University of São Paulo
Institute of Astronomy, Geophysics and Atmospheric Sciences
Geophysics Department

Victor Salles Araujo

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

São Paulo
2023

VICTOR SALLES ARAUJO

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

Thesis presented to the Department of Geophysics of the Institute of Astronomy, Geophysics and Atmospheric Sciences of the University of São Paulo, as a partial requirement for obtaining the title of Master of Science.

Area of concentration: Geophysics.

Supervisor: Prof. Marcelo B. de Bianchi.

São Paulo
2023

NON DVCOR DVCO

Abstract

SALLES, V. **Optimization of the SeisComP software for automatic detection and location of regional seismic events in Brazil.** 2023. Master's thesis – Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, São Paulo.

The Brazilian Seismographic Network (RSBR) began operating in 2009 and is currently composed of almost 100 broadband seismographic stations with the capacity for real-time data transmission. The automatic monitoring of seismic events is carried out using the SeisComP software, with few modifications in its original detection and location parameters. More than 90% of the events in the RSBR seismic catalog were detected manually by analysts in daily seismogram analysis routines. Aiming to improve the automatic detectability of seismic events in Brazil, we analyzed the automatic detection capacity of the RSBR based on its seismic catalog. Then, a study was conducted to optimize the detection parameters (frequency filters and STA/LTA windows, limits for detection, and picker) using known events. We also sought to optimize the origins' nucleation and association procedures, including those that are part of the software source code, in order to maximize the detection of real events and minimize false positives. Finally, a new velocity model was proposed, based on the NewBR model and using automated processing routines to find the best set of parameters that minimize the RMS of the events used to obtain the model.

As optimized parameters, the analyzes indicated a band-pass filter with cutoff frequencies of 4.5 Hz and 10 Hz, an AIC picker, and time windows of 0.2s (STA) and 45s (LTA) to maximize P-wave arrivals from regional events. Regarding the source code of the *scautoloc* module, new ways of verifying the validity of nucleated origins were established. Furthermore, parameters and arbitrary flows that hinder the automatic location of regional events were removed. We also propose using grid points whose minimum number of wave arrivals and maximum distances for nucleation depend on the density of stations within a radius of 10 degrees from each grid node. The results indicate that the implemented modifications improve the automatic detection of seismic events in Brazil: the detections that can be located increased from 1024 to 5981 between 2014 and 2021. Of these detections, the events also present in the RSBR seismic catalog increased from 78 to 292, in addition to real events that are not in the official catalog. The origins' parameters showed significant improvements, such as the decrease in the travel-time residuals of wave arrivals, the possibility of nucleation with the contribution of a few stations, and the detection of events with low magnitudes ($\sim$M1.5). It is also worth mentioning that, after the modifications, it is possible to detect blasts in regions of known mines and quarries.

Keywords: Seismology, SeisComP, Automatic Detection, Regional Seismicity.

Contents

List of Abbreviations and Terms

AIC - *Akaike Information Criterion*. Algorithm used to refine detections of amplitude anomalies performed by the STA/LTA ratio algorithm, acting as a picker. It cannot be used as a detector since it does not process continuous data.

Arrival - Wave arrival detection associated with a seismic event origin.

Locatable detection - An origin automatically calculated by SeisComP but not yet reviewed manually for validation. A "locatable detection" indicates that the origin has passed all the software's tests and criteria, and is defined by the software as a real seismic event. However, to confirm that it is a seismic event, validation by a seismologist is required.

Grid - Grid used for nucleating events. Initially, the hypocenters of nucleated origins are defined at one of the points discretized by the grid and are subsequently refined through relocations.

Hard-coded - Parameter or process defined in the software's source code, not accessible to the user. To modify something hard-coded, it is necessary to edit the source code and recompile the software.

Origin - Set of parameters that determine the origin of a seismic event (geographic coordinates, origin time, arrivals, etc.).

Preferred origin - A seismic event may contain different origins due to, for example, the addition of new picks or manual refinement. The preferred origin is the one chosen to be the representative of the event, reported in seismic bulletins and catalogs.

Pick - Detection of amplitude anomaly (transient signal) in the seismogram. A pick can correspond to a wave arrival from seismic events or a high-amplitude noise that stands out from the local background vibrations.

Picker - Algorithm used to perform picks on the seismogram.

***RMS* -** *Root mean square*.

SNR - *Signal-to-noise ratio*. Signal-to-noise ratio of a detection on the seismogram (pick).

STA/LTA - Short-term average through long-term average. Algorithm used to detect amplitude anomalies (transient signals) on the seismogram. It can be used as a "detector and picker" or only as a "detector". In the second case, the STA/LTA algorithm provides a time window for the detection (pick) to be performed by a more robust picker (*e.g.*, AIC).

List of Figures

List of Tables

1. 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 1.1.

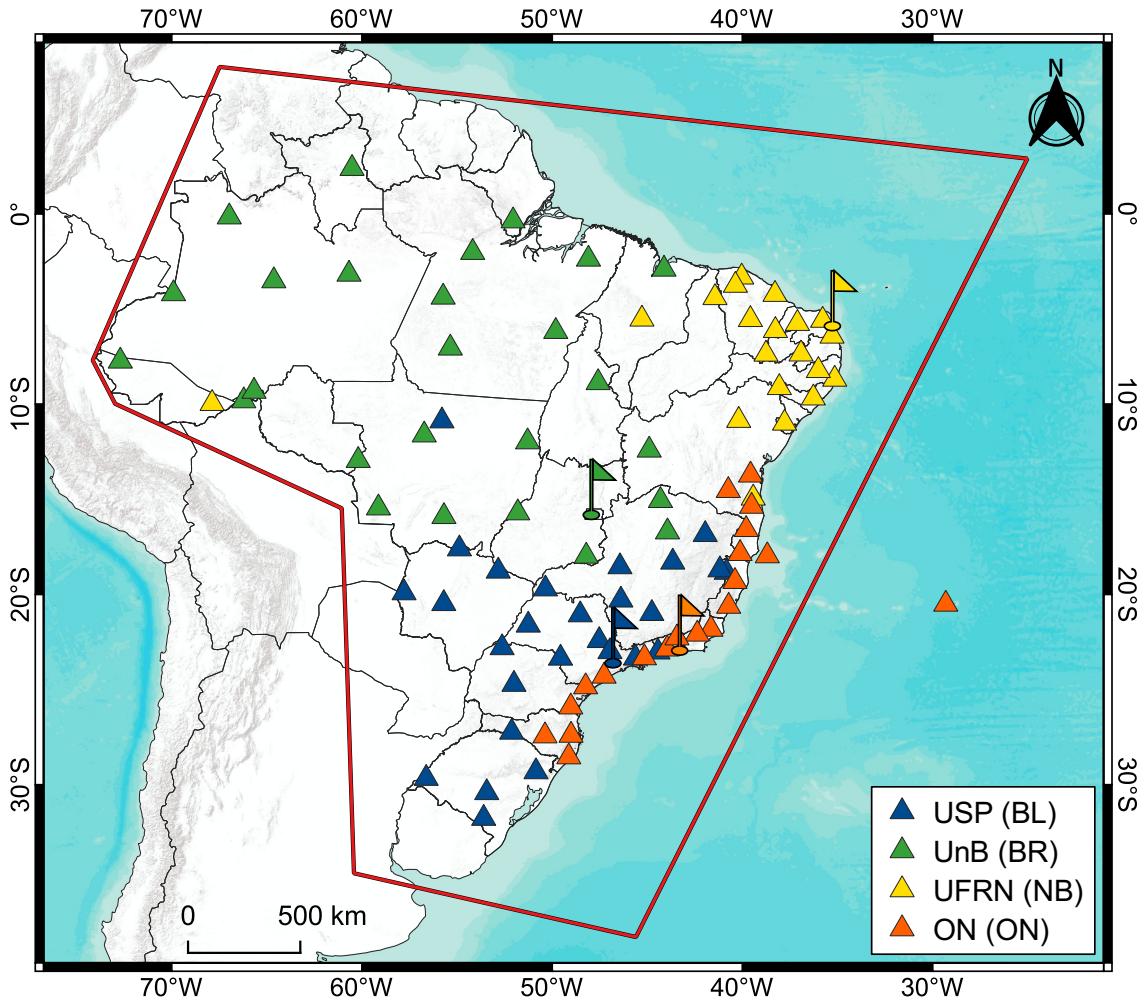


Figure 1.1: Seismological stations (triangles) and institutions (flags) participating in RSBR. Subnetworks and their respective responsible institutions are indicated by different colors. Source of information: IAG-USP database (October/2022).

Each seismological center is responsible for the installation, operation and maintenance of seismographic stations in certain regions of the national territory. The Brazilian network is composed of four subnetworks (BL, BR, ON and NB), which present variations regarding the instrumentation sets and technologies, but maintain a previously agreed minimum standard (Bianchi et al., 2012). In addition to the permanent stations, there are temporary stations that are operated by each group and act in a complementary way to RSBR.

Currently, the Institute of Astronomy, Geophysics and Atmospheric Sciences of the University of São Paulo (IAG-USP) performs the location of seismic events for RSBR, while the National Observatory (ON) carries out the disclosure of these events. All events recorded by RSBR stations undergo a manual review, carried out by IAG, to define their "preferred origin". The revised data is sent to the International Seismological Centre (ISC) to compose its regularly published bulletin, together with quality parameters of the networks.

Over time and with the installation of new RSBR seismographic stations, the detectability of seismic events has improved. Because of that, Assumpção et al. (2014) inferred the possibility of detecting earthquakes in the national territory with magnitudes above M3.5, after manual review of the data, from the year 1980 onwards. On the other hand, Bianchi et al. (2018) points out that, although this value is adequate for most of Brazil, it is probably overestimated for the Amazon region, where there are fewer seismographic stations in operation.

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, as shown in Figure 1.2. According to the author, the installation of RSBR stations in the Amazon region in 2014 allowed the detection of events with magnitudes as low as M3.5 for all of Brazil.

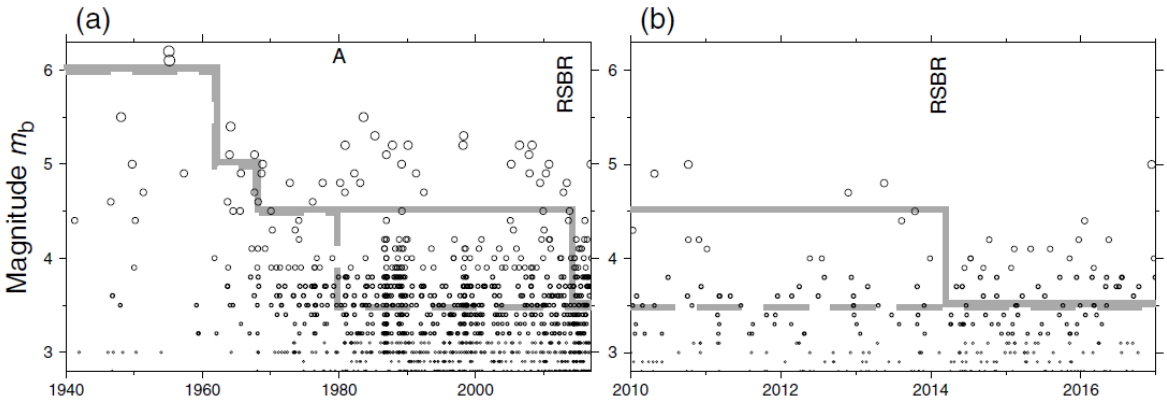


Figure 1.2: Temporal distribution of magnitudes from the RSBR catalog. (a) Data between 1940 and 2016, (b) Expansion between 2010 and 2016. "A" indicates the increase in the number of stations installed in Brazil, especially in the Northeast, Southeast and Amazon. "RSBR" indicates the installation of RSBR stations in the Amazon in 2014. The dashed gray line indicates detection limits inferred by Assumpção et al. (2014), while the continuous gray line indicates detection limits inferred by the author. Source: Bianchi et al. (2018).

Bianchi et al. (2018) points out that, 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 (Figure 1.3). For some regions (Southeast and Northeast), where the density of stations is higher, the detection limit is M3.0 or even lower.

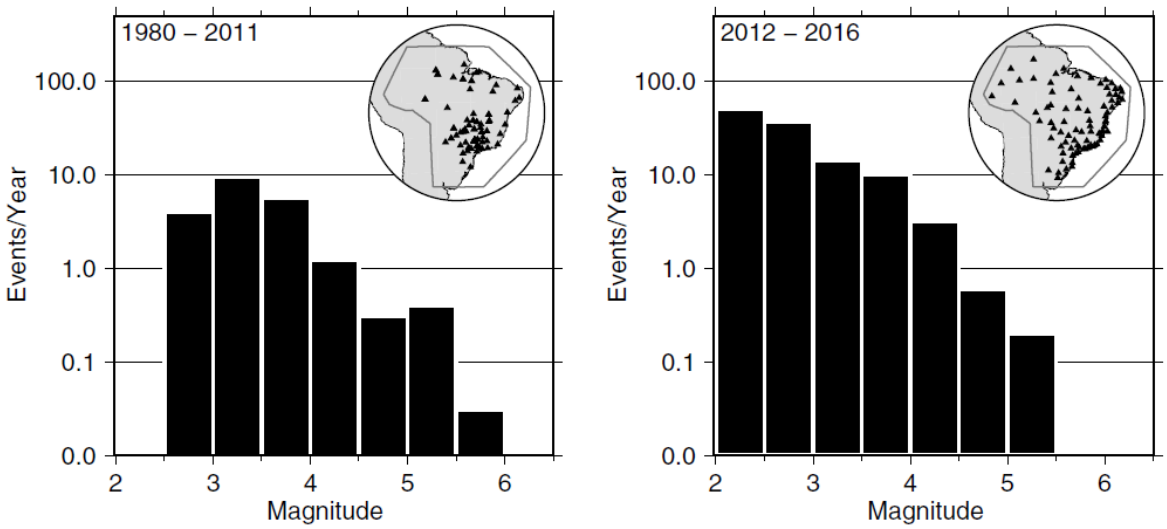


Figure 1.3: Annual average of the number of seismic events by magnitude included in the RSBR catalog, from 1980 to 2011 (left) and from 2012 to 2016 (right) - before and after the beginning of the installation of RSBR stations, respectively. In detail, the main stations operating in each period. Source: Bianchi et al. (2018).

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

Lopez (2021) performed an optimization of the wave arrival detection (picking) parameters for part of the RSBR stations, but without considering the continuous records, that is, using only snippets of waveforms that contain known seismic events. According to the author, the optimizations increased the number of automatic picks, contributing to the increase in the automatic location of events in the selected time windows. However, there is also an increase in the amount of picks associated with noise, which impair the nucleation of events in a context of real-time processing.

In tests carried out using the parameters proposed by Lopez (2021) in continuous waveform records, it was not possible to locate seismic events due to the large number of "false picks" generated. This indicates that, to optimize the location of seismic events in real time, the new parameters must be obtained considering all continuous records and not only time windows that contain known events.

2. Motivation: RSBR seismic catalog

The International Federation of Digital Seismograph Networks (FDSN) regulates the operation of seismological networks around the world, in order to standardize the storage and distribution of parametric and waveform data (Clark et al., 2014). It is responsible for issuing unique and individual network codes, as well as coordinating cooperation between international seismology groups.

RSBR registers its network codes with the FDSN and adheres to its rules, seeking to distribute its data using such standards. Waveform data stored by RSBR is in "SEED" format (Standard for Exchange of Earthquake Data) (IRIS, 2010), while event metadata is stored in *QuakeML* format (Schorlemmer et al., 2011), described below. Access to data and metadata is carried out in a simple and standardized way through the FDSN Web Services (FDSNWS), an interface for requesting data from seismological centers that adhere to the services offered by the FDSN.

With this, it is possible to obtain the necessary information to evaluate the effectiveness of the automatic detection system of seismic events in Brazil. Figure 2.1 presents two maps with the events that occurred between 2014 and 2021 in the region of interest, as well as the magnitude histograms considering the detection modes.

The seismic events included in the RSBR catalog have different magnitude scales (*e.g.*, MLv, m_R, mb, Mw). Each scale is adopted according to the epicentral parameters of the events and is based on different aspects of the seismogram, such as P-wave amplitudes, event duration, among others (Duda and Nuttli, 1974; Kanamori, 1983). In Brazil, local events have only the local magnitude (MLv) computed, while regional events use the magnitude m_R (Assumpção, 1983) which, in the context of South America, is equivalent to the teleseismic mb scale. In this work, for comparison purposes, the magnitudes of all seismic events are referred to as M, regardless of the type of associated magnitude.

Most of the confirmed earthquakes that were automatically detected and located in the region of interest have origins associated with plate boundaries, either in the region of the state of Acre (due to the subduction of the Nazca Plate under the South American Plate) or in the region of the Mid-Atlantic Ridge. Considering only intraplate events, 1717 earthquakes were recorded during the period, of which only 106 (6.2%) have at least one automatically calculated origin. The main aspects that can cause the low effectiveness of the automatic detection system are *(i)* the average spacing between RSBR stations, which prevents a low magnitude seismic event from being recorded by several seismographs and *(ii)* parameters and processing flows of the system used for automatic detection, which must be configured and optimized considering the context of RSBR.

To improve the first highlighted fact, it is necessary to add new stations to RSBR. For this, it is essential that financial investments be made, in the short and long term, for the installation and maintenance of such stations, in order to provide high quality data records for long periods of time. Considering the second aspect, the automatic detection of seismic events can be improved even with the current set of stations in the network. For this, it is necessary to evaluate and restructure the various parameters and processes carried out by the software that manages and processes the data registered by RSBR, SeisComP (Helmholtz-Centre Potsdam - GFZ German Research Centre for Geosciences and gempa GmbH, 2008).

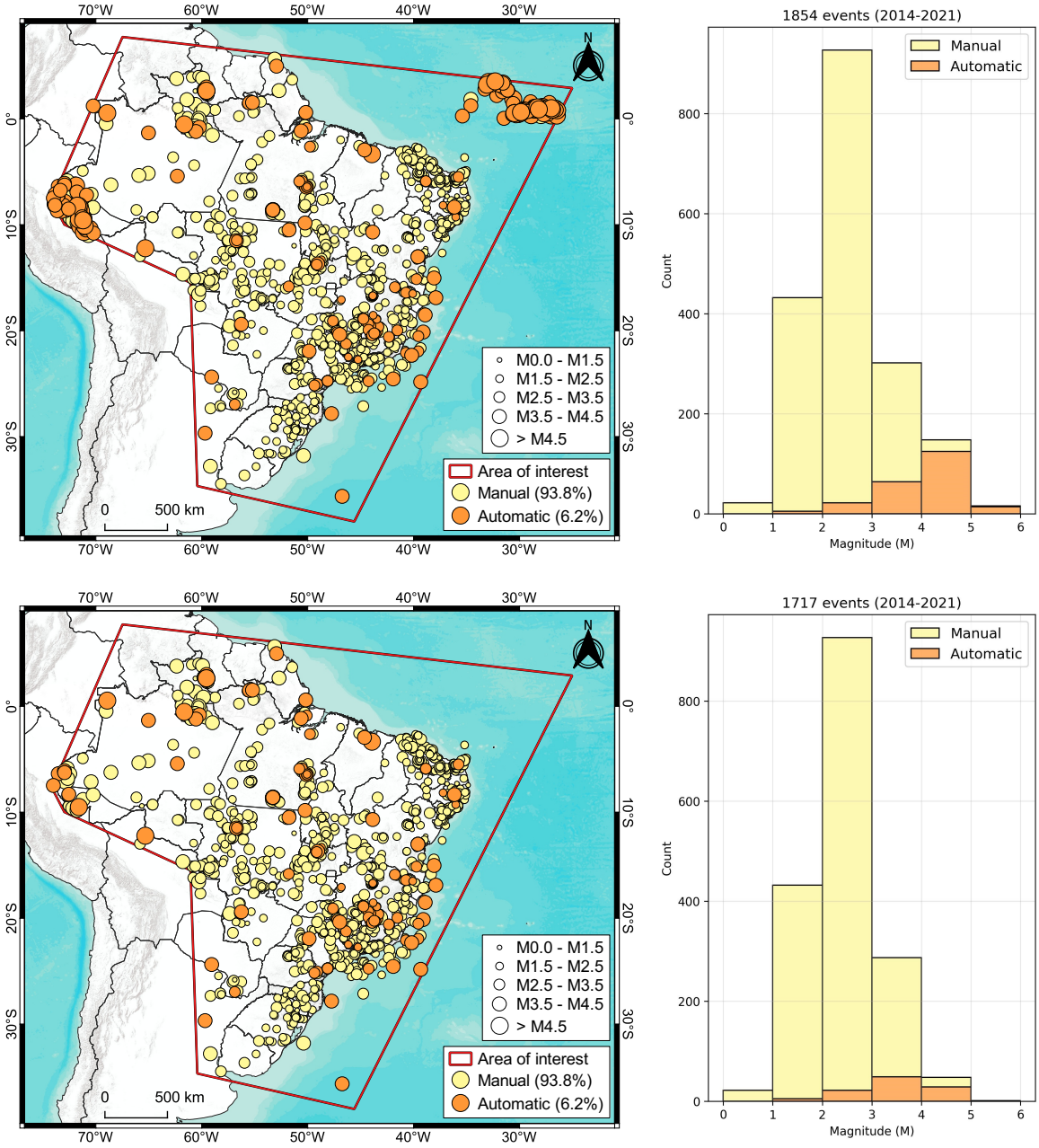


Figure 2.1: All seismic events that occurred between January 2014 and December 2021 in the region of interest (above) and excluding deep events associated with plate boundaries (below). Source of information: IAG-USP (October/2022).

An estimate of the maximum detectability that can be achieved with the current number of RSBR stations can be inferred by analyzing the catalog of the IAG-USP Seismological Center. For this, the graph on the left in Figure 2.2 shows the relationship between magnitude and the number of arrivals of the preferred origins of all seismic events recorded between January 2014 and December 2021 in the region of interest shown in Figure 2.1. On the right, the same relationship is showed, however, considering only the wave arrivals of the last automatic origins of the events (if any).

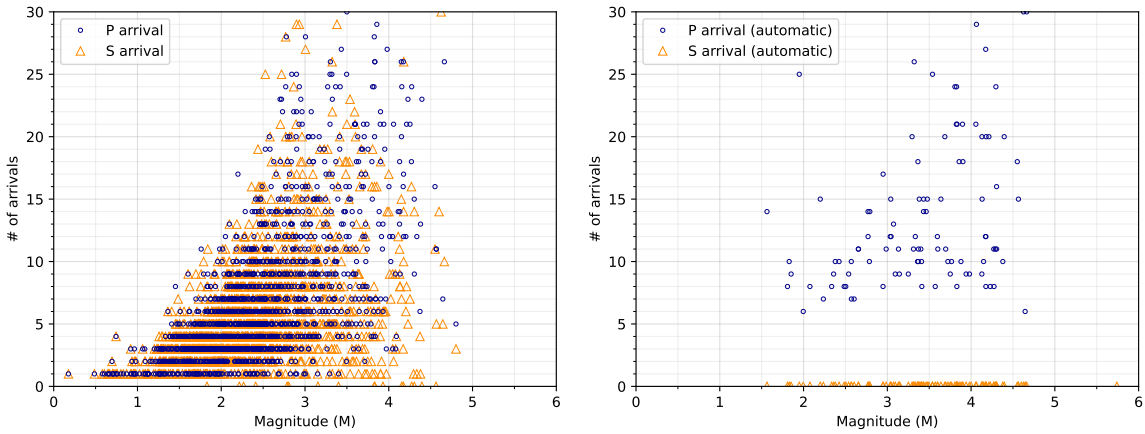


Figure 2.2: Left: relationship between magnitude and quantity of arrivals of preferred origins of seismic events occurring between 2014 and 2021 in the region of interest. Right: same relationship, but using only the last automatic origins of events, if any.

By default, SeisComP needs at least 6 P-wave arrival detections at different stations to be able to nucleate an event automatically. For manual processing, the software may be able to find events with fewer stations, as shown in the graph on the left in Figure 2.2.

By analyzing the graph on the right, it is possible to see that the events located automatically have, for the most part, magnitudes greater than M2.5. However, when comparing it with the graph on the left, it is possible to verify that several events with magnitudes lower than M2.5 were not automatically detected by SeisComP, even if they presented 6 or more detections in their preferred origins. This fact may indicate the failure of the system to automatically identify wave arrivals from real seismic events, currently requiring manual verification of waveform data for these events to be located.

It is also important to highlight the large number of events with relatively high magnitudes that have few detections, which possibly indicates an error in the calculation of their magnitudes (*e.g.*, events with magnitudes greater than M3.0 and with only one P-wave arrival detection).

Another way to evaluate the detectability threshold of seismic events in Brazil is by analyzing the relationship between magnitude and maximum recorded distance, since the density of stations in the national territory is not uniform and, for this reason, events that are not recorded at large distances tend not to be located because they do not present wave arrivals at several stations.

Bianchi et al. (2012) estimates that a seismic event of magnitude M2.5 is recorded at a maximum distance of 150 km, while an event of magnitude M3.5 is recorded at a maximum distance of 500 km. Finally, according to the author, an event of magnitude M4.0 can be recorded at a distance of up to 1200 km.

To verify Bianchi et al. (2012)'s estimates, Figure 2.3 shows the relationship between magnitude and maximum recorded distance, obtained through data of all seismic events from the RSBR catalog, which occurred between 2014 and 2021 in the region of interest.

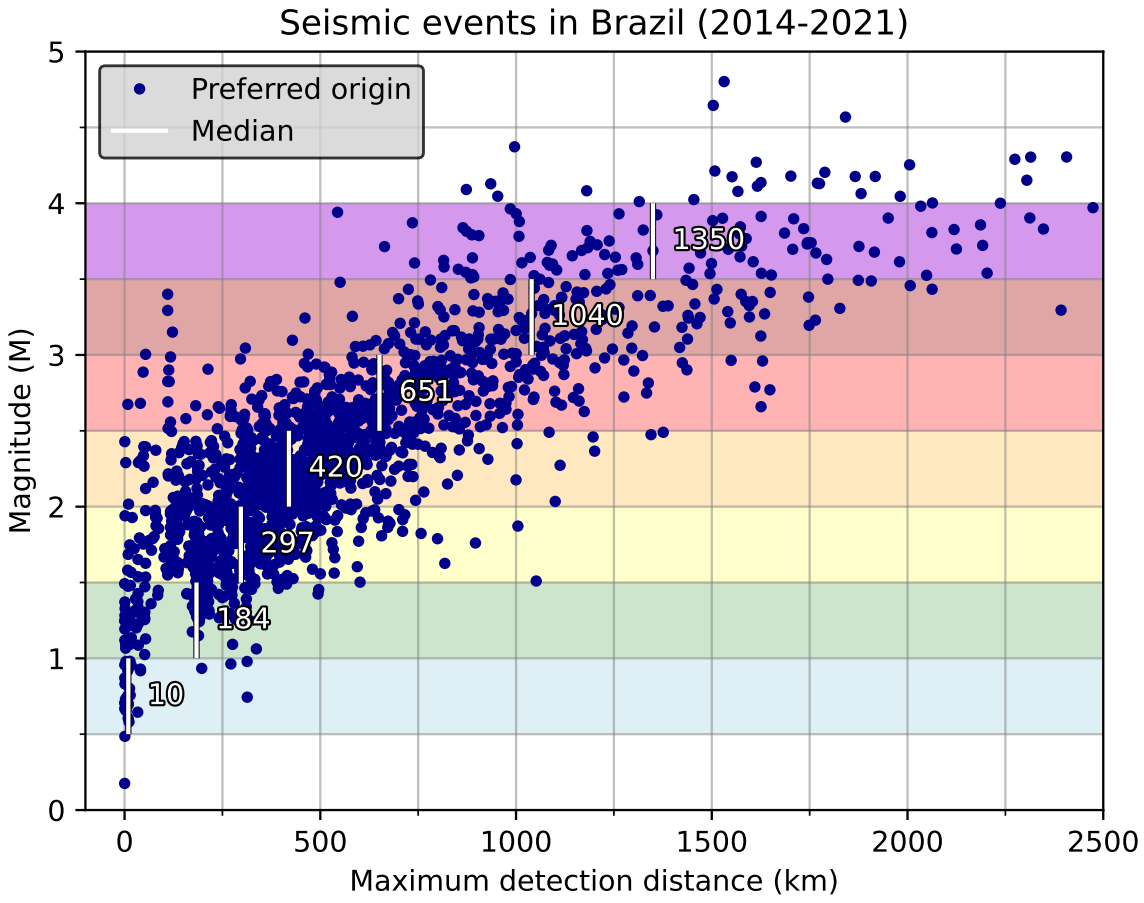


Figure 2.3: Relationships between magnitudes of the 1717 seismic events that occurred between 2014 and 2021 and the maximum distances at which they were recorded. The vertical white lines indicate the medians of distances for each magnitude range between M0.5 and M4.0.

The median of maximum distances for recording seismic events with magnitudes between M1.5 and M2.0 is 300 km, while seismic events with magnitudes between M2.0 and M2.5 have a median value of 420 km. Therefore, the graph indicates that seismic events in Brazilian territory are detected at greater distances than Bianchi et al. (2012) estimated. This data must be considered when defining the maximum distances at which seismological stations can contribute wave arrival detections, aiming to mitigate the addition of possible noise in very distant stations, in which records of a given event were not expected.

Additionally, it is convenient to measure the distances between the stations that make up RSBR subnetworks. For this, Figure 2.4 shows two boxplots: the first refers to all distances between stations in each subnetwork, while the second considers only the shortest distances for each station.

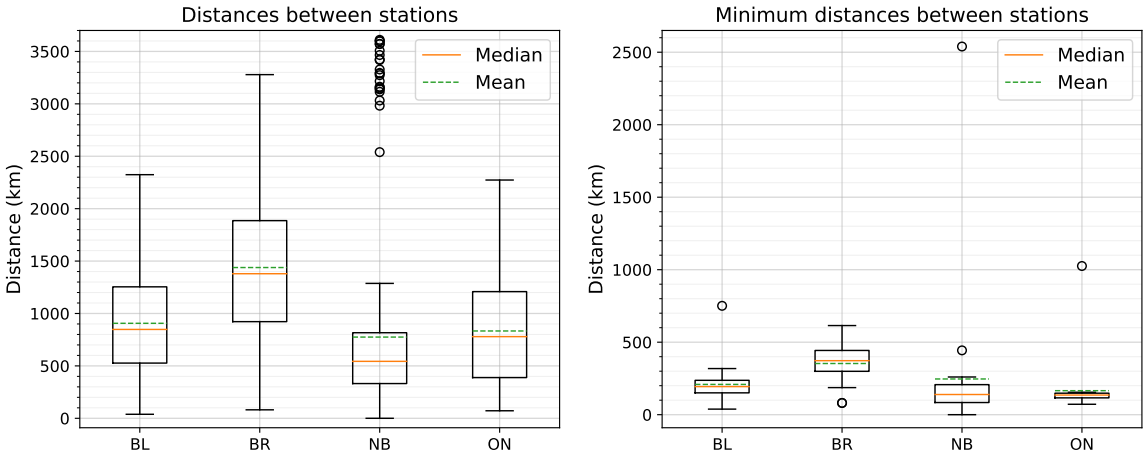


Figure 2.4: Boxplots of distances between RSBR stations. Left: all distances between stations in each subnetwork. Right: shortest distances between two stations of each subnetwork.

Table 2.1 shows the medians of the distances between stations on RSBR subnetworks, graphically presented in Figure 2.4.

Table 2.1: Medians of distances between stations on RSBR subnetworks. "Total" refers to the medians of the distances between all pairs of stations on the same subnetwork, while "Minimum" indicates the medians of the shortest distances between two stations on the subnetwork.

	BL	**BR**	**NB**	**ON**
Total (km)	906	1379	543	779
Minimum (km)	194	373	139	135

When considering RSBR as a whole, there is a median of approximately 159 km of minimum distance between stations. Comparing this result with the graph in Figure 2.3, it is suggested that the current RSBR set should be able to automatically detect more seismic events with magnitudes lower than M3.5. When analyzing subnetworks individually, it is inferred that BL, NB, and ON must have similar detection capabilities, while BR, responsible for covering the North Region and part of the Midwest, presents stations that are more widely spaced and is, consequently, less capable of detecting events with smaller magnitudes.

Thus, this work aimed to optimize parameters and procedures performed during the automatic detection and location of seismic events in the context of RSBR, improving the efficiency in the automatic detection of regional seismic events in Brazil.

3. SeisComP

In the first stages of establishing RSBR, IAG-USP focused its efforts on finding software solutions that would facilitate cooperation among different universities and institutes in the operation of the seismographic network. The software chosen for real-time acquisition and processing was SeisComP, developed at the German Research Center for Geosciences (*GeoForschungsZentrum* - GFZ), which has some versatility, such as the possibility of improving tools and developing new solutions (Pirchiner et al., 2011). The software version currently used by the USP Seismological Center and explored in this work is "*SeisComP3 Jakarta 2020.330.06*", however, subsequent versions have not undergone changes in the processing modules. Thus, it was found that the latest version currently available (*SeisComP 5.3.0*) can also be optimized using the same procedures described here.

SeisComP's operation 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 3.1.

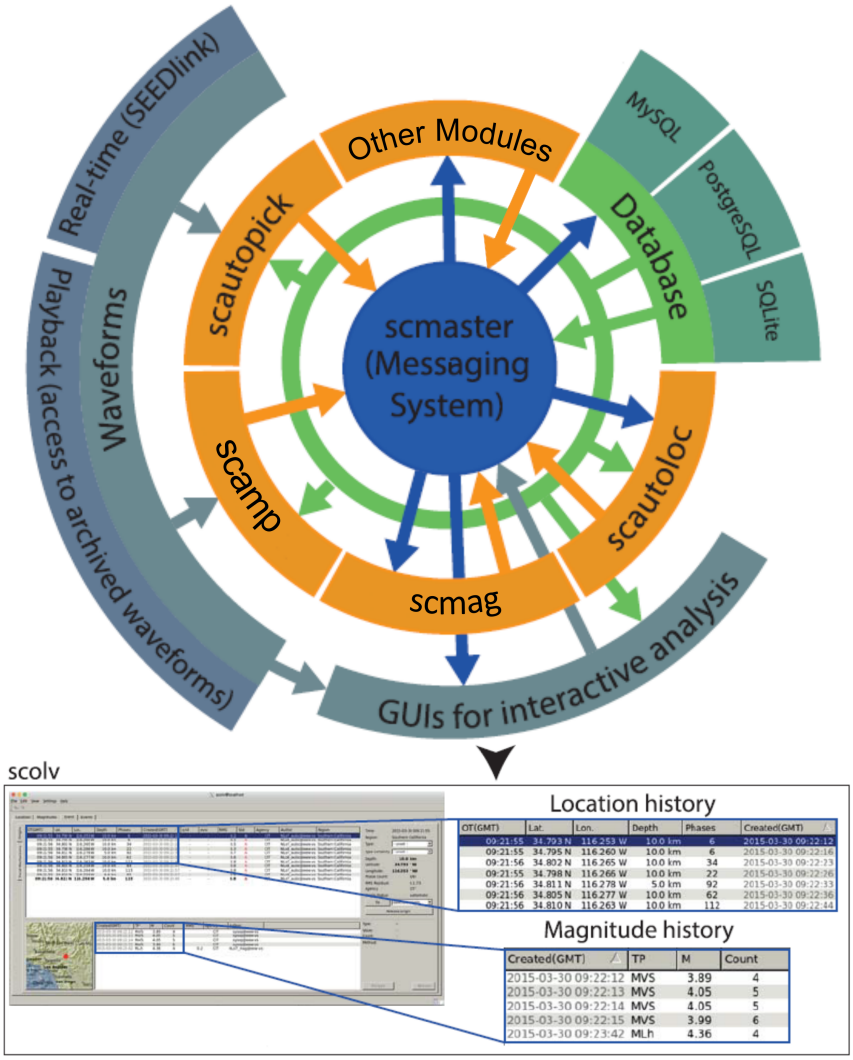


Figure 3.1: Operation flowchart of SeisComP and its modules. The arrows indicate the direction of the information flow, colored according to those responsible for sending it. Source: adapted from Behr et al. (2016).

Figure 3.1 shows that the detection, location and cataloguing of seismic events are carried out through a succession of operations performed by the software modules. The modules represented by the color orange are processing modules, coordinated by the control module (*scmaster*). Data are read by all modules and written only by the control module to different supported database types. Interactive analysis and review of the data can be performed through the *scolv* module, a graphical user interface (GUI) that provides access to the database and waveforms, among other information.

Some of the main processing modules and their brief descriptions are:

- ***scautopick*:** looks for transient signals in form of amplitude anomalies in the waveform data. It applies a STA/LTA (short-time-average through long-time-average) ratio algorithm (Allen, 1978) which, when it reaches a certain value, creates a wave arrival detection (pick) at the instant of time when this limit is exceeded. Optionally, it can use more robust algorithms to refine the initial detection;
- ***scautoloc*:** responsible for nucleating origins, associating detections and locating seismic events automatically in real time. It performs continuous analysis of picks, trying to nucleate a new origin, that is, to identify combinations of picks that correspond to the same seismic event. If a location consistent with certain criteria is produced, it is reported and passed on to other modules that receive origins as input data;
- ***scamp*:** calculates waveform amplitudes using different time windows, based on previously obtained picks;
- ***scmag*:** calculates different types of magnitudes using the amplitudes calculated by the *scamp* module. At IAG-USP, this module is also configured to calculate magnitudes on m_R scale (Assumpção, 1983); and
- ***scevent*:** associates a previously calculated origin to a known event or, if compatibility is not found, creates a new event and associates the origin to it, provided that the established criteria are met. It is important to point out that a single seismic event can contain several calculated origins, since new origins can be obtained from arrivals processed after the first nucleation.

3.1 Processing modules: *scautopick* and *scautoloc*

The *scautopick* and *scautoloc* modules 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. Figure 3.2 presents a flowchart of the main aspects of both modules, as well as the relationship between them.

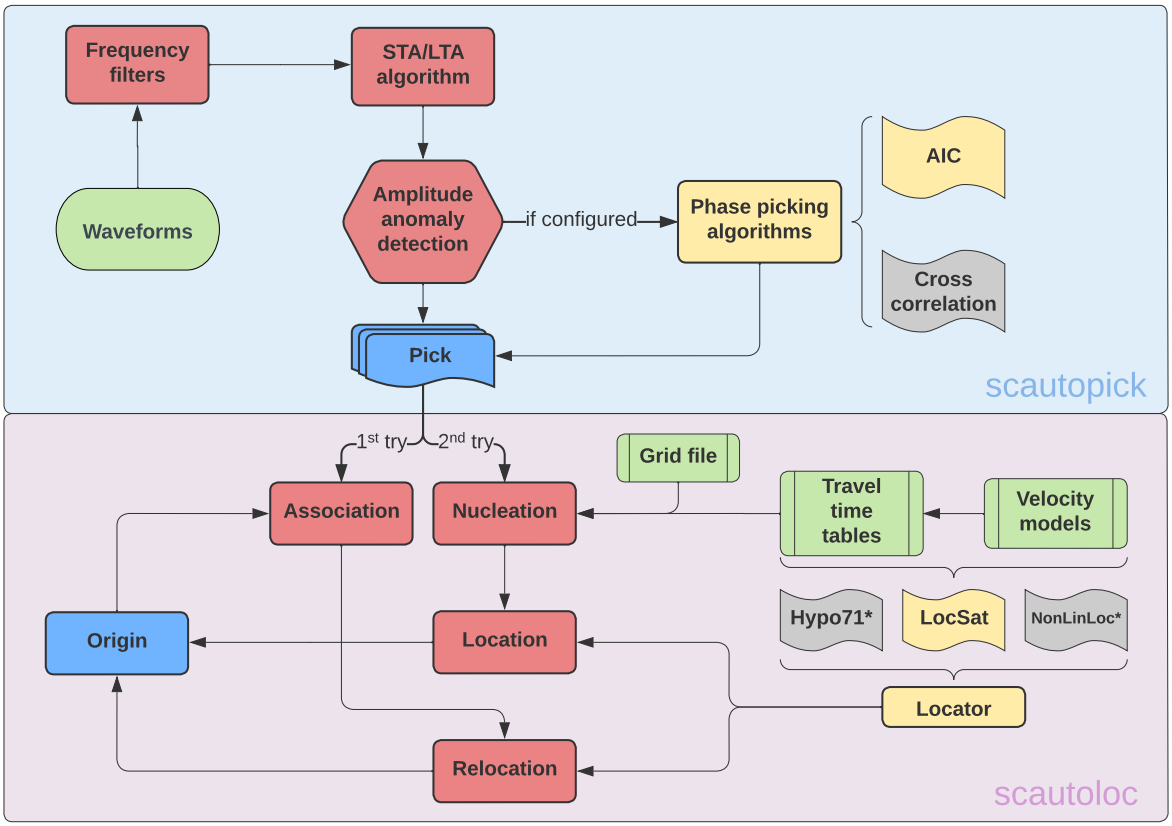


Figure 3.2: Flowchart of the main aspects of *scautopick* and *scautoloc* modules. In green, input data required for processing are represented. In red, the main processes carried out by each module are shown. In yellow, algorithms implemented in the software in order to contribute to the detection and location process and, in grey, optional algorithms to be used, if configured. In blue, the products generated by each module are presented. (*) Locators used only when relocating origins.

The waveform data are read by *scautopick*, which applies a frequency filter and then uses an STA/LTA algorithm (Allen, 1978) to detect transient signals (in the form of amplitude anomalies) in the seismograms. When such anomalies are detected, a timestamp (pick) is created and stored to be used when locating the events.

Then, if configured, an AIC algorithm (*Akaike Information Criterion*) (Akaike, 1971) can be used to improve the pick, adjusting the marking in time to coincide with the moment of P-wave arrival at the seismographic station. AIC algorithms are commonly used in seismology

to detect wave arrivals (Li et al., 2016; Yadav and Mishra, 2014), providing more accurate results than those obtained by STA/LTA algorithms by detecting stationarity changes, that is, identifying changes in behavior of the signal within the indicated window.

The pick is received by the location module (*scautoloc*), which initially tries to associate it with a known origin. If this is not possible, the module uses this and other unassociated picks to try to nucleate a new origin, identifying combinations that correspond to the same seismic event. This stage is carried out through a grid search, in which all picks not associated with an origin are projected onto all grid points. For each pick and each grid point, the origin time of the event is projected, using the travel time table provided together with the distance from the station that registered the pick to the grid point. In this way, it tries to find possible hypocenters with the lowest associated travel time residuals. This is a computationally costly process, as it performs thousands of calculations involving all picks and grid points.

The main processes carried out by *scautoloc* can be summarized in the following stages, according to the official SeisComP documentation:

- **Association:** initially, *scautoloc* tries to associate an incoming pick with a known origin, especially with origins that already have a large amount of arrivals, in order to improve its locations. If it is not possible to make this association due to incompatibility of criteria, the nucleation process is carried out;
- **Nucleation:** if the association process fails, *scautoloc* tries to create a new origin out of this and other detections previously received and unassociated with an origin. The nucleation process is a grid search, in terms of space and time, on a grid covering the area of interest. This grid can be modified according to the needs of the seismological monitoring. During the grid search, each grid point is considered a possible hypocenter, at the same time that the compatibility of the origin time projections made for each pick is verified, considering the travel time table provided. In this process, the software continuously analyzes, over time, the density of detections at each point. If the set of projected detections meets all previously configured requirements and it is possible to obtain a location, it is accepted as a new valid origin of a seismic event;
- **Origin refinement:** origins created through nucleation or even association may be contaminated by phases wrongly interpreted as P-wave arrivals. In the origin refinement process, *scautoloc* tries to improve the origins based, for example, on the signal-to-noise ratios of wave arrivals and their amplitudes, in addition to carrying out new attempts to associate detections that were not previously associated; and
- **Origin filtering:** a last check of consistency and compatibility with certain criteria of the newly created or updated origins. During this process, the origins are not modified, only discarded or validated.

During the processes carried out by *scautoloc*, the pick goes through several stages, which contain predefined flows and parameters in the module's configuration files and in the software source code. Therefore, to optimize such processing, it is necessary to optimize parameters available for modification by the user, as well as perform adjustments directly in SeisComP's source code and recompile it. It is also extremely important to use a grid appropriate for the region of interest, which is used during the nucleation of new events.

3.2 Configuration module: *scconfig*

SeisComP has a configuration module (*scconfig*) with a graphical user interface. Through it, it is possible to adjust configurations and parameters of all the software modules in an easy and intuitive way.

Both *scautopick* and *scautoloc* modules have their main parameters accessible to the user through *scconfig*, under the path *Modules* ⇒ *Processing* ⇒ *scautopick/scautoloc*, as shown in Figure 3.3.

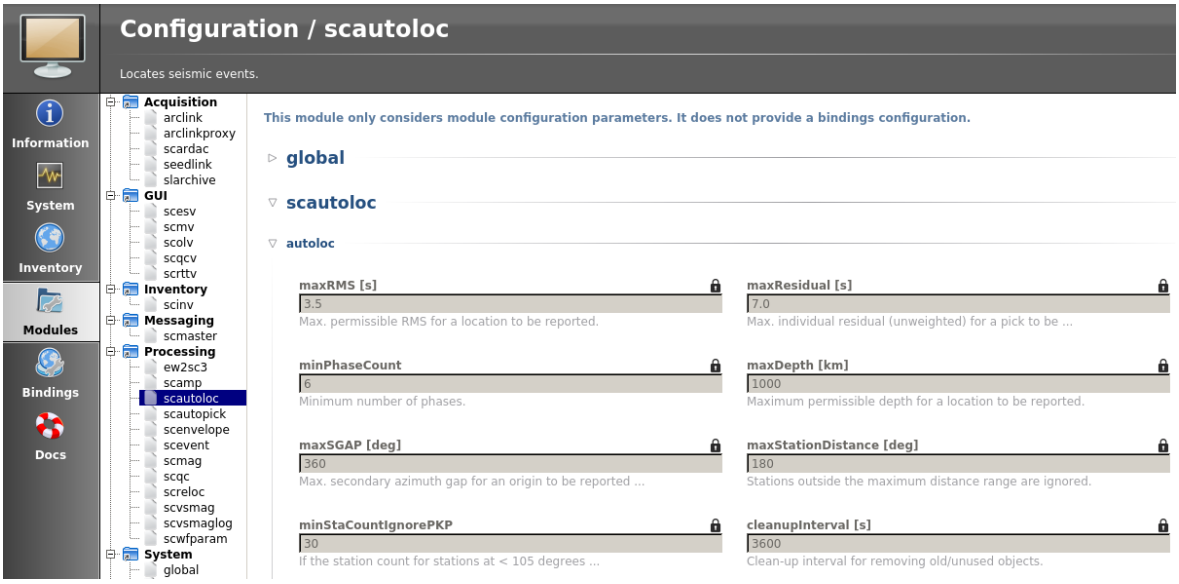


Figure 3.3: Example of graphical user interface of the *scconfig* configuration module, with part of *scautoloc*'s parameter configuration.

The configurations included in these paths of *scconfig* are applied to all seismograph stations within the inventory, affecting all processing performed by both modules. It is also possible to modify these parameters directly in the configuration files of each module (contained in "*/seiscomp/etc/*").

Complementarily, there are station "bindings" , which contain specific information and parameters for each station in the inventory, such as channel codes used (*e.g.*, HH, BH, EH). Through the bindings, individual adjustments are made to global parameters aiming, for example, to reduce the amount of picks generated by a noisy station. *Binding* parameters can be modified through *scconfig* or the individual files of each station, located in the "*seiscomp/etc/key/*" directory.

3.3 Hard-coded parameters and flows

As previously shown, the *scautoloc* 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 in locating regional events and, in the way they are defined in the original code, 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 configurations are not actually used during the processes. Figure 3.4 presents examples of source code snippets from *scautoloc* in which user-defined parameters are modified, in addition to performing arbitrary normalizations on information related to wave arrival detections. An example of arbitrary source code modification of a user-defined parameter is the minimum time residual that is allowed for a wave arrival detection to be used, which is set by the user through the *scconfig* module, but is multiplied by arbitrary numbers in different parts of the processing flow.

Among the most important procedures for the automatic location of seismic events, the score assigned to each nucleated origin stands out. This score (*originScore*) aims to assess whether the main parameters of the origin (time residuals, signal amplitudes and epicentral distances) are consistent and expected for a real seismic event. Calculations of individual scores for each parameter, as well as the minimum score for an origin to be accepted (*minScore*), are arbitrary and based on global scale events. Aiming to adapt the processes to the reality of RSBR, it is also necessary to make adjustments in the calculation of scores, in addition to the minimum value for an origin to be accepted by the software.

(a)

```
double minResidual = -minFactor*_config.maxResidualUse;
double maxResidual =  maxFactor*_config.maxResidualUse;

if ( _config.aggressivePKP && is_PKP_arrival(arr) ) {
        minResidual *= 2;
        maxResidual *= 2;
}

if ( is_P_arrival(arr) ) {
        // Autoloc 2 hack for regional phases, to allow use of Pg (sometimes even S)
        // as Pn by increasing maxResidual. Which --in principle-- is bad but may be
        // better than leaving those phases out completely.
        double regionalWeight = 1.+0.7*exp(-arr.distance*arr.distance/50.);
        maxResidual *= regionalWeight;
}
```

(b)

```
// Weight residuals at regional distances "a bit" lower
// This is quite hackish!
x = 1 + 0.6*exp(-0.003*delta*delta) + 0.5*exp(-0.03*(15-delta)*(15-delta));
```

(c)

```
// test minimum number of picks
if (origin->arrivals.size() < 6) // TODO: make this limit configurable
        continue;
```

(d)

```
double normalizationAmplitude = 2000.; // arbitrary choice
const_cast<Pick*>(pick)->normamp = pick->amp/normalizationAmplitude;

double normalizedResidual = arr.residual/_config.maxResidualUse;
// add penalty for positive residuals  // TODO: make it configurable?
if (normalizedResidual > 0) normalizedResidual *= 1.5;
```

Figure 3.4: *scautoloc* source code snippets in which there are internal modifications of user-defined parameters. (a) Modification of the maximum residual time allowed for a wave arrival detection; (b) Arbitrary weight for wave arrival detections; (c) Minimum number of detections to define an origin being a fixed value; (d) and (e) Normalizations and arbitrary penalties of amplitudes and residuals.

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, among other implementations described in the Methods and procedures section.

3.4 Velocity models

During the process of locating seismic events, SeisComP uses a travel time table to perform a grid search (possible hypocenters). Traditionally, SeisComP is supplied with a travel times table calculated using the IASP91 velocity model (Kennett and Engdahl, 1991). In RSBR monitoring routines, a second model has been used to relocate regional events, the NewBR model (Assumpção et al., 2010).

3.4.1 IASP91

The IASP91 velocity model (Kennett and Engdahl, 1991) was elaborated by the International Association of Seismology and Physics of the Earth's Interior (IASPEI) and is a 1D model that summarizes the global behavior of travel times of P and S waves for an average Earth radius of 6371 km. From the IASP91 velocity model, travel time tables were proposed in order to update the data published by Jeffreys and Bullen (1940), widely used until then.

In the IASP91 model, the Earth's crust has a thickness of 35 km, with a discontinuity at 20 km depth. The velocities in each layer are represented by a parametric model and the main mantle discontinuities were defined at 410 km and 660 km.

IASP91 is the standard velocity model used by SeisComP for earthquake location. However, as it is a global model, using it to locate regional seismic events may result in larger time residuals, especially when stations at local, regional and teleseismic distances are combined. To avoid the need to accept a larger temporal residual per detection and, consequently, a larger RMS of the time residuals of the event, it is convenient to use a regional velocity model to guarantee more precise locations with smaller time residuals.

3.4.2 NewBR

The NewBR velocity model was proposed by Assumpção et al. (2010) to be a model that takes into account the predominance of an old and cratonic lithosphere in most of Brazil. Fifteen well-known events in Brazilian territory were used, with magnitudes $\geq$ M4.0 and recorded in regional distances, so that their travel times could be corrected using their respective depths. The spherical NewBR model presents a crustal thickness of 42 km.

To develop NewBR, the authors adopted the model proposed by Herrin (1968) as the initial model since, according to Ardito (2009), this model has travel times closer to the times observed in Brazil. According to the author, the velocities of the NewBR model were adjusted through direct modeling by "trial and error".

3.5 Parametric data

During the development of this work, the parametric data of the seismic events was widely used to generate the analyzes presented. For these metadata, SeisComP implements a representation structure called *Sc3ML*, compatible with the *QuakeML* format (Schorlemmer et al., 2011), as shown in the example in Figure 3.5.

```
<?xml version="1.0" encoding="UTF-8"?>
<seiscomp xmlns="http://geofon.gfz-potsdam.de/ns/seiscomp3-schema/0.11" version="0.11">
  <EventParameters>
    <pick publicID="20190128.135521.56-AIC-NB.NBPB..HHZ">
                        [...]
    <amplitude publicID="20190128.135521.56-AIC-NB.NBPB..HHZ.snr">
                        [...]
    <origin publicID="Origin/20220729165733.122099.15657">
      <time>
        <value>2019-01-28T14:05:22.432703Z</value>
        <uncertainty>0.600353539</uncertainty>
      </time>
      <latitude>
        <value>-16.63494873</value>
        <uncertainty>3.464893818</uncertainty>
      </latitude>
      <longitude>
      <depth>
      <methodID>LOCSAT</methodID>
      <earthModelID>newbr3</earthModelID>
      <quality>
        <associatedPhaseCount>7</associatedPhaseCount>
        <usedPhaseCount>6</usedPhaseCount>
        <associatedStationCount>7</associatedStationCount>
        <usedStationCount>6</usedStationCount>
        <standardError>0.2539726416</standardError>
        <azimuthalGap>183.9771471</azimuthalGap>
        <maximumDistance>2.752312899</maximumDistance>
        <minimumDistance>0.08704286814</minimumDistance>
        <medianDistance>2.379316688</medianDistance>
      </quality>
      <evaluationMode>automatic</evaluationMode>
      <creationInfo>
      <arrival>
         [...]
```

Figure 3.5: Example of a seismic event's hierarchical organization of parametric data in *Sc3ML* format. "[...]" indicates the existence of content not shown in the figure.

The creation of the *QuakeML* format aimed to standardize the metadata of seismic events and their sharing. There are several applications that use this format, among them some seismology software and catalog services of major seismological centers in the world. Parametric data in these formats are easily readable by both humans and algorithms because its elements are standardized and well defined.

In the *QuakeML* and *Sc3ML* formats, the parametric data of a seismic event describes its elements in a hierarchical way, for example, associating picks with the origins that are associated with the event. In this way, events and origins are elements related to each other through a "data tree". One of its most important features is the storage of the event's complete history since the nucleation of its first origin. Any parameter associated with the event at some point is kept in its history and is accessible through its parametric file.

4. Methods and procedures

For the development of this work, waveform data from IAG-USP Seismological Center's database were used. Only data recorded by stations from SeisComP's inventory of the Seismological Center were considered, that is, no new stations were added.

Figure 4.1 presents a flowchart indicating the data used during each optimization stage 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 shown below.

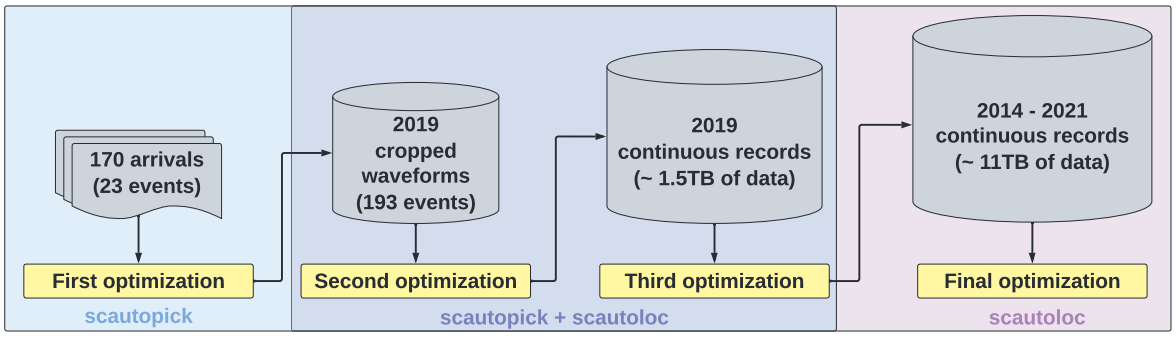


Figure 4.1: Flowchart indicating the data used during each optimization stage of *scautopick* and *scautoloc* parameters.

Initially, 170 arrivals from 23 Brazilian regional events from the RSBR catalog were selected for the initial optimization stage of *scautopick*'s two main parameters: 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, as described in subsection 4.2. Other *scautopick* parameters also went through the initial optimization stage, such as the value of the STA/LTA ratio needed to trigger the picker.

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

In a third optimization stage, 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 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 permanent and temporary stations.

For comparison purposes, the original SeisComP (*i.e.*, without any modification to its detection and location parameters) 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, *scautopick* optimization and *scautoloc* optimization.

4.1 Development of a velocity model - BRA23

To minimize the time residuals of P-wave arrival detections at stations, it is extremely important to use an adequate velocity model. For this, the NewBR regional model was used as a basis for the generation of a new 1D velocity model for the Brazilian territory, called BRA23. Two more recent seismic events were also considered for the elaboration of this model.

The data used are travel time detections for regional stations, combined with events' hypocenters, both accurately determined. The 17 regional events used to derive the BRA23 model were:

- Pacajus/CE (1980);
- Codajás/AM (1983);
- João Câmara/RN (2x 1986 e 1989);
- Sobral/CE (1988 e 2008);
- Palhano/CE (1988 e 1989);
- Taperuaba/CE (1991);
- Porto dos Gaúchos/MT (1998 e 2005);
- Nova Ponte/MG (1998);
- São Caetano/PE (2006);
- Caraíbas/MG (2007);
- Mara Rosa/GO (2010); and
- Montes Claros/MG (2012).

The picks used to generate the BRA23 model were obtained and reviewed by the International Seismological Centre (ISC) and were provided through personal communication with Prof. Dr. Marcelo Assumpção.

The travel times for each station and each event are shown in Figure 4.2, overlapped with the times predicted by the NewBR model in each case. It is possible to observe that the NewBR model tends to overestimate travel times, especially for intermediate distances (up to approximately 2200 km). Thus, the calculated time residuals ($t_{model} - t_{obs}$) tend to be positive, as shown in Figure 4.3.

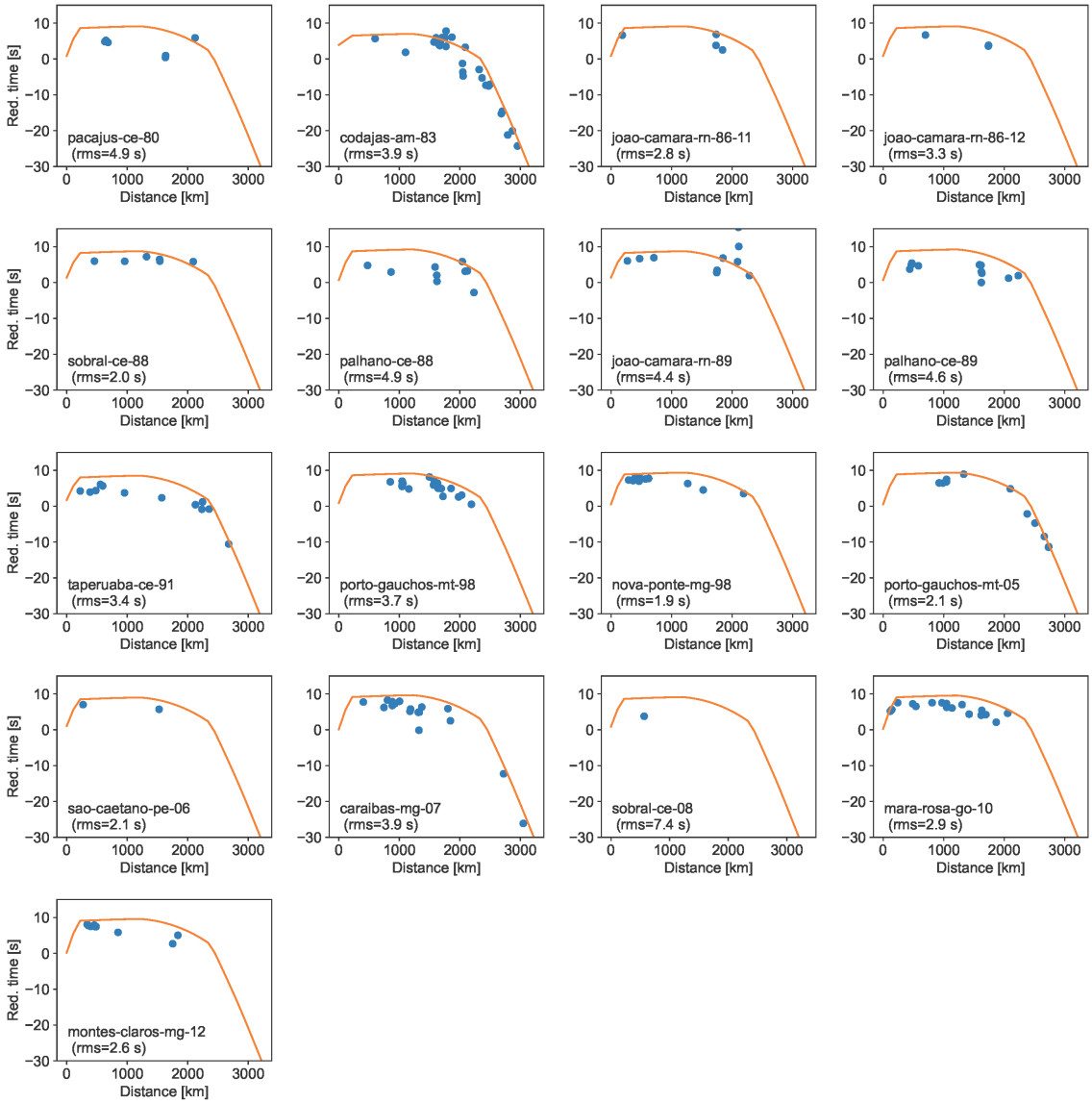


Figure 4.2: Reduced travel times for a velocity of 8.2 km/s for each of the 17 events used to derive the BRA23 model. The orange curves overlapped with the data correspond to the travel times estimated using the NewBR model and considering the depth of each event. The estimated RMS in each case is indicated below the name/year of the event.

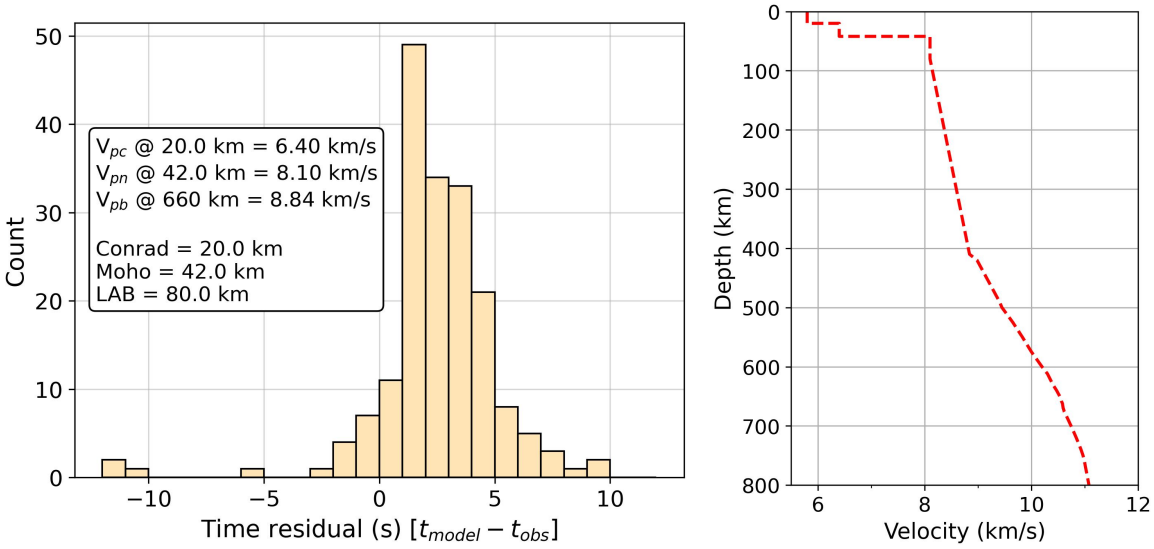


Figure 4.3: Histograms of residuals (left) of wave arrival detections using the NewBR model (right), whose main parameters are indicated. "V_{pc}": P-wave velocity in the lower crust, "V_{pn}": Pn wave velocity, "V_{pb}": P-wave velocity at the base of the upper mantle. In this model, the Conrad discontinuity, the Moho discontinuity and the LAB (lithosphere-asthenosphere boundary) are 20 km, 42 km and 80 km deep, respectively.

The new model of P-wave velocities was obtained from the parameterization of a model composed of two layers of uniform velocities in the crust, a mantle parameterized by a layer of uniform velocity and a velocity gradient that extends to a depth of 660 km. Below 660 km, the AK135-F model (Kennett et al., 1995; Montagner and Kennett, 1996) was adopted.

The travel times for each arrival and each event were computed individually, using the *TauP* package (Crotwell et al., 1999) through the *ObsPy* package interface (Beyreuther et al., 2010), not applying any correction to the data. For each event-station pair, two travel times (t_1 and t_2) were calculated. In the first, it was considered that the 1D model has the same crustal thickness as that of the hypocenter region and, in the second, the crustal thickness of the model was defined as being the same as that of the station's region, as shown in Figure 4.4. The final travel time for the event-station pair was obtained through the simple average of the two cases. The crustal thicknesses considered were proposed by Rivadeneyra et al. (2019) and rounded to one decimal place, in order to reduce the number of possible combinations (tested models) and enable an inversion.

To obtain the BRA23 model, the main constants included in the parameterization proposed by the NewBR model were optimized, aiming to minimize the sum of all calculated residuals. The optimizations were carried out in a similar way to the process of elaborating the NewBR model, but in an automated way, through the development of processing algorithms. The optimized constants, as well as their descriptions and adopted search intervals, are presented in Table 4.1.

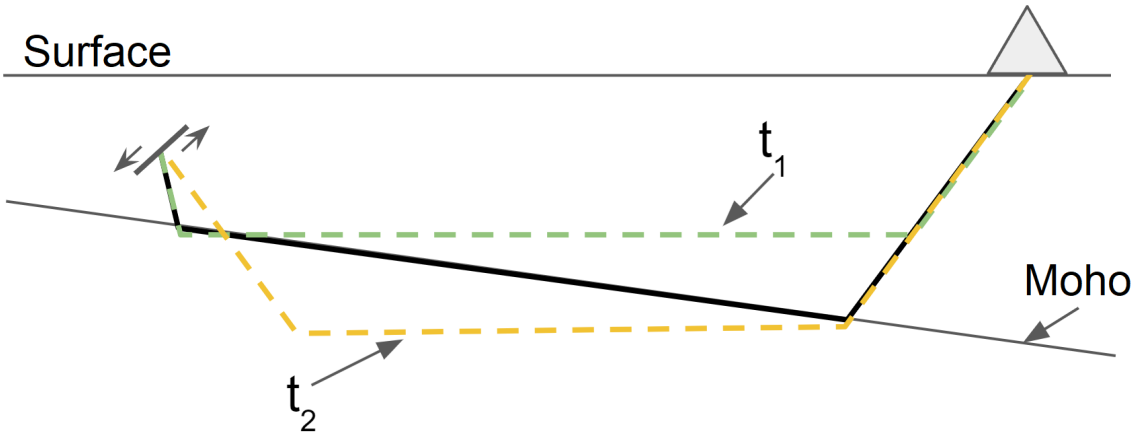


Figure 4.4: Schematization of wavefront travel time calculation between the event's hypocenter and the station. As the thickness of the crust is not constant, an average of the times was adopted considering the thicknesses in the event's region and the station, simulated by two different 1D models.

In order to obtain a stable solution, parameter optimization was performed in two stages: a global optimization, using all the data and the Differential Evolution method implemented by the *SciPy* package (Virtanen et al., 2020), followed by a local optimization by the Nelder-Mead method also implemented by the *SciPy* package. Before local optimization, outliers (points below and above 1.5 the interquartile distance subtracted and added to the first and third quartiles) were removed from the data. Furthermore, the global optimization solution was considered as a starting point for the local optimization.

Table 4.1: Parameters optimized in the process of obtaining the BRA23 model. The parameter optimization was carried out considering the intervals specified in the column "Search range".

Parameter	Search range	Description
$\mathbf{V}_{pc}$	6.4 km/s - 7.5 km/s	P-wave velocity in the lower crust, below the Conrad discontinuity.
$\mathbf{V}_{pn}$	7.5 km/s - 8.6 km/s	P-wave velocity in the upper mantle, inside the LAB.
$\mathbf{V}_{pb}$	8.5 km/s - 10.4 km/s	P-wave velocity at the base of the mantle (660 km depth).
Conrad	10 km - 20 km	Conrad discontinuity depth.
LAB	60 km - 250 km	LAB depth (or the constant velocity layer on top of the upper mantle).

4.2 *scautopick* optimization

The first stage of improving SeisComP involved an evaluation of *scautopick*'s default parameters. As it is responsible for detecting wave arrivals in seismograms, *scautopick* must be able to identify amplitude anomalies originated from seismic events, recording them as accurately as possible, that is, in the first arrival of P-waves.

Inevitably, incorrect picks associated with noise in the signal are recorded and passed on to the next processing modules. Ideal optimization seeks to increase the number of correct detections while decreasing incorrect detections and, for this, several parameters are used in order to enhance seismic events on seismograms and attenuate unrelated noise.

To optimize wave arrival detections performed by *scautopick*, the main parameters evaluated and optimized are described below, using SeisComP syntax[1][2].

- **RMHP(*timespan*)**: moving average equivalent to a high-pass filter, used to remove *offset* from data;
- **ITAPER(*timespan*)**: taper, a one-sided cosine-tapered window, aimed at gradually decreasing the amplitudes at the edges of the data to avoid discontinuities in the calculations;
- **BW(*or,hp,lp*)**: butterworth band-pass frequency filter applied to continuous data. It is possible to modify its order (or) and its lower (hp) and upper (lp) limits;
- **STALTA(*sta,lta*)**: parameters for detecting transient signals, composed by the STA (short-time window) and LTA (long-time window);
- **picker**: algorithm used to perform wave arrival detections (pickings);
- **triggerOn / triggerOff**: activation and deactivation thresholds used by the picker;
- **deadTime**: period, in seconds, in which there will be no new picks after a detection; and
- **amplMaxTimeWindow**: time window, in seconds, used to calculate the signal-to-noise ratio (SNR) of a detection.

These and other parameters can be set for all stations through the *scconfig* configuration module or individually for each station using the bindings, as shown in subsection 3.2.

[1] https://www.seiscomp.de/doc/base/filter-grammar.html
[2] https://www.seiscomp.de/doc/base/concepts/configuration.html

4.2.1 Frequency filter

scautoloc uses only P-wave arrivals from seismic events to perform the location of origins, therefore, a frequency filter must be sought that enhances this phase of Brazilian regional earthquakes. For this, different band-pass filters were initially tested on waveform data from 170 arrivals of 23 seismic events, covering most of the magnitude range of Brazilian regional earthquakes. These wave arrival detections were manually refined individually, seeking the highest accuracy possible. Using an algorithm developed in *Python*, thousands of combinations of band-pass filter limits were performed, seeking those that provided the highest signal-to-noise ratio for each event.

Figure 4.5 presents an example of frequency filter analysis, carried out on data recorded by the NB.NBMA station referring to the event "*usp2021load*" from the RSBR catalog, which occurred near the city of Quixeramobim/CE with magnitude M1.5.

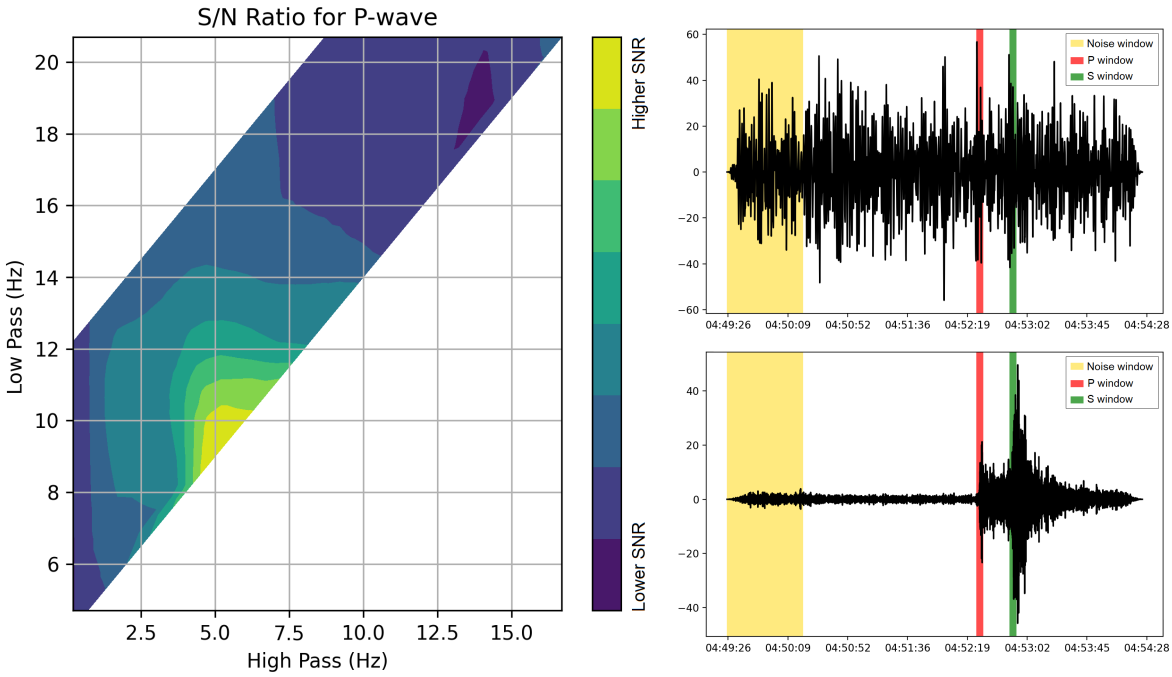


Figure 4.5: Left: Analysis of frequency filter combinations for a seismic event of magnitude M1.5 recorded by the NB.NBMA station. Right: Example of enhancement of seismic waves from the same event through the appropriate frequency filter. The seismogram at the top shows the waveforms of the event using SeisComP's default filter, while the seismogram at the bottom presents the filtered data with the suggested limits for this event after the analyses.

In the graph on the left in Figure 4.5, it is noticeable that the event presented a higher signal-to-noise ratio of P-wave arrivals at the NB.NBMA station when the band-pass filter was applied with limits of 5.2 Hz and 9.5 Hz approximately. The seismograms on the right show the waveforms of this event with SeisComP's default filter (at the top) and the suggested band-pass filter for this event (at the bottom), based on the results of the analysis.

4.2.2 STA/LTA parameters

After applying the frequency filter, *scautopick* uses an STA/LTA algorithm to detect wave arrivals from seismic events. By default, the time windows used are 2 and 80 seconds for STA and LTA, respectively.

To find the best STA and LTA parameters, the same procedure used to define the frequency filter was adopted. Thus, an analysis of several combinations of time windows was carried out, in order to find which one maximizes the STA/LTA ratio of the P-wave arrival, after applying the suggested frequency filter. Figure 4.6 presents an example of this analysis, also based on data referring to the event "*usp2021load*" and recorded by the NB.NBMA station.

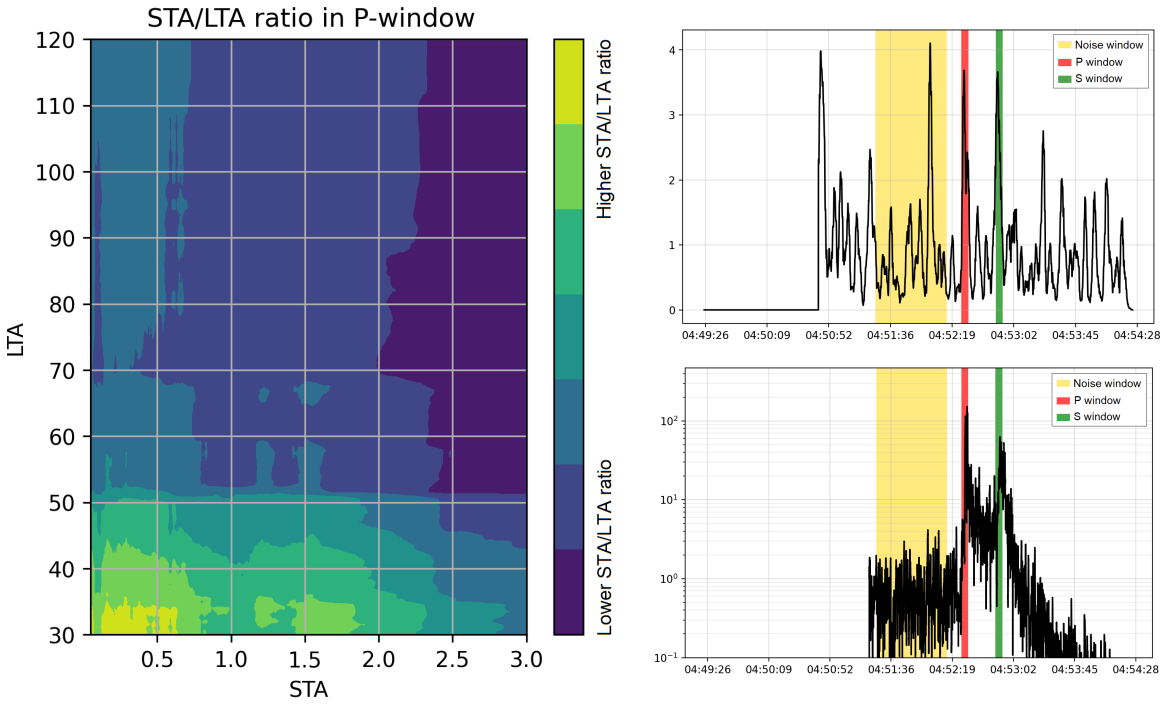


Figure 4.6: Left: Analysis of time window combinations of the STA/LTA algorithm for a seismic event of magnitude M1.5 recorded by the NB.NBMA station. Right: Example of enhancement of seismic waves from the same event through the appropriate frequency filter and STA/LTA windows. The seismogram at the top shows the event's waveforms using SeisComP's default parameters, while the seismogram at the bottom (in logarithmic scale) presents the data with the suggested parameters for this event after the analyses.

The graph on the left in Figure 4.6 shows that the event presented a higher signal-to-noise ratio of P-wave arrivals at the NB.NBMA station when using an STA window between 0.1 and 0.4 seconds and an LTA window between 30 and 35 seconds. On the right, the seismogram at the top shows the waveforms with SeisComP's default parameters (0.7 Hz - 2 Hz; 2 s - 80 s), while the seismogram at the bottom presents the data using the suggested parameters for this event (5.2 Hz - 9.5 Hz; 0.25 s - 33 s).

4.2.3 Picker

In seismology, the most used method for detecting P-wave arrivals in a seismogram is the STA/LTA ratio, as proposed by Allen (1978). Considering the fact that this ratio only provides perceptions of signal amplitude variations, several studies have already been carried out aiming to propose more efficient pickers, that is, ones that are able to evaluate other waveform information such as, for example, the frequency. Due to the fact that the STA/LTA ratio relates only to the signal amplitudes, it is not possible for the algorithm to notice the difference between noise-related pulses and real seismic events, generating false positives in noisy stations.

Akaike Information Criterion (AIC) is a method proposed by Akaike (1971), who at the time was studying the quality of adjustments to a statistical model, trying to find the lowest order that best adjusts to the observed data. One of the observed side benefits is that this criterion is able to separate a time series into two segments with different properties. The implementation of this algorithm also presents robust results even with the presence of high amplitude noise (St-Onge, 2011).

An AIC algorithm is implemented into *scautopick*, although it is not used by default. When activated, the STA/LTA method continues to be used as a detector, providing a time window (which contains the detection of a transient signal) for the AIC algorithm to act as a picker, refining the initial detection. When AIC is deactivated, the STA/LTA method is used as a detector and picker simultaneously, being responsible for generating the picks used for event location.

The AIC algorithm cannot be used as a detector, as it is not possible to apply it to continuous waveform data. Thus, it is necessary to use the STA/LTA algorithm to perform an initial detection and, in a second stage, the AIC acts in a user-defined time window which contains the detection to be refined.

In the present work, the AIC algorithm was used as a picker, which can be configured through the stations' bindings. The configuration parameters of filters used by this picker followed the same ones obtained for the STA/LTA. More details about all the proposed modifications and parameters used are presented in the Results section.

4.3 *scautoloc* optimization

Next, the improvements made to *scautoloc* are detailed, including the modifications made through the configuration module (*scconfig*) and hard-coded processes or parameters, that is, contained in the source code of the software.

As shown in subsection 3.3, some parameters editable by the user through *scconfig* are arbitrarily changed in the original source code of SeisComP. To remove such arbitrariness, modifications were made directly to the source code and, therefore, the parameters set by the user are respected.

Thus, the main location parameters available in *scconfig* and evaluated in this work were:

- **defaultDepth**: default depth of located events, used if the RMS of the resulting origin is lower than the RMS of the origin with the estimated depth. Commonly used for shallow events without depth resolution;
- **minimumDepth**: minimum depth the locator can associate with an origin;
- **maxDepth**: maximum depth the locator can associate with an origin;
- **maxRMS**: maximum RMS (*root mean square*) allowed for an origin to be accepted;
- **maxResidual**: maximum time residual of an arrival for it to contribute to the location of an origin;
- **minPhaseCount**: minimum number of phases (arrivals) contributing to the location of an origin for it to be accepted;
- **maxStationDistance**: maximum distance allowed for a station to provide picks to an origin;
- **amplTypeAbs**: amplitude type used by *scautoloc* (*e.g.*, mb, MLv); and
- **profile**: velocity model associated with the travel time table used to locate events.

During the *scautoloc* optimization process, optimizations were also performed on the grid used to nucleate new origins of seismic events (described below), in addition to the development of the velocity model used by the software (subsection 4.1).

4.3.1 Grid file (*grid.conf*)

The grid used by *scautoloc* during the nucleation process can be modified by the user through the file *grid.conf* (gridfile) within the software directories. As described in the official SeisComP documentation (Helmholtz-Centre Potsdam - GFZ German Research Centre for Geosciences and gempa GmbH, 2008), this file consists of a line for each grid point and they are defined by 6 columns:

```
-10.00 105.00 20.0 5.0 180.0 8
```

The columns are, respectively, the coordinates of the point (latitude, longitude and depth), radius, maximum distance and minimum number of picks. The example line above creates a point at coordinates 10°S/105°E, with a depth of 20 km. This grid point is sensitive to seismic events at a distance of 5° from its center, while stations with distances of up to 180° can provide picks for the nucleation of events. A minimum of 8 picks need to be used to create a valid origin at that location.

All parameters of the grid are extremely important to optimize the automatic location of regional seismic events. For example, a low magnitude seismic event (*e.g.*, $<$ M3.0) occurring in the North Region of Brazil will hardly be registered by a seismic station in the South Region of the country. In this way, it is necessary to limit the maximum distance allowed for each grid point, avoiding the erroneous nucleation of eventual noises.

The density of points in the region of interest allows seismic events to be nucleated more easily, since, considering that the grid points are "initial guesses" of a hypocenter, the more points, the greater the chances of finding one in which the travel time residuals are close to reality. In the same way, an exaggerated density of points might allow noise to be nucleated as seismic events, as it also increases the chances of finding a point in which the noise time residuals are compatible with the expected residuals for a real event.

By default, SeisComP's gridfile has a grid comprising the entire globe area, with spacing of approximately 5°, with some additional points comprising greater depths in places where deep hypocenters occur most frequently.

Seeking to improve the automatic detection of regional seismic events in Brazil, the grid was reformulated in order to increase the number of points in this region, and establish parameters, for each point, consistent with the Brazilian seismicity and the set of seismographic stations in operation.

The proposed grid is composed of 330 points equispaced 2° apart (mostly in Brazilian territory) and 12 points covering the Andean region. Although Andean events are not the focus of this work, it is important that the software consider grid points in this region, preventing those seismic events from being erroneously nucleated in Brazil.

For the grid points arranged in the Brazilian region and its surroundings, the number of stations within a 10° radius from each grid point was the criterion adopted to define the maximum distance (d_{max}) at which a station can provide picks and the minimum number of picks needed to nucleate an event at the point (n_{picks}). n_{sta} refers to the number of stations in a radius of 10° centered on the point. The rules used to define parameters were:

$$d_{max} = \begin{cases} 15, & \text{if } n_{sta} \leq 5 \\ 10, & \text{if } 5 < n_{sta} \leq 15 \\ 8, & \text{if } 15 < n_{sta} \leq 20 \\ 6, & \text{if } 20 < n_{sta} \leq 30 \\ 5, & \text{if } n_{sta} > 30 \end{cases} \qquad n_{picks} = \begin{cases} 5, & \text{if } n_{sta} \leq 10 \\ 6, & \text{if } n_{sta} > 10 \end{cases}$$

Figure 4.7 presents two maps of the South American region, one with SeisComP's default grid and the other with the grid proposed in this work.

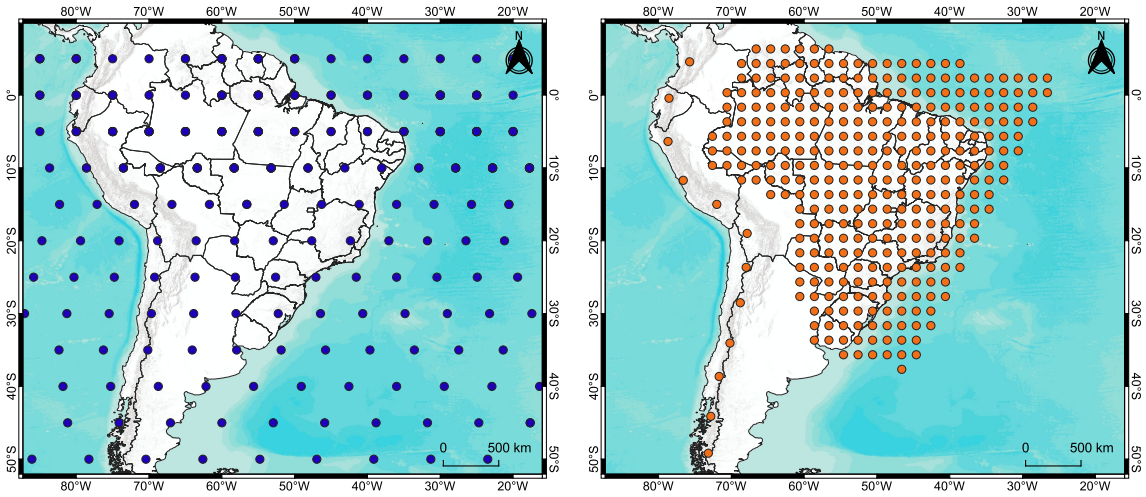


Figure 4.7: SeisComP's default grid (left) compared to the proposed grid (right).

The 12 points arranged over the Andean region do not follow the same rules for defining their parameters. As these are events with origins associated with plate boundaries, with magnitudes mostly greater than those presented by intraplate events, it was defined a maximum distance (d_{max}) of 35° and a minimum of 9 picks (n_{picks}) for an event to be nucleated at these points.

Considering that the number of seismographic stations in Brazil has significantly changed in recent years, eight grids with different parameters were created, referring to each year between 2014 and 2021. To calculate the density of seismographic stations at each point, only the stations that presented at least 50% of data recorded during each year were considered. Further on, in the Results section and in the Appendices section, the grids used for each year are presented, as well as the annual variations in the density of stations for each point.

4.3.2 Station configurations (*station.conf*)

Additionally, it was necessary to change the *station.conf* file, responsible for configuring some parameters of the stations contained in SeisComP's inventory. The content of this file is defined by 4 columns, as described in the official documentation:

```
* * 1 90
GE * 1 180
GE HLG 1 10
TE RGN 0 10
```

The columns represent, respectively: network code, station code, indication of use (0 or 1), and maximum distance for nucleation. An indication of use defined by the number 1 indicates that the station can be used during the event location process. The last column indicates the maximum distance, in degrees, in which that station can contribute to the generation of new origins. The example above indicates that all stations on all networks can provide detections of wave arrivals at events up to 90° away. GE network stations can contribute at any distance, except for the HLG station which is limited to 10°. The RGN station on the TE network will not be used by *scautoloc*.

The file *station.conf* has been modified to contain a single line:

```
* * 1 35
```

Such parameters indicate that all stations in the inventory must be used and that they can contribute detections of wave arrivals of events at up to 35° away. Thus, it is established that the maximum distances associated with the origins located by SeisComP will be defined only by the grid, configured through gridfile.

4.3.3 Hard-coded parameters and processes

As previously described, in order to improve the detectability of regional seismic events in Brazilian territory, it is also necessary to evaluate and restructure the hard-coded parameters and flows, which are not accessible to the user. During the present work, *scautoloc*'s source code has been studied in order to understand the processes involved in the location of seismic events and propose changes consistent with the seismicity in Brazil.

Figure 4.8 presents a simplified flowchart of the creation of an origin, with the functions in *scautoloc*'s source code that underwent the main modifications throughout this work.

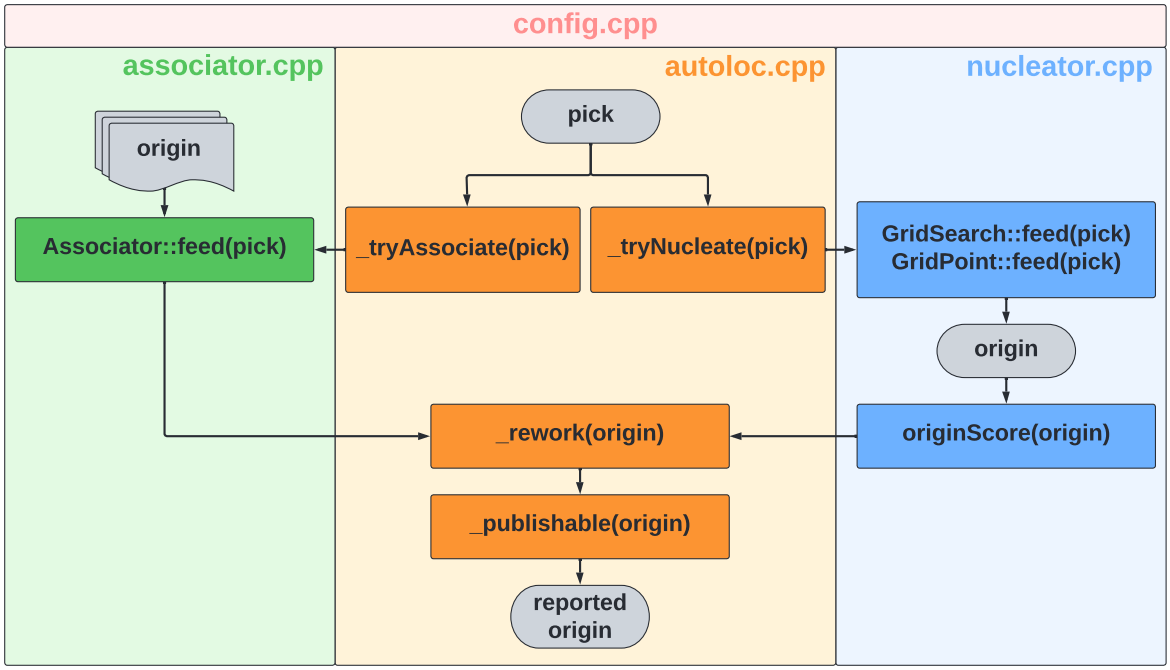


Figure 4.8: Flowchart of the processes in *scautoloc* that underwent the main modifications during the work. The information in parentheses indicates the object received by the function. The "config.cpp" file has configuration parameters that are used throughout the process.

There are several other functions involved in the processes carried out by *scautoloc*, in addition to those presented in Figure 4.8. Thus, this simplification summarizes the functions that underwent the main modifications in parameters and/or flows in the source code, allowing SeisComP to detect more regional seismic events when compared to the original software. Such modifications are presented in more detail in the following topics.

4.3.3.1 Configurations (*config.cpp*)

Among *scautoloc*'s code files, there is a configuration file (*config.cpp*) with parameters used during the location of a seismic event. Some of the parameters defined in this file are not available to be modified through the *scconfig* configuration module and therefore need to be adjusted directly in the software source code to obtain a better success rate in locating regional seismic events in Brazil.

Below are the details of each parameter that underwent changes directly in the *config.cpp* file. The results of their changes are presented in the Results section.

- ***goodRMS***: defined in the source code as "typically good RMS in our network", this parameter is only used in two moments during the processes performed by *scautoloc*, both when there is already at least one nucleated origin. At first, this parameter is used to test whether, when modifying the origin's depth to the default value (*defaultDepth*), the new RMS is better than the RMS with the calculated depth. In the second moment, the *goodRMS* parameter is used to verify the possibility of adding new picks to an origin, performing the validation by comparing this parameter (arbitrarily multiplied by 2) with the time residuals of each pick. However, after the proposed modifications, this verification was changed and started to use *maxRMS* for validation;
- ***dynamicPickThresholdInterval***: threshold interval, in seconds, in which *scautoloc* starts discarding picks of similar amplitudes from a certain station, aiming to reduce the possible use of detections arising from problems in the data (*e.g.*, noisy station);
- ***maxResidualKeep***: maximum temporal residual of a pick for it to be associated with an origin. Unlike *maxResidualUse* (defined in *scconfig*), *maxResidualKeep* does not refer to the maximum residual for a pick to contribute to the location of the origin. Therefore, in the case where $maxResidualUse < RMS \leq maxResidualKeep$, the pick will still be associated with the origin, but will not be used in the location process. It should be noted that, although there is a possibility of changing the value of this variable in the *config.cpp* file, it is reset to its original value in the *app.cpp* file. Thus, it is necessary to exclude the unnecessary redefinition of this variable in the *app.cpp* file so that the change is actually carried out;
- ***defaultMaxNucDist***: default maximum distance for a station to contribute to origin nucleation. This parameter is overridden by the value defined in the *station.conf* file, as shown in "Station configurations (*station.conf*)";
- ***minScore***: originally, it is the minimum acceptable value for an origin to be accepted, referring to the arbitrary score performed by *scautoloc*. During the proposed modifications, this parameter was renamed to *minScoreTolerance* due to the new score calculations, described in more detail in "*_score(origin) function*" and "*_publishable(origin) function*";

- ***defaultDepth***: default value to be used as the depth of an origin. The depth can be recalculated throughout the process if the origin quality parameters are improved with the relocation;
- ***minimumDepth***: minimum depth allowed for an origin. Considering the Brazilian intraplate shallow seismicity, the value was redefined to 0;
- ***minScoreBypassNucleator***: *scautoloc* always tries to associate a new pick with a known origin before adding it in the nucleation process. If the pick is already associated with an origin, *minScoreBypassNucleator* defines a minimum value of score that prevents the pick from going on to the nucleation process, as the software understands that it actually belongs to the origin that presented a good score.

4.3.3.2 Associator (*associator.cpp*)

The process of associating a new pick to an existing origin is started by the *_tryAssociate(pick)* function, included in the *autoloc.cpp* file. This function feeds the associator (*associator.cpp*) with the pick, following a flow in which some of its main information is evaluated.

Among the processes, the calculation of an arbitrary parameter called "affinity" stands out, which determines whether the pick should be associated with the origin. For the pick to be associated, this parameter must be greater than or equal to 0.1, and is defined by:

$$affinity = cos^2\left(\frac{x \times \pi}{2}\right) \tag{1}$$

where x is a parameter obtained through an equation involving the temporal residual of the pick (in seconds) and the epicentral distance (in degrees):

$$\begin{gathered} x = \frac{residual}{10 \times w} \\ w = 1 + 0.6 \times \exp(-0.003 \times dist^2) + 0.5 \times \exp(-0.03 \times (15 - dist)^2) \end{gathered} \tag{2}$$

To reformulate the method that defines the viability of a pick being associated with an origin, it was decided to use as a reference the maximum residual allowed (*maxResidualUse*, set by the user) and the maximum distance allowed for the origin (defined by the grid point at which it was generated). Thus, the original equation was replaced by the following condition:

$$\begin{cases} \text{associate,} & \text{if } |residual| \leq maxResidualUse \text{ and } dist \leq gridMaxDist \\ \text{do not associate,} & \text{if } |residual| > maxResidualUse \text{ or } dist > gridMaxDist \end{cases} \tag{3}$$

That way, if the pick meets the distance and temporal residual requirements, it will be associated with the origin. This was the main modification carried out in the process of associating picks, aiming to allow any new wave arrival detections to be associated with the origins in a less arbitrary way and also obeying the conditions imposed by the user, instead of arbitrary calculations.

4.3.3.3 Nucleator (*nucleator.cpp*)

As previously described, if the association process fails, the pick is introduced into the nucleation process flow through the *_tryNucleate(pick)* function, which triggers the nucleator (*nucleator.cpp*) to attempt to create new origins.

GridPoint::feed(pick) and *GridSearch::feed(pick)* functions

Originally, the nucleator projects the pick into every grid point (via the *GridPoint::feed(pick)* function), while checking if there is a minimum number of picks grouped at each point. This initial requirement is hard-coded and set to the value 6, that is, at least 6 picks are needed for nucleating a new origin. Although there is an editable parameter in *scconfig* called *minPhaseCount*, it is not used in this part of the process.

Considering that RSBR has a low density of stations in several regions of the national territory, it is prudent to modify the minimum number necessary to nucleate new origins. However, the adverse effect is the probable increase in the amount of false nucleated events, that is, groups of noise picks that were mistakenly interpreted as wave arrivals from real seismic events. Thus, the minimum number of picks grouped at a given point to proceed with the nucleation process has been redefined to 5.

When the *GridPoint::feed(pick)* function identifies a group of picks that meet the initial criteria, a "candidate origin" is created. Since the new picks are projected into every grid point, it is common for more than one point to meet the minimum number criterion, producing more than one candidate origin. Once a candidate origin is created, the picks associated with it are called "arrivals".

Candidate origins are returned to the *GridSearch::feed(pick)* function, which uses several criteria to define whether a new origin will be reported. In this function, many criteria were modified in order to allow greater detectability of seismic events in Brazil, since, again, several original parameters and processes are arbitrary and defined aiming the detection of global earthquakes.

One of the main reasons why potentially promising candidate origins (*i.e.*, those that could turn out to be an origin of an actual seismic event) are discarded at an early stage is that there are too many unnecessary parameter checks throughout the entire process. Several times, *GridSearch::feed(pick)* checks whether the preliminary origin already meets the minimum number of picks to be published, discarding it if not. As there is the possibility of new picks being associated with the origin after its creation, it is not ideal to discard it at that moment. It is prudent to only carry out this verification in the moment immediately before the publishing of the origin, that is, when it will actually be accepted as a seismic event origin.

A second crucial aspect in this stage of the location process is the fact that the nucleator returns only one origin among all the candidates, choosing the best one based on some criteria, including an arbitrary score that will be explained later. This issue is commented in the nucleator source code, showing that it is a point to be revisited by developers in the future:

> *"Now for all 'candidate' origins in tempOrigins try to find the 'best' one. This is a bit problematic as we don't go back and retry using the second-best but give up here. Certainly scope for improvement."* [3]

In this way, the processes performed by the nucleator were reformulated, aiming to remove unnecessary checks carried out in preliminary stages, as well as to allow more than one candidate origin to be nucleated. With these changes, all candidate origins are passed on and, if they do not meet the requirements defined by the user, they are discarded at an opportune moment in the process, when there is no further possibility of refinement (in the *_publishable(origin)* function).

[3] Comment taken from *scautoloc*'s source code (nucleator.cpp - line 653).

_originScore(origin) function

The nucleator has a function that assigns an arbitrary score to the origin, using it as a parameter to keep or discard the origin during the process. This function is also accessed outside the nucleator, in order to update the score when necessary. The score is based on key information from each origin's arrival:

- Epicentral distance and depth;
- Amplitude and SNR; and
- Time residual.

Again, all calculations performed to obtain the score are arbitrary, as well as the threshold defined for the origin to be discarded (*minScore*, by default, set to 8). The score is originally calculated as follows:

$$
\begin{aligned}
originScore &= \left(\sum_{n=0}^{N_{arr}} arrivalScore[n] \right) \times depthFactor \\
arrivalScore &= distScore \times amplScore \times timeScore \times phaseScore \\
distScore &= 1.5 \times \exp\left(\frac{-d \times d}{r^2} \right) \\
amplScore &= 1.0 + 0.8 \times \left(1.0 + 0.5 \times \log_{10}\left(\frac{amp}{2000} \right) \right) \times \log_{10}(snr) \\
timeScore &= 0.5 \times \left(cos\left(\frac{\left(\frac{|residual|}{2 \times maxRMS} - 0.2 \right) \times \pi}{1.0 - 0.2} \right) + 1.0 \right)^2 \\
phaseScore &= \begin{cases} 1.0 \text{ , if arrival is P} \\ 0.3 \text{ , if arrival is PKP} \end{cases} \\
depthFactor &= 1 + 0.0005 \times (200 - depth)
\end{aligned}
\tag{4}
$$

In the equations above, d is the distance from the station that produced the pick, r is the maximum allowed nucleation distance for the station (defined in the *station.conf* file), amp and snr are the mb and SNR amplitudes calculated by *scautopick*, while $maxRMS$ is the maximum RMS allowed for an origin to be reported.

All constants, as well as the equations themselves, are arbitrary and defined by SeisComP developers, aiming to locate earthquakes of global scales. There is also no normalization of the score of an origin based on its number of arrivals, which allows an origin with many "bad" arrivals to be accepted, for example, in a scenario in which the sum of the low scores of its arrivals reaches the minimum score required for acceptance.

The equations used to calculate a score for the origins were completely redefined, following rules that would help to choose an acceptable limit for the origin to be defined as belonging to a real seismic event. Thus, the following premises were adopted:

- Each arrival from an origin contributes a score between 0 and 1;
- Penalties on scores should preferably be based on parameters predefined by the user;
- Equations should be simplified as much as possible; and
- The score of each arrival is the average of the scores of each parameter, while the score of the origin is the average of the scores of all arrivals.

The new equations were proposed using a database of Brazilian seismic events, whose picks were previously verified and consolidated. Several rounds of event nucleation tests were performed in order to find the set of equations that allowed the highest detectability of real events while minimizing false positives.

With that, all parts of the score calculation were reformulated and are described below. Originally, the minimum score value for an origin to be accepted can only be changed by the user through the command line. In the source code, this parameter is defined in the *config.cpp* file (*minScore*) and has a value of 8.

The new minimum score value is obtained through a calculation performed at the end of the process of creating an origin, during the evaluation of the main parameters of the origin to be published, and is detailed in "*_publishable(origin)* function". A new option was also added in the *scconfig* configuration module so that the user can have access to the minimum score in a simplified way, through the graphical interface.

• ***Distance Score***

$$distScore = 1.0 - \left(0.25 \times \frac{arrivalDistance}{gridMaxDist}\right) \tag{5}$$

To calculate the *distScore*, it was decided to use an equation that penalized the score as the distance from the origin to the station that recorded the arrival approached the maximum distance allowed by the grid point where the origin was nucleated. Arrivals with distances greater than the maximum allowed by the grid point are not used in the location process, that is, the smallest *distScore* possible to be obtained by an arrival is 0.75.

Figure 4.9 presents the graph referring to the *distScore*'s calculation.

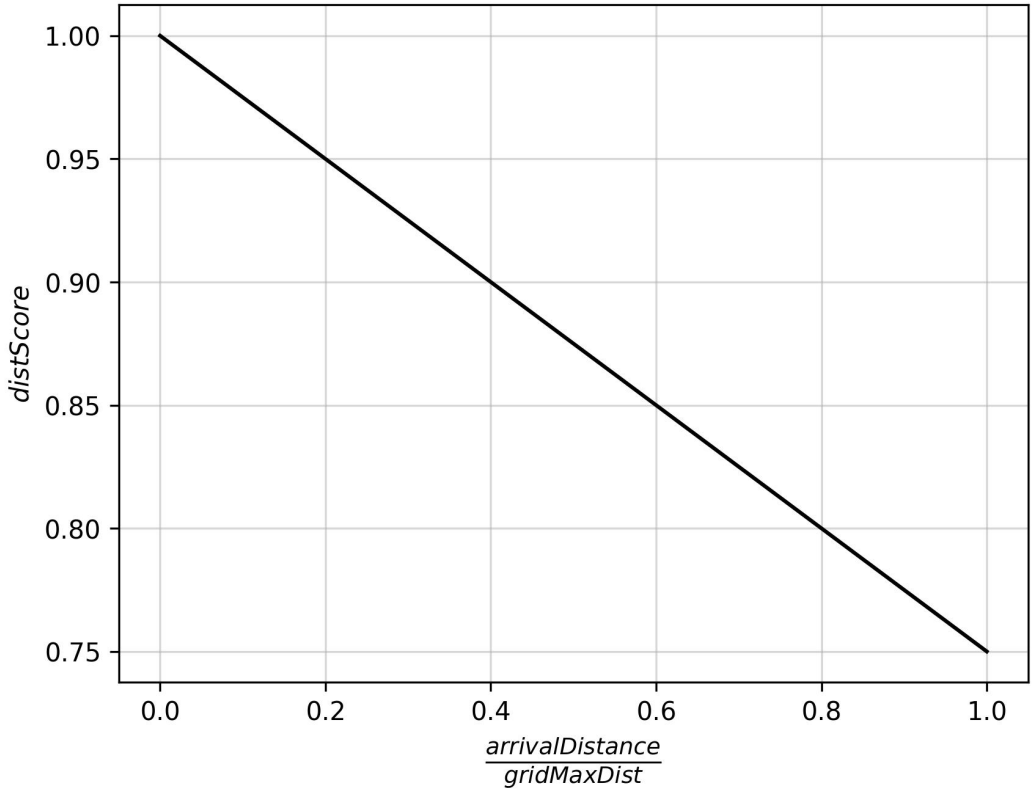


Figure 4.9: Graphical representation of *distScore*'s new calculation. The highest scores are obtained with the smallest distances, while the penalty becomes more severe as the arrival distance approaches the maximum allowed distance (*gridMaxDist*).

• ***Amplitude Score***

The calculation of the score related to the wave arrival amplitude (*amplScore*) has been severely reformulated in order to introduce the concept of magnitude. The calculation of the magnitude of a seismic event takes into account the distance from the station to the epicenter and, therefore, is a parameter that helps in the validation of an arrival.

Originally, as seen in the set of Equations 4, the *amplScore* calculation takes into account the largest amplitude of the seismogram during P-wave arrivals, in addition to normalizations and arbitrary operations. In this work, we propose to use the amplitude and epicentral distance of each arrival to calculate a "preliminary MLv magnitude" of the event.

To calculate the MLv magnitude at each station, the equation derived from Richter (1935) was implemented, together with the calibration function proposed by Alsaker et al. (1991), as shown in Equation 6.

$$\begin{aligned} MLv &= \log_{10}(amp) - \log_{10}(A_0) \\ -\log_{10}(A_0) &= 0.91 \times \log_{10}(dist) + (0.00087 \times dist) + 1.01 \end{aligned} \tag{6}$$

Figure 4.10 illustrates an example of the calculated MLv magnitude using data from 10 seismographic stations. Each station contributes a magnitude based on their amplitudes and epicentral distances, and the event's "preliminary magnitude" corresponds to the median of the values.

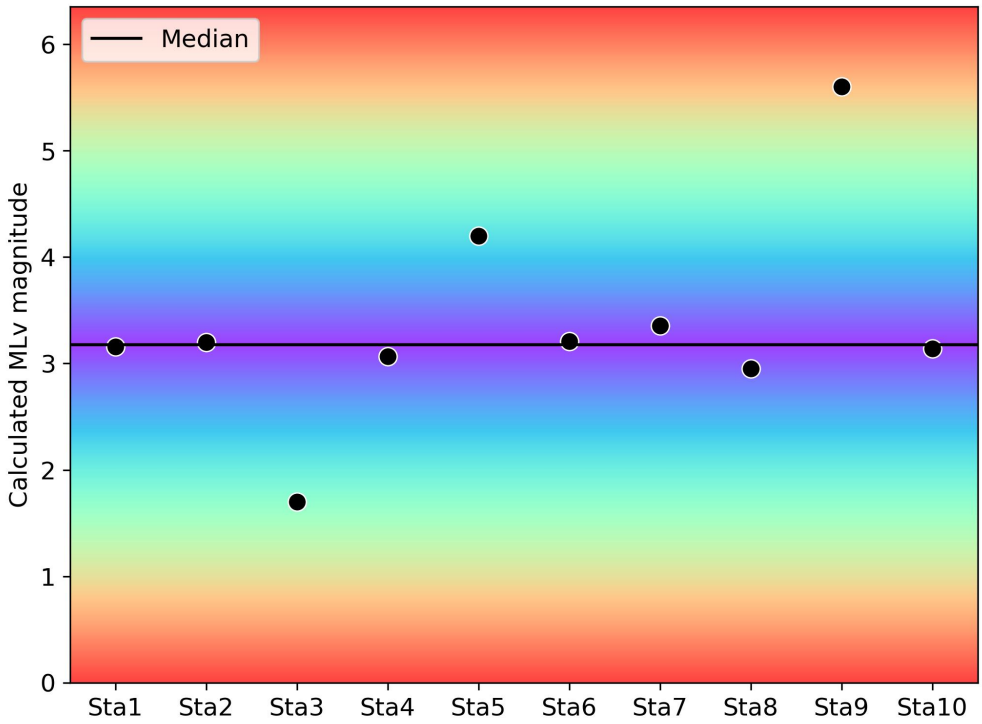


Figure 4.10: Graphical representation of an example of obtaining the preliminary MLv magnitude (median). The farther from the median, the greater the penalty on the *amplScore* of the arrival. The color scale represents the penalty in a qualitative way, with the purple color representing the smallest penalties and the red color the largest ones.

After calculating the MLv magnitude at each station, the *amplScore* is obtained through Equation 7, which penalizes arrivals with magnitudes far from the median, as illustrated by the color scale of the example from Figure 4.10.

$$amplScore = \begin{cases} 0 \text{ , if } |arrivalMag - medianMag| > 4/3 \\ 1.0 - (0.75 \times (|arrivalMag - medianMag|)) \text{ , otherwise} \end{cases} \quad (7)$$

This procedure was adopted using the premise that picks associated with the same real seismic event tend to present calculated magnitudes close to each other and, therefore, picks that present magnitudes that differ significantly from the median of the group are more likely to be false.

Figure 4.11 shows the graph referring to the *amplScore*'s calculation.

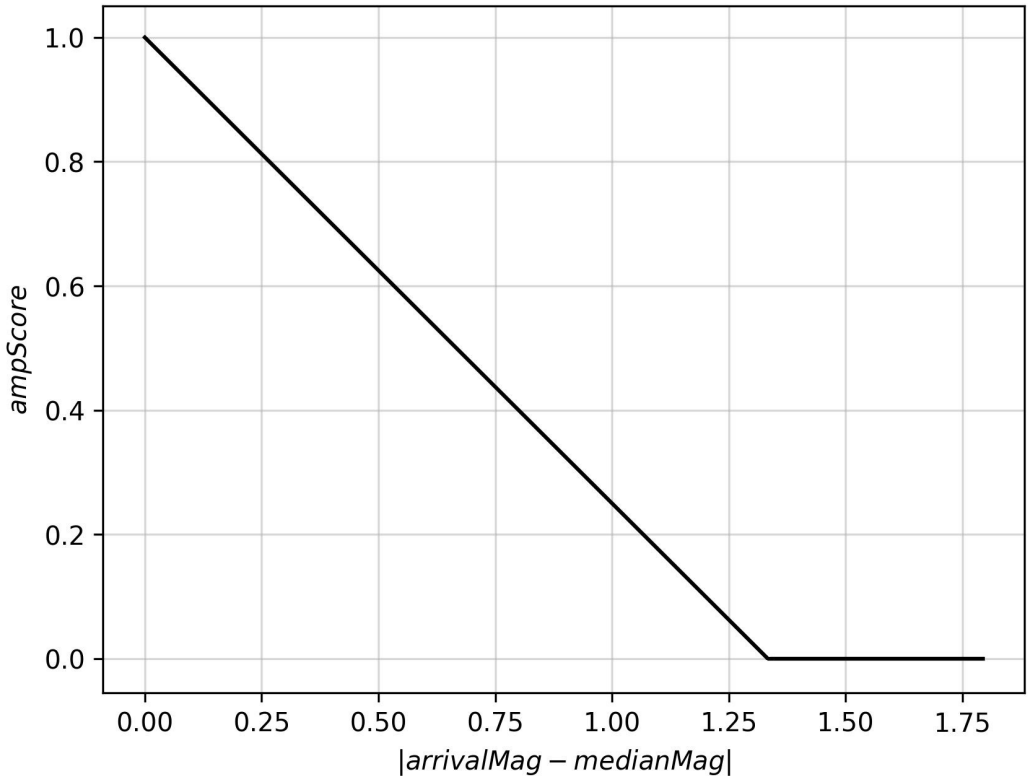


Figure 4.11: Graphical representation of *amplScore*'s new calculation. The highest scores are obtained with the smallest differences between the magnitude of the arrival and the median of the set. From a certain difference between the magnitudes $(4/3)$, the score becomes zero.

For the proposed changes in *amplScore*'s calculation to achieve the expected result, it is necessary to ensure that the MLv amplitude is calculated by *scautopick* and that *scautoloc* is configured to use it as preferred amplitude, as shown in the Results section.

• ***Time Score***

$$timeScore = 1.0 - \left(0.25 \times \frac{|residual|}{maxRMS}\right) \tag{8}$$

To calculate the *timeScore*, it was decided to use an equation that penalized the score as the arrival residual approaches the maximum allowed residual. Arrivals with time residuals greater than the maximum allowed are not used in the location process, that is, the smallest possible *timeScore* to be obtained by an arrival is 0.75.

Figure 4.12 presents the graph referring to the *timeScore*'s calculation.

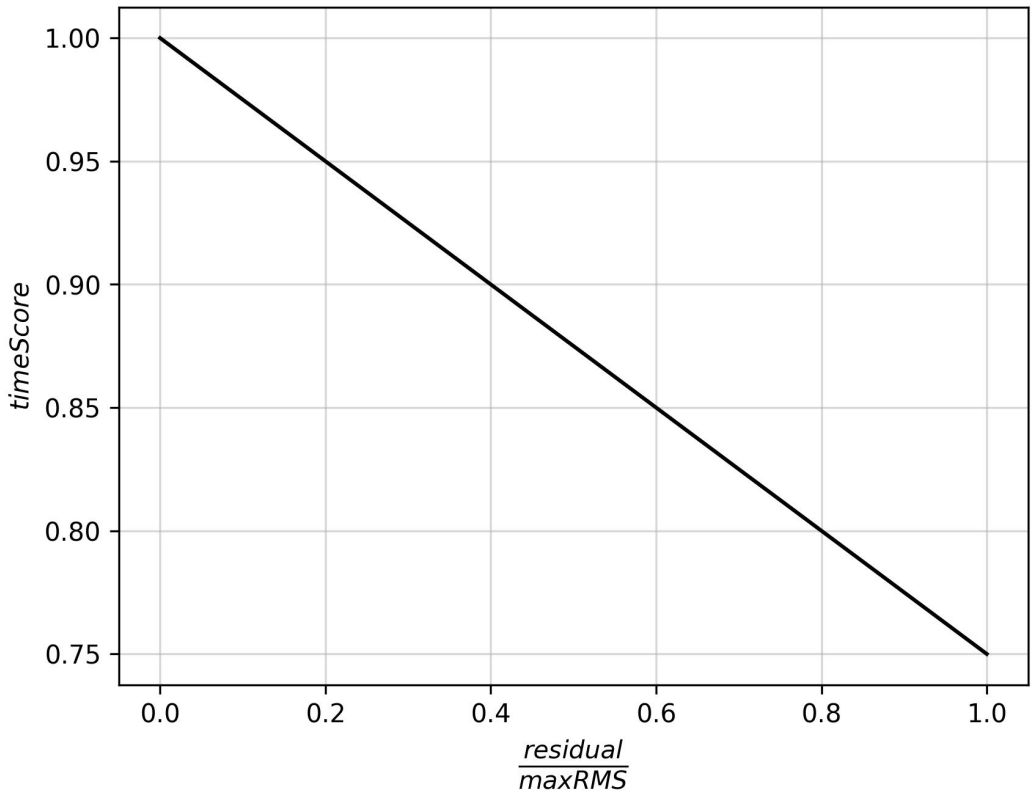


Figure 4.12: Graphical representation of *timeScore*'s new calculation. The highest scores are obtained with the smallest time residuals, while the penalty becomes more severe as the arrival residual approaches the maximum allowed residual (*maxRMS*).

The use of the BRA23 velocity model has a direct relationship with the *timeScore* parameter since, when using it to locate seismic events in Brazil, there is a tendency for the time residuals of arrivals to decrease. As a result, the proposed penalties for the *timeScore* tend not to cause major impacts on the scores related to real seismic events.

• *Origin Score*

After reformulating the calculation of each parameter score, the final origin score (*originScore*) is now calculated through the following equation, in which N_{arr} refers to the number of arrivals that contribute to the location of the origin.

$$originScore = \frac{\left(\sum_{n=1}^{N_{arr}} arrivalScore[n]\right)}{N_{arr}}$$
$$arrivalScore = \frac{distScore + amplScore + timeScore}{3.0} \tag{9}$$

Additionally, the following penalties to the final origin score are applied, if necessary:

$$originScore = originScore \times depthFactor$$

$$depthFactor = \begin{cases} 1.00 \text{ , if } depth < 20\ km \\ 0.95 \text{ , if } 20\ km < depth \leq 30\ km \\ 0.80 \text{ , if } depth > 30\ km \end{cases} \tag{10}$$

$$originScore = \begin{cases} originScore - 0.10 \text{ , if } distancePenalty \text{ is triggered} \\ originScore - 0.05 \text{ , if } phaseCountPenalty \text{ is triggered} \end{cases}$$

Regarding the distances of arrivals in relation to the origin, a new penalty was added to the score (*distancePenalty*). This penalty is applied *(i)* if the first wave arrival was registered in a station located at a distance greater than half of the maximum distance allowed by the grid point that originated it (*gridMaxDist*) or *(ii)* when the average distance of all arrivals is greater than $0.65 \times gridMaxDist$.

Based on the premise that stations close to events are the ones that record them in greater quantity, this penalty aims to prevent the nucleation of noise in distant stations from generating a false origin. Arrivals with distances greater than the maximum allowed are not used in the location process, that is, the lowest possible *distScore* to be obtained by an arrival is 0.75.

Regarding the number of phases associated with the origin (*i.e.*, number of stations that contributed to the location), we sought to penalize the origins that present the minimum number allowed for nucleation (*minPhaseCount*). This optional penalty considers that allowing nucleation with fewer stations can lead to an increase in false positives, and can be disabled by the user through the "*phaseCountPenalty*" parameter, implemented in *scconfig* after modifications.

_rework(origin) function

One of the most important processes when creating an origin is its relocation. Initially, the nucleation of an origin generates a hypocenter at some of the points defined in the grid, and it is necessary to refine this initial location so that the quality parameters are optimized. Removing arrivals from the origin due to poor quality parameters and modifying the event depth are examples of moments when it becomes necessary to relocate an origin.

In a simplified way, the *_rework(origin)* function is responsible for relocating an origin. As in other parts of the code, in this function there is also an excessive amount of minimum requirement checks that can invalidate the continuation of an origin in the process, such as the count of arrivals. Aiming to keep such validations concentrated only at the end of the process, the *_rework(origin)* function was also restructured.

In the relocation process, arrivals with time residuals above the allowed value are also excluded, in addition to trying to add other possible picks to the solution. Detections performed by distant stations are also discarded, in addition to possible PkP phases identified by the software. At the end of the relocation, the origin criteria referring to the maximum distance (d_{max}) and minimum number of picks (n_{picks}) are redefined based on the grid point closest to the new hypocenter.

_publishable(origin) function

As previously mentioned, one of the biggest setbacks to a promising origin being classified as a valid origin and consequently reported is the excessive amount of quality tests carried out too early. In its initial stages of formation, a potentially good origin may present quality parameters below those configured by the user for it to be accepted.

For example, due to the fact that the nucleation is based on an arbitrarily discretized grid, the point at which the origin was initially located can result in large time residuals because it is not exactly the real epicenter of the event. In this case, it is not convenient to eliminate the origin because of its high RMS, as it should present an improvement in this parameter when it is relocated closer to its real epicenter.

A second example is the case where a promising origin is initially nucleated with fewer stations than the user-defined number for an origin to be accepted. In the original source code, the number of stations test is performed several times, including immediately after nucleation. By understanding that new picks can be associated with the origin at different times throughout the process, it was considered unwise to eliminate the origin in the initial moments due to the minimum number of arrivals. Similar situations occur throughout the entire process, until the actual publishing of the origin.

In this way, we sought to remove from the source code several origin validation checks performed at times considered unnecessary, focusing them preferably on the moment when the origin is about to be published, that is, accepted as a true origin. The *_publishable(origin)* function became responsible for making the last tests and parameter checks, in order to verify if the origin is "publishable". It is worth highlighting the following processes, carried out shortly after the last effective relocation:

- Calculation of distances and azimuths for each arrival of the origin, ensuring the storage of correct values;
- Validation of the minimum number of arrivals based on the value referring to the grid point closest to the hypocenter;
- Origin's azimuth gap validation;
- Origin's score validation, whose calculation is explained in "*_score(origin)*". The minimum score value for an origin to be accepted is explained below;
- Origin's RMS validation; and
- Origin's depth validation.

As mentioned in "*_score(origin)*", the score reformulation also included the minimum score required for an origin to be accepted. Previously, this value was fixed and defined in the source code (*config.cpp*) or set by the user through the command line (argument *--min-score*).

With the proposed modifications, the new minimum score became dynamic and is now linked to the maximum distance at which a station can contribute to the origin (d_{max}, parameter defined in the *grid.conf* file and presented in "Grid file (*grid.conf*)"). This is due to the fact that origins that allow the contribution of more distant stations are more likely to nucleate noise and, therefore, greater strictness in the evaluation of the score becomes necessary.

The *minScore* variable was renamed to *minScoreTolerance* and assigned a value of 0.75, adding an additional value according to the maximum allowed distance, as shown in Equation 11.

$$minScore = minScoreTolerance + \frac{gridMaxDist}{100} \tag{11}$$

Thus, an origin that has a maximum nucleation distance of 10° must present a score of at least 0.85 to be accepted. On the other hand, origins with exclusively closer arrivals (*e.g.*, maximum of 5°) can be accepted if they present a score value of 0.80.

5. Results

As mentioned in the Methods and procedures section, the parameters were adjusted following processing and analysis stages, starting from a scenario with little data (170 arrivals from 23 selected events) until all continuous data from 2014 to 2021 ($\approx$ 11TB of data) were covered. Figure 5.1 presents a corner plot of results regarding the first stage of optimizing the frequency filter parameters and STA/LTA, as described in section 4.

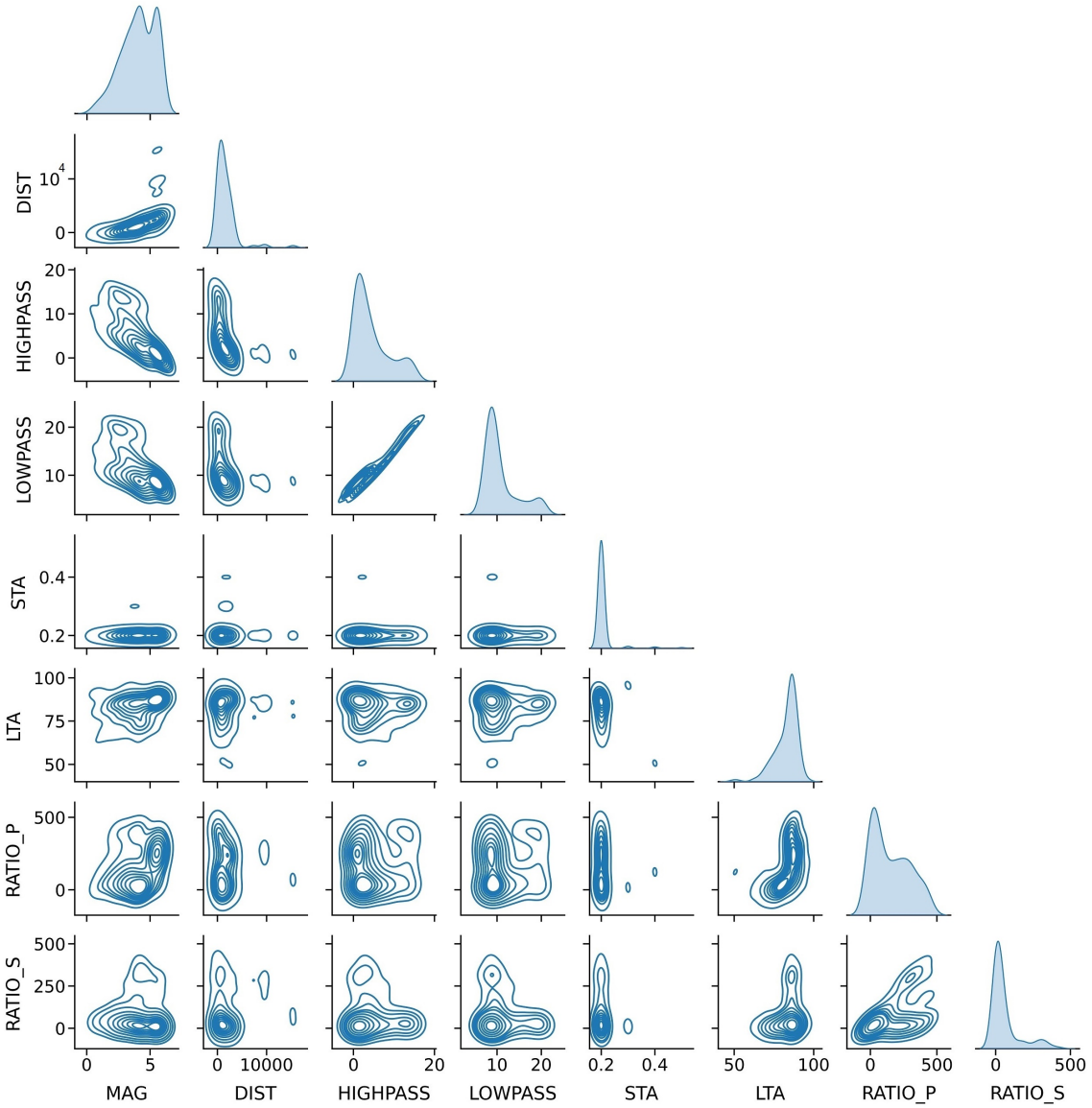


Figure 5.1: Corner plot regarding the main parameters of the 170 arrivals used in the first stage of optimizing detection parameters.

Aiming at the automatic detection of P-wave arrivals from regional seismic events, combinations of band-pass filters and STA and LTA windows were sought to provide the highest possible signal-to-noise ratio for these arrivals. From the diagonal graphs in Figure 5.1, it was possible to obtain the initial values for these parameters, namely: band-pass filter limits of approximately 4 Hz and 9 Hz and STA and LTA windows of

approximately 0.2 and 80 seconds, respectively. Then, several processing stages and test rounds were performed in order to refine these detection parameters, together with other parameters and processes mentioned below.

From the graphs, some aspects of the results obtained initially also stand out, such as the relationship between the epicentral distance and the limits of the frequency filter, which corroborates what was expected, since events with smaller epicentral distances tend to have higher frequencies and vice versa.

5.1 Frequency filter and STA/LTA parameters

After the parameter adjustment stages, the limits of the ideal frequency filter were defined as 4.5 Hz and 10 Hz, values that provided greater enhancement of P-waves in most of the seismic events evaluated. Figure 5.2 shows the comparison between the waveforms filtered with SeisComP's default parameters and with the proposed parameters.

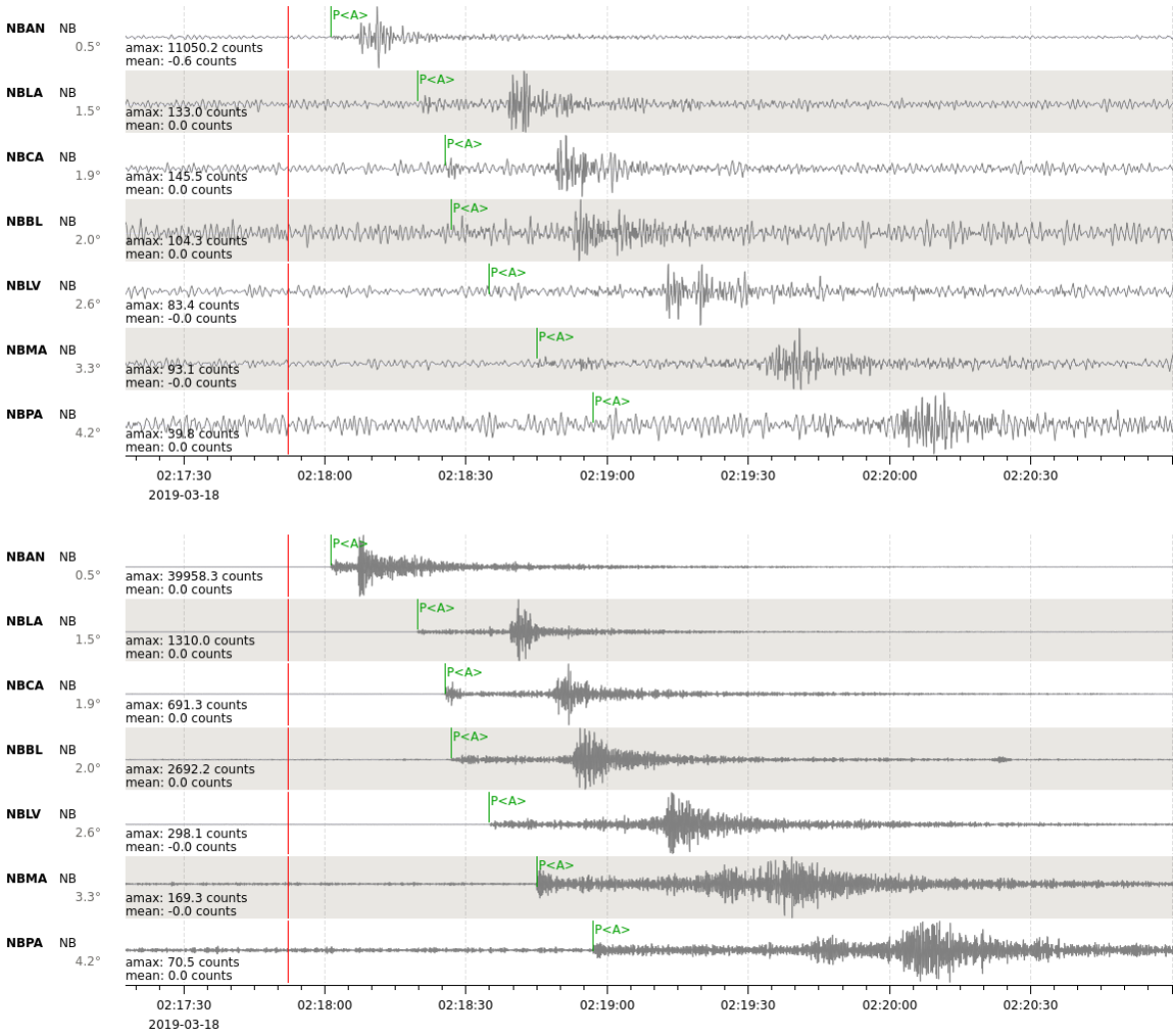


Figure 5.2: Examples of waveform filtering referring to a seismic event of magnitude M2.5 in Olho d'Água Grande/AL. Above, the default limits of SeisComP (0.7 Hz and 2 Hz) were used. Below, the data filtered with the limits proposed in this work (4.5 Hz and 10 Hz).

The time windows for the STA/LTA algorithm were defined as 0.2 and 45 seconds, respectively. Figure 5.3 shows the difference between the waveforms filtered with the combination of SeisComP's default parameters and with the combination of the proposed parameters.

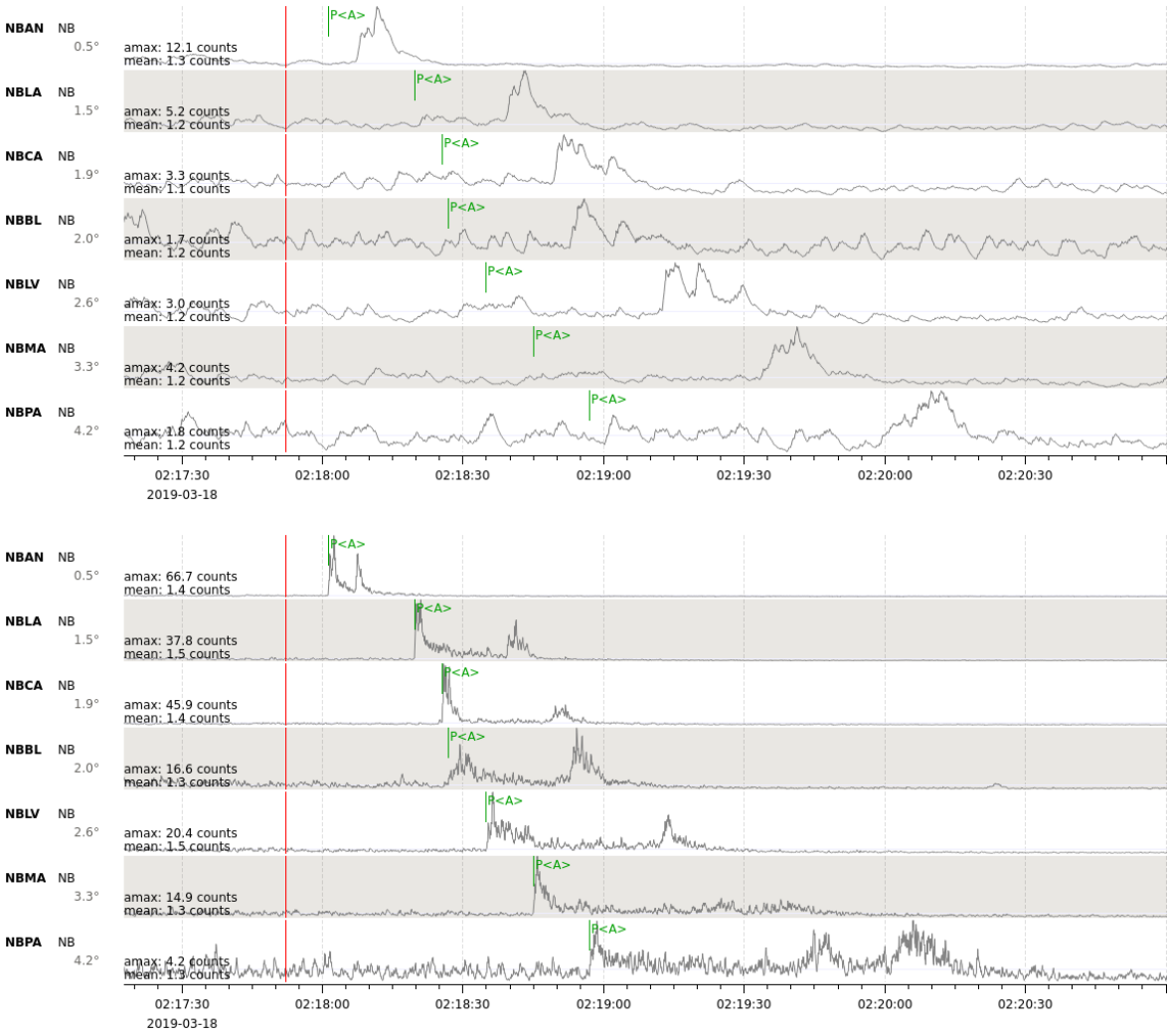


Figure 5.3: Examples of filtering and application of STA/LTA in waveforms referring to a seismic event of magnitude M2.5 in Olho d'Água Grande/AL. Above, SeisComP's default parameters were used (frequencies from 0.7 Hz to 2 Hz and STA/LTA of 2 and 80 seconds). Below, the data processed with parameters proposed in this work (frequencies from 4.5 Hz to 10 Hz and STA/LTA of 0.2 and 45 seconds).

The waveform processed with SeisComP's default parameters show a poor STA/LTA ratio at the time of the P-wave arrivals, which is practically undetectable by the software. In this case, only S or surface waves could be detected by the automatic detecting system. In several other cases, even these arrivals are not identifiable by the system.

Using the proposed parameters, the P-wave arrival and its STA/LTA ratio are considerably enhanced, which allows its correct detection in the seismograms.

5.2 Configuration parameters

In this subsection, the proposed modifications for the various configuration parameters are exposed, whether they are accessible through *scconfig* or only through the software source code. The tables present the modified parameters using SeisComP syntax, the same one adopted in the configuration module and in the software source.

5.2.1 *scautopick* and *scautoloc*

Table 5.1: Configuration parameters of *scautopick* accessible through *scconfig*'s graphical interface.

	Original	**Proposed**
filter	RMHP(10) ITAPER(30) BW(4,0.7,2) STALTA(2,80)	RMHP(10) BW(4,4.5,10) STALTA(0.2,45)
timeCorrection	-0.8	0
picker		AIC
thresholds.triggerOn	3.0	3.0
thresholds.triggerOff	1.5	0.7
thresholds.maxGapLength	4.5	1
thresholds.amplMaxTimeWindow	10	3
thresholds.deadTime	30	2.5

In Table 5.1, we highlight the changes made to the frequency filter and the STA/LTA windows for wave arrival detection (picking), as well as the activation of the AIC algorithm as picker. In the frequency filter, it was decided to remove the *taper* (*ITAPER(30)*) as it was observed that this keeps the picker inactive for a while after a detection.

The *timeCorrection* parameter was set to zero, since the AIC picker becomes responsible for correcting the detection performed by the STA/LTA algorithm. The *amplMaxTimeWindow* parameter refers to the size of the time window, in seconds, used to calculate the signal-to-noise ratio of the pick. As the data was filtered between 4.5 Hz and 10 Hz, there is no need for a large window to perform this calculation, in addition to the fact that its decrease also contributes to the pick being reported faster.

Although the value of the *triggerOn* parameter remains the same, the value of *triggerOff* has been decreased to 0.7. Together with the new *deadTime* of 2.5 seconds. Both metrics prevent the picker from performing several consecutive detections during the same seismic event, deactivating it while the STA/LTA ratio remains greater than 0.7 or for the configured time interval. As regional events tend to have a short duration, there is no need to keep the picker deactivated for 30 seconds as originally configured.

Table 5.2 presents parameters related to *scautoloc* that were modified through the *scconfig* configuration module.

Table 5.2: Configuration parameters of *scautoloc* accessible through *scconfig*'s graphical interface.

	Original	**Proposed**
locator.defaultDepth	10	0
locator.minimumDepth	5	0
autoloc.maxDepth	1000	50
autoloc.maxRMS	3.5	0.8
autoloc.maxResidual	7.0	1.2
autoloc.minPhaseCount	6	5
autoloc.maxStationDistance	180	35
autoloc.amplTypeAbs	mb	MLv
locator.profile	iasp91	BRA23
autoloc.phaseCountPenalty*		true
autoloc.dropAndeanOrigins*		true
autoloc.minScoreTolerance*		0.75

Considering that Brazilian regional seismic events occur mostly at shallow depths, we sought to refine the range of depths allowed for *scautoloc* to associate with the origins created. Additionally, in order to reduce the number of false origins, the maximum RMS allowed was changed from 3.5 to 0.8. Besides that, the maximum temporal residual allowed for each arrival was decreased from 7.0 to 1.2. Thus, origins with large time residuals are discarded because they are probably associated with noise.

To allow origins with less than 6 picks to be accepted, the *minPhaseCount* parameter has been set to 5, in addition to being included in the hard-coded processes for checking the minimum amount of arrivals. In the original SeisComP, even if this parameter is modified to a smaller number, the minimum value of 6 arrivals is defined in the source code and overrides the user's choice during the process.

The *amplTypeAbs* parameter defines the type of amplitude used by *scautoloc*. For the calculation of *Amplitude Score* to be performed correctly, it is necessary to use the MLv amplitude. Finally, the module was configured to use the travel time table based on the BRA23 velocity model (subsection 5.4).

Parameters marked with (*) are not present in the original SeisComP and were proposed in this work. The *phaseCountPenalty* parameter enables the penalty, in the *originScore*, of the origins that have the minimum number of wave arrivals allowed (*Origin Score*). On the other hand, the *dropAndeanOrigins* parameter causes events with origins in the Andean region (close to any of the 12 points shown in "Grid file (*grid.conf*)") not to be reported.

5.2.2 *Bindings*

Table 5.3 presents the *binding* parameters of the stations, configured through *scconfig*.

Table 5.3: Stations' individual configuration parameters ("bindings"), accessible through *scconfig*'s graphical interface.

	Proposed
amplitudes.enableResponses	true
picker.AIC.noiseBegin	-10
picker.AIC.signalBegin	-3
picker.AIC.signalEnd	2
picker.AIC.filter	RMHP(10) BW(4,4.5,10)
picker.AIC.minSNR	3.0
detecFilter*	RMHP(10) BW(4,4.5,10) STALTA(0.2,45)
trigOn*	3.0
trigOff*	0.7
timeCorr*	0

In the bindings of each station, we sought to configure parameters of the AIC picker, in addition to activating the removal of the instrument response in the data, a necessary procedure for the MLv amplitudes to be correctly calculated.

Parameters marked with (*) are the same ones that can be defined for all stations through *scautopick*'s settings, as shown previously in Table 5.1. However, it may be necessary to make individual adjustments, for example in noisy stations or networks. During the development of this work, we noticed the need to increase the *trigOn* and *picker.AIC.minSNR* values of the XC network stations, due to the fact that these stations produce many picks associated with noise. Thus, these parameters were changed from 3.0 to 4.5.

5.2.3 Hard-coded configuration parameters (*config.cpp*)

Table 5.4 presents *scautoloc*'s hard-coded parameters, directly modified in the software source code (*config.cpp* file).

Table 5.4: Hard-coded parameters within *config.cpp* file. The other parameters of this file, not shown in this table, are accessible to the user through the *scconfig* configuration module and it is not necessary to modify them in the code.

	Original	**Proposed**
goodRMS	1.5	0.4
dynamicPickThresholdInterval	3600	600
maxResidualKeep	21	3
defaultMaxNucDist	180	35
~~**minScore**~~ **minScoreTolerance**	8	0.75
minScoreBypassNucleator	40	0.97

The *dynamicPickThresholdInterval* parameter has been reduced to 600 seconds, that is, *scautoloc* will discard detections from a given station after 10 picks of similar amplitudes within 600 seconds. The maximum residual allowed for an arrival to be associated with an origin (*maxResidualKeep*), even without contributing to its location, has been reduced from 21 to 3 seconds.

It is worth highlighting the renaming of the *minScore* parameter to *minScoreTolerance*, since the minimum score for an origin to be accepted is now dynamic and no longer a fixed value, as detailed in "Function *_publishable(origin)*". The minimum value of this dynamic variable was defined as 0.75 considering that, after modifications, the origin score ranges from 0 to 1.

Originally, the *minScore* parameter can only be modified through command line options (argument *--min-score*) or directly in *scautoloc*'s configuration file. However, in order to facilitate the change of this variable by the user, the modifications proposed in this work include its availability through the *scconfig* configuration module (as shown in Table 5.2). The addition of new parameters to *scconfig* can be done through the "descriptions" files of each module, contained in "*seiscomp/etc/descriptions*".

The *config.cpp* file also contains the definition of other parameters used during the process of locating seismic events. However, such parameters (not shown in Table 5.4) are accessible through *scconfig* and do not need to be modified directly in the source code. These are presented in Table 5.2.

5.3 Proposed grids

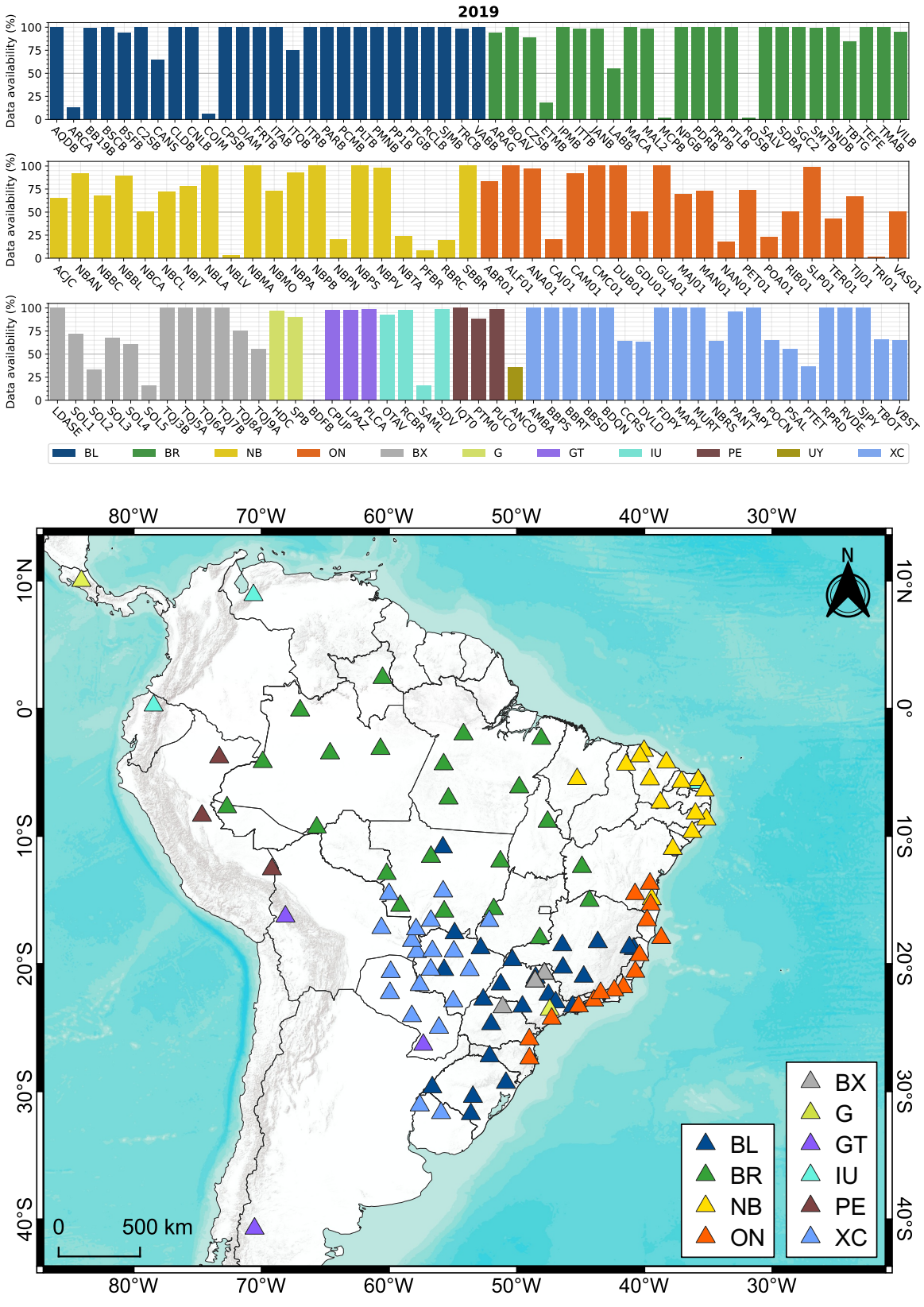


Figure 5.4: Above: availability of waveform data recorded in 2019 by the stations included in the IAG-USP SeisComP inventory. Below: locations of stations with at least 50% of data available in the period.

The grid creation process is exemplified in Figure 5.4: 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 5.5 shows the grid for the year 2019, with a color scale based on the number of stations in 10° radius circles centered on each point.

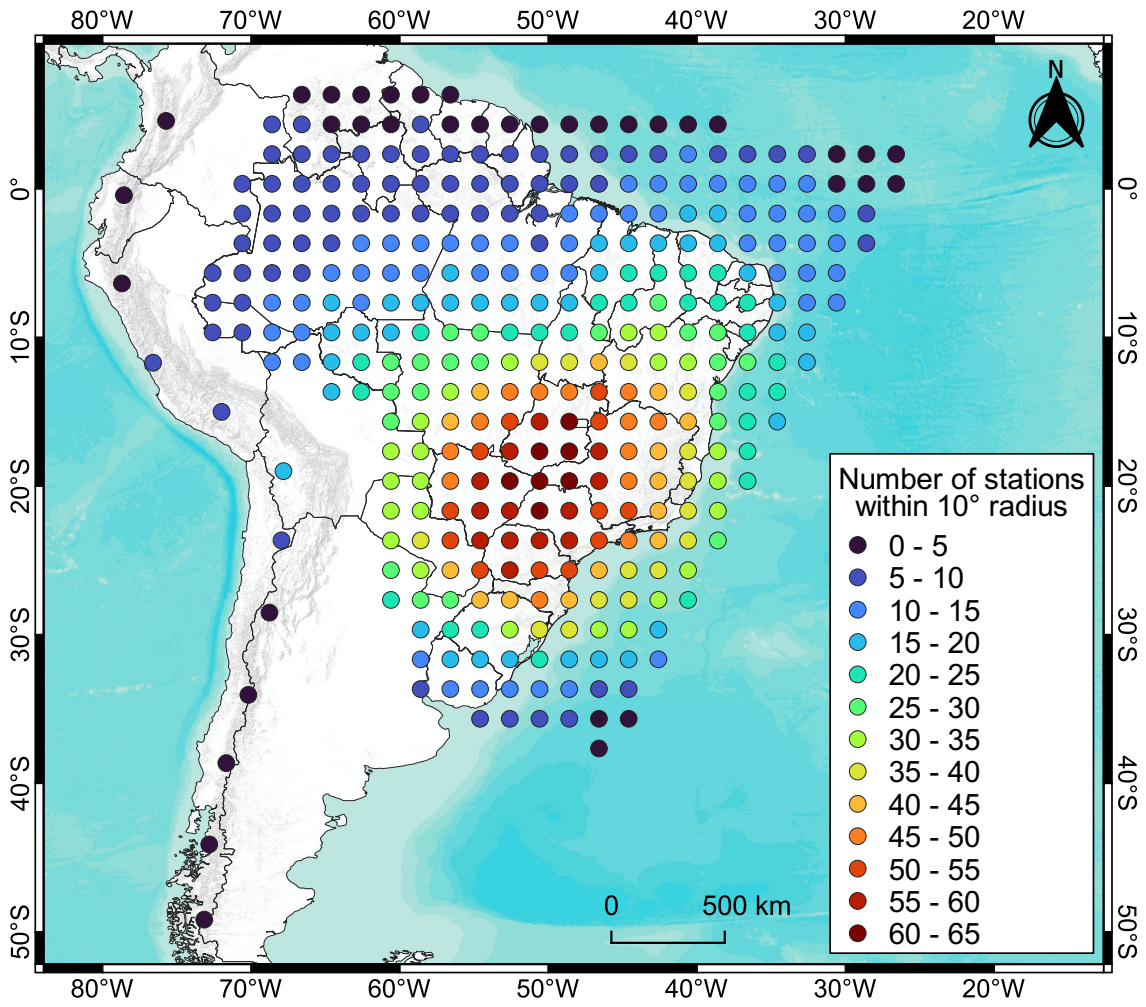


Figure 5.5: Grid for the year 2019. The color scale refers to the number of stations within a radius of 10° with at least 50% of data recorded in the year.

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. As detailed in "Grid file (*grid.conf*)", the number of stations for each point determines the maximum distance allowed for a station to contribute to an origin nucleated at the point as well as the minimum number of picks for this origin to be accepted.

Grids for each year are presented in the Appendices, as well as the annual data availability for each station and their geographic locations.

5.4 New velocity model - BRA23

During the process of nucleating new origins, 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.

As previously presented, the BRA23 model was obtained through an optimization of the same data used in the construction of the NewBR model, added to two more recent regional events (subsection 4.1). In total, data from 17 events were considered, with 183 P-wave detections from different regions of Brazil. Figure 5.6 presents a map with the distribution of the events used, as well as the associations with the stations used for each event.

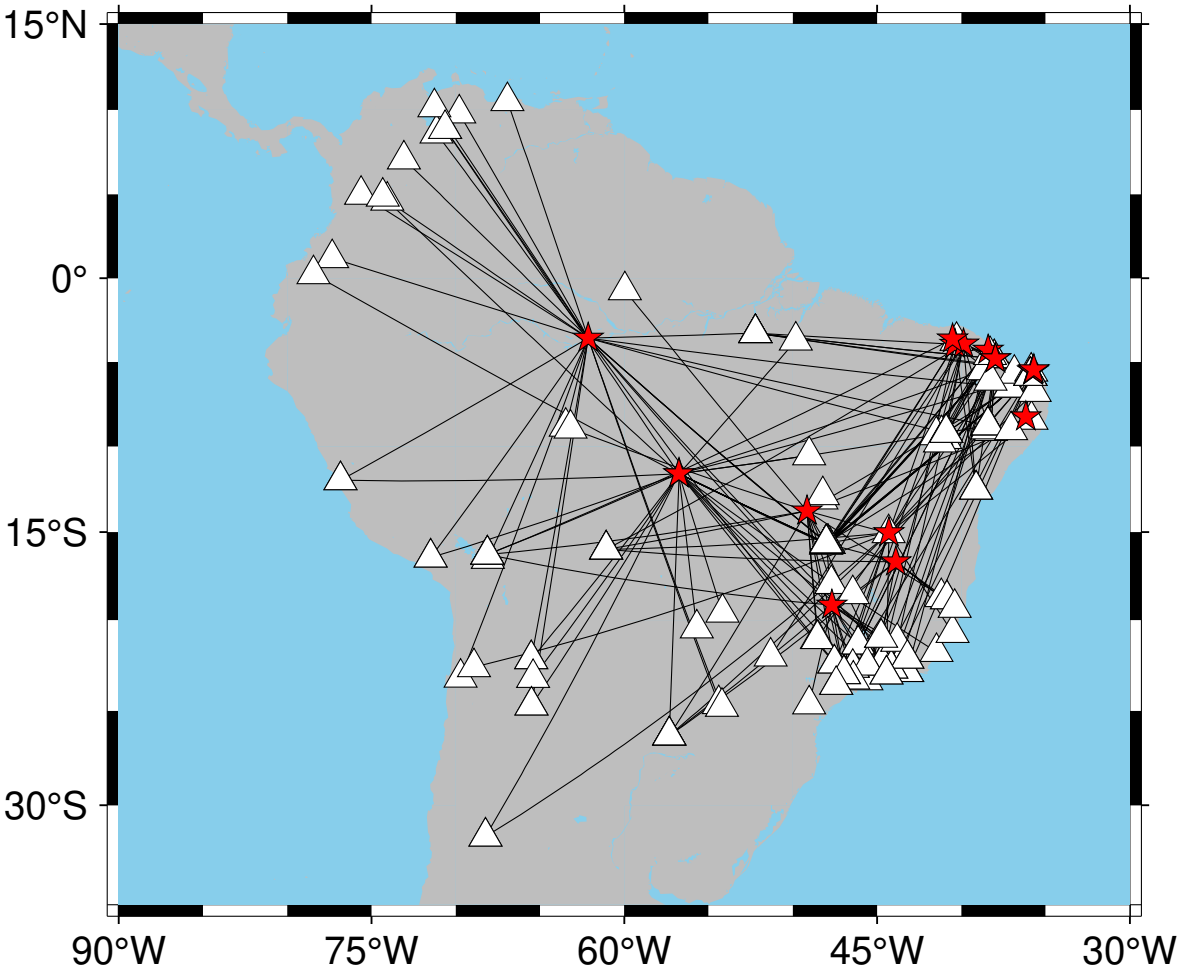


Figure 5.6: Map with the 17 regional events (stars) and stations (triangles) used in the construction of the BRA23 model.

The optimization of the new model consisted of minimizing the travel time residuals for the paths shown in Figure 5.6. For each event-station pair, the event depth was taken into account during the travel time calculation, as well as the crust thicknesses in the station and event region. Crust thicknesses used to correct travel times were obtained from Rivadeneyra et al. (2019). To minimize the average residual of all paths, the following parameters were optimized:

- **V_{pc}:** P-wave velocity in the lower crust;
- **V_{pn}:** P-wave velocity in the subcrustal mantle (extrapolated to the LAB);
- **V_{pb}:** P-wave velocity at the base of the mantle;
- **Conrad:** Depth of the Conrad discontinuity, separating the upper crust (with fixed velocity of 5.8 km/s) from the lower crust (with velocity V_{pc}); and
- **LAB:** Depth of the lithosphere-asthenosphere boundary which, in the BRA23 model, corresponds to the depth of influence of the optimized V_{pn} velocity and in which the mantle velocity was imposed as constant.

Because it is a highly non-linear problem, and in order to seek the absolute minimum of residuals, it was decided to divide the optimization into two stages. In the first stage, 10 rounds were performed using a global optimization method (Differential Evolution), which is a stochastic method capable of searching, from search intervals (Table 4.1), the global minimum of a function. As a result of this process, 10 different models were generated, presented on the left in Figure 5.7.

The models obtained are close to each other and follow, approximately, the NewBR model. Optimized models show a certain variance in both crustal and mantle parameters. Then, in order to refine the optimization, each of the 10 global solutions were used as initial parameters of a local optimization, based on the Nelder-Mead method. This method has a convergence to the local minimum of the objective function, being also subject to the search limits, and presents a faster convergence than a stochastic search algorithm.

Unlike global optimization, local optimization did not use all the data. In each case of local optimization, picks inconsistent with the initial model being tested (outliers) were discarded. The amount of discarded data varied between 6% and 9.2% (11 and 17 detections out of a total of 183) in the 10 cases.

The 10 models resulting from the local optimization are presented on the right in Figure 5.7. Note that all the final models obtained present a good convergence to an average model, especially for the parameters V_{pn}, V_{pb} and LAB. The parameters V_{pc} and Conrad present greater variability.

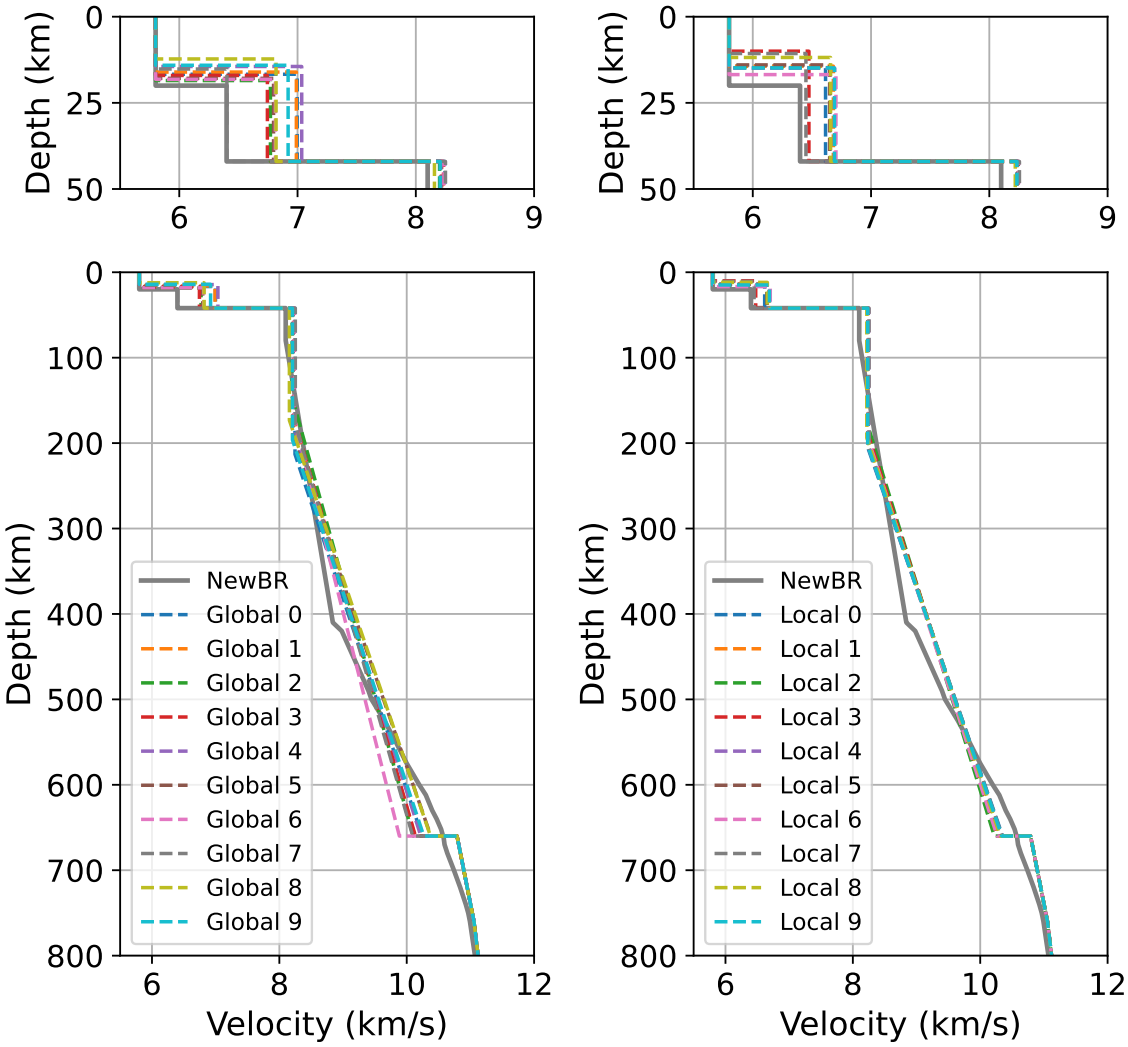


Figure 5.7: P-wave velocity models derived from global (left) and local (right) inversion. In total, 10 global inversions were performed and, for each global inversion, a local inversion was derived. The upper panels show a detailing of the crustal region. For comparison, the original NewBR model is also presented. In general, global models show greater variability in the mantle when compared to local models, which show greater convergence of values in this region.

Since it was unfeasible to test more models due to the total computing time required in view of the deadline determined for the conclusion of this work, the final model adopted was obtained through the median of each parameter established by the 10 local optimizations. The final values for each parameter, as well as their uncertainties (given by the standard deviation of the solutions), are shown in Table 5.5.

Table 5.5: Final solution of the optimized parameters for the BRA23 model.

$\mathbf{V}_{pc}$	$\mathbf{V}_{pn}$	$\mathbf{V}_{pb}$	**Conrad**	**LAB**
6.66 ± 0.09 km/s	8.24 ± 0.01 km/s	10.32 ± 0.04 km/s	14 ± 2 km	198 ± 8 km

In general, the values obtained for all parameters were consistent with expectations. A last parameter necessary for the construction of the model is the thickness of the crust, for which the value of 40 km was adopted, the average value proposed by Rivadeneyra et al. (2019). As previously detailed, during the optimizations this value was considered known and variable for each event, which is not true in real cases where a velocity model is used.

During the global optimization process, it was possible to monitor the tested models and their respective RMS. Figure 5.8 shows the histogram, for each of the parameters, of the number of models tested, as well as a histogram of the RMS obtained. The Differential Evolution Method, derived from genetic algorithm methods (Storn and Price, 1997), tends to concentrate the generations of models to be tested around the most attractive solutions for the objective function. In the graphs, the red horizontal bars represent the region corresponding to the median value adding and subtracting one standard deviation, calculated from the 10 solutions of local optimizations in each parameter. Note that, for all variables, there is a tendency for the average solution to be robust in relation to the optimization process, as well as the estimated uncertainties themselves.

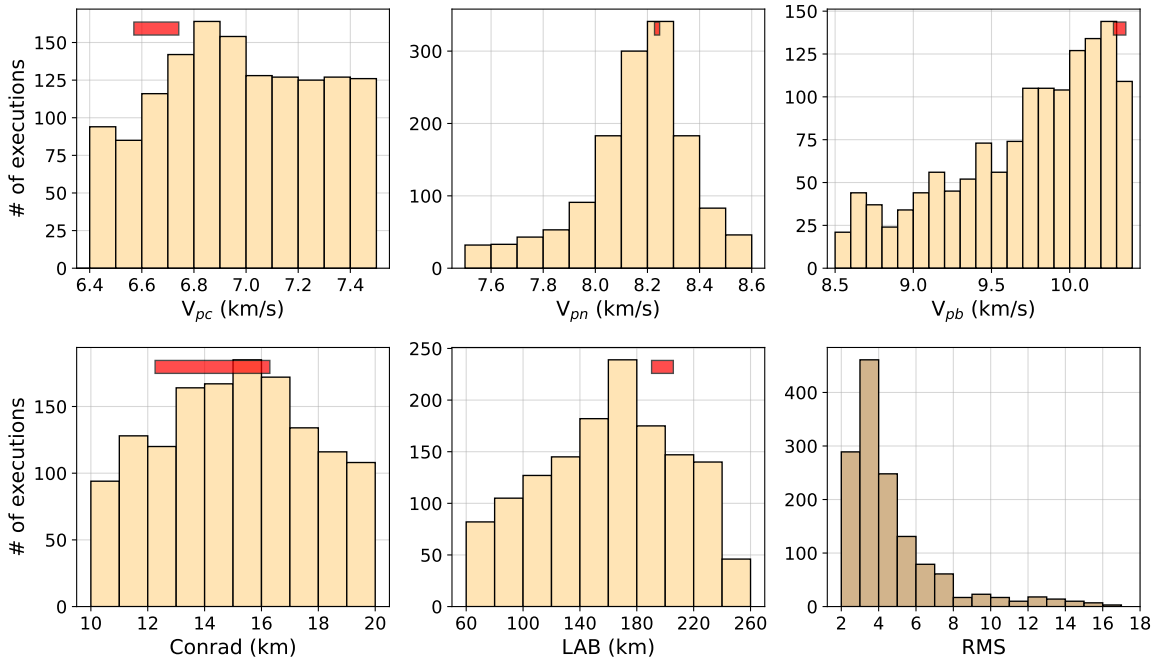


Figure 5.8: Histograms of the number of tests performed during global optimizations, referring to each of the optimized variables. Each red bar indicates the region of the median value regarding the 10 local optimizations, added and subtracted one standard deviation.

Figure 5.8 also presents the histogram of the RMS of all the solutions obtained through global optimization, which was able to find solutions with a minimum RMS of 2.5 seconds. Although this value is better than the RMS value of the NewBR model (3.39 s), it was still higher than the average RMS obtained by local optimizations (1.34 s), which varied between 1.26 s and 1.48 s.

For a final validation of the BRA23 model, Figure 5.9 presents the graph of 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 individual residuals presented in the same figure. The BRA23 velocity model has a final RMS value of 2.462 s, compared to the RMS value of 3.390 s for the NewBR model.

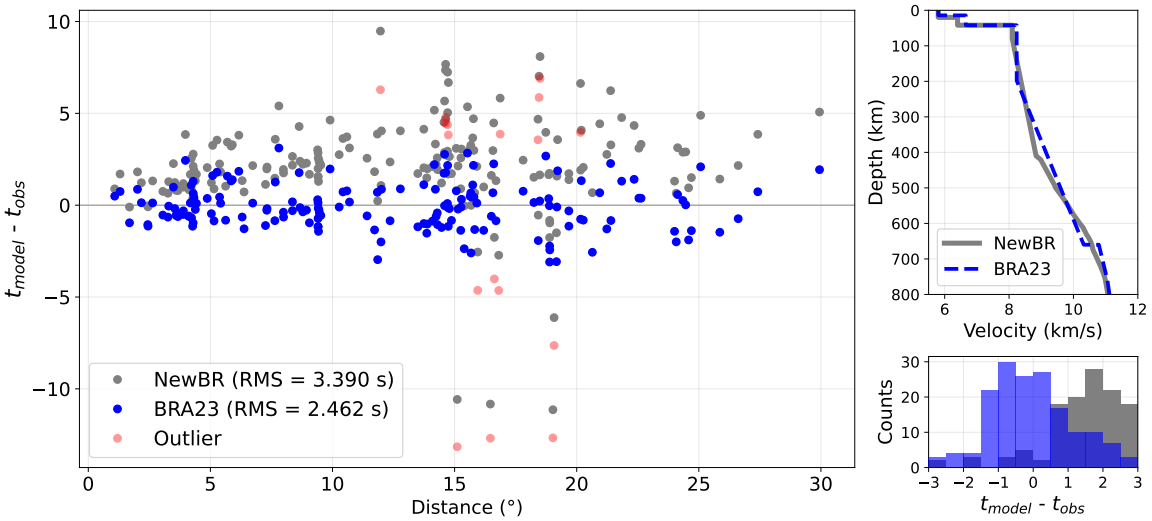


Figure 5.9: Comparison of time residuals ($t_{model} - t_{obs}$) using models BRA23 and NewBR. The proposed new model tends to present an average close to zero, unlike the NewBR model, which presents a trend of positive residuals. The points in red correspond to outliers, not used during local optimization processes.

The dispersion of points is explained by the great variability of data, which corresponds to events in different regions of Brazil. As it is a continental country, fitting all the data into a single 1D model results in greater dispersion, but makes it feasible to use it in the process of nucleating new origins in SeisComP.

On the other hand, even though it is a model based only on travel times, there is a compatibility between the lithosphere thickness of the model and the tomography results in Brazil (Ciardelli et al., 2022; Heit et al., 2007). The velocities of the lower crust and upper mantle, as well as the thickness of the upper/lower crust boundary, were compatible with the results of Mooney et al. (1998), being the same as the values proposed by the author for Archean terrains.

The 1D BRA23 velocity model is included in the Appendices. The travel time table used by SeisComP can be obtained through the package *TauP* (Crotwell et al., 1999).

5.5 Locatable detections (2014 - 2021)

A "locatable detection" is an automatic event generated by the software, whether it is real or false (as presented in "List of Abbreviations and Terms"). It is 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 5.10 shows the locatable detections obtained through the original SeisComP, using the database from 2014 to 2021.

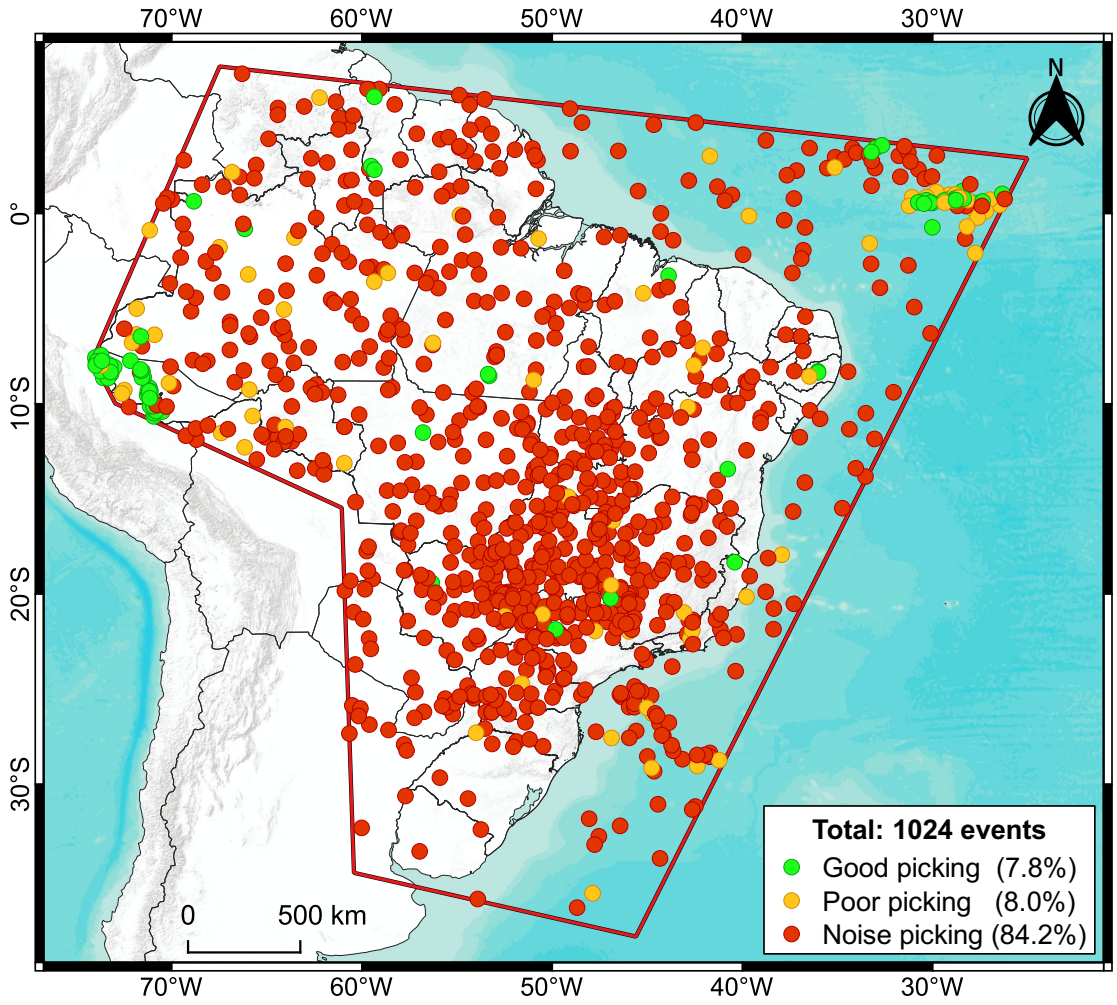


Figure 5.10: Locatable detections (2014-2021) obtained using the original SeisComP.

Of the total of 1024 locatable detections in the period from 2014 to 2021, only 80 (7.8%) were evaluated as satisfactorily located seismic events, which are mostly associated with events with plate 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 5.11 presents the locatable detections obtained through the software with the proposed modifications, using the same database that generated the previous figure.

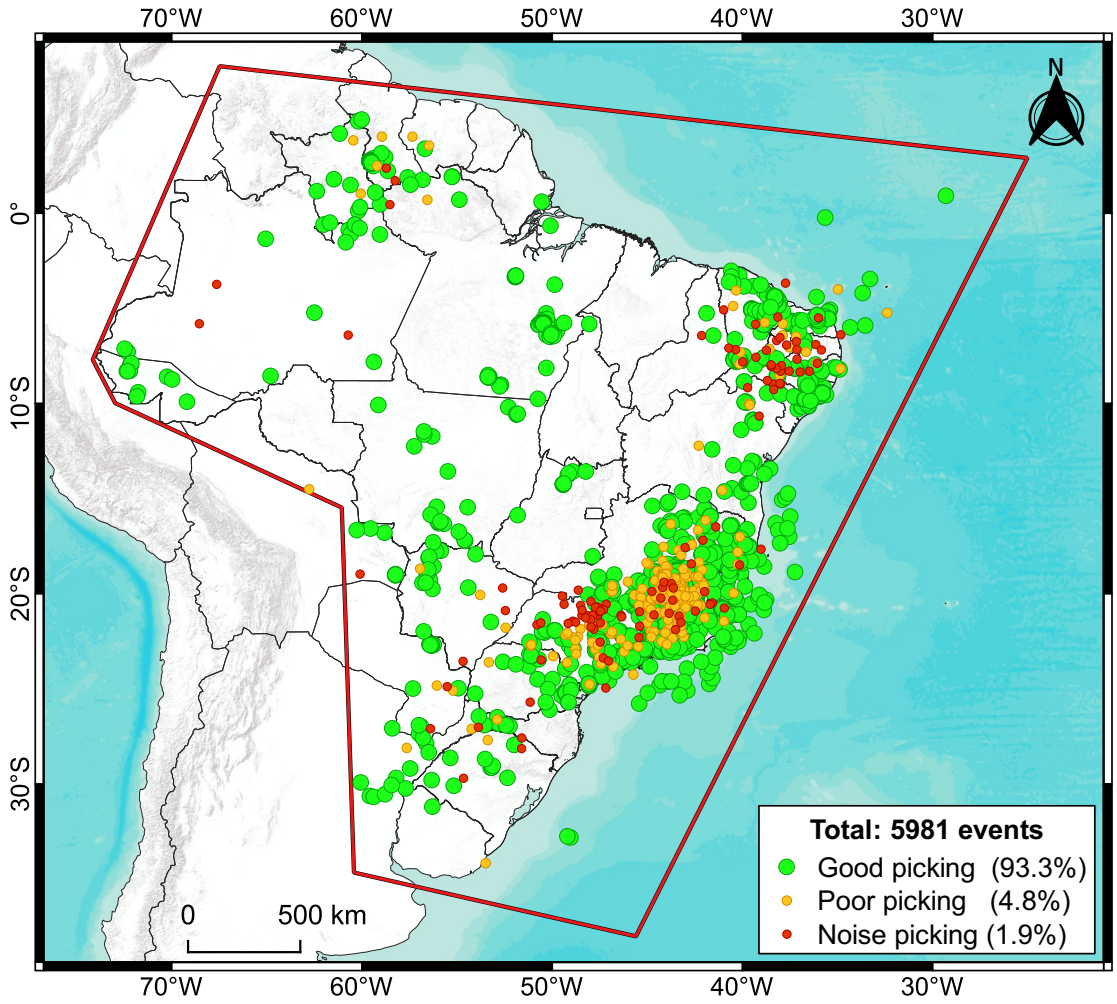


Figure 5.11: Locatable detections (2014-2021) obtained using the modified SeisComP.

Of the total of 5981 locatable detections from 2014 to 2021 using the modified SeisComP, 5580 (93.3%) were considered real events, with good wave arrival detections 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 negatively affected their epicenters. Finally, 112 detections (1.9%) were classified as false positives ("*noise picking*"), that is, were not related to real seismic events and with picks associated with local noises at each station.

5.6 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 with the catalog following criteria based on the times of origin and geographic coordinates. Events concurrent with 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 5.12 presents the map with the locatable detections obtained through the original SeisComP and that are concurrent with events in the RSBR seismic catalog.

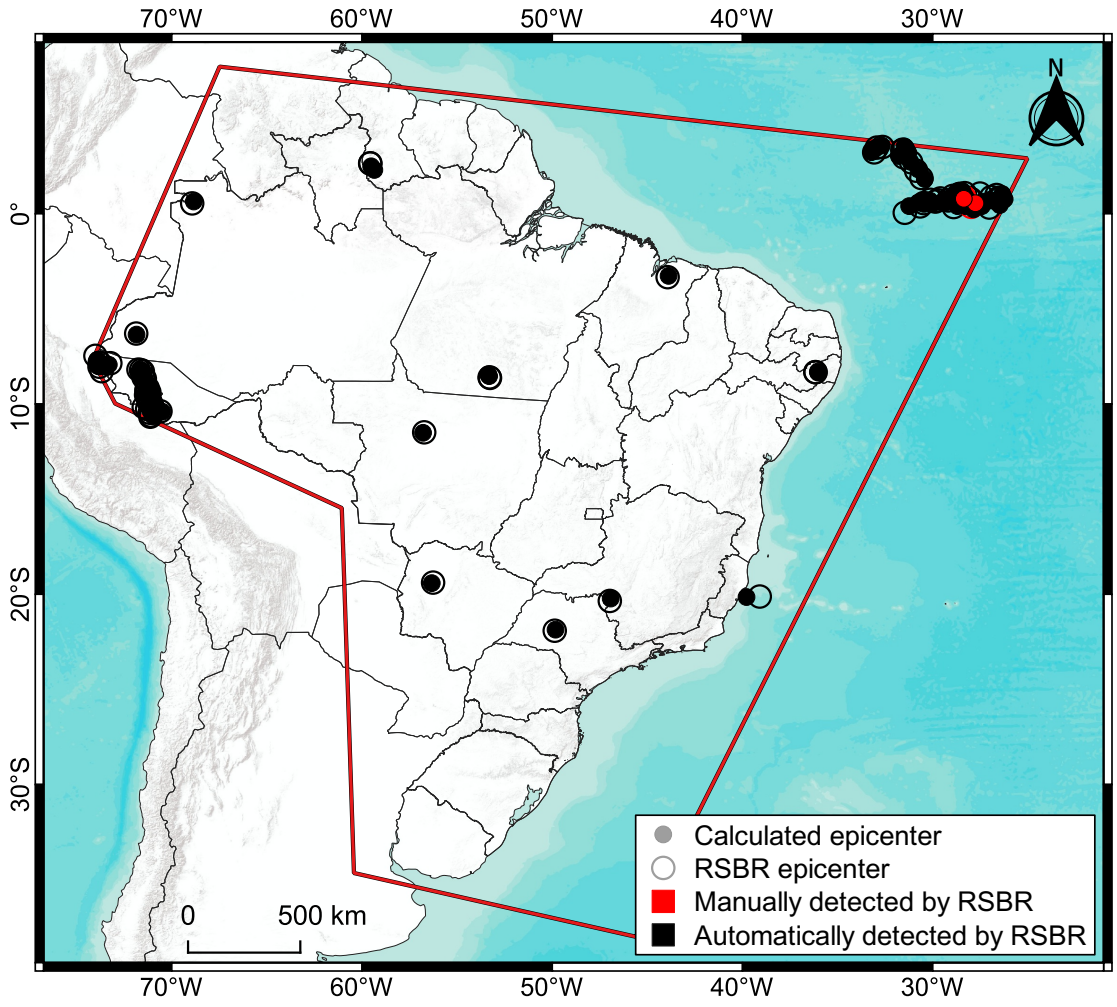


Figure 5.12: Epicenters of the 78 locatable detections concurrent with the RSBR seismic catalog from 2014 to 2021, using the original SeisComP.

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 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 the Motivation section, it was mentioned that the RSBR catalog has 106 automatic events, a greater number than shown in Figure 5.12. This is due to the fact that, for a certain period, the SeisComP at the IAG-USP Seismological Center had two pickers operating simultaneously, one of them with filter parameters modified by the system maintainers. Because of that, it was possible to detect a few more events than would have been detected with only the original picker.

Figure 5.13 shows the map with the locatable detections obtained through the modified SeisComP and which are concurrent with events in the RSBR seismic catalog.

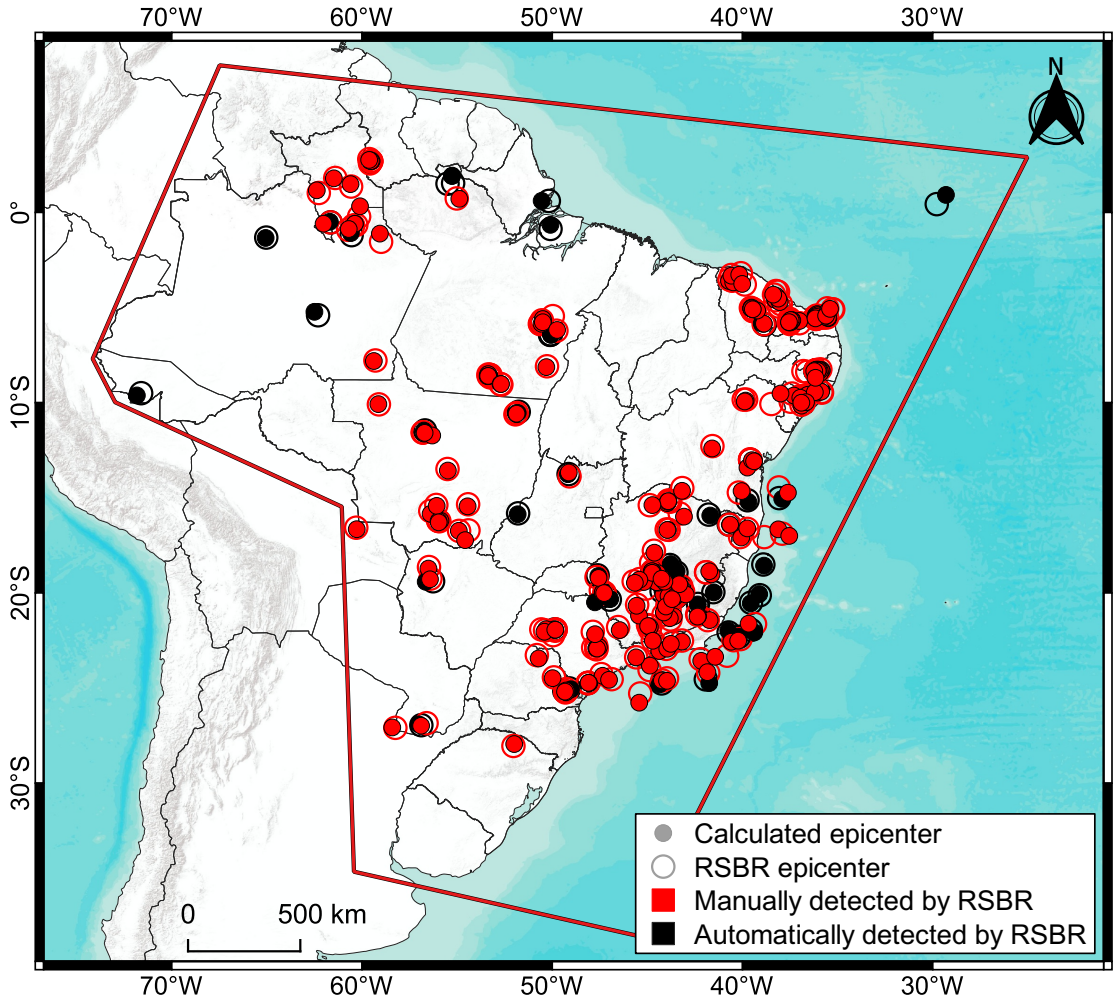


Figure 5.13: Epicenters of the 292 locatable detections concurrent with the RSBR seismic catalog from 2014 to 2021, using the modified SeisComP.

In total, there are 292 seismic events concurrent with 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 recent seismic events that occurred in Guyana.

6. Discussions

6.1 Locatable detections

With the proposed modifications, it was possible to observe a significant increase in the number of locatable detections during the period from 2014 to 2021, as well as events concurrent with the RSBR seismic catalog. Figure 6.1 compares the percentages of each qualitative classification of the locatable detections, considering the modified SeisComP and its original version.

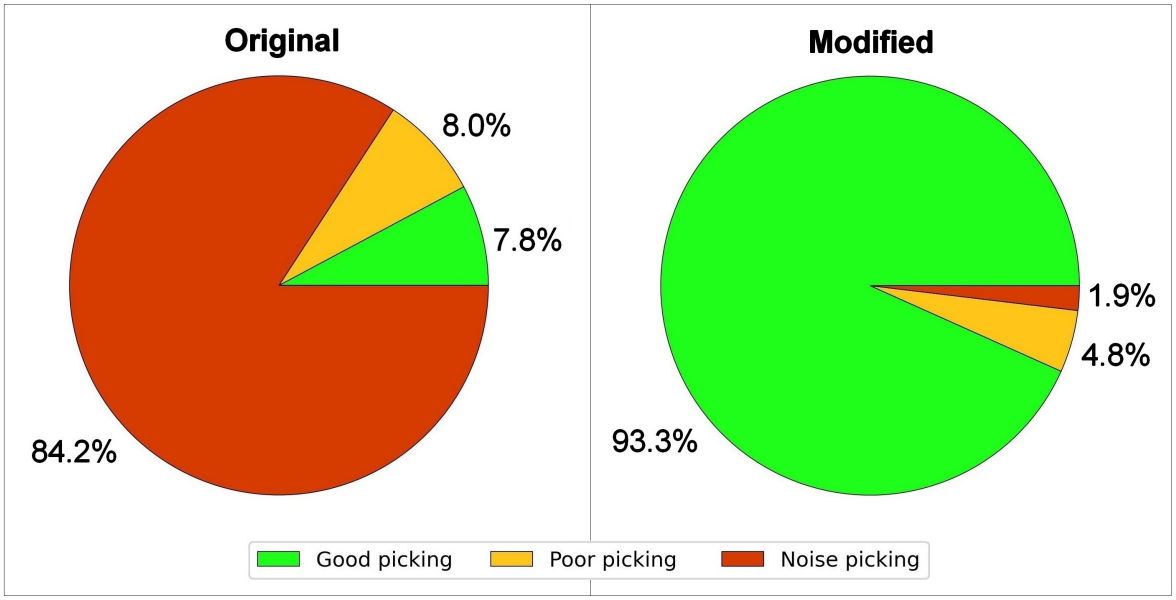


Figure 6.1: Pie charts regarding percentages of classified locatable detections in the period from 2014 to 2021. On the left, the percentages referring to the 5981 detections obtained through the modified SeisComP. On the right, the percentages referring to the 1024 detections obtained through the original SeisComP.

In addition to the significant increase in the number of locatable detections, it is also worth mentioning that the proposed modifications increased the number of positively classified events. In total, detections classified as "*good picking*" increased from 80 to 5580, while detections classified as "*noise picking*" decreased from 862 to 112.

With the decrease in the percentage of false positives, the annual average of erroneous events decreased from 108 to 14, which can facilitate the work of operators responsible for verifying the events located automatically. In the same way, the increase of real events located in a satisfactory way or even those that have incorrect picks (but are still associated with real events) provides a better effectiveness in the work of manual validation of events.

6.2 Events concurrent with the RSBR catalog

Regarding the events concurrent with the RSBR catalog, the modifications allowed an increase from 78 to 292 events located automatically and which are also included in the catalog. As previously mentioned, the criteria for defining the concurrency of a locatable detection with an event from the RSBR seismic catalog were its geographic coordinates and its origin time. It was defined that concurrent events would be considered those that presented a distance of up to 100 km in relation to the epicenters of the events in the catalog, as well as a difference of up to 15 seconds in the origin times.

To verify that these criteria are stable, that is, if there is no significant variation in the number of concurrent events when varying such metrics, Figure 6.2 presents a graph indicating the number of events concurrent with the catalog according to the variation of both criteria.

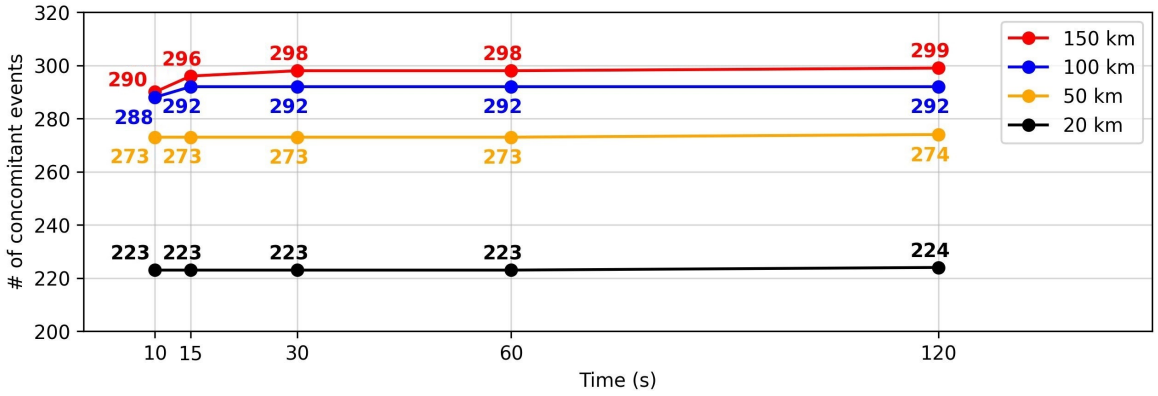


Figure 6.2: Number of events concurrent with the RSBR catalog after changes to SeisComP, considering different combinations of concurrency criteria.

The variation in the maximum allowed difference between the origin times does not show significant changes in the number of concurrent events, while the variation in the maximum allowed distance between the epicenters has a greater impact on the number of events. Therefore, the definition of the premise that a locatable detection is concurrent with the catalog becomes valid if there is a maximum distance of 100 km between its epicenters and a difference of up to 15 seconds in its origin times.

The concurrency of the events was considered using only the results obtained automatically, with no manual changes in the origins after their locations. Poorly positioned picks in real events can significantly influence both parameters (coordinates and origin time), therefore, it is understood that after a manual evaluation of the automatically located origins, the number of events concurrent with the catalog should increase.

To verify some of the parameters of the events concurrent with the catalog, the following figures show comparisons between the 292 concurrent events obtained through the modified SeisComP and the 78 concurrent events located with the original SeisComP.

Figure 6.3 shows that the modifications in SeisComP allowed the detection of seismic events with minimum magnitudes of up to M1.5, while most of the events concurrent with the catalog have magnitudes between M2.0 and M3.5. Without the modifications, the software was only able to detect earthquakes of magnitudes greater than M3.5.

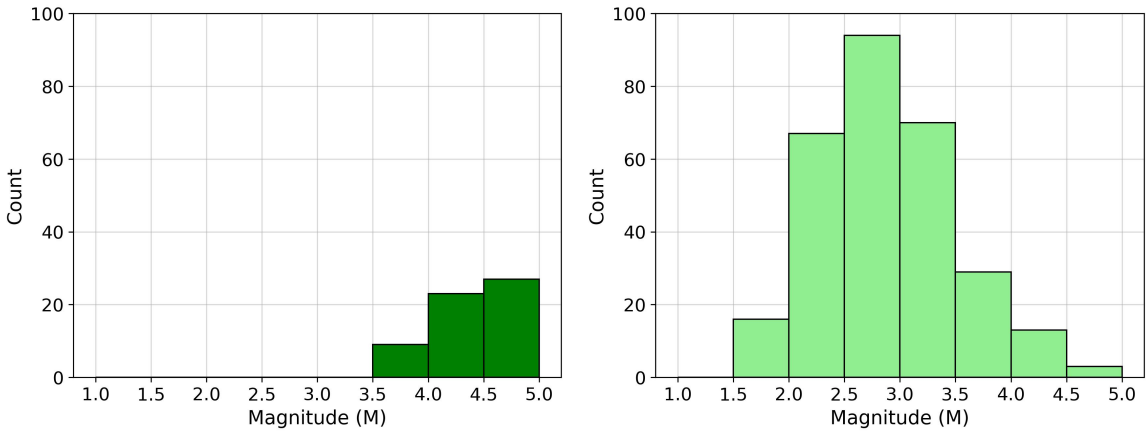


Figure 6.3: Magnitude histograms of the events concurrent with the RSBR catalog obtained through the original SeisComP (left) and the modified version (right).

Figure 6.4 and Figure 6.5 show the distributions of the RMS of origins and time residuals of their wave arrivals, respectively.

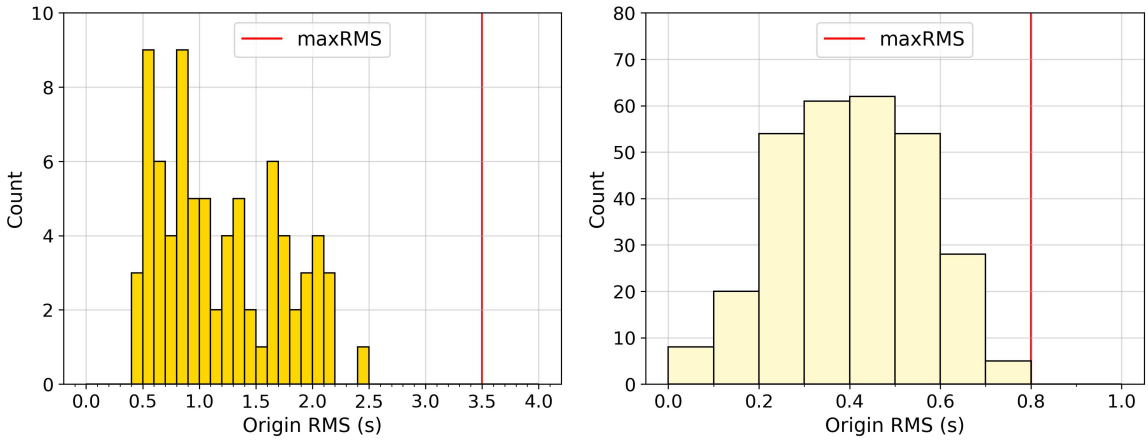


Figure 6.4: RMS histograms of the origins of events concurrent with the RSBR catalog obtained through the original SeisComP (left) and the modified version (right). Bars are 0.1s wide in both graphs.

The modified SeisComP generated origins concurrent with the catalog with RMS between 0 and 0.8 seconds, while the original SeisComP located concurrent origins with RMS between 0.4 and 2.5 seconds (Figure 6.4). Regarding the time residuals of arrivals (Figure 6.5), the same pattern is repeated: after the modifications, the arrivals mostly presented absolute residuals below 1 second, while the unmodified SeisComP presented several wave arrivals with residuals of a few seconds. The improvement in the residuals of arrivals converges with the results obtained during the development of the BRA23 model, as previously shown in Figure 5.9.

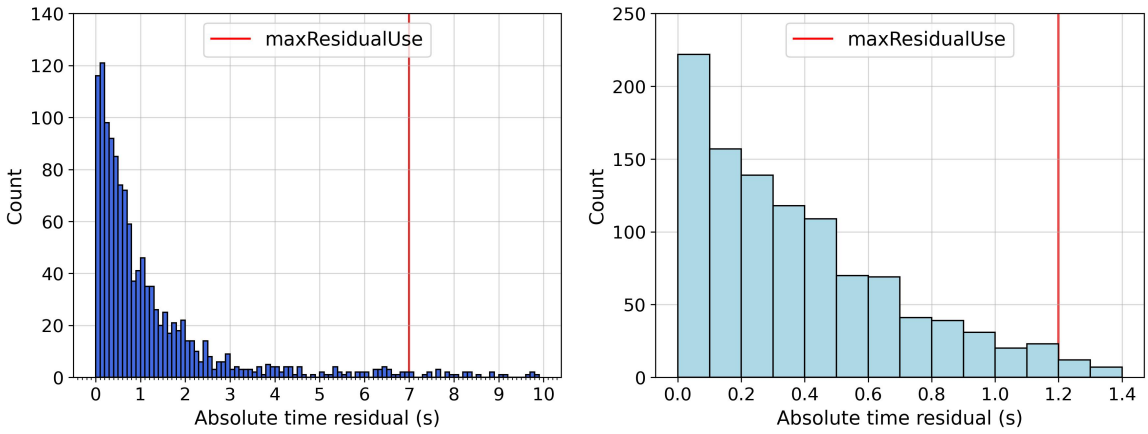


Figure 6.5: Absolute time residual histograms of arrivals of events concurrent with the RSBR catalog, obtained through the original SeisComP (left) and the modified versiom (right). Bars are 0.1s wide in both graphs.

The new travel times table used by SeisComP (BRA23 velocity model) and the proposed grids reduced the arrivals' time residuals and, consequently, the RMS of the located automatically origins. The greater accuracy in picking, as a result of changes made to *scautopick*, also played a significant role in reducing these residuals.

The red vertical lines in Figures 6.4 and 6.5 correspond, respectively, to the parameters *maxRMS* (maximum residual for an origin to be accepted) and *maxResidualUse* (maximum residual of an arrival so it can contribute to locating an origin). These parameters were considerably reduced after the modifications.

Finally, Figure 6.6 presents the distribution of numbers of arrivals used for locating each origin. With the proposed modifications, it was possible to locate events concurrent with the catalog with only 5 wave arrivals, something that was previously not possible due to limitations defined in the software source code.

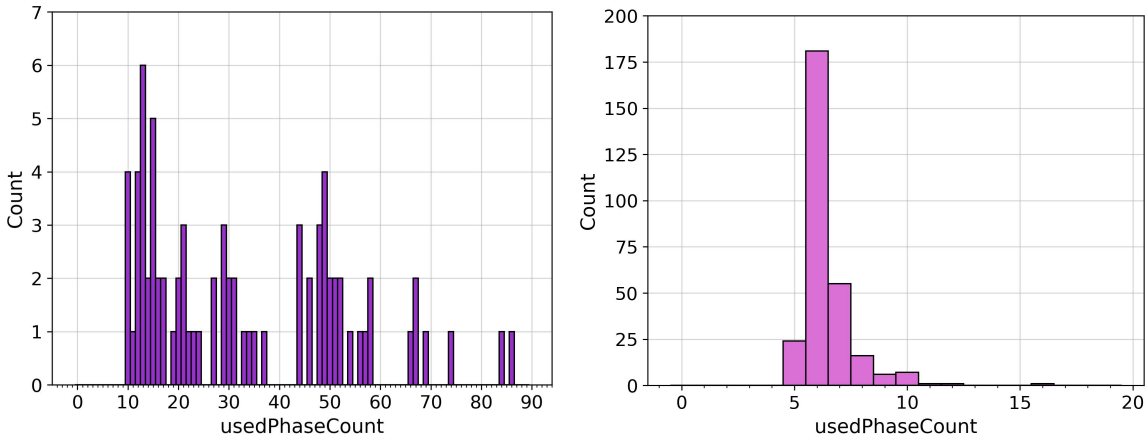


Figure 6.6: Histograms of the numbers of arrivals used (*usedPhaseCount*) in the location of events concurrent with the RSBR catalog, obtained through the original SeisComP (left) and the modified version (right).

6.3 Estimates of automatic detectability after modifications

Combining the medians of the maximum distances at which seismic events in the area of interest were recorded (Figure 2.3) with the density of seismographic stations in Brazil in the year 2021 (Appendices), it was possible to estimate the current automatic detectability of seismic events in the region of interest, based on their magnitudes (Figure 6.7).

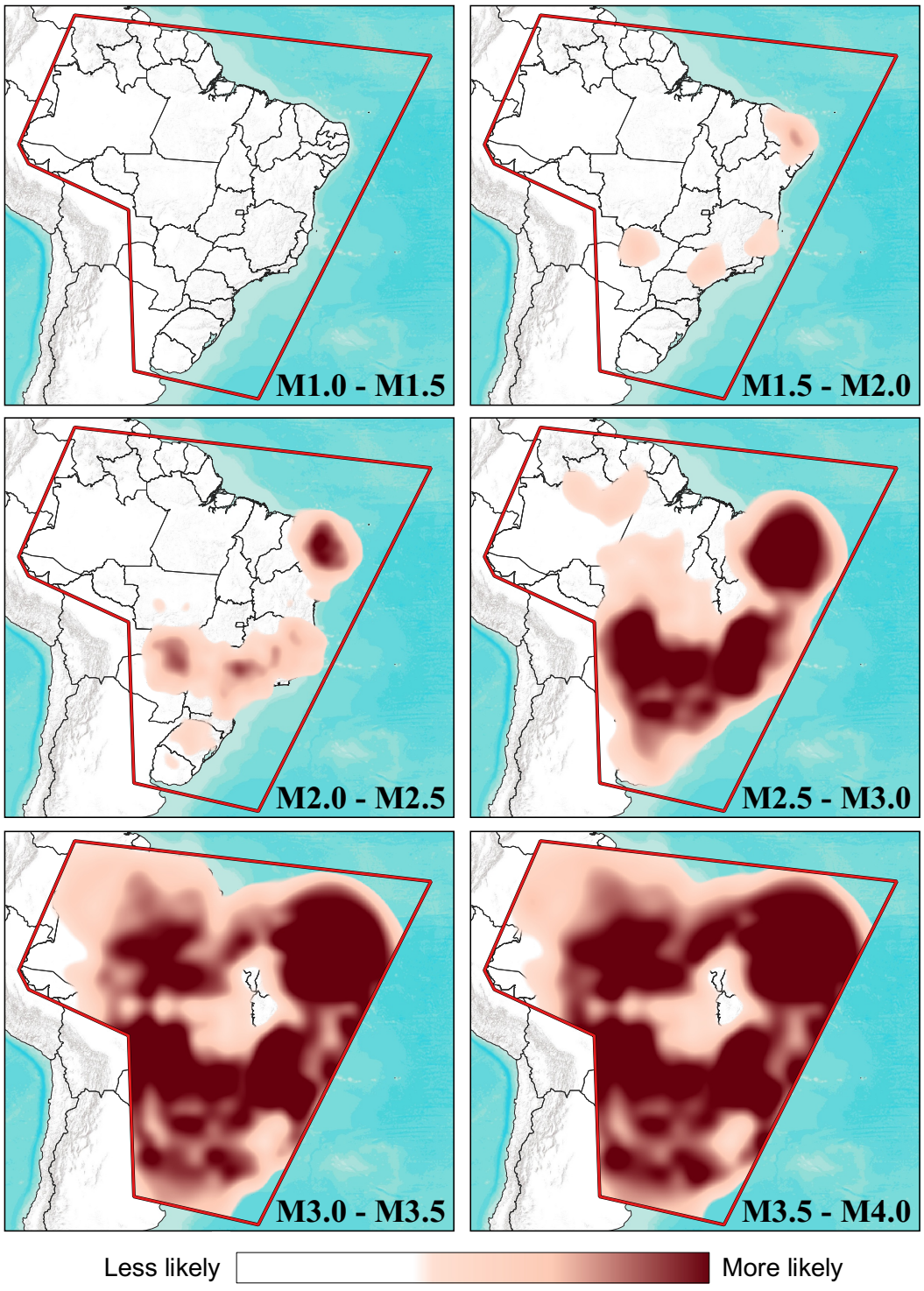


Figure 6.7: Estimates of automatic detectability of seismic events in Brazil. Qualitative color scale indicates probability of automatic detection.

To check the consistency between the estimates of automatic detectability and the results obtained, Figure 6.8 presents the same maps as in the previous figure, but with the overlapping of events from 2014 to 2021 concurrent with 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.

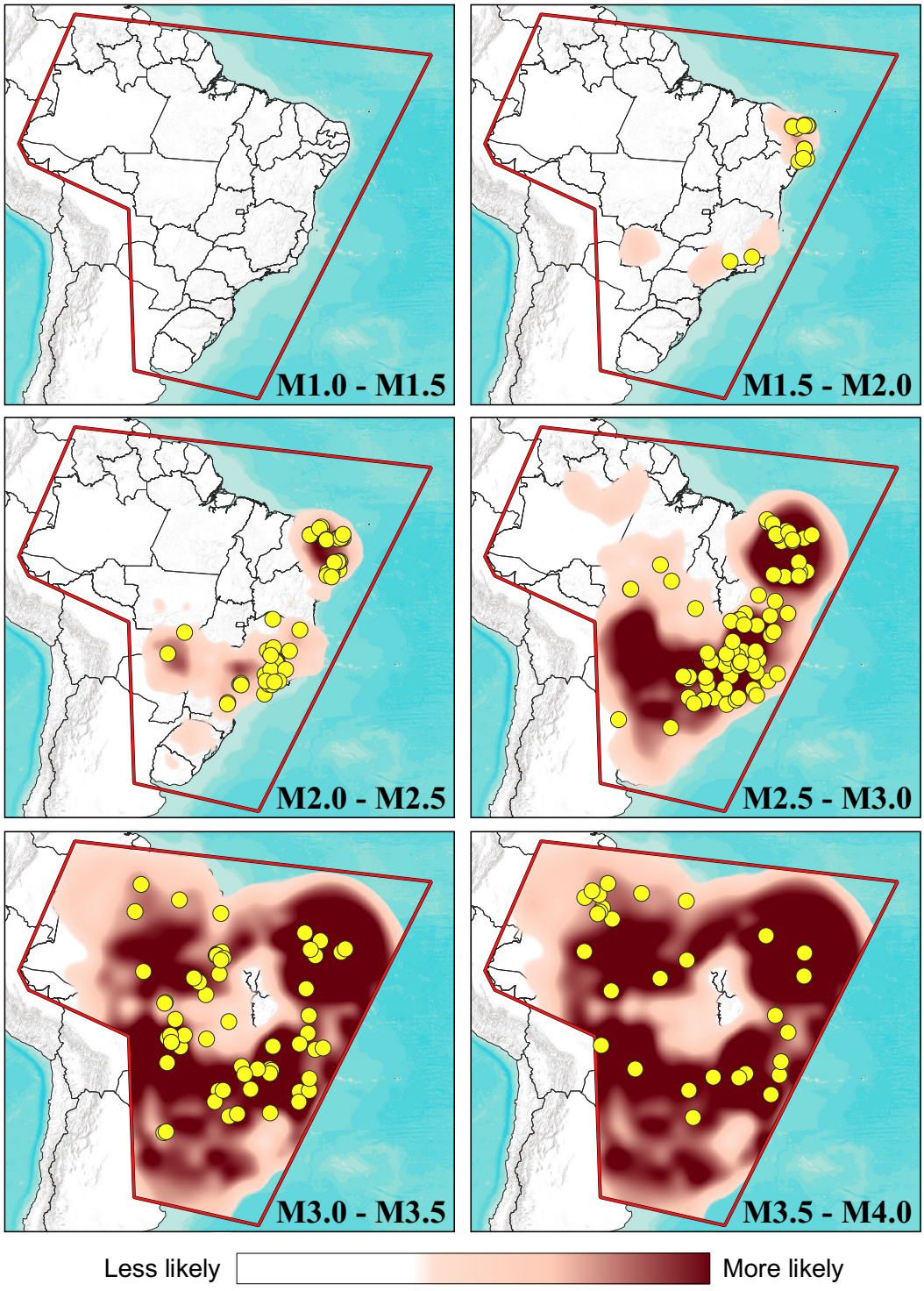


Figure 6.8: Estimates of automatic detectability of seismic events in Brazil, with the overlapping of events concurrent with the RSBR catalog obtained after modifications in SeisComP. Qualitative color scale indicates probability of automatic detection.

The automatic detectability estimation maps were obtained through some processing stages. In the first, the region of interest was discretized into points spaced 0.5° apart. It was considered that there are two "limiting distances" for an event to be detected by a station: *(i)* the epicentral distance itself and *(ii)* the maximum distance allowed by SeisComP's grid, as discussed in "Grid file (*grid.conf*)". Thus, for each magnitude range, circles centered on these points were drawn, with radii equal to the medians of the maximum recorded distances included in the RSBR catalog (Figure 2.3) or equal to the maximum distance allowed by the closest point in the grid (d_{max}), whichever is smaller (limiting).

Then, the number of stations within each circumference was associated with its central point. In other words, if an event occurs at a given point, we sought to calculate how many stations would be able to detect it, considering the limitations of its epicentral distance and the maximum distance imposed by SeisComP for that point. Finally, the maps were obtained through krigings using the number of stations associated with each point. The qualitative color scale of the maps does not associate any color to regions with counts smaller than 5 distinct stations (the minimum number for an event to be automatically detected).

The automatic detectability estimation maps of seismic events in Brazil were obtained using the 2021 set of stations. Considering that the database used in this work covers the period from 2014 to 2021, it was not expected that the comparison between what was automatically detected and the estimates obtained would be fully adherent, since the density of stations varied considerably over this period (see maps in Appendices).

For example, the estimated detectability map for a magnitude range between 2.5 and 3.0 shows a higher probability of detection in the Midwest region, mainly in the state of Mato Grosso do Sul. However, after the modifications, no seismic events of such magnitudes were recorded in this region (Figure 6.8), possibly due to the fact that the temporary subnetwork XC (responsible for increasing the density of stations in this region) was not in operation throughout the period from 2014 to 2021. Another factor that may have impacted the detection efficiency in this region was the increase in the *triggerOn* value of noisy stations in the XC subnetwork, as discussed in "*Bindings*".

Also, the automatic detectability maps do not take into account the seismic activity of each region of the Brazilian territory. From these maps, it is inferred the regions in which there are greater possibilities of automatic detection when using the modified SeisComP if a seismic event occurs, based only on the number of stations within "limiting distances".

Despite these facts, the coherence between the maps of automatic detectability estimates and the epicenters of the 292 events concurrent with the catalog (obtained after the modifications) shows that such estimates are compatible with the new detection reality of SeisComP.

6.4 Events in the RSBR catalog not automatically located

Despite the significant increase in locatable detections after the proposed modifications, it was not expected that all seismic events from the RSBR catalog would be located automatically. The main reason is the fact that many events have emergent wave arrivals that are difficult to differentiate from local noise, even after applying frequency filters, as the example in Figure 6.9 shows. These events were located and added to the catalog manually and in many cases less than five stations were used to locate them.

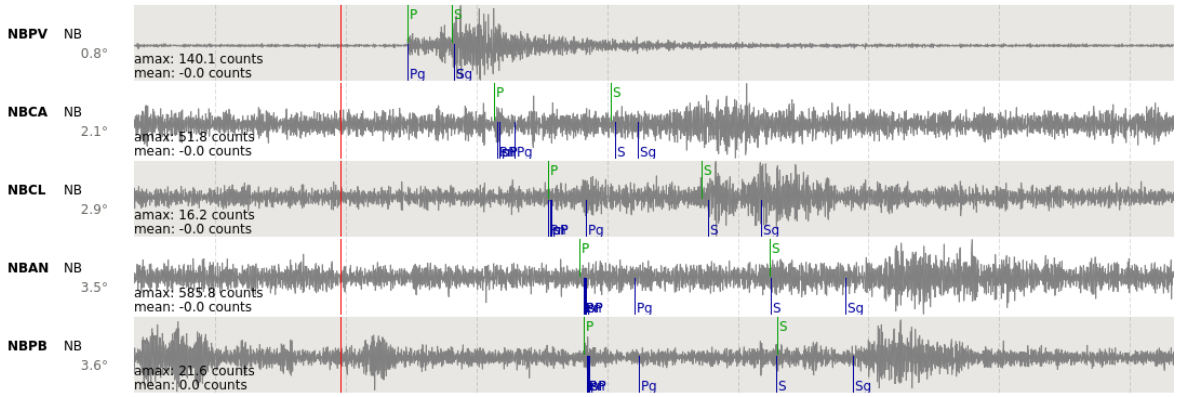


Figure 6.9: Example of emergent wave arrivals from event "usp2019jnud" (M2.1), which are difficult to differentiate from local noise. Picks (green) were carried out manually, possibly with the aid of theoretical arrivals (blue), based on the travel time table used.

A second explanation for the lack of expectation of detection of all seismic events from the RSBR catalog is the absence of part of the waveform data in the IAG-USP database. During the work, synchronizations were carried out with the individual databases maintained by the participating RSBR institutions, in order to obtain the greatest possible amount of data during the period of interest. However, it was found that several seismic events from the RSBR catalog occurred at times when some stations did not have data stored in the database used, which makes their automatic detection difficult.

The lack of data may be associated with a period of non-operation/transmission of the station's data, or the unavailability of such data by the responsible institutions. Figure 6.10 presents an example of lack of data from five stations of the NB network (managed by UFRN) at the time of the occurrence of a seismic event from the RSBR catalog.

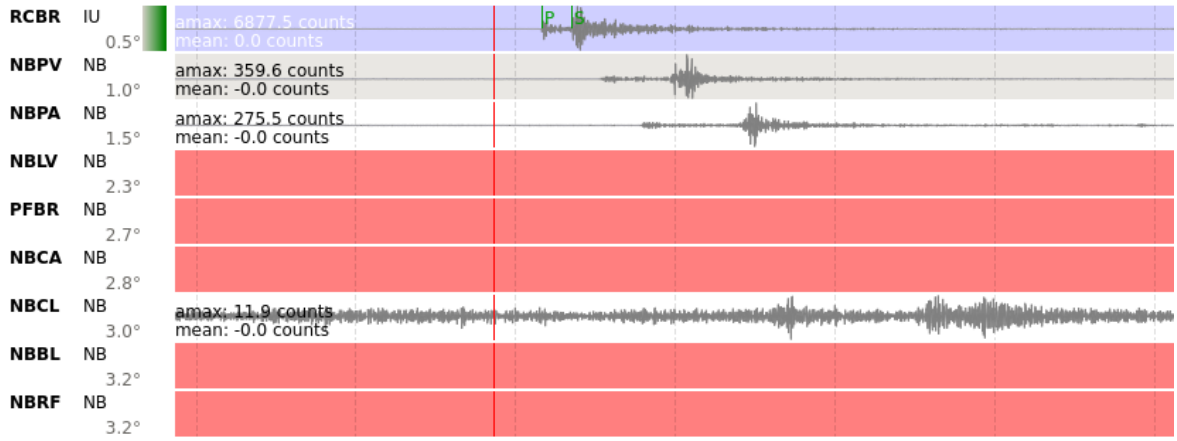


Figure 6.10: Example of lack of waveform records in the IAG-USP database, resulting in a smaller number of detectable events. The event is identified by the code "*usp2019scqm*" and occurred on 09/15/2019, with magnitude M1.3, in João Câmara/RN.

Through an analysis of the seismographic records of the 193 events that occurred in 2019 included in the RSBR catalog, it was estimated that 52 (27%) could be detected automatically after the proposed modifications. Events presumably not possible to be detected have records in few stations or emergent P-wave arrivals (with low signal-to-noise ratio). After the modifications to SeisComP, 44 concurrent events were detected in the catalog in 2019, while the original software was able to detect only 7. It should be emphasized that poorly positioned automatic picks significantly impact the event concurrency criteria, therefore, a manual evaluation of all automatically registered events can increase the number of events concurrent with the catalog.

Considering the catalog from 2014 to 2021 (1854 events in the region of interest), the proposed modifications resulted in 16% concurrency (292 events), while the original software recorded only 4% (78 events). This result indicates that the modifications allowed the initial estimate to be satisfactorily achieved, reinforcing the idea that it was not expected that all events in the catalog could have been located automatically.

The maps in Figure 6.11 show the epicenters of the events from 2014 to 2021 in the RSBR catalog that were not automatically detected by the modified SeisComP, compared with the previously presented automatic detectability estimation maps.

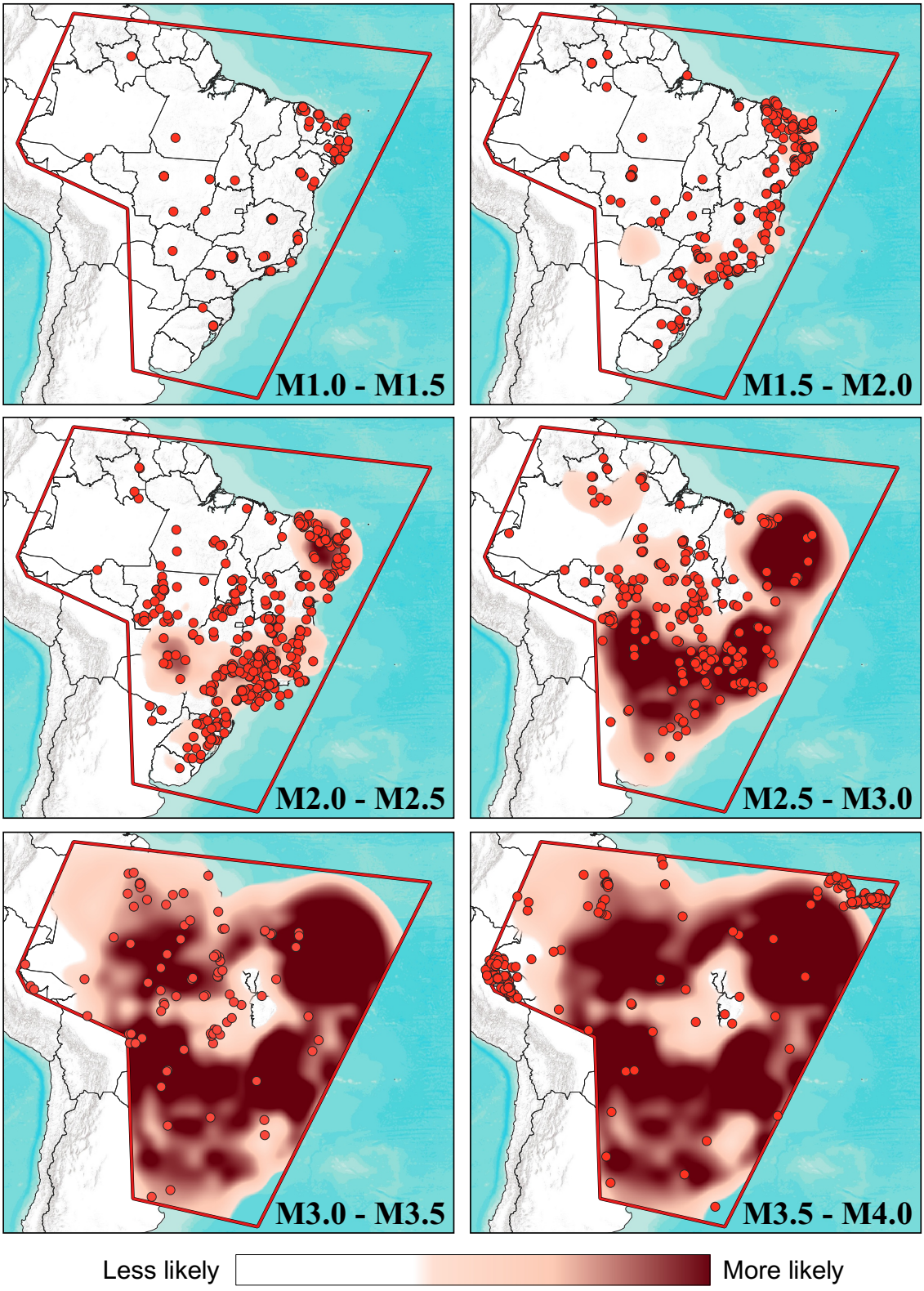


Figure 6.11: Estimates of automatic detectability of seismic events in Brazil, with overlapping of events from the RSBR catalog that were not automatically located after modifications to SeisComP. The qualitative color scale indicates the probability of automatic detection. The magnitude ranges indicated on each map refer to the events included in the RSBR catalog.

It was found that approximately 50% of the events not automatically detected are outside the regions of higher probability of automatic detection by the modified SeisComP. Once again, it is worth emphasizing that the maps of detectability estimates were calculated based on the set of stations in 2021. Thus, some regions in which the maps indicate a greater probability of automatic detection may not correspond to the reality of events that occurred in other years.

A large number of events in the Northeast Region were not automatically detected due to the lack of data from stations in that region, as shown in Figure 6.10. The other events, mainly in the North and Midwest regions, were not well recorded in a satisfactory number of stations.

The histogram in Figure 6.12 shows the distribution of the magnitudes of events from the RSBR catalog that were not automatically detected after modifications to SeisComP.

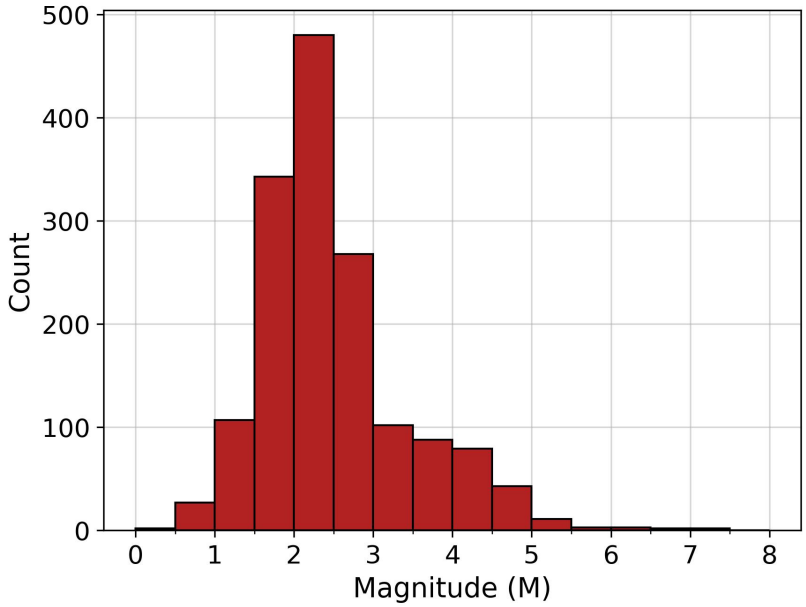


Figure 6.12: Magnitude histogram of events from the RSBR catalog that were not detected automatically after modifications to SeisComP.

Most events not automatically detected have magnitudes between M1.5 and M3.0, as seen in Figure 6.11. Despite the modifications providing greater detection of events in this magnitude range, the previously highlighted facts made it impossible for the software to be more efficient.

Modifications to SeisComP were proposed in order to find a satisfactory balance between true and false events. The number of real events automatically detected can increase if, for example, the end user sets a *minScore* value lower than 0.75, as defined in this work and discussed in "Function *_publishable(origin)*". The side effect is the greater probability of false origins being accepted, and the user must determine their tolerance level for false positives.

6.5 Possible real events not included in the catalog

We verified the detection of possible real seismic events that are not included in the RSBR catalog, indicating that it may be possible to increase the number of events in Brazil within the RSBR catalog when using the results obtained through modified SeisComP. Figure 6.13 presents two examples of seismic events located after the proposed modifications that are not included in the RSBR catalog.

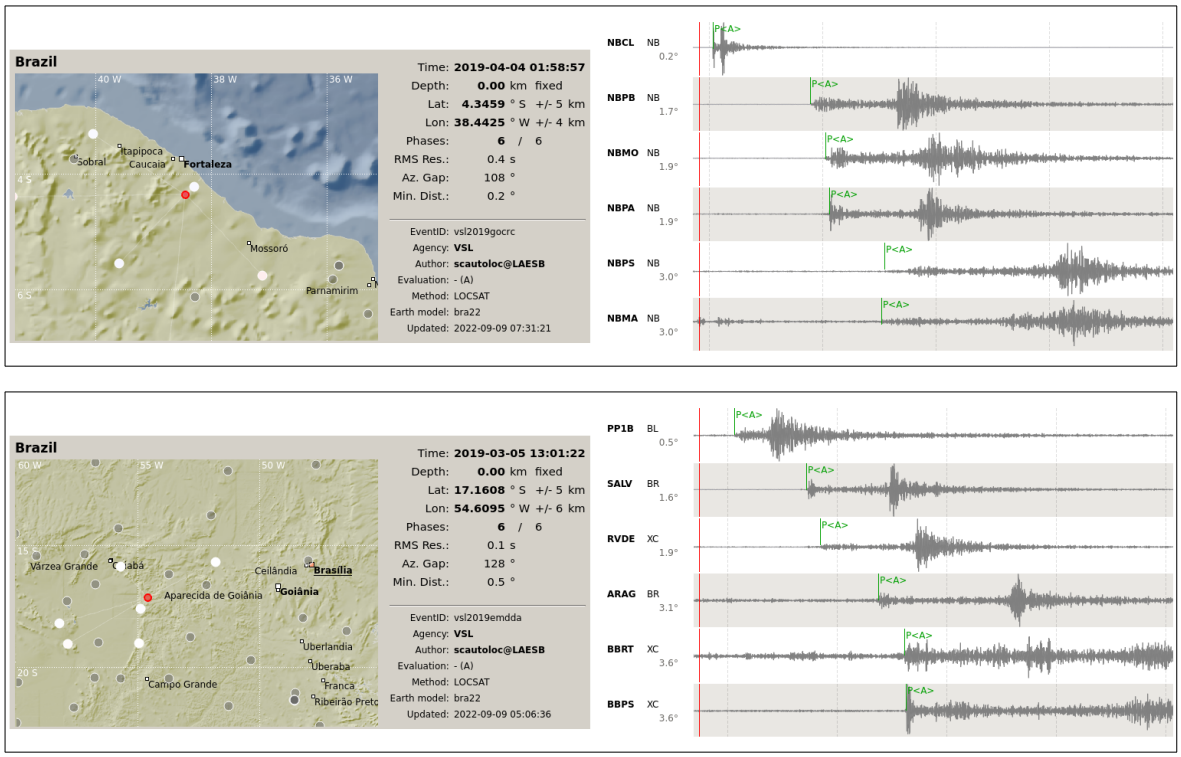


Figure 6.13: Example of two seismic events located automatically after the proposed modifications that are not included in the official RSBR catalog. Above, an event occurred on 04/04/2019 near Fortaleza/CE. Below, an event occurred on 03/05/2019 near Aparecida de Goiânia/GO.

Finally, the majority of the 5981 locatable detections obtained after modifications in the software refer to blasts in mines and quarries, which are not included in the RSBR seismic catalog since they are not natural or induced seismic events. Of these, most refer to blasts close to the city of Belo Horizonte/MG.

Figure 6.14 presents clusters of locatable detections considered "*good picking*" and probably associated with blasts, which occurred close to the cities of Parauapebas/PA and Belo Horizonte/MG, in known mine regions. On the top map, there are 124 locatable detections, while the map of the region near Belo Horizonte/MG presents a total of 4051. Together, the two maps represent about 70% of the origins obtained after the modifications, which is the main reason for the difference between the number of events included in the RSBR catalog (1854) and the number of origins obtained (5981).

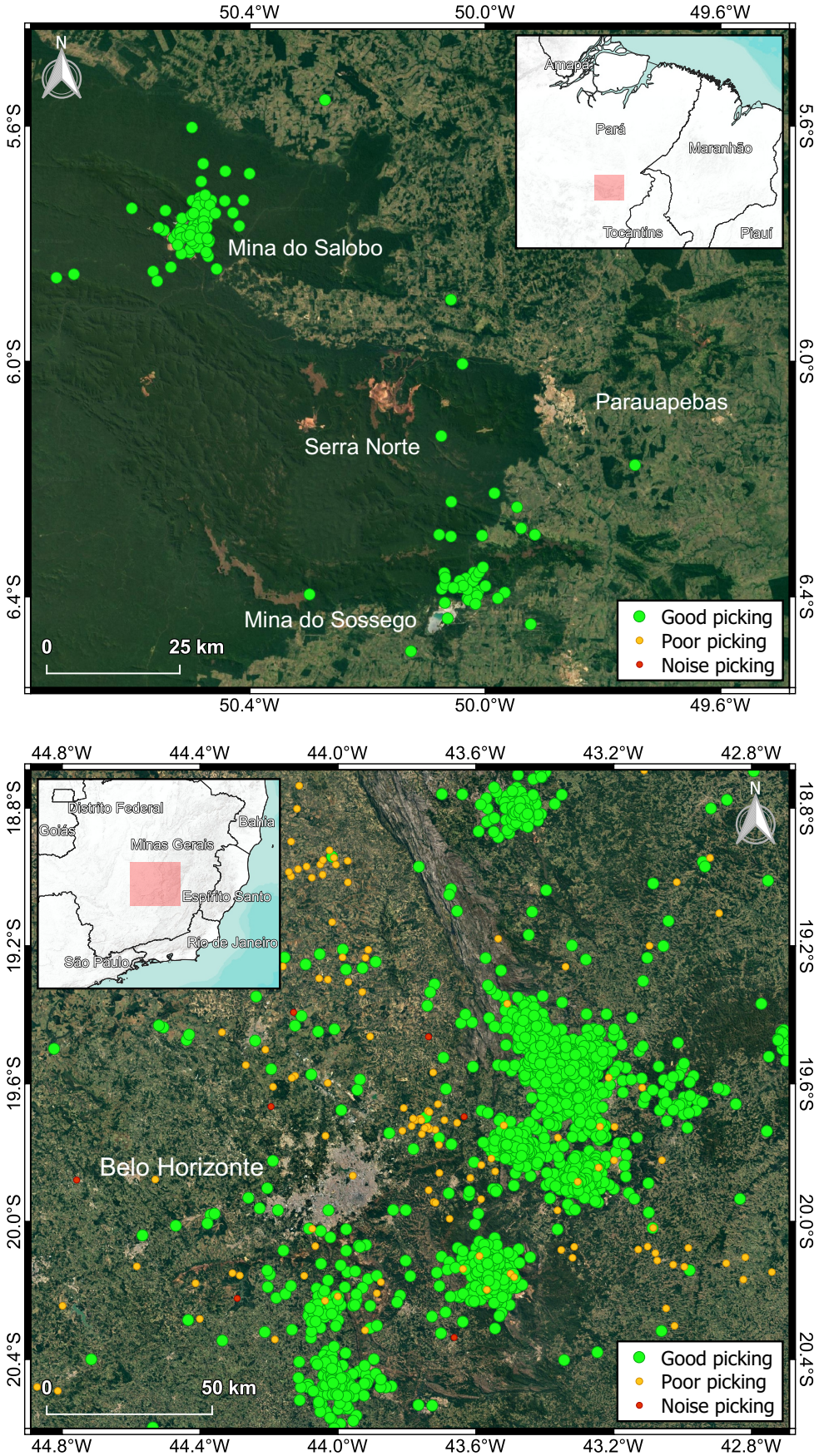


Figure 6.14: Clusters of locatable detections possibly associated with blasts in known mine regions. Above are the mines close to Parauapebas/PA and, below, clusters related to the various mines close to Belo Horizonte/MG.

7. Conclusions

SeisComP is a software widely used in research centers around the world, a reference in the acquisition, processing and storage of seismological data. Due to the fact that its development aimed at the automatic detection and location of seismic events on global scales, it is necessary to carry out several adjustments so that the software operates more adequately in a context of regional seismicity, as is the case of Brazilian seismicity.

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 calculations of scores that define whether an origin is accepted (*originScore*) have been reformulated to be based on Brazilian events and to remove arbitrary complexities from the equations. Among the origin scores, it is worth mentioning the changes made in the calculation of the *amplScore* parameter, related to the amplitude of a wave arrival within the origin. The proposed modifications introduce the concept of magnitude in the calculation of this score, in order to take into account the fact that the amplitudes in seismograms decrease with the increase of the epicentral distance. Still in the context of the nucleator, it is also proposed to use grids whose minimum number of picks and maximum distances for nucleation depend on the number of stations within a radius of 10 degrees from each grid point.

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).

Comparing the geographic coordinates and the origin times of the locatable detections with the same parameters of seismic events from the RSBR catalog, it was verified the concurrency of 292 events after the modifications. With the original software, only 78 detections concurrent with the catalog were registered. The main difference between both scenarios, in addition to the significant increase after the modifications, is the fact that the modified SeisComP allowed the detection of many more events in Brazilian territory, while the original SeisComP detected, for the most part, events with origins associated with plate boundaries (deep seismicity in the state of Acre and in the Mid-Atlantic ridge).

Still considering the events concurrent with the RSBR catalog, the modified SeisComP automatically detected events with magnitudes between M1.5 and M5.0 between 2014 and 2021. In comparison, the original SeisComP was only able to detect seismic events with magnitudes between M3.5 and M5.0 in the same region. The time residuals of arrivals and the RMS of origins showed significant reductions after the modifications, in addition to the possibility of recording origins with few stations, a common scenario in the context of the national seismographic network, which does not have a good density of stations throughout the Brazilian territory.

An estimate of the automatic detectability of events with the current set of stations (year 2021) was inferred, considering the fact that, with the proposed modifications, it is possible to automatically locate events with at least 5 stations. Qualitative maps of detectability were developed for different magnitude ranges, which were later compared with events concurrent with the RSBR catalog obtained after modifications. The overlapping of these events on the detectability maps showed a great concordance with the regions with the highest probability of detection for each magnitude range, indicating that the automatic detectability estimates are consistent with the new automatic detection capability of SeisComP.

The events included in the RSBR catalog which were not automatically detected refer mainly to events with emergent wave arrivals that do not differ from local noise, even after applying filters. Also, there was no data from some stations at the time some seismic events in the catalog occurred, another important factor that impairs the automatic detection and location of these events. This lack of data may be associated with a period of non-operation or data transmission, or even the unavailability of these data by the responsible institutions. However, when comparing the locations of events not automatically detected with the maps of automatic detectability estimates, most of these events are located in regions where RSBR is not expected to be able to perform automatic detections.

Between 2014 and 2021, the modified SeisComP detected thousands of blasts in known quarries and mines, which are not included in the RSBR catalog since they are not natural or induced seismic events. At least 70% of the origins obtained after the modifications are possibly associated with blasts, which is the main reason for the difference between the number of events in the RSBR catalog (1854) and the number of origins obtained after the modifications (5981). On the other hand, after the modifications, we identified possible

automatically detected natural seismic events that are not included in the RSBR catalog, indicating that there is the possibility of increasing the number of events in Brazil by using the results obtained in this work.

Finally, 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 data of the region of interest, and it is possible to reproduce them in 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.

The list of all modifications applied directly to SeisComP's source code, as well as the dissertation (both in Portuguese and English), the BRA23 velocity model, the grids calculated for each year and other files resulted from this work are available in a public repository[4] on the platform *Zenodo* (Salles, 2023).

[4]https://zenodo.org/record/7412143

References

Akaike, H. (1971). Autoregressive Model Fitting for Control. *Annals of The Institute of Statistical Mathematics - Ann. Inst. Statist. Math*, 23:163, DOI: 10.1007/BF02479221.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5):1521, ISSN: 0037-1106, DOI: 10.1785/BSSA0680051521.

Alsaker, A. et al. (1991). The ML scale in Norway. *Bulletin of the Seismological Society of America*, 81:379, DOI: 10.1785/BSSA0810020379.

Ardito, J. C. (2009). Determinação de epicentros regionais. *Trabalho de Graduação. Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo (IAG-USP)*.

Assumpção, M. (1983). A regional magnitude scale for Brazil. *Bulletin of the Seismological Society of America*, 73(1):237, ISSN: 0037-1106.

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

Assumpção, M. et al. (2014). Intraplate seismicity in Brazil. *Intraplate earthquakes*, pages 50–71.

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

Beyreuther, M. et al. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, 81(3):530.

Bianchi, M. et al. (2012). The Brazilian Seismographic Network: Historical Overview and Current Status. *Summary of the Bulletin of the International Seismological Centre*, 49:70, DOI: 10.5281/zenodo.998851.

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

Ciardelli, C. et al. (2022). Adjoint waveform tomography of South America. *Journal of Geophysical Research: Solid Earth*, 127(2):e2021JB022575.

Clark, A., Evans, P., and Strollo, A. (2014). FDSN recommendations for seismic network DOIs and related FDSN services. *version 1.0. Technical report. International Federation of Digital Seismograph Networks*.

Crotwell, H. P., Owens, T. J., and Ritsema, J. (1999). The TauP Toolkit: Flexible seismic travel-time and ray-path utilities. *Seismological Research Letters*, 70(2):154, ISSN: 0895-0695, DOI: 10.1785/gssrl.70.2.154.

Duda, S. J. and Nuttli, O. W. (1974). Earthquake magnitude scales. *Geophysical surveys*, 1(4):429.

Heit, B. et al. (2007). An S receiver function analysis of the lithospheric structure in South America. *Geophysical Research Letters*, 34(14).

Helmholtz-Centre Potsdam - GFZ German Research Centre for Geosciences and gempa GmbH (2008). The SeisComP seismological software package. GFZ Data Services. DOI: 10.5880/GFZ.2.4.2020.003.

Herrin, E. (1968). Seismological tables for P-phases. *Bulletin of the Seismological Society of America*, 60:461.

IRIS (2010). SEED Manual - Version 2.4. FDSN Website. Accessed: March 12, 2021.

Jeffreys, H. and Bullen, K. E. (1940). Seismological tables.

Kanamori, H. (1983). Magnitude scale and quantification of earthquakes. *Tectonophysics*, 93(3):185, ISSN: 0040-1951, DOI: 10.1016/0040-1951(83)90273-1. Quantification of Earthquakes.

Kennett, B. and Engdahl, E. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105:429, DOI: 10.1111/j.1365-246X.1991.tb06724.x.

Kennett, B. L., Engdahl, E., and Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1):108.

Li, X. et al. (2016). Identifying P phase arrival of weak events: The Akaike Information Criterion picking application based on the Empirical Mode Decomposition. *Computers and Geosciences*, 100, DOI: 10.1016/j.cageo.2016.12.005.

Lopez, C. M. (2021). Parameter optimization of automatic phase detection and picking algorithms - Application in São Paulo University Seismological Center and Colombian National Seismic Network. *Dissertação de Mestrado. Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo (IAG-USP).*

Montagner, J.-P. and Kennett, B. (1996). How to reconcile body-wave and normal-mode reference Earth models. *Geophysical Journal International*, 125(1):229.

Mooney, W. D., Laske, G., and Masters, T. G. (1998). CRUST 5.1: A global crustal model at 5× 5. *Journal of Geophysical Research: Solid Earth*, 103(B1):727.

Pirchiner, M. et al. (2011). The BRAzilian Seismographic Integrated Systems (BRASIS): Infrastructure and Data Management. *Annals of Geophysics*, 54(1):17, DOI: 10.4401/ag-4865.

Richter, C. F. (1935). An instrumental earthquake magnitude scale. *Bulletin of the Seismological Society of America*, 25(1):1, ISSN: 0037-1106, DOI: 10.1785/BSSA0250010001.

Rivadeneyra, C. et al. (2019). An updated crustal thickness map of Central South America based on receiver function measurements in the region of the Chaco, Pantanal, and Paraná Basins, Southwestern Brazil. *Journal of Geophysical Research: Solid Earth*, 124, DOI: 10.1029/2018JB016811.

Salles, V. (2023). SeisComP software optimization focused on automatic detection and localization of regional seismic events in Brazil. (Complementary Data) to the Master's Thesis, DOI: 10.5281/zenodo.7412143. Zenodo Public Repository.

Schorlemmer, D. et al. (2011). QuakeML: Status of the XML-based seismological data exchange format. 54:59.

St-Onge, A. (2011). Akaike Information Criterion applied to detecting first arrival times on microseismic data. *Society of Exploration Geophysicists*, page 1658.

Storn, R. and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341.

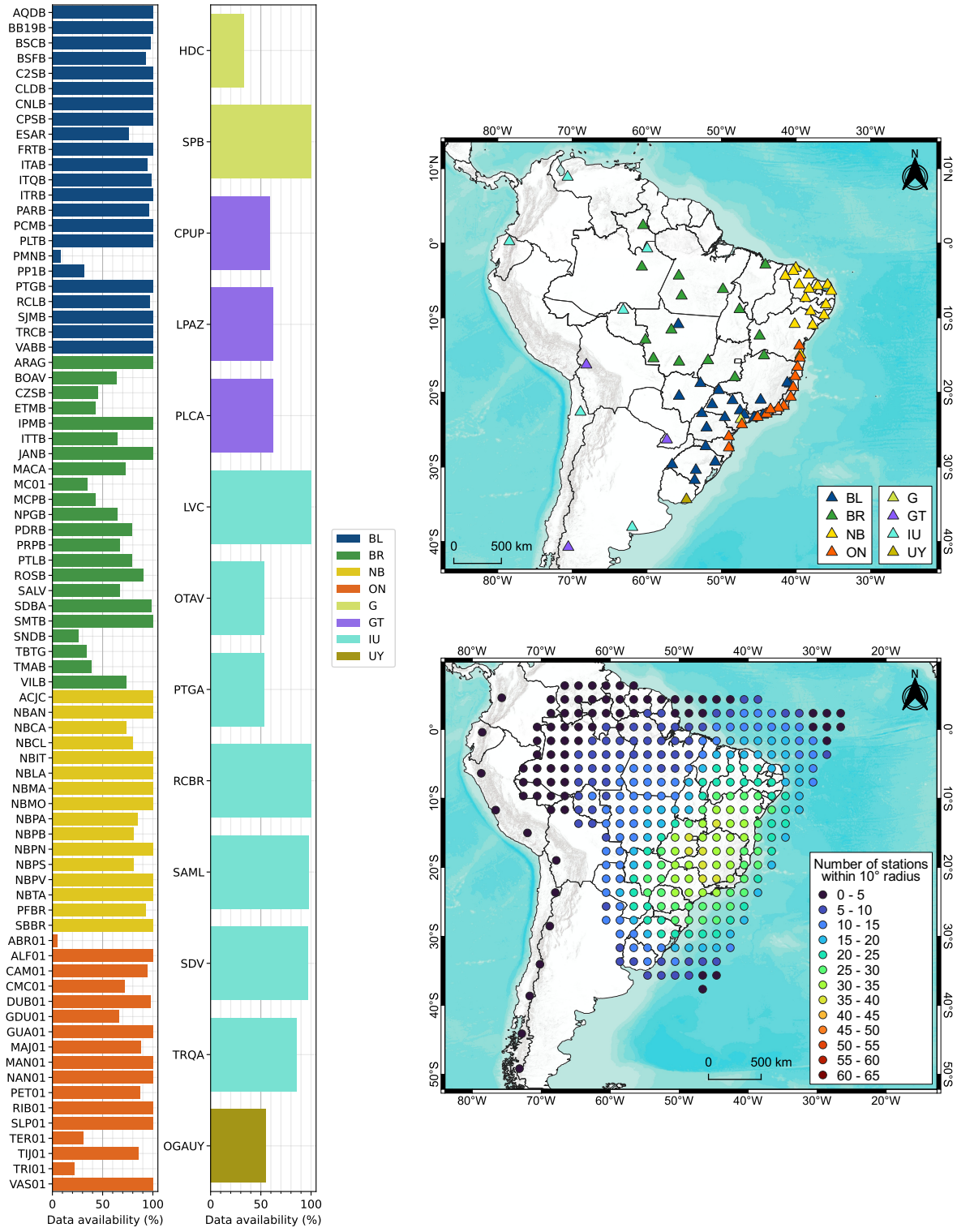
Virtanen, P. et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261, DOI: 10.1038/s41592-019-0686-2.

Yadav, A. K. and Mishra, P. K. (2014). Method for detection of first phase in seismogram based on Akaike Information Criterion (AIC) function. *International journal of engineering research and technology*, 3.

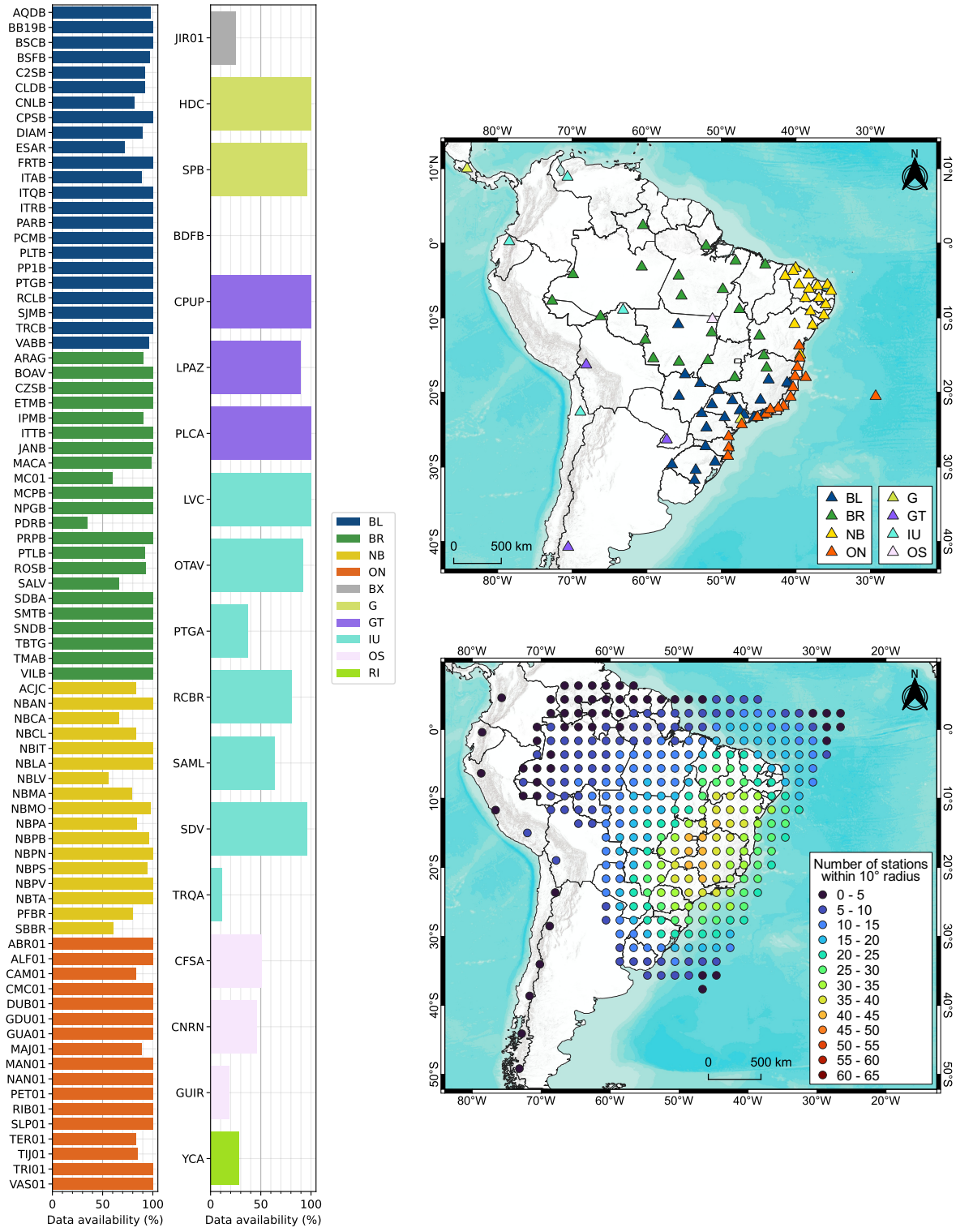
Appendices

APPENDIX 1

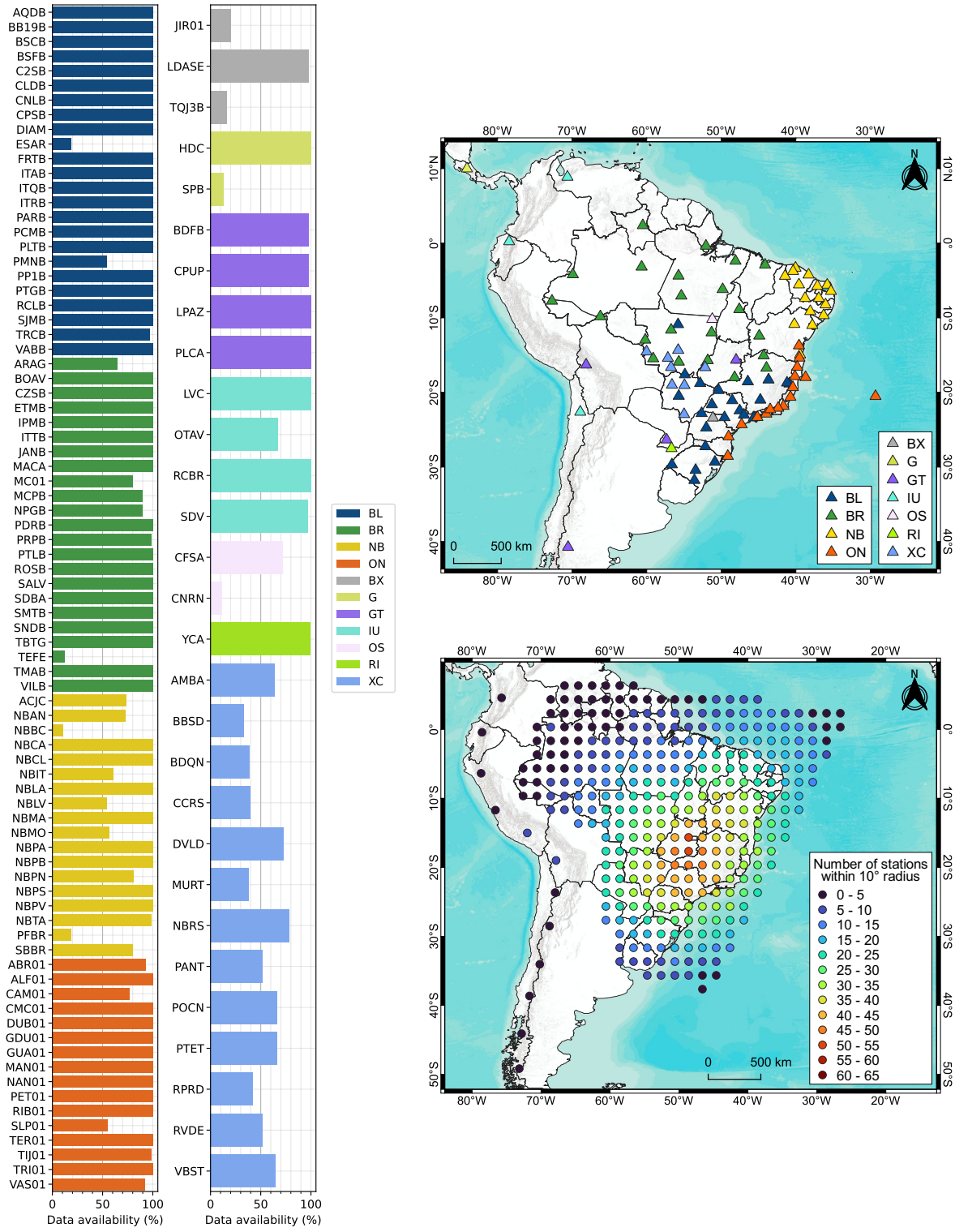
Grids - 2014 to 2021



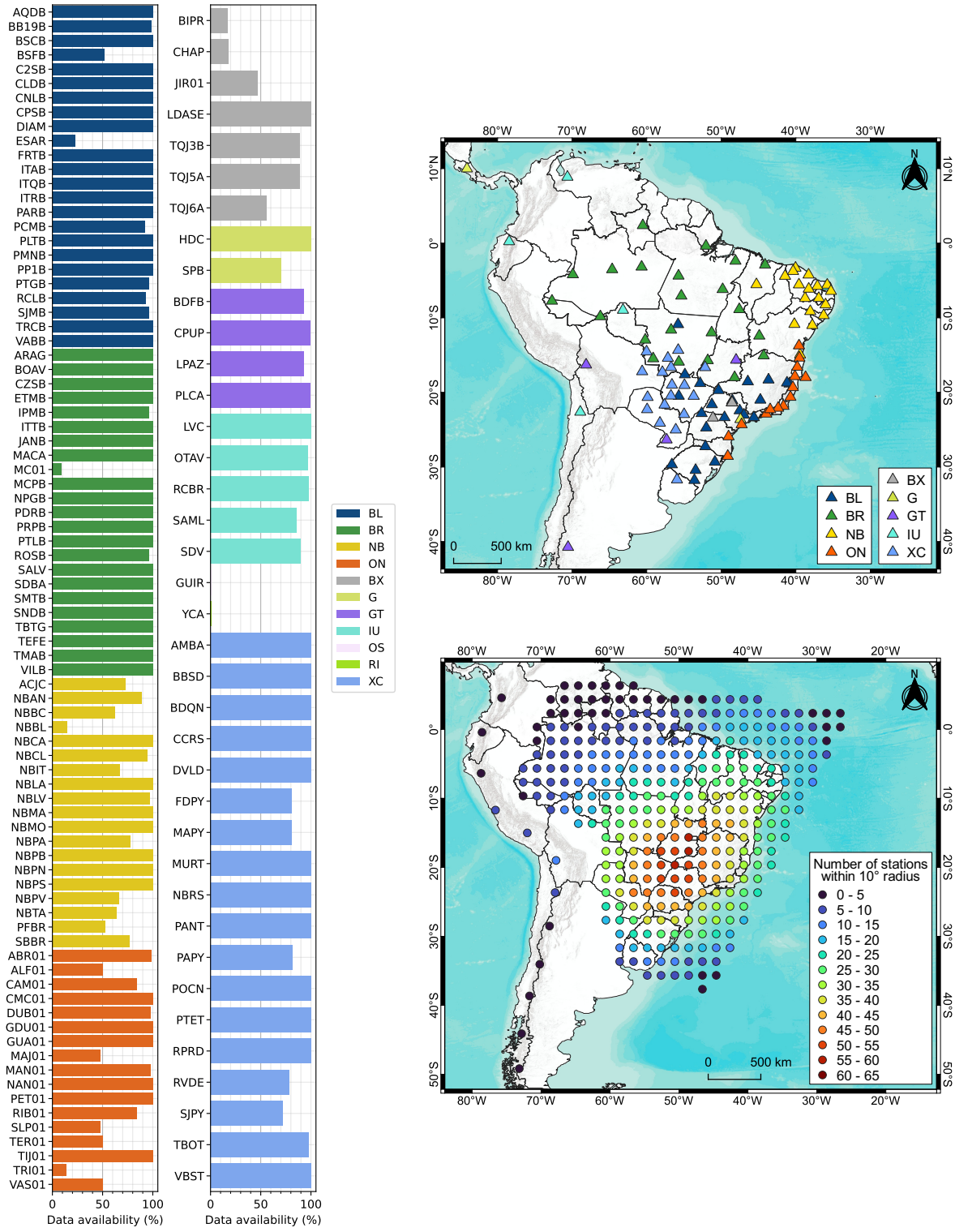
Left: Availability of data in the form of waves, recorded in 2014, from each station included in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



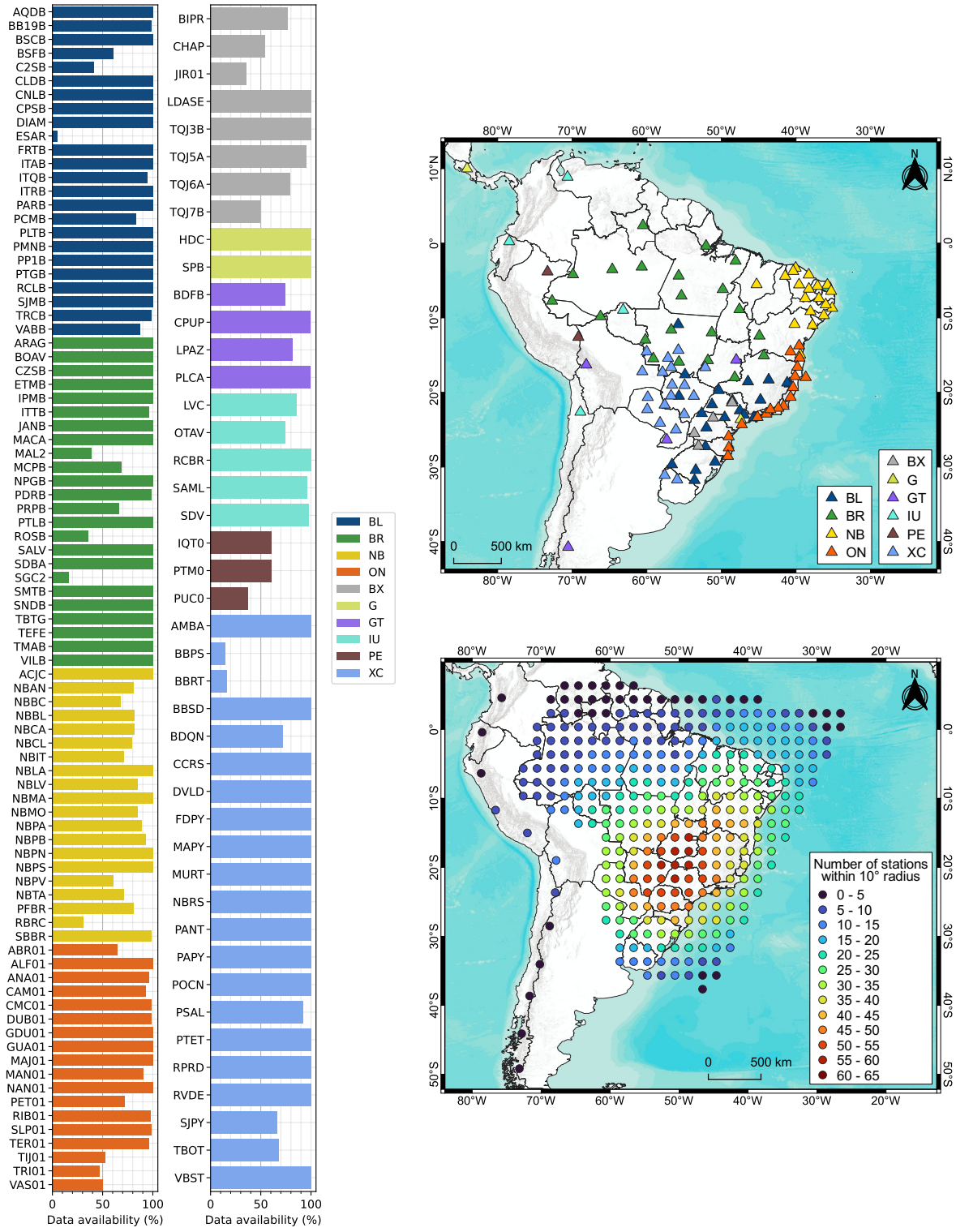
Left: Availability of data in the form of waves, recorded in 2015, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



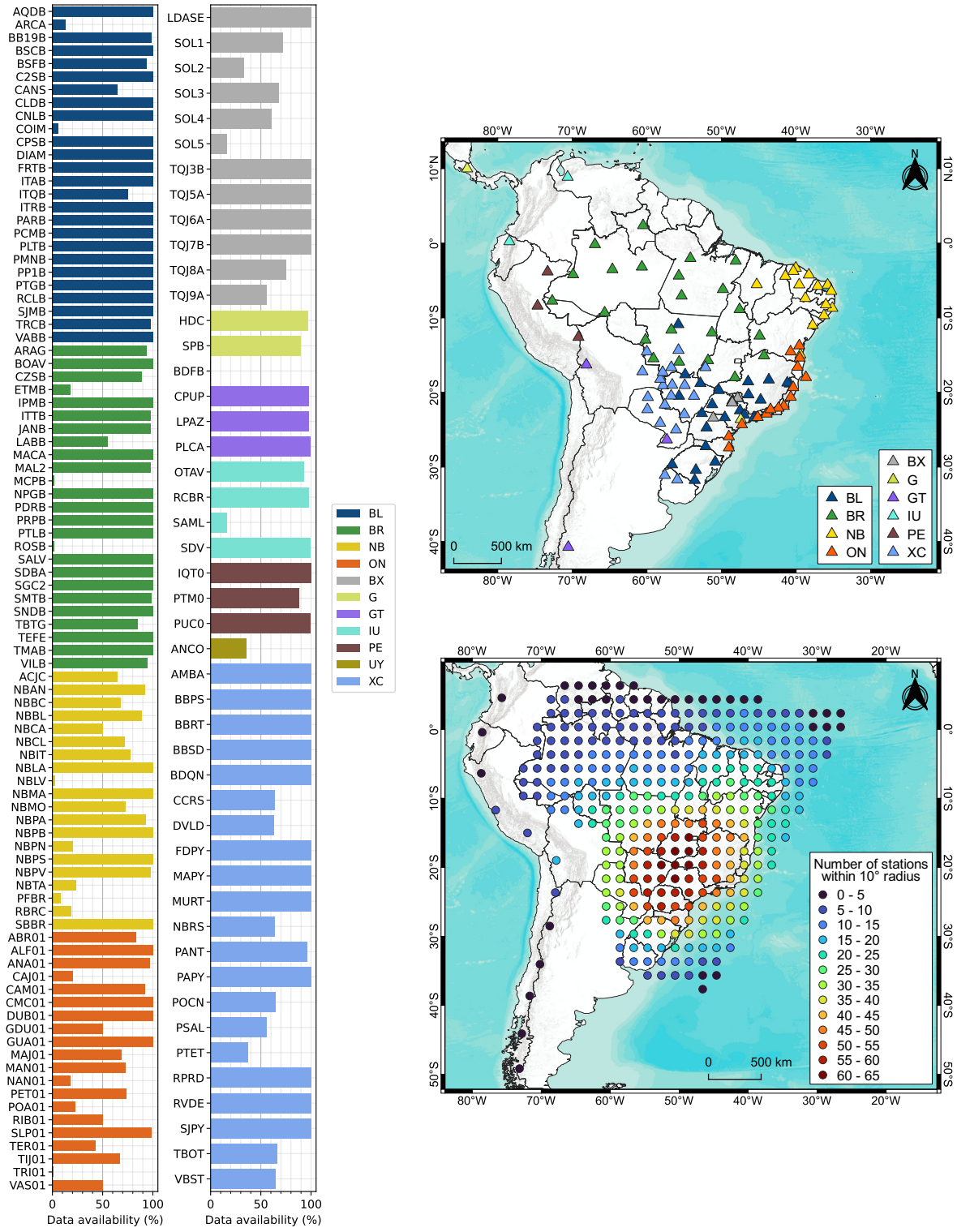
Left: Availability of data in the form of waves, recorded in 2016, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



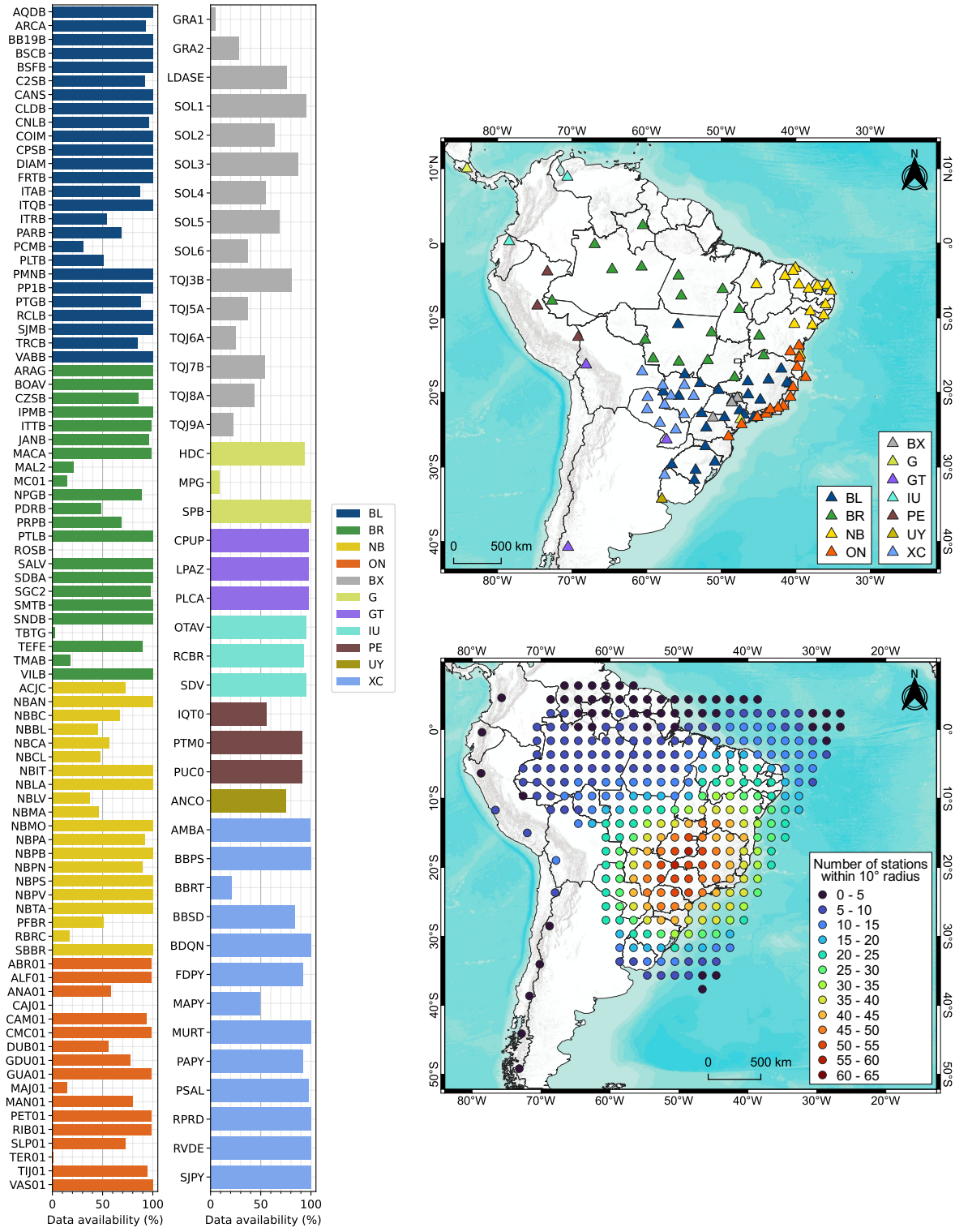
Left: Availability of data in the form of waves, recorded in 2017, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



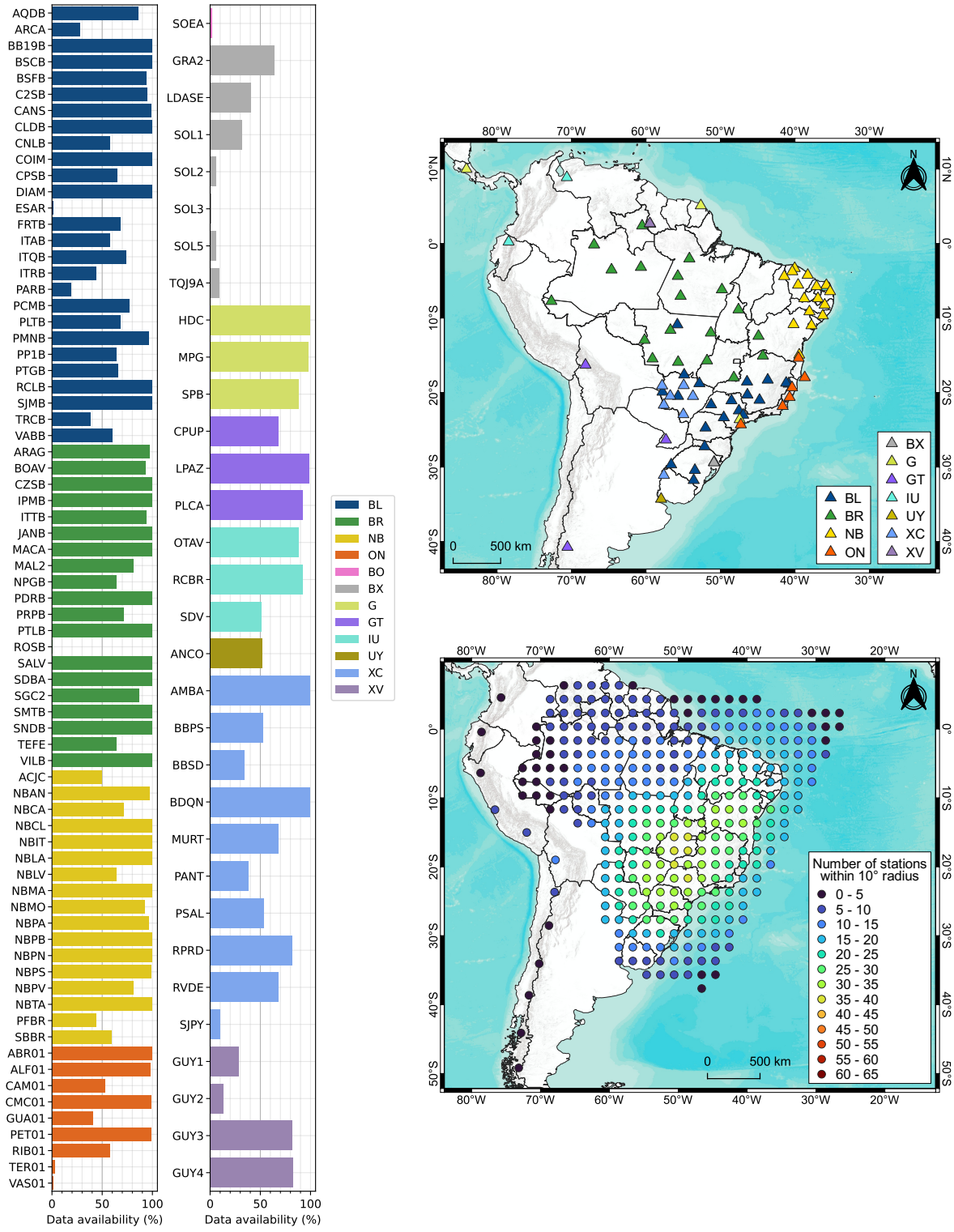
Left: Availability of data in the form of waves, recorded in 2018, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



Left: Availability of data in the form of waves, recorded in 2019, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



Left: Availability of data in the form of waves, recorded in 2020, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.



Left: Availability of data in the form of waves, recorded in 2021, from each station contained in the IAG-USP SeisComP inventory in operation that year. Right: Locations of stations with at least 50% of data available in the period and the grid obtained from the spatial density of these stations.

APPENDIX 2

1D Velocity model

BRA23

Depth (km)	**V_P (km/s)**	**V_S (km/s)**	**Density (g/cm^3)**
0.0000	5.8000	3.4600	2.7200
14.3000	5.8000	3.4600	2.7200
14.3000	6.6556	3.8500	2.9200
40.0000	6.6556	3.8500	2.9200
40.0000	8.2369	4.4800	3.3198
198.0000	8.2369	4.5000	3.3713
660.0000	10.3194	5.6100	4.0646
660.0000	10.7900	5.9600	4.3714
710.0000	10.9229	6.0897	4.4010
760.0000	11.0558	6.2095	4.4305
809.5000	11.1353	6.2426	4.4596
859.0000	11.2221	6.2798	4.4885
908.5000	11.3068	6.3160	4.5173
958.0000	11.3896	6.3512	4.5459
1007.5000	11.4705	6.3854	4.5744
1057.0000	11.5495	6.4187	4.6028
1106.5000	11.6269	6.4510	4.6310
1156.0000	11.7026	6.4828	4.6591
1205.5000	11.7766	6.5138	4.6870
1255.0000	11.8491	6.5439	4.7148
1304.5000	11.9200	6.5727	4.7424
1354.0000	11.9895	6.6008	4.7699
1403.5000	12.0577	6.6285	4.7973
1453.0000	12.1245	6.6555	4.8245
1502.5000	12.1912	6.6815	4.8515
1552.0000	12.2550	6.7073	4.8785
1601.5000	12.3185	6.7326	4.9052
1651.0000	12.3819	6.7573	4.9319
1700.5000	12.4426	6.7815	4.9584
1750.0000	12.5031	6.8052	4.9847
1799.5000	12.5631	6.8286	5.0109
1849.0000	12.6221	6.8515	5.0370
1898.5000	12.6804	6.8742	5.0629
1948.0000	12.7382	6.8972	5.0887
1997.5000	12.7956	6.9194	5.1143
2047.0000	12.8526	6.9418	5.1398
2096.5000	12.9096	6.9627	5.1652

Depth (km)	**V_P (km/s)**	**V_S (km/s)**	**Density (g/cm^3)**
2146.0000	12.9668	6.9855	5.1904
2195.5000	13.0222	7.0063	5.2154
2245.0000	13.0783	7.0281	5.2403
2294.5000	13.1336	7.0500	5.2651
2344.0000	13.1894	7.0720	5.2898
2393.5000	13.2465	7.0931	5.3142
2443.0000	13.3018	7.1144	5.3386
2492.5000	13.3585	7.1369	5.3628
2542.0000	13.4156	7.1586	5.3869
2591.5000	13.4741	7.1807	5.4108
2640.0000	13.5312	7.2031	5.4345
2690.0000	13.5900	7.2258	5.4582
2740.0000	13.6494	7.2490	5.4817
2740.0000	13.6494	7.2490	5.4817
2789.6700	13.6530	7.2597	5.5051
2839.3300	13.6566	7.2704	5.5284
2891.5000	13.6602	7.2811	5.5515
2891.5000	8.0000	0.0000	9.9145
2939.3300	8.0382	0.0000	9.9942
2989.6600	8.1283	0.0000	10.0722
3039.9900	8.2213	0.0000	10.1485
3090.3200	8.3122	0.0000	10.2233
3140.6600	8.4001	0.0000	10.2964
3190.9900	8.4861	0.0000	10.3679
3241.3200	8.5692	0.0000	10.4378
3291.6500	8.6496	0.0000	10.5062
3341.9800	8.7283	0.0000	10.5731
3392.3100	8.8036	0.0000	10.6385
3442.6400	8.8761	0.0000	10.7023
3492.9700	8.9461	0.0000	10.7647
3543.3000	9.0138	0.0000	10.8257
3593.6400	9.0792	0.0000	10.8852
3643.9700	9.1426	0.0000	10.9434
3694.3000	9.2042	0.0000	11.0001
3744.6300	9.2634	0.0000	11.0555
3794.9600	9.3205	0.0000	11.1095
3845.2900	9.3760	0.0000	11.1623

Depth (km)	**V_P (km/s)**	**V_S (km/s)**	**Density (g/cm^3)**
3895.6200	9.4297	0.0000	11.2137
3945.9500	9.4814	0.0000	11.2639
3996.2800	9.5306	0.0000	11.3127
4046.6200	9.5777	0.0000	11.3604
4096.9500	9.6232	0.0000	11.4069
4147.2800	9.6673	0.0000	11.4521
4197.6100	9.7100	0.0000	11.4962
4247.9400	9.7513	0.0000	11.5391
4298.2700	9.7914	0.0000	11.5809
4348.6000	9.8304	0.0000	11.6216
4398.9300	9.8682	0.0000	11.6612
4449.2600	9.9051	0.0000	11.6998
4499.6000	9.9410	0.0000	11.7373
4549.9300	9.9761	0.0000	11.7737
4600.2600	10.0103	0.0000	11.8092
4650.5900	10.0439	0.0000	11.8437
4700.9200	10.0768	0.0000	11.8772
4801.5800	10.1415	0.0000	11.9414
4851.9100	10.1739	0.0000	11.9722
4902.2400	10.2049	0.0000	12.0001
4952.5800	10.2329	0.0000	12.0311
5002.9100	10.2565	0.0000	12.0593
5053.2400	10.2745	0.0000	12.0867
5103.5700	10.2854	0.0000	12.1133
5153.5000	10.2890	0.0000	12.1391
5153.5000	11.0427	3.5043	12.7037
5204.6100	11.0585	3.5187	12.7289
5255.3200	11.0718	3.5314	12.7530
5306.0400	11.0850	3.5435	12.7760
5356.7500	11.0983	3.5551	12.7980
5407.4600	11.1166	3.5661	12.8188
5458.1700	11.1316	3.5765	12.8387
5508.8900	11.1457	3.5864	12.8574
5559.6000	11.1590	3.5957	12.8751
5610.3100	11.1715	3.6044	12.8917
5661.0200	11.1832	3.6126	12.9072
5711.7400	11.1941	3.6202	12.9217

Depth (km)	**V$_P$ (km/s)**	**V$_S$ (km/s)**	**Density (g/cm^3)**
5813.1600	11.2134	3.6337	12.9474
5863.8700	11.2219	3.6396	12.9586
5914.5900	11.2295	3.6450	12.9688
5965.3000	11.2364	3.6498	12.9779
6016.0100	11.2424	3.6540	12.9859
6066.7200	11.2477	3.6577	12.9929
6117.4400	11.2521	3.6608	12.9988
6168.1500	11.2557	3.6633	13.0036
6218.8600	11.2586	3.6653	13.0074
6269.5700	11.2606	3.6667	13.0100
6320.2900	11.2618	3.6675	13.0117
6371.0000	11.2622	3.6678	13.0122