Optimization of the SeisComP software for automatic detection and location of regional seismic events in Brazil

Salles, V. and Bianchi, M. (University of São Paulo)

Abstract

We analyzed 11TB of seismic data from the Brazilian Seismographic Network (RSBR) and additional monitoring stations to optimize the software and workflows used for detecting and locating seismic events in Brazil. The RSBR operators have been using the SeisComP software since the beginning of its operation 10 years ago. While the software offers a wide range of configuration options, no optimization had been done so far. By analyzing the catalog of seismic events in the period of 2014-2021, we found that more than 90% of the existing events were manually detected. To optimize the parameters and workflows used for detecting and locating seismic events, we used four increasing volume datasets. We used known events as a reference for wave detection parameter optimization, including a band-pass and an STA/LTA filters. Results indicate that the best parameters to detect regional P-wave arrivals are a band-pass filter with cutoff frequencies between 4.5 Hz and 10 Hz followed by an STA/LTA filter with time windows of 0.2s (STA) and 45s (LTA). After detecting the waves, an AIC picker was used to define the precise arrival time. The nucleation grid was constructed considering the density of stations within a radius of 10 degrees from each grid node. We also implemented modifications that included new ways of verifying the validity of nucleated origins (using preliminary magnitudes) and removal of unnecessary arbitrary flows that hinder the automatic location of regional events. Our modifications significantly improved the automatic detection of seismic events in Brazil. The number of locatable detections increased from 1024 to 5981 between 2014 and 2021, while reducing the number of false events from 84.2% to 6.7%. We also detected real events that are not in the official RSBR catalog. The events that are also present in the RSBR seismic catalog increased from 78 to 292. Furthermore, we were able to detect events with low magnitudes (approximately M1.5) and improve the origin parameters, including the decrease in travel time residuals and the possibility of nucleation with the contribution of a few stations.

Introduction

The Brazilian Seismographic Network (Rede Sismográfica Brasileira - RSBR) began its operations after a few decades of attempts to establish a continuous study of seismicity in Brazil and South America. Several initiatives took place from the beginning of the 1900s, until seismology had a breakthrough in Brazil in 1970 with the formation of seismological groups in different regions of the country, with the purpose of studying the seismic hazard related to nuclear power plants and occurrences of induced seismicity (Bianchi et al., 2018).

Currently, RSBR has almost 100 broadband seismographic stations with the capacity for real-time data transmission, continuously feeding the processing computers of the four main seismological centers in the country: the University of São Paulo (USP), the University of Brasília (UnB), the Federal University of Rio Grande do Norte (UFRN) and the National Observatory (ON), as shown in Figure 1.

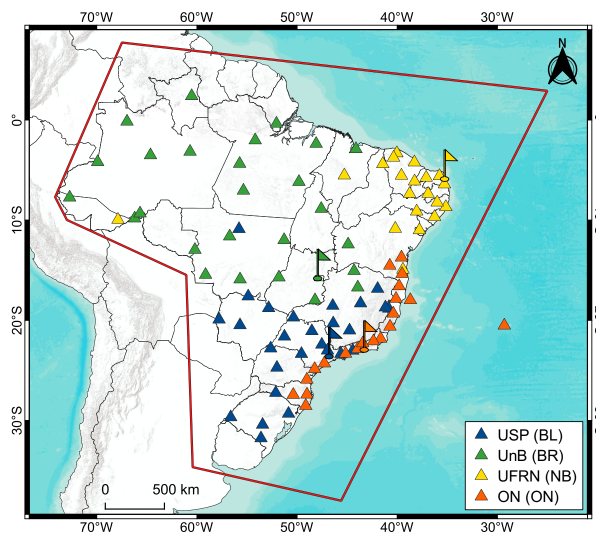

Figure 1 – Seismological stations (triangles) and institutions (flags) participating in RSBR. Subnetworks and their respective responsible institutions are indicated by different colors. Source of information: IAG-USP database (October/2022).

By analyzing the temporal magnitude distribution from the RSBR catalog, Bianchi et al. (2018) estimated the detectability of seismic events in Brazil between the years 1940 and 2016. According to the author, the installation of stations in the Amazon region in 2014 allowed the detection of events with magnitudes as low as M3.5 for all of Brazil.

Since 2012, RSBR has annually detected twice as many events with magnitudes between M3.5 and M4.5, when compared to what used to be detected prior to the establishment of the national seismographic network. For some regions (Southeast and Northeast), where the density of stations is higher, the detection limit is M3.0 or even lower.

Although the RSBR has been using the SeisComP software since 2009, only a minority of the events in the Brazilian seismic bulletin were detected automatically. The majority were detected manually through analysis of dayplots from strategically selected seismographic stations.

One of the factors that can increase the detectability of events by RSBR is the evaluation of the processes carried out by the automatic detection system used to manage the network and its parameters, in order to optimize the detection and location of events in Brazilian territory.

SeisComP

The operation of SeisComP is based on several independent modules, each one performing specific processing that contribute to the location of a seismic event. An operation flowchart of these modules is shown in Figure 2.

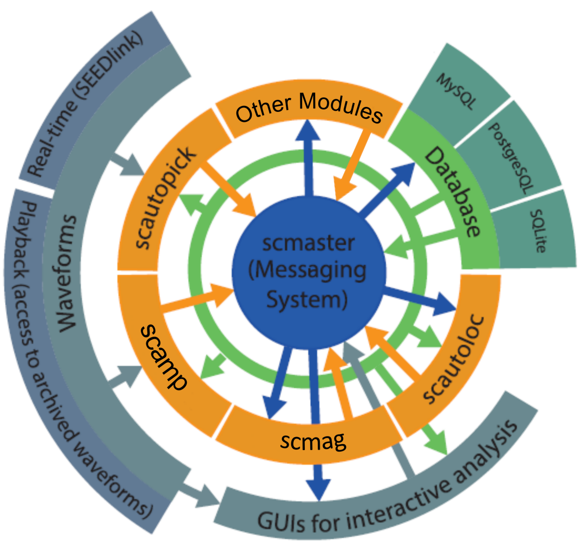

Figure 2 – Operation flowchart of SeisComP and its modules. The arrows indicate the direction of information flow, colored according to those responsible for sending it. Source: adapted from Behr et al. (2016).

The modules *scautopick* and *scautoloc* are responsible for a series of important processes in the automatic detection of seismic events. Within them, several parameters are defined so that the automation is as effective as possible, avoiding false-positives and false-negatives. It is also in these modules that a large part of the capability for detecting regional events with relatively low magnitudes ($\leq$ M3.5) can be improved, adjusting parameters and processing methods that allow SeisComP, initially developed for detecting earthquakes of global scales, to also satisfactorily locate regional events.

SeisComP's configuration module provides the user with the possibility of modifying several parameters used throughout the process of locating seismic events. However, many parameters and processing flows are not accessible to the user and can only be modified through SeisComP's source code. Many of these parameters and flows are fundamental while locating regional events. SeisComp default options and adopted flows focus on the teleseismic location and sometimes prevent the software from being successful in locating this type of seismic event. In addition, it was found that some user-editable parameters are overwritten or modified throughout the source code, meaning that the chosen settings are not actually used during the processes.

To establish a processing flow and set of parameters that optimize the location of seismic events in Brazil, a thorough examination of all the procedures executed by *scautoloc* was conducted. This analysis allowed the identification of improvement points as well places where significant modifications were necessary. New methodologies for evaluating the quality of nucleated origins were defined, in addition to new parameters that must be followed during the nucleation process.

Methods and procedures

This work focused on manually optimizing sets of parameters and procedures for an incremental series of datasets. Figure 3 presents a flowchart indicating the four datasets used during each stage of optimization of *scautopick* and *scautoloc* parameters. Several rounds of processing were carried out in order to verify the contribution of each parameter and process in the detection and location of Brazilian regional earthquakes. Concomitantly with the procedures for optimizing the processing modules, a new velocity model was also developed, as further described.

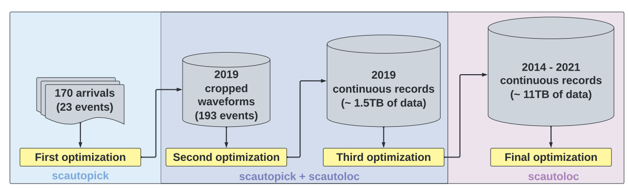

Figure 3 – Flowchart indicating all four datasets used during each stage of optimization of parameters of scautopick and scautoloc modules. As the datasets progressed, the amount of data under consideration increased. The first dataset was used solely for optimizing the picking parameters (scautopick), while the final dataset served to test the system's overall performance considering its new parameters and adapted procedures.

Initially, 170 arrivals from 23 Brazilian regional events from the RSBR catalog were selected for the initial stage of optimization of the two main parameters of the *scautopick* module: the frequency filter and the time windows of the STA/LTA algorithm. We sought to cover most of the magnitude ranges included in the catalog, choosing events with good quality seismographic records. Then, several routines were developed in *Python* to carry out the optimizations, favoring the detection of P-wave arrivals from seismic events. Other parameters of *scautopick* also went through the initial optimization stage, such as the trigger on and off values.

Subsequently, *scautopick*'s parameters underwent a second optimization process, using snippets of waveforms from 193 seismic events that occurred in 2019. At this second moment, tests were also conducted to improve parameters of the *scautoloc* module, including its hard-coded processes.

In a third optimization step, the detection and location parameters were refined through several rounds of tests, this time using continuous data recorded throughout 2019. In this way, we sought to verify the behavior of the modified software in the context of processing with continuous data, not just waveform snippets. Again, algorithms were developed in *Python* and *Bash* to execute *scautopick* and *scautoloc* modules in an efficient and systematic way. The parametric data of the automatically located events, in *Sc3ML* format, were constantly evaluated through graphs and statistical analyzes.

Finally, the entire database for the period from 2014 to 2021 was processed, in order to verify whether the parameters refined with data from the year 2019 also generated the expected results throughout the entire period, in addition to making final adjustments to *scautoloc*'s source code. At this stage, more than 11TB of continuous data were processed, recorded by RSBR and additional temporary and permanent stations (Figure 5).

For comparison purposes, the original SeisComP (*i.e.*, without any modification) was used to generate a catalog of events during the same period. In that way, three catalogs of seismic events were obtained to be compared with each other: *(i)* official RSBR catalog, *(ii)* modified SeisComP catalog and *(iii)* original SeisComP catalog. It should be noted that the SeisComP used at IAG-USP is very similar to the original version, presenting few modifications in its detection and location parameters and no modifications in its source code.

The work carried out can be condensed into three main topics: development of a new velocity model, optimization of *scautopick* and optimization of *scautoloc*.

Results

Frequency filter and STA/LTA parameters

The limits of the ideal frequency filter were defined as 4.5 Hz and 10 Hz, values that provided greater enhancement of the P-waves of most of the seismic events evaluated. The time windows for the STA/LTA algorithm were defined as 0.2 and 45 seconds, respectively. Figure 4 shows the difference between the waveforms filtered with the combination of default SeisComP parameters and with the combination of the proposed parameters.

The waveform processed with the default parameters of SeisComP show a poor STA/LTA ratio for the P-wave arrival, which is practically undetectable by the software. Using the proposed parameters, the arrival of the P-wave and its STA/LTA ratio is considerably enhanced, which allows its correct detection in the seismograms.

Finally, we set the “trigger on” and “trigger off” values as 3.0 and 0.7, respectively. For the XC network, the “trigger on” was set to 4.5 since this network presents a higher noise level when compared to other stations.

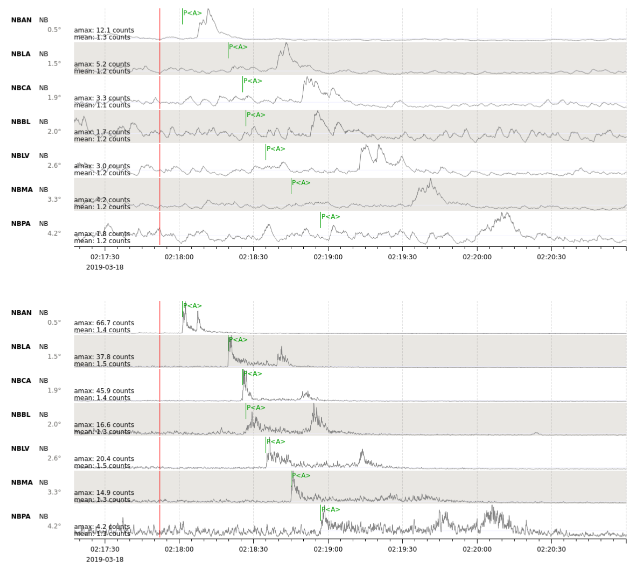

Figure 4 – Examples of filtering and application of STA/LTA in waveforms referring to a seismic event of magnitude M2.5. Above, the default parameters of SeisComP were used (frequencies from 0.7 Hz to 2 Hz and STA/LTA of 2 and 80 seconds). Below, the data processed with parameters proposed in this work (frequencies from 4.5 Hz to 10 Hz and STA/LTA of 0.2 and 45 seconds).

Proposed grids

The *scautoloc* module depends on the usage of a nucleation grid, which in the default distribution assumes a nucleation point per degree for the whole Earth. To improve the location of regional events, a regional grid must be constructed. Figure 5 exemplifies its proposed creation process: for each station, the percentages of data available in 2019 are presented, as well as the geographic locations of the stations that have at least 50% of data recorded in the year, which are used to generate the grid for that period. The map in Figure 5 shows the grid for the year 2019, with a color scale based on the number of stations contained in 10° radius circles centered on each point.

In all years, the Southeast Region and part of the Midwest had the highest station densities, which directly reflects the number of events detected in these regions. The North Region of the country has the lowest number of stations in operation in the analyzed period of time, followed by the Northeast Region. The number of stations for each point determines the maximum distance allowed for a station to contribute to an origin and the minimum number of picks for this origin to be accepted. Both parameters are encoded in the grid description, which has also been optimized.

New velocity model - BRA23

During the process of origin's nucleation, the main way to reduce false events is through the use of consistent values of maximum temporal residual allowed for an arrival and maximum RMS of the origin, considering a regional scale seismicity. Time residuals depend on the velocity model used during location, requiring an optimized model (BRA23) to reduce the residuals and RMS of the solutions. The new model needed to provide consistent travel times for both local and regional waves.

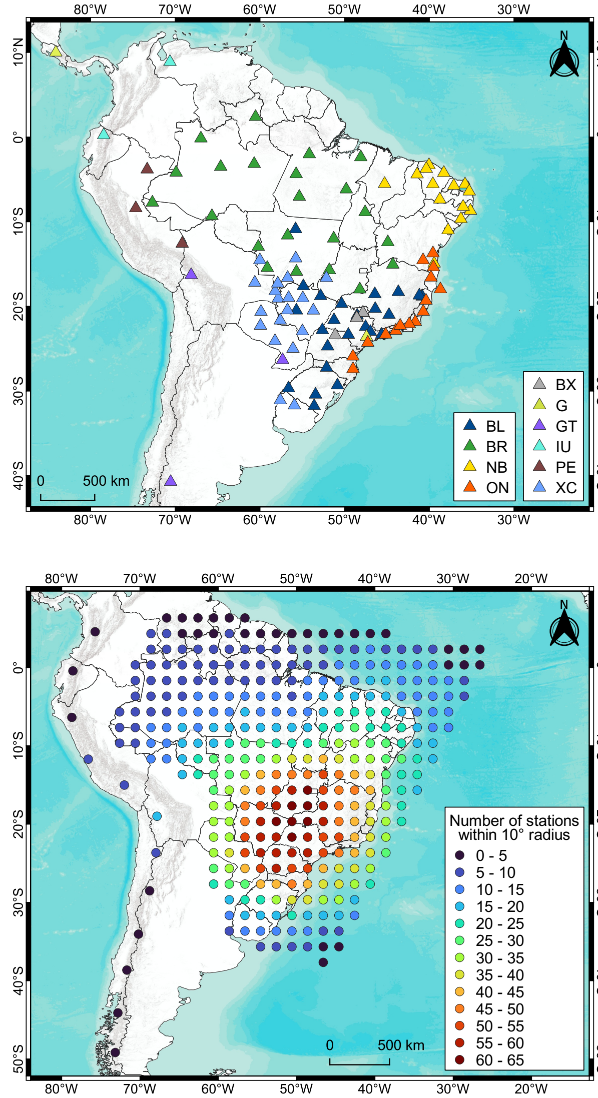

Figure 5 – Above: stations with at least 50% of data available in 2019. Below: proposed grid for the year 2019.

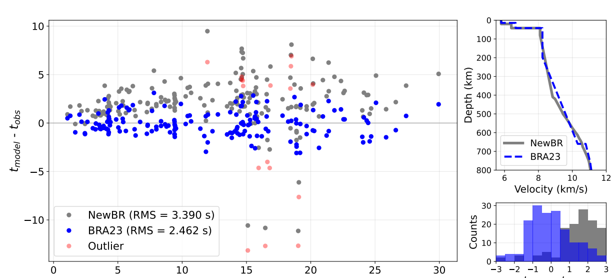

Figure 6 – Comparison of time residuals ($t_{model} - t_{obs}$) using models BRA23 and NewBR. The proposed new model tends to present an average close to zero, unlike the NewBR model, which presents a trend of positive residuals. The points in red correspond to outliers, not used during local optimization processes.

The BRA23 model was obtained through an optimization of the same data used in the construction of the NewBR model Assumpção et al. (2010), added to two more recent regional events. In total, data from 17 events were considered, with 183 P-wave detections from different regions of Brazil.

For a validation of the BRA23 model, Figure 6 presents the time residuals of all detections used to create the model, comparing them with the residuals considering the NewBR model. In general, the travel time residuals for the different events have an average close to zero, as shown in the histogram of the individual residuals presented in the same figure. The BRA23 velocity model has a final RMS value of 2.462 s, which is slightly better when compared to the RMS value of 3.390 s for the NewBR model.

Locatable detections (2014 - 2021)

A "locatable detection" is an automatic event generated by the software, whether it is real or fake. It is represented by a set of picks that presented acceptable quality parameters, which passed through all the tests and checks carried out for an event to be created. Each locatable detection was visually inspected, in a qualitative way, to define whether it has good picks (*good picking*), poorly positioned picks, but which still refer to a wave arrival of a real seismic event (*poor picking*), or noisy picks unrelated to seismic events (*noise picking*).

Figure 7 shows the locatable detections obtained through the original SeisComP with data from 2014 to 2021.

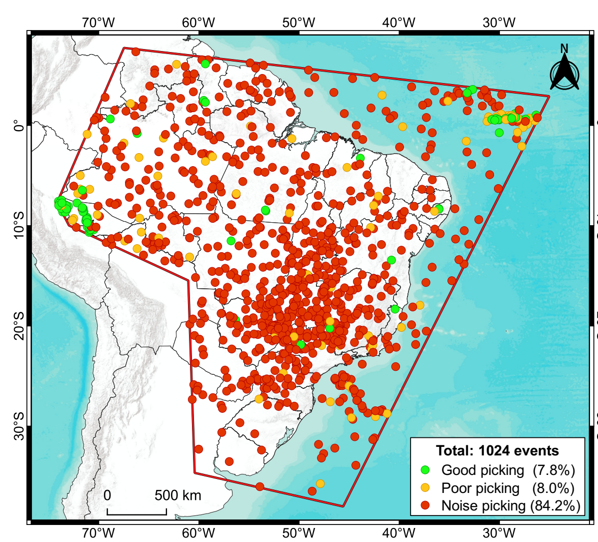

Figure 7 – Locatable detections (2014-2021) obtained using the original SeisComP.

Of the total of 1024 locatable detections from 2014 to 2021, only 80 (7.8%) were evaluated as satisfactorily located seismic events, which are mostly associated with events with plate tectonics boundaries (deep earthquakes in the state of Acre and seismicity in the Mid-Atlantic Ridge), in addition to events with greater magnitudes in Brazilian territory. Also noteworthy is the fact that 862 detections (84.2%) were considered false, associated with the nucleation of local noises.

To evaluate the performance of the modified SeisComP, Figure 8 presents the same map as in Figure 7, but with the locatable detections obtained through the software after the proposed modifications.

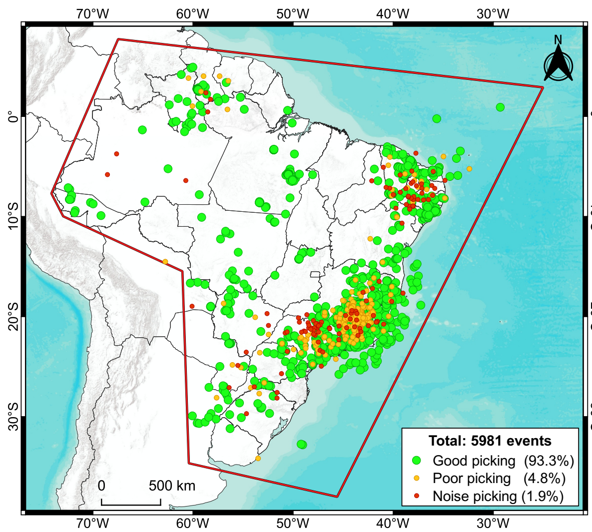

Figure 8 – Locatable detections (2014-2021) obtained using the modified SeisComP.

Of the total of 5981 locatable detections from 2014 to 2021 using the modified SeisComP, 5580 (93.3%) were considered real events, with good detections of wave arrivals and, consequently, good epicentral locations ("*good picking*").

A total of 289 detections (4.8%) were categorized as "*poor picking*", *i.e.*, real events, but with some incorrect picks which possibly affected their epicenters. Finally, 112 detections (1.9%) were classified as false positives ("*noise picking*"), that is, not related to real seismic events and with picks associated with local noises at each station.

Events concurrent with the RSBR catalog

In order to verify the concordance of the locatable detections obtained in both scenarios (modified and original SeisComP) with the events included in the RSBR seismic catalog, we sought to find the events concurrent to the catalog following criteria based on the origin times and coordinates. Events concurrent to the RSBR catalog are the ones that present up to 15 seconds of difference in the origin times and a maximum distance of 100 km between the epicenters.

Figure 9 presents the maps with the locatable detections obtained through the original and modified SeisComP that are concurrent to events in the RSBR seismic catalog.

As expected, the original SeisComP was able to detect, for the most part, only seismic events that were also automatically detected by RSBR. There were 78 concurrent events in the RSBR catalog, only 2 of which were originally located manually. Most of these events are of plate tectonics boundaries origin, such as the deep earthquakes in Acre and those in the Mid-Atlantic Ridge. Some events of greater magnitude in Brazilian territory were also detected automatically.

In total, there are 292 concurrent events to the catalog, 65 of which were automatically detected by RSBR and 227 were located only after manual analysis of the recorded data. The events in the Northeast Region stand out, most of which were automatically detected only after the proposed modifications. There is also a significant increase in automatic detections in the Southeast and Midwest regions, in addition to events in the North of the country and the 2021 seismic events occurred in Guyana.

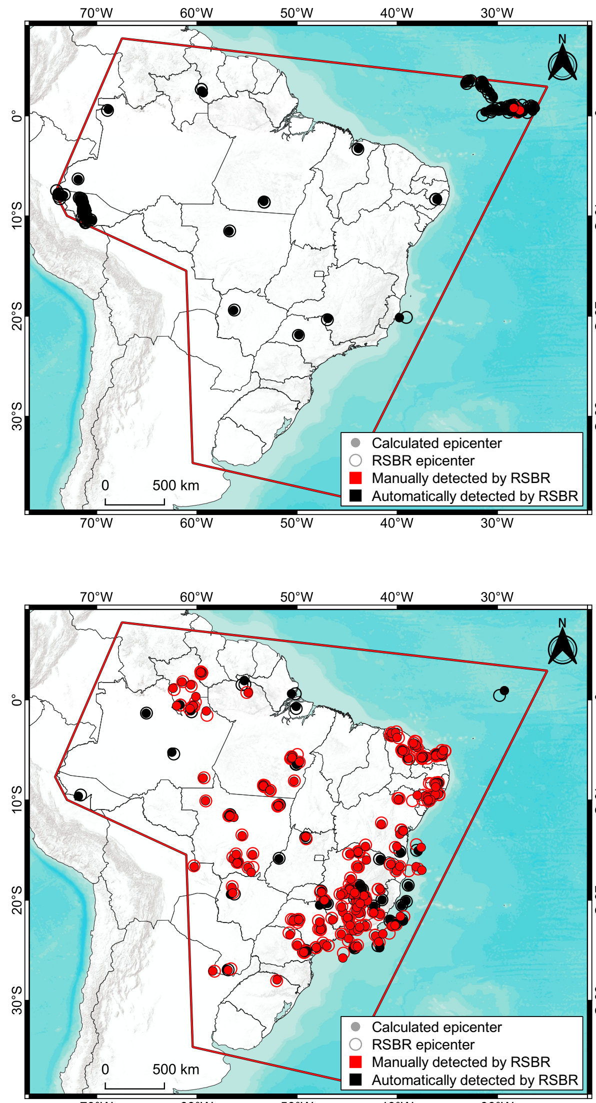

Figure 9 – Above: Epicenters of the 78 locatable detections concurrent to the RSBR seismic catalog from 2014 to 2021, using the original SeisComP. Below: Epicenters of the 292 locatable detections concurrent to the RSBR seismic catalog from 2014 to 2021, using the modified SeisComP.

Estimates of automatic detectability after modifications

Combining the medians of the maximum distance at which RSBR manually located events were picked with the density of seismographic stations in Brazil in the year 2021, it was possible to estimate the current automatic detectability of seismic events in the region of interest, based on their magnitudes.

To check the consistency between the estimates of automatic detectability and the results obtained, Figure 10 presents the estimates of automatic detectability with the overlapping of events from 2014 to 2021 concurrent to the RSBR catalog, detected after the modifications to SeisComP. The distributions of concurrent events are in accordance with the regions where the highest detection probabilities were estimated.

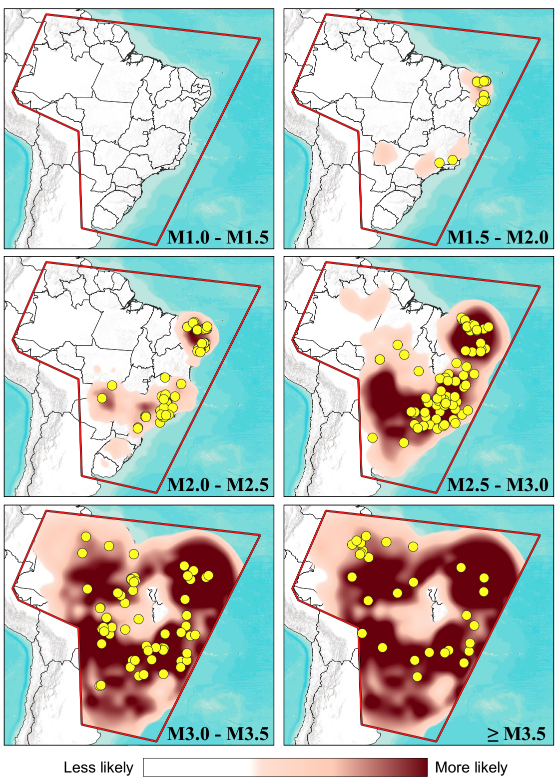

Figure 10 – Estimates of automatic detectability of seismic events in Brazil, with the overlapping of events concurrent to the RSBR catalog obtained after modifications in SeisComP. Qualitative color scale indicates probability of automatic detection.

Conclusions

Although several parameters are modifiable through SeisComP's configuration module, some are arbitrarily changed in the software source code, not actually being used as the end user initially defined them. Furthermore, there are also several processing flows in the source code that impair the detection and location of regional and low magnitude events, which need to be evaluated and changed, recompiling the software at the end of the process.

The analyses carried out in this work indicated as optimized parameters for wave arrival detection, a band-pass filter with cut-off frequencies of 4.5 Hz and 10 Hz, together with an AIC picker and time windows of 0.2 s (STA) and 45 s (LTA), aiming to enhance the P-waves of the events. Regarding *scautoloc*'s hard-coded processes, new ways of verifying the validity of nucleated origins were established, in addition to the removal of arbitrary parameters and flows that hinder the automatic location of regional events.

The development of a new 1D velocity model aiming at minimizing the residuals of well-known Brazilian regional events was an important factor to allow the automatic detection of regional earthquakes in Brazil. The new velocity model proposed (BRA23) was based on the NewBR model and more recent regional events, and was obtained through parameter optimization processing routines, implemented in *Python*. The optimized parameters are compatible with results in the literature and allowed the reduction of the RMS of the events, compared to the NewBR model.

The proposed modifications allowed an increase of almost 600% in the total amount of locatable detections in the region of interest, which jumped from 1024 (with the original SeisComP) to 5981 during the period from 2014 to 2021. The satisfactory locations increased from 80 (out of 1024) to 5580 (out of 5981) after the modifications. Likewise, origins associated with noise decreased from 862 (out of 1024) to 112 (out of 5981).

The results obtained in this work indicated that SeisComP can be used more efficiently in a context of regional seismicity, with relatively low magnitude events. In order to obtain a greater effectiveness of the system, it is necessary to make the proposed adjustments based on the region of interest and its data availability and quality. It is possible to reproduce the proposed modifications considering any region of the world. In the Brazilian seismological monitoring scenario, the adoption of the modified version of SeisComP could considerably facilitate the work of those responsible for manual reviews of data generated by the software, in addition to allowing the detection of events that could be overlooked in these reviews.

References

Assumpção, M., Ardito, J. C. & Barbosa, J. R., 2010. An improved velocity model for regional epicentre determination in Brazil, IV Simpósio Brasileiro de Geofísica: 13–16.

Behr, Y. et al., 2016. The Virtual Seismologist in SeisComP3: A New Implementation Strategy for Earthquake Early Warning Algorithms, Seismological Research Letters, vol. 87: 363., doi:10.1785/0220150235.

Bianchi, M. et al., 2018. The Brazilian Seismographic Network (RSBR): Improving Seismic Monitoring in Brazil, Seismological Research Letters, vol. 89(2A): 452., doi: 10.1785/0220170227.