PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

USE OF MACHINE LEARNING TECHNIQUES FOR THE ESTIMATION OF THE SEISMIC SUSCEPTIBILITY AND THE QUANTITATIVE ASSESSMENT OF THE SEISMIC HAZARD. APPLICATION TO THE MAIPO RIVER BASIN, CHILE

JUAN PABLO DÍAZ FERNÁNDEZ

Thesis submitted to fulfill the requirement for the Degree of
Master of Science in Engineering

Advisor:
ESTEBAN PATRICIO SÁEZ ROBERT

Santiago de Chile, (December, 2021)

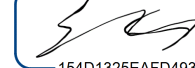PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

USE OF MACHINE LEARNING TECHNIQUES FOR THE ESTIMATION OF THE SEISMIC SUSCEPTIBILITY AND THE QUANTITATIVE ASSESSMENT OF THE SEISMIC HAZARD. APPLICATION TO THE MAIPO RIVER BASIN, CHILE

JUAN PABLO DÍAZ FERNÁNDEZ

Members of the Committee

ESTEBAN PATRICIO SÁEZ ROBERT

GONZALO ALEJANDRO YÁÑEZ CARRIZO

CÉSAR RODRIGO PASTÉN PUCHI

ENZO ENRIQUE SAUMA SANTIS

To fulfill the requirement for the Degree of

Master of Science in Engineering

Santiago de Chile, (December 2021)

To María Olivia and Hernán,
my beloved parents

ACKNOWLEDGMENTS

First, I would like to give a special thanks to my advisor, Professor Esteban Sáez, for have trusted me from the beginning, always supported and guided me with great disposition and for showing me the beauty behind the science of geotechnics. I will always thank you for all the opportunities and learnings that you gave me during these last two years.

Many thanks to Professor Mauricio Monsalve, who provided fundamental contributions for the development and achievements of this research. I will always appreciate his constant support, guidance and great disposition.

I am very grateful to my parents, who with their infinite love and unconditional support have always inspired me to be better. My achievements would not have been possible without your company, protection and example. You are part of this too. To my entire family, who in their own way always supported me unconditionally. To my dear Cata, who lived this beautiful stage with me, she understood me and accompanied me in the most complex moments.

To my great friends who were there from the beginning: Luco, Matías, Harold, Jose, Heber and Diego. You were the best of this stage and I hope to have you close forever. To my friends and colleagues that I met during the investigation with whom we shared pleasant and not so pleasant moments: Felipe, Javier, Valeria, Ricardo and Laura, I will thank you forever for your valuable advice and company during this challenging process.

GENERAL INDEX

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Seismic site amplification and seismic hazard maps are crucial inputs for decision making and risk evaluation in places where seismicity imposes a significant risk to human life and infrastructure. In Chile, one of the most seismically active countries on Earth, the state-of-the-art techniques used for constructing seismic hazard maps rely primarily on qualitative or indirect, geologic data, with a limited consideration of the dynamic parameters dictating the site's seismic response. In this work, we propose a novel methodology for integrating qualitative and quantitative data to map the seismic susceptibility and dynamic amplification, using machine learning (ML). Our method uses measurements of surface shear wave velocities (V_{s30}) and predominant frequencies (f_0) on randomly distributed sites, combined with gravity anomaly maps to update the geographic extension of seismic amplification units. Additionally, we trained the predictive models to interpolate and extrapolate V_{s30} and f_0 in the unsampled sites. We selected the Maipo River Basin as our testing region for this model because of its strategic importance and the large population and critical infrastructure exposed to seismic hazards. Applying this method to the case study resulted in (i) a refined seismic susceptibility map, and (ii) maps of V_{s30} and f_0 estimated with great precision in the study area. The best predictions, obtained by ML techniques and validated through cross-validation, are possibly due to the inclusion of spatial covariates for algorithm training, enhancing the ability of the model to capture the spatial correlations of geological, geophysical and geotechnical data. The estimation of predominant frequencies is considerably improved by including gravity as a covariant. The accuracy of the f_0 predictions apparently depends more on the choice of covariates than on the algorithm used, while the V_{s30} predictions are more sensitive to the chosen algorithm. These results illustrate the great potential of machine learning predictive algorithms in digital soil mapping, which surpass traditional geostatistical techniques. Additionally, and using the best predictions, seismic hazard maps were generated for the case study through an open-source software, designed to generate strong motion indicators for different seismicity models and ground motion prediction equations under different scenarios. The major contribution of this work is to introduce a novel

estimation methodology based on artificial intelligence models to extend local measurements of site's dynamic properties in an area of interest. This information can be used to quantitatively estimate the seismic hazard at regional scale.

Keywords: seismic site amplification, seismic susceptibility, machine learning, seismic hazard.

RESUMEN

La amplificación sísmica de sitios y los mapas de amenaza sísmica son insumos cruciales para la toma de decisiones y evaluación del riesgo en lugares donde la sismicidad impone un riesgo significativo para la vida humana y la infraestructura. En Chile, uno de los países con mayor actividad sísmica de la Tierra, las técnicas de vanguardia utilizadas para construir mapas de amenaza sísmica se basan principalmente en datos geológicos cualitativos o indirectos, con una consideración limitada de los parámetros dinámicos que dictan la respuesta sísmica del sitio. En este trabajo, proponemos una metodología novedosa para la integración de datos de naturaleza cualitativa y cuantitativa para mapear la susceptibilidad sísmica y la amplificación dinámica mediante aprendizaje automático (AA). Nuestro método utiliza mediciones de velocidades de onda de corte superficial (V_{s30}) y frecuencias predominantes (f_0) en sitios aleatoriamente distribuidos, combinados con mapas de anomalía gravimétrica para para actualizar la extensión geográfica de las unidades de amplificación sísmica. Seleccionamos la Cuenca del Río Maipo como nuestra región de prueba para este modelo debido a su importancia estratégica y a la numerosa población e infraestructura critica expuestas a amenazas sísmicas. La aplicación de este método al caso de estudio resultó en (i) un mapa de susceptibilidad sísmica refinado, y (ii) mapas de V_{s30} y f_0 estimados con gran precisión en el área de estudio. Las mejores predicciones, obtenidas mediante técnicas de AA y validadas mediante validación cruzada, se deben posiblemente a la inclusión de covariantes espaciales en el entrenamiento de los algoritmos, mejorando la capacidad del modelo para capturar las correlaciones espaciales entre datos geológicos, geofísicos y geotécnicos. La estimación de las frecuencias predominantes es considerablemente mejorada al incluir la gravedad como covariante. La precisión de las predicciones de f_0 aparentemente depende más de la elección de las covariantes que del algoritmo utilizado, mientras que las predicciones de V_{s30} son más sensibles al algoritmo escogido. Estos resultados ilustran el gran potencial de los algoritmos predictivos de aprendizaje automático en el mapeo digital de suelos, que superan a las técnicas geoestadísticas tradicionales. Adicionalmente, y usando las mejores predicciones, se generaron mapas de amenaza sísmica para el caso de estudio, a través de

un *software* de acceso libre diseñado para generar indicadores de movimiento fuerte para diferentes modelos de sismicidad y ecuaciones de predicción de movimiento de suelo bajo distintos escenarios. La principal contribución de este trabajo es introducir una metodología novedosa de estimación basada en modelos de inteligencia artificial para extender mediciones locales de las propiedades dinámicas de los sitios en un área de interés. Esta información se puede utilizar para estimar cuantitativamente la amenaza sísmica a una escala regional.

Palabras Clave: amplificación sísmica de sitios, susceptibilidad sísmica, aprendizaje automático, amenaza sísmica.

1 INTRODUCTION

Chile is a country highly exposed to natural disasters. Earthquakes, tsunamis, volcanic eruptions, forest fires, landslides and many other natural hazards have impacted the territory since ancient times and remains as a permanent threat. The 2017 Organization for Economic Co-operation and Development report shows that Chile has an average of 4 natural disasters per year, ranking it as the fourteenth country in this group with the most natural disasters per year, and as the second country with the most economic losses due to these disasters. Between 1980 and 2011, in Chile there have been annual losses close to 1.2% of gross domestic product due to natural disasters (United Nations, 2015). This is equivalent to 3 times the percentage of gross domestic product that the country spends on investment in science. These large costs considerably limit the development of the country, i.e. strategies are required to mitigate the economic and social impacts resulting from natural disasters.

On the other hand, even though Chile has made progress in the study and individual characterization of the physical processes that constitute natural hazards, there is not an integrated vision of these processes, which is technically known as multi-hazard. Numerous recent events have highlighted the interrelation of the different hazards, showing how its chain effect can increase their impact. For example, in September 2018, a Mw 6.6 earthquake occurred under the Hidaka mountain area on Hokkaido Island, Japan. The earthquake generated a massive landslide, killing 41 people, hundreds injured, and numerous houses destroyed. The earthquake occurred just days after Jebi Typhoon, so there are well-founded suspicions that the devastating result of the earthquake was amplified by the softening of the soils because of the rains associated with the typhoon. Also in September 2018, a Mw 7.7 earthquake and tsunami occurred in Palu, Indonesia. The events left as a result, only in the Palu area, 1,750 deaths, 800 missing and more than 70,000 residential buildings destroyed. The great level of devastation was enhanced by the cascading effect of the earthquake, tsunami and liquefaction that affected the subsoil

of the city of Palu. Another example corresponds to the model assess the impact of different volcanic hazards on the building structures, developed by Zuccaro et al. (2008). They showed that the sequence of events derived from a volcanic eruption (earthquakes, ash flows and pyroclastic flows) can cause the progressive deterioration of the building's resistance characteristics, even causing their partial failure. Similarly, research in other countries have studied the relationships and mutual influences between different types of natural hazards (e.g. Tarvainen et al., 2006; Marzocchi et al., 2009). Then, it becomes clear that the multi-hazard dimension could be a key component in efforts to mitigate the impact of natural disasters in Chile.

Chile's high exposure to natural hazards together with the inadequate perception of multi-hazard risk constitute a great opportunity for applied scientific research to foster transfer to the public system and to develop technological improvements. The Fondef project ID19|10021 "Proposal for a guide, mapping and multi-hazard platform for critical decision-making and adaptation to climate change in metropolitan regions and large conurbations of Chile" aims to take this opportunity and solve the needs mentioned above. The project's proof of concept is developed in the pilot case of the Maipo River Basin (MRB), which includes the Santiago Metropolitan Region (SMR) and part of the Valparaíso Region. Specifically, the project seeks to (i) identify the individual and concatenated hazards present in the MRB, (ii) generate a multi-hazard map, (iii) generate a platform and guidelines for the development and use of the multi-hazard map, and (iv) disseminate the map and the platform.

The natural hazards that will be addressed by the project are seismic, volcanic, tsunamis, landslides, rockslides, forest fires and floods. As part of these threats, the seismic hazard is a particularly relevant component. Indeed, the tectonic context of the Chilean territory places it as one of the most seismic country in the world (Scholz, 2002), where an oceanic plate (Nazca plate) subducts to a continental plate (South American plate). Under this tectonic environment, Chile is affected mainly by 3 types of earthquakes: contact between

plate or interplate, intermediate depth intraplate and shallow crustal. Its impacts can generate great economic and human losses, and large earthquakes are usually triggers of other potentially devastating natural events such as tsunamis, rockfall, liquefaction, etc. This thesis addresses the seismic hazard in the MRB, as a component of the Fondef project ID19|10021.

The ways in which the seismic hazard is assessed in Chile are diverse and operate at different scales. At the regional scale, the seismic hazard is usually assessed through seismic response maps, generated mainly with qualitative information, such as available geological information and perceptions of previously recorded earthquakes (e.g. Fernández, 2003; Von Igel et al., 2004). At a more local level, the seismic hazard is assessed through the seismic classification of sites, according to the Chilean seismic design code for residential buildings, NCh433. This classification is generated using mainly quantitative information, such as surface shear wave velocities in the first 30 m (V_{s30}), rock quality designation (RQD), undrained shear strength (S_u), etc. In this study, the main idea will be to integrate information of different nature and scale, to achieve a more complete, detailed and accurate seismic hazard assessment than those achieved previously. For this, we will give special prominence to computational artificial intelligence techniques applied to digital soil mapping (DSM). Information from the areas of geology (seismic response maps), geophysics (gravimetry and results from surface wave geophysical methods), seismic geotechnical engineering (seismic classification), computation (artificial intelligence models) and topography (digital elevation models) will be combined to achieve an innovative and original result.

This study provides one qualitative and many quantitative evaluations of the seismic hazard in the MRB, consisting in the following results. First, a refined and updated seismic response map of the MRB was generated for the qualitative assessment of the seismic hazard. This map was generated from available geological information, previous seismic response maps, an extensive compilation and execution of dynamic soil characterizations,

gravimetric studies, and topographic information of the area. Second, seismic hazard maps were generated for different scenarios, in terms of the peak ground acceleration (PGA). These maps are the result of a methodology that combines computational artificial intelligence tools with the refined seismic response map mentioned above, to predict dynamic properties of sites with great accuracy. These predictions subsequently allow evaluating the expected seismic intensity in a future scenario through a state-of-the-art seismic hazard analysis platform. Finally, and to allow the combination of the seismic hazard with other natural hazards considered in the Fondef project ID19|10021, a map of probabilities of exceedance of a given intensity is provided. This exceedance probability map is obtained from the same inputs as the seismic hazard map using the same platform but applying a different criterion.

Chapter 2 shows the background compilation and details of the field work. Chapter 3 shows the results of this study at the MRB scale. Chapter 4 corresponds to a copy of the article "On the use of machine learning techniques for the estimation of seismic susceptibility and the quantitative assessment of the seismic hazard. Application to the Santiago Basin, Chile", submitted to Engineering Geology journal. That chapter details the methodologies and main results of this study but applied only to the Santiago basin to simplify the presentation and discussion. Finally, Chapter 5 provides the main conclusions of this study and recommendations for future work. Annexes with detailed research results are also included.

1.1 Hypothesis

The use of machine learning predictive models in combination with geological, geophysical, geotechnical and topographic information can significantly improve the accuracy of quantitative assessments of the seismic hazard in the Maipo River Basin.

1.2 Objectives

This study has two main objectives: i) Estimate the seismic response of the sites of the Maipo River Basin using qualitative and quantitative information available in different areas and; ii) To assess the seismic threat in the Maipo River Basin for a scenario compatible with Chilean regulations using predictive machine learning models. Specific objectives are the following:

- Compile seismic response maps in the study area, collect available information of dynamic characterization of sites and carry out complementary geophysical surveys.

- Seismically classify the sites where dynamic characterizations is available.

- Redefine the seismic response units in the study area.

- Predict dynamic properties of sites in the study area using machine learning techniques based on digital soil mapping tools.

- Calculate and generate maps of the seismic hazard in the study area from the best predictions of the dynamic properties of the sites.

2   BACKGROUND AND FIELD WORK

2.1   Study Area

The study area of this work corresponds to the Maipo River Basin. It is located in central Chile, between the 256,000 m to 426,600 m E and the 6,205,500 m to 6,355,900 m N (UTM coordinates) and involves a large part of the SMR and part of the Valparaíso Region, specifically the district of San Antonio (see Figure 2-1). The selection of this study area is due firstly to the fact that the MRB hydrodynamically integrates the Main Cordillera with the coastline, secondly, the SMR concentrates 50% of the Chilean population and 42% of Chile's gross domestic product, and third, the port of San Antonio is a key sector for international trade.

In general terms, the study area can be described around three well differentiated morphostructural units, which are from west to east: Cordillera de la Costa, Depresión Central and Cordillera Principal. The Cordillera de la Costa consists mainly of the Paleozoic coastal batholith and the central Mesozoic batholith, in addition to Paleozoic metamorphic basement outcrops (Wall et al., 1996). The Depresión Central corresponds to a sedimentary valley in which the city of Santiago is located, on the basin of the same name. The Santiago basin is mainly composed of Pleistocene to Holocene alluvial and fluvial sediments (Sellés & Gana, 2001). It is an irregular depression with variable depths ranging from a few meters to over 500 m (Yáñez et al., 2015) and some isolated hills. The surface geology of the Santiago basin's soils can be divided into 4 main units (e.g. Fernández, 2003): (i) deposits of alluvial fans composed by gravels in a sandy matrix located in the central and southern zone of the Santiago basin; (ii) deposits of alluvial fans composed by gravels with sandy-clay matrices located mainly in the eastern part of the basin; (iii) deposits of volcanic ash with lithics and pumice, known as Pudahuel Ignimbrite, located in the west and northwest of the Santiago basin, and (iv) deposits of

alluvial fans consisting mainly of silts and clays, located in the northwest area of the basin. The Cordillera Principal is composed of Cenozoic volcanic and volcanic-sedimentary rocks and Neogenic intrusives in its western side, while on its eastern side it is composed of sedimentary and volcanic Mesozoic continental and marine rocks (e.g. Fock, 2005).

On the other hand, the district of San Antonio is mainly composed by marine terraces that give rise to the Formación Navidad to the northeast of the district, by sandy wind deposits (dunes) in its central area, and by rocky massifs to the north and south (Wall et al., 1996).
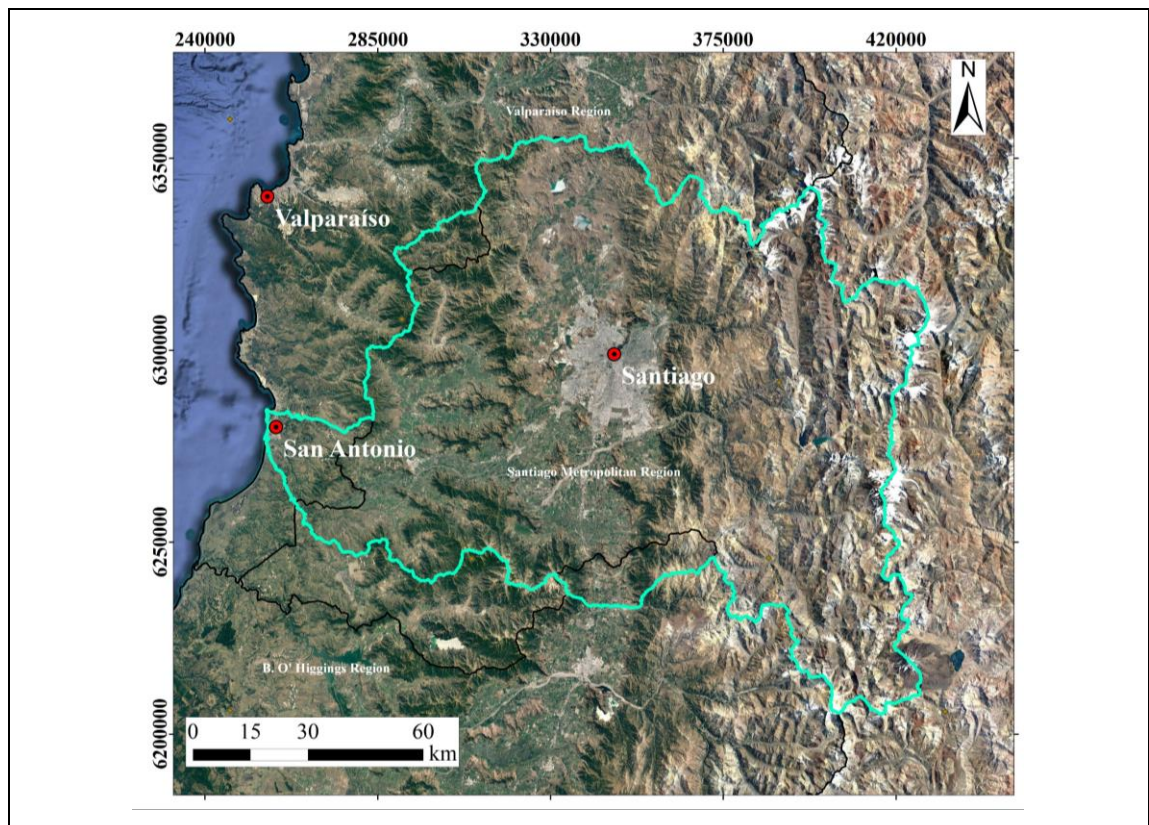
Figure 2-1: Study area of this work (cyan perimeter), corresponding to the Maipo River Basin. The main cities of area are also shown.

2.2   Data Sources

This chapter describes the sources of the data and information used as input to generate the seismic response map for this study. The digital elevation model used to predict properties of the sites of the study area is presented. The geological and seismic zoning maps in the study area prior to this work are shown. Additionally, the techniques used to obtain different parameters through geophysical methods are briefly described.

2.2.1   Digital elevation model

The first data used for this study corresponds to a digital elevation model (DEM). Digital elevation models are visual and mathematical representations of height values relative to mean sea level. These models allow to characterize the relief and the elements that are on the surface. Multiple DEMs of 12.5 m resolution available in public satellite data (ALOS palsar, for more information visit https://asf.alaska.edu) were used and were merged using ArcGIS Desktop software to obtain a DEM over the entire study area, as shown in Figure 2-2. From this DEM, and using ArcGIS tools, it is possible to obtain the elevation at any point in the study area, as well as a slope map, as shown in Figure 2-3. To obtain the slope map, the ArcGIS "Slope" tool was used, this tool adjusts a representative plane of the surrounding topography to a point of interest and calculates its slope using the maximum average technique (Burrough & McDonnel, 1998). The topographic elevation and slope of the MRB sites are key parameters that will be used in Chapter 4 of this study.
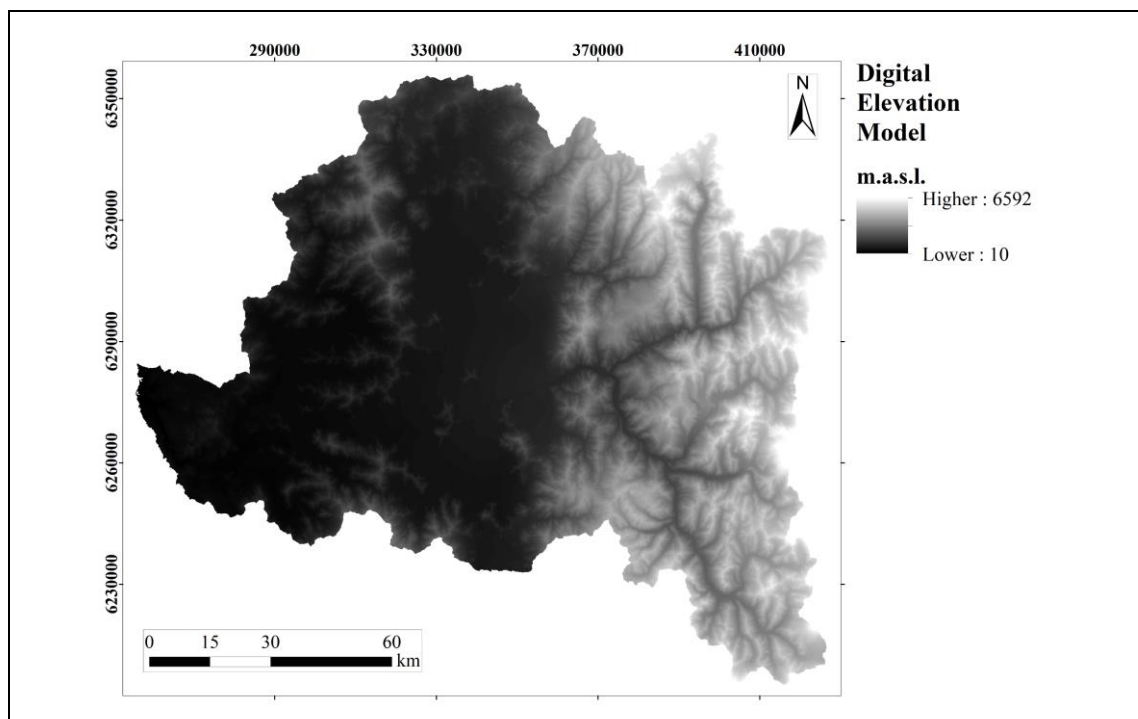
Figure 2-2: Digital elevation model of the study area.
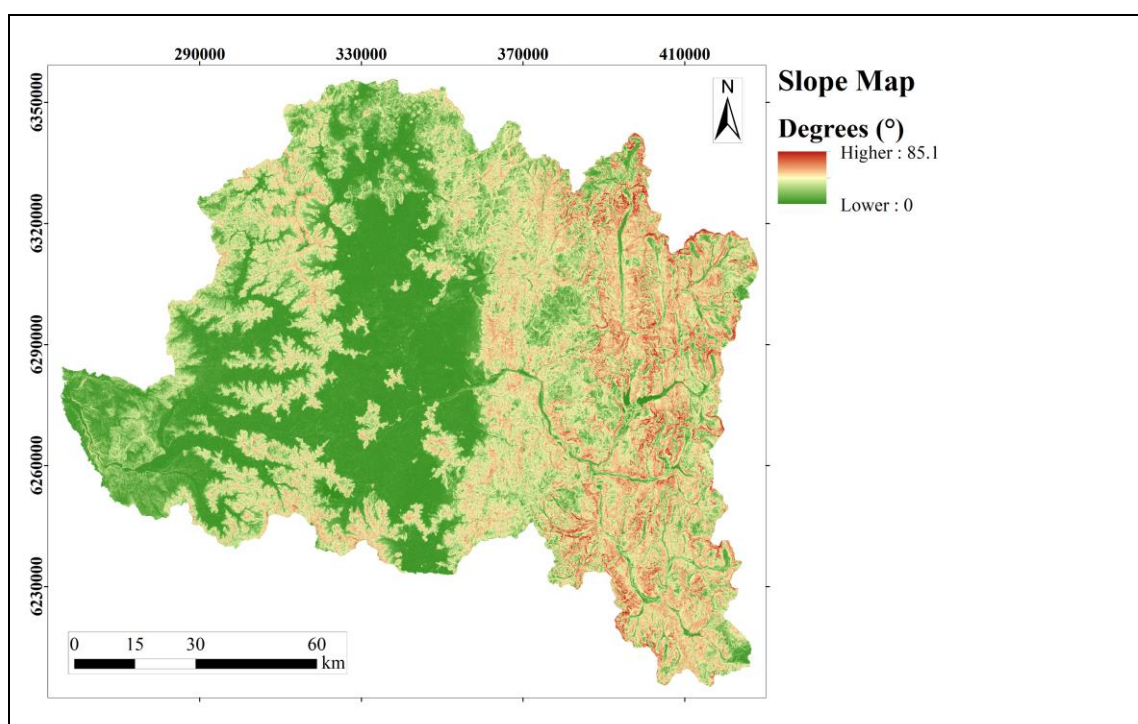
Figure 2-3: Slope map of the study area, obtained from the DEM of Figure 2-2.

2.2.2   Seismic response of the Santiago Metropolitan Region

Von Igel et al. (2004), through the Servicio Nacional de Geología y Minería (Sernageomin), generated the seismic response map of the SMR shown in Figure 2-4. This map corresponds to an extrapolation of the seismic response map of the Santiago basin, generated by Fernández (2003), using the same methodology. This map provides a qualitative estimate of the seismic response of the different geological units of the SMR, for a similar scenario to that of the Algarrobo Ms 7.8 earthquake of March 3, 1985. For each seismic unit, an intensity interval was estimated, associated with the damage observed for this earthquake.
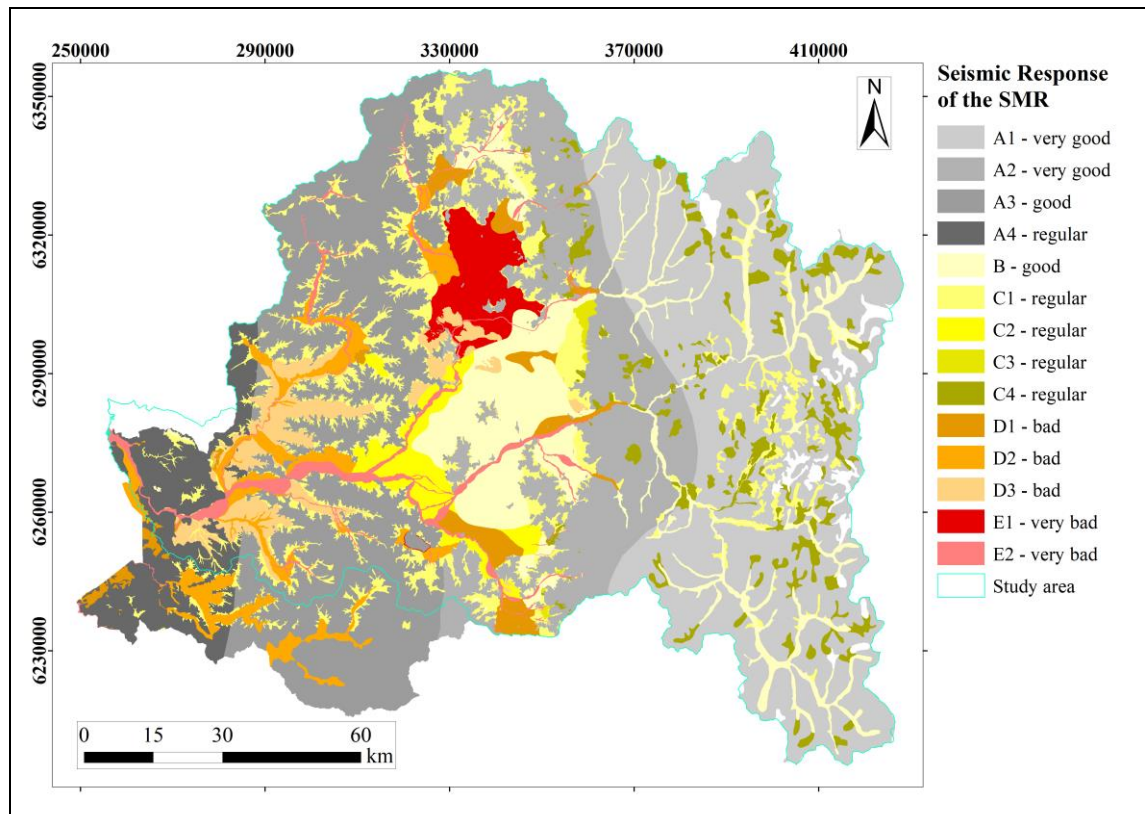
Figure 2-4: Seismic response map of the SMR. Modified from von Igel et al. (2004).

Due to the nature of the information used to generate this map, it has several limitations:

- There are very few instrumental quantitative data (records of accelerograms) from the 1985 Algarrobo earthquake available to develop the seismic zoning of the RMS.

- The intensities used are subjective and depend on who made the observations. In addition, they depend on the characteristics of the buildings and not only on the properties of the sites.

- As the thickness of the sedimentary fill in E units is unknown, there is much uncertainty about the seismic response in these units.

- Topographic effects that may affect the seismic response of the sites are not considered. Neither seismic amplification effects nor associated secondary effects (mass removals) were considered.

- The boundaries between seismic units are mostly inferred.

- In areas with limited observations, the seismic response was extrapolated mainly based on geological information.

Despite these limitations, this map was used as a first-order input in this study, because its information is of great value for the combined analysis with quantitative information. In addition, its extension covers almost the entire study area (see Figure 2-4).

2.2.3 Seismic zonation of the Santiago Basin

Leyton et al. (2011) made one of the first efforts to generate an updated seismic zoning that includes geological, geotechnical and geophysical information in the Santiago Basin. The work of Leyton et al. combined the following information: updated surface geology of the Santiago basin (Figure 2-5), intensity records and damage cadasters from the 1985 Algarrobo earthquake and 2010 Maule earthquake, surface shear wave velocities in the first 30 m (V_{s30}) and fundamental period of different sites from the Santiago basin. These last two parameters are indicators of the local effects of the site that can cause dynamic amplifications under the action of seismic waves (Tokimatsu, 1997; Pastén, 2007). As a result, the seismic zoning shown in Figure 2-6 was proposed. This seismic zoning identifies two main areas, each with a homogeneous seismic demand.

The main limitations of this work are:

- The data for V_{s30} and fundamental periods are not all in the same places, which restricts an overall characterization of the sites.

- There is not enough information to improve the seismic classification in places where gravel soils (rigid soils) presented high seismic intensities in prior earthquakes.
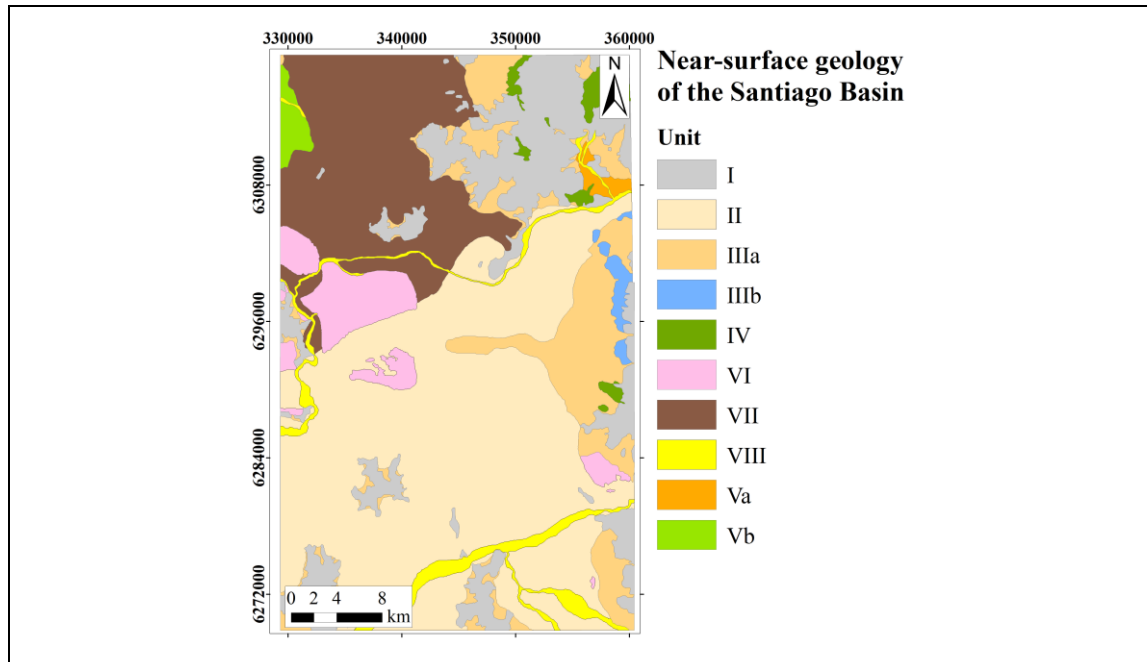
Figure 2-5: Near-surface geology of the Santiago basin. Modified from Leyton et al. (2011).
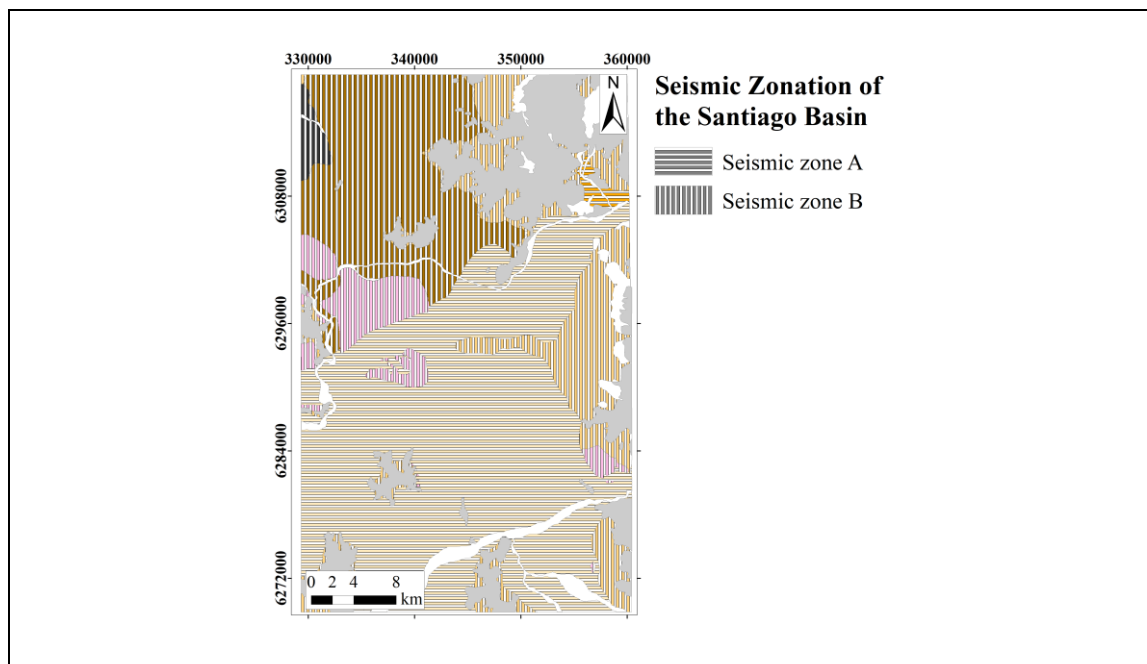
Figure 2-6: Seismic zonation of the Santiago basin. Modified from Leyton et al. (2011).

2.2.4 Seismic response of San Antonio, Valparaíso Region

The Sernageomin in collaboration with the Fondef D10|1027 project published in 2019 a seismic response map of the urban area of San Antonio - Lololleo, Valparaíso Region, presented in Figure 2-7. This map was generated including information and variables from different sources: geology, geomorphology, slopes, V_{s30} and fundamental period measurements. For the elaboration of this map, a subduction seismogenic source was considered, taking into consideration the Algarrobo earthquake of March 3, 1985 (Mw 8) and the Maule earthquake of February 27, 2010 (Mw 8.8). Like other maps, this map classifies the seismic response of sites into 5 categories, where a site A has the best seismic response and a site E the worst.

This map contains unpublished and recent information on the zone of the study area and much of this data was generated in a previous collaboration between Sernageomin and the research group in which this research belongs.
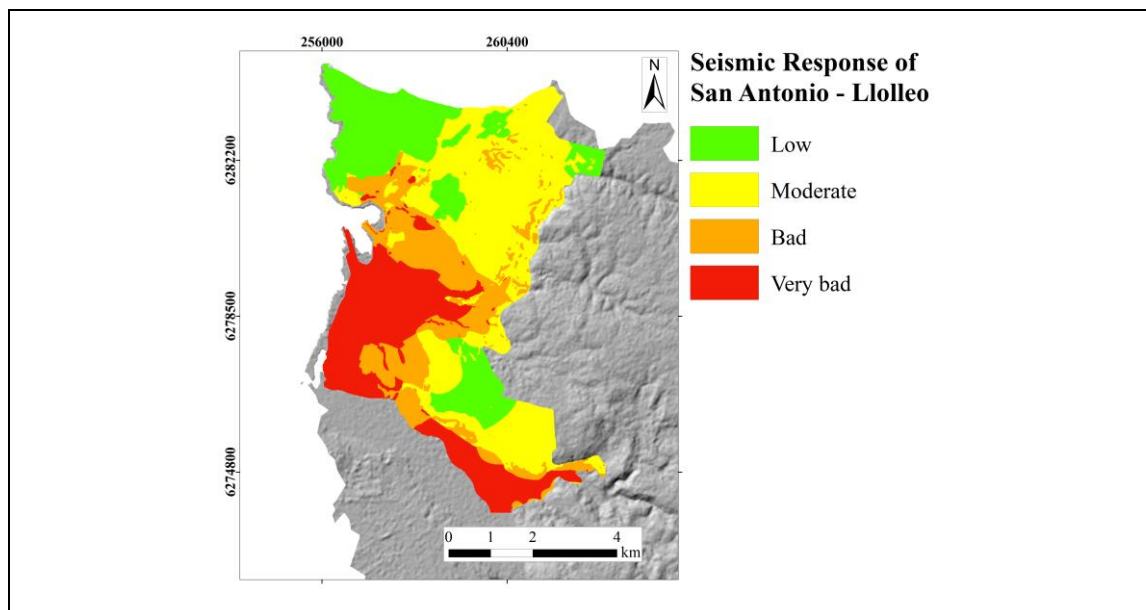
Figure 2-7: Seismic response of the urban area of San Antonio and Llolleo.
Modified from Sernageomin (2019).

2.2.5   Dynamic characterization of sites

The seismic motion on the surface of a soil deposit under the action of an earthquake does not depends only on the magnitude of the event and the proximity to the source. Numerous studies have shown that the local geotechnical conditions of a site can produce dynamic amplifications of seismic waves, which is known as site effects. Different technical standards such as ASCE7-10 in the United States, Eurocode 8 for the countries of the European Union, and Decreto Supremo No. 61 in Chile establish a seismic classification of sites to anticipate their seismic response under the action of an earthquake. All these standards use V_{s30} as the main seismic classification parameter. This parameter was included in the Chilean regulations in 2011, after the devastating consequences of the Mw8.8 Maule earthquake of February 27, 2010. This parameter is obtained by

$$V_{s30} = \frac{\sum h_i}{\sum (\frac{h_i}{V_{s-i}})} \tag{1}$$

where V_{s-i} is the shear wave velocity of stratum i, in m/s; h_i is the thickness of stratum i, in meters and n is the number of strata in the upper 30 meters of the site. However, in this study a modified seismic classification was used, in which V_{s30} is replaced by $V_{s<900}$, defined as the average of shear wave velocities in soils up to the depth where V_s is less than 900 m/s, before reaching a depth 30 m.

Currently in Chile, in addition to V_{s30}, the measurement of other static parameters is required to classify a soil, such as rock quality designation (RQD), undrained shear strength (S_u), unconfined compressive strength (q_u) and the blow counts in the standard penetration test (N_{SPT}). A major disadvantage of the current Chilean seismic classification system is that the static classification parameters require expensive and time-consuming surveys, in addition to not considering other important parameters, such as depth to the

basement or the predominant period of a site ($T_0 = 1/f_0$). This last dynamic parameter is key in the seismic characterization of a site, since is the period associated with the impedance contrast that predominates when the site vibrates (Podestá, 2017). f_0 is commonly measured through the Nakamura method (1989).

Current Chilean regulations for seismic classification are in a period of transition towards a new seismic classification system, based solely on the dynamic parameters V_{s30} and T_0. The great advantage of this new classification system is that the tests required for its measurement can be much cheaper and faster than traditional tests if geophysical methods are used. In addition, dynamic parameters are more directly related to the seismic response of a site than static parameters.

The V_{s30} data and the predominant frequency of the sites (f_0) used in this study come from different sources. It is important to mention that in all the dynamically characterized sites in this study, measurements of both V_{s30} and f_0 were carried out to obtain a more complete seismic characterization of the sites. Data was compiled from measurements carried out in a previous research project (Fondef D10|1027), geophysical explorations carried out by Dictuc S.A. (which is a subsidiary of the Pontificia Universidad Católica de Chile, hereinafter PUC), courtesy of the RyV Ingenieros company, a master's thesis research from the Universidad de Chile (Acevedo, 2021), surveys carried out in the framework of a prior study oriented to updating the Santiago Metropolitan Regulatory Plan, the website of the Centro Sismológico Nacional and surveys carried out specifically for the present Fondef project. Even though these data were well characterized and georeferenced, the first problem with this database presented is that not all surveys had carried out f_0 measurements as they were conducted as independent measurements. Only since 2015 the research group systematically begin to measure V_{s30} and f_0 together. However, all these surveys were intended to measure V_{s30}, this parameter was available in all measurements in the initial database. Figure 2-8 shows the distribution of the database available at the beginning of this study and the location of "incomplete" surveys (those without f_0

measurements). To solve this issue, the first field work was to complete these measurements. Between November 2019 and January 2020, f_0 measurement campaigns were carried out for sites where this parameter was not available. In February 2020 this initial campaign was completed and measurements of V_{s30} and f_0 were already available in all the explored sites. In all the explored sites, at least one V_{s30} measurement and 3 to 5 f_0 measurements were performed to estimate the uncertainty of the measurements and to avoid problems associated with erroneous records.

Details of the methods used to obtain these parameters and an example result are presented in Annex 1.
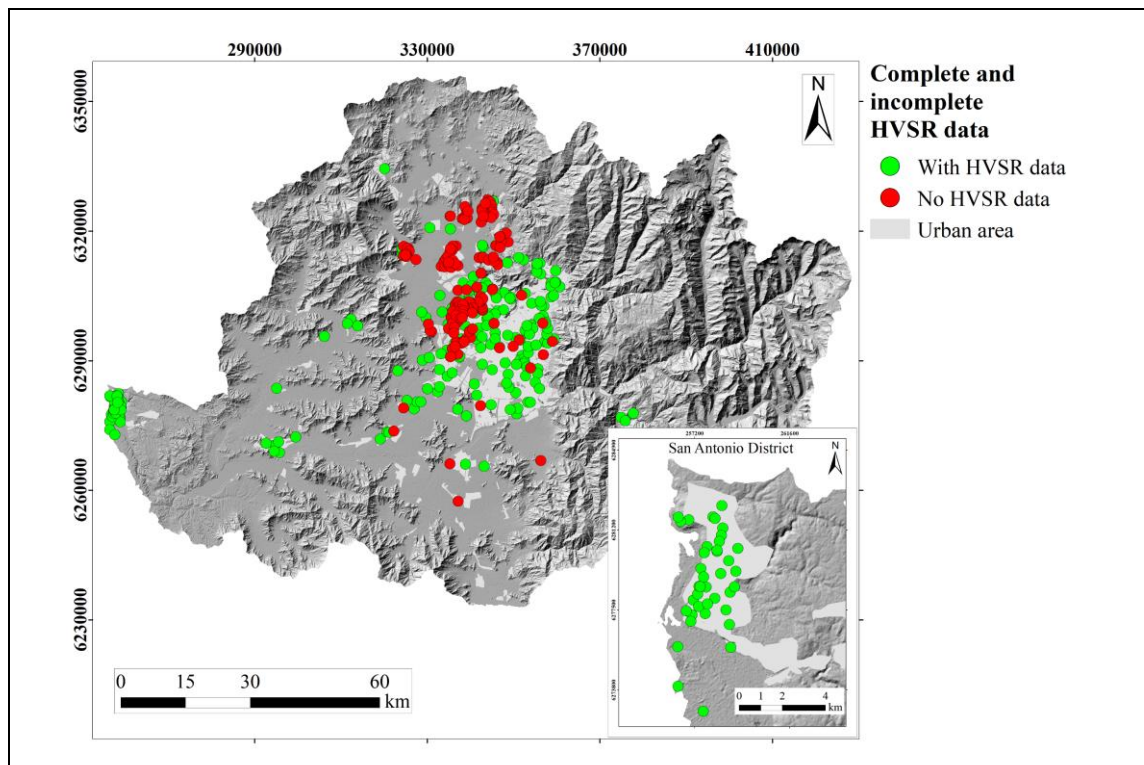
Figure 2-8: Location of sites without initial HVSR measurement.

As can be seen in Figure 2-8, most of the geophysical measurements are in urban areas. Due to the large extension of the study area, there are many sparsely populated areas of geophysical information, hence a new geophysical survey campaign was planned with the aim of covering areas of interest that were not well sampled.

The main objective of this new measurement campaign was to populate the study area with new data to achieve a more complete and detailed seismic zoning. The criteria for selecting sites to perform the geophysical measurements were based mainly on covering urban areas where there was little or no data, such as the districts of San José de Maipo to the east of the study area, Curacaví to the west of the study area and Buin, to the south. Measurements were also made in non-urban sectors far from the areas where the data are concentrated, to obtain measurements in different types of soil (based on surface geology), always trying to balance the amount of data for each type of site. Additionally, measurements in poorly sampled sites and close to the contacts between geological units were prioritized. This criterion aimed to define these limits in more detail from a site effect perspective, since current limits are mostly inferred (von Igel et al., 2004). In the final stage of the study, the measurements were aimed to map mainly rock units since available data in this units are very scarce due to their high slopes and difficulty to access. This last survey was carried out between February 2020 to October 2021.

The geophysical information obtained in the field for V_{s30} and f_0 is summarized in "site cards". These "site cards" summarizes the empirical and adjusted dispersion curves, the profile shear wave velocity from the inversion process, the detailed shear wave profile and the associated V_{s30} values, and the corresponding HVSR curves. Annex 2 shows the "site cards" of the geophysical surveys carried out during this project.

Finally, complete database used in this project consisted of 365 sites with V_{s30} and f_0 data in the study area, as shown in Figure 2-9.
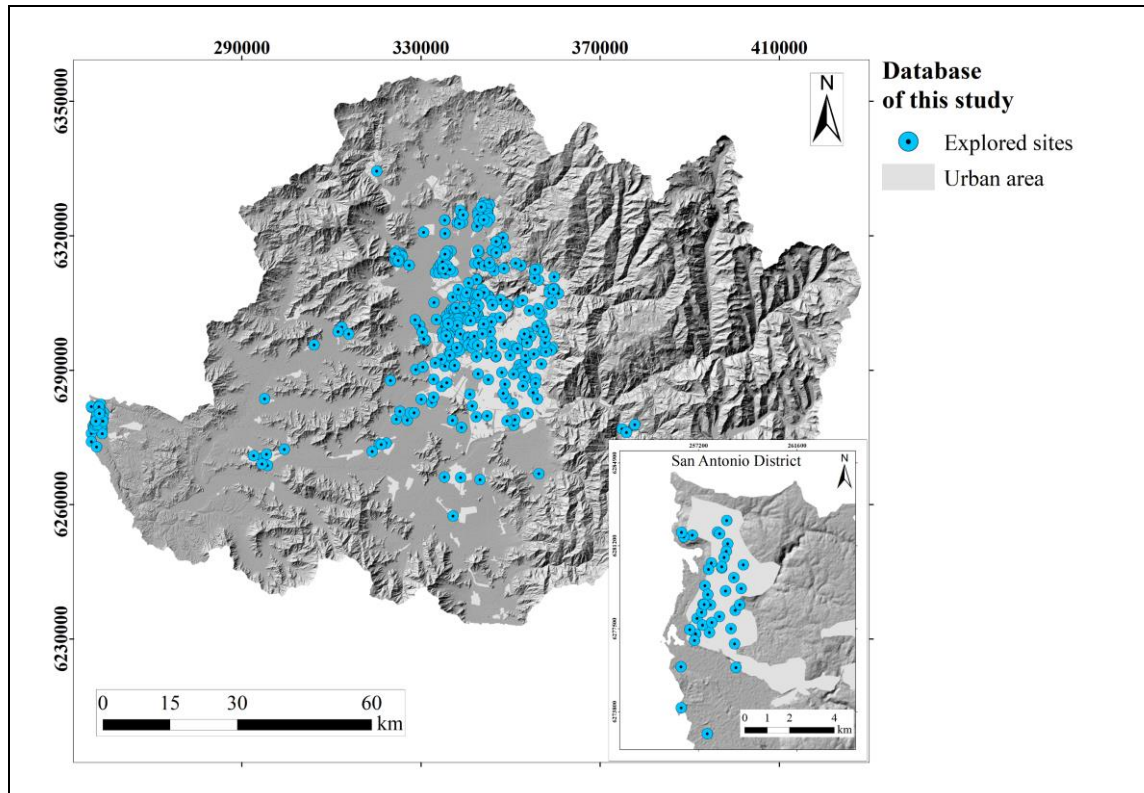
Figure 2-9: Distribution of the 365 sites dynamically characterized in this study.

2.2.6 Gravimetry

The last source of data that was very useful for this study corresponds to the gravity records compiled by Yáñez et al. (2015). In that study, 1,115 gravity stations were used to study gravimetric fluctuations in the Santiago basin. Yáñez and coworkers integrated this information with boreholes, geological surveys and petrophysical measurements to generate a well-constrained model of the depth to the basin's basement.

Specifically, the data from the article by Yáñez et al. (2015) used in this study were the measurements processed in terms of the residual Bouguer anomaly, whose results were interpolated as shown in Figure 2-10. As can be seen, the gravimetric database only covers part of the Santiago basin, i.e. a reduced portion of the study area. Nevertheless, it was

very useful to develop the prediction model in SRM as described in Chapter 4 of this document.
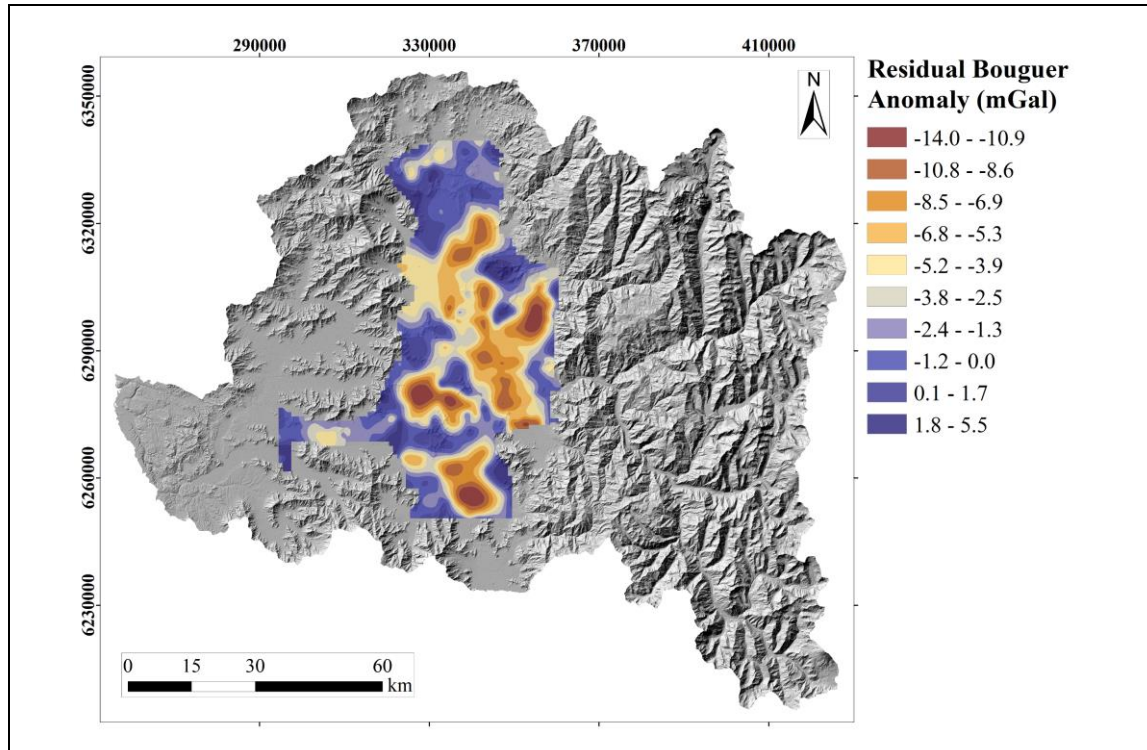
Figure 2-10: Residual Bouguer anomaly interpolated in the SMR. Modified from Yáñez et al. (2015).

3   RESULTS

This chapter presents the results of this study at the scale of the MRB. Firstly, the distribution of the measured data is shown, secondly, the resulting seismic susceptibility map is presented. Then, the predictions obtained by the developed machine learning models are discussed. Finally, hazard maps and exceedance probability maps are presented for the study area.

3.1   Data distribution

Figure 3-1 shows the distribution of V_{s30} measured in the study area, including a closer look to the district of San Antonio to improve the visualization of the sites. Similarly, Figure 3-2 shows the distribution of f_0 and A_0 measured in the study area.
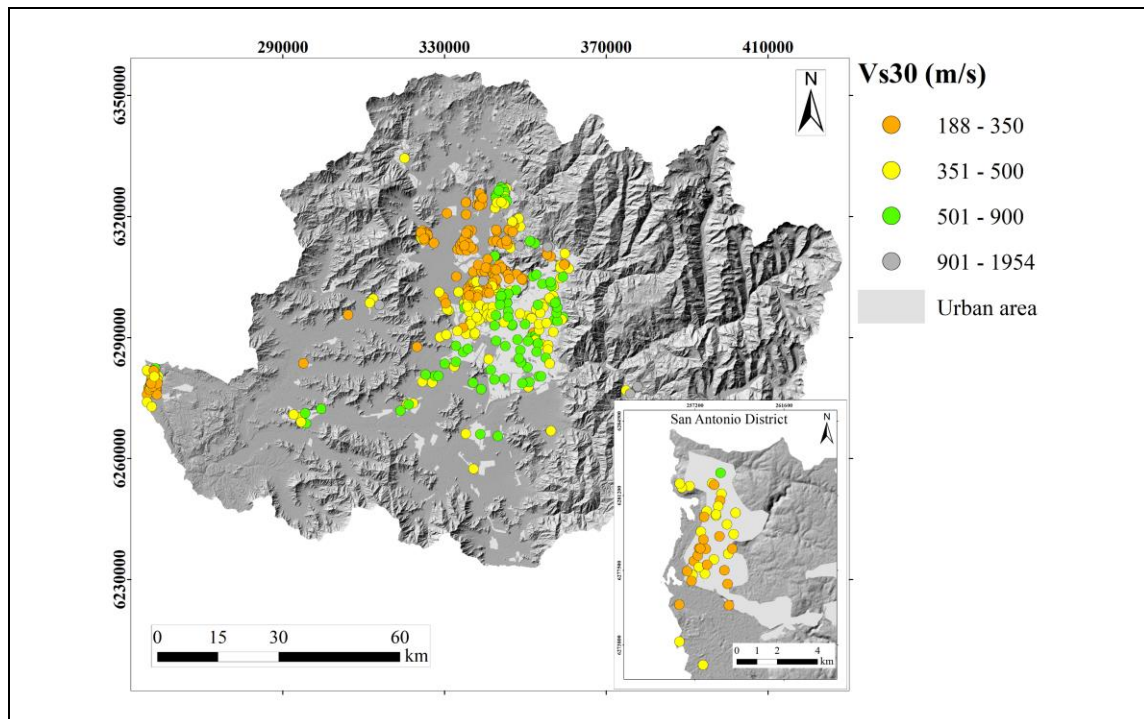
Figure 3-1: Distribution of V_{s30} in the study area.
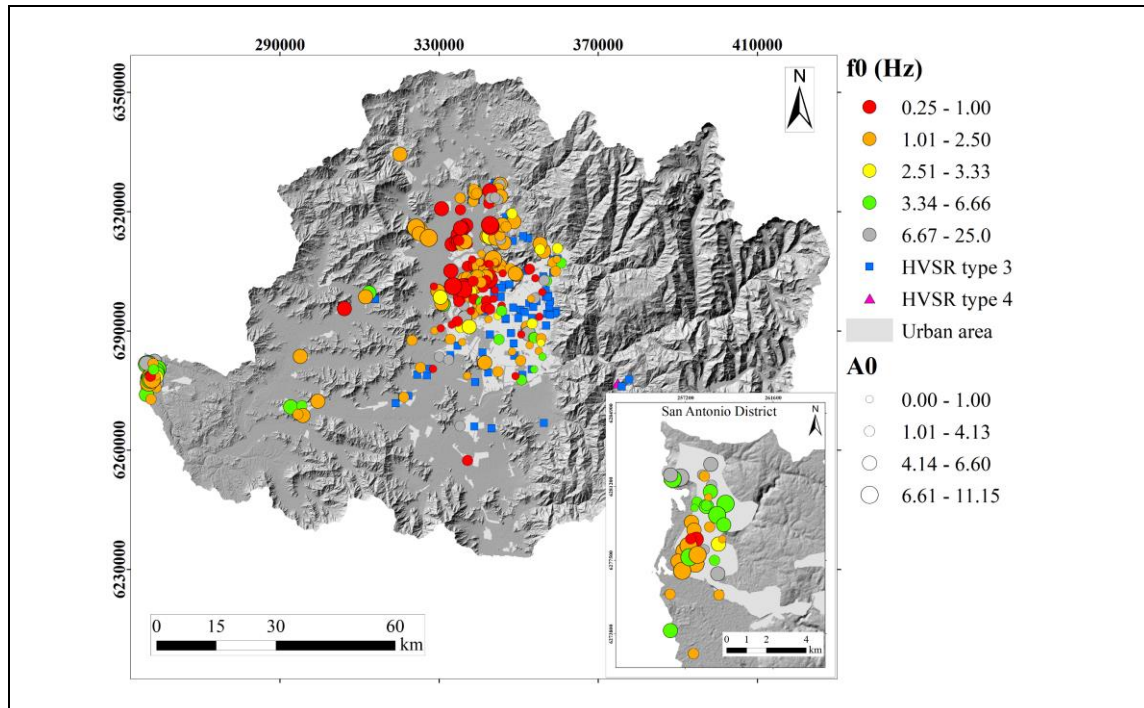
Figure 3-2: Distribution of f_0 and A_0 in the study area.

3.2   Seismic susceptibility index distribution

A seismic susceptibility index was assigned to each site in the study area, as indicated in Table 3-1. It should be noted that if criterion 1 of Table 3-1 is met at a site, but criterion 2 is not met, the site must be downgraded by one category. Although this rule was designed to penalize deep, long-period sites, it was also applied if a HVSR curve of type 4 was obtained (see the types of HVSR curves in Annex 1). Figure 3-3 shows the distribution of the seismic susceptibility in the study area. All the data collected and measured used in this study is provided in Table 7-1 in Annex 2. For those sites with more than one measurement of V_{s30} and/or f_0, we selected the combination giving the most conservative seismic classification.

Table 3-1: Definition of the seismic susceptibility index.

Index category	First criterion: V_{s30} (m/s)	Second criterion: T_0 (s)
A	≥ 900	< 0.15 or flat HVSR
B	≥ 500	< 0.30 or flat HVSR
C	≥ 350	< 0.40 or flat HVSR
D	≥ 180	< 1.00 or flat HVSR
E	< 180	
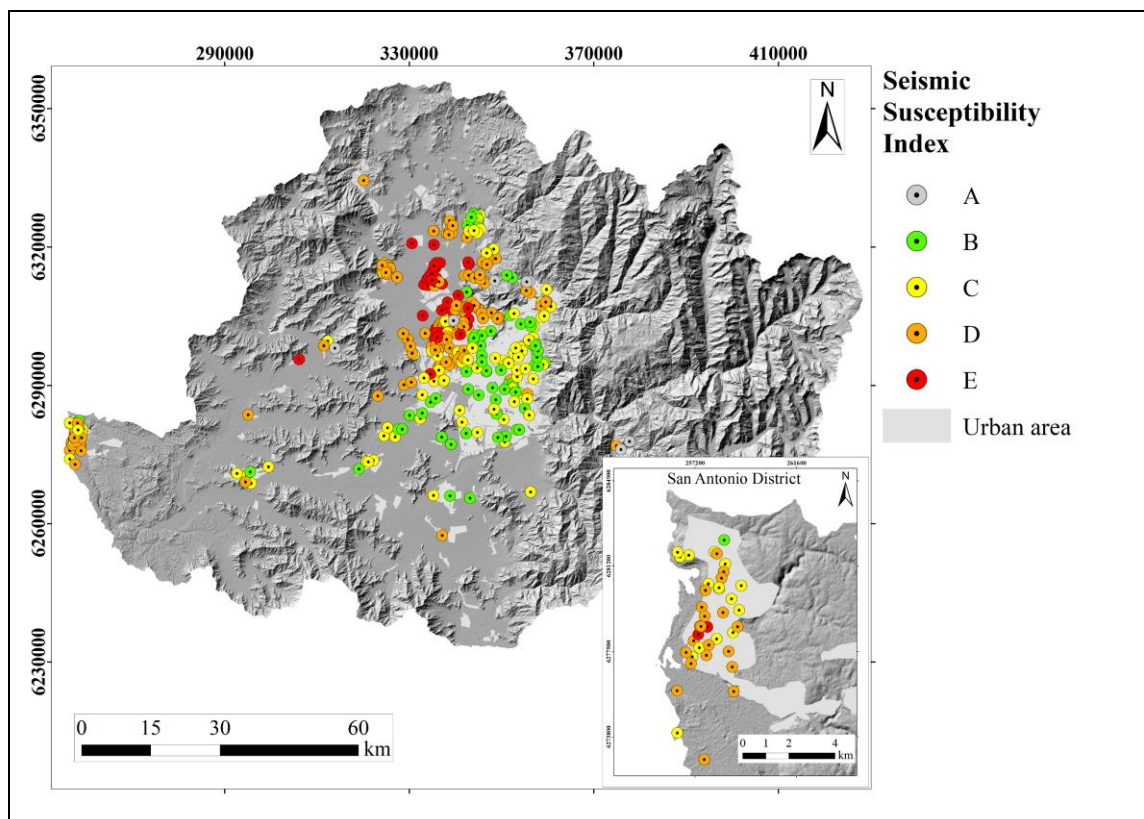
Figure 3-3: Seismic susceptibility index assigned to the sites in the study area.

3.3   **Seismic susceptibility map**

The joint analysis of the site classification of the available seismic response maps and the gravimetry, allowed the generation of an updated seismic microzoning of the MRB (Figure 3-4). The refined microzoning shows that the units with the best seismic response (A) correspond mainly to rock units present in a large part of the MRB, these units are located firstly in the Cordillera Principal, in the Cordillera de la Costa and in some isolated hills within the Santiago basin. Sites with the best seismic response (B and C) are located almost entirely in the Depresión Central, where alluvial, fluvial and volcanic sediments are abundant. Type C sites are also observed in the northeast of the Santo Domingo district. The units with the worst seismic susceptibility index are located specifically in three sectors of the study area: (i) in the northwest of the Santiago basin, where soft soils are abundant, (ii) in the valleys of the Cordillera de la Costa, where alluvial fans of sandy-clay matrices dominate, and (iii) in the coastal zone of the district of San Antonio, composed of fine fluvial and eolian deposits.

Some minor changes are observed in the seismic units with respect to prior maps, mainly in the Santiago basin and in the district of San Antonio, where most geophysical data is concentrated. Nevertheless, despite the seismic unit updates, the refined seismic microzoning shows consistency and is similar to previous maps.
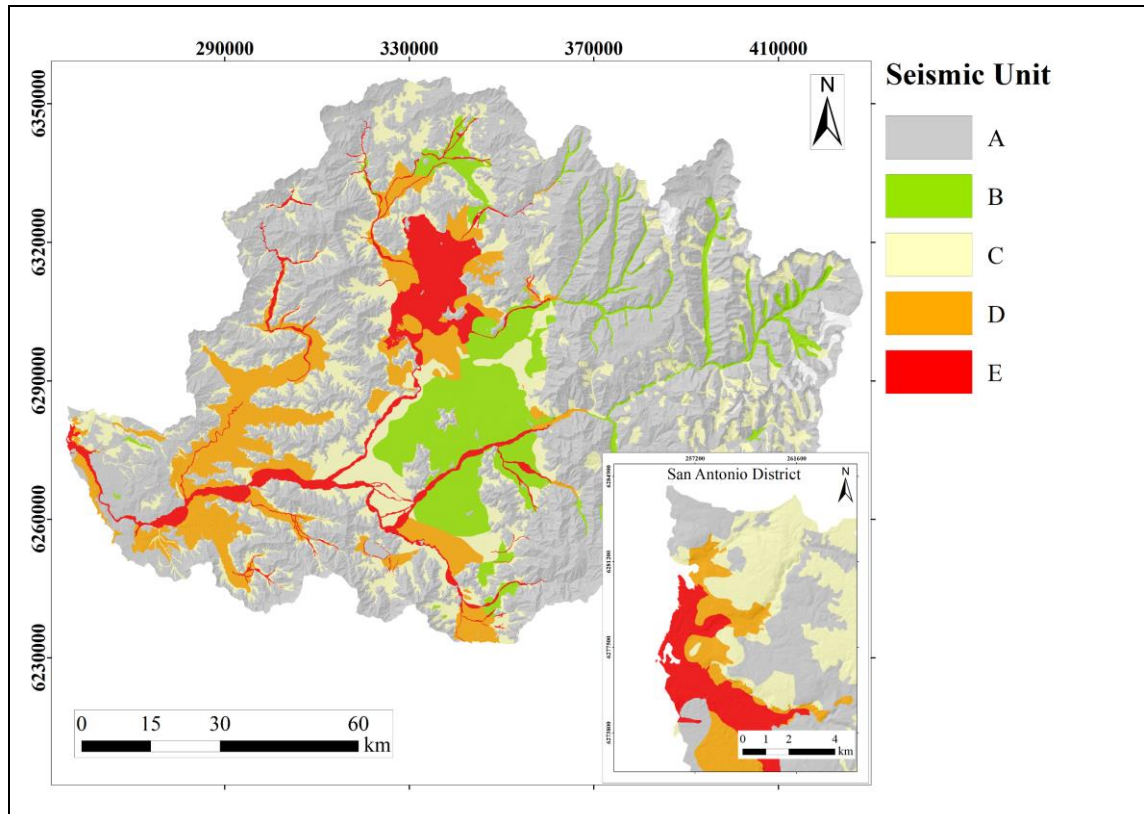
Figure 3-4: Refined microzoning of the Maipo River Basin.

3.4   Maps of predicted dynamic properties of sites

The predictive algorithms presented in Chapter 4 of this study were applied to the complete MRB database to estimate V_{s30}. The settings of the algorithms, the treatment of the data and the validation of the estimates was identical to that presented in Chapter 4 of this document. To avoid duplication, in this chapter we will only show the results applied to the entire study area. Table 3-2 shows the prediction performances of the best models of each algorithm tested to predict V_{s30}.

For the sites where the gravimetric covariant is available, the results show that Linear Regression (LR) was the most robust algorithm to predict V_{s30}, with an RMSE of 69.3 m/s

and an RRMSE of 17.5%. The second-best predictor was Random Forest (RF), with an RMSE of 76.8 m/s and an RRMSE of 18%, while the third-best predictor was Decision Trees (DT) with an RMSE of 104.7 m/s and an RRMSE of 24.9%.

For the sites where the gravimetric covariant is not available, the results show that LR was the most robust algorithm to predict V_{s30}, with an RMSE of 190.9 m/s and an RRMSE of 23.7%. The second-best predictor was RF, with an RMSE of 167.7 m/s and an RRMSE of 25.6%, while the third-best predictor was EN with an RMSE of 238.8 m/s and an RRMSE of 37.3%.

The estimation of f_0 in the study area is shown in Chapter 4 of this document since acceptable results were only obtained in the Santiago basin. Figures 3-5, 3-6 and 3-7 show the predictions of V_{s30} obtained with the best predictive algorithms. The observed V_{s30} are also shown in the maps for a visual comparison of the results.

Table 3-2: Cross-validation of the best models for each algorithm for the V_{s30} predictions.

	Gravimetric covariant included		Gravimetric covariant not included	
Algorithm	RMSE (m/s)	RRMSE (%)	RMSE (m/s)	RRMSE (%)
SK	176.0	35.6	339.4	38.8
LR	**69.3**	**17.5**	190.9	**23.7**
EN	120.8	27.1	238.8	37.3
RF	76.8	18.0	**167.7**	25.6
ANN	199.3	39.3	379.3	67.3
DT	104.7	24.9	242.0	46.0

Note: The bolds show the best performances obtained among the algorithms.
SK: Simple Kriging; LR: Linear Regression; EN: Elastic Net; RF: Random Forests; ANN: Artificial Neural Network; DT: Decision Trees
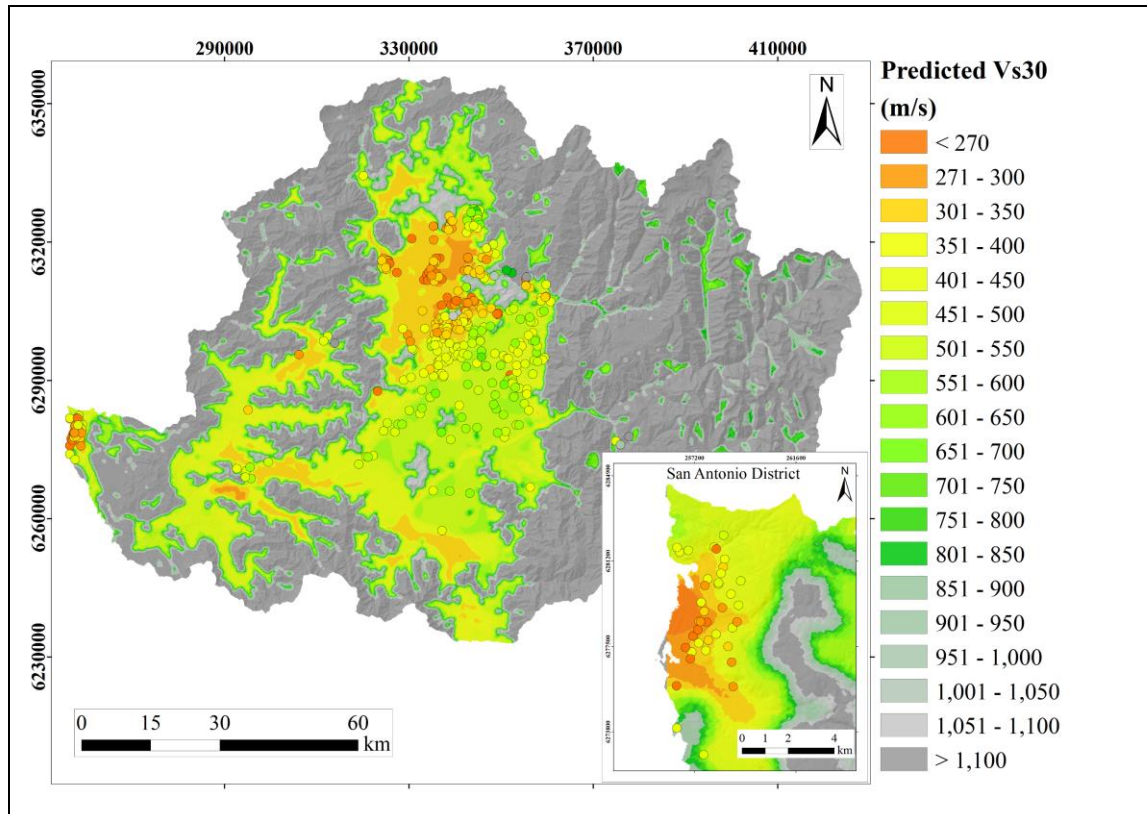
Figure 3-5: V_{s30} prediction map using the Linear Regression model in the study area.
Circles show the observed V_{s30} values in the same color scale.
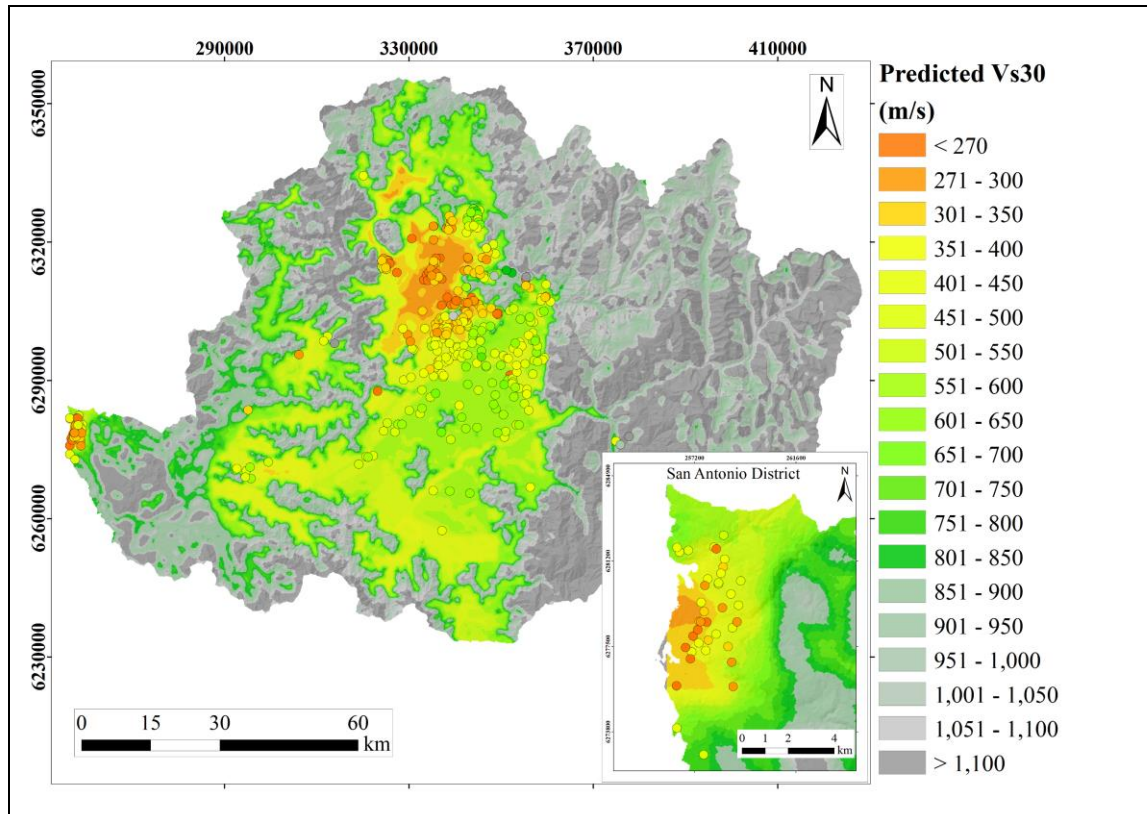
Figure 3-6: V_{s30} prediction map using the best Random Forests model in the study area. Circles show the observed V_{s30} values in the same color scale.
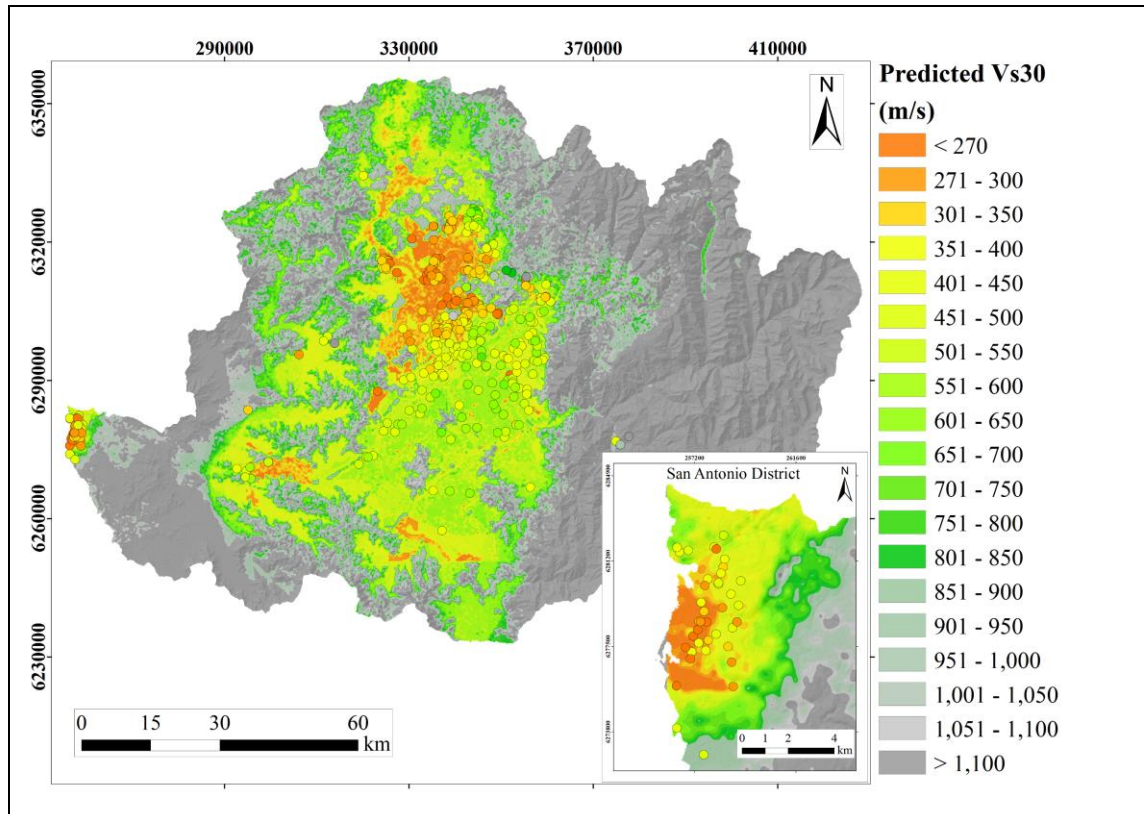
Figure 3-7: V_{s30} prediction map using the Decision Trees model in the SMR and Elastic Net out of it. Circles show the observed V_{s30} values in the same color scale.

3.5 Seismic hazard maps

Once a dynamic characterization of sites for the entire study was generated, it is possible to proceed with the Seismic Hazard assessment to obtain different strong motion indicator. We selected the peak ground acceleration (PGA) as main indicator to simplify combination with other natural hazard that are part of the project founding this research. Figures 3-8, 3-9 and 3-10 show the estimate of PGA in the study area for the best predictions of V_{s30} and using the settings described in Chapter 4.
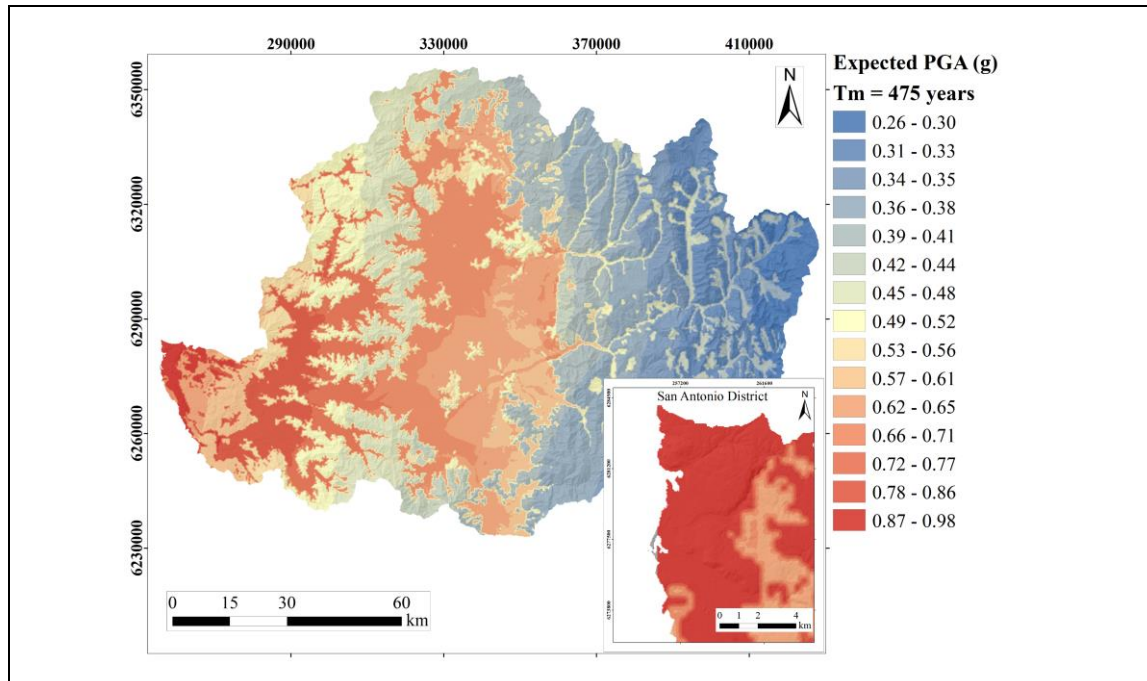
Figure 3-8: Expected PGA map for events with T_m = 475 years, based on the estimates of V_{s30} resulting from the LR model.
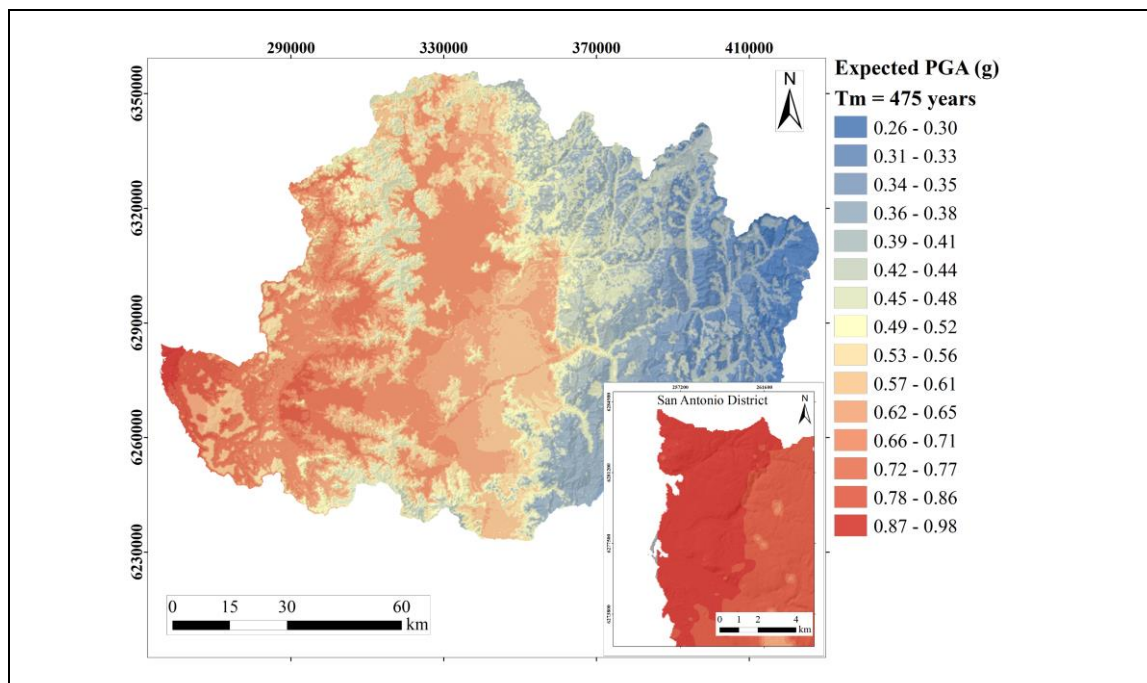
Figure 3-9: Expected PGA map for events with T_m = 475 years, based on the estimates of V_{s30} resulting from the RF model.
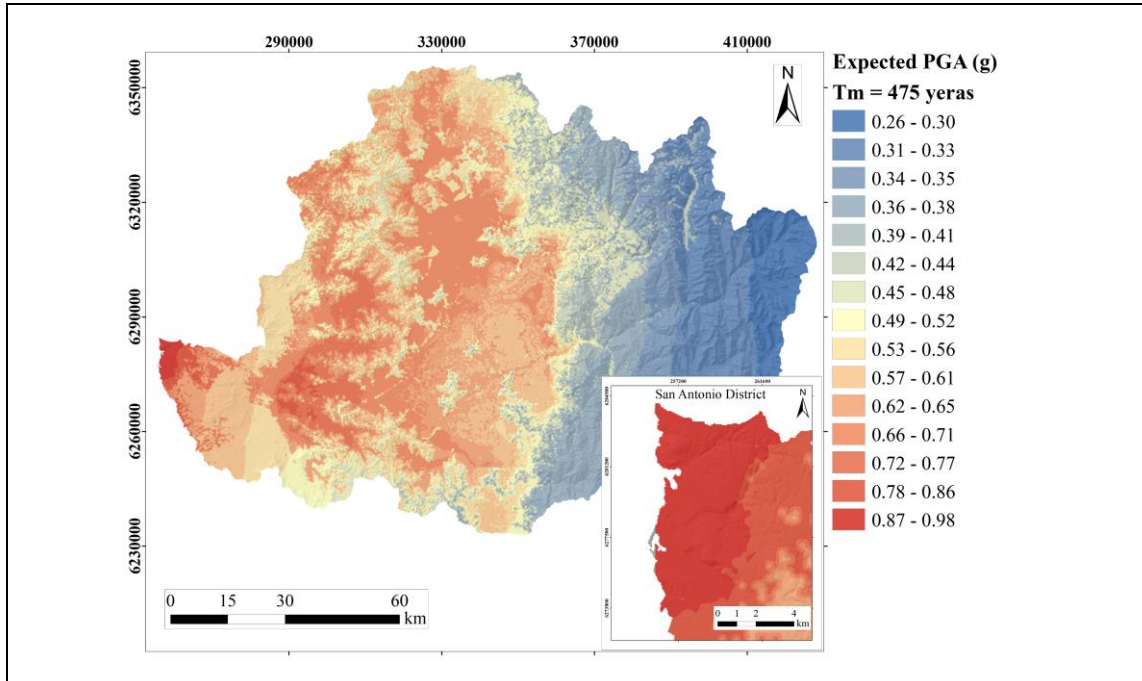
Figure 3-10: Expected PGA map for events with $T_m = 475$ years, based on the estimates of V_{s30} resulting from the DT model.

3.6 Additional applications

V_{s30} distribution estimated in this work can also be used for complementary analysis. For example, a probabilistic seismic hazard analysis (PSHA) which can be combined with a vulnerability analysis. With this purpose, we calculate the probability that an intensity measure IM exceeds the value im during an observation time window T. Then, if the temporal distribution of earthquakes follows a Poisson statistical process, this probability can be calculated as

$$P(IM > im) = 1 - \exp(-\lambda_{IM}T) \tag{2}$$

where λ_{IM} is the annual exceedance rate for the intensity measure IM. This annual exceedance rate can be calculated as

$$\lambda_{IM} = \sum_{i=1}^{N_s} \lambda_0^i \iint P(IM < im|m, r) f_M(m) f_R(m, r) \, dm \, dr \qquad (3)$$

where N_s is the number of seismic sources considered, λ_0^i is the activity of the i-th seismic source, $P(IM < im|m, r)$ is the probability that the intensity measure IM exceeds the value im conditioned to the occurrence of an event of magnitude m at a distance r from a site with known dynamic characteristics (V$_{s30}$ in this case), and $f_M(m)$ and $f_R(m, r)$ are the probability density functions of the magnitude of the event and the location of the source, respectively. Then, if λ_{IM} is known and T is defined, it is possible to compute $P(IM > im)$ using the equation 7. The calculation of this type of indicator was already implemented in the platform used for this research (Candia et al., 2019). Our contribution was to perform this computation supported on a more robust estimation of the dynamic properties of the sites across the study area. Figure 3-11 shows, as an example of this application, the probability distribution that the intensity measure of $Sa(T_0 = 1.62s) = 0.43g$ is exceeded in the Maipo basin, for a time window of $T = 1000$ years. $Sa(T_0 = 1.62s) = 0.43\,g$ was chosen as example because it corresponds to an extensive damage threshold proposed by Ugalde et al. (2020) for 17-story residential buildings with fundamental period $T_0 = 1.62\,s$ in Chile.

This map of exceedance probabilities can be used as input for a subsequent vulnerability analysis. For example, in the project of which this research is part, it was combined with information on susceptibility to landslides triggered by earthquakes to quantify this risk.
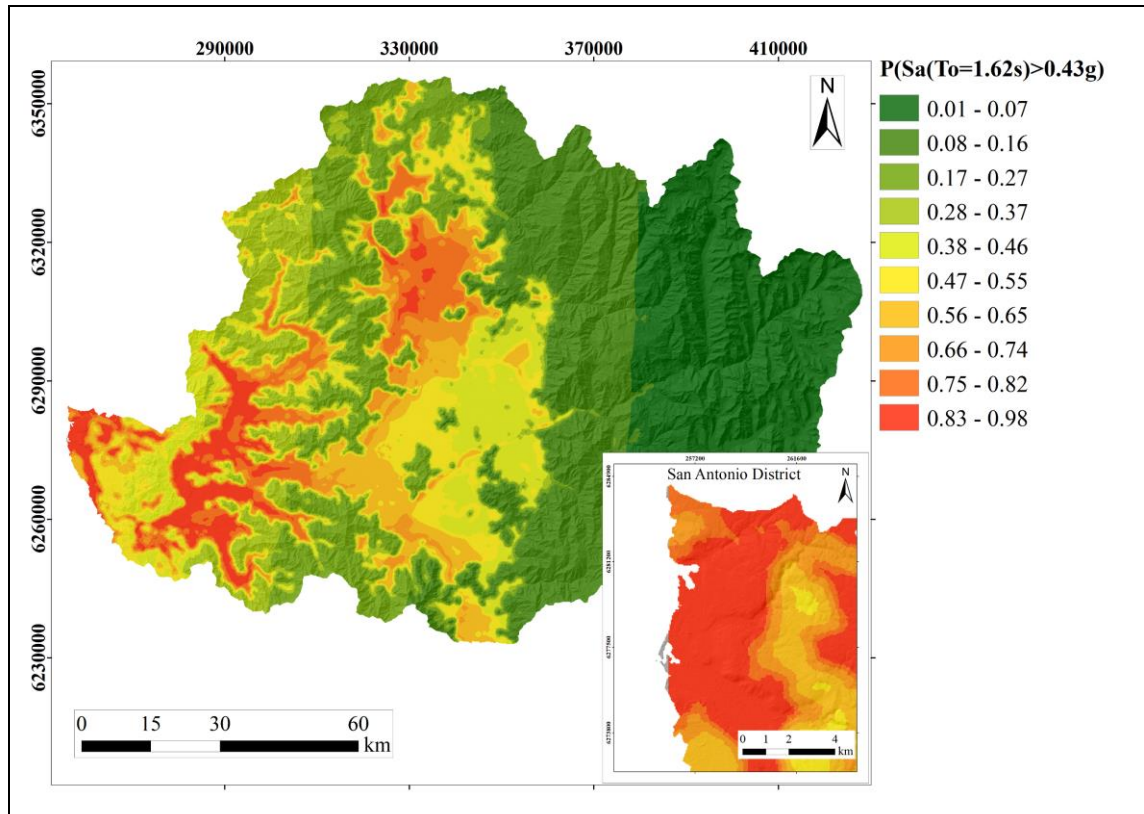
Figure 3-11: Probability of exceeding a measurement intensity of $Sa(T_0 = 1.62s) = 0.43g$ in an observation time window of $T = 1000$ years in the MRB.

4 MACHINE LEARNING TECHNIQUES FOR ESTIMATING SEISMIC HAZARD SUSCEPTIBILITY AND AMPLIFICATION IN THE SANTIAGO DE CHILE BASIN.

4.1 Introduction

Chile is one of the most seismically active countries in the world. The active continental margin where the oceanic plate (Nazca plate) subducts under the continental plate (South American plate), extends between 18° and 47° S. This active margin has generated some of the largest subduction earthquakes on record (e.g. 9.5 Mw Valdivia 1960, Cifuentes, 1989). Observational data show that, along the country, the impact of seismic waves increases in the areas closest to the trench and decreases with increasing distance to this seismogenic source (i.e. Leyton et al., 2010). In addition, the configuration of the national territory and the distribution of the main Chilean cities over the subduction margin locate the urban areas in zones permanently susceptible to major earthquakes. Then, this type of threat is one of the main natural hazards in the country and its urban areas.

Estimations of the recurrence times of large historical earthquakes of Mw > 8 indicate that these events occur approximately every 80 years in front of the Santiago Metropolitan Region (SMR) (Figure 4-1). The last major earthquake that affected much of this area was the 8.0 Mw Algarrobo earthquake of 1985.

The SMR concentrates 50% of the Chilean population and 42% of Chile's gross domestic product, so efforts focused on predicting and minimizing the damage caused by major seismic events are extremely necessary in this region.
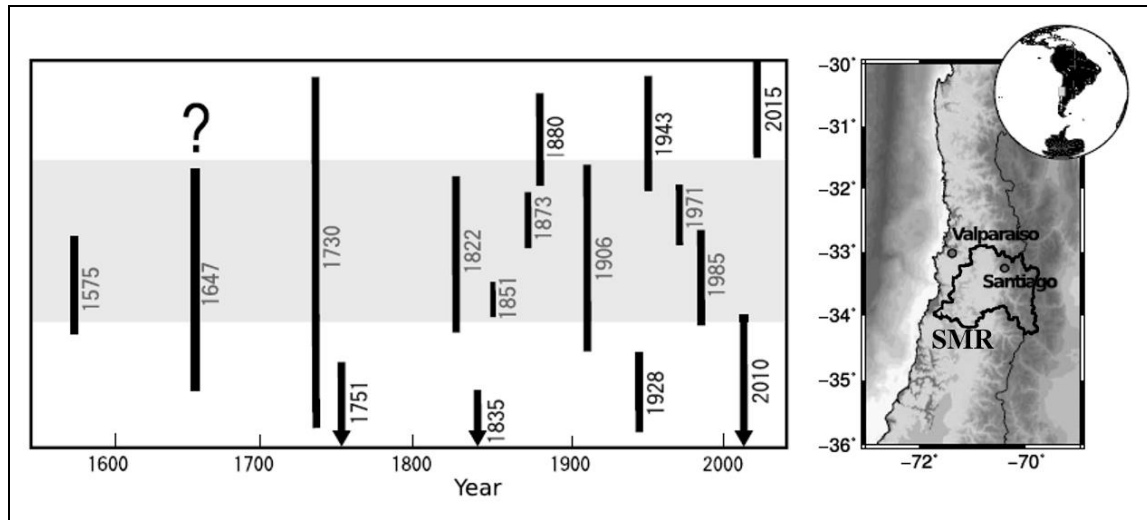
Figure 4-1: Historical recurrence of large earthquakes in central Chile. The length of the bars indicates the approximate extent of the rupture that generated each event, while their widths are proportional to the registered magnitudes. The map to the right indicates the location of the Santiago Metropolitan Region (SMR). (Modified from Bravo et al., 2019).

Several works have shown that the local geotechnical conditions of a site can induce important seismic amplifications (i.e. Aki, 1988), known as site effects. Site effects can be evaluated through the dynamic characterization of the subsoil, measuring dynamic parameters of the site's response to the pass of seismic waves. A key parameter to achieve a correct geotechnical characterization is the shear wave velocity profile, which allows a primary evaluation of the dynamic response of a site (Tokimatsu, 1997). Another very important dynamic property of soils to estimate site effects is the predominant frequency of sites, f_0, defined as the frequency associated with the impedance contrast that predominates in a site (Maringue et al., 2021). As in many other countries, in Chile the seismic code classifies a site based on the value of the harmonic average of the propagation velocity of shear waves in the first 30 m depth, V_{s30}. However, other parameters from in situ tests, such as drilling, standard penetration and laboratory tests, are also required for this classification. Regardless, geophysical techniques based on surface waves allow determining the V_{s30} and f_0 parameters of a site in a noninvasive, fast

and low-cost manner (Becerra et al., 2015). Several investigations in different cities have shown that dynamic characterization based on geophysical surveys using surface wave methods are reliable techniques to evaluate site effects (Chavez-Garcia & Cuenca, 1998; Scott et al., 2006; Tuladhar et al. 2004).

Seismic site amplification and seismic hazard maps are crucial inputs for decision making and risk evaluation in places where seismicity imposes a significant risk to human life and infrastructure. Therefore, it is useful to have accurate and reliable maps that account for the seismic response of the sites. However, despite the knowledge of the advantages offered by geophysical methods for the dynamic characterization of sites, in Chile, the use of dynamic properties in the generation of seismic response maps is an incipient practice. In the case of SMR, most of the available seismic response maps do not consider the dynamic characteristics of the subsoil and have been developed based on observations of seismic damage distribution. Only a few studies have incorporated site dynamic properties and made the first efforts to generate this type of maps (e.g., Leyton et al., 2011).

Digital soil mapping (DSM) uses statistical models to generate digital representations of the spatial distribution of soil properties using point soil observations and spatially exhaustive environmental covariates (proxies or independent variables) (McBratney et al., 2003; Scull et al., 2003). In recent decades, DSM has proven to be quite successful in producing soil property maps, capturing their main patterns of soil spatial variation (e.g. Moore et al., 1993; McBratney et al., 2000). However, the application of the DSM in seismic geotechnical engineering in Chile is hampered by limited dynamic site characterizations and data availability.

Spatial interpolation of natural variables is important in many scientific fields, and throughout history numerous interpolation techniques have been developed to achieve these purposes. In the 1980s, the krigging geostatistical interpolation technique was

introduced, gaining popularity as it had the advantage - unlike previous techniques - of considering the spatial correlation of the data and being able to quantify the interpolation error (Matheron, 1963). Besides, the statistical approach of data mining has recently been providing useful tools for DSM. This approach identifies patterns in datasets through statistical methods, transforming this information into a perceptible structure for further use (Khaledian & Miller, 2020).

Machine learning (ML) is a data analysis method and a field of artificial intelligence that uses data mining to learn and build a model that is capable of discovering and quantifying common patterns revealed by the data (Clifton, 2010). In recent years, ML techniques have been increasingly used for spatial interpolation in fields such as soil science and geology (e.g. Li & Heap, 2014). ML is highly dependent on the relationship between the target variable and its associated covariates and can produce remarkably accurate results if this correlation is strong (Sekulic et al., 2020). A great advantage of ML prediction models over traditional techniques is their ability to capture the non-linear interaction between the data without having to assume a functional form of the relation between the input and output data (Kohestani et al., 2015).

However, there have been few attempts to use these techniques in the area of seismic geotechnical engineering. Pokherl (2013) used kriging to estimate liquefaction potential in alluvial soil from Saitama, Japan. Kohestani et al. (2015) used ML tools to predict liquefaction potential in soils based on cone penetration tests. Thomson and coworkers used variants of kriging to estimate V_{s30} in Kobe and California, in 2010 and 2014, respectively. Though there have been several attempts to predict V_{s30} by geostatistical methods, until now - at least in Chile - no tested techniques of ML have been used for spatially predicting the dynamic site properties (V_{s30} and f_0). Thus, we believe that ML techniques can reasonably predict V_{s30} and f_0 values and improve the accuracy of quantitative seismic hazard assessments in the Santiago de Chile basin.

This paper aims to improve the quality and accuracy of seismic response and hazard maps in the Santiago Metropolitan Region (SMR). We present two results: (i) an updated seismic microzoning of the SMR based on recent measurements of dynamic properties of sites through the basin, and (ii) a methodology that uses regression machine learning computational tools, which integrates the geology, geophysical data and seismic geotechnical engineering concepts, to predict V_{s30} and f_0 accurately in unsampled areas. We compared six predictive algorithms to estimate V_{s30} and f_0: simple krigging, linear regression, elastic net, random forests, artificial neural networks and decision trees. The best predictions obtained were used to generate seismic hazard maps in the study area, through a state-of-the-art software that uses ground motion prediction equations (GMPE), seismicity models and seismic scenarios to assess the seismic hazard due to subduction and cortical earthquakes. The major contribution of this work is to introduce a novel estimation methodology based on artificial intelligence models to extend local measurements of site's dynamic properties in an area of interest. This information can be used to quantitatively estimate the seismic hazard at regional scale.

4.2 Methodology

4.2.1 Study area and available seismic response maps

The study area of this research is the Santiago de Chile basin, located in the center of the SMR (see Figure 4-2). This area contains a sedimentary infill of the Santiago basin. The alluvial sedimentary infill has accumulated between the Main Cordillera and the Coastal Cordillera, reaching maximum depths in the range of 350 - 500 m (Araneda et al., 2000; Yáñez et al., 2015). There are two seismic hazard maps for this area which are seismic response maps developed after the events of 1985 and 2010. These maps define seismic units in the basin representing a seismic microzonation:

- Seismic response of the Santiago Metropolitan Region (von Igel et al., 2004): in this work the seismic response of the geological units of the SMR is qualitatively determined based on seismic intensities collected after the March 3rd earthquake of 1985 (Ms 7.8) and available geological information. This map does not incorporate quantitative information on the dynamic characteristics of sites and was developed mainly based on damage observations.

- Seismic zoning of the SMR, Chile (Leyton et al., 2011): in this work, a seismic zoning of the Santiago basin is carried out based on the surficial geology, available measurements of the predominant period (Bonnefoy-Claudet et al., 2009) and the distribution of the damage observed for the Maule earthquake.

4.2.2 V_{s30} and f_0 measurements

To obtain the dynamic properties of the sites in the basin, 312 measurements of V_{s30} and f_0 were recorded using surface wave geophysical methods. In cases in which more than one measurement of V_{s30} and f_0 was available, the uncertainty in the values of these parameters was evaluated to report the level of accuracy of field experiments.

V_{s30} measurements were obtained through an inversion process of the empirical dispersion curve of each site. These curves were obtained using a methodology that combines active sources (hammer) with passive sources (ambient noise), with a multichannel analysis approach (Humire et al., 2015). To obtain V_{s30}, the Geode® equipment manufactured by the Geometrics® company was used, while to obtain f_0, a triaxial Tromino® seismometer was used. One-dimensional arrays were used for both active and passive measurements, while two-dimensional arrays were intended for passive measurements. As a general rule, the objective of the surveys was to describe the dispersion curve between a wavelength of 10 m and 90 m using a combination of active

and passive methods. The methods used in this investigation were the f-k method (frequency-wave number method) (Lacoss et al., 1969; Kvaerna & Ringdahl, 1986) for active 1D and passive 2D arrays, the SPAC method (spatial autocorrelation method) (Aki, 1957) for passive 2D arrays, and the ESPAC (Hayashi et al., 2008) method (extended spatial autocorrelation method) for passive 1D arrays. In sites with more than one estimate of V_{s30}, measurements showed coefficients of variation (CV) between 0% and 35.8%. Furthermore, in 50% of the cases the CV was less than 2.6%.

To obtain f_0, the horizontal-to-vertical spectral ratio (HVSR) or Nakamura's technique was used (Nakamura, 1989; Molnar et al., 2018). This technique estimates the ratio of the Fourier amplitude spectrum between horizontal components and vertical component produced by environmental vibrations. Then, the predominant period ($T_0=1/f_0$) is defined by the peak of the HVSR curve (Pastén, 2007), and the amplitude of this peak is defined as A_0. Flat HVSR curves are associated with rigid sites where no significant stiffness contrast is observed between sediment and rock (e.g. gravels). In this investigation, the analysis was carried out following a variation of the Nakamura method (Leyton et al., 2012) which considers fixed windows of 60 seconds, applying the Stockwell Transform (S-transform) in each of these windows. For f_0, CV between 0% and 46.4% were obtained, and in 50% of the cases this value was lower than 5.6%. For A_0, a CV between 0% and 50.8% were obtained, and in 50% of the cases this value was less than 12%.

Figure 4-2a shows the spatial distribution of the V_{s30} values, while Figure 4-3a shows simultaneously the distribution of f_0 and A_0. These sites concentrate near the urban areas of the study area. Figures 4-2b and 4-3b show the histograms of the data distribution of V_{s30} and f_0, respectively.
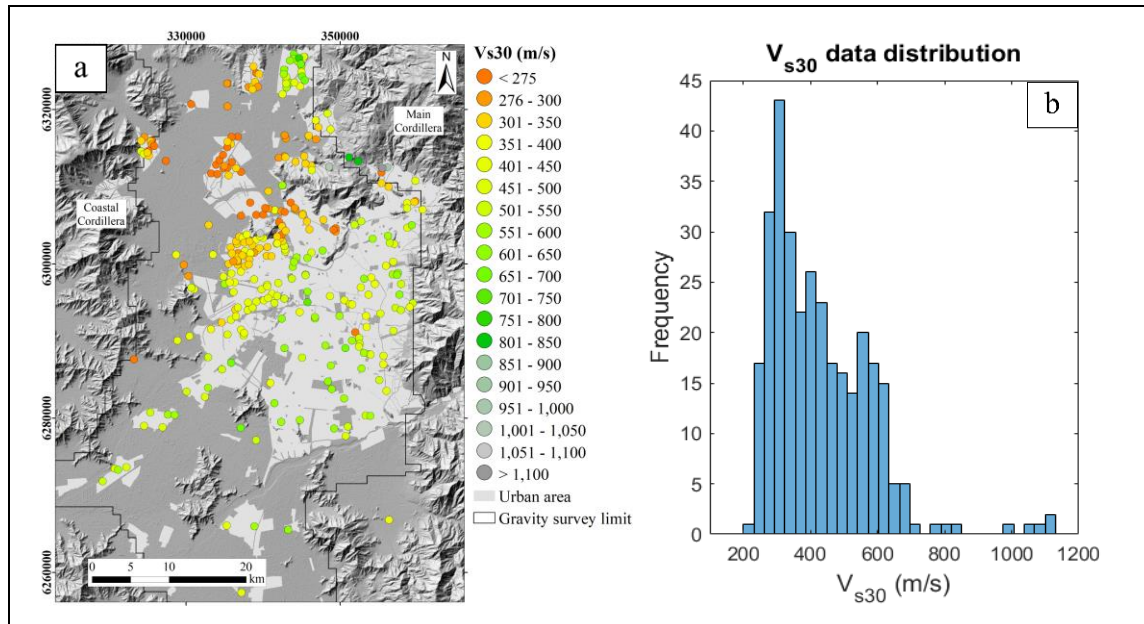
Figure 4-2: a) Distribution of V_{s30} in the study area, and b) Distribution histogram of V_{s30}.
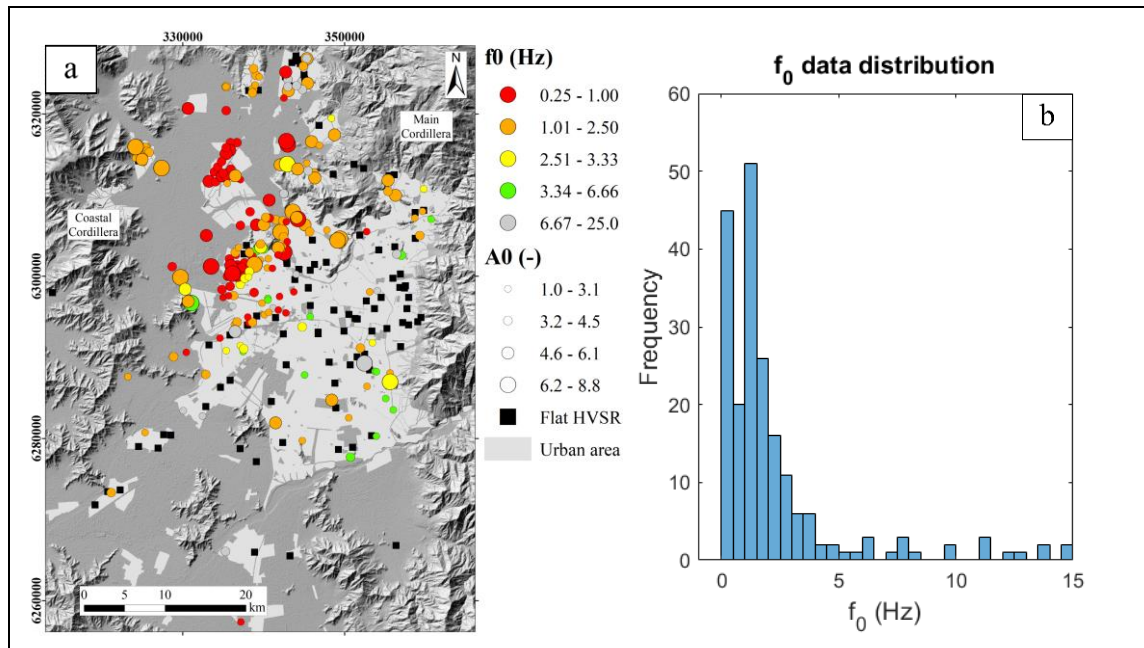
Figure 4-3: a) Distribution of f_0 in the study area, and b) Distribution histogram of f_0.

Among the 312 sites considered, 101 sites reported flat HVSR curves

4.2.3 Seismic susceptibility index

Chile is currently in the process of improving its seismic classification system for residential buildings. The new classification simultaneously uses the value of V_{s30} and the estimate of the predominant period T_0, which were used as a susceptibility index in this work. Five indices were defined: A, B, C, D and E, according to the criteria shown in Table 4-1 (modified from Verdugo et al., 2019).

Table 4-1: Definition of the seismic susceptibility index.

Index category	First criterion: V_{s30} (m/s)	Second criterion: T_0 (s)
A	≥ 900	< 0.15 or flat HVSR
B	≥ 500	< 0.30 or flat HVSR
C	≥ 350	< 0.40 or flat HVSR
D	≥ 180	< 1.00 or flat HVSR
E	< 180	

4.2.4 Seismic units update

To update the seismic units of the study area, the current seismic response maps were compared against the distribution of seismic susceptibility indices from Table 4-1, to solve the limits of the units. Five seismic units were defined in this investigation, following the seismic susceptibility indices (A, B, C, D and E), where a seismic unit classified as A is the one with the best seismic response (i.e. rock) and a unit classified as E is the one with the worst response in terms of seismic amplification expected due to site effects (i.e. very soft site).

Additionally, gravity models of the Santiago de Chile basin published by Yáñez et al. (2015) were incorporated to fill the information gaps in areas where there was not enough data from V_{s30} and f_0 measurements to update limits between seismic units determined by geologic techniques. The area in which this study is available is shown in Figure 4-2a. As can be seen, the gravity model covers most of the Santiago de Chile Basin.

The direct gravimetric residual is also expected to have a good correlation with f_0, mainly in soft soils, because it provides an idea of depth to a significant change in density or gravimetric contrast (Maringue et al., 2021). Then, units with a considerable gravimetric anomaly, low frequencies ($< 1\ Hz$), fine granulometries and/or presence of surficial volcanic ash were classified as low seismic response units (D or E). The updated seismic units are shown in Figure 4-6.

4.2.5   Prediction of V_{s30} and f_0

This chapter describes the procedures and considerations used to generate a predictive model of V_{s30} and f_0 in the study area. First, the database and the covariates used to train the predictive models for each explored algorithm are presented; secondly, the algorithms used are briefly described. Additionally, the methods for validating and evaluating the predictive performance of the models are detailed.

4.2.5.1   Data and choice of covariates

As mentioned earlier, proper choice of training covariates from ML predictive models is key to obtaining reasonable and accurate estimates in DSM. In this work, we chose the covariates shown in Table 4-2. The decision to use punctual covariates such as slope, topographic elevation, and geological typology is based on previous work that showed an improvement in the performance of predictive models in predicting V_{s30} (Wills & Clahan,

2006; Wald & Allen, 2007), so these three covariates were included in all the models for training. The topographic slope and the topographic elevation were obtained from a digital elevation model (DEM) of 12.5 m resolution available in public satellite data (https://asf.alaska.edu), while the surficial geology covariate was obtained directly from the seismic susceptibility map (updated seismic units shown in Figure 4-6).

Despite the success of ML predictive models in DSM, most of these approaches do not consider the possible spatial correlation between the observed data and focus mostly on punctual covariates, thus they do not fully exploit the available spatial information. Numerous recent investigations have shown that the inclusion of spatial covariates (in addition to punctual ones), such as distance and inverse of distance to neighboring observations, considerably improves the predictive capacity of ML models in DSM (Beguin et al. 2017; Deng 2020; Sekulic et al., 2020). Therefore, to take full advantage of the spatial information of the available observation, we used a combination of punctual and spatial covariates to train the models, as indicated in Table 4-2. Figure 4-4 generally describes the spatial covariates chosen in this work.

Note that since gravimetry does not cover the entire study area (see Figure 4-2a), we worked in two independent areas: a zone with gravimetric information and another zone without it. The predictive models for both areas only differed in the inclusion of the gravimetric covariate in their training.

Although several measurements were available at some sites which made it possible to generate some uncertainty indicators, only a smaller set of sites (35%) had this information available. To simplify the training of the models, the pair V_{s30} and f_0, leading to the most conservative classification according to the criteria in Table 4-1, was selected for training.

Table 4-2: Description of the covariates used to train the models and predict the values of V_{s30} and f_0 in Santiago basin.

Covariate	Definition	Unit
Slope	Maximum elevation change rate of each pixel.	°
Elevation	Elevation above sea level according to the DEM used.	m
Seismic unit	Seismic unit of the Seismic Susceptibility Map that contains the evaluation point.	-
Gravity *	Residual Bouguer anomaly measured at site.	mGal
External seismic unit	Seismic unit of the Seismic Susceptibility Map that does not contain the evaluation point but is the closest to it.	-
Edge distance	Inverse of the minimum distance between the evaluation point and the seismic unit that contains it.	1 / km
Distance to closest observations * *	Inverse of the minimum distance between the evaluation point and the observation closest to it.	1 / km
V_{s30} closest observations * *	Value of V_{s30} in the closest observation to the evaluation point.	m / s
HVSR peak in the closest observations * *	Closest observation to the evaluation point that has a peak in the HVSR curve (1) or does not present a peak (0).	Binary
Nearest predominant frequency * *	Predominant frequency (f_0) measured at the closest observation to the evaluation point.	Hz
HVSR amplitude * *	Amplitude of the HVSR curve (A_0) at the closest observation to the evaluation point.	-

* Covariate was used only in the Santiago basin area where the gravimetric study was carried out.

* * Covariates were calculated for the 6 closest observations to each evaluation point.
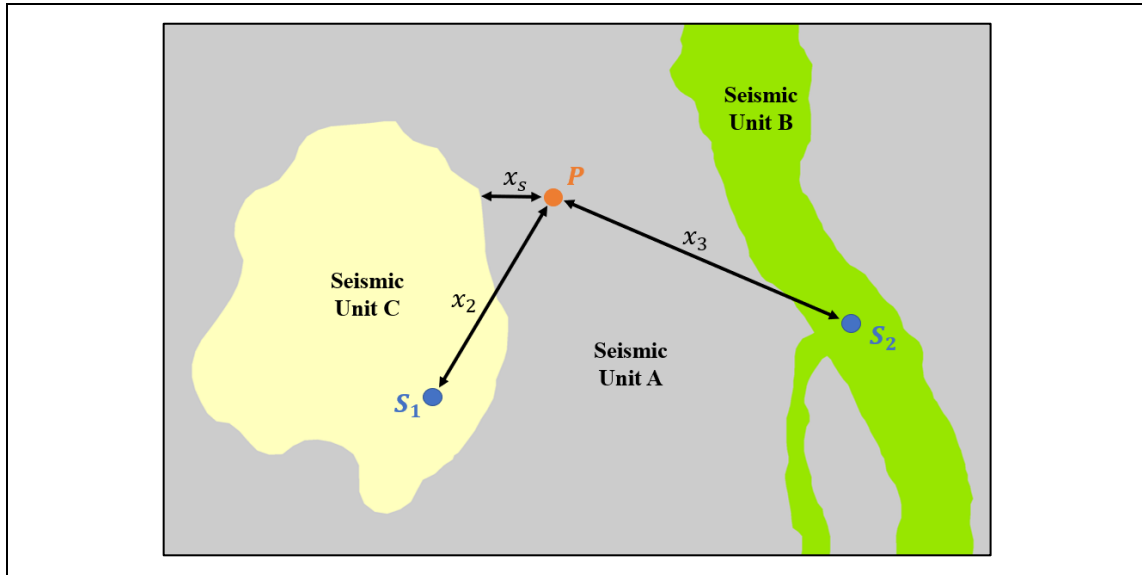
Figure 4-4: Example of the spatial covariates associated with a point P in the study area. In this case, point P is in seismic unit A, x_g is the shortest distance from P to the boundaries of this seismic unit. Seismic unit C is the closest unit to point P. On the other hand, x_2 and x_3 are the two shortest distances to P of S_1 and S_2, where V_{s30} and f_0 are known and are considered spatial covariates associated with the point P.

4.2.5.2   Predictive methods

As mentioned above, there are numerous methods for predicting soil properties from a sample data set. In this paper, geostatistical and ML predictive methods are compared. The geostatistical method tested was Simple Kriging (SK), while the ML methods were Linear Regression (LR), Elastic-Net (EN), Random Forests (RF), Artificial Neural Networks (ANN) and Decision Trees (DT). It is also important to note that the hyperparameters (tuning parameters) are the variables that are used to setting the algorithms. These variables can be adjusted by trial and error until a minimum amount of error is obtained when the predictions are validated. The way in which the predictive algorithms used in this article work will be briefly described below.

- SK: This is a generalized least-squares regression algorithm that assigns weights for the surrounding measured values to derive a prediction for each location. These weights, in addition to being based on the distance between the measured points and the prediction site, are based on the general spatial arrangement of the measured points (Augusto Filho et al., 2017). This method considers that the spatial fluctuation of the mean of the observations is unknown, but constant (Thomson et al., 2010).

- LR: This algorithm fits a linear model with coefficients to minimize the sum of the squares in the difference between the observed and predicted values by the linear approximation (Hutcheson & Sofroniou, 1999).

- EN: Is a linear regression algorithm that combines two linear models: (i) the Ridge method, that addresses some of the problems of Linear Regression by imposing a penalty on the size of the coefficients; and (ii) the Lasso method, that estimates sparse coefficients. EN learns from its shortcomings to improve the regularization of statistical models and is useful when there are multiple features correlated with one another (Friedman et al., 2010).

- RF: It is a set learning algorithm that randomly selects a group of observations from the larger set, to build a decision tree that is associated with this group. The process is repeated to build multiple decision trees based on different observation sets. Typically, two-thirds of the observations are used for algorithm training, and the rest are used to test model error. RF randomly permutes the arrangement of the covariates in the selection of the observation groups, considering all the possibilities of arrangement of covariates. Finally, the predictions are based on the average of the results produced from thousands of decision trees. It is currently the most widely used ML algorithm in DSM, and it often shows excellent potential when it comes to spatial data (Boulesteix, 2012; Deng, 2020).

- ANN: These are algorithms that mimic biological neural networks. They build a set of nodes called artificial neurons, forming a network. Through multiple layers of the network, information is transmitted from one neuron to another. The connection between neurons consists of weights that define the network architecture, organize the layers, and adjust the parameters to learn from the data. Training the network consists of comparing the input to the output and calculating a residual, then the algorithm goes back through the layers to fit the equation of the network and recalculate the residual. This process is repeated until a minimum residual is reached. It is a common and longstanding algorithm used in DSM (Behrens, 2005; Were et al., 2015).

- DT: These models divide the data space and fit a simple prediction model within each partition. A decision tree is the graphical result of each partition. DT are intended for dependent variables that take continuous or ordered discrete values (Breiman et al., 1984).

To predict V_{s30} and f_0, a total of 47 models were tested. The models and their hyperparameter settings are shown in Table 4-3. In this Table, each tested model corresponds to a combination of the hyperparameter values shown in the setting column.

Table 4-3: Tested models and hyperparameter settings. The total of tested models corresponds to the possible combinations of the values of the presented hyperparameters.

Algorithm	Number of tested models	Hyperparameter	Settings
SK	1	Power of the inverted distances	1
LR	1	-	-
EN	2	Total penalty value (α)	1, 2
		Penalty ratio (ρ)	0.5
RF	25	Number of trees	20, 40, 50, 80, 100
		Proportion of variables considered (%)	10, 20, 40, 60, 80
ANN	17	Number of neurons per layer	10, 20, 30, 50, 75, 100
		Maximum number of layers	5
		Maximum number of neurons	100
		Trigger function	tanh*
		Training method	lbfgs*
DT	1	-	-

* 'tanh' refers to the hyperbolic tangent function; 'lbfgs' refers to an implementation of the BFGS quasi-Newton method for nonlinear optimization. For more information about the hyperparameters of the ML models used in this paper visit https://scikit-learn.org/stable/.

4.2.5.3   Predictive performance evaluation

All models were run in the Python environment. The first treatment to the original database was the application of the Data Augmentation technique. This technique consists in artificially increasing the initial number of observations, as well as their covariates, to obtain an increased training dataset, while preserving the associations present in the original data (Padarian et al., 2019). The use of Data Augmentation has been shown to reduce the variance and overfitting of ML models, improve their robustness, and avoid biased results (Shorten & Khoshgoftaar, 2019; Padarian et al., 2019; Roudier et al., 2020; Zhong et al., 2020). This technique has recently been adapted to ML for cases where the amount of training data is very limited; it is especially beneficial for ANN models, since,

unlike RF, ANN are sensitive to small sample sizes (Khaledian & Miller, 2020). Additionally, this technique is expected to allow training of predictor models under difficult prediction scenarios, such as sites with few or no measurements in their proximity (which is the enabling assumption of spatial interpolation).

After testing with data augmentations of 10, 20, 30, 40 and 50 times the size of the database, the largest increase was applied because with it marginally better results were consistently obtained for both V_{s30} and f_0. All models were trained with 90% of the augmented database (training sets) by cross-validation and validated with the remaining 10% (testing sets). To ensure that the comparisons between the models were valid, the same set of covariates was kept for the training of all predictive models (see Table 4-2). For all models, the root mean squared error (RMSE) and the root relative mean squared error (RRMSE) were calculated. Additionally, the predictive models of f_0 allowed evaluating the probability of a peak in the HVSR curve. For those obtained probabilities less than 60%, a flat HVSR curve was assumed. The error rate in the prediction of this probability (ErrRate) was also quantified. These errors were calculated as

$$RMSE = \sqrt{\frac{1}{n}\sum(y'_i - y_i)^2} \tag{4}$$

$$RRMSE = \sqrt{\frac{1}{n}\sum\left(\frac{y'_i - y_i}{y_i}\right)^2} \tag{5}$$

$$ErrRate = \frac{\sum|b'_i - b_i|}{n} \tag{6}$$

where $i = 1..n$ is the i-th iteration and n is the total number of tests, y_i' is the i-th predicted value and y_i is the i-th observed value. b_i' is a binary value equal to 1 if the i-th prediction has a peak in the HVSR curve and equal to 0 otherwise, while b_i is also a binary value equal to 1 if the i-th observation has a peak in the HVSR curve and equal to 0 otherwise. Figure 4-5 shows the main steps in the training of the predictive models of V_{s30} and f_0.
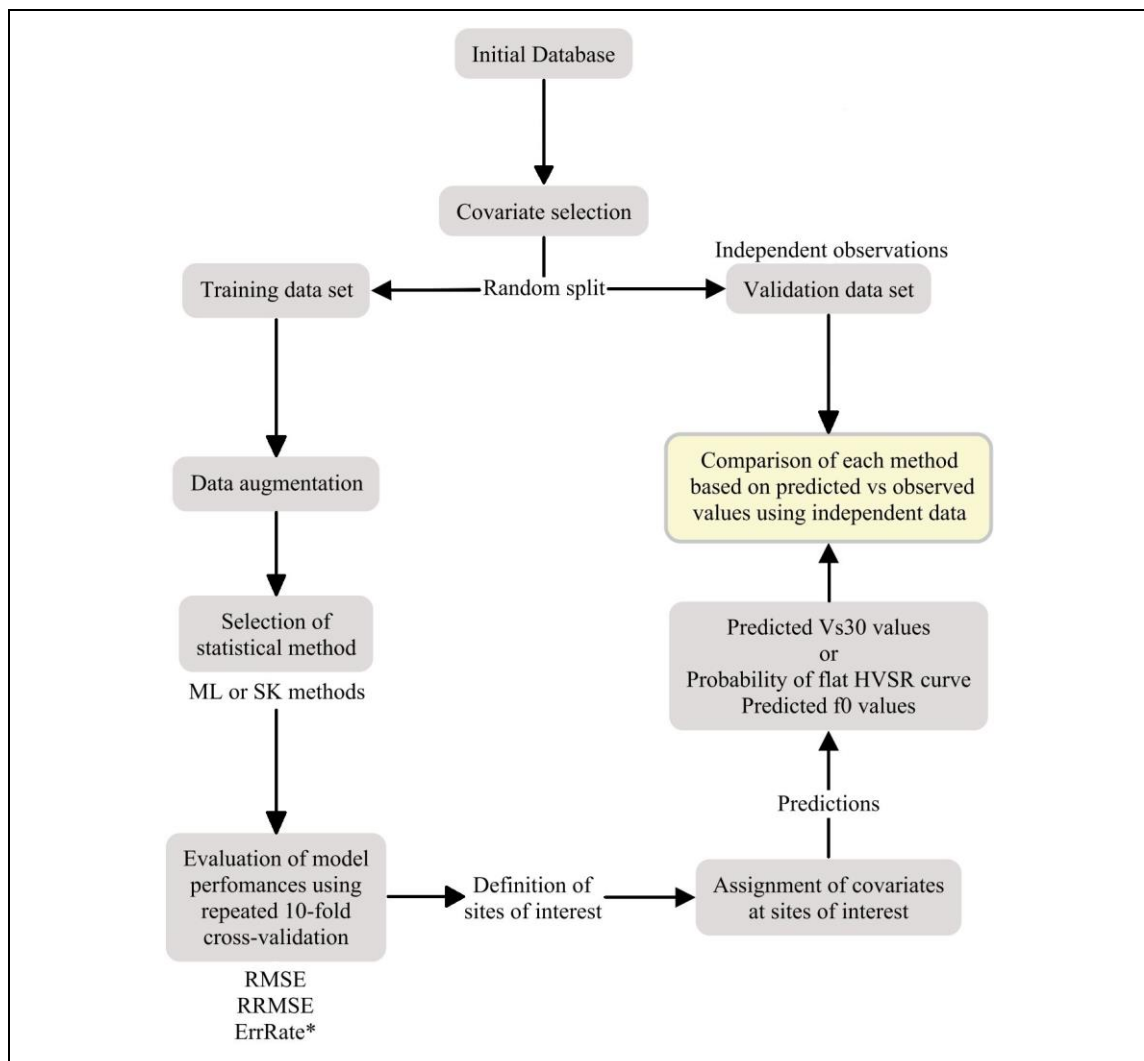

Figure 4-5: Flowchart showing the main steps of the modeling process of V_{s30} and f_0 used in this study. *ErrRate is only calculated in the modeling of f_0.

4.2.5.4   Probabilistic hazard assessment

A PGA map consistent with a return period of 475 years was developed in the PSHA software *Seismic-Hazard* (Candia et al., 2019). This software allows us to compute hazard-consistent ground motion parameters (e.g., PGA, PSA) at a single site or distributed over a large region, using state-of-art seismicity models and rigorous account of scientific uncertainties. In the current study we adopted the Poulos et al. (2019) source model for subduction earthquakes, which uses the Slab 1.0 model (Hayes et al., 2012) to account for the contact surface between the Nazca and South American plates. Additionally, the seismicity model includes four crustal faults reported in the GEM global Active Faults catalog (Styron, 2020) located within a 200 km radius of the study area and the Diablo Fault (also known as Baños Morales Fault) located towards the east boundary of the basin. A logic tree of 3 ground motion models was defined, giving greater weight to the Montalva et al. (2017) model, as it collects local knowledge and incorporates into its development the mega earthquakes (Mw> 8.0) that occurred in Chile in the period 2010-2017, and use V_{s30} in its calculations to estimate PGA. On the other hand, the seismicity from crust sources was modeled with the PCEnga attenuation law (Macedo & Candia, 2020). Three PGA maps are presented to illustrate the influence of different V_{s30} realizations in the PGA distribution. Finally, to discuss PGA changes due to uncertainty of the predictive models, a sensitivity analysis is performed on 30 observed and estimated V_{s30} values not used for training.

4.3   Results

4.3.1   Geophysical survey

This chapter presents the results of the dynamic characterization and seismic classification of the sites of the Santiago basin. The combination of the parameters V_{s30} and f_0 in the

sampled sites allows us to assign an index of seismic susceptibility to each site, as indicated in Table 4-1. The seismically classified sites are shown in Figure 4-6. Sites rated A are generally rock outcrops, with very high V_{s30} (> 900 m/s) and flat HVSR curves (without a clear peak). The A_0 value can be considered as an indicator of the predominant impedance contrast of the site (Pilz et al., 2010; Leyton et al., 2013). It should be noted that there were only a few in rock sites (A), because of the challenges involved in accessing remote areas with flat rock outcrops, required to deploy large arrays of sensors (~100 m long) carrying geophysical equipment. Sites rated B show high values of V_{s30} (exceeding 500 m/s) in the south, center, and northeast of Santiago, where the soils correspond mainly to alluvial fans and fluvial gravels. These sites also have mostly flat HVSR curves. C sites correspond typically to alluvial fans composed by gravels with a higher content of fines and sandy sites. These sites are located mainly nearby the Main Cordillera and the Mapocho river, to the east and southwest of the study area, respectively. In these sectors, the HVSR curves are also mostly flat, showing that in general there are no predominant frequencies or clearly defined impedance contrasts. The sites more prone to seismic amplification are located in the northwestern side of the study area, where V_{s30} tends to be less than 350 m/s, f_0 show low values (< 1 Hz), and large values of A_0 are observed. These sites, classified as D and E, are composed by fine-grain, sedimentary deposits.

4.3.2 Seismic units zoning map

The joint analysis of the site classification with the collected maps of seismic response and gravity model, resulted in an updated seismic microzoning of the Santiago basin (Figure 4-6). The refined microzoning shows that the units with the best seismic response (A) correspond mainly to rock units that surround the sedimentary valley of Santiago and to some isolated hills within the valley. The sites with the best seismic response (B and C) are in the south and east of the study area, while the sites with the worst seismic response (D and E) are located in the center and northwest of the study area. These results

are clearly consistent with previous seismic response maps available in the Santiago basin (i.e. von Igel et al., 2004; Leyton et al., 2011).
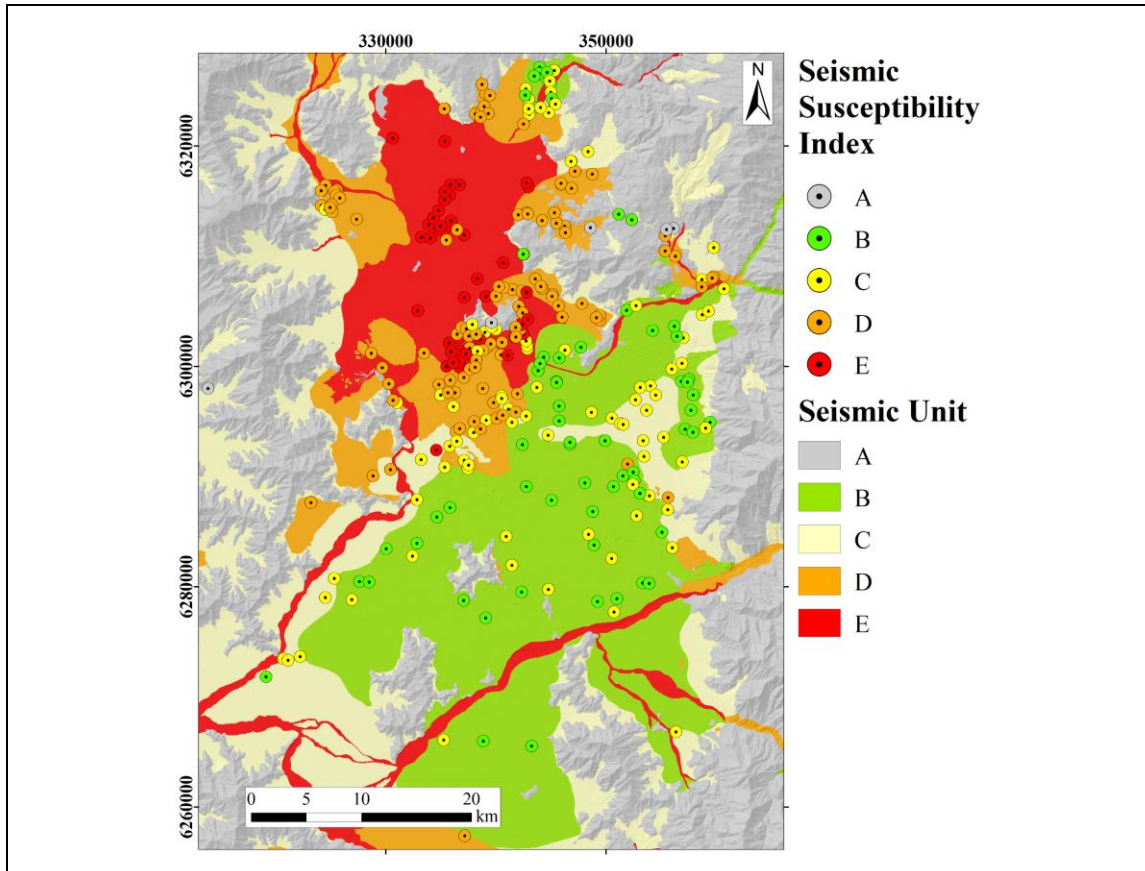

Figure 4-6: Seismic susceptibility indexes and microzoning of Santiago de Chile basin.

4.3.3 Predictive models comparison and resulting maps

This chapter shows the prediction performances of the 6 predictive algorithms of V_{s30} and f_0, shown in Tables 4-4 and 4-5. These results were obtained from the test sets defined for cross-validation, as explained in Chapter 4.2.5.3.

In the prediction of V_{s30}, LR is the best performing algorithm in those sites where the gravimetric covariant is available, with an RMSE of 68.4 m/s and an RRMSE of 17.6%, followed by RF and DT. Similarly, when the gravimetric covariant is not available, the best performing algorithm was LR, with an RMSE of 70.5 m/s and an RRMSE of 17.8%, followed by RF and DT.

In the prediction of f_0, RF is the best performing algorithm in those sites where the gravimetric covariant is available, with an RRMSE of 45.6%, an RMSE of 0.13 Hz and an ErrRate of 21.3%, followed by EN and LR (ANN it is discarded as having too large an ErrRate). On the other hand, when the gravimetric covariant was unknown, the best performing algorithm was LR, with an RRMSE of 164.8%, an RMSE of 2.43 Hz and an ErrRate of 22.3%, followed by RF and SK. It can be noted that if the gravimetry is available, the estimation error is considerably reduced.

Table 4-4: Cross-validation of the best models for each algorithm for the V_{s30} predictions.

Algorithm	Gravimetric covariant included		Gravimetric covariant not included	
	RMSE (m/s)	RRMSE (%)	RMSE (m/s)	RRMSE (%)
SK	185.6	37.5	233.4	41.1
LR	**68.4**	**17.6**	**70.5**	**17.8**
EN	124.6	27.5	141.9	30.0
RF	79.6	19.4	86.3	20.4
ANN	208.7	39.7	234.5	46.5
DT	98.4	22.9	116.7	26.7

Note: The bolds show the best performances obtained among the algorithms.

Table 4-5: Cross-validation of the best models for each algorithm for the f_0 predictions.

Algorithm	Gravimetric covariant included			Gravimetric covariant not included		
	RMSE (Hz)	RRMSE (%)	ErrRate (%)	RMSE (Hz)	RRMSE (%)	ErrRate (%)
SK	0.155	54.8	28.7	2.917	217.9	29.1
LR	0.132	50.8	**20.8**	**2.432**	**164.8**	22.3
EN	**0.111**	47.2	32.1	2.938	224.1	32.2
RF	0.130	**45.6**	21.3	2.502	183.1	**20.9**
ANN	0.135	48.7	41.3	3.716	214.3	45.5
DT	0.162	58.0	27.0	3.976	423.0	31.3

Note: The bolds show the best performances obtained among the algorithms.

Figures 4-7 and 4-8 show the distribution of V_{s30} and f_0 in the study area predicted by the best resulting models, respectively. In the case of f_0 prediction, only the area where gravity modeling is available is shown because the error outside of this zone is too high. The distribution of the values of both parameters is consistent with the observations shown in Chapter 4.2.2. Once the dynamic characterization of sites for the entire study area are available, it is possible to proceed with the Seismic Hazard assessment to obtain the PGA map. Figure 4-9 shows the estimate of PGA in the study area for the settings described in Chapter 4.2.5.4.

4.3.4 PGA sensitivity

To measure the uncertainty associated with the prediction of V_{s30} with the three best predictive algorithms, we randomly chose 30 sites where this parameter was measured (Figure 4-10). None of them were used in the training of the models. In general, the predictions of V_{s30} show a good fit for V_{s30} less than 500 m/s. Above these values, the three models tend to underestimate V_{s30}. This underestimation of V_{s30} for rigid sites is reflected in an overestimation of PGA for a design scenario when PGA is calculated with the predictive algorithms. This overestimation grows when the sites become more rigid, reaching PGA values of about 15% higher than those calculated from the measured V_{s30}.

The main reason is that the number of sites used to train the model under 500 m/s is much larger (96% of the database) than the data available over this value of V_{s30}.

Table 4-6 shows the prediction error indicators for the set of 30 randomly chosen test sites. The results corroborate, using independent data, that the best predictive model to estimate V_{s30} was LR, followed by RF and DT.

Table 4-6: V_{s30} prediction error indicators using independent data.

Algorithm	$RMSE\left(\frac{m}{s}\right)$	$RRMSE$ (%)
LR	79.5	14.5
RF	82.9	15.4
DT	124.9	22.4

4.4   Discussion

4.4.1   Seismic microzoning

The results shown in Figure 4-6 were obtained from an integrated approach that uses geology, geophysics, and earthquake geotechnical engineering information, combining geophysical characterization of sites with seismic response maps. The result was a refined seismic microzoning that considers site effects on the seismic response of the soils in the study area (Figure 4-6). Seismically classified sites and the distribution of high gravimetric anomalies in soft soils helped to redefine the boundaries of seismic units, previously drawn only based on inferred geological information. In sites well covered by dynamic characterizations, we attached greater importance to quantitative information, while in poorly sampled sites, we attached greater importance to geological information.

Some changes are observed in the seismic units with respect to prior maps, mainly in the central zone of the study area (see von Igel et al., 2004 and Leyton et al., 2011 references). Nevertheless, despite the seismic unit updates, the refined seismic microzoning shows consistency with previous maps. The great difference of the zoning presented in this research is the combination of qualitative (i.e. geology) with a large amount of quantitative information obtained using geophysical techniques.

4.4.2 Predictions of V_{s30} and f_0

Regarding the prediction of V_{s30}, for the sites where the gravimetric covariant is available, the three best models had similar performances. Between the first and third best model, the difference of RMSE and RRMSE is only about 30 m/s and 5 %, respectively. The rest of the models show larger errors. Nevertheless, the best performance was obtained with LR, followed by RF and DT. For those sites where gravity data were not available, the models have more dispersion. Between the first and third best model, the difference of RMSE and RRMSE is about 46 m/s and 8.9 %, respectively. Apparently, the V_{s30} predictions are sensitive to the chosen algorithm, however ML algorithms continue to outperform traditional SK.

These results show that the use of ML algorithms to predict V_{s30} provide reliable approximations with reasonable uncertainty, improving the capabilities of the SK geostatistical algorithm.
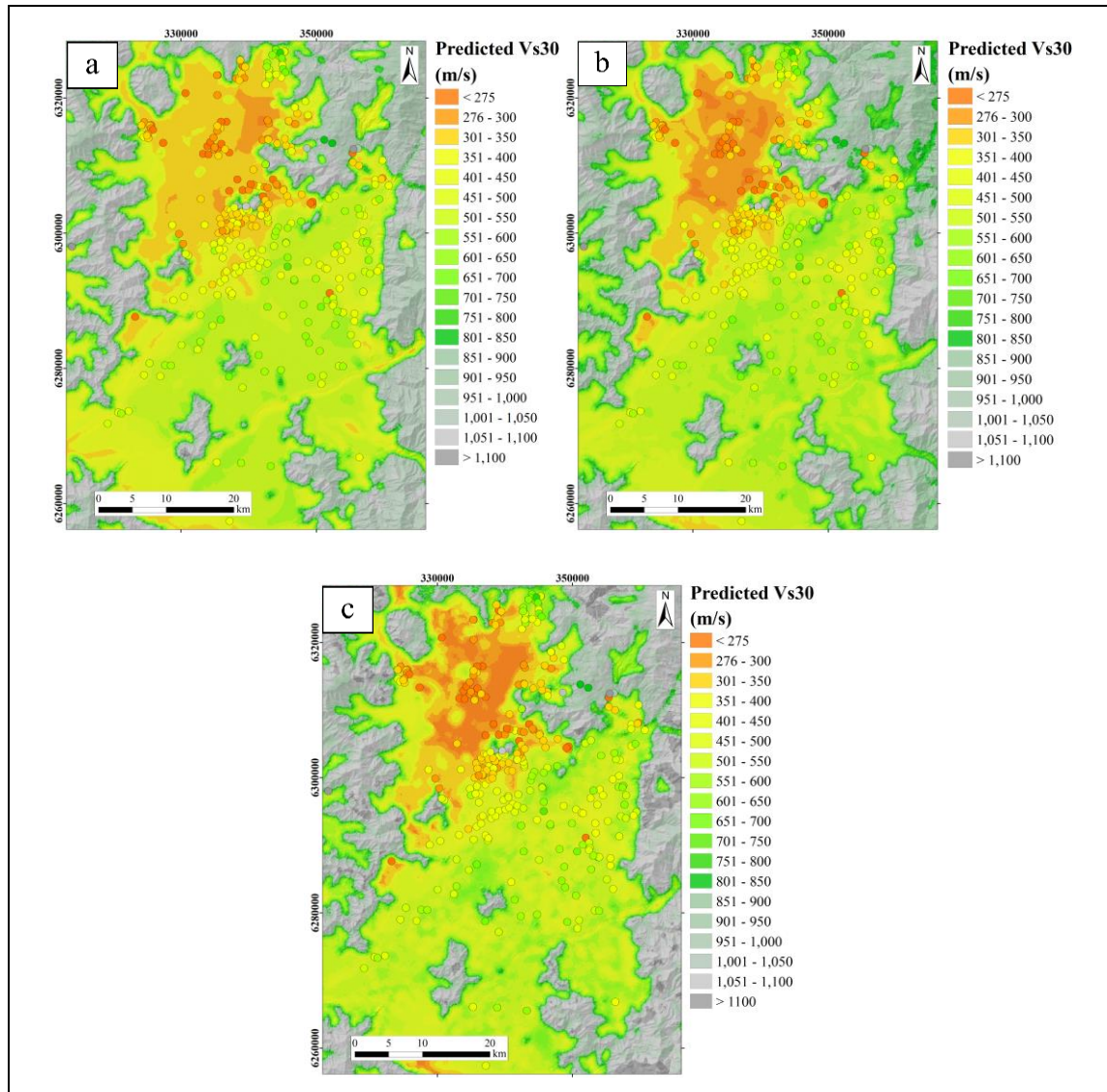
Figure 4-7: V$_{s30}$ prediction maps using a) LR model, b) best RF and c) DT model. Circles show the observed V$_{s30}$ in the same color scale.
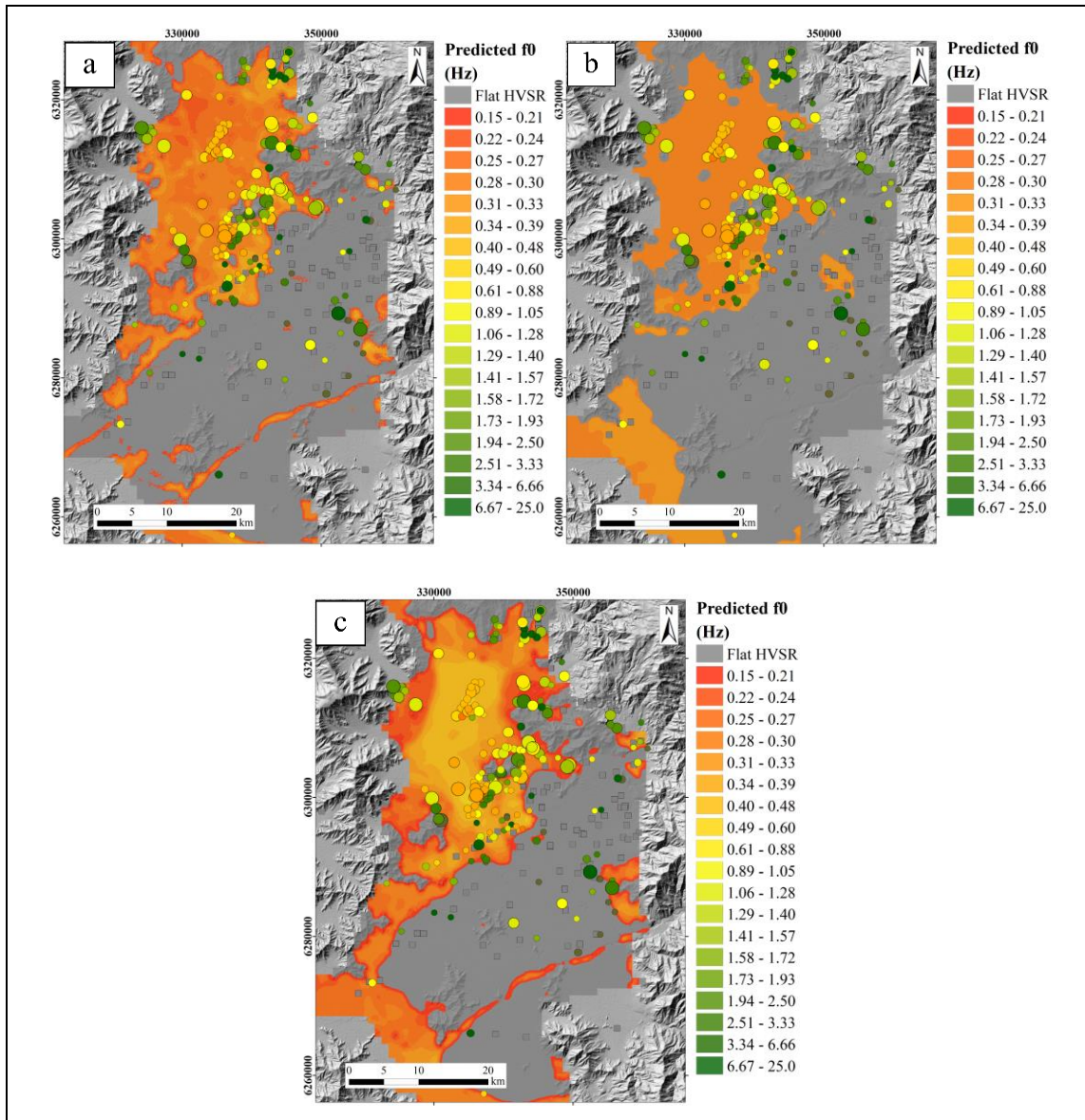
Figure 4-8: f_0 prediction maps using a) best RF model, b) best EN model and c) LR model. Circles show the observed f_0 in the same color scale and their sizes are proportional to A_0.
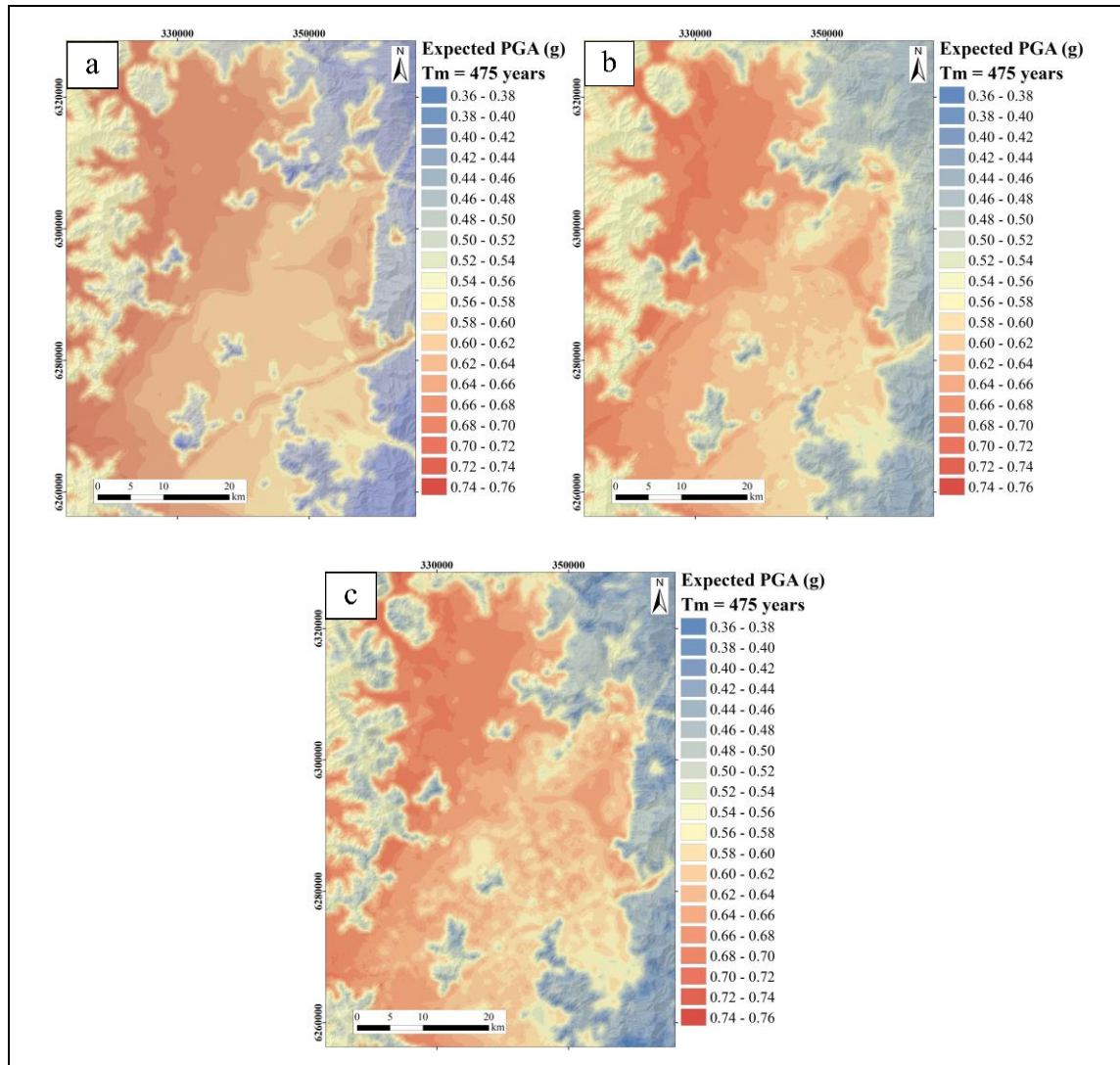
Figure 4-9: Expected PGA map for events with $T_m = 475$ years, based on the estimates of V_{s30} resulting from the a) LR model, b) best RF model, and c) DT model.
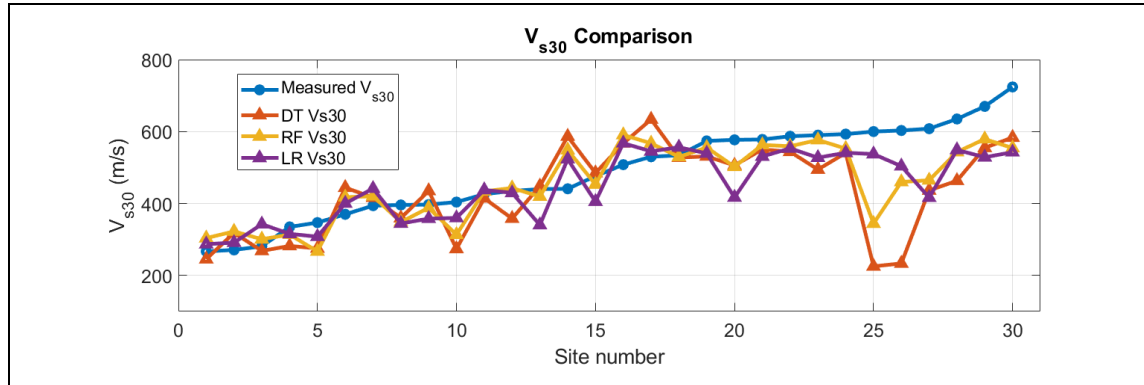
Figure 4-10: Uncertainty associated with the predictive models of V_{s30}. Results based on measured V_{s30} are shown in blue.

Regarding the prediction of f_0, for those sites within the boundaries of the gravity model, the algorithms provide quite stable results. Between the best and worst model, the difference of RMSE and RRMSE is only about 0.05 Hz and 12%, respectively. Similarly to the V_{s30} estimations, ML algorithms outperform SK. The best performance was obtained with RF, followed by EN and LR. Nevertheless, the RRMSE values are high (about 46% in the best case). This would be explained by the previous step of prediction of the peak of the HVSR curve. This previous step inherently increases the errors in the final prediction of f_0, since there are sites where the real HVSR curve is flat, but the algorithm is not able to identify this situation and, erroneously, provides a numeric estimate of f_0. The predictions of f_0 were highly influenced by this initial step since it was made using regression algorithms and not classification algorithms, which discriminate between finite categories or classes. The use of classification algorithms is outside the scope of this article, but it could be a good opportunity to improve predictions in future work. Unlike the case of V_{s30}, for those sites where gravity information is not available, the predictions of f_0 fail, displaying RRMSE values higher than 100% in all models. Thus, these results are not reliable and are considered unacceptable.

The good performance of the ML algorithm is explained by several reasons: a good density of samples in the study area, which allowed properly characterizing most of the

types of sites; a correct choice of the covariates used for training, because original covariates were developed focused on including information on the spatial distribution of the data, capturing correlations between geological, geophysical and geotechnical information. The use of the Data Augmentation technique allowed expanding the original database, avoiding overfitting the models and training them to achieve reasonable predictions in complex scenarios. Another possible reason is that some ML models (e.g. RF and DT) are not limited to using only linear combinations of the observations, and can model the nonlinearity between the target variable and the covariates (Appelhans et al., 2015), i.e. inverse distances used in training probably play a non-linear role. Unexpectedly, it was observed that the LR model performed slightly better than the rest, probably because a large number of covariates were used for the training (44 and 45), which would facilitate the prediction as a linear combination of the covariates.

It is also interesting to note that the improvement in predictions when including the gravimetric covariant is substantially greater when predicting f_0 than when predicting V_{s30}, even though the predictions with and without gravity were not made in the same sites. In the best models for predicting V_{s30}, RMSE and RRMSE decreased from 70.5 m/s to 68.4 m/s and from 17.8% to 17.6%, respectively. While, when predicting f_0, RMSE and RRMSE decreased from 2.5 Hz to 0.13 Hz and from 183.1% to 45.6%, respectively. This would be explained because gravimetric residual has a much closer correlation with f_0 than with V_{s30} since it provides an indicator of sediment thickness which often coincides with the depth at which the predominant impedance contrast is located (Maringue et al., 2021). These results suggest that for the same area of interest, including a gravimetric covariate considerably improves the predictions of f_0, and that the predictive capacity of f_0 depends more on the considered covariates than on the algorithms used.

Figure 4-7 shows the fit of the predictions to the V_{s30} observations. Quite similar predictions are observed among the best 3 predictive models. All 3 models were able to correctly distinguish the rock units and predict V_{s30} within the basin. In the northwest area, LR shows to be the least conservative predictor, followed by RF, while DT shows the

lowest values of V_{s30}. In the central and southern sector of the basin, LR is also less conservative than RF and DT. These differences are reflected in the PGA maps (Figure 4-9), where the highest expected PGA values are located within the Santiago sedimentary basin and are associated with the most conservative prediction of V_{s30} values (DT model), and the lowest expected PGA values are associated with the model that predicted the highest V_{s30} values (LR model). The rock units present the lowest expected PGA values, which is consistent with the V_{s30} values observed in these units. In general, the rock units to the east of the study area show higher expected PGA values than those located to the east, because they are closer to the main seismogenic source (subduction zone).

The predicted f_0 maps are quite similar. They differ mainly in their ability to predict where the HVSR curve is flat and in the values of f_0 when the HVSR curve has a clear peak. RF is the most accurate, since it identifies the rock units and the most rigid soils in the study area reasonably well, assigning a flat HVSR curve. It is also the one that best fits the f_0 values observed in rigid and soft soils. EN adequately identifies rock units but was only able to predict a narrow range of f_0, resulting in an almost bimodal map. LR also identifies rock units correctly and fits the observations well but is more conservative in the southwestern zone of the study area, delivering low predominant frequencies (deep sites) where soils are known to be rigid. Among the models, RF is the one that best fits the sites with non-flat HVSR curves, correctly identifying the sites classified as D and E, delivering the lowest observed values of f_0 and therefore more prone to seismic amplifications. In general, all models were only able to predict a narrow range of f_0 (0.2 to 0.40 Hz). This is probably due to the previous step that defines the shape of the HVSR curve and to the observed range of f_0 values, which is mostly at frequencies less than 2 Hz (see Figure 4-3b).

The sensitivity analysis performed in Chapter 4.3.4 shows that the estimates of V_{s30} are better for values of V_{s30} less than 500 m/s, and they get worse when this limit is exceeded. This is because the database contains less information on stiffer sites compared to softer

sites, making it difficult for algorithms to predict the V_{s30} value of the most rigid sites. Despite this, the error in the PGA calculation associated with the estimation error of V_{s30} is small for rigid sites, reaching maximum values of only 15%.

Finally, the results of our work show that: (i) it is possible to generate a refined seismic microzoning in the Santiago basin incorporating quantitative and qualitative information that allows evaluating site effects on soils. And (ii) it is possible to obtain a reasonably good prediction of the dynamic properties of the sites of the Santiago basin using ML predictive algorithms, surpassing the capabilities of traditional geostatistical predictive models.

4.4.3 Extensions and improvements

The main limitation in the generation of the refined seismic microzoning is related to a very heterogeneous distribution of data from geophysical techniques. This caused that there were very wide areas in which only geological information was known, losing the opportunity to combine qualitative with quantitative information. This situation was mainly observed away from urban areas. Therefore, for future stages of this research, performing dynamic characterizations in the poorly sampled area is recommended, where the uncertainty of dynamic site properties is greater.

The satisfactory results obtained suggest that this methodology could be replicated in other regions of Chile or the world, combining the dynamic properties of sites with information from surface geology, other geophysical techniques and digital elevation models to improve the accuracy of qualitative seismic response maps.

Regarding the predictions of V_{s30} and f_0, the main limitations were: (i) the few measurements available in rock, which made it difficult to train the models in this type of

sites, (ii) the limited spatial extend of gravimetric data, and (iii) the difficulty of predicting the existence or not of a peak in the HVSR curve, increasing the errors in the prediction of f_0. The best estimate of f_0 is available as a tool to classify the unsampled sites of the study area, and the methodology to obtain these estimates remains available to be used in future GMPE that include f_0 in their seismic hazard estimates.

Therefore, for future research it is recommended to have more measurements in rock, at least in an amount comparable to observations in other types of soil. It will also be very useful to design a gravimetric experiment that covers the entire area of interest, since it has been shown that it correlates very well with f_0. It would also be useful to study the performance of ML classification algorithms to decide whether the HVSR curves are flat or not, since better results will probably be obtained than using only regression algorithms. The importance of choosing and testing the training covariates was confirmed, since best predictions are obtained when there is a strong correlation between the covariates and the objective variable.

It is important to mention that there should be other combinations of covariates that further improve the estimates of V_{s30} or f_0. This implies the possibility of removing or adding new covariates to the training database of the predictive models. Exploring new combinations of covariates may be necessary for geological contexts other than the sedimentary basins of Santiago, for example, to generate a predictive model on a much larger spatial scale.

Finally, the distribution of f_0 presented in this article was not directly used to assess the seismic hazard in the study area, because there is not a sufficiently validated GMPE applicable to Chile that includes this parameter in its calculations. However, GMPE that include f_0 have already been developed (e.g. Kwak et al., 2020), using a two-stage nonlinear site amplification model derived empirically from records of strong earthquakes in Japan. Those models show that the residual, associated with GMPE that only include V_{s30} as a site parameter, decreases considerably when including the observed values of f_0

at the sites, mainly for spectral periods > 0.1 s. In this way, the use of f_0 could strengthen seismic hazard estimates.

4.5   Conclusions

This work presents two results: (i) a refined seismic microzoning that provides a qualitative estimate of the seismic hazard in the Santiago basin, and (ii) a methodology that uses ML computational tools to estimate dynamic properties of the soils in the unsampled sectors of the Santiago basin, allowing the assessment of site effects and quantitatively estimate the seismic hazard in terms of PGA. A rationale is presented to generate these estimates for both V_{s30} and the predominant period f_0.

The integration of qualitative information with quantitative data based on geophysical exploration made it possible to generate an update of the existing seismic microzoning map for the study area. This integration of information also allowed the generation of more complete predictive models of dynamic properties of sites integrating both types of sources of information.

Regarding the predictive algorithms of V_{s30} and f_0, the following can be concluded:

- Five ML algorithms (LR, EN, RF and DT) were compared with a traditional geostatistical algorithm (SK). For predicting V_{s30}, the most robust algorithm was LR, followed by RF and DT. For predicting f_0, the best algorithm was RF, followed by EN and LR.

- The results of all models were verified by cross-validation, obtaining a RMSE in the best prediction of V_{s30} and f_0 of 68.4. m/s and 0.13 Hz, respectively, and a RRMSE in the best prediction of V_{s30} and f_0 of 17.6% and 45.6%, respectively.

The spatial distribution of estimated V_{s30} and f_0 is consistent with the available observations.

- The improvement in the estimates of V_{s30} and f_0 by ML algorithms are explained by the inclusion of spatial covariates for algorithm training, helping the techniques capture the spatial correlations of geological, geophysical and geotechnical data. Similar results are well documented in related literature.

- By including the gravimetric residual covariate in the training of the predictive models, a significant improvement was observed in the prediction of f_0, which suggests that both parameters have a strong correlation in sedimentary contexts.

- The predictive capacity of f_0 apparently depends more on the choice of covariates than on the algorithm used, while the V_{s30} predictions are more sensitive to the chosen algorithm.

ML algorithms have shown to be promising tools in the prediction of dynamic properties of sites. Future work should be focused on increasing the database, exploring which combination of covariates gives better predictions in more general geological contexts, testing ML classification tools to reduce uncertainty when estimating f_0, and including f_0 in the estimation of the seismic hazard through GMPE that include this parameter.

5   CONCLUSIONS AND FUTURE WORK

This work aimed to develop a methodology that combines qualitative information from geology, with quantitative data from geophysics and a digital elevation model through computational techniques of machine learning to develop a predictive model of dynamic properties of sites. The final goal is to achieve a qualitative estimation of the seismic susceptibility and a quantitative assessment of the seismic hazard in the Maipo River Basin (MRB).

A refined seismic susceptibility map was generated based on previous geological and seismic response maps, combining this information with the dynamic characterization of 365 sites in the study area in terms of their surface shear wave velocities (V_{s30}) and their predominant frequencies (f_0). Gravimetric information available in the study area was also incorporated. The integrated analysis of this information allowed to identify zones in which different levels of dynamic amplification are expected due to seismic action.

In a second stage, different predictive algorithms were trained to estimate V_{s30} and f_0 in the area with gravimetry available. The algorithms were trained using local and spatial covariates related to the observed data, supported in the susceptibility map generated in this study. Five predictive machine learning algorithms were tested: artificial neural networks (ANN), random forests (RF), linear regression (LR), decision trees (DT), and elastic net (EN). Additionally, a simple kriging (SK) algorithm was tested as the only geostatistical method.

Independent models were developed for two domains: within the Santiago metropolitan region (SMR) and in the complete MRB.

Regarding the predictive algorithms of V_{s30} and f_0 within the SMR, the following can be concluded:

- For predicting V_{s30}, the most robust algorithm was LR, followed by RF and DT.

- For predicting f_0, the best algorithm was RF, followed by EN and LR.

- The results of all models were verified by cross-validation, obtaining a RMSE in the best prediction of V_{s30} and f_0 of 68.4. m/s and 0.13 Hz, respectively, and a RRMSE in the best prediction of V_{s30} and f_0 of 17.6% and 45.6%, respectively. The spatial distribution of estimated V_{s30} and f_0 is consistent with the available observations.

- The improvement in the estimates of V_{s30} and f_0 by ML algorithms would be explained by the inclusion of spatial covariates for training, helping the techniques to capture the spatial correlations of geological, geophysical and geotechnical data. Similar results are well documented in related literature.

- By including the gravimetric residual covariate in the training, a significant improvement was observed in the prediction of f_0, which suggests that both parameters have a strong correlation in sedimentary contexts.

- The precision of the f_0 predictions apparently depends more on the choice of covariates than on the algorithm used, while the V_{s30} predictions are more sensitive to the chosen algorithm.

Regarding the predictive algorithms of V_{s30} and f_0 within the Maipo River Basin, the following can be concluded:

- For predicting V_{s30}, the most robust algorithm was LR, followed by RF and DT.

- Acceptable results for f_0 were only obtained at sites where gravimetry is available (SRM).

- The results of all models were verified by cross-validation, obtaining a RMSE in the best prediction of V_{s30} of 69.3 m/s, and a RRMSE of 17.5%. The spatial distribution of estimated V_{s30} is consistent with the available observations.

The hypothesis of this work was corroborated since notable improvements were obtained in the qualitative estimation of seismic susceptibility and in the quantitative evaluation of the seismic hazard through the integration of information from different sources and the use of machine learning tools. Furthermore, very good estimates of V_{s30} were achieved in the study area. Nevertheless, it was not possible to achieve a reasonable estimate of f_0 in the entire study area due to the limited spatial extent of the gravimetric data, which proved to be key to estimating f_0 accurately. Furthermore, the prediction of f_0 was more complex than V_{s30} due to the different ways of interpreting the shapes of the available HVSR curves.

Recommendations for future works include:

- Perform dynamic characterizations in the poorly sampled area, where the uncertainty of dynamical site properties is greater.

- Include gravimetric measurements that cover a greater extension of the study area: it was found that this covariant improves the predictive capacity of ML algorithms, especially when predicting f_0.

- Study new combinations of training covariates: although good results were obtained with the covariates proposed in this work, there might be new

combinations of attributes (local and/or spatial) that further improve the performance of predictive algorithms.

- Include f_0 in seismic hazard estimating algorithms: as has been proven in other parts of the world, including f_0 in ground motion prediction equations (GMPE) could reduce the uncertainty in estimating the seismic hazard.

- Explore predictive algorithms not necessarily scalar for site characterization, but also vector ones, i.e. the dispersion and/or the HVSR curves instead of V_{s30} and f_0.

- Study the performance of predictive classification algorithms: although classification algorithms are beyond the scope of this study, they could possibly achieve better results to predict the existence or not of flat HVSR curves, improving the final estimate of f_0.

- Extend the application of these results to perform vulnerability analyzes.

Finally, the results obtained in this study suggest that the proposed methodology could be replicated in other sedimentary basins of Chile or worldwide.

REFERENCES

Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute, 35*, 415-456.

Appelhans, T., Mwangomo, E., Hardy, D. R., Hemp, A., & Nauss, T. (2015). Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. *Spatial Statistics*, *14*, 91-113.

Araneda, M., Avendaño, M. & Merlo, C. (2000). Modelo gravimétrico de la cuenca de Santiago con el ensayo de Refracción Sísmica. *Proc. 5to Congreso de Ingeniería Geotécnica*, Universidad de Chile, Santiago, Chile.

Augusto Filho, O., Soares, W., & Irigaray, C. (2017). Mapping of compactness by depth in a quaternary geological formation using deterministic and geostatistical interpolation models. *Environmental Earth Sciences, 76*(17), 1-12.

Becerra, A., Podestá, L., Monetta, R., Sáez, E., Leyton, F., & Yanez, G. (2015). Seismic microzoning of Arica and Iquique, Chile. *Natural Hazards, 79*(1), 567-586.

Beguin, J., Fuglstad, G. A., Mansuy, N., & Paré, D. (2017). Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma, 306*, 195-205.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E. D., & Goldschmitt, M. (2005). Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science, 168*(1), 21-33.

Bonnefoy-Claudet, S., Baize, S., Bonilla, L. F., Berge-Thierry, C., Pasten, C., Campos, J., ... & Verdugo, R. (2009). Site effect evaluation in the basin of Santiago de Chile using ambient noise measurements. *Geophysical Journal International*, 176(3), 925-937.

Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2*(6), 493-507.

Bravo, F., Koch, P., Riquelme, S., Fuentes, M., & Campos, J. (2019). Slip Distribution of the 1985 Valparaíso Earthquake Constrained with Seismic and Deformation Data. *Seismological Research Letters, 90*(5), 1792-1800.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. (1st ed.). Chapman & Hall/Crc.

Burrough, P. A., & McDonnell, R. A. (1998). Principles of geographical information systems: Oxford University Press. *New York*, 190.

Candia, G., Macedo, J., Jaimes, M. A., & Magna-Verdugo, C. (2019). A new state-of-the-art platform for probabilistic and deterministic seismic hazard assessment. *Seismological Research Letters, 90*(6), 2262-2275.

Chávez-García FJ, Cuenca J. (1998). Site effects and microzonation in Acapulco. *Earthquake Spectra 14*(1):75–93.

Chávez-García, F. J., Rodríguez, M., & Stephenson, W. R. (2005). An alternative approach to the SPAC analysis of microtremors: exploiting stationarity of noise. *Bulletin of the Seismological Society of America, 95*(1), 277-293.

Chile. Servicio Nacional de Geología y Minería, & Wall, R. (1996). *Mapa geológico del área de San Antonio-Melipilla.*

Cifuentes, I. (1989). The 1960 Chilean earthquakes. *Journal of Geophysical Research: Solid Earth, 94*(B1), 665-680.

Clifton, C. (2010). Encyclopedia Britannica: Definition of data mining. *Retrieved on*, 9(12), 2010.

Deng, L., Adjouadi, M., & Rishe, N. (2020, October). Inverse Distance Weighted Random Forests: Modeling Unevenly Distributed Non-Stationary Geographic Data. In *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 41-46). IEEE.

Fernández, J. (2003). Respuesta sísmica de la cuenca de Santiago. Servicio Nacional de Geología y Minería. *Carta Geológica de Chile*, Serie Geología Ambiental N°1.

Fock, A. (2005). *Cronología y tectónica de la exhumación en el Neógeno de los Andes en Chile central entre los 33º y 34ºS.* (Unpublished master´s thesis). Universidad de Chile, Santiago, Chile.

Foti, S. (2000). *Multistation Methods for Geotechnical Characterization using Surface Waves.* (Unpublished doctoral thesis). Politecnico di Torino, Torino, Italy.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1.

Hayashi, K. (2008). *Development of the Surface-wave Methods and Its Application to Site Investigations. Report*. Kyoto University. Retrieved from http://repository.kulib.kyoto-u.ac.jp/dspace/handle/2433/57255

Hayes, G. P., Wald, D. J., & Johnson, R. L. (2012). Slab1. 0: A three-dimensional model of global subduction zone geometries. *Journal of Geophysical Research: Solid Earth, 117*(B1).

Humire, F. (2013). *Aplicación de Métodos Geofísicos Basados en Ondas Superficiales para la Caracterización Sísmica de suelos. Aplicación a la microzonificación sísmica del Norte y Poniente de Santiago*. (Unpublished master's thesis). Pontificia Universidad Católica de Chile, Santiago, Chile.

Humire, F., Sáez, E., Leyton, F., & Yañez, G. (2015). Combining active and passive multi-channel analysis of surface waves to improve reliability of Vs30 estimation using standard equipment. *Bulletin of Earthquake Engineering, 13*(5), 1303-1321.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. (1st ed.). Sage.

Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling, 81*, 401-418.

Kohestani, V. R., Hassanlourad, M., & Ardakani, A. J. N. H. (2015). Evaluation of liquefaction potential based on CPT data using random forest. *Natural Hazards, 79*(2), 1079-1089.

Kvaerna, T., & Ringdahl, F. (1986). Stability of various fk estimation techniques. *NORSAR Semiannual technical summary, 1*, 1-86.

Kwak, D. Y., & Seyhan, E. (2020). Two-stage nonlinear site amplification modeling for Japan with VS 30 and fundamental frequency dependency. *Earthquake Spectra, 36*(3), 1359-1385.

Lacoss, R. T., Kelly, E. J., & Toksöz, M. N. (1969). Estimation of seismic noise structure using arrays. *Geophysics, 34*(1), 21-38.

Lermo, J., & Chávez-García, F. J. (1993). Site effect evaluation using spectral ratios with only one station. *Bulletin of the seismological society of America*, *83*(5), 1574-1594.

Leyton, F., Ruiz, S., & Sepúlveda, S. A. (2010). Reevaluación del peligro sísmico probabilístico en Chile central. *Andean Geology, 37*(2), 455-472.

Leyton, F., Sepúlveda, S. A., Astroza, M., Rebolledo, S., Acevedo, P., Ruiz, S., ... & Foncea, C. (2011). Seismic zonation of the Santiago basin, Chile. In *5th International Conference on Earthquake Geotechnical Engineering.* Santiago, Chile.

Leyton, F., Ramírez, S., & Vásquez, A. (2012). Uso y limitaciones de la técnica de microvibraciones (RHV) en la clasificación sísmica de suelos. *VII Congreso Chileno de Geotecnia.*

Leyton, F., Ruiz, S., Sepúlveda, S. A., Contreras, J. P., Rebolledo, S., & Astroza, M. (2013). Microtremors' HVSR and its correlation with surface geology and damage observed after the 2010 Maule earthquake (Mw 8.8) at Talca and Curicó, Central Chile. *Engineering geology, 161*, 26-33.

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software, 26*(12), 1647-1659.

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software, 53*, 173-189.

Macedo, J., & Candia, G. (2020). Performance-based assessment of the seismic pseudo-static coefficient used in slope stability analysis. *Soil Dynamics and Earthquake Engineering, 133*, 106109.

Maringue, J., Sáez, E., & Yañez, G. (2021). An Empirical Correlation between the Residual Gravity Anomaly and the H/V Predominant Period in Urban Areas and Its Dependence on Geology in Andean Forearc Basins. *Applied Sciences*, 11(20), 9462.

Marzocchi, W., Garcia-Aristizabal, A., Gasparini, P., Mastellone, M. L., & Di Ruocco, A. (2012). Basic principles of multi-risk assessment: a case study in Italy. *Natural Hazards, 62*(2), 551-573.

McBratney, A. B., Odeh, I. O., Bishop, T. F., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma, 97*(3-4), 293-327.

McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1-2), 3-52.

Mendoza, L. (2021). *Estimación cuantitativa de la amenaza sísmica en base a métodos geofísicos.* (Unpublished master's thesis). Pontificia Universidad Católica de Chile, Santiago, Chile.

Molnar, S., Cassidy, J. F., Castellaro, S., Cornou, C., Crow, H., Hunter, J. A., ... & Yong, A. (2018). Application of microtremor horizontal-to-vertical spectral ratio (MHVSR) analysis for site characterization: State of the art. *Surveys in Geophysics*, *39*(4), 613-631.

Montalva, G. A., Bastías, N., & Rodriguez-Marek, A. (2017). Ground-motion prediction equation for the Chilean subduction zone. *Bulletin of the Seismological Society of America, 107*(2), 901-911.

Moore, I. D., Gessler, P. E., Nielsen, G. A. E., & Peterson, G. A. (1993). Soil attribute prediction using terrain analysis. *Soil science society of America journal, 57*(2), 443-452.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Railway Technical Research Institute, Quarterly Reports, 30*(1), 25-33.

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil, 5*(1), 79-89.

Pastén, C. (2007). Respuesta Sísmica de la Cuenca de Santiago. [Master's thesis, Universidad de Chile]. Repositorio académico de la Universidad de Chile. http://repositorio.uchile.cl/handle/2250/102937

Pilz, M., Parolai, S., Picozzi, M., Wang, R., Leyton, F., Campos, J., & Zschau, J. (2010). Shear wave velocity model of the Santiago de Chile basin derived from ambient noise measurements: a comparison of proxies for seismic site conditions and amplification. *Geophysical Journal International, 182*(1), 355-367.

Podestá, L. F. (2017). Estudio geofísico y modelación de efectos de sitio 3D en la ciudad Viña del Mar, Chile. [Master's thesis, Pontificia Universidad Católica de Chile]. Repositorio UC. https://repositorio.uc.cl/handle/11534/21340

Pokhrel, R. M., Kuwano, J., & Tachibana, S. (2013). A kriging method of interpolation used to map liquefaction potential over alluvial ground. *Engineering Geology, 152*(1), 26-37.

Poulos, A., Monsalve, M., Zamora, N., & de la Llera, J. C. (2019). An Updated Recurrence Model for Chilean Subduction Seismicity and Statistical Validation of Its Poisson Nature. *Bulletin of the Seismological Society of America, 109*(1), 66-74.

Roudier, P., Burge, O. R., Richardson, S. J., McCarthy, J. K., Grealish, G. J., & Ausseil, A. G. (2020). National scale 3D mapping of soil pH using a data augmentation approach. *Remote Sensing, 12*(18), 2872.

Samuel-Rosa, A., Heuvelink, G. B. M., Vasques, G. M., & Anjos, L. H. C. (2015). Do more detailed environmental covariates deliver more accurate soil maps?. *Geoderma, 243*, 214-227.

Scholz, C. (2002). *The Mechanics of Earthquakes and Faulting.* Cambridge, England: Cambridge University Press.

Scott, J. B., Rasmussen, T., Luke, B., Taylor, W. J., Wagoner, J. L., Smith, S. B., & Louie, J. N. (2006). Shallow shear velocity and seismic microzonation of the urban Las Vegas, Nevada, basin. *Bulletin of the Seismological Society of America, 96*(3), 1068-1077.

Scull, P., Franklin, J., Chadwick, O. A., & McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical geography, 27*(2), 171-197.

Sekulić, A., Kilibarda, M., Heuvelink, G., Nikolić, M., & Bajat, B. (2020). Random forest spatial interpolation. *Remote Sensing, 12*(10), 1687.

Sekulić, A., Kilibarda, M., Protić, D., & Bajat, B. (2021). A high-resolution daily gridded meteorological dataset for Serbia made by Random Forest Spatial Interpolation. *Scientific Data, 8*(1), 1-12.

Sellés, D., & Gana, P. (2001). Geología del área Talagante-San Francisco de Mostazal. *Servicio Nacional de Geología y Minería, Carta Geológica de Chile, Serie Geología Básica, 74*, 30.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), 1-48.

Soto, V. (2019). *3D modeling of site-city effects using spectral element method. Application to Viña del Mar city, Chile.* (Unpublished master's thesis). Pontificia Universidad Católica de Chile, Santiago, Chile.

Styron, R., & Pagani, M. (2020). The GEM global active faults database. *Earthquake Spectra, 36*(1_suppl), 160-180.

Tarvainen, T., Jarva, J., & Greiving, S. (2006). Spatial pattern of hazards and hazard interactions in Europe. *SPECIAL PAPER-GEOLOGICAL SURVEY OF FINLAND, 42*, 83.

Thompson, E. M., Baise, L. G., Kayen, R. E., Tanaka, Y., & Tanaka, H. (2010). A geostatistical approach to mapping site response spectral amplifications. *Engineering Geology, 114*(3-4), 330-342.

Thompson, E. M., Wald, D. J., & Worden, C. B. (2014). A Vs30 map for California with geologic and topographic constraints. *Bulletin of the Seismological Society of America, 104*(5), 2313-2321.

Tokimatsu, K. (1997). Geotechnical site characterization using surface waves. In Ishihara (Ed.), *Proc. 1st Intl. Conf. Earthquake Geotechnical Engineering, 3*, pp. 1333–1368. Balkema.

Tuladhar, R., Yamazaki, F., Warnitchai, P., & Saita, J. (2004). Seismic microzonation of the greater Bangkok area using microtremor observations. *Earthquake engineering & structural dynamics, 33*(2), 211-225.

Ugalde, D., López-García, D., & Parra, P. F. (2020). Fragility-based analysis of the influence of effective stiffness of reinforced concrete members in shear wall buildings. *Bulletin of Earthquake Engineering*, *18*(5), 2061-2082.

United Nations. (2015). *Global Assessment Report on Disaster Risk Reduction.* Recovered from: https://reliefweb.int/sites/reliefweb.int/files/resources/GAR15_Pocket_EN.pdf

Verdugo, R., Ochoa-Cornejo, F., González, J., & Valladares, G. (2019). Site effect and site classification in areas with large earthquakes. *Soil Dynamics and Earthquake Engineering*, 126, 105071.

Von Igel, B., Naranjo, J., Wall R. (2004). *Respuesta Sísmica de la Región Metropolitana de Santiago*. (Report IR-04-25). Servicio Nacional de Geología y Minería.

Wald, D. J., & Allen, T. I. (2007). Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America, 97*(5), 1379-1395.

Wills, C. J., & Clahan, K. B. (2006). Developing a map of geologically defined site-condition categories for California. *Bulletin of the Seismological Society of America, 96*(4A), 1483-1501.

Yáñez, G., Muñoz, M., Flores, V., & Bosch, A. (2015). Gravity derived depth to basement in Santiago Basin, Chile: implications for its geological evolution, hydrogeology, low enthalpy geothermal, soil characterization and geo-hazards. *Andean Geology, 42*(2), 147-172.

Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020). Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(07), 13001-13008.

Zuccaro, G., Cacace, F., Spence, R. J. S., & Baxter, P. J. (2008). Impact of explosive eruption scenarios at Vesuvius. *Journal of Volcanology and Geothermal Research, 178*(3), 416-453.

ANNEXES

Annex 1: Field work

This Annex describes the methods and analyzes performed in the geophysical seismic surveys carried out in this study. General criteria used methodologies and example results are described.

Annex 1.1: Surface wave geophysical methods

There are various methods to measure V_{s30}, which can be classified into direct and indirect methods. Direct methods have the advantage of being reliable and accurate, but require one or more borehole at the site of interest, which can often be impractical, expensive, and time-consuming. On the other hand, indirect methods are much more efficient in time and cost since they use the recording of surface waves to characterize the soil. These methods take advantage of the dispersive nature of surface waves (Tokimatsu, 1997), i.e. the velocity of propagation or phase velocity (v) depends on both its frequency (f) and wavelength (λ) and are related as

$$v(f) = \lambda(f)f \tag{7}$$

Therefore, long wavelengths allow exploration of greater depths of the subsoil and are associated with low frequencies, while short wavelengths allow exploration of shallower soil layers and are associated with high frequencies. The graphical representation of the dispersive properties of a soil is the dispersion curve as it shows the relationship between the phase velocity and frequency of surface waves.

There are different approaches to obtain and analyze the scattering of surface waves at a site. The procedure for this characterization can be summarized as follows (Tokimatsu, 1997; Foti, 2000):

- Observation and recording of surface waves

- Determination of the characteristic dispersion curve of the site

- Shear wave velocity profile estimation, based on an inversion process of the dispersion curve

In this study and based on the experience of previous similar works (Humire, 2013; Podestá, 2017; Soto, 2019; Mendoza, 2021), V_{s30} measurements were performed using a multichannel analysis approach. For this, a set of 24 vertical geophones (4.5 Hz natural frequency) was used to record seismic waves from different sources, using linear and 2D arrays. Active seismic sources (hammer blows) of known location and passive seismic sources (ambient noise) of unknown incidence were used. Linear arrays were used for the methods based on active and passive sources, while 2D arrays were intended only for ambient noise recording. The different methods of obtaining seismic waves allow exploring different portions of the characteristic dispersion curve of a site, as shown in Figure 7-1. Therefore, it is necessary to combine the results of the different methods to obtain a reliable dispersion curve that allows explore at least 30 m deep and get V_{s30}.

The approaches used to obtain the dispersion curves in this investigation were the F-K method (frequency-wave number method) (Lacoss et al., 1969; Kvaerna & Ringdahl, 1986) for active 1D and passive 2D arrays, the SPAC method (spatial autocorrelation method) (Aki, 1957) for passive 2D arrays, and the ESPAC method (Chávez-García et al., 2005) (extended spatial autocorrelation method) for passive 1D arrays. These methods are briefly described below. A detailed description of these seismic methods is available in Humire (2015).

- Spectral analysis F-K: This method assumes that the receiver array is crossed by a p wavefront of known frequency, speed, and direction of propagation. The signals are delayed according to the geometry of the array and under a hypothesis of direction of incidence of waves, thus, the arrival times to the receivers have a

common reference and the signals can be combined. Using appropriate integral transform, the information is analyzed in the domain of frequencies and wave number, allowing the construction of an amplitude spectrum based on records of the studied array. Finally, from the spectrum obtained, the dispersion curve is constructed for each combination of frequency and phase velocity. This method can be applied to both active and passive tests (Tokimatsu, 1997).

- SPAC: The spatial autocorrelation method was proposed by Aki (1957) and assumes that the field of waves related to ambient vibrations is a stochastic process, stationary in time and space, and it is mainly composed of surface waves. Using micro-tremors measurements, it is possible to calculate the spatial autocorrelation function between each pair of sensors of the array. From these functions an autocorrelation coefficient is obtained for different frequency values to define the autocorrelation curve. These autocorrelation curves are used to construct the dispersion curve associated with the SPAC method. This method requires two-dimensional arrangements, and its main advantage is that it uses the same information recorded in a passive F-K test, hence it does not require any extra effort in field.

- ESPAC: It is a method derived from the SPAC method applied to linear arrays. It assumes that the wave field is stationary in time and that there is no predominant direction of propagation. This method can be very useful if there no space available to deploy two-dimensional arrays.

Figure 7-1 shows an example of the combination of dispersion curves from the different analysis methods used in this study, for a sample site in Colina district. The dispersion curves from the different methods are combined to solve a global inversion process using the tools available in Geopsy® package. This process uses the neighborhood algorithm, proposed by Wathelet (2008), which adjusts velocity profiles of 1D shear waves

associated with theoretical dispersion curves matching to those observed. The algorithm searches the one that minimizes the misfit or mismatch between the theoretical and empirical dispersion curves given an initial parametrization. Figure 7-2 shows the theoretical dispersion curves adjusted for the Colina site in Figure 7-1. Figure 7-3a shows the profile of shear wave velocities associated with the theoretical dispersion curve associated to the lowest misfit, while the Figure 7-3b shows the shear wave profile and the resulting V_{s30} value.
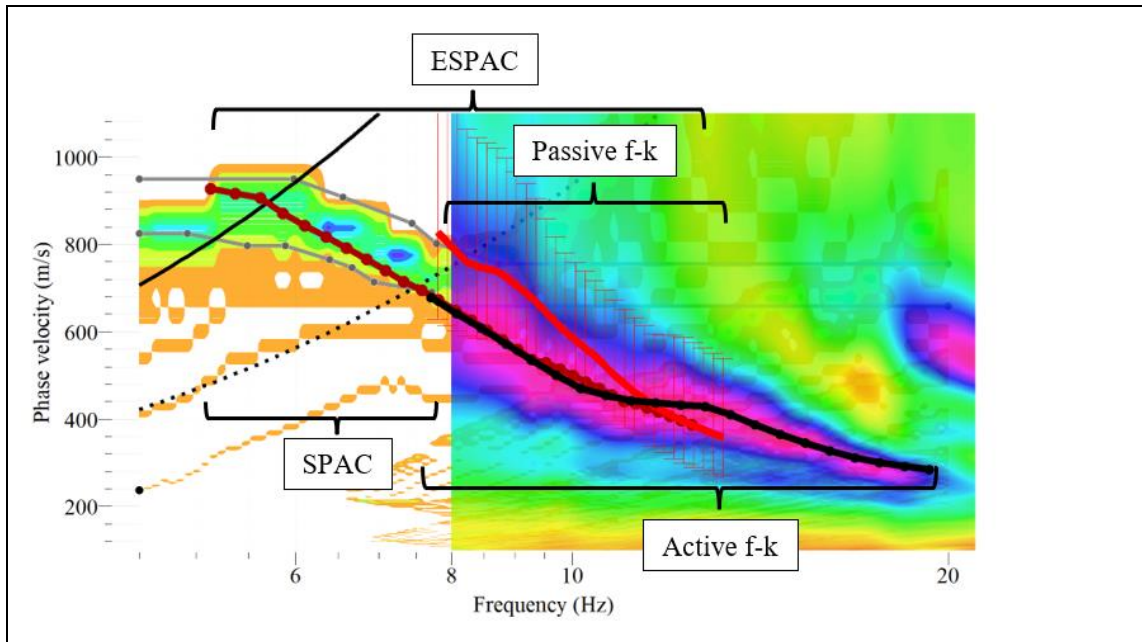
Figure 7-1: Empirical dispersion curves obtained by different methods: Active f-k (7.7 – 19.3 Hz), Passive f-k (7.8 – 13.1 Hz), ESPAC (5.1 – 12.5 Hz) and SPAC (5.7 – 7.8 Hz).
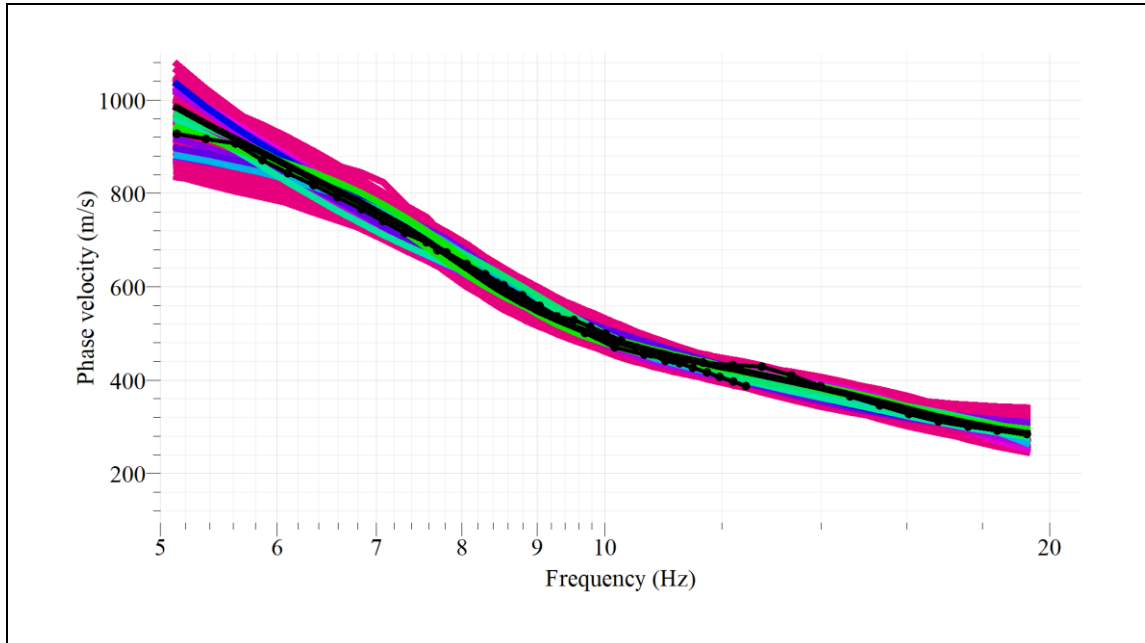
Figure 7-2: Fitted dispersion curve in software Geopsy®.

H_i (m)	H_f (m)	V_s (m/s)
0	1.3	161
1.3	6.2	270
6.2	9	382
9	23.9	526
23.9	30	899
	V_{s30}	437
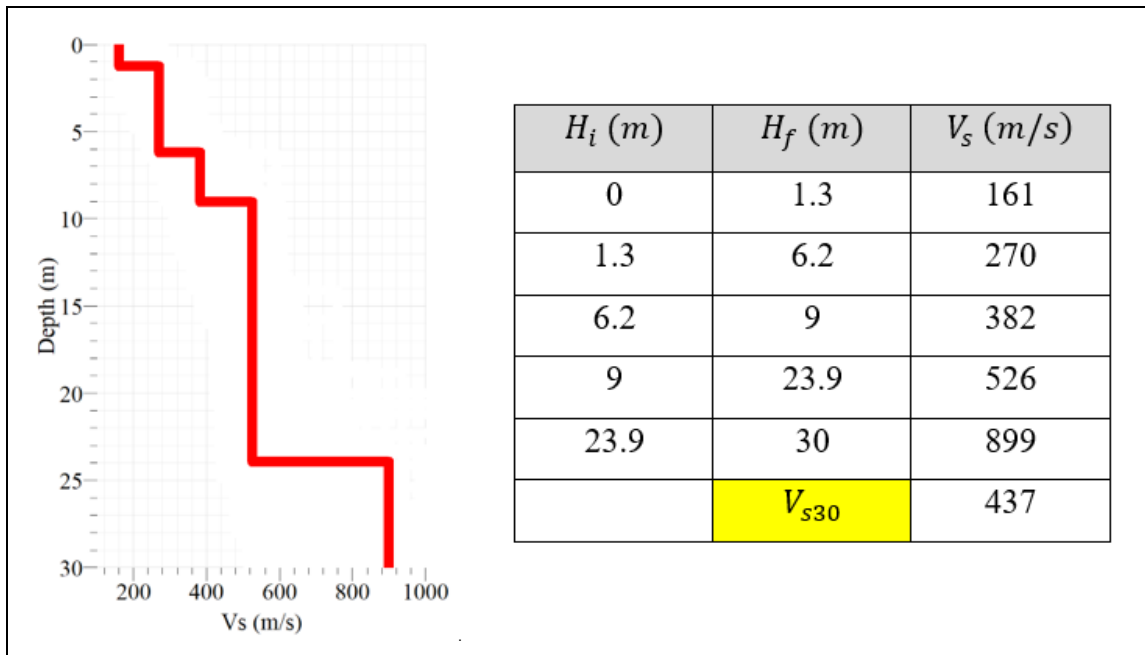
Figure 7-3: (a) Shear wave velocity profile resulting from the inversion process. (b) Detailed shear wave profile and V_{s30} value.

Annex 1.2: HVSR method

The usual horizontal to vertical spectral ratio (HVSR) or Nakamura method (1989) consists of estimating the ratio between the Fourier amplitude spectrum of the horizontal and vertical components of the ambient vibrations (H/V spectral ratio). It allows estimating the predominant frequency of a site f_0, which can be identified as the peak value of the H/V spectrum (Pastén, 2007).

The method assumes that the micro-tremors are essentially Rayleigh surface waves and that the amplifications due to site effects are due to the presence of a soil deposit over a more rigid half-space (Lermo & Chávez-García, 1993). Then, assuming that the vertical component of the movement does not undergo amplification due to the soil deposits, the effect of the Rayleigh waves on the vertical component (A_R) is calculated as:

$$A_R = \frac{V_S}{V_B} \tag{8}$$

where V_S and V_B are the amplitude of the spectra of the vertical components on the surface and in the rocky basement, respectively.

Furthermore, the method defines an estimate of the site effect A_E, as the ratio between the horizontal components at the surface (H_S) and at the basement (H_B):

$$A_E = \frac{H_S}{H_B} \tag{9}$$

If it is assumed that the vertical and horizontal components are modified in the same way by Rayleigh waves, the method proposes a modified site effect function (A_M):

$$A_M = \frac{A_E}{A_F} = \frac{\dfrac{H_S}{V_S}}{\dfrac{H_B}{V_B}} \tag{10}$$

Because the environmental vibrations in a rocky outcrop are approximately the same in any direction (Nakamura, 1989), $H_B/V_B \approx 1$ and:

$$A_M = \frac{H_S}{V_S} \tag{11}$$

The ambient noise measurements to obtain f_0 in this study were recorded with Tromino® triaxial geophones. Data analysis was performed using a Matlab® implementation, where the records were divided into 60-second windows and the Stockwell transform was calculated for each window (Leyton, Ramírez & Vásquez, 2012). Finally, the implementation combines the horizontal components and calculates their ratio with respect to the vertical component. According to the obtained results, the spectral ratios from this method were classified into 4 types:

- Type 1: Spectral ratios where a peak with an amplitude greater than 2 is clearly identified.
- Type 2: Spectral ratios where more than one peak can be clearly identified with amplitudes greater than 2, and therefore, more than one possible predominant frequency.
- Type 3: Flat spectral ratios, without predominant frequencies.
- Type 4: Spectral ratios without peaks or clear predominant frequencies, but with amplitudes greater than 2.

Figure 7-4 shows examples of the 4 types of spectral ratios obtained in this study. It is important to mention that the amplitude of the HVSR spectrum (A_0) can be a relevant indicator of the impedance contrast between two or more soil strata (Pilz et al., 2010;

Leyton et al., 2013). This value was not explicitly used in this study, but its use is considered in future stages of the research.
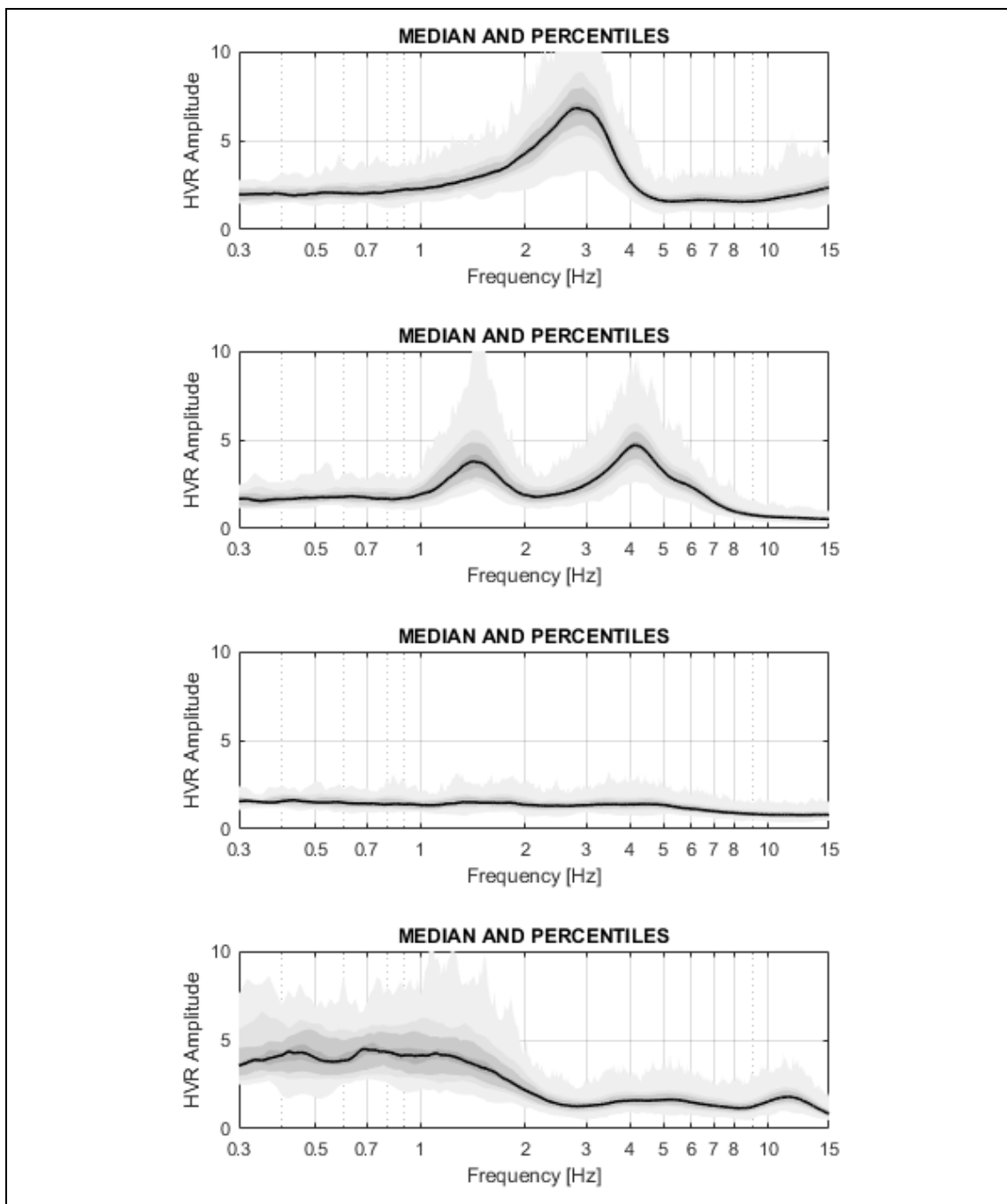
Figure 7-4: Examples of HVSR curve types in the localities of: a) Colina, b) Huechuraba, c) San Miguel and d) Buin.

Annex 2: Site dynamic characterization data

Table A-1: Summary of the geophysical parameters compiled and measured in this study.

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T001	294777	6270384	498	P	P	C
T002	292687	6270828	383	3.50	4.90	C
T003	295130	6283553	341	1.81	6.32	D
T004	306198	6295565	297	0.98	6.06	E
T005	319159	6271813	531	P	P	B
T006	320743	6273464	489	P	P	C
T007	322258	6273637	477	P	P	C
T008	324550	6279012	480	P	P	C
T009	326957	6278809	492	P	P	C
T010	323229	6287627	217	1.57	2.92	D
T011	330079	6283434	530	13.60	2.52	B
T012	332441	6282757	477	15.00	2.31	C
T013	332856	6283934	558	P	P	B
T014	332844	6287891	510	1.72	3.20	C
T015	334677	6286315	517	P	P	B
T016	337134	6291459	386	2.92	2.79	C
T017	336400	6294048	341	1.15	2.41	D
T018	336749	6294335	411	1.10	3.58	D
T019	339140	6295113	434	P	P	C
T020	340103	6295279	405	2.47	2.14	D
T021	338187	6294280	398	0.37	3.00	D
T022	337989	6293973	425	P	P	C
T023	338610	6294320	423	1.23	3.27	D
T024	335823	6292743	419	P	P	C
T025	336480	6293208	426	14.90	5.14	C
T026	335381	6290836	481	3.15	2.61	C
T027	329719	6299876	293	1.15	6.39	D
T028	334846	6298352	441	0.43	3.53	D
T029	336712	6299889	330	0.37	3.35	E

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T030	335555	6299991	308	0.37	3.98	E
T031	337127	6298994	369	2.60	3.12	D
T032	335876	6298754	429	0.38	3.38	D
T033	334996	6297386	554	0.48	2.65	C
T034	331035	6296691	410	4.58	6.53	C
T035	330681	6296909	453	2.25	4.82	D
T036	330301	6298419	286	3.32	4.78	D
T037	336318	6297601	377	1.05	4.13	D
T038	336177	6296337	434	12.00	2.65	C
T039	339812	6296705	414	1.80	2.50	D
T040	324164	6314589	478	2.22	2.81	D
T041	325554	6315509	275	1.72	3.50	D
T042	324181	6315936	282	2.48	7.75	D
T043	330659	6320709	237	0.71	4.97	E
T044	335375	6323377	268	1.35	3.16	D
T045	324566	6316415	281	1.37	2.70	D
T046	325588	6315999	256	2.35	2.99	D
T047	325400	6315868	303	1.65	3.73	D
T048	325555	6315534	288	1.66	2.56	D
T049	325848	6315280	246	1.50	3.43	D
T050	325094	6314106	327	1.45	2.63	D
T051	325121	6314658	271	1.27	3.34	D
T052	324516	6314231	353	P	P	C
T053	325155	6314021	326	1.53	2.66	D
T054	324158	6315945	304	2.48	7.75	D
T055	336726	6316464	256	0.42	4.37	E
T056	335905	6316466	257	0.45	3.40	E
T057	335831	6315527	326	0.40	5.00	E
T058	335351	6315118	275	0.40	5.22	E
T059	333254	6311732	264	0.55	4.80	E
T060	333893	6312842	266	0.45	4.13	E
T061	334321	6313472	264	0.42	3.52	E
T062	334806	6314132	272	0.45	4.25	E

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T063	334077	6311638	243	0.33	3.12	E
T064	334876	6312359	275	0.43	6.02	E
T065	335414	6312481	244	0.38	5.39	E
T066	335946	6313193	241	0.41	3.35	E
T067	335381	6315852	340	0.42	4.43	E
T068	335534	6311470	312	1.58	2.90	D
T069	337150	6311925	264	0.50	2.50	E
T070	336505	6312403	302	1.02	5.82	D
T071	324962	6314422	321	1.47	5.29	D
T072	327383	6313340	254	1.12	7.14	D
T073	334968	6312716	276	0.38	3.64	E
T074	345356	6326845	605	1.50	6.09	C
T075	345327	6326862	491	11.40	3.20	C
T076	344538	6326224	519	P	P	B
T077	344970	6325617	578	15.00	2.70	B
T078	342792	6313943	277	2.85	5.04	D
T079	342038	6313758	280	1.68	4.67	D
T080	342910	6313819	305	2.88	6.97	D
T081	344214	6313228	320	1.03	5.67	D
T082	344224	6313230	319	1.03	5.67	D
T083	348766	6317439	415	1.03	5.04	D
T084	345965	6316493	310	1.80	3.00	D
T085	347189	6317713	331	N	N	D
T086	348376	6319501	495	2.98	2.85	C
T087	346861	6318613	385	P	P	C
T088	346839	6318622	389	P	P	C
T089	346261	6312729	342	1.70	3.25	D
T090	346331	6312127	432	2.43	4.69	D
T091	345497	6312983	306	14.60	4.05	D
T092	345329	6313937	341	1.78	2.80	D
T093	344013	6327171	673	P	P	B
T094	343962	6326563	646	P	P	B
T095	344662	6326697	779	P	P	B

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T096	344910	6325909	689	1.33	3.23	C
T097	343046	6323805	484	P	P	C
T098	343515	6323477	610	P	P	B
T099	343054	6322882	575	1.18	4.66	C
T100	342530	6321997	380	0.73	2.47	D
T101	344994	6324874	352	P	P	C
T102	345072	6324315	661	P	P	B
T103	345437	6323793	663	1.45	5.00	C
T104	344818	6323014	378	9.60	3.63	C
T105	338149	6322891	282	P	P	D
T106	338696	6323043	306	2.08	3.80	D
T107	338955	6323599	308	1.90	3.00	D
T108	339328	6322940	278	P	P	D
T109	342994	6324248	473	13.90	3.88	C
T110	342740	6325186	610	0.83	5.95	C
T111	342687	6324611	587	P	P	B
T112	343498	6326350	629	P	P	B
T113	338622	6322621	334	2.40	3.12	D
T114	339089	6324756	246	1.60	4.32	D
T115	338767	6325594	335	2.38	3.19	D
T116	339475	6324586	325	1.40	2.53	D
T117	342994	6323371	480	8.28	3.72	C
T118	345932	6316618	304	1.30	4.80	D
T119	346887	6316163	300	1.50	2.74	D
T120	337135	6306245	254	0.43	3.42	E
T121	339131	6306335	245	0.80	5.56	E
T122	340014	6306380	275	1.10	5.98	D
T123	342108	6305450	261	2.01	8.15	D
T124	342556	6310154	486	7.53	3.92	C
T125	342504	6310192	565	7.53	3.92	B
T126	341516	6306953	359	1.15	3.76	D
T127	335774	6302122	312	0.43	4.13	E
T128	339234	6301777	356	1.32	3.48	D

Table A-1: Summary of the geophysical parameters compiled and measured in this study
(cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T129	341791	6302636	308	0.75	2.22	E
T130	342880	6301576	462	P	P	C
T131	342864	6301872	407	P	P	C
T132	342767	6302299	453	P	P	C
T133	342522	6302929	328	0.47	7.11	E
T134	342235	6303372	302	1.05	2.41	D
T135	342383	6303784	272	1.00	3.06	E
T136	342456	6304841	315	2.10	3.74	D
T137	342930	6304284	328	0.88	2.45	E
T138	341814	6303527	301	1.63	5.99	D
T139	340031	6303347	445	3.78	4.03	C
T140	339018	6303088	392	4.05	4.02	C
T141	339645	6303582	422	3.07	5.19	C
T142	341830	6302645	363	0.75	2.22	D
T143	340578	6302184	336	1.77	2.39	D
T144	339538	6302029	320	1.52	2.49	D
T145	338567	6302434	345	0.98	3.48	E
T146	336619	6302584	328	0.58	2.76	E
T147	336491	6302920	322	1.03	4.24	D
T148	336915	6303544	344	1.25	2.47	D
T149	337345	6303383	396	N	N	D
T150	337579	6302760	301	P	P	D
T151	338228	6302906	304	1.93	3.91	D
T152	336743	6300311	296	0.25	6.57	E
T153	337653	6301662	367	0.42	3.77	D
T154	336757	6301393	297	0.42	4.46	E
T155	338864	6301432	292	1.17	7.11	D
T156	337777	6300922	301	0.43	2.69	E
T157	336668	6300880	310	0.40	5.27	E
T158	335944	6301350	323	0.38	5.36	E
T159	337300	6301107	302	0.38	4.02	E
T160	338354	6301361	363	P	P	C
T161	340486	6301025	318	1.13	2.64	D

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T162	338244	6300565	344	2.97	3.33	D
T163	336645	6300446	343	0.42	3.52	E
T164	337669	6299725	340	3.10	3.16	D
T165	338147	6299926	347	2.85	2.69	D
T166	341094	6300981	309	0.46	2.03	E
T167	338825	6297979	408	0.38	4.07	D
T168	340514	6297299	494	3.50	2.60	C
T169	340557	6297081	438	3.37	2.72	C
T170	341160	6296108	385	11.40	2.15	C
T171	340645	6295564	409	1.10	3.07	D
T172	341484	6294907	460	P	P	C
T173	342789	6289082	615	P	P	B
T174	340960	6284556	393	P	P	C
T175	342362	6279510	563	P	P	B
T176	339086	6277141	539	P	P	B
T177	335279	6266080	426	11.20	3.42	C
T178	338880	6265969	572	P	P	B
T179	337182	6257389	424	0.50	2.47	D
T181	356352	6266808	397	P	P	C
T182	350750	6277687	461	3.81	3.26	C
T183	350993	6278918	544	P	P	B
T184	353345	6280328	548	P	P	B
T185	350521	6282550	589	1.22	2.54	C
T186	348809	6286814	573	P	P	B
T187	345080	6287850	502	5.41	2.74	B
T188	348093	6289416	583	P	P	B
T189	346749	6292902	570	P	P	B
T190	346748	6293102	567	P	P	B
T191	345517	6298593	602	P	P	B
T192	344352	6300839	552	P	P	B
T193	346296	6301472	465	P	P	C
T194	347744	6301723	616	P	P	B
T195	349432	6304181	288	1.47	5.45	D

Table A-1: Summary of the geophysical parameters compiled and measured in this study
(cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T196	345160	6306521	334	1.43	2.08	D
T197	345130	6306351	303	1.50	4.50	D
T198	351884	6305078	515	P	P	B
T199	356304	6310000	323	2.03	5.50	D
T200	358713	6307874	356	P	P	C
T201	358705	6307197	471	1.20	2.70	D
T202	360732	6307041	383	6.45	2.70	C
T203	358745	6304681	419	P	P	C
T204	357039	6302582	408	6.40	3.74	C
T205	356833	6298622	563	P	P	B
T206	354058	6298218	382	9.60	2.32	C
T207	354522	6297399	395	P	P	C
T208	352698	6296957	429	P	P	C
T209	350553	6295304	403	P	P	C
T210	351322	6294791	361	P	P	C
T211	351564	6294718	355	P	P	C
T212	349932	6293244	561	P	P	B
T213	351992	6291122	233	2.42	3.78	D
T214	351545	6290047	564	P	P	B
T215	352478	6290388	542	P	P	B
T216	352580	6289633	542	P	P	B
T217	352450	6289274	477	12.73	6.60	C
T218	353778	6288301	473	4.97	2.38	C
T219	353947	6288251	489	6.05	2.76	C
T220	355647	6288074	420	2.40	2.07	D
T221	356925	6291318	487	P	P	C
T222	353466	6291811	372	2.80	2.10	C
T223	353388	6293256	364	P	P	C
T224	353696	6296018	406	P	P	C
T225	357187	6294287	503	P	P	B
T226	359471	6294912	542	P	P	B
T227	359052	6294388	400	P	P	C
T228	258815	6276825	280	7.77	6.47	D

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T229	257084	6277262	398	7.38	2.47	C
T230	257144	6277954	268	1.55	8.06	D
T231	256820	6277462	251	2.35	6.21	D
T232	256425	6275798	239	2.32	3.29	D
T233	256438	6273966	415	3.95	5.88	C
T234	345516	6298538	541	P	P	B
T235	356464	6302720	522	25.00	3.40	B
T236	347845	6305691	326	1.23	2.88	D
T237	335378	6320418	284	0.43	3.17	E
T238	327647	6280445	569	P	P	B
T239	320132	6334329	387	2.02	4.74	D
T240	359793	6308141	286	P	P	D
T241	325353	6280750	511	1.93	2.12	C
T242	357919	6294011	535	P	P	B
T243	344083	6323499	437	7.80	3.77	C
T244	341485	6281937	619	1.18	5.09	C
T245	258465	6282325	545	9.72	4.29	B
T246	256938	6281670	475	8.38	7.09	C
T247	256550	6281597	388	6.02	9.15	C
T248	258053	6281811	441	8.50	2.87	C
T249	258509	6281289	360	7.70	3.28	C
T250	258448	6280948	345	4.50	5.92	D
T251	257778	6280429	383	6.25	3.23	C
T252	259212	6280352	464	3.45	7.87	C
T253	258247	6280208	375	5.18	6.36	C
T254	258792	6279774	360	3.65	9.38	C
T255	259112	6279286	419	4.00	5.75	C
T256	258415	6279183	289	1.02	3.38	D
T257	257499	6279417	380	1.05	5.25	D
T258	257620	6279023	303	1.88	5.29	D
T259	257741	6278560	245	0.97	5.69	E
T260	257347	6278229	248	1.00	8.71	E
T261	258138	6278037	446	7.38	4.11	C

Table A-1: Summary of the geophysical parameters compiled and measured in this study

(cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T262	258848	6278313	442	3.30	5.51	C
T263	257696	6277327	361	2.15	7.61	D
T264	258660	6277495	345	4.83	3.58	D
T268	258885	6275764	286	1.25	3.10	D
T269	257032	6276968	188	1.25	10.40	D
T270	257389	6277662	389	3.99	11.15	C
T271	257803	6277776	306	1.14	10.94	D
T272	257407	6278595	320	1.07	0.08	D
T273	259048	6278567	280	1.65	0.17	D
T274	257649	6280150	278	4.10	0.18	D
T275	258350	6280680	369	2.43	0.05	D
T276	256450	6281797	367	10.45	5.22	C
T277	258152	6281733	269	1.33	3.41	D
T278	357716	6296014	616	P	P	B
T279	357937	6297394	510	P	P	B
T280	356906	6300282	448	P	P	C
T281	359793	6310764	433	2.55	2.93	C
T282	355377	6311857	253	1.45	5.78	D
T283	355396	6310467	313	2.47	4.06	D
T284	341977	6297517	459	0.37	2.94	D
T285	341845	6295842	491	0.37	2.63	D
T286	342768	6295482	510	0.33	2.83	C
T287	295542	6271181	586	4.10	2.82	B
T288	295768	6268700	564	1.07	4.94	C
T289	299509	6272256	511	1.32	4.52	C
T290	312365	6299547	456	4.50	4.32	C
T291	311527	6298504	429	1.50	5.29	D
T292	313903	6297992	1117	P	P	A
T293	336135	6300325	298	0.36	7.37	E
T294	342986	6316261	298	0.82	7.96	E
T295	342819	6316648	295	0.78	7.73	E
T296	348933	6283758	604	P	P	B
T297	359317	6305002	604	1.18	3.39	C

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T298	343606	6307951	254	1.20	8.77	D
T299	337890	6303795	390	P	P	C
T300	343252	6265525	615	P	P	B
T301	345757	6300754	566	P	P	B
T302	353100	6288456	555	P	P	B
T303	357542	6298820	613	P	P	B
T304	344797	6279750	576	1.68	2.76	C
T305	343860	6299616	552	P	P	B
T306	334608	6292390	305	0.45	2.94	E
T307	333500	6301210	360	0.37	7.96	D
T308	374903	6276811	409	N	N	D
T309	352359	6313288	839	P	P	B
T310	351168	6313802	836	P	P	B
T311	375938	6276024	993	P	P	A
T312	377787	6277795	1954	P	P	A
T313	356152	6312510	1000	P	P	A
T314	355506	6312411	1108	P	P	A
T315	355245	6293564	457	P	P	C
T316	355636	6286990	380	3.26	8.37	C
T317	355975	6299752	395	P	P	C
T318	344770	6293735	378	2.60	3.84	C
T319	338337	6307933	250	0.40	3.50	E
T320	340680	6309387	302	0.88	4.49	E
T321	340707	6307148	279	0.85	4.33	D
T322	340280	6307249	269	1.13	4.36	D
T323	342383	6306823	296	1.04	4.30	D
T324	344249	6307063	286	0.88	7.45	D
T325	344105	6307253	296	1.15	5.49	D
T326	345692	6305495	325	1.28	3.60	D
T327	346027	6304460	335	P	P	D
T328	349432	6304626	265	1.14	7.81	D
T329	328510	6280412	573	P	P	B
T330	335845	6287162	675	P	P	B
T331	337096	6278732	670	P	P	B
T332	342441	6292896	620	P	P	B
T333	345772	6296382	634	P	P	B

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T334	348436	6284743	600	1.00	5.85	C
T335	349254	6278654	603	P	P	B
T336	350700	6289077	670	P	P	B
T337	353132	6298069	608	1.00	2.00	C
T338	352793	6286428	587	1.69	2.26	C
T339	354245	6303250	578	P	P	B
T340	355108	6284922	508	5.50	2.15	B
T341	344023	6300237	590	7.01	2.94	B
T342	343738	6298070	530	0.48	2.41	C
T343	328875	6290048	404	1.42	3.44	D
T344	330449	6290627	426	0.60	2.25	D
T345	333248	6291532	394	P	P	C
T346	338043	6294977	440	0.37	2.33	D
T347	328705	6301184	396	0.41	3.95	D
T348	335661	6297612	397	0.42	2.84	D
T349	337474	6290717	476	3.50	2.50	C
T350	257604	6272808	485	2.20	2.00	D
T351	335327	6323376	281	1.34	3.00	D
T352	349140	6304412	267	1.50	6.40	D
T353	332939	6305033	347	0.39	4.50	E
T354	337548	6291025	370	2.73	4.40	C
T355	357414	6298548	635	P	P	B
T356	294526	6269011	471	1.40	3.80	D
T357	356222	6303632	574	P	P	B
T358	345754	6295042	724	3.60	3.00	B
T359	321152	6273318	577	1.00	3.30	C
T360	257458	6278576	281	0.95	2.00	D
T361	258252	6280261	477	5.60	2.00	C
T362	353926	6280301	593	4.20	1.00	B
T363	342801	6306723	271	0.93	1.00	E
T364	356020	6283518	435	3.56	1.00	C
T365	359644	6308019	335	1.45	1.00	D

Table A-1: Summary of the geophysical parameters compiled and measured in this study (cont.).

Site ID	East (m)	North (m)	V_{s30} (m/s)	f_0 (Hz)	A_0 (-)	Sus. Index
T366	352742	6305521	534	0.83	2.70	C
T367	348689	6295824	441	P	P	C
T368	339623	6303974	1092	1.12	3.75	A
T369	348596	6312588	1050	P	P	A

Annex 3: Site cards of the explored sites in this research.

Peñalolén – T278
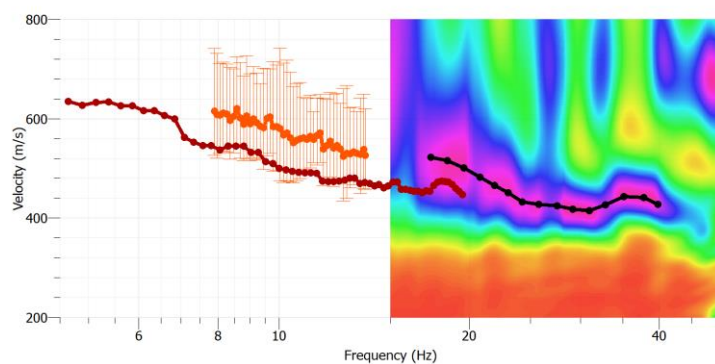
Dispersión Empírica
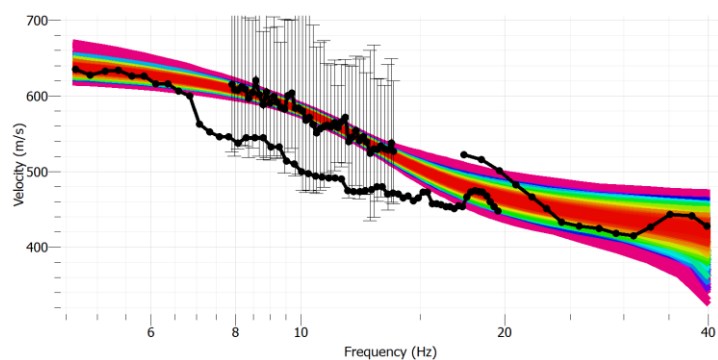
Dispersión Ajustada
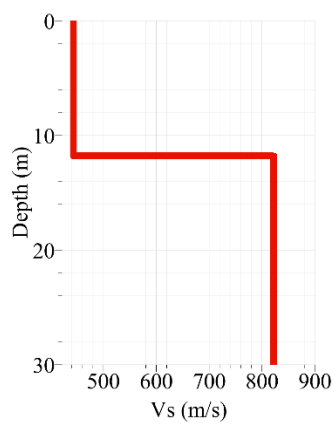
Hi (m)	Hf(m)	Vs (m/s)
0	11.7	443.7
11.7	30.0	822.3
	v_{s30}	616

Perfil Vs
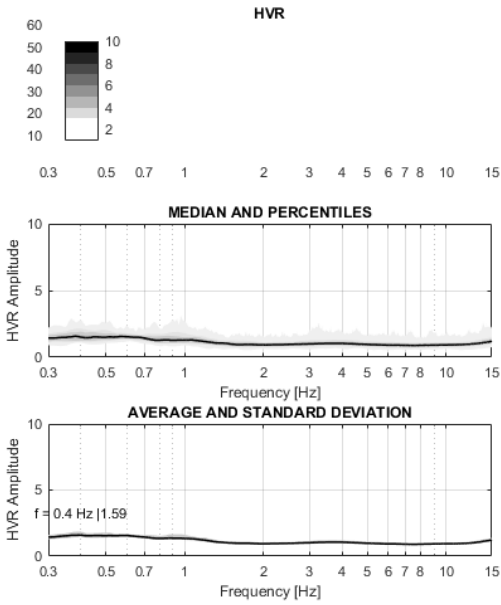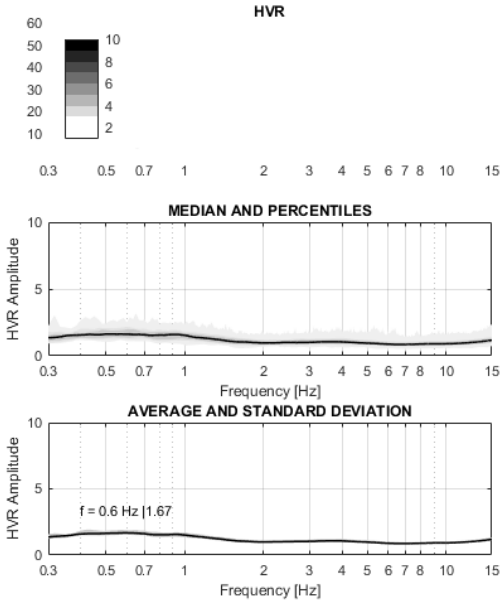
HV

La Reina – T279
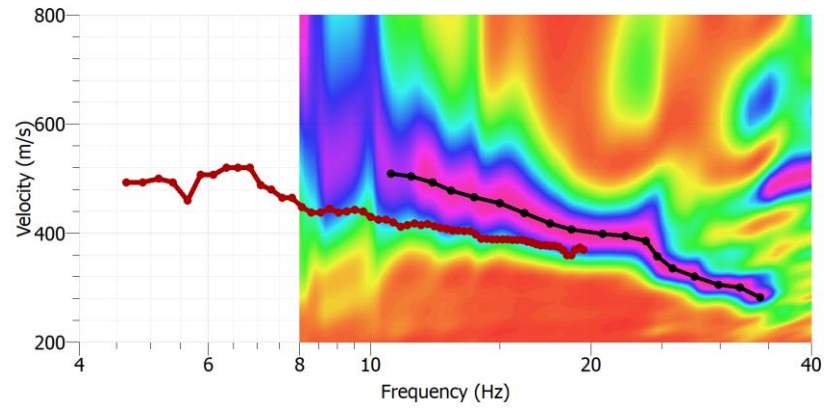
Dispersión Empírica
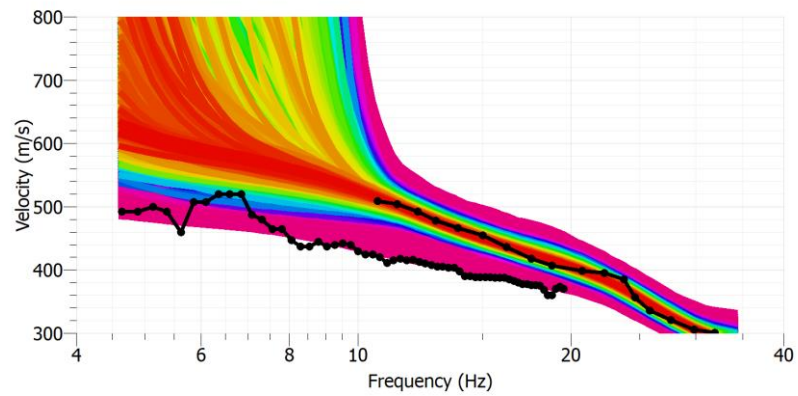
Dispersión Ajustada
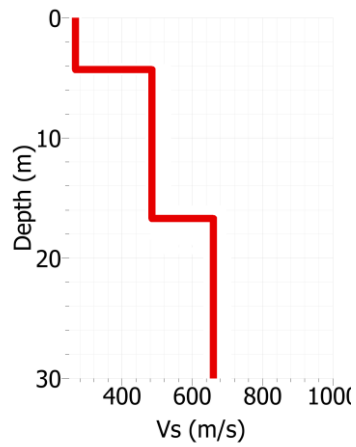
Hi (m)	Hf(m)	Vs (m/s)
0.0	4.2	269
4.2	16.7	491
16.7	30.0	663
	v_{s30}	510

Perfil Vs

HVR

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 0.4 Hz |1.53

f = 0.5 Hz |1.51
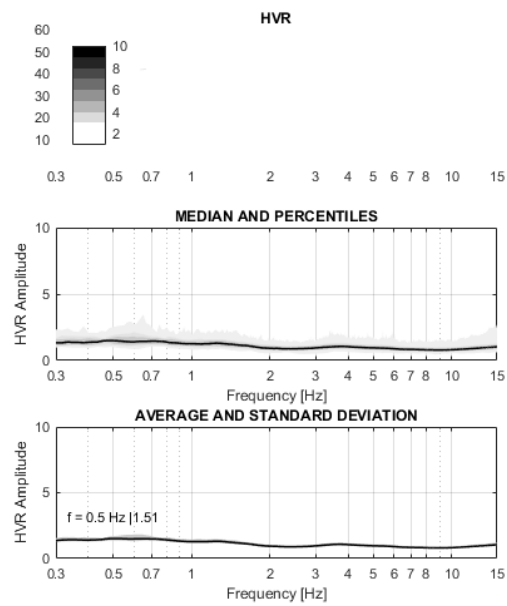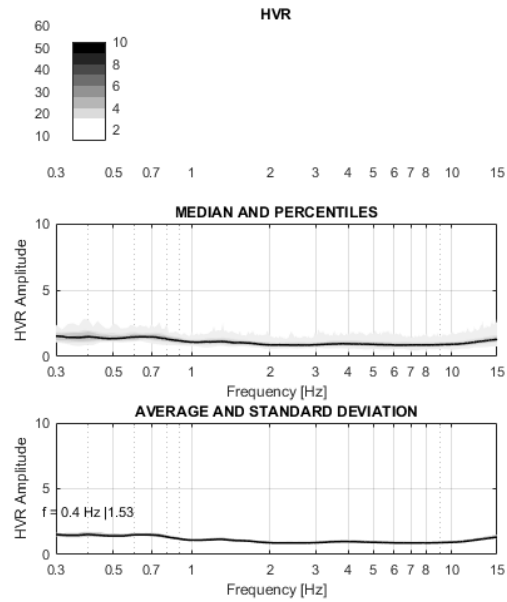
HV

Las Condes – T280
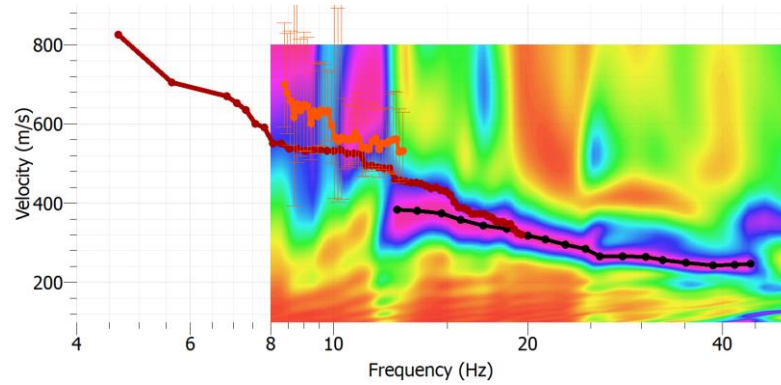
Dispersión Empírica
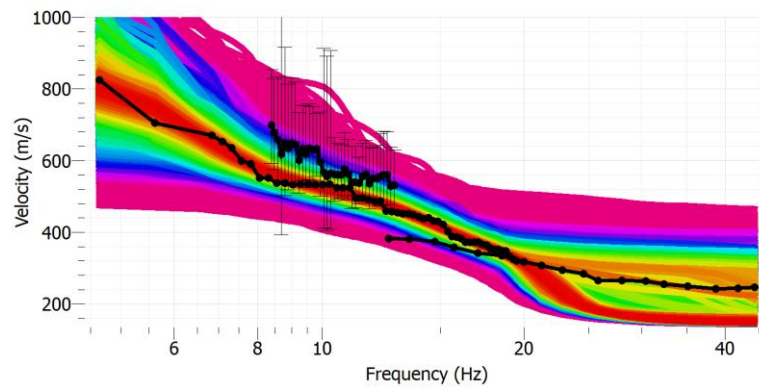
Dispersión Ajustada
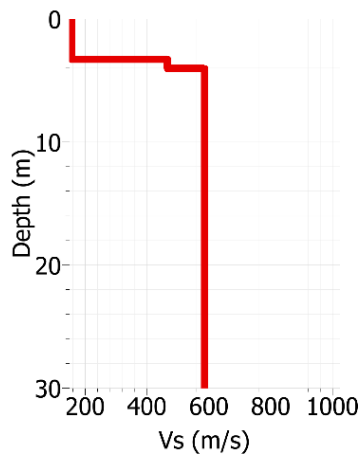
Hi (m)	Hf(m)	Vs (m/s)
0.0	3.3	156
3.3	4.0	466
4.0	30.0	586
	v_{s30}	448

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 15 Hz |2.05

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 0.333 Hz |1.73
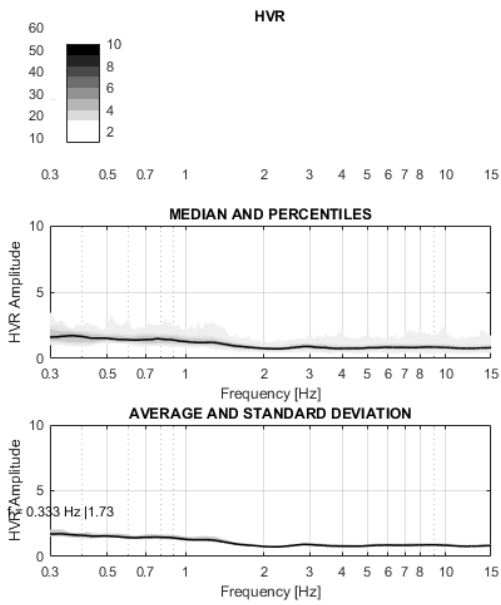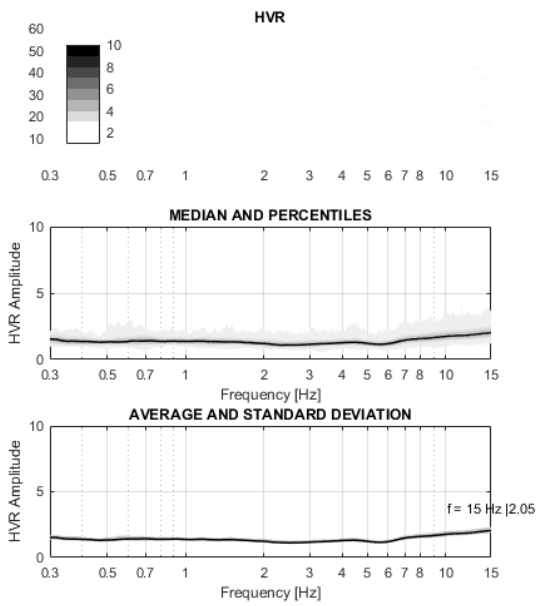
HV
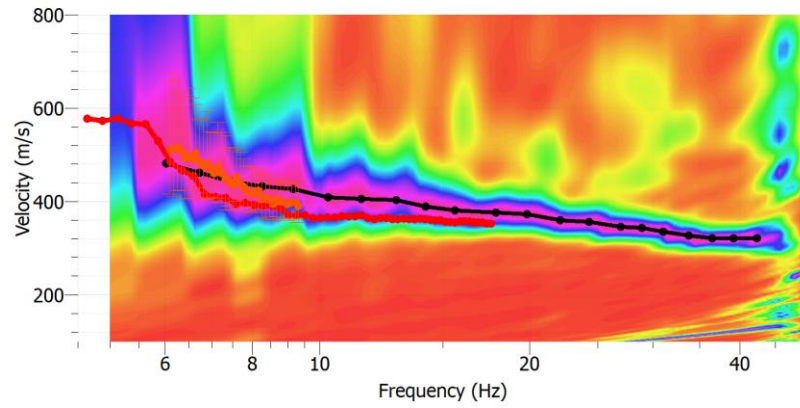
Lo Barnechea – T281

Dispersión Empírica
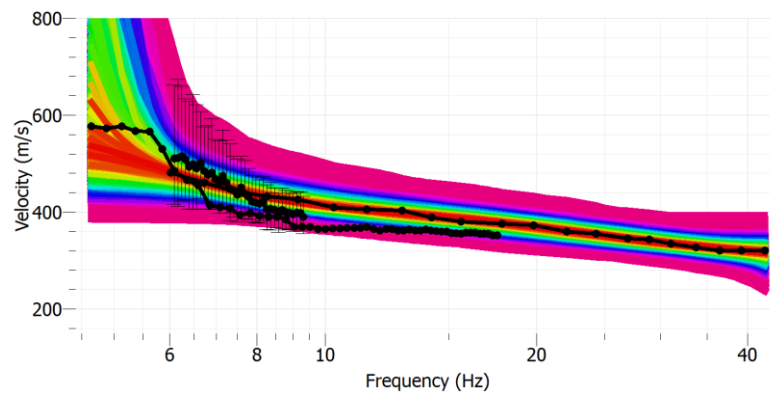
Dispersión Ajustada
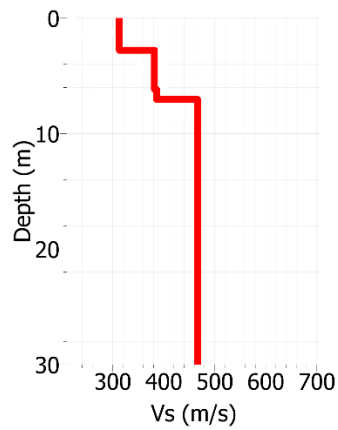
Hi (m)	Hf(m)	Vs (m/s)
0.0	2.8	313
2.8	6.2	385
6.2	7.1	389
7.1	30.0	469
	v_{s30}	433

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 2.55 Hz |2.93

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 2.83 Hz |2.15
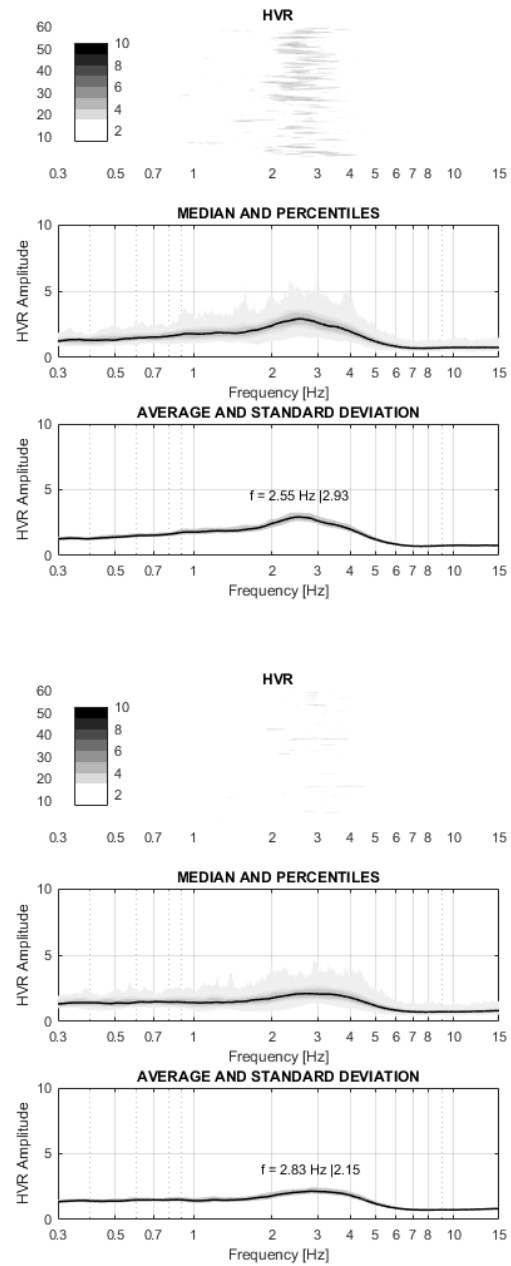
HV
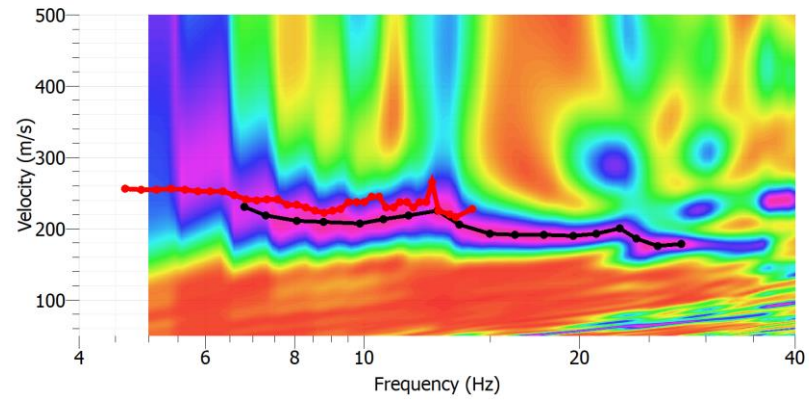
Lo Barnechea – T282

Dispersión Empírica
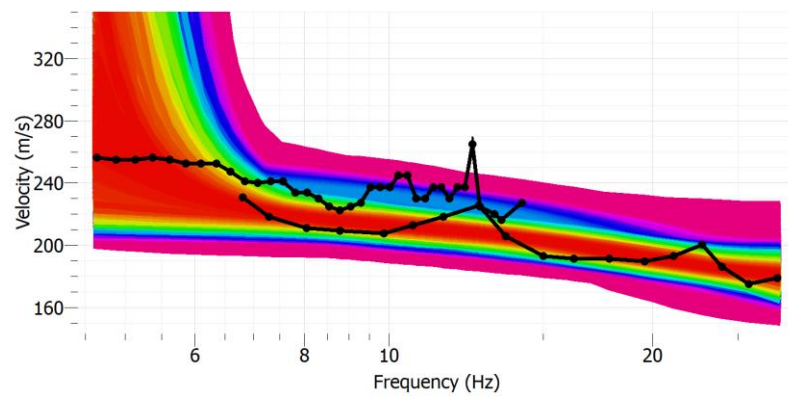
Dispersión Ajustada
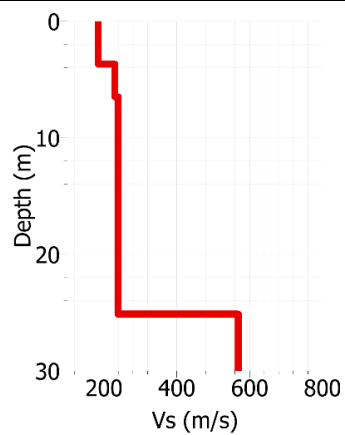
Hi (m)	Hf(m)	Vs (m/s)
0.0	3.7	185
3.7	6.5	230
6.5	25.1	240
25.1	30.0	569
	v_{s30}	253

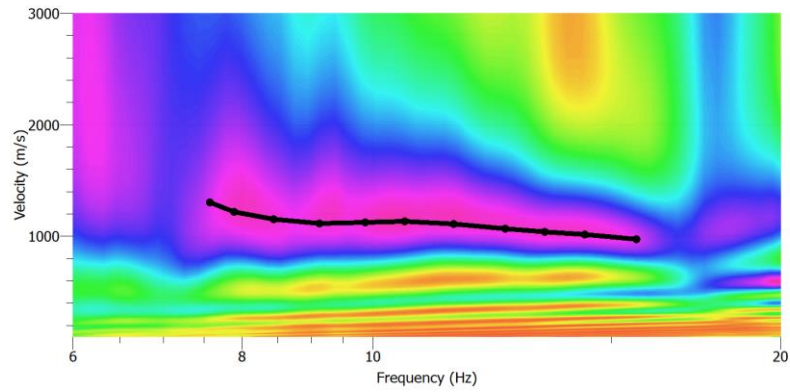Perfil Vs

HVR
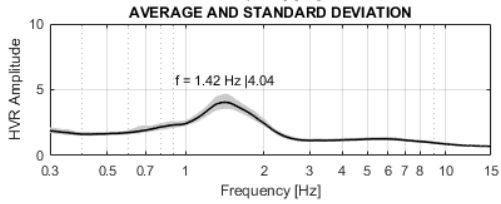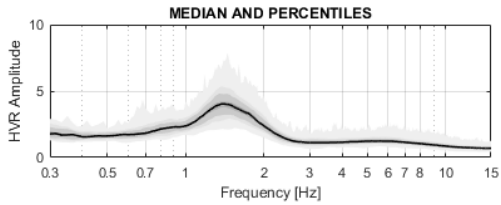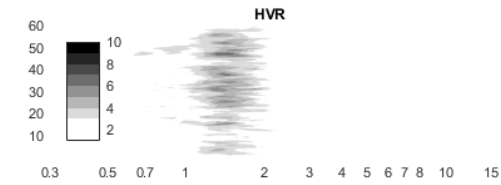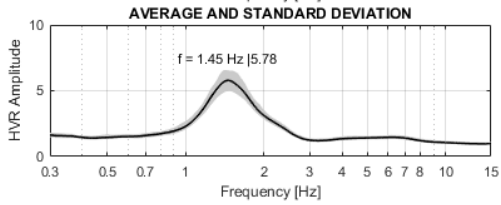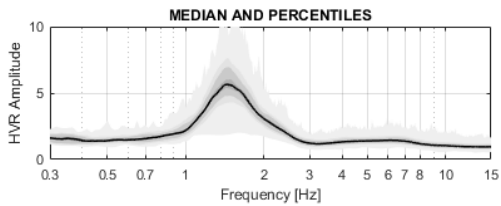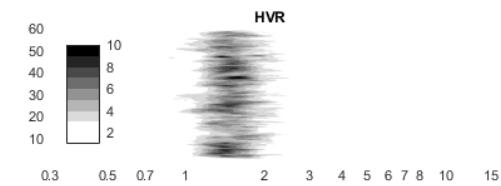
Lo Barnechea – T283
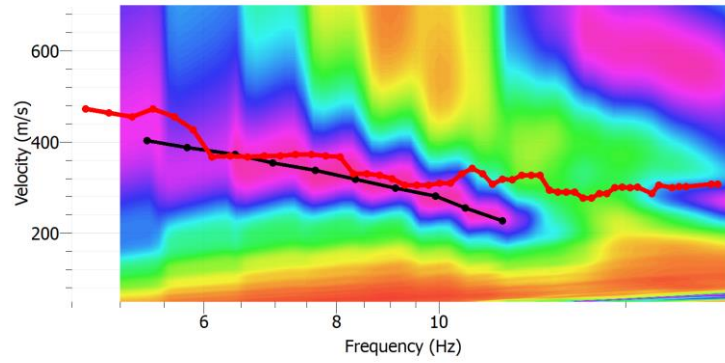
Dispersión Empírica
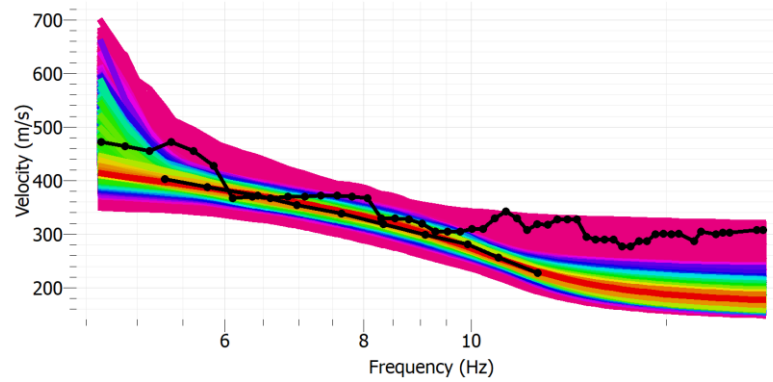
Dispersión Ajustada
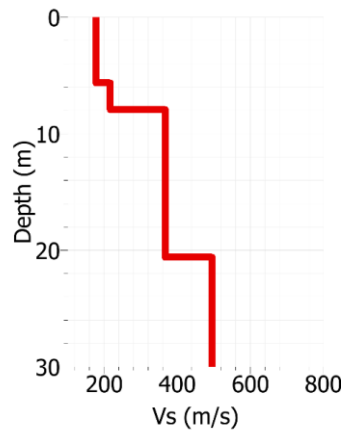
Hi (m)	Hf(m)	Vs (m/s)
0.0	5.6	178
5.6	7.9	217
7.9	20.6	367
20.6	30.0	495
	v_{s30}	313

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 2.47 Hz |4.06
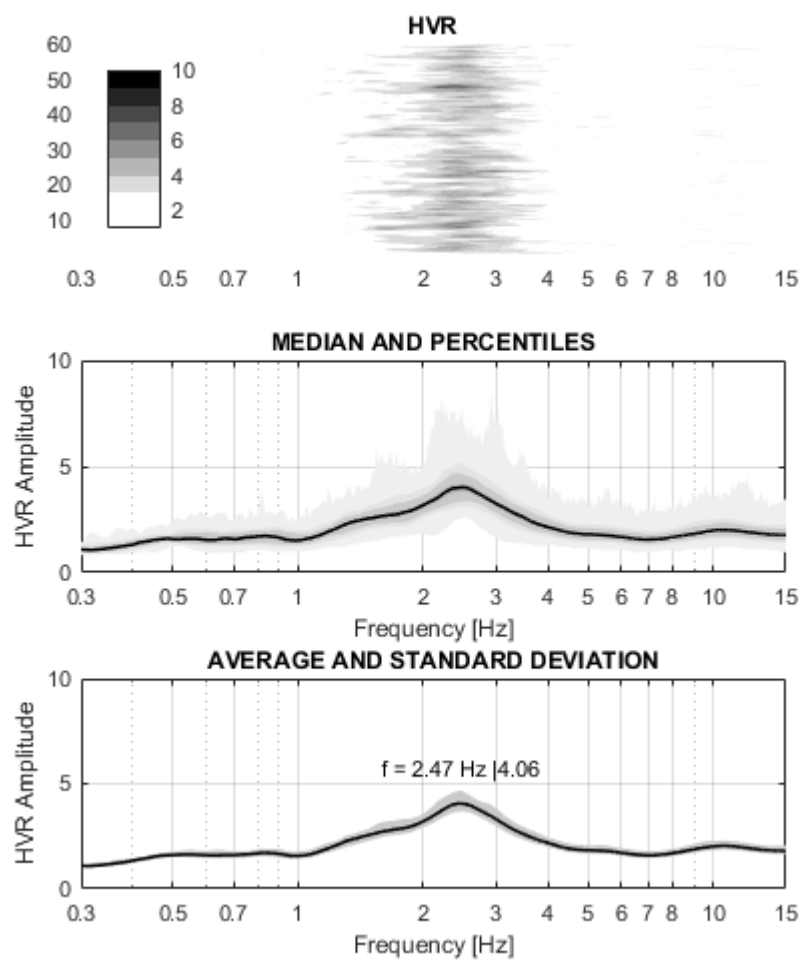
Estación Central – T284
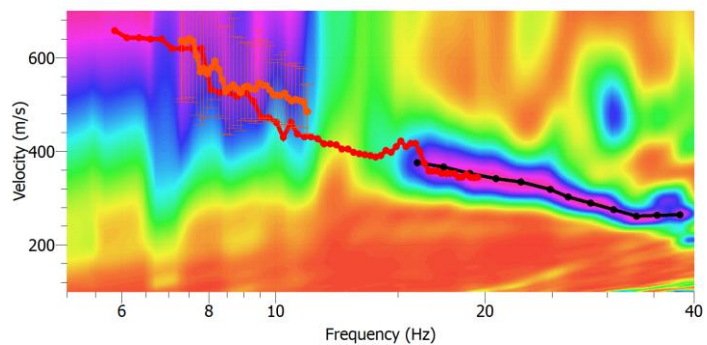
Dispersión Empírica
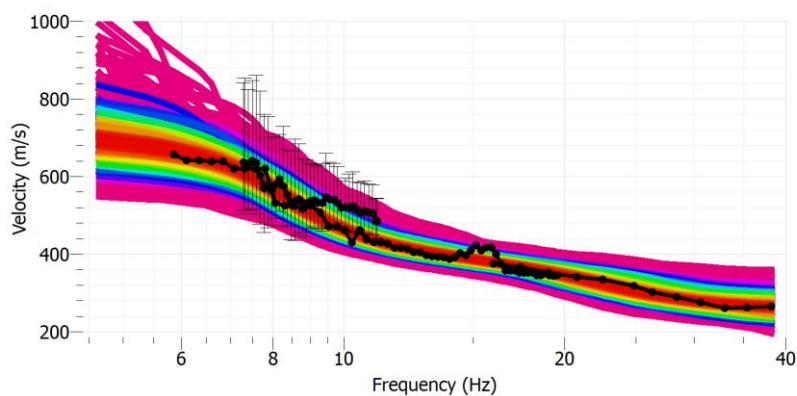
Dispersión Ajustada
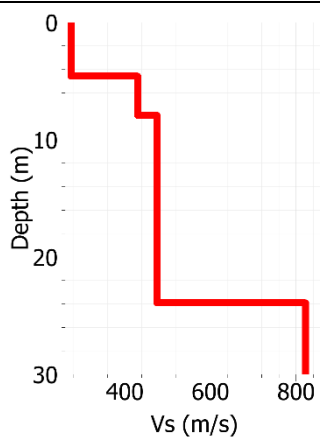
Hi (m)	Hf(m)	Vs (m/s)
0.0	4.7	272
4.7	8.0	434
8.0	24.0	479
24.0	30.0	825
	v_{s30}	459

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 0.367 Hz | 2.94

HV
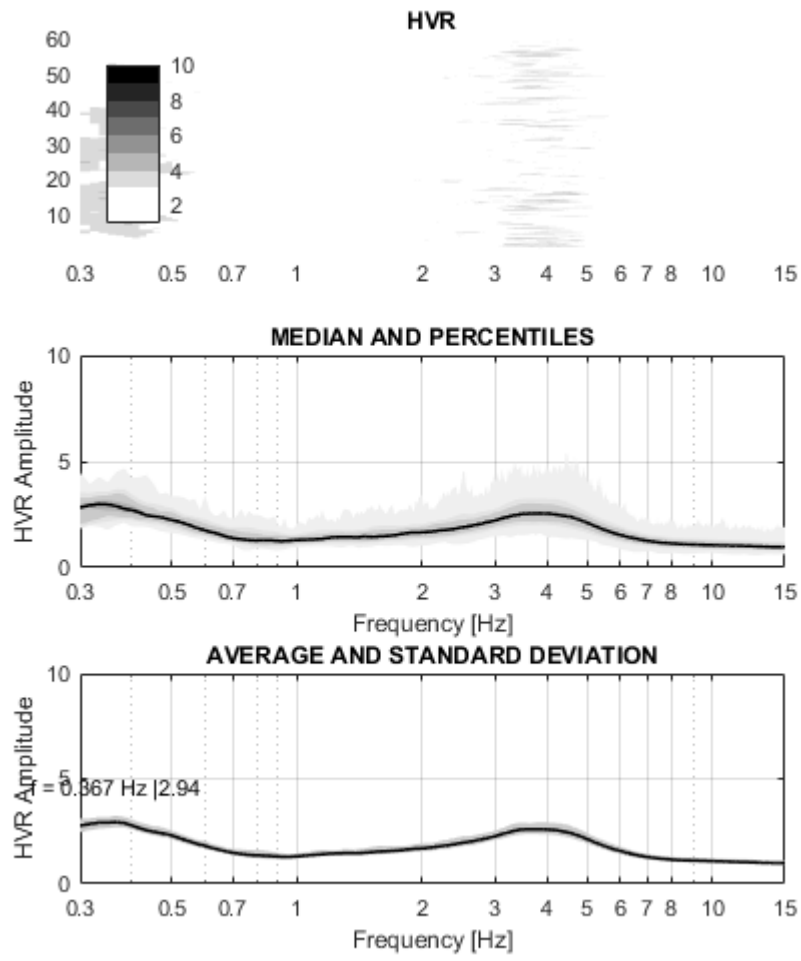
Estación Central – T285
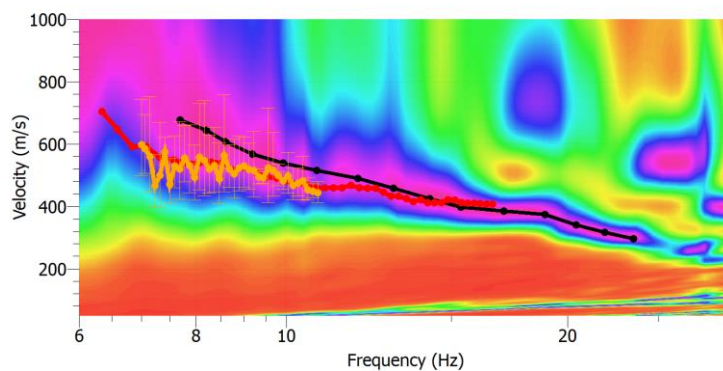
Dispersión Empírica
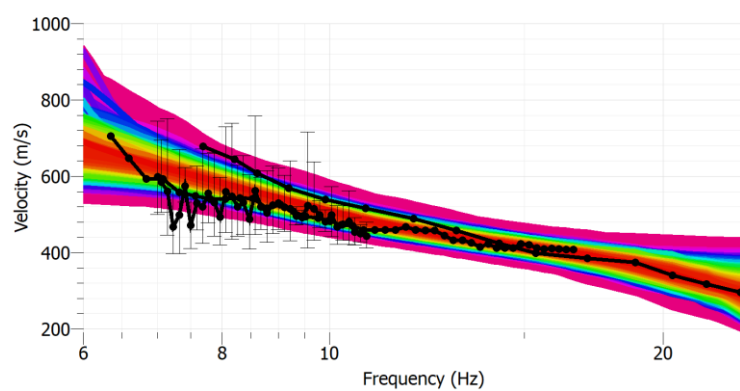
Dispersión Ajustada
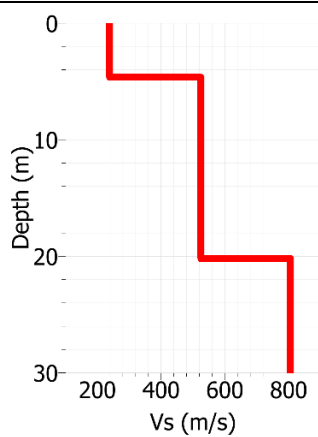
Hi (m)	Hf(m)	Vs (m/s)
0.0	4.6	239
4.6	20.2	526
20.2	30.0	806
	v_{s30}	491

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 0.367 Hz |2.63

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 0.4 Hz |2.79

HV
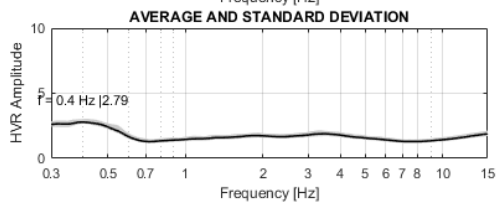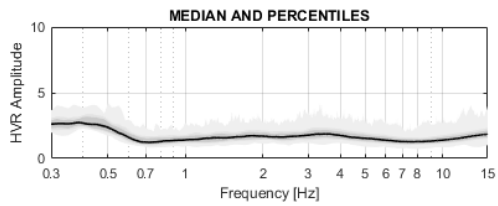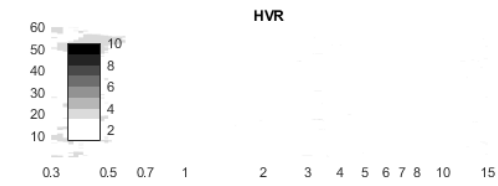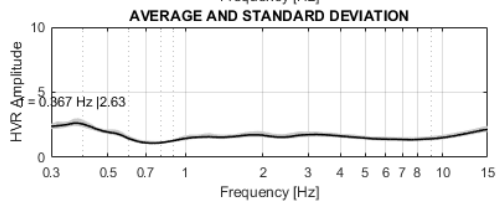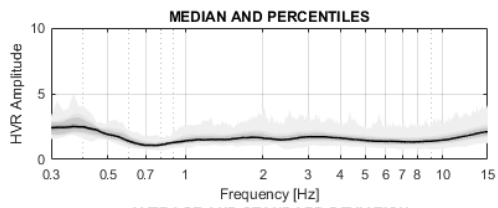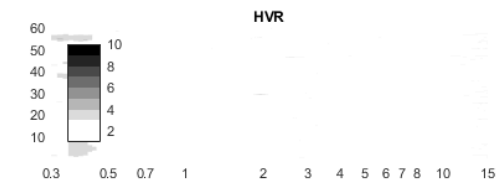
Estación Central – T286
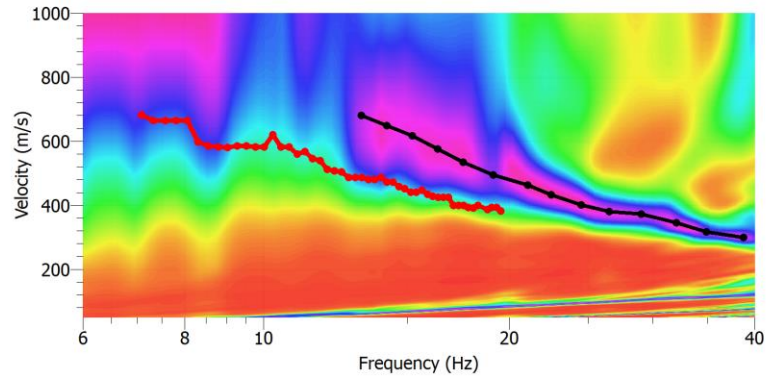
Dispersión Empírica
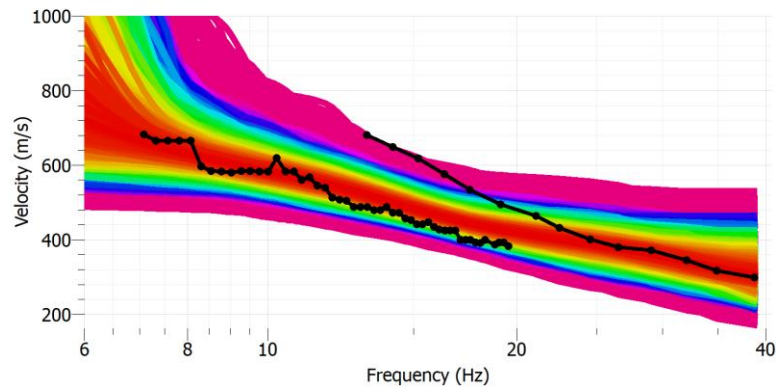
Dispersión Ajustada
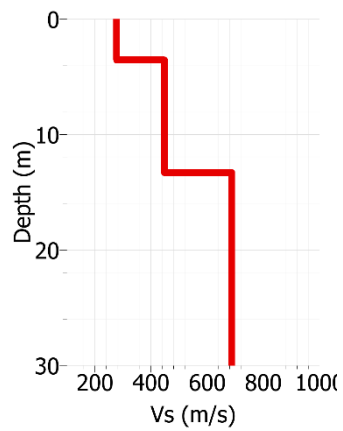
Hi (m)	Hf(m)	Vs (m/s)
0.0	3.5	278
3.5	9.8	448
9.8	30.0	687
	v_{s30}	510

Perfil Vs
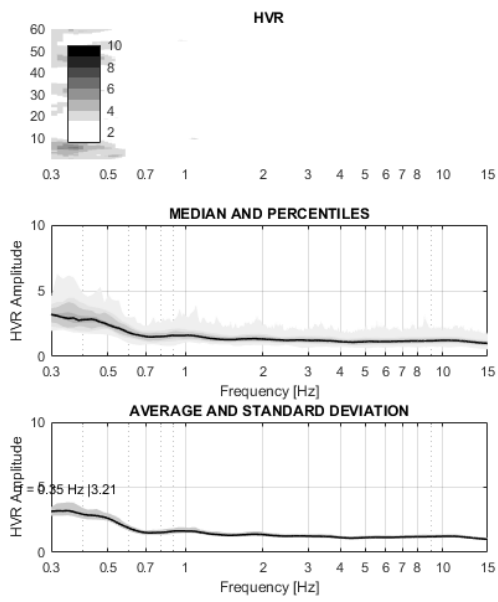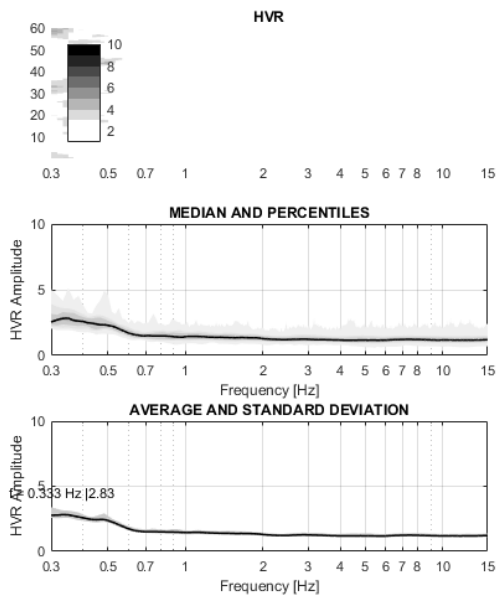
HV

Melipilla – T287
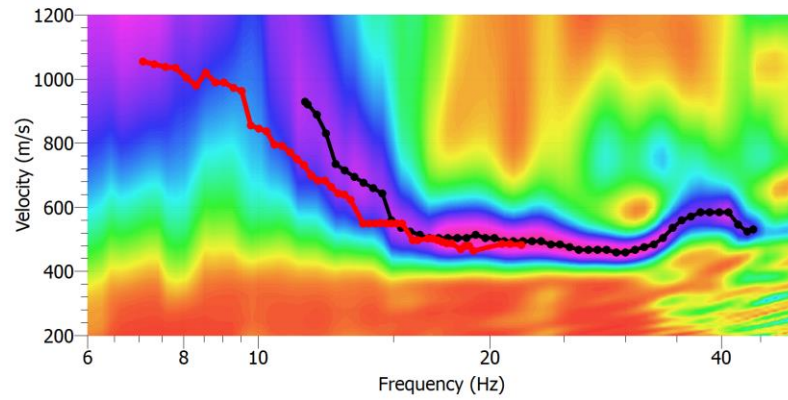
Dispersión Empírica
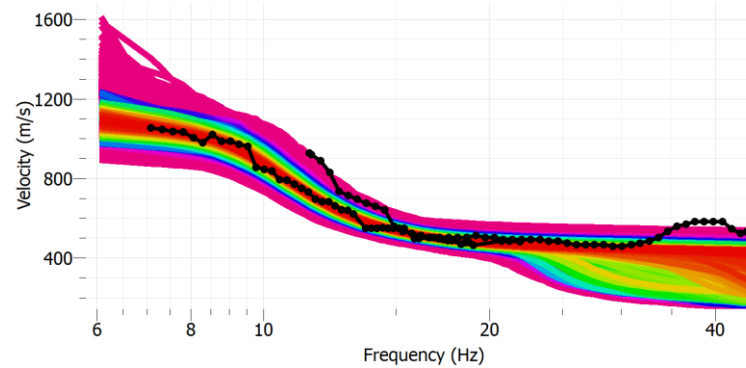
Dispersión Ajustada
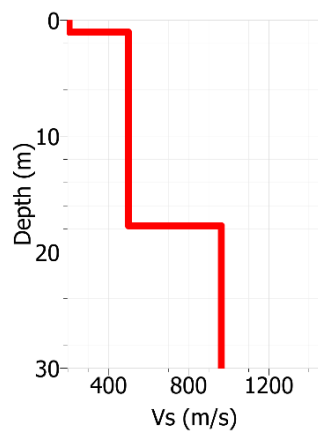
Hi (m)	Hf(m)	Vs (m/s)
0.0	1.1	212
1.1	17.8	503
17.8	30.0	965
	v_{s30}	586

Perfil Vs
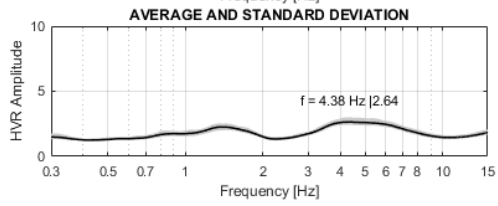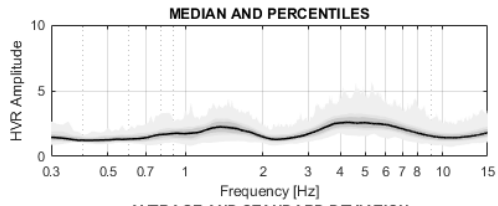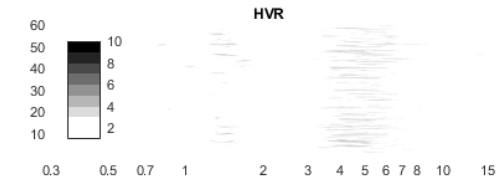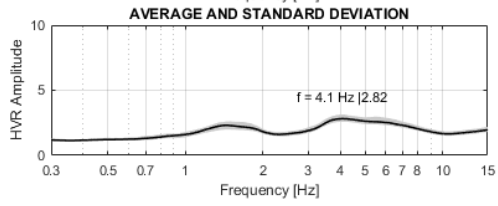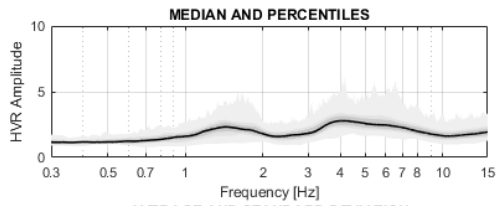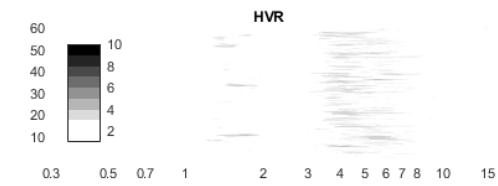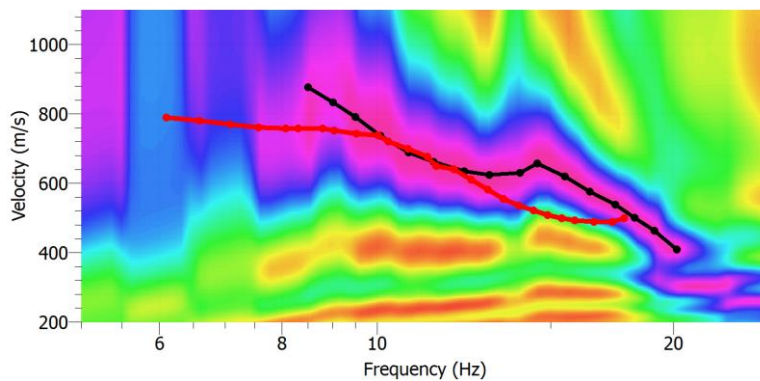
Melipilla – T288

Dispersión Empírica
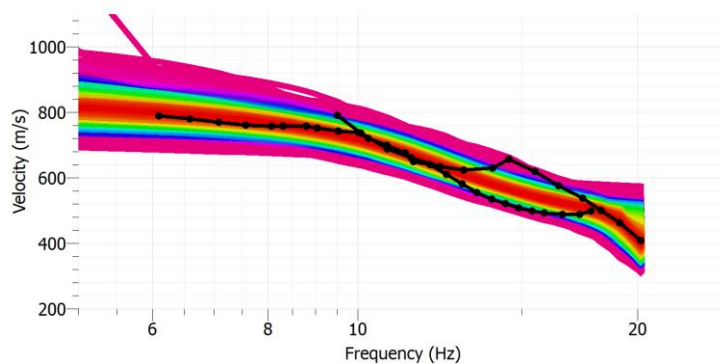
Dispersión Ajustada
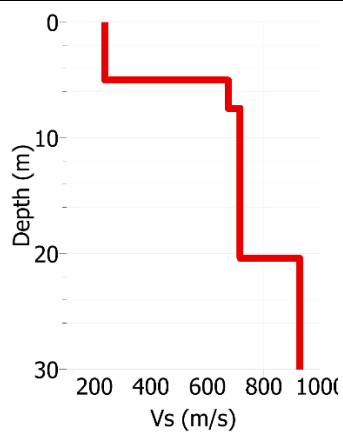
Hi (m)	Hf(m)	Vs (m/s)
0.0	5.0	239
5.0	7.5	679
7.5	20.4	716
20.4	30.0	926
	v_{s30}	564

Perfil Vs

HVR
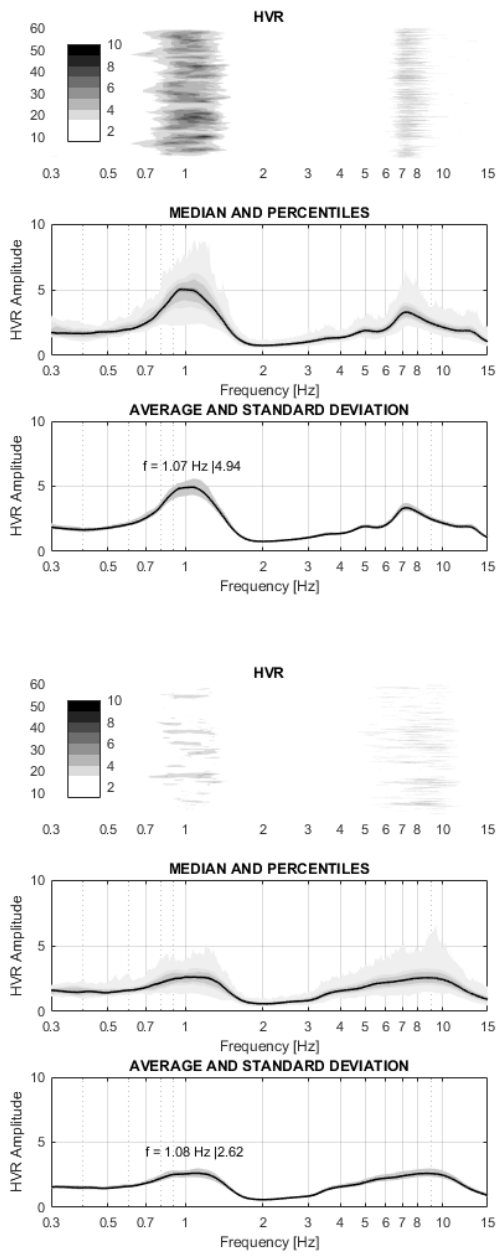
Melipilla – T289
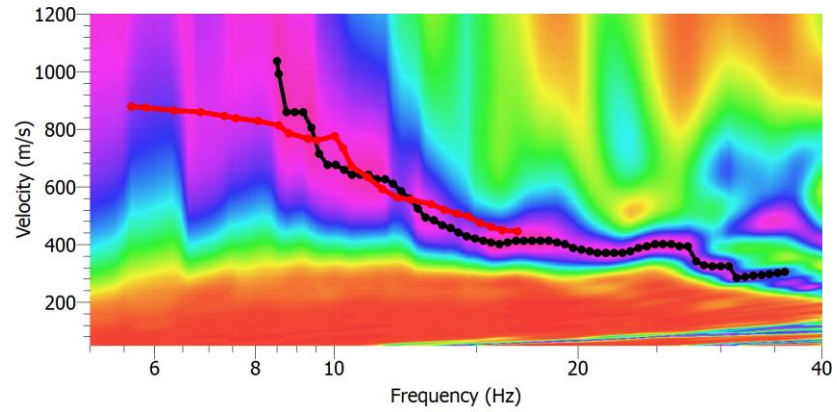
Dispersión Empírica
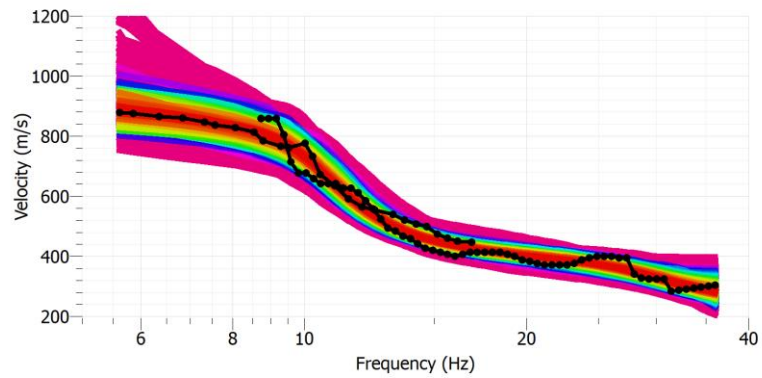
Dispersión Ajustada
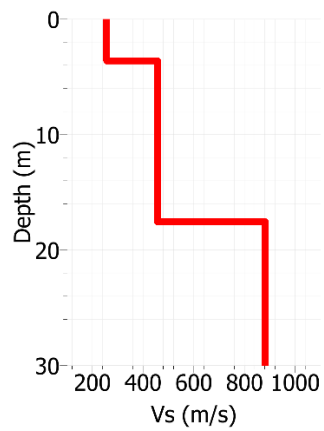
Hi (m)	Hf(m)	Vs (m/s)
0.0	3.6	257
3.6	17.5	457
17.5	30.0	882
	v_{s30}	511

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 1.32 Hz |4.52

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 1.53 Hz |3.33

HV
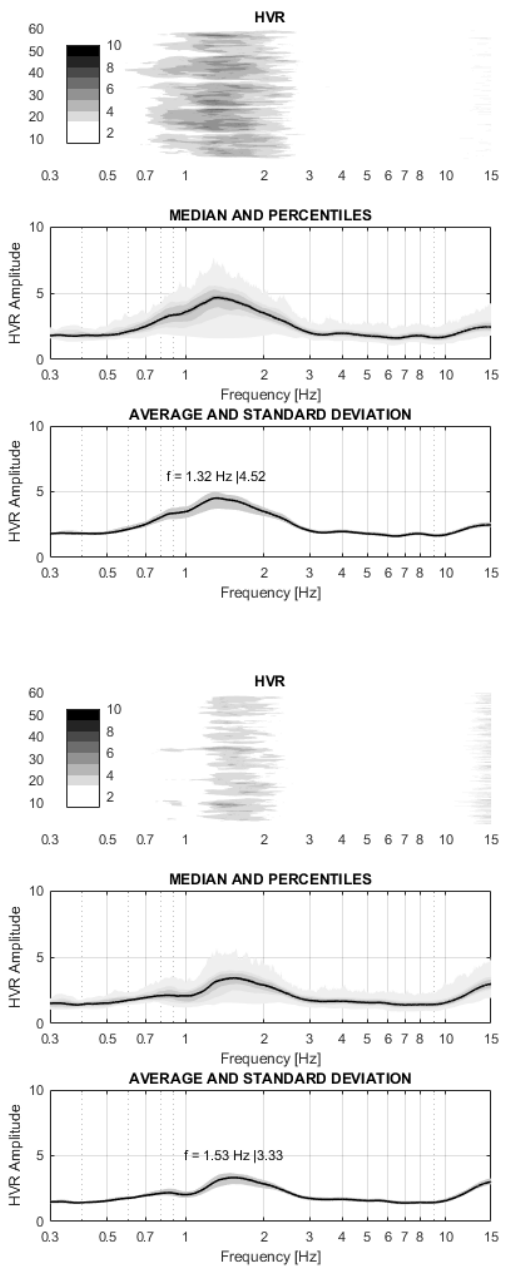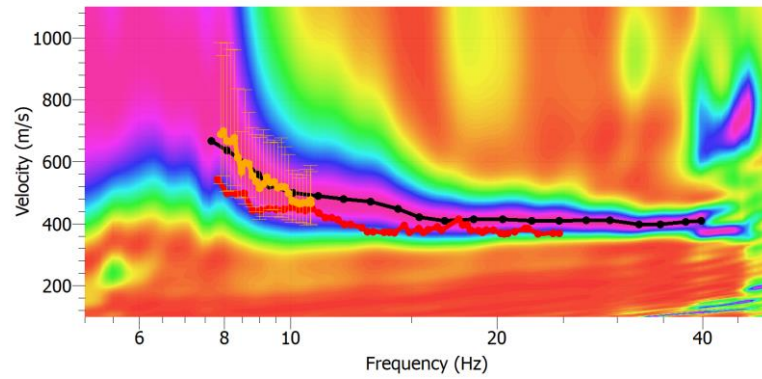
Curacaví – T290

Dispersión Empírica
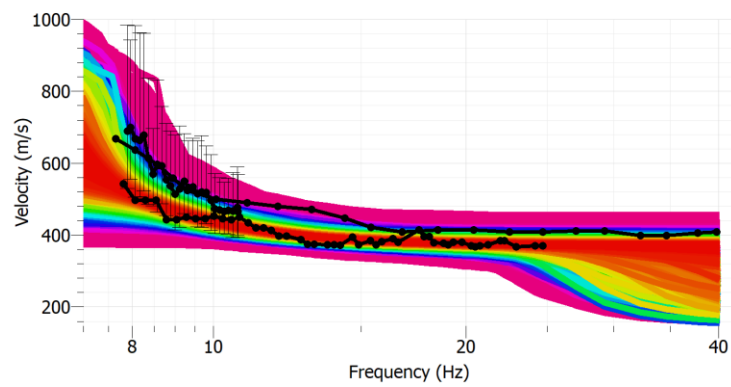
Dispersión Ajustada
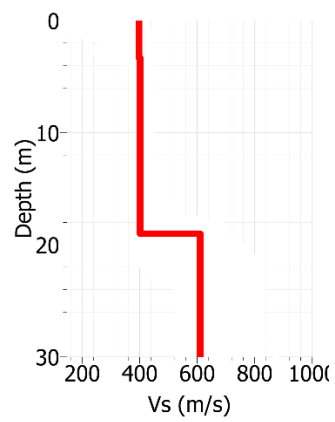
Hi (m)	Hf(m)	Vs (m/s)
0.0	19	402
19	30.0	610
	v_{s30}	456

Perfil Vs

HVR
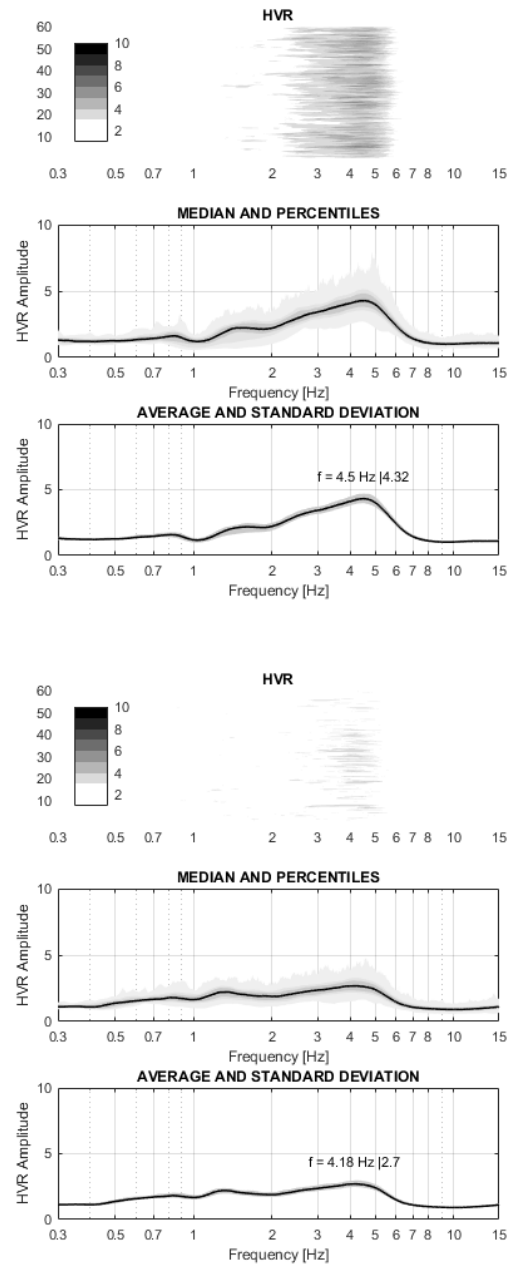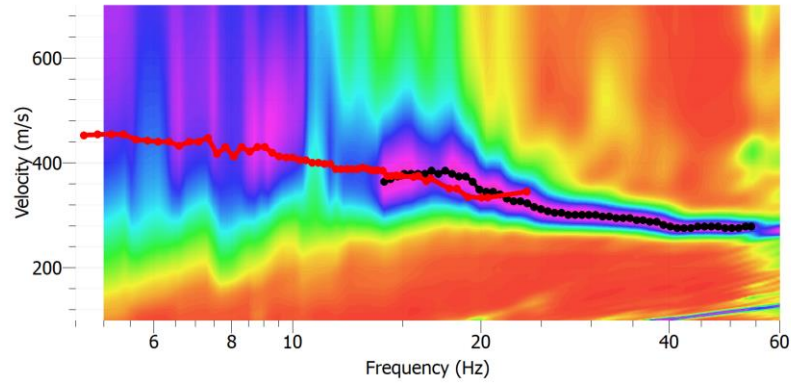
Curacaví – T291

Dispersión Empírica
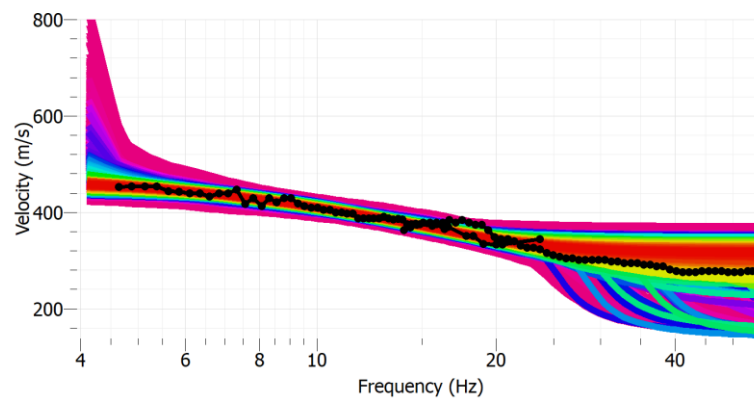
Dispersión Ajustada
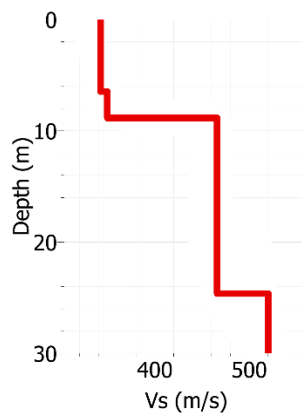
Hi (m)	Hf(m)	Vs (m/s)
0.0	6.6	342
6.6	8.9	350
8.9	24.7	468
24.7	30.0	522
	v_{s30}	429

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 1.5 Hz |5.29

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 1.48 Hz |3.52
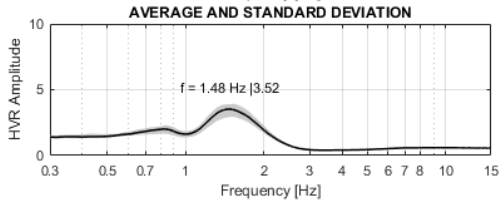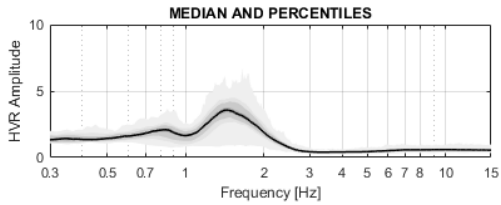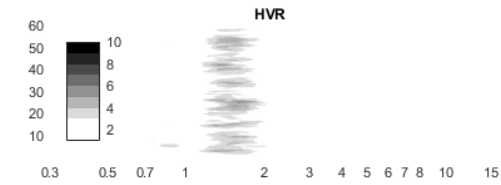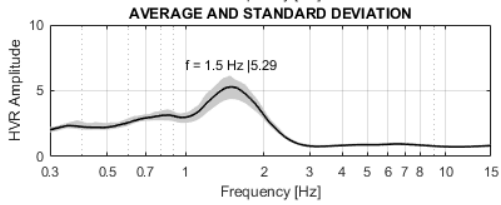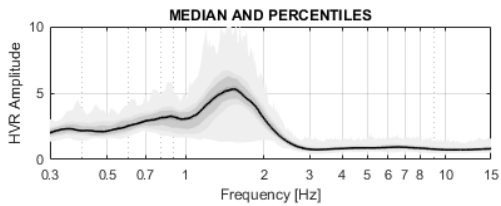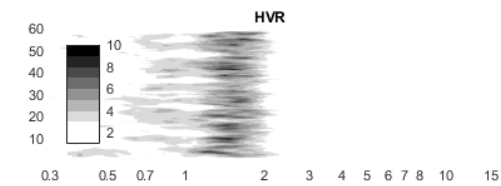
HV

Curacaví – T292
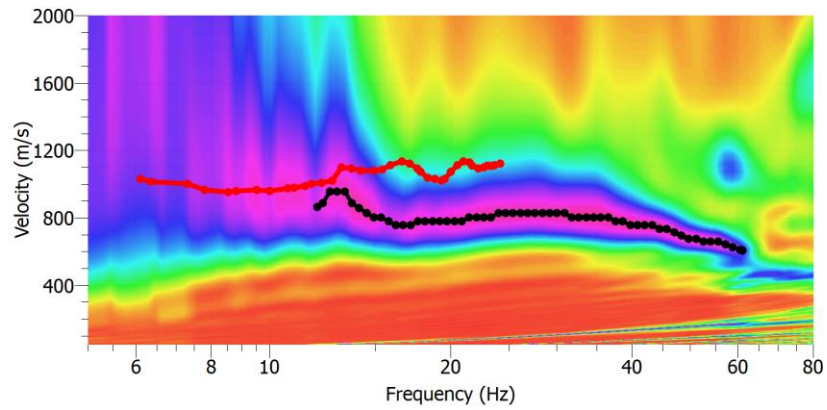
Dispersión Empírica
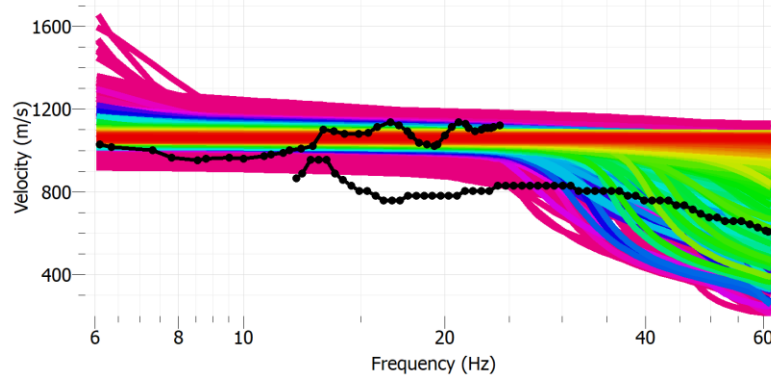
Dispersión Ajustada
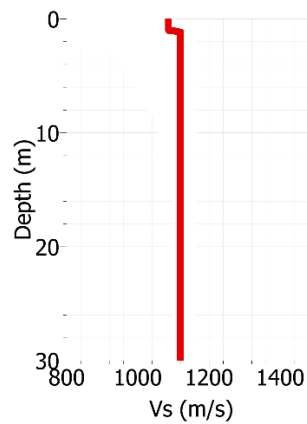
Hi (m)	Hf(m)	Vs (m/s)
0.0	1.1	1090
1.1	30.0	1122
	v_{s30}	1117

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 13.8 Hz |1.92
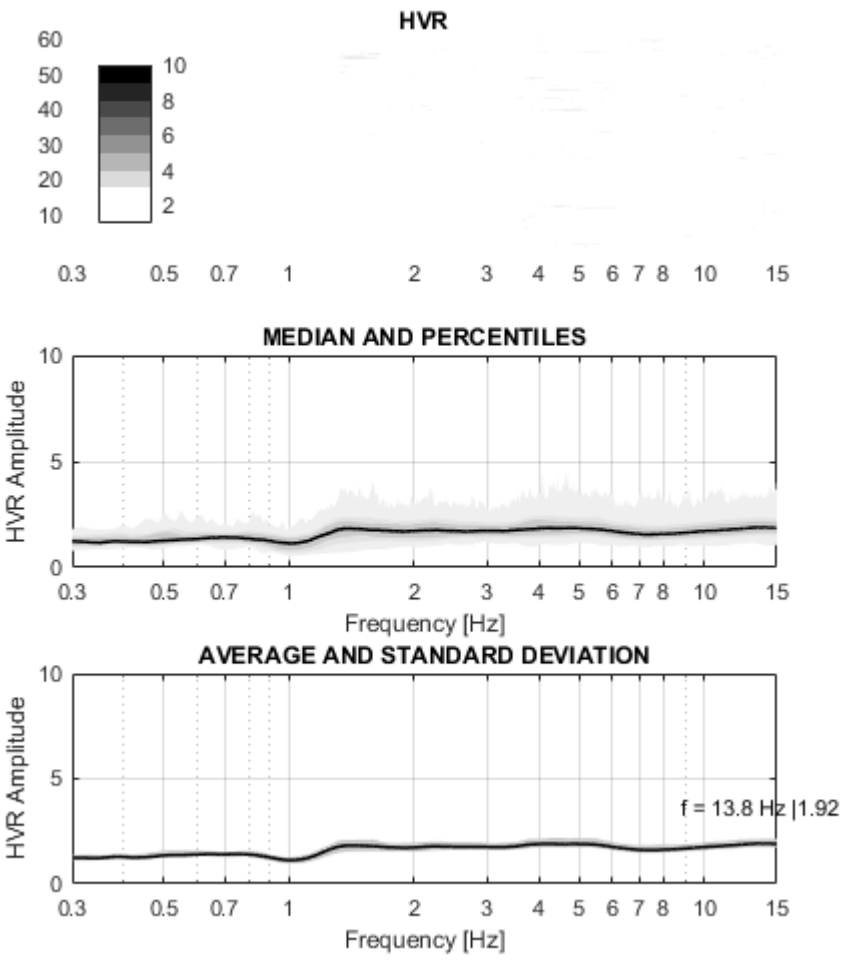
San José de Maipo – T308

Dispersión Empírica
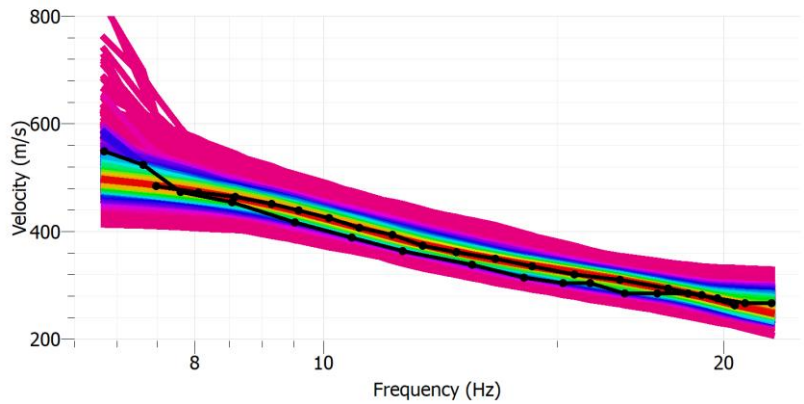
Dispersión Ajustada
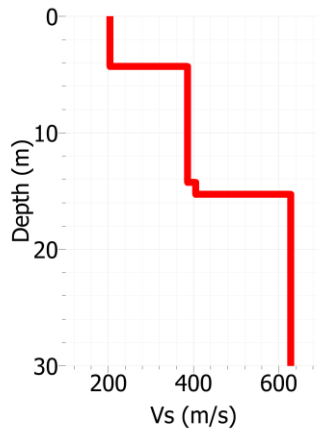
Hi (m)	Hf(m)	Vs (m/s)
0.0	4.2	206
4.2	14.2	385
14.2	15.3	408
15.3	30.0	631
	v_{s30}	409

Perfil Vs
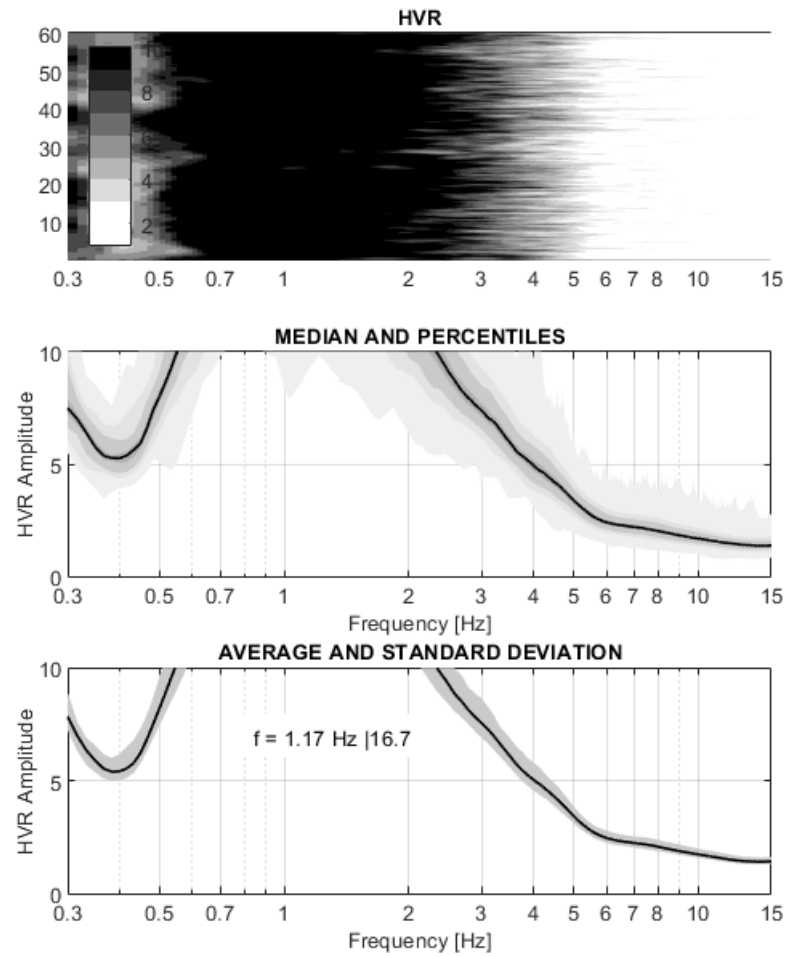
HVSR no concluyente.

HV

Lo Barnechea – T309
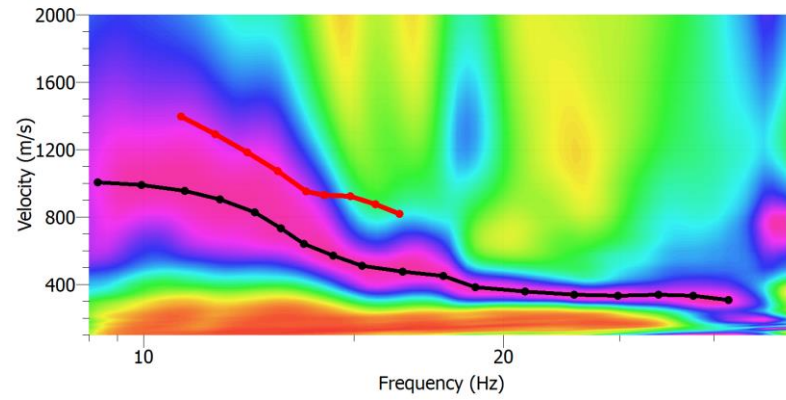
Dispersión Empírica
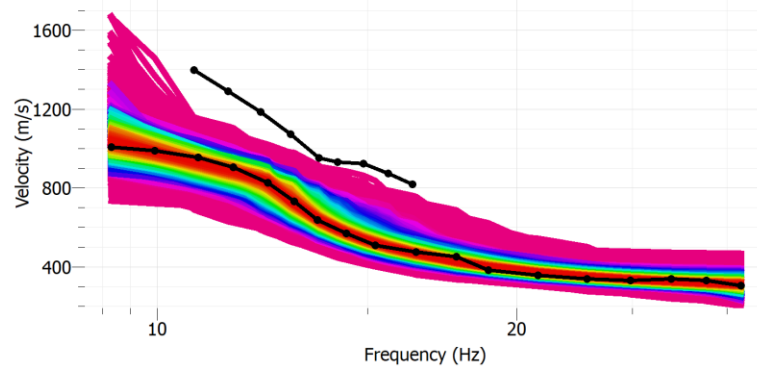
Dispersión Ajustada
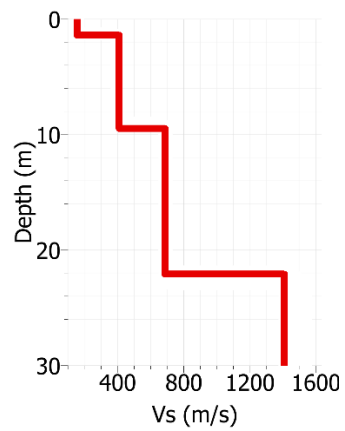
Hi (m)	Hf(m)	Vs (m/s)
0.0	1.4	163
1.4	9.5	418
9.5	22.2	691
22.2	30.0	1407
	v_{s30}	839

Perfil Vs
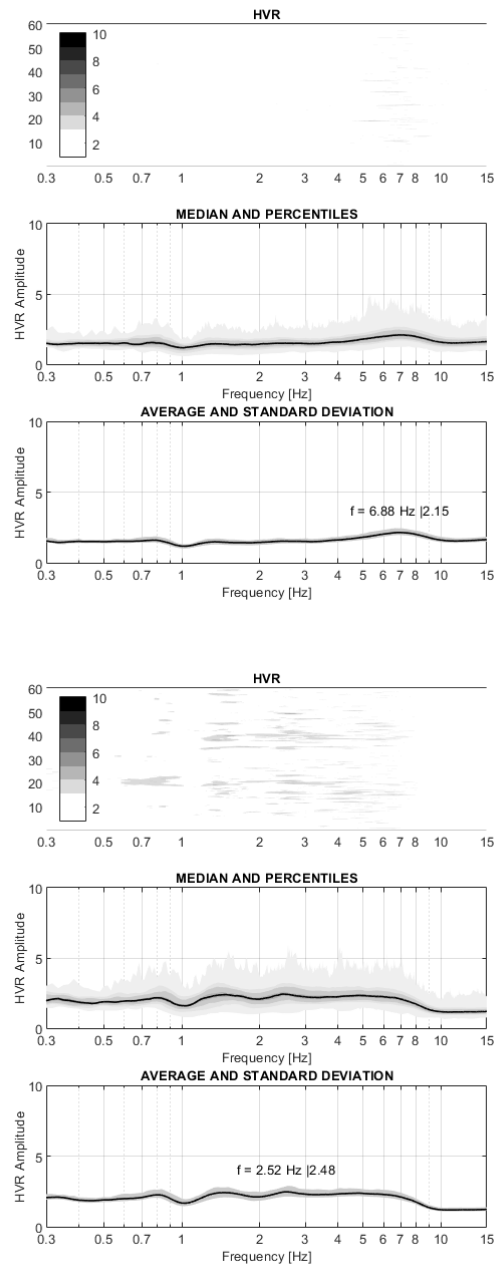
Lo Barnechea – T310

Dispersión Empírica
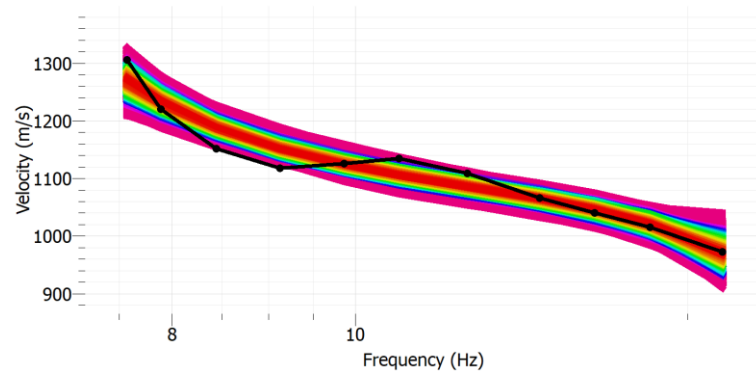
Dispersión Ajustada
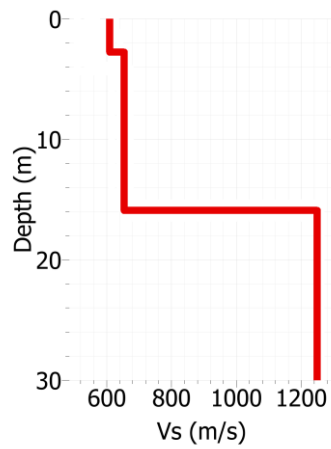
Hi (m)	Hf(m)	Vs (m/s)
0.0	2.8	610
2.8	16.0	654
16.0	30.0	1248
	v_{s30}	836

Perfil Vs
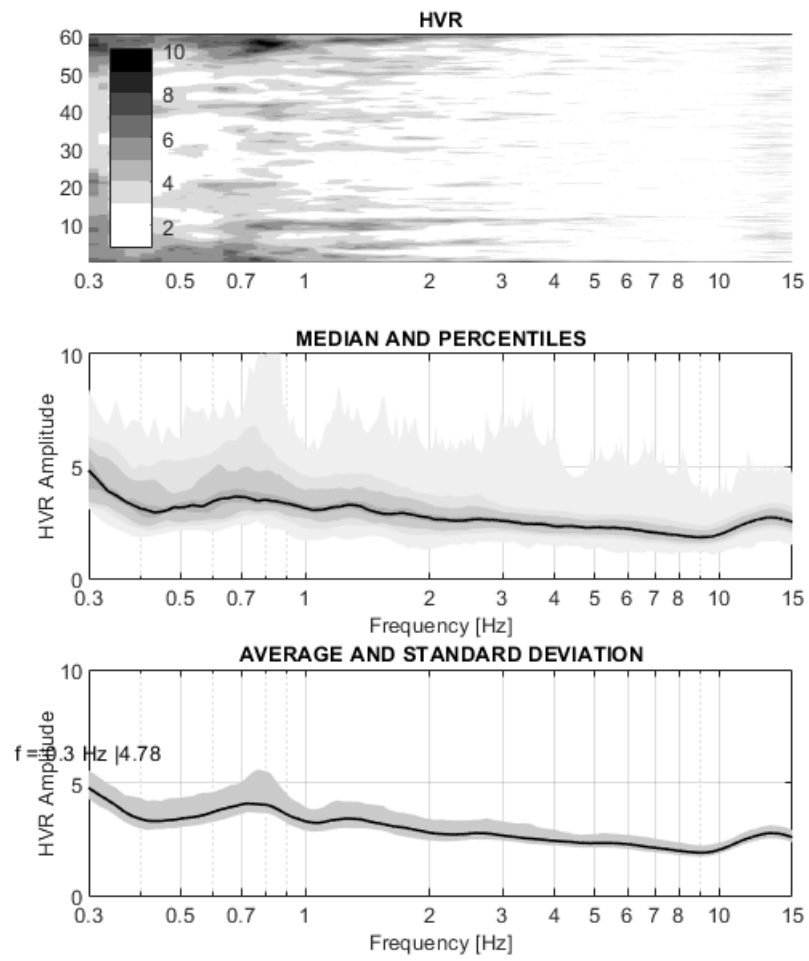
San José de Maipo – T311
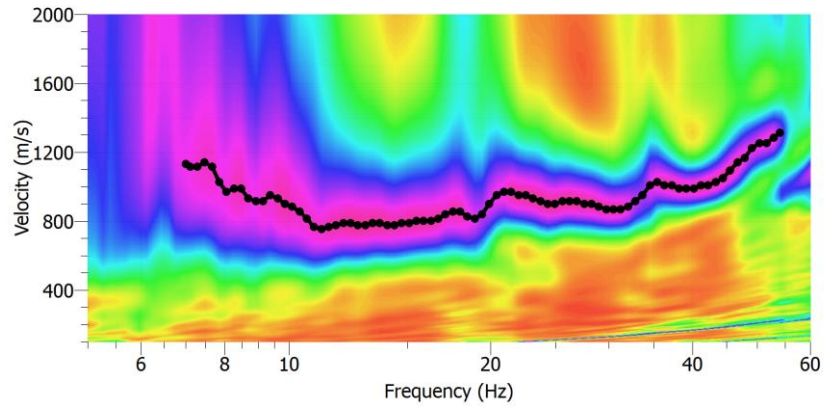
Dispersión Empírica
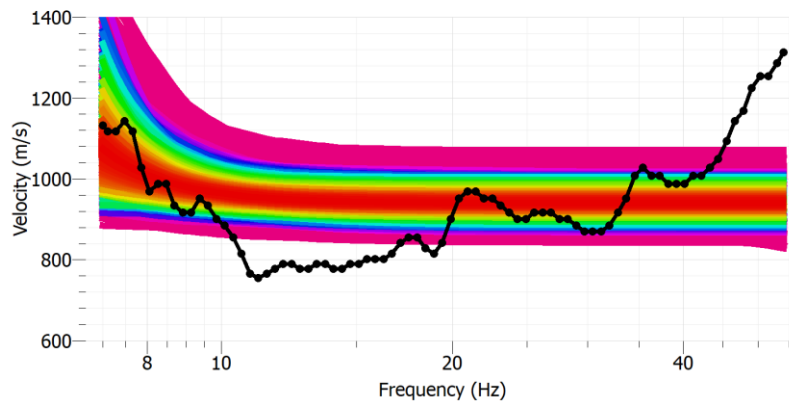
Dispersión Ajustada
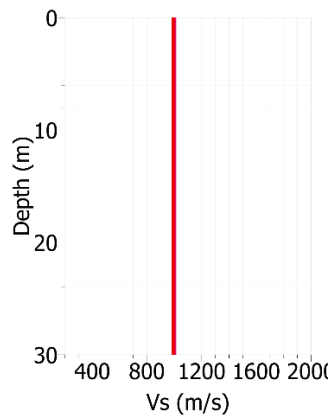
Hi (m)	Hf(m)	Vs (m/s)
0.0	30.0	993
	v_{s30}	993

Perfil Vs

HVR

MEDIAN AND PERCENTILES

AVERAGE AND STANDARD DEVIATION

f = 15 Hz |1.6
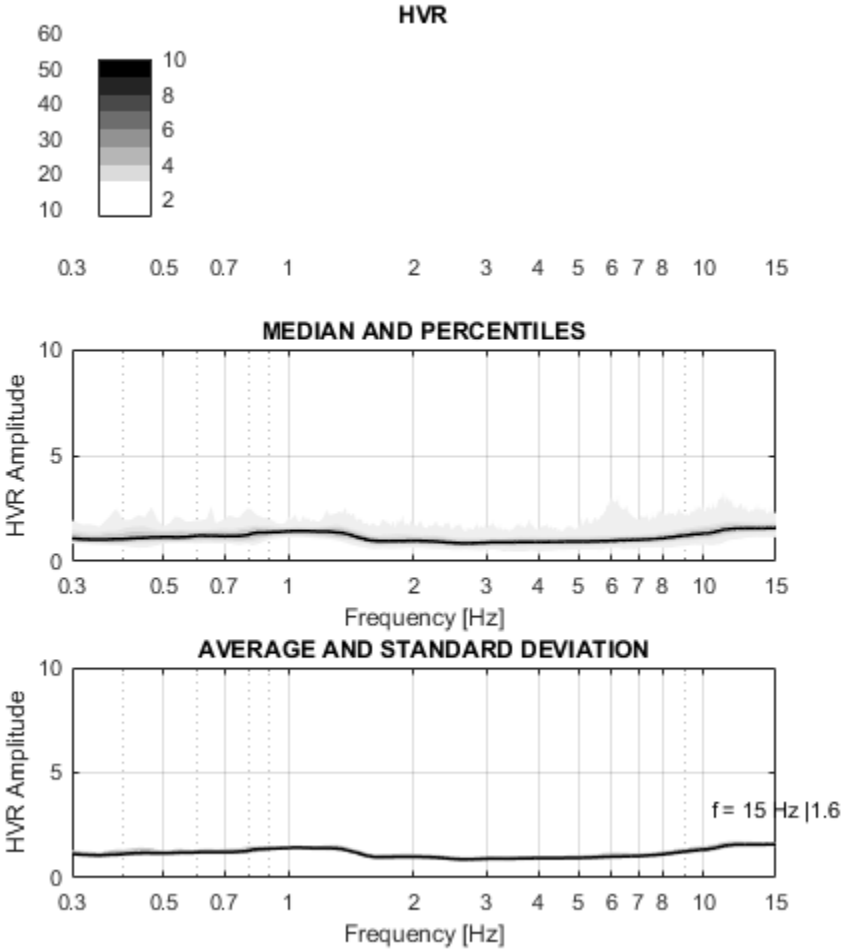
San José de Maipo – T312
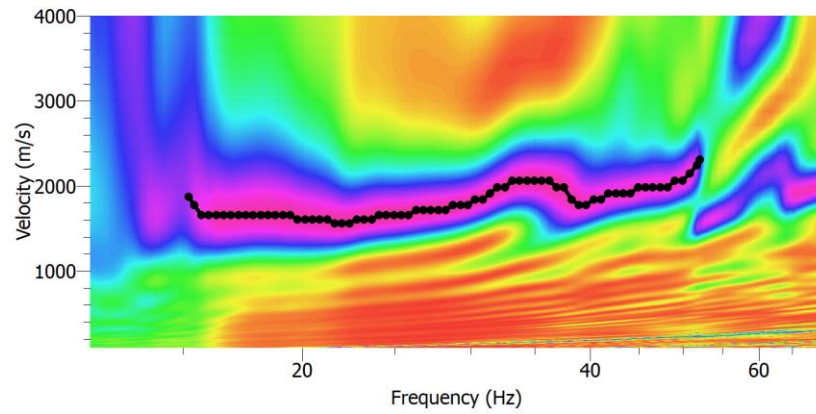
Dispersión Empírica
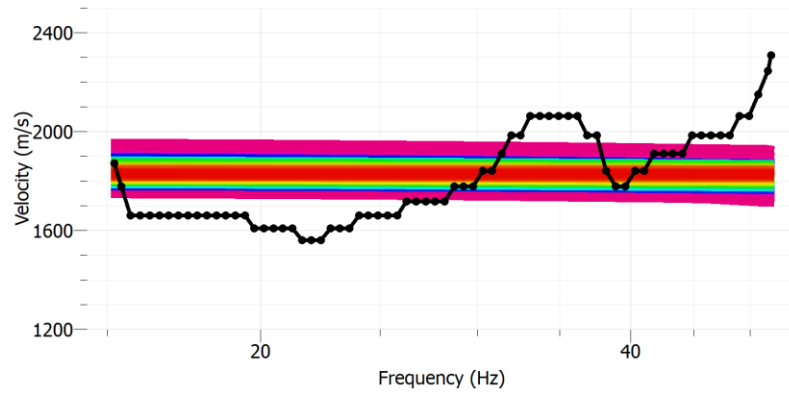
Dispersión Ajustada
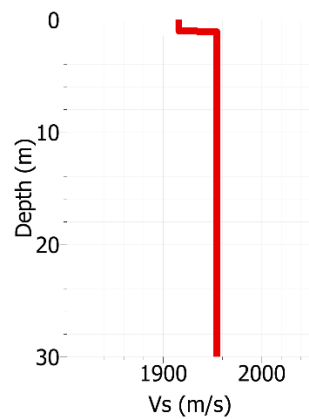
Hi (m)	Hf(m)	Vs (m/s)
0.0	1.2	1917
1.2	30.0	1956
	v_{s30}	1954

Perfil Vs
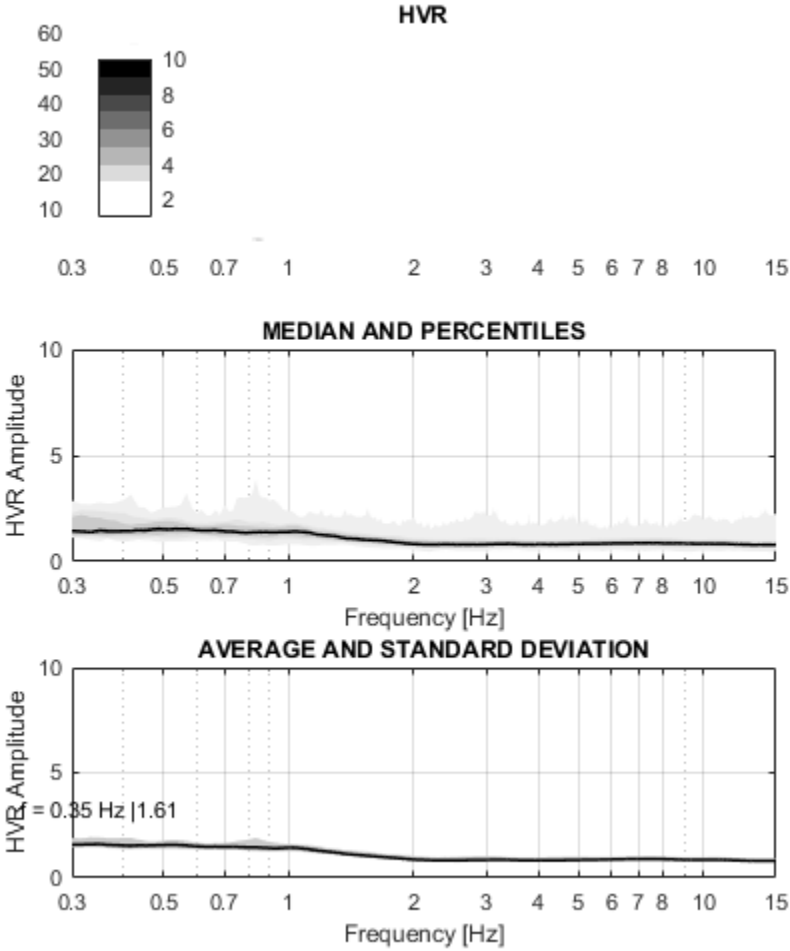
Renca – T368
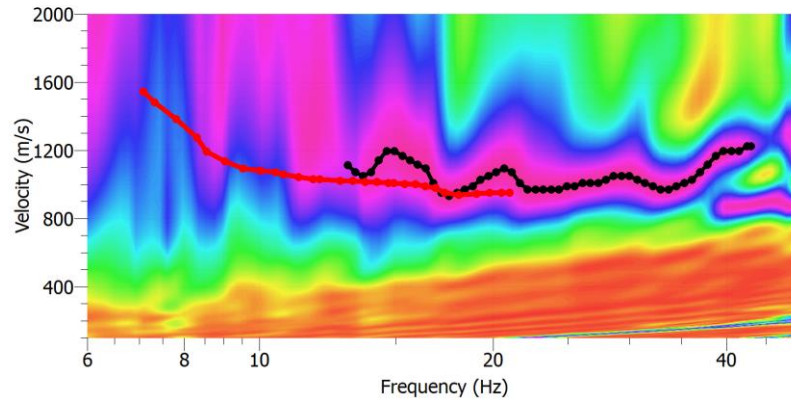
Dispersión Empírica
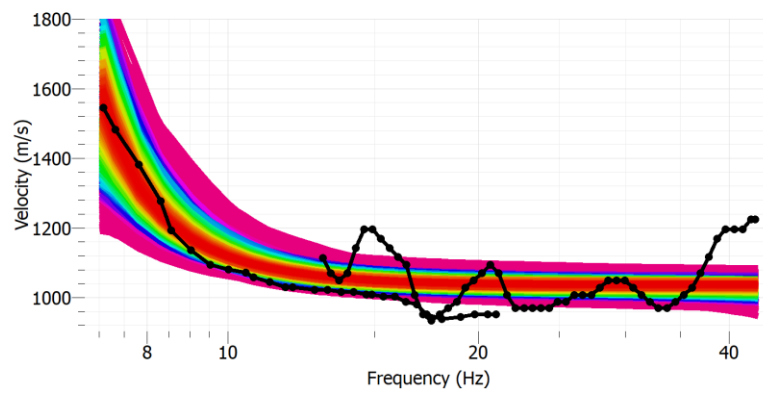
Dispersión Ajustada
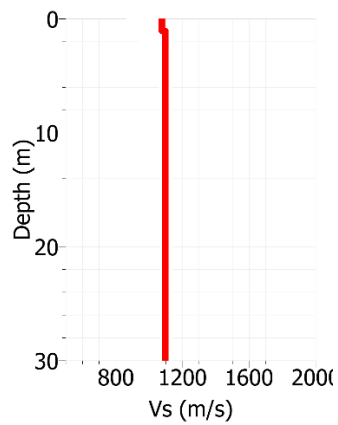
Hi (m)	Hf(m)	Vs (m/s)
0.0	1.0	1086
1.0	30.0	1093
	v_{s30}	1092

Perfil Vs
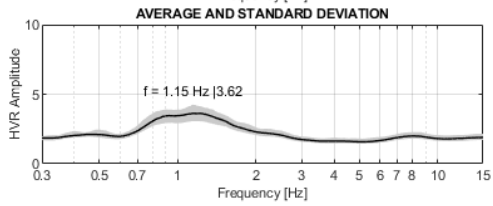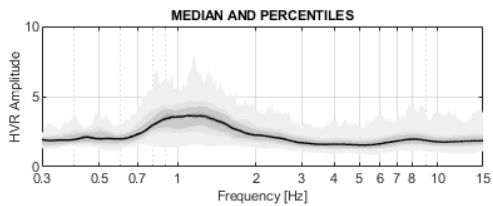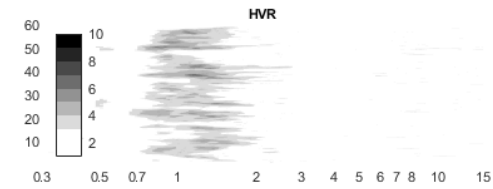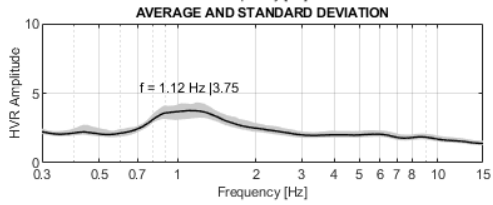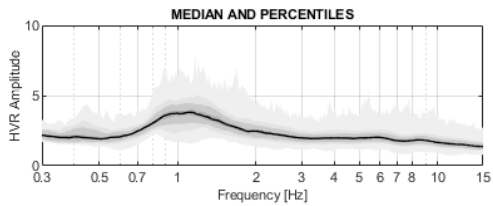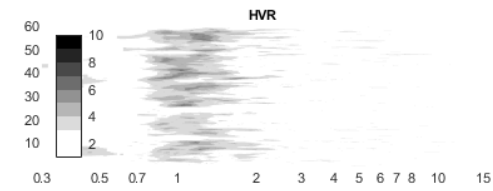
HV

Colina – T369
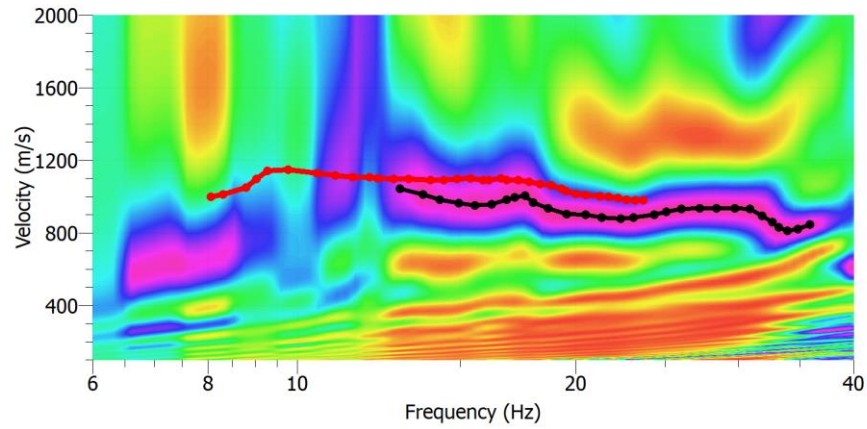
Dispersión Empírica
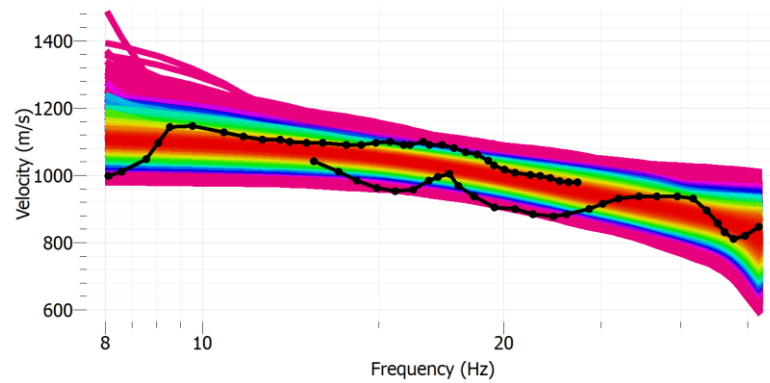
Dispersión Ajustada
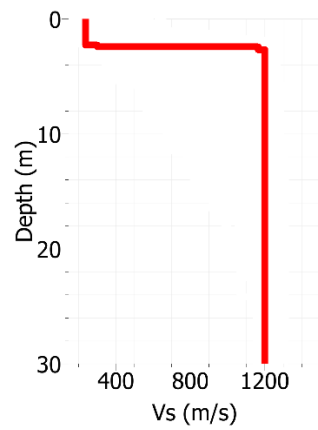
Hi (m)	Hf(m)	Vs (m/s)
0.0	1.2	243
1.2	30.0	1218
	v_{s30}	1050

Perfil Vs

HV