Freie Universität Berlin

GFZ Helmholtz Centre Potsdam

From Classical Earthquake Location to Deep Learning-based Phase Detection and Picking using Seismic Data from the Northern Chile Subduction Zone

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fachbereich Geowissenschaften

an der Freien Universität Berlin

vorgelegt von

Hugo Enrique Soto Parada

Berlin, 2021

Erstgutachter: **Prof. Dr. Onno Oncken**
Freie Universität Berlin, GeoForschungsZentrum Potsdam

Zweitgutachter: **Prof. Dr. Frederik Tilmann**
Freie Universität Berlin, GeoForschungsZentrum Potsdam

Prüfungskommission:

***Vorsitzende*: Prof. Dr. Serge A. Shapiro**
Freie Universität Berlin

Prof. Dr. Onno Oncken
Freie Universität Berlin, GeoForschungsZentrum Potsdam

Prof. Dr. Frederik Tilmann
Freie Universität Berlin, GeoForschungsZentrum Potsdam

Dr. Bernd Schurr
GeoForschungsZentrum Potsdam

Prof. Dr. Timm John
Freie Universität Berlin

Prof. Dr. Andreas Rietbrock
Karlsruher Institut für Technologie

eingereicht: 30.10.2020
verteidigt: 26.02.2021

Utopia is on the horizon.
I move two steps closer; it moves two steps further away.
I walk another ten steps and the horizon runs ten steps ahead.
No matter how much I may walk, I will never reach it.
What, then, is the purpose of utopia?
It is to cause us to keep walking.

Eduardo Galeano

Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Literatur angefertigt wurde. Die Stellen der Arbeit, die anderen Werken wörtlich oder inhaltlich entnommen sind, wurden durch entsprechende Angaben der Quellen kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Berlin, October 2020

Abstract

Recent development in seismological instrumentation has made possible the deployment of temporary and permanent seismic networks with large numbers of seismometer stations. These networks register an extensive amount of seismic data that is then available to the scientific community for carrying out different types of seismological studies. Producing reliable earthquake catalogs, based on the implementation of efficient automatic event location procedures, which can handle such extensive amount of seismic data, is one of the critical steps prior to perform most studies. Precise and maximally complete event catalogs become essential in regions on the Earth where tectonic plate subduction takes place, such as the Chilean convergent margin. In subduction zones, seismicity analysis can help to shed light on the different tectonic processes involved in the generation of great tsunamigenic megathrust earthquakes, which are usually devastating for the inhabitants of these hazardous regions. Particularly, insights on the conditions characterizing the frictional behavior of a subduction interplate fault, can be investigated in detail from seismicity. This ultimately leads to a better assessment of the earthquake hazard potential in a subduction zone.

In my thesis, I apply a multistage automatic earthquake detection and location workflow to generate a high-resolution earthquake catalog for the northern Chile region. This automatic workflow is applied to seismic data recorded by the Integrated Plate Boundary Observatory Chile (IPOC) permanent seismic network, as well as by a number of complementary temporary deployments, adding up to >100 seismic stations. The resulting event catalog contains ∼19,000 foreshocks, aftershocks and background seismicity occurring in the time interval between one month preceding and nine months following the 1 April 2014 M8.1 Iquique earthquake. I analyse the seismicity features, in combination with modelled coseismic slip and modelled static stress changes, as well as geodetically-derived afterslip and interseismic locking models, in order to investigate the small-scale frictional heterogeneities on the plate interface and interpret the seismotectonic behavior of the overlying continental forearc in northern Chile.

Results from this analysis point towards a primarily along-dip segmentation of the frictional behavior along the plate interface. In the downdip direction, an aseismic velocity-strengthening segment of the interface and a frictionally transitional region underlay, respectively, an aseismic frontal prism and a transitional zone in the outer continental wedge. Deeper on the plate interface, the seismogenic segment coincides with the highest coseismic slip underlying the inner continental wedge. The velocity-weakening segment connects downdip to a frictionally heterogeneous region where aftershock seismicity (interpreted as conditionally stable) and highest afterslip patches (velocity-strengthening) anticorrelate. A sparsity of events, which may be a generic feature along the Chilean subduction, separates this heterogenous segment from the

deepest interplate events, where presumably frictionally transitional behaviour predominates.

With the aim of improving the efficiency and simplifying the implementation of the earthquake catalog generation workflow, I additionally present DeepPhasePick, an automatic method entirely based on systematically optimized deep neural networks that carries out the first stages involved in this workflow: detection and picking of P and S phases originating from local earthquakes. DeepPhasePick makes use of a convolutional neural network architecture to perform the phase detection on three-component seismograms. It then uses two recurrent neural networks to conduct the phase picking on the vertical component for P phases and on the two-horizontal components for S phases. The phase time onsets and their corresponding onset uncertainties returned by the phase picking stage can feed a phase associator algorithm in the next step of the earthquake location workflow.

DeepPhasePick architectures are optimized and trained using >30,000 manually-picked seismic records extracted from two sets of event waveforms occurring ∼7 and ∼17 years before the 2014 M8.1 Iquique earthquake, in a region of northern Chile that partially overlaps with the area covered by the automatically-derived ∼19,000 earthquake catalog described above. The algorithm is then tested on different test sets: a set of manually-picked records independent from the training set, >1,000,000 automatically-picked records taken from the ∼19,000 earthquake catalog, and a set of >200,000 automatically-picked records taken from another recently published earthquake catalog for the northern Chile region. Results from these tests show that DeepPhasePick is able to detect seismic phases with high accuracy, as well as to predict phase time onsets with a precision comparable to the more conventional phase picking methods, be they manual or automatic. Moreover, I demonstrate that DeepPhasePick's detection and picking abilities perform effectively not only on the largely lower-seismic noise data recorded in northern Chile, but also generalize to higher-seismic noise data recorded from a different tectonic regime in an urban region of Albania.

Zusammenfassung

Durch die fortlaufende Weiterentwicklung seismologischer Sensoren ist es heute möglich, temporär oder dauerhaft installierte seismische Netzwerke bestehend aus vielen Einzelsensoren aufzubauen. Diese Netzwerke akkumulieren große Datenmengen, welche sodann der seismologischen Community zur Verfügung stehen, um mit ihnen verschiedenste Studien durchzuführen. Um solche Studien mithilfe dieser umfangreichen Datenmengen in Angriff nehmen zu können, bedarf es der Erstellung akkurater Bebenkataloge mittels effizienter automatisierter Erdbeben-Lokationsmethoden. Solche präzise lokalisierten und möglichst kompletten Bebenkataloge sind vor Allem in Regionen der Erde essentiell, wo die Subduktion einer tektonischen Platte auftritt, so zum Beispiel entlang des Chilenischen Kontinentalrands. Seismizitätsanalyse in Subduktionszonen kann eine wichtige Rolle bei der Erforschung der verschiedenen aktiven Prozesse spielen, die große Erdbeben entlang der Megathrust-Verwerfung hervorrufen. Solche Beben, die wiederum oft Tsunamis auslösen, haben oft katastrophale Auswirkungen für die Bewohner der betroffenen Gegenden. Eine detaillierte Untersuchung der örtlichen Seismizität kann insbesondere zu Erkenntnissen über die Reibungseigenschaften der Plattengrenze und den Bedingungen, von denen diese abhängen, führen.

In meiner Dissertationsschrift stelle ich die Anwendung eines mehrstufigen, automatisierten Bebendetektions- und Lokationsablaufs vor, der zur Erstellung eines hochaufgelösten Bebenkatalogs für Nordchile benutzt wurde. Dieser automatische Workflow wurde mit Daten angewendet, die vom permanenten seismischen Netzwerk des IPOC-Konsortiums (IPOC = Integrated Plate Boundary Observatory Chile) sowie von einigen zusätzlichen temporär installierten Netzwerken aufgezeichnet wurden, so dass insgesamt mehr als 100 Stationen verwendet wurden. Der erstellte Bebenkatalog enthält etwa 19,000 Beben, darunter viele Vor- und Nachbeben des Iquique-Bebens vom 1. April 2014 (Momentenmagnitude 8.1) sowie Hintergrundaktivität, Alles in einem Zeitraum von einem Monat vor bis neun Monate nach dem Iquiquebeben. Ich werte die Seismizitätsmerkmale zusammen mit Modellierungsergebnissen des coseismischen Versatzes und der Spannungsänderungen durch das Beben aus, und beziehe auch geodätisch ermittelte Modelle von Afterslip und der interseismischen Kopplung zwischen den Platten mit in die Analyse ein. So kann ich die kleinskaligen Heterogenitäten in der Reibungsfestigkeit der Plattengrenze untersuchen und die seismotektonischen Charakteristika der darüber liegenden Region des kontinentalen Forearcs Nordchiles interpretieren.

Die Ergebnisse dieser Forschungstätigkeit zeigen eine Segmentierung in den Reibungeigenschaften der Plattengrenze primär in Richtung des Einfallens der abtauchenden Platte. Mit ansteigender Tiefe entlang der Plattengrenze treten dort zunächst eine aseismische Region auf, deren Rheologie velocity-strengthening ist, gefolgt von einer Region die als transitional beze-

ichnet werden kann. Diese befinden sich unterhalb eines aseismischen Frontal Prisms sowie des äußeren kontinentalen Keils, der konditional stabile Reibungseigenschaften aufweist. In größerer Tiefe befindet sich die seismogene Zone der Plattengrenze, die die größten coseismischen Versätze aufweist und unterhalb des inneren kontinentalen Keils liegt. Auf dieses Segment, das velocity-weakening aufweist, folgt eine heterogene Region, in der Nachbebenaktivität und die höchsten Afterslip-Werte räumlich antikorreliert zu beobachten sind. Darunter liegt eine Region deutlich verminderter Bebenaktivität, die möglicherweise überall entlang des chilenischen Kontinentalrands auftritt, sowie schlussendlich die tiefsten Interplattenbeben, die wohl in einer Tiefenregion mit transitionalen Reibungseigen-schaften aufreten.

Zusätzlich stelle ich DeepPhasePick vor, eine neue automatisierte Methode zum Picken der Ersteinsätze von P- und S-Wellen lokaler Beben, die komplett auf systematisch optimierten neuronalen Netzwerken beruht. Das Ziel dieser Methode ist es, die Effizienz der automatisierten Bebenkatalogsproduktion zu verbessern sowie diese Prozedur zu vereinfachen. DeepPhasePick benutzt tiefe neuronale Netzwerkarchitektur, um ankommende Phasen in Drei-Komponenten-Seismogrammen zu detektieren. Daraufhin verwendet es zwei rekurrente neuronale Netze, um die P-Phase auf der Vertikalkomponente und die S-Phase auf beiden Horizontalkomponenten zu picken. Die Ersteinsatzzeiten sowie ihre Unsicherheiten, die vom Programm ausgegeben werden, können sodann in einen Phasen-Assoziierungs-Algorithmus eingespeist werden, der den nächsten Schritt im Workflow der automatisierten Bebenlokation darstellt.

Die Architektur von DeepPhasePick wird optimiert und trainiert mithilfe von >30,000 handgepickter Phasenankünfte, die aus zwei Datensätzen aus Nordchile extrahiert wurden, die Zeiträume ca. 7 und ca. 17 Jahre vor dem Iquiquebeben umfassen. Örtlich überlappen die Regionen dieser Datensätze mit dem oben beschriebenen Bebenkatalog, der ca. 19,000 automatisch ermittelte Bebenlokationen umfasst. Der Algorithmus wird dann mit verschiedenen Datensätzen getestet: mit einem Satz handgepickter Datenspuren, der unabhängig vom Trainingsdatensatz ist, mit >1,000,000 automatisch gepickter Spuren aus dem oben beschriebenen, etwa 19,000 Beben umfassenden Katalog, sowie mit einem weiteren Satz von >200,000 automatisch gepickter Spuren aus einem anderen, kürzlich publizierten Bebenkatalog aus der Region Nordchile. Die Ergebnisse dieser Tests zeigen, dass DeepPhasePick seismische Phasen sehr akkurat detektiert sowie die Ersteinsätze mit einer Präzision bestimmt, die derer konventioneller Phasenpick-Methoden (seien sie automatisch oder manuell) gleichkommt. Darüber hinaus zeige ich, dass die Detektions- und Pickingfähigkeiten von DeepPhasePick nicht nur für die Daten aus Nordchile, die sich durch sehr niedriges Hintergrundrauschen auszeichnen, gut funktionieren. Auch Daten mit höherem Noise-Niveau und aus einem gänzlich anderen tektonischen Kontext, nämlich aus einer urbanen Region in Albanien, werden von DeepPhasePick präzise ausgewertet.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

On 1 April 2014, the M8.1 Iquique earthquake broke a central piece of the northern Chile-southern Peru seismic gap, generating a tsunami of ~2 m height along the northern Chile coast.

In the first part of this thesis I use data from >100 permanent and temporary seismic stations covering the region of the Iquique earthquake sequence and, implementing classical commonly used automatic seismic phase detection and earthquake location procedures, I generate a catalog of ~19,000 earthquakes. I use this dataset to analyze the seismicity patterns related to the Iquique earthquake sequence and to investigate how this seismicity can give some clues on the heterogeneous frictional behavior exhibited by the subduction plate boundary in this region.

In the second part of this thesis I develop DeepPhasePick, an algorithm entirely based on deep neural networks, which I optimize and train for automatic detection and picking of seismic phases originating from local earthquakes. DeepPhasePick aims at providing a more efficient phase detection and picking method as an alternative to classical approaches, such as the one I apply in the first part of this work.

In the next section I give a brief overview of the tectonic processes taking place in subduction zones, particularly in the northern Chile convergent margin. In section 1.2, I present the research questions of this thesis and how I address them in the following chapters. Chapter 2 introduces the classical earthquake location and seismic phase detection methods which I use in this work and presents the key ideas behind deep learning which I leverage to develop the DeepPhasePick algorithm.

1.1 Essential Concepts

1.1.1 Subduction Zone

Plate tectonics theory explains the motion of the different rigid tectonic plates, formed by oceanic and thicker continental lithospheres (crust + uppermost solid mantle), on top of the mechanically weaker asthenosphere (e.g., Turcotte and Schubert, 2014). This theory has been

accepted since the seafloor spreading process was proposed about 60 years ago, and has since incorporated the preceding idea of continental drift, proposed by Alfred Wegener in the early 20th century.

The area where two tectonic plates meet is called the plate boundary and can be of convergent, divergent or transform type, depending on whether the plates move towards, away from, or slide past each other. Convergent margins form either continental collisional zones (e.g., Indian-Eurasian plates collision) or subduction zones (e.g., Nazca-South American or Pacific-North American plates subduction). Divergent boundaries are places where either new oceanic crust is being formed (e.g., the Mid-Atlantic Ridge) or new ocean basins are formed by continental rifting (e.g., Africa's East African Rift). Transform boundaries can exhibit right-lateral (e.g., San Andreas Fault in USA, or North Anatolian Fault in Turkey) or left-lateral (e.g., East Anatolian Fault) relative plate motion.

Most of the seismic moment of the Earth (~95%) is released in subduction zones. At these margins, oceanic lithosphere of a subducting plate descends beneath the continental or oceanic lithosphere of an overriding plate (Turcotte and Schubert, 2014). The sporadic sudden stress releases generated at the plate interface by this process have given rise to the largest ever recorded earthquakes on Earth, such as the 1960 M~9.5 Valdivia (Chile), the 1964 M9.2 Alaska (USA), the 2004 M~9.2 Sumatra (Indonesia), and the 2011 M9.1 Tohoku (Japan) megathrust earthquakes (e.g., Hayes et al., 2011; Kanamori, 1977; Lay and Bilek, 2007; Pacheco and Sykes, 1992; Wang and Liu, 2006).

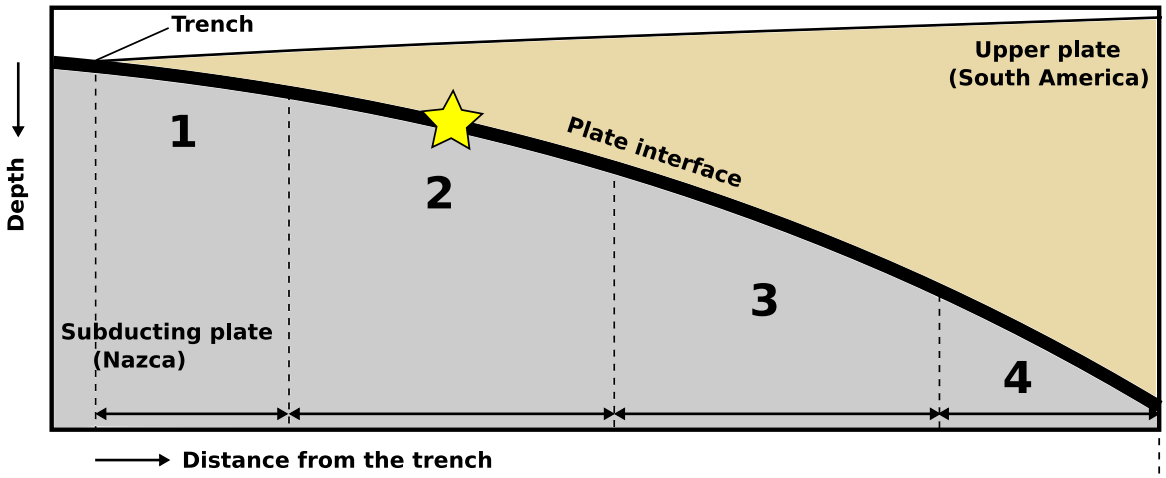

Figure 1.1: Non-scaled sketch of slip mode segmentation on the Chilean subduction plate interface, adapted from Lay et al. (2012). It has been suggested that along dip the plate interface features a shallowest portion (**1**) which slide aseismically and where large tsunamigenic earthquakes can eventually propagate. Further downdip, an intermediate region (**2**) where large megathrust earthquakes are expected to occur is followed by a segment (**3**) where condition stability can generate moderate slip. The deepest interface region (**4**) may slide primarily aseismically and feature slow slip, which may rupture small seismic patches that populate this region.

The accelerated slip on the plate interface resulting from a megathrust earthquake can last from seconds to a few minutes and corresponds to the coseismic stage of the seismic cycle in a subduction zone. In the postseismic stage that follows, the plate interface experiences a transient deformation induced by the relaxation of coseismic stress changes. The interplate fault deformation can occur seismically (as aftershocks), aseismically (as *afterslip*), or in a

coupled aseismic-seismic process where afterslip may drive aftershock seismicity that decreases over time (Hsu et al., 2006; Marone et al., 1991; Perfettini et al., 2010). Following and usually overlapping the postseismic period, the interseismic stage of the seismic cycle can take decades or even centuries, during which time elastic strain is accumulated once again on the plate interface before a new earthquake ruptures (e.g., Aki, 1979; Moreno et al., 2011).

The exact mode in which the plate interface slips during a megathrust earthquake is intrinsically related to the heterogeneous frictional behavior along the subduction megathrust (Figure 1.1; e.g., Lay et al., 2012; Perfettini et al., 2010). Understanding this relationship ultimately determines the earthquake hazard potential in a specific subduction zone, and is therefore a relevant active matter of study.

1.1.2 Nazca - South American Convergent Margin (Northern Chile)

One of the most active subduction regions on the planet is the western margin of South America, where the subduction of the oceanic Nazca plate beneath the continental South American plate has generated numerous M>8.0 great megathrust earthquakes which frequently strike the Chilean coast, the most recent examples being the 1995 M8.0 Antofagasta, the 2010 M8.8 Maule, the 2014 M8.1 Iquique, and the 2015 M8.3 Illapel earthquakes (e.g., Moreno et al., 2010; Ruegg et al., 1996; Schurr et al., 2014; Tilmann et al., 2016). In addition to recurrent seismic activity, subduction along this margin has given rise to observed active volcanism and also to the Andes mountain range, the highest non-collisional orogen with an average altitude of ~4,000 m.a.s.l. along >8,000 km (Ramos, 1999).

From its northernmost limit (~4°N) at the Cocos-Nazca spreading centre to its southernmost edge (~46°S) at the Chile Triple Junction (CTJ), the Nazca-South America plate boundary exhibits several remarkable bathymetric features (Figure 1.2). At the northern end of the Chilean subduction, the Nazca Ridge is currently subducting at ~15°S (Beck and Ruff, 1989). The Iquique Ridge has been interpreted to be subducting at ~20°S (Geersen et al., 2015). Further south, the volcanic chain forming the Juan Fernandez Ridge (JFR) collides with the oceanic trench at ~32.5°S and the Chile Ridge is being subducted at the CTJ at ~46.5°S (Herron et al., 1981).

Offshore northern Chile, in the region where the M8.1 Iquique earthquake ruptured, the incoming oceanic plate subducts at ~67 mm/yr with a convergence angle of ~76° N, exhibiting prominent relief consisting of horst-and-grabens, caused by plate bending, and spreading fabrics together with the subducting Iquique Ridge and its several noticeable seamounts (Geersen et al., 2018a). From indirect geophysical observations, it has been suggested that other seamounts are subducting in this area as well (Geersen et al., 2015). Other bathymetric features such as embayments, antiforms, and synforms, which have been mapped on the overriding continental plate seafloor, provide further indirect clues that rugged relief is being subducted in this region (Geersen et al., 2018a).

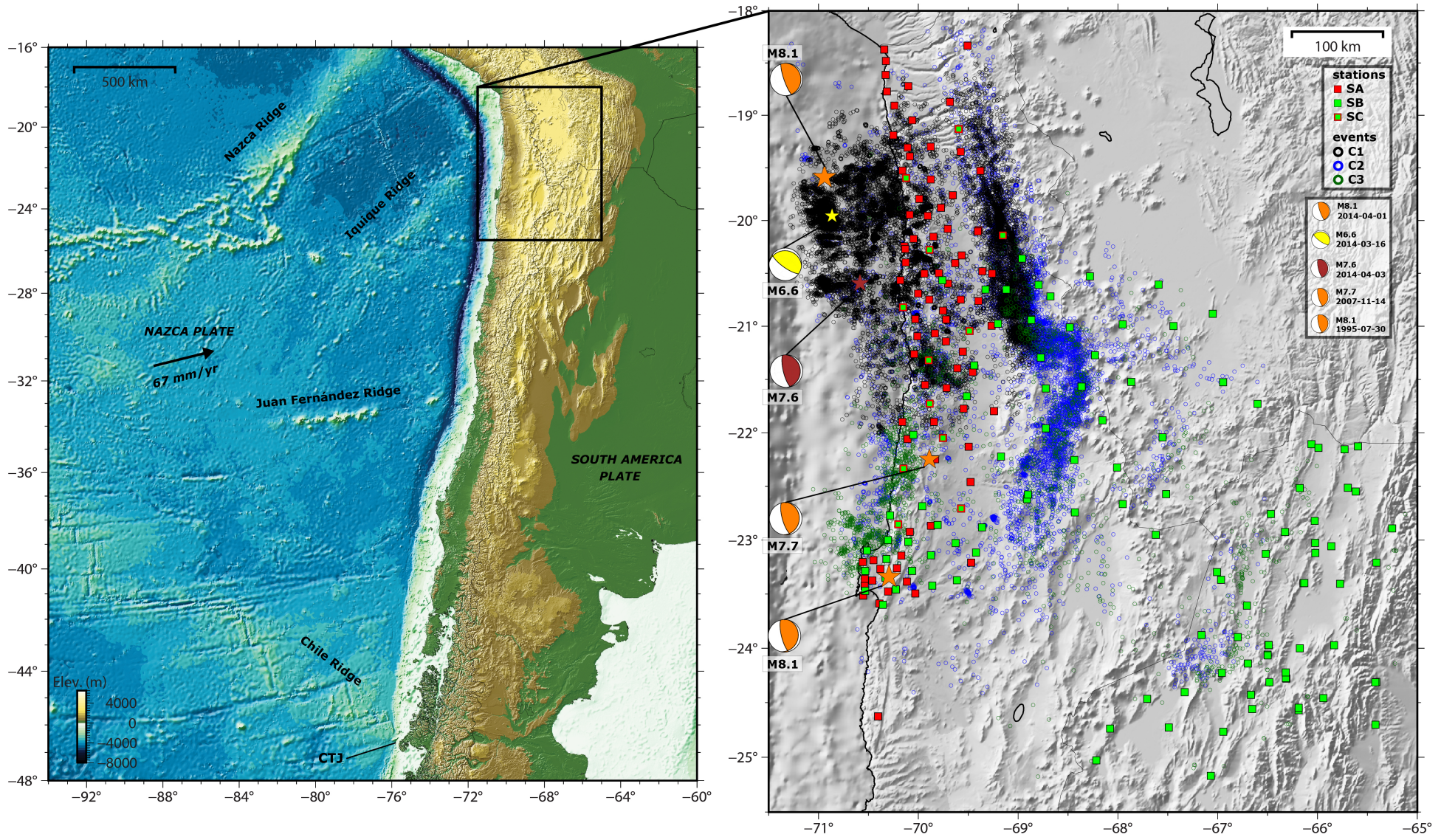

Figure 1.2: (Left) Chilean convergent margin, outlining the most important bathymetric features observed on the incoming oceanic plate and the location of the study region (black box), which is shown in more detail in the right-hand panel. (Right) The various sets of earthquake catalogs and seismic stations used in this work, which are described in the thesis. Seismic records taken from stations SA and SC are used to automatically derive the earthquake catalog C1, as I describe in detail in Chapter 3. Manually-picked event records included in catalog C3 are taken from stations SB and SC to train and test the deep learning-based DeepPhasePick algorithm in the tasks of P- and S-phase detection and picking, as I explain in Chapter 4. Automatically-picked event records extracted from catalogs C1 and C2 are used for further testing the performance of DeepPhasePick (see Chapter 4). The most important earthquakes occurring in the region covered by the different event catalogs are also shown for reference.

Tectonically, northern Chile is considered to be an erosive convergent margin, featuring a trench starving of sediments. Less than 500 m of sediment thickness accumulated within the trench cannot cover up the rugged subducting oceanic seafloor, which favors subduction erosion processes (e.g., von Huene and Scholl, 1991; Völker et al., 2006). This lack of sediment is primarily caused by the extreme arid conditions (rainfall rate <5 cm/yr) of the Puna-Altiplano Plateau and the western forearc, and intensified by the JFR presence suppressing the transport of sediment northward along the trench axis (Yañez et al., 2001). In agreement with observations in other erosive convergent margins, northern Chile has been also shown to feature a small frontal prism (~5-30 km) and normal faulting in the outer forearc (Contreras-Reyes et al., 2012; Geersen et al., 2015; Maksymowicz et al., 2018; Ranero et al., 2006; Sallarès and Ranero, 2005).

1.1.3 Northern Chile–Southern Peru Seismic Gap and the 2014 M8.1 Iquique Earthquake Sequence

In northern Chile, where the recurrence time of megathrust earthquakes is approximately 100 years, no rupture had occurred in the so-called northern Chile-southern Peru seismic gap since 1877, where the rupture of a M~8.8 extended from near Arica to the Mejillones peninsula (Comte and Pardo, 1991).

On 1 April 2014, the M8.1 Iquique earthquake broke a central piece of the northern Chile-southern Peru seismic gap, generating a tsunami of ~2 m height along the northern Chile coast (Hayes et al., 2014). Several coseismic rupture models of the Iquique mainshock have been published to date, most of them reporting an ellipsoidal rupture of ~120 km length, associated with a seismic moment of ~1.7 x 1021 Nm (M = 8.1) and a maximum slip of up to ~12 m (e.g., Duputel et al., 2015; Hayes et al., 2014; Lay et al., 2014; Liu et al., 2015; Ruiz et al., 2014; Schurr et al., 2014).

Seismicity preceding the Iquique mainshock occurring seemingly on the plate boundary in three weeks-long clusters in July 2013, January 2014, and March 2014, and accompanying transient aseismic deformation, have been suggested to progressively weaken an intermediate locking area, which could have finally triggered the mainshock rupture (e.g., Bedford et al., 2015; Bürgmann, 2014; Ruiz et al., 2014; Schurr et al., 2014, 2020).

The largest foreshock (M6.6, 16 March 2014) occurred in the upper plate (Cesca et al., 2016; Schurr et al., 2020) and has been suggested to have reactivated a trench-oblique upper plate reverse fault in the forearc (González et al., 2015). After this M6.6 event, foreshocks migrated northward onto the megathrust, registering three more M>6 events and closing the updip edge of a distinct Mogi doughnut pattern whose downdip edge had been activated during the past several years, before the mainshock was finally triggered (Kato and Nakagawa, 2014; Schurr et al., 2020). It has been proposed that this final stage of events in the foreshock series was caused by either a slow-slip event (Kato and Nakagawa, 2014; Maksymowicz et al., 2018; Yagi et al., 2014) or successive multi-event afterslips (Bedford et al., 2015).

The intense aftershock series following the Iquique mainshock included an M7.6, the largest aftershock, occurring 2 days later about 100 km south of the mainshock epicenter (Duputel et al., 2015; Schurr et al., 2014) and nucleated at a depth comparable to the mainshock. Evidence suggests that the M7.6 aftershock, and the mainshock, only ruptured a partly locked segment of the northern Chile-southern Peru seismic gap since 1) a correlation is observed between the downdip ends of both event ruptures and the interseismic coupling, and 2) both events show an accelerated rupture propagation toward higher locking, as calculated by mapping their high-frequency seismic waves radiation (Schurr et al., 2014).

The fact that a rupture of the northern Chile-southern Peru seismic gap had been expected to occur for several years fostered an international effort to set up the Integrated Plate Boundary Observatory Chile (IPOC) in the region. Operative since 2006, the IPOC network made it possible to exceptionally capture the whole Iquique earthquake sequence. This effort was further bolstered by the addition of several temporary deployments soon after the beginning of the more immediate foreshock sequence, as well as a few days after the Iquique mainshock ocurrence.

As described in Figure 1.2, the large amount of seismic data collected from Northern Chile allows me to address the relevant questions on which the chapters of my thesis are focused, which are introduced in the following section. Using this seismic data, I first aim to shed light on the seismotectonic processes that take place in the northern Chilean subduction zone (Chapter 3). Secondly, I present a novel solution to the classical procedures used for detecting and picking seismic phases (Chapter 4).

1.2 Research Questions and Outline

This thesis addresses the following research questions:

1.1 How do observed seismicity patterns of the Iquique earthquake sequence correlate with static stress changes induced by this earthquake, as well as with coseismic, afterslip, and interseismic locking?

1.2 What does this relationship tell us about the frictional behavior of the northern Chilean portion of the megathrust and possibly of other subduction regions?

Chapter 3 addresses these two research questions. I present a high-resolution earthquake catalog containing ~19,000 foreshocks, aftershocks, and background seismicity which occurred in a time interval covering 1 month preceding and 9 months following the 1 April 2014 M8.1 Iquique earthquake in northern Chile. I describe in detail the foreshock and aftershock spatio-temporal patterns observed from this seismicity dataset. I further explore the correlation between the observed seismicity features with calculated induced static stress changes as well as with different published solutions of coseismic slip, and geodetically inferred afterslip and interseismic locking. I then interpret, based on the results from this analysis, the different frictional behaviors that may take place on the plate interface offshore northern Chile.

2. What role may the forearc and incoming plate structures of the northern Chile convergent margin play in producing the seismicity pattern observed in the Iquique mainshock sequence?

I address this research question in Chapter 3, by investigating how the Iquique sequence seismicity and the inferred frictional behavior on the plate interface can be interpreted in relation to both the seismotectonic processes occurring in the continental wedge overlying the megathrust as well as the diverse underthrusting relief structures, which seemingly populate the incoming oceanic plate.

3. Can we make the seismic phase detection and picking procedures involved in an earthquake location workflow more precise and efficient using the huge amount of seismic data collected in northern Chile?

In Chapter 4, I address this question by presenting DeepPhasePick, a newly developed deep-learning based two-stage automatic algorithm for detecting and picking P and S seismic phases from local earthquakes, trained and optimized using seismic records from earthquakes occurring in northern Chile. I describe the adaptive convolutional and recurrent neural network architectures used in the two stages involved in DeepPhasePick, as well as the systematic hyperparameter optimization which I implement for training these models. I show that DeepPhasePick is capable of predicting P- and S-phase time onsets with analyst level of precision as well as computing the associated time onset uncertainties. Hence, the predicted onset and uncertainties can feed a phase associator algorithm in a next stage of an automatic earthquake location workflow, as a compelling alternative to other conventional automatic picking methods.

4. Can a potentially alternative phase detection and picking method be generalized to other regions where a different tectonic regime is observed?

I aim to respond this question in Chapter 4, by demonstrating the ability of the deep learning-based DeepPhasePick method to generalize to a seismotectonic regime other than northern Chile, from where the data used to train the algorithm was obtained. I do this by testing DeepPhasePick's performance on continuous seismic data from a 30-station network deployed in the rupture area of a recent M6.4 earthquake in a region of convergence between Adriatic and Eurasian plates, near Durres, Albania. Unlike the northern Chile region, the Albanian network is located in an urban area, which is therefore more likely to record higher seismic-noise signals.

1.3 Author's Contributions and Publications

I have reported the outcomes from the analysis of the seismicity patterns related to the 1 April 2014 M8.1 Iquique earthquake presented in Chapter 3 in the publication (Soto et al., 2019b):

Soto, H., Sippl, C., Schurr, B., Kummerow, J., Asch, G., Tilmann, F., Comte, D., Ruiz, S., and Oncken, O. (2019), Probing the northern Chile megathrust with seismicity: the 2014 M8.1

iquique earthquake sequence. Journal of Geophysical Research: Solid Earth, Volume 124, Issue p. 12,935-12,954. https://doi.org/10.1029/2019JB017794

Particularly, I want to acknowledge the contribution of Dr. Jörn Kummerow to this publication, who calculated the set of repeating earthquakes which I have used to complement and substantiate some of the findings resulting from the analysis of the seismicity patterns observed during the aftershock series of the Iquique earthquake.

The earthquake catalog generated in this publication has been made available from (Soto et al., 2019a):

Soto, H., Sippl, C., Schurr, B., Kummerow, J., Asch, G., Tilmann, F., Comte, D., Ruiz, S., and Oncken, O. (2019), Catalog of Hypocenters for the 2014 M8.1 Iquique Earthquake Sequence, recorded by IPOC (plus additional) seismic stations. GFZ Data Services.
https://doi.org/10.5880/GFZ.4.1.2019.009

This relocated catalog contains 18,964 earthquakes occurring in the time period between March and December 2014; 1,137 and 17,826 events before and after the Iquique mainshock, respectively. The first six columns in the downloadable file represent event origin times (year, month, day, hour, minute, second). The next four columns indicate: longitude, latitude, depth (in km) and local magnitude (ML) of each event.

The seismicity contained in this catalog has been partly used to characterize the postseismic deformation following the Iquique mainshock in the following publication, to which I contributed as co-author:

Shrivastava, M. N., González, G., Moreno, M., Soto, H., Schurr, B., Salazar, P., and Báez, J. C. (2019), Earthquake segmentation in northern Chile correlates with curved plate geometry. Scientific Reports, 9(1), 4403. https://doi.org/10.1038/s41598-019-40282-6

The characteristics of the newly developed deep learning-based automatic algorithm DeepPhasePick trained for detecting and picking seismic phases from local earthquakes, as well as the findings resulting from applying it to seismic data from two different tectonic regimes, are presented in Chapter 4 and have been included in a publication submitted at *Geophysical Journal International*, which is available as a preprint version from *EarthArxiv*:

Soto, H. and Schurr, B. (2020), DeepPhasePick: A method for Detecting and Picking Seismic Phases from Local Earthquakes based on highly optimized Convolutional and Recurrent Deep Neural Networks. https://doi.org/10.31223/X5BC8B

The optimized DeepPhasePick models trained for phase detection and phase picking tasks, as well as an example script which applies the DeepPhasePick algorithm for both tasks on continuous waveforms, will be made available via the GitHub repository: *https://github.com/hsotoparada/DeepPhasePick*.

The Section S2 in the Chapter Supplementary Information of this thesis provides additional figures and details regarding the multistage automatic procedure, which I have adapted from

Sippl et al. (2013) in order to conduct the phase detection and picking, as well as the subsequent localization of the ~19,000 earthquakes used for the analysis carried out in Chapter 3.

The Section S3 in this chapter further includes figures and tables that complement the results obtained from the application of the DeepPhasePick algorithm introduced in Chapter 4. Details related to some conditions that can be optionally applied to enhance the performance of DeepPhasePick in phase detection and picking tasks are also included there.

Chapter 2

Methods

Primarily, my thesis deals with the detection, location and interpretation of a large number of earthquakes. This chapter introduces the conceptual ideas behind the various methods I use for performing classical earthquake detection and location. I also present the key concepts related to deep learning, which I apply in the development of a new automatic seismic phase detection and picking algorithm.

2.1 Classical Earthquake Detection

2.1.1 Detection and picking of seismic phases

Classical earthquake location methods rely ultimately on precise identification and picking of P- and S-phase onsets. This has usually been a task performed manually by analysts, based on their experience and training. The use of numerous automatic detection techniques developed in the last half of the past century, however, has recently increased due to the exponential growth in the seismic data available at hand.

Fundamentally, there exists three groups of classical automatic methods, performing detections based on: (1) energy or frequency content of the seismic waveforms (e.g., Aldersons, 2004; Allen, 1978; Baer and Kradolfer, 1987; Diehl et al., 2009; Di Stefano et al., 2006), (2) correlations between template and continuous seismic waveforms (e.g., Gibbons and Ringdal, 2006; Harris, 1991; Van Trees, 1968), and (3) linear combination of orthogonal basis seismic waveforms (Harris, 1997, 2001; Scharf and Friedlander, 1994).

In this thesis I use three algorithms implementing automatic energy-based detections, namely: *STA/LTA* (Allen, 1978; Baer and Kradolfer, 1987), *MPX* (Aldersons, 2004; Di Stefano et al., 2006), and *Spicker* (Diehl et al., 2009). Below I describe the main characteristics of two of them, which I use to obtain P-phase (*MPX*) and S-phase (*Spicker*) time onsets. The *STA/LTA* algorithm is an in-house developed version of the STA/LTA trigger algorithm described by Withers et al. (1998), which I apply to obtain preliminary P time arrivals.

I do not describe correlation based detectors in detail since this type of detector is only used to calculate the repeating earthquakes which are provided to complement the seismicity

analysis performed in Chapter 3 (see also Section 1.3). Detectors based on orthogonal basis waveforms are not used in this study at all and for this reason I will not be discussing these types of detectors further.

2.1.1.1 MPX

The *MPX* algorithm (Aldersons, 2004; Di Stefano et al., 2006) performs picking of P seismic phases in three steps.

First, an adaptive Wiener filter (Aldersons, 2004; Douglas, 1997) is applied to the raw seismic waveforms. This avoids the distortion of the waveforms, which usually occurs when other filters, such as a standard Butterworth bandpass filter, are used (Douglas, 1997). The applied Wiener filter adapts to the frequency content of each analyzed seismogram in order to attenuate the single or multi frequency bands where noise is predominant.

MPX requires preliminary P picks which have been previously associated to an earthquake location, from which theoretical P arrival times can be calculated. *MPX* then defines a pair of time windows located around the theoretical arrival time, one in the noise and one in the signal portion of the wavefrom, where spectral analysis is performed to obtain the noise-dominated frequencies (Aldersons, 2004; Di Stefano et al., 2006).

Following the filtering and spectral analysis, the second stage in *MPX* consists of running a picking engine (Baer and Kradolfer, 1987) on the filtered data, where P onsets and their respective polarities are determined. In a final step, *MPX* computes time arrival uncertainties based on a calibrated weighting algorithm that takes into account the waveform sampling rate and the results from the spectral analysis (Aldersons, 2004; Di Stefano et al., 2006).

2.1.1.2 Spicker

The *Spicker* algorithm (Diehl et al., 2009) combines three different detection and picking approaches in order to compute S-phase time arrivals, based on existing manually- or automatically-derived P-phase picks (tP_{obs}) and theoretical S-phase arrival times (tS_{pre}) determined from a given velocity model.

In the first of these approaches, *Spicker* applies an *STA/LTA* detector on the two-horizontal seismogram components and S-phase search windows are then defined based on tP_{obs} and tS_{pre}. A picking algorithm similar to the Baer and Kradolfer (1987) method is applied on the S-phase search windows of the *STA/LTA* characteristic function in order to obtain a minimum (tS^1_{min}) and maximum (tS^1_{thr}) possible S onsets.

The second approach used in *Spicker* consists of applying a polarization detector, based on the method proposed by Cichowicz (1993), where the degree of linear polarization, the directivity of the particle motion, and the ratio between transverse and total energy are combined to define a characteristic function (CF_S) that is primarily sensitive only to S-wave energy and neglects P-waves. Then, a picking algorithm similar to that used in the first approach is applied to CF_S, so as to compute two additional minimum (tS^2_{min}) and maximum (tS^2_{thr}) possible S-phase onsets.

In the third approach, *Spicker* applies an autoregressive Akaike Information Criterion (AR-AIC) picking algorithm, based on the method proposed by Takanami and Kitagawa (1988), on a search window defined by the minimum and maximum possible S-phase onsets resulting from the detectors applied in the first two approaches.

Once the three implemented approaches are performed, *Spicker* makes use of their combined results to finally calculate S-phase arrival times as well as their corresponding interval of uncertainty.

2.1.2 Seismic phase association

In the next step of a classical earthquake detection stage, the phase time onsets resulting from the identification and subsequent picking of the seismic phases are associated to the earthquakes that originated them. As for the phase detection and picking tasks, the use of automatic phase association methods has become more common than the conventional manual approach performed by analysts.

Most of the existent association algorithms follow a similar approach. First, travel times are computed from each node in a gridded region of interest to each station in the network, given a presumed seismic velocity structure. Then, for each node in the grid, subsets of phase time onsets picked at different stations are examined by back-propagating their travel times through the grid, searching for coherent origin times (e.g., Draelos et al., 2015; Johnson et al., 1995; LeBras et al., 1994; Stewart, 1977).

In this thesis I use the algorithm *Binder* (Rietbrock and Heath, pers. comm., 2010), in order to associate preliminary P-phase picks obtained by applying the STA/LTA detector. In *Binder* a migration approach on an orthorhombic grid is implemented, and, following the procedure described above, events are declared when coherent origin times at one grid node are found for the back-propagated travel times, within a user-defined error.

2.2 Classical Earthquake Location

An earthquake is a sudden release of energy in the Earth's crust or mantle induced by tectonic stress, creating seismic waves. The absolute location (hypocenter) of an earthquake (x, y, z) and its origin time (t) are related to the arrival time (t_i) of the earthquake at an ith network station, according to:

$$t_i = t + \frac{\sqrt{((x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2)}}{v} \tag{2.1}$$

for a homogeneous material with seismic velocity v.

Earthquake location is an inverse problem that is most commonly solved by minimizing the misfit between the P- and S- seismic phase arrival times observed at stations and the arrival times calculated based on a given seismic wave velocity model. In order to solve this problem,

equation 2.1 can be written in a more general manner as

$$\boldsymbol{F}(\boldsymbol{m}) = \boldsymbol{d} \tag{2.2}$$

where $\boldsymbol{m}(x, y, z, t)$ is a vector defining the earthquake source parameters and $\boldsymbol{F}$ is an operator that calculates the data vector $\boldsymbol{d} = (t_1, t_2, \ldots, t_n)$ from the source parameters $\boldsymbol{m}$.

Equation 2.1, and therefore also 2.2, is inherently nonlinear, since a change in the hypocentral coordinates produces a nonlinear perturbation in the observed arrival time. Therefore, the solution cannot be obtained by simply solving a system of linear equations, and hence numerical approaches are required. Numerical methods which have been developed to solve equation 2.2, can be primarily classified as linearized or nonlinearized (direct-search) methods.

2.2.1 Linearized Absolute Earthquake Location

Linearized approaches aim to solve the earthquake location problem by applying Taylor series expansion to linearize equation 2.2. In this way, $\boldsymbol{m}$ can be iteratively improved by incremental changes in an initial guess $\boldsymbol{m}_0$. In particular, the technique developed by Geiger (1910) iteratively minimizes the root mean square (RMS) of the travel time residuals t_r, which correspond to the difference between the observed and calculated travel times. In the iterative Taylor series expansion, the time residuals are obtained from

$$t_r = \frac{\partial t_i}{\partial m_j} \delta m_j \tag{2.3}$$

where δm_j represents the small perturbations on the hypocentral parameters m_j (Lay and Wallace, 1995).

Geiger's method is valid if errors produced by a simplistic assumed velocity model, which are introduced in the misfits between observed and calculated travel times, are neglected. Since this approach is relatively computationally inexpensive, it is used in most of the existing software to calculate earthquake location, such as *HYPOELLIPSE* (Lahr, 1999) and *HYPO71* (Lee and Lahr, 1975) (Lee and Lahr, 1975).

More accurate earthquake locations, which additionally take into account the velocity structure in the region and its uncertainty, can be obtained by solving equation 2.2 as a coupled hypocenter-velocity model problem (Kissling, 1988; Kissling et al., 1994). A solution to the coupled problem, which is based on damped least square estimation, has been implemented in the software *VELEST* (Kissling et al., 1994, 1995) in order to invert for a so-called minimum 1D velocity model. This is the preferred initial 1D velocity model in procedures involving earthquake location or seismic tomography, especially in regions where the seismic velocity structure is not well known. A different version of the damped least squares solution, obtained by applying the parameter separation method (Pavlis and Booker, 1980), is implemented in the software *simulps* (Eberhart-Phillips, 1993; Evans et al., 1994; Thurber, 1983, 1993) in order to invert for a 2D or 3D velocity model. This usually represents the next step after a *VELEST*-derived 1D velocity model is obtained in detailed studies of local earthquake tomography.

2.2.2 Nonlinearized Absolute Earthquake Location

In the nonlinearized earthquake location approach, the solution is probabilistic and does not involve the explicit calculation of partial derivatives as in equation 2.3. Instead, an optimal solution is usually found by applying deterministic or stochastic searches over the space of possible earthquake hypocenters and origin times (Lomax et al., 2009).

As proposed by Tarantola and Valette (1982), the result of this search can be expressed as a complete solution, given by a posterior probability density function (posterior PDF) that represents the likelihood of occurrence of all possible solutions for earthquake hypocenters and origin times. Unlike linearized solutions, which produce a single *best-fit* earthquake location and an associated uncertainty, the nonlinearized method returns the full probabilistic solution representing the discrepancy between calculated and observed phase arrival times.

The most commonly used software that solves the nonlinearized eathquake location problem is *NonLinLoc* (Lomax et al., 2000). *NonLinLoc* estimates the posterior PDF for the hypocenter coordinates and origin time of events by searching over the solution space based on a given velocity model and a calculated 3D P- and S-phase travel-time grids.

For sources within the seismic network and far from significant velocity structure gradients, linearized methods provide stable location and uncertainty estimations (e.g., Husen et al., 2003; Lomax et al., 2000). However, these methods are less consistent when sources are outside the network. In such cases nonlinearized solutions can be more precise.

2.2.3 Relative Earthquake Location

Unlike absolute locations, relative earthquake locations are obtained with respect to other events, or groups of events, and are commonly used to refine the precision of already located hypocenters.

One of the techniques used to obtain relative locations is the double difference (DD) method, which has been implemented in the commonly used software *HypoDD* (Waldhauser and Ellsworth, 2000). This approach relies on an important fact: the difference in travel times between two events at a common station can be associated to their inter-event distance, as long as the latter is small compared to the distance between each event and the station.

The DD technique solves an extended version of equation 2.3, where time residual differences for pairs of earthquakes are minimized by iteratively adjusting the vector difference between hypocenters in each event pair. By using observed P- and S-phase travel time differences obtained from both cross-correlation and catalog picks, DD is able to relocate earthquakes relative to each other with high precision. This allows the determination of relative distances between near-located events in a cluster with the accuracy level of the cross-correlation data, while relative distances between different event clusters can be simultaneously obtained with the accuracy of the catalog travel times data (Waldhauser and Ellsworth, 2000).

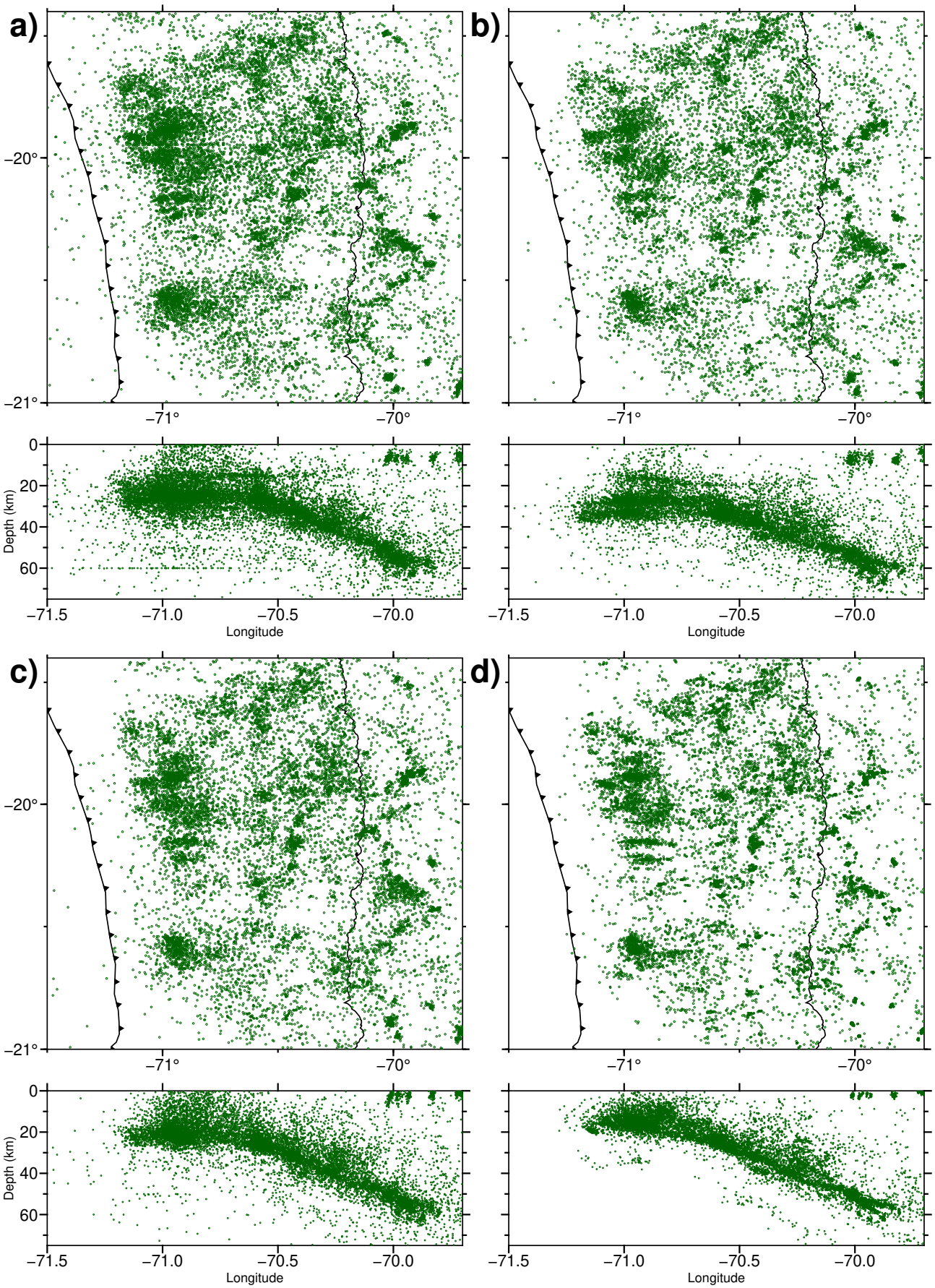

Figure 2.1: Comparison between successive earthquake relocations carried out by some of the methods described in this section, as part of the automatic relocation workflow implemented in Chapter 3. (a) Events relocated by *HYPO71*, after picking P and S phases. (b) Events relocated by using the 1D velocity model derived in *Velest*. (c) Events relocated using the 2D velocity model calculated in *simulps*. (d) Events relocated by applying the Double Difference method in *HypoDD*. As hypocenter locations improve through these stages, the various seismicity features progressively become more compact and better defined.

2.3 Machine Learning: Key Concepts

Machine learning is a field that aims to build algorithms which can *learn* from data and make predictions without being explicitly programmed. These algorithms usually implement either an unsupervised or a supervised learning. In the unsupervised case, there is only non-labeled input data and no corresponding output data, hence the goal is to learn the underlying structure in the input data, frequently applying clusterization or association methods. In the supervised case, which will be further discussed in this section since it is the approach I use in this study, all data is labeled and the goal is learning to predict the output from the input data, by solving tasks than can be expressed as:

$$\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{m}) \tag{2.4}$$

where f is a function or model mapping an input tensor $\boldsymbol{x}$ to an output tensor $\boldsymbol{y}$, and $\boldsymbol{m}$ is a tensor including the model parameters which are *learned* in order to solve the problem at hand. A common practice in supervised approaches is to subdivide the available data into three different sets: training, validation and test datasets. The data samples in the training set are used to fit or train the mapping model. The validation set provides an unbiased performance evaluation of the trained model and is often used for optimizing the model hyperparameters, which define the external configuration of the model. Once the model is optimized, a final unbiased performance can then be evaluated on the test set.

Generally, the supervised tasks to be solved are either regression or classification (binary or categorical). Regression is the task of finding (or learning) a mapping function f from the input data to a continuous real-valued predicted output. For instance, predicting the next amplitude value in a time series based on various related observations. In categorical or multi-class classification, f is a model learned for mapping input data samples to a vector of N numbers, each representing the probability of a sample belonging to one of the N tested classes. For example, classifying event seismic waveforms into P-phase, S-phase and perhaps also additional phase classes. A binary classification is a special case of multi-class classification where $N = 2$, hence f is learned to map uncategorized data samples to numbers between 0 and 1, which represent the probability of a sample belonging to one of two classes. An example of this task would be the classification of each timestep in a time series as corresponding to one or another characteristic (e.g., noise or signal).

In each of the abovementioned tasks, the learning consists of finding the set of parameters that minimize a *loss function* L, which represents how well a particular model f and set of parameters $\boldsymbol{m}$ solve the task, and can be expressed as:

$$L(\boldsymbol{x}, \boldsymbol{m}, \boldsymbol{y_t}) = \sum_{(\boldsymbol{x}, \boldsymbol{y_t})} l(f(\boldsymbol{x}, \boldsymbol{m}), \boldsymbol{y_t}) = \sum_{(\boldsymbol{x}, \boldsymbol{y_t})} l(\boldsymbol{y}, \boldsymbol{y_t}) \tag{2.5}$$

where $l(\boldsymbol{y}, \boldsymbol{y_t})$ is a *penalty* function that quantifies the agreement between the model predictions $\boldsymbol{y}$ and the known target values $\boldsymbol{y_t}$.

For a regression task, common choices of the penalty function are the mean squared error (MSE), the mean squared logarithmic error (MSLE), or the mean absolute error (MAE). In classification tasks, the conventionally used penalty function is the *cross-entropy*, which measures the distance between assumed probability distributions of the training samples, and can be the binary cross-entropy

$$l(y, y_t) = -y_t \log(y) - (1 - y_t) \log(1 - y) \tag{2.6}$$

for binary classification, and the categorical cross-entropy

$$l(\boldsymbol{y}, \boldsymbol{y_t}) = -\sum_i y_{t_i} \log(y_i) \tag{2.7}$$

for multi-class classification.

Equations 2.6 and 2.7 compute cross-entropies for a single prediction, where predicted and target categorical values are *hot-encoded* (as required by most machine learning algorithms) in binary variables $\boldsymbol{y}$ and $\boldsymbol{y_t}$ representing their probability distributions. For multi-class classification, the classes are modeled as a categorical distribution. For example, in a three-class classification task, $\boldsymbol{y_t} = [y_{t_1}, y_{t_2}, y_{t_3}] = [1, 0, 0]$ could represent a sample with a label for the first class, which may be predicted by the model as $\boldsymbol{y} = [y_1, y_2, y_3] = [0.7, 0.2, 0.1]$. For binary classification, equation 2.6 is obtained as a especial case of 2.7 for two classes ($i = 2$) modeled as a Bernoulli distribution so that $\boldsymbol{y_t} = [y_{t_1}, y_{t_2}] = [y_t, 1 - y_t]$ and $\boldsymbol{y} = [y_1, y_2] = [y, 1 - y]$.

2.3.1 Deep Neural Networks

Deep learning makes use of special tunable models f to solve the equation 2.4, which are formed by multiple stacked nonlinear operations that activate weighted connections of units. These tunable functions are called *deep* neural networks, because they were originally designed with the aim of modeling the strength (weights) of the firing rate (activation) of the spikes generated by the neurons (units) synapses. Stacking nonlinear operations allows the network to learn incremental abstractions from the input, so that inputs can be mapped to output predictions for solving very complex tasks.

Each of the stacking nonlinear functions in a neural network corresponds to a *layer* f_L, in which an input vector $\boldsymbol{x}$ is mapped to an output vector by applying a nonlinear function g (so-called *activation* function) through weighted sums of the input, so that:

$$f_L(\boldsymbol{x}, \boldsymbol{m_L}, g) = g(\boldsymbol{b_L} + \boldsymbol{w_L}^\top \boldsymbol{x}) \tag{2.8}$$

where $\boldsymbol{m_L} = (\boldsymbol{w_L}, \boldsymbol{b_L})$ represents the learnable parameters of the model layer given by the weight $\boldsymbol{w_L}$ and the bias $\boldsymbol{b_L}$ tensors, and $\boldsymbol{w_L}^\top$ is the transpose weight tensor.

Typically, the input data is fed to the input layer of the network and then propagated to the output layer by the activations applied in the intermediate *hidden* layers. Depending on the

task at hand, the activation function g applied can be one of the following commonly predefined functions (see Figure 2.2):

$$\text{rectified linear function (ReLU):} \quad g(t) = \max(t, 0) \tag{2.9}$$

$$\text{logistic sigmoid function:} \quad g(t) = \frac{1}{1 + e^{-t}} \tag{2.10}$$

$$\text{softmax function:} \quad g_i(\boldsymbol{t}) = \frac{e^{t_i}}{\sum_{j=1}^{N} e^{t_j}} \tag{2.11}$$

$$\text{hyperbolic tangent (tanh):} \quad g(t) = \frac{e^{t} - e^{-t}}{e^{t} + e^{-t}} \tag{2.12}$$

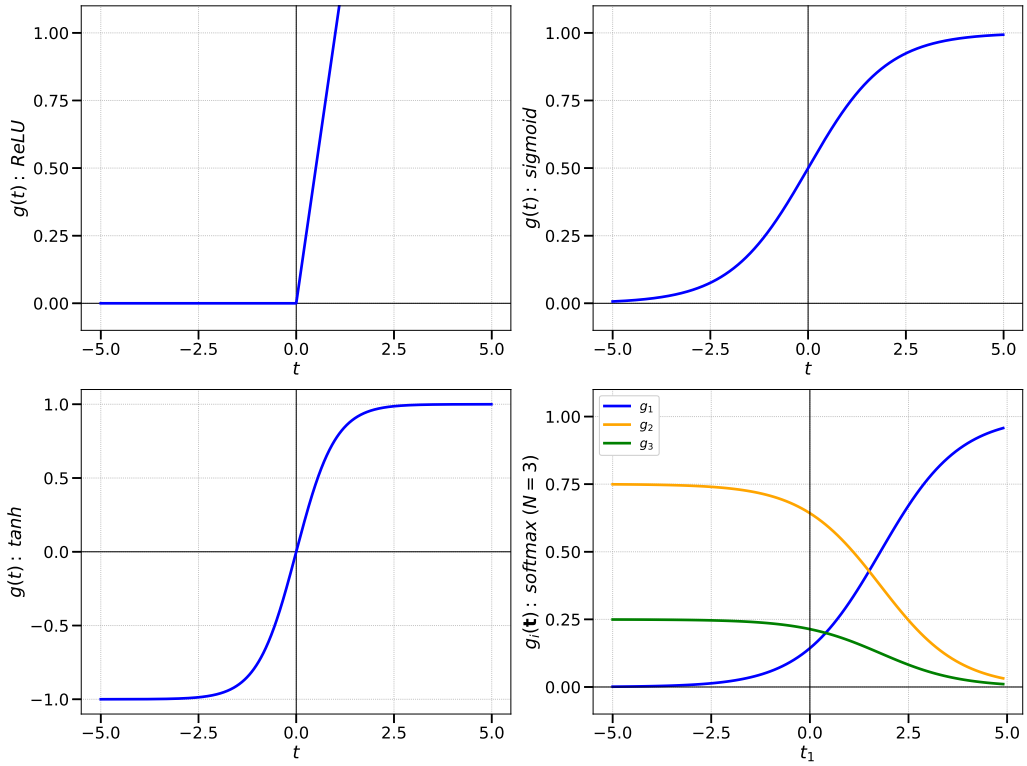

Figure 2.2: Activation functions defined in equations 2.9 to 2.12. For the multivariable softmax function, the plot shows the vector components resulting from applying the activation on input vectors $\boldsymbol{t} = [t_1, t_2, t_3] = [t_1, 1.5, 0.4]$, so considering t_2 and t_3 as constant values.

ReLU (Nair and Hinton, 2010) is the activation most commonly applied in the hidden layers of modern neural networks, although sigmoid and tanh functions are also widely used for this purpose. The sigmoid function is the preferred activation in a binary classification task, where it is usually applied to a 1-unit output layer of a network and returns a value between 0 and 1, which can be interpreted as a class probability. The softmax function is the preferred activation function in a multi-class classification task, where it is typically applied to an N-unit output layer of a network, returning an output vector of values that sum to 1, representing the predicted probabilities for each of the N classes. The tanh activation squashes a real-valued input to the

range between -1 and 1, and is mostly used in the memory blocks of LSTM networks (see below).

The most frequently used neural networks may or may not include feedback or recurrent connections between their units. Networks that have feedback connections are called Recurrent Neural Networks (RNNs; Rumelhart et al., 1986), which include the Long-Short Term Memory (LSTM; Hochreiter and Schmidhuber, 1997) network. Networks with no feedback connections are known as Feedforward Neural Networks (FNNs), the most common of which are the Multi-Layer Perceptron (MLP; Rumelhart et al., 1986; Werbos, 1988) and the Convolutional Neural Network (CNN; e.g., Goodfellow et al., 2016).

The simplest type of neural network is the MLP (Figure 2.3), which is formed by one or more stacked hidden layers. In the simplest case of one hidden layer, MLP is defined by:

$$f_{L_2}(f_{L_1}(\boldsymbol{x}, \boldsymbol{m_{L_1}}, g_{L_1}), \boldsymbol{m_{L_2}}, g_{L_2}) = g_{L_2}(\boldsymbol{b_{L_2}} + \boldsymbol{w}_{\boldsymbol{L_2}}^{\top}\, g_{L_1}(\boldsymbol{b_{L_1}} + \boldsymbol{w}_{\boldsymbol{L_1}}^{\top}\, \boldsymbol{x})) \tag{2.13}$$

where $\boldsymbol{x}$, f_{L_1}, and f_{L_2} are usually called the input, the hidden, and the output layers, respectively. Depending on the problem, the depth of the model can be increased by stacking more hidden layers. Moreover, the size of each hidden layer can be modified by adjusting the dimensions of the weight $\boldsymbol{w}$ and bias $\boldsymbol{b}$ tensors.

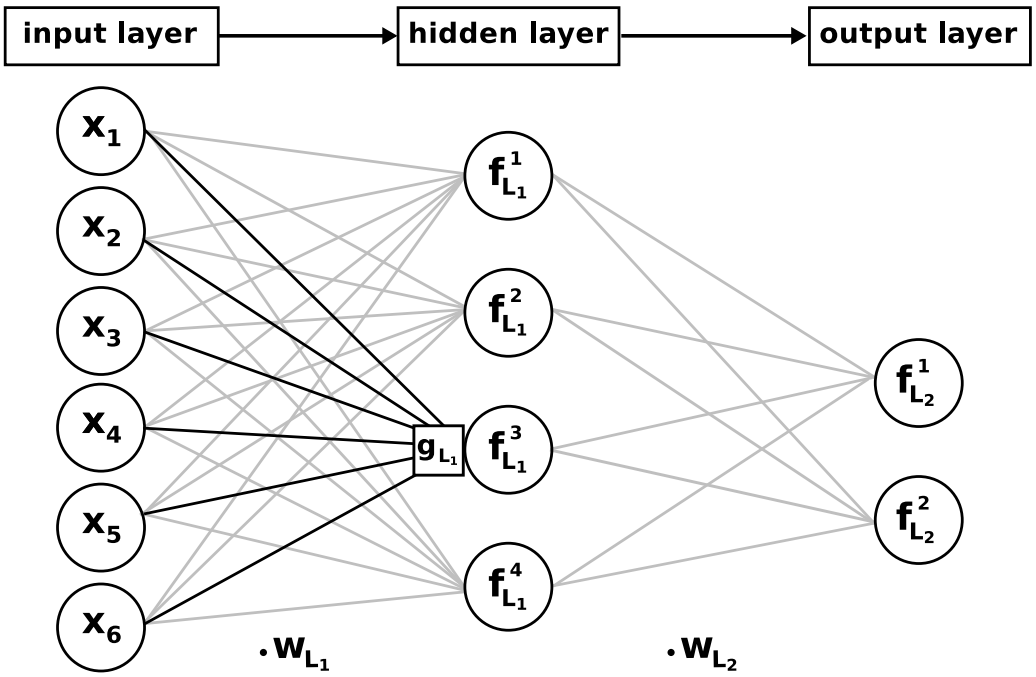

Figure 2.3: Visualization of a MLP with one 4-unit hidden layer f_{L_1}, 6-unit input layer $\boldsymbol{X}$, and 2-unit output layer f_{L_2}. Bold connections represent the activation of one of the hidden units ($f_{L_1}^3$).

The MLP model is formed by fully connected layers, where each unit in a layer is connected with each other unit in the previous layer. This has proven to be inefficient for solving modern complex tasks, due to the usually high number of parameters required. Moreover, MLP disregards inherent structure in the input data which can be determinant to learn certain features, such as 2D pixel distributions forming shapes in images or temporal patterns in time series.

These issues have been overcome with the development of CNNs (Figure 2.4), which are

networks formed by intermediate layers f_L where the input $\boldsymbol{X}$ is convolved with a weight tensor $\boldsymbol{W}$:

$$f_L(\boldsymbol{X}, \boldsymbol{m_L}, g) = g(\boldsymbol{b_L} + \sum_i \boldsymbol{X_i} * \boldsymbol{W_i}) \tag{2.14}$$

In a CNN, $\boldsymbol{W}$ is called *filter* or *kernel*. The number of layers, the amount of convolutional units per layer, and the size of the filters are examples of hyperparameters, which need to be selected by the user. The input tensor $\boldsymbol{X}$ is usually a matrix which can represent, for instance, the red, green and blue intensities for respective color channels in an image, or the amplitudes of different movement components in a sensor time series.

The efficiency of a CNN relies on its particular properties of *local connectivity* and *weight sharing*. Local connectivity means that the elements in the output of a convolutional layer are related to a certain *receptive field* area in the vicinity of each of the elements in the input. This implies that a CNN network can learn local, strongly correlated features in the data, which in turn reduces the number of learnable parameters needed compared to the fully connected layers in MLP networks. Weight sharing refers to the fact that the same filters in a CNN layer are convolved through all the input data, which prevents the filters from learning features that only exist in certain locations in the data. This additionally reduces the number of learnable parameters.

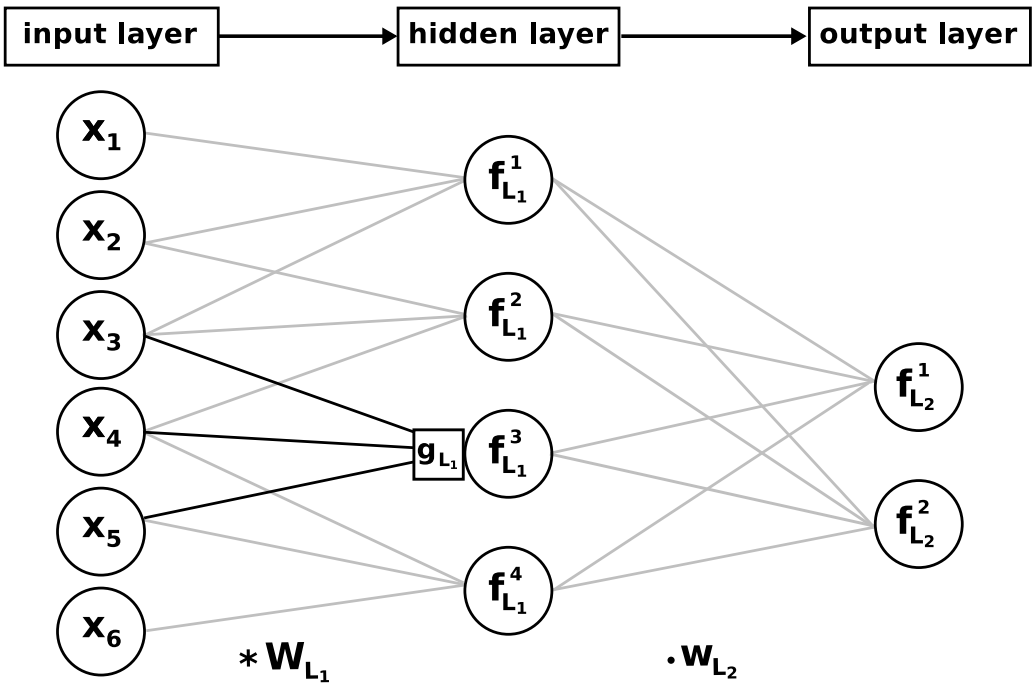

Figure 2.4: Visualization of a CNN with one 4-unit hidden layer f_{L_1}, 6-unit input layer $\boldsymbol{X}$, and 2-unit output layer f_{L_2}. Bold connections represent the activation of one of the hidden units ($f^3_{L_1}$), with a convolutional filter W_{L_1} of size 3.

Even though local connectivity and weight sharing makes CNNs highly efficient, the size and number of filters required to solve complex tasks can still produce a very high amount of learnable parameters. One of the common methods that helps to further reduce the number of parameters is to introduce *pooling layers*, which summarize some local-scale statistics in the

input data. For instance, a max-pooling layer computes the maximum in a certain local region, while average-pooling layers calculate the mean.

MLPs and CNNs can map input patterns to output predictions, however this mapping does not take into account the history of previous inputs, hence these networks cannot learn contextual information from input data which is sequentially structured. This limitation is overcome by the RNNs (Figures 2.5a and 2.5b), which are designed to maintain an internal memory state of previous inputs, thanks to their recurrently connected units. A recurrent layer can be expressed as a MLP layer, where however the activation is not only defined in function of the current external input, but it also takes into account the layer activation from the previous timestep:

$$f_L(\boldsymbol{x}, \boldsymbol{m_L}, g) = g(\underbrace{\boldsymbol{b_L^t} + \boldsymbol{w_L}^\top \boldsymbol{x^t}}_{\text{current input}} + \hat{\boldsymbol{w}}_{\boldsymbol{L}}^\top \underbrace{g(\boldsymbol{b_L^{t-1}} + \boldsymbol{w_L}^{\top(\boldsymbol{t-1})} \boldsymbol{x^{t-1}})}_{\text{previous timestep activation}}) \tag{2.15}$$

However, RNNs are only able to capture a limited range of context, due to the fact that the effect of an input element on units in hidden and output layers in recurrent networks can exponentially decay or increase. This is known as the vanishing/exploding gradient problem (Bengio et al., 1994; Hochreiter, 1991; Hochreiter et al., 2001), which can make the network forget long-term past inputs and thus stop it from learning enough contextual information.

The solution to this problem was achieved with the development of the LSTM network, a special type of recurrent network that makes use of multiplicative gates contained in memory blocks to store information that can be available to the network over long periods of time. Basically, the LSTM architecture is a RNN where memory blocks are used as the hidden units. A memory block is formed by at least one self-connected memory cell and three multiplicative gates (forget, input, and output), which control the activation of the cell. The forget gate multiplies the previous cell state, controlling what information is important to keep from a previous timestep. The input gate controls which relevant information from the current input and previous hidden state is added to the cell state. The new cell state is then passed through a tanh activation and the result is multiplied by the output gate to determine the next hidden state. Finally, the new cell and hidden states are passed on to be processed in the next timestep (see Figure 2.5c).

Even though LSTM networks can learn patterns from very long sequences, they only process sequential data in chronological order, thus they are limited when learning contextual information from both past and future data matters. Bidirectional RNNs (BRNNs; Baldi et al., 1999; Schuster and Paliwal, 1997; Schuster, 1999) overcame this shortcoming by processing input sequences in both forward and backward directions through two separate recurrent layers, which are connected to a common output layer. Combining LSTM and BRNN architectures leads to Bidirectional LSTM networks (Chen and Chaudhari, 2005; Graves and Schmidhuber, 2005; Thireou and Reczko, 2007), which are capable of learning very long sequences by taking into account a more comprehensive temporal contextual information.

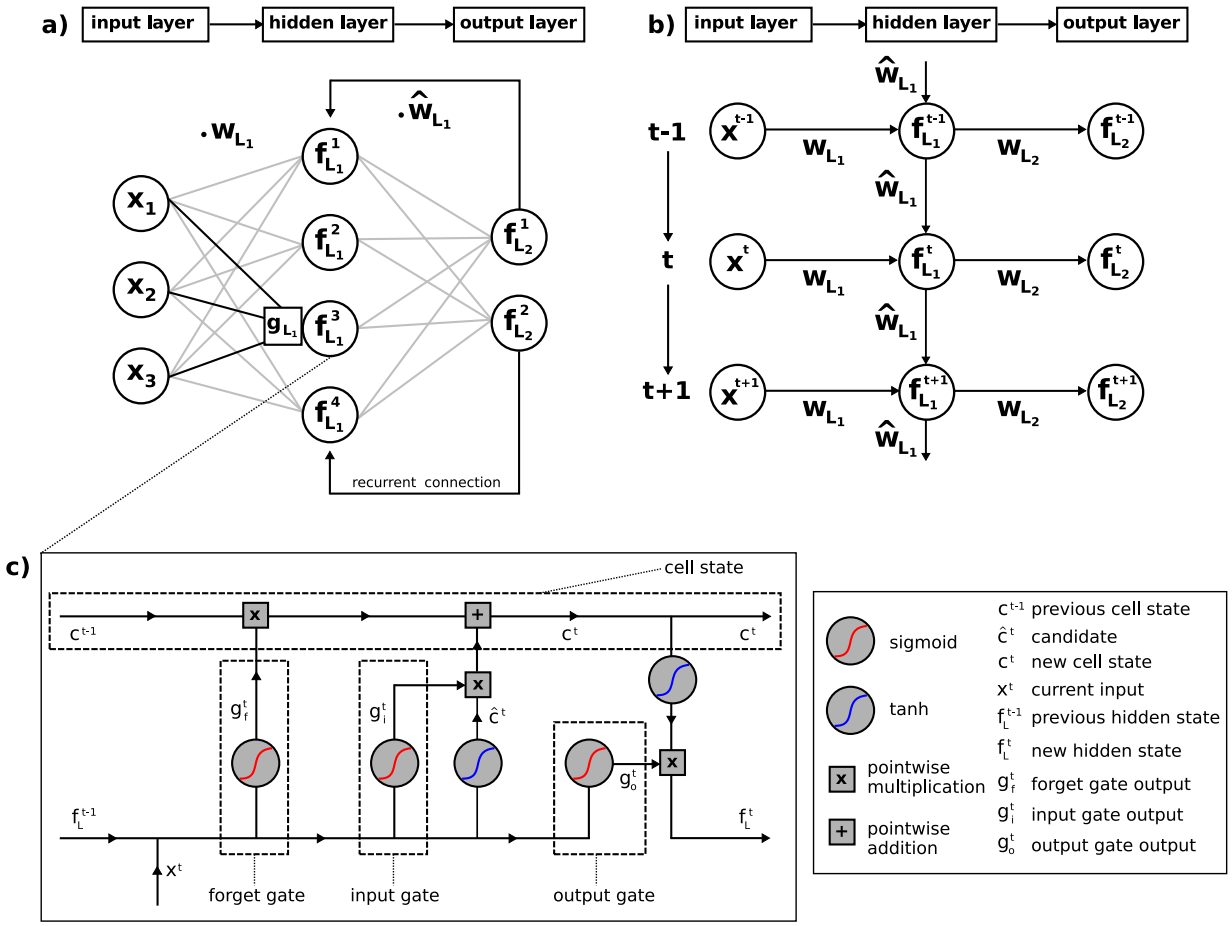

Figure 2.5: (a) Visualization of a RNN with one 4-unit hidden layer f_{L_1}, 3-unit input layer $\boldsymbol{X}$, and 2-unit output layer f_{L_2}. Bold connections represent the activation of one of the hidden units ($f^3_{L_1}$). (b) Visualization of recurrent connections (external loop in (a)) in the unfolded RNN over three timesteps. Each node represents a layer of units at each timestep. (c) Memory block representing a hidden unit in a LSTM network. The flow of relevant information is controlled by the memory cell state and the forget, input and ouput gates, based on the current input and the previous hidden unit state.

2.3.2 Model Optimization and Generalization

As in machine learning, finding an optimal model in deep learning basically means minimizing the loss function L in equation 2.5, or in other words finding the model parameters $\boldsymbol{m}$ which satisfy

$$\frac{\partial}{\partial \boldsymbol{m}} L(\boldsymbol{m}, f, \boldsymbol{d}) = 0 \tag{2.16}$$

Deep neural networks have been designed so that all the operations involved in their layers are differentiable, hence it would be possible, in theory, to analytically solve the equation 2.16. In practice, however, this becomes unfeasible in modern neural networks, that are usually formed by thousands or millions of learnable parameters.

Therefore, an alternative approach is normally used, in which initial model parameters $\boldsymbol{m_0}$ (usually random values) are iteratively adjusted in the direction that reduces L. This is achieved by adjusting the initial parameters in the opposite direction of the gradient of L with respect to $\boldsymbol{m_0}$, as this gradient represents a local linear approximation of L around $\boldsymbol{m_0}$. That is, a

parameter is decreased if the gradient is positive and increased if the gradient is negative. A small positive number λ, called the learning rate, controls how much the initial parameters are changed in each iteration.

The complete iterative process consists basically of three steps: 1) a forward pass where network predictions are computed for the training data, 2) the estimation of the gradient of the loss function L in equation 2.5 with respect to the network output, and 3) a backward pass where this gradient is propagated from the output to the input layer of the network and all the learnable network parameters are adjusted so that L is reduced. This learning cycle is repeated until L does not further decrease.

The backward pass is performed by computing the gradient of the loss function with respect to the model parameters:

$$\frac{\partial}{\partial \boldsymbol{m}} L(\boldsymbol{m}, f, \boldsymbol{d}) = \sum_{(\boldsymbol{x}, \boldsymbol{y_t})} \frac{\partial}{\partial \boldsymbol{m}} l(f(\boldsymbol{x}, \boldsymbol{m}), \boldsymbol{y_t}) = \sum_{(\boldsymbol{x}, \boldsymbol{y_t})} \frac{\partial}{\partial \boldsymbol{y}} l(\boldsymbol{y}, \boldsymbol{y_t}) \frac{\partial}{\partial \boldsymbol{m}} f(\boldsymbol{x}, \boldsymbol{m}) \tag{2.17}$$

This is performed by the backpropagation or *backprop* (Rumelhart et al., 1986) algorithm in FNNs and by the backpropagation through time (BPTT; Werbos, 1990; Williams and Zipser, 1995) algorithm in RNNs. The difference between both methods is basically that in BPTT the dependency of the loss function on the hidden layer activations needs to take into account the influence from previous timesteps as well (see equation 2.15).

Several different methods, called optimization algorithms, have been developed to decide how exactly the model parameters are to be adjusted. In the *Gradient Descent* (GD) method, the parameters are adjusted in the direction of the fastest reduction of the loss function L, that is, toward a negative gradient of L:

$$\boldsymbol{m_0} - \lambda \frac{\partial L}{\partial \boldsymbol{m_0}} \rightarrow \boldsymbol{m_1} \tag{2.18}$$

An alternative version of GD is the *Stochastic Gradient Descent* (SGD) method, in which GD is applied to only a subset (*mini batch*) of the input data. Other variants, which use the history of past gradient evaluations to further speed up the convergence to a global minimum and avoid local minima, include *Momentum*, *Adaptive Moment Estimation* (ADAM; Kingma and Ba, 2014), and RMSprop.

A well-performing deep learning model should have enough parameters in order to learn representations and predict well on the data used for the training (avoiding underfitting). Moreover, the model solution should generalize to data the model has not seen before (avoiding overfitting). The simplest way of fighting overfitting is adding more training samples to solve the task, which would incorporate additional constraints to the learnable model parameters. This additional data could include either new samples or sample copies generated by applying certain transformations on the input data (*Data augmentation*). For instance, these transformations can be rotations and horizontal or vertical flips in an image classification task. Alternatively,

the complexity of the model could be reduced, hence decreasing the number of learnable model parameters required to solve the problem. An example of this applied to CNNs is the use of depthwise separable convolutional layers (Chollet, 2017), instead of the conventional convolution.

Additional approaches to avoid overfitting implement methods such as *Early stopping* or regularization techniques such as *Weight regularization* and *Dropout* (Srivastava et al., 2014). As its name suggests, Early stopping interrupts the model training when the validation loss has stopped decreasing, even if the training loss continues decreasing after that point. Weight regularization techniques add a penalty term to the loss function. This has the effect of down-weighting the learnable model parameters (weights in fully connected layers and filters in convolutional layers) in the backward pass, so preventing small changes in the input from producing large changes in the output of a layer.

Dropout is a regularization technique that consists of randomly setting a fraction of the output features of a layer (*dropout rate*) to zero during the training. This introduces noise into the learned model parameters, so preventing the model from learning patterns which do not lead to a generalized solution. It has been shown that when using dropout during training, the predictions made by the trained model in testing data approximate Bayesian variational inference estimates (Gal and Ghahramani, 2016a,b). This is leveraged by the stochastic regularization technique Monte Carlo Dropout (MCD) in order to obtain uncertainties associated to the model predictions.

Chapter 3

Probing the Northern Chile Megathrust with Seismicity – The 2014 M8.1 Iquique Earthquake Sequence

Note

This chapter is a reformatted version of a paper with the same title published in *Journal of Geophysical Research: Solid Earth, Volume 124, Issue p. 12,935-12,954.* *https://doi.org/10.1029/2019JB017794*

Abstract

We used data from >100 permanent and temporary seismic stations to investigate seismicity patterns related to the 1 April 2014 M8.1 Iquique earthquake in northern Chile. Applying a multistage automatic event location procedure to the seismic data, we detected and located ~19,000 foreshocks, aftershocks and background seismicity for 1 month preceding and 9 month following the mainshock. Foreshocks skirt around the updip limit of the mainshock asperity; aftershocks occur mainly in two belts updip and downdip of it. The updip seismicity primarily locates in a zone of transitional friction on the megathrust and can be explained by preseismic stress loading due to slow-slip processes and afterslip driven by increased Coulomb failure stress due to the mainshock and its largest aftershock. Afterslip further south also triggered aftershocks and repeating earthquakes in several EW striking streaks. We interpret the streaks as markers of surrounding creep that could indicate a change in fault mechanics and may have structural origin, caused by fluid-induced failure along presumed megathrust corrugations. Megathrust aftershocks terminate updip below the seaward frontal prism in the outer continental wedge that probably behaves aseismically under velocity-strengthening conditions. The inner wedge locates further landward overlying the megathrust's seismogenic zone. Further downdip, aftershocks anticorrelate with the two major afterslip patches resolved geodetically and partially correlate with increased Coulomb failure stress, overall indicating heterogeneous

frictional behavior. A region of sparse seismicity at ~40- to 50-km depth is followed by the deepest plate interface aftershocks at ~55- to 65-km depth, which occur in two clusters of significantly different dip.

3.1 Introduction

The frictional behavior of subduction megathrusts is heterogeneous, exhibiting regions slipping either in a stable, that is, aseismic, manner or unstably during earthquakes (e.g., Lay et al., 2012; Perfettini et al., 2010). Mapping and understanding this variability is crucial because it ultimately determines the earthquake hazard potential. Measuring and inverting interseismic surface strain patterns due to variable plate locking based on GNSS measurements is one approach to achieve this (e.g., Bürgmann et al., 2005; Métois et al., 2013; Savage, 1983). However, sparse and one-sided data acquisition limits the resolution of such models. While the interseismic period is often seismically quiet, the coseismic and postseismic periods open a window in which the full spectrum of slip modes can be observed. Whereas coseismic slip inversions map the main asperities, aftershocks illuminate mainly the marginal areas (e.g., Asano et al., 2011; Beroza and Zoback, 1993; Das and Henry, 2003; Hsu et al., 2006; Kato et al., 2010; Mendoza and Hartzell, 1988; Perfettini et al., 2010; Wetzler et al., 2018; Woessner et al., 2006) for which slip behavior is varied and often not known.

Large earthquakes induce Coulomb stress changes (e.g., Harris and Simpson, 1998), which are often relaxed as afterslip (Hsu et al., 2006; Miyazaki et al., 2004; Perfettini and Avouac, 2007; Perfettini et al., 2010), poroelastic rebound (Hughes et al., 2010; Wang, 2000) and viscoelastic relaxation (Hergert and Heidbach, 2006; Hu et al., 2004; Rundle, 1978). Afterslip (Marone et al., 1991) in turn has been suggested as the primary mechanism driving aftershock seismicity (Hsu et al., 2006; Perfettini et al., 2010). Postseismic slip can be estimated from geodetic observations (e.g., Hoffmann et al., 2018, for the Iquique earthquake), albeit with the same limitations due to data coverage as for interseismic observations. Other indicators of aseismic slip may be repeating earthquakes (e.g., Igarashi et al., 2003), which are interpreted as repeating stick-slip ruptures on a fault that otherwise creeps (e.g., Nadeau and Johnson, 1998; Schaff et al., 1998). Hence, repeating earthquakes may directly image the spatiotemporal distribution of aseismic creep in the interseismic or afterslip in the postseismic stage (e.g., Huang et al., 2017; Meng et al., 2015).

On 1 April 2014, the M8.1 Iquique earthquake broke a central piece of the long-standing, >500-km-long northern Chile seismic gap, where no megathrust earthquake $M > 8$ had occurred since 1877 (e.g., Comte and Pardo, 1991; Ruiz and Madariaga, 2018). Seismic activity started several months before the mainshock, with events $M > 5$ in August 2013 and January 2014 (Bedford et al., 2015). After a large foreshock (M6.6, 16 March 2014), which occurred in the upper plate (Cesca et al., 2016), the more immediate foreshock series migrated down onto the megathrust, where three more events of $M > 6$ were registered, before the mainshock broke a compact asperity (An et al., 2014; Duputel et al., 2015; Gusman et al., 2015; Hayes et al., 2014;

Herman et al., 2016; Lay et al., 2014; Liu et al., 2015; Meng et al., 2015; Ruiz et al., 2014; Schurr et al., 2014; Yagi et al., 2014). The mainshock was followed by a vigorous aftershock sequence, including an M7.6 event, the largest aftershock, ~100 km south of the mainshock approximately two days later (Duputel et al., 2015; Schurr et al., 2014). The whole sequence was well recorded by the permanent seismic network of the Integrated Plate Boundary Observatory Chile (IPOC), complemented by several temporary deployments (see Section 3.2).

We use these data to derive a 10-month earthquake catalog covering the Iquique earthquake sequence. This catalog depicts the preseismic and postseismic deformation pattern in unprecedented detail, thus allowing us to correlate seismicity features with coseismic slip, geodetically inferred afterslip and interseismic locking, and induced static stress changes.

3.2 Data and Methods

We used data from 104 seismic stations located throughout northern Chile (Figure 3.1). 19 of these stations belong to the IPOC network, which was complemented soon after the first large Iquique foreshock by 30 additional temporary seismic stations deployed by the Oficina Nacional de Emergencia (ONEMI), the Departamento de Geofísica of the Universidad de Chile (DGF) and the Chilean National Seismological Center (CSN). Beginning on 11 April, the GFZ German Research Centre for Geosciences installed 23 additional stations. We further used data from several other temporary networks (see a list of data sources in Section S1). The continuous seismic data streams were scanned for earthquakes using a multistage automatic earthquake detection and location procedure adapted from Sippl et al. (2013), which is detailed in the supporting information (see Section S2) and summarized next. First, picking and repicking of P phases (Aldersons, 2004; Di Stefano et al., 2006; Lomax et al., 2000) based on an existing 1D velocity model (Husen et al., 1999), and picking of S phases (Diehl et al., 2009) were carried out alternated by event relocations (Lee and Lahr, 1975) and followed by calculation of local magnitudes (Bormann and Dewey, 2014). We then relocated the events in a 1D and a 2D velocity models (Eberhart-Phillips, 1993; Evans et al., 1994; Kissling et al., 1994; Thurber, 1983, 1993) determined from a subset of the earthquake catalog, as well as in a double-difference scheme (Waldhauser and Ellsworth, 2000) using catalog and cross correlation derived arrival time differences.

We finally associated several sets of published focal mechanisms (Cesca et al., 2016; Hayes et al., 2014; León-Ríos et al., 2016; GEOFON Data Centre, 2013 and Global CMT project (www.globalcmt.org) with our relocated events in order to better constrain hypocentral depths of earthquakes with unfavorable event-station geometry and to inform our interpretations of fault kinematics. Event uncertainties were estimated by relocating our catalog in the calculated 2D velocity model with a probabilistic scheme (Lomax et al., 2000).

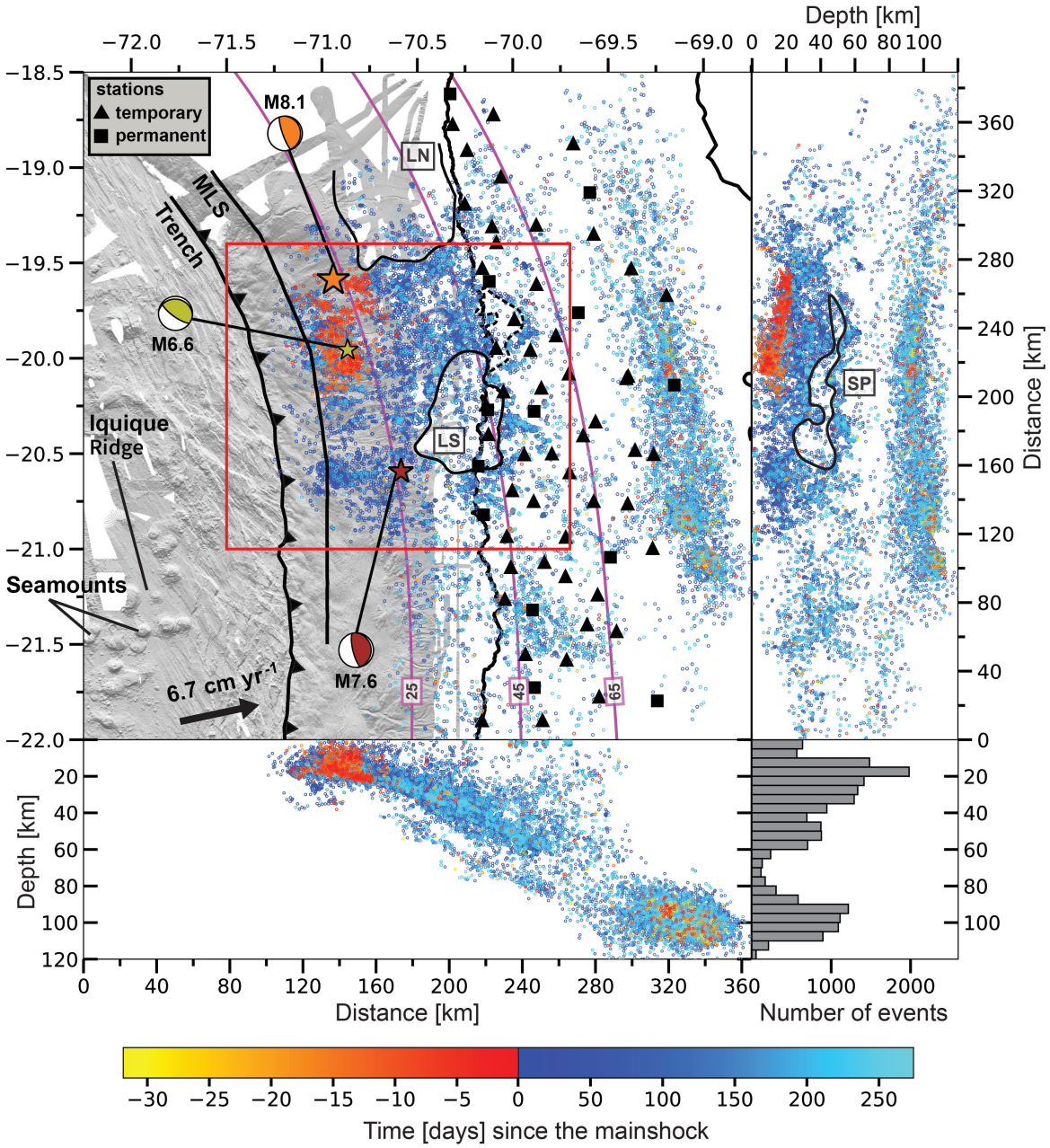

Figure 3.1: Seismicity in the area of the 2014 M8.1 Iquique earthquake. Seismic stations are shown as black symbols. Focal mechanisms (beachballs) and epicenters (stars) of the 2014 Iquique M8.1 mainshock, its largest M6.6 foreshock and its largest M7.6 aftershock (Hayes et al., 2014) are plotted in orange, yellow and brown color, respectively. Gray shaded bathymetry is from Geersen et al. (2018a). The limit between the lower and middle continental slopes (MLS) is taken from (Maksymowicz et al., 2018). Magenta contours show the slab surface (Hayes et al., 2018) at 25-, 45- and 65-km depth. Relocated earthquakes are presented in map view and as projections onto longitudinal and latitudinal planes, colored by days since the mainshock (events before the mainshock on top of events after). Lower right subplot shows a histogram of the event depths. Labeled regions outline an area of clearly decreased seismicity (SP) in the latitudinal section (dashed line in map view) and areas of low seismic activity (LN and LS) in the map view, which are discussed in the text. Red box outlines the region shown in Figure 3.2a.

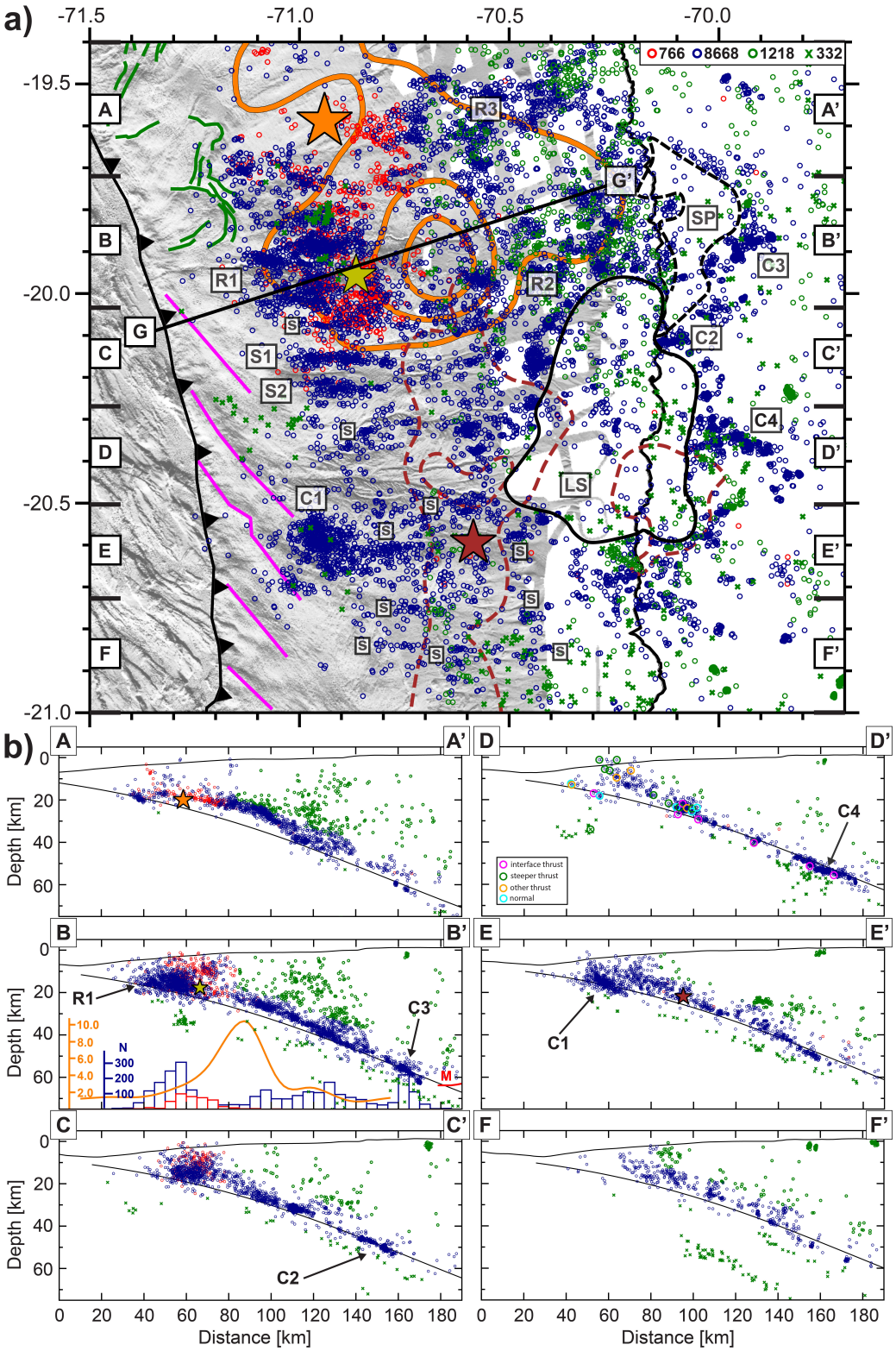

Figure 3.2: (a) Zoom into the study region, showing different seismicity clusters (labels C1–C4), streaks (labels S) and regions of high (labels R1-R3) and low (labels LS and SP) seismic activity discussed in the text. Cross Sections AA' to FF' of subfigure (b) are marked. Profile GG' is shown in Figure 3.8. Slip contours of the mainshock (orange, 2, 6, and 10 m) and the M7.6 aftershock (brown-dashed, 0.56 and 0.85 m) are from Duputel et al. (2015). Epicenter of these events and the M6.6 foreshock are shown as in Figure 3.1. Bathymetry features are marked with green (embayments) and magenta (NW-SE antiforms and synforms; Geersen et al., 2018a). Aftershocks are plotted as blue circles, on top of foreshocks in red. Presumed noninterface seismicity is plotted in green for upper-plate (circles) and intraslab (crosses) events. (b) Seismicity Cross Sections AA' to FF'. Slab surface (Hayes et al., 2018) and swath bathymetry (upper curve; topography at depth > 0 km) are drawn in black. In Profile BB', we also plot a histogram of event numbers, the coseismic slip (orange curve), and the location of the continental Moho (label M, Sodoudi et al., 2011). In Profile DD', focal mechanisms are plotted as open colored circles (see Figure 3.3 for a definition of the event classes).

3.3 Results

The distribution of seismicity is plotted in Figure 3.1. The catalog covers the time period from March to December 2014. The magnitude distribution of the earthquake catalog is shown in Figure S3. The magnitude of completeness is 2.7 for the foreshock sequence (before April 1) and 1.7 for the aftershock sequence (after 1 April), where the differences arise due to the improved station density provided by the temporary networks. The vast majority of events were recorded after the Iquique mainshock (17,826 vs. 1,137 before).

3.3.1 Spatial Distribution of Events

The events define two main depth ranges, the first of which covers an offshore region of shallower seismicity (hypocentral depths of 0-65 km) concentrated between 19.5°S and 21°S and forms an eastward dipping triple seismic zone (Bloch et al., 2014; Sippl et al., 2018), which we subdivided into upper plate (here defined as events located more than 10 km above the slab surface and east of 70.7°W), intraslab (events more than 5 km below the slab surface) and interface (remaining events) seismicity. The latter group of earthquakes belong mostly to the foreshock and aftershock sequences of the Iquique earthquake and will be analyzed in detail in this article (see Figure 3.2a). Since west of 70.7°W depth uncertainties are higher (>6 km; see Figure S2), we do not distinguish events located in this region and for simplicity color them as interface seismicity but will treat them with caution in our interpretation. The deeper onshore (east of ~69.5°W) band of intermediate-depth seismicity (80- to 120-km depth) corresponds to intraslab background seismicity. These events have been previously examined in great detail (Sippl et al., 2018, 2019), so we omit them from our analysis here. In the following, we will provide a more detailed description of observed seismicity features (clusters, streaks, and regions of high and low seismic activity), moving from west to east, that is, in the downdip direction, through three different sub-regions. We will adopt the coseismic slip model from Duputel et al. (2015) for comparison with the seismicity patterns that we observe, as it was derived from the most complete set of observations to date (seismic strong motion and high-rate GNSS time series, InSAR- and GNSS-derived static surface displacements, tide gauge, and DART buoy measurements of tsunami height) using a Bayesian inversion scheme, which does not apply smoothing. We will further discuss some features of this model relevant to our interpretations in section 3.4.1.

3.3.1.1 Updip Seismicity (0-25-km Depth)

Earthquake mechanisms in the region of the shallow plate interface just landward of the trench show a predominance of thrust faulting striking in a similar plane as the mainshock mechanism, but featuring a variety of dip angles. There are also smaller populations of differently oriented thrust events as well as some strike-slip and normal-faulting earthquakes (Figure 3.3).

In the northern part of the shallowest region of the megathrust, foreshocks and aftershocks occur in dense clusters that are located at the updip edge of the moderate-to-high (6-10 m)

coseismic slip zone, with aftershocks slightly but systematically displaced updip with respect to the foreshocks (region labeled R1 in map view and in cross section BB' in Figure 3.2).

Further south, shallow megathrust events locate along ~EW striking streaks (indicated with S in Figure 3.2a), with the prominent exception of a compact cluster (C1) situated directly west of the M7.6 aftershock epicenter. The two most distinct streaks locate SSW of the mainshock main rupture (coseismic slip > 6 m), containing mostly aftershocks and some foreshocks. In total, the northern (S1) and southern (S2) streaks are composed of 210 ($1.5 < M < 5.9$) and 148 ($1.4 < M < 5.4$) earthquakes, respectively. The Western Cluster C1 contains 452 earthquakes ($1.4 < M < 4.9$).

Figure 3.4a and S5a show the location of earthquakes at the latitudes of the two main streaks (S1 and S2) in map view and cross sections. The continuous, eastward dipping geometry outlined by these events and most of the focal mechanisms in this area indicate thrust earthquakes presumably located along the plate interface.

Most of the seismicity updip of the mainshock main rupture appears to locate on the plate interface in the northern cross sections (Figure 3.2b). However, some events appear to branch away upward from the slab surface and intersect the seafloor landward of the trench in at least one of the southern sections, indicating possible activation of a splay fault (Profile DD'; Figure 3.2b).

3.3.1.2 Mainshock Main Rupture and Central Seismicity (25- to 45-km Depth)

Deeper on the megathrust, the highest coseismic slip area of the Iquique earthquake contains only a few foreshocks and aftershocks (Figures 3.2a and 3.2b, cross section BB'). In the along-strike direction, this region is continuous into the area of highest slip of the M7.6 aftershock and its hypocenter to the south, which is surrounded by several minor ~EW striking seismicity streaks, similar to the ones observed at shallower depths.

Downdip of the mainshock main rupture, a ~SW-NE oriented region of increased aftershock seismicity that partially overlaps moderate-to-low (2-6 m) coseismic slip can be observed at depths of 25-45 km in the Cross Sections BB' and CC' and is labeled R2 in map view (Figure 3.2). Interestingly, no foreshock activity is observed at such depths. A similar region of aftershocks (R3), slightly more oriented toward E, extends from north of the area of maximum coseismic slip to ~19.4°S. Aftershocks in R3 connect to the concentration of foreshocks located updip of the mainshock's main slip and extend downdip to a similar depth as events in R2 (Cross Section AA'). In fact, between about 35 and 45 km, R2 and R3 merge, forming a belt of aftershocks just around the downdip edge of the deepest significant coseismic slip. Considerable upper plate seismicity is also noticeable above these two regions (Cross Sections AA', BB' and CC').

To the south and north of Regions R2 and R3, respectively, two largely aseismic areas (labeled LS and LN in Figures 3.1 and 3.2) are observed that extend at least 50 km in both directions. Northward, low seismicity rates (LN) prevail at this depth until the limit of the region covered by our catalog, whereas a number of isolated clusters and small ~EW seismicity streaks of

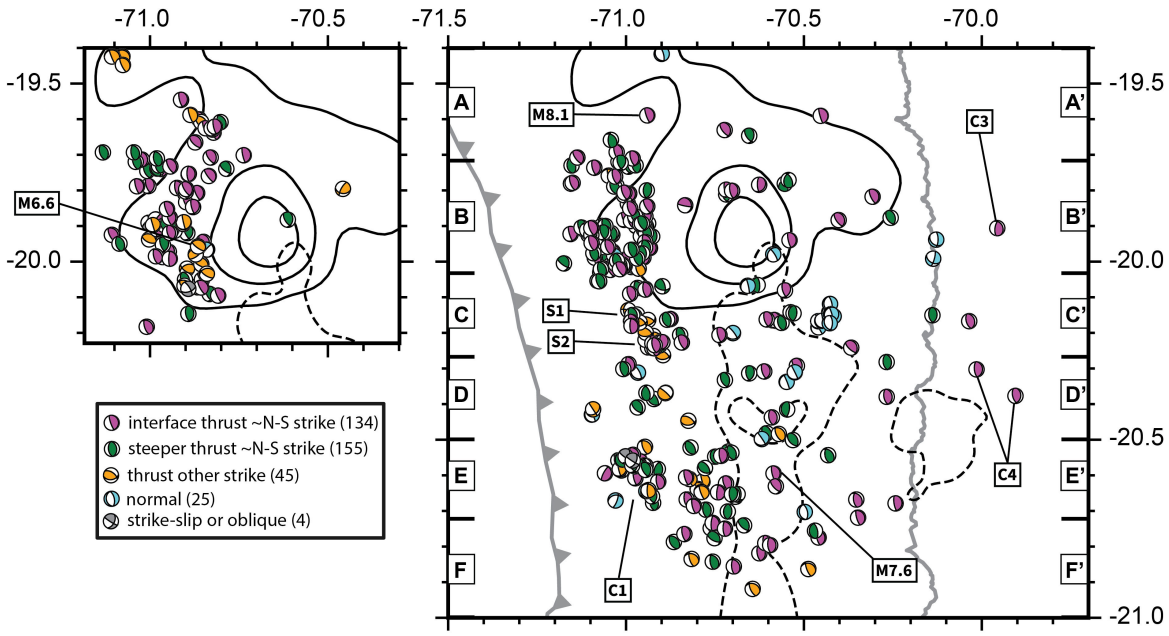

Figure 3.3: Focal mechanisms (beachballs) related to the Iquique earthquake sequence, compiled from global catalogs (GEOFON Data Centre and Global CMT project) as well as from Cesca et al. (2016), Hayes et al. (2014), and León-Ríos et al. (2016). We classified the mechanisms by their rake angles into thrust with strike ±45° from north (magenta and green), thrust with significant rotation (orange; deviation >45° from north), normal (cyan) and strike-slip or oblique faulting (gray). We further subdivided the first group into events with dip differences ≤15° ("interface", magenta) and >15° ("steeper", green) between the mechanisms and the plate interface plunge (Slab2; Hayes et al., 2018). Large map covers the time period after the Iquique earthquake, smaller panel before the mainshock. Other features as in Figure 3.2a. The same colors and definitions are used in Figure 3.2b, profile DD'.

aftershocks are observed to define the southern termination of the southern low seismicity area (LS) at ~20.5°S (Figures 3.1 and 3.2).

At this depth range, the belt of intense aftershock activity formed by R2 and R3 downdip of the main rupture outlines a sharply defined inclined plane in cross section (the plate interface), which is shallower and shows significant topography relative to the slab surface model Slab2 (Hayes et al., 2018; profiles AA' and BB' in Figure 3.2b). Focal mechanisms in this region are mostly consistent with rupture along the east dipping plate interface (Figure 3.3). There is, however, a cluster of presumed normal faulting events observed at ~35-km depth in swath C, at the western edge of the low seismicity area LS (Figures 3.2a, 3.3 and 3.4a).

3.3.1.3 Deep Interface Seismicity (45- to 65-km Depth)

Further downdip, a depth interval with only sparse aftershock seismicity is observed at ~40- to 50-km depth; it is labeled SP in Figure 3.1 (map view and latitudinal profile) and is also visible as a subtle minimum in the event depth histogram. Its depth extent is not constant along strike; it appears wider to the south (up to 25 km), where it connects to the southern postseismically aseismic patch LS described in Section 3.3.1.2.

This region of low activity separates the previously described offshore lower part of the central seismicity from the deepest interplate events, which are situated east of the coastline (Figures 3.1 and 3.2a). Here seismic activity concentrates in three clusters of aftershocks,

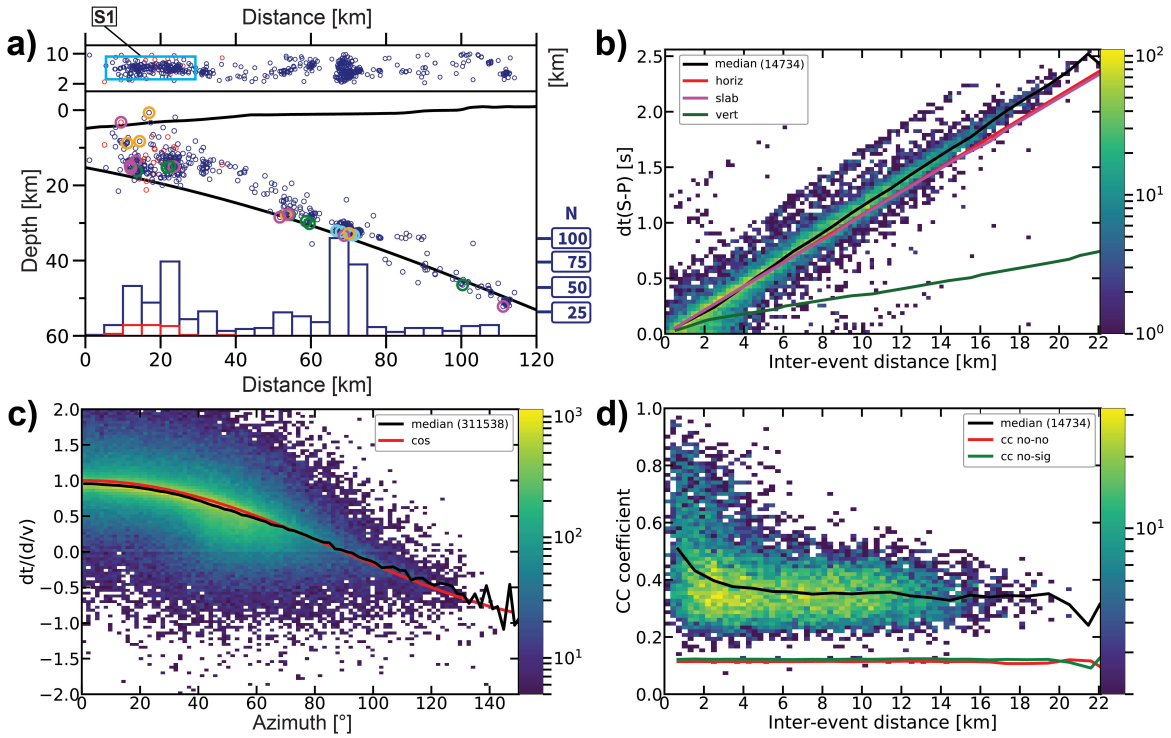

Figure 3.4: Detailed analysis of one seismicity streak (labeled S1 in Figure 3.2a). (a) Map and depth profile of the events in the region of the streak (red: foreshocks, blue: aftershocks). Focal mechanisms types are shown in cross section as bold colored circles (as in Figure 3.3). (b-d) 2D histograms for events in the streak (within cyan box shown in map view of subfigure (a)). Black curves show the results for the median of the data, which was calculated for a moving window with width of 1 km in subfigures (b) and (d) and 2° in subfigure (c). The number of data used in each plot is shown in parentheses. (b) *dt(S-P)* versus interevent distance. Colored curves show the results for the horizontally (horiz), along-dip (slab) and vertically (vert) oriented artificial streaks. (c) P and S wave arrival time differences (dt) for all event pairs within S1, normalized by the inverevent distance d divided by velocity V, versus azimuth. Red curve shows a theoretical cosine (coss) function expected for a streak oriented in EW direction. (d) CC coefficient versus interevent distance. Colored curves represent the noise level calculated from noise-noise (no-no) and noise-signal (no-sig) CC. Similar plots for the second major seismicity streak S2 are shown in Figure S5.

marked C2, C3 and C4 in Figure 3.2. They each contain a large number of events (168, 320 and 539 events, respectively) produced by low-magnitude earthquakes ($0.9 < M < 3.3$, $0.7 < M < 3.9$ and $0.5 < M < 4.5$, respectively), which are well detected under the footprint of the network. The two deeper clusters form two fingers beneath the Coastal Cordillera to depths of 55-65 km.

Even though we cannot be sure that events within C3 and C4 are located along the plate interface, they appear to connect further updip to the already described inclined planes of central-depth seismicity, but their on-profile plunges are observed to be slightly (C4) or clearly (C3) steeper than suggested by the Slab2 model (Figure 3.2b, Cross Sections BB' and DD'), implying a strong plate curvature. Alternatively, they could also indicate faulting in the lower plate, particularly the steeper dipping cluster C3. However, this alternative interpretation would imply a small angle between the plate surface and the presumed fault plane in the subducting oceanic plate, whereas known examples of elongated or sheet-like clusters in the lower plate tend to indicate faults at high angle to the plate interface (e.g., Fuenzalida et al., 2013), which also agrees with the focal mechanisms typically observed for intraplate events (Bloch et al., 2018). In addition, the preexisting fabric of the subducting plate is generally

characterized by the high angle normal faults formed at the outer rise.

3.3.2 Temporal Evolution of Seismicity

The rich seismicity content of our catalog allows us to inspect the time evolution of the foreshock and aftershock series with a high level of detail.

We observe that the foreshocks (from 15 March to 1 April, the mainshock origin time) started updip and NW from where the mainshock rupture later nucleated. Then the region south of the future mainshock epicenter was rapidly activated, with the occurrence of the largest M6.6 foreshock (Figure 3.5a and Movie S1). Events next propagated southward and northward for a couple of days, activating in the south the two most prominent observed seismicity streaks (S1 and S2) on 17-18 March (events occurred 15.5-14 days before the mainshock; Figures 3.5a and 3.5b). From this point onward, seismicity migrated almost continuously northward to where the Iquique earthquake finally ruptured (Figure 3.5a and Movie S1). These patterns of foreshock seismicity migration have been associated with slow-slip events (Ruiz et al., 2014; Socquet et al., 2017), which would have increased the Coulomb stress in the future mainshock rupture area and thus contributed to its nucleation and propagation (Kato and Nakagawa, 2014; Yagi et al., 2014).

In the first 2 days after the mainshock, seismicity gradually progressed southward until the M7.6 aftershock nucleated on 3 April, while aftershocks also occurred in the updip and central parts of the main coseismic rupture, including a few events in the two main seismicity streaks (Figure 3.5c and Movie S2). After the M7.6 event, the region surrounding its epicenter, the minor streaks there as well as the cluster (C1) westward of it were rapidly activated. The southern and northern onshore cluster (C4 and C3), seismically inactive until then, started to form just few hours later (Figure 3.5d and 3.5e and Movie S3).

During the following days and up to the 16th day after the mainshock (17 April), while the major streaks (S1 and S2) and clusters (C1, C3 and C4) remained active (Figure 3.5d), several small clusters of aftershocks progressively populated the region updip of the main rupture (R1) and formed the regions R2 and R3, previously described in Section 3.3.1.2 (Movie S3). From 18 April until the end of the study period, all the major streaks and clusters, as well as regions updip and downdip of the mainshock asperity, remained active (Figure 3.5f). Movie S4 shows a more detailed view of these aftershocks until 30 April. We note that earthquakes along the major streaks S1 and S2 did not propagate in any particular direction but occurred randomly over time (Movie S5).

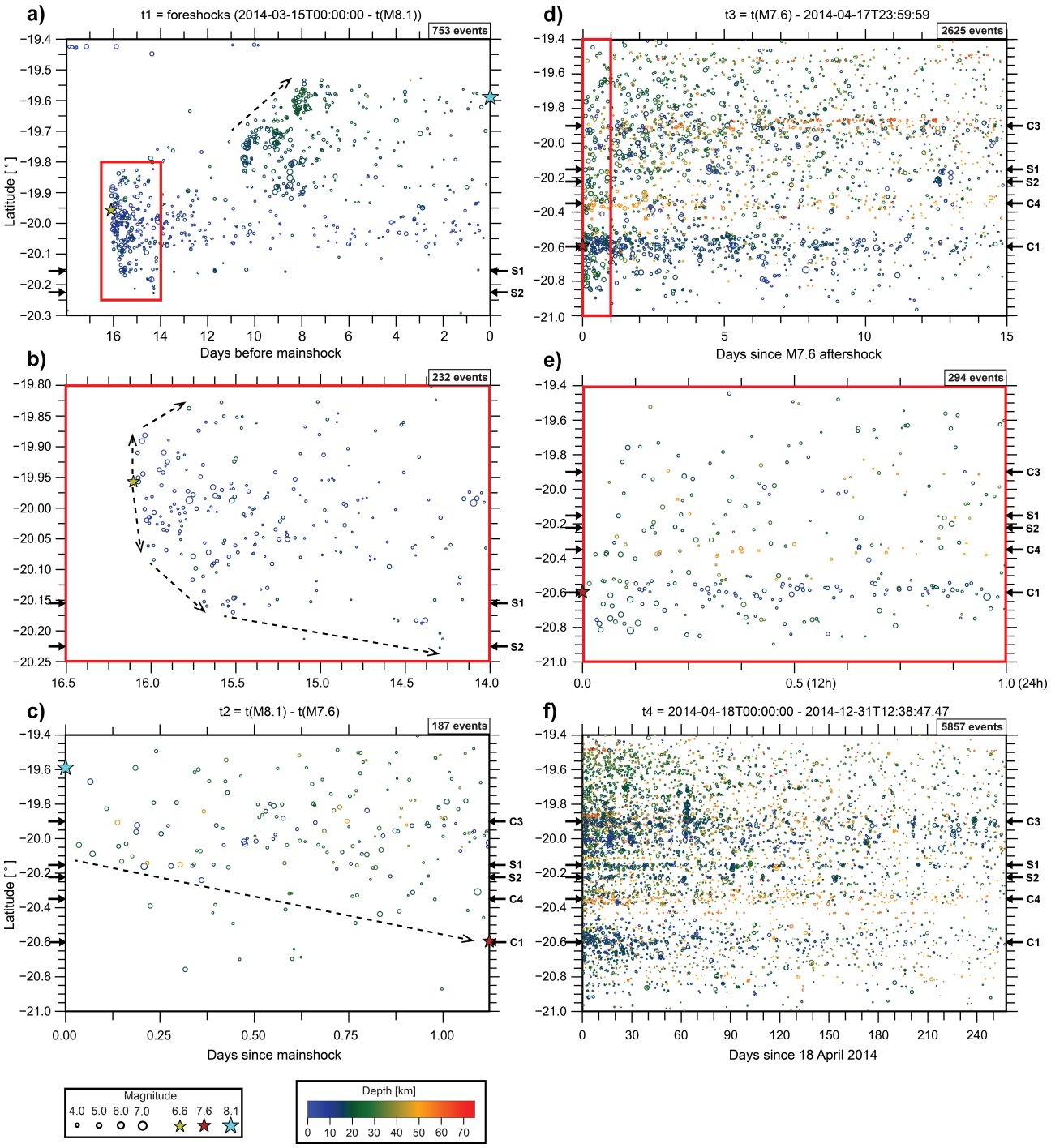

Figure 3.5: Plots of latitude versus time showing the temporal evolution of the foreshock (red circles in Figure 3.2, since 15 March 2014) and aftershock (blue circles in Figure 3.2) seismicity, subdivided in four time periods. Earthquakes are colored by depth and scaled by magnitude. The mainshock, the M6.6 foreshock, and the M7.6 aftershock are plotted as colored stars. The latitudes where the major seismicity streaks (S1 and S2) and clusters (C1, C3 and C4) are located are marked with small black arrows. Dashed black arrows indicate earthquake migration fronts in subperiods shown in subfigures (a)-(c). (a) t1: foreshocks between 15 March 2014 and the time of the mainshock, which are also shown in Movie S1. (b) More detailed view of the events in region shown in red box of subfigure (a), for the period between 16.5 and 14 days before the mainshock. (c) t2: time period between the mainshock and the largest M7.6 aftershock, which is also shown in Movie S2. (d) t3: time period between the M7.6 aftershock and 17 April 2014, which is also shown in Movie S3. (e) More detailed view of the events in region shown in red box of subfigure (d), during the 24 hr after the M7.6 aftershock. (f) t4: time interval between 18 April 2014 and the end of the study period (31 December 2014). Seismicity in this subperiod is further discussed in the main text and plotted in Figures 3.6, 3.7a and S6. A subsample of these events until 30 April 2014 is shown in Movie S4.

3.4 Discussion

3.4.1 Relation between Coseismic Slip, Afterslip and Aftershock Seismicity

Figure 3.6 shows an overlay of the aftershocks (since Day 17 after the mainshock) on a model of Coulomb Failure Stress change (ΔCFS) and the geodetically derived afterslip model (cumulative between Day 17 and 334 after the mainshock) of Hoffmann et al. (2018). The time period was selected for the best overlap of geodetic and seismic stations coverage. Afterslip is concentrated within two aseismic regions (see Section 3.3.1.2) NNE and SSE of the interseismically highly locked and coseismically ruptured asperity (Li et al., 2015; Schurr et al., 2014; Figures 3.6 and S6). Modeled afterslip is dominant in the downdip part of the rupture; however, synthetic tests (Hoffmann et al., 2018) indicate that farther offshore afterslip is not well resolved. The observed anticorrelation between aftershock seismicity and regions of high afterslip occurs mostly where afterslip is well resolved (Figure S15d of Hoffmann et al., 2018). For the ΔCFS model (Figure 3.6), we projected the coseismic slip from the mainshock and the M7.6 aftershock (Duputel et al., 2015) onto a representation of the megathrust (Hayes et al., 2018) that consisted of 232 individual fault patches. We used the software Coulomb3.4 (Lin and Stein, 2004; Toda et al., 2005) and chose the following parameters: Young's modulus 8×10^4 MPa, Poisson's ratio 0.25 (Lin and Stein, 2004) and friction coefficient 0.1 (Lamb, 2006). The receiver fault orientation was prescribed to be identical to the plate interface model.

The regions that experienced coseismic slip collocate with decreased Coulomb stress. They are surrounded by two lobes of elevated Coulomb stress updip and downdip of the two coseismic ruptures. Both of these lobes extend primarily from ~20°S southward, although the shallower branch also extends northward to the location of the mainshock epicenter (Figure 3.6). South of 20°S, the updip distribution of seismicity correlates well with the region of elevated Coulomb stress, whereas parts of the shallow seismicity directly updip of the mainshock's highest coseismic slip appears to have occurred in areas that experienced negligible Coulomb stress change due to an apparent shallow slip patch in the Duputel et al.'s (2015) model (Region R1; Figure 3.6). Nevertheless, the near-trench part of subduction earthquake slip models is notoriously underconstrained, and published slip models for the Iquique earthquake also show significant variability here. For example, the models of Liu et al. (2015) and Gusman et al. (2015) show a similar region of shallow slip as Duputel et al.'s (2015) model, whereas, for example, An et al. (2014), Hayes et al. (2014), or Ruiz et al. (2014) do not (see Duputel et al.'s Figure S12 for comparison of slip models). The most sensitive data to shallow slip are offshore DART buoy tsunami height measurements, which are incorporated in Duputel et al.'s (2015) and Gusman et al.'s (2015) models, but also in the model of An et al. (2014). ΔCFS on the megathrust is strongly linked to the assumed slip model, and hence, interpretations of details in the pattern should be made with due caution considering these uncertainties. What seems to be robust is a general pattern of increased ΔCFS updip and downdip of the mainshock rupture in regions of high aftershock seismicity.

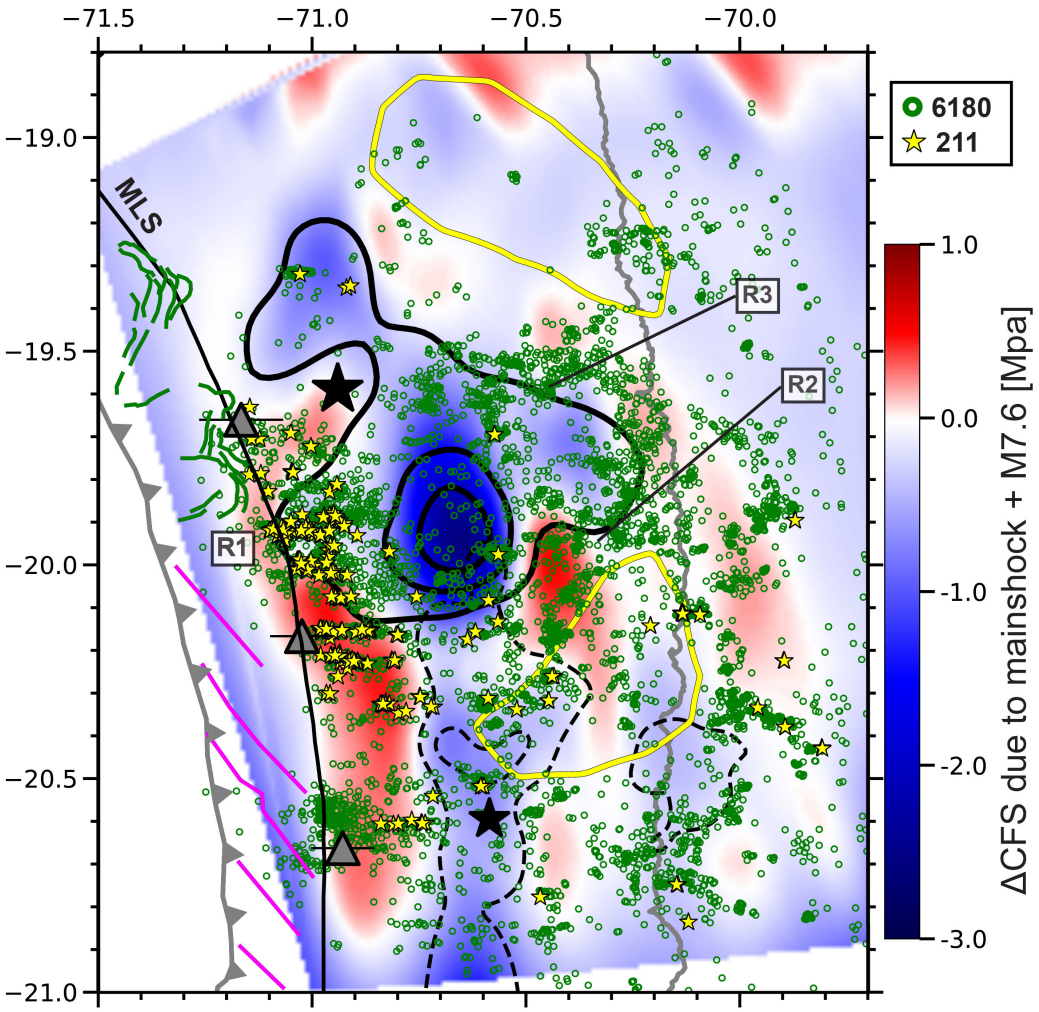

Figure 3.6: Comparison between presumed plate interface aftershock seismicity since Day 17 after the mainshock (18 April 2014, green circles) and other features in the study region. Small yellow stars represent repeating earthquake sequences (CC $\geq$ 0.95; see text for more details) with at least two events detected within the same time period. Gray triangles and lines across denote the location and width of subducting seamounts identified by Geersen et al. (2015). High afterslip patches are shown with yellow contours (0.4 m) for the period between Days 17 and 334 after the mainshock (Hoffmann et al., 2018). The calculated Coulomb Failure Stress changes (ΔCFS) due to the mainshock and the M7.6 aftershock are plotted into the map as blue (stress decrease) and red (stress increase) colors. MLS is shown as in Figure 3.1. Other features are plotted as in Figure 3.2a.

The concentration of aftershocks in a belt updip of the Iquique earthquake main rupture may be related to the updip extent reached by the mainshock slip. Similar patterns of aftershocks have been shown for other subduction earthquakes where strong postseismic activation updip of the main coseismic slip occurred and the rupture did not or did only slightly propagate into the shallowest domain A (in the classification of Lay et al., 2012; e.g., 2005 Sumatra [Nias] M8.7 and southernmost part of 2004 Sumatra [Aceh-Andaman] M9.2: Hsu et al., 2006; Tilmann et al., 2010. 2007 Peru [Pisco] M8.0: Perfettini et al., 2010. 2007 Sumatra [Bengkulu] M8.4: Collings et al., 2012; Konca et al., 2008). In contrast, updip postseismic activation was weak for coseismic ruptures that extended significantly into the shallowest domain A (e.g., 2011 Japan [Tohoku] M9.0: Asano et al., 2011. 2015 Chile [Illapel] M8.3: Lange et al., 2012; Tilmann et al., 2016).

Although it could not be resolved geodetically, this shallow region of the megathrust, which extends southward along the shallower lobe of positive ΔCFS, might have undergone significant

afterslip, driven by the static stress increase due to the mainshock and its largest aftershock. As we move further south, the observation of decreasing interplate locking (Li et al., 2015; Figure S6) leads to the expectation of afterslip after a large event; under partial locking, creep must occur during the interseismic phase, making the region also susceptible to coseismic stress changes. This spatial correlation between regions of reduced interseismic coupling, significant afterslip, and increased Coulomb stress has been previously observed in the areas surrounding main ruptures associated with the 2005 M8.7 Nias-Simeulue (Chlieh et al., 2008; Hsu et al., 2006) and the 2015 M8.3 Illapel (Huang et al., 2017; Tilmann et al., 2016) earthquakes, which could point to a similar explanation. Consequently, we interpret our results as indicative of an upper portion of the megathrust that predominantly undergoes aseismic afterslip. The occurrence of repeating earthquakes there may be another indication that this part of the fault is dominated by stable slip except where the seismic clusters and streaks occur (see Sections 3.4.2 and 3.4.3). To a certain degree, the two deepest seismicity clusters C3 and C4 may resemble the shallower seismicity streaks in fault mechanics and slip properties (Figure S6; see also Section 3.4.2). They occur close to the upper plate Moho (Figure 3.2b, profiles BB' and DD') and thus where the downdip transition to aseismic behavior was hypothesized to occur, except for fast and young plates, due to the contact of the lower plate with a serpentinized mantle wedge (e.g., Oleskevich et al., 1999). In other subduction zones, for example, Sumatra, however, the seismogenic zone has been shown to extend into the forearc mantle and the mantle wedge does not always seem to be serpentinized (Dessa et al., 2009; Simoes et al., 2004).

We observe more complicated seismicity patterns in the region downdip and south of the mainshock main rupture, where aftershocks located in regions R2 and R3 form a belt of seismicity located between two nearly aseismic patches of high afterslip (> 0.4 m), which are found in the vicinity of the M7.6 epicenter in the south and NE of the mainshock in the northern part of the study region (Figure 3.6). This belt of aftershocks locates in a region of increased Coulomb stress, significant locking (0.7-0.9) and moderate afterslip (0.3-0.4 m), possibly dominated by conditionally stable frictional behavior. However, considerable seismicity can also be observed in areas of negligible or negative ΔCFS. Near R2 high afterslip overlaps both with the coseismic rupture of the M7.6 aftershock and with high locking (> 0.9), partly accompanied by aftershocks, partly not. This region is difficult to interpret in detail and maybe best explained by a velocity-strengthening behavior with small-scale frictional heterogeneity that is not well represented in the overly smooth coseismic, interseismic, and postseismic models as well as in the resulting ΔCFS calculations. High frictional heterogeneity seems to be a hallmark of the deepest part of the megathrust expressed, for example, in localized coherent short-period seismic radiation from the mainshock rupture (Lay et al., 2012) and aftershock accumulation (Schurr et al., 2012). In fact, the deepest interface seismicity was observed to closely coincide with the short-period emission points in the 2010 M8.8 Maule earthquake aftershock time series (Palo et al., 2014). The northern patch of high afterslip observed north of 19.5°S has slightly reduced CFS in our model due to some coseismic slip. This region may be conditionally stable, where coseismic slip may propagate to but that behaves as velocity-strengthening at interseis-

mic low strain rates. This is corroborated by a moderate degree of interseismic locking observed in this patch (< 0.7; Figure S6).

Downdip of the belt of aftershocks formed by R2 and R3, we see a region of very sparse aftershock seismicity at ~40- to 50-km depth, approximately delineated by the coastline (Figures 3.1 and 3.2a). A similar along-dip seismicity gap, albeit more pronounced, was observed in the aftershock series of the 2010 M8.8 Maule earthquake (Lange et al., 2012) as well as before and after the 2015 M8.3 Illapel earthquake, and has been suggested to be a general pattern of the megathrust seismicity in central Chile (Lange et al., 2016). The aseismic area separating the shallower and deeper clusters has been shown to approximately correlate with the continental Moho in south-central Chile (Lange et al., 2012, 2016). Our relocated seismicity confirms the occurrence of an aseismic zone, suggesting that this particular depth segmentation of the megathrust seismicity may be generic along most of the Chilean subduction zone. However, in the northern Chile case, the continental Moho (Sodoudi et al., 2011) is imaged at a level below even the deepest interplate events, downdip of where the seismicity gap occurs (Figure 3.2b, cross section BB').

3.4.2 Linear Seismicity Streaks on the Shallow Plate Interface as Markers of Aseismic Creep

A characteristic feature of our relocated catalog is several ~EW striking seismicity streaks (Figure 3.2). Similar alignments of microearthquakes have been observed in continental strike-slip systems like the San Andreas fault (e.g., Rubin et al., 1999), where they were interpreted as indicative of small regions of brittle failure in a predominantly creeping section of the fault. Since these streaks are located offshore, that is, outside the seismometer network, the event-station geometry is unfavorable and location errors in EW direction are nonnegligible (Figure S2). In order to check whether these alignments are not artifacts of the location process, we performed a number of tests using travel time differences of P and S phases from the catalog as well as cross-correlation (CC) time delays between event pairs of streak earthquakes. In particular, we analyzed streaks labeled S1 and S2 in Figure 3.2a (cyan boxes in Figures 3.4a and S5a).

We calculated differential S-P travel times for event pairs ($e1$ and $e2$) detected at the same station, given by

$$dt_{(S-P)} = dt_S - dt_P = (t_{S(e2)} - t_{S(e1)}) - (t_{P(e2)} - t_{P(e1)}), \tag{3.1}$$

and plotted them against interevent distances (Figures 3.4b and S5b). $dt_{(S-P)}$ should be independent of a possible trade-off between event location and origin time, thus is a suitable proxy for the distance between event and station. For this calculation, we used travel times to stations within a small angle (10°) around the streak, so that the azimuth of the seismic rays connecting source and receiver is relatively uniform. We compared the observed linear trend to synthetically calculated $dt_{(S-P)}$ values for a perfectly horizontal chain of earthquakes,

an alignment along the slab surface as well as a vertical alignment. While we can rule out the last of these geometries, we can not distinguish between the two other scenarios based on our data. However, since most available focal mechanisms for intrastreak or neighboring earthquakes consistently show low-angle thrusting consistent with displacement along the plate interface (Figures 3.3 and 3.7c), we prefer to infer that these streaks are not purely horizontal but oriented along the megathrust.

Following Rubin et al. (1999), we also created plots of arrival time differences (dt) between event pairs in the streak registered at one station against the angle between interevent vector and recording station (azimuth), normalized by the interevent distance d divided by the seismic velocity V in the source region. In this calculation, we used P and S time delays for all the stations located between 19°S and 22°S that were calculated from CC, and we complemented those with corresponding delays obtained from catalog picks when a threshold CC coefficient (CC $>$ 0.5) was not reached for the CC delays. In case of a true linear arrangement, the observed pattern of time delays should resemble a cosine function, with maximum positive time delay at an azimuth of 0°, zero time delay at 90° and maximum negative delay at 180°. Figures 3.4c and S5c show such plots for the two main streaks in our data set, which suggest that the seismicity streaks we observe are most likely real features.

We additionally searched for repeating earthquakes, that is, sequences of very closely located seismic events with highly similar waveforms (CC $>$ 0.95 at at least three stations out of a set of five: PB01, PB02, PB08, PB11, and PB12). The cross correlation utilizes a 35-s time window that includes both P and S waves and applies a 1- to 4-Hz bandpass filter). We found a total of 211 such sequences containing at least two events occurred since Day 17 after the mainshock. The large majority of these sequences is situated in the shallow part of the plate interface, updip of the highest coseismic slip (Region R1) and within the major ~EW oriented seismicity streaks, although some are also found in various of the minor lineaments of microearthquakes that surround the M7.6 aftershock epicenter as well as in the most distinct (C2, C3 and C4) and some minor clusters located at the deepest part of the plate interface (Figure 3.6). Due to the fact that waveform similarity between two earthquakes increases the more closely located they are in space, CC coefficients for pairs of earthquakes along the main streaks show a tendency to decay with the distance between event epicenters (Figures 3.4d and S5d); thus, repeaters within these streaks lie in the uppermost left part of these plots.

The observation of numerous repeating earthquakes occurring inside seismicity streaks on the shallow portion of the megathrust most likely hints at the presence of afterslip (e.g., Huang et al., 2017; Igarashi et al., 2003; Meng et al., 2015), which loads a number of velocity-weakening patches on the interface, that then rupture seismically (Nadeau and Johnson, 1998; Schaff et al., 1998). In this sense, seismicity streaks would be markers of surrounding creep along the plate interface. A number of repeaters were also found where the deepest onshore clusters C3 and C4 are located, in a region of moderate afterslip. This could suggest similar frictional properties in the deepest and shallowest parts of the plate interface. As previously seen by Waldhauser et al. (2004) in a strike-slip environment, the seismicity streaks observed in our catalog occur

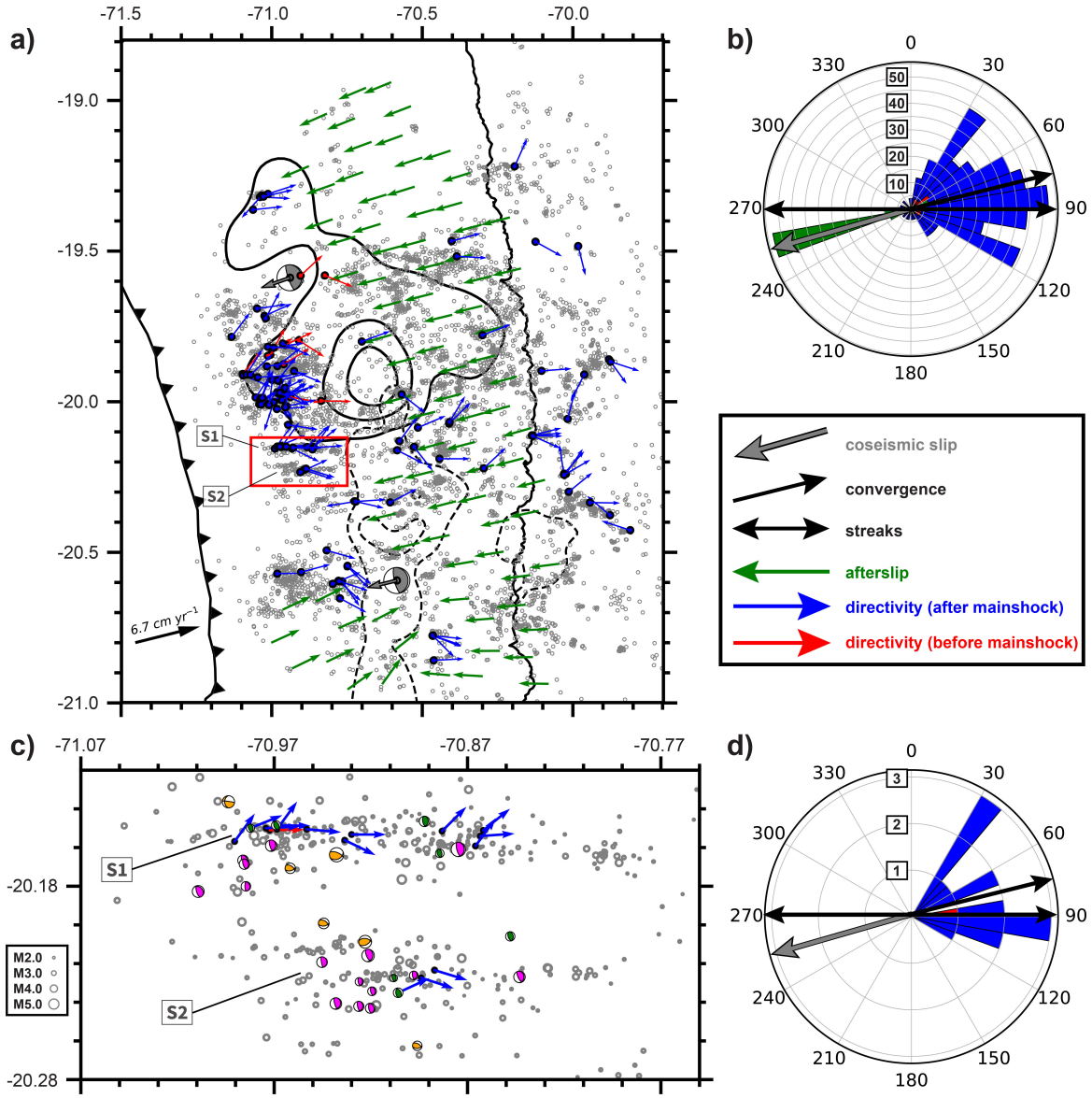

Figure 3.7: (a) Comparison of direction vectors of different processes. Vectors represent the direction of plate convergence (black; Angermann et al., 1999), the rake (gray; movement of the hanging wall) of the mainshock and M7.6 aftershock focal mechanisms (Hayes et al., 2014), the afterslip (green; Hoffmann et al., 2018), and rupture direction estimates for repeating events that occurred before (red) and after (blue) the Iquique earthquake (Folesky et al., 2018). The red box outlines a region surrounding the major streaks, shown in subfigure (c). Interface aftershocks are plotted as gray circles. (b) Rose diagram of directions plotted in the map of subfigure (a). Peripheral numbers are direction azimuths measured in degrees clockwise from the north (0°), and radial numbers are counts of afterslip directions and augmented counts of rupture directions (x4, for enhancing visibility). Gray arrow points toward the average (253.7°) rake direction between the mainshock and the M7.6 aftershock. (c) Map view zoomed in the region of the major streaks (labeled S1 and S2). Focal mechanisms collected in this area are shown as colored beachballs (as in Figure 3.3). Interface foreshocks and aftershocks and focal mechanisms are scaled by magnitude as shown in inset on the left. (d) Rose diagram for the distribution of observed rupture directions in the region of subfigure (c). The correlation between orientations of the major streaks and the preferential ~EW rupture direction observed in subfigure (b) is maintained at this local scale.

along the edges of the slip patches (both for the mainshock and the M7.6 aftershock), thus in regions with significant slip gradient. From this observation, and following Bennington et al. (2011), we speculate that such linear features may also be indicators of a change in the mechanical properties of the fault and consequently could have played an important role

in limiting the updip and along-strike extents of both the Iquique mainshock and its largest aftershock ruptures.

If the streaks were caused by rock destruction induced, for example, by subducting seamounts, we would expect them to be subparallel to the plate convergence vector (76°; Angermann et al., 1999), coseismic slip vectors (253.7° direction of movement of the hanging wall, average between mainshock and M7.6 aftershock; Hayes et al., 2014) or afterslip vectors (250-260°; Hoffmann et al., 2018). However, the ~EW orientation of the seismicity streaks deviates slightly but consistently from all of these (Figures 3.7a and 3.7b). Recently published rupture direction estimates of Iquique earthquake aftershocks (Folesky et al., 2018) appear to strike also predominantly toward east, that is, in line with the streaks (Figure 3.7c and 3.7d). This could imply that streaks are caused by a local persistent mechanism, rather than related to coseismic or postseismic processes.

3.4.3 Role of Forearc and Incoming Plate Structures

The northern Chile convergent margin is sediment starved (< 500 m of sediment thickness) due to the aridity of the Atacama desert and has been classified as erosive (von Huene and Scholl, 1991). Erosive-dominated margins generally have continental wedges composed of two segments: (1) an outer wedge at the most seaward part, composed mainly of framework rock of the upper plate and with a small frontal prism of slump debris derived from land and possibly some accreted seafloor sediments, and (2) an inner wedge further landward made of more consolidated framework rock (von Huene et al., 2004; Wang and Hu, 2006). Small frontal prisms formed of mostly sedimentary debris eroded from the continental slope have been suggested throughout northern Chile based on reflection seismics, bathymetry and gravity data: ~5 km wide at ~23°S, the latitude of the Mejillones Peninsula (Sallarès and Ranero, 2005); ~30 km wide at ~22°S (Contreras-Reyes et al., 2012); ~10-20 km wide at the latitudes of the Iquique earthquake rupture (Geersen et al., 2015); and on average ~20-30 km wide, estimated from Arica to Mejillones Peninsula (Maksymowicz et al., 2018).

We observe that the updip limit of seismicity (foreshocks and aftershocks) seems to correlate well with the approximate location of the limit between lower and middle continental slopes (MLS, Maksymowicz et al., 2018; Figure 1), which thus marks the boundary between a seaward aseismic (20-30 km wide) and a landward seismically activated portion of the plate interface (Figures 3.6 and 3.8a; Maksymowicz et al., 2018). The gradual increase in coseismic slip landward of this boundary (Figure 3.8b) also supports the idea of a small aseismic frontal prism. It has been suggested that this frontal prism is composed of highly fractured and hydrated rocks and acted as a barrier for the updip propagation of the coseismic deformation during the 2014 Iquique (Contreras-Reyes et al., 2012; Maksymowicz et al., 2018), the 2010 Maule and the 1960 Valdivia earthquakes (Contreras-Reyes et al., 2010, 2017). A transitional zone may be present landward of the MLS, where seismicity rate and coseismic slip increase landward, and the degree of fracturing smoothly decreases in the same direction (Maksymowicz et al., 2018; Wang and Hu, 2006).

In the context of dynamic Coulomb wedge theory (Wang and Hu, 2006), we envisage that the frontal prism and the transitional region form the outer wedge, overlying an updip velocity-strengthening and a transitional (frictionally heterogeneous) segment of the plate interface, respectively. The inner wedge is located further landward, overlying the velocity-weakening (seismogenic) segment of the megathrust, where we observe the highest coseismic slip (Figure 3.8a). During the Iquique earthquake, the outer wedge was elastically compressed and brought to or beyond the critical state, with an increase in basal and internal stresses and pore fluid pressure. After the earthquake, once the seismogenic segment is again locked, the outer wedge is expected to creep in order to relax the coseismically induced stress, returning to a stable state.

The concentration of aftershocks located in the postulated transitional region updip of the main coseismic rupture (region R1; Figure 3.6) may be mostly driven by afterslip due to postseismic stress relaxation, in a region of the plate interface with dominant conditionally stable behavior. Given the higher uncertainties of the shallow coseismic slip, it is also plausible that afterslip occurs in small-scale velocity-strengthening ("stable") areas hidden in the stress shadow of the nearby velocity-weakening coseismic asperity. In this case, the stable areas would not slip interseismically and thus would appear with high degree of interseismic locking (> 0.9, Figure S6); however, they would be able to undergo considerable postseismic slip (Bürgmann et al., 2005; Hetland and Simons, 2010; Métois et al., 2012). The significant concentration of ~NS oriented thrust faulting aftershocks with dip angles considerably steeper than the plate interface in this region (Figure 3.3) may indicate additional activation of offshore upper plate faults, associated with the outer wedge deformation due to its postseismic stress relaxation.

Stress accumulation caused by slow-slip prior to the mainshock has been suggested to account for the concentration of foreshocks in the same region on the plate boundary (Kato and Nakagawa, 2014; Maksymowicz et al., 2018; Yagi et al., 2014). And a number of thrust mechanisms, with significantly rotated planes relative to the plate interface (like the largest foreshock, M6.6; Figure 3.3), has additionally been interpreted as indication of some degree of preseismic upper plate deformation (Hayes et al., 2014). Both interplate and upper plate events (foreshocks and aftershocks) form the seismically active wedge volume observed here (Figure 3.8a). However, this spread seismicity also reflects very poor event depth resolution, as the depth uncertainties in this area are the highest (Figure S2).

The incoming Nazca plate features substantial relief, both bending-related horst-and-grabens and spreading fabrics, as well as single prominent seamounts associated with the Iquique Ridge offshore our study area (Figure 3.1; Geersen et al., 2018a). Rugged subducting seafloor has been associated with low interplate locking and aseismic creep on megathrusts (Wang and Bilek, 2014), and we further believe that it may have also had a strong influence on some characteristic features observed in the aftershock pattern of the Iquique earthquake. In addition to seamounts visible in the bathymetry, several currently subducting seamounts have been postulated based on the analysis of reflection seismics (Figure 3.6; Geersen et al., 2015). Further indications of underthrusting lower plate relief come from large embayments and NW-SE trending antiforms

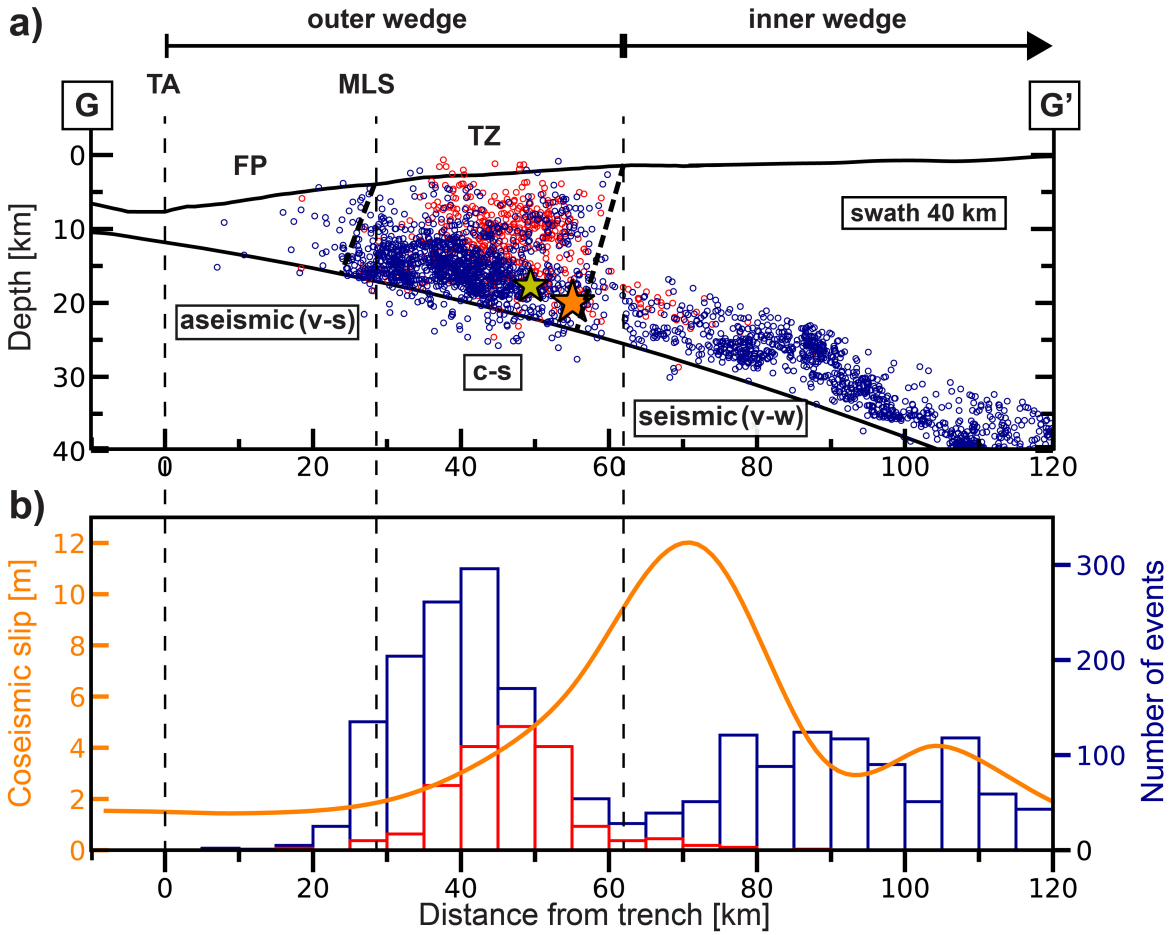

Figure 3.8: Cross Section GG' (for location see Figure 3.2a) through the seismicity in the region of the Iquique earthquake rupture. (a) Seismotectonic interpretation of the profile, in terms of frictional behavior along the plate interface and the structural composition of the continental wedge. Slab surface (lower curve; Hayes et al., 2018) and swath bathymetry (upper curve) are indicated. Hypocenters of the mainshock (orange star), largest M6.6 foreshock (yellow star), foreshocks (red circles) and aftershocks (blue circles) are plotted. Dashed vertical lines mark the position of the trench axis (TA), MLS and the presumed boundary between the outer and inner wedges. Dashed oblique lines separate different regions of seismicity. We interpret the outer wedge as composed of a small frontal prism (FP) and a transition zone (TZ). Along-dip, the shallow megathrust is subdivided in an uppermost aseismic velocity-strengthening (v-s) portion, an intermediate conditionally-stable (c-s) area and a deeper seismic velocity-weakening (v-w) region possibly located below the inner wedge. (b) Coseismic slip distribution along dip (orange line) plotted on top of a downdip histogram of plate interface event (foreshock and aftershock) numbers.

and synforms in the lower continental slope seen in high-resolution bathymetry (Figures 3.2a and 3.6; Geersen et al., 2018a). Embayments are observed in the northern part of the study region, partially at the location of the northernmost proposed seamount (Figure 3.6). Antiforms and synforms, located southward of 20°S and westward of the shallowest aftershock seismicity, have been interpreted as possible expressions of creep on the plate interface and internal deformation of the upper plate, related to subducting lower plate relief (Geersen et al., 2018b). A second suggested seamount appears to correlate with where we see the major seismicity streaks in our relocated earthquake catalog, whereas the southernmost such seamount locates at the position of the large offshore seismicity cluster C1 in the south, where we did not find repeating earthquakes (Figures 3.2 and 3.6). Focal mechanism solutions of the events in the earthquake cluster C1 are significantly more diverse than elsewhere (Figure 3.3), which indicates complex

deformation involving the upper plate, possibly caused by increased stress due to the nearby M7.6 aftershock event activating a network of fractures that may have been generated by the subducting seamount. Correlations regarding subducting seamounts, however, should be considered with care. The positions of seamounts in Geersen et al. (2015) are prescribed by the three analysed active seismic transects; seamounts may easily also be present elsewhere, where they could not be imaged. Further, the proposed seamounts were suggested based on images of two-way travel time, which can either be explained with plate interface topography or with strong lateral variations of seismic wavespeed. As observed in the region updip of the main coseismic rupture, several thrust events, which are oriented ~NS and feature dips steeper than the slab surface, likely also indicate the activation of offshore upper plate faults, possibly splay faults forking off from the megathrust (Profile DD', Figure 3.2b, may display such a geometry). Even though location uncertainties are largest in this shallowest segment of the megathrust (Figure S2), this interpretation is supported by the occurrence of two thrust events with rotated planes and steeper dip angles within the suspected splay (Figure 3.3). Furthermore, a similar splay was observed in the aftershock sequence of the 2010 Maule earthquake in central Chile based on ocean bottom data (Lieser et al., 2014), that is, not subject to the same location uncertainty, such that aftershock activity on a large splay fault is certainly plausible.

Recently, as has been commonly observed at exhumed fault surfaces (Engelder, 1975; Petit, 1987), a 3D seismic reflection study of the Costa Rica margin has shown for the first time well-defined "corrugations" on a subduction zone megathrust (Edwards et al., 2018). They are expressed as regular shallow plate interface relief with a dominant wavelength, and an orientation deviating slightly from the plate convergence direction. From higher reflection amplitudes along the corrugated portions, Edwards et al. (2018) inferred a higher fluid content/pressure there and suggested that corrugations may act as fluid conduits. These features were seen in a region where seamounts, plateaus and ridges are being subducted, which has been shown to cause extensive fracturing and faulting of the upper plate (e.g., Sak et al., 2009; Zhu et al., 2009), and also appears to inhibit seismic rupture propagation (DeShon et al., 2003; Wang and Bilek, 2011). Subducting rough topography and resulting fracturing and modification of the plate interface in the direction of convergence have also been linked to the formation of linear seismicity features observed offshore the Aleutian-Alaska subduction zone (Abers et al., 1995). Structural features aligned to the fault slip direction, such as corrugations or blocks of resistant host rock, have also been proposed as possible causes of seismicity streaks in strike-slip faulting environments (Rubin et al., 1999).

We envision that the megathrust offshore Iquique is not only being deformed and fractured by the effect of underthrusting rough relief, but it may also be corrugated as imaged in the Costa Rica subduction zone. Numerous fractures produced by subducting seamounts may allow the escape of fluids from the overpressured oceanic crust to the plate interface, breaking the sealed plate boundary (Audet et al., 2009). Similar to a proposed possible mechanism for generating slip-parallel streaking of tremors in the Cascadia subduction zone (Ghosh et al., 2010), the released fluids at the subduction interface may flow through corrugated conduits along the

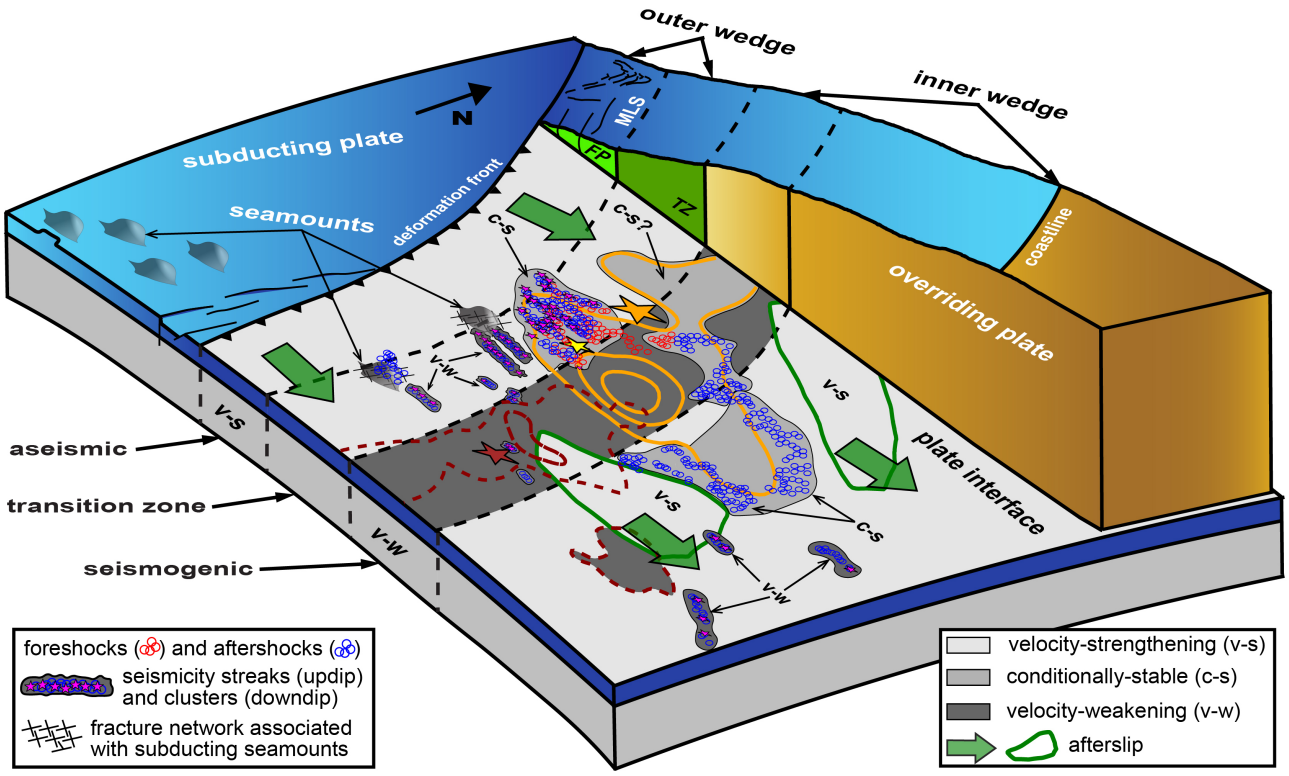

Figure 3.9: Synoptic model summarizing the proposed frictional heterogeneity along the plate interface and the seismotectonic segmentation of the continental wedge in the study area. The shallowest portion of the plate interface is aseismic (velocity-strengthening), underlying a highly fractured and deformed frontal prism (FP) in the outer wedge. Further downdip, a transitional zone is characterized by seismically unstable (velocity-weakening, along seismicity streaks) and conditionally-stable (updip of mainshock asperity) regions embedded in aseismically creeping (velocity-strengthening) areas. Here fracture networks caused by subducting seamounts and repeating earthquakes hint at afterslip loading unstable/conditionally-stable regions. Above this segment, the outer wedge forms a transitional zone (TZ) approximately below the middle continental slope, where fracturing of the upper plate likely gradually decreases landward. Deeper downdip, the seismogenic (velocity-weakening) segment of the plate interface correlates with the highest coseismic slips, below the inner continental wedge. This seismically unstable region changes in the downdip direction to a frictionally heterogeneous behavior, where we observe an anticorrelation between aftershock seismicity (presumably located in conditionally stable areas) and the two (velocity-strengthening) areas with the highest afterslip. Finally, seismicity clusters are observed in the deepest portion of the plate interface. Features not described in the legend are plotted as in Figures 3.2 and 3.6.

megathrust. By altering the effective normal stress due to the pressure exerted on the conduit walls, fluids may eventually cause shear failure on the interface and trigger earthquakes along the conduits, forming seismic structures such as the ~E-W striking streaks in our relocated catalog. Some of these ruptures may give rise to the repeating earthquakes along the streaks that we observe, in case a single fault patch is sequentially activated several times.

In Section 3.4.2, we noted that the streaks are aligned with a preferential eastward rupture direction. Folesky et al. (2018) speculated that directivity may be caused by the bimaterial nature of the subduction process, which predicts downdip rupture direction, due to the higher compliance of the underlying plate material. They estimated that this effect (Weertman, 1980) would be enhanced due to the geometrical limitation of possible rupture directions given the along-dip elongated shape of the observed repeater asperities. We believe that the existence of corrugations is a plausible structural explanation for the observed ~EW oriented seismicity streaks and thus for the associated elongated repeater asperities. Besides, the corrugations may

behave as a pathway along the megathrust through which favored rupture directions could be reinforced.

The presence of corrugations is suggested by the earthquake mechanisms observed in the region of the major streaks. Although most of them indicate ~NS striking thrust faulting occurring at the plate interface, a number of them are small ($M \leq 4$) thrust events with dips steeper than the slab surface (Figures 3.4a, S5a and 3.7c), indicating local deviations in the rupture planes that could reflect structural features such as corrugations. Some shallow events located at the edges of both major streaks, with thrust focal mechanisms striking along planes rotated with respect to the plate interface orientation, may additionally hint at upper plate deformation taking place in areas surrounding the observed streaks and presumed megathrust corrugations (Figures 3.4a, S5a and 3.7c). Nevertheless, the scarce number of focal mechanisms in this region is far from being sufficient, and additional observations are required to confirm the existence of such features.

3.5 Conclusions

We conducted a detailed seismicity study focused on the aftershock sequence following the 2014 M8.1 Iquique earthquake. Our results allow us to illuminate small-scale heterogeneity on the plate interface and the seismotectonic behavior of the overlying continental wedge. We present our key results in a conceptual model in Figure 3.9. The continental forearc in the study region can be divided into an outer and an inner wedge. The outer wedge is composed of (1) a highly fractured and deformed frontal prism (FP) overlying the shallowest aseismic segment of the plate interface and (2) a transitional region (TZ) where fracturing of the upper plate gradually decreases landward, located above a transitional frictional segment of the megathrust. The inner wedge overlays the seismogenic segment of the megathrust, where we observe the highest coseismic slip (Figures 3.8a and 3.9).

The FP may have acted as a barrier for the updip propagation of the mainshock rupture. Further downdip, the plate interface is characterized by velocity-weakening and conditionally stable regions embedded in a primarily velocity-strengthening area, where aseismic afterslip driven by increased static Coulomb stresses takes place. Afterslip is further favored by subducted seafloor relief, which may impose conditions such as fracturing of plate interface rocks, upper plate deformation, and reduction of plate coupling. The presence of earthquakes with rupture planes significantly different from those expected for megathrust events may indicate upper plate deformation and the activation of splay faults. A prominent cluster of seismicity (C1) may be linked to the fracture network developed by a subducting seamount. Numerous repeating earthquakes updip of the main coseismic rupture and inside seismicity streaks support the occurrence of afterslip, despite it not being detectable geodetically (Figure 3.9).

We interpret the earthquake streaks and their repeating earthquakes as markers of creep that takes place in the surrounding aseismic regions of the plate interface. The streaks may indicate a change in mechanical properties of the fault and consequently could have played an important

role in limiting the updip and along-strike extents of the ruptures of both the Iquique mainshock and its largest aftershock. We speculate that the streaks may have a structural origin, possibly being triggered by shear failure induced by fluid transported along conduits formed due to presumed corrugations present on the plate interface.

In the central portion of the megathrust, aftershock seismicity surrounds the asperities of the mainshock and the M7.6 aftershock, which can mostly be explained by the Coulomb stress redistribution triggered by these events in a region dominated by mixed frictional behavior. Moreover, maximum afterslip is located in two patches flanking and below the main coseismic rupture, which correlate with areas of low aftershock seismicity. These areas are thus interpreted as dominated by a velocity-strengthening frictional regime (Figure 3.9).

The southern low seismicity area connects downdip to a depth interval of sparse aftershock seismicity at ~40- to 50-km depth, which supports previous observations that hint at a possible depth segmentation of the interface seismicity along the Chilean margin. Downdip of this region, the deepest interplate earthquakes reach ~55- to 65-km depth in two individual clusters beneath the coastal cordillera, which feature a significant difference in dip from each other. At this depth, the observation of repeating earthquakes and moderate afterslip could imply similar frictional conditions as proposed for the shallow megathrust transitional zone (Figure 3.9).

Chapter 4

DeepPhasePick: A Method for Detecting and Picking Seismic Phases from Local Earthquakes based on Highly Optimized Convolutional and Recurrent Deep Neural Networks

Note

This chapter is a reformatted version of a paper with the same title submitted at *Geophysical Journal International*, which is available as a preprint version from *EarthArxiv* under *https://doi.org/10.31223/X5BC8B*

Abstract

Seismic phase detection, identification and first-onset picking are basic but essential routines to analyse earthquake data. As both the number of seismic stations, globally and regionally, and the number of experiments greatly increase due to ever greater availability of instrumentation, automated data processing becomes more and more essential. E.g., for modern seismic experiments involving 100s to even 1,000s instruments, conventional human analyst-based identification and picking of seismic phases is becoming unfeasible, and the introduction of automatic algorithms mandatory. In this paper, we introduce DeepPhasePick, an automatic two-stage method that detects and picks P and S seismic phases from local earthquakes. The method is entirely based on highly optimized deep neural networks, consisting of a first stage that detects the phases using a convolutional neural network, and a second stage that uses two recurrent neural networks to pick both phases. Detection is performed on three-component seismograms. P- and S-picking is then conducted on the vertical and the two-horizontal components, respectively. Systematic hyperparameter optimization was applied to select the best model architectures and to define both the filter applied to preprocess the seismic data as well as the characteristics of the window sample used to feed the models. We trained DeepPhasePick using seismic records extracted from two sets of manually-picked event waveforms originating

from northern Chile (~39,000 records for detection and ~36,000 records for picking). In different tectonic regimes, DeepPhasePick demonstrated the ability to both detect P and S phases from local earthquakes with high accuracy, as well as predict P- and S-phase time onsets with an analyst level of precision. DeepPhasePick additionally computes onset uncertainties based on the Monte Carlo Dropout technique as an approximation of Bayesian inference. This information can then further feed an associator algorithm in an earthquake location procedure.

4.1 Introduction

One of the most fundamental components in any earthquake hypocenter estimation routine is the identification and picking of seismic phases, primarily P and S phases from local earthquakes. In the past, this task was commonly performed manually by analysts, who identified each phase arrival based on their training and experience. However, as the available seismic data has rapidly increased over time, the use of automatic phase detection algorithms has become increasingly necessary.

These automatic algorithms encompass detectors which are based on the energy or frequency content of the seismic waveforms such as STA/LTA (e.g., Aldersons, 2004; Allen, 1978; Baer and Kradolfer, 1987; Diehl et al., 2009; Di Stefano et al., 2006; Earle and Shearer, 1994; Sleeman and Van Eck, 1999), those based on correlations of template waveforms against continuous seismic data (e.g., Gibbons and Ringdal, 2006; Harris, 1991; Van Trees, 1968), and detectors based on the representation of seismic data as a linear combination of orthogonal basis waveforms (Harris, 1997, 2001; Scharf and Friedlander, 1994).

Phase detectors based on frequency or energy have been used in the past as part of multi-stage automatic earthquake location procedures that allowed the creation of high-quality earthquake catalogs, e.g., for the Northern Chile region (Sippl et al., 2018; Soto et al., 2019b). Correlation detectors, such as the matched filter method (Van Trees, 1968), rely on the similarity of known template waveforms and have been widely used for detecting repeating earthquakes (e.g., Folesky et al., 2018; Huang et al., 2017; Igarashi et al., 2003; Kato and Igarashi, 2012; Nadeau and Johnson, 1998) or searching for missing events (e.g., Peng and Zhao, 2009; Ross et al., 2019a; Shelly et al., 2007) in different tectonic regimes. Subspace detectors, which are based on orthogonal basis waveforms, have been used for identifying earthquakes associated to aftershock sequences and low-frequency tremors (Harris and Dodge, 2011; MacEira et al., 2010).

Despite the fact that energy-based phase detectors do not require strong prior waveform knowledge, an increase in their ability to detect small onsets also implies higher false positive rates. Correlation detectors can achieve very low false positive rates, but they can solely detect similar signals to the ones already present in the selected template waveforms (Harris, 1991). Subspace detectors can further extend the range of detected signals, depending on which detection threshold and subspace dimension parameters are used. However, they are not easy to implement efficiently since they require a high computational cost (see e.g., Harris and Paik,

2006).

All the above described methods exploit a priori assumed attributes of the signals such as energy or waveform similarity. Instead, deep learning is a representation-learning method that learns multiple layers of features (a so called "neural network") directly from input data by applying non-linear transformations sequentially (Goodfellow et al., 2016; LeCun et al., 2015). In the supervised version of deep learning, the goal is to reduce the calculated error (objective function) between the network predictions and the known labels that the network is fed with. This is achieved by adjusting the units (weights) in each layer after backpropagating the gradients of the objective function, computed over the units in the last layer, to the input layer of the network. The network training consists of repeating the whole process until there is no further decrease in error (Goodfellow et al., 2016; LeCun et al., 2015).

Thanks to the enormous increase in computational calculation capacity, deep learning has in recent years shown stunning results in diverse fields such as image recognition (Krizhevsky et al., 2012), speech recognition (Hinton et al., 2012), language translation (Sutskever et al., 2014) and particle Physics (Kaggle challenge, 2014).

Convolutional neural networks (CNNs) are a class of deep neural networks specialized for processing grid-like data (Goodfellow et al., 2016). CNNs stand out due to the fact that they are computationally more efficient, easier to train, and have proven to effectively generalize learned features in many supervised tasks, ranging from image and document recognition (Krizhevsky et al., 2012; LeCun et al., 1990, 1998; Simard et al., 2003; Taigman et al., 2014) to seismic waves simulation (Moseley et al., 2020), or volcanic ash particles classification (Shoji et al., 2018).

Originally inspired by the properties observed in the primary visual cortex (PVC) of the mammalian brains (Hubel and Wiesel, 1959, 1962), modern CNN architectures are usually formed by several stages of consecutive operations of convolution, non-linear transformation, pooling, and regularization. In the first stage, a filter (kernel) performs local weighted sums (convolutions) through the input data. This linearly activates or detects local features, emulating the behavior observed in the so-called simple cells of the PVC. The resulting feature maps are transformed by applying a non-linear function. A further function then summarizes its statistics in a local scale (pooling). Usually the maximum (max pooling) within a sub-region is computed in this second function, inspired by the function of PVC complex cells (Goodfellow et al., 2016; LeCun et al., 2015). Convolution is highly efficient and effective because it takes advantage of the local connectivity and invariance to location exhibited by meaningful learnable features in natural signals. Firstly, it detects meaningful locally-connected features by using kernels smaller than the input, which allows units in deeper layers to preserve information from the input layer. Secondly, it applies the same filters all over the input, thus permitting the learning of only one common set of parameters instead of many, as well as ensuring that the same learned representations can be found at different locations in the data. In addition to convolution, pooling in neighboring units in a layer can reduce the dimensions of the learned representation and makes it invariant to small perturbations in the input (Goodfellow et al., 2016; LeCun et al., 2015). Furthermore, regularization techniques help the learned features to

better generalize to new unseen data, preventing model overfitting. One of the more effective and most frequently applied examples of regularization is Dropout (Srivastava et al., 2014), where noise is introduced in the output features of a layer aiming at inhibiting the network from learning non-meaningful patterns.

Traditional deep neural networks, including CNNs, are limited by having no memory, hence they cannot extract meaningful contextual information from sequentially structured data. Recurrent neural networks (RNNs; Rumelhart et al., 1986) overcome this constraint by simplistically mimicking human beings' mechanism for processing external information. They achieve this by implementing an internal loop that iterates over a sequence of data, element by element, while keeping an internal memory state of data they have already processed. The development of the Long-Short Term Memory RNNs (LSTMs) was one of the major breakthroughs in RNN architecture, since it solved the so-called vanishing gradient problem, whereby simple RNNs proved to be incapable of preserving long-term dependent information (Hochreiter and Schmidhuber, 1997). By carrying information to later points in the sequence, LSTMs are capable of retaining patterns, and therefore learning, from very long sequences. LSTM models also make use of internal dropout and recurrent dropout that help prevent overfitting, so improving model performance. Another significant advancement was the invention of Bidirectional RNNs (BRNNs; Schuster and Paliwal, 1997), which make use of two recurrent layers to process sequences in both forward and backward directions. BRNNs can make learning more effective in sequences where both the past and future contexts can provide valuable insights.

In seismology, CNN models have been recently applied for detection (Dokht et al., 2019; Ross et al., 2018a,b; Woollam et al., 2019; Zhu and Beroza, 2019) and association (McBrearty et al., 2019) of P and S wave arrivals, as well as for earthquake localization (Kriegerowski et al., 2019; Perol et al., 2018; Zhang et al., 2020). RNN-based networks have been used for predicting approximate earthquake times and locations (Panakkat and Adeli, 2009), and for seismic phase association (Ross et al., 2019b).

The present work adds to these previous studies, and introduces DeepPhasePick, a new automatized two-stage method for detecting and picking seismic P and S phases from local earthquakes, entirely based on highly optimized deep neural networks. The first stage in DeepPhasePick consists of an adaptive CNN architecture trained for detecting the phases. Here the type of data preprocessing, as well as the length and position of the seismic phase windows used to train the network, were included among the optimizable hyperparameters. Phase picking is conducted in the second stage by applying two additional adaptive Bidirectional LSTM (BLSTM) networks, which were trained specifically to predict P- and S-phase time onsets. The onsets, and their uncertainties, are determined on time windows defined based on the predicted probabilities of the P and S phases in the detection stage.

Here we demonstrate how the optimized CNN network in the first stage of DeepPhasePick, trained on a rather small dataset of labeled phases in comparison to previous studies, is able to detect with high accuracy P and S phases from local earthquakes of different tectonic regimes. We also show how, by leveraging the information of the detected seismic phases, the optimized

BLSTM models trained for picking are able to predict P- and S-phase time onsets with analyst levels of precision, while also avoiding inherent human bias.

4.2 Data and Methods

4.2.1 Earthquake Catalog Datasets

In this study, we used two sets of manually picked event waveforms (Figure 4.1). The first set (S1) consists of 1,125 events from the time period between 1996-12-02T00:54:33.89 and 1997-11-20T02:24:59.41 (Schurr et al., 2006). The second set (S2) contains 1,196 events which occurred from 2007-06-14T01:18:52.24 to 2007-12-13T07:23:39.24 in the area of the 2007 M7.7 Tocopilla earthquake (Schurr et al., 2012). S1 contains mostly plate interface events, whereas S2 contains mostly intermediate-depth intraplate events.

From these earthquake catalogs, we extracted three-component seismograms which we subdivided into three window classes: 25,647 P-phase, 25,647 Noise (N), and 14,397 S-phase windows. 16,234 P, 16,234 N and 8,061 S of these samples were obtained from S1, and 9,413 P, 9,413 N and 6,336 S samples from S2. We used these seismic windows as input data to train adaptive neural networks in phase detection and phase picking tasks, as will be described in the next sections.

4.2.2 Hyperparameter Optimization of Adaptive Neural Networks

The architecture of a neural network is defined by its hyperparameters, such as the number of layers in the network, the training learning rate, and the batch size used during training. The selection of the model hyperparameters is key when implementing a supervised deep learning task, since it may lead to a significant improved performance of the trained model, especially when the available data is limited.

Commonly used hyperparameter optimization approaches, which rely on grid or manual search, have been shown to be less efficient than an optimization based on random search (Bergstra and Bengio, 2012). However, all the above mentioned methods select the subsequently sampled values without an informed criterion. This makes the optimization less effective, since the sampling of hyperparameter values which do not lead to improved performance tends to require a significant amount of time. In contrast, Bayesian optimization selects the next sampled hyperparameters based on previous evaluations. This has proven more efficient in terms of balancing exploration-exploitation of the search space, time consumption, and model performance results, compared to random search (Bergstra et al., 2013b; Hinz et al., 2018).

In Bayesian optimization, an objective function is minimized by mapping past evaluations of the hyperparameters to a surrogate probabilistic model of the objective function, which is then more simply optimized instead. In this work, we used a Bayesian approach that optimizes a surrogate model defined by the Tree Parzen Estimator (TPE) algorithm, as implemented in the Python library Hyperopt (Bergstra et al., 2013a). Basically, TPE algorithm applies the Bayes

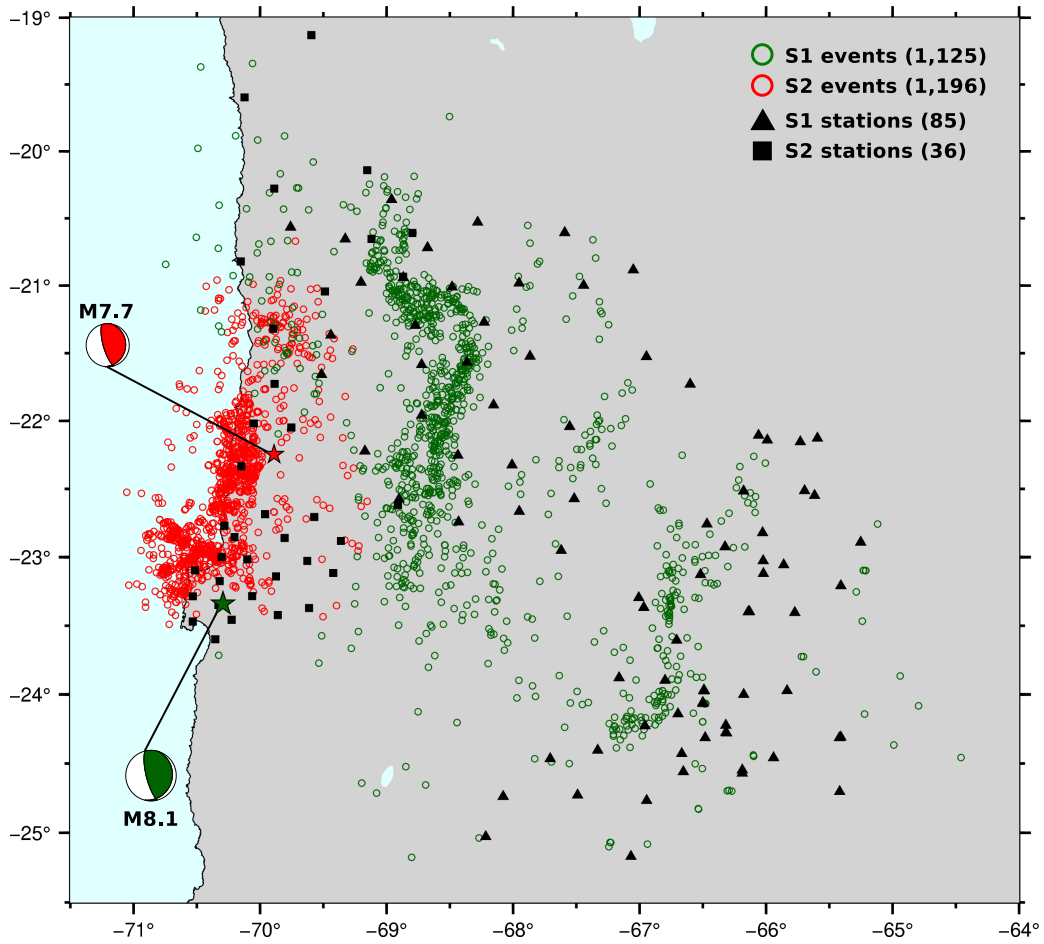

Figure 4.1: Events datasets from which window samples picked by analysts were extracted for training the phase detection and picking models in DeepPhasePick. Events in S1 and S2 datasets are plotted as green and red circles, respectively. Focal mechanisms (beachballs) and epicenters (stars) of the 1995-07-30 M8.1 Antofagasta and the 2007-11-14 M7.7 Tocopilla earthquakes are plotted in green and red, for reference. Window samples used in this work come from the picked stations plotted as black triangles (from S1) and squares (from S2).

rules to create two different probability distributions for the hyperparameters, depending on the score reached by the objective function. The next set of sampled hyperparameters is then selected based on the expected improvement in the objective function (Bergstra et al., 2011).

We implemented two types of objective functions, which are minimized in order to optimize the training of different adaptive neural network architectures designed to solve two tasks: phase detection and phase picking. These architectures will be described in detail in the next sections. The hyperparameter optimization process was performed using one GPU NVIDIA GeForce RTX 2080. Multiple iterations (trials) were run for each task, with a different hyperparameter configuration being attempted in each trial, until a best-performing trained model, i.e., a model with minimum error, was found. The adaptive neural networks were built in Python using the machine learning framework Keras (Chollet, 2015) with Tensorflow as backend (Abadi et al., 2015).

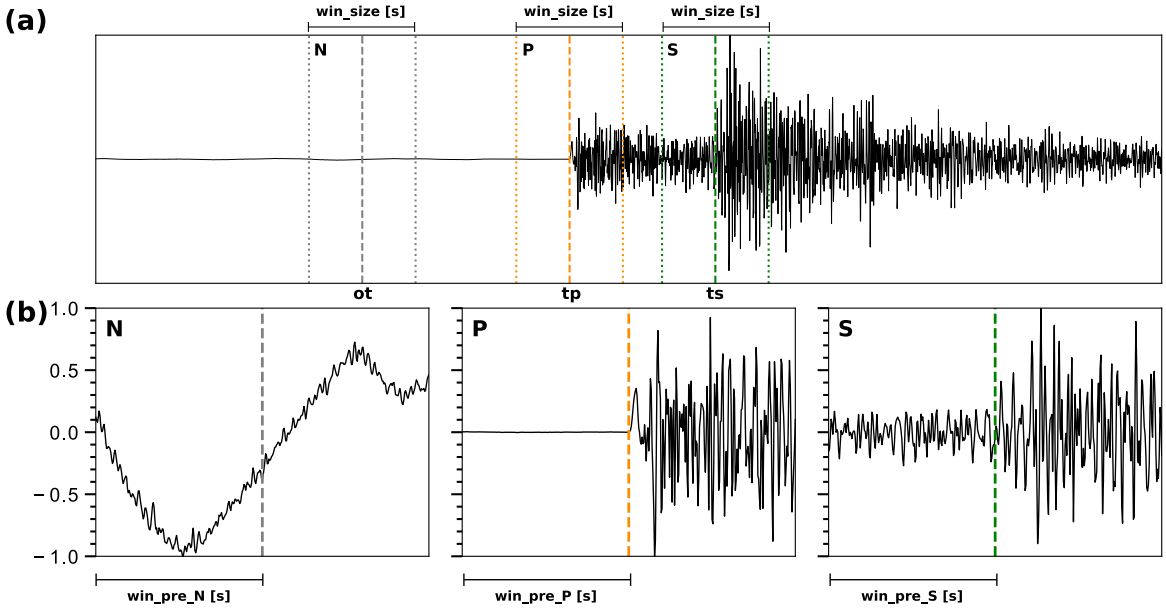

Figure 4.2: Seismic window extraction. (a) Example of an event waveform, outlining how the seismic window of three classes (N, P, S) are extracted for the optimized training of the phase detection model. N-, P-, and S-class windows are extracted, respectively, in the vicinity of the event origin time (ot), and the true P- (tp) and S- (ts) onset times found by analysts. The hyperparameter win_size [s] defines the length of the extracted windows, which is equal for the three classes. (b) Zoom into the three-class windows. The starting time of the window before ot, tp and ts is defined by the hyperparameters $frac\_pre\_N$, $frac\_pre\_P$, and $frac\_pre\_S$ for classes N, P and S, respectively. Here $win\_pre\_C = win\_size \times frac\_pre\_C$, where $C = N$, P or S.

4.2.3 Phase Detection as a Supervised Multi-class Classification Task

We implemented the phase detection stage as a supervised multi-class classification task, based on an adaptive neural network formed by two blocks of deep layers. The first block is made up of between one and five convolutional layers, which extract abstract representations (features) from three-component seismograms input (samples) and help reduce their dimensionality. The output of the convolutional layers is then flattened before entering the second block, formed by between one and four fully-connected dense layers. All the units in a fully-connected layer are connected to all the units in the preceding layer. This allows them to better learn

correlated features throughout the input data, rather than only the locally-connected features learned by convolutional layers. Hence, dense layers are well suited for the final classification stage in the adaptive network. Since the network is adaptive, the number of layers in each block, as well as specific layer variables such as the number of convolutional filters or dense units, are adjusted during the hyperparameter optimization.

We chose to use depthwise separable 1D convolutional layers for the feature extraction in the first block. Depthwise separable convolutional layers implement a 2-step convolution process, first performing independent convolutions on each channel of the input data and then combining individual channel outputs through a pointwise (1x1) convolution (Chollet, 2017). This type of convolution is appropriate for learning patterns from multi-component seismic data, such as 1D amplitude time series from seismic waveforms, since different features can be extracted from the three input channels (the three components of each seismogram) independently. This may help the network extract specific patterns from, e.g., the two horizontal components, in order to better identify S phases. Furthermore, compared to standard convolutional layers, depthwise separable layers have the additional advantage of helping reduce model overfitting, since they have fewer weights to adjust, and therefore require fewer calculations. This further reduces the overall computational cost, since less calculation time is required to complete the multiple iterations over different network architectures involved in the hyperparameter optimization.

Outputs from each of the convolutional and dense layers in the network, except from the output of the final dense layer, are passed through a layer that applies a non-linear activation function that can be either a rectified linear unit (ReLU; Nair and Hinton, 2010) or a sigmoid, according to the hyperparameter selection. Additional Batch Normalization (Ioffe and Szegedy, 2015) and Dropout (Srivastava et al., 2014) layers are stacked to the activation outputs in both blocks. In the convolutional block, a further 1D Max Pooling layer is applied between each Batch Normalization and Dropout layer. The output of the final dense layer is passed instead through a softmax activation function. This outputs a vector of three probabilities, adding up to 1.0, each expressing the likelihood of a sample belonging to one of the three possible classes tested: P, S, or N. The greatest among these probabilities determines the predicted class of the three-component record.

Prior to initializing the optimized training, we randomly extracted an independent test set consisting of 1,440 P, S and N three-component time series windows, made up of 888 records from S1 and 552 records from S2 datasets for each class, so as to keep the proportion of samples present in each dataset. In order to reduce the possibility of biased phase classification due to an imbalanced class distribution in our dataset (39% P, 39% N, 22% S), the remaining data was balanced out at each optimization trial by randomly discarding the surplus samples of the over-represented P and N classes. Then, the balanced data was shuffled and assigned to the training and validation sets before carrying out the model training. The resulting training, validation, and test sets added up to 75% (32,393), 15% (6,478), and 10% (4,320) of the total balanced data samples respectively. During the training, features of the waveforms were learned from the training set and the weights in the network layers were adjusted based on this. The

updated model was then used to predict the sample classes in the validation set, and adjust the model hyperparameters accordingly. The final performance of the best trained model was evaluated using the test set. We assigned to each sample in the training, validation and test sets, binary vectors representing the sample class as follows: [1, 0, 0] for P class [0, 1, 0] for S class, and [0, 0, 1] for N class. These vectors were used as the known labels which were compared with the vector of class probabilities predicted by the model.

Before entering the network, the seismic records in the training, validation and test sets were linearly detrended, resampled at 100 Hz, and then normalized by the maximum amplitude across the three waveform components. We included additional preprocessing criteria as optimizable hyperparameters. First, the hyperparameter *pre_mode* controls the type of filter applied to the seismic data, which can be band-pass (2-10 Hz), high-pass (>0.2 Hz) or no filter. The length of the extracted seismic windows is given by the hyperparameter *win_size* (2-5 [s]), which we imposed to be equal for the three classes. This makes data manageable by the network during the training. Finally, three additional hyperparameters varying between 0.2 and 1.0 ($frac\_pre\_N$, $frac\_pre\_P$, and $frac\_pre\_S$), define the fraction of the extracted windows placed before the event origin time, the true P-phase onset and the true S-phase onset for classes N, P, and S, respectively (see Figure 4.2).

The total space of hyperparameters searched during the optimization is presented in Table 4.1. In order to search for the best-performing network trained in detecting seismic phases, we ran 1,000 hyperparameter optimization trials (~9 minutes per trial). In each trial we used the validation accuracy as the optimizable metric and we trained the model for up to 60 epochs using the categorical cross-entropy loss function. To speed up the training process, we additionally applied an early stopping callback that stopped the training if the validation accuracy did not increase in 6 epochs.

4.2.4 Phase Picking as a Supervised Sequence Binary Classification Task

As mentioned previously, phase picking in DeepPhasePick uses two optimized adaptive network architectures, each formed by one or two BLSTM layers, which we implemented as a supervised sequence binary classification task.

We trained a first model for picking P phases, using as input the amplitude time series taken from vertical-component seismograms. A second model was trained for picking S phases, based on the amplitude time series of the two horizontal-component records. We trained each model to learn patterns from seismic data that account for the transition from noise to signal in both P and S phases from local earthquakes, therefore allowing identification of P and S time onsets. BLSTM-based networks are suitable for such task, since they are able to extract and retain meaningful features while processing input seismic sequences in both chronological and anti-chronological order, hence learning dependencies between phase onsets and neighboring seismic patterns. A final dense layer in both picking models applies a sigmoid activation function, which outputs the probability of the timesteps in the time series corresponding to either noise or seismic signal.

Table 4.1: Hyperparameter search space optimized during the training of phase detection models. The second column lists the hyperparameter values that could be sampled during the optimization. The third column shows the best-performing value found after 1,000 trials. The hyperparameter *pre_mode* defines the type of filter applied to the seismic data before entering the network as follows: *pre_mode* = 1, 2, 3 correspond to no filter applied, band-pass filter (2-10 Hz), and high-pass filter (>0.2 Hz), respectively. Hyperparameters *win_size*, $frac\_pre\_P$, $frac\_pre\_S$, and $frac\_pre\_N$ control the length and time position of the extracted window samples used for the training, as described in the main text and Figure 4.2. The range tested for the number of filters shown here relates only to the first convolutional layer. Another hyperparameter, not shown here, allows the number of filters in subsequent convolutional layers to be equal to or twice the value in the current layer.

Hyperparameter	Range tested	Best-performing value
	Block of convolutional layers	
Number of layers	[1, 2, 3, 4, 5]	5
Number of filters	[2, 4, . . . , 30, 32]	12, 24, 48, 96, 192
Kernel size	[3, 5, 7, 9, 11, 13, 15, 17, 19, 21]	17, 11, 5, 9, 17
Activation function	[ReLU, sigmoid]	ReLU, ReLU, ReLU, ReLU, sigmoid
Dropout	[0.2, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5]	0.25, 0.25, 0.3, 0.4, 0.25
	Block of dense layers	
Number of layers	[1, 2, 3, 4]	1
Number of units	[50, 100, 150, 200, 250, 300]	50
Activation function	[ReLU, sigmoid]	ReLU
Dropout	[0.2, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5]	0.2
	Model training	
Optimizer	[Adam, SGD, RMSprop]	RMSprop
Learning rate	[1e-05, 1e-04, 1e-03, 1e-02, 1e-01]	1e-03
Batch size	[50, 60, . . . , 190, 200]	50
	Data preprocessing and seismic window extraction	
pre_mode	[1, 2, 3]	1
win_size [*s*]	[2.0, 2.2, . . . , 4.8, 5.0]	4.8
$frac\_pre\_P$	[0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	0.7
$frac\_pre\_S$	[0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	0.5
$frac\_pre\_N$	[0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	0.9

Before carrying out the model training, 10% of the window samples were randomly extracted as an independent test set. As in the phase detection stage, the test set includes a proportional amount of records from S1 and S2 datasets. The remaining samples were first shuffled and then subdivided into training (65%) and validation (25%) sets. Altogether, this corresponds to 16,670 training, 6,412 validation, and 2,565 (1,625 from S1, and 940 from S2) test one-

component samples for P-picking; and 9,358 training, 3,599 validation, and 1,440 (808 from S1 and 632 from S2) test two-component samples for S-picking. As in the phase detection stage, weights and hyperparameters of the models were adjusted based on the training set and validation set, respectively. The test set was then used to evaluate the performance of the best trained models. The window samples were defined based on the optimized hyperparameter values for the phase detection model, namely the window length (win_size) and its relative position with respect to the true onset ($frac\_pre\_P$ and $frac\_pre\_S$). This allows the integration of phase detection and phase picking in a common two-stage workflow, with the aim of making the overall process more efficient, since both stages share already optimized information.

However, instead of using the above-described seismic windows, we trained and validated the picking models on versions of them, which were randomly shifted in time by a range between -5.0 and 5.0 [s] around the true P and S onsets. This allows the networks to better learn data patterns in the vicinity of the phase onsets, and so preventing the model from memorizing only fixed onset times. A similar approach has been applied by Ross et al. (2018a) by using an artificially augmented number of samples to train a CNN network as regressor for P-phase picking. In the test set, we kept both the original time windows and their shifted versions, in order to test the detection capacity of the trained models in both scenarios.

So as to perform the model training, we assigned to each input sample a binary vector of length equal to the sample window length. This vector is a binary representation of the ground truth, where all the 0s correspond to timesteps associated to noise, and all the 1s representing signal, so that the first 1 represents the timestep of the true phase onset. Once trained, the models output a vector of the same length, which contains numbers between 0 and 1.0 corresponding to binary class probabilities: <0.5 for timesteps predicted as noise and >0.5 for timesteps predicted as signal. Thus, we can determine the predicted phase time onset in a given sample as the timestep where the output probabilities first rise above 0.5, indicating a change from noise to signal.

We ran 50 hyperparameter optimization trials for the P- (~130 minutes per trial) and S-phase (~50 minutes per trial) picking tasks. Table 4.2 presents the hyperparameter space searched during the optimization. We optimized on the validation accuracy metric, using the binary cross-entropy loss function and the Adam stochastic optimization algorithm (Kingma and Ba, 2014) in the training. The training was carried out using the same number of epochs and similar early stopping as in the phase detection stage.

4.2.5 From Predicted Phase Probabilities to Preliminary Onsets

By applying our best-performing models trained for both phase detection and phase picking on continuous seismic data, we were able to obtain the time onsets of P and S phases originating from local earthquakes. In this section we explain how we first obtained preliminary time onsets in the phase detection stage. The next section describes how we refined them in the phase picking stage to estimate the final phase onsets and their associated uncertainties.

Table 4.2: Hyperparameter search space optimized during the training of P- and S-phase picking models. The second column lists the hyperparameter values that could be sampled during the optimization. The third and fourth columns show the best-performing values found after running 50 P- and S-phase optimization trials.

Hyperparameter	Range tested	Best value (P-picking)	Best value (S-picking)
Recurrent BLSTM layers			
Number of layers	[1, 2]	2	2
Units	[10, 20, ..., 190, 200]	100, 160	20, 30
Dropout	[0.2, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5]	0.2, 0.35	0.25, 0.45
Recurrent Dropout	[0.2, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5]	0.2, 0.25	0.35, 0.25
Model training			
Learning rate	[1e-05, 1e-04, 1e-03, 1e-02, 1e-01]	1e-04	1e-02
Batch size	[50, 60, ..., 190, 200]	60	50

First, we used the optimized phase detection model to predict the class probabilities of three-component moving windows. By iteratively shifting these windows along the continuous seismic waveforms, we obtained a discrete probability time series of each class. The length of the moving window is given by the optimized hyperparameter win_size. We assigned the predicted probability at the position within the window defined by the optimized hyperparameters $frac\_pre\_N$, $frac\_pre\_P$, and $frac\_pre\_S$ for the respective classes (see Table 4.1 and Figure 4.2). The timestep between consecutive moving windows is a user-defined parameter that controls the resolution of the probability time series.

Next, we declared a P-phase search window between the time where the P-class probability time series (pb_P) rises above certain trigger threshold pb_P_th1 and the time where pb_P decreases below a certain detrigger threshold pb_P_th2. We estimated a preliminary P-phase onset as the time of the maximum pb_P within the search window. For waveforms where a preliminary P onset is found, we followed a similar approach to define a preliminary S-phase onset at the time of the maximum S-class probability time series (pb_S), within an S-phase search window delimited by the time interval that satisfies the condition $pb\_S\_th1 < pb\_S < pb\_S\_th2$. Cases where pb_P and pb_S overlap or are close enough in time at values above the trigger thresholds cannot be handled solely with the previous criteria. We included few additional conditions, based on the predicted probabilities, to deal with this and other particular scenarios. These conditions can be optionally activated by the user and depend on a few user-defined parameters (see Section S3).

4.2.6 Estimation of Final Phase Onset Times and Their Uncertainties

The preliminary P and S onsets were defined based on probabilities returned by a model trained for recognizing seismic phases, rather than for actually picking them, and therefore need to be refined. To this end, we used the search windows corresponding to the preliminary

P- and S-phase onsets as "picking windows". On these picking windows, we then applied the optimized models trained for P- and S-picking to estimate the respective phase onsets as well as the associated uncertainties.

Uncertainties were determined by applying the stochastic regularization technique Monte Carlo Dropout (MCD). It has been shown that using dropout for training a deep neural network can be interpreted as a Bayesian approximation of a Gaussian process, hence MCD can be used as an approximation of Bayesian variational inference (Gal and Ghahramani, 2016a,b). This means that a model posterior distribution for a given sample can be represented by multiple (T) predictions made by the network with dropout. Thus, the value and uncertainty of the predicted sample can be obtained as the mean and standard deviation of the T inferences.

In practice, we implemented MCD by performing T model predictions with dropout on. In this case, since phase picking is treated as a sequence binary classification task, MCD generates a collection of predicted class probabilities of the timesteps in an input time series. This collection is an approximation of the model posterior distribution for that input sample, which is described by the mean and standard deviation of the T class probabilities at each of the sample timesteps.

Then, we estimated the predicted phase time onset (*tpred*) for the sample as the first timestep in which the mean of the class probabilities rises above 0.5 (*pb*), which indicates the predicted transition from noise to signal. The standard deviation of the class probabilities at *tpred* (*pb_std*) represents the variability of the predicted probability at the assumed time onset. We propose that a reasonable first order estimation for the time onset uncertainty can be determined as the interval of timesteps delimited before and after *tpred* by the projection of *pb_std* on the mean class probability (see Figure 4.3). The resulting time uncertainty is asymmetric and inversely proportional to the slope of the mean predicted probability curve around *tpred*. That is, the steeper the decrease/increase in the predicted probability before/after *tpred*, the lower the time uncertainty, as one might intuitively expect from an abrupt, easy-to-pick change from noise to signal. Based on these estimated pick uncertainties, we defined the weighting class scheme for the phases P and S shown in Table 4.3, which was adapted from (Sippl et al., 2013).

Table 4.3: Weighting scheme for P and S predicted pick classes. Where $terr = \frac{1}{2}(|tons\_err^-| + |tons\_err^+|)$ is the mean of the absolute time uncertainties calculated before and after the phase onset shown in insets in Figures 4.3b and 4.3d.

Quality Class	Time onset error (*terr*)
P0, S0	$terr < 0.05$
P1, S1	$0.05 < terr < 0.1$
P2, S2	$0.1 < terr < 0.2$
P3, S3	$terr > 0.2$

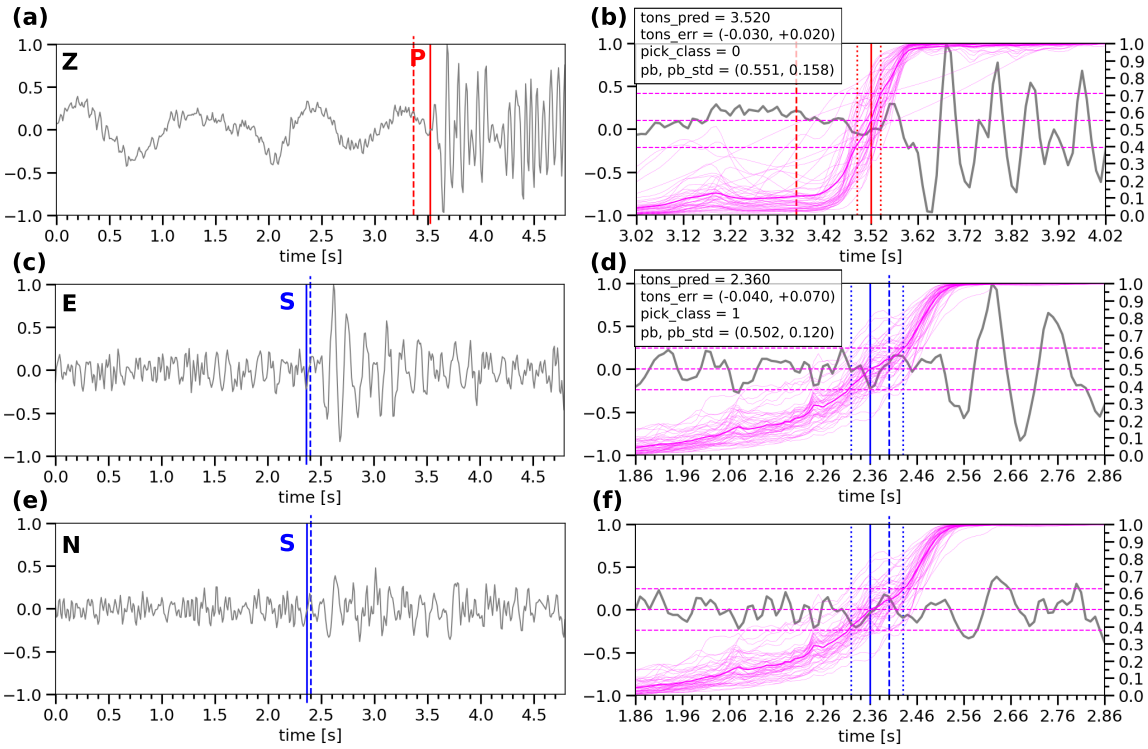

Figure 4.3: Estimation of time onset and associated uncertainty for predicted P (a-b) and S (c-f) phases. (a), (c), (e) Examples of predicted Z, E, N (Vertical, East, North) window samples. Red and blue dashed lines represent the preliminary P and S onsets predicted in the phase detection stage, which were refined in the phase picking stage (red and blue solid lines). (b), (d), (f) Zoom in centered on the refined P and S time onsets (*tons_pred* [s]), estimated by applying MCD (T=50) inference. Predicted class probabilities for each MCD prediction are plotted as magenta curves (right y-axis); bold curve represents the mean probability. Magenta horizontal dashed lines indicate the mean (*pb*) and one standard deviation (*pb_std*) of the class probabilities at the onset time. As explained in the main text, from *pb* and *pb_std*, we can obtain *tons_pred* and its uncertainty (*tons_err* [s]), outlined by red and blue vertical dotted lines before and after the refined P- and S-phase onsets. Based on the weighting class scheme defined in Table 4.3, P and S phases shown here correspond to picks of class 0 and 1, respectively.

4.3 Results

4.3.1 Phase Detection Model Optimization

Figure 4.4 summarizes how the six main hyperparameters in the phase detection model evolved over the 1,000 optimization trials attempted. For the majority of the main hyperparameters, the most frequently sampled hyperparameter value (higher histogram bins) did not coincide with the value that led to the highest model accuracy found after 1,000 trials (shaded histogram bin), although this correlation was observed for the number of convolutional layers and $frac\_pre\_P$.

Remarkably, as can be seen from the subplots in Figure 4.5, the best hyperparameter values or the most frequently sampled values (when both did not coincide) formed clusters at several times (around trials 150, 260, 400, 590, and 800) during the optimization. This suggests that the same set of hyperparameter values was systematically found to be effective in producing highly accurate models during the optimization process, while the rest of the hyperparameters in the network were probably still being adjusted.

The best-performing hyperparameter values indicate that models in which the convolutional

block was considerably deeper (five convolutional layers) than the dense block (only one or two dense layers) proved to be more accurate. Furthermore, models that learned from band-pass filtered data reached lower training accuracy in comparison to when high-pass filter or no filter was applied, with the latter approach performing the best. Thus, using the whole frequency content proved to be more effective for model training than removing very long period background seismic noise or substantially limiting the frequency content. We note here that quite a few of the data are from short-period seismometers, hence inherently high-pass filtered (>1 Hz), which could account for the most frequently sampled value of *pre_mode*.

Another important constraint in the optimization process seems to be the amount of data available for the network to learn, since longer window samples (480-500 s) clearly led to more correctly predicted phases. P phases were more accurately predicted when the sample windows include a relatively larger portion of noise (70%) compared to signal, whereas S phases were better predicted from windows containing similar amount of noise relative to signal (50%).

Table 4.1 lists the best-performing values found for each of the optimized hyperparameters, which were used to train the best phase detection model architecture outlined in Figure 4.6.

4.3.2 Predicted Phase Classes compared to Analyst Labels

Figures 4.7a and 4.7b show that the accuracy of the best phase detection model smoothly increased on the training set and reached 0.986 after 35 epochs, when the early stopping condition stopped the training. The training loss progressively decreased and reached 0.044 at the end of the training. The accuracy and loss calculated on the validation set followed a similar behavior to the training curves, although showing overall higher accuracy and lower loss due to the use of dropout regularization during training. Dropout disables a certain number of layer units, therefore some information about the input samples is lost and the network attempts to learn based on incomplete data representations. However, during validation all the units are available, thus the network uses its full computational power.

The confusion matrix in Figure 4.7c shows the overall performance of the best trained model in classifying the classes of the independent test set of 4,320 samples. The cells in this matrix are filled with the predicted classes, which correspond to the highest of the three-class probabilities returned by the model for each sample. An overall very high ratio (0.99) of correct predictions is observed for the three classes, as can be seen from the numbers in the matrix diagonal. Figure S7 show waveform examples of these correctly predicted P-, S- and N-class samples.

Of the few misclassified samples, only three correspond to S phases predicted as P, whereas no P phases were mispredicted as S, indicating that the network has a high capability to discriminate between both classes. Figure S8 shows one of these misclassified S phases, where the pick made by the analyst was somewhat earlier than the most impulsive change in amplitude. This could be one possible source of misclassification between the two phases, since it may have led to the network misinterpreting the sample as a P phase.

The remaining mispredictions correspond to either actual P or S phases predicted as noise (24 samples) or presumed noise predicted as P or S phases (21 samples). From the samples

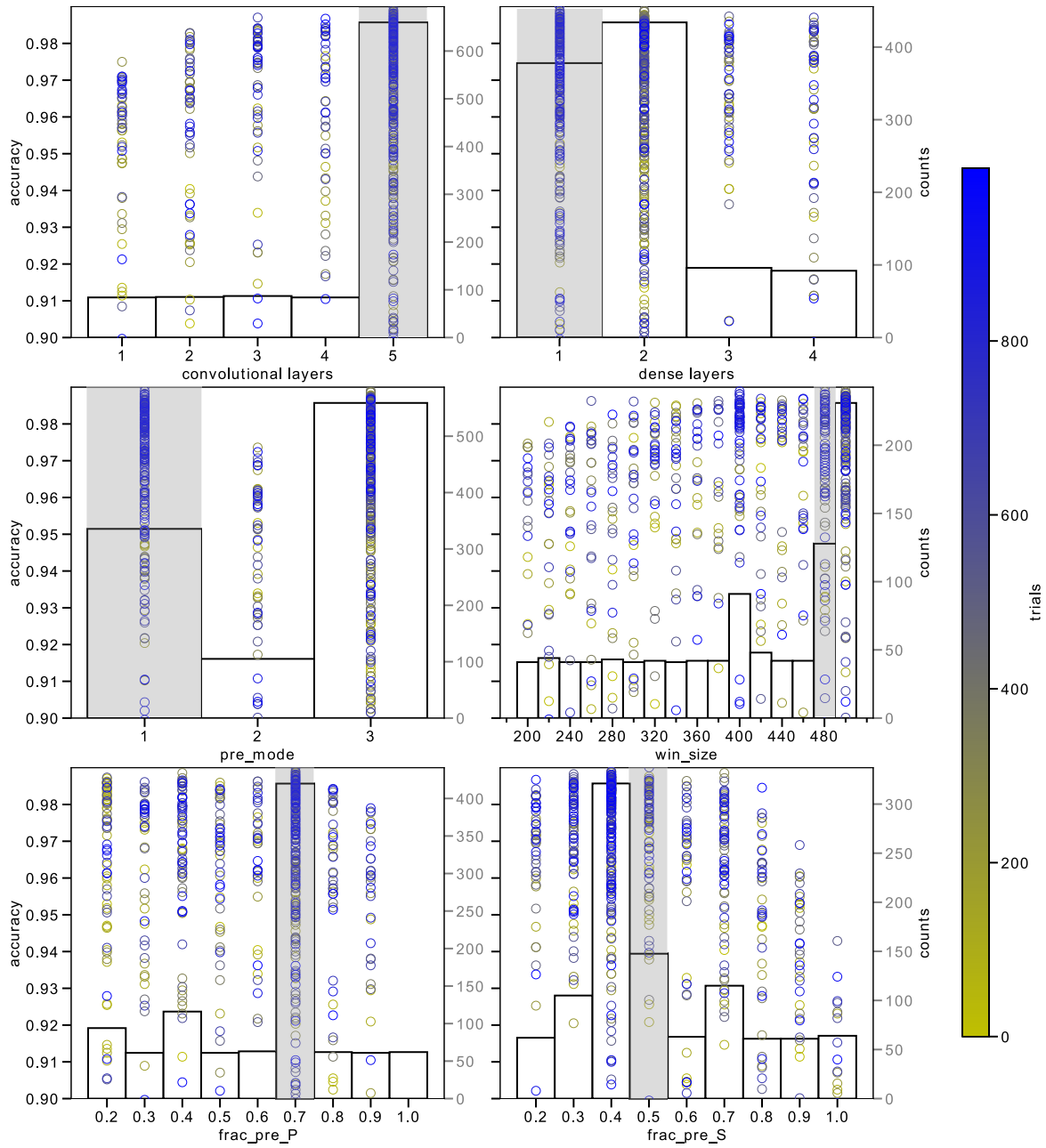

Figure 4.4: Distribution over the 1,000 optimization trials of the six main hyperparameters used in the phase detection model training. Subplots show sampled hyperparameter values colored by trial number and plotted versus the model accuracy reached at that trial (left y-axis). The distribution of the hyperparameter values sampled over the trials is presented as a histogram (right y-axis). A gray shaded bin in the histogram represents the best-performing hyperparameter value, that is, the value with which the model reached the highest accuracy.

in the former group, the majority of the P (90%) and S (~71%) samples were assigned pick weights higher or equal than 1 (<75% confidence) by analysts (Figure 4.7e). This is indicative of waveforms with relatively low signal-to-noise ratio (SNR), as can be seen from the examples of mispredicted P and S phases shown in Figures 4.8a and 4.8b, which may cause the model misclassification. In the latter group, a significant majority of the noise samples were predicted as signal with relatively high probability. In fact, of these, 87.5% of the P and ~77% of the S predictions had a probability higher than 0.7, with 75% of the P and 38% of the S predictions having a probability higher than 0.9 (Figure 4.7e). The examples in Figures 4.9a and 4.9b demonstrate that these apparently mispredicted samples actually correspond to previously un-

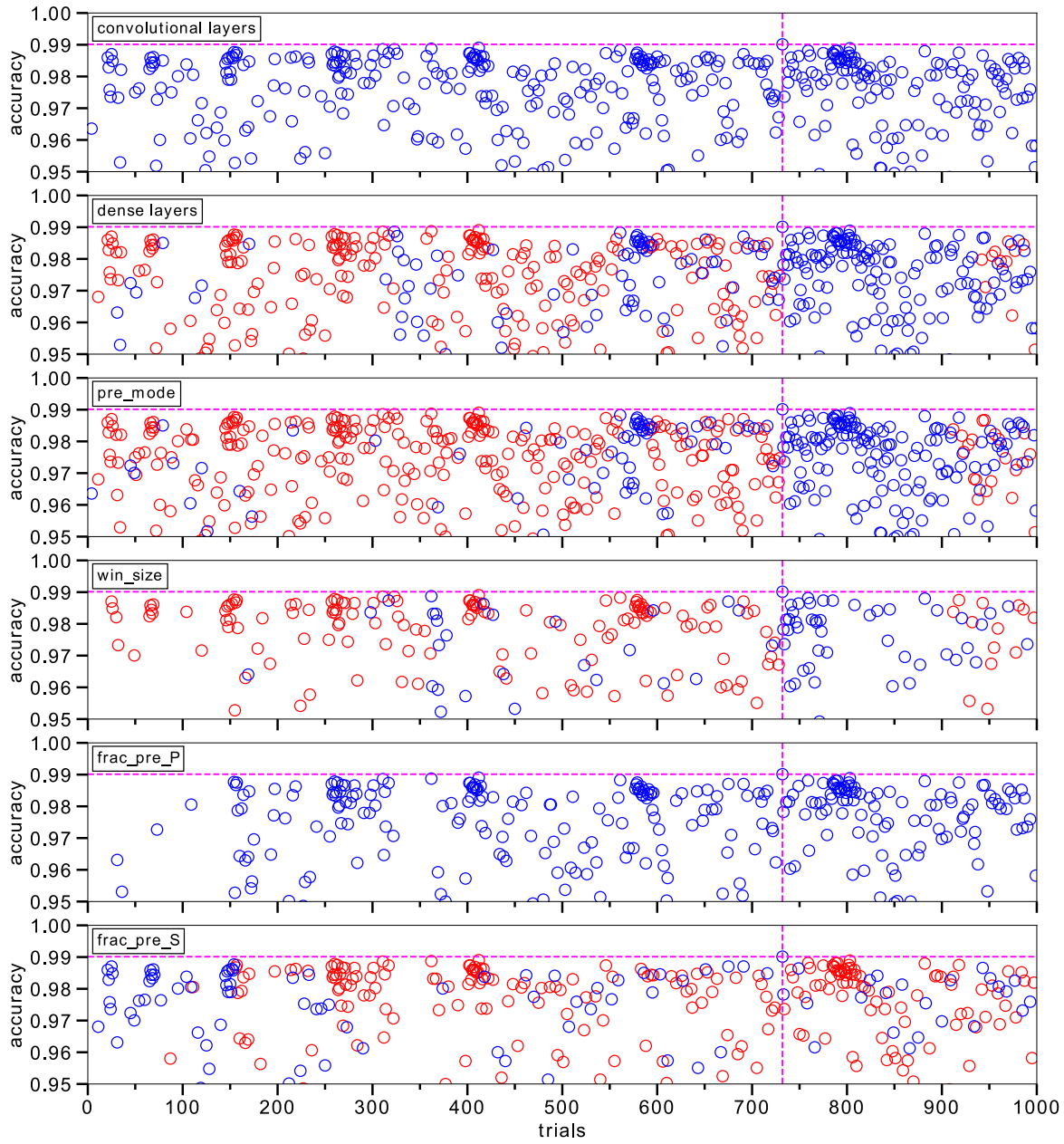

Figure 4.5: Each subplot depict in detail the distribution of the best-performing hyperparameter values (gray shaded bins in histograms of Figure 4.4; blue circles) that led to higher model accuracies (>0.95) over the trials. Values plotted are the accuracies on the validation set, reached at the end of each trial training. For the hyperparameters in which the best-performing value does not coincide with the most frequently sampled value, the latter are plotted as well (red circles). For reference, magenta horizontal and vertical dashed lines mark the highest model accuracy (0.990) reached during the optimization and the trial (732) at which it was achieved, respectively.

detected phases, hence evidencing the powerful detection capacity of the model.

The detection capability of the trained model can be further assessed by calculating the precision and recall metrics. Precision, given by TP / (TP + FP), describes how well the model predicts a specific class, by comparing the correct predictions of the class (true positives: TP) with all the predictions made for that class (TP + FP; where FP means false positives). Recall is computed as TP / (TP + FN) and reports the ability of the model to identify all the samples of a given class, by comparing TP with all the samples of that class in the test set (TP + FN; where FN are false negatives).

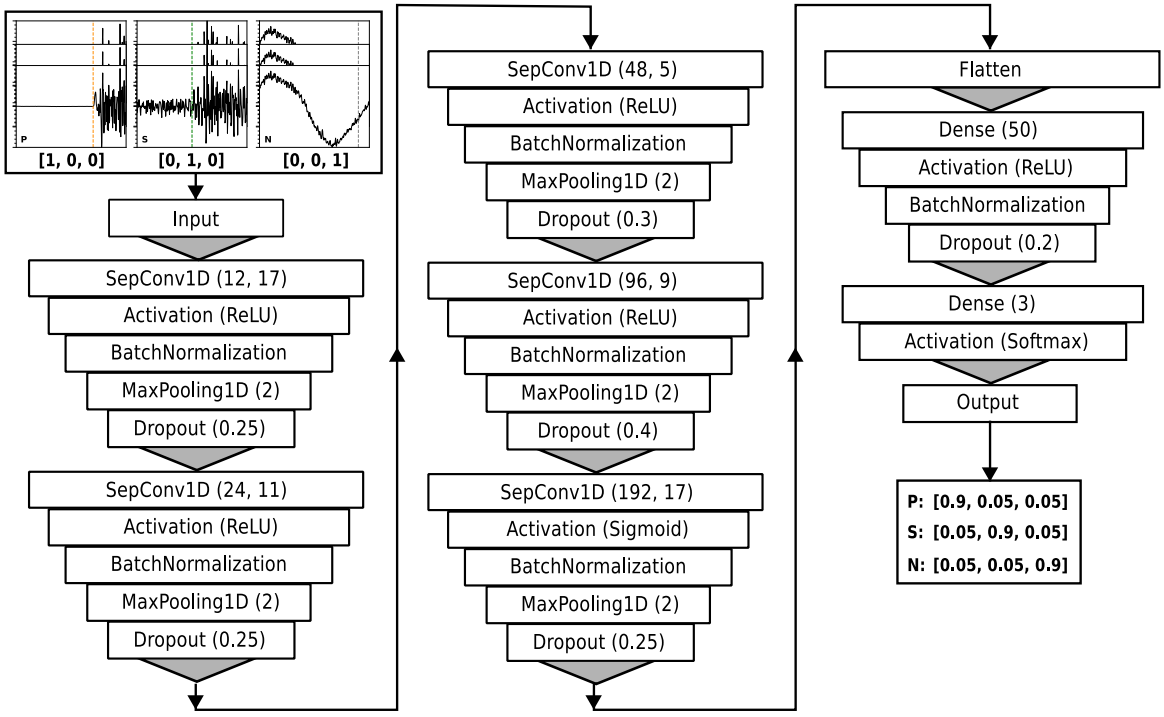

Figure 4.6: Optimized model architecture trained for phase detection task. The network receives as input class vectors representing the respective accompanying three-component seismogram samples and outputs three-probability vectors indicating the predicted classes (P, S, or N). The optimized hyperparameters in each block of convolutional and dense layers are shown here and also presented in Table 4.1.

Precision and recall vary between 0 and 1, where higher values are associated with better performing models. In a precision-recall curve, precision is plotted against recall for different probability thresholds, i.e., the probability above which a sample is predicted to be of a certain class. The more the curve bends toward the point (1, 1), and therefore the larger the area under the curve (AUC), the better the model performance. Figure 4.7d shows the precision-recall curve obtained for the samples in the independent test set. As can be seen here the performance of the model in predicting both P- and S-class samples was very high, with the performance being slightly better for the P than for the S class. The model performance on N-class samples was somewhat lower. However, this calculated performance does not take into account any presumed N-class samples where the model detected a real seismic signal. As previously discussed by Ross et al. (2018b), the probability threshold can be tuned so as to improve precision (reducing false positives) or recall (aiming at reducing missed detections), depending on the application.

4.3.3 Predicted Phase Time Onsets compared to Analyst Picks

Table 4.2 shows the best-performing hyperparameter values found during the optimization of the P- and S-phase picking models. These parameters were used to train the best picking model architectures outlined in Figures 4.10a and 4.10b.

At the end of the training, the best P-picking model reached higher accuracies (0.983 in training) than the accuracies of the S-picking model (0.972 in training). This is also reflected in the training and validation loss functions, which reached a lower value for P phases (0.041

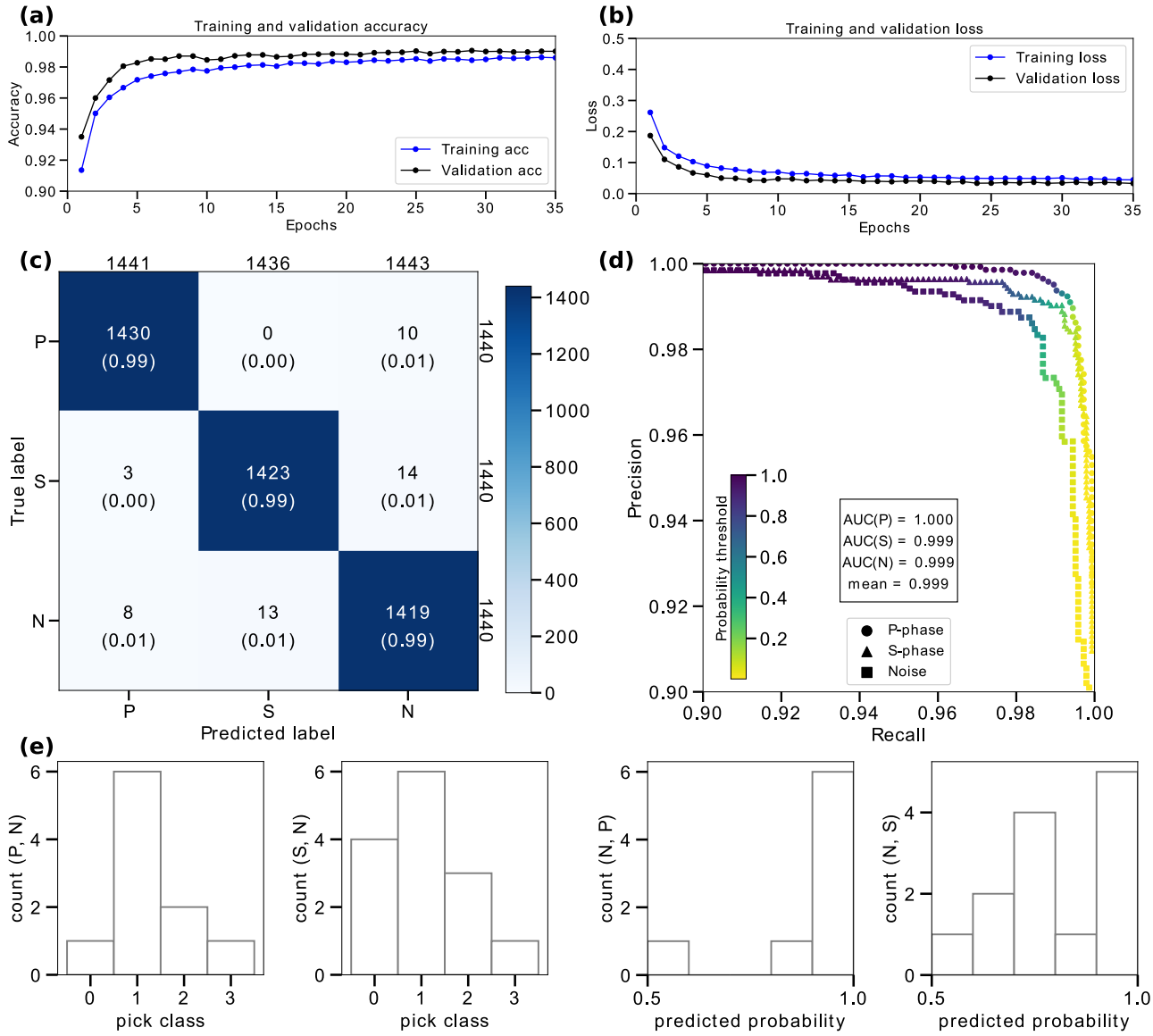

Figure 4.7: (a-b) Evolution of the training and validation accuracy (a) and loss function (b) over the training epochs of the optimized model trained for phase detection. (c-d) Results from evaluation of the best phase detection model on the independent test set of 4,320 samples, represented in a confusion matrix (c) and a precision-recall curve (d). In (d), AUC represents the area under the precision-recall curve for each class. (e) disaggregated histograms of incorrectly predicted P or S phases as noise and presumably misclassified noise samples as P or S phases. (P, N) means P-phase samples predicted as Noise, and similarly for other cells in the confusion matrix.

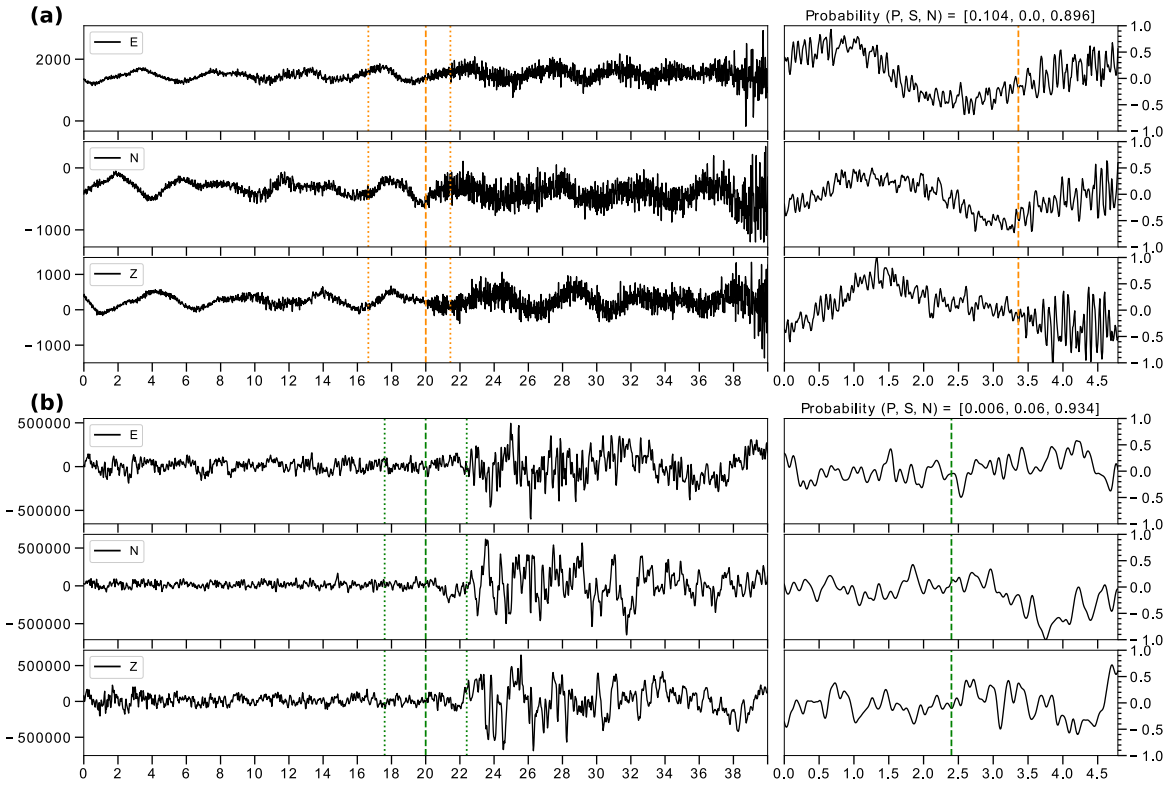

Figure 4.8: (a) Example of P-phase sample in the test set of 4,320 samples, mispredicted as noise (N). (b) Example of S-phase sample in this test set, mispredicted as noise. The left subplots in (a) and (b) show the three-component seismic wavefrom from which the samples were extracted. The extracted samples, as received by the network, are shown in the right subplots. Examples shown here correspond to P-phase and S-phase samples with analyst's pick weights 2 (50% confidence) and 1 (75% confidence), respectively.

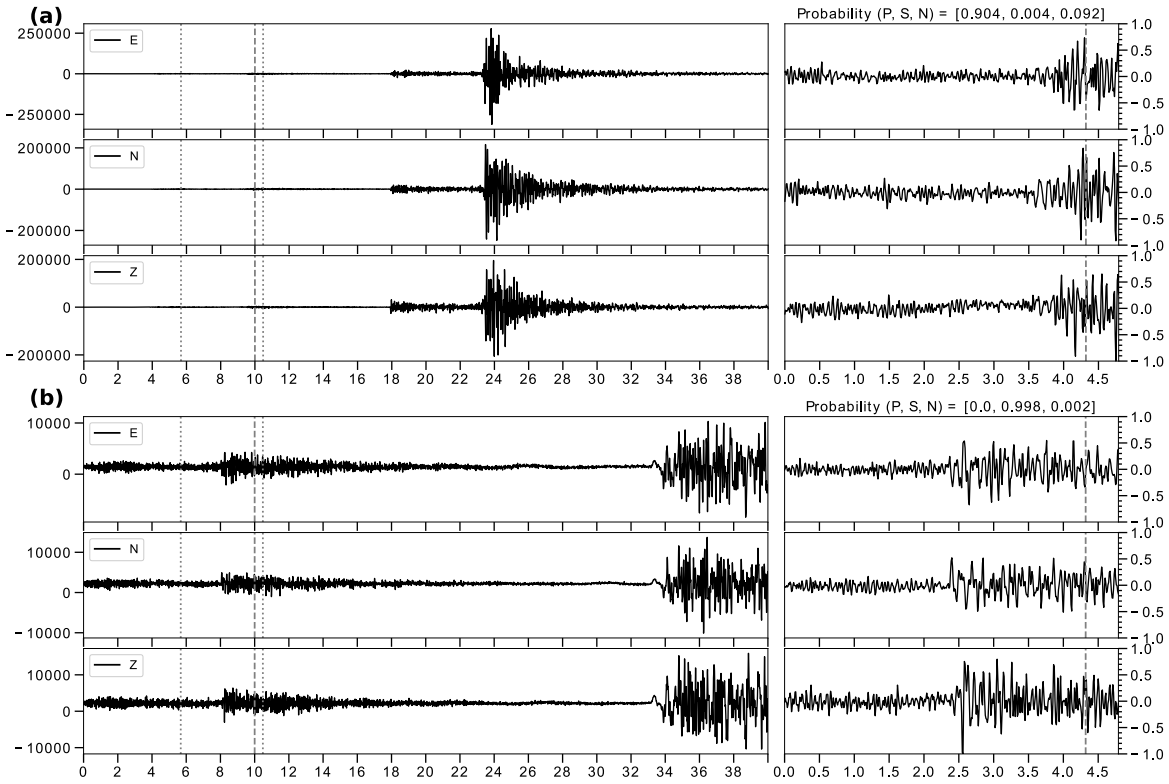

Figure 4.9: (a) Example of noise (N) sample in the test set of 4,320 samples, seemingly mispredicted as a P-phase. (b) Example of noise sample in this test set, seemingly mispredicted as a S-phase. Subplots are plotted the same way as in Figure 4.3. The apparent misclassifications shown here are examples of non-picked phases, which were however detected by the network.

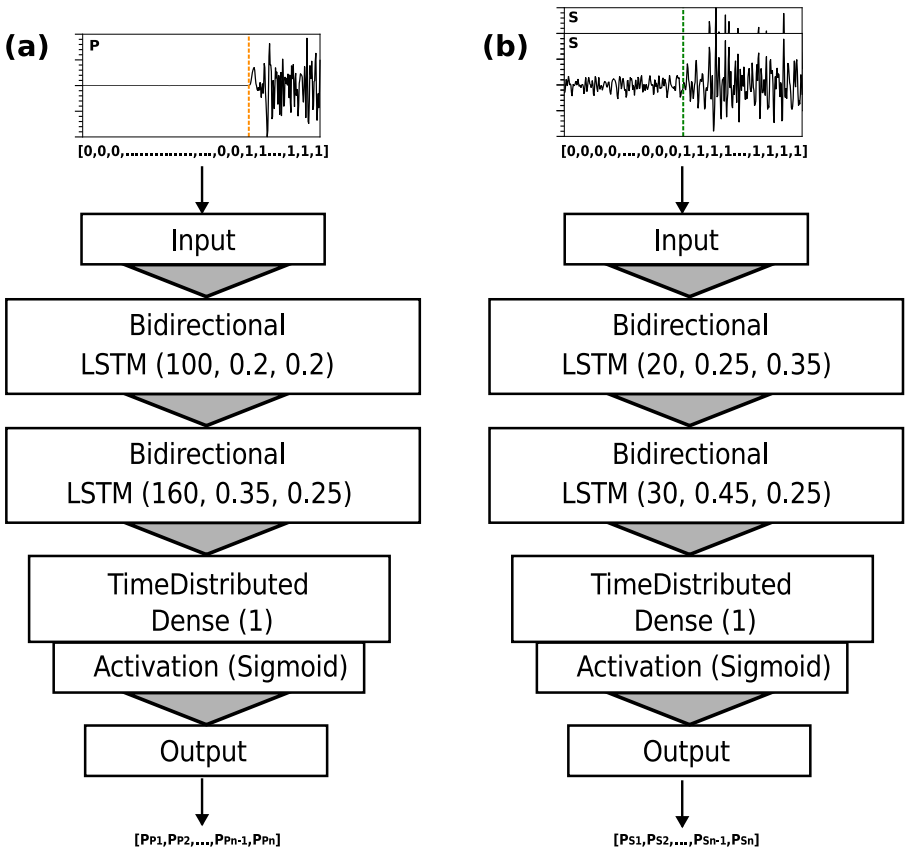

Figure 4.10: Optimized model architectures trained for P-phase (a) and S-phase (b) picking tasks. The networks receive as input binary vectors representing the noise (0) and signal (1) content in accompanying time series taken from the vertical-component (a) and the two horizontal-component (b) seismograms of the samples used for training. The models output is the probability of the timesteps in input samples corresponding to either noise or seismic signal. The optimized hyperparameters in each BLSTM layer are shown as (Units, Dropout, Recurrent dropout), which are also reported in Table 4.2 for both models.

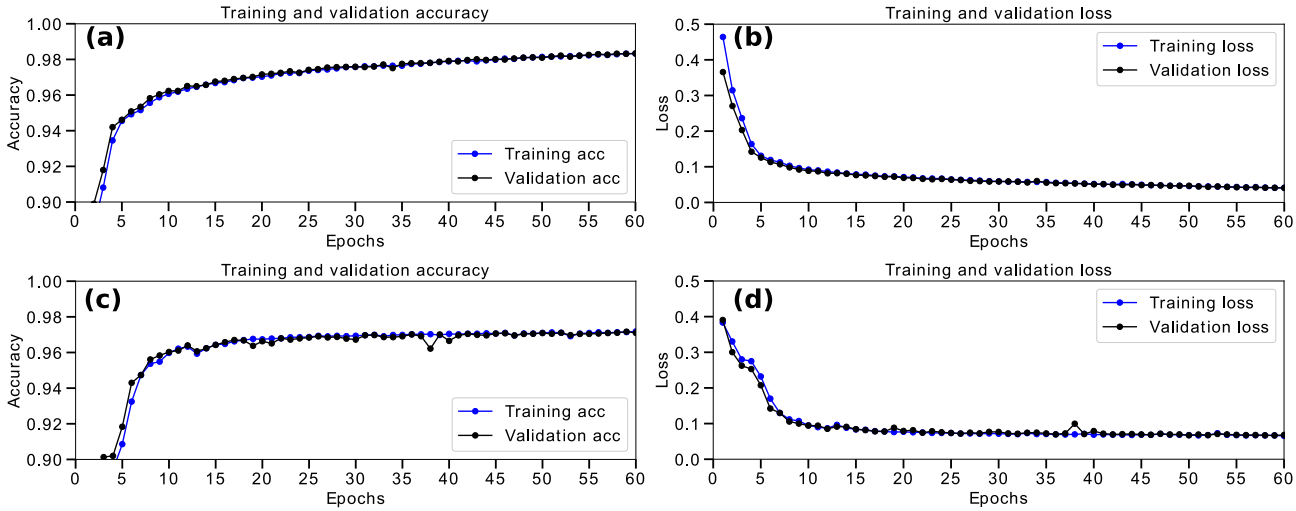

Figure 4.11: Evolution of the training and validation accuracy and loss function over the training epochs of the optimized model trained for P-phase (a-b) and S-phase (c-d) picking.

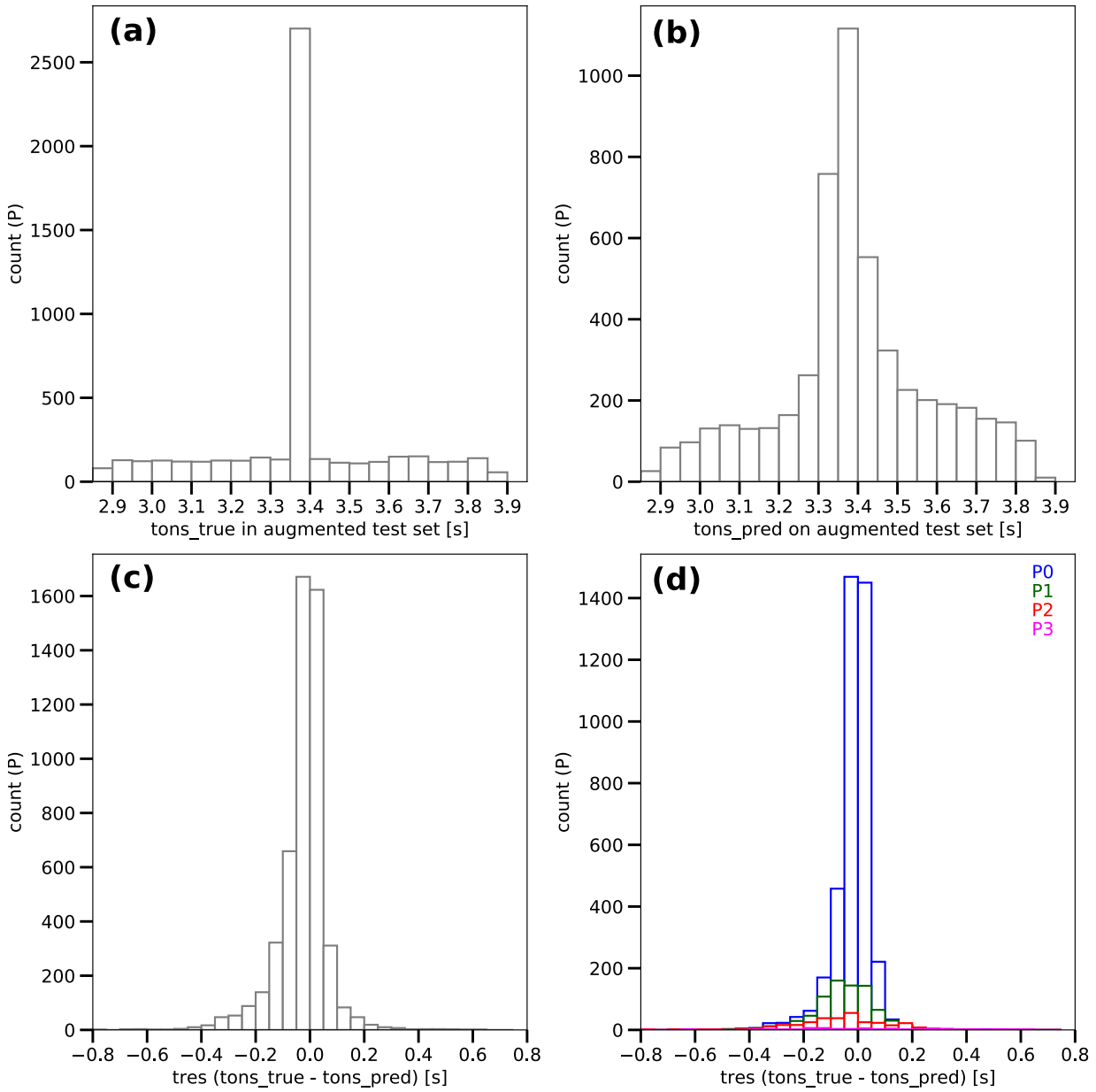

Figure 4.12: Evaluation of the optimized model trained for P-phase picking on the independent augmented test set of 2,565 x 2 one-component waveform samples. (a) True time onsets (*tons_true*) from both the actual and shifted samples forming the augmented test set. (b) Predicted time onsets (*tons_pred*, defined as in Figure 4.3) by the best model trained for P-phase picking. (c) Time residuals (*tres*) distribution, defined as the difference between the true and predicted time onsets. (d) Predicted pick quality classes, as defined by weighting scheme defined in Table 4.3.

in training) than for S phases (0.066 in training) (Figure 4.11). As previously explained, the MCD estimation of predicted phase time onsets and their uncertainties requires that the picking models are trained with dropout enabled. For this reason the training and validation curves do not show the difference in accuracy and loss observed for the best phase detection model.

Figures 4.12 and 4.13 summarize the performance of the optimized models trained for picking P and S phases, when evaluated on the independent test sets of 2,565 one-component and 1,440 two-component samples respectively. In order to evaluate how the models perform in different possible picking scenarios, we used an augmented test set formed by two groups of samples: 1) the actual waveforms in the original test set, as defined by the optimized hyperparameters of

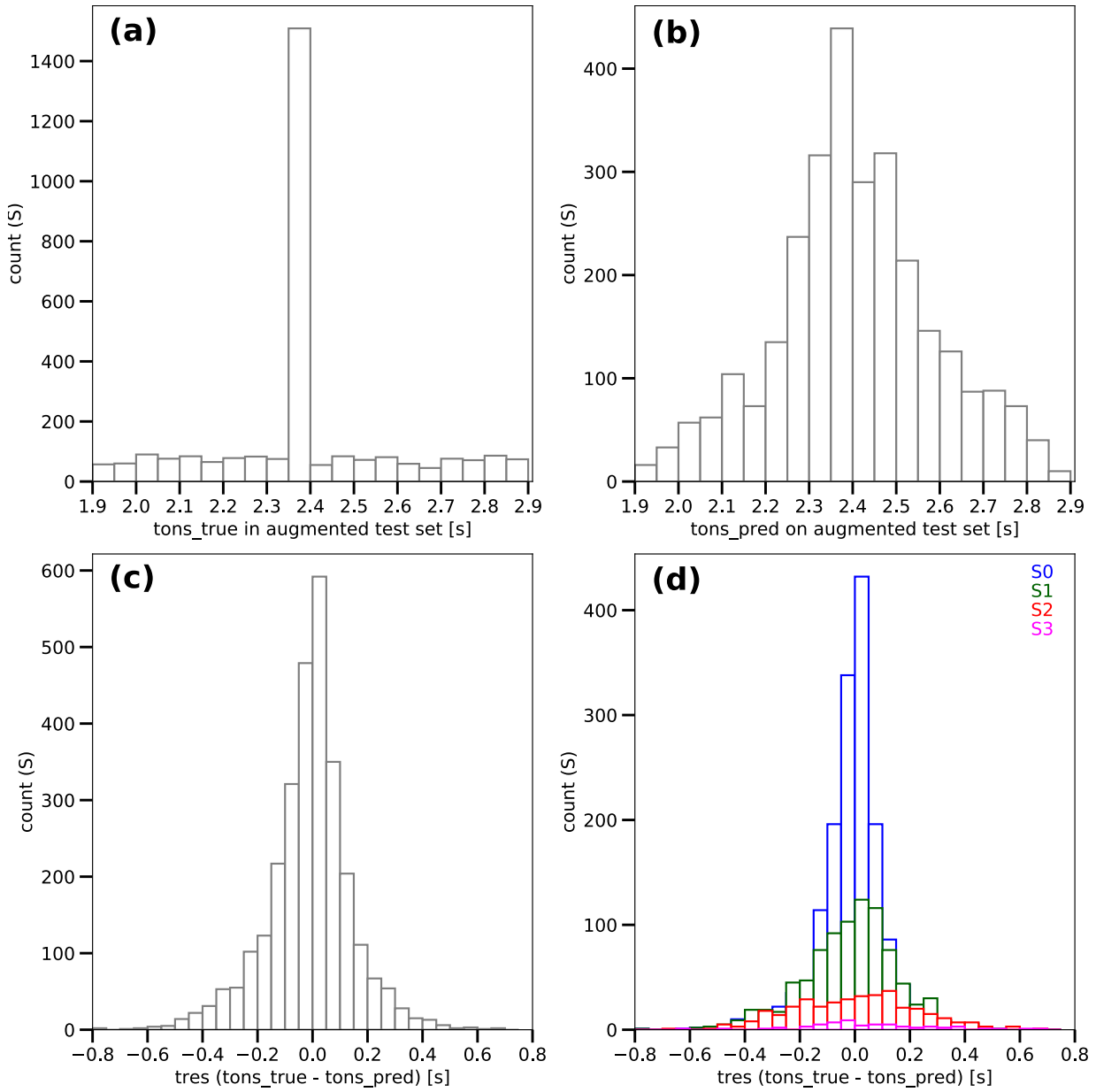

Figure 4.13: Evaluation of the optimized model trained for S-phase picking on the independent augmented test set of 1,440 x 2 two-component waveform samples. Subplots are similar as in Figure 4.12.

the best phase detection model (*win_size*, *frac_pre_P*, and *frac_pre_S*), and 2) artificial versions of the same waveforms, randomly shifted between -0.5 and 0.5 s around the true phase onsets made by the analysts.

The slightly higher accuracy reached by the P-phase picking model is reflected in the shape of the true and predicted time onset distributions, which more closely resemble each other for the P-phase (Figures 4.12a and 4.12b) compared to the S-phase (Figures 4.13a and 4.13b). Consequently, a narrower distribution of time residuals was obtained for P phases compared to S phases (Figures 4.12c and 4.13c). Overall, however, the time residual distributions for both P and S phases show a remarkable compliance with analyst picks. Based on the weighting class scheme defined in Table 4.3, we observe that a significant majority of P- and S-phase onsets were predicted with lower uncertainties (pick quality classes 0 or 1 in Figures 4.12d and 4.13d). Even though the distributions of all pick quality classes span over a similar range

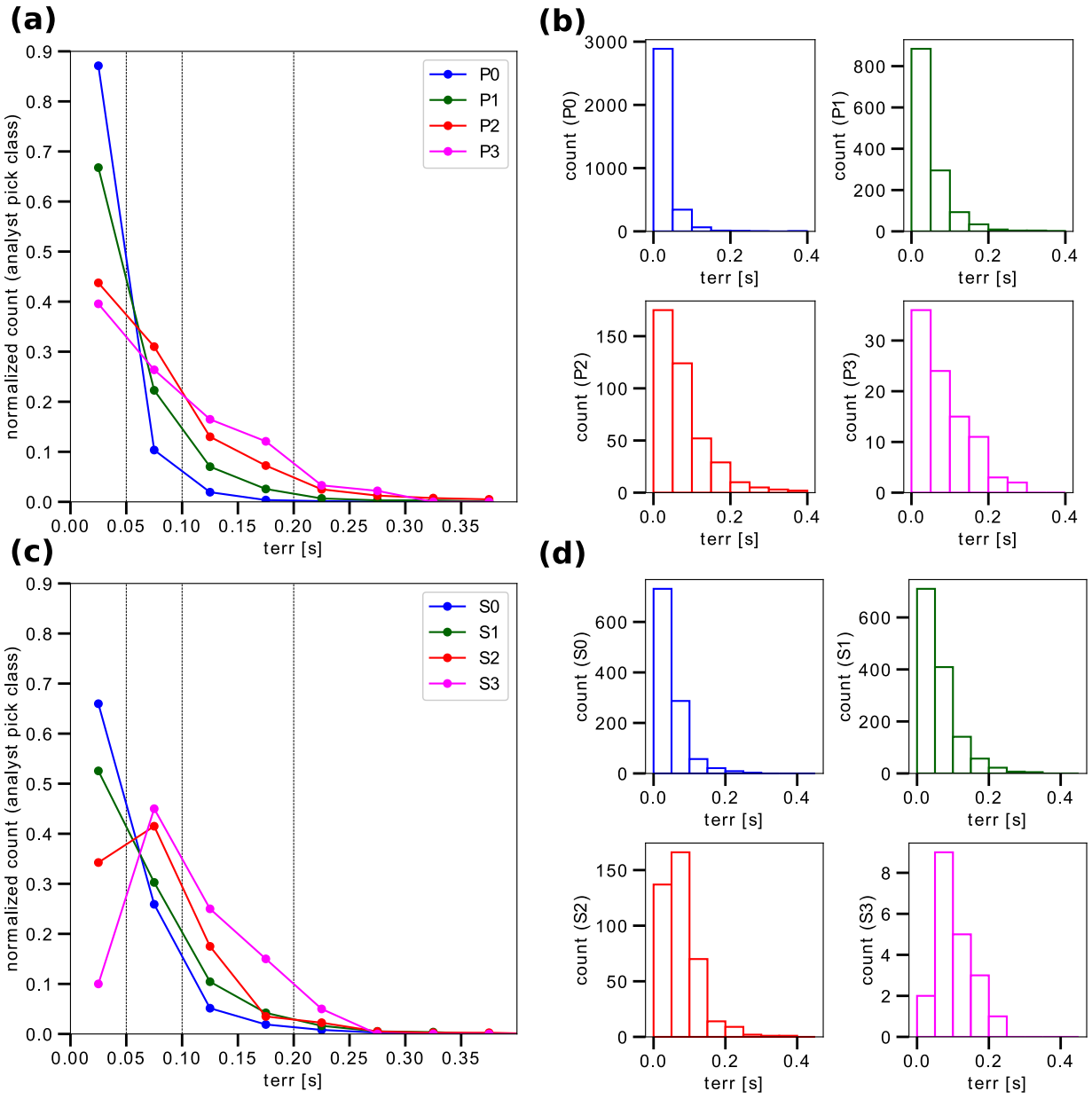

Figure 4.14: (a), (c) Normalized distribution of analyst P- and S-pick quality classes over the range of predicted time onset errors (*terr* in Table 4.3). Time errors were predicted on the augmented test sets, so analyst quality classes are counted twice. Points defining each colored curve are plotted at the center of the histogram bins corresponding to each pick class distribution. Dotted vertical black lines mark the time confidence interval limits used in the weighting scheme to assign the predicted pick classes (Table 4.3). (b), (d) Histograms (bin size = 0.05 [s]) showing disaggregated distributions of analyst P- and S-pick classes shown in (a) and (c), respectively.

of time residuals, the distribution is more flattened for higher uncertainties. This may suggest that picks predicted with lower uncertainties by the network were also easier to pick by the analyst, resulting in lower time residuals since both the network and analyst would have picked a similar onset. Conversely, picks that were harder to pick by the analyst (e.g., due to a low SNR), were probably predicted with higher uncertainties by the network, which would lead to higher absolute residuals.

So as to investigate this correlation between predicted and analyst pick classes, we plotted the distribution of the analyst pick classes against the predicted time onset errors (*terr*) for both test sets (Figure 4.14). Essentially, these distributions show that analyst classes do not always

correlate with equivalent predicted classes, but instead span the entire range of predicted time onset errors.

However, a couple of observations from the normalized distributions in Figures 4.14a and 4.14c may indicate that at least some degree of correlation does exist. First, lower-uncertainty analyst picks (classes 0 and 1) concentrate in predicted lower-error intervals ($terr < 0.05$) in a considerably higher proportion, compared with higher-uncertainty analyst picks (classes 2 and 3). Second, analyst picks of higher uncertainty decrease more gradually and rise above picks of lower uncertainty when $terr > 0.05$.

4.3.4 Evaluation on Independent Test Sets from Automatically-derived Earthquake Catalogs

We further analyzed the performance of the best trained models of DeepPhasePick on two additional independent test sets. These test sets consisted of three-component window samples extracted from two earthquake catalogs recently published for the region where the 2014 M8.1 Iquique megathrust earthquake ruptured the northern Chilean subduction on 1 April 2014 (Figure 4.15). The first catalog (T1) contains 8,278 events occurring in a time window of nine months before the Iquique mainshock (between 2013-06-01 and 2014-02-28) (Sippl et al., 2018), from which we extracted 228,230 records (P: 88,449; N: 88,449; S: 51,332). The second catalog (T2) includes 18,963 events originating between one month before and nine months after the mainshock (between 2014-03-01 and 2014-12-31) (Soto et al., 2019b), from which we obtained 1,361,544 records (P: 545,746, N: 545,746, S: 270,052). Event records extracted from T1 and T2 were automatically picked by applying conventional state-of-the-art phase picking algorithms (Aldersons, 2004; Diehl et al., 2009; Di Stefano et al., 2006), thus provide a valuable quantitative assessment when compared to DeepPhasePick predictions. We first analyzed the performance of the optimized model trained for phase detection. For this analysis, we used test sets formed by the entire set of samples available from catalogs T1 and T2. Figure 4.16 summarizes the classification results from the evaluation of the model on both test sets. In the confusion matrices in Figures 4.16a and 4.16c, we observe the highest ratio of correct predictions for noise samples, comparable to the evaluation performed on the test set extracted from datasets S1 and S2 (Figure 4.7c). The percentage of correct classifications decreased by 8-15% for P samples and 17-19% for S samples.

From the mispredicted P and S records, only a small fraction corresponds to P phase mispredicted as S, or viceversa, compared to the number of P and S records which were predicted as noise. As we observed for the results in Figure 4.7c, the optimized model can mispredict S phases as P phases if the true S time onsets are somewhat earlier. Conversely, it is conceivable that the network classifies a P phase as an S phase if the true P time onset is late.

P and S phases predicted as noise correspond almost entirely to samples which were assigned automatic pick quality class 1 or worse (<75% confidence), with most of them being assigned the poorest pick quality class (class 3; <25% confidence) (upper subplots in Figures 4.16b and 4.16d). As we have discussed, the classification of these hard-to-pick phases by the network

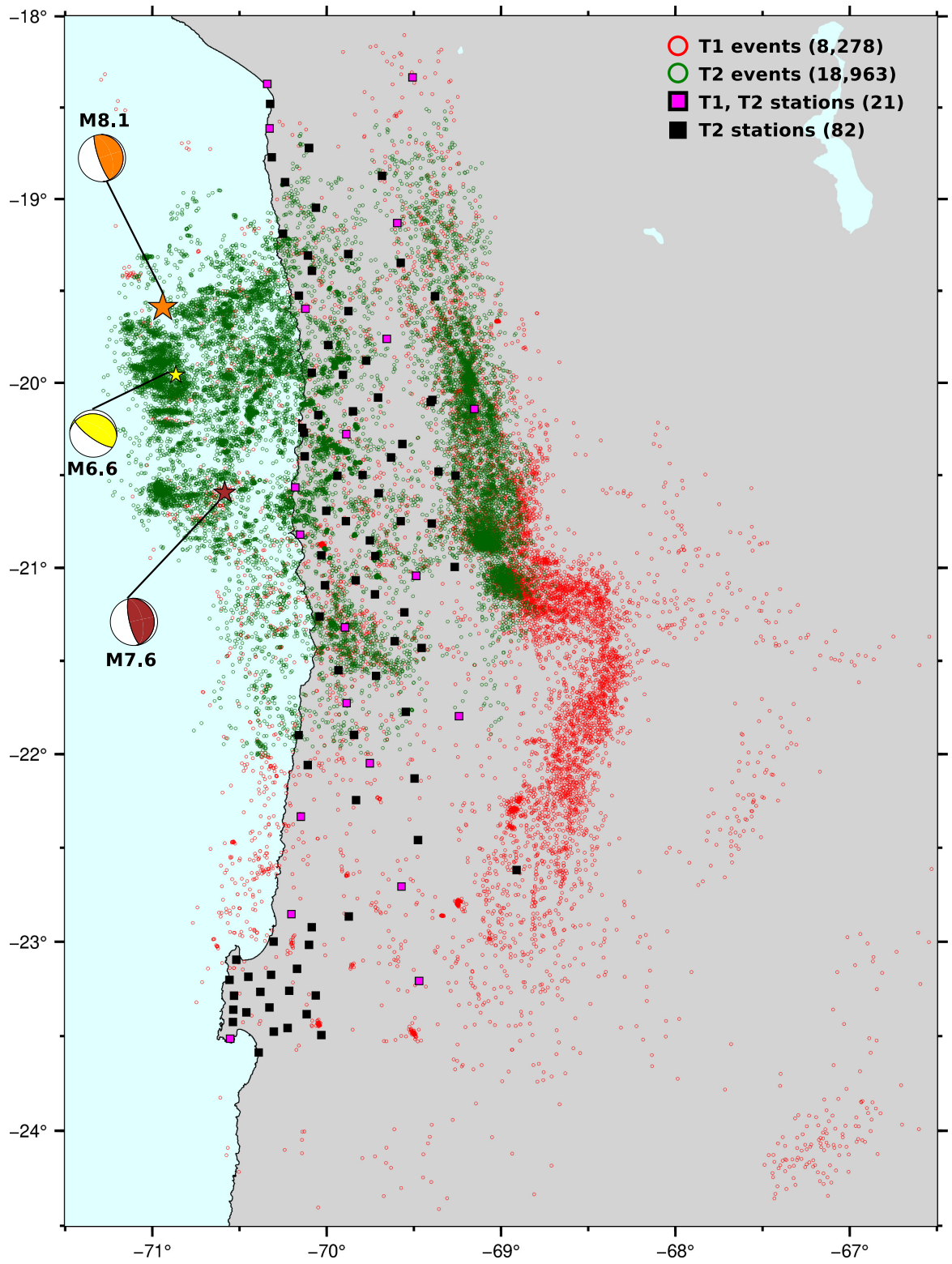

Figure 4.15: Automatically-derived earthquake catalogs in northern Chile, from which two independent test sets of three-component samples were used to evaluate the DeepPhasePick performance. A first test set of 228,230 samples was extracted from events occurring between 2013-06-01 and 2014-02-28 (T1: red circles; Sippl et al., 2018). A second test set of 1,361,544 samples was taken from events occurring between 2014-03-01 and 2014-12-31 (T2: green circles; Soto et al., 2019b). Window samples obtained from T1 and T2 were registered by the 21 and 82 stations shown on the map. Focal mechanisms (beachballs) and epicenters (stars) of the Iquique M8.1 mainshock, its largest M6.6 foreshock and its largest M7.6 aftershock (Hayes et al., 2014) are plotted for reference in orange, yellow, and brown color, respectively.

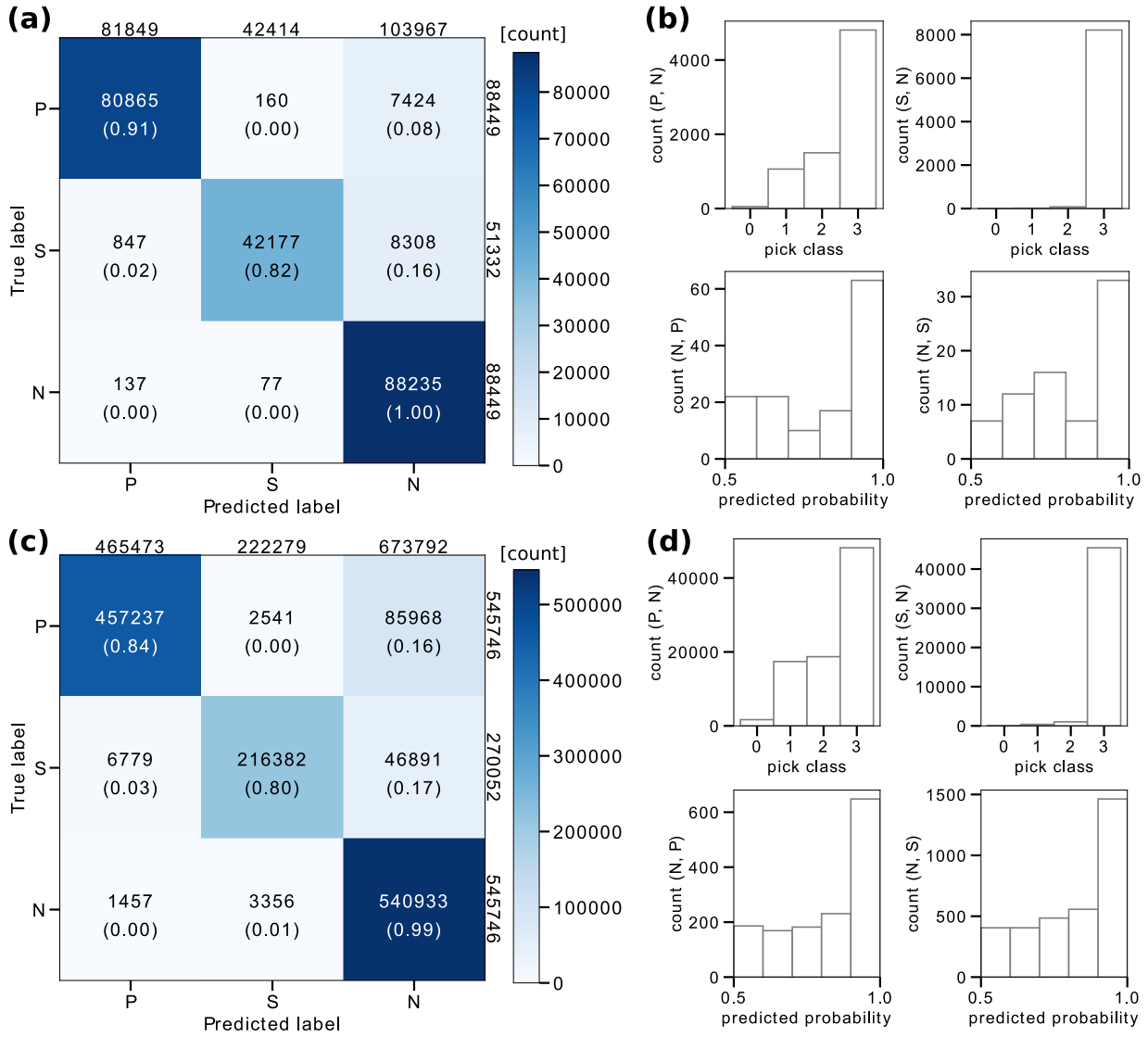

Figure 4.16: Results from the evaluation of the best model trained for phase detection on independent test sets taken from automatically-derived earthquake catalogs T1 (a-b) and T2 (c-d). (a), (c) Confusion matrices. (b), (d) disaggregated histograms of incorrectly predicted P or S phases as noise (upper subplots) and presumably misclassified noise samples as P or S phases (lower subplots). (P, N) means P-phase samples predicted as Noise, and similarly for other cells in the confusion matrix.

appears to be challenging. Moreover, about half of the few apparently misclassified noise samples were predicted as P or S phases with a probability of at least 90% (lower subplots in Figure 4.16b and d), suggesting that these presumed noise records might actually represent phases previously undetected by the automatic detection procedure, similar to the examples shown in Figure 4.9.

We next analyzed the performance of the optimized models trained for P- and S-phase picking, using test sets formed by 1,000 P- and S-phase samples randomly selected from T1 and T2 catalogs. As described in the previous section, the picking models were evaluated on augmented versions of these test sets.

Results from the picking predictions are summarized in Figure 4.17. Time residual distributions, which are narrower for P phases (Figures 4.17a and 4.17c) compared to S phases (Figures 4.17b and 4.17d), are observed in both test sets. Similar distribution shapes were

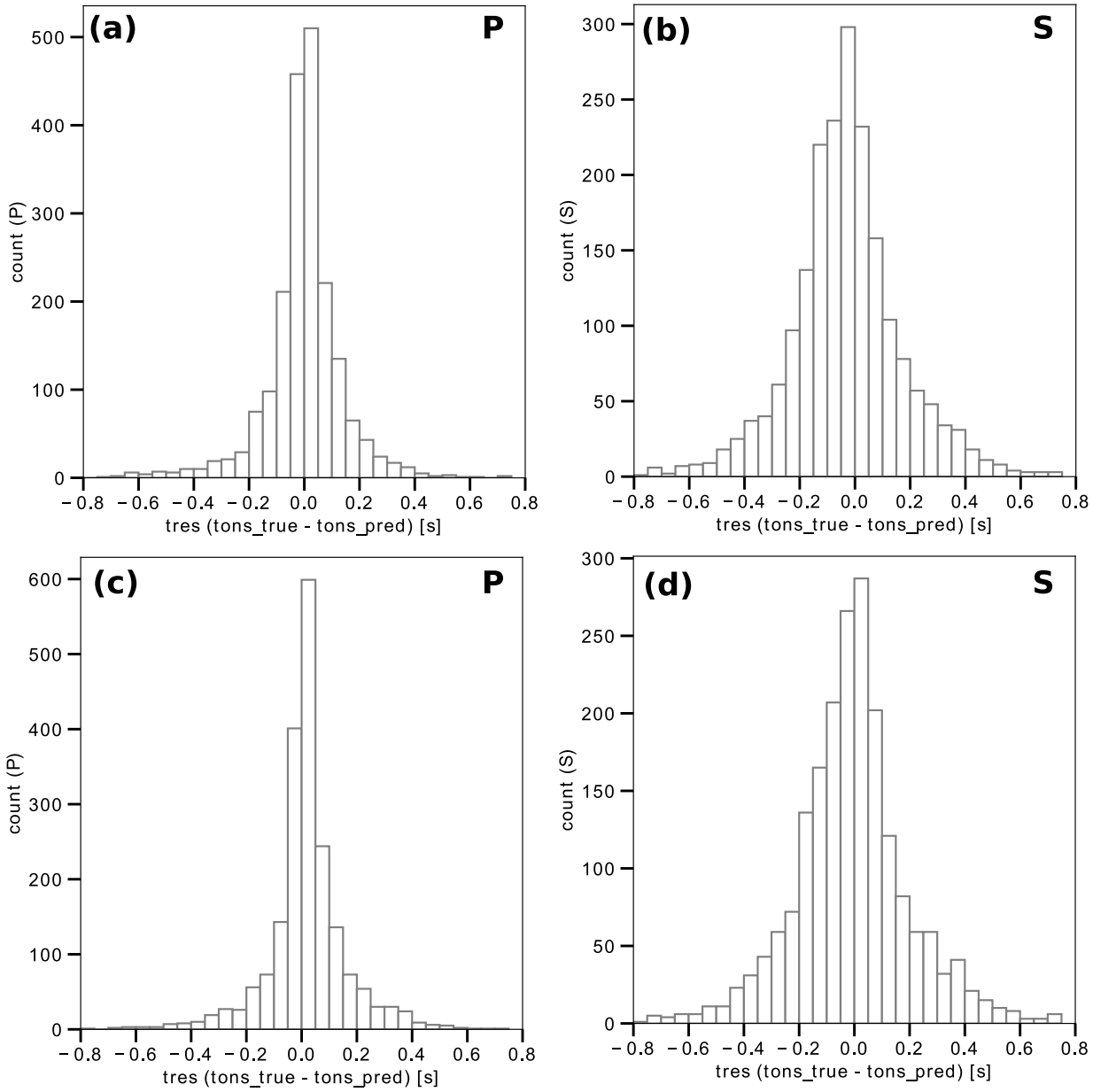

Figure 4.17: Results from the evaluation of the optimized model trained for P- and S-phase picking on independent augmented test sets formed by 1,000 x 2 samples taken from the automatically-derived earthquake catalogs T1 (a-b) and T2 (c-d). (a) P-phase time residuals distribution for test samples taken from T1 (as in Figure 4.12c). (b) S-phase time residuals distribution for test samples taken from T1 (as in Figure 4.13c). (c) P-phase time residuals distribution for test samples taken from T2. (b) S-phase time residuals distribution for test samples taken from T2.

obtained for the augmented test set derived from S1 and S2 datasets (Figures 4.12c and 4.13c), hence evidencing that the trained picking models generalize effectively to different datasets.

4.3.5 Prediction on Continuous Seismic Waveform Data

After analyzing the performance of DeepPhasePick with different independent sets, we used the best-performing trained models in the algorithm to carry out the phase detection and picking tasks on continuous seismic data recorded in two different tectonic environments. First, we applied DeepPhasePick on data which covers the time of two megathrust earthquake sequences that occurred in the northern Chile subduction zone (2007-11-14 M7.7 Tocopilla and 2014-04-

01 M8.1 Iquique mainshocks). In this case, the data used was recorded by network stations situated in desertic areas, far from cities, and therefore exposed to low background seismic noise. Next, DeepPhasePick was applied on seismic data that partly covers the aftershock series of a recent M6.4 earthquake which occurred on 2019-11-26 near the port town in Durres, Albania, in a region of convergence between Adriatic and Eurasian plates. Here we used seismic data from a 30-station network deployed in the rupture area of the M6.4 earthquake, about one month after its occurrence. This station network covers an urban area, thus being subject to higher seismic-noise signals compared to northern Chile.

We numerically compared the prediction performance of DeepPhasePick in both tectonic regimes with the detections done by an in-house developed STA/LTA trigger algorithm applied to all vertical component waveforms. For this test, we used STA and LTA window lengths of 1.0 and 20.0 s; STA/LTA trigger and detrigger ratios of 8.0 and 1.5; and bandpass filter lower and upper corner frequencies of 2.0 and 10.0 Hz.

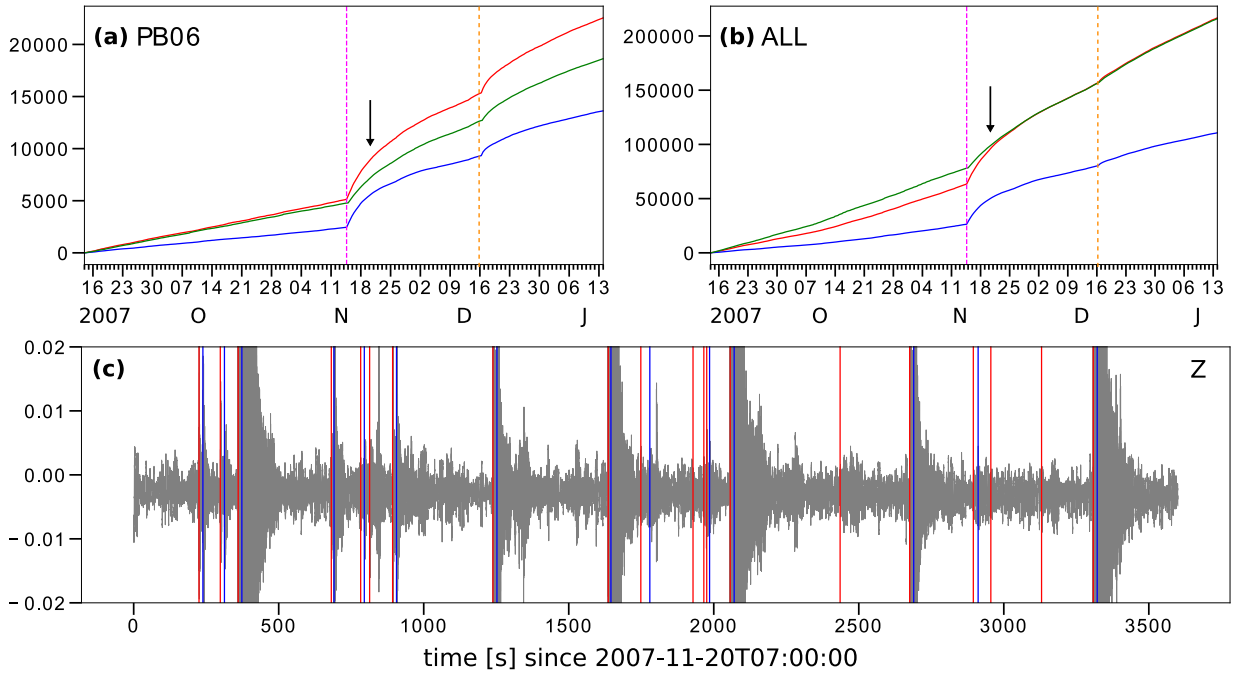

Figure 4.18: (a), (b) Cumulative number of P (red curve) and S (blue curve) phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$), and triggered detections by the STA/LTA algorithm (green curve) on continuous seismic data recorded at the PB06 station (a) and across all IPOC stations with available data (b). Cumulative curves are plotted for the time interval covering two months before up to two months after the 2007 M7.7 Tocopilla mainshock (2007-11-14), which is marked as a magenta dashed line. An orange dashed line indicates the time of occurrence of a M7.1 aftershock (2007-12-16) ~50 km south of the mainshock epicenter. (c) One-hour waveform snippet of PB06 station data, depicting the phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$) at the time indicated by the arrows in (a) and (b): red and blue lines represent P and S phases, respectively.

4.3.5.1 Performance in a Lower-seismic Noise Region: Northern Chile Subduction Zone

The cumulative number of P and S phases predicted by DeepPhasePick on data from one representative as well as all available stations in northern Chile, depict an abrupt increase right after the occurrence of the 2007 M7.7 Tocopilla mainshock (Figures 4.18a and 4.18b) and the

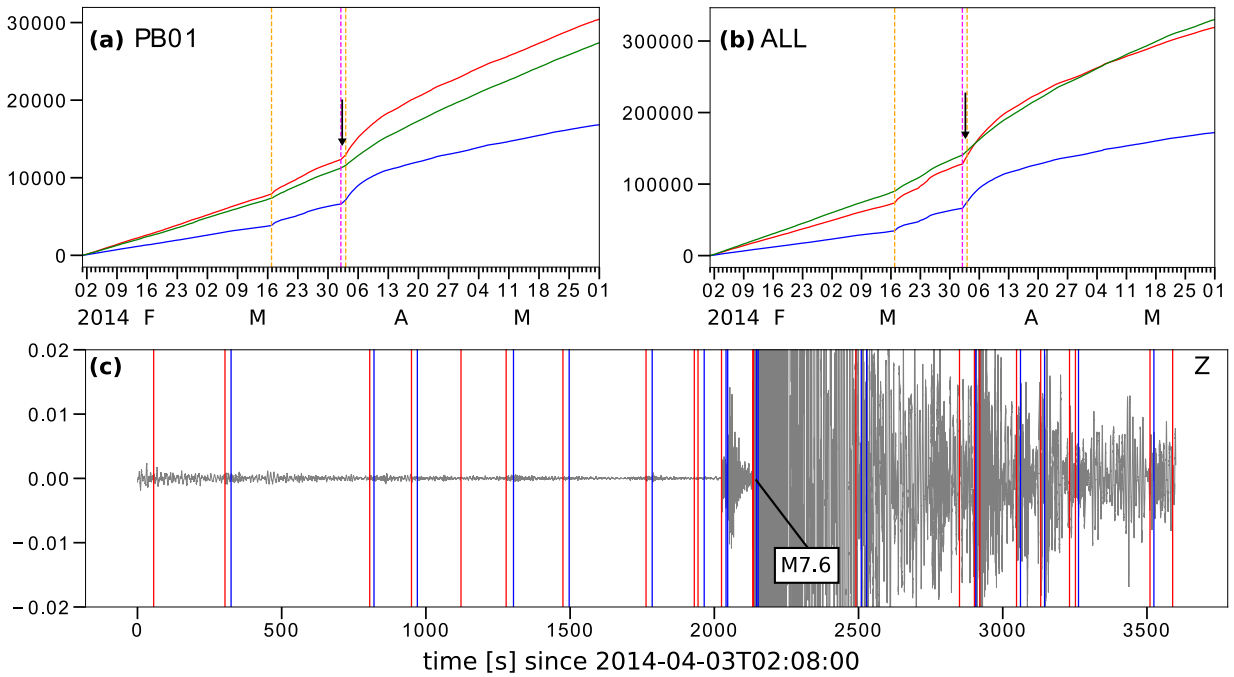

Figure 4.19: (a), (b) Cumulative number of P (red curve) and S (blue curve) phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$), and triggered detections by the STA/LTA algorithm (green curve) on continuous seismic data recorded at the PB01 station (a) and across all IPOC stations with available data (b). Cumulative curves are plotted for the time interval covering two months before up to two months after the 2014 M8.1 Iquique mainshock (2014-04-01), which is marked as a magenta dashed line. Two additional orange dashed lines indicate the time of occurrence of the largest foreshock (M6.7; 2014-03-16) and aftershock (M7.6; 2014-04-03) in the Iquique sequence. (c) One-hour waveform snippet of PB01 station data, depicting the phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.95$) at the time indicated by the arrows in (a) and (b): red and blue lines represent P and S phases, respectively.

2014 M8.1 Iquique mainshock (Figures 4.19a and 4.19b), corresponding to the beginning of the aftershock series of these large megathrust earthquakes. Less pronounced increments in cumulative predictions are observed at other times as well: after a M7.1 aftershock occurring approximately one month following the Tocopilla mainshock (Figures 4.18a and 4.18b), and after the occurrence of the largest foreshock (M6.7; 2014-03-16) and aftershock (M7.6; 2014-04-03) of the Iquique sequence (Figures 4.19a and 4.19b).

DeepPhasePick successfully captured a large majority of the various sized events occurring in the early postseismic stage of the Tocopilla and Iquique mainshocks (Figures 4.18c and 4.19c), which demonstrates the high resolution power of the algorithm. In particular, DeepPhasePick was capable of detecting P and S phases of several events occurring only a few minutes after the M7.6 aftershock in the Iquique earthquake sequence (Figure 4.19c).

4.3.5.2 Performance in a Higher-seismic Noise Region: Albania

For most of the stations in northern Chile, the cumulative number of P phases predicted by DeepPhasePick is comparable to the cumulative detections made by a classical STA/LTA algorithm (Figures S9 and S10), as one might expect from high-SNR seismic data. However, this does not hold true for several stations in Albania (Figures S11 and S12), which presumably recorded lower-SNR seismic data, though the overall cumulative number of predicted P phases and STA/LTA detections across all the stations in this region is still comparable (Figure 4.20b).

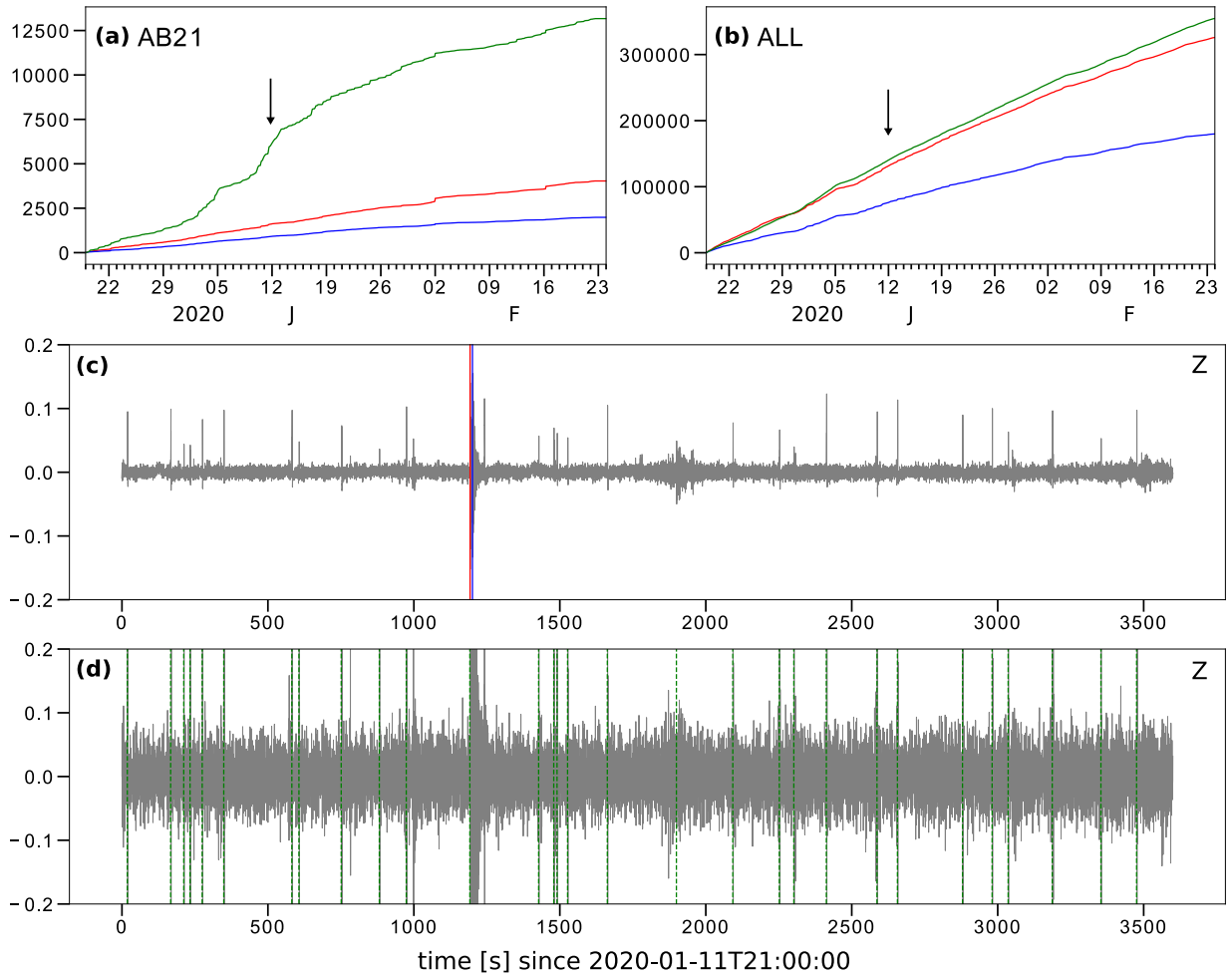

Figure 4.20: (a), (b) Cumulative number of P (red curve) and S (blue curve) phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$), and triggered detections by the STA/LTA algorithm (green curve) on continuous seismic data recorded at the AB21 station (a) and across all Albanian stations with available data (b). Cumulative curves are plotted for the time interval between 2019-12-19 and 2020-02-24, during the aftershock sequence of the 2019 M6.4 Albania earthquake (2019-11-26). (c) One-hour waveform snippet of AB21 station data, depicting the phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$) at the time indicated by the arrows in (a) and (b): red and blue lines represent P and S phases, respectively. (d) Green dashed lines depict the STA/LTA detections produced on the one-hour waveform plotted in (c).

The discrepancy between the cumulative STA/LTA detections and DeepPhasePick predictions in the Albanian data arises mostly in two distinguishable scenarios. In the first scenario, cumulative STA/LTA detections greatly surpass P- and S-phase predictions (e.g., stations AB05, AB12, or AB21). Figures 4.20c and 4.20d illustrate one example of this, where the STA/LTA algorithm detects numerous false positives in the noisy one-hour waveform, whereas DeepPhasePick predicts only one P- and S-phase occurrence. In the second scenario, cumulative P-phase predictions are considerably higher than STA/LTA detections (e.g., stations AB10 or AB27) (Figures S11 and S12). In this case, presumed false positive predicted by DeepPhasePick could be discarded, for instance, if they are not detected in a minimum number of stations when applying a phase associator algorithm.

In spite of eventual mispredictions generated due to noisy data, DeepPhasePick is generally able to detect multiple events across a station network. One example of multi-station prediction is shown in Figure 4.21, where six events are well detected in at least five stations within a 400-s

window. Figure 4.22 shows the corresponding time onsets for one of these events, generated by applying the MCD method in the picking stage. Similar results for two more events shown in Figure 4.21 are presented in Figures S13 and S14. From the statistics of the picks predicted for the three events shown in Figures 4.22, S13, and S14, which are reported in Tables S1, S2, and S3, we observe that the difference between the preliminary phase time onsets (*tons_prelim*) obtained in the phase detection stage and the refined ones (*tons_pred*) computed in the picking stage can be up to 0.13 [s] for P phases and up to 0.17 [s] for S phases.

4.4 Discussion

We have demonstrated that DeepPhasePick can be successfully applied to accurately detect and pick P and S phases originating from local earthquakes. The accurate predictions produced by DeepPhasePick result from the highly optimized set of hyperparameters defining its convolutional and recurrent deep neural networks trained for the tasks of seismic phase detection and picking, respectively. The systematic optimization process implemented aimed at compensating the limited seismic data used for training the models.

Results from the phase detection model optimization give us clues on the network architecture as well as input seismic samples characteristics that make the model perform more efficiently. Firstly, the optimization indicates that the network detection capacity is enhanced by using a deeper block of convolutional layers. This can be explained by the fact that the model's ability to learn patterns from the input data, as well as the model's capacity to generalize so as to identify those patterns in new unseen data, are usually enhanced by adding convolutional layers to a CNN-based model. A large number of dense layers in the network, which aim at learning non-local relationships in the input data, does not seem to be required for improving network performance. Secondly, the network learns features in P and S phases more effectively when it is fed with relatively long input seismic windows which have not been filtered. Longer windows may ease the extraction of relevant seismic features in order to better discriminate between P and S phases and identify noise, particularly if very long-period background seismic noise is present. Interestingly, the network detects P phases better when sample windows contain a larger fraction of noise prior to a shorter portion of signal, in a way that resembles how analysts commonly identify this phase. Unlike P phases, optimized detection of S phases is obtained on sample windows consisting of a similar fraction of noise relative to signal, which may imply that the model learns patterns from features such as the S coda in order to recognize this phase.

Predictions performed on seismic samples from two independent test sets show that DeepPhasePick is capable of recognizing manually as well as automatically picked P and S phases with high accuracy, although it decreases for lower-quality automatic picks. These results also demonstrate that DeepPhasePick predicts phase time onsets which are comparable to those picked by analysts, as can be seen from the narrow time residual distributions in Figures 4.12c and 4.13c. These residuals are comparable to those obtained in previous studies that implemented a deep learning-based picking approach (Ross et al., 2018a; Zhu and Beroza, 2019),

even though a much smaller dataset was used for training DeepPhasePick. Moreover, some degree of correlation between predicted and analyst pick classes was found (see Figure 4.14). We note here that phase-pick weighting made by analysts is not an exact measure, but usually based on experience and intuition. Hence we would not necessarily expect a perfect correlation, and the weighting provided by DeepPhasePick might in fact be more objective.

In particular, we have shown that DeepPhasePick predicts P- and S-phase time onsets which are at least comparable to the time onsets obtained by applying dedicated automatic picking algorithms such as *MPX* (Aldersons, 2004; Di Stefano et al., 2006) and *Spicker* (Diehl et al., 2009), which served to derive the high resolution catalogs for the Northern Chile region shown in Figure 4.15. Since DeepPhasePick also computes uncertainties and weights associated to the predicted picks, it can be used as a compelling alternative to those methods.

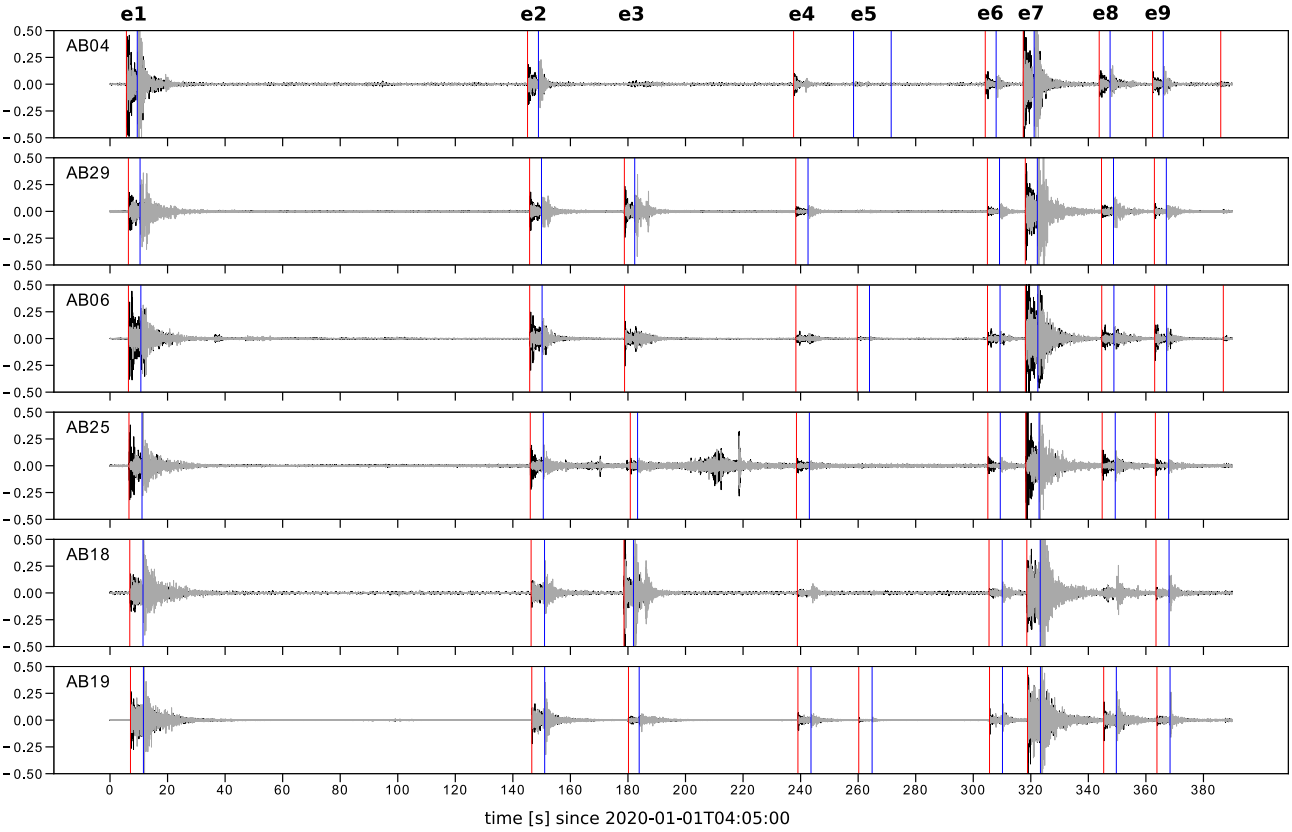

Figure 4.21: Example of multi-station phase prediction made by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$) on continuous seismic data in Albania. Red and blue lines mark predicted P and S phase time onsets, respectively. Horizontal north component (gray) is plotted on top of vertical component (black) in each station waveform. Several events are clearly detected across the station network.

However, DeepPhasePick is not restricted for use in the northern Chilean subduction zone, where the seismic samples used in training the models were originated. The algorithm is also able to predict P and S phases from local earthquakes occurring in a different tectonic regime, as we demonstrated for seismic data recorded by a 30-station network in a region in Albania (e.g., Figure 4.21).

DeepPhasePick has been designed to internally share the knowledge learned from the optimization of the phase detection model to the phase picking models, allowing the algorithm to perform both tasks in a joint two-stage process. Consequently, DeepPhasePick can be applied

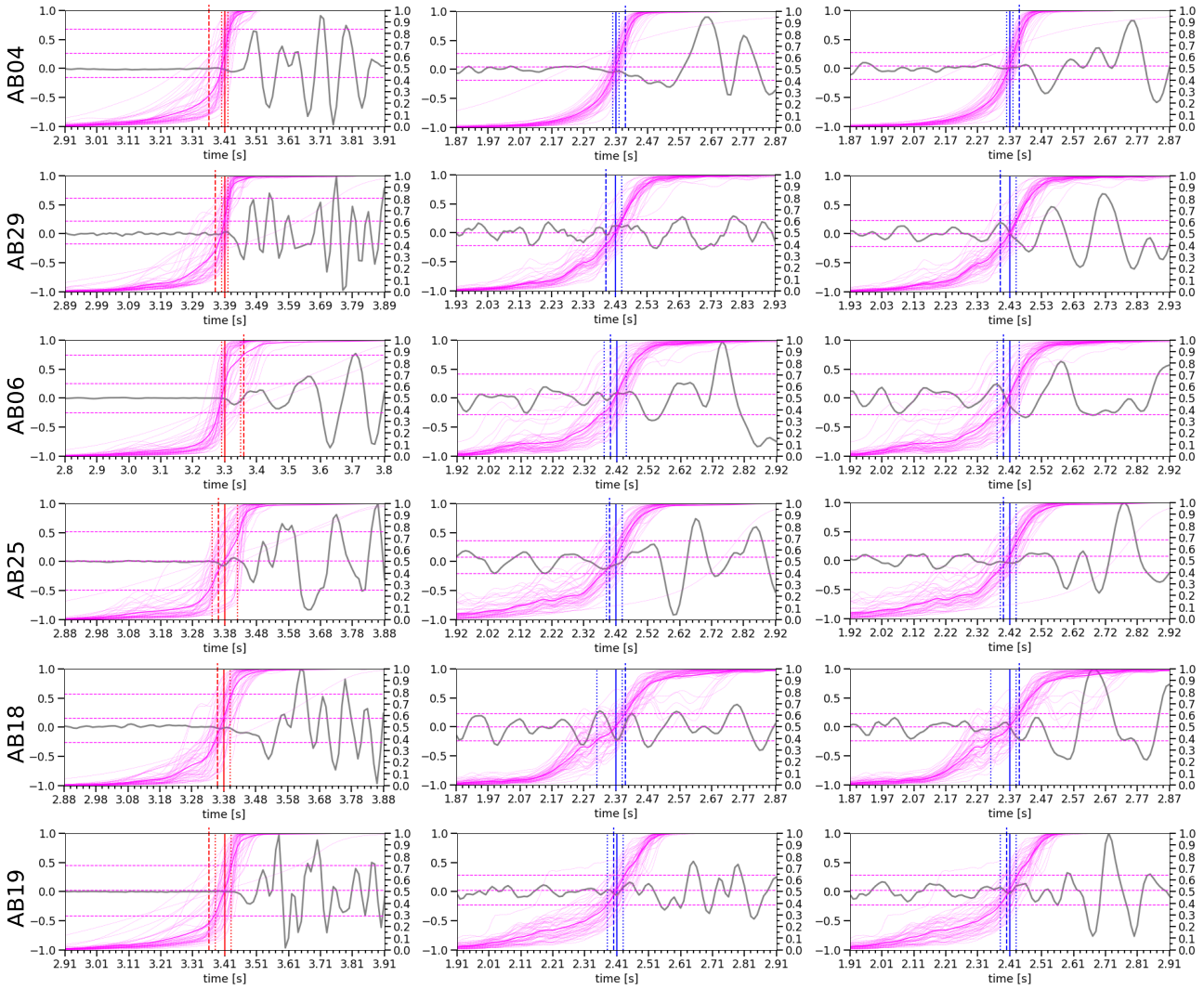

Figure 4.22: Estimation of time onset and associated uncertainty for predicted P (left) and S (center: E component, left: N component) phases of the event **e1** in Figure 4.21. Results from applying MCD method to each seismogram component are shown as in Figure 4.3. Pick statistics for this event are reported in Table S1.

directly on continuous seismic waveforms with minor preprocessing involved, so as to determine accurate P and S time arrivals that can then feed a phase associator algorithm in the next stage of an automatic earthquake location workflow.

Further improvement of DeepPhasePick can be addressed in future work, for instance by retraining the CNN-based phase detection model with additional manually-picked samples from a new region of interest. A simple way to achieve this upgrade would be to use the new data for applying techniques such as feature extraction and fine tuning on the optimized model here presented. Further, the RNN-based phase picking models can be retrained by using more than one shifted version of each available seismic record (see e.g., Ross et al., 2018a). This would provide a broader spectrum of learnable sequential data during the training, which might improve the picking models performance, though with the caveat that the computing time required to train the models would further increase.

4.5 Conclusion

The fast development in deep learning algorithms has made it possible to reach super-human performance levels in tasks involving huge amount of data, such as image classification and natural language processing. In this work, we leveraged this computational progress to develop DeepPhasePick, a novel two-stage algorithm for detection and picking of P and S seismic phases originating from local earthquakes.

DeepPhasePick has been built based on highly optimized convolutional and recurrent deep neural network architectures trained for such tasks. In a first stage, DeepPhasePick reports probabilities of waveform samples belonging to three phase classes (P, S or Noise). Preliminary phase onsets obtained from these probabilities are refined in a second step, in which final time onsets and their associated uncertainty are obtained by applying the Monte Carlo Dropout regularization technique, as an approximation of Bayesian variational inference.

DeepPhasePick has proven capable of recognizing both manually and automatically picked P and S seismic phases with high accuracy. It can also predict phase time onsets, which are comparable to those picked by analysts or derived from conventional, dedicated automatic phase picking algorithms. The P- and S-phase time onsets, as well as their associated uncertainties, predicted by DeepPhasePick from continuous seismic data can be directly used to feed a phase associator algorithm as part of an automatic earthquake location workflow.

Chapter 5

Conclusions and Outlook

5.1 Synthesis and Conclusions

Chapter 1 introduced the following research questions:

1.1 How do observed seismicity patterns of the Iquique mainshock sequence correlate with static stress changes induced by this earthquake, as well as with coseismic, afterslip, and interseismic locking?

1.2 What does this relationship tell us about the frictional behavior of the northern Chilean portion of the megathrust and possibly of other subduction regions?

2. What role may the forearc and incoming plate structures of the northern Chile convergent margin play in producing the seismicity pattern observed in the Iquique mainshock sequence?

3. Can we make the seismic phase detection and picking procedures involved in an earthquake location workflow more precise and efficient using the huge amount of seismic data collected in northern Chile?

4. Can a potentially alternative phase detection and picking method be generalized to other regions where a different tectonic regime is observed?

The subsequent sections summarise the main findings of my thesis regarding these research questions.

5.1.1 Correlation between Seismicity and Other Data during The Iquique Earthquake Sequence, and Insights about the Frictional Behavior along the Plate Interface

A primarily along-dip segmentation of small-scale frictional heterogeneity on the plate interface is illuminated from the ~19,000 events catalog, which was derived by applying an automatic earthquake location workflow using data from >100 stations in northern Chile.

An aseismic updip-most portion of the megathrust is followed by a transitional region further downdip interpreted as being predominantly characterized by aseismic creeping (velocity-strengthening) embedding small-scale conditionally stable and seismically unstable (velocity-weakening) areas. Conditional stability is proposed for the region updip of the main coseismic rupture, where coseismic slip is allowed to propagate, though it behaves intrinsically aseismic. This is supported by the presence of afterslip, which is likely to occur at this depth level and could explain the high aftershock activity as well as the numerous repeating earthquakes found there. Afterslip in this region may be due to postseismic stress relaxation or occur in velocity-strengthening areas under the stress shadow caused by the adjacent mainshock asperity. Repeating earthquakes are also found further south, along several seemingly real ~EW striking streaks, which are then interpreted as velocity-weakening patches, indicators of surrounding aseismic creep on the plate interface.

Deeper downdip, the central portion of the plate interface is interpreted as the seismogenic velocity-weakening segment where the main coseismic rupture of both the Iquique mainshock and its largest M7.6 aftershock occurred. Aftershock seismicity forms a belt that flanks the downdip edge of the mainshock asperity. This seismicity belt partially correlates with increased Coulomb failure stress and anticorrelates with the two major geodetically-derived afterslip patches. This portion is interpreted as featuring a heterogeneous frictional behavior; presumably conditionally stable where the belt of aftershock seismicity occurs and velocity-strengthening correlating with the two afterslip patches.

The central segment of the plate interface is connected to the deepest interplate events through a region of remarkably sparse aftershock seismicity. Similar event sparsities, observed after the 2010 M8.8 Maule earthquake (Lange et al., 2012), and before and after the 2015 M8.3 Illapel earthquake (Lange et al., 2016), hint to a presumably generic depth segmentation of the plate interface seismicity along at least a significant part of the Chilean subduction.

Two clusters formed by the deepest plate interface seismicity exhibit significantly different dip and are accompanied by several repeater earthquakes. The vicinity of these clusters to the imaged upper plate Moho, where a downdip transition to aseismic behavior has been hypothesized to occur at least in cool old subduction zones (e.g., Oleskevich et al., 1999), may indicate similar fault mechanics and slip properties at this depth level to those interpreted for the streaks observed in the updip transition zone.

5.1.2 Role of the Forearc and Subducting Plate Structures in the Observed Seismicity Pattern

The continental forearc in the study region is interpreted as being composed of two main bodies: 1) An outer wedge, formed by a ~25-30 km wide aseismic frontal prism (FP) and a transitional zone (TZ). Whereas the presumed highly fractured and deformed FP overlies the updip-most aseismic portion of the plate interface, the TZ presumably features an upper plate fracturing gradually reduced landward and is located approximately below the middle conti-

nental slope and above the transitional frictional portion of the interface. 2) An inner wedge, presumably formed by more consolidated framework rock and which approximately overlies the seismogenic velocity-weakening segment of the plate interface, where the highest coseismic slip occurs.

The hypothesis that the FP may have acted as a barrier for the updip propagation of the mainshock rupture (Maksymowicz et al., 2018) is supported by the lack of foreshock and aftershock seismicity observed in the updip-most portion of the plate interface and outer wedge, as well as the gradual increase in coseismic slip landward.

Possible activation of offshore upper plate faults is evidenced by numerous ~NS oriented thrust faulting aftershocks with dips steeper than the plate interface. These aftershocks, which occur in the transitional region updip of the main coseismic rupture of both the Iquique mainshock and its largest aftershock, could have been associated with the outer wedge deformation resulting from its postseismic stress relaxation. Presumed preseismic upper plate deformation in the same region (Hayes et al., 2014) is supported by a number of thrust foreshocks, which feature fault planes significantly rotated relative to the plate interface.

The substantially rugged seafloor of the incoming oceanic plate may cause fracturing of the plate interface, upper plate deformation, and reduction of the plate coupling, all of which can favor afterslip on the plate interface (Wang and Bilek, 2014). This hypothesis is supported by the existence of the Iquique Ridge offshore northern Chile, with its numerous seamounts, and the horst-and-grabens and spreading fabrics which form the incoming plate relief. Besides, underthrusting plate relief is suggested by large embayments and NW-SE trending antiforms and synforms seen in the lower continental slope from the seafloor bathymetry (Figures 3.2a and 3.6; Geersen et al., 2018a), and by presumed subducting seamounts (Figure 3.6; Geersen et al., 2015). One of these subducting seamounts coincides with the location of the major ~EW striking streaks observed. Another seamount correlates with a large cluster of earthquakes with significantly diverse faulting.

The plate interface offshore Iquique is interpreted to not only be deformed and fractured due to the rugged subducting relief, but possibly also corrugated. Corrugations have been imaged in an offshore portion of the Costa Rica subduction plate boundary (Edwards et al., 2018), where seamounts and rigdes are also being subducted. Presumed plate interface corrugations may be a structural explanation for the observed ~EW striking streaks, as they can serve as conduits for the transportation of fluids, which eventually induce shear failure that triggers seismicity and repeating earthquakes along the streaks. Possible indirect evidence for a corrugated megathrust could be provided by several small thrust earthquakes featuring fault planes steeper or rotated relative to the interplate fault surface orientation suggesting local upper plate deformation in the area where the major streaks are observed.

5.1.3 Using Deep Learning as an Efficient and Generalized Alternative Method to Classical Phase Detection and Picking Procedures

Instead of exploiting presumed characteristics of seismic signals by conventional automatic phase detection and picking methods, a deep learning-based approach is capable of extracting meaningful features directly from a training set of seismic samples. The learned features can then in turn be used to make generalized predictions on new seismic data.

Convolutional neural networks' ability of learning spatial features from data, as well as the adeptness of recurrent neural networks at learning contextual sequential data patterns, are combined in DeepPhasePick; the automatic algorithm presented in this thesis trained for detecting and picking P and S seismic phases originating from local earthquakes.

The systematic optimization of the hyperparameters defining DeepPhasePick network architectures suggests that the accuracy of the method's predictions improves by increasing the number of both the convolutional layers used for detecting the seismic phases and the number of LSTM recurrent layers used for P- and S-phase picking. Prediction performance also improves when the algorithm is fed with relatively long, non-filtered seismic windows, where the fraction of noise is larger than (for P phase) or similar to (for S phase) the subsequent signal. These preferred choices are somewhat intuitive. Shortening and filtering signals might discard some meaningful seismic features. Also, whereas analysts usually better identify P phases by using a larger portion of noise before the phase onset, S phases may be better recognized by taking into account features such as the S phase coda.

Even though the DeepPhasePick models were trained using a limited amount of seismic records (~39,000 records for detection and ~36,000 records for picking), the models' highly optimized architectures and internal sharing of learned knowledge facilitate the presented method's effective performance, which has been demonstrated in both of the trained tasks. For the phase detection task, DeepPhasePick reaches an accuracy of 99% in predicting the class (P, S, or Noise) of manually-picked seismic records. This is comparable to the accuracy reached when detecting the class of high-quality automatically-picked samples, although it decreases by up to 15% and 19% respectively for lower-quality P and S samples. For the phase picking task, DeepPhasePick has proven capable of predicting manually-picked onsets with very high precision (for the most part within 0.2 seconds). The precision just slighty decreases when predicting automatically-picked onsets obtained from conventional, dedicated automatic phase picking algorithms such as *MPX* or *Spicker* (Aldersons, 2004; Diehl et al., 2009; Di Stefano et al., 2006).

The fact that DeepPhasePick computes probabilistic derived asymmetric uncertainties accompanying the predicted phase time onsets, and that it does not require preliminary input phase picks but rather uses internally predicted ones, makes the presented method a compelling alternative to the above mentioned conventional automatic picking algorithms. Thus, DeepPhasePick can be integrated in an automatic earthquake location procedure, by using its predicted phase arrivals and uncertainties to feed a subsequent associator algorithm.

DeepPhasePick's detection and picking abilities have proven to perform effectively on both

lower- and higher-seismic noise data. The lower-seismic noise data used for this test was recorded in northern Chile, hence presenting similar signal content to the data used to train the algorithm. On the other hand, the higher-seismic noise data was registered in an urban region in Albania, which features different tectonic characteristics from the training data set. This demonstrates the method's ability to effectively generalize.

DeepPhasePick is also efficient, in the sense that although many hours are required to optimize and train the algorithm, this is only carried out one time. In production mode, predictions performed on continuous seismic waveforms only take seconds to minutes, depending on the length of the seismic traces and the number of stations on which the algorithm is applied.

5.2 Outlook

My thesis demonstrates how a high-resolution automatically-derived earthquake catalog can be used to illuminate the small-scale frictional heterogeneity observed along subduction plate boundaries, where regions of stable aseismic, conditionally stable, or unstable seismic behavior can be observed during megathrust earthquakes. Analysing this seismicity, in combination with different datasets such as geodetically inferred afterslip and interseismic locking, and coseismic slip solutions, can help to better understand the tectonic processes and better assess the earthquake hazard potential in highly active seismic regions such as the Chilean subduction zone.

However, in order to obtain a comprehensive understanding of the frictional and tectonic variability, both seismological and geodetic observations registered at shallow depths in the offshore region of subduction zone plate interfaces need to be included in the analysis. In particular, future seismological studies should ideally integrate submarine ocean-bottom seismometer (OBS) data with inland data, which would provide further constraints especially for the hypocenter depths of shallow offshore earthquakes. This would allow for more reliable interpretations regarding relevant features such as the activation of upper plate splay faults or the existence of presumed corrugations on the plate interface.

Future studies in subduction zones aiming at investigating small-scale seismicity distributions will most likely have to include additional data either from OBSs or from a higher number of inland seismic stations. Other studies aiming to generate earthquake catalogs could cover longer time periods. In both scenarios, the amount of waveform seismic records may easily increase up to the point where it becomes unmanageable, since processing the entire earthquake catalog generation workflow would become too computationally expensive. Hence, a more efficient automated data processing approach, for instance based on deep learning, becomes critical.

DeepPhasePick accomplishes the first of the stages involved in an automatic earthquake location workflow, where P and S phases are detected and picked. This method can be easily upgraded by retraining its CNN and RNN constituent architectures, with the goal of enhancing prediction generalizability. The CNN model performing the phase detection can be retrained by applying feature extraction and fine tuning techniques on new manually-picked seismic records

obtained from a different study region. These new labelled samples can also be used to retrain the RNN model performing the phase picking. Optionally, the number of shifted copies of the records can be increased compared to that used in this study, which would further broaden the spectrum of possible learnable features, that in turn can improve the picking performance.

In the near future, I envisage that the phase arrivals together with their uncertainties generated by DeepPhasePick can be used to feed a phase associator algorithm which is also built based on optimized neural networks (e.g., Ross et al., 2019b). Further, it is conceivable that the phase arrivals already associated to events can then be provided to an additional deep learning-based algorithm trained specifically for locating earthquakes. I am convinced that, ultimately, having an automatic earthquake location workflow entirely built on optimized deep learning methods would significantly facilitate and improve the efficiency in processing the fast-growing available seismic datasets.

Supplementary Information

This supplement contains seismic waveform data information, video captions, figures and tables in support of Chapter 3 and Chapter 4.

S1 Data Availability

The catalog of relocated seismicity generated in Chapter 3 and analyzed in this study can be accessed from Soto et al. (2019a). Seismic waveform data used in Chapter 3 were taken from networks CX (GFZ German Research Centre for Geosciences, and Institut des Sciences de l'Univers-Centre National de la Recherche CNRS-INSU, 2006), IQ (Cesca et al., 2009), 3D (Asch et al., 2014), and GE (GEOFON Data Centre, 2013) accessed via EIDA webservices (e.g., *https://geofon.gfz-potsdam.de/*), as well as from Chilean Seismological Network (C, C1) stations (Barrientos, 2018) accessed via IRIS webservices (*http://ds.iris.edu/SeismiQuery/*). Additional waveform data were used from the MEJIPE temporary network deployed by FU Berlin (Salazar et al., 2013) accessed via EIDA web services as well as from a temporary network deployed by the Chilean ONEMI, DGF, and CSN institutions accessed from CSN upon request.

For training DeepPhasePick models for phase detection and picking tasks presented in Chapter 4, we used waveform earthquake data collected from networks ZB (Schurr et al., 1997), ZE (Haberland et al., 1996), 8F (Wigger et al., 1996), Y9 (Sobiesiak and Schurr, 2007), CX, and GE.

For the evaluation of the performance of DeepPhasePick in independent test sets built from automatically-derived earthquake catalogs, we used waveform earthquake data taken from networks CX, IQ, 3D, and GE, as well as from Chilean Seismological Network (C, C1) stations, the MEJIPE temporary network, and a temporary network deployed by the Chilean ONEMI, DGF, and CSN institutions accessed from CSN upon request. The performance of DeepPhasePick predictions was further evaluated on continuous waveform data taken from networks CX and 9K (obtained from GEOFON data centre).

The optimized DeepPhasePick models trained for phase detection and phase picking tasks, together with an example script that applies DeepPhasePick method for both tasks on continuous waveforms, can be accessed via the GitHub repository: *https://github.com/hsotoparada/DeepPhasePick*.

S2 Supplementary Information for Chapter 3

Text S1: Multistage Automatic Earthquake Localization.

I used seismic data consisting of three-component velocity seismograms registered by 104 seismic stations spread throughout the Northern Chile region (Figure 3.1). The raw seismic data was processed in a multistage automatic procedure, which has been adapted from Sippl et al. (2013) and is summarized in the following.

Estimation of preliminary P picks and phase association.

An in-house developed STA/LTA trigger algorithm was applied to all vertical component waveform data. After some testing with a small data set, the parameters were set as follows: the lengths of STA and LTA windows to 1.0 and 20.0 s; the STA/LTA trigger and detrigger ratio to 8.0 and 1.5; the lower and upper corner frequencies of the bandpass filter to 2.0 and 8.0 Hz. These preliminary phase alerts were then associated to preliminary events using *Binder* (Rietbrock and Heath, personal communication, 2010), requiring at least 8 picks to form an event. As input for Binder, I used an existing 1D velocity model estimated for a region somewhat south of the M8.1 Iquique earthquake epicenter (Husen et al., 1999). After this stage, the resulting events were relocated using Geiger's least squares inversion method implemented in *HYPO71* (Lee and Lahr, 1975).

Repicking of P phases.

Once the preliminary event locations and picks were determined, P onsets were repicked using the *MPX* algorithm (Aldersons, 2004; Di Stefano et al., 2006). By following a similar scheme previously applied to seismicity in the region of the Pamir in Central Asia (Sippl et al., 2013), *MPX* was run on vertical waveform data, classifying the resulting phase onsets into four quality classes. Waveform data from stations with no preliminary picks were also examined with *MPX*, using in this case reference P onsets calculated from raytracing in *NonLinLoc* (*NonLinLoc* package, Lomax et al., 2000) based on the preliminary event location and Husen's 1D velocity model. The refined picks were then used to once more relocate all the events in *HYPO71*.

Repicking of S phases.

Based on *MPX* P picks and subsequent event relocations, S onsets were automatically determined by applying the software *Spicker* (Diehl et al., 2009) and using preliminary S picks estimated by raytracing in *NonLinLoc*, in a similar way as in (Sippl et al., 2013). A new relocation in *HYPO71* was performed afterwards.

Calculation of ML.

Local magnitudes (ML) were calculated for all the events in the catalog based on the formula recommended by the IASPEI (Bormann and Dewey, 2014).

Filtering and completion of event catalog.

As I am focused on the seismicity related to the Iquique 2014 earthquake, I removed from the catalog events located outside of the study region indicated in Figure 3.1 (also see text). In order to ensure the reliability of our catalog, I further remove possible double events and

limit further analysis to events that achieve a high enough score in a weighted pick sum scheme that was also used by Sippl et al. (2013), where P picks of quality classes 0, 1, 2, or 3 add 4, 3, 2 and 1 point(s) to this sum, respectively. And S picks of same classes add 8, 6, 4 and 2 point(s), respectively. I accept only events with a minimum total sum of 24 points and a minimum sum of 12 points counting only S-picks in order to remove events that feature only P-onsets and may thus have badly constrained depths. Furthermore, I compared the catalog to the Chilean National Seismological Center (CSN) catalog and searched for possibly missed events by comparing origin times. This led to the additional inclusion of a small number (324) of previously not detected earthquakes that were then processed in the same way as outlined above.

Estimation of a minimum 1D velocity model.

I performed the inversion for a 1D velocity model based on the procedure described in Kissling et al. (1994). First I relocated the events in *VELEST* using three iterations during which I removed picks with residuals that deviated by more than two standard deviations from the event mean residual as well as had an absolute residual above 1.5 s. I then selected a subset of 2,000 events in the catalog by subdividing the study area into several cells and selecting only the best located events in each cell based on pick quality classes and excluding events with an azimuthal gap $> 230°$. The selected events were simultaneously inverted for a minimum 1D velocity model, hypocentral parameters and static station corrections, prescribing an 11-layer structure in depth, with spacing of 10 km and 20 km above and below 70 km depth, respectively, up to a maximum depth of 130 km. The minimum 1D velocity model (Figure S1a) was found after trying several initial models (modified from Husen's 1D velocity model) and diverse combinations of the key parameters in an iterative approach. Finally, the entire set of events was once more relocated in Velest using the newly calculated minimum 1D velocity model.

Estimation of a 2D velocity model.

The determination of a 2D velocity model was accomplished with the software simulps (Eberhart-Phillips, 1993; Evans et al., 1994; Thurber, 1983, 1993). I defined an inversion grid according to the density of stations and earthquakes, with a horizontal node spacing of 15-20 km that increases from the center toward the edges of the study region and the same layers in depth as used for the inversion in *VELEST*. I used the same subset of selected events as before in *VELEST* (relocated in the minimum 1D velocity model) and the minimum 1D model extended horizontally as the initial 2D velocity model for Vp inversion. I performed two consecutive inversions for Vp and Vp/Vs. In the first stage, I searched for the best P velocity model, keeping Vp/Vs fixed at the value 1.73, after which I inverted for Vp/Vs and kept Vp at the best previously determined model. The S velocity model was determined from this last inversion. The optimal roughness of the final velocity model, i.e. the balance between data and model variance, is highly dependent on the choice of the damping parameter of the inversion. Therefore, I calculated tradeoff (L-) curves between these variances in both inversions and chose the damping parameter accordingly (Figure S1b). Lastly, the final 2D velocity model

(Figure S1c) was used to once again relocate the entire event data set in *simulps*.

Double-differences relocation.

A further refinement of the hypocenter locations was accomplished using the program *HypoDD* (Waldhauser and Ellsworth, 2000) that implements a double-difference (DD) relocation scheme. Pairs of events registered at common stations and separated by small distances (clusters) are relocated relative to each other, making use of only their travel time differences and therefore avoiding path-dependent effects or model errors. I implemented an inversion based on a weighting scheme using differential travel times from both catalog and cross correlation (CC) phase pairs. CC travel times differences were calculated for event pairs separated by a maximum distance of 8 km that already had phase picks, by using windows of 2/4 s for P/S phases and retaining only measurements with correlation coefficient > 0.7. The inversion was started with highly weighted catalog phase pairs, CC data weights were gradually increased in consecutive steps whereas catalog data were down-weighted accordingly.

Relocation by focal mechanisms.

For some of the shallow events that are located far offshore, the event-station geometry was highly unfavorable. In order to independently constrain hypocentral depths, I used centroid depths of collected focal mechanisms as initial depths in a final *HypoDD* relocation run. After this relocation, the events in this area became more clustered and better depicted the continuity of the plate interface compared to the previous run without the additional constrain.

Event locations uncertainties.

I estimated absolute location uncertainties by relocating the entire catalog with a probabilistic scheme (*NonLinLoc*), using the 2D velocity model calculated in this work. The distribution of the obtained location errors, i.e. the extent of ellipses that contain 68% of the probability density function, was averaged in x-, y- and z-direction for all the events within each of the spatial 0.1° x 0.1° grid cell in which I subdivided the study region. The resulting errors are shown in Figure S2 for the interval between 0-65 km depth.

Movie S1: Foreshock series, in time period from 15 March 2014 to the mainshock occurrence (2014-04-01T23:46:45.720). Time step (time window shown in each frame) = 6 hours. Filename ms01.mp4

Movie S2: Aftershock series, in time period from the mainshock to the occurrence of the largest M7.6 aftershock (2014-04-03T02:43:13.940). Time step = 30 minutes. Filename ms02.mp4

Movie S3: Aftershock series, in time period from the occurrence of the largest M7.6 aftershock to day 16 (17 April 2014) since the mainshock. Time step = 2 hours. Filename ms03.mp4

Movie S4: Aftershock series, in time period from day 17 (18 April 2014) since the mainshock to 30 April 2014. Time step = 4 hours. Here the two high afterslip patches are plotted as in Figure 3.6. Filename ms04.mp4

Movie S5: Earthquakes in the region of major streaks S1 and S2 (same region shown in Figure 3.6c; main text) in time period between 15 March and 30 April 2014. Time step = 4 hours. Filename ms05.mp4

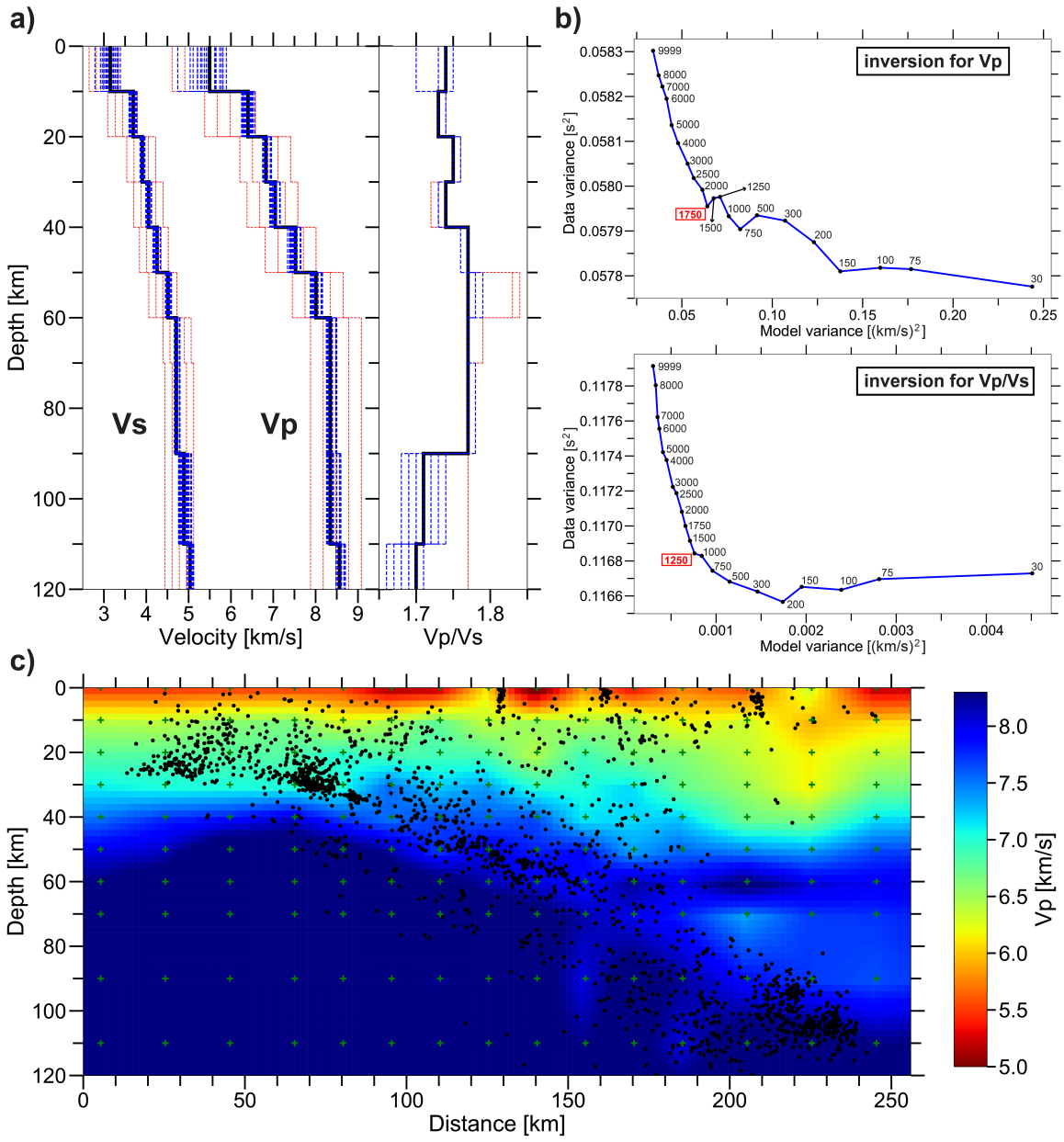

Figure S1: (a) Minimum 1D model derived using *VELEST*. The range of input models is shown with red dashed lines, the different output models with blue dashed lines. The final model is indicated with a thick black line. (b) L-curves obtained from the inversions for Vp and Vp/Vs in the determination of a 2D velocity model. The chosen damping parameters are indicated in red in each curve. (c) 2D velocity model calculated for the study region using *simulps*. Black dots denote earthquake hypocenters of the events whose traveltimes were used for deriving the model. Green plus signs mark inversion nodes.

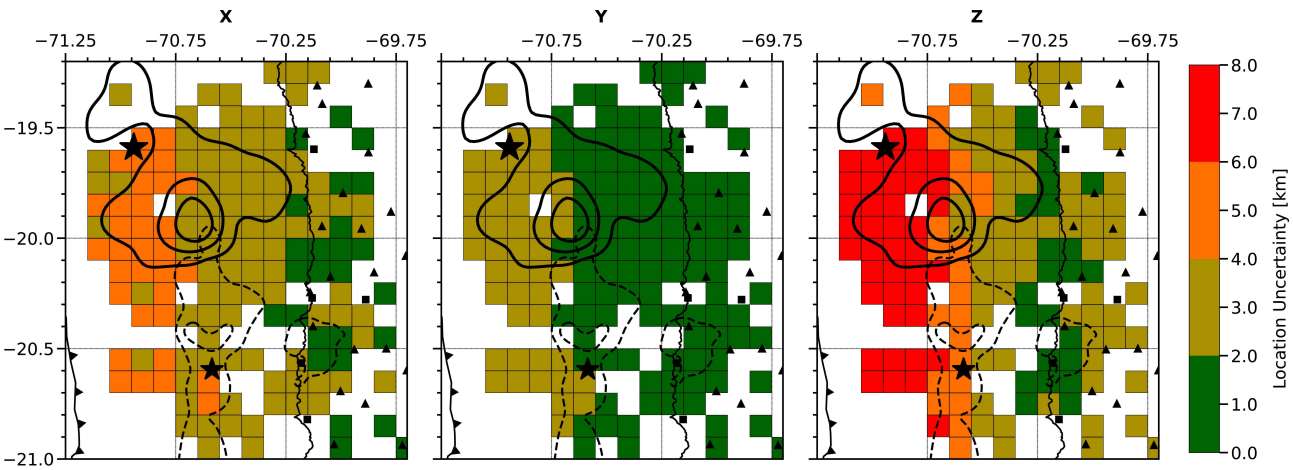

Figure S2: Average event location uncertainties (for events at 0-65 km depth) determined using a probabilistic location scheme (*NonLinLoc*). All events were relocated with this scheme, and the projection of the error ellipsoid (68%) on the x-, y- and z-direction was averaged for each spatial grid cell. Cells that contain less than 20 events were left blank. Location errors are smaller inland, where all the used stations (marked as in Figure 3.1) are located, but increase offshore, farther from the footprint of the network. Everywhere, the uncertainties in N-S direction (y) are smaller than those in E-W direction (x). In the offshore region, uncertainties in depth (z) are considerably larger than both horizontal errors (x, y).

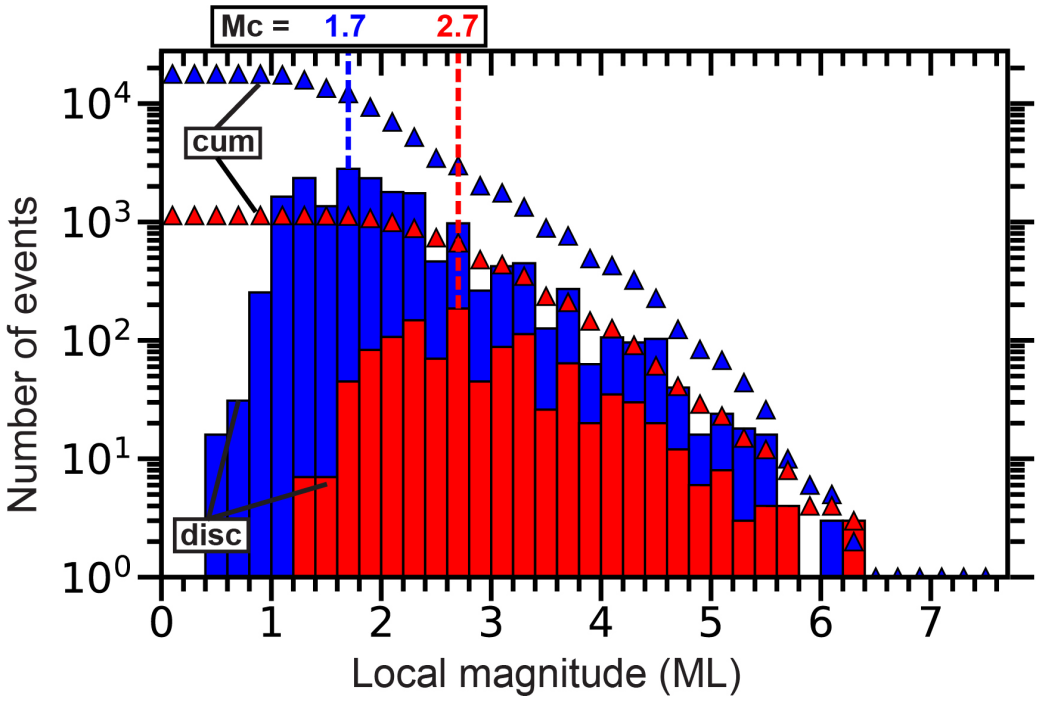

Figure S3: Discrete (disc) and cumulative (cum) distribution of local magnitudes in the relocated earthquake catalog (plotted in Figure 3.1). Red and blue histograms show the discrete distributions for earthquakes before and after the Iquique mainshock, respectively. The magnitude of completeness (Mc) for both time periods is shown as the highest frequency of events in each discrete magnitude distribution. The clearly lower Mc for the postseismic period is due to changes in network geometry: most of the temporary stations were not in place during the foreshock series, whereas most aftershocks were recorded with close to the maximum station density.

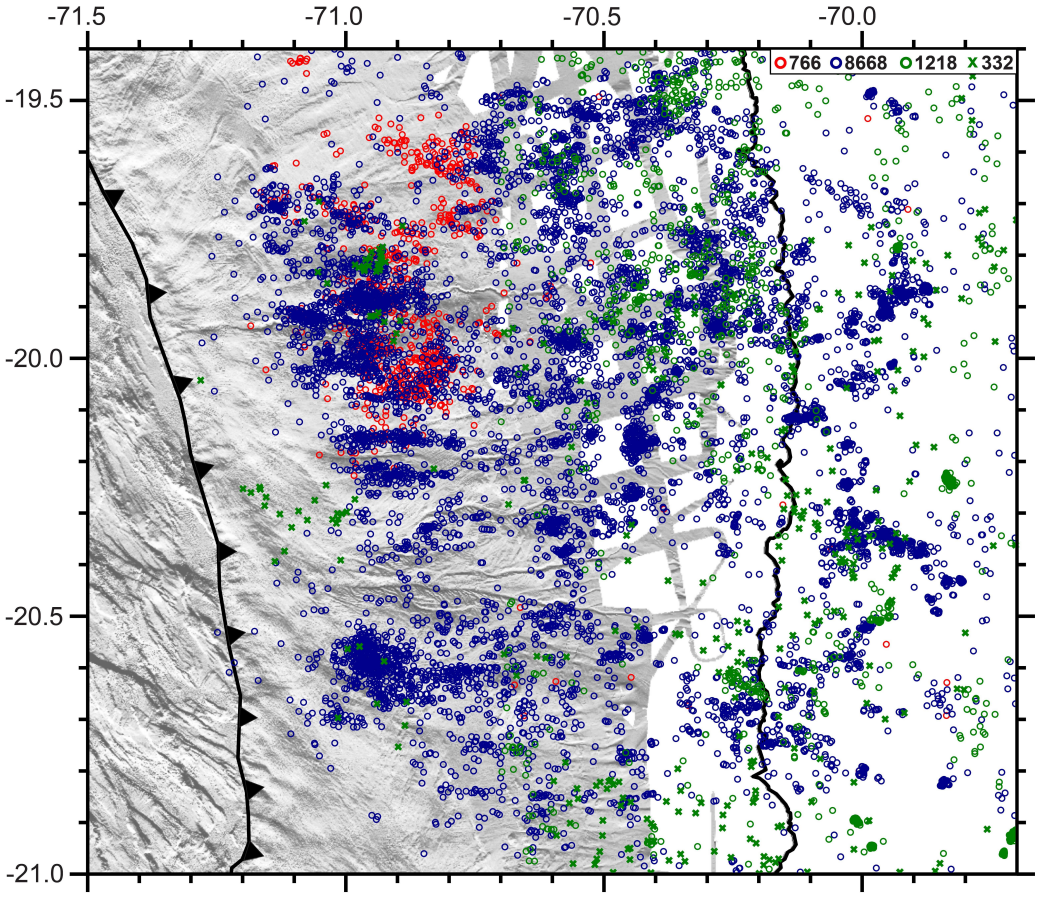

Figure S4: Zoom into the study region (red box shown in Figure 3.1), showing only the relocated seismicity as in Figure 3.2a, allowing a more detailed inspection.

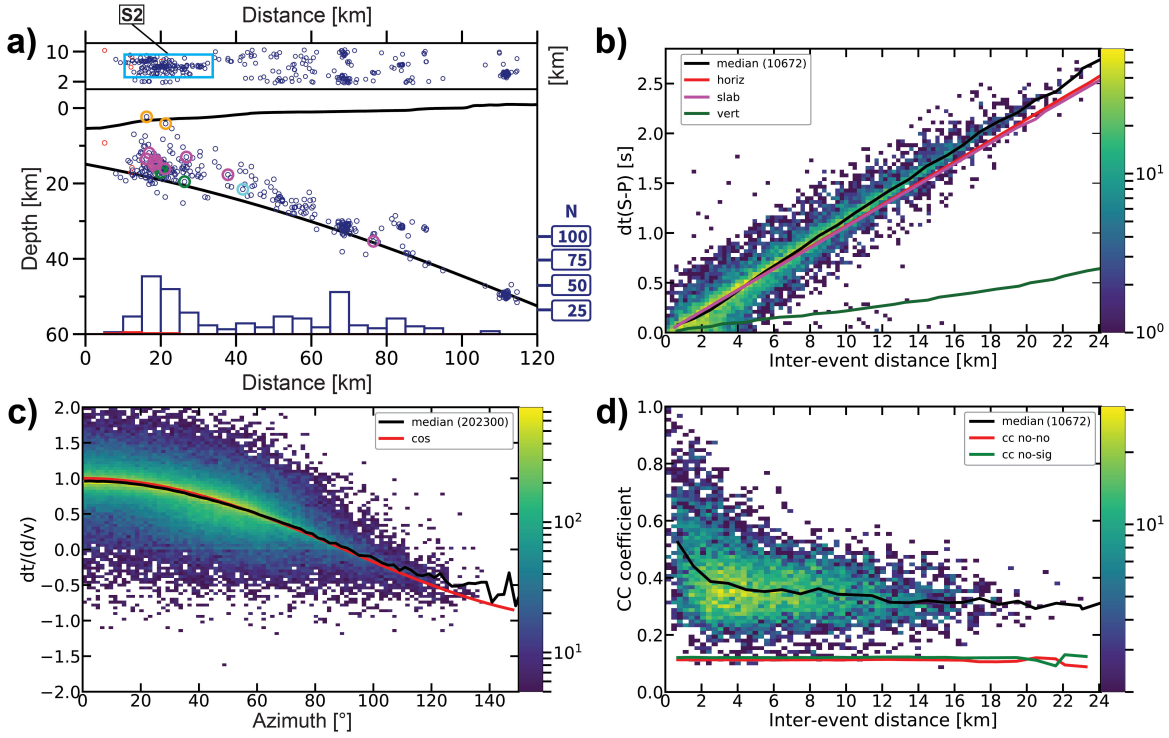

Figure S5: Detailed analysis of seismicity streak labeled S2 in Figure 3.2a. Subfigures (a-d) show similar plots as in Figure 3.4 for streak S1.

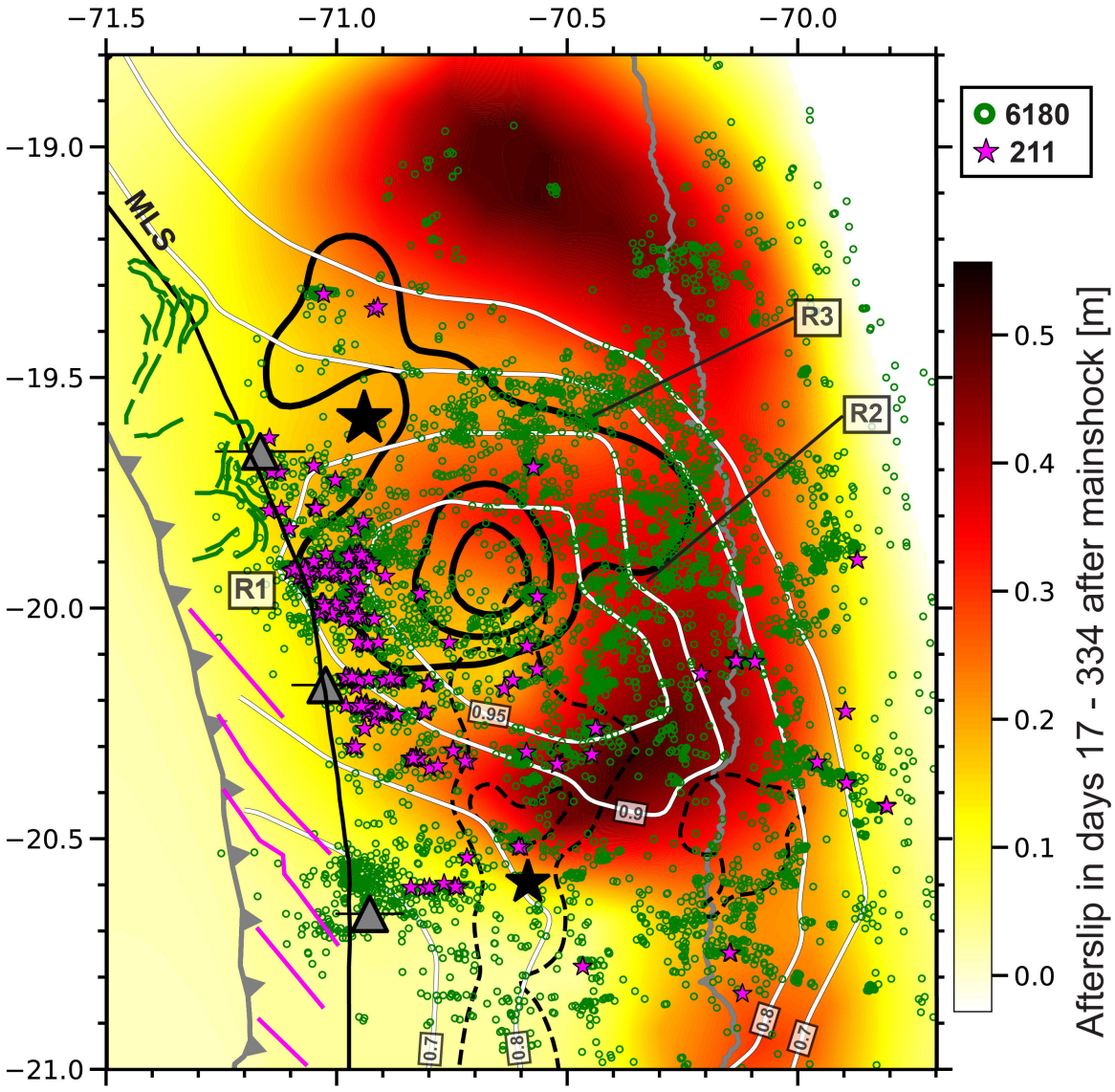

Figure S6: Plate interface aftershock seismicity since day 17 after the mainshock (18 April 2014, green circles) plotted on top of the afterslip model of Hoffmann et al. (2018) for the period between day 17-334 after the mainshock. White lines are interseismic locking contours (0.7, 0.8, 0.9, and 0.95) from the locking model of Li et al. (2015). Small magenta stars represent repeating earthquake sequences (CC $\geq$ 0.95) with at least 2 events (as in Figure 3.6). Other features are plotted as in Figure 3.6.

S3 Supplementary Information for Chapter 4

Text S2: Additional Optional Conditions for Improving Phase Detection.

In order to enhance the performance of DeepPhasePick in correctly detecting P and S phases, some additional conditions to the procedure described in Section 4.2.5 of the manuscript are included. These conditions apply criteria based on the P- and S-phase predicted probabilities in order to resolve some special cases, where discriminating between P and S phases is not trivial. The different conditions can be optionally activated by the user and depend on a few user-defined parameters, as described below.

(1) Resolve between P and S phases predicted close in time.

(i) For each predicted P onset (tP), several threshold time intervals are defined as follows:

$\Delta t_{th} \in [tP\_th\_pre, tP\_th\_pos]$,

$\Delta t_{th\_pre} \in [tP - dt\_PS\_max, tP\_th\_pre]$,

$\Delta t_{th\_pos} \in [tP\_th\_pos, tP + dt\_PS\_max]$.

Here:

$tP\_th\_pre = tP - t(pb\_P\_pre > 0.5)$,

$tP\_th\_pos = tP + t(pb\_P\_pos > 0.5)$,

where $t(pb\_P\_pre > 0.5)$ and $t(pb\_P\_pos > 0.5)$ are the times at which the predicted probabilities before (pb_P_pre) and after (pb_P_pos) tP rise above 0.5, and dt_PS_max [s] is an user-defined parameter.

(ii) If a predicted S onset (tS) is found at $t \in \Delta t_{th}$, we resolve between P and S as follows:

• tP is kept and tS is discarded, if

→ $pb\_P(tP) \geq pb\_S(tS)$,

→ $pb\_S(tS) > pb\_P(tP)$, but no P is found at $t \in \Delta t_{th\_pre}$, and at least one P or S is found at $t \in \Delta t_{th\_pos}$

• tS is kept and tP is discarded, if

→ $pb\_S(tS) > pb\_P(tP)$, at least one P is found at $t \in \Delta t_{th\_pre}$, and at least one P or S is found at $t \in \Delta t_{th\_pos}$

→ $pb\_S(tS) > pb\_P(tP)$, at least one P is found at $t \in \Delta t_{th\_pre}$, and no P or S are found at $t \in \Delta t_{th\_pos}$

• both tS and tP are discarded, if

→ $pb\_S(tS) > pb\_P(tP)$, but no P is found at $t \in \Delta t_{th\_pre}$, and no P or S are found at $t \in \Delta t_{th\_pos}$

(2) Discard predicted S phases for which there is no earlier P or P-S phases predicted.

(i) For each predicted S onset (tS), the following threshold time intervals are defined:

$\Delta t^S_{th\_pre\_1} \in [tS - dt\_PS\_max \times dt\_PS\_frac, tS]$, where dt_PS_frac is an user-defined parameter.

$\Delta t^S_{th\_pre\_2} \in [tP\_pre, tS]$, where $tP\_pre \in \Delta t^S_{th\_pre\_1}$

(ii) tS is discarded, if
$\rightarrow$ no $tP\_pre \in \Delta t^S_{th\_pre\_1}$ is found.
$\rightarrow$ at least one $tP\_pre \in \Delta t^S_{th\_pre\_1}$ is found, but more than one $tS\_pre \in \Delta t^S_{th\_pre\_2}$ are found.
$\rightarrow$ at least one $tP\_pre \in \Delta t^S_{th\_pre\_1}$ is found, one $tS\_pre \in \Delta t^S_{th\_pre\_2}$ is found, but $tS > tS\_pre + |tP\_pre - tS\_pre|$.

(3) Resolve between possible predicted duplicated S phases.
(i) For each predicted S onset tS, the following threshold time intervals are defined:
$tS\_th\_pos = tS + t(pb\_S\_pos > 0.5)$,
$t\hat{S}\_th\_pre = t\hat{S} - t(pb\_\hat{S}\_pre > 0.5)$,
for each subsequent predicted S onset $t\hat{S}$ ($t\hat{S} > tS$). Here $t(pb\_S\_pos > 0.5)$ and $t(pb\_\hat{S}\_pre > 0.5)$ are the times at which the predicted probabilities after tS (pb_S_pos) and before $t\hat{S}$ ($pb\_\hat{S}\_pre$) rise above 0.5, respectively.

(ii) If $|tS\_th\_pos - t\hat{S}\_th\_pre| < dt\_Sdup\_max$,
we only keep the S onset with the maximum predicted probability between $pb\_S(tS)$ and $pb\_\hat{S}(t\hat{S})$. Here dt_Sdup_max [s] is an user-defined parameter.

(4) Resolve between P and S phases predicted close in time, for special cases not handled in (1).
(i) For each predicted P onset (tP) and each subsequent predicted S onset (tS), the following threshold time intervals are defined:

$tP\_th\_pre = tP - t(pb\_P\_pre > 0.5)$,
$tP\_th\_pos = tP + t(pb\_P\_pos > 0.5)$,
$tS\_th\_pre = tS - t(pb\_S\_pre > 0.5)$,
$tS\_th\_pos = tS + t(pb\_S\_pos > 0.5)$.

Here $t(pb\_P\_pre > 0.5)$ and $t(pb\_P\_pos > 0.5)$ are the times at which the predicted probabilities before (pb_P_pre) and after (pb_P_pos) tP rise above 0.5. Similarly, $t(pb\_S\_pre > 0.5)$ and $t(pb\_S\_pos > 0.5)$ are the times at which the predicted probabilities before (pb_S_pre) and after (pb_S_pos) tS rise above 0.5.
We also consider the threshold time differences between successive S-P or P-S predicted phases:
$dt\_SP\_th = |tS\_th\_pos - tP\_th\_pre|$,
$dt\_PS\_th = |tP\_th\_pos - tS\_th\_pre|$

(ii) If $dt\_SP\_th < dt\_SP\_near$ or $dt\_PS\_th < dt\_SP\_near$, we only keep the onset with the maximum predicted probability between $pb\_P(tP)$ and $pb\_S(tS)$. Here dt_SP_near [s] is an user-defined parameter.

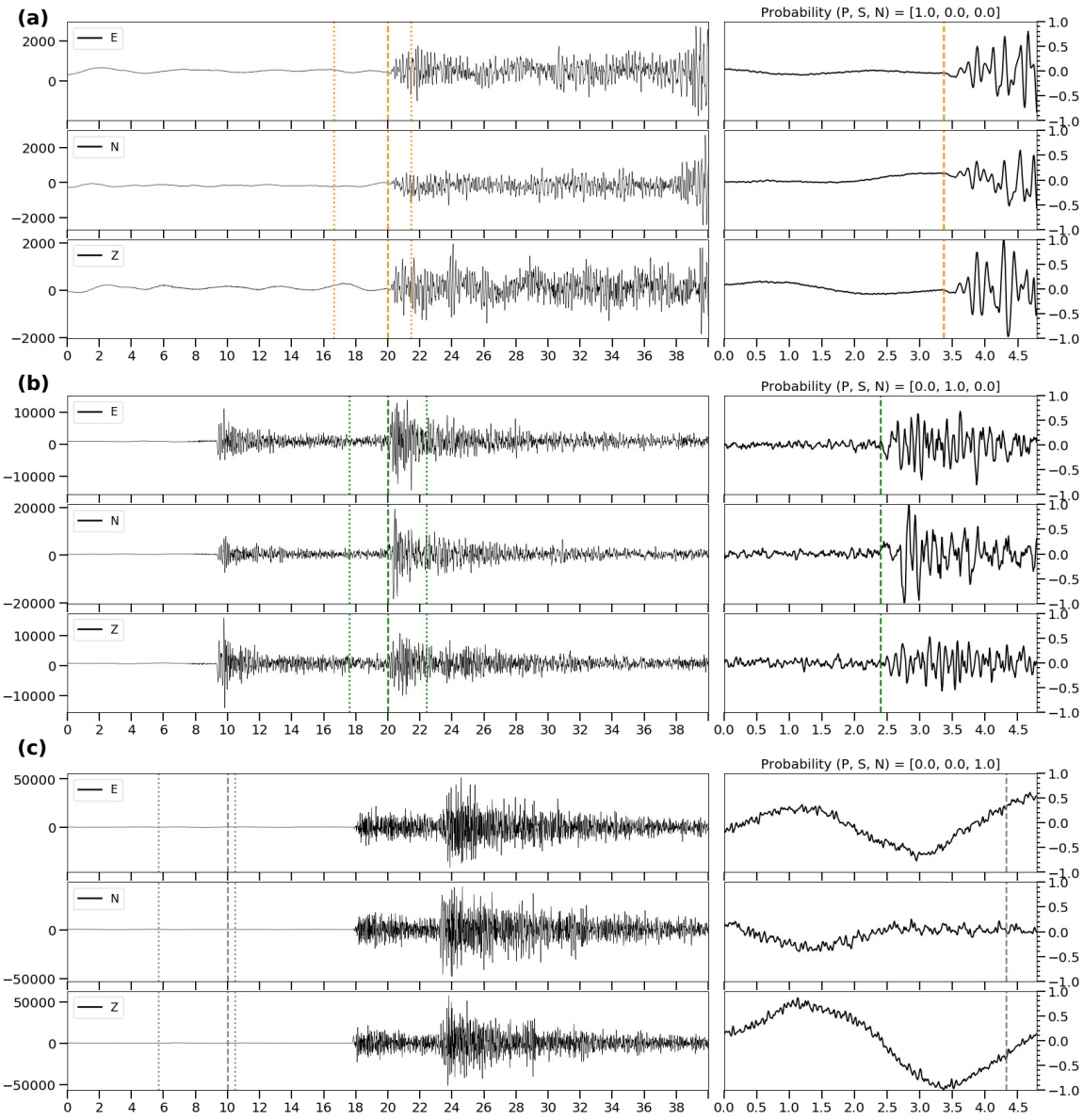

Figure S7: Examples of correctly predicted P- (a), S- (b) and N-class (c) samples in the test set of 4,320 samples. Subplots are plotted the same way as in Figure 4.8.

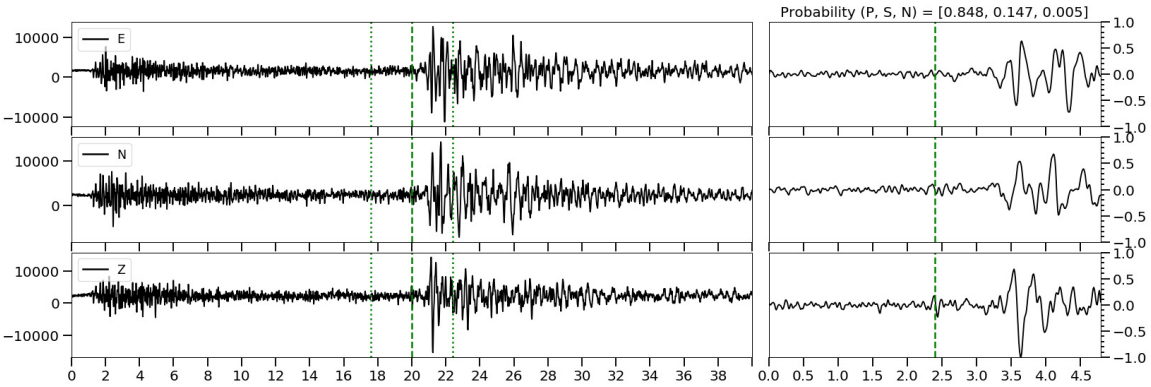

Figure S8: One example of S phase in the test set of 4,320 samples, which was misclassified by the network as P phase. Subplots are plotted the same way as in Figure 4.8.

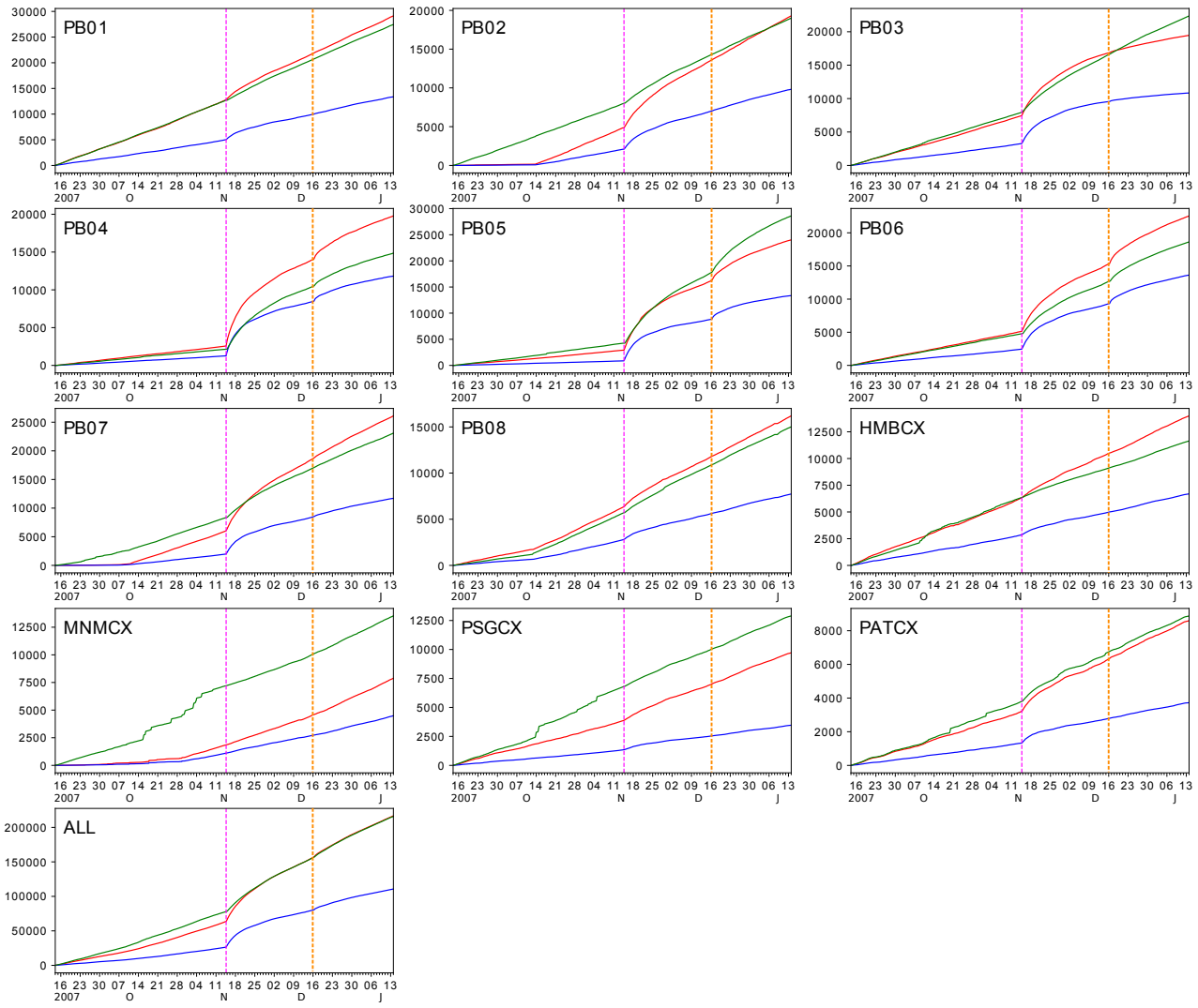

Figure S9: Cumulative number of P (red curve) and S (blue curve) phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$), and triggered detections by the STA/LTA algorithm (green curve) on continuous seismic data recorded by the stations with available data in northern Chile. Last subplot shows the combined results for all the stations in the network. Cumulative curves are plotted for the time interval between two months before and two months after the 2007 M7.7 Tocopilla mainshock (magenta dashed line). An orange dashed line indicates the time of occurrence of a M7.1 aftershock (2007-12-16) ~50 km south of the mainshock epicenter.

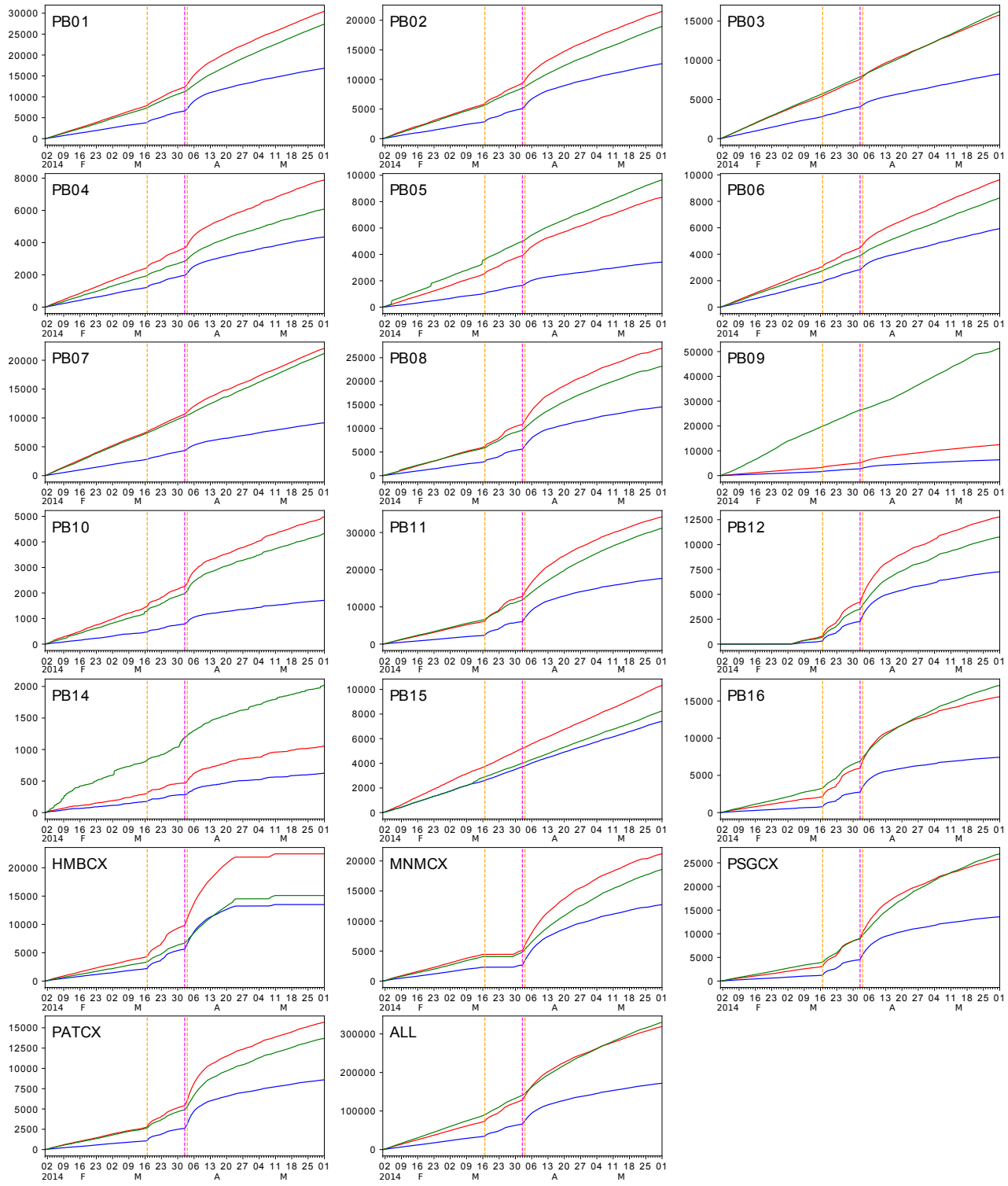

Figure S10: Cumulative number of P (red curve) and S (blue curve) phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$), and triggered detections by the STA/LTA algorithm (green curve) on continuous seismic data recorded by the stations with available data in northern Chile. Last subplot shows the combined results for all the stations in the network. Cumulative curves are plotted for the time interval between two months before and two months after the 2014 M8.1 Iquique mainshock (magenta dashed line). Two additional orange dashed lines indicate the time of occurrence of the M6.7 foreshock (2014-03-16) and the M7.6 aftershock (2014-04-03) in the Iquique sequence.

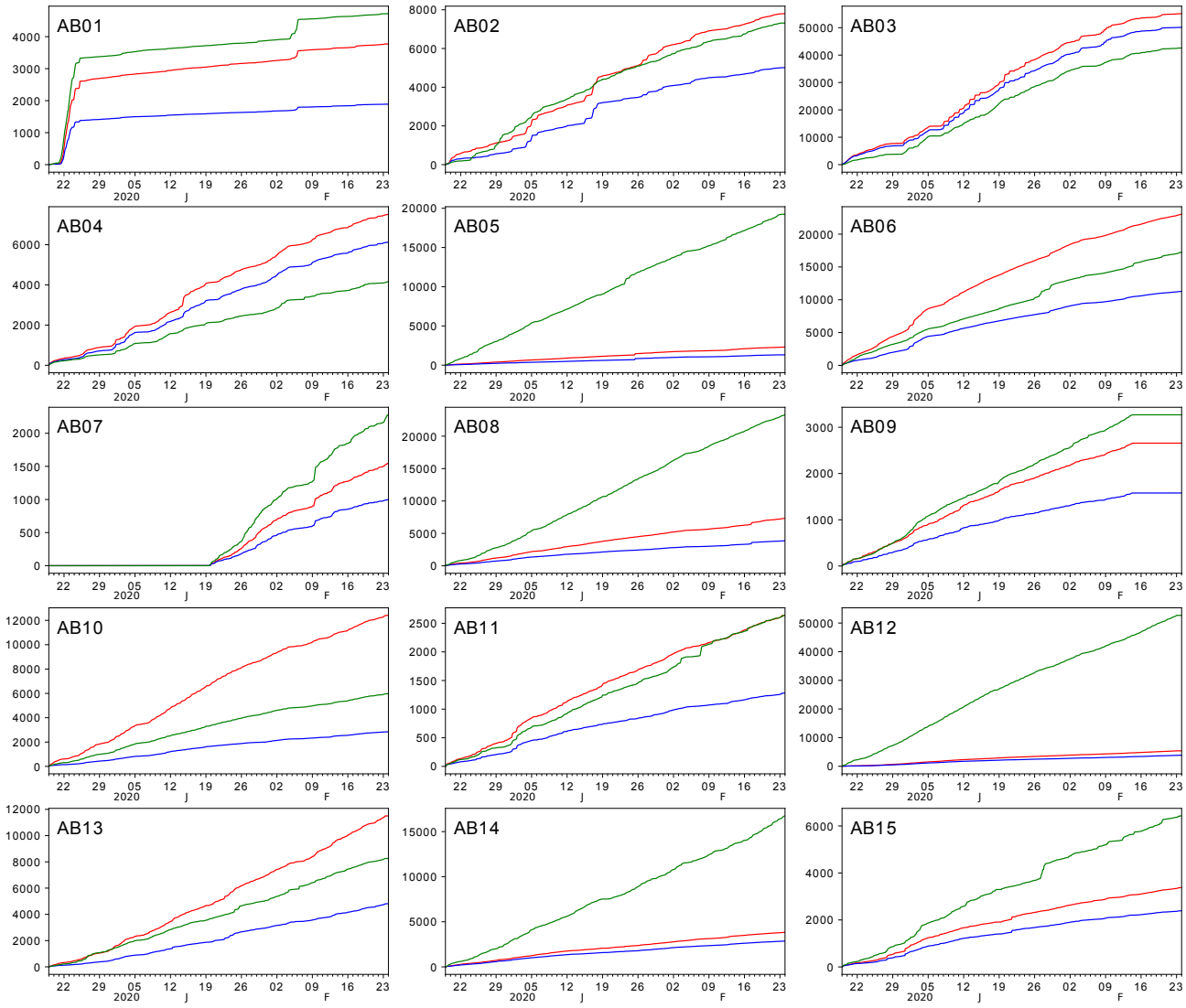

Figure S11: Cumulative number of P (red curve) and S (blue curve) phases predicted by DeepPhasePick ($pb\_P\_th1 = pb\_S\_th1 = 0.98$), and triggered detections by the STA/LTA algorithm (green curve) on continuous seismic data recorded by the stations with available data in Albania. Cumulative curves are plotted for the time interval between 2019-12-19 and 2020-02-24, during the aftershock sequence of the 2019 M6.4 Albania earthquake. First 15 stations in the network are shown here.

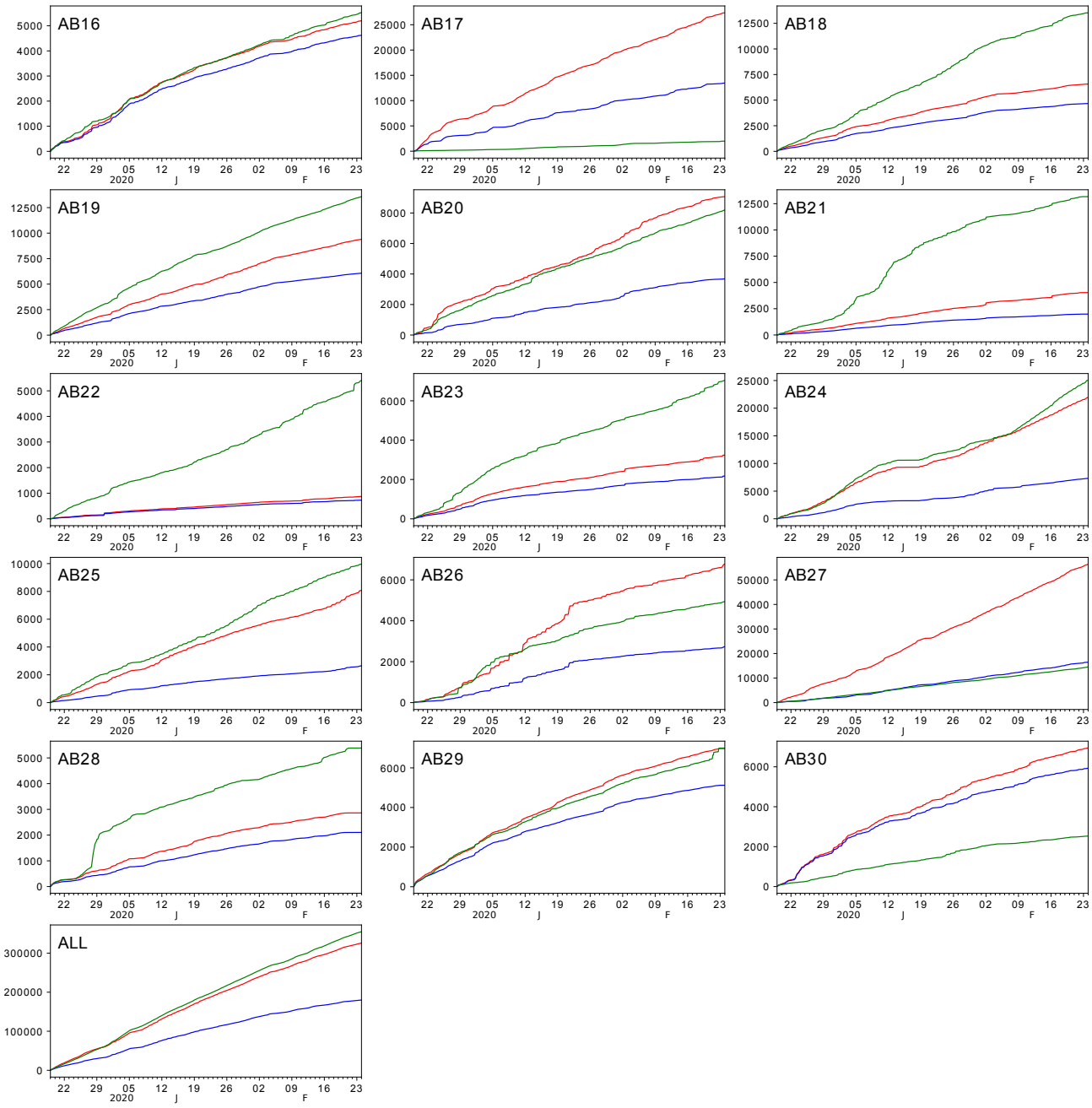

Figure S12: Continuation to Figure S11, showing results from remaining stations in the Albanian network. Last subplot shows the combined results for all the stations in the network.

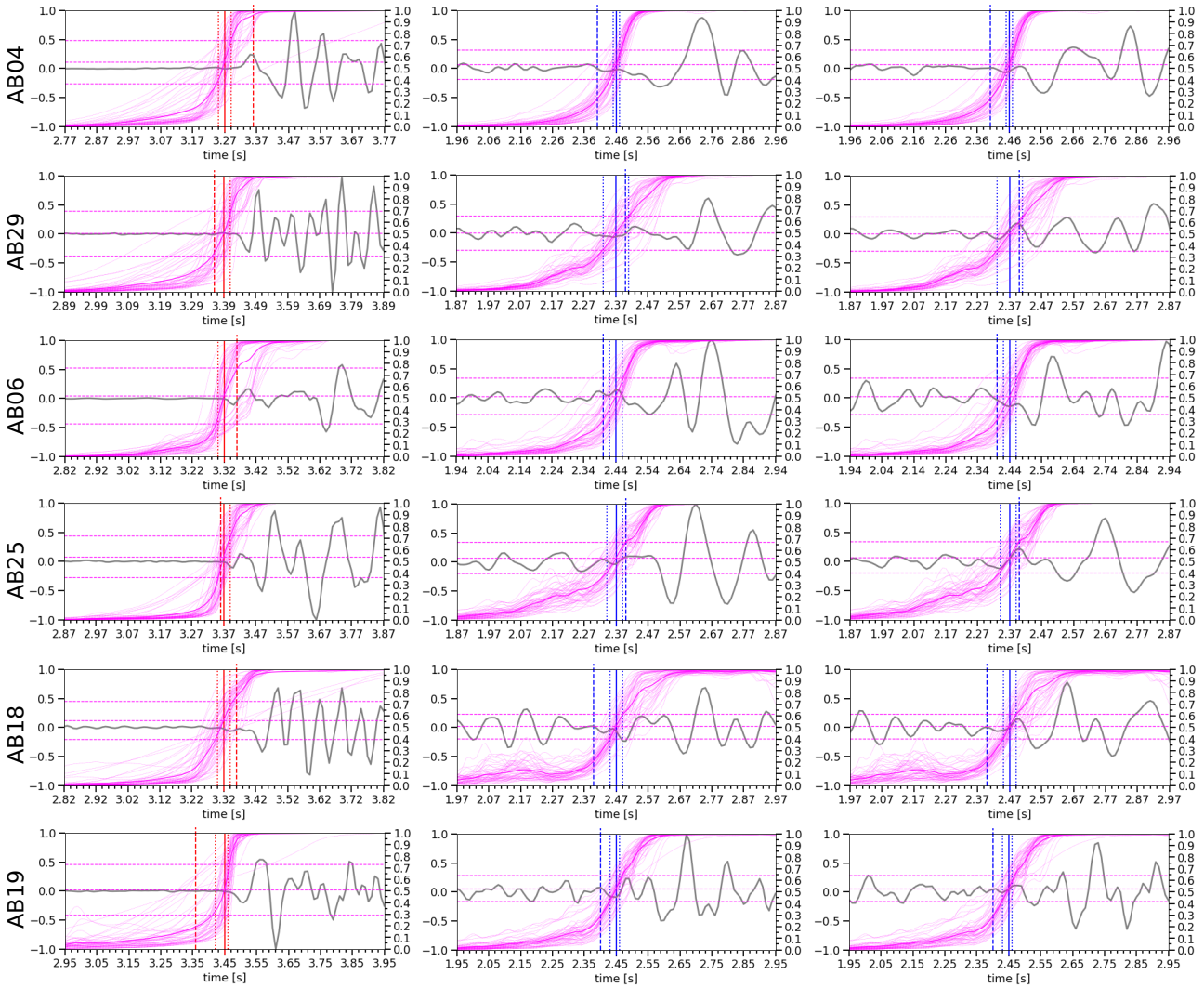

Figure S13: Similar to Figure 4.22, for detected event **e7** in Figure 4.21. Pick statistics for this event are reported in Table S2.

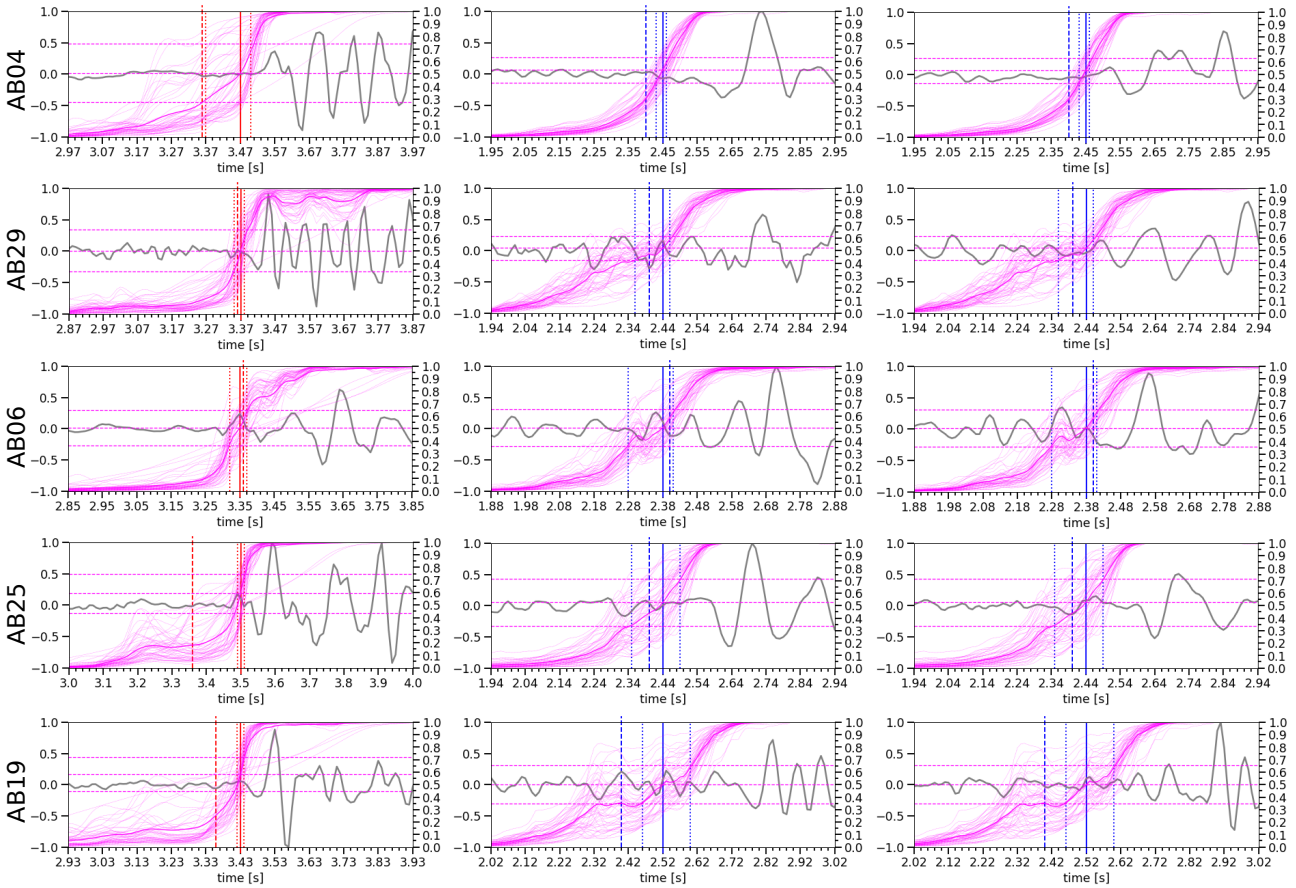

Figure S14: Similar to Figure 4.22, for detected event **e8** in Figure 4.21. Pick statistics for this event are reported in Table S3.

Table S1: Statistics of predicted picks shown in Figure 4.22 for event **e1** in Figure 4.21.

station	phase	*tons_prelim* [s]	*tons_pred* [s]	*tons_err* $(-,+)$ [s]	*pick_class*	*pb*	*pb_std*
AB04	P	3.36	3.41	(0.010, 0.010)	0	0.635	0.207
AB04	S	2.40	2.37	(0.010, 0.010)	0	0.523	0.114
AB29	P	3.36	3.39	(0.010, 0.010)	0	0.611	0.198
AB29	S	2.40	2.43	(0.030, 0.020)	0	0.503	0.111
AB06	P	3.36	3.30	(0.010, 0.050)	0	0.625	0.249
AB06	S	2.40	2.42	(0.040, 0.030)	0	0.536	0.175
AB25	P	3.36	3.38	(0.040, 0.040)	0	0.507	0.253
AB25	S	2.40	2.42	(0.030, 0.020)	0	0.539	0.140
AB18	P	3.36	3.38	(0.020, 0.020)	0	0.580	0.208
AB18	S	2.40	2.37	(0.060, 0.020)	0	0.500	0.117
AB19	P	3.36	3.41	(0.030, 0.020)	0	0.511	0.215
AB19	S	2.40	2.41	(0.030, 0.020)	0	0.513	0.128

Table S2: Statistics of predicted picks shown in Figure S13 for event **e7** in Figure 4.21.

station	phase	*tons_prelim* [s]	*tons_pred* [s]	*tons_err* $(-,+)$ [s]	*pick_class*	*pb*	*pb_std*
AB04	P	3.36	3.27	(0.020, 0.010)	0	0.547	0.171
AB04	S	2.40	2.46	(0.010, 0.010)	0	0.564	0.109
AB29	P	3.36	3.41	(0.020, 0.010)	0	0.591	0.176
AB29	S	2.40	2.38	(0.040, 0.040)	0	0.527	0.175
AB06	P	3.36	3.32	(0.010, 0.020)	0	0.551	0.229
AB06	S	2.40	2.44	(0.020, 0.020)	0	0.529	0.139
AB25	P	3.36	3.38	(0.010, 0.010)	0	0.561	0.152
AB25	S	2.40	2.36	(0.030, 0.020)	0	0.519	0.117
AB18	P	3.36	3.30	(0.020, 0.020)	0	0.556	0.154
AB18	S	2.40	2.48	(0.020, 0.040)	0	0.544	0.148
AB19	P	3.36	3.45	(0.020, 0.010)	0	0.592	0.218
AB19	S	2.40	2.45	(0.020, 0.020)	0	0.539	0.109

Table S3: Statistics of predicted picks shown in Figure S14 for event **e8** in Figure 4.21.

station	phase	*tons_prelim* [s]	*tons_pred* [s]	*tons_err* $(-,+)$ [s]	*pick_class*	*pb*	*pb_std*
AB04	P	3.36	3.49	(0.020, 0.010)	0	0.562	0.150
AB04	S	2.40	2.45	(0.020, 0.020)	0	0.515	0.120
AB29	P	3.36	3.38	(0.010, 0.020)	0	0.604	0.127
AB29	S	2.40	2.44	(0.120, 0.030)	1	0.514	0.117
AB06	P	3.36	3.35	(0.030, 0.020)	0	0.524	0.157
AB06	S	2.40	2.38	(0.100, 0.020)	1	0.521	0.138
AB25	P	3.36	3.49	(0.020, 0.010)	0	0.518	0.172
AB25	S	2.40	2.40	(0.070, 0.080)	1	0.510	0.180
AB19	P	3.36	3.42	(0.020, 0.010)	0	0.521	0.153
AB19	S	2.40	2.57	(0.080, 0.040)	1	0.506	0.145

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., and Citro, C. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems. Software available from *https://www.tensorflow.org/*.

Abers, G. A., Hu, X., and Sykes, L. R. (1995). Source scaling of earthquakes in the shumagin region, Alaska: time-domain inversions of regional waveforms. *Geophysical Journal International*, 123(1):41–58.

Aki, K. (1979). Characterization of barriers on an earthquake fault. *Journal of Geophysical Research*, 84(B11):6140–6148.

Aldersons, F. (2004). Toward a Three-Dimensional Crustal Structure of the Dead Sea region from Local Earthquake Tomography. *Ph.D. thesis. Tel-Aviv University, Israel*.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5):1521–1532.

An, C., Sepúlveda, I., and Liu, P. L. (2014). Tsunami source and its validation of the 2014 Iquique, Chile, earthquake. *Geophysical Research Letters*, 41(11):3988–3994.

Angermann, D., Klotz, J., and Reigber, C. (1999). Space-geodetic estimation of the Nazca-South America Euler vector. *Earth and Planetary Science Letters*, 171(3):329–334.

Asano, Y., Saito, T., Ito, Y., Shiomi, K., Hirose, H., Matsumoto, T., Aoi, S., Hori, S., and Sekiguchi, S. (2011). Spatial distribution and focal mechanisms of aftershocks of the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, 63(7):669–673.

Asch, G., Tilmann, F., Heit, B., and Schurr, B. (2014). HART-PISAGUA Project Chile). *GFZ Data Services. Other/Seismic Network*. https://doi.org/10.14470/8Q7569558037.

Audet, P., Bostock, M. G., Christensen, N. I., and Peacock, S. M. (2009). Seismic evidence for overpressured subducted oceanic crust and megathrust fault sealing. *Nature*, 457(7225):76–78.

Baer, M. and Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4):1437–1445.

Baldi, P., Brunak, S., Frasconi, P., Soda, G., and Pollastri, G. (1999). Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics*, 15(11):937–946.

Barrientos, S. (2018). The Seismic Network of Chile. *Seismological Research Letters*, 89(2A):467–474.

Beck, S. L. and Ruff, L. J. (1989). Great earthquakes and subduction along the Peru trench. *Physics of the Earth and Planetary Interiors*, 57(3-4):199–224.

Bedford, J., Moreno, M., Schurr, B., Bartsch, M., and Oncken, O. (2015). Investigating the final seismic swarm before the Iquique-Pisagua 2014 Mw 8.1 by comparison of continuous GPS and seismic foreshock data. *Geophysical Research Letters*, 42(10):3820–3828.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, 10(2):157–166.

Bennington, N., Thurber, C., Feigl, K. L., and Murray-Moraleda, J. (2011). Aftershock distribution as a constraint on the Geodetic Model of Coseismic Slip for the 2004 Parkfield Earthquake. *Pure and Applied Geophysics*, 168(10):1553–1565.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems 24*, pages 2546–2554.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305.

Bergstra, J., Yamins, D., and Cox, D. (2013a). Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. In *Proceedings of the 12th Python in Science Conference*, pages 13–19.

Bergstra, J., Yamins, D., and Cox, D. (2013b). Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the International Conference on Machine Learning 30*, pages 115–123.

Beroza, G. C. and Zoback, M. D. (1993). Mechanism diversity of the Loma Prieta aftershocks and the mechanics of mainshock-aftershock interaction. *Science*, 259(5092):210–213.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010). ObsPy: A python toolbox for seismology. *Seismological Research Letters*, 81(3):530–533.

Bloch, W., Kummerow, J., Salazar, P., Wigger, P., and Shapiro, S. A. (2014). High-resolution image of the North Chilean subduction zone: Seismicity, reflectivity and fluids. *Geophysical Journal International*, 197(3):1744–1749.

Bloch, W., Schurr, B., Kummerow, J., Salazar, P., and Shapiro, S. A. (2018). From Slab Coupling to Slab Pull: Stress Segmentation in the Subducting Nazca Plate. *Geophysical Research Letters*, 45(11):5407–5416.

Bormann, P. and Dewey, J. W. (2014). The new IASPEI standards for determining magnitudes from digital data and their relation to classical magnitudes. In Bormann, P., editor, *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pages 1–44. Deutsches GeoForschungsZentrum GFZ, Potsdam.

Bürgmann, R. (2014). Earth science: Warning signs of the Iquique earthquake. *Nature*, 512(7514):258–259.

Bürgmann, R., Kogan, M. G., Steblov, G. M., Hilley, G., Levin, V. E., and Apel, E. (2005). Interseismic coupling and asperity distribution along the Kamchatka subduction zone. *Journal of Geophysical Research: Solid Earth*, 110(7):1–17.

Cesca, S., Grigoli, F., Heimann, S., Dahm, T., Kriegerowski, M., Sobiesiak, M., Tassara, C., and Olcay, M. (2016). The Mw 8.1 2014 Iquique, Chile, seismic sequence: A tale of foreshocks and aftershocks. *Geophysical Journal International*, 204(3):1766–1780.

Cesca, S., Sobiesiak, M., Tassara, A., Olcay, M., Günther, E., Mikulla, S., and Dahm, T. (2009). The Iquique Local Network and PicArray). *GFZ Data Services. Other/Seismic Network*. https://doi.org/10.14470/vd070092.

Chen, J. and Chaudhari, N. (2005). Protein Secondary Structure Prediction with bidirectional LSTM networks. In *International Joint Conference on Neural Networks: Post-Conference Workshop on Computational Intelligence Approaches for the Analysis of Bio-data (CI-BIO)*.

Chlieh, M., Avouac, J. P., Sieh, K., Natawidjaja, D. H., and Galetzka, J. (2008). Heterogeneous coupling of the Sumatran megathrust constrained by geodetic and paleogeodetic measurements. *Journal of Geophysical Research: Solid Earth*, 113(5):1–31.

Chollet, F. (2015). Keras. *https://github.com/fchollet/keras*.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, volume 2017-January, pages 1800–1807.

Cichowicz, A. (1993). An automatic S-phase picker. *Bulletin - Seismological Society of America*, 83(1):180–189.

Collings, R., Lange, D., Rietbrock, A., Tilmann, F., Natawidjaja, D., Suwargadi, B., Miller, M., and Saul, J. (2012). Structure and seismogenic properties of the Mentawai segment of the Sumatra subduction zone revealed by local earthquake traveltime tomography. *Journal of Geophysical Research: Solid Earth*, 117:B01312.

Comte, D. and Pardo, M. (1991). Reappraisal of great historical earthquakes in the northern Chile and southern Peru seismic gaps. *Natural Hazards*, 4(1):23–44.

Contreras-Reyes, E., Flueh, E. R., and Grevemeyer, I. (2010). Tectonic control on sediment accretion and subduction off south central Chile: Implications for coseismic rupture processes of the 1960 and 2010 megathrust earthquakes. *Tectonics*, 29(6):1–27.

Contreras-Reyes, E., Jara, J., Grevemeyer, I., Ruiz, S., and Carrizo, D. (2012). Abrupt change in the dip of the subducting plate beneath north Chile. *Nature Geoscience*, 5(5):342–345.

Contreras-Reyes, E., Maksymowicz, A., Lange, D., Grevemeyer, I., Muñoz-Linford, P., and Moscoso, E. (2017). On the relationship between structure, morphology and large coseismic slip: A case study of the Mw 8.8 Maule, Chile 2010 earthquake. *Earth and Planetary Science Letters*, 478:27–39.

Das, S. and Henry, C. (2003). Spatial relation between main earthquake slip and its aftershock distribution. *Reviews of Geophysics*, 41(3):1013.

DeShon, H. R., Schwartz, S. Y., Bilek, S. L., Dorman, L. M., Gonzalez, V., Protti, J. M., Flueh, E. R., and Dixon, T. H. (2003). Seismogenic zone structure of the southern Middle America Trench, Costa Rica. *Journal of Geophysical Research: Solid Earth*, 108(B10):2491.

Dessa, J. X., Klingelhoefer, F., Graindorge, D., André, C., Permana, H., Gutscher, M. A., Chauhan, A., and Singh, S. C. (2009). Megathrust earthquakes can nucleate in the forearc mantle: Evidence from the 2004 Sumatra event. *Geology*, 37(7):659–662.

Di Stefano, R., Aldersons, F., Kissling, E., Baccheschi, P., Chiarabba, C., and Giardini, D. (2006). Automatic seismic phase picking and consistent observation error assessment: Application to the Italian seismicity. *Geophysical Journal International*, 165(1):121–134.

Diehl, T., Deichmann, N., Kissling, E., and Husen, S. (2009). Automatic S-wave picker for local earthquake tomography. *Bulletin of the Seismological Society of America*, 99(3):1906–1920.

Dokht, R. M., Kao, H., Visser, R., and Smith, B. (2019). Seismic event and phase detection using time-frequency representation and convolutional neural networks. *Seismological Research Letters*, 90(2A):481–490.

Douglas, A. (1997). Bandpass filtering to reduce noise on seismograms: Is there a better way? *Bulletin of the Seismological Society of America*, 87(3):770–777.

Draelos, T. J., Ballard, S., Young, C. J., and Brogan, R. (2015). A new method for producing automated seismic bulletins: Probabilistic event detection, association, and location. *Bulletin of the Seismological Society of America*, 105(5):2453–2467.

Duputel, Z., Jiang, J., Jolivet, R., Simons, M., Rivera, L., Ampuero, J. P., Riel, B., Owen, S. E., Moore, A. W., Samsonov, S. V., Ortega Culaciati, F., and Minson, S. E. (2015). The Iquique earthquake sequence of April 2014: Bayesian modeling accounting for prediction uncertainty. *Geophysical Research Letters*, 42(19):7949–7957.

Earle, P. S. and Shearer, P. M. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin - Seismological Society of America*, 84(2):366–376.

Eberhart-Phillips, D. (1993). Local earthquake tomography: earthquake source regions. In Iyer, H. and Hirahara, K., editors, *Seismic Tomography: Theory and Practice*, pages 614–643. Chapman and Hall, London.

Edwards, J. H., Kluesner, J. W., Silver, E. A., Brodsky, E. E., Brothers, D. S., Bangs, N. L., Kirkpatrick, J. D., Wood, R., and Okamoto, K. (2018). Corrugated megathrust revealed offshore from Costa Rica. *Nature Geoscience*, 11(3):197–202.

Engelder, J. T. (1975). Correction [to "Microscopic wear grooves on slickensides: Indicators of paleoseismicity"]. *Journal of Geophysical Research*, 80(2):286–288.

Evans, J., Eberhart-Phillips, D., and Thurber, C. (1994). User's manual for simulps12 for imaging vp and vp/vs: A derivative of the "Thurber" tomographic inversion simul3 for local earthquakes and explosions. In *Tech. rep., Open File Report 94-431*, pages 94–431. U.S. Geological Survey, London.

Folesky, J., Kummerow, J., and Shapiro, S. A. (2018). Patterns of Rupture Directivity of Subduction Zone Earthquakes in Northern Chile. *Journal of Geophysical Research: Solid Earth*, 123(12):10,785–10,796.

Fuenzalida, A., Schurr, B., Lancieri, M., Sobiesiak, M., and Madariaga, R. (2013). High-resolution relocation and mechanism of aftershocks of the 2007 Tocopilla (Chile) earthquake. *Geophysical Journal International*, 194(2):1216–1228.

Gal, Y. and Ghahramani, Z. (2016a). A theoretically grounded application of dropout in recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 1027–1035.

Gal, Y. and Ghahramani, Z. (2016b). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *33rd International Conference on Machine Learning, ICML 2016*, volume 3, pages 1651–1660.

Geersen, J., Ranero, C. R., Barckhausen, U., and Reichert, C. (2015). Subducting seamounts control interplate coupling and seismic rupture in the 2014 Iquique earthquake area. *Nature Communications*, 6, 8267.

Geersen, J., Ranero, C. R., Klaucke, I., Behrmann, J. H., Kopp, H., Tréhu, A. M., Contreras-Reyes, E., Barckhausen, U., and Reichert, C. (2018a). Active Tectonics of the North Chilean Marine Forearc and Adjacent Oceanic Nazca Plate. *Tectonics*, 37(11):4194–4211.

Geersen, J., Ranero, C. R., Kopp, H., Behrmann, J. H., Lange, D., Klaucke, I., Barrientos, S., Diaz-Naveas, J., Barckhausen, U., and Reichert, C. (2018b). Does permanent extensional

deformation in lower forearc slopes indicate shallow plate-boundary rupture? *Earth and Planetary Science Letters*, 489:17–27.

Geiger, L. (1910). Herdbestimmung bei Erdbeben aus den Ankunftszeiten. *K. Gesellschaft d. Wiss. Göttingen*, 4:331–349.

GEOFON Data Centre (2013). GEOFON Data Centre (1993). GEOFON seismic network. *Deutsches GeoForschungsZentrum GFZ. Other/Seismic Network.* https://doi.org/10.14470/TR560404.

GFZ German Research Centre for Geosciences, and Institut des Sciences de l'Univers-Centre National de la Recherche CNRS-INSU (2006). IPOC Seismic Network. Integrated Plate boundary Observatory Chile - IPOC. *Other/Seismic Network.* https://doi.org/10.14470/PK615318.

Ghosh, A., Vidale, J. E., Sweet, J. R., Creager, K. C., Wech, A. G., Houston, H., and Brodsky, E. E. (2010). Rapid, continuous streaking of tremor in Cascadia. *Geochemistry, Geophysics, Geosystems*, 11:Q12010.

Gibbons, S. J. and Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1):149–166.

González, G., Salazar, P., Loveless, J. P., Allmendinger, R. W., Aron, F., and Shrivastava, M. (2015). Upper plate reverse fault reactivation and the unclamping of the megathrust during the 2014 northern Chile earthquake sequence. *Geology*, 43(8):671–674.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning (Vol. 1).* MIT Press, Cambridge, Massachusetts.

Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18:602–610.

Gusman, A. R., Murotani, S., Satake, K., Heidarzadeh, M., Gunawan, E., Watada, S., and Schurr, B. (2015). Fault slip distribution of the 2014 Iquique, Chile, earthquake estimated from ocean-wide tsunami waveforms and GPS data. *Geophysical Research Letters*, 42(4):1053–1060.

Haberland, C., Rietbrock, A., Asch, G., and Chong, G. (1996). The ANCORP Seismic Network. *GFZ Data Services. Other/Seismic Network.* https://doi.org/10.14470/MR6441682066.

Harris, D. B. (1991). A waveform correlation method for identifying quarry explosions. *Bulletin - Seismological Society of America*, 81(6):2395–2418.

Harris, D. B. (1997). Waveform correlation methods for identifying populations of calibration events. In *Proceedings of the 19th Annual Seismic Research Symposium on Monitoring a CTBT*, pages 604–614.

Harris, D. B. (2001). Subspace Techniques for Detecting Repeating Events, poster, SSA 2001 Annual Meeting, San Francisco, CA, 18-20 April 2001. *abstract in Seism. Res. Lett.*, 72(2):245.

Harris, D. B. and Dodge, D. A. (2011). An autonomous system for grouping events in a developing aftershock sequence. *Bulletin of the Seismological Society of America*, 101(2):763–774.

Harris, D. B. and Paik, T. (2006). Subspace Detectors: Efficient Implementation. *Department of Energy, United States*.

Harris, R. A. and Simpson, R. W. (1998). Suppression of large earthquakes by stress shadows: A comparison of Coulomb and rate-and-state failure. *Journal of Geophysical Research: Solid Earth*, 103(10):24439–24451.

Hayes, G. P., Earle, P. S., Benz, H. M., Wald, D. J., and Briggs, R. W. (2011). 88 Hours: The U.S. Geological survey national earthquake information center response to the 11 march 2011 Mw 9.0 tohoku earthquake. *Seismological Research Letters*, 82(4):481–493.

Hayes, G. P., Herman, M. W., Barnhart, W. D., Furlong, K. P., Riquelme, S., Benz, H. M., Bergman, E., Barrientos, S., Earle, P. S., and Samsonov, S. (2014). Continuing megathrust earthquake potential in Chile after the 2014 Iquique earthquake. *Nature*, 512(7514):295–298.

Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., and Smoczyk, G. M. (2018). Slab2, a comprehensive subduction zone geometry model. *Science*, 362(6410):58–61.

Hergert, T. and Heidbach, O. (2006). New insights into the mechanism of postseismic stress relaxation exemplified by the 23 June 2001 Mw = 8.4 earthquake in southern Peru. *Geophysical Research Letters*, 33(2):3–6.

Herman, M. W., Furlong, K. P., Hayes, G. P., and Benz, H. M. (2016). Foreshock triggering of the 1 April 2014 Mw 8.2 Iquique, Chile, earthquake. *Earth and Planetary Science Letters*, 447:119–129.

Herron, E. M., Cande, S. C., and Hall, B. R. (1981). An active spreading center collides with a subduction zone: A geophysical survey of the Chile Margin triple junction. *Memoir of the Geological Society of America*.

Hetland, E. A. and Simons, M. (2010). Post-seismic and interseismic fault creep II: transient creep and interseismic stress shadows on megathrusts. *Geophysical Journal International*, 181(1):99–112.

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B. (2012). Deep neural networks for acoustic

modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97.

Hinz, T., Navarro-Guerrero, N., Magg, S., and Wermter, S. (2018). Speeding up the Hyperparameter Optimization of Deep Convolutional Neural Networks. *International Journal of Computational Intelligence and Applications*, 17(2):1850008.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. *Unpublished doctoral dissertation, Institut für Informatik, Technische Universität, Munchen.*

Hochreiter, S., Bengio, Y., Frasconi, P., and Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In Kremer, S. C. and Kolen, J. F., editors, *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hoffmann, F., Metzger, S., Moreno, M., Deng, Z., Sippl, C., Ortega-Culaciati, F., and Oncken, O. (2018). Characterizing Afterslip and Ground Displacement Rate Increase Following the 2014 Iquique-Pisagua Mw 8.1 Earthquake, Northern Chile. *Journal of Geophysical Research: Solid Earth*, 123(5):4171–4192.

Hsu, Y. J., Simons, M., Avouac, J. P., Galeteka, J., Sieh, K., Chlieh, M., Natawidjaja, D., Prawirodirdjo, L., and Bock, Y. (2006). Frictional afterslip following the 2005 Nias-Simeulue earthquake, Sumatra. *Science*, 312(5782):1921–1926.

Hu, Y., Wang, K., He, J., Klotz, J., and Khazaradze, G. (2004). Three-dimensional viscoelastic finite element model for postseismic deformation of the great 1960 Chile earthquake. *Journal of Geophysical Research: Solid Earth*, 109(12):1–14.

Huang, H., Xu, W., Meng, L., Bürgmann, R., and Baez, J. C. (2017). Early aftershocks and afterslip surrounding the 2015 Mw 8.4 Illapel rupture. *Earth and Planetary Science Letters*, 457:282–291.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.

Hughes, K. L., Masterlark, T., and Mooney, W. D. (2010). Poroelastic stress-triggering of the 2005 M8.7 Nias earthquake by the 2004 M9.2 Sumatra-Andaman earthquake. *Earth and Planetary Science Letters*, 293(3-4):289–299.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3):90–95.

Husen, S., Kissling, E., Deichmann, N., Wiemer, S., Giardini, D., and Baer, M. (2003). Probabilistic earthquake location in complex three-dimensional velocity models: Application to Switzerland. *Journal of Geophysical Research: Solid Earth*, 108:2077.

Husen, S., Kissling, E., Flueh, E., and Asch, G. (1999). Accurate hypocentre determination in the seismogenic zone of the subducting Nazca Plate in northern Chile using a combined on-/offshore network. *Geophysical Journal International*, 138(3):687–701.

Igarashi, T., Matsuzawa, T., and Hasegawa, A. (2003). Repeating earthquakes and interplate aseismic slip in the northeastern Japan subduction zone. *Journal of Geophysical Research: Solid Earth*, 108(B5):1–9.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167*.

Johnson, C. E., Bittenbinder, A., Bogaert, B., Dietz, L., and Kohler, W. (1995). Earthworm: A flexible approach to seismic network processing. *Iris newsletter*, 14(2):1–4.

Kaggle challenge (2014). Higgs boson machine learning challenge. Kaggle. https://www.kaggle.com/c/higgs-boson.

Kanamori, H. (1977). The energy release in great earthquakes. *Journal of Geophysical Research*, 82(20):2981–2987.

Kato, A. and Igarashi, T. (2012). Regional extent of the large coseismic slip zone of the 2011 Mw 9.0 Tohoku-Oki earthquake delineated by on-fault aftershocks. *Geophysical Research Letters*, 39(15):2–7.

Kato, A., Miyatake, T., and Hirata, N. (2010). Asperity and barriers of the 2004 Mid-Niigata prefecture earthquake revealed by highly dense seismic observations. *Bulletin of the Seismological Society of America*, 100(1):298–306.

Kato, A. and Nakagawa, S. (2014). Multiple slow-slip events during a foreshock sequence of the 2014 Iquique, Chile Mw 8.1 earthquake. *Geophysical Research Letters*, 41(15):5420–5427.

Kingma, D. P. and Ba, J. L. (2014). Adam: A method for stochastic optimization. *ArXiv:1412.6980 [Cs]. Retrieved from* http://arxiv.org/abs/1412.6980.

Kissling, E. (1988). Geotomography with local earthquake data. *Reviews of Geophysics*, 26(4):659–698.

Kissling, E., Ellsworth, W. L., Eberhart-Phillips, D., and Kradolfer, U. (1994). Initial reference models in local earthquake tomography. *Journal of Geophysical Research*, 99(B10):19635–19646.

Kissling, E., Kradolfer, U., and Maurer, H. (1995). *VELEST User's Guide: Short Introduction*. Institute of Geophysics and Swiss Seismological Service, ETH, Zurich.

Konca, A. O., Avouac, J. P., Sladen, A., Meltzner, A. J., Sieh, K., Fang, P., Li, Z., Galetzka, J., Genrich, J., Chlieh, M., Natawidjaja, D. H., Bock, Y., Fielding, E. J., Ji, C., and Helmberger, D. V. (2008). Partial rupture of a locked patch of the Sumatra megathrust during the 2007 earthquake sequence. *Nature*, 456(7222):631–635.

Kriegerowski, M., Petersen, G. M., Vasyura-Bathke, H., and Ohrnberger, M. (2019). A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. *Seismological Research Letters*, 90(2A):510–516.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems 25*, pages 1097–1105.

Lahr, J. C. (1999). HYPOELLIPSE: A Computer Program for Determining Local Earthquake Hypocentral Parameters, Magnitude, and First-Motion Pattern. *U.S. Geological Survey Professional Paper*.

Lamb, S. (2006). Shear stresses on megathrusts: Implications for mountain building behind subduction zones. *Journal of Geophysical Research*, 111(B7):B07401.

Lange, D., Geersen, J., Barrientos, S., Moreno, M., Grevemeyer, I., Contreras-Reyes, E., and Kopp, H. (2016). Aftershock seismicity and tectonic setting of the 2015 September 16 Mw 8.3 Illapel earthquake, Central Chile. *Geophysical Journal International*, 206(2):1424–1430.

Lange, D., Tilmann, F., Barrientos, S. E., Contreras-Reyes, E., Methe, P., Moreno, M., Heit, B., Agurto, H., Bernard, P., Vilotte, J. P., and Beck, S. (2012). Aftershock seismicity of the 27 February 2010 Mw 8.8 Maule earthquake rupture zone. *Earth and Planetary Science Letters*, 317-318:413–425.

Lay, T. and Bilek, S. (2007). Anomalous earthquake ruptures at shallow depths on subduction zone megathrusts. In Dixon, T. and Moore, J., editors, *The Seismogenic Zone of Subduction Thrust Faults*, pages 476–511. Colombia Press, New York.

Lay, T., Kanamori, H., Ammon, C. J., Koper, K. D., Hutko, A. R., Ye, L., Yue, H., and Rushing, T. M. (2012). Depth-varying rupture properties of subduction zone megathrust faults. *Journal of Geophysical Research: Solid Earth*, 117(4):1–21.

Lay, T. and Wallace, T. C. (1995). *Modern Global Seismology*. San Diego: Academic Press.

Lay, T., Yue, H., Brodsky, E. E., and An, C. (2014). The 1 April 2014 Iquique, Chile, Mw 8.1 earthquake rupture sequence. *Geophysical Research Letters*, 41(11):3818–3825.

LeBras, R., Swanger, H., Sereno, T., Beall, G., Jenkins, R., and Nagy, W. (1994). *Global association. Design Document and User's Manual*. Science Applications International Corporation, California.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., and Hubbard, W. (1990). Handwritten digit recognition with a backpropagation network. In *Proceedings of Advances in Neural Information Processing Systems 2*, pages 396–404.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323.

Lee, W. H. K. and Lahr, J. C. (1975). Hypo71 (revised): A computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes. *US Geol. Surv. Open File Report*, 75-311:1–113.

León-Ríos, S., Ruiz, S., Maksymowicz, A., Leyton, F., Fuenzalida, A., and Madariaga, R. (2016). Diversity of the 2014 Iquique's foreshocks and aftershocks: clues about the complex rupture process of a Mw 8.1 earthquake. *Journal of Seismology*, 20(4):1059–1073.

Li, S., Moreno, M., Bedford, J., Rosenau, M., and Oncken, O. (2015). Revisiting viscoelastic effects on interseismic deformation and locking degree: A case study of the Peru-North Chile subduction zone. *Journal of Geophysical Research: Solid Earth*, 120(6):4522–4538.

Lieser, K., Grevemeyer, I., Lange, D., Flueh, E., Tilmann, F., and Contreras-Reyes, E. (2014). Splay fault activity revealed by aftershocks of the 2010 Mw 8.8 Maule earthquake, central chile. *Geology*, 42(9):823–826.

Lin, J. and Stein, R. S. (2004). Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults. *Journal of Geophysical Research: Solid Earth*, 109(B2):1–19.

Liu, C., Zheng, Y., Wang, R., and Xiong, X. (2015). Kinematic rupture process of the 2014 Chile Mw 8.1 earthquake constrained by strong-motion, GPS static offsets and teleseismic data. *Geophysical Journal International*, 202(2):1137–1145.

Lomax, A., Michelini, A., and Curtis, A. (2009). *Encyclopedia of Complexity and Systems Science*. Springer, New York.

Lomax, A., Virieux, J., Volant, P., and Berge-Thierry, C. (2000). Probabilistic Earthquake Location in 3D and Layered Models. In Thurber, C. and Rabinowitz, N., editors, *Advances in Seismic Event Location. Modern Approaches in Geophysics, vol 18*. Springer, Dordrecht.

MacEira, M., Rowe, C. A., Beroza, G., and Anderson, D. (2010). Identification of low-frequency earthquakes in non-volcanic tremor using the subspace detector method. *Geophysical Research Letters*, 37(6):1–5.

Maksymowicz, A., Ruiz, J., Vera, E., Contreras-Reyes, E., Ruiz, S., Arraigada, C., Bonvalot, S., and Bascuñan, S. (2018). Heterogeneous structure of the Northern Chile marine forearc and

its implications for megathrust earthquakes. *Geophysical Journal International*, 215(2):1080–1097.

Marone, C. J., Scholtz, C. H., and Bilham, R. (1991). On the mechanics of earthquake afterslip. *Journal of Geophysical Research*, 96(B5):8441–8452.

McBrearty, I. W., Delorey, A. A., and Johnson, P. A. (2019). Pairwise association of seismic arrivals with convolutional neural networks. *Seismological Research Letters*, 90(2A):503–509.

Mendoza, C. and Hartzell, S. H. (1988). Aftershock patterns and main shock faulting. *Bulletin - Seismological Society of America*, 78(4):1438–1449.

Meng, L., Huang, H., Bürgmann, R., Ampuero, J. P., and Strader, A. (2015). Dual megathrust slip behaviors of the 2014 Iquique earthquake sequence. *Earth and Planetary Science Letters*, 411:177–187.

Métois, M., Socquet, A., and Vigny, C. (2012). Interseismic coupling, segmentation and mechanical behavior of the central Chile subduction zone. *Journal of Geophysical Research: Solid Earth*, 117(3):1–16.

Métois, M., Socquet, A., Vigny, C., Carrizo, D., Peyrat, S., Delorme, A., Maureira, E., Valderas-Bermejo, M. C., and Ortega, I. (2013). Revisiting the North Chile seismic gap segmentation using GPS-derived interseismic coupling. *Geophysical Journal International*, 194(3):1283–1294.

Miyazaki, S., Segall, P., Fukuda, J., and Kato, T. (2004). Space time distribution of afterslip following the 2003 Tokachi-oki earthquake: Implications for variations in fault zone frictional properties. *Geophysical Research Letters*, 31(6):4–7.

Moreno, M., Melnick, D., Rosenau, M., Bolte, J., Klotz, J., Echtler, H., Baez, J., Bataille, K., Chen, J., Bevis, M., Hase, H., and Oncken, O. (2011). Heterogeneous plate locking in the South-Central Chile subduction zone: Building up the next great earthquake. *Earth and Planetary Science Letters*, 305(3-4):413–424.

Moreno, M., Rosenau, M., and Oncken, O. (2010). 2010 Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone. *Nature*, 467:198–202.

Moseley, B., Nissen-Meyer, T., and Markham, A. (2020). Deep learning for fast simulation of seismic waves in complex media. *Solid Earth*, 11(4):1527–1549.

Nadeau, R. M. and Johnson, L. R. (1998). Seismological studies at Parkfield VI: moment release rates and estimates of source parameters for small repeating earthquakes. *Bulletin of the Seismological Society of America*, 88(3):790–814.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve Restricted Boltzmann machines. In *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pages 807–814.

Oleskevich, D. A., Hyndman, R. D., and Wang, K. (1999). The updip and downdip limits to great subduction earthquakes: Thermal and structural models of Cascadia, south Alaska, SW Japan, and Chile. *Journal of Geophysical Research: Solid Earth*, 104(B7):14965–14991.

Pacheco, J. F. and Sykes, L. R. (1992). Seismic moment catalog of large shallow earthquakes, 1900 to 1989. *Bulletin - Seismological Society of America*, 82(3):1306–1349.

Palo, M., Tilmann, F., Krüger, F., Ehlert, L., and Lange, D. (2014). High-frequency seismic radiation from Maule earthquake (Mw 8.8, 2010 February 27) inferred from high-resolution backprojection analysis. *Geophysical Journal International*, 199(2):1058–1077.

Panakkat, A. and Adeli, H. (2009). Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. *Computer-Aided Civil and Infrastructure Engineering*, 24(4):280–292.

Pavlis, G. L. and Booker, J. R. (1980). The mixed discrete- continuous inverse problem: application to the simultaneous determination of earthquake hypocenters and velocity structure. *Journal of Geophysical Research*, 85(B9):4801–4810.

Peng, Z. and Zhao, P. (2009). Migration of early aftershocks following the 2004 Parkfield earthquake. *Nature Geoscience*, 2(12):877–881.

Perfettini, H. and Avouac, J. P. (2007). Modeling aftership and aftershocks following the 1992 Landers earthquake. *Journal of Geophysical Research: Solid Earth*, 112(7):1–19.

Perfettini, H., Avouac, J. P., Tavera, H., Kositsky, A., Nocquet, J. M., Bondoux, F., Chlieh, M., Sladen, A., Audin, L., Farber, D. L., and Soler, P. (2010). Seismic and aseismic slip on the Central Peru megathrust. *Nature*, 465(7294):78–81.

Perol, T., Gharbi, M., and Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2):e1700578.

Petit, J. P. (1987). Criteria for the sense of movement on fault surfaces in brittle rocks. *Journal of Structural Geology*, 9(5-6):597–608.

Ramos, V. A. (1999). Plate tectonic setting of the Andean Cordillera. *Episodes*, 22(3):183–190.

Ranero, C. R., von Huene, R., Weinrebe, W., and Reichert, C. (2006). Tectonic Processes along the Chile Convergent Margin. In Oncken, O., Chong, G., Franz, G., Giese, P., Götze, H.-J., Ramos, V., Strecker, M., and Wigger, P., editors, *The Andes – Active Subduction Orogeny. Frontiers in Earth Science Series, Part II*, pages 91–121. Springer, Berlin Heidelberg New York.

Ross, Z. E., Meier, M. A., and Hauksson, E. (2018a). P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning. *Journal of Geophysical Research: Solid Earth*, 123(6):5120–5129.

Ross, Z. E., Meier, M. A., Hauksson, E., and Heaton, T. H. (2018b). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5):2894–2901.

Ross, Z. E., Trugman, D. T., Hauksson, E., and Shearer, P. M. (2019a). Searching for hidden earthquakes in Southern California. *Science*, 771(May):767–771.

Ross, Z. E., Yue, Y., Meier, M. A., Hauksson, E., and Heaton, T. H. (2019b). PhaseLink: A Deep Learning Approach to Seismic Phase Association. *Journal of Geophysical Research: Solid Earth*, 124(1):856–869.

Rubin, A. M., Gillard, D., and Got, J. L. (1999). Streaks of microearthquakes along creeping faults. *Nature*, 400(6745):635–641.

Ruegg, J. C., Campos, J., Armijo, R., Barrientos, S., Briole, P., Thiele, R., Arancibia, M., Cañuta, J., Duquesnoy, T., Chang, M., Lazo, D., Lyon-Caen, H., Ortlieb, L., Rossignol, J. C., and Serrurier, L. (1996). The Mw=8.1 Antofagasta (North Chile) Earthquake of July 30, 1995: First results from teleseismic and geodetic data. *Geophysical Research Letters*, 23:917–920.

Ruiz, S. and Madariaga, R. (2018). Historical and recent large megathrust earthquakes in Chile. *Tectonophysics*, 733:37–56.

Ruiz, S., Metois, M., Fuenzalida, A., Ruiz, J., Leyton, F., Grandin, R., Vigny, C., Madariaga, R., and Campos, J. (2014). Intense foreshocks and a slow slip event preceded the 2014 Iquique Mw8.1 earthquake. *Science*, 345(6201):1165–1169.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Rundle, J. B. (1978). Viscoelastic crustal deformation by finite quasi-static sources. *Journal of Geophysical Research: Solid Earth*, 83(B12):5937–5945.

Sak, P. B., Fisher, D. M., Gardner, T. W., Marshall, J. S., and LaFemina, P. C. (2009). Rough crust subduction, forearc kinematics, and Quaternary uplift rates, Costa Rican segment of the Middle American Trench. *Bulletin of the Geological Society of America*, 121(7-8):992–1012.

Salazar, P., Wigger, P., Bloch, W., Asch, G., Shapiro, S. A., and Kummerow, J. (2013). MEJIPE. *GFZ Data Services. Other/Seismic Network.* https://doi.org/10.7914/SN/8G_2013.

Sallarès, V. and Ranero, C. R. (2005). Structure and tectonics of the erosional convergent margin off Antofagasta, north Chile (23°30'S). *Journal of Geophysical Research: Solid Earth*, 110(6):1–19.

Savage, J. C. (1983). A dislocation model of strain accumulation and release at a subduction zone. *Journal of Geophysical Research*, 88(B6):4984–4996.

Schaff, D. P., Beroza, G. C., and Shaw, B. E. (1998). Postseismic response of repeating aftershocks. *Geophysical Research Letters*, 25(24):4549–4552.

Scharf, L. L. and Friedlander, B. (1994). Matched Subspace Detectors. *IEEE Transactions on Signal Processing*, 42(8):2146–2157.

Schurr, B., Asch, G., Hainzl, S., Bedford, J., Hoechner, A., Palo, M., Wang, R., Moreno, M., Bartsch, M., Zhang, Y., Oncken, O., Tilmann, F., Dahm, T., Victor, P., Barrientos, S., and Vilotte, J. P. (2014). Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake. *Nature*, 512(7514):299–302.

Schurr, B., Asch, G., Rosenau, M., Wang, R., Oncken, O., Barrientos, S., Salazar, P., and Vilotte, J. P. (2012). The 2007 M7.7 Tocopilla northern Chile earthquake sequence: Implications for along-strike and downdip rupture segmentation and megathrust frictional behavior. *Journal of Geophysical Research: Solid Earth*, 117(5):1–19.

Schurr, B., Asch, G., and Wigger, P. (1997). The ZB Seismic Network. *GFZ Data Services. Other/Seismic Network.* https://doi.org/10.14470/MO6442843258.

Schurr, B., Moreno, M., Tréhu, A. M., Bedford, J., Kummerow, J., Li, S., and Oncken, O. (2020). Forming a Mogi Doughnut in the Years Prior to and Immediately Before the 2014 M8.1 Iquique, Northern Chile, Earthquake. *Geophysical Research Letters*, 47(16):e2020GL088351.

Schurr, B., Rietbrock, A., Asch, G., Kind, R., and Oncken, O. (2006). Evidence for lithospheric detachment in the central Andes from local earthquake tomography. *Tectonophysics*, 415(1-4):203–223.

Schuster, M. (1999). On supervised learning from sequential data with applications for speech recognition. *Ph.D. thesis. Nara Institute of Science and Technology, Kyoto, Japan.*

Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45:2673–2681.

Shelly, D. R., Beroza, G. C., and Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, 446(7133):305–307.

Shoji, D., Noguchi, R., Otsuki, S., and Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Scientific Reports*, 8:8111.

Simard, D., Steinkraus, P. Y., and Platt, J. C. (2003). Best practices for convolutional neural networks. In *Proceedings of International Conference on Document Analysis and Recognition 7*, pages 958–963.

Simoes, M., Avouac, J. P., Cattin, R., and Henry, P. (2004). The Sumatra subduction zone: A case for a locked fault zone extending into the mantle. *Journal of Geophysical Research: Solid Earth*, 109:B10402.

Sippl, C., Schurr, B., Asch, G., and Kummerow, J. (2018). Seismicity Structure of the Northern Chile Forearc From >100,000 Double-Difference Relocated Hypocenters. *Journal of Geophysical Research: Solid Earth*, 123(5):4063–4087.

Sippl, C., Schurr, B., John, T., and Hainzl, S. (2019). Filling the gap in a double seismic zone: Intraslab seismicity in Northern Chile. *Lithos*, 346-347.

Sippl, C., Schurr, B., Yuan, X., Mechie, J., Schneider, F. M., Gadoev, M., Orunbaev, S., Oimahmadov, I., Haberland, C., Abdybachaev, U., Minaev, V., Negmatullaev, S., and Radjabov, N. (2013). Geometry of the Pamir-Hindu Kush intermediate-depth earthquake zone from local seismic data. *Journal of Geophysical Research: Solid Earth*, 118(4):1438–1457.

Sleeman, R. and Van Eck, T. (1999). Robust automatic P-phase picking: An on-line implementation in the analysis of broadband seismogram recordings. In *Physics of the Earth and Planetary Interiors*, volume 113, pages 265–275.

Sobiesiak, M. and Schurr, B. (2007). The Y9 Seismic Network. *GFZ Data Services. Other/Seismic Network.* https://doi.org/10.14470/L47552843299.

Socquet, A., Valdes, J. P., Jara, J., Cotton, F., Walpersdorf, A., Cotte, N., Specht, S., Ortega-Culaciati, F., Carrizo, D., and Norabuena, E. (2017). An 8 month slow slip event triggers progressive nucleation of the 2014 Chile megathrust. *Geophysical Research Letters*, 44(9):4046–4053.

Sodoudi, F., Yuan, X., Asch, G., and Kind, R. (2011). High-resolution image of the geometry and thickness of the subducting Nazca lithosphere beneath northern Chile. *Journal of Geophysical Research: Solid Earth*, 116(4):1–11.

Soto, H., Sippl, C., Schurr, B., Kummerow, J., Asch, G., Tilmann, F., Comte, D., Ruiz, S., and Oncken, O. (2019a). Catalog of Hypocenters for the 2014 M8.1 Iquique Earthquake Sequence, recorded by IPOC (plus additional) seismic stations. *GFZ Data Services.* https://doi.org/10.5880/GFZ.4.1.2019.009.

Soto, H., Sippl, C., Schurr, B., Kummerow, J., Asch, G., Tilmann, F., Comte, D., Ruiz, S., and Oncken, O. (2019b). Probing the Northern Chile Megathrust With Seismicity: The 2014 M8.1 Iquique Earthquake Sequence. *Journal of Geophysical Research: Solid Earth*, 124(12):12935–12954.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958.

Stewart, S. W. (1977). Real-Time Detection and Location of Local Seismic Events in Central California. *Bulletin of the Seismological Society of America*, 67(2):433–452.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, volume 4, pages 3104–3112.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1701–1708.

Takanami, T. and Kitagawa, G. (1988). A new efficient procedure for the estimation of onset times of seismic waves. *Journal of Physics of the Earth*, 36(6):267–290.

Tarantola, A. and Valette, B. (1982). Inverse problems = quest for information. *Journal of Geophysics - Zeitschrift fur Geophysik*.

Thireou, T. and Reczko, M. (2007). Bidirectional long short-term memory networks for predicting the subcellular localization of eukaryotic proteins. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California (USA). *Journal of Geophysical Research*, 88(B10):8226–8236.

Thurber, C. H. (1993). Local earthquake tomography: Velocities and vp/vs-theory. In Iyer, H. and Hirahara, K., editors, *Seismic Tomography: Theory and Practice*, pages 563–583. Chapman and Hall, London.

Tilmann, F., Zhang, Y., Moreno, M., Saul, J., Eckelmann, F., Palo, M., Deng, Z., Babeyko, A., Chen, K., Baez, J. C., Schurr, B., Wang, R., and Dahm, T. (2016). The 2015 Illapel earthquake, central Chile: A type case for a characteristic earthquake? *Geophysical Research Letters*, 43(2):574–583.

Tilmann, F. J., Craig, T. J., Grevemeyer, I., Suwargadi, B., Kopp, H., and Flueh, E. (2010). The updip seismic/aseismic transition of the Sumatra megathrust illuminated by aftershocks of the 2004 Aceh-Andaman and 2005 Nias events. *Geophysical Journal International*, 181(3):1261–1274.

Toda, S., Stein, R. S., Richards-Dinger, K., and Bozkurt, S. B. (2005). Forecasting the evolution of seismicity in southern California: Animations built on earthquake stress transfer. *Journal of Geophysical Research: Solid Earth*, 110(5):1–17.

Turcotte, D. and Schubert, G. (2014). *Geodynamics*. Cambridge University, New York, 3 edition.

Van Trees, H. (1968). *Detection, Estimation and Modulation Theory*. John Wiley and Sons, Inc., New York.

Völker, D., Wiedicke, M., Ladage, S., Gaedicke, C., Reichert, C., Rauch, K., Kramer, W., and Heubeck, C. (2006). Latitudinal Variation in Sedimentary Processes in the Peru-Chile Trench off Central Chile. In Oncken, O., Chong, G., Franz, G., Giese, P., Götze, H.-J., Ramos, V., Strecker, M., and Wigger, P., editors, *The Andes – Active Subduction Orogeny. Frontiers in Earth Science Series, Part II*, pages 193–216. Springer, Berlin Heidelberg New York.

von Huene, R., Ranero, C. R., and Vannucchi, P. (2004). Generic model of subduction erosion. *Geology*, 32(10):913–916.

von Huene, R. and Scholl, D. W. (1991). Observations at convergent margins concerning sediment subduction, subduction erosion, and the growth of continental crust. *Reviews of Geophysics*, 29(3):279–316.

Waldhauser, F. and Ellsworth, W. L. (2000). A Double-difference Earthquake location algorithm: Method and application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, 90(6):1353–1368.

Waldhauser, F., Ellsworth, W. L., Schaff, D. P., and Cole, A. (2004). Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity. *Geophysical Research Letters*, 31(18):2–5.

Wang, H. F. (2000). *Theory of linear poroelasticity with applications to geomechanics and hydrogeology*. Princeton University Press, Princeton, NJ.

Wang, K. and Bilek, S. L. (2011). Do subducting seamounts generate or stop large earthquakes? *Geology*, 39(9):819–822.

Wang, K. and Bilek, S. L. (2014). Invited review paper: Fault creep caused by subduction of rough seafloor relief. *Tectonophysics*, 610:1–24.

Wang, K. and Hu, Y. (2006). Accretionary prisms in subduction earthquake cycles: The theory of dynamic Coulomb wedge. *Journal of Geophysical Research: Solid Earth*, 111(6):1–16.

Wang, X. and Liu, P. L. (2006). An analysis of 2004 Sumatra earthquake fault plane mechanisms and Indian Ocean tsunami. *Journal of Hydraulic Research*, 44(2):147–154.

Weertman, J. (1980). Unstable slippage across a fault that separates elastic media of different elastic constants. *Journal of Geophysical Research*, 85(B3):1455–1461.

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*.

Werbos, P. J. (1990). Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE*, 78(10):1550–1560.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H., and Tian, D. (2019). The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*, 20(11):5556–5564.

Wetzler, N., Lay, T., Brodsky, E. E., and Kanamori, H. (2018). Systematic deficiency of aftershocks in areas of high coseismic slip for large subduction zone earthquakes. *Science Advances*, 4(2):1–10.

Wigger, P., Salazar, P., Kummerow, J., Bloch, W., Asch, G., and Shapiro, S. (1996). West–Fissure- and Atacama-Fault Seismic Network (2005/2012). *GFZ Data Services. Other/Seismic Network.* https://doi.org/10.14470/3S7550699980.

Williams, R. J. and Zipser, D. (1995). Gradient-based learning algorithms for recurrent networks and their computational complexity. In Chauvin, Y. and Rumelhart, D., editors, *Back-propagation: Theory, Architectures and Applications*. Erlbaum, Hillsdale, New Jersey.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, 88(1):95–106.

Woessner, J., Schorlemmer, D., Wiemer, S., and Mai, P. M. (2006). Spatial correlation of aftershock locations and on-fault main shock properties. *Journal of Geophysical Research: Solid Earth*, 111(8):1–17.

Woollam, J., Rietbrock, A., Bueno, A., and De Angelis, S. (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. *Seismological Research Letters*, 90(2A):491–502.

Yagi, Y., Okuwaki, R., Enescu, B., Hirano, S., Yamagami, Y., Endo, S., and Komoro, T. (2014). Rupture process of the 2014 Iquique Chile earthquake in relation with the foreshock activity. *Geophysical Research Letters*, 41(12):4201–4206.

Yañez, G., Ranero, C., von Huene, R., and Díaz, J. (2001). A tectonic interpretation of magnetic anomalies across a segment of the convergent margin of the Southern Central Andes (32°–34°S). *J Geophys Res*, 106(3):6325–6345.

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., and Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific Reports*, 10, 1941.

Zhu, J., Kopp, H., Flueh, E. R., Klaeschen, D., Papenberg, C., and Planert, L. (2009). Crustal structure of the central Costa Rica subduction zone: Implications for basal erosion from seismic wide-angle data. *Geophysical Journal International*, 178(2):1112–1131.

Zhu, W. and Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273.

Acknowledgements

First of all, I would like to thank Dr. Bernd Schurr, for supporting my application to my doctoral scholarship and for giving me the opportunity to work under his supervision. Many thanks also for helping me to keep focused when necessary and for giving me feedback about my writing and when discussing my results, as well as for letting me take the initiative in implementing my own ideas in the last part of this research.

I wish to also thank Prof. Onno Oncken and Prof. Frederik Tilmann who further provided their knowledge and wisdom when needed during this research. They are also acknowledged for their supervision duties and in their role as reviewers of this thesis.

A big thanks goes to our Head Assistant Franziska Alberg and Secretary Regina Prero at Section 4.1 Lithosphere Dynamics of the GFZ German Research Centre for Geosciences for their support with all sorts of administrative and personal issues. Whenever I had a problem, I knew for sure that they would promptly manage to solve it.

Thanks Prof. Andres Tassara and Prof. Klaus Bataille from University of Concepcion in Chile for encouraging me to pursue my PhD in Berlin. Many thanks to my Chilean colleagues and friends, who are currently working in several of the many sections of GFZ, for supporting me in one way or another during these years: Ayleen Gaete, Begoña Parraguez, Carlos Peña, Violeta Veliz, Felipe Vera, and Isabel Urrutia. I would also like to thank some of the ones who already left GFZ, for sharing valuable information from their own past experiences: Jaime Araya, Jorge Bernales, and Catalina Ramos.

Many thanks go also to my non-Chilean colleagues currently or previously working at GFZ: Antonio Bayona, Jon Bedford, Silvia Crosetto, Ehsan Kosari, Shaoyang Li, Armel Menant, Alejandro Mora, and all my colleagues at Section 4.1 for exchanging scientific and non-scientific ideas through various interesting conversations.

Special thanks go to Christian Sippl, for sharing his coding skills (and his code scripts too) and seismological knowledge, as well as for being always generously supportive to all my numerous questions and for giving me valuable advice through all stages of my PhD. Thanks also to Hans Agurto for helping me from a distance to solve practical issues and for his assertive advising on scientific matters. Further thanks go to Diego Sáez for sharing his knowledge on Machine Learning, that helped me in the decision making process during the last part of my research.

I acknowledge the organizations National Commission for Scientific and Technological Research (CONICYT) and The Deutscher Akademischer Austauschdienst (DAAD) for financial support during most of these years as a PhD student. The HAzard and Risk Team (HART) initiative of the GFZ Centre in collaboration with the Institute of GeoSciences, Energy, Water and Environment of the Polytechnic University Tirana, Albania and the KIT Karlsruhe Insti-

tute of Technology for the additional financial support that allowed me to complete the last part of my research. The GFZ Centre, in particular, for providing me with a workstation and the necessary resources to develop this research project. Institutions and people that support the Integrated Plate Boundary Observatory Chile (IPOC), as well as the various other temporary seismic networks, which collected the large amount of the seismic data without which this thesis would not have been possible. The Geophysical Instrument Pool Potsdam (GIPP), in particular, for providing the seismic instruments used at several of these networks.

I thank all the Machine Learning community, including both their members at GFZ and the ones who actively give their support online, from whom I got the necessary inspiration to carry out the second part of my research. I specially thank the developers of the Python library Hyperopt (Bergstra et al., 2013a) and the dedicated machine learning frameworks Keras (Chollet, 2015) and Tensorflow (Abadi et al., 2015), without which I would not have been able to implement my deep learning ideas. Developers of Python packages Obspy (*http://obspy.org*; Beyreuther et al., 2010) and Matplotlib (Hunter, 2007) and Generic Mapping Tools GMT (Wessel et al., 2019) are acknowledged as well, since I use these tools to create many of the figures and perform most of the data analysis presented in this thesis.

I want to finally thank my family for their constant love and encouragement, especially my parents Héctor and Elizabeth, my sister Mariely, and my nephew Gabriel, for cheering me up from the other side of the pond during all my PhD, especially in this tough global pandemic. I also warmly thank Maria for helping me with my English grammar issues and for listening when I most needed it in the last stage of this research adventure.

Curriculum Vitae

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.