributions, strike, dip and rate probabilities to the rupture planes on each source, its top and bottom limits on the subsurface, magnitude-area scale relationships and also the depth distributions must be assigned in other ways to proceed the hazard computation and take advantages of the latest generation GMPEs.

In this work, to keep our focus on compute the hazard given a grid of punctual seismic sources computed from smoothed seismicity models, we compute the hazard probabilities using a single GMPE [EXPOSE WHICH ONE] for a stable continental crust tectonic environment.

The earthquake catalog data comes from the Brazilian seismological research authorities [BSB, 2014].

1.5. Specific purposes

The specific purposes of this paper are: (i) to present the *BSB* [2014] earthquake catalog, (ii) review three distinct smoothing (zoneless) methods to spatially model the seismic rate, (iii) evaluate each model by its information scores and error diagrams, (iv) add and share into the HMTK two more smoothing methods to be used by the hazard modeling community, (v) use the GEM-Openquake tools to perform hazard calculation and (vi) update the hazard studies at national scale by proposing a simple hazard map based on these smoothed seismicity models.

Is not on our purposes at this time: (i) make an exhaustive evaluation of completeness magnitude in dependence of space and time, (ii) go deep on the GMPE selection methodology and neither (iii) create the definitive Brazilian hazard map which we handle as a continuous process.

To aim this, the Brazilian seismicity data is presented and some preprocessing steps are described. Next the theory behind this smoothed seismicity methodologies are presented as the results are discussed in terms of its scores. Then the hazard is computed from these source models in a logic-tree and presented to allow us take some lessons, try to conclude or even suggest something.

2. Brazilian seismicity data

Data on intraplate seismicity in Brazil is currently the Brazilian Seismic Bulletin maintained mainly by the Seismological Center of the University of São Paulo. This catalog is a joint compilation from different sources [BSB, 2014].

2.1. Hypocenters

For events up to 1981, the catalog corresponds to *Berrocal et al.* [1984] which is a comprehensive compilation of historic and instrumental data with participation from Universidade de São Paulo (USP), Universidade de Brasilia (UnB), Universidade Federal do Rio Grande do Norte (UFRN), Observatório Nacional (ON) and Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT). From 1982 to 1995 the information comes from the annual bulletins published by Revista Brasileira de Geofísica. Since 1995, the Bulletin is maintained mainly by USP under the same cooperation network in addition to Universidade Estadual Paulista (UNESP). Earlier versions of this catalog were used to compose the well-known *CERESIS* [1985, 1995] catalogs.

Today the catalog is distributed in two ways. One, which could be called *raw*, contains all compiled information, such as: events outside the country which were felt in Brazil, errors in previous versions and also non-seismic events (known quarry blasts, sonic boom, etc.). The other way could be called *clean*, where all non-seismic and low quality events were removed. This last one was used in this paper. There are 1941 earthquakes recorded from 1724 to the end of 2013.

This catalog only covers crustal events. Andean subduction earthquakes deeper than 50 km in the Brazil-Peru border were discarded as well as earthquakes from the Mid-Atlantic ridge. Epicenters into the catalog are shown in figure 1.

2.2. Magnitudes

Most magnitudes were computed as *mb* type or the equivalent *mR* [Assumpção, 1983] regional magnitude. For historical events magnitudes were estimated from macroseismic data (felt area in km^2 , A_f and maximum intensity, I_0) [*Berrocal et al.*, 1984].

Since most GMPEs are based on M_W moment scale magnitude, in this present study, we converted all mb magnitudes to M_W values following the guidelines on table 1. REFERENCIA PARA STEPHANE/ASSUMPCAO WORK IN PROGRESS. (Johnston)

Other proxies for intraplate regions like *Scordilis* [2006] was discarded by its magnitude range definition, over than 6, whilst most magnitudes on *BSB* [2014] catalog had values lower than this.

2.3. Catalog checking

Figure 2 shows the distribution of depths, day of the week and origin hour following *Julia et al.* [2012]. Most events have no depth determination and the well determined depths (from pP phases or local networks) are mostly less than 10km.

Figures 2b and 2c show an almost uniform distribution of day of the week, and a slightly trend to record earthquakes during the night hours. This could suggest some influence of daily noise level on the station records.

The annual number of earthquakes since 1900 (figure 3) shows a steady increase since 1960 and two peaks in 1986 and 1998 derived from two large seismic sequences: João Câmara [*Takeya et al.*, 1989] and Porto dos Gaúchos [*Barros et al.*, 2009]. The increasing number of records after 1960 is due to the operation of global, regional and local stations.

2.4. Declustering procedures

To preserve the Poissonian assumption about the independence of the events, several declustering algorithms were tried and the results as shown in figure 4a. The window-method [*Gardner and Knopoff*, 1974] was tested with three different windows: *Gardner and Knopoff* [1974], *Uhrhammer* [1986] and Grüenthal [*van Stiphout et al.*, 2012]. In addition the algorithm (AFTERAN) proposed by *Musson* [1999] was also tested using Grüenthal distance window.

In the case of Brazilian catalog the methods did not present large differences at the final number of earthquakes (fig. 4a) and the Gardner-Knopoff/Uhrhammer combination was chosen to preserve the maximum number of events.

The map of figure 4b locates all clusters in the catalog. It is an *a-priori* evidence of regions in Brazil where earthquake sequences tend to occur more frequently.

2.5. Catalog completeness

The completeness of Brazilian seismicity data has evident time and space dependence. Brazil just recently installed its permanent seismographic network [Pirchiner *et al.*, 2011]. The number and quality of stations increased over time. The population density, previously concentrated along the coast increased also on other regions changing the historical network coverage. There are many methodologies to assessment of the temporal and spatial completeness [Stepp, 1972; Mignan and Wöessner, 2012; Ogata and Katsura, 1993; Wiemer and Wyss, 2000; Cao and Gao, 2002; Stucchi *et al.*, 2004; Woessner and Wiemer, 2005; Mignan *et al.*, 2011; Mignan, 2012; Vorobieva *et al.*, 2013; Mignan *et al.*, 2013; Nasir *et al.*, 2013; Mignan and Chouliaras, 2014], or explaining how to compute the real Probability of Detecting an Earthquake (PDE) from the waveform records [Schorlemmer and Woessner, 2008], but the time and efforts to proceed this exhaustive study of the completeness magnitude, despite its relevance is beyond the scope of this work. The waveforms from earliest years was not in digital form. The histories of most regional (temporary) stations are not easy available. In addition, the low number of earthquakes in the catalog also difficult some high resolution statistical methods. All of this hinders the direct application of most that methodologies. For this reason, the spatial completeness was not considered and just simple temporal M_c

parameters for the whole country was defined by an easy cumulative rate criteria using historical information about the records evolution and the *Stepp* [1972] plot (fig. 5a).

Figure 5b shows the magnitude by time distribution in a scatter plot as well the cumulative number of events which is important to perform the temporal course of earthquake frequency (TCEF) [*Nasir et al.*, 2013]. The linear behavior of cumulative rate on the recent years, gives an idea about the constant record trend. For magnitudes greater than 3, this threshold occurs about 1970. For the magnitudes greater than 4 and 5, these limits could be extrapolated to 1950 and 1940, respective and roughly. Figure 5b also show the completeness tables from the *Stepp* [1972] test using $\Delta_{mag} = 0.5$ and $\Delta_{time} = 2$ as magnitude and time bins, and from the literature [*Assumpção et al.*, 2014]. Tables 2 and 3 show each one of these completeness tables.

2.6. Frequency-magnitude distribution

Figure 6 presents the catalog Magnitude Frequency Distribution (MFD). For the minimum magnitude threshold was 2.9. The blue exes and red crosses are the incremental and cumulative recurrence distributions binned by 0.2 magnitude unity. Stepp completeness (table 2) and *BSB* [2014] declustered catalogue was used.

Simple *Gutenberg and Richter* [1944] incremental MFDs,

$$\log N(m|a, b) = a - b \times m, \quad (1)$$

were fitted using a fixed b -value ($b=1$), showed as the blue continuous line (fig. 6), by the maximum-likelihood method (magenta, $b = 0.91 \pm 0.07$), *Weichert* [1980] (green, $b = 0.88 \pm 0.04$) and *Kijko and Smit* [2012] methods (orange, $b = 0.76 \pm 0.04$) are shown as continuous lines. The respective a -values was 3.7 , 3.8 ± 0.2 , 3.5 ± 0.2 , 3.2 ± 0.1 .

From these distributions over the incremental catalog data, its possible to note just few records with magnitudes greater than 5, even considering different magnitude com-

plteness on time, due probably the small observation time (few than 100 years) on Brazil stable crustal set.

Other important feature extracted from figure 6, is the low values of the b -values relative to the unity. Incremental $b=1$ MFD underestimate the earthquake occurrences, specially under great magnitudes. From now on, we will assume the b -value of 0.9 as better representative for the whole region. On the future more investigations on the spatial b -value variations should be undertaken.

3. Smoothing methods

The smoothing methods considered in this article are now briefly explained and their results are discussed.

3.1. Frankel: fixed bandwidth

Frankel [1995] smoothing seismicity proposal consisted originally in using a “correlation distance” d_F as the *fixed* kernel bandwidth and applying the Nadaraya-Watson [Nadaraya, 1964; Watson, 1964] estimator to smooth a 2D seismicity histogram using a Gaussian kernel:

$$\tilde{n}_j = \frac{\sum_i n_i e^{-\left(\frac{d_{ij}}{d_F}\right)^2}}{\sum_i e^{-\left(\frac{d_{ij}}{d_F}\right)^2}}, \quad (2)$$

where \tilde{n}_j is the smoothed seismicity (number of earthquakes with magnitude m above the minimum magnitude M_d in the catalog) on cell j . n_i is the earthquake counting in each other cell i and d_{ij} is distance between grid cells i and j .

3.2. Woo: magnitude-dependent bandwidth

Woo [1996] suggested to evaluate the contribution of each earthquake i , located at \mathbf{r}_i , to the seismicity rate R (number of earthquakes per year and unity area) on the cell

centered in \mathbf{r} , depending on the magnitude m :

$$R(\mathbf{r}, m) = \sum_{i=1}^N \frac{K(\mathbf{r} - \mathbf{r}_i, m)}{T(\mathbf{r}_i)}, \quad (3)$$

where N is the number of earthquakes i in the magnitude bin $m \pm dm$, and $T(\mathbf{r}_i)$ is the completeness time of magnitude m observed on \mathbf{r}_i .

Any kernel could be applied on that definition. I used used the *Kagan and Knopoff* [1980] kernel for infinite spatial domain as described by *Woo* [1996]:

$$K_{KK}(\mathbf{r}, m | a_W) = \frac{a_W - 1}{\pi h(m)^2} \left[1 + \frac{\mathbf{r}^2}{h(m)^2} \right]^{-a_W}, \quad (4)$$

where a_W is fractal dimension factor, generally about 1.5 and 2 [*Vere-Jones*, 1992].

VERIFICAR SE É O CASO DE APRESENTAR OUTRAS FORMAS DE KERNEL.
DAR PREFERENCIAS AOS REALMENTE UTILIZADOS.

To compute the magnitude-dependent bandwidth function $h(m)$, *Woo* [1996] suggested the follow relation

$$h(m | a_0, a_1) = a_0 e^{a_1 m}, \quad (5)$$

where a_0 and a_1 are computed by the regression of the mean nearest distance h from each magnitude bin $m \pm dm$, as illustrated on figure 7.

Just to compare, in figure 7 it is shown the functions computed by *Beauval* [2003] for Norway and Spain. The angular coefficient (a_1) is in the same range of the others but the general mean distance between events into the same magnitude bin is higher than other countries. It is in some way expected. Brazil is quite large than mentioned countries.

3.3. Helmstetter and Werner: locally-adaptive space and time bandwidths

Even work on a forecast perspective, the background seismicity rate for a long-term time-independent forecast could be used to characterize a diffused seismicity under the common assumption that this background seismic rate is invariant over time.

Helmstetter and Werner [2012] proposes a seismicity model space and time dependent using kernels for space and time independently:

$$R(\mathbf{r}, t) = \sum_{i=1}^N \frac{1}{h_i d_i^2} K_t \left(\frac{t - t_i}{h_i} \right) K_r \left(\frac{\|\mathbf{r} - \mathbf{r}_i\|}{d_i} \right), \quad (6)$$

where $R(\mathbf{r}, t)$ is the seismic rate located \mathbf{r} distant from the earthquake occurred on the instant t , K_t is the time domain kernel function, where t_i is the time location of earthquake i and h_i is the temporal bandwidth for earthquake i , K_r is the space domain kernel function, where \mathbf{r}_i is the spatial location of earthquake i and d_i is the spatial bandwidth for earthquake i .

As they are interested in forecast, just past time $t_i < t$ need to be considered. Also the observation completeness is taken in account by a set of weights w in this follow way:

$$R(\mathbf{r}, t) = R_{min} + \sum_{t_i < t} \frac{2 w(\mathbf{r}_i, t_i)}{h_i d_i^2} K_t \left(\frac{t - t_i}{h_i} \right) K_r \left(\frac{\|\mathbf{r} - \mathbf{r}_i\|}{d_i} \right), \quad (7)$$

where R_{min} is the minimum seismic rate, positive, allowing earthquakes to occur where their never occurred yet.

The weights $w(\mathbf{r}_i, t_i)$, computed for each earthquake i , are the Gutenberg-Richter a -value projection. This is the expression used to compute it:

$$w(\mathbf{r}, t) = 10^{b(\mathbf{r}, t)[M_c(\mathbf{r}, t) - M_d]}, \quad (8)$$

where w is the weight factor computed on location \mathbf{r} on instant t , $b(\mathbf{r}, t)$ is the space and time dependent b-value, $M_c(\mathbf{r}, t)$ is the completeness magnitude on location \mathbf{r} and t , M_d is the minimum magnitude value on the catalogue.

These weights increase the seismicity contribution from earthquakes occurred where and when the completeness magnitude M_c was greater than minimum catalog magnitude M_d . They also could easily take into account the space and time completeness and b -value fluctuations.

3.3.1. Local bandwidth computation

The method implemented to compute the space and time bandwidths for each earthquake was the Coupled-Nearest-Neighbor (CNN) [Helmstetter and Werner, 2012] expressed as:

$$h_i, d_i = \arg \min_{\substack{h_i \geq h_k \\ d_i \geq d_k}} [s(h_i, d_i | k_{cnn}, a_{cnn}) := h_i + a_{cnn}d_i], \quad (9)$$

where k_{cnn} is the k^{th} nearest neighbor, a_{cnn} is a space-time coupling factor, d_k is the $\max \{d_j\}, j = 1, \dots, k_{cnn}$ and h_k is the $\max \{h_j\}, j = 1, \dots, k_{cnn}$.

The bandwidths are defined locally by this simple optimization process. It could be small on high earthquake density regions and higher where earthquakes are rarely or regions with information lack.

3.3.2. Stationary seismic rate

After compute the model parameters it is completely defined, and the time-independent or stationary seismic rate \bar{R} on each location \mathbf{r}_0 can be computed on the way proposed by Helmstetter and Werner [2012] taking the median value for some location \mathbf{r}_0 over all considered time window:

$$\bar{R}(\mathbf{r}_0) = \text{Median}[R(\mathbf{r}_0, t)]. \quad (10)$$

The median should avoid the seismicity rate fluctuations derived by fore and aftershocks and consequently the decluster procedures. This is one of the most important achievement of this methodology.

3.3.3. Maximum likelihood optimization

The model is completely defined by k_{cnn} , a_{cnn} and R_{min} . To optimize these parameters the catalog is divided in two parts: one for *learning* about the parameters and other to test them. The *testing* catalog will able us to check the model prediction performance

derived from its chosen parameters. The best model parameters will maximize the prediction model capacity likely the target catalog.

If the earthquake occurrence could be modeled by a Poisson process with rate N_p , then the probability to observe exactly n events on the considered time frame is

$$p(N_p, n) = \frac{N_p^n e^{-N_p}}{n!}. \quad (11)$$

Then, over all cells, the log-likelihood, to be maximized, between model prediction (from the learning catalog) and the observed earthquakes (on the testing catalog) is written as

$$L = \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} \log p [N_p(i_x, i_y), n(i_x, i_y)] \quad (12)$$

where $N_p(i_x, i_y)$ is the predicted seismic rate on cell (i_x, i_y) and $n(i_x, i_y)$ is the number of target earthquakes observed on cell (i_x, i_y) .

The model parameters R_{min} , a_{cnn} and k_{cnn} should be optimized by the maximization of log-likelihood L .

4. Results

4.1. Simulations

4.1.1. Simulation 1

5. Discussion

Appendix A: Appendix Title

Acknowledgments. The author thanks specially to all IAG-USP seismological team.

It is also grateful to Stéphane Drouet and Vincent Guigues for all helpful discussions.

References

- ABNT, A. (2006), NBR 15421: Design of seismic resistant structures - procedures.
- Almeida, A., M. Assumpção, J. Berrocal, J. Bommer, S. Drouet, L. Ferrari, C. Prates, C. Riccomini, and J. Riera (2013), Developing a logic-tree for updating the probabilistic seismic hazard assessment for the Angra dos Reis nuclear power plant in Brazil., in *22nd International Conference on Structural Mechanics in Reactor Technology (SMiRT22)*, San Francisco.
- Assumpção, M. (1983), A regional magnitude scale for brazil, *Bulletin of the Seismological Society of America*, 73(1), 237–246.
- Assumpção, M., and V. Sacek (2013), Intra-plate seismicity and flexural stresses in central Brazil, *Geophysical Research Letters*, 40(3), 487–491, doi:10.1002/grl.50142.
- Assumpção, M., M. Schimmel, C. Escalante, J. Roberto Barbosa, M. Rocha, and L. V. Barros (2004), Intraplate seismicity in SE brazil: stress concentration in lithospheric thin spots, *Geophysical Journal International*, 159(1), 390–399, doi:10.1111/j.1365-246X.2004.02357.x.
- Assumpção, M., J. Ferreira, L. Barros, H. Bezerra, G. S. França, J. R. Barbosa, E. Menezes, L. C. Ribotta, M. Pirchiner, A. do Nascimento, and J. C. Dourado (2014), Intraplate seismicity in Brazil, in *Intraplate Earthquakes*, pp. 50–71, Cambridge University Press.
- Barros, L. V., M. Assumpção, R. Quintero, and D. Caixeta (2009), The intraplate porto dos gaúchos seismic zone in the amazon craton — brazil, *Tectonophysics*, 469(1-4), 37–47, doi:10.1016/j.tecto.2009.01.006.

Beauval, C. (2003), Analyse des incertitudes dans une estimation probabiliste de l'aléa sismique, exemple de la france, Ph.D. thesis, Université Joseph-Fourier-Grenoble I.

Berrocal, J., U. de São Paulo. Instituto Astronômico e Geofísico, and B. C. N. de Energia Nuclear (1984), *Sismicidade do Brasil*, Instituto Astronômico e Geofísico, Universidade de São Paulo.

Berrocal, J., C. Fernandes, A. Bassini, and J. R. Barbosa (1996), Earthquake hazard assessment in southeastern brazil, *Geofísica Internacional*, 35(3), 257–272.

BSB (2014), Boletim Sísmico Brasileiro (BSB) version 2014.11. Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, IAG-USP.

Cao, A., and S. S. Gao (2002), Temporal variation of seismic b-values beneath northeastern japan island arc, *Geophysical Research Letters*, 29(9), 48–1–48–3, doi: 10.1029/2001GL013775.

CERESIS (1995), Catalogue for south america and the caribbean prepared in the framework of gshap, *Tech. rep.*, Centro Regional de Sismología para América del Sur.

CERESIS, S. T. A., Bonny L. Askew (1985), Catalog of earthquakes for south america, *Tech. rep.*, CERESIS (Centro Regional de Sismología para América del Sur). Earthquake mitigation program in the Andean Region (SISRA project).

Cornell, C. A. (1968), Engineering seismic risk analysis, *Bulletin of the Seismological Society of America*, 58, 1583–1606.

de Azevedo, P. A., M. P. Rocha, J. E. P. Soares, and R. A. Fuck (2015), Thin lithosphere between the Amazonian and São Francisco cratons, in central Brazil, revealed by seismic P-wave tomography, *Geophysical Journal International*, 201(1), 61–69, doi: 10.1093/gji/ggv003.

Dourado, J. (2014), Mapa de ameaça sísmica do brasil, in *Congresso Brasileiro de Geologia*.

Frankel, A. (1995), Mapping seismic hazard in the central and eastern united states,

Seismological Research Letters, 66(4), 8–21.

Gardner, J. K., and L. Knopoff (1974), Is the sequence of earthquakes in southern California, with aftershocks removed, poissonian?, *Bulletin of the Seismological Society of America*, 64(5), 1363 – 1367.

Giardini, D., G. Grünthal, K. M. Shedlock, and P. Zhang (1999), The GSHAP global seismic hazard map, *Annali di Geofisica*, 42(6), 1225–1230.

Julia, L., S. Wiemer, and M. Wyss (2012), Catalog artifacts and quality control, *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa-93722864.

Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, 34, 185 – 188.

Hampshire De C. Santos, S., L. Zanaica, C. Bucur, and S. De Souza Lima (2013), Comparative Study of Codes for Seismic Design of Structures, *Mathematical Modelling in Civil Engineering*, 9(1), 1–12, doi:10.2478/mmce-2013-0001.

Helmstetter, A., and M. J. Werner (2012), Adaptive spatiotemporal smoothing of seismicity for long-term earthquake forecasts in California, *Bulletin of the Seismological Society of America*, 102(6), 2518–2529.

Johnston, A. C. (1996), Seismic moment assessment of earthquakes in stable continental regions — i. instrumental seismicity, *Geophysical Journal International*, 124(2), 381–414, doi:10.1111/j.1365-246X.1996.tb07028.x.

Kagan, Y. Y. (2007), On Earthquake Predictability Measurement: Information Score and Error Diagram, *Pure and Applied Geophysics*, 164(10), 1947–1962, doi: 10.1007/s00024-007-0260-1.

- Kagan, Y. Y. (2009), Testing long-term earthquake forecasts: likelihood methods and error diagrams, *Geophysical Journal International*, 177(2), 532–542, doi:10.1111/j.1365-246X.2008.04064.x.
- Kagan, Y. Y., and L. Knopoff (1980), Spatial distribution of earthquakes: the two point correlation function, *Geophys. J. R. Astr. Soc.*, 62.
- Kijko, A., and A. Smit (2012), Extension of the aki-utsu b-value estimator for incomplete catalogs, *Bulletin of the Seismological Society of America*, 3, 1283 – 1287.
- Marzocchi, W., and J. D. Zechar (2011), Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model, *Seismological Research Letters*, 82(3), 442–448.
- McGuire, K. K. (1976), FORTRAN computer program for seismic risk analysis, *Open-File report 76-67*, United States Department of the Interior, Geological Survey, 102 pages.
- Mignan, A. (2012), Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude, *Journal of Geophysical Research*, 117(B8), doi: 10.1029/2012JB009347.
- Mignan, A., and G. Chouliaras (2014), Fifty years of seismic network performance in greece (1964–2013): Spatiotemporal evolution of the completeness magnitude, *Seismological Research Letters*, 85(3), 657–667, doi:10.1785/0220130209.
- Mignan, A., and J. Wöessner (2012), Corssa - theme IV - estimating the magnitude of completeness for earthquake catalogs, *Tech. Rep. doi: 10.5078/corssa-00180805*, Community Online Resource for Statistical Seismicity Analysis.
- Mignan, A., M. J. Werner, S. Wiemer, C.-C. Chen, and Y.-M. Wu (2011), Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs, *Bulletin of the Seismological Society of America*, 101(3), 1371–1385, doi:

10.1785/0120100223.

Mignan, A., C. Jiang, J. D. Zechar, S. Wiemer, Z. Wu, and Z. Huang (2013), Completeness of the mainland china earthquake catalog and implications for the setup of the china earthquake forecast testing center, *Bulletin of the Seismological Society of America*, 103(2A), 845–859, doi:10.1785/0120120052.

Molchan, G. M., and Y. Y. Kagan (1992), Earthquake prediction and its optimization, *Journal of Geophysical Research*, 97(B4), 4823, doi:10.1029/91JB03095.

Musson, R. M. W. (1999), Probabilistic seismic hazard maps for the north balkan region, *Annali di Geofisica*, 42(2), 1109 – 1124.

Nadaraya, E. A. (1964), On estimating regression, *Theory of Probability & Its Applications*, 9(1), 141–142, doi:10.1137/1109020.

Nasir, A., W. Lenhardt, E. Hintersberger, and K. Decker (2013), Assessing the completeness of historical and instrumental earthquake data in austria and the surrounding areas, *Austrian Journal of Earth Sciences*, 106(1), 90–102.

Ogata, Y., and K. Katsura (1993), Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophysical Journal International*, 113(3), 727–738, doi:10.1111/j.1365-246X.1993.tb04663.x.

Pagani, M., D. Monelli, G. Weatherill, L. Danciu, H. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, M. Simionato, and D. Vigano (2014), OpenQuake engine: An open hazard (and risk) software for the global earthquake model, *Seismological Research Letters*, 85(3), 692–702, doi:10.1785/0220130087.

Pirchiner, M., B. Collaço, J. Calhau, M. Assumpção, and J. C. Dourado (2011), The BRAzilian seismographic integrated systems (BRASIS): infrastructure and data management, *Annals of Geophysics*, doi:10.4401/ag-4865.

Riznichenko, J. V. (1959), On quantitative determination and mapping of seismic activity, *Annals of Geophysics*, 12(2).

Santos, S., and S. Lima (2008), The new brazilian standard for seismic design, in *The 14 World Conference Earthquake Engineering 14WCCEE*, Beijing, China.

Santos, S. H. C., S. d. S. Lima, and F. C. M. Silva (2010), Seismic hazard for brazilian northeastern region, *IBRACON Structures and Materials Journal*, 3(3).

Schorlemmer, D., and J. Woessner (2008), Probability of detecting an earthquake, *Bulletin of the Seismological Society of America*, 98(5), 2103 – 2117.

Schorlemmer, D., M. C. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007), Earthquake likelihood model testing, *Seismological Research Letters*, 78(1), 17–29, doi:10.1785/gssrl.78.1.17.

Scordilis, E. (2006), Empirical global relations converting m s and m b to moment magnitude, *Journal of Seismology*, 10(2), 225–236, doi:10.1007/s10950-006-9012-4.

Shedlock, K. M., and J. G. Tanner (1999), Seismic hazard map of the western hemisphere, *Annals of Geophysics*, 42(6).

Stepp, J. C. (1972), Analysis of completeness of the earthquake sample in the Puget Sound area and its effects on statistical estimates of earthquake hazard, in *International Conference on Microzonation for Safer Construction Research and Application*, vol. 2, pp. 897–909, University of Seattle.

Stucchi, M., P. Albini, M. Mirto, and A. Rebez (2004), Assessing the completeness of italian historical earthquake data, *Annals of Geophysics*, 47(2-3).

Takeya, M., J. Ferreira, R. Pearce, M. Assumpção, J. Costa, and C. Sophia (1989), The 1986–1988 intraplate earthquake sequence near João Câmara, northeast Brazil: evolution of seismicity, *Tectonophysics*, 167(2-4), 117–131, doi:10.1016/0040-1951(89)90062-0.

Tanner, J. G., and J. B. Shepherd (1997), Seismic hazard in latin america and caribe,

Tech. rep., PAIGH, Pan-American Institute of Geography and History.

Uhrhammer, R. (1986), Characteristics of northern and central california seismicity,

Earthquake Notes, 57(1), 21.

van Stiphout, T., J. Zhuang, and D. Marsan (2012), Theme V: Models and techniques for analysing seismicity, *Tech. rep.*, Community Online Resource for Statistical Seismicity Analysis.

Veloso, A. V. (2014), On the footprints of a major Brazilian Amazon earthquake, *Anais da Academia Brasileira de Ciências*, 86, 1115 – 1129.

Vere-Jones, D. (1992), Statistical methods for the description and display of earthquake catalogs, *Statistics in the Environmental and Earth Sciences*.

Vorobieva, I., C. Narteau, P. Shebalin, F. Beauducel, A. Nercessian, V. Clouard, and M.-P. Bouin (2013), Multiscale mapping of completeness magnitude of earthquake catalogs, *Bulletin of the Seismological Society of America*, 103(4), 2188–2202, doi: 10.1785/0120120132.

Watson, G. S. (1964), Smooth regression analysis, *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4), pp. 359–372.

Weatherill, G. (2014), Hazard modeller's toolkit - user guide, *Tech. rep.*, Global Earthquake Model (GEM).

Weatherill, G., and M. Pagani (2014), From smoothed seismicity forecasts to probabilistic seismic hazard: insights and challenges from a global perspective, in *2nd European Conference on Earthquake Engineering and Seismology*, ECEES.

Weatherill, G. A., M. Pagani, and D. Monelli (2012), The hazard component of the GEM modeller's toolkit: A framework for the preparation and analysis of probabilistic seismic hazard (PSHA) input tools, in *Proceedings of the 15th World Conference on*

Earthquake Engineering, Lisbon, Portugal, paper.

Weichert, D. H. (1980), Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, *Bulletin of the Seismological Society of America*, 70(4), 1337 – 1346.

Wiemer, S., and M. Wyss (2000), Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan, *Bulletin of the Seismological Society of America*, 90(4), 859–869, doi:10.1785/0119990114.

Woessner, J., and S. Wiemer (2005), Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty, *Bulletin of the Seismological Society of America*, 95(2), 684 – 698, doi:10.1785/0120040007.

Woo, G. (1996), Kernel estimation methods for seismic hazard area source modeling, *Bulletin of the Seismological Society of America*, 86(2), 353–362.

Zechar, J. D., and T. H. Jordan (2010), Simple smoothed seismicity earthquake forecasts for italy, *Annals of Geophysics*, 53(3), 99–105.

Zoback, M. L., and R. M. Richardson (1996), Stress perturbation associated with the Amazonas and other ancient continental rifts, *Journal of Geophysical Research*, 101(B3), 5459, doi:10.1029/95JB03256.

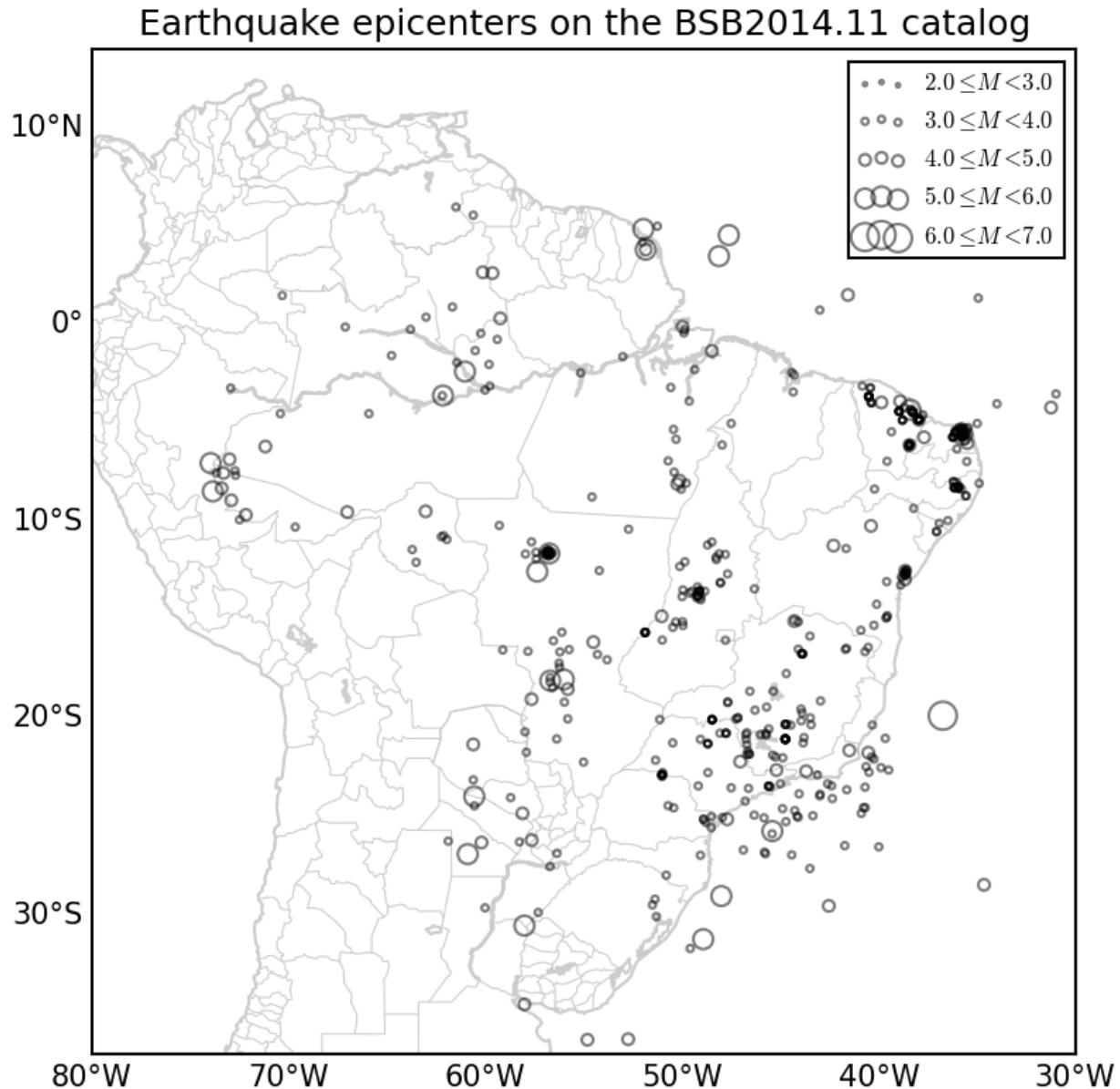


Figure 1. Brazilian seismicity from 1767 to 2014-01-01. Source: [BSB, 2014].

Table 1. BSB 2014.11 M_W proxies *ad-hoc* guidelines.

magnitude source	rule	uncertainty
mb or mR	$M_W(m) = 1.121m - 0.76$	0.3
Felt area A_f	$M_W(A_f) = 0.81 + 0.639 \log(A_f) + 0.00084\sqrt{A_f}$	0.4
Maximum intensity I_0	$mb(I_0) = 1.21 + 0.45I_0$ then $M_W(m)$	0.6
mb and A_f	$M_W(m, A_f) = 0.7M_W(m) + 0.3M_W(A_f)$	0.33

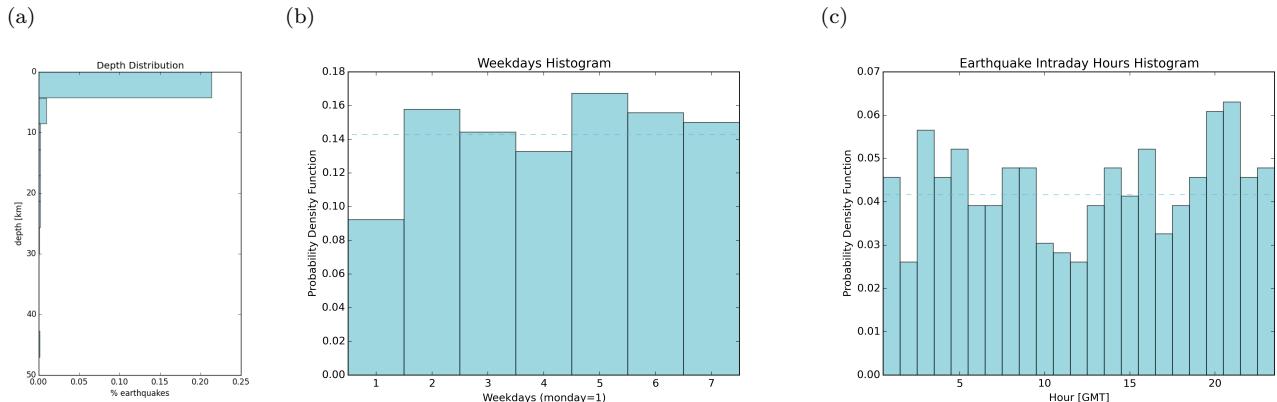


Figure 2. Catalogue overview. The histogram (a) shows the depth distribution. The histograms (b) and (c) represent the distributions of weekday and hour which earthquakes occur respectively. Dashed lines represent the mean value.

Table 2. Temporal completeness. For each magnitude interval, the completeness was computed by the *Stepp* [1972] method.

magnitude	year (Stepp)
3.0	1970
4.0	1959
4.5	1951
6.0	1933

Table 3. Temporal completeness. For each magnitude interval, the completeness comes from *Assumpção et al.* [2014] converted to M_W from m_b following table 1.

magnitude	year (Assumpção)
3.1	1980
4.3	1968
4.8	1962
6.0	1940

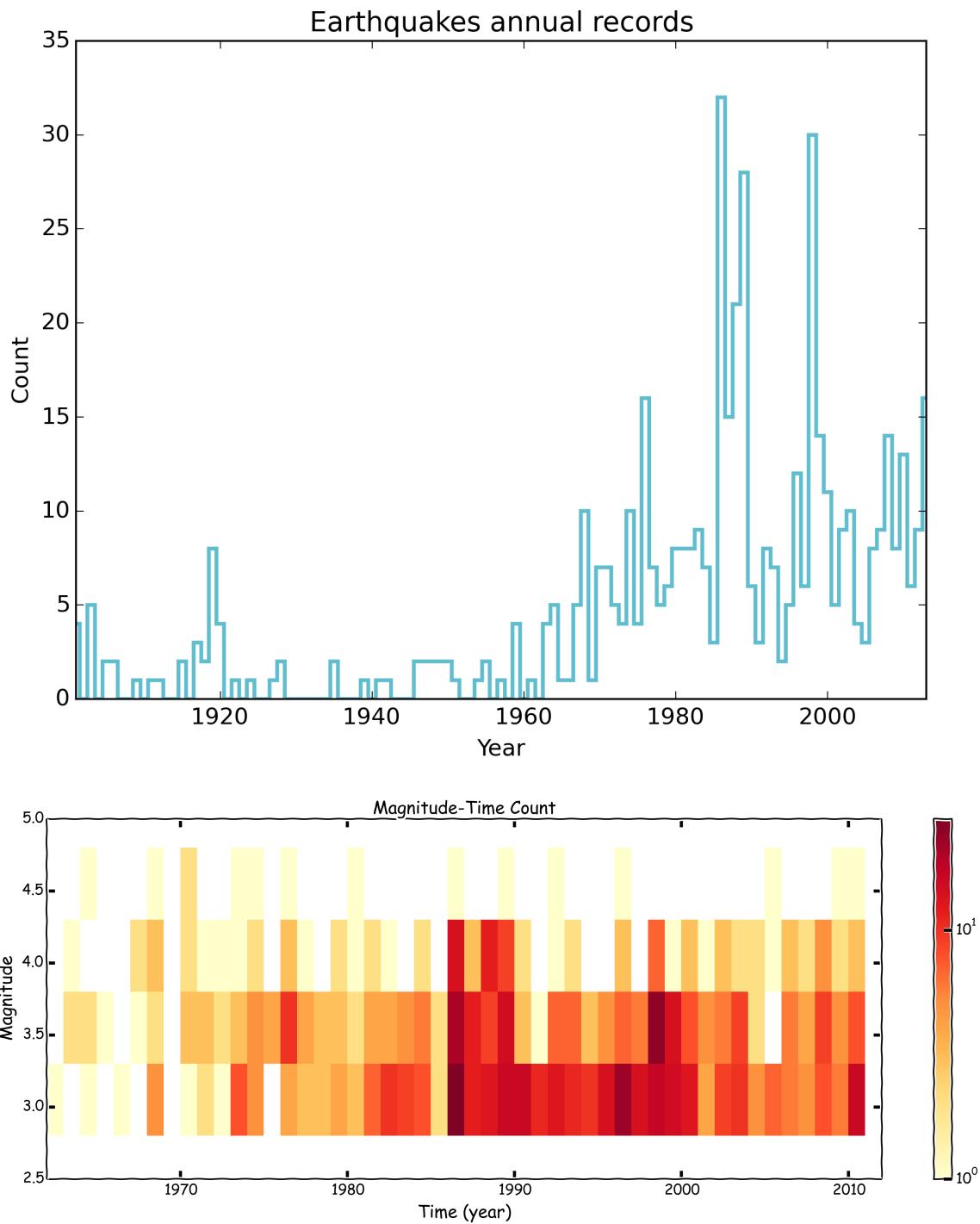


Figure 3. Earthquakes by year (a) and discriminated by magnitude (b).

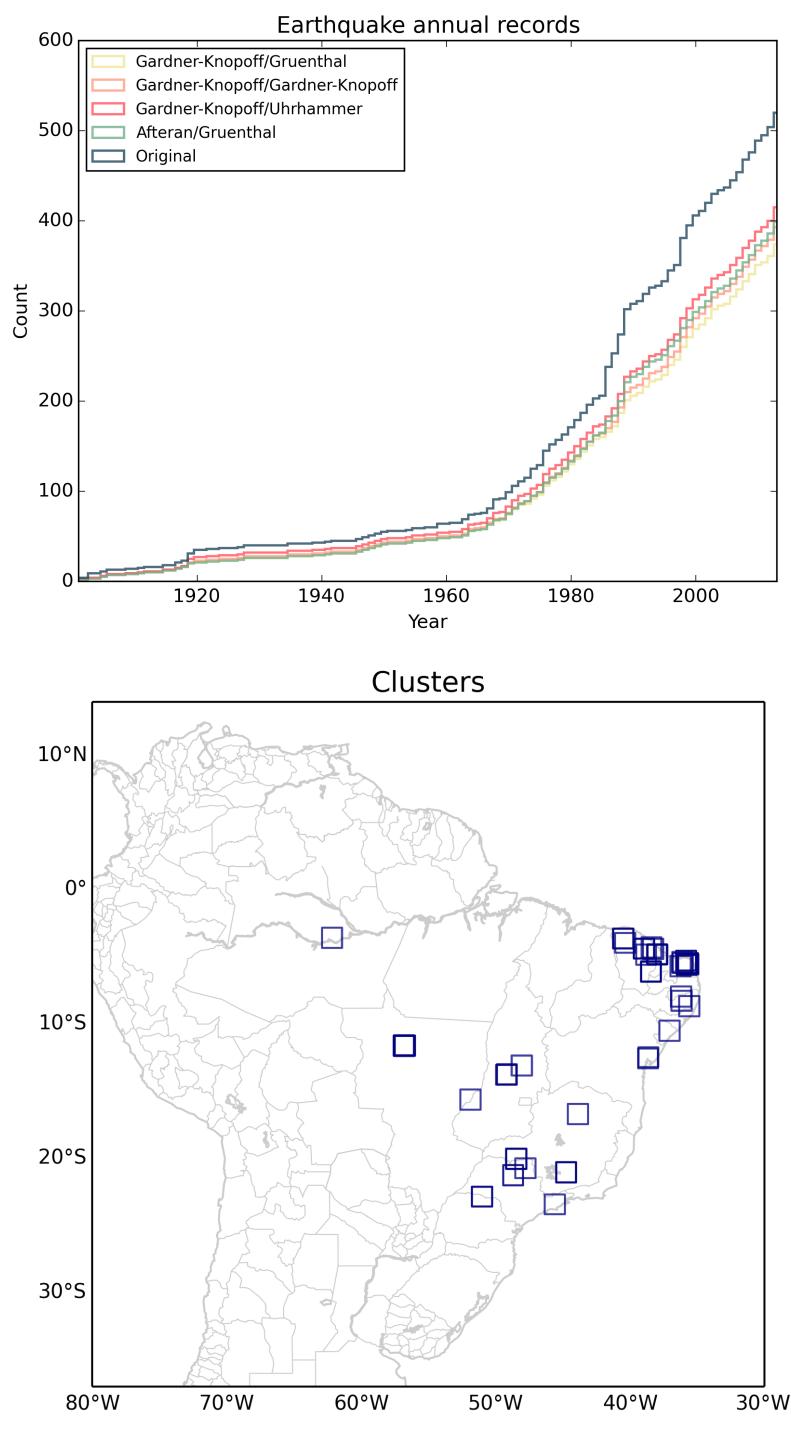


Figure 4. Decluster evaluation (a) using distinct algorithms and windows. There is no significant differences between final results. The map (b) of clusters from selected method shows places where earthquake sequences tends to occur.

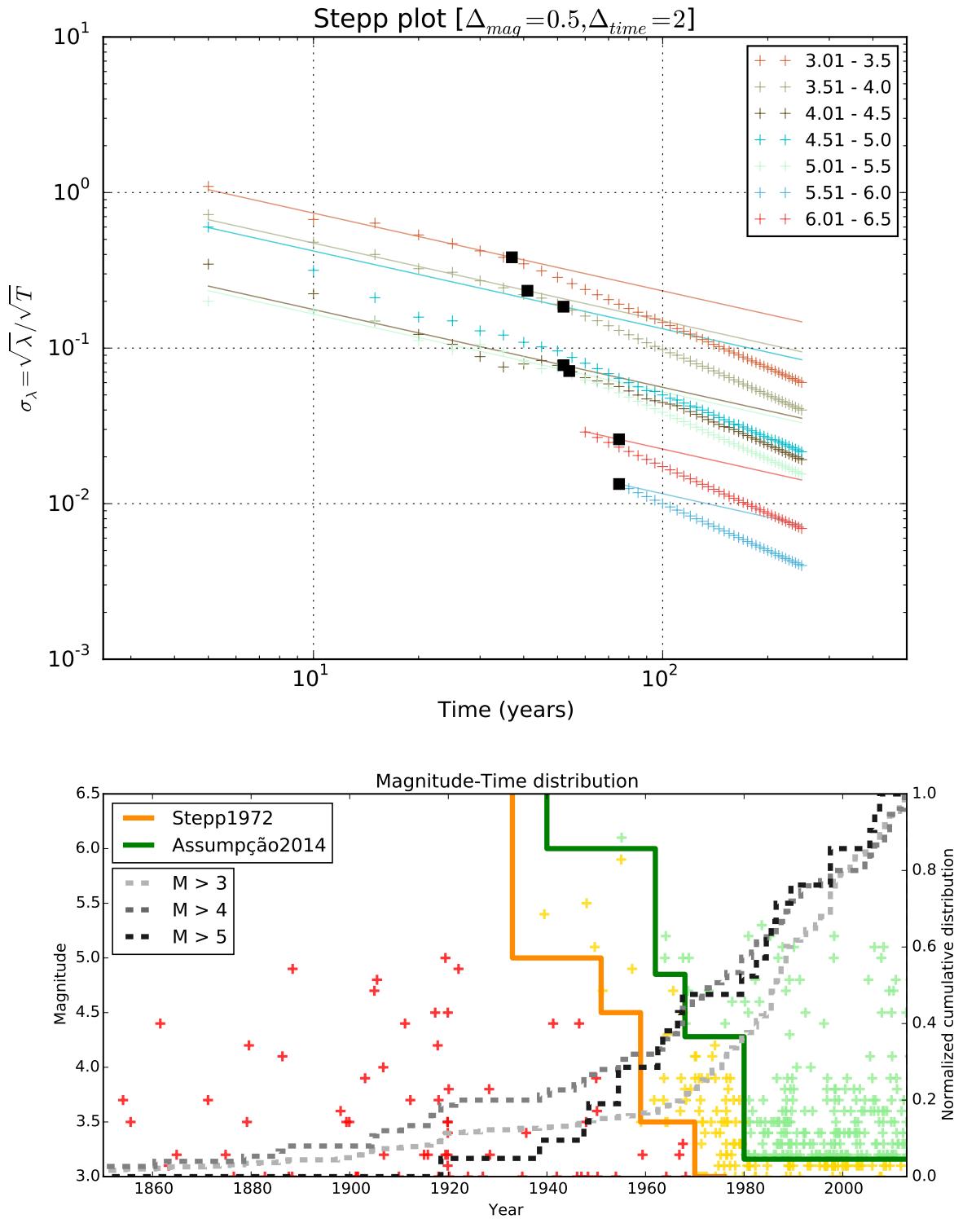


Figure 5. Completeness evaluation. The Stepp plot (a) was made using $\Delta_{mag} = 0.5$ and $\Delta_{time} = 2$ as magnitude and time bins. The magnitude-time scatter plot (b) shows in addition, three different temporal completeness models and some cumulative rates to better perform a visual analysis.

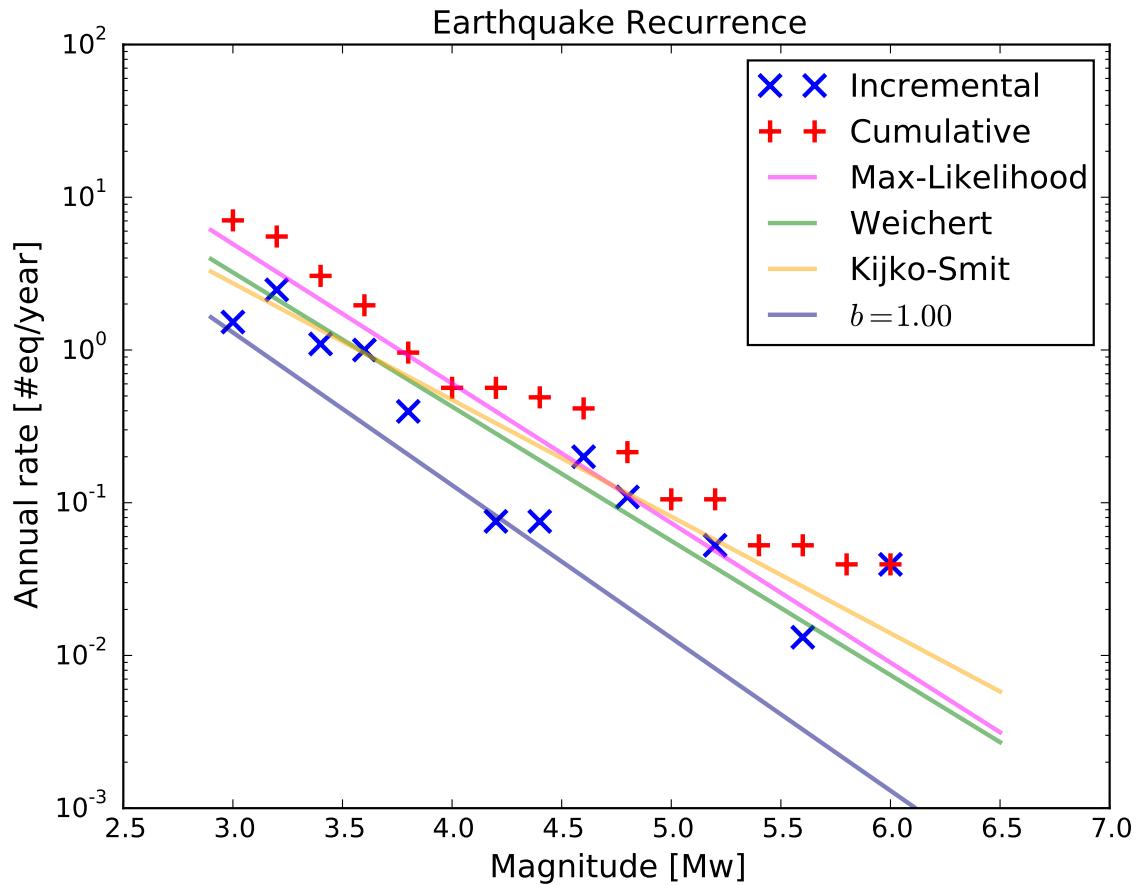


Figure 6. Annual earthquake recurrence. The blue x and red crosses are the incremental and cumulative recurrence distributions binned by 0.2 magnitude unity. Stepp completeness table and *BSB* [2014] declustered catalogue was used. Also a fixed b-value ($b=1$) incremental distribution is shown in blue continuous line, and the fitted by maximum-likelihood method (magenta, $b = 0.91 \pm 0.07$), *Weichert* [1980] (green, $b = 0.88 \pm 0.04$) and *Kijko and Smit* [2012] (orange, 0.76 ± 0.04) are shown as continuous lines. The respective 10^a was 3.7, 3.8 ± 0.2 , 3.5 ± 0.2 , 3.2 ± 0.1 .

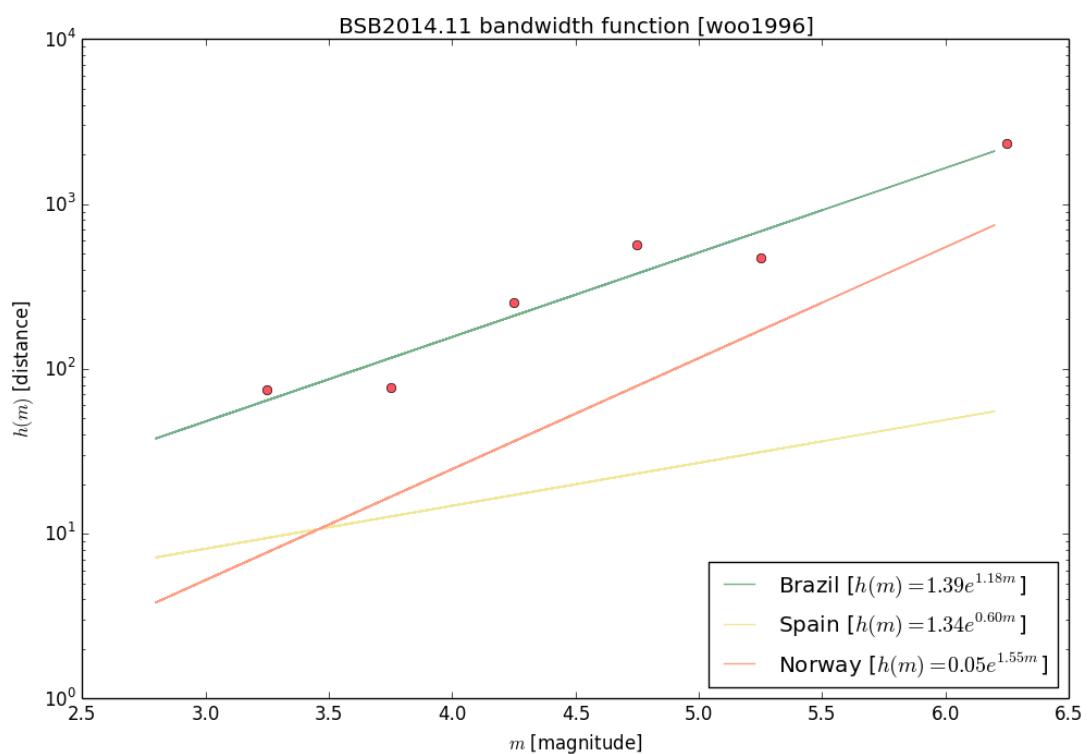


Figure 7. Magnitude dependence bandwidth.

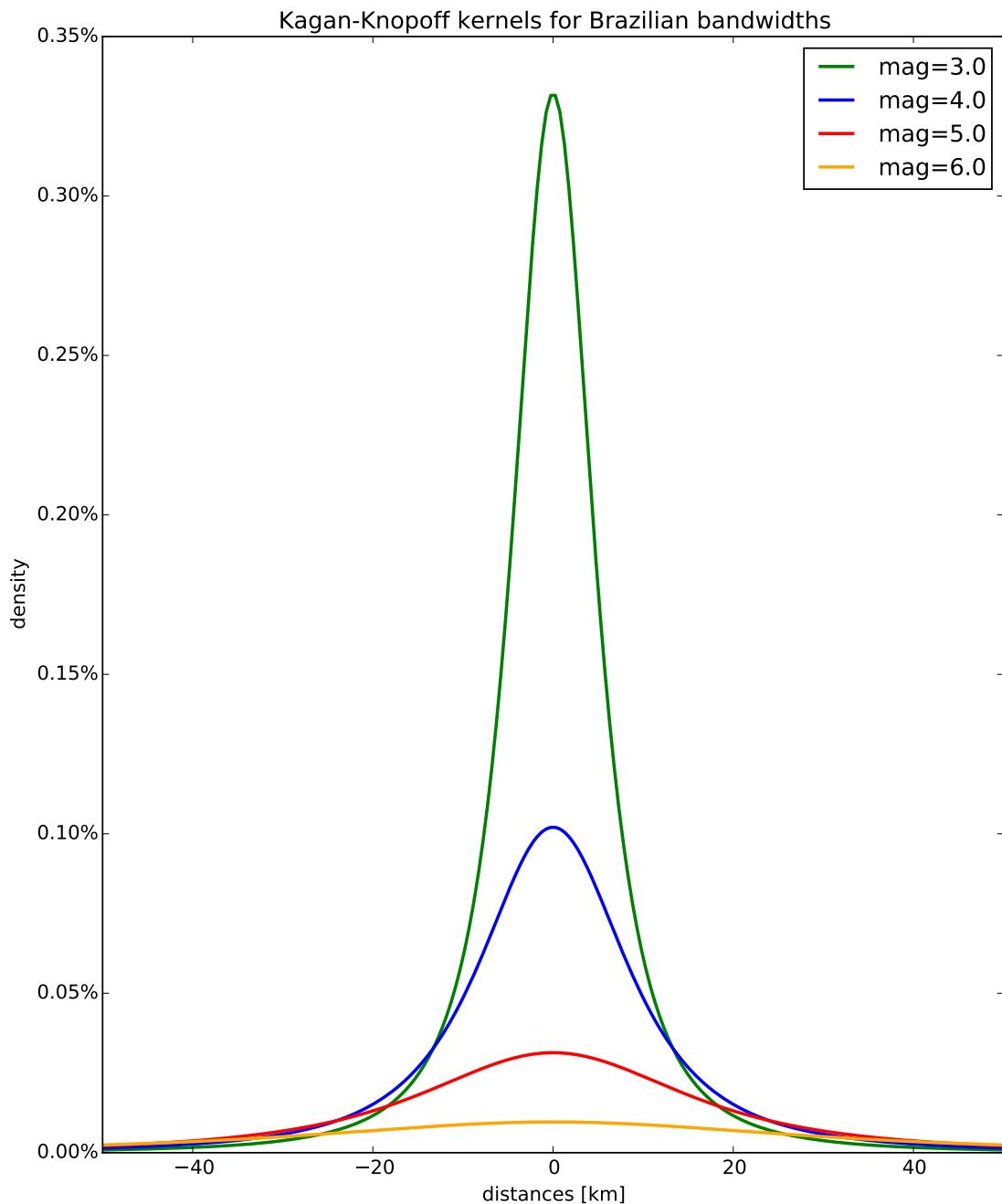


Figure 8. *Kagan and Knopoff* [1980] kernel shapes for some magnitude values.

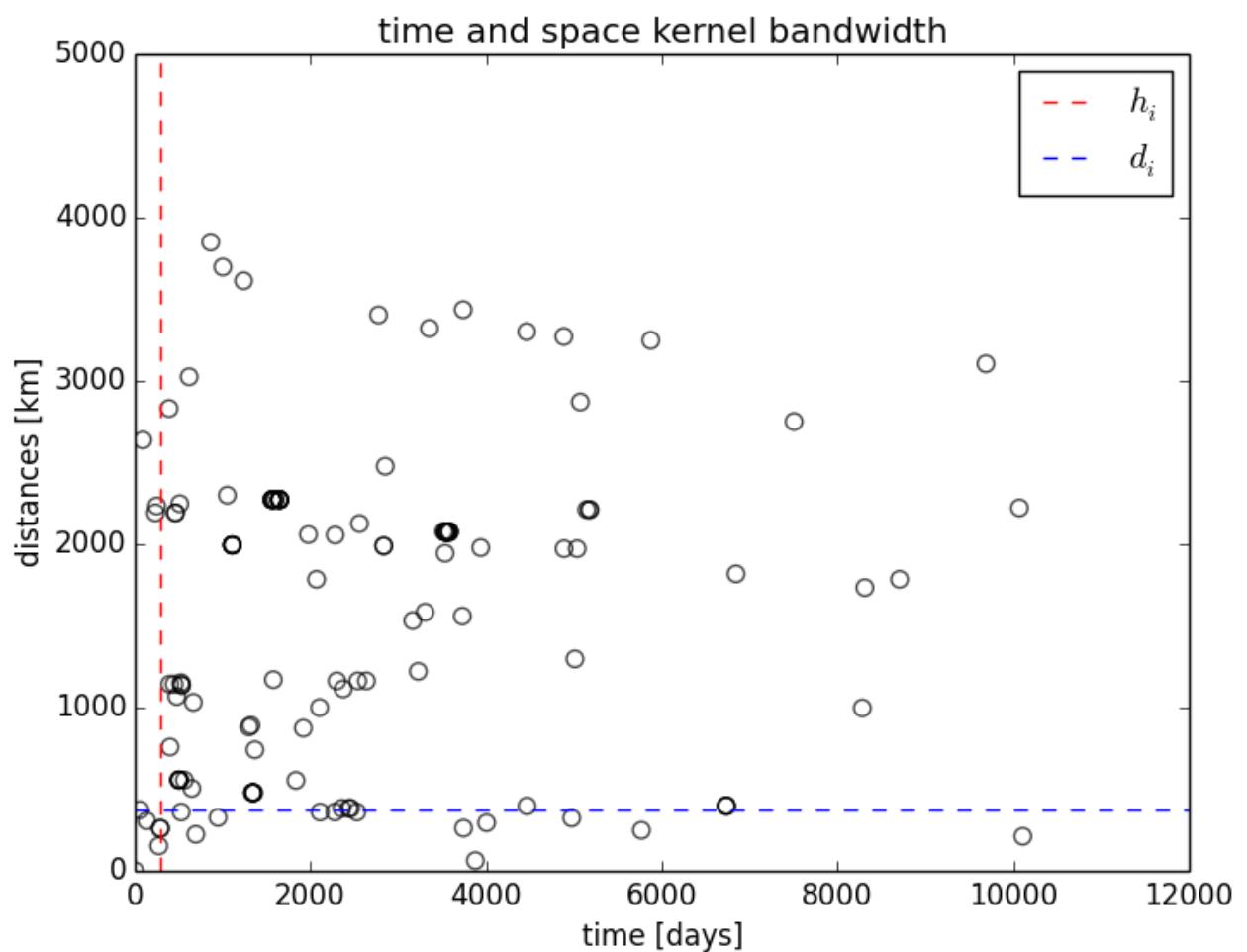
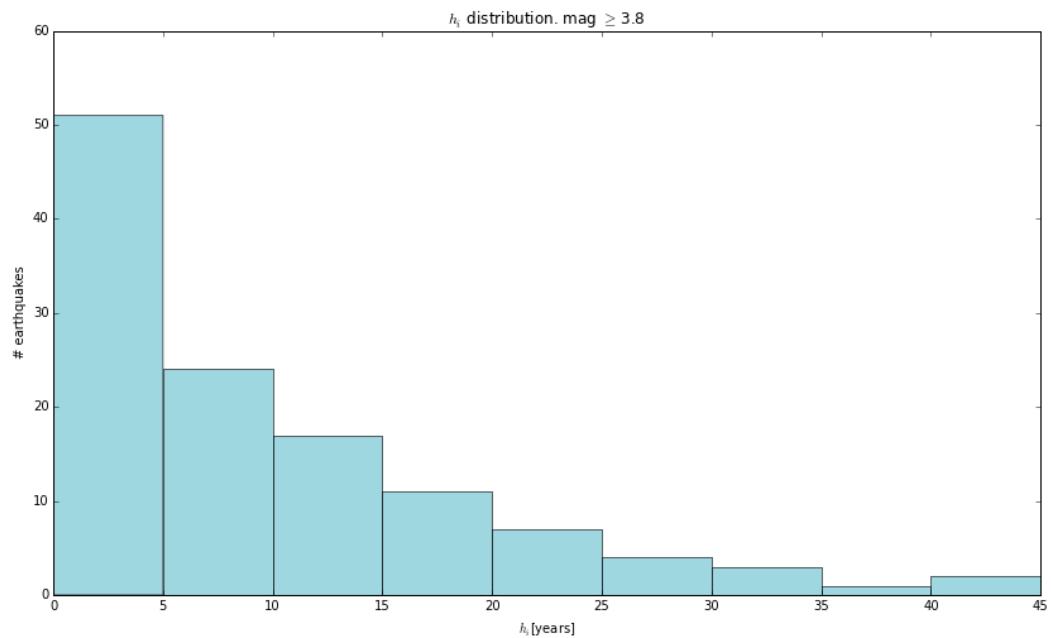
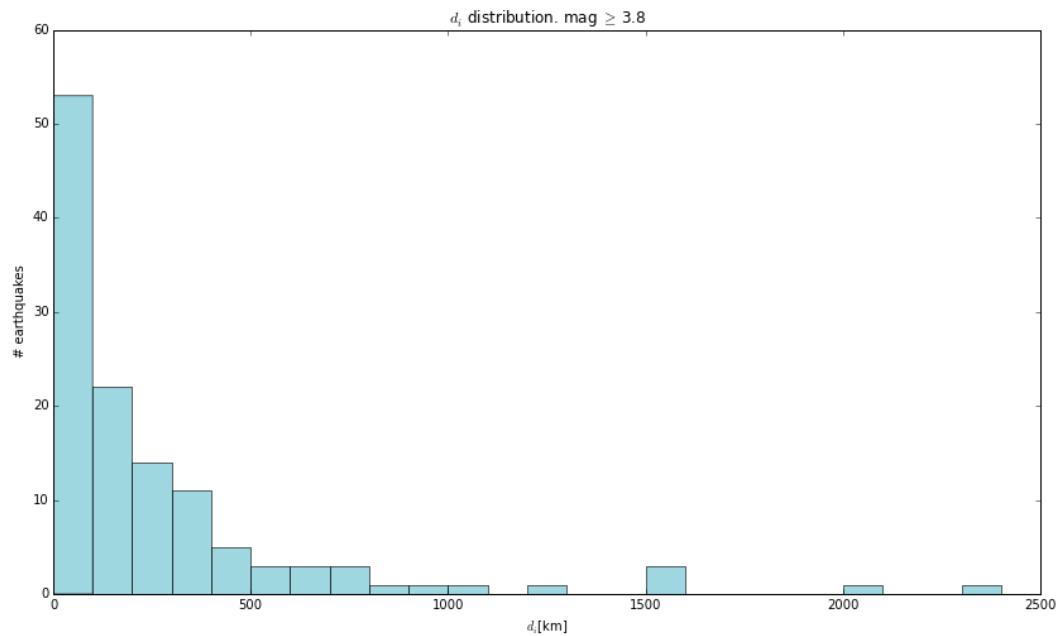


Figure 9. Local bandwidth example.



(a)



(b)

Figure 10. (a) h_i and (b) d_i distributions from BSB.

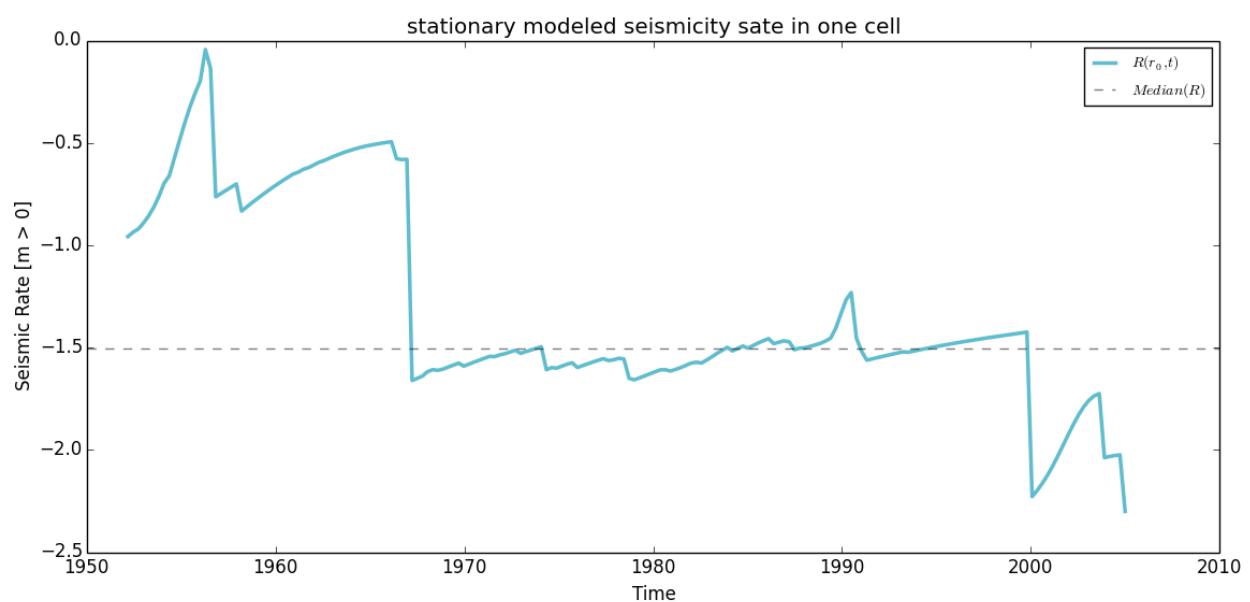


Figure 11. Stationary seismic rate

BSB2014.11 learning and target catalogs

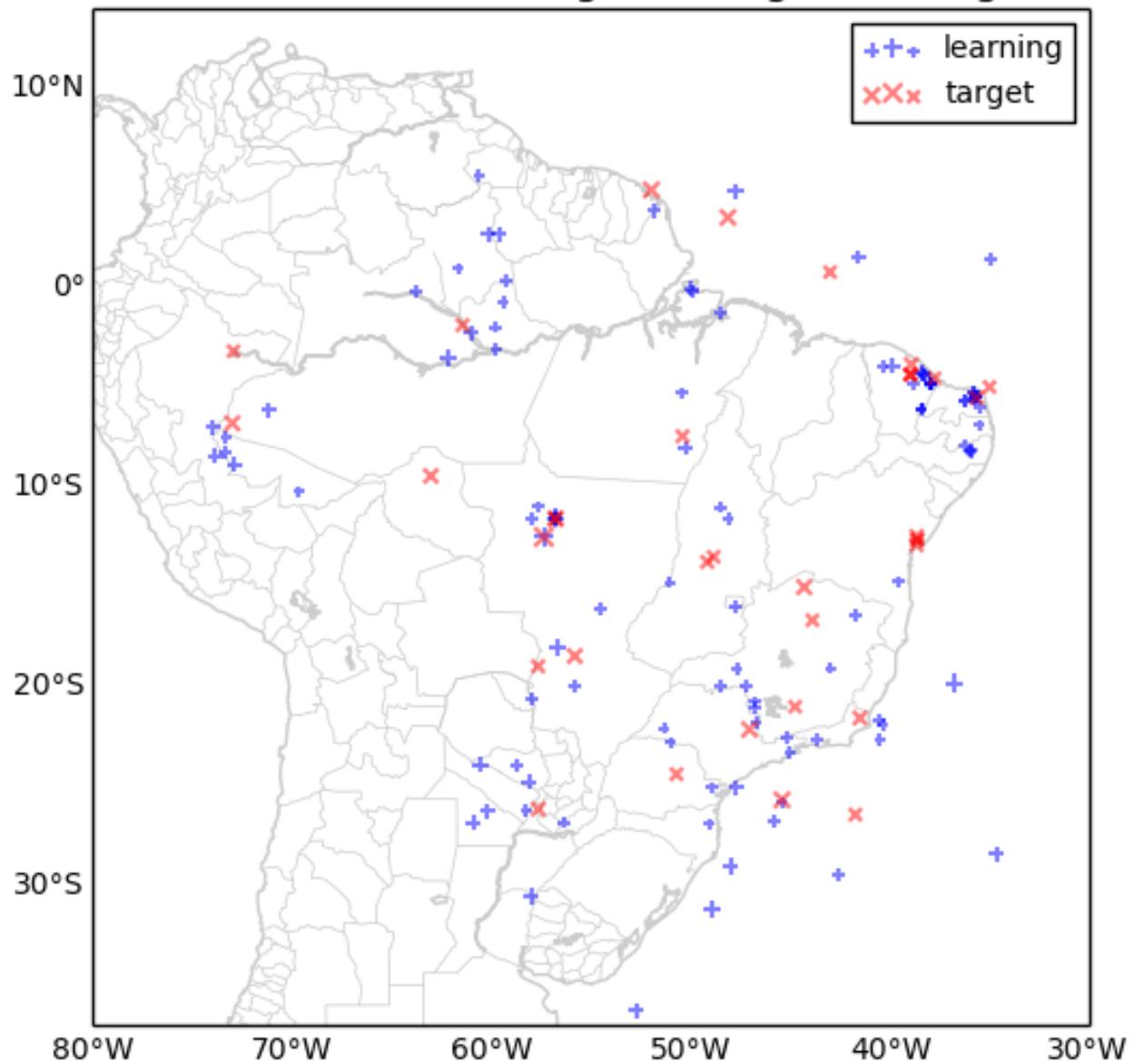


Figure 12. Learning and testing catalogs.

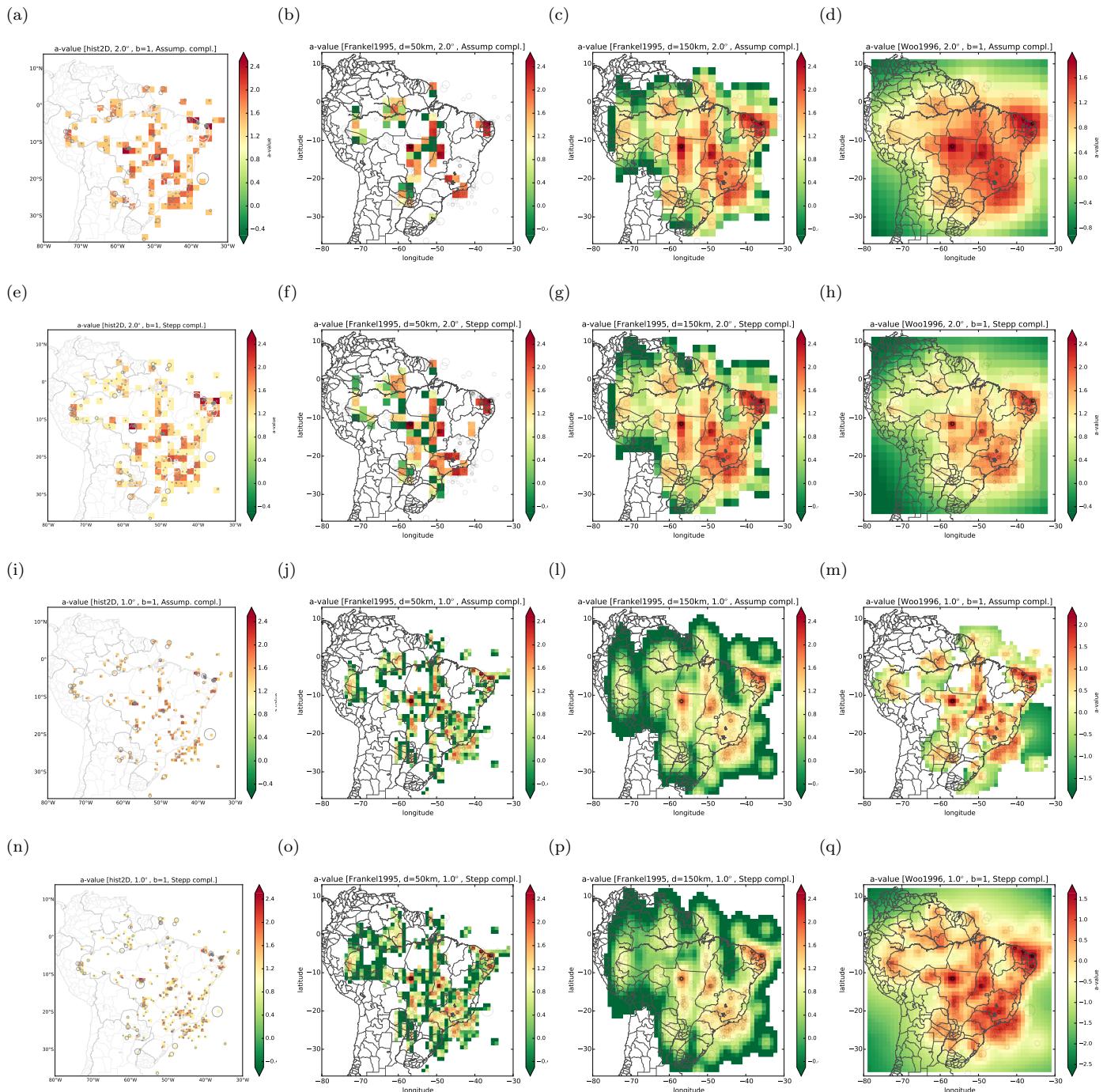


Figure 13. Smoothed rates comparission: (a) shows *Frankel* [1995] method smoothed rate results, (b) and (c) represent rates smoothed by *Woo* [1996] and *Helmstetter and Werner* [2012] methods respectively.

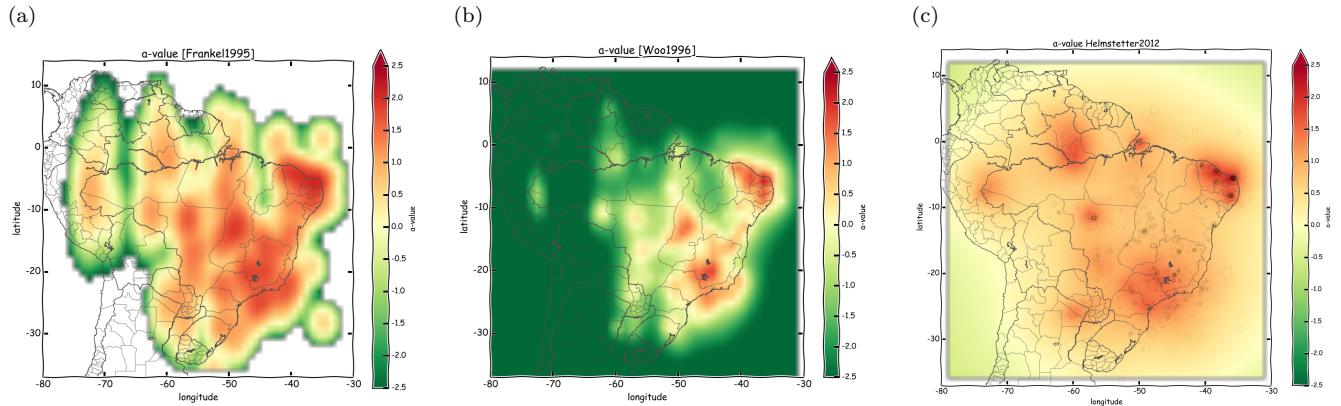


Figure 14. Smoothed rates comparission: (a) shows *Frankel* [1995] method smoothed rate results, (b) and (c) represent rates smoothed by *Woo* [1996] and *Helmstetter and Werner* [2012] methods respectively.

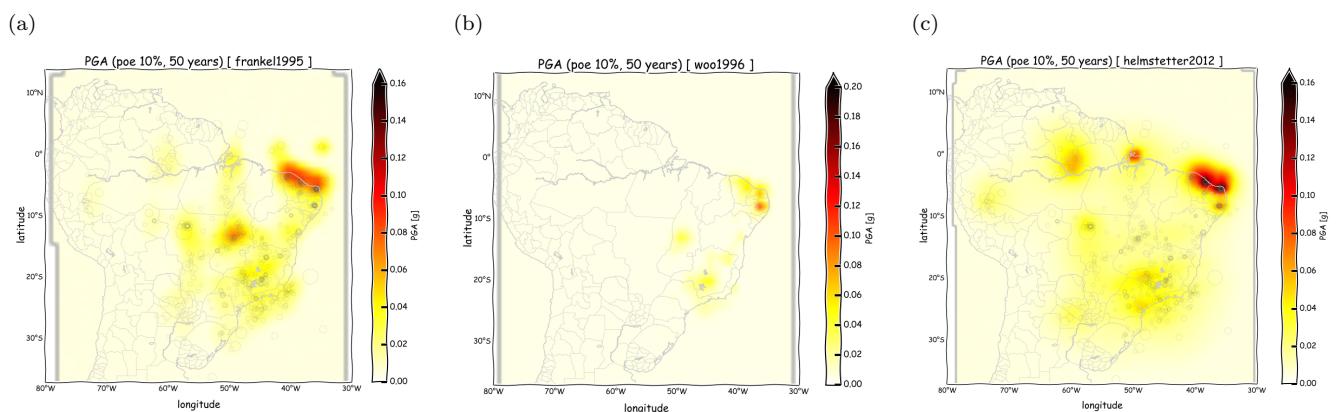


Figure 15. PGA (poe 10%/50y) comparission for (a) *Frankel* [1995], (b) *Woo* [1996] and (c) *Helmstetter and Werner* [2012] methods.

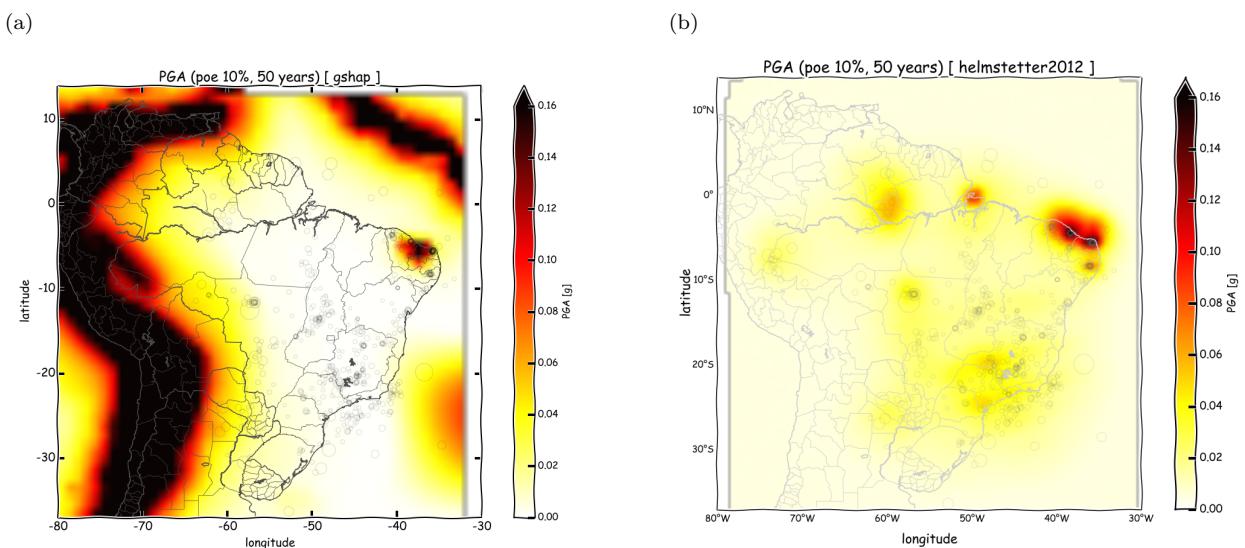


Figure 16. Previous available Brazilian PSHA from (a) GSHAP project and the PSHA proposed by this work (b).