UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE GEOFÍSICA

APLICACIÓN DE MÉTODOS BAYESIANOS DE INVERSIÓN AL ANÁLISIS DEL
TERREMOTO DE TOCOPILLA (M_W 7.7) 2007

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS, MENCIÓN GEOFÍSICA

NATALIA PATRICIA DÍAZ SALAZAR

PROFESOR GUÍA:
FRANCISCO ORTEGA CULACIATI

PROFESOR CO-GUÍA:
JAVIER RUIZ PAREDES

MIEMBROS DE LA COMISIÓN:
MARCOS MORENO SWITT
MARIO PARDO PEDEMONTE

SANTIAGO DE CHILE
2019

APLICACIÓN DE MÉTODOS BAYESIANOS DE INVERSIÓN AL ANÁLISIS DEL TERREMOTO DE TOCOPILLA (M_W 7.7) 2007

El terremoto de Tocopilla (M_w 7.7) 2007 fue el primer gran evento interplaca ocurrido en la parte sur de la brecha sísmica definida por el gran terremoto de 1877 (M_w 8.8), en la interfaz entre las placas de Nazca y Sudamericana en el norte de Chile. Este evento fue seguido por el terremoto de Pisagua (M_w 8.2) en 2014 y su réplica principal (M_w 7.7), ambos rompieron la parte central de la brecha, produciendo una pequeña fracción del déficit de deslizamiento potencialmente acumulado en la región desde 1877 sin reducir significativamente la amenaza sísmica en la zona. Como el problema de la estimación de parámetros de la fuente sísmica a partir de observaciones en superficie es mal condicionado, se reanaliza el terremoto de Tocopilla utilizando un enfoque Bayesiano de inversión, que tiene la ventaja de caracterizar completamente la incertidumbre de los parámetros de la fuente inferidos sin introducir una regularización adicional a la impuesta por la parametrización de la fuente y utilizando sólo información a priori justificable físicamente.

En esta tesis se utiliza el algoritmo de muestreo CATMIP (*Cascading Adaptive Metropolis in Parallel*) implementado en el *software* AlTar, junto con los datos geodésicos disponibles (GPS & InSAR) y una geometría detallada de la placa subductante, para producir una caracterización completa del deslizamiento cuasiestático de la fuente del terremoto de Tocopilla. A diferencia de otros estudios, en este se incluye incertezas observacionales y de la predicción del modelo físico directo debido a errores en la estructura elástica del medio. Estudios previos sugieren que una característica geométrica o reológica podría estar actuando como una barrera a la propagación de la ruptura. En este estudio se incorpora un cambio abrupto en el manteo del plano de falla basado en modelos recientes de la interfaz de subducción en esa zona. De la inversión Bayesiana se obtiene un conjunto de modelos con sus probabilidades asociadas, el cual se analiza calculando modelos representativos, tales como el modelo promedio, mediana y de máxima verosimilitud. Estos modelos, concuerdan con estudios previos en que la ruptura ocurrió en la parte más profunda del contacto sismogénico, con el mayor deslizamiento localizado en dos asperezas principales siendo el máximo de $2 \pm 0,6$ m. Sin embargo, el patrón de deslizamiento de estos modelos es más compacto y parte de la zona de mayor deslizamiento se ubica costa afuera comparado con estudios previos. La solución encontrada sugiere variaciones abruptas en las propiedades reológicas sobre el contacto de falla, lo que concuerda con la hipótesis planteada.

Este estudio difiere en el énfasis en la caracterización de las incertezas, correlaciones y/o dependencias espaciales entre la dislocación en las distintas subfallas. Por ejemplo, la forma de la distribución en los elementos cercanos a la fosa sugiere una resolución más baja en esa zona, dado que esos elementos están peor constreñidos por las observaciones.

El método empleado es costoso en términos computacionales, dado que muchos modelos directos ($\sim 10^{10}$) son evaluados para representar la solución del problema inverso. Sin embargo, este método ofrece una caracterización completa de las soluciones y sus incertezas.

APPLICATION OF BAYESIAN INVERSE METHODS IN THE TOCOPILLA M_W 7.7 2007 EARTHQUAKE ANALYSIS

The Tocopilla (Mw 7.7) 2007 earthquake was the first large interplate event to occur within the southern end of the seismic gap as defined by the great 1877 (Mw 8.8) earthquake, at the Nazca - South American plate interface in Northern Chile. This event was followed by the Pisagua (Mw 8.2) 2014 earthquake and its main aftershock (Mw 7.7) which only ruptured the central portion of the gap. Those events produced a small fraction of the slip that has potentially accumulated in the region since 1877, and therefore do not significantly reduce concerns for earthquake hazard in the region. As the problem of estimating the seismic source from surface observations is ill-conditioned, we reanalyze the Tocopilla earthquake using a Bayesian approach, which has the advantage of producing a complete characterization of the uncertainties of model parameter estimates and being able to find a solution to the inverse problem without requiring additional regularization, other than that imposed by the model parameterization itself, and using only physically justifiable a priori information.

We use the sampling algorithm CATMIP (Cascading Adaptive Transitional Metropolis in Parallel) implemented in the AlTar software, along with available geodetic data (GPS & InSAR) and a detailed geometry of the subducting plate to produce a complete characterization of quasi-static slip at the Tocopilla earthquake source. Differing from other studies, ours includes observational and physical model prediction uncertainties due to errors in the elastic structure of the media. Previous studies suggest that a geometric or rheological features might be acting as a rupture propagation barrier updip. Here we incorporate a kink in the fault plane based in more recent studies of the subduction interface in that zone. The result in a Bayesian inversion is an ensemble of models with their associated probabilities, which can be analized using representative models such as the mean, median and the maximum likelihood model. These models agree with previous studies in which the rupture occurred in the deepest part of the seismogenic contact, with largest slip located at two main asperities, obtaining $2 \pm 0.6m$ as the maximum slip.

However, the rupture is more compact and it is partially offshore compared with the previous published studies. The more compact solution found suggests abrupt changes in rheology. We differ too in the emphasis given to slip uncertainty characterization, spatial correlations and dependencies between the dislocation on different subfaults. For example, the shaped slip probability distribution for elements close to the trench suggests a lower resolution, since these elements are less constrained by the observations.

The method employed is costly in computational resources, since many models ($\sim 10^{10}$) are evaluated to find the ensemble of slip models that define the solution of the inverse problem. However, it offers a complete characterization of the solutions and their uncertainties.

Dedicada a mi mamá

Acknowledgements

Contents

List of figures

List of tables

List of Algorithms

Chapter 1

Introduction

1.1　The Tocopilla Earthquake

The 2007 M_w 7.7 Tocopilla earthquake had a rupture zone located in the plate interface of the subduction of the Nazca plate under the South - American plate, one of the most active margins in the world. The Tocopilla earthquake occurred inside the rupture zone of the 1877 great earthquake in northern Chile, an event with estimated magnitude Mw 8.8. (*Comte & Pardo*, 1991) that defines the seismic gap of the northern Chile, a region in which large earthquakes had not ocurred for more than a century (Figure 1.1). Due to the plate tectonic activity in the region, with a convergence rate near 6.7cm/y, that zone is expected to have accumulated a large slip deficit, although it is not straightforward how much of it would manifests as earthquakes (*Wyss & Wiemer*, 1999), considering the occurrence of other type of phenomena such as aseismic slip (*Pritchard & Simons*, 2006).

The Tocopilla (M_w 7.7) earthquake is the first large event occurred since 1877 within the northern Chile seismic gap, taking place at the southern edge of the gap in the deepest section of the subduction megathrust ($> 30km$). Following the Tocopilla earthquake, the 2014 (M_w 8.2) Pisagua earthquake partially broke the central segment of the gap. The 2014 earthquake was studied using optimization inversion techniques by several authors (e.g. *Jara et al.*, 2018; *Yagi et al.*, 2014; *Gusman et al.*, 2015) and using Bayesian methods by *Duputel et al.* (2015). The advantage of the Bayesian approach is that it does not impose a priori restrictions to the solution, allows to obtain a complete ensemble of solutions and make a more complete characterization of the uncertainties. The Pisagua earthquake was followed by a large aftershock of magnitude M_w 7.7 (e.g. *Duputel et al.*, 2015; *Schurr et al.*, 2014; *Leon Rios et al.*, 2016). The Tocopilla and Pisagua events have released a small fraction of the slip deficit accumulated in the region (*Duputel et al.*, 2015), remaining still a great potential for the occurrence of earthquakes of great magnitude.

In order to better understand the mechanical behavior of the subduction interface at the 1877 seismic gap, we use a fully Bayesian inversion technique to analyze the source of the 2007 M_w 7.7 Tocopilla earthquake. This method allows to fully explore plausible slip distributions and characterize uncertainties of estimated model parameters, without requiring additional regularization (that would bias our estimates), other than that imposed by the model parameterization and physically justifiable a priori information.

The Tocopilla earthquake has been characterized by a comprehensive set of geodetic and seismological observations, allowing a detailed study of the source of the earthquake. Some of these studies were those conducted by *Delouis et al.* (2009) from the analysis of seismological data, and *Béjar-Pizarro et al.* (2010), using GPS and InSAR observations. These works conclude that the slip in the source of the earthquake occurs in two main asperities, coinciding in the lower limit of the rupture zone(50 km). However, these studies differ slightly in position and dimensions of these asperities. *Béjar-Pizarro et al.* (2010) puts the upper limit in a slightly more shallow location compared to *Delouis et al.* (2009) (depths of 30 km and 35 km, respectively). *Minson* (2010) studied the Tocopilla earthquake using Bayesian methods, that as we will see later, yield an ensemble of models solving the inverse problem. The representative models show solutions with a maximum slip near 3.0 m.

The work of *Béjar-Pizarro et al.* (2010) indicates that the fitting between observed deformation and the results of the model is not very accurate for the asperity located in the southern part of the study area. *Minson* (2010) solved the inverse problem using Bayesian methods to model the Tocopilla earthquake including the observational source of error, however, this analysis can be improved by incorporating errors in the elastic structure of the propagative medium, as we do in this work.
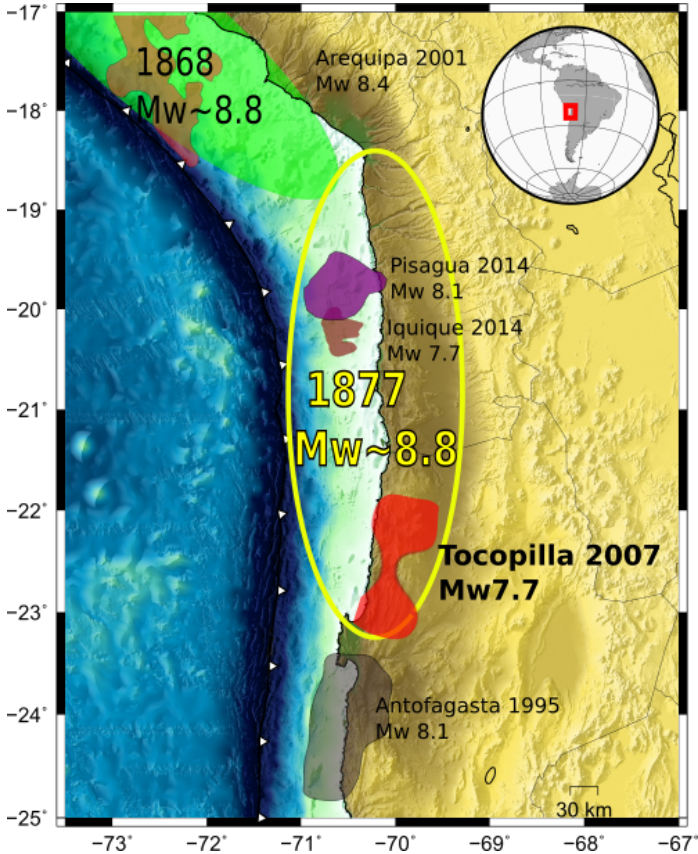
Figure 1.1: Large subduction earthquakes in southern Peru and northern Chile ocurred since 1868.

In general, with the exception of *Minson* (2010), the techniques used so far to analyze the Tocopilla earthquake (e.g., regularized least squares) must impose a bias on the slip solution,

for example, via the smoothing of slip on the fault. This requires a regularization term to find the solution of an inverse problem that is inherently ill-conditioned.

More robust results can be achieved using Bayesian inverse methods, which, in contrast with optimization methods (e.g., least squares), do not require additional regularization to stabilize the solutions of the inverse problem. This means that the Bayesian methodology can be done using only physically justifiable a priori information on model parameters, such as type of fault (e.g inverse fault slip), but does not require imposing artificial biases to the solution, such as smoothing constraints over model parameters. The Bayesian methodology allows to deeply explore the possible solutions of the ill-posed inverse problem, as well as to better characterize uncertainties in the model parameter estimates.

1.2 Bayesian Methods

Bayesian methods are based on Bayes' Theorem. To solve the inverse problem, it can be formulated in terms of a set of parameters $\boldsymbol{\theta}$ and a set of observations $\mathbf{D}$, as

$$P(\boldsymbol{\theta}|\mathbf{D}) = \frac{P(\mathbf{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{D})} \tag{1.1}$$

Formula 1.1 implies that the probability of a model $\boldsymbol{\theta}$ given the observed data $\mathbf{D}$ is proportional to the probability of such data $\mathbf{D}$ given the model $\boldsymbol{\theta}$, i.e $P(\mathbf{D}|\boldsymbol{\theta})$ (the likelihood function) and a probability distribution of models $P(\boldsymbol{\theta})$, representing the prior knowledge about the model space. $P(\mathbf{D})$ is the probability of the observed data $\mathbf{D}$, which is a constant for the entire problem. The last two expressions are much more accessible, since calculating these simply requires to evaluate the likelihood function, and the probability density function (pdf) defining a priori information on model parameters. The solution of the inverse problem is the probability distribution of the model parameters given the observed data $\mathbf{D}$. $P(\boldsymbol{\theta}|\mathbf{D})$, called the a posteriori pdf of model parameters, represents the solution of the inverse problem.

Bayesian methods solve an inverse problem by sampling the a posteriori probability distribution of model parameters, which produces an ensemble of models, i.e. a set of models with their probability of being a solution of the inverse problem.

This Bayesian methodology allows to solve problems of both types, linear and nonlinear, without imposing extra conditions other than physical restrictions. They are costly in computational terms, but have become more manageable in recent decades thanks to advances in parallel computing, and available hardware such as large scale computational facilities with thousands of CPU's or GPGPU hardware.

The additional advantage of Bayesian methods is that they allow to describe in a more extensive way the errors in the modeling. The uncertainties associated to model parameter estimates come from several sources, two of the main are: observational errors, associated to the measurement process of the data used to constrain our models, and second, errors inherent to the used physical model used (epistemic error). The latter error comes from the selection of the theory used to represent the physical phenomena being studied and the variables that constitute it (*Tarantola*, 2005). Usually the error analysis considers only the

source of observational error, which may be less significant than epistemic errors present in the problem (*Duputel et al.*, 2014; *Minson*, 2010). Not considering epistemic errors leads to solutions whose predictions overfit the observations, i.e., fitting the errors. Then, model parameter estimates may contain spurious signals that try to explain the aforementioned errors. Bayesian methods provide a way to incorporate the epistemic error in the modeling, which is not an easy task when using other methods.

Exploring the model parameter solution space requires significant computational resources, since the number of calculations grows enormously with the number of parameters, due to the so called 'curse of dimensionality' (*Tarantola*, 2005). For this reason, it is extremely important to use methods that allow efficient sampling of the a posteriori pdf of model parameters and take advantage of current parallel computing capabilities. Numerous algorithms have been developed in order to characterize the solution space of the inverse problem through samples of the a posteriori pdf, most of these based on MCMC (Markov Chain Monte Carlo) sampling techniques. In this context, the algorithm developed by *Minson* (2010) in her doctoral thesis, CATMIP, shows to be the most efficient for geophysical applications.

1.3   CATMIP Algorithm and AlTar

The CATMIP algorithm (Cascading Adaptive Transitional Metropolis in Parallel) is a Bayesian sampling algorithm that iteratively approximates the a posteriori distribution of model parameters of the inverse problem. CATMIP has the advantage of allowing to solve problems with a large number of parameters, such as the modeling of the seismic source for large earthquakes (*Duputel et al.*, 2015; *Simons et al.*, 2011; *Minson et al.*, 2013). Its implementation incorporates elements of sampling algorithms such as Metropolis-Hastings, resampling techniques and transitioning from a priori to a posterior pdf's of model parameters. The CATMIP algorithm is a technique that allows more complex distributions in the space of model parameters; for example, with more than one maximum value. The particularity of this method is that it allows us to perform calculations efficiently using parallel processing.

We use an implementation of CATMIP, the software AlTar, developed in the Seismological Laboratory of Caltech, which allows to solve large slip inversion problems by taking advantage of parallel processing using GPGPU hardware.

The purpose of this thesis is to use Bayesian methods of inversion, in particular AlTar software, to characterize the solutions of a specific type of problem in Geophysics: the inversion of the source of earthquakes. The main objective is to apply this methodology to the detailed re-analysis of the 2007 Tocopilla Earthquake (Mw 7.7), in which we incorporate epistemic or model prediction errors due to the elastic structure of the media, while there is a wide data registry that consists of both, spatial geodesy and seismological data.

1.4   Objectives

1.4.1   General

To apply Bayesian inverse methods to the analysis of the seismic source of the Tocopilla earthquake 2007 (Mw 7.7), to characterize in a robust way, the solutions and the errors associated with the parameters of the seismic source model and to analize the source of the Tocopilla (Mw 7.7) earthquake through a robust characterization of plausible static slip distributions and associated uncertainties, using fully Bayesian inversion techniques.

1.4.2   Specifics

- To understand how the CATMIP algorithm works, performing tests of its different stages through analysis based on synthetic models. Also, to understand the scope and limitations of these methods.

- To test the algorithm applying it to an inverse problem with few parameters, specifically the hypocenter location of an earthquake, and to analyze the results obtained by the method.

- To model the seismic source of the 2007 Tocopilla earthquake using the AlTar software to perform the Bayesian inversion of this problem. The results of this modeling will be compared with previous works.

Outline of this Thesis

This Thesis has been organized in 5 chapters. A summary of the contents of each chapter is presented below.

- **Chapter 1: Introduction**
  Introduction to the 2007 Tocopilla earthquake (Mw 7.7) and brief description of the Bayesian methodology and the CATMIP algorithm. Here we present the problem to be solved, the inversion of the static slip distribution of the earthquake using the Bayesian method CATMIP, including a more complete analysis of uncertainties including epistemic or model prediction errors.

- **Chapter 2: Bayesian Methods**
  This chapter presents first a brief history and formulation of the Bayesian methods and a summary of the more relevant Bayesian methods to understand the methodology employed in this thesis, the CATMIP algorithm. Finally, a section dedicated to graphically understanding this algorithm and to present tests using synthetic and real data, showing the stages and performance of the method when solving a simplified version of the hypocenter determination problem.

- **Chapter 3: The Tocopilla earthquake**
  This chapter summarize previous studies of the 2007 Tocopilla earthquake (Mw 7.7) and the seismotectonic context of this event.

- **Chapter 4: Modeling the Tocopilla earthquake**
  In this chapter we present first the modeling of the 2007 Tocopilla (Mw 7.7) earthquake, including the setting of the fault geometry, the data set and the inverse problem formulation. Later, in the section dedicated to the results, we show the marginal distributions of static slip obtained from the fully Bayesian inversion of the 2007 Tocopilla earthquake. The results are presented in conjunction with an analysis that includes aspects as the sensibility and the uncertainty of the solutions.

- **Chapter 6: Conclusions**
  We summarize the main results, the relevant aspects as the maximum slip, updip, downdip and the relation with the regional seismotectonic context.

Chapter 2

Bayesian Methods

Bayesian methods are based on Bayes' Theorem, named after Thomas Bayes (1702 - 1761), a philosopher and presbyterate minister whose work in the field of mathematics was rediscovered posthumously. It was near two years after his death when Richard Price a colleage and a friend of Bayes corrected his work, making his own contributions and publishing the article "An Essay towards solving a Problem in the Doctrine of Chances" (*Bayes* (1763)) highlighting its importance and laying the foundations of Bayesian statistics. However, it was Pierre-Simon Laplace who in 1774 gave a Bayesian interpretation to the theory of probability (e.g.*McGrayne* (2012)), in which the Bayes' theorem proposes a relationship between events based on conditional probabilities, relating two events A and B in the form:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.1}$$

Thus raising a relationship between these events, in the sense that the probability of event A given event B is proportional to the probability of event B given event A, multiplied by the probability of A.

One of the intriguing aspects of Bayes' theorem is that it makes it possible to relate the probability of a cause given its effect with that of the effect given its cause.

For data modeling purposes, if $\boldsymbol{\theta}$ is a set of parameters and $\mathbf{D}$ the observations. Bayes' theorem can be formulated as:

$$P(\boldsymbol{\theta}|\mathbf{D}) = \frac{P(\mathbf{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{D})} \tag{2.2}$$

$P(\boldsymbol{\theta}|\mathbf{D})$ is called the posterior distribution of parameters and it is the probability of a model $\boldsymbol{\theta}$ given the observed data $\mathbf{D}$. $P(\mathbf{D}|\boldsymbol{\theta})$ is the probability of observed data given a model or likelihood function. This function gives a notion of how well data predicted by a model $\boldsymbol{\theta}$ fits the observed data. Finally $P(\boldsymbol{\theta})$ is the probability of a model in absolute terms and represents the prior knowledge about the model parameters. Evaluating $P(\boldsymbol{\theta})$ and $P(\mathbf{D}|\boldsymbol{\theta})$ is typically much more accessible than $P(\boldsymbol{\theta}|\mathbf{D})$. Bayes' theorem provides tools to obtain information about phenomena when calculating $P(\boldsymbol{\theta}|\mathbf{D})$ absolute probabilities is a difficult or impossible task.

Here, Bayesian methods represent $P(\boldsymbol{\theta}|\mathbf{D})$ by generating samples of such pdf, which can become very costly in computational terms for large inverse problems, but have become more manageable in recent decades thanks to advances in parallel computing.

To use the Bayesian scheme to solve inverse problems it is necessary to define the likelihood function and the prior probability distribution (prior distribution for short) of the model parameters. The likelihood function is assumed to have a maximum when the residual between observed and predicted data has a minimum. If $\mathbf{G}(\boldsymbol{\theta})$ is the predicted data as a consequence of applying a physical model $\mathbf{G}$ over a set of parameters $\boldsymbol{\theta}$, $\mathbf{d_{obs}}$ is the set of observed data, and the L_2 norm is chosen, the aim is to minimize the function $||\mathbf{G}(\boldsymbol{\theta})-\mathbf{d_{obs}}||_2$. The choice of the L_2 norm is consistent with the maximum information entropy information principle (*Jaynes*, 1957), which indicates that the normal distribution maximizes entropy among all other probability distributions. The likelihood function that brings together these conditions and usually used in this context is:

$$L(\boldsymbol{\theta}) = \gamma e^{-\frac{1}{2}(\mathbf{G}(\boldsymbol{\theta}) - \mathbf{d_{obs}})^T \mathbf{C}_\chi^{-1}(\mathbf{G}(\boldsymbol{\theta}) - \mathbf{d_{obs}})}, \tag{2.3}$$

where γ is normalization constant coming from the setting of the inverse problem (in this case a probability of data, coming from Bayes' theorem). $\mathbf{C}_\chi$ is the misfit covariance, summarizing the uncertanties in the modeling. This includes observational ($\mathbf{C_d}$) and model prediction ($\mathbf{C_p}$) errors. $\mathbf{C_d}$ comes from the measurement process and $\mathbf{C_p}$ expresses the uncertainty in the prediction of the model as a consequence of uncertainties in the physical model.

The continuous formulation of the Bayes' theorem is in terms of the probability density functions (pdf's) as shown in equation 2.4. A posteriori pdf of model parameters $f()$; is proposed as a solution of the inverse problem, and the problem emerging from this point is to generate samples of f.

$$f(\boldsymbol{\theta}|\mathbf{D}) = f_{prior}(\boldsymbol{\theta})L(\mathbf{D}|\boldsymbol{\theta}) \tag{2.4}$$

The essential tool to solve problems in the Bayesian inversion is sampling, in which an ensemble of models that are samples of $f(\boldsymbol{\theta}|\mathbf{D})$ is generated. Some of the more important sampling methods will be described in the next section.

2.1 Sampling Methods

In the context of this thesis, we show here some of the most useful sampling methods to generate samples of a given probability distribution.

2.1.1 Rejection Method

A Rejection method or Acceptance-Rejection method is a technique that generates a set of samples of a probability distribution by evaluating every sample in an acceptance-rejection

scheme. Defining f as the probability density function whose distribution is going to be sampled, this method uses a pdf g as a proposal distribution, such that g multiplied by a constant c works as an envelope for f, that is, $f(x) \leq cg(x)$ for every x in the domain (see Figure 2.1). The first step is to generate a sample $x*$ of a distribution g, that sample will be accepted with probability $p = f(x*)/cg(x*)$. To test a sample it is necessary to generate a uniform random number u in $U[0,1]$. The sample is accepted if $u < p$ and rejected otherwise. This process is repeated until enough samples have been generated (*Tarantola*, 2005).
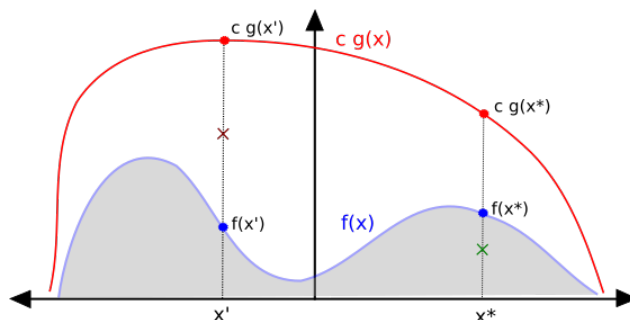
Figure 2.1: Scheme showing how a rejection method works. The figure shows two samples $x*$ and x' of the distribution whose pdf is g. A random number is generated between zero and $cg(x)$. For the sample $x*$, this random number fell under $f(x*)$ (green x), thus the sample is accepted. In the second case the random number fell outside $f(x')$ (red x), and then the sample is rejected.

The following table summarizes the acceptance - rejection method.

Algorithm 2.1 Rejection algorithm

---
1. Generate a sample x* of g, and evaluate $p = f(x*)/cg(x*)$
2. Generate a random number u in $U[0,1]$. If $u \leq p$ accept the test, and collect the test sample as part of the ensemble
3. Repeat 1 and 2 until a certain number of samples N is reached.

2.1.2 Metropolis-Hastings Algorithm (MH)

The Metropolis-Hastings algorithm is a sampling method that allows to represent a probability density function obtaining samples by designing a random walk that produces samples of the parameter space. This random walk is a sequence in which every step depends on the previous one, which obeys the definition of a Markov chain. This method was developed by Nicholas Metropolis, who published along with his team the article called *Equation of State Calculations by Fast Computing Machines* (*Metropolis & Teller*, 1953). Then, in 1970 Wilfred Keith Hastings complemented the technique allowing its use in a wider range of distributions. The advantage of this algorithm is that it explores the parameter space in a more efficient way, because it uses the previous accepted sample as a starting point to make a test step in order to search for the next samples, as shown in algorithm 2.1 below.

Schematically, the Metropolis algorithm can be represented by figure 2.2, where a sequence of samples is illustrated. In each iteration the algorithm generates a test sample starting from

---

Algorithm 2.2 Metropolis-Hastings algorithm

1. Generate a seed sample $\boldsymbol{\theta}_0$.
2. Use a proposal distribution q() to generate a test sample $\boldsymbol{\theta}_{test}$.
3. Evaluate the acceptance function $\alpha = min\left\{\frac{f(\boldsymbol{\theta}_{test})}{f(\boldsymbol{\theta}_{current})}, 1\right\}$
4. Generate a random number u in $U[0,1]$. If $u \leq \alpha$ accept the test, collect the test sample as part of the ensemble and redefine $\boldsymbol{\theta}_{current}$ as $\boldsymbol{\theta}_{current} = \boldsymbol{\theta}_{test}$. If $u > \alpha$ reject the test sample and repeat the current sample as part of the ensemble.
5. Repeat steps 2-4 until a number of samples is reached.

---

a current sample (the last accepted one) defining a perturbation from it drawn from a proposal distribution in order to generate a test model, candidate to be accepted as a sample of the pdf. Then, it compares the probability density function of the generated candidate with the corresponding to the current sample. If the pdf value of the test model is greater than that of the current sample the test model is accepted, as a sample of the pdf but, if this is not the case, the test model still might be accepted with some probability. This probability is proportional to the ratio between the value of the pdf of the test model and the current sample. In this sequence, if the test model is accepted it becomes a new link in the chain, but, if it is rejected, the last accepted model is repeated (the current sample). When a number of steps is reached, the chain itself defines an ensemble of models containing the samples of the pdf and its associated probability.

An advantage of the Metropolis-Hastings algorithm is that it allows to sample multidimensional distributions efficiently. Nevertheless, an issue with this method is that the obtained samples are correlated, as the sampling process depends on the last sampled model. Another issue is the tendency to accumulate samples around local maxima of the posterior pdf. A solution for these problems is using an alternative form of this algorithm, the Metropolis-Hastings in Parallel algorithm (MIP).
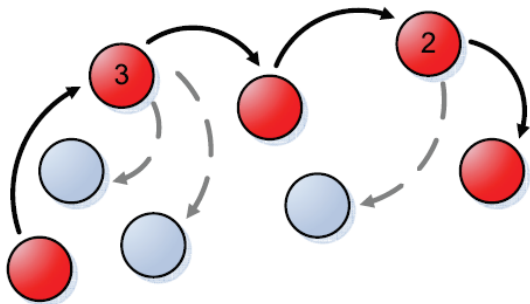
Figure 2.2: Scheme from *Minson* (2010) explaining the Metropolis Algorithm. The figure shows that a sequence of samples is generated and the accepted ones become part of the chain (in red). If a sample is rejected, the chain repeats the last accepted one and uses it as a starting point to find the next candidate. The numbers indicate how many times a repeated sample appear in the chain.

2.1.3 Metropolis-Hastings in Parallel (MIP)

In this scheme, multiple Metropolis-Hastings chains are run in parallel, saving only the last sample of each chain. Even when the net effect of this is actually increasing the number of calculations, having the proper computational resources, such as parallel processing, for instance allows to sample the whole pdf and not being stuck in local maxima. The Metropolis - Hastings in Parallel algorithm allows to distribute the seed models and the exploration regions of the Markov chains. Another effect of this processing is over the independence of the obtained samples, this in the sense that every chain is generating samples independently by contrast with the case in which we use correlated samples from the same chain. Also, if the Markov Chain MonteCarlo (MCMC) chains are long enough the last sample of the chain can be considered independent of the initial seed model of the chain.

2.1.4 Transitional Markov Chain Monte Carlo (TMCMC)

When a pdf has some pathological aspects, for example, multimodal pdfs, very peaked pdf, and pdfs with flat regions, the Metropolis algorithm may not be the best option to generate an ensemble of sampled models that represent the target probability distribution. Basically, the Metropolis algorithm, even in a parallel scheme, is prone to oversample some regions of the model parameter space in a way that makes difficult to explore other regions, as will be detailed in the next section. This problem can be adressed by adding a transitional scheme to the previous algorithms to avoid sampling directly pdfs with sharp local or global maxima, as it does the algorithm called Transitional Markov Chain Monte Carlo (TMCMC). This method was developed by *Ching & Chen* (2007) and allows to handle distributions with extreme characteristics. The algorithm transitions from the prior pdf to posterior pdf through a series of intermediate pdfs that converge to the posterior pdf. Applied to solve an inverse problem following a modified version of the Bayes rule, the algorithm achieves the posterior probability density through the following formulation:

$$f(\boldsymbol{\theta}|\mathbf{D})_m \propto P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta})^{\beta_m},$$
$$m \in 0, ..., M \tag{2.5}$$
$$0 \leq \beta_m \leq 1, \beta_0 = 0, \beta_M = 1$$

Where $f(\boldsymbol{\theta}|\mathbf{D})_m$ is the intermediate pdf of the solution in the parameter space given the observed data, β_m is the annealing variable, increasing its value from 0 to 1. When this latter value is reached, the expression corresponds to Bayes' theorem.

This algorithm starts from a prior distribution $P(\boldsymbol{\theta})$ according to the knowledge of the range of values expected for the parameters $\boldsymbol{\theta}$. Thus, the algorithm needs a function $P(\boldsymbol{\theta})$ describing this a priori knowledge. In the case of homogeneous information $P(\boldsymbol{\theta})$ represents a uniform distribution over a domain in the dimension of $\boldsymbol{\theta}$. The distribution in the step $m = 0$ will have the form of this a priori pdf. In a certain step the aim is to obtain the samples of the pdf intermediate posterior pdf from samples of the pdf corresponding to the one of the previous step. To choose a β for the next iteration value it is desirable that the distribution changes gradually. So, *Ching & Chen* (2007) define this condition in terms of the variation coefficient:

$$C_v(\boldsymbol{\omega}) = \frac{\sigma(\boldsymbol{\omega})}{\bar{\boldsymbol{\omega}}} \tag{2.6}$$

Where $\boldsymbol{\omega}$ is a set of weights associated with the sample in the current iteration obtained from the ratio between the value of the posterior pdf in the current step and the same function associated to the previous step. $\sigma(\boldsymbol{\omega})$ and $\bar{\boldsymbol{\omega}}$ are the standard deviation and the mean of those weights, respectively.

Here, ω defines the weight that will be used for resampling. This process generates a set of samples from the samples in the previous annealing iteration, repeating those according to their density in the intermediate distribution compared with their density in the preceding. Samples with higher probability appear more often in the set of samples. Naming $\boldsymbol{\theta}_{m,i}$ as the i-th sample of the set in the iteration m, the probability to generate a certain sample in the process of resampling is:

$$\omega(\boldsymbol{\theta}_{m,i}) = \frac{p(\boldsymbol{\theta}_{m,i})p(\mathbf{D}|\boldsymbol{\theta}_{m,i})^{\beta_{m+1}}}{p(\boldsymbol{\theta}_{m,i})p(\mathbf{D}|\boldsymbol{\theta}_{m,i})^{\beta_m}} = p(\mathbf{D}|\boldsymbol{\theta}_{m,i})^{\beta_{m+1}-\beta_m}. \tag{2.7}$$

Then, this method explores the surroundings of the samples. This, by using the Metropolis algorithm in such a way that repeated samples will be replaced by a Markov Chain in which the length of this chain is the number of times that a sample is repeated.

A remarkable aspect of this method is how it defines the proposal distribution for the step in the random walk. It defines a Covariance matrix:

$$
\begin{aligned}
\Sigma_m &= c^2 \sum_{i=1}^{N} \frac{\omega(\boldsymbol{\theta}_i)}{\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_i)}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^T, \\
\bar{\boldsymbol{\theta}} &= \sum_{i=1}^{N} \frac{\omega(\boldsymbol{\theta}_i)\boldsymbol{\theta}_i}{\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_i)}.
\end{aligned}
\tag{2.8}
$$

This is a weighted covariance matrix of the sample's set, where c is a scaling factor to modulate the size of the step. The previous expression means that the prefered direction for a step of the proposal distribution will be that in which more samples are present, and the size of the step will be larger in the directions having more spread sampled model parameter values at the current iteration of the method.

The following table describe briefly the TMCMC algorithm.

Algorithm 2.3 Transitional Markov Chain MonteCarlo

---

1. Set $m = 0$. Generate N samples $\{\boldsymbol{\theta}_m\} = \{\boldsymbol{\theta}_1, ... \boldsymbol{\theta}_N\}$ of the prior PDF $f_0 = P(\boldsymbol{\theta})$.
2. Set $m = m + 1$. Choose m such that $COV[\omega]$ is close to 1.
3. Calculate Σ_m and c_m.
4. Draw N samples from $\{\boldsymbol{\theta}_{m-1}\}$ with probability ω. The set of resampled models is $\left\{\hat{\boldsymbol{\theta}}_m\right\}$.
5. Count the number N_{rep} of repetitions of every sample and starting from one of these samples generate N_{rep} models using the Metropolis algorithm.
6. Repeat steps 2 to 5 until $\beta_M = 1$ is reached.

2.1.5 CATMIP

Cascading Adaptive Transitional Metropolis in Parallel (CATMIP) is a Bayesian sampling algorithm developed by *Minson* (2010) in the context of her doctoral thesis. It takes many elements of TMCMC, making some important improvements to this scheme, the main one of them is to add a mechanism for parallelization of the sampling process. Another important aspect is that the exploration of the parameter space uses a proposal distribution that depends on the acceptance rate of the previous set of samples obtained using Metropolis Hastings algorithm. Thus, allowing a larger step on the random walk if the efficiency of the last stage is higher, but a smaller size in the opposite case. Additionally it incorporates a tool to handle different sets of data, allowing to use the solution of an inversion using fewer data sets to obtain a more complete solution of the problem in a procedure called Cascading.

Let be $\boldsymbol{\theta}$ the set of parameters, $\mathbf{D}$ the observations and $f(\boldsymbol{\theta}|\mathbf{D})$ the posterior probability distribution of the parameters, the fundamental formulation of this algorithm is the same of TMCMC, defining intermediate distributions of the form:

$$f(\boldsymbol{\theta}|\mathbf{D})_m \propto P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta})^{\beta_m},$$
$$m \in 0, ..., M \tag{2.9}$$
$$0 \leq \beta_m \leq 1, \beta_0 = 0, \beta_M = 1$$

For the case $\beta = 1$, this formula corresponds to Bayes's theorem and expresses the fact that the probability of a certain model $\boldsymbol{\theta}$ given the data $\mathbf{D}$ is proportional to the probability of the data given the model. The expression $P(\mathbf{D}|\boldsymbol{\theta})$ corresponds to the likelihood function, which gives a notion of how well the test parameters model the observed data. This function is usually expressed as $L(\boldsymbol{\theta})$.

The variable β_m plays an important role as a transitional variable increasing its value gradually from 0 to 1. This means that starting from an a priori distribution, where $\beta_m = 0$, it progressively reaches the desired distribution given by Bayes's theorem (the case $\beta = 1$).

This algorithm consists of 3 main stages. The first one is the setting of an initial set of model parameter samples, that is consistent with a priori information over the model parameters. Then, the second stage follows an iterative sequence, where the algorithm generates intermediate distributions. The algorithm reaches the final stage when $\beta = 1$ and it obtains the a posteriori distribution of model parameters.

The CATMIP Algorithm

The inversion requires to define the number of samples N of the a priori pdf $f_{prior}(\boldsymbol{\theta})$ describing the previous knowledge on the model parameter space, whose determination, in turn, could be a problem by itself. This opens the question of how many samples are necessary to adequately sample the distribution, which requires to define some convergence criteria. Once these properties are settled, the algorithm works as follows.

- **Initiation:** This stage generates N samples of the $f_{prior}(\boldsymbol{\theta})$ distribution. Here, the samples can obey additional conditions that are consistent with the definition of $f_{prior}(\boldsymbol{\theta})$, for example that the summation of all parameters is close to a certain value.

- **Transition:** This stage is the iterative part of the program. At the $k - th$ iteration we have a set of samples θ_k, associated with an annealing variable β_k. The three steps that follow are repeated sequentially:

  1. **Choose β_{k+1}:** β_{k+1} is the variable that models the convergence of the model to the posterior distribution. Each β_k is chosen adaptively according to the condition described in *Beck & Zuev* (2013):

  $$C_v[\omega(\mathbf{\theta}_{k,i})] = 1, \tag{2.10}$$

  $$\omega(\mathbf{\theta}_{k,i}) = \frac{p(\mathbf{\theta}_{k,i})p(\mathbf{D}|\mathbf{\theta}_{k,i})^{\beta_{k+1}}}{p(\mathbf{\theta}_{k,i})p(\mathbf{D}|\mathbf{\theta}_{k,i})^{\beta_k}} = p(\mathbf{D}|\mathbf{\theta}_{k,i})^{\beta_{k+1}-\beta_k}, \tag{2.11}$$

  where $C_v(.)$ is the coefficient of variation and $\omega(\mathbf{\theta}_{k,i})$, the importance weight associated to every sample (see equations 2.6 and 2.7 in section about TMCMC). Then, the condition for β_k is:

  $$\frac{\sum_{i=1}^{N_k} L(\mathbf{\theta}_k^{(i)})^{2(\beta_{k+1}-\beta_k)}}{(\sum_{i=1}^{N_{k-1}} L(\mathbf{\theta}_k^{(i)})^{(\beta_{k+1}-\beta_k)})^2} = \frac{1}{\gamma N_k}, \tag{2.12}$$

  where N_k is the number of samples at step k, and γ is a threshold for the effective sample size $\gamma = \hat{N}_k^{eff}/N_k$, being $\hat{N}^{eff}$ an estimator of the effective sample size. The initial value of β is $\beta_0 = 0$ and the one of the final iteration is $\beta_k = 1$.

  2. **Resampling:** A resampling is performed by assigning a weight to each sample of the distribution according to its probability compared to the previous step of the algorithm. The likelihood function provides a notion of how well the used parameters used match the observed data. This stage will be based on the work of *Ching & Chen* (2007) and consists in generating the set of samples representing the $(k + 1) - th$ iteration from the set of samples obtained as a result of the $k - th$ iteration. This process repeats more times those samples with higher weight according to equation 2.11, as it is represented in figure 2.3.
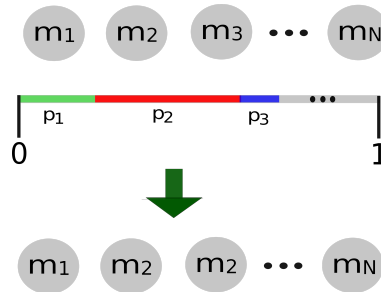
Figure 2.3: Scheme of operation of the resampling stage. This assigns a probability to each sample and then generates a new set of samples from them, repeating more times those with a greater weight.

  3. **Metropolis Sampler:** The main difference between CATMIP and TMCMC is in this part. In CATMIP, every sample of the set obtained by resampling is

used as a starting point for a random walk. So this stage uses the Metropolis algorithm, where the number of steps of the chain and the proposal distribution must be previously defined. This scheme, in which multiple chains are generated independently, allows to parallelize the algorithm in the sense that each MCMC chain can be run independently (in parallel). The proposal distribution for the step size in the random walk is a Gaussian distribution. The covariance matrix Σ_m modulating the step is very similar to the TMCMC (*Ching & Chen*, 2007), but it has a scaling factor that depends on the acceptance rate R. If the acceptance rate is high then the step will be larger, allowing more exploration of the model's space. The covariance matrix, $\boldsymbol{\Sigma}_k$ is defined as,

$$\boldsymbol{\Sigma}_k = c_k{}^2 \sum_{i=1}^{N} \frac{\omega(\boldsymbol{\theta}_i)}{\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_i)} (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^T$$

$$\bar{\boldsymbol{\theta}} = \sum_{i=1}^{N} \frac{\omega(\boldsymbol{\theta}_i)\boldsymbol{\theta}_i}{\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_i)} \tag{2.13}$$

$$c_k = a + bR$$

where R is the acceptance rate of the Metropolis sampling and a and b are constant values chosen by performance tests.

- **Final:** Once $\beta = 1$ is reached, the distribution of the solutions space is obtained. The a posteriori probability density function of the parameters allows to quantify the error associated with the inversion. this could be represented, for example, through the sample covariance matrix of each distribution or by quantifying marginal errors from the sampled models.

The following table summarizes the CATMIP algorithm.

Algorithm 2.4 CATMIP algorithm. Table from *Minson* (2010)

---

1. Set $k = 0$. Generate N samples $\{\boldsymbol{\theta}_k\} = \{\boldsymbol{\theta}_1, ...\boldsymbol{\theta}_N\}$ of the prior pdf $f_0 = P(\boldsymbol{\theta})$
2. Set $k = k + 1$. Choose β_k such that the $COV[\omega]$ equals some target value
3. Calculate $\boldsymbol{\Sigma_k}$ and c_k
4. Draw N samples from $\{\boldsymbol{\theta}_{k-1}\}$ with probability ω. The set of resampled models is $\left\{\hat{\boldsymbol{\theta}}_{k-1}\right\}$
5. Use each resampled model $\hat{\boldsymbol{\theta}}_{k-1}$ in $\left\{\hat{\boldsymbol{\theta}}_{k-1}\right\}$ as the seed for generating N_{steps} models from the Metropolis algorithm with proposal density Σ_m
6. $\{\boldsymbol{\theta}_k\}$ is comprised of the final model from each Markov chain. Thus the total number of samples is unchanged.
7. Repeat steps 2 to 6 until $\beta_k = 1$ is reached

A diagram describing each major iteration of the algorithm is provided in Figure 2.4.

Until now, this chapter has explained why this scheme is adaptive, transitional and how it uses Metropolis in parallel. Another important aspect of CATMIP is that it allows to deal efficiently with problems with more than one set of data by using the concept of cascading, when certain conditions are met. The simplest case is when using two sets of data, where
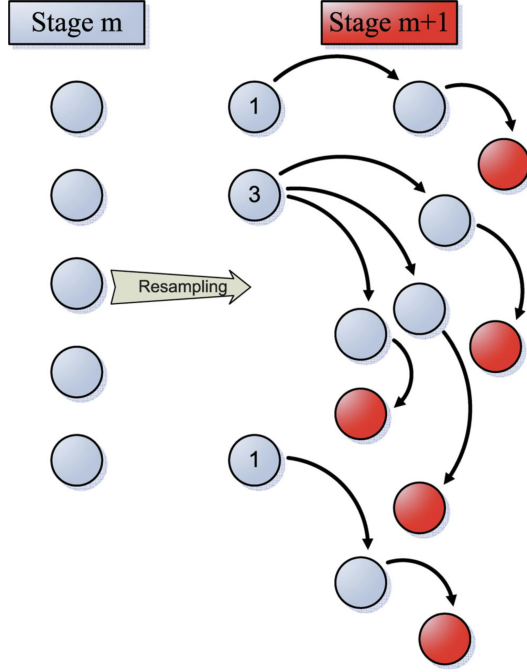
Figure 2.4: Diagram of the CATMIP algorithm. This cartoon illustrates one complete cooling stage of the CATMIP algorithm. The five samples from the previous iteration are resampled and then an instance of the Metropolis algorithm is run for each of the resulting samples. Numbers indicate the frequency of each model after resampling. The five red samples comprise the posterior distribution for the next iteration of the algorithm. The algorithm is plotted with very short Markov chains and a 100 percent acceptance rate for simplicity. In applications, the Markov chains would be much longer and the acceptance rate much lower. CATMIP scheme taken from *Minson et al.* (2013).

one set of data $\mathbf{D_1}$ depends only of one set of model parameters $\boldsymbol{\theta}_1$ and the other set of data $\mathbf{D_2}$ depends on both set of model parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. An example of that setting is the inversion of the source mechanism of an earthquake considering the static and the kinematic problem, where $\mathbf{D_1}$ and $\boldsymbol{\theta}_1$ are static crustal offsets and static slip amplitudes, respectively, and $\mathbf{D_2}$, $\boldsymbol{\theta}_2$ are seismograms and model parameters defining the kinematics of the source, such as rupture speed, rise time, etc.

To asses this problem the formulation of the posterior pdf can be expressed as a decomposition:

$$
\begin{aligned}
P(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{D_1}, \mathbf{D_2}) &\propto P(\mathbf{D_1}, \mathbf{D_2} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) P(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&\propto P(\mathbf{D_1} | \theta_1) P(\mathbf{D_2}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) P(\boldsymbol{\theta}_1) P(\boldsymbol{\theta}_2) \\
&\propto P(\mathbf{D_1} | \boldsymbol{\theta}_1) P(\boldsymbol{\theta}_1) P(\mathbf{D_2}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) P(\boldsymbol{\theta}_2) \\
&\propto P(\boldsymbol{\theta}_1 | \mathbf{D_1}) P(\mathbf{D_2} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) P(\boldsymbol{\theta}_2).
\end{aligned}
\tag{2.14}
$$

The above expression is proportional to $P(\boldsymbol{\theta}_1 | \mathbf{D_1})$ the posterior pdf of the simpler problem. In this manner the problem can be assessed by solving first a simpler problem and expanding later the problem to add more data and calculate the additional parameters.

The algorithm implements this procedure using the transitional scheme in the following

16

manner:

$$P(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{D_1}, \mathbf{D_2})_{m,n} \propto P(\boldsymbol{\theta}_1, \mathbf{D_1})_m P(\mathbf{D_2} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^{\alpha_n} P(\boldsymbol{\theta}_2),$$
$$0 \leq \alpha_n \leq 1, \tag{2.15}$$

where $P(\boldsymbol{\theta}_1 | \mathbf{D_1})_m$ is an intermediate distribution of the single problem. This expression means that the pdf obtained from the simplest problem is used to get the pdf of the complete problem. Then, the iterative part of this problem can be divided in two steps, as algorithm 2.5 shows.

Algorithm 2.5 Cascading in CATMIP algorithm . Table from *Minson* (2010)

---
1. Sample $P(\boldsymbol{\theta} | \mathbf{D})_m \propto P(\boldsymbol{\theta}_1) P(\mathbf{D_1} | \boldsymbol{\theta}_1)_m, 0 < \beta_m < 1$
2. Sample $P(\boldsymbol{\theta} | \mathbf{D})_n \propto P(\boldsymbol{\theta}_1) P(\boldsymbol{\theta}_2) P(\mathbf{D_1} | \boldsymbol{\theta}_1) P(\mathbf{D_2} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^{\alpha_n}, 0 < \alpha_n < 1$

2.2 Understanding the CATMIP Algorithm

This section illustrates how the CATMIP algorithm works. With that purpose we implemented a serial version of the algorithm using Python. First, we show an example in 1D, which makes possible to visualize the effect of each stage on the distribution and the advantages of CATMIP over more classical sampling scheme. Then, the algorithm will be applied over a four-dimensional inverse problem, concretely, to find the hypocenter of an earthquake.

2.2.1 One Dimensional Tests

The CATMIP algorithm (Cascading Adaptive Tempered Metropolis in Parallel) is a Bayesian method designed to solve inverse problems in an accurate and efficient form. It applies especially well when the posterior probability distribution of the parameters is difficult and the number of parameters is big. The kind of difficulty that can appear at the moment of solving a problem is the presence of extreme characteristics, as multiplicity of local maximums, very peaked maximums or flat regions with very low probability.

In this section, we will illustrate graphically some key aspects of CATMIP by producing samples of a known probability density function pdf, that includes some pathological characteristics as explained in the following.

Recalling Bayes' theorem, the posterior distribution has the form seen in equation 2.16. In this equation, the prior probability distribution is $P(\boldsymbol{\theta})$ and the likelihood function is $P(\mathbf{D} | \boldsymbol{\theta})$.

$$f(\boldsymbol{\theta} | \mathbf{D}) = \gamma f_{prior}(\boldsymbol{\theta}) L(\mathbf{D} | \boldsymbol{\theta}), \tag{2.16}$$

The prior information considered is a uniform distribution for $P(\theta)$ in a given interval for θ. The likelihood function $P(D | \theta)$ is chosen as a 1D function with two local maxima, as a sum of a Gaussian and a Laplace function:

$$f(\theta|\mu, b) = \alpha_1 e^{-(\frac{\theta - \mu_1}{2\sigma_1})^2} + \frac{1}{2b} e^{-|\theta - \mu|/b}. \tag{2.17}$$

The aim is to sample the posterior distribution illustrating the effect of the processing in the different stages of the algorithm. Here we consider the values of parameters α_1, μ_1, α_2, μ_2, that can produce a posteriori pdf's that can have regions of very low probability between local maxima.

Metropolis-Hastings (Sequential and Parallel)

To understand the role of Metropolis-Hastings in Parallel algorithm in CATMIP, it is useful to start from the Metropolis algorithm itself, reminding that it works through sampling by Markov Chains. It generates a chain through successive steps, where the starting point to generate the next candidate model to be sample of $P(\boldsymbol{\theta}|\mathbf{D})$ is the last accepted sampled model. The size of this step follows a proposal distribution, usually a Gaussian distribution. If the size of the step is too small, the process tends to obtain samples that are very close from the last one. This makes easy to clusterize samples around a local maximum, as we can see in figure 2.5 in an application of the classical Metropolis algorithm. Here, most of the sampled models (if not all) will lie around to the local maxima of the pdf $P(\theta|D)$ that is closer to the initial model if the average size of the step implied by the proposal distribution is small compared with the length of the valley of low probability between the two local maxima of $P(\theta|D)$ (see figure 2.5).

(a) (b)
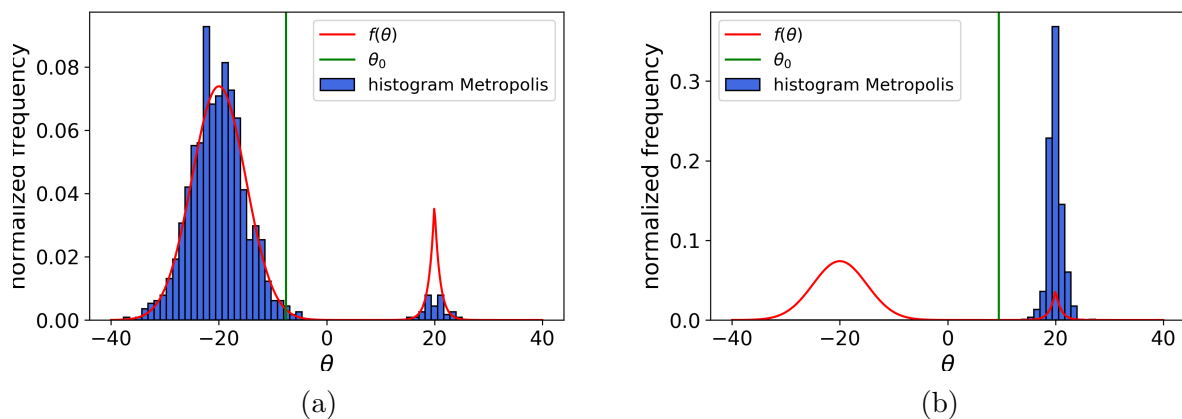
Figure 2.5: Histograms showing the posterior pdfs obtained in two different runs of the Metropolis algorithm considering different initial models when the distribution has multiple local maxima. If the size of the step is small, samples tend to cluster in one of them. This clusterization can be partially solved by increasing the number of elements in the chain or growing the step for the proposal distribution. In this example the number of elements in the chain is $N = 1000$ where the realization is clustering the sampling near a) the Gaussian b) the Laplacian function. Green vertical lines indicate initial θ model for MH algorithm.

The problem of exploring samples only near a local maxima of $P(\theta|D)$ problem can be addressed by increasing the number of elements in the chain or increasing the size of the

steps, but at the potential cost of drastically reducing the acceptance rate of the MCMC chain. Another way to solve this problem is applying a parallel scheme to the algorithm.
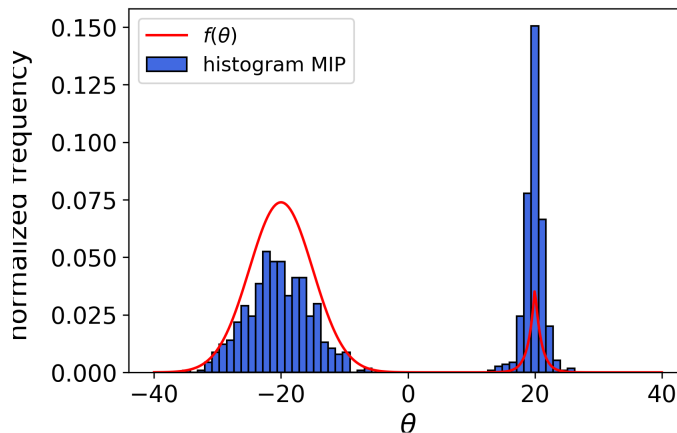
(a)

Figure 2.6: Posterior distribution obtained applying the Metropolis algorithm in Parallel (MIP) to sample the synthetic 1D pdf. This example uses 1000 chains, each one generated 100 steps where only the final sample is saved.

When it is applied, Metropolis in parallel distributes samples in both maxima, but the proportion is not necessarily balanced. Figure 2.6 shows that the peaked function in the right side of the graph has more samples. This can be explained thinking in terms of the Markov chains, that at the begining are distributed almost uniformly and then, when new steps are tested, the samples move to regions with higher probability. This process is very sensitive to the width of the region associated to each other local maxima, since more chains start in these regions, and it is more probable that the chains remain in that neighborhood. But if the function is very peaked, once a sample is close to the peak it is difficult to escape from that region. A test sample that moves away from the peak makes very difficult to accept that sample as part of the chain, because the probability of getting away from the maximum is proportional to the ratio between the likelihood of the test and the current sample. One way to solve the problem of unbalanced samples is applying resampling, a process that generates a set of samples from an existing set, repeating more times the samples with more likelihood, given by $P(\theta|D)$.

Resampling

The resampling is a process that uses a set of samples of a given pdf as a repository to sample a different pdf, repeating more often the samples that presents higher pdf value according to the new pdf, and eventually discarding those with lower pdf values. In order to set the expected frequency of appearance of each sample in the new distribution it is necessary to assign a weight associated to each sample. This weight is evaluated from the ratio between the current pdf value and the value of previous pdf. When we incorporate a resampling previous to Metropolis-Hastings in Parallel, the obtained graph is the following (Figure 2.7).

(a) without resampling   (b) with resampling
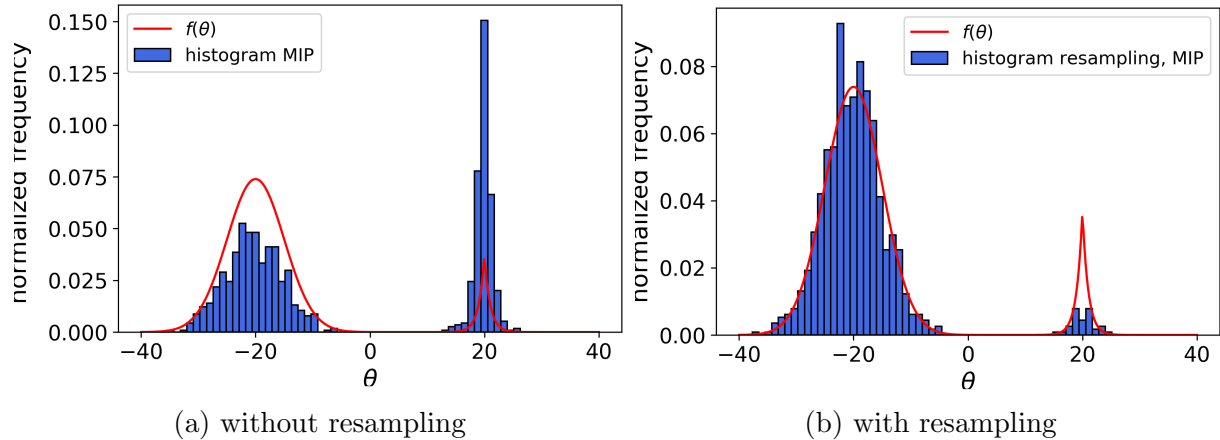
Figure 2.7: Histograms showing the posterior probability distribution obtained applying a) Metropolis in Parallel, b) Resampling and Metropolis in Parallel (MIP).

The resampling redistributes the samples according to their probability in the posterior distribution.

Figure 2.7 shows that incorporating this process to the studied example, the resulting ensemble of samples is such that its histogram (normalized) represents much better the target a posteriori pdf. It is important to note that the resampling is applied before Metropolis in parallel, as the CATMIP algorithm state.

A relevant aspect that can be noticed from the figure is that the sample set histogram delineates much better the Gaussian than the Laplacian function.

An intuitive explanation for this phenomenon can be given in terms of the probability that a sample falls in the vicinity of a rough peak of the pdf. Even when the process of resampling generates more samples where the samples have more likelihood, the probability that they reach the most acute point is lower. An improvement to the previously showed stages is to apply a transitional technique to gradually move samples to those regions.

Transition

The transitional scheme from a priori to a posteriori pdf described in section 2.1.4 refines the preceding Bayesian methods by changing gradually the sampled distribution from a prior distribution to the posterior distribution in a sequence of intermediate steps, that produce intermediate posterior pdf's.

Incorporating the transitional scheme to the stages previously showed we obtain the structure of the CATMIP algorithm (without cascading, which applies when the inversion considers more than one set of data). When we implement the transitional procedure, the resampling and the Metropolis-Hastings sampling are applied in a sequence of intermediate steps, where an intermediate pdf is sampled from the models obtained in the previous step. The obtained ensemble of models for the 1D study case incorporating transition in the form of logarithmic increasing variable β is shown in the Figure 2.8.

(a) MIP

(b) Resampling and MIP
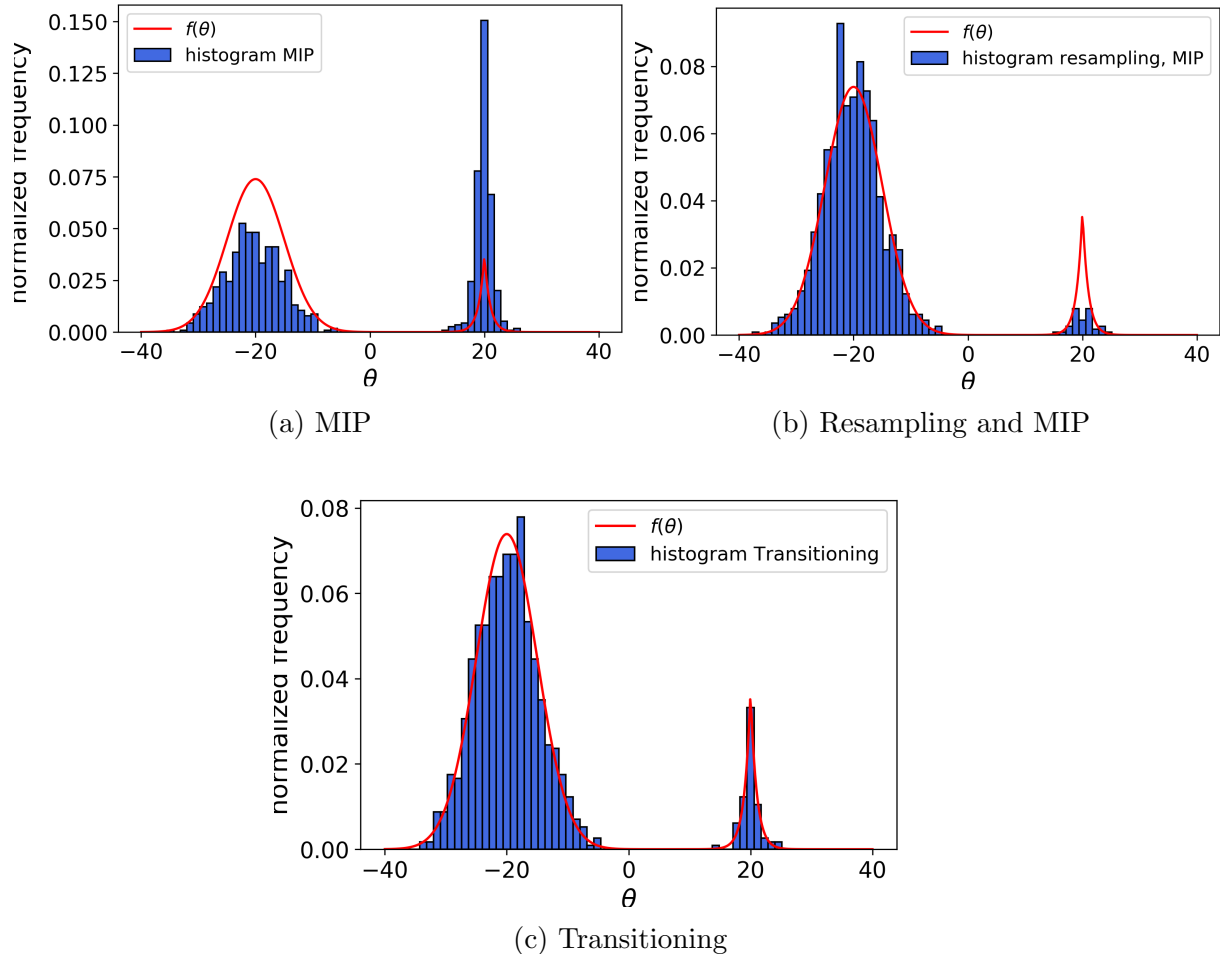
(c) Transitioning

Figure 2.8: Comparison between posterior probability distributions obtained applying different sampling schemes. Transitional scheme is compared with MIP (Metropolis-Hastings in Parallel) and MIP adding resampling. The number of samples is $N = 1000$, in every case, while the number of steps in MCMC is $N_{mcmc} = 100$.

CATMIP

The CATMIP sampling algorithm use the transitional scheme in an adaptive form, such that the distribution of samples change gradually (equation 2.10) and efficiently. Here, we incorporate the adaptive transition to the Resampling and Metropolis-Hastings in Parallel on the resolution of the 1D synthetic problem. The result is shown in figure 2.9, where the intermediate pdfs obtained through the CATMIP algorithm are shown. The prior distribution was chosen as a uniform $\beta = 0$ and then the algorithm calculates in an adaptive form the next β values to generate the intermediate distributions. Finally the posterior pdf is reproduced when $\beta = 1$. The cascading part of the algorithm is not tested in these examples because only one data set was used.
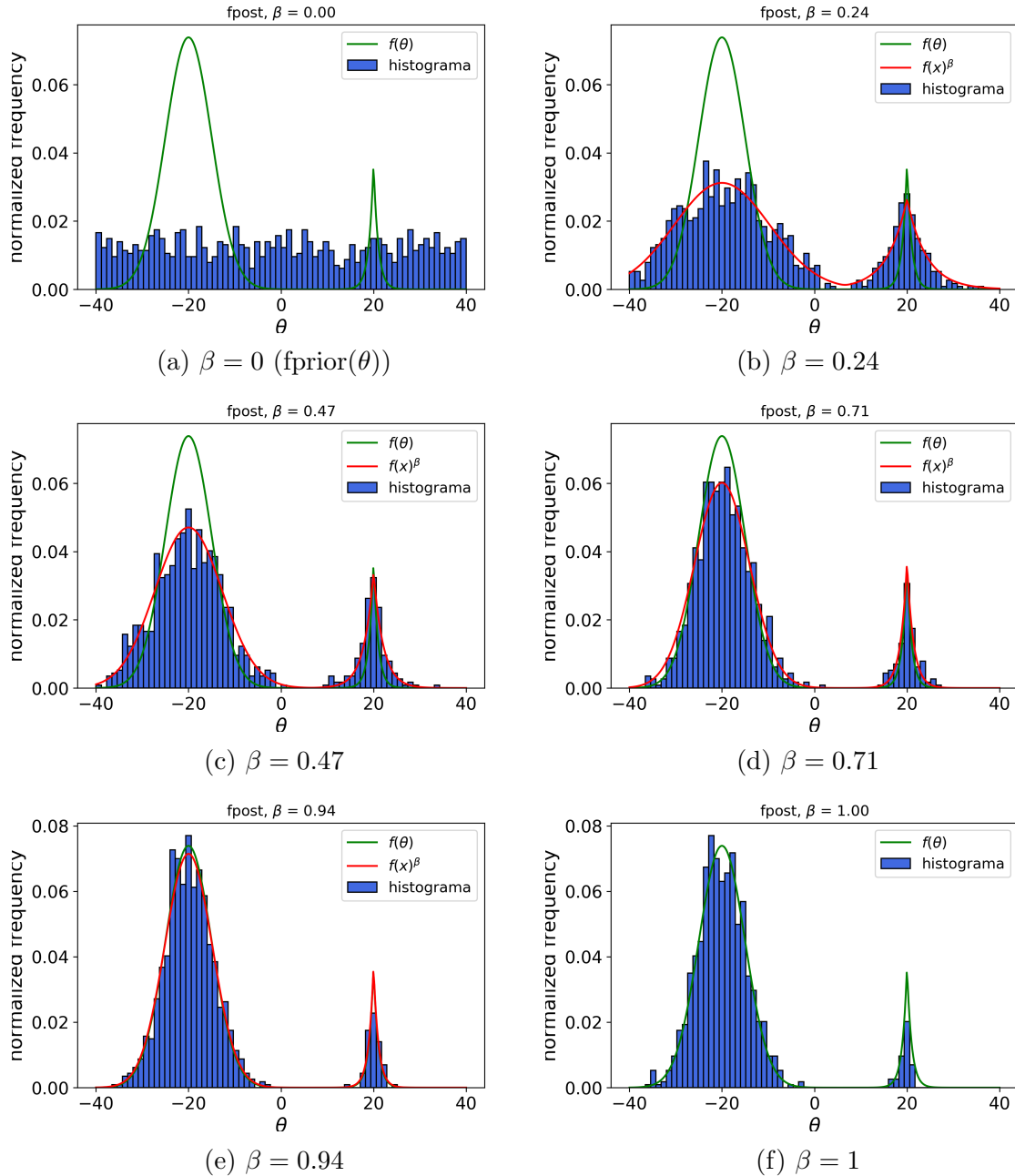
Figure 2.9: Posterior probability distribution obtained applying the CATMIP sampling algorithm to a 1D synthetic problem for different β values. (1000 samples, 100 steps MCMC)

When we use the CATMIP algorithm, or in general any sampling method, for the 1D synthetic study case with a larger number of samples, we obtain figure 2.10. A larger number of samples translates in achieving a better representation of the intermediate pdf's as well the posterior pdf, in the sense that the histogram of the set of samples resembles in a much better way the pdf of the distribution being sampled. Thus, by a process of trial and error, one should determine the best number of samples by analyzing the stability of the marginal histograms that they produce.
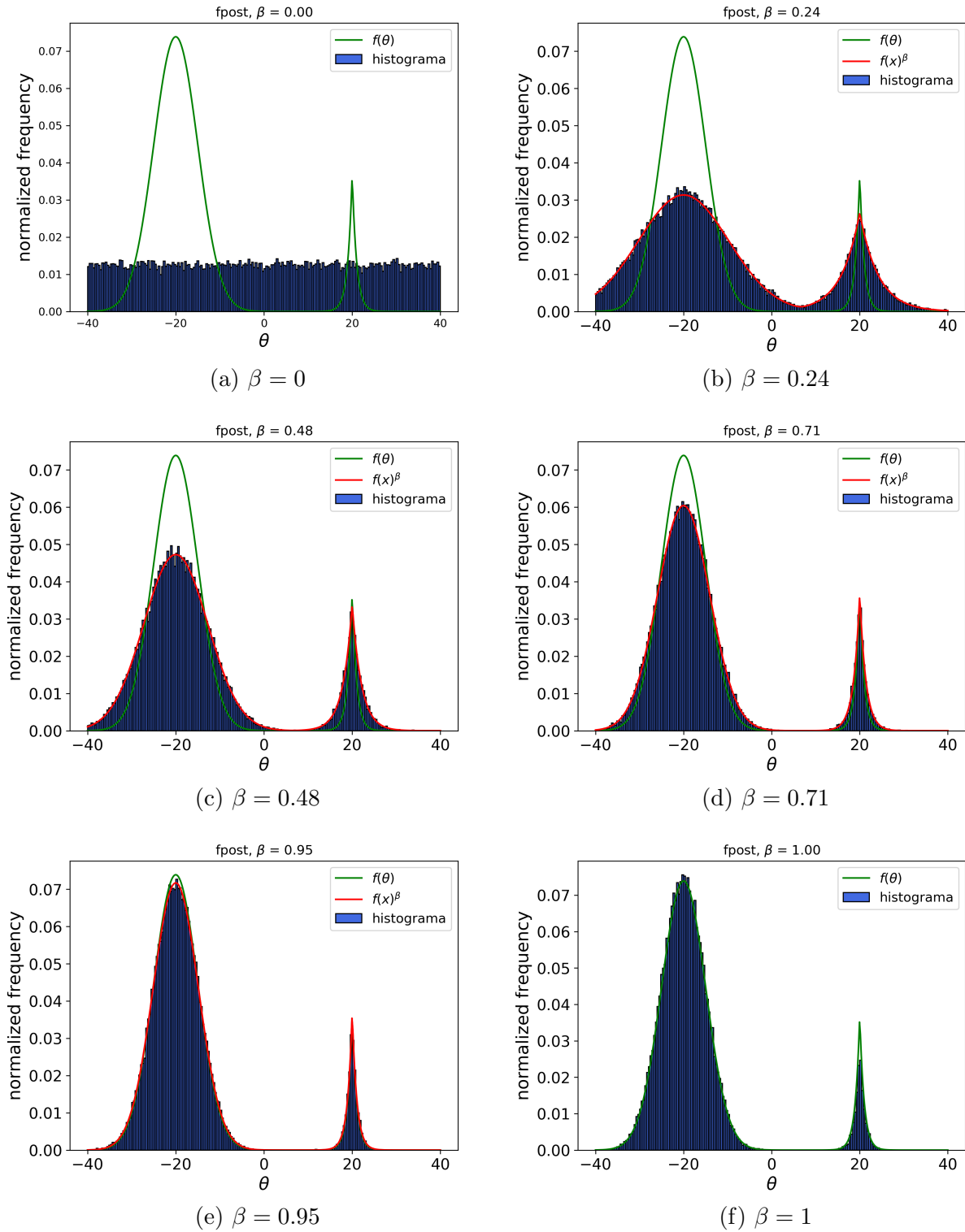
Figure 2.10: Sequence of distributions obtained through the CATMIP algorithm applied to a 1D synthetic problem. The function shown corresponds to the sum of a Gaussian and a Laplace function. The number of samples is 50000 and β goes from 0 to 1.

2.2.2 Test in 4D: Hypocenter location of an earthquake

As part of the initial tests of our implementation of the CATMIP algorithm, we chose a problem with only a few parameters, which allows to illustrate the way this method works. We analyze here the inversion of the hypocenter of an earthquake, which is a non-linear problem where the parameters to be inverted are those that define the hypocenter, that is, the three spatial coordinates and the origin time. The observations used to solve this problem are the arrival times of the P and S waves. The formulation of this problem is in terms of the travel times of the seismic waves, from the source to receivers which is the sum of travel times on paths in different propagation media. The earthquake location problem can be modelled using complex velocity models from 3D tomography, however, we will analize the case of a simple homogeneous halfspace, as the goal is to illustrate in a simple form the inversion methodology and compare CATMIP with other methods. The forward model that relates the hypocenter (E_h, N_h, Z_h, T_h) with the travel times T_p, T_s is described in equations 2.18 and 2.19.

$$T_p^{station} = T_h + \frac{\sqrt{(E_h - E^{station})^2 + (N_h - N^{station})^2 + (Z_h - Z^{station})^2)}}{v_p} \tag{2.18}$$

$$T_s^{station} = T_h + \frac{\sqrt{(E_h - E^{station})^2 + (N_h - N^{station})^2 + (Z_h - Z^{station})^2)}}{v_s}, \tag{2.19}$$

where (E_h, N_h, Z_h) and $(E^{station}, N^{station}, Z^{station})$ denote coordinates of the source and stations, respectively, T_h corresponds to origin time, v_p and v_s are velocities of p and s waves.

Test with synthetic data: Circular network of seismometers

This example is the synthetic case in which we locate a circular array of stations whose center has the location of the epicenter of a seismic event (see Figure 2.11), with the source at a depth of 30 km and that for the inversion we use only the P wave arrival time to determine the hypocenter location. Because of the symmetry of this problem it has infinite solutions along the z axis (having a trade-off between source depth and origin time). A particularity of this problem, is that the region of maximum likelihood (where admissible solutions are) spreads with depth due to observational uncertainties (we assumed an error of 0.5 seconds for the P - wave arrival time). Also, we note the peaked form of the spatial distribution of feasible hypocenters as we approach to the true hypocenter used to generate synthetic data that constraints our models during the inversion. These conditions make this synthetic example a good study case to compare the results using different sampling algorithms: Metropolis-Hastings, Metropolis-Hastings in Parallel (MIP) and CATMIP.

The obtained ensembles of samples resulting from the hypocenter determination inverse problem using Metropolis-Hastings, Metropolis-Hastings in Parallel (MIP) and CATMIP are shown in figure 2.12 a, b and c respectively. In this figure, all the presented solutions have the same amount of samples ($N = 1000$ samples), so we can compare the obtained distributions in

similar conditions. The expected behavior is a peaked distribution along the z-axis, spreading in depth as we move away from origin z-coordinate. The solution obtained using Metropolis-Hastings for that amount of samples does not represent the expected distribution of the parameter models. The peaked behavior near the origin of the event is better recovered when we use CATMIP, as well as the conical shaped form of the distribution representing the uncertainties on hypocenter location.

The marginal distributions obtained using Metropolis-Hastings, Metropolis-Hastings in Parallel and CATMIP are shown in figure 2.13, 2.14 and 2.15, respectively.

A relevant aspect that can be noticed from figures 2.13, 2.14 and 2.15 is the dependence between the vertical component of hypocenter location and the origin time. This dependence, which is the expected behavior of the solution, is poorly recovered using the sequential Methopolis-Hastings algorithm, and better described by the results obtained using Metropolis-Hastings in Parallel and even much better with CATMIP (for the same amount of samples). A similar observation can be done in terms of the quality of the recovered marginal histograms for each model parameter, in which CATMIP does a better job by obtaining smoother or less noisy histograms for the given discretization size used to compute each one of them.
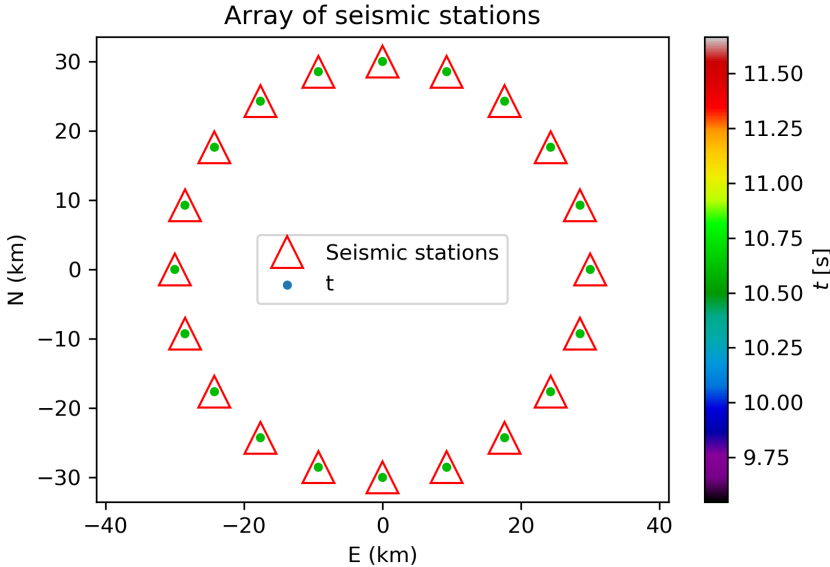
Figure 2.11: Circular array of stations centered at the epicenter of the synthetic event. Colored dots indicate time of arrival of P waves, relative to hypocenter time. Note that for this special case, all stations have the same arrival time of P-waves.
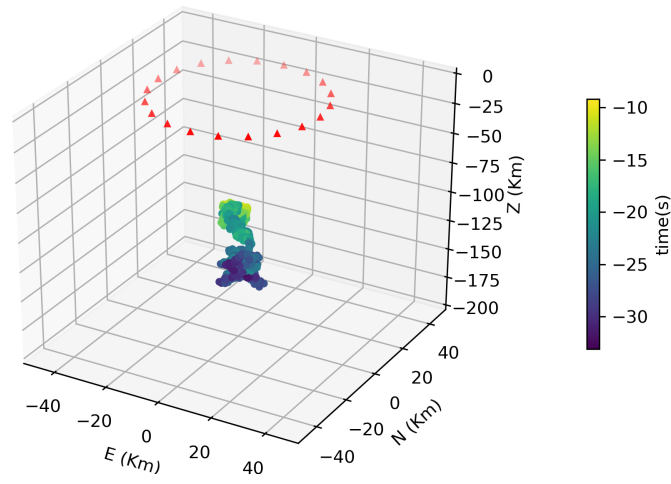
(a) Metropolis
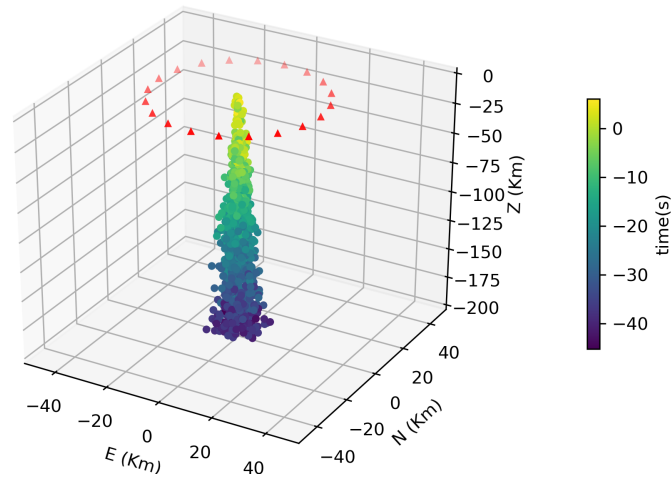
(b) Metropolis in Parallel
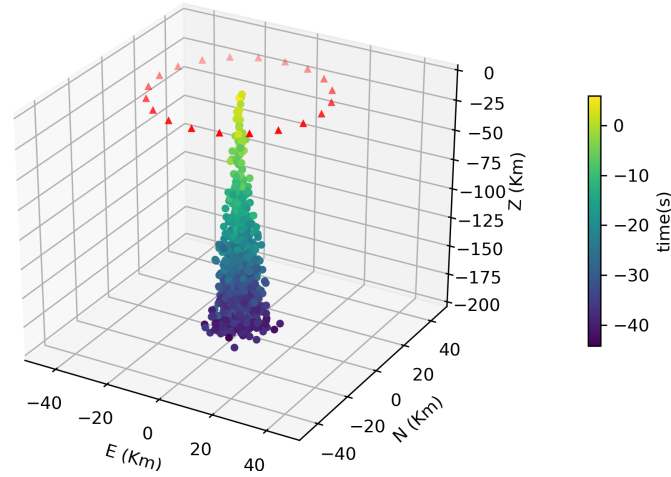
(c) CATMIP

Figure 2.12: 3D plot of hypocentral location for all sampled models using the Metropolis-Hastings, MIP and CATMIP sampling algorithm. Color shows origin time associated to each sample. The samples correspond to the $N = 1000$, in the case of Metropolis-Hastings these are the final samples of a single Markov chain, while in MIP and CATMIP the resulting samples corresponds to $N = 1000$ independent chains, where the number of steps is $N_{mcmc} = 100$.
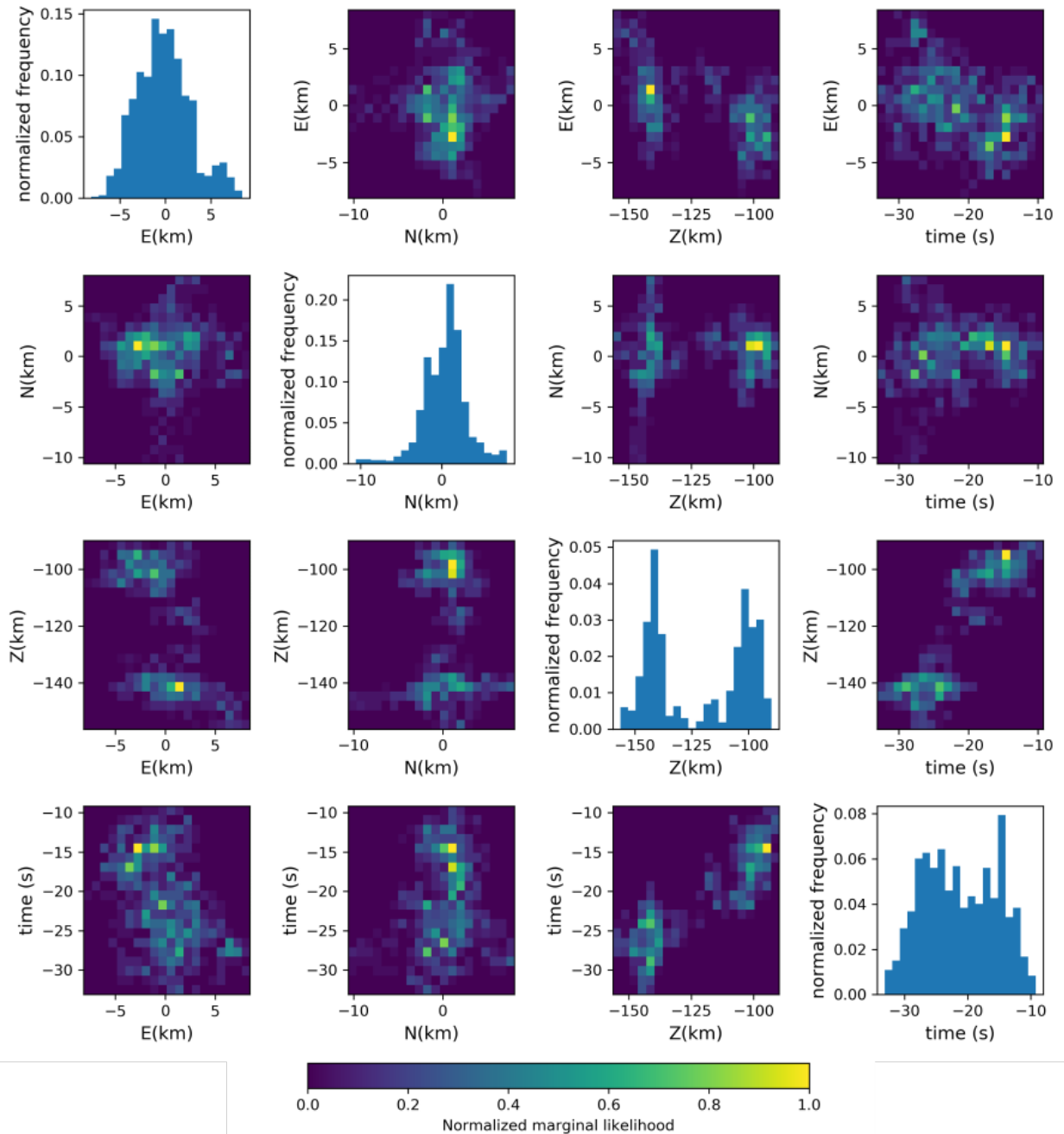
Figure 2.13: Marginal distributions of the obtained sampled models for the hypocentral determination problem using the Metropolis-Hastings algorithm. The parameters shown are E, N, Z coordinates (in km) and origin time T [s]. Diagonal elements show marginal 1D histograms corresponding to each parameter, and out of diagonal elements show the 2D histograms associated to parameter pairs. The samples are the final 1000 steps of a Markov chain of $N_{mcmc} = 5000$.
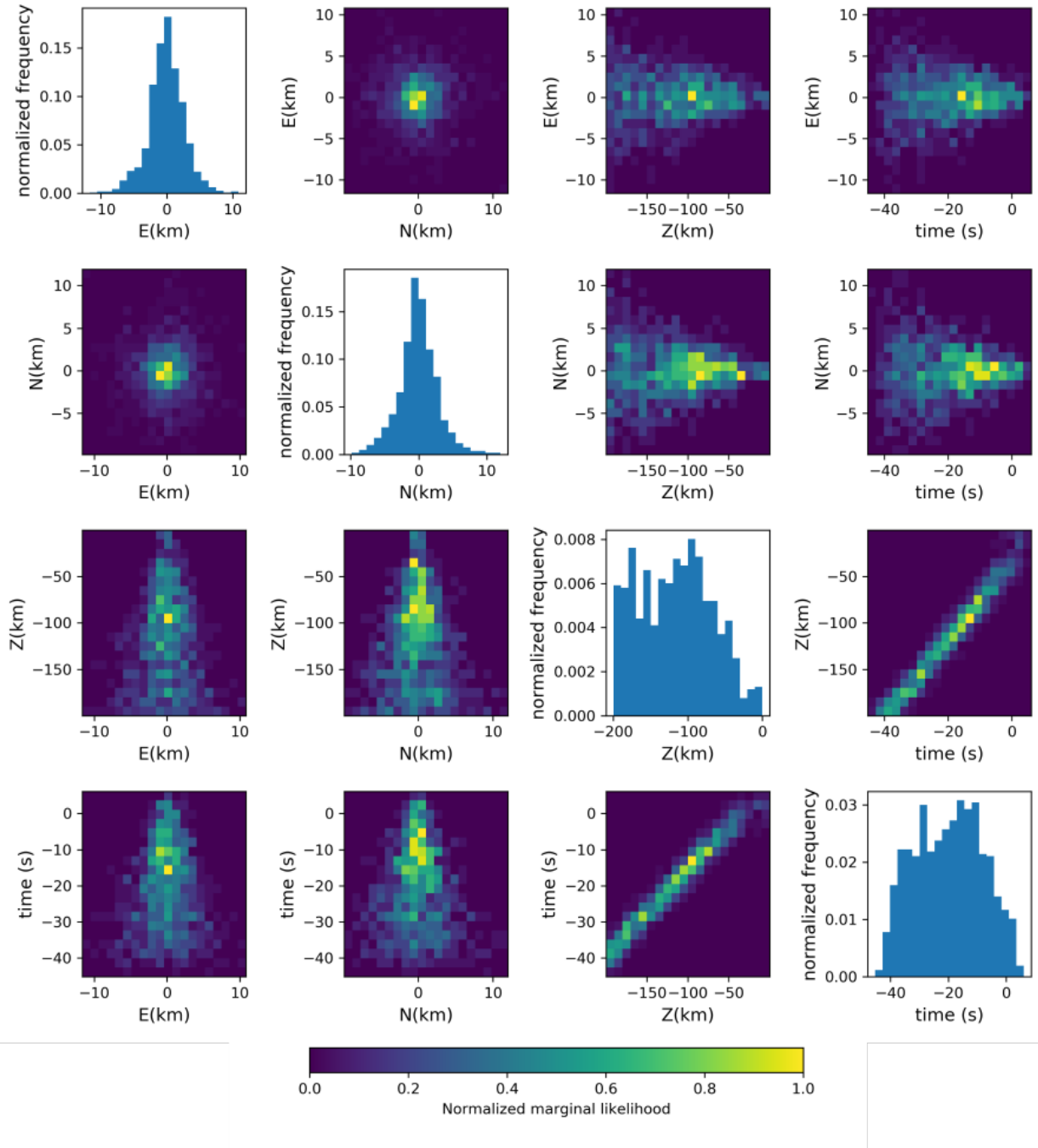
Figure 2.14: Marginal distributions of the obtained sampled models for the hypocentral determination problem using the Metropolis-Hastings in Parallel algorithm. The parameters shown are E, N, Z coordinates (in km) and origin time T [s]. Diagonal elements show marginal 1D histograms corresponding to each parameter, and out of diagonal elements show the 2D histograms associated to parameter pairs. The number of samples employed was $N = 1000$, where each Markov chain has $N_{mcmc} = 500$ steps.
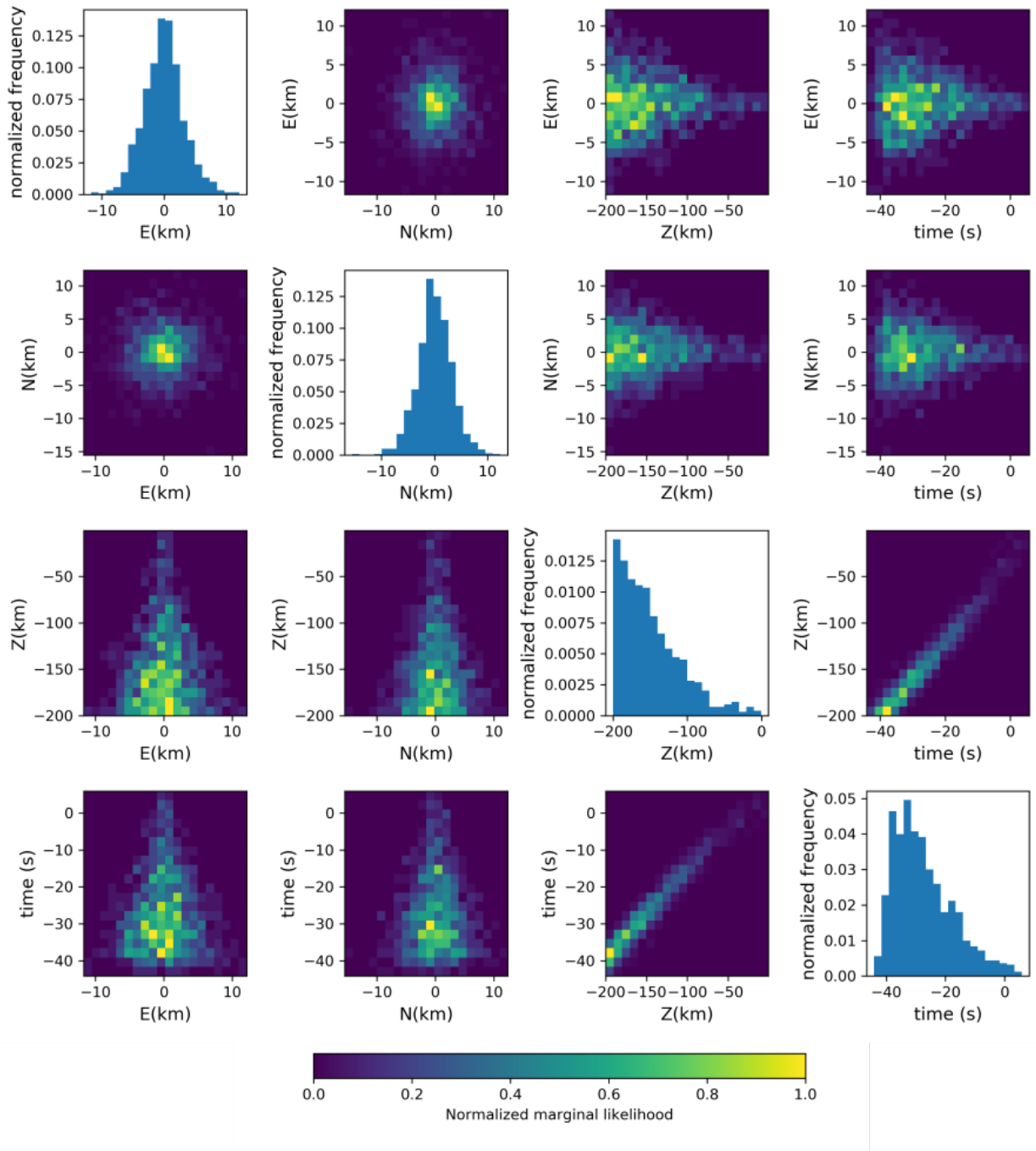
Figure 2.15: Histograms showing marginal pdfs (diagonal elements) and correlation between parameters (out of diagonal elements) for the hypocentral determination problem using CATMIP. The parameters are the components E, N, Z, and the time of occurrence of the earthquake. The number of samples employed was $N = 1000$, where each Markov chain has $N_{mcmc} = 500$ steps.

Test with real data: Intermediate depth event following the Pisagua earthquake.

This example used real data coming from a seismic event within a series of earthquakes registered after the 2014 (Mw 8.2) Pisagua earthquake (**?**, . Diana Comte, personal communication)). The observations used to solve this problem are the arrival times of the P and S waves at several seismic stations, whose location is indicated in Figure 2.16. Here, we assume an elastic homogeneous halfspace with $v_p = 4$ km/s and $v_s = 2.2$ km/s.
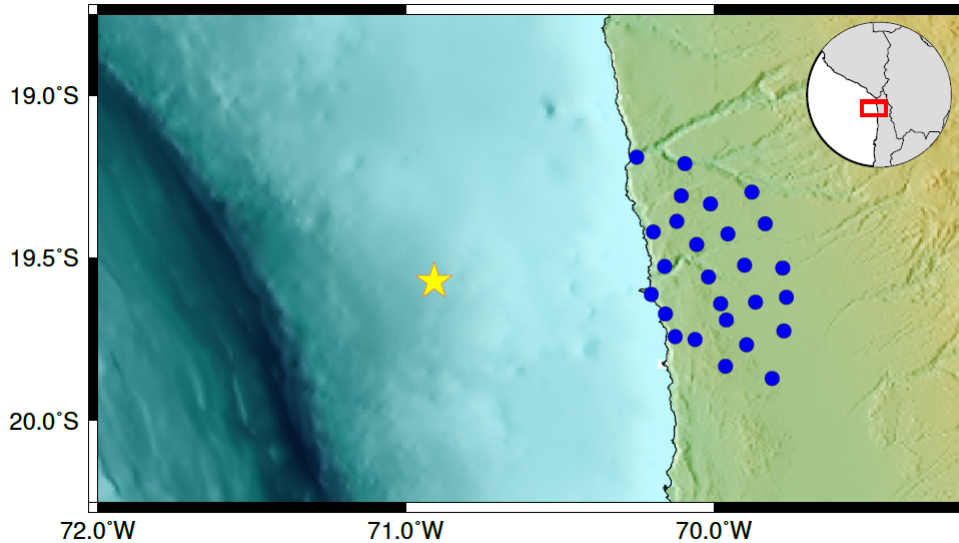
Figure 2.16: Distribution of seismic stations used for the inversion of the hypocenter location for the aftershock of the Pisagua earthquake.

The result when we use CATMIP algorithm can be seen in Figure 2.17 where the model parameters obtained are an ensemble of hypocenter coordinates that better fits the arrival times. The diagonal boxes of the figure shows 1D histograms representing the marginal distributions of the model parameters, and the off diagonal boxes are 2D histograms showing the correlations between pairs of model parameters, all constructed from the set of sampled models. This illustrates how Bayesian methods allow to characterize the solution, where relevant characteristics can be obtained, for example, the position of the center, degree of asymmetry, the size of uncertainty bounds, and the correlation and dependence between model parameters (*Tarantola*, 2005).
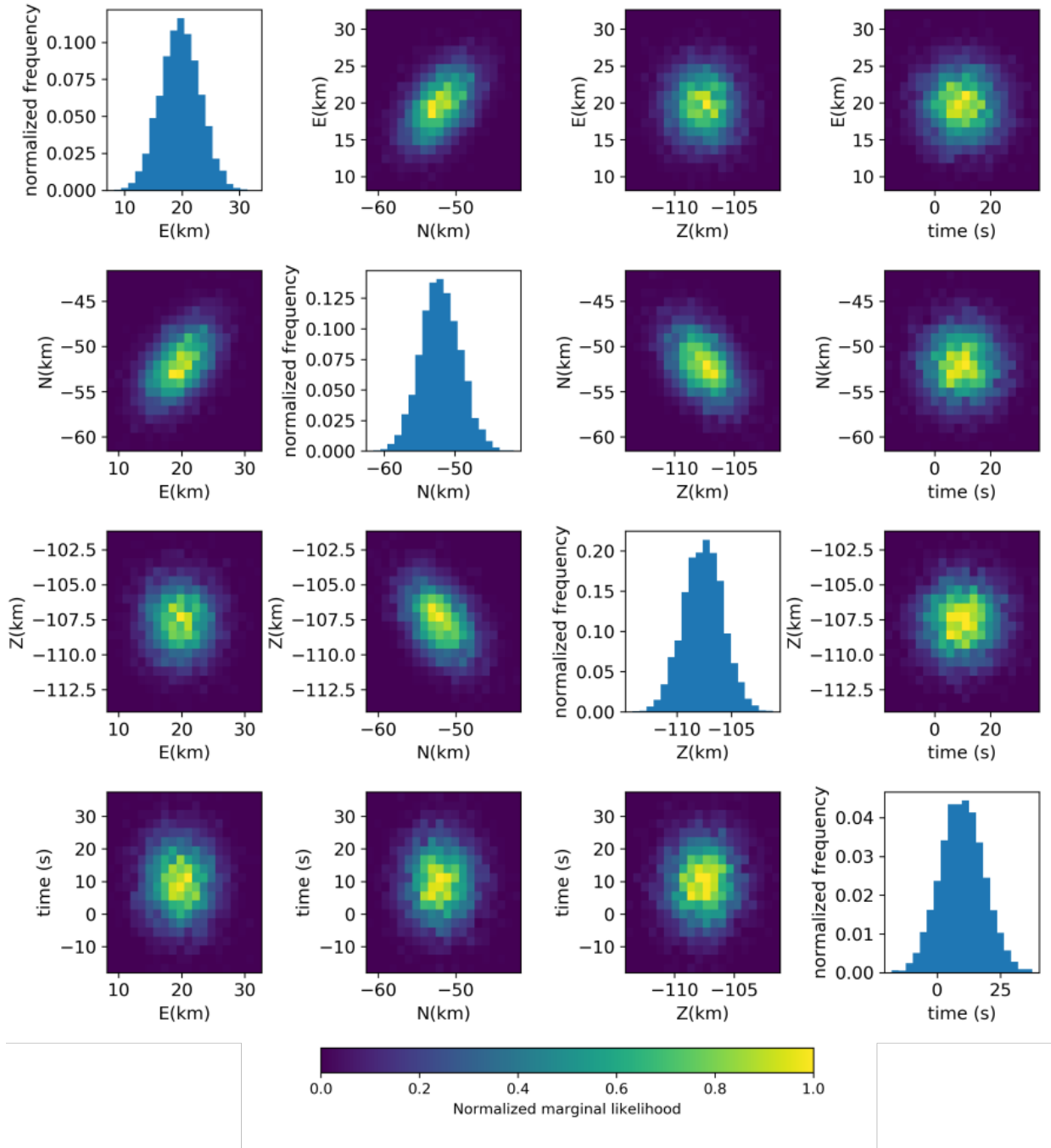
Figure 2.17: Histograms showing marginals (diagonal elements) and correlation between parameters (out of diagonal elements) resulting from the inversion of real data for an aftershock of the Pisagua earthquake using CATMIP. The parameters are the components E, N, Z, and the time of ocurrence of the earthquake. The number of samples employed is $N = 5000$ and the number of steps of the Markov chains is $N_{mcmc} = 500$.

Although, this particular example does not show the full potential of this method (the distribution does not show pathological characteristics, and we did not incorporate additional sources of error), it allows to illustrate some key points of the Bayesian formulation of the inverse problem that will be used in the next chapters. For instance, to better understand

model parameter uncertainties, and statistical dependence and/or correlations between them.

A limitation of this method and the sampling methods in general, is to determine aspects as the convergence of the solution, especially when different number of samples are used. For example, the inversion of the aftershock of the Pisagua earthquake used 5.000 samples and the distribution changed very little when 10.000 samples were employed, thus ensuring that 5.000 samples is enough to describe the statistics of the posterior distribution of model parameters. Such a criteria will used in this thesis to test that the main characteristics of the solution of the inverse problem.

Chapter 3

The 2007 (M_w 7.7) Tocopilla Earthquake

3.1 Seismotectonic context

The Nazca-South America subduction margin is one of the zones with the highest seismicity in the world. The north of Chile, located in this margin, is characterized by a high convergence rate, with a value close to 6.7 cm/y (*Kendrick et al.*, 2003). Consequently, that zone presents a high seismicity rate, that includes the occurrence of large magnitude earthquakes (e.g., *Ide*, 2013), such as the 1877 event, which ruptured a segment more than 500 km long, with a magnitude close to Mw 8.8 (*Comte & Pardo*, 1991; *Ide*, 2013). After the Tocopilla earthquake many smaller events have occurred in that region, most of them with magnitude lower than Mw 7 and a few with magnitude higher than Mw 7.4 (*Béjar-Pizarro et al.*, 2010; *Ruiz & Madariaga*, 2018), corresponding to the Tocopilla (Mw 7.7) 2007, the Pisagua (Mw 8.2) 2014, and its largest aftershock (Mw 7.7) of 2014. The events occurred after 1877 have not released the energy potentially accumulated in the region, one of the reasons why that zone has been characterized as a seismic gap (e.g. *Duputel et al.*, 2015), therefore do not significantly reduce concerns for continious seismic hazard in the region.

The Tocopilla earthquake spreads as far as the southern edge of the seismic gap at the Mejillones Peninsula, a zone that apparently acts as a seismic barrier (*Béjar-Pizarro et al.*, 2010; *Victor et al.*, 2011). Two more events have occurred in the central segment of the 1877 seismic gap, the 2014 (Mw 8.2) Pisagua earthquake and its largest aftershock with magnitude Mw 7.7 (*Duputel et al.*, 2015). None of the earthquakes occurred in the region have reached the shallow part near the trench.

Before the 1877 event, in 1868, another large earthquake of similar magnitude occurred in southern Peru, northwards of the region of 1877 event. Defining a seismic gap that remained unbroken until 2004 when the Mw 8.4 Arequipa earthquake occurred, rupturing the northern region of the 1868 seismic gap.

Outside and just south of the 1877 seismic gap, the last big event was the 1995 (Mw 8.1) Antofagasta earthquake (*Pritchard et al.*, 2002). Here for both, 1877 and 1995 events, the Mejillones Peninsula acted as a barrier for the northward rupture propagation of the 1995

and southwards propagation of the 1877 event. Thus, strengthening the idea of the Mejillones peninsula behaving as a seismic barrier for rupture propagation (*Victor et al.*, 2011; *Song & Simons*, 2003) of big earthquakes in that zone.

The continental wedge geometry on the Chilean margin shows a latitudinal segmentation at the Mejillones peninsula. To a first order differentiating an erosive and accrecionary margin north and south of the peninsula, respectively. The erosive segment, where the Tocopilla earthquake occurred, presents low sedimentation and, consequently, the absence of an accretionary prism, in addition to the presence of normal faults and a steep slope (*Maksymowiczk*, 2015).

The region of the subduction megathrust in front of the Mejillones Peninsula shows low coupling inferred from geodesy (*Métois et al.*, 2013), which is consistent with studies indicating the occurrence of seismic slip in such region after the 1995 Antofagasta earthquake. (*Pritchard & Simons*, 2006).

The seismogenic behavior under the Mejillones peninsula is controlled by the presence of fluids (*Pasten-Araya et al.*, 2018). The Nazca plate in this portion has fractured zones, whose permeability allows the transportation of water, which in turn modifies the rheologic properties in the subduction interface. The northern zone of the Mejillones peninsula has a specially higher concentration of fluids, a fact that can be explained as a consequence of the subduction of the Mejillones fracture zone (*Pasten-Araya et al.*, 2018). Figure 3.1, from *Pasten-Araya et al.* (2018), shows the scheme of these fracture zones. Also, in the north and central portion of the Mejillones peninsula the seismicity observed is higher, characterized by the presence of slow earthquakes and repeaters (*Poli et al.*, 2017; *Ruiz et al.*, 2014).
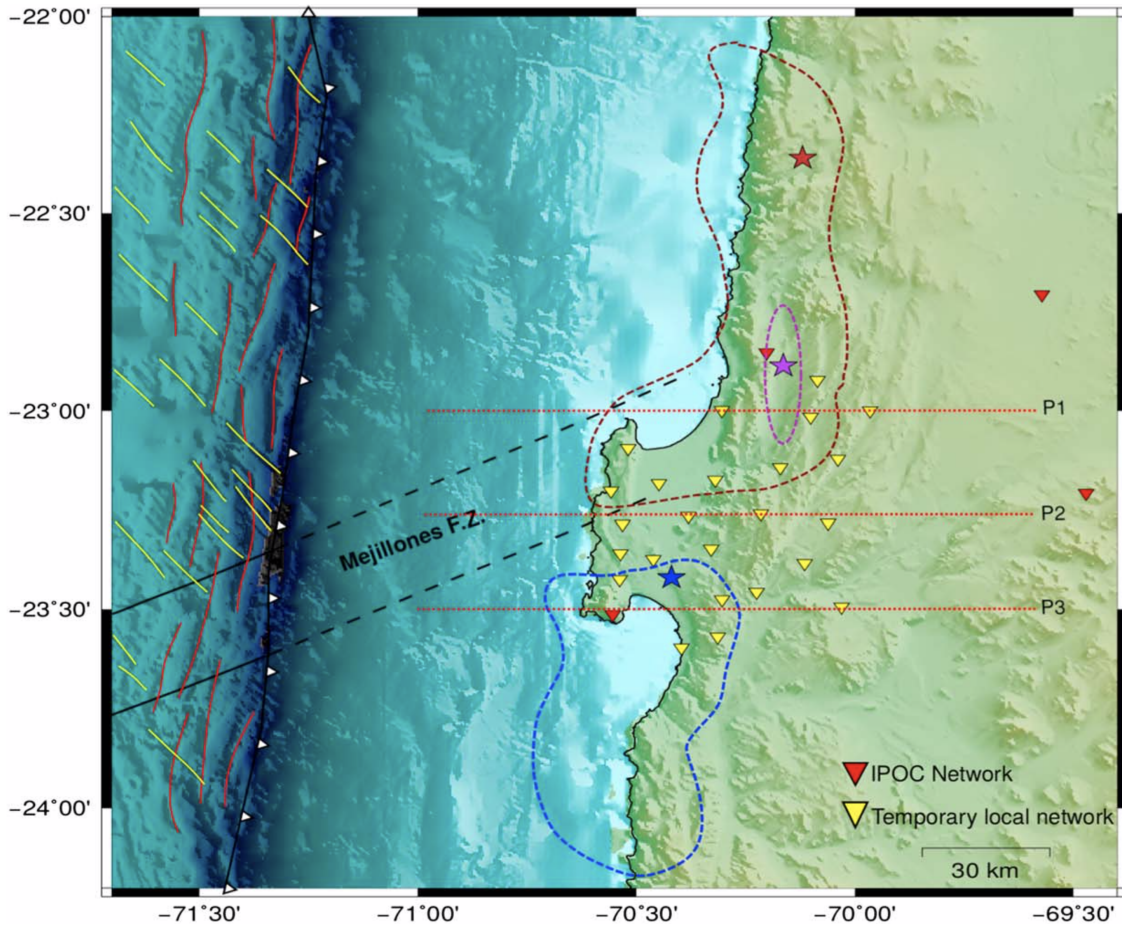
Figure 3.1: Seismotectonic setting of the Mejillones Peninsula from *Pasten-Araya et al.* (2018). Black dashed lines represent the Mejillones fault zone. Dashed blue contours represent the rupture zone of the 1995 Antofagasta earthquake (Mw 8.1) and red dashed contour the rupture of the 2007 Tocopilla earthquake (Mw 7.7)

The main differences between the profiles studied by *Pasten-Araya et al.* (2018) over the Mejillones peninsula is the fluid content. The seismicity over the Mejillones peninsula was also studied by *Pasten-Araya et al.* (2018), as is shown in Figure 3.2. The downdip limit for seismicity is shallower here than in the rest of the subduction contact zone, being here ∼30km, shallower than typical values ∼45-55 km in depth (*Schurr et al.*, 2012; *Tong et al.*, 2010).
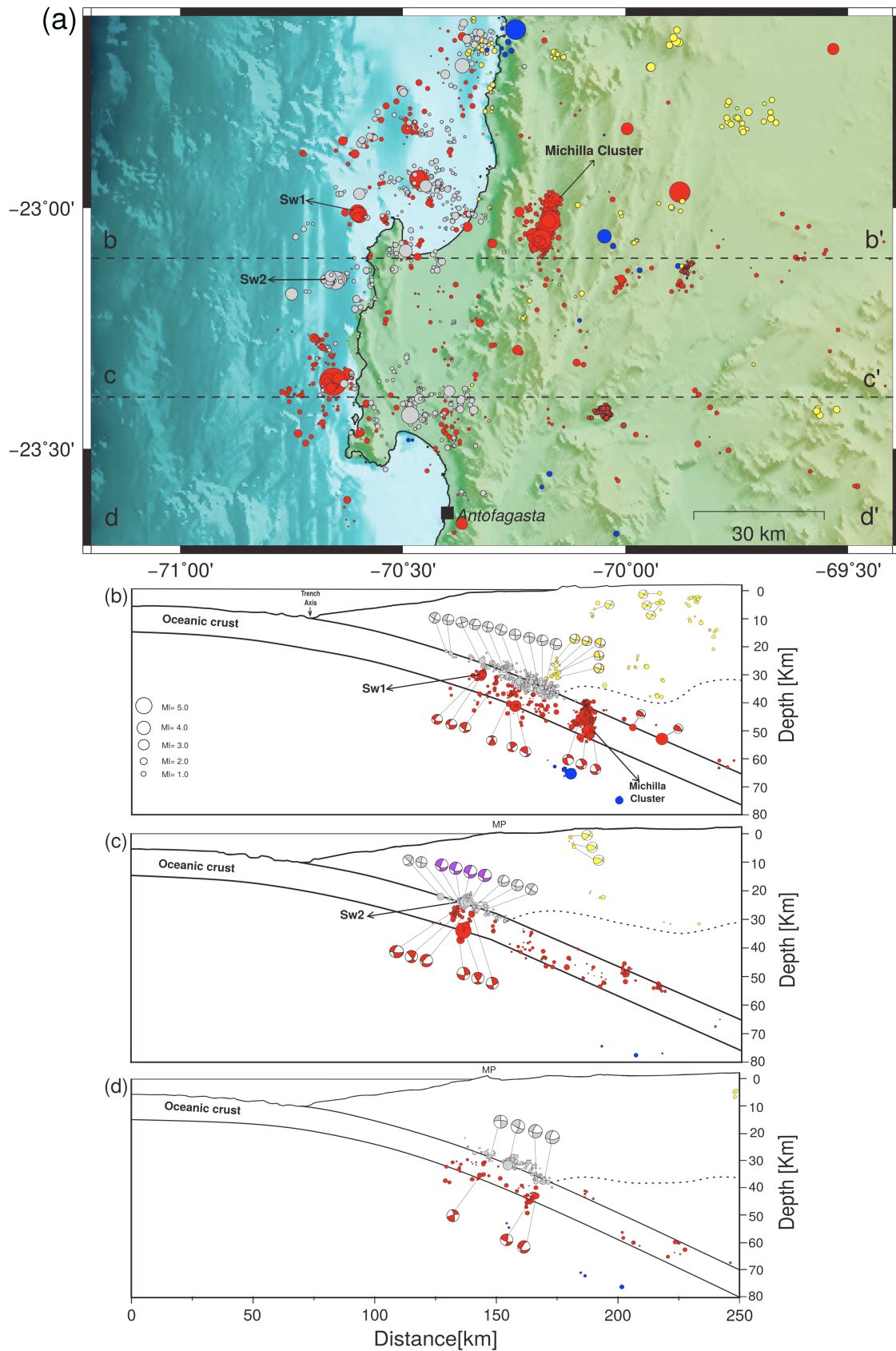
Figure 3.2: Seismicity and focal mechanisms under the Mejillones peninsula from *Pasten-Araya et al.* (2018). Gray circles represent interplate events, while red circles show intraslab seismicity. The dashed line corresponds to the mantle wedge. a) Map with seismicity and b), c) and d) are cross sections showing distribution of seismicity and focal mechanism at depth northward, on the Mejillones peninsula and southward from the Mejillones peninsula, respectively.

The geometry of the slab in the region is characterized by an average subduction angle close to 20° (*Béjar-Pizarro et al.*, 2010; *Delouis et al.*, 2009). A more precise geometry can be modeled including the results of *Contreras-Reyes et al.* (2012) from a seismic tomography profile made in the region close to Tocopilla (profile shown in Figure 3.3). This study concludes that the slab in this zone has a kink where the dip changes abruptly at depths of about 20 km. Figure 3.4 taken from *Contreras-Reyes et al.* (2012) shows the inferred slab geometry and the kink proposed by the authors. Note the approximate change in dip from 10° to 22°. This geometry will be used to represent the megathrust fault contact for the Bayesian modeling of slip at the source of the 2007 Tocopilla earthquake.
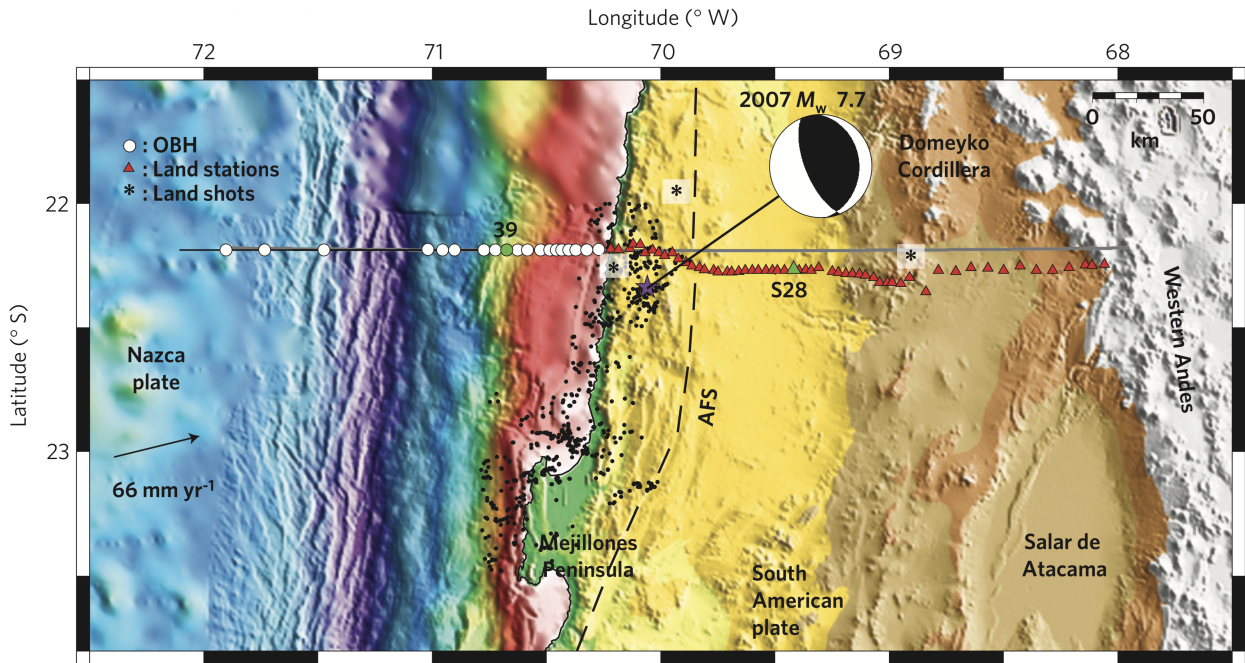
Figure 3.3: Seismotectonic setting of north Chile. Swath bathymetry and topography and location of the seismic experiment made by *Contreras-Reyes et al.* (2012) at the region affected by the Tocopilla earthquake.
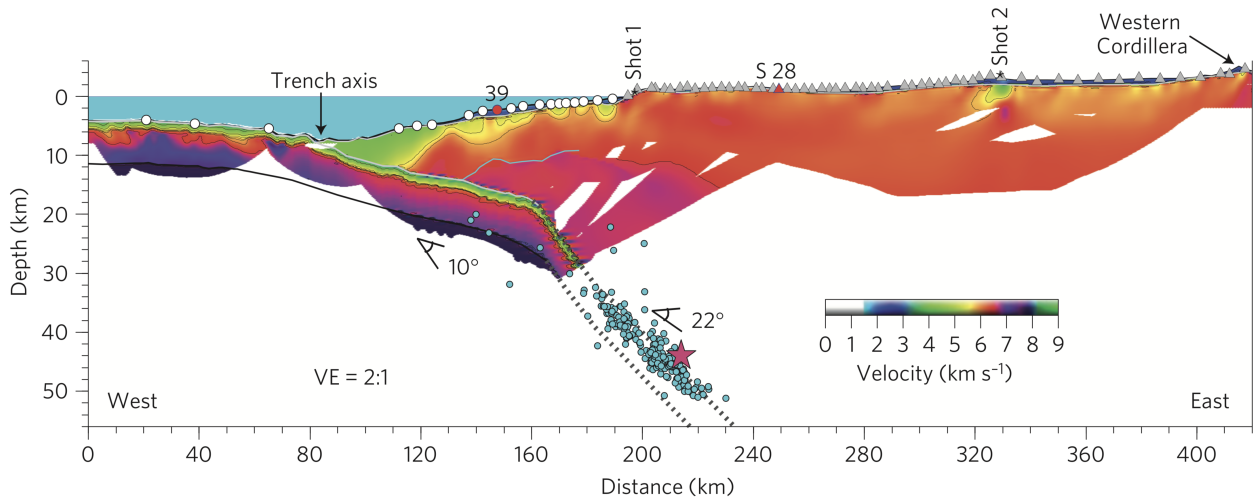
Figure 3.4: Seismic velocity structure of the upper subduction zone. Figure from *Contreras-Reyes et al.* (2012) showing two-dimensional velocity–depth model obtained from tomographic inversion using travel time data.

3.2 Previous studies of the 2007 Tocopilla earthquake

The source mechanism of the 2007 November 14 Tocopilla (Mw 7.7) earthquake has been analyzed by several authors (e.g. *Béjar-Pizarro et al.*, 2010; *Delouis et al.*, 2009; *Peyrat et al.*, 2010; *Minson*, 2010). In the following we will detail more about the work made by *Béjar-Pizarro et al.* (2010), *Peyrat et al.* (2010) and *Delouis et al.* (2009). For the earthquake source, all of the aforementioned studies identify two main patches of rupture, each having a peak slip reaching between 2.6 m to 3.0 m. The main slip patches are located along a 130 km segment, and spans depths from about 25 km to 55 km. The rupture started near the northern patch and then propagated to the south (*Peyrat et al.*, 2010; *Schurr et al.*, 2012). The essential aspects of these studies such as data, method, geometry, prior assumptions and maximum slip obtained are summarized in Table 3.1.

Summary				
Study	*Minson* (2010)	*Peyrat et al.* (2010)	*Béjar-Pizarro et al.* (2010)	*Delouis et al.* (2009)
Data	GPS and INSAR	Teleseismic and strong motion data	GPS and InSAR	Far field teleseismic data, accelerograms
Method	CATMIP (static and kinematic)	kinematic non linear inversion	regularized least squares with positivity constraint	non-linear kinematic inversion, simulated annealing
Max Slip	3.015 ± 0.273 m	1.5 m(teleseismic), 2.6 m (accelerometric)	2.5 m (estimated from image), 1.1 m (average)	1.2 m mean, 3.0 m peak slip
Geometry	plane strike 5°, dip 20°, rake 99°	strike 358°, dip 26°, rake 109° , 160×80 km^2 (20×20 km^2 subfault size)	plane strike , dip 20°, rake 105°, 19×12 subfaults (14×14 km^2 subfault size)	strike 0°, dip 20°, rake 105°, 176 subfaults (12×12 km^2 subfault size)
Prior assumptions	498 free params. Uniform prior distributions for the parallel to rake components, Gaussian distributions for the perpendicular component, Dirichlet distribution for the initial distributions.	Two elliptical slip patches that ruptured at constant speed starting from the hypocentre. 14 independent parameters, seven for each elliptical patch.	228 patches, rms similar for both data sets (GPS and InSAR)	the parameters to be inverted for are the slip onset time, the rake angle, and the amplitudes of the two triangular functions. Bounding values for free parameters

Table 3.1: Comparison of previous studies of the 2007 Tocopilla event.

These studies employ different methodologies and assumptions for the inversion of the 2007 Tocopilla 7.7 earthquake. For example, *Béjar-Pizarro et al.* (2010) employed a regularized least squares with positivity constraints method. The formulation of the inverse problem using this kind of methodology requires the use of a regularization term, which imposes a smoothing constraint to the solution. This additional restriction is necessary to mitigate the instability of the inverse problem, but as a result it imposes a bias in the solution, disregarding potential sharper models. Figure 3.5 shows the slip models obtained by *Béjar-Pizarro et al.* (2010) when using GPS and InSAR data separately, and in a join manner. The overall slip pattern is similar in the three cases, being shallower in the case when using InSAR only. Additionaly, this study does not present an analysis of the errors on the estimated slip model.

On the other hand, *Minson* (2010) developed the Bayesian sampling method CATMIP and applied it to the study of the 2007 Tocopilla earthquake (Mw 7.7), where the static slip inversion was made using GPS and InSAR data as constraints on slip. The resulting slip distribution is shown in Figure 3.6, which represents the ensemble of samples as contoured lines, indicating the intervals in which a portion of the samples resides, and presents (in gray) the mean model solution from all sampled models (Figure 3.6). Compared to the optimization least squares approach, fully Bayesian methods does not require to set smoothing constraints as prior information in order to stabilize the ill posed inverse problem, as it obtains the complete ensemble of models that represent all possible solutions (with associated probability) to the inverse problem, including the model variations that results from the ill conditioning of the inverse problem. CATMIP, in particular, sample the space of solutions efficiently and reliably, allowing also a direct characterization of the errors of the estimated solutions.
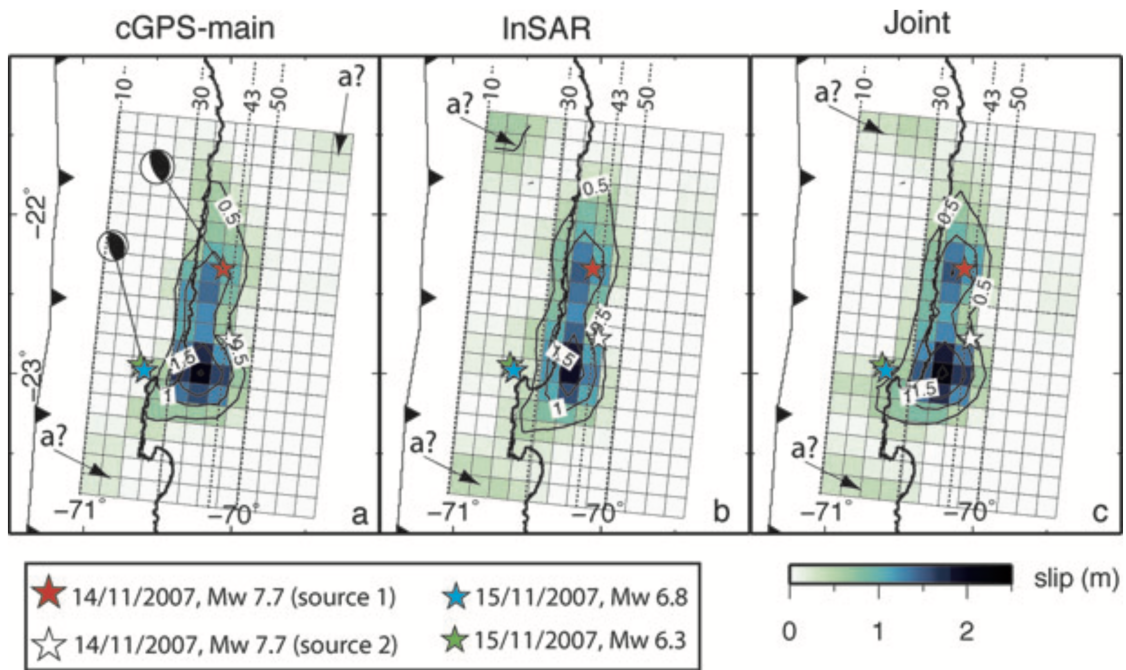
Figure 3.5: Slip models obtained by *Béjar-Pizarro et al.* (2010) for the Tocopilla earthquake using different data sets. (a) Slip inverted from coseismic cGPS data. (b) Slip inverted from InSAR data. (c) Joint GPS-InSAR inversion. GPS in this case includes 26 d of post-seismic deformation to span a period comparable to InSAR.
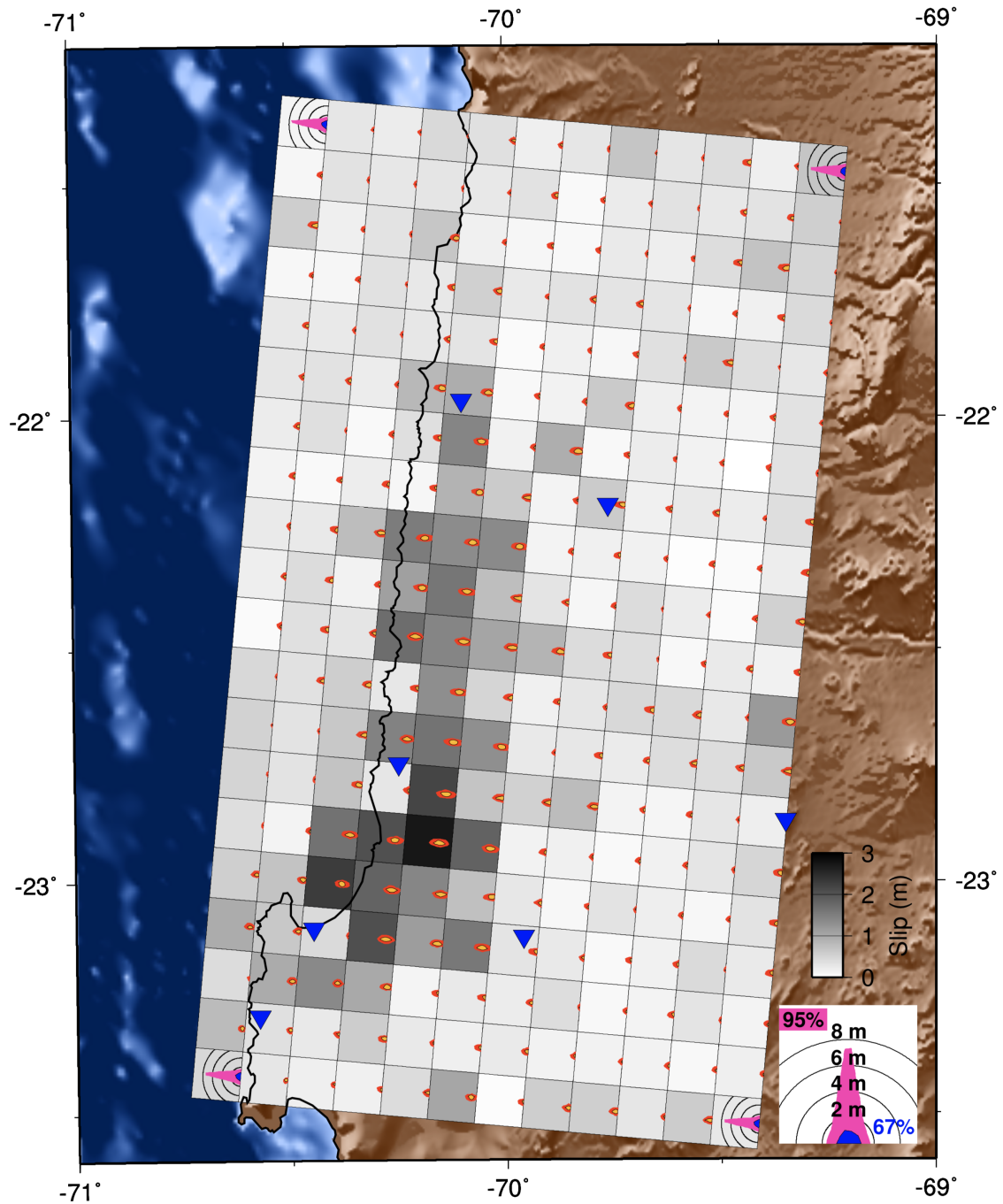
Figure 3.6: Posterior slip distribution from static modeling from *Minson* (2010). Two contour lines are drawn for each patch. Each contour illustrates the range of slip values which contains 67% and 95% of all posterior models. The patches at the corners illustrate the range of slip values found in the prior distribution. The background grayscale intensity of each patch is the mean of the slip distribution for that patch. Locations of GPS stations are indicated by triangles.

In contrast to what is new in this study of the Tocopilla earthquake, the analysis made by *Minson* (2010) incorporate only the observational source of error $\mathbf{C_d}$, ignoring the errors associated to the prediction of the forward model. After that, *Duputel et al.* (2014) formulate

a method to calculate the uncertainties in the prediction of the model $\mathbf{C_p}$ from the uncertainties in the variables of the physical model. This error $\mathbf{C_p}$ is linearly related to the vector of model parameters $\boldsymbol{\theta}$, thus, for the inversion of slip of big earthquakes, the uncertainties of the forward model prediction can be larger than the observational ones. The CATMIP algorithm allows to incorporate the errors on the forward model, updating its estimation based on the ensemble of model solutions on each iteration of the algorithm, as it approaches the posterior pdf, it uses the ensemble of models to calculate a representative value of $\mathbf{C_p}$. As an example of the use of model prediction uncertainties on slip inversion, *Duputel et al.* (2015) employed CATMIP to estimate the rupture process of the 2014 (Mw 8.2) Pisagua earthquake incorporating the errors coming from the definition of the elastic structure of the medium. In this thesis, we study the 2007 Tocopilla earthquake based on the study made by *Duputel et al.* (2014, 2015) to invert the slip distribution on the interplate contact, incorporating the errors in the determination of wave propagation speeds v_p, v_s and density ρ.

Chapter 4

Modeling the 2007 ($M_w7.7$) Tocopilla Earthquake

Fault slip associated to earthquake occurrence can cause measurable motion at the surface of the Earth, such as dynamic transients associated to the passage of seismic waves, as well as remanent (static) displacements of the Earth's crust. In order to model and analyze the source of such surface displacements, we need to define a model of the source and of the response of the media.

For the Tocopilla earthquake, we will use surface static displacements from space geodesy measurements to constrain static slip distributions of the earthquake at the causative megathrust fault. Here we explain how we set the fault geometry where the source of the earthquake takes place, the structure of the considered elastic media, and the theoretical relation between fault slip and crustal surface displacements through the definition of the Green's functions of the slip estimation problem.

4.1   Data

The observations used to constrain slip at the source of the Tocopilla earthquake emerge from the coseismic crustal deformation caused by the earthquake. This deformation can be measured at the surface of the Earth's crust by space geodetic instruments, in this case, by a network of instruments of the Global Positioning System (GPS) and InSAR (Interferometric Synthetic Aperture Radar) satellites.

By one hand, continuous GPS data measure the 3D displacement field (East, North and Vertical components), with respect to a given reference frame, on stations of a GPS network established in the region affected by the Tocopilla earthquake. The position of the antenna of a GPS station is obtained by a process of trilateration between known coordinates of the GPS satellites and the coordinates of the antenna to be determined at a given time given by the reception of an electromagnetic signal transmitted by such satellites (see Figure 4.1a and *Blewitt*, 2015, for further details in the GPS measurement technique).

On the other hand, InSAR technique uses interferometry data to measure the relative deformation along a Line of Sight (LOS) between thousands of points on the Earth's crust

and the InSAR satellite antenna (see Figure 4.1b and *Simons & Rosen*, 2015a, for further details in the InSAR measurement technique).

(a) GPS acquisition

(b) InSAR acquisition
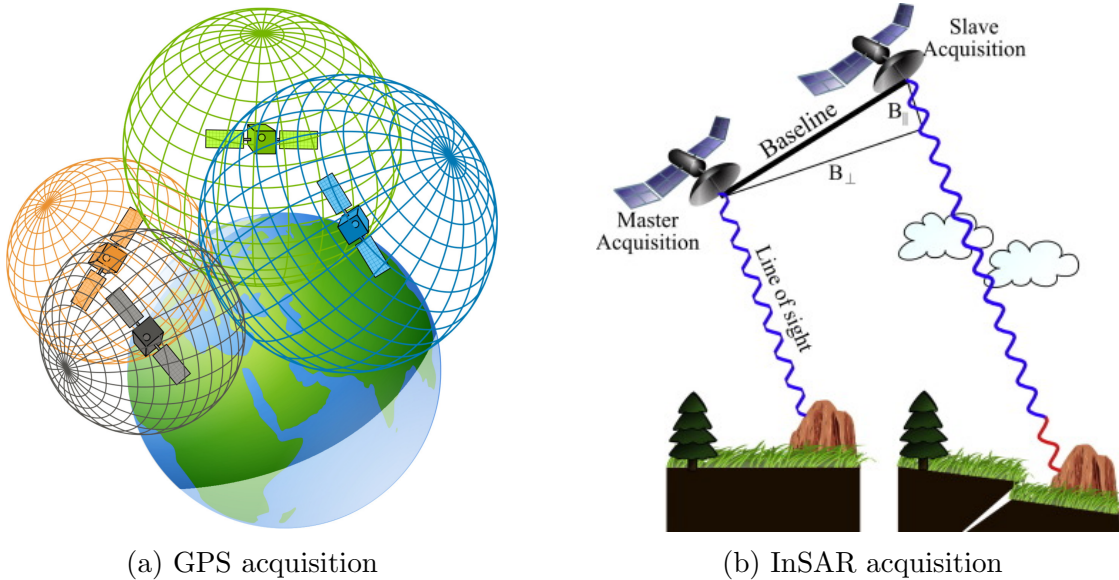
Figure 4.1: a) Scheme of trilateration in which distances between GPS (receiver) antenna and Satellites is used for positioning of the ground antenna from known satellite coordinates (Source of image https://openclipart.org/download/191659/GPS-3D- trilateration.svg). b) Schematic illustration from *Osmanoglu et al.* (2016) explaining the functioning of the InSAR acquisition. Undulating lines represent propagation of electromagnetic waves though the Line of Sight direction. Red portion of the line indicate the extra travelled path due to crustal motions, from which the displacement along LOS is estimated by the InSAR technique.

4.1.1 GPS data

The GPS data is obtained from the work of *Béjar-Pizarro et al.* (2010) and corresponds to co-seismic crustal displacements (three components: east, north and vertical) measured at 11 cGPS stations located between the coast line and nearly 350 km eastwards of the Nazca-SouthAmerican trench. The remanent co-seismic displacements obtained by analyzing GPS data within a period between 10 days before and after the Tocopilla earthquake main shock. Figure 4.2 shows the coseismic displacements measured at the GPS network, as well as their observational uncertainties.
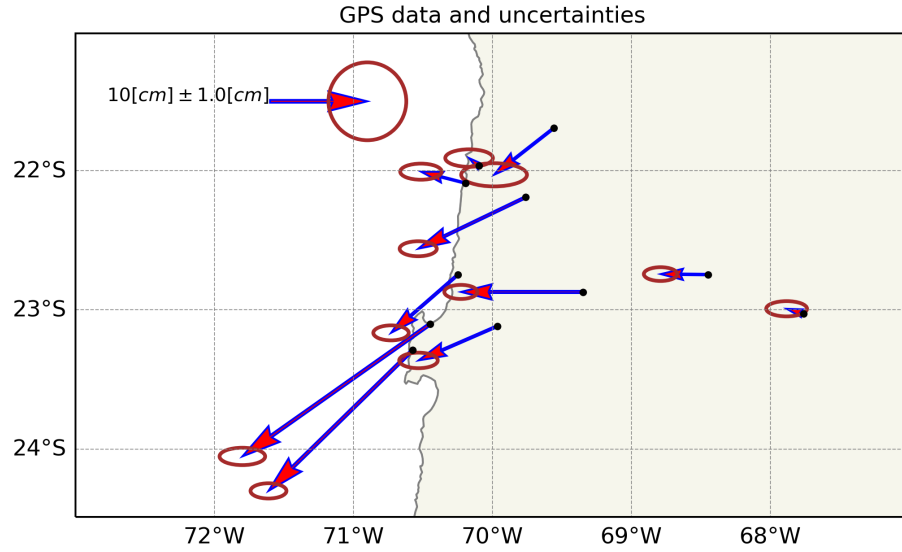
(a) GPS inferred coseismic quasiestatic crustal displacements.
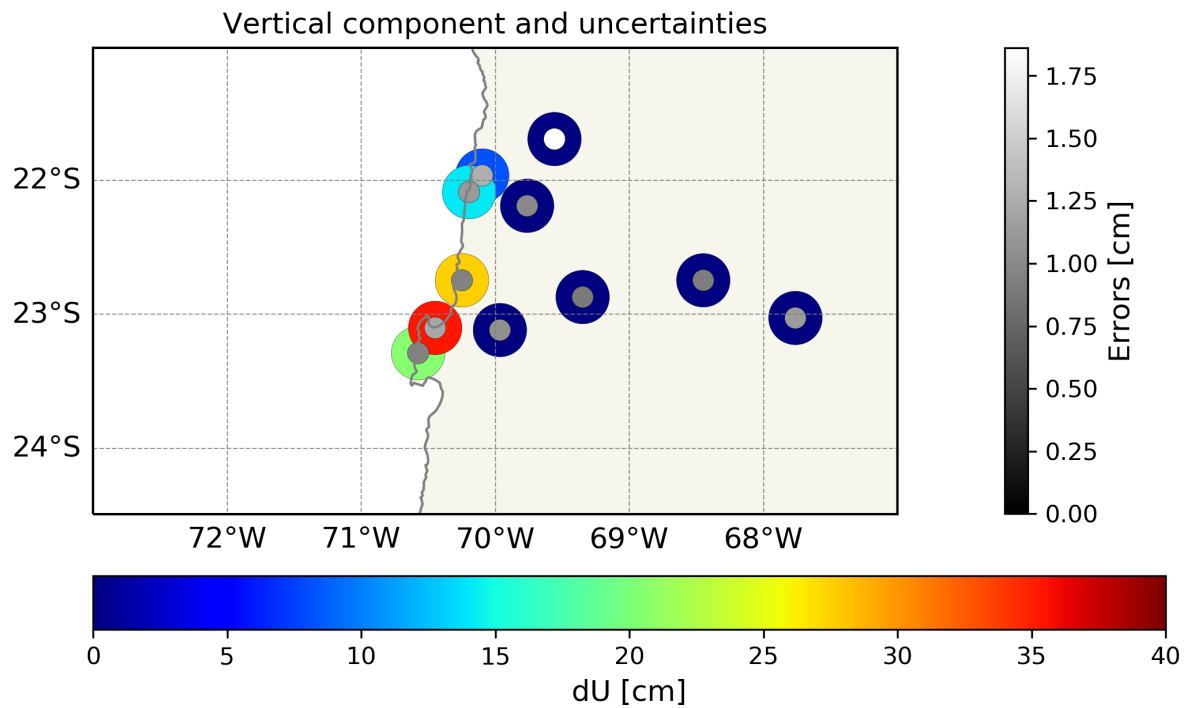
(b) Vertical component.

Figure 4.2: GPS data processed by *Béjar-Pizarro et al.* (2010) and standard deviation associated to the GPS data. a) Horizontal components indicated with arrows and standard deviation using ellipses b) Vertical displacement and its standard deviation indicated in colors.

4.1.2 InSAR data

We use 1 track of InSAR data, courtesy of Sarah Minson. The interferogram used is T096 (descending), from the satellite ENVISAT. As each InSAR image has several thousands of pixels that are locally highly correlated in space, the data was resampled by Romain Jolivet using the method developed in *Lohman & Simons* (2005), in order to produce a reduced, representative set of a few hundreds of data points (794 data points) resembling coseismic displacements of the Tocopilla earthquake in the LOS direction. The interferograms covers the coastal region near Tocopilla in periods that spans a few months before and after the mainshock event, between April and December 2007.

Figure 4.3 show the resampled InSAR measured displacements in the LOS direction, while Figures 4.4 and **??** show the observational uncertainties in the form of standard deviation and covariance for a specific data index. The resampled observations and uncertainties come from the processing of the InSAR data described in *Lohman & Simons* (2005).
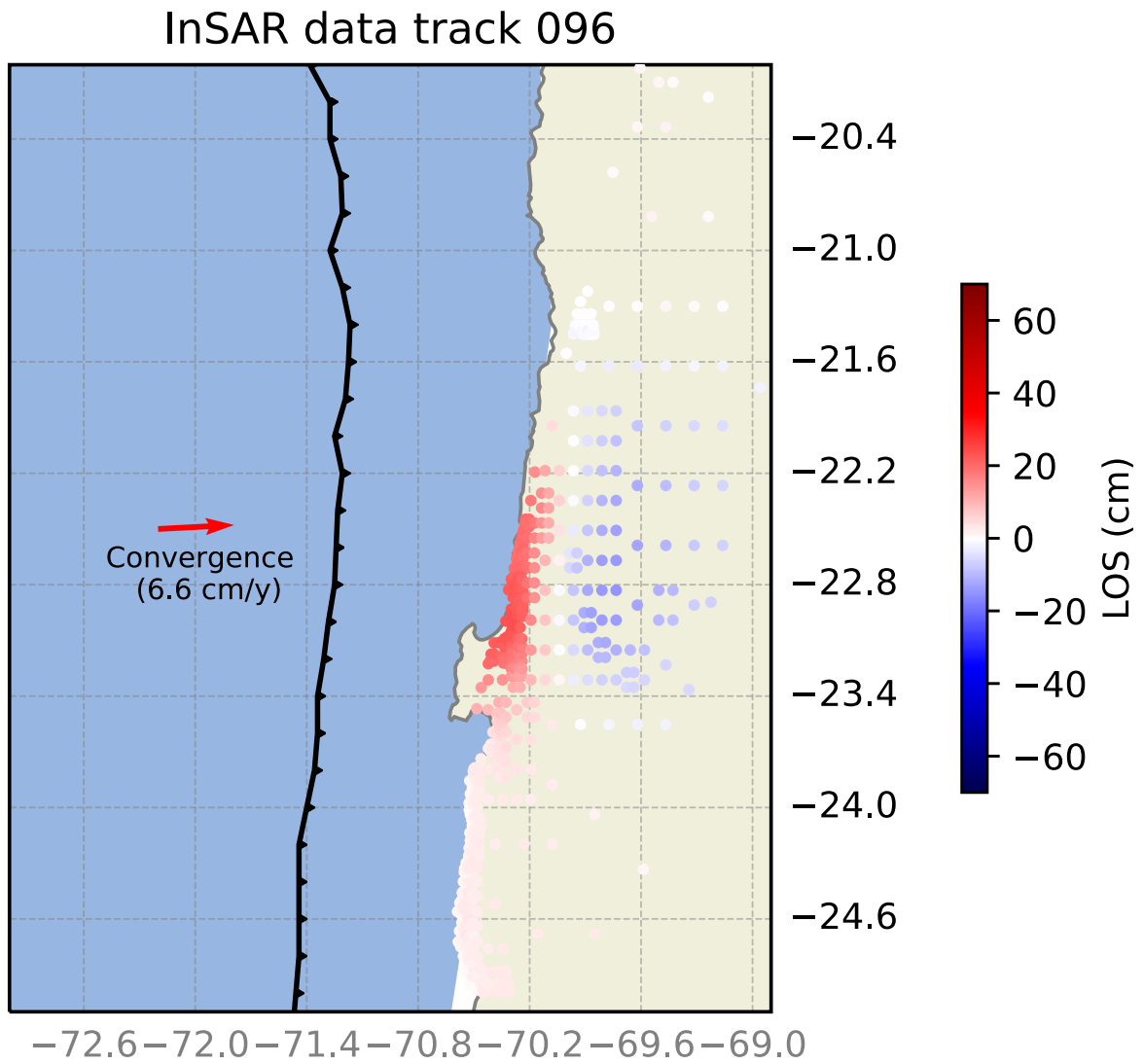
Figure 4.3: InSAR data (track 96 - April - Dic 2007) (courtesy of Sarah Minson) (mean LOS vector ($LOS_E = 0.406$, $LOS_N = -0.099$, $LOS_U = 0.907$)).
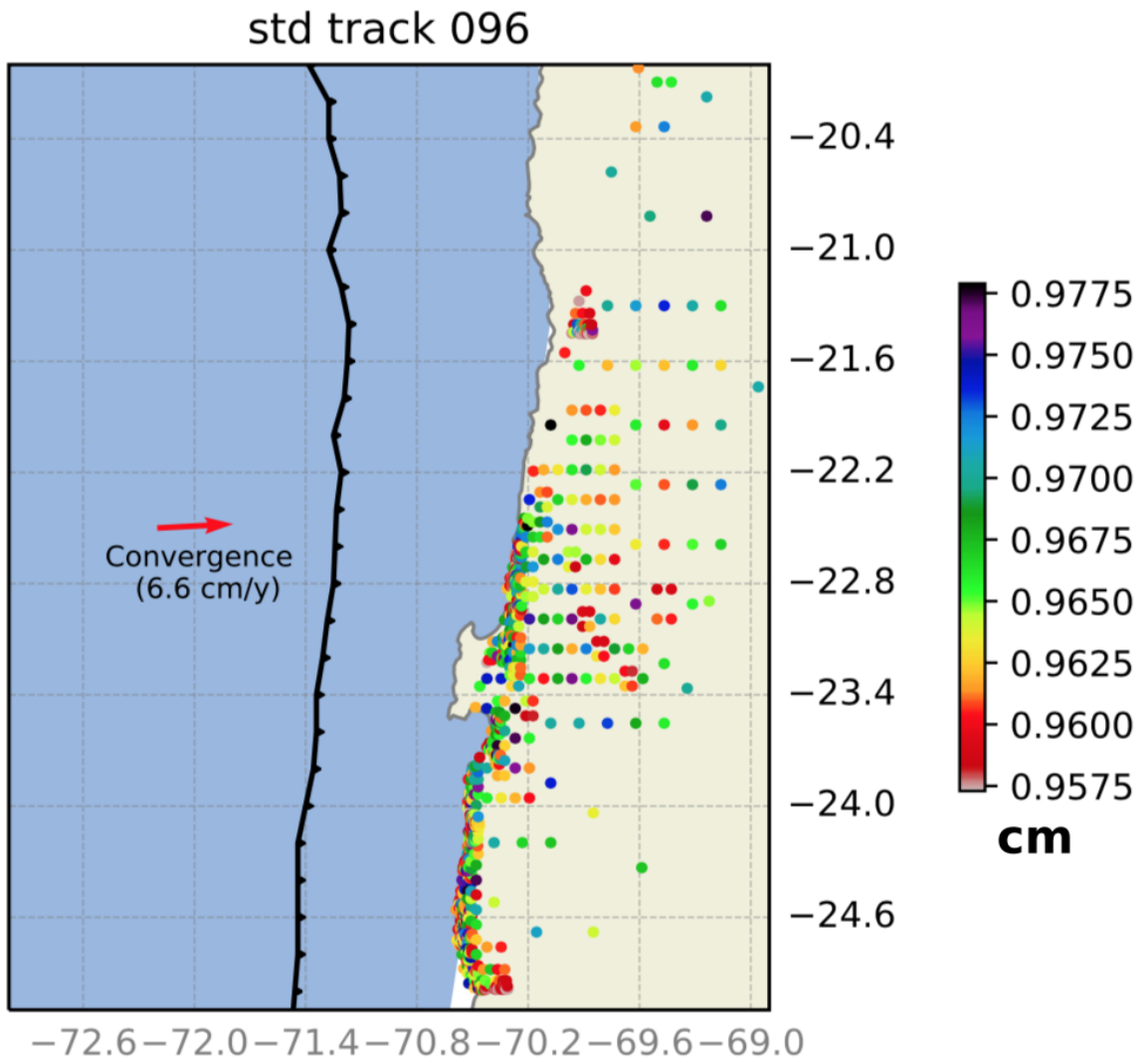
Figure 4.4: Standard deviations in centimeters associated to InSAR data (track 96 d). The errors are all about 1 cm.
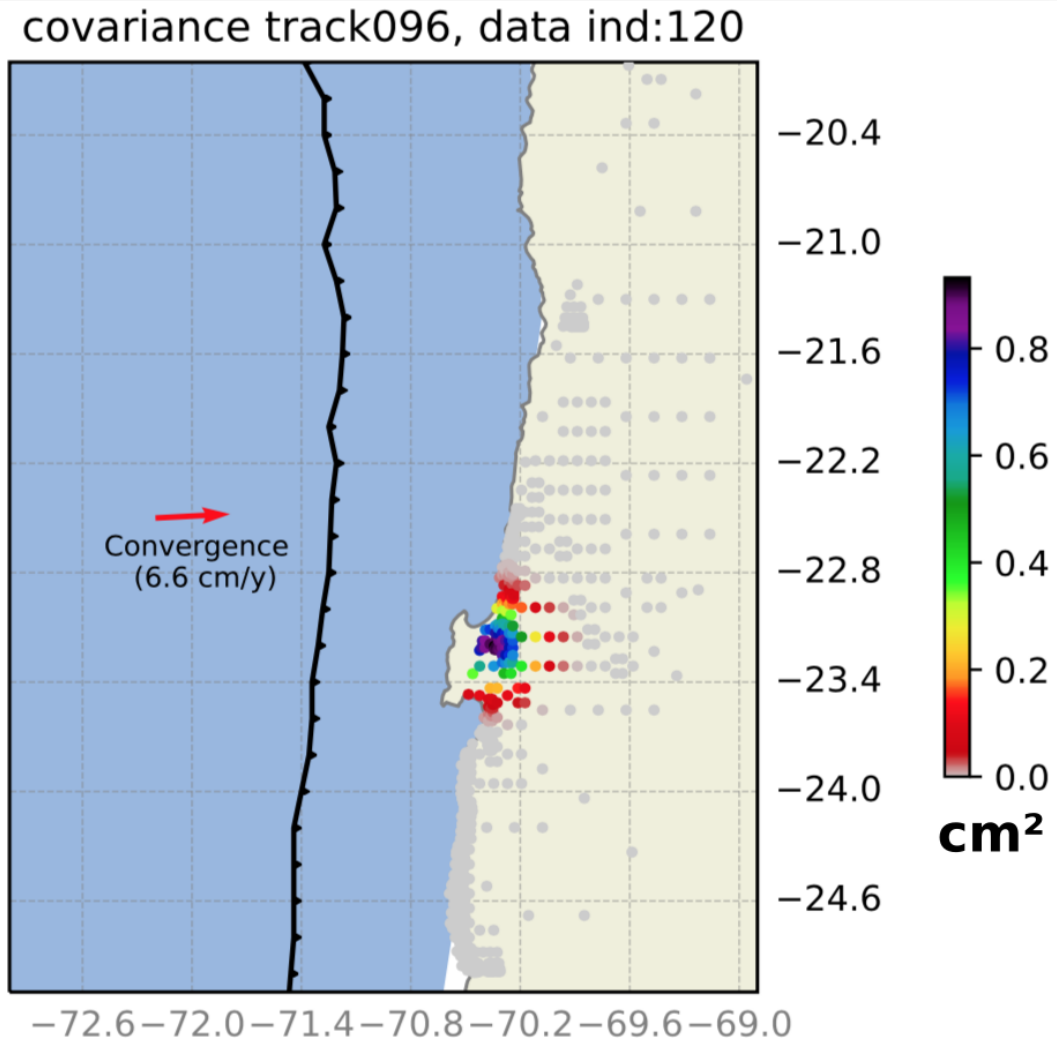
Figure 4.5: Covariance of the InSAR data set for track 96 d. The figure shows the row 120 of the covariance matrix C_d.

4.2 Fault Geometry and Slip Parameterization

The fault geometry used in this study was modeled from a tomographic profile made by *Contreras-Reyes et al.* (2012) in the northern part of the study zone. Besides other models of the subduction interface (e.g., SLAB 1.0 - *Hayes et al.*, 2012), the *Contreras-Reyes et al.* (2012) found an abrupt change in dip in the subducted slab at about 20 km depth, where the subduction dip angle changes from 10° to 22° (see Figure 4.6).

We modeled the geometry of the slab (Figure 4.7) by extruding the profile from *Contreras-Reyes et al.* (2012) along a latitudinal direction defined by the average strike angle of the Nazca-SouthAmerican trench (strike of 5°) in the region affected by the Tocopilla earthquake. The defined geometry spans 260 km along strike, 180 km along dip, and was discretized in 13×9 subfaults, along strike and dip directions, respectively.
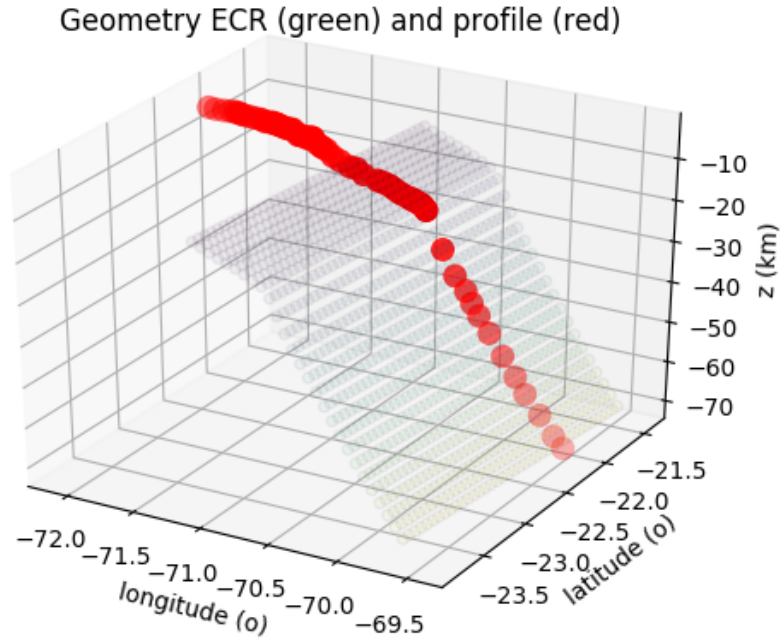
Figure 4.6: Profile obtained by *Contreras-Reyes et al.* (2012) in red and the fault geometry extrapolated to model the Tocopilla earthquake (translucent dots).
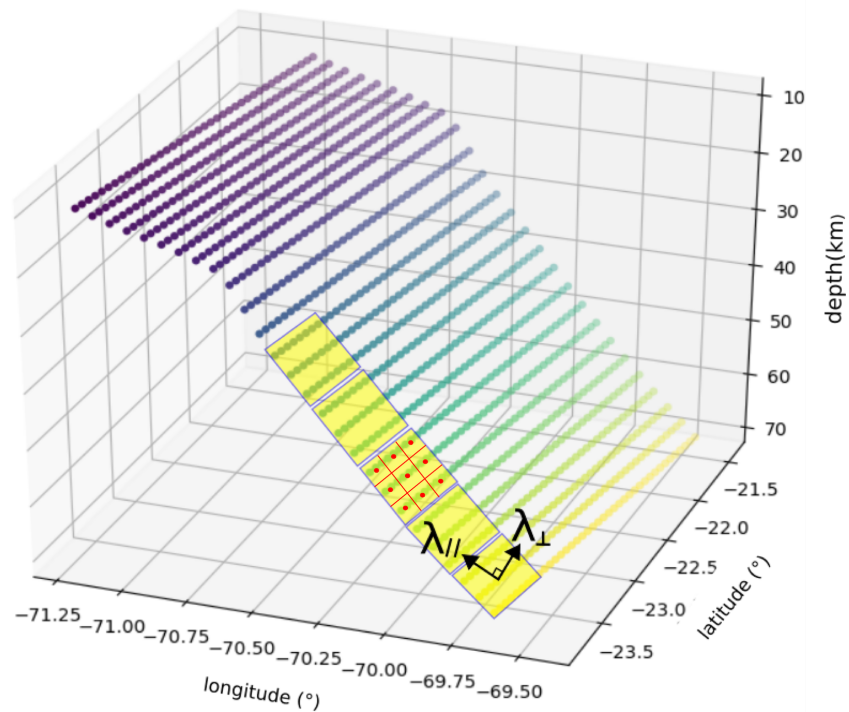
Figure 4.7: Fault geometry from *Contreras-Reyes et al.* (2012), parameterized into 13×9 subfaults along strike and dip, respectively (only some patches are shown as yellow rectangles). Slip directions are shown as rake parallel ($\lambda_\parallel$) and rake perpendicular ($\lambda_\perp$) directions in one yellow subfault. Red lines and dots inside one of the yellow subfaults, resemble the sub-subfault subdivision and point sources at the center of each, respectively.

At each subfault (yellow rectangles in Figure 4.7), slip is assumed to be uniformly distributed (same value of slip across each subfault). Also, slip on each subfault is parameterized to occur in 2 orthogonal directions, rake-parallel and rake-perpendicular directions. As inferred directly by their names, the rake parallel direction models slip occurring in a direction that is parallel to the slip direction implied by the focal mechanism of the Tocopilla earthquake obtained from GCMT (rake 105°). The rake perpendicular direction, is a direction that is orthogonal to the rake parallel one, that also lies within the fault surface.

Each subfault is composed of 3×3 sub-subfaults, in which a point source is defined at their center (see example of red dots indicating point sources in one of the subfaults in Figure 4.7). The number of sub-subfaults is chosen to ensure that the minimum distance between point sources is at least a quarter of the distance from a point source to an observation point located at the surface of the Earth. According to the Saint-Venant principle, under such configuration, the point source representation is enough to model the finite rectangular sub-subfault. Given the subfault subdivision, the response of the media, for instance static crustal displacements generated due to slip on every subfault, is computed as the superposition of the response due to slip on each of the point sources conforming all sub-subfaults of the given subfault. The representation of these point sources, as well as the discretization over the fault geometry are shown in Figure 4.7.

4.3 Structural Model of the Medium

We construct a 1D velocity model consisting in a layered elastic half space in order to model the static displacements of the 2007 (Mw 7.7) Tocopilla earthquake. We use the 1D layered medium defined by *Husen et al.* (1999), describing elastic parameters for 15 homogeneous layers. Figure 4.8 shows the depth dependent values of P-wave speed, v_p, S-wave speed, v_s, and density, ρ according to Husen's model. Particularly, in this study we account for the uncertainties on the values of the elastic parameters in our modeling. As such model parameters are strictly positive quantities, and we lack further information about their uncertainty, we consider a fractional error of 10% for each value of the elastic parameters.
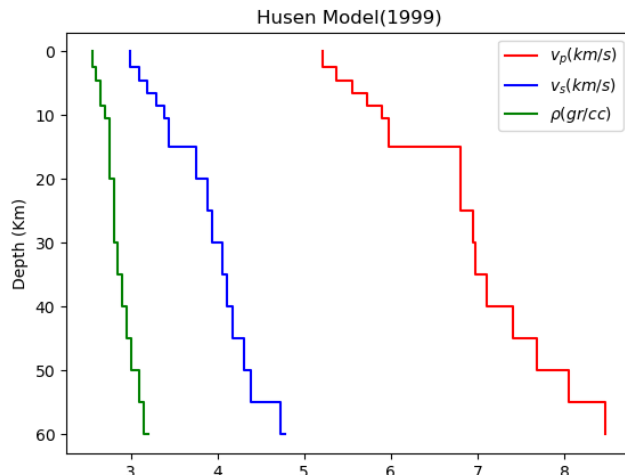
Figure 4.8: 1D velocity model: body waves P-wave speed v_p, S-wave speed v_s, and density ρ from *Husen et al.* (1999).

4.4 Forward and Inverse Problem Formulation for the Static Slip Inversion Problem

The deformation signal caused by an earthquake can be studied by decomposing it into a transient motion associated to the passing of seismic waves and a static component representing the remanent crustal displacements caused by the earthquake. Here, we study the remanent or static part of the deformation, which we model in terms of the elastic static response of the crust due to static dislocations occurring at the causative fault of the earthquake. The elastic static response depends on the rheological properties of the propagating medium representing the Earth's lithosphere. For the purpose of this work, we use the layered elastic medium described in section 4.3.

Here we seek to model the static coseismic slip of the Tocopilla earthquake that occurs at the causative megathrust fault. As described in section 4.2, the fault geometry is discretized into rectangular subfault elements, in which slip is assumed to be uniform. At each subfault, the slip vector is parameterized using 2 orthogonal directions along the subfault, a rake parallel and a rake perpendicular direction (see Figure 4.7). Thus, we represent slip occurring along each subfault by the components of the slip vector $\mathbf{m_{slip}} = [\mathbf{m}_\lambda^\parallel, \mathbf{m}_\lambda^\perp]^T$ of unknowns to be determined during the inversion procedure.

The prediction of the slip model $\mathbf{m_{slip}}$, is carried out by calculating the static displacements at selected points on the surface of the Earth's crust, either the location of GPS stations or pixels of InSAR resampled interferograms, caused by a given slip field occurring at the causative fault. The calculation of the surface displacements is performed using the CPS software package (Computer Programs in Seismology; *Herrmann*, 2013)), a package of algorithms designed to study problems in seismology. The functions of this package allows to calculate the displacements (static and dynamic) at a point on the surface of the Earth due to a point source causing a perturbation of a stratified medium, using the discrete wavenumber method of *Bouchon* (1981). By calculating the displacements at all points of interest at the surface of the Earth due to a unitary dislocation at each subfault, one can construct a Green's functions matrix $\mathbf{G^{slip}}$, and define the forward model that predicts Earth's surface coseismic displacements, $\mathbf{d_{slip}^{pred}}$, given a fault slip distribution parameterized in $\mathbf{m_{slip}}$ as,

$$\mathbf{d_{slip}^{pred}}(\mathbf{m_{slip}}) = \mathbf{G^{slip}} \mathbf{m_{slip}} \tag{4.1}$$

Depending on the dataset used, some extra parameters need to be considered in order to incorporate corrections that need to be performed on the data. For instance, when using InSAR data, as explained in the following, a linear (or sometimes quadratic) ramp needs to be removed from the InSAR LOS(Line of sight) displacements in order to correct for errors in the orientation of the radar antenna of the InSAR satellite. In such case, the model prediction will need to incorporate such data corrections, and thus the vector of model parameters (that will generally named θ) will include not only the slip values at the fault, but also the coefficients that define the ramp used to correct InSAR data, i.e.,

$$\theta = [\mathbf{m_{slip}^T}, \mathbf{m_{ramp}^T}]^T \tag{4.2}$$

In general, we will name the vector of model parameters as $\boldsymbol{\theta}$, regardless that if includes only the parameters that define slip at the fault, or some other parameters that are needed to perform the inversion constrained by a certain dataset. Thus, in general we refer to the forward model prediction using the linear formulation

$$\mathbf{d^{pred}}(\boldsymbol{\theta}) = \mathbf{G}\boldsymbol{\theta} \tag{4.3}$$

where $\mathbf{G}$ and $\boldsymbol{\theta}$ are the Green's function matrix and the vector of unknown model parameteres, respectively. Both include the components associated to the prediction of crustal displacements due to fault slip and potentially some data corrections if needed, and will be explained in each case analyzed later in this work.

Once the forward model is set, the inverse problem is defined in terms of a Bayesian framework (see chapter 2), in which we seek to find an ensemble of model parameters that represent the a posteriori joint probability density function of model parameters,

$$f^{post}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) f^{prior}(\boldsymbol{\theta}) \tag{4.4}$$

where $L(\boldsymbol{\theta})$ is the likelihood function of a given model $\boldsymbol{\theta}$ measuring how well the model prediction fits the observations, in which data and model prediction errors are represented as Gaussianly distributed (see equation 2.3). $f^{post}(\boldsymbol{\theta})$ and $f^{prior}(\boldsymbol{\theta})$ are the joint probability distributions representing a posteriori and a priori information states in model parameters, understanding information states as a representation of the knowledge we have on such model parameters.

The Bayesian inversion problem is solved using a sampling technique, the CATMIP algorithm (see chapter 2) that allows to find an ensemble of model parameters consisting in a set of model $\boldsymbol{\theta}$ with their associated probabilities to be solution of the posed inverse problem. Such models are samples of the a posteriori probability density function $f^{post}(\boldsymbol{\theta})$ in the sense that, for a large enough number of samples, their multidimensional histogram is proportional to the values of $f^{post}(\boldsymbol{\theta})$. We analyze our inversion results for 2 cases, in which coseismic slip is constrained by GPS inferred crustal displacements of the Tocopilla earthquake, as well as a case in which, additionally, InSAR inferred LOS displacements are used to constrain fault slip. In the following we define our model setup, in which we define the Green's function matrix $\mathbf{G}$, the vector of data $\mathbf{d}$, the uncertainties of data and forward model prediction, as well as the corrections needed in order to properly incorporate InSAR data in our analyses. Particularly, the quantification of uncertainty of the forward model prediction is explained in section 4.5.

So far, a function package called ESM (Earthquake Source Modeling) has been programmed to calculate the static Green's function of the problem and build the Green's matrix. Within this package are the necessary modules to model the fault geometry, to give format to the data and to visualize the results. Besides, this package contains the modules necessary to make the inversion using the method of regularized least squares, with a positivity restriction, perform synthetic tests such as the checkerboard test, to prepare all the information needed for the Bayesian inversion using CATMIP as well as for the analysis and interpretation of the ensemble of models generated by CATMIP as the solution of the inverse problem.

4.4.1 Case 1: Fault Slip Constrained by GPS Data

The Green's function is the response of the system to a unitary perturbation, in this case, it is the displacement on the surface of the Earth's crust at all GPS stations generated by a unitary slip on every slip direction of each subfault in the discretized geometry. This problem is parameterized such that slip is decomposed in a rake parallel component as well as a perpendicular to rake component. Then, the linear modeling of the problem in terms of a **G** matrix considers the calculation of displacements generated by a unitary slip in this two components on every subfault.

Let be M the number of subfaults discretizing the fault geometry and N the number observations. In this case, N corresponds to the total number of measured displacements in components East, North and Z on all GPS stations. Lets define $\mathbf{G^{SS}_{ij}}$ and $\mathbf{G^{DS}_{ij}}$ the Green function matrix associated to the prediction of the i-th observation due to slip on the j-th subfault at the strike slip (SS) and dip slip (DS) directions, respectively. In this case, the strike used is 5° according the slip parametrization seen in section 4.1.

The Green function matrix, for slip parametrized on strike and updip directions can be written as:

$$\mathbf{G_0^{GPS}} = \left[\begin{array}{cccc|cccc} G_{11}^{SS} & G_{12}^{SS} & \cdots & G_{1M}^{SS} & G_{11}^{DS} & G_{12}^{DS} & \cdots & G_{1M}^{DS} \\ G_{21}^{SS} & G_{22}^{SS} & \cdots & G_{2M}^{SS} & G_{21}^{DS} & G_{22}^{DS} & \cdots & G_{2M}^{DS} \\ \vdots & & \ddots & \vdots * \vdots & \vdots & & \ddots & \vdots * \vdots \\ G_{N1}^{SS} & G_{N2}^{SS} & \cdots & G_{NM}^{SS} & G_{N1}^{DS} & G_{N2}^{DS} & \cdots & G_{NM}^{DS} \end{array} \right] \tag{4.5}$$

Now we can express the Green's function matrix for slip ocurring at a rotated direction, in which one of the orthogonal slip axes coincides with the rake direction obtained from the GCMT solution for the Tocopilla earthquake ($\lambda = 105°$). The resultant Green's function matrix becomes

$$\mathbf{G^{GPS}} = \left[\begin{array}{cccc|cccc} G_{11}^{\parallel} & G_{12}^{\parallel} & \cdots & G_{1M}^{\parallel} & G_{11}^{\perp} & G_{12}^{\perp} & \cdots & G_{1M}^{\perp} \\ G_{21}^{\parallel} & G_{22}^{\parallel} & \cdots & G_{2M}^{\parallel} & G_{21}^{\perp} & G_{22}^{\perp} & \cdots & G_{2M}^{\perp} \\ \vdots & & \ddots & \vdots * \vdots & \vdots & & \ddots & \vdots * \vdots \\ G_{N1}^{\parallel} & G_{N2}^{\parallel} & \cdots & G_{NM}^{\parallel} & G_{N1}^{\perp} & G_{N2}^{\perp} & \cdots & G_{NM}^{\perp} \end{array} \right] \tag{4.6}$$

$$\mathbf{d^{GPS}} = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_N \end{pmatrix} \qquad \mathbf{\theta} = \begin{pmatrix} m_{slip}^{\parallel} \\ m_{slip}^{\perp} \end{pmatrix} \tag{4.7}$$

Where $G_{11}^{\parallel}$ is the first observation predicted by a unitary slip along the rake direction on the first subfault and $G_{11}^{\perp}$ the analog for the component perpendicular to rake.

4.4.2 Case 2: Fault Slip Constrained by GPS and InSAR Data

When using InSAR data it is necessary to incorporate additional parameters to be determined during the inversion, because the data has not only the deformation signal, but also a signal that can be usually modeled as a linear or quadratic ramp, representing errors on InSAR satellite orientation and location during data acquisition. In this case, the ramp will be modeled as bi-linear, in the form $ax + by + c$ (*Simons & Rosen*, 2015a). Thus, we incorporate three additional parameters to be determined during the inversion for the InSAR dataset.

To perform the inversion the matrix $\mathbf{G}$ can be structured using matrix blocks, where each block model GPS and InSAR data, are:

$$
\mathbf{G} =
\begin{bmatrix}
\mathbf{G_1^{GPS}} & \mathbf{0} \\[2ex]
\mathbf{G_1^{InSAR}} & \mathbf{G_{ramp1}}
\end{bmatrix}
\tag{4.8}
$$

For the InSAR data, the observations and forward model are chosen so that we seek for the following relation to hold

$$
\mathbf{d_{obs}^{InSAR}} = \mathbf{d_{slip}^{InSAR}} + \mathbf{d_{ramp}^{InSAR}}
\tag{4.9}
$$

where, $\mathbf{d_{obs}^{InSAR}}$ are the LOS displacements at the pixels of the resampled interferograms; $\mathbf{d_{slip}^{InSAR}} = \mathbf{G^{InSAR}}\mathbf{m_{slip}}$ is the prediction of LOS displacements due to fault slip, and the prediction of the linear ramp $(a, b, c \in \mathbb{R})$ is $\mathbf{d_{ramp}^{InSAR}} = \mathbf{G^{ramp}}\mathbf{m_{ramp}}$; $\mathbf{m_{ramp}} = [a, b, c]$; $a, b, c \in \mathbb{R}$ are constants to be determined during the inversion. Then, we have $\mathbf{d} = \mathbf{G\theta}$ where $\mathbf{G}$ is the one in 4.8, and N_{SAR} is the number of resampled pixels of the interferogram. So,

$$
\mathbf{d_{ramp}} = a\mathbf{x} + b\mathbf{y} + c
\tag{4.10}
$$

$$
\mathbf{G^{ramp}} =
\begin{bmatrix}
x_1 & y_1 & 1 \\
x_2 & y_2 & 1 \\
\vdots & \vdots & \vdots \\
x_{N_{SAR}} & y_{N_{SAR}} & 1
\end{bmatrix}
\tag{4.11}
$$

$$
\mathbf{d} = \begin{pmatrix} \mathbf{d_{GPS}} \\ \mathbf{d_{InSAR1}} \end{pmatrix}
\qquad
\mathbf{\theta} = \begin{pmatrix} \mathbf{m_{slip}} \\ a \\ b \\ c \end{pmatrix}
\tag{4.12}
$$

Note that here, $\mathbf{G^{GPS}}$ is a matrix block associated with GPS data. This would be the $\mathbf{G}$ matrix if the data came only from GPS instruments, same as equation 4.6. When using InSAR data, the matrix is expanded, so additional blocks appear. In the first place, blocks associated with the response of a unitary slip in the discretized elements of the fault plane,

as well as the block for the linear ramp correction.

4.5   Calculation of Uncertainties in the Prediction of the Forward Model

When we are modeling an inverse problem it is necessary to take into account sources of error that prevent the solution to overfit the observed data. These uncertainties are unavoidable and they limit the range where solutions are admissible as a result of the problem.

In the Bayesian framework these errors prescribe the proximity of predictions respect to the observations that will be tolerable to accept for a model in the sampling process. Seen in the data space, where model predictions occur, these errors represent how wide is the likelihood function at its local or global maximum values. Underestimating errors leads to model estimates whose predictions overfit the observed data with the consequence that part of the estimated models are fitting observational and model prediction errors.

The main sources of error are the observational error, coming from the process of measurement, and the epistemic error, coming from the uncertainties related with the construction of the physical model used, which typically is a simplified version of the physical process being modeled.

The observational error is easier to evaluate and obtaining it typically consists in calculating the standard deviation of the measurements, used as observational constraints for the inverse problem. However, the epistemic error requires a better understanding of the behavior of the forward model predictions with uncertainties in the variables that constitute the physical model, and that are not estimated during the inversion (e.g; values of the elastic parameters of the media)

Duputel et al. (2014) developed an approach that allows to quantify the epistemic or forward model prediction error, which can be used to calculate the error associated with the physical model prediction in the case of earthquake source determination. The formulation starts by defining the model $\mathbf{G}$ as a transformation that depends on other variables ν_i describing the properties of the physical model that are not estimated during the inversion process and therefore assumed to be known and fixed. The predicted model is the result of applying $\mathbf{G}$ over a set of parameters $\boldsymbol{\theta}$. In the case of this work, ν_i represents the elastic coefficients of the layered medium expressed in terms of the propagation speed of body waves and density (v_p, v_s and ρ, respectively). Assuming that the true value of these quantities are ν_{0i} we can express the first order Taylor expansion around these values as,

$$G(\boldsymbol{\theta}, \nu_i) = G(\boldsymbol{\theta}, \nu_{0i}) + \Delta G(\boldsymbol{\theta}, \nu_{0i}) = G(\boldsymbol{\theta}, \nu_{0i}) + \frac{\partial G}{\partial \nu_i}(\boldsymbol{\theta}, \nu_{0i})(\nu_i - \nu_{0i}). \qquad (4.13)$$

To determine how $\Delta G(\boldsymbol{\theta}, \nu_0)$ is distributed it is necessary to find a covariance matrix $\mathbf{C}_\nu$, representing uncertainties on the known values of elastic parameters.

As this problem is linear with respect to $\boldsymbol{\theta}$, $\mathbf{G}(\boldsymbol{\theta}, \nu_i) = \mathbf{G}(\nu_i)\boldsymbol{\theta}$ and the above expression can be written as:

$$\frac{\partial \mathbf{G}(\boldsymbol{\theta}, \nu_{0i})}{\partial \nu_i} = \frac{\partial \mathbf{G}(\nu_{0i})}{\partial \nu_i} \boldsymbol{\theta} \tag{4.14}$$

and $\partial \mathbf{G}(\nu_{0i})/\partial \nu_i$ is the sensitivity kernel of the problem $\mathbf{K}$, which can be written as

$$K_{ik,j} = \frac{\partial G_{ik}}{\partial \nu_j} \tag{4.15}$$

Defining $\mathbf{K}^\nu$ as a matrix where each column is a perturbation in the data predicted as a result of a perturbation in a variable ν

$$K^\nu{}_{ij} = K_{ik,j}\theta_k \tag{4.16}$$

If errors on elastic parameters are distributed according a Gaussian distribution, we can write the covariance matrix of the model prediction perturbations as

$$\mathbf{C_p} = \mathbf{K}_\nu \mathbf{C}_\nu \mathbf{K}_\nu^\mathbf{T} \tag{4.17}$$

Now, if the variables ν_i are strictly positive, as is the case for elastic parameters, then the problem can be re-formulated as:

$$\nu_i = ln\mu_i \tag{4.18}$$

Where μ_i is the actual value of the elastic parameter that is strictly positive (i.e, a Jeffreys parameter, see *Tarantola* (2005), section 1.2.4) we obtain

$$\mathbf{C_p} = \mathbf{K}^\mu \mathbf{C}_\mu \mathbf{K}^{\mu\mathbf{T}}, \tag{4.19}$$

where

$$K^\mu{}_{ik,j} = \frac{\partial G_{ik}}{\partial ln(\mu_j)} \tag{4.20}$$

is the sensitivity kernel due to logarithmic perturbation of elastic parameters. Using the latter formulation, we calculate the sensitivity kernels using a numerical approximation of the derivatives of $\mathbf{G}$ due to logarithmic perturbations of the elastic variables on all layers around the values defined by *Husen et al.* (1999). The amount of the logarithmic perturbation applied to the elastic parameters of the structural model is $lnpert = 0.1$ (10%).

To calculate the covariance matrix $\mathbf{C_p}$ associated to the prediction of the model it is necessary to evaluate a particular solution for the vector of parameters in every iteration, until the final distribution is reached. For the first iteration of the algorithm, we use the model obtained using least squares with positivity constraint using GPS data (see appendix A).

4.6   Software AlTar

The software AlTar, named as an acronym for Albert Tarantola, is designed as a framework to work efficiently on fully Bayesian inversions. It was developed at the Seismological Laboratory of Caltech and was especially designed to deal with large scale problems (hundreds to thousands of model parameters depending on the computational resources). It applies particularly for the inversion of earthquake rupture, through the Bayesian sampler defined by the CATMIP algorithm (*Minson et al.*, 2013). This software uses parallel processing to make the necessary evaluations to solve the inverse problem. For this purpose it takes advantage of the ability of Beowulf clusters and GP-GPUs (General Purpose Graphics Processing Unit) to perform parallel calculations through the use of thousands of processing cores. For this work we use a computer node with 2 Nvidia Tesla K40 GPGPU processing units with 12GB of internal RAM and 2800 cuda processors cores, acquired by the proyect 'Proyecto Fondecyt 11140904'.

4.6.1   AlTar Bayesian Inversion

To use this software, from a practical point of view, it is necessary to create a directory, which in a fixed format contains all the information concerning the inversion. This information includes the observations, their associated errors and the parameters of the inversion, as well as the Green's functions, expressed in a matrix form. All the information about the inversion must be placed in a subdirectory named config_dir. Some of the basic files that work as input of the program are the following:

data_file: File with observations in vector format.

gf_file: Green's functions corresponding to the forward model of the problem in matrix form.

Cd_file: File with the covariance matrix associated to the observational uncertainties of the entire data set, in the same sequence as the data file.

Optionally we can include the error in the prediction of the forward model. When it is the case, we need to provide the files with the sensitivity kernels (see next section for its definition) of the prediction of the forward model representing the uncertainties of each fixed quantity (e.g. elastic parameters) used in the definition of the Green's function matrix $\mathbf{G}$.

model.cfg: The configuration file where we set the characteristics of the inversion, as the number of samples, the number of steps in the chains, and the distributions to be used as initial and run priors. The initial priors define the initial set of samples used by the program, the starting models from which model parameter space will be explored during the sampling process. Instead, the run priors define the distributions used as $f_{prior}(\theta)$. An initial prior can be more restrictive that a run prior, requiring to satisfy a certain condition over the initial samples, as in this study case, where we impose that the magnitude of the event lies in a certain range (*Minson et al.*, 2013, using the Dirichlet distribution).

The inversion directory must contain the files altar.cfg specifying how altar works the problem, for example, if it will be treated in a sequential scheme, using GPU or MPI parallel.

In the next sections we will analyze the information used as input for AlTar to solve the inverse problem for the static slip distribution of the Tocopilla earthquake.

4.6.2   Sensitivity Kernels for GPS and InSAR data used in this study

An optional input for AlTar is a folder with the sensitivity kernels according to the formalism developed by *Duputel et al.* (2014). These are used to calculate the uncertainties associated to the prediction of the physical model.

The matrix $\mathbf{C_p}$ expresses how the uncertainties in the definition of $\mathbf{G}$, due to the use of fixed values for uncertain parameters (e.g., fault geometry, elastic structure, etc.), affects the predicted data. To calculate $\mathbf{C_p}$ the program requires the sensitivity kernels respect to the uncertain variables of the physical model. These kernels are evaluated using numerical derivatives over $\mathbf{G}$ with respect to uncertain elastic properties of the medium. Once calculated, the covariance in the predicted data associated to these variables (the $\mathbf{C_p}$ matrix) is partially known, being necessary to evaluate a particular model to fully complete the calculation. The model used is updated in each annealing stage in AlTar, so the error is updated too. The particular model implemented in AlTar is the mean solution of all the ensemble of sampled model at a given iteration of the algorithm.

The sensitivity kernels used in this thesis come from the variables v_p, v_s and ρ in the 15 layers conforming the structure of the medium as defined by *Husen et al.* (1999). As we lack the knowledge of the uncertainties of the estimation of the elastic structure of the medium, we compute $\mathbf{C_p}$ using a 10% logarithmic perturbation on actual values of v_p, v_s and ρ in order to represent their uncertainties.

A preliminary estimation of $\mathbf{C_p}$ can be made by using the described kernels and a particular model solution obtained using another methodology, in this case, the regularized least squares method (See appendix A). The estimated model prediction error standard deviation can be seen in Figure 4.9 for GPS and InSAR (track 96). From the figure we can see that standard deviations associated to this source of error are close to 2 cm, larger than the observational errors, which are closer to 1 cm as can be seen in Figures 4.4 and 4.5, for GPS and InSAR data, respectively.
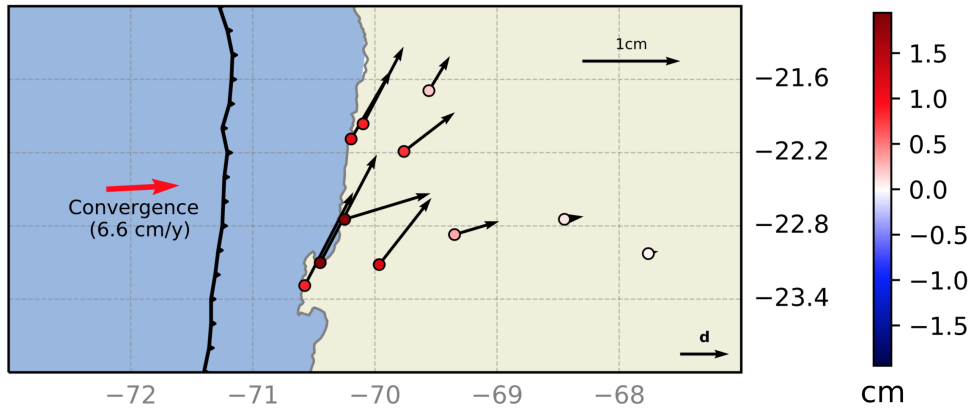
(a) Model prediction uncertainties for GPS data
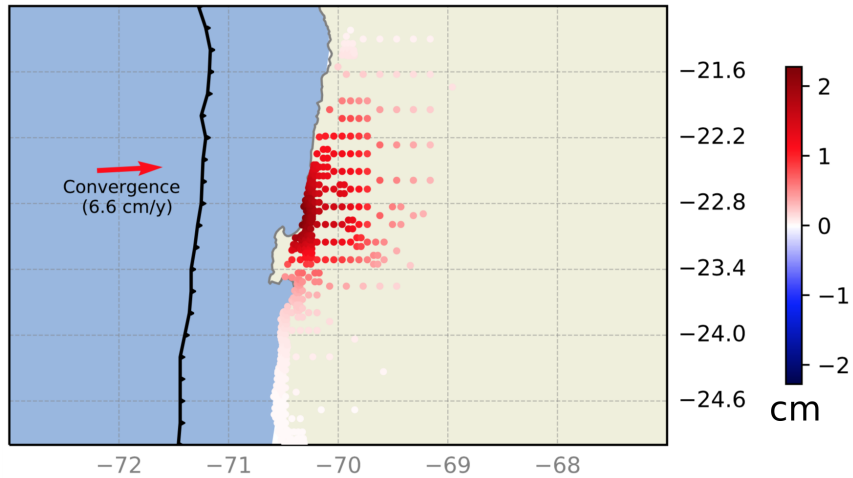
(b) Model prediction uncertainties for InSAR, track 96
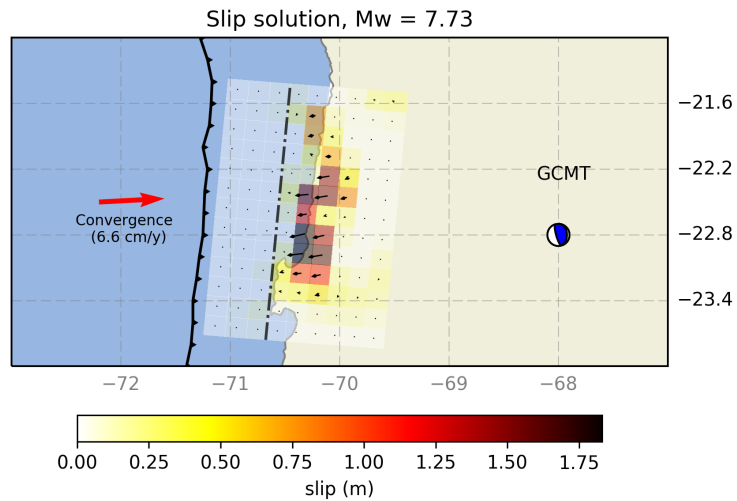
(c) Slip model used to compute $\mathbf{C_p}$.

Figure 4.9: Standard deviation of errors in the forward model prediction of GPS (a) and InSAR (b) data, calculated considering a 10% of error on values of v_p, v_s, ρ and using a slip model of the Tocopilla earthquake (c) obtained by a regularized least squares inversion (See Appendix A).

4.6.3 Definition of Initial Distributions and Prior Distributions

Before performing an inversion, AlTar requires to define prior distributions. It allows to generate an initial set of samples using an initial distribution and separately set the prior distribution of the model parameters that will be used during the sampling process. These distributions can be settled for the different subsets of parameters based on the knowledge about the slip components (parallel to rake and perpendicular to rake), as well as the additional parameters associated with the ramp defined for InSAR data.

The initial distribution chosen for the parameters in the parallel component of slip is the Dirichlet distribution. This choice incorporates previous knowledge about the magnitude of the earthquake since it ensures that magnitudes of initial samples will be close to the previously calculated magnitude for this earthquake, namely M_w 7.7. These samples are a starting point for the algorithm and do not restrict the moment magnitude of the sampled models during the next steps of the sampling process.

From the initial distribution the samples are redistributed and new samples are generated according to the intermediate distributions that converge to the a posteriori distribution of the model parameters θ. To generate samples they must obey the prior distribution associated with model parameters in the running process $f_{prior}(\theta)$ (as well as the restrictions imposed by the likelihood function that represents the physics of the forward model). These run priors must be satisfied during all the sampling process. We will discuss some choices here, particularly the uniform, Gaussian, and truncated Gaussian distributions.

The uniform distribution means that the samples does not have a bias for a certain value in an interval. Consequently, all values in the interval have the same probability. This is usually called state of homogeneous information for cartesian model parameters (*Tarantola*, 2005, , section 2.4).

The Gaussian distribution has a symmetrical pdf with a maximum value about its mean μ, and characterized by a variance σ^2 limiting the region where most of the probability resides. For instance, the confidence interval defined by $[\mu - 3\sigma, \mu + 3\sigma]$ contains near 99% of the total probability.

Based on previous studies of this earthquake we chose a mean $\mu = 0$ m and $\sigma = 1$ m, therefore the region with higher probability resides in ± 3 m. The symmetry of this function allows to consider a certain amount of backslip as a possible solution, therefore we use a truncated Gaussian distribution, that able us to limit the backslip set to 0.5 m.

Meanwhile, as we expect little slip to occur in such direction distribution, the distribution chosen for the perpendicular to rake component of slip is the Gaussian $N(\mu = 0$ m, $\sigma = 0.1$ m) for both, the initial prior and the running process prior. It should be noted at this point that the acceptance rate depends strongly of the interval chosen for the parallel component of slip, as will be discussed in the next sections.

Figures 4.10 and 4.11 show the marginal distributions for slip parameters over the fault geometry for the parallel to rake and perpendicular to rake directions, respectively. Figure 4.10 shows the initial samples of the parallel to rake component of slip, and follow the beta distribution, which is the marginal of a Dirichlet. In this case, parallel slip amplitudes are restricted so that the magnitude of the initial samples is close to Mw 7.7. Figure 4.11 shows

the initial samples for the perpendicular to rake component of slip, which was chosen as Gaussian.

Pdfs slip parallel to rake (positive backdip)(median in colors, Mw = 7.56)
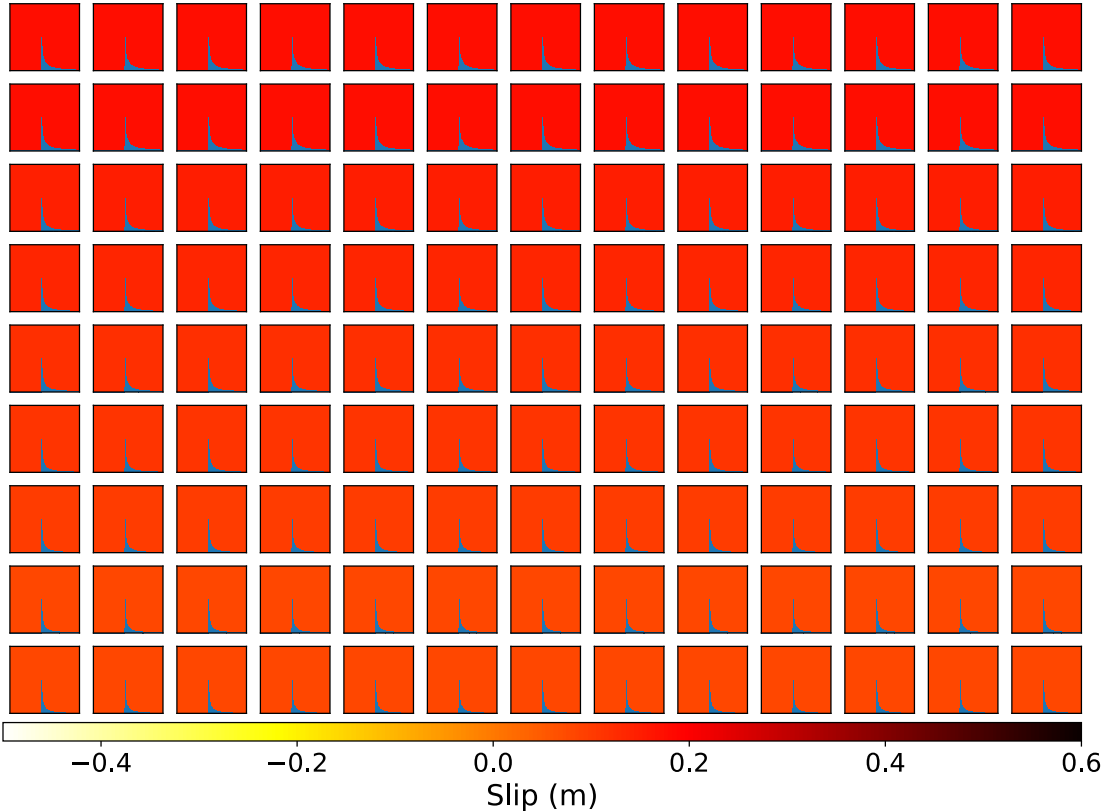
(a) Marginal histograms in the parallel to rake direction
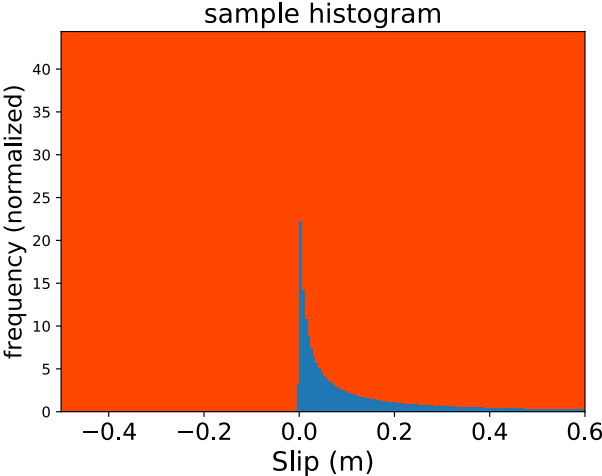
(b) Lower right corner histogram from a)

Figure 4.10: Marginal histograms representing the distributions of the initial samples generated for the parallel to rake component of slip a) for the whole fault, and b) for a particular subfault (lower right corner). The histograms follow the Dirichlet distribution with $\mu = 7.7$ and $\sigma = 0.5$. Colorscale represent slip amplitude.
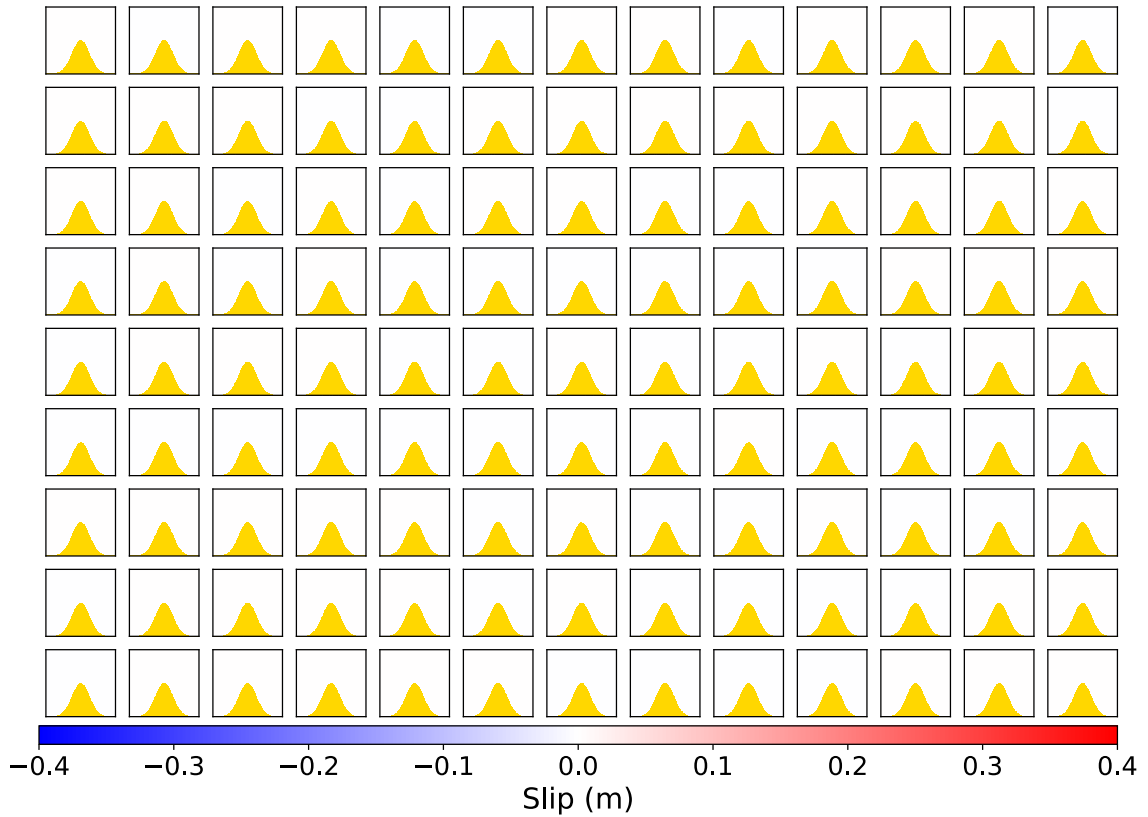
(a) Marginal histograms in the perpendicular to rake direction
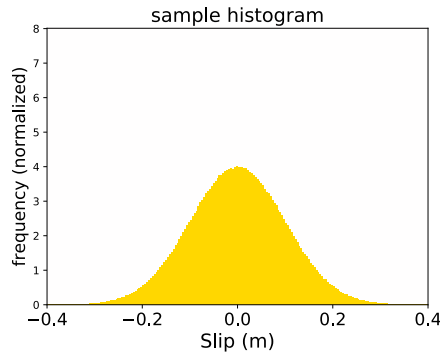
(b) Lower right corner histogram from a)

Figure 4.11: Initial samples generated for the perpendicular to rake component of slip. The marginal histograms shown are Gaussians with $\mu = 0$ m and $\sigma = 0.1$ m. The number of samples used is 3×10^6. a) Histogram for the whole fault and b) a particular subfault. Colorscale represent slip amplitude.

It is important to notice that the slip in the perpendicular to rake component is chosen as symmetrical and centered in 0 m, by contrast with the parallel to rake component, where we favored a positive slip. The information we are using here is that the slip corresponds predominantly to an inverse fault.

In this section we will discuss the inversion results obtained using the Bayesian methodology implemented in AlTar for estimating static slip distribution of the Tocopilla (Mw 7.7) 2007 earthquake. We contrast the solutions of the static problem obtained in different cases, incorporating GPS and InSAR data, and using both observational errors and those corresponding to the prediction of the (forward) physical model.

Recall that, regardless of the linear forward (physical) model, the part of the inverse problem formulation in which the data has a role is the Likelihood function

$$L(\boldsymbol{\theta}) = \gamma e^{-\frac{1}{2}(\mathbf{G}\boldsymbol{\theta}-\mathbf{d_{obs}})^T \mathbf{C}_\chi^{-1}(\mathbf{G}\boldsymbol{\theta}-\mathbf{d_{obs}})}, \tag{4.21}$$

where $\mathbf{C}_\chi$ is the misfit covariance matrix, i.e., associated to the errors of the difference $\mathbf{G}\boldsymbol{\theta} - \mathbf{d_{obs}}$. As mentioned in previous chapters, the misfit covariance matrix is composed as

$$\mathbf{C}_\chi = \mathbf{C_d} + \mathbf{C_p} \tag{4.22}$$

where $\mathbf{C_d}$ and $\mathbf{C_p}$ are the covariance matrix associated to observational and forward model prediction (epistemic) errors, respectively.

Therefore, when performing an inversion accounting only for observational uncertainties, we assume that the forward model is perfect, thus setting $\mathbf{C_p} = \mathbf{0}$. On the contrary, when considering both, observational and epistemic uncertainties, $\mathbf{C_p} = \mathbf{C_p}(\mathbf{m_{slip}})$ and is calculated using the formalism of *Duputel et al.* (2014). In both cases, we assemble $\mathbf{C_d}$ directly from the formal uncertainties obtained from the GPS or InSAR measurements, when corresponds.

4.7   Results

4.7.1   Slip Estimates Constrained by GPS Data

In this section we obtain slip estimates constrained solely by GPS observations of static coseismic surface crustal displacements of the Tocopilla earthquake. Here we will compare the slip solutions obtained when we account only for observational errors, as well as when we account for both, observational and epistemic (forward model prediction) errors during the inversion in order to compare their effect in the solution of the slip estimation problem.

Inversion Accounting only for Observational Uncertainties

The solution of the inverse problem in the Bayesian framework is the joint posterior distribution of the slip parameters. Particularly, we obtain a numerical approximation of the inverse problem solution, as we represent the joint posterior distribution by a finite set of its samples, with their associated posterior distribution pdf values. Regarding the sampling process, the number of samples is 3.000.000 and each Markov chain has 5.000 steps and we used a uniform $U(-0.5 \text{ m}, 10 \text{ m})$ distribution for rake parallel slip as the run prior and a Gaussian distribution $N(0 \text{ m}, 0.1 \text{ m})$ for the rake perpendicular slip prior. The convergence of the CATMIP algorithm took 41 iterations to reach the posterior pdf of the slip distribution.

From the ensemble of sampled models, constrained in this case by the GPS observations,

three particular models can be used to represent the whole ensemble: the mean, median and ML (maximum likelihood) models. These models are shown in Figure 4.12, as the distribution of slip amplitude for each of the three aforementioned models.

From Figure 4.12, we observe that mean and median slip models are very similar, while the ML model looks more scattered. Such a behavior is not surprising, as mean and median models represent the behavior of the ensemble of models, while the maximum likelihood model is chosen as the slip model that has maximum likelihood value from the ensemble of sampled models. The mean and median slip models exhibit zones with larger slip distributed in two main patches located near the coastline in map view, and in the corners of the parameterized fault surface that are closer to the trench. As we analyze in the following, the first features (near the coastline) can be interpreted as part of the source of the Tocopilla earthquake, and the latter features (near trench corners) correspond to inversion artifacts, usually seen in the edges of the solutions of inverse problems, typically associated to over fitting of observational or epistemic errors.

On one hand, the slip pattern off the edges of the parameterized fault surface, near the coastline, exhibit two zones of larger slip that coincide with the two main asperities ruptured by the earthquake that were also identified by other authors (e.g., *Minson*, 2010; *Béjar-Pizarro et al.*, 2010; *Delouis et al.*, 2009; *Peyrat et al.*, 2010).

On the other hand, the aforementioned artifacts can also be related to the finitude of the fault geometry, in the sense of the sizes of the subfaults in which the fault surface is discretized. Since there is no data offshore, as we only have onland GPS data, such observations poorly constrain the slip located at fault portions closer to the trench, because slip in these zones have displacement predictions that have a small contribution to predicted data of a whole slip model. So, slip models with high amplitudes near the trench are easily allowed during the inversion. Such effect will be seen more clearly in the next sections, where we explain the concept of *Sensitivity* (*Ortega*, 2013), which is a measure of how sensitive are the predicted data to the slip of each element of the fault plane, as well as when we explore the marginal distributions of slip at each discretized element of the fault surface.

As we are solving the slip estimation problem using a fully Bayesian framework, the solution of the inverse problem is characterized by the obtained ensemble of samples representing the joint a posteriori probability distribution of slip at the fault. The obtained ensemble of samples can also be analyzed by computing the marginal pdf's of slip on each direction and subfault and the correlation between such model parameters. For instance, Figure 4.13 shows the marginal histograms (approximating the corresponding marginal pdf) of rake parallel slip at each subfault from the ensemble of sampled models obtained during the inversion process. The zones with higher slip in the inner part of the fault area tend to have more symmetrical and peaked distributions, which resemble the pdf of a Gaussian or truncated Gaussian distribution. Such fault regions are surrounded by zones with lower probability for large slip, in which the slip histogram is more similar to the pdf of a Beta distribution, which is the marginal of the Dirichlet distribution. This is similar to the marginal distributions seen in figure 4.10, which can be related to the poor constraint that the data have on slip on such subfaults with their a priori chosen size, or that the actual slip value is very constrained to have very low values. The histograms at the corners of the fault show broader distributions (especially near the trench), indicating a larger error in the determination of slip at such fault elements.

The median model was chosen to represent the solutions, since it gives a notion of the central value of the distribution (slip values have the same probability of being larger or smaller than the median value). It better recovers the expected behavior of the slip solution in the sense that slip at the edges is less likely to show inversion artifacts. The median model fits the observations as can be seen in Figure 4.14, where the magnitude obtained is M_w 7.87.
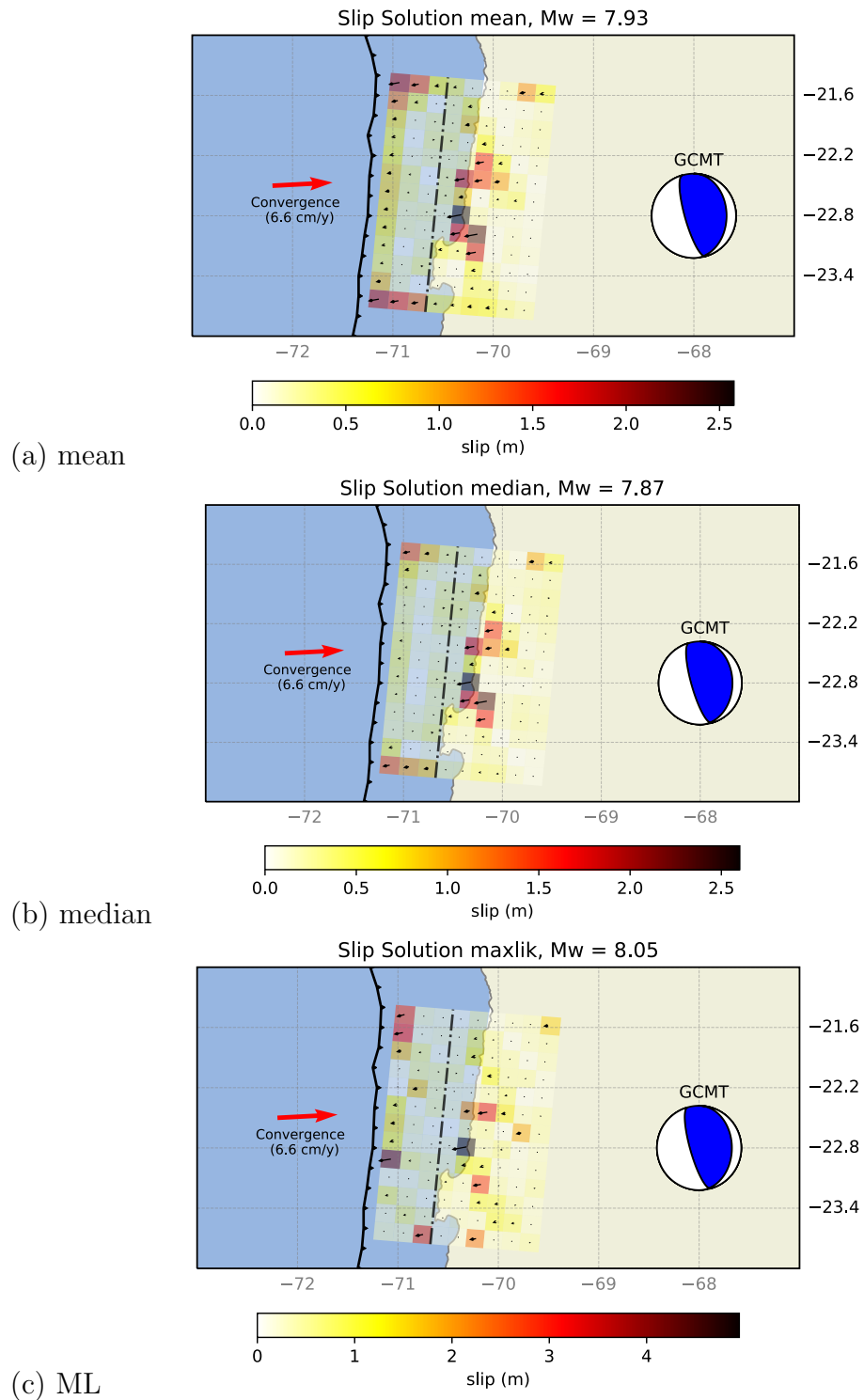
Figure 4.12: Map view for the mean of the solutions, median and ML (maximum likelihood) solution for the slip inverted using GPS data and only the observational source of error. The dashed line indicates the kink in the fault geometry.
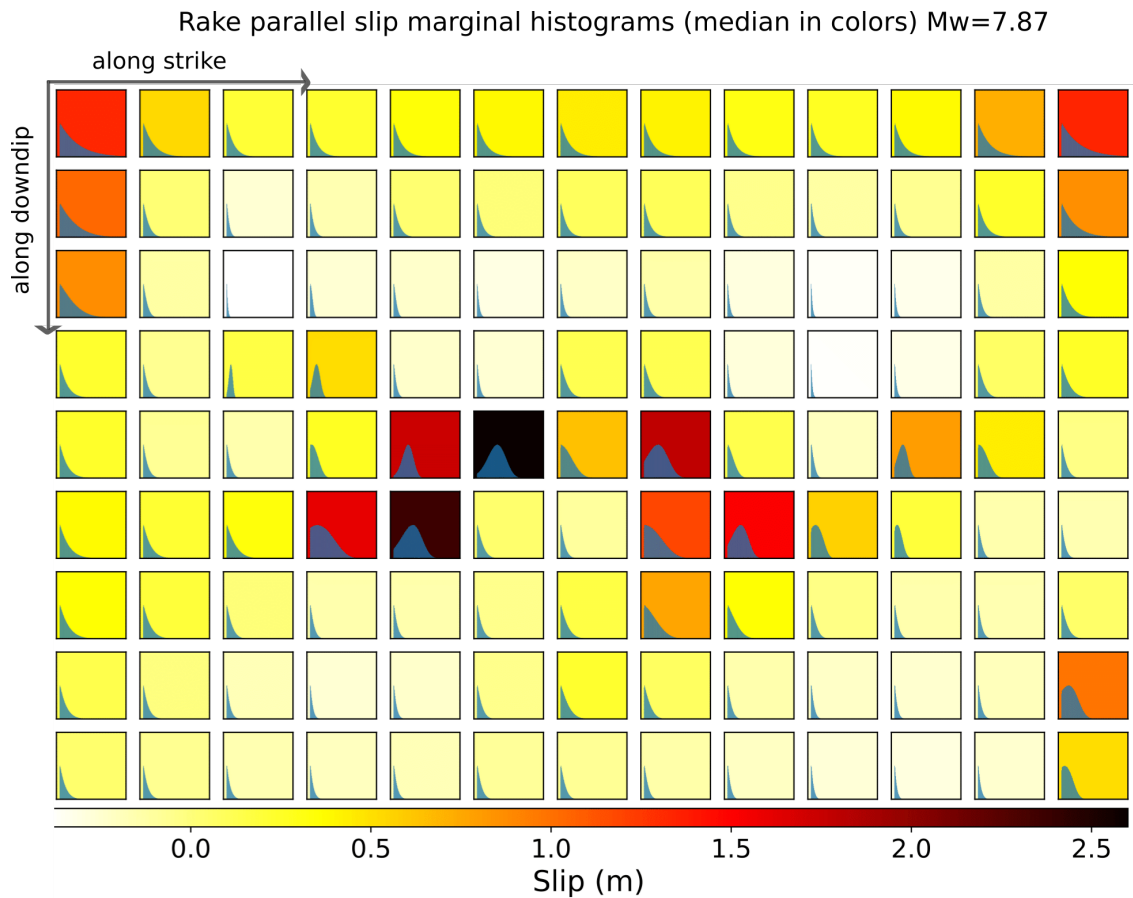
(a) Marginal histograms in the parallel to rake direction
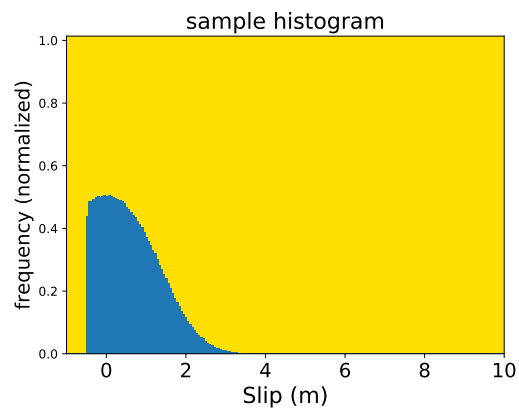
(b) Histogram of lower right corner subfault

Figure 4.13: a) Marginal histograms of the slip parameters over the fault geometry in the case where we consider only the observational source of error in the inversion using GPS data. The background color indicates the median model obtained from the ensemble of sampled models. b) Histogram showing the lower right corner element (all the subfaults have the same interval as b). Horizontal and vertical directions on each plot along strike and along updip directions respectively.
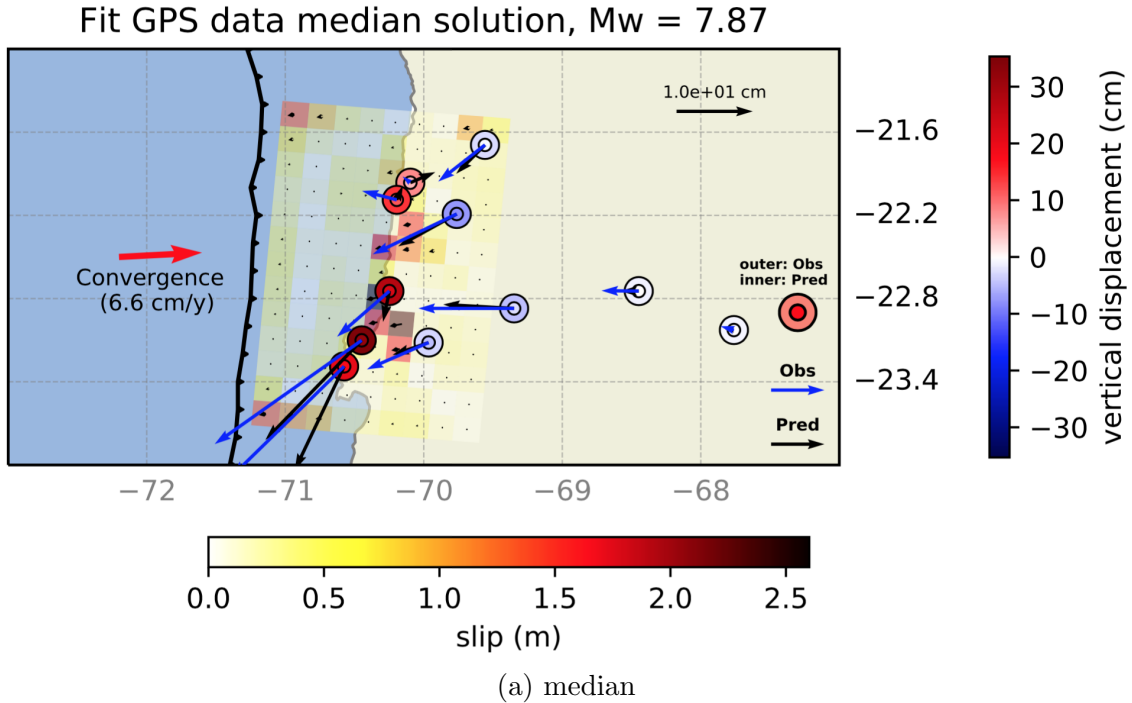
Figure 4.14: Comparison between observed data and predicted data for the median solution obtained in the Bayesian inversion using GPS data and incorporating only observational error. Arrows indicate horizontal displacements and the color of circles is scaled to the vertical displacement seen in the colorbar (in m).

Inversion Accounting for Observational and Epistemic Uncertainties

Here we present the inversion results for models constrained by GPS data and considering both, observational and epistemic uncertainties. The latter is represented by the covariance matrix $\mathbf{C_p}$, that models the uncertainty from the 1D velocity model. Regarding the inversion algorithm, the inversion was performed using 3.000.000 samples and 5.000 steps in the Markov chains. We used a uniform $U(-0.5 \text{ m}, 10 \text{ m})$ distribution for rake parallel slip as the run prior and a Gaussian distribution $N(0, 0.1 \text{ m})$ for the rake perpendicular slip prior. The process took 38 annealing steps to reach the posterior pdf of the slip distribution. As in the previous case, we show mean, median and ML models in Figure 4.15. We must note that during the processing of inversions using GPS data the scaling factor regulating the weight of the covariance matrix of the samples in the model was modified. This scaling factor $c = a + bR$ modulates the step size in the random walk, where R is the acceptance rate of the Metropolis sampling associated to certain chain while a and b are constants defined previously which depend on the particular problem solved. Applied to earthquake modeling these values were chosen as $a = 1/9$ and $b = 8/9$ by *Minson* (2010) obtained by performance tests, and those were the values used by default in AlTar. The configuration of the inversion was modified in such a way that the scaling factor acceptance weight is higher. We chose a value $b = 0.95$ and a rejection weight 0.05. That is because we chose a uniform prior distribution for parallel to rake components in an interval that makes extremely easy to fall outside the region for some component of the sample, favoring the rejection. So, making the step size smaller we increase the probability of acceptance.

The mean and the median are very similar but differ more notoriously at the edges and corners of the fault area. The slip pattern of these solutions agree with those describing two zones of higher slip, with a maximum slip near 2.6 m. This, reached at the southern end of the rupture zone. These two large slip patches are connected by a shallow slip zone, but the main rupture occurred deeper than the depth of the abrupt variation in dip in the fault geometry.

The ML solution presents some extreme characteristics. This is rougher than the other models and slip at the corners of the discretized fault is even more notorious (Figure 4.15)

Regarding the choice of accounting for uncertainties in the prediction of the forward model through the covariance matrix $\mathbf{C_p}$, a comparison of mean and median slip distributions from Figures 4.12 and 4.15, reveal that accounting for epistemic errors drastically reduces the amount of slip at the edges and corners of the parameterized fault surface, compared to the case in which we only account for observational uncertainties where we overfit the data.

Also, one of the main differences between the slip models obtained using and not using the covariance matrix $\mathbf{C_p}$ is in the portion connecting the two patches with higher slip at the source region of the earthquake. Particularly, slip connecting the two main asperities seems to be shifted offshore with respect to that shown in previously published solutions. As inferred by this study and previous authors, the downdip kink in the slab geometry proposed by *Contreras-Reyes et al.* (2012) might act as a barrier to rupture propagation, assertion that is true for the case of this earthquake, but needs to be further investigated when a larger earthquake occurs in the future in such seismogenic region. In the case where we neglect the epistemic error, we can see two isolated regions of higher slip in the median and mean slip model, while in the case where we consider this source of error there is a path of non null slip that joins these two portions.

As in the previous case, the obtained ensemble of sampled models can also be analyzed by computing the marginal pdf's of slip on each direction and subfault and the correlation between such model parameters. For instance, the marginal histograms for rake parallel slip are shown in Figure 4.16 along with the median model. Figure 4.17 shows the prediction of the median solution compared with the observed data.

There are differences in the marginal distribution shape depending on the patch location over the fault contact (seen in Figure 4.16). In zones where the median model shows maximum slip, the histogram shape is more symmetrical and wider than most of the distributions located at the surroundings. Most of the patches outside the zone of maximimum slip are more peaked, narrower and asymmetrical with samples clusterized close to left side of the slip interval. The shape of these histogramas resemble a beta distribution (see Appendix B). Histograms in patches close to the trench tend to be wider than in their inner neighbors, especially at the corners.

From the perspective of finding slip models that represents observations well, the inclusion of epistemic errors in the Bayesian inversion allows for a much better fit to the GPS coseismic crustal displacements, as can be observed when comparing the data prediction of the mean models for the case in which only observational errors are considered (Figure 4.14) and for the case in which we account also for the forward model prediction uncertainties (Figure 4.17).

(a) mean

(b) median

(c) maximum likelihood
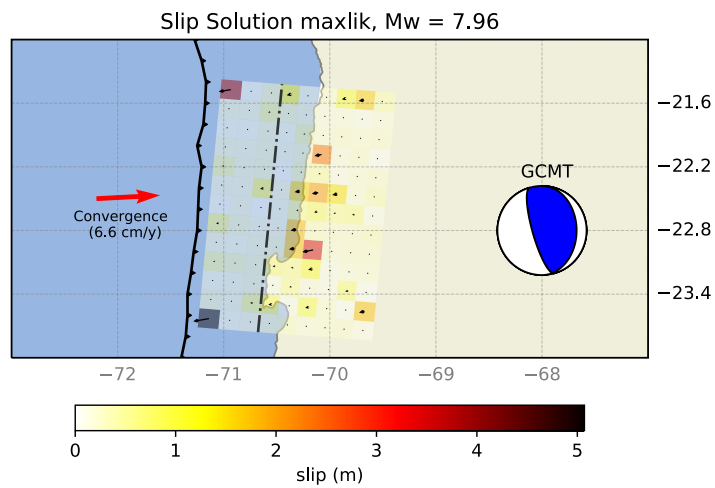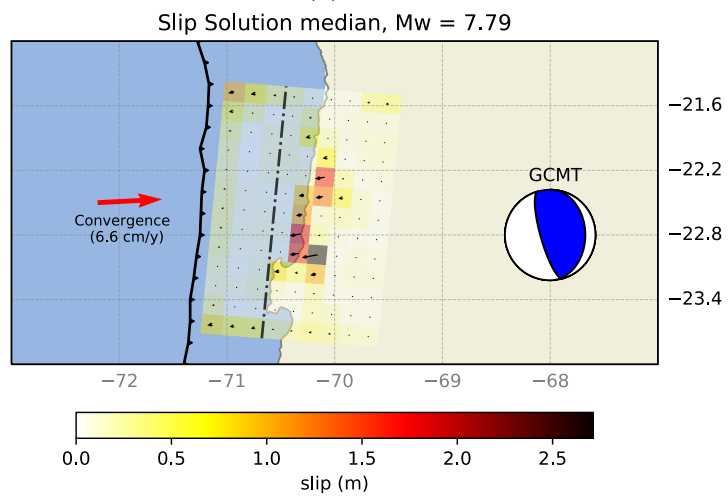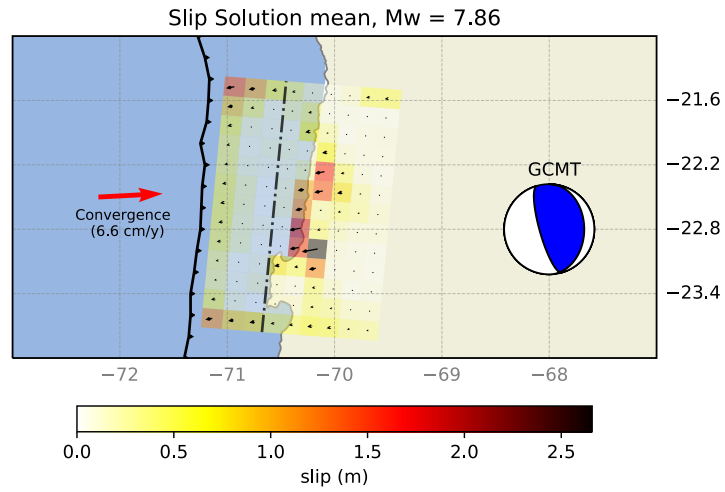
Figure 4.15: Mean, median and maximum likelihood solution for the slip obtained in the Bayesian inversion using only GPS data and incorporating both, the observational and epistemic error. Dashed- dotted line indicates the kink in the fault geometry along dip. Moment magnitude is indicated at the top of each panel.
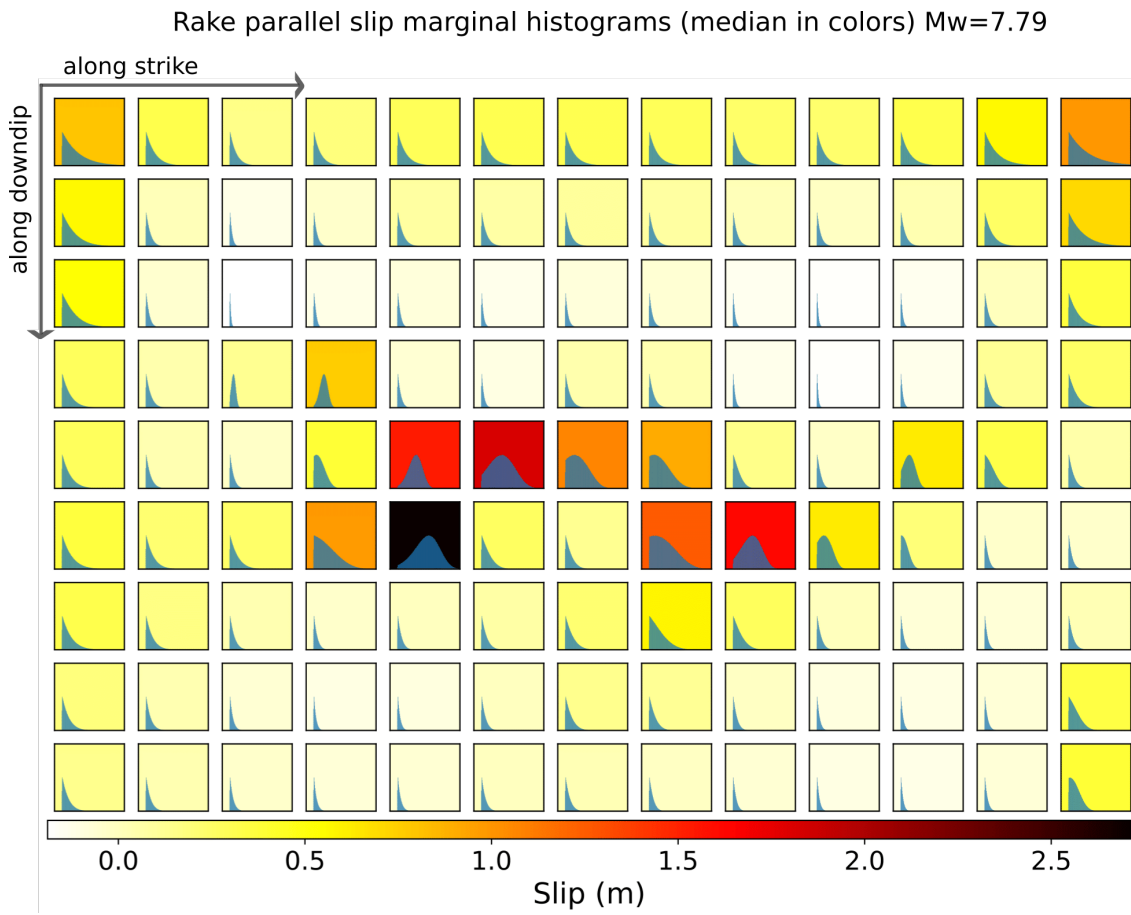
(a) Marginal histograms in the parallel to rake direction.
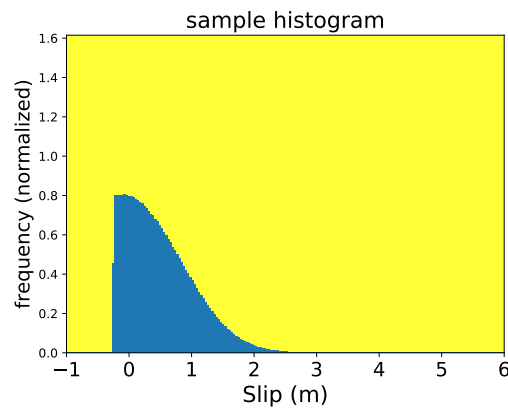
(b) Histogram of lower right corner subfault.

Figure 4.16: a) Marginal histograms of the slip parameters (rake parallel direction) over the fault geometry in the case when we consider the observational as well as the epistemic source of error. The color in background indicates the median solution. b) Histogram showing the lower right corner element (all the subfaults have the same interval as b along the x axis).
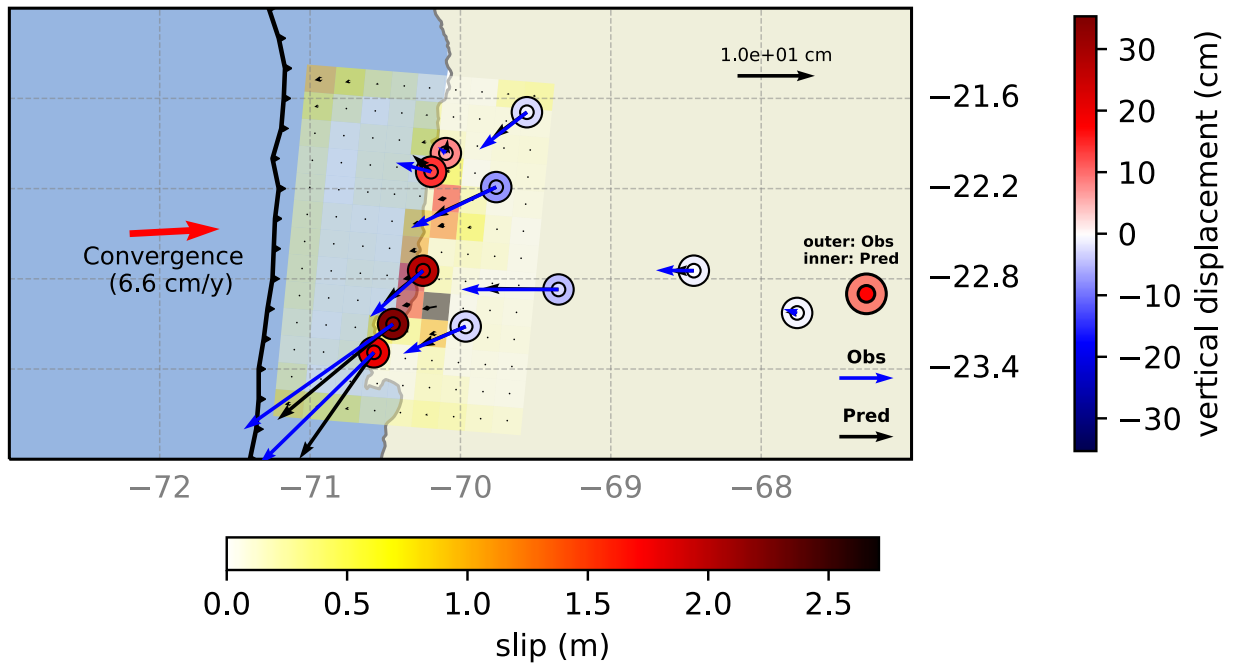
(a) median

Figure 4.17: Comparison between observed data and predicted data for the median solution obtained in the Bayesian inversion of the 2007 Tocopilla earthquake using GPS data and observational and epistemic error.

4.7.2 Results using GPS and InSAR data

When we add InSAR data, the formulation of the problem changes only in relation to the additional parameters that need to be estimated, as consequence of accounting for corrections to the InSAR data. Such correction arises because an InSAR image incorporates a systematic component that is unrelated to deformation, but to the misplacing and uncertainties in the repeating orbits of the satellites taking the pair of images needed to obtain the InSAR interferogram. The aforementioned errors add a long wavelength ramp to the estimated LOS displacements, which is usually modeled as a linear ramp (*Simons & Rosen*, 2015b). In this thesis we consider the case in which we add only one track of InSAR (track 96).

As in the previous section, as a first choice we consider a uniform distribution for the run prior fprior(θ) for rake parallel slip, with limits allowing only little back slip (-0.5 m) and a maximum amplitude of slip of 10 meters. Then, as a second step we consider instead a truncated Gaussian distribution[1] for rake parallel slip run prior TN(0m, 1m, -0.5m, 10m), in which we observe substantial improvements in efficiency and quality of solution of the inverse problem.

In both cases we reweighted the data errors in such a way that the error covariance matrix of the two sets have the same information content (*Minson et al.*, 2013, and references in there). We do this in order to fit the two datasets considering that the number of InSAR observations is larger than the GPS observations, so both datasets have the same importance in the inversion.

Assuming Gaussian distributed errors, the entropy of a distribution with mean μ and covariance matrix $\mathbf{C}_\chi$ is defined as

$$H(N(\mu, \mathbf{C}_\chi)) = \frac{1}{2} \log_2((2\pi e)^N |\mathbf{C}_\chi|) \tag{4.23}$$

where N is the dimension of the space of the random variable describing the errors.

The weighting factor α applied to the covariance matrix $\mathbf{C}_\chi^{\mathbf{InSAR}}$ seek to equally represent both data sets in terms of their entropy (using entropy as a measure of information content, see *Ortega* (2013) and references therein), then satisfy the following relationship:

$$\ln(|\mathbf{C}_\chi^{\mathbf{GPS}}|) = \ln(|\alpha\mathbf{C}_\chi^{\mathbf{InSAR}}|) \tag{4.24}$$

A similar strategy have been used by other authors considering aspects as the rms of the different data sets as *Béjar-Pizarro et al.* (2010). However, InSAR data spans a period of 1 month before and after the coseismic deformation, thus the InSAR data may include other sources of deformation, such as part of inter-seismic which might be negligible due to the short time span, but also can contain early post-seismic deformation that might be more relevant. Therefore, we do not expect to fully explain InSAR data with our modeling.

---

[1] A truncated Gaussian distribution is similar to a Gaussian distribution, but values of the random variable are bounded to a given interval. For instance, TN(0, 1, -1 , 1) is similar to a Gaussian distribution N(0,1) but with the additional restraint that the value of the random variable can only take values in the interval [-1, 1].

Slip Constrained by GPS and InSAR data : Uniform Distribution as Rake Parallel fprior(θ)

In this case, we perform the inversion using a uniform distribution in $U[-0.5m, 10m]$ as a prior pdf for the running process (run prior). The number of samples used is 2.000.000 and the number of steps in the Markov chains is 20.000. Note that here we use a reduced number of samples, 2/3 of the used for the inversion constrained only with GPS data, due to a memory limitation of the computational resources (available RAM on the Tesla K40 GPGPU cards). The process took 40 annealing steps to reach the posterior pdf.

As we are solving the slip estimation problem using a fully Bayesian framework, the solution of the inverse problem is characterized by the obtained ensemble of samples representing the joint a posteriori probability distribution of slip at the fault. The obtained ensemble of samples can also be analyzed by computing the marginal pdf's of slip on each direction and subfault and the correlation between such model parameters. From the ensemble of sampled models, we compute the mean, median models and select the ML (Maximum Likelihood) models, that are shown in Figure 4.18. Although those three models are more similar for this inversion case, probably because we are using longer Markov chains during the sampling process, for comparison purposes we analyze the median model. The estimated moment magnitudes are closer to M_w 7.9. The marginal histograms of slip and the median model are shown in Figure 4.19, and the fit to the GPS and InSAR data, given by the median slip model, is shown in Figure 4.20. The mean values of the ramp parameters obtained are $a = 3.0 \times 10^{-1} cm/100km$, $b = -5.1 \times 10^{-1} cm/100km$, $c = -2.1 \times 10^0 cm$, with standard deviations $\sigma_a = 2.8 \times 10^0 cm$, $\sigma_b = 8.1 \times 10^{-1} cm$, $\sigma_c = 3.5 \times 10^0 cm$.

As a general comment for all representative models shown in Figure 4.18, we retrieve a similar spatial slip pattern as in the previous inversion. Slip is basically characterized by two large slip patches. The marginal histograms in these zones of larger slip present a more symmetrical behavior and higher width than those patches on the surroundings. We observe that slip at the top corners of the parameterized fault surface is poorly constrained, which can be seen at the histograms on Figure 4.19. Here, marginal distributions at the top corners resemble the pdf of a beta distribution (the marginal of a Dirichlet), meaning that slip at the size of the discretized subfault can not be well resolved, and thus one should use a larger subfault size at those regions to obtain a better resolved slip value (see *Duputel et al.*, 2014; *Minson et al.*, 2013, and Appendix in this work.). As an alternative, one could use some additional prior information that allows to reduce the uncertainty on slip estimation as well as to reduce the size of the model parameter space in order to make it easier to sample. In the following section, we use the latter option, as given the available computational resources and the size of the model parameter space, it is harder for the CATMIP algorithm to find sampled models that explain well all the GPS and InSAR data, as evidenced in Figure 4.20. The difference is evidenced also in the acceptance rate of samples being lower when we use a uniform distribution compared with the case when we use, for example, a truncated Gaussian (these acceptance rates are 0.7% and 3%, respectively).
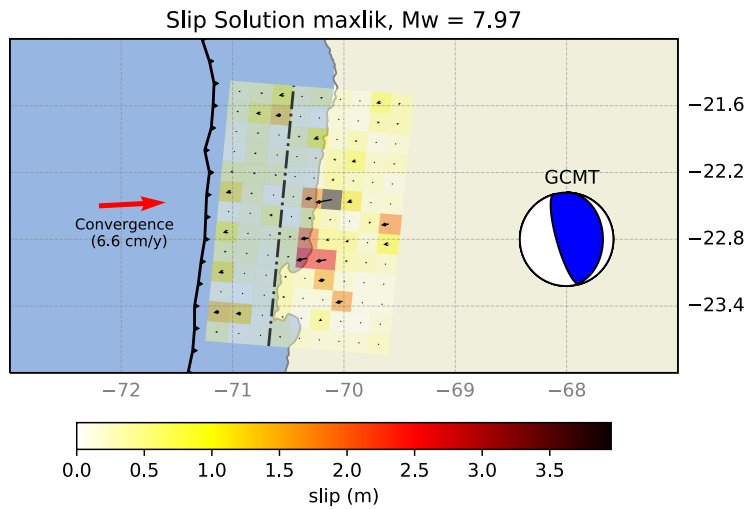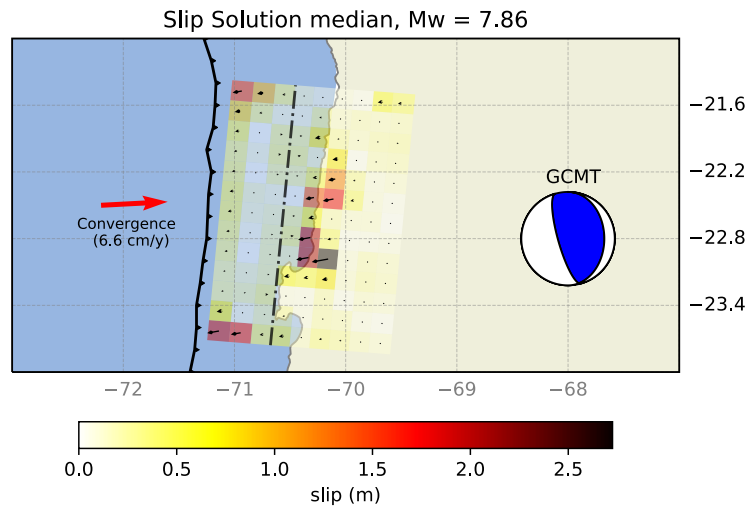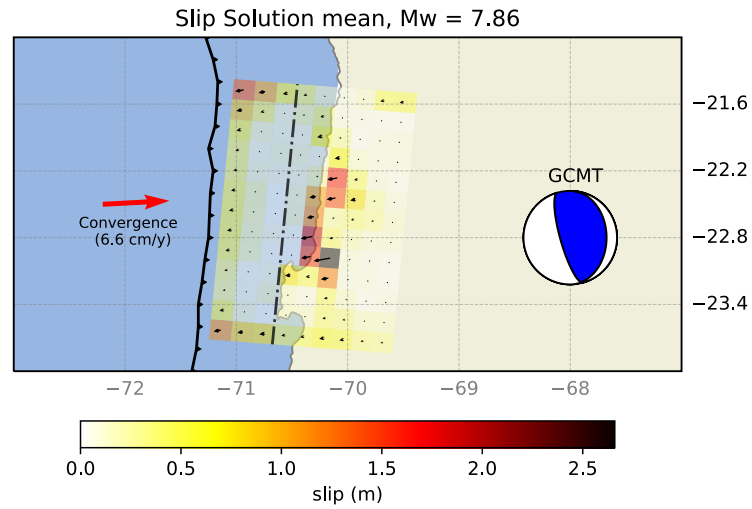
Figure 4.18: Mean, median and ML solution for the slip obtained from the inversion of the 2007 Tocopilla earthquake using GPS and InSAR data. The slip pattern was obtained using both, the observational and model prediction error. Moment magnitude is indicated at the top of each panel.
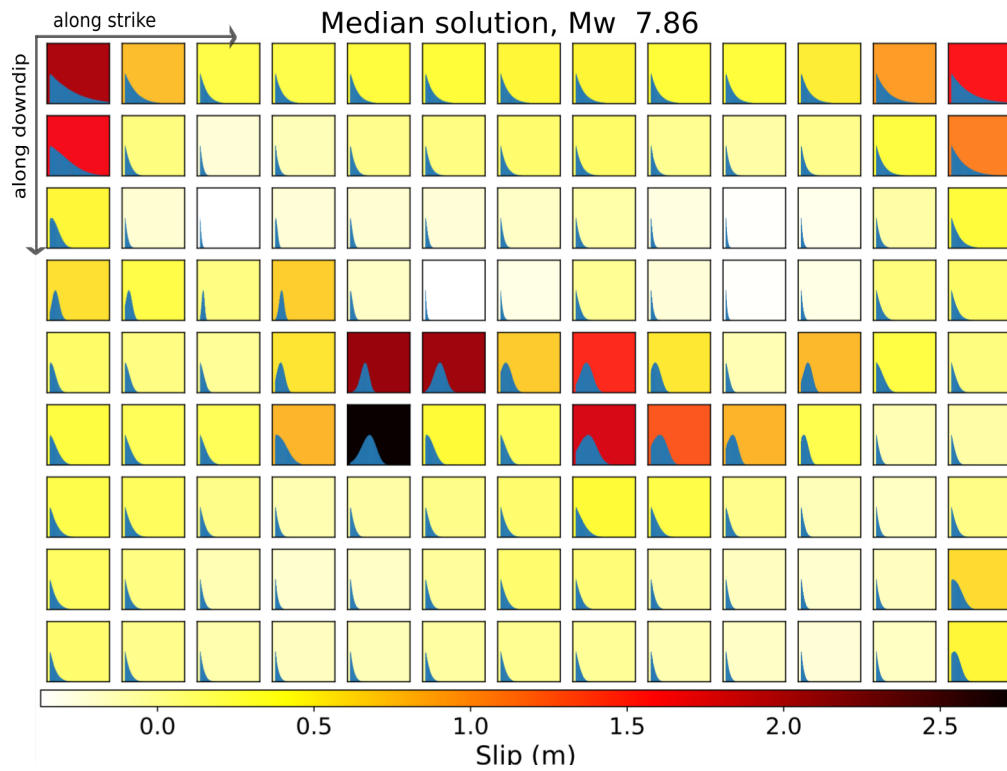
(a) Rake Parallel Slip Marginal Histograms
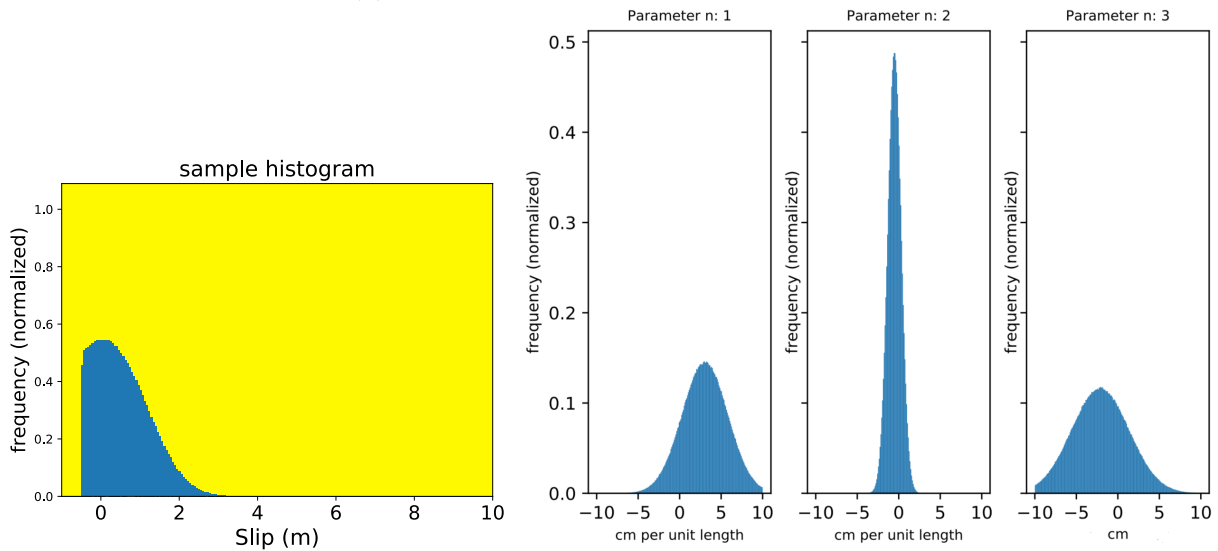
(b) Histogram of lower-right corner sub-fault

(c) Histograms Ramp parameters

Figure 4.19: a) Marginal histograms of the slip parameters (rake perpendicular direction) over the fault geometry in the case where we consider the observational as well as the epistemic source of error in the inversion using GPS and InSAR data with a run prior U(-0.5 m, 10 m). The color in background indicates the median solution. b) Histogram showing the lower right corner element. c) Marginal normalized histograms for ramp parameters, a, b and c, respectively.
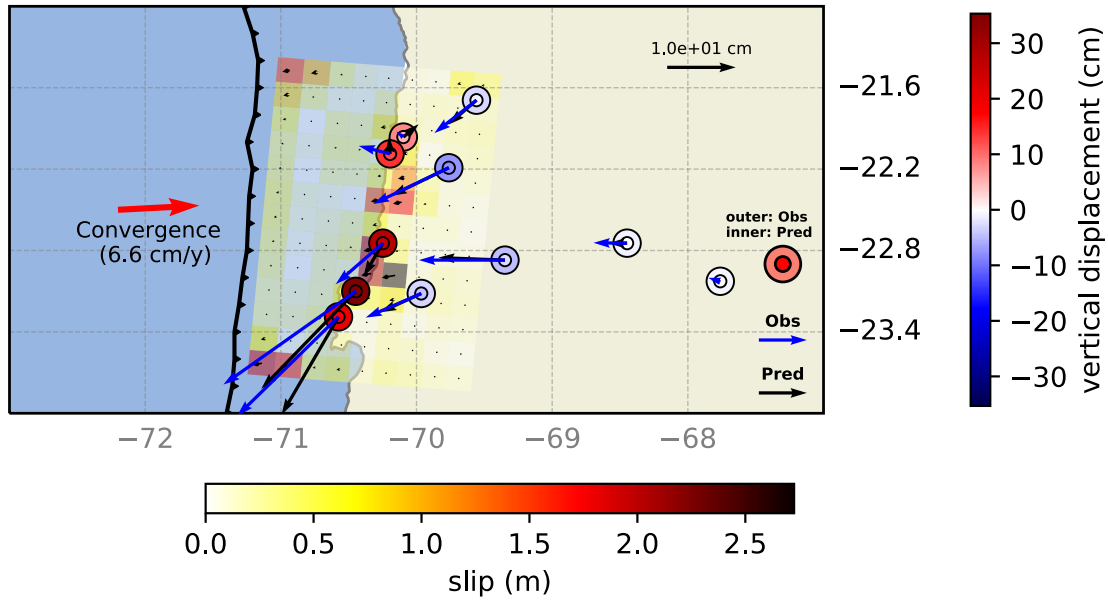
(a) Fit of GPS data
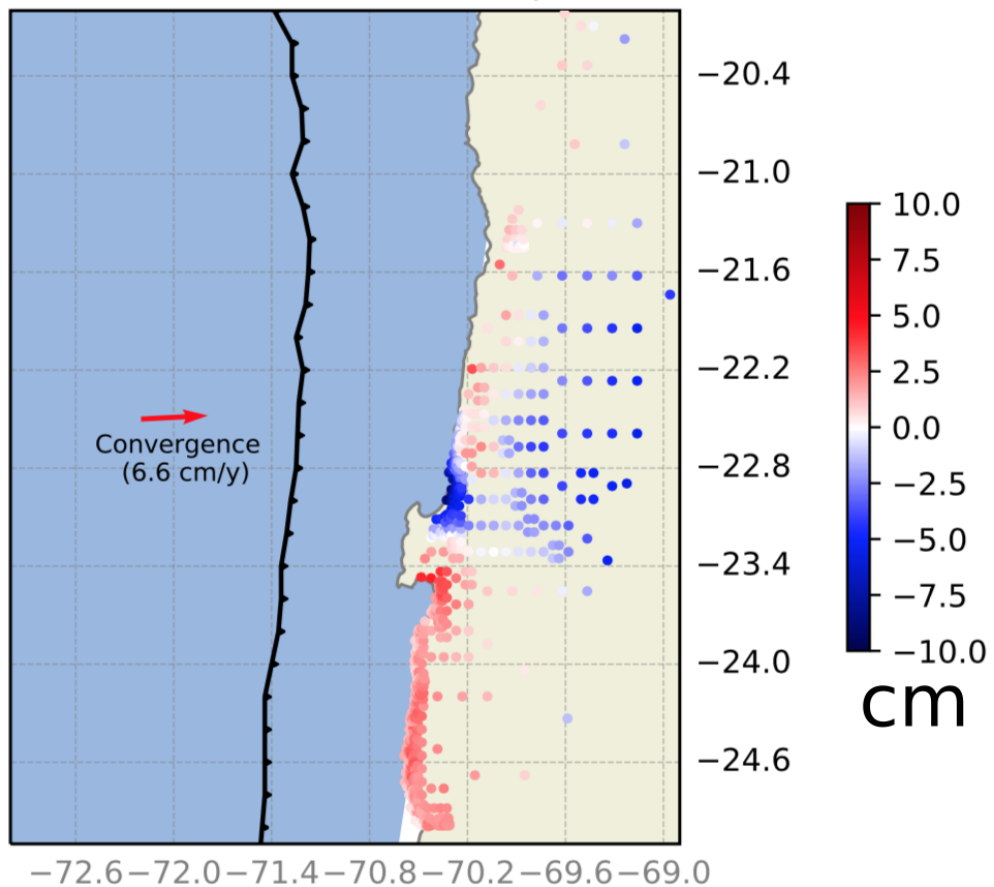
(b) Residual InSAR data

Figure 4.20: Fit for horizontal and vertical displacement components of the median solution resulting from the inversion using GPS and InSAR data with a run prior U(-0.5 m, 10 m) a) Comparison between the predicted GPS data and the observed GPS data b) Residual representation of InSAR data in map view with colorbar in cm.

Slip Constrained by GPS and InSAR Data : Truncated Gaussian Distribution as Rake Parallel fprior(θ)

In this section we analyze slip inversion results in which a truncated Gaussian distribution with a priori mean value $\mu = 0$ m, variance $\sigma^2 = 1$ m and truncation interval [-0.5 m, 10 m] is used as prior information for rake parallel slip (run prior). There are 2 main reasons for this choice, that there is a bounded region for slip where the higher probability resides, around 0, so subfaults poorly constrained by the observations will tend to have slip values closer to 0. Also, as we limit rake parallel slip to be reverse (allowing only a small fraction of normal slip for sampling purposes), when using a uniform prior, the rejection rate of the algorithm is high as we have several subfaults with slip near 0, thus using a truncated Gaussian, as it gives larger probability to slip values near 0 (larger than those values at the edges of the uniform pdf) it considerably reduces the rejection rate of the algorithm as more models tend to have rake parallel slip values near 0 and less near the edge of the interval (-0.1 m). As a consequence, we observed that using shorter Markov chains, one could achieve a better inversion result, thus improving as well the efficiency of the CATMIP sampler to solve for the slip inversion problem.

For this inversion case, the number of samples is 2.200.000 and the number of steps in the Markov chains is reduced to 5000. The sampling algorithm took 29 annealing steps to reach the posterior pdf of slip and InSAR ramp model parameters. Note that the total number of samples is the maximum allowed by the available computational resources, and the length of the Markov chains is selected using a trial and error process, so that it is the minimum number that allows for stability in the estimation of slip marginal distributions.

As we are solving the slip estimation problem using a fully Bayesian framework, the solution of the inverse problem is characterized by the obtained ensemble of samples representing the joint a posteriori probability distribution of slip at the fault. The obtained ensemble of samples can also be analyzed by computing the marginal pdf's of slip on each direction and subfault and the correlation between such model parameters. From the ensemble of sampled models, we compute the mean, median models and select the ML (Maximum Likelihood) models, that are shown in Figure 4.21. As those three models are very similar for this inversion case, for comparison purposes we analyze the mean model. All three estimated models have magnitudes closer to Mw 7.8. The major differences, comparing with the results obtained using a Uniform run prior for rake parallel slip, are at the edges of the fault geometry, especially in the fault elements close to the trench. By using the truncated Gaussian prior, the median slip model at those subfault is much closer to zero, which indicates again that such subfaults are poorly constrained by the observations dominated by the used prior information.

The marginal histograms of slip and the mean model are shown in Figure 4.22 for rake parallel slip and in Figure 4.23 for rake perpendicular slip, respectively. The fit to the GPS and InSAR data, given by the mean slip model, is shown in Figure 4.24.

The maximum likelihood (ML) model exhibits some rougher characteristics in the edges of the parameterized fault surface, with some patches with slip higher than 1.0 m. Thus, suggesting that a larger number of samples must be used during the sampling process in CATMIP, which is unachievable in this case as we are using all the available computational resources to perform the Bayesian slip inversion.

(a) mean

(b) median

(c) maximum likelihood
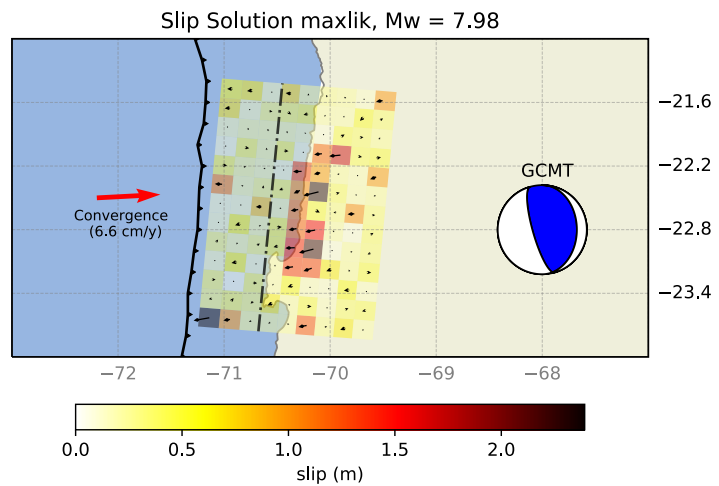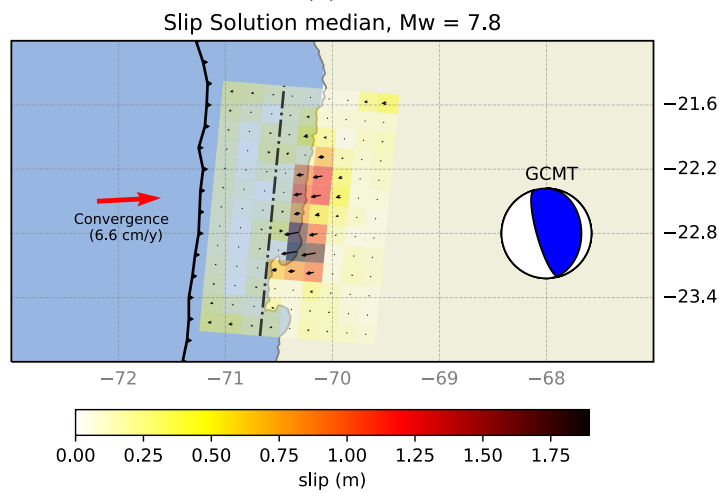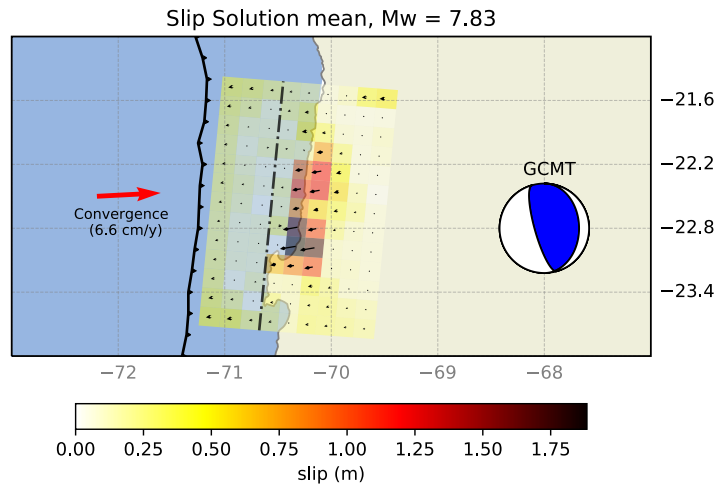
Figure 4.21: Solutions for mean, median and ML inverted slip using GPS and InSAR data and a truncated Gaussian as a prior pdf in the running process ($\mu = 1$ m, $\sigma = 1$ m and limits in an interval [-0.5m, 10m]).
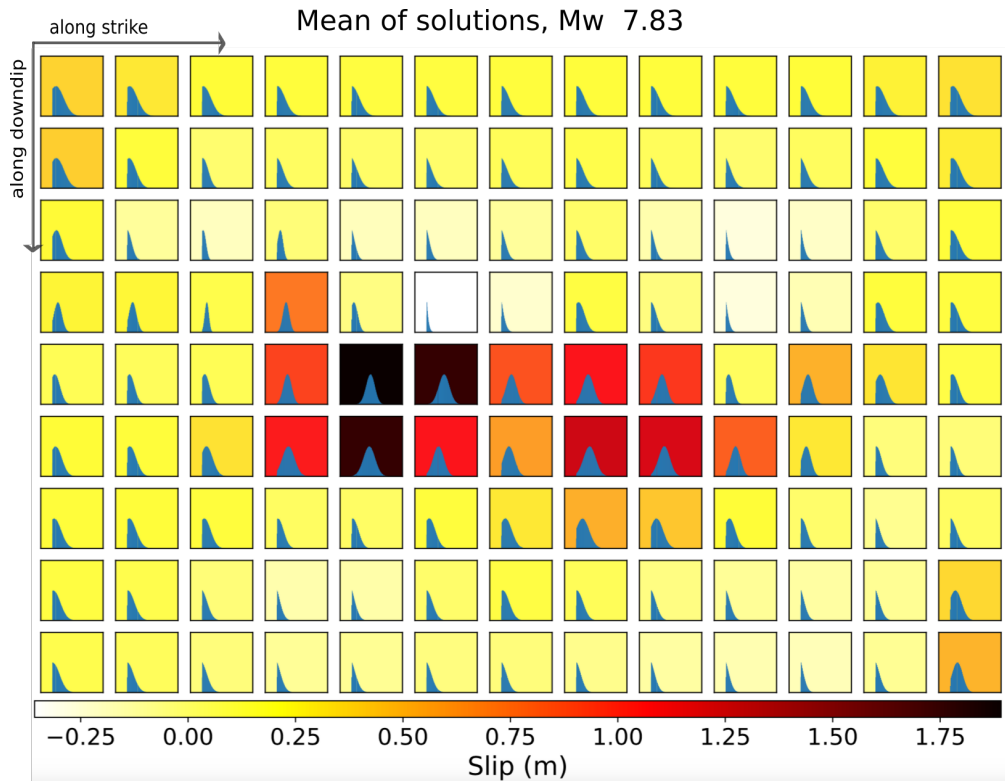
(a) Rake Parallel Slip Marginal Histograms.
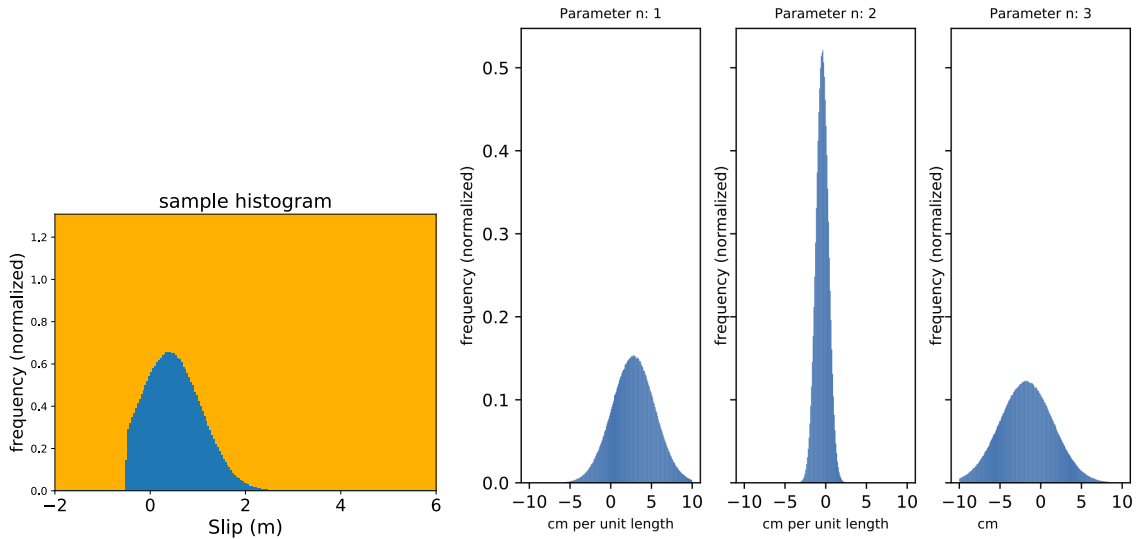
(b) Histogram of lower right corner subfault.

(c) Histograms for ramp parameters.

Figure 4.22: a) Marginal Histograms for parallel to rake component of slip. This inversion use the epistemic error $\mathbf{C_p}$ and a truncated Gaussian as a prior for the running process (not initial and run priors for the ramp, because results are basically the same after removing the ramp and using priors obtained using least squares). The background color indicates the mean solution. b) rake parallel slip histogram for the lower right corner element (all the subfaults have the same slip intervals along the x axis as in b). c) Histograms for the parameters associated to the linear ramp a, b and c. Horizontal and vertical directions in (a) and (b) resemble along strike and along updip directions respectively.
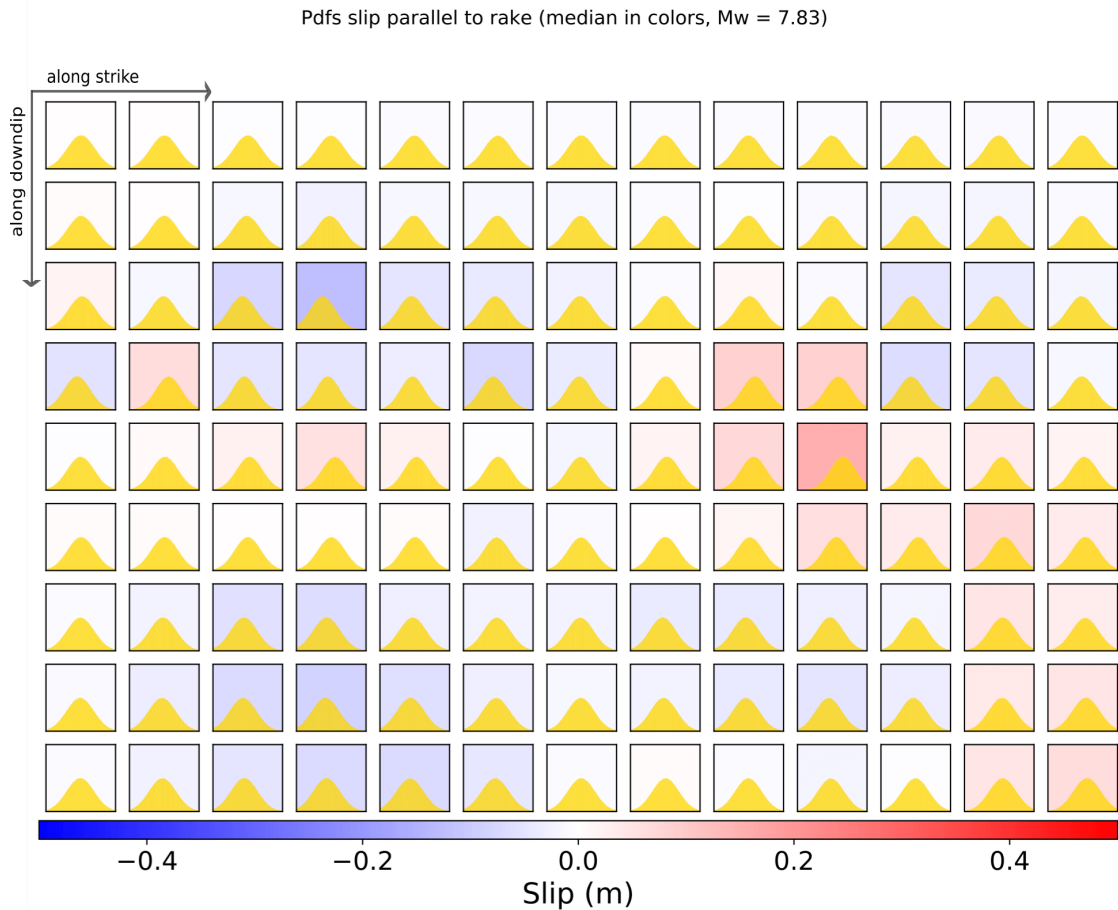
(a) Rake Perpendicular Slip Marginal Histograms
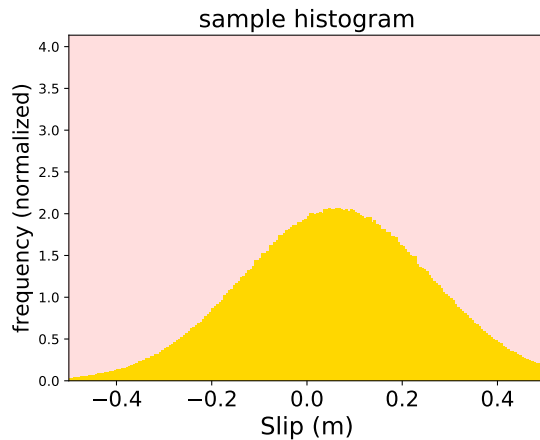
(b) Histogram of lower right corner subfault

Figure 4.23: a) Marginal Histograms for the perpendicular to rake component of slip. This inversion use the epistemic error $\mathbf{C_p}$ and a truncated Gaussian as a prior for the running process (not initial and run priors for the ramp, because results are basically the same after removing the ramp and using priors obtained using least squares). The background color indicates the mean solution of slip in m. b) Rake perpendicular slip histogram for the lower right corner element (all the subfaults have the same slip intervals along the x-axis as in b). Horizontal and vertical directions in (a) and (b) resemble along strike and along updip directions respectively.

From the previous figures, fault regions with higher slip have the highest information gain in the sense that a posteriori marginal pdf's for rake parallel slip differentiate the most from a priori ones (the ones that are most different from the beta pdf's). If we represent the slip distribution as the mean slip of the sampled models (Fig 4.21.a), there are 2 regions at the interior of the fault surface, located north and south of the slip distribution, that concentrate most of coseismic slip. Such regions are then identified as the 2 main asperities ruptured by the Tocopilla earthquake.

At the aforementioned asperities, maximum slip value is close to 2 m and was reached at the southern asperity, which is a slightly lower value than the slip solutions inferred by other authors. The northern asperity, located closer to the hypocenter, reached a maximum slip of about 1.7 m. Both slip regions are connected by a zone of lower slip amplitude compared to the maximum peak slip.

The representative slip distribution estimated in this work is different from the obtained by other studies, possibly due to incorporation of epistemic (forward model prediction) errors and the use of a more detailed geometry for the subduction interface. Most of the rupture process occurred deeper than the limit of abrupt variation in dip.

The ramp parameters have marginal pdf's that can be well approximated by the one of a Gaussian distribution (see Figure 4.23c). The mean value of the ramp parameters are $a = 1.6 \times 10^{-1}$ cm/100 km, $b = -1.29 \times 10^{-1}$ cm/100km, $c = 1.08 \times 10^{0}$cm, with standard deviations $\sigma_a = 6.1 \times 10^{-1}$ cm, $\sigma_b = 1.6 \times 10^{-1}$ cm, $\sigma_c = 7.5 \times 10^{-1}$ cm.

Regarding the fit to the GPS and InSAR datasets, observed and predicted GPS data as well as misfit residuals for InSAR data ($\mathbf{d^{InSAR} - G^{InSAR}m^{slip} - G^{ramp}m^{ramp}}$) obtained using the mean model and the observed data are shown in figure 4.24 a,b, respectively. In addition, InSAR data LOS displacement prediction of the mean ramp values, ramp corrected InSAR observations of LOS displacements and mean slip model InSAR LOS displacements prediction are shown in Figures 4.25, 4.26 and 4.27, respectively. Here we must note, that for this inversion case, in which we use the truncated Gaussian distribution as (run) prior information, the mean slip model estimated from the ensemble of sampled models fits much better the GPS and InSAR data, compared with the slip estimates obtained with uniform prior. The moment magnitude in this case is $M_w = 7.83$ We analyze the origin of such differences in the next section of this work.
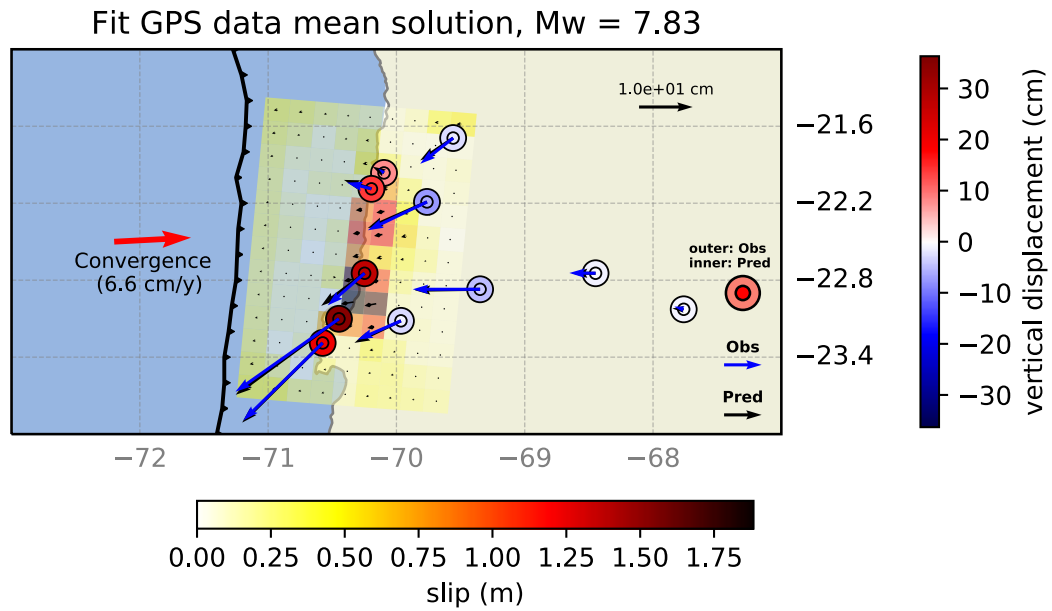
(a) Fit to GPS data, mean model from the ensemble of sampled models constrained by GPS and InSAR data.
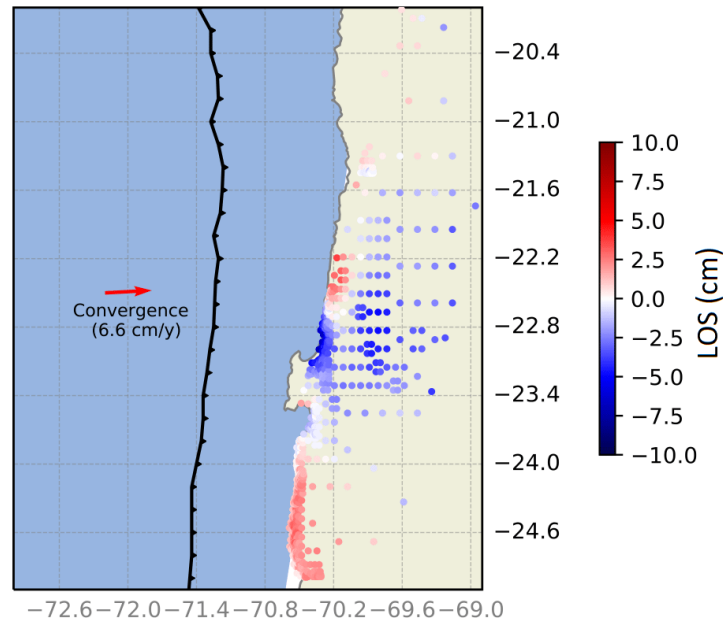
(b) Residual track 96, using mean model from the ensemble of sampled models constrained by GPS and InSAR data.

Figure 4.24: a) Comparison between observed and predicted GPS data for the horizontal and vertical static displacements. b) Residual of InSAR data (track 96).
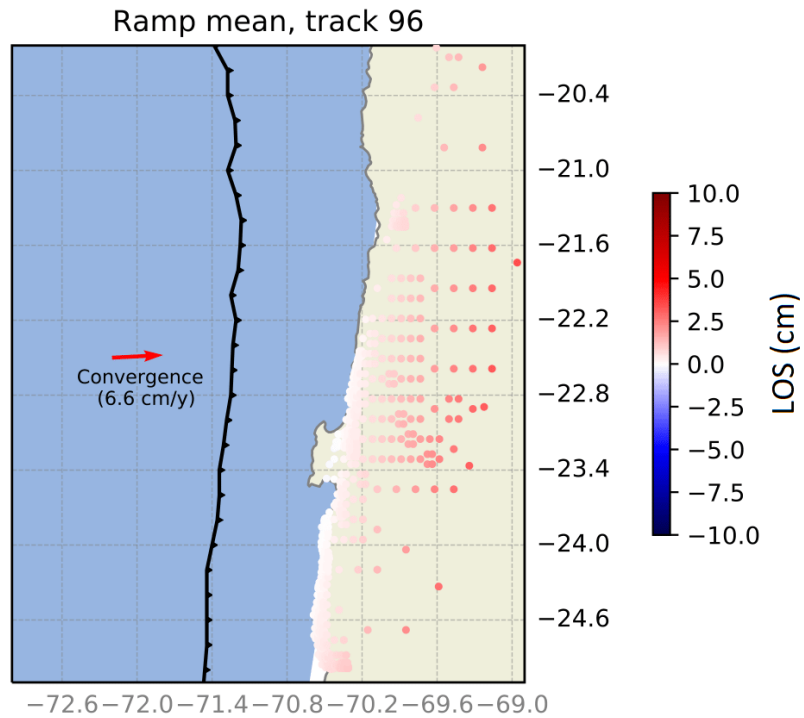
Figure 4.25: Predicted InSAR LOS displacements due to the mean linear ramp estimated from the ensemble of sampled models.
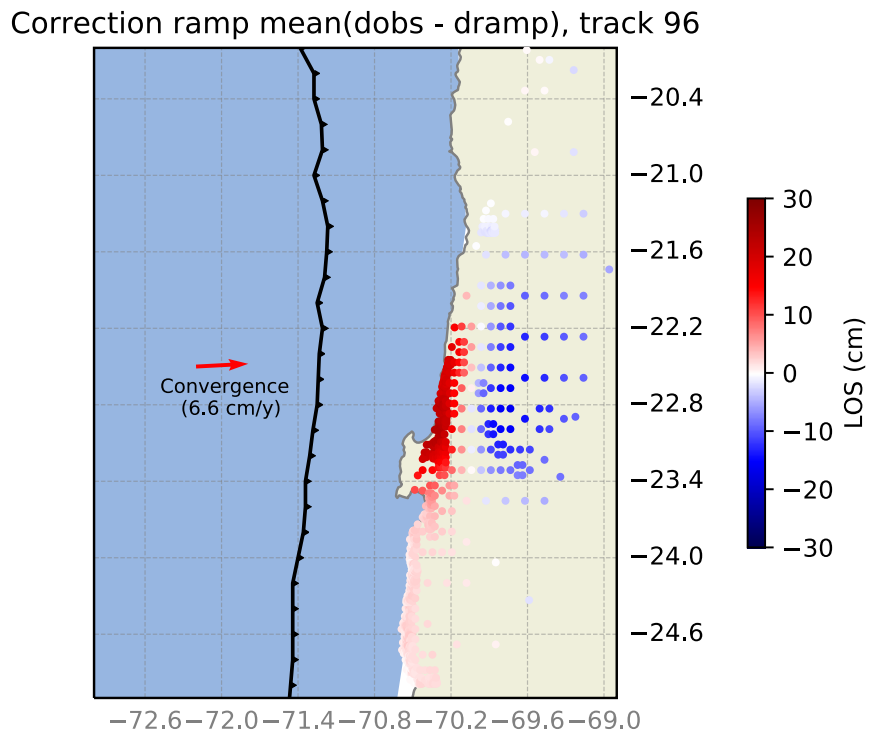
Figure 4.26: InSAR LOS displacements resulting from subtracting the mean linear ramp obtained by the inversion from the InSAR data (track 96).
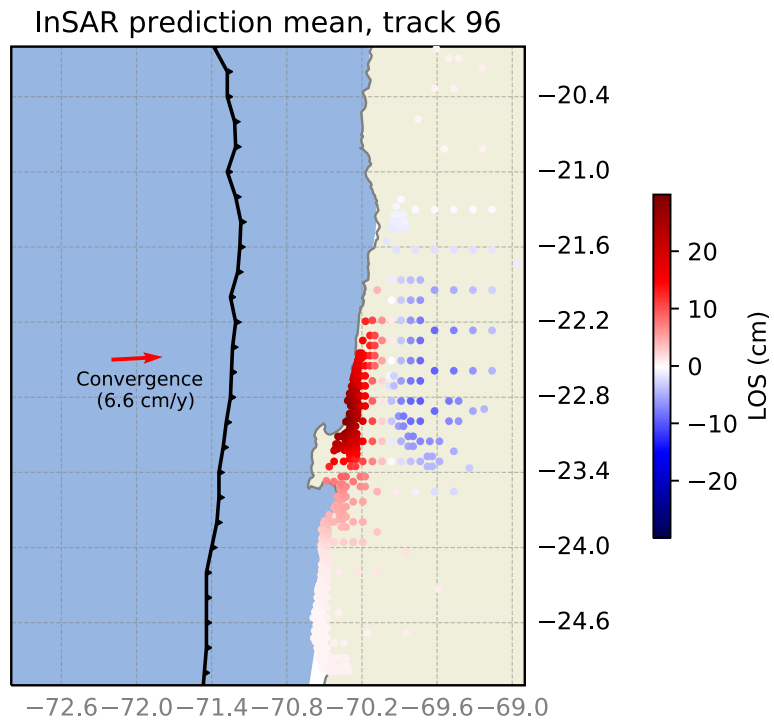
Figure 4.27: Predicted InSAR LOS displacements by the mean slip model

4.8 Discussion

In this section we analyze and discuss the results obtained applying the Bayesian methodology to the inversion of the 2007 M_w 7.7 Tocopilla earthquake in terms of the characterization of the uncertainties of model parameter estimates, fault slip and InSAR ramp correction.

4.8.1 Error and Resolution Analysis

In this section we show the full potential of the Bayesian inversion methodology, the ability to fully characterize and analyze uncertainties on model parameter estimates without making any assumptions on the functional form of their probability distribution pdf's. Here we analyze the distribution of errors of estimated slip and InSAR ramp parameters in terms of their marginal errors as well as to understand the dependence, correlations and trade-offs between estimated model parameters.

On one hand, estimated slip models presented in the previous sections show broader rake parallel slip pdf's in the edges of the fault that are closer to the trench, zones that show mean or median models with slip in those regions that can be interpreted as artifacts (see Figures 4.19a and 4.22a) due to the size of the discretization of the fault geometry in such regions and the poor constraint that onland data has on slip in those fault edges near the trench.

On the other hand, the zones with higher slip located apart from the edges, near the center of the fault surface, have higher values of co-seismic slip, associated to the main asperities ruptured by the Tocopilla earthquake (see Figures 4.19a and 4.22a). Also, such region exhibits higher uncertainties, possibly due to the use of epistemic or forward model prediction errors, that are larger for larger values of fault slip (see Figure 4.28). This region with higher uncertainty is spread out compared with the solutions shown previously, especially for the northern asperity. The maximum uncertainty is about 0.6 m for rake parallel slip and 0.9 m for rake perpendicular slip, being the latter similar to the prior uncertainty set for such slip direction across the fault (1.0 m).

The marginal uncertainties allow to characterize confidence intervals for slip and ramp parameters in terms of the mean/median values and the standard deviation of model parameters. However, in order to make a complete characterization of errors we need also to understand dependency and/or correlations between model parameter estimates. For instance, covariances associated to some particular fault slip parameters can be seen in figures 4.29, 4.30 and 4.31, where we show the behavior for the parallel to rake component of slip as well as the perpendicular to rake component. In the latter figures, we can see that the perpendicular to rake solutions are not correlated with the neighbors, while in the case of the parallel to rake component the amount of dependence/correlation depends on the location of the fault slip patch. The errors in these patches exhibits an anticorrelation with the neighbors, as can be seen from the direction of the arrows in the above mentioned figures. Patches close to the zone with larger slip show more anticorrelation with the immediate neighbors. That means that we have a trade-off between slip on adjacent fault patches. For instance, for a given representative slip model (mean or median) perturbations on such model due to observational or epistemic errors that produce higher slip on a given fault patch, tend to produce also perturbation that leads to smaller slip at the adjacent fault slip patch.

Figure 4.32 shows correlation between ramp parameters as 2D histograms. Ramp parameters describe the linear plane $ax+by+c$ incorporated in the modeling to take into account the additional offset in observed LOS because of the processing of InSAR data. Parameters a and b are the longitudinal and latitudinal slope describing the plane, respectively, and c is a constant value describing a uniform plane. Diagonal elements in Figure 4.32 show the 2D histogram of every ramp parameter (represented in a 2D histogram as a 45° colored line), while out of diagonal elements show 2D histograms between different ramp parameters. The figure shows a low anticorrelation between a and b, a larger anticorrelation between a and c, and low positive correlation between b and c. To better understand some of these results it is necessary to consider the distribution of measurement points, which is not balanced in the longitudinal and latitudinal directions, being more spread in latitude. This distribution shows a constrain to a greater degree of the linear ramp slope in latitudinal direction b compared to the longitudinal slope a (as can be seen in Figure 4.22.c). Most of the InSAR observations are accumulated near the coast line, especially near the Mejillones Peninsula, then, this region contributes to a greater degree to fit the ramp parameters. The larger anticorrelation between a and c indicates a tradeoff between the longitudinal slope and the constant plane to explain the observations. Each longitudinal slope a value is associated to a plane which keeps constant values in latitude, and locally behaves like a constant plane which compites with the contant plane associated to c. The correlation between parameters can be visualized as 2D histograms resulting from applying the Bayesian methodology as can be seen in Figure 4.32
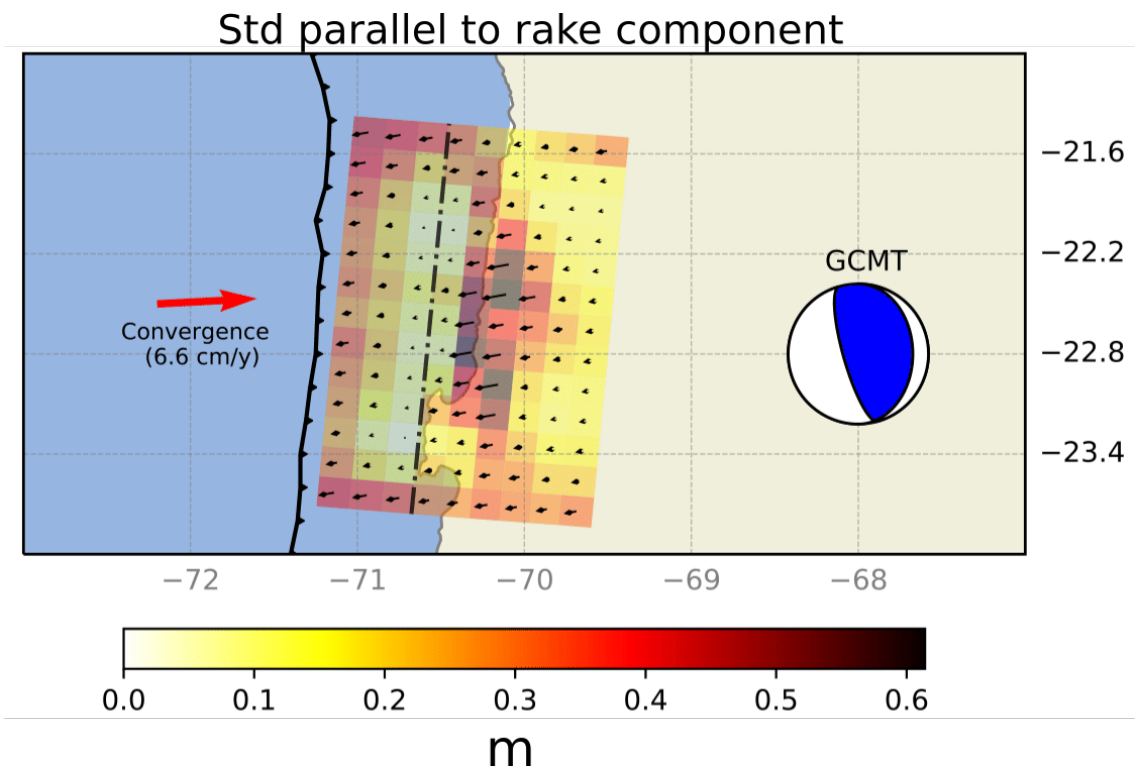
(a) parallel
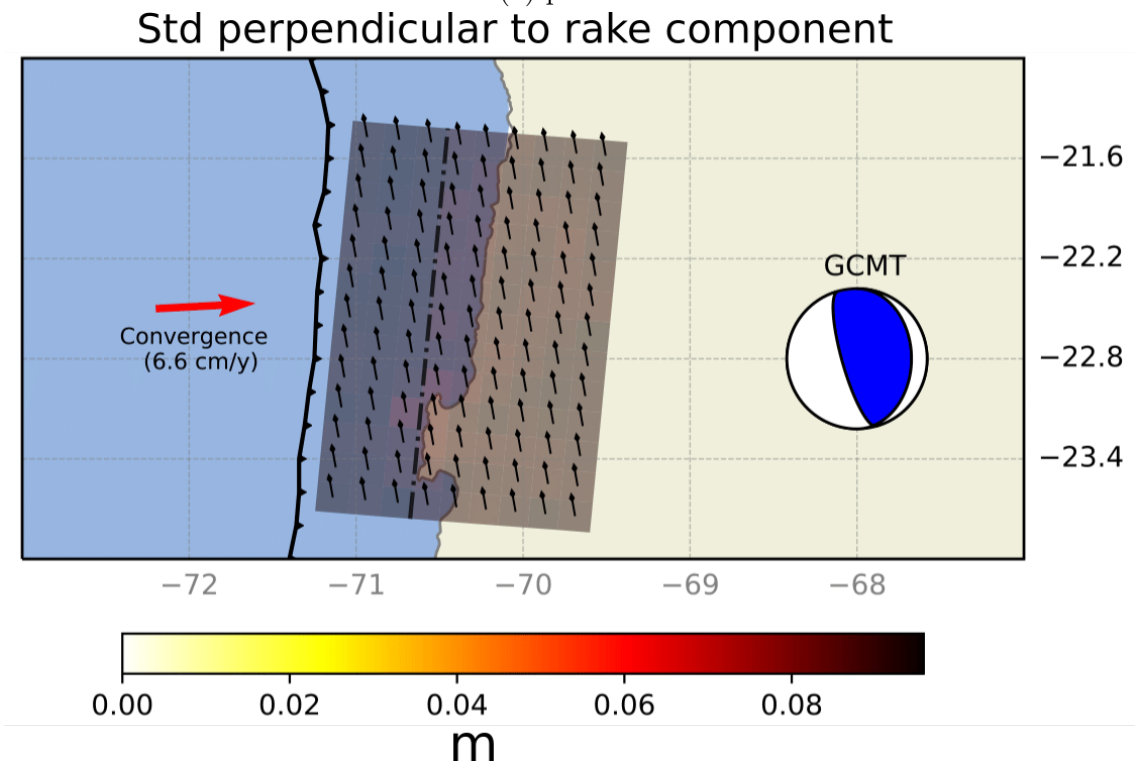
(b) perpendicular

Figure 4.28: Standard deviation associated to the slip estimated parameters over the fault geometry for the case constrained with GPS and InSAR data using a Truncated Gaussian slip prior. a) Parallel to rake component of slip b) Perpendicular to rake component of slip.
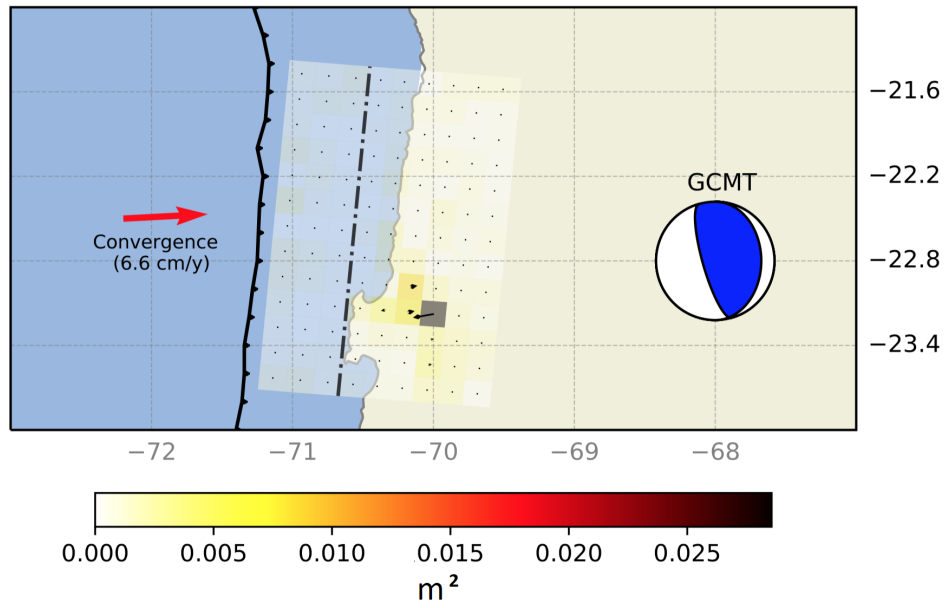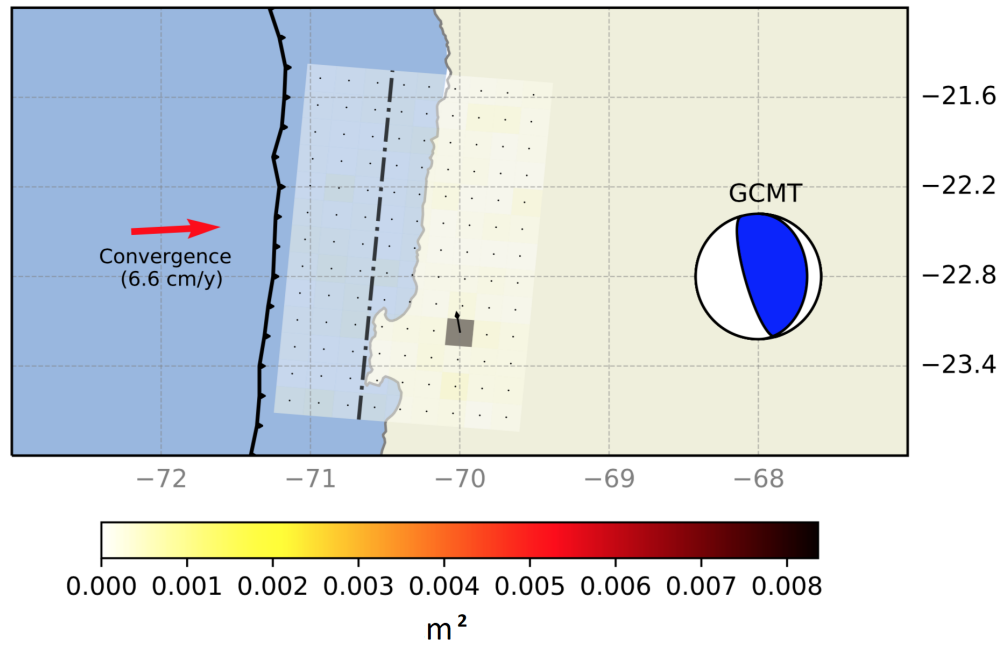
(a) parallel

(b) perpendicular

Figure 4.29: Covariance for subfault 33 considering both, the a) parallel to rake and b) perpendicular to rake component.
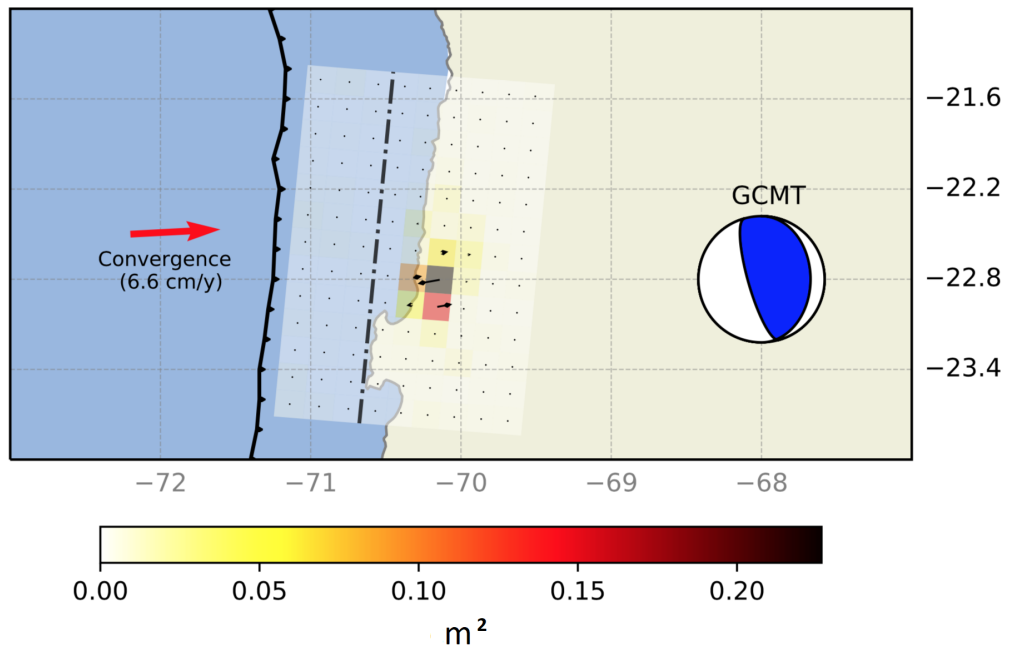
(a) parallel
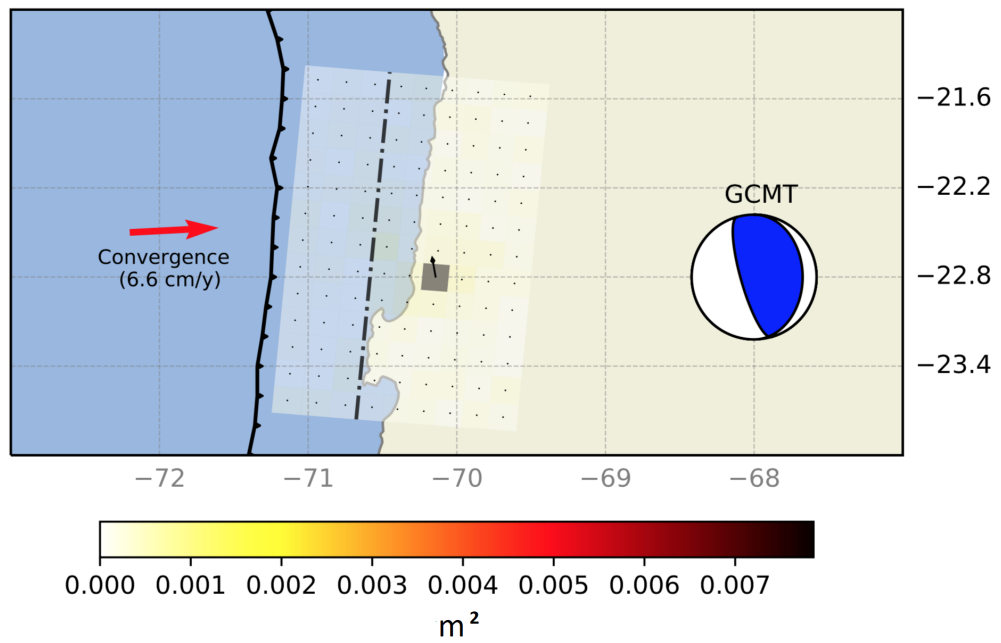
(b) perpendicular

Figure 4.30: Covariance for subfault 50 considering both, a) the parallel to rake and b) perpendicular to rake component of slip.
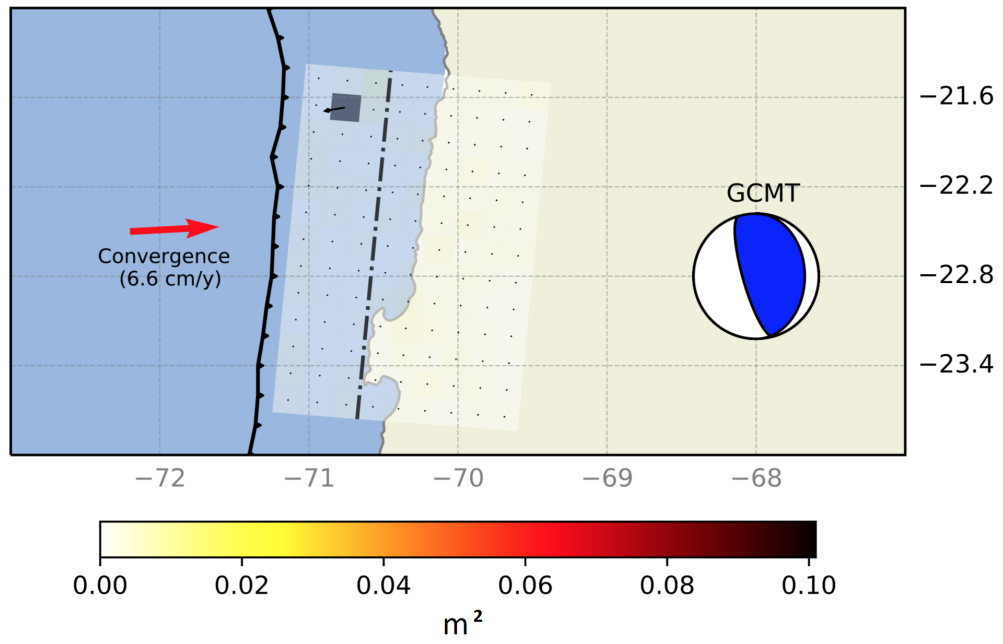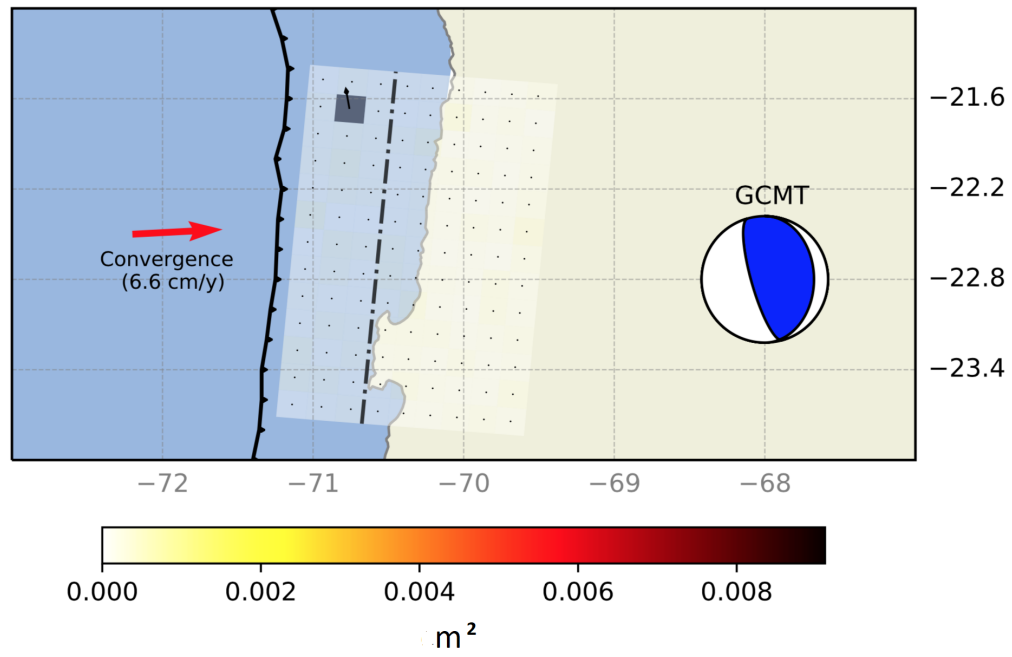
(a) parallel

(b) perpendicular

Figure 4.31: Covariance for subfault 100 considering both, a) the parallel to rake and b) perpendicular to rake component of slip.
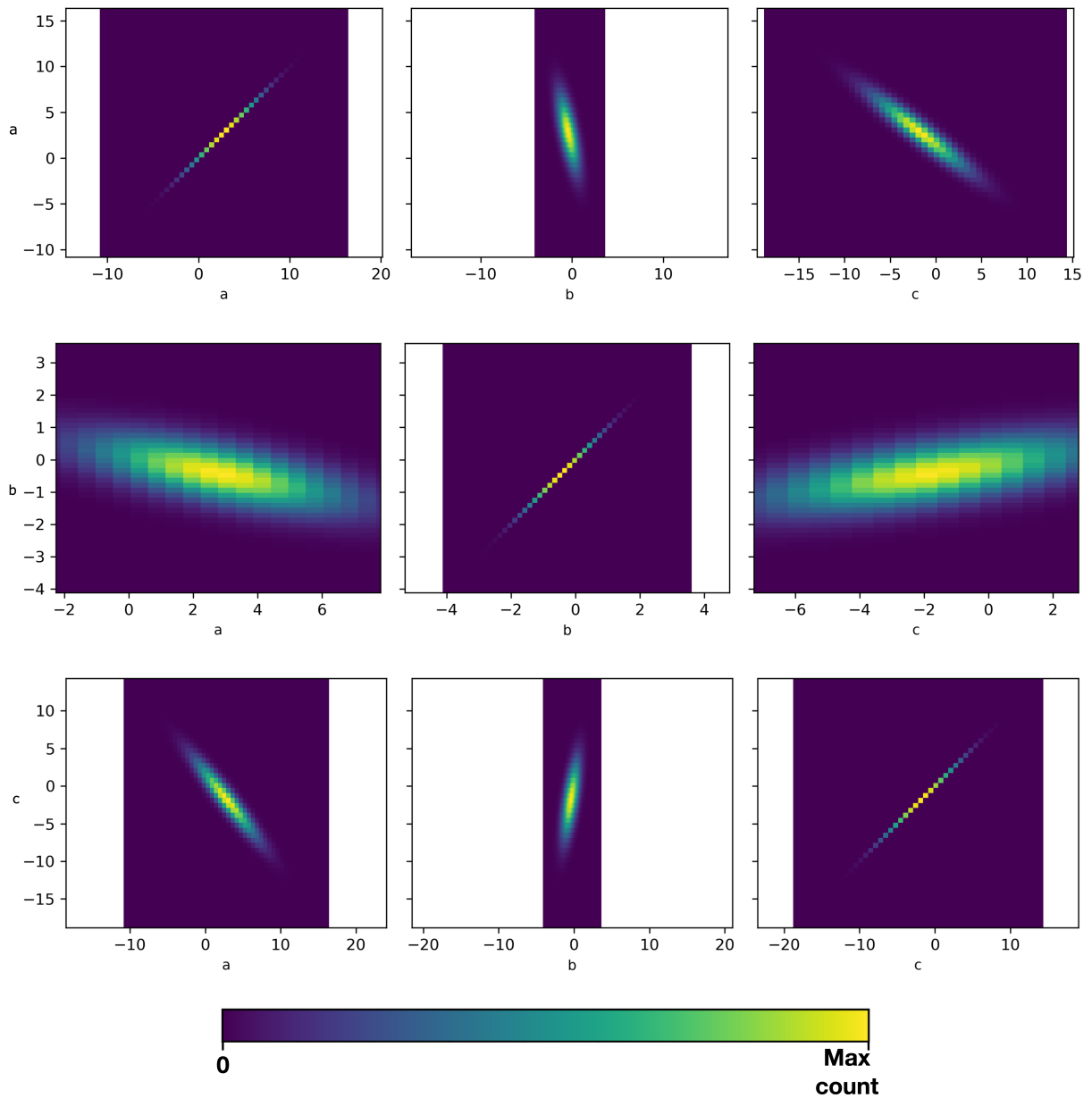
Figure 4.32: Correlation between ramp parameters forming a linear ramp ax + by + c shown as 2D marginal histograms. Color scale indicates counts of model parameters on each bin of the 2D histogram normalized to the maximum count on each plot. Aspect ratio for plots on each row of the figure is 1:1 in order to compare each model parameter with all others in the row.

Figures 4.33 and 4.34 show the anticorrelation of slip when plotting 2D marginal histogram for slip at two adjacent fault patches located at the top left corner of the fault (near trench) as well as at the center of the fault (near coast), where the maximum co-seismic slip is estimated, for the inversion cases in which Uniform and Truncated Gaussian distributions are set for run priors, respectively. 2D marginal distributions for pairs of neighboring fault patches near trench and near the center of the fault, are shown in Figures 4.33b, 4.34b and 4.33c, 4.34c, respectively. In all cases we observe anticorrelation between slip at adjacent subfaults, where

the amplitude of the uncertainties is larger for slip at subfaults closer to the trench. Also, we observed that when using the Truncated Gaussian distribution for the prior distribution, uncertainties are much more reduced at the fault elements near the trench, implying that slip at the near trench fault regions is much less constrained by the data, as the change of prior information, from Uniform to Truncated Gaussian distribution, influenced greatly the outcome of the inversion. This phenomenon is barely noticeable for fault elements near the coast, implying that the inverse problem solution is dominated by the information provided by the data in those regions.

An aspect that can be noticed from the mean slip model, is that it has a fault region, between the trench and depths where maximum slip occurs, exhibiting a very small amount of normal slip (See Figure 4.22). This is not surprising, as we allow little backslip in our prior distribution. However, the estimated normal slip may be related to the appearance of small reverse slip at the near trench subfaults. As both of them are further away from onland GPS and InSAR observations, each of the described slip features produce surface displacement fields with long wavelengths that might be canceling each other, i.e., there is a trade off between slip at the aforementioned neighboring regions, allowing to have very low normal and reverse slip.

The marginal histograms in the edges show a behavior similar to a Beta function (which is the marginal of a Dirichlet distribution). The Dirichlet distribution describes problems where the sum of variables lie in a certain range, so the zones where Beta distributions appear can be associated to a tradeoff between parameters in neighboring fault patches. As described in *Duputel et al.* (2014), a possible interpretation for this behavior is that the resolution length (the size of slip features that we can solve with the available data) is larger than the size of the subfaults in these regions (20 km in this case). *Duputel et al.* (2014) suggests that the resolution length can be found by fitting the estimated marginal distribution to the pdf defining a Beta distribution, allowing to find the number of adjacent subfaults so that slip at such set of faults follows a Dirichlet distribution in which the sum of slip at the set of subfaults is well constrained. In this case, we have a 2D geometry where we identify this behavior outside the zone with larger slip associated to the coseismic slip, near the edges of the parameterized fault surface. The results could be re-processed so that the histograms in these zones represent the mobile average in a set of patches, where such mobile average is calculated over the ensemble of sampled models. Figure 4.35 shows the marginal histograms resulting from this operation. We can see from the figure that those averaged patches present a behavior more similar to a Gaussian, with a mean value closer to zero, thus defining a slip model that is better constrained across the fault. This process could be considered as a spatially variable posterior smoothing of slip across the fault by calculating local spatial averages of sampled models, in which the amount of smoothing is based on the amount of the constraint on slip provided by the observations. Here we must note that smoothing is applied only at the fault portions where is needed, in opposition to regularized least squares inversions, where smoothing is needed across all the fault in order to be able to obtain a single (smoothed) slip estimate.

In both cases, whether using $\mathbf{C_p}$ or not, the fitting of the GPS data is far from the expected considering their errors represented by 1-σ ellipses in Figure 4.2. This could be related to the high slip values observed in the corners of the fault slip, and it is discussed later when we incorporate InSAR data in our analysis.
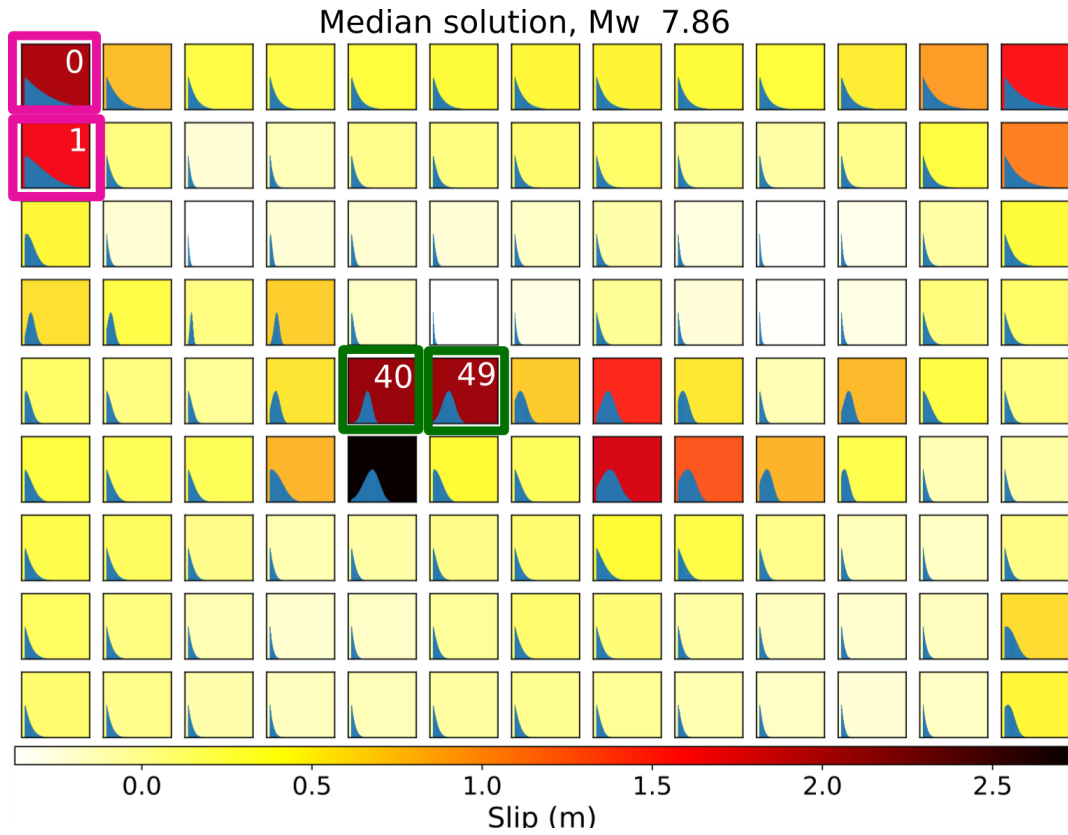
(a) Marginal Histograms for parallel to rake direction
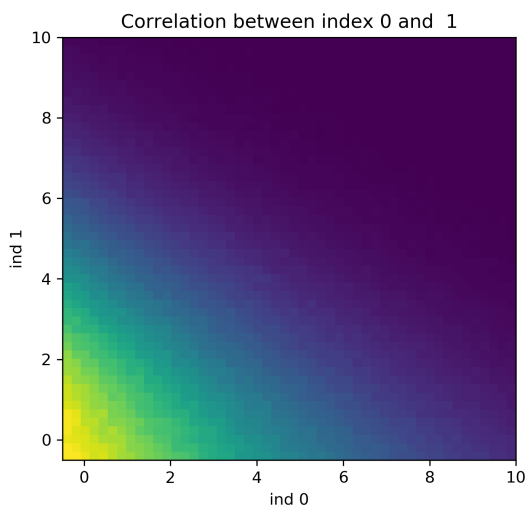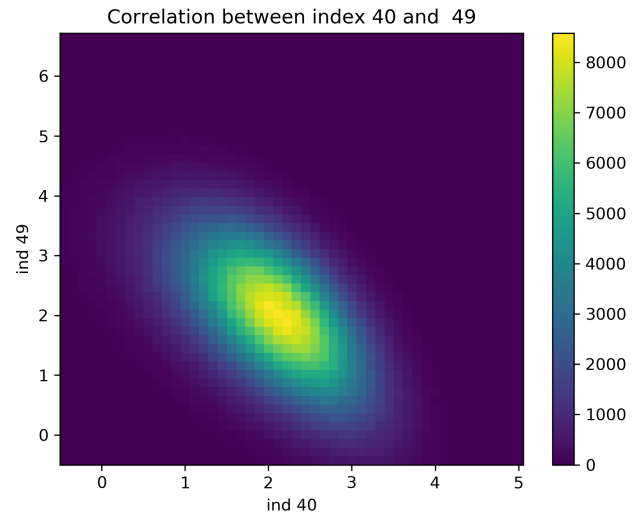
(b) 2D histogram

(c) 2D histogram

Figure 4.33: Marginal histograms for rake parallel slip when using Uniform priors. (a) marginal rake parallel slip histogram for each subfault. (b) and (c) show 2D histograms between slip at pairs of adjacent subfaults near the trench (b) and near the region of larger co-seismic slip (c) indicated by purple and green squares in (a), respectively.
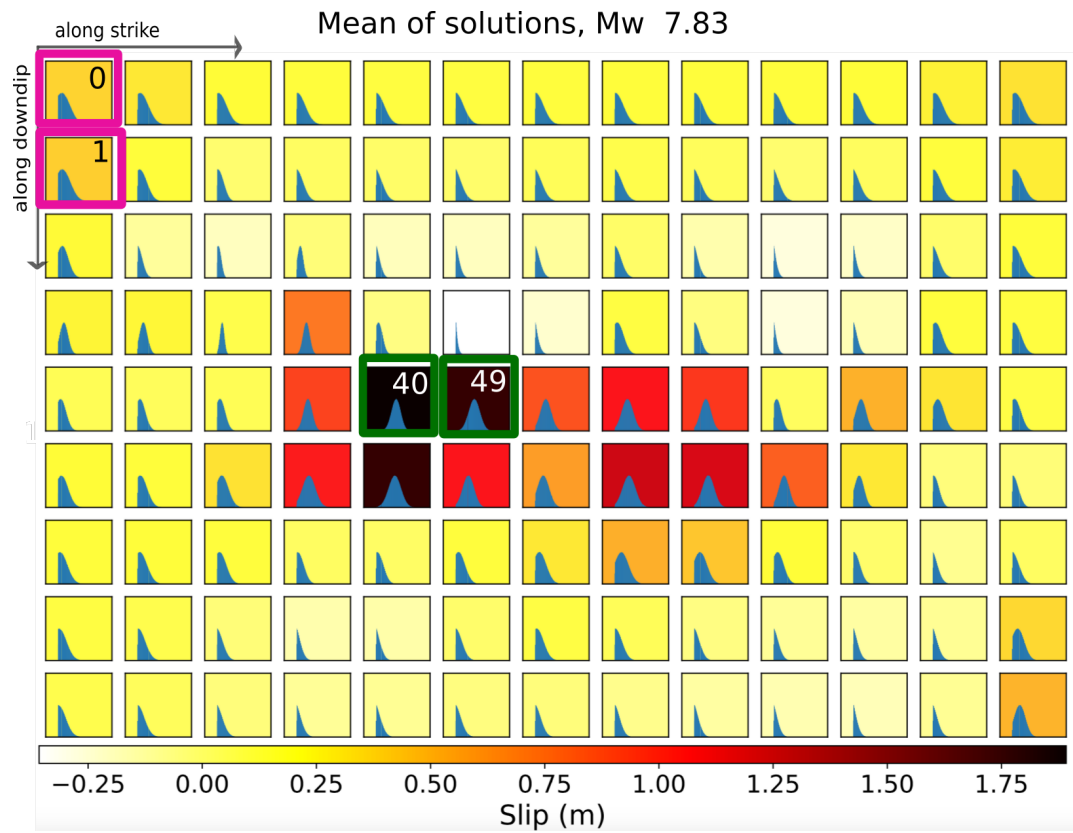
(a) Marginal Histograms for parallel to rake direction
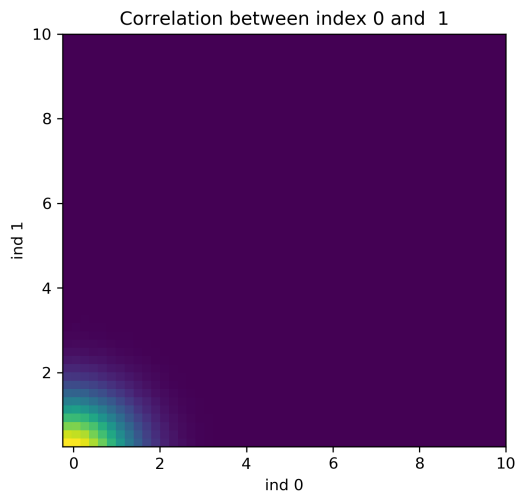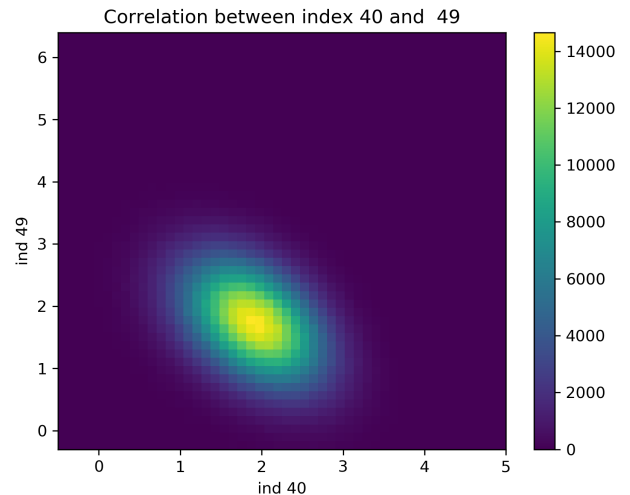
(b) 2D histogram

(c) 2D histogram

Figure 4.34: Marginal histograms for rake parallel slip when using Truncated Gaussian priors. (a) marginal rake parallel slip histogram for each subfault. (b) and (c) show 2D histograms between slip at pairs of adjacent subfaults near the trench (b) and near the region of larger co-seismic slip (c) indicated by purple and green squares in (a), respectively.
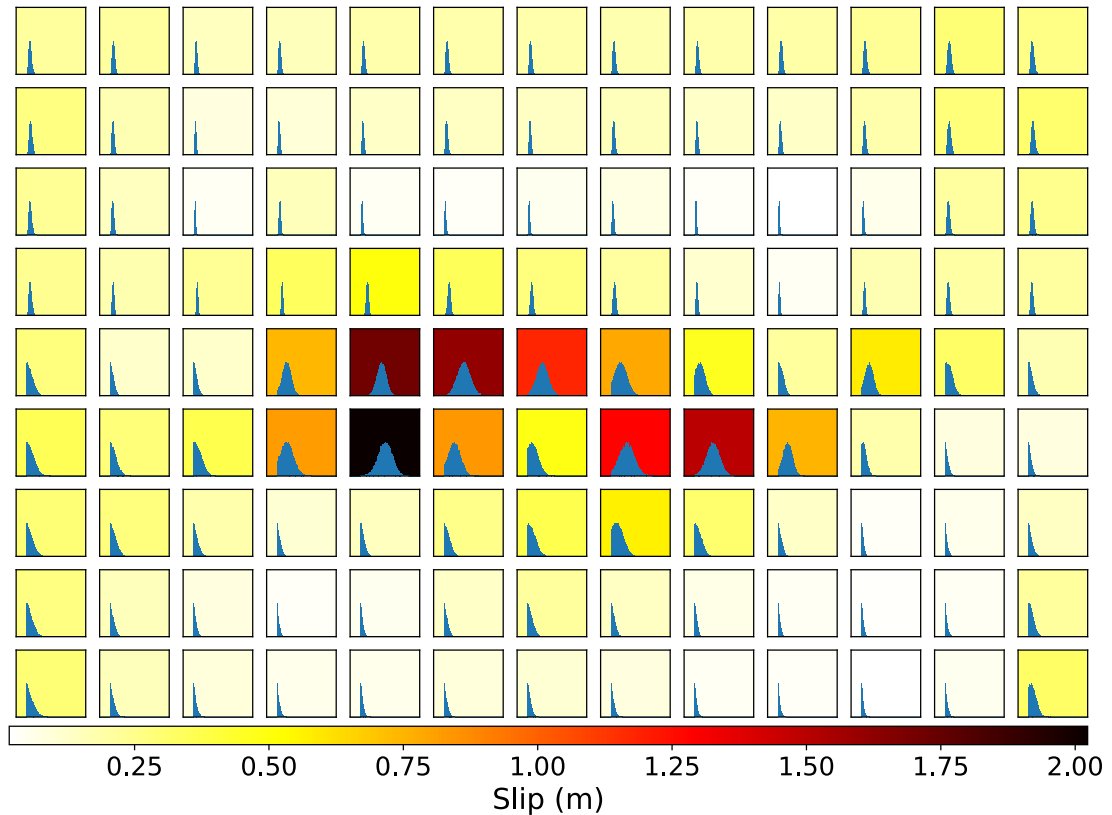
(a) Marginal Histograms parallel to rake component
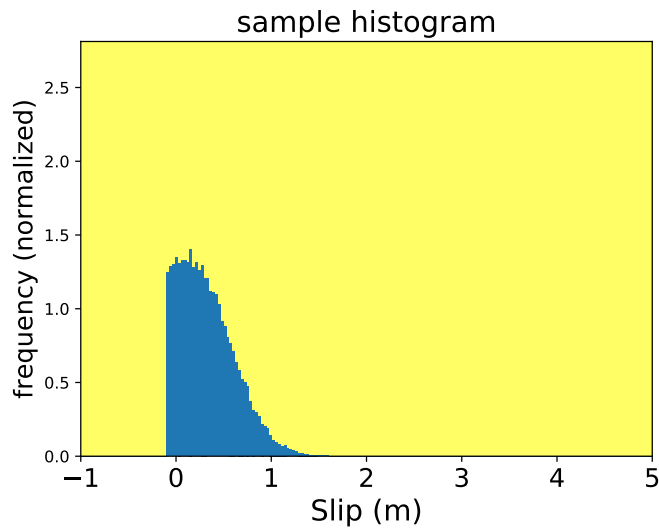
(b) Histogram of lower right corner

Figure 4.35: a) Marginal histograms obtained averaging patches in the trench border. The first four cells along dip show histograms averaging the neighbor patches that form a cross centered in the original patch. b) rake parallel slip histogram for the lower right corner element (all the subfaults have the same slip intervals along x axis as in b).

4.8.2 Limitations of the Bayesian Inversion and the Option of a Sensitivity Analysis

An important limitation of the Bayesian methodology is the requirement of costly computational resources, since billions of forward model evaluations are needed to obtain the posterior ensemble of models resolving the inverse problem. The number of evaluations for the inversions performed in this thesis was $\sim 10^{11} - 10^{12}$, with execution times between 15 and 70 hours. The slip estimates presented in this work were obtained using a computer node with 2 Tesla K40 GPGPU cards, that cost about 12.000.000 Chilean pesos.

The determination of the number of samples and the length of chains takes into account the number of data sets included, since the incorporation of a new data set makes necessary to reduce the number of samples due to the hardware limitations. For example, using only GPS data, the number of samples employed was 3.000.000, while this number decreased to 2.000.000 when 1 InSAR track was incorporated.

Observing examples where Bayesian modeling have been used, we can notice that these methods are expensive in terms of hardware and execution time. For example, *Duputel et al.* (2015) modeled the 2014 Pisagua earthquake using static GPS, InSAR, tsunami, tide gauges, high-rate GPS, and strong motion data, employing $\sim 10^{10}$ models to obtain the ensemble of parameter model solutions using 90 GPGPU and 24 hrs of execution time.

Given those limitations, an stability analysis can be restrictive in terms of hardware and execution time. Therefore, we will discuss the obtained solutions in terms of a Sensitivity analysis (*Ortega*, 2013) that can be done in a single desktop workstation. Also, this analysis provides a tool to evaluate the importance of each data set in the prediction of the model.

As it was discussed in previous sections, estimated slip at fault elements near the trench present greater variability compared with elements closer to the coastline. Also, at the near trench fault regions, estimated slip models present artifacts that can be related to the sensitivity of slip at the border elements of the fault.

From *Ortega* (2013), the sensitivity of fault slip is defined as the diagonal matrix

$$\mathbf{S} = \sqrt{diag(\mathbf{G^T C_\chi^{-1} G})} \tag{4.25}$$

where each diagonal element of $\mathbf{S}$ quantifies the sensitivity of slip at each discretized element of the fault. In order to easily interpret sensitivity, we will assume a simplification in which all observational and model prediction errors are independent and identically distributed (iid), i.e., the misfit covariance matrix is the diagonal matrix $\mathbf{C}_\chi = \sigma_o^2 \mathbf{I}$. With this simplification, we can write the sensitivity of rake parallel slip at the i-th fault discretization as

$$s_i = \frac{1}{\sigma_o} \sqrt{\mathbf{G}_{:i}^{\parallel \mathbf{T}} \mathbf{G}_{:i}^{\parallel}} = \frac{1}{\sigma_o} \left\| \mathbf{G}_{:i}^{\parallel} \right\|_2 \tag{4.26}$$

where $\mathbf{G}_{:i}^{\parallel}$ is the vector formed by the i-th column of the Green's function matrix for rake parallel slip $\mathbf{G}^{\parallel}$. Figures 4.36 and 4.37 show the slip sensitivity calculated at each element of the fault surface for rake parallel and rake perpendicular slip directions, respectively. Note

that slip sensitivity varies about one order of magnitude between trench and coastline regions of the fault, being lower in the edges of the fault surface closer to the trench, and higher for those closer to the zones where we obtained the maximum slip (near the coastline). The lower sensitivity at the fault elements closer to the trench can be explained as a consequence of the lack of offshore observations, as all of them are located on land.

On one hand, slip sensitivity s_i can be interpreted from a model prediction viewpoint. As s_i is the norm of the Green's functions for slip on a given slip direction and subfault, it is a measure of how much 1 meter of slip in a single subfault "moves" (causes displacements) the whole GPS network and/or InSAR pixels location. So, as the sensitivity for fault elements near the trench is about 10 times smaller than at the elements near the coast, we need 10 meters of slip at a near trench subfault to produce a similar amount of displacement caused by 1 meter of slip at a near coast subfault. Thus, slip perturbation located near the trench, have a much lower impact on the data misfit ($\left\| \mathbf{G\theta} - \mathbf{d^{obs}} \right\|_2^2$), than slip near the coast, implying that near trench estimated slip will have larger uncertainties. Such a behavior can be observed in Figure 4.33, where the spread of the marginal distribution of rake parallel slip is much larger at the fault elements near the trench. Note that we do not observe this behavior when we use a Truncated Gaussian distribution for the run priors (see Figure 4.34), as such prior tries to set slip as small as possible, providing a constraint on slip amplitude on each subfault.

On the other hand, we can see sensitivity from a perspective based on the amount of the constraint on fault slip provided by the data. Here, in order to explain small variations in data, due to observational or model prediction errors, the amount of slip required to explain such data variations is 10 times larger if using fault elements near the trench, compared to using near coast fault elements. Thus, data errors, might produce larger slip variations at fault portions near the trench, than near the coast, such as the artifacts observed in some of our slip model estimates.

It is important to note that the sensitivity is not the same for the two data sets. An explanation is that they do not incorporate the same components of deformation, therefore they do not constrain fault slip in the same manner. GPS data measures 3 displacement components E, N and Vertical, while InSAR measures only those in the direction of LOS.
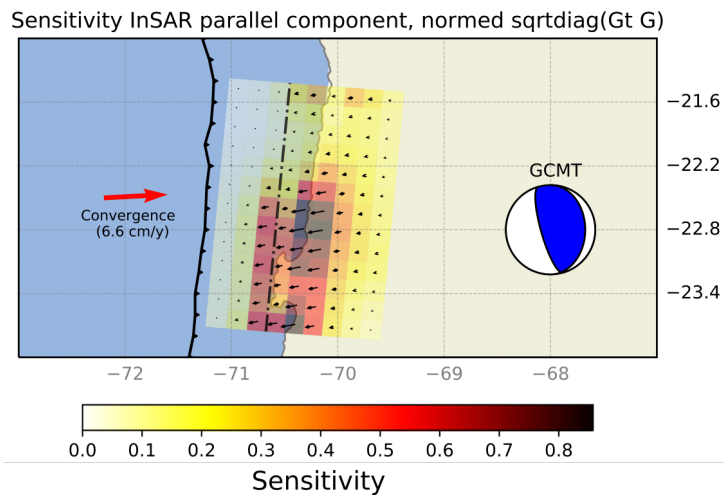
(a) All data
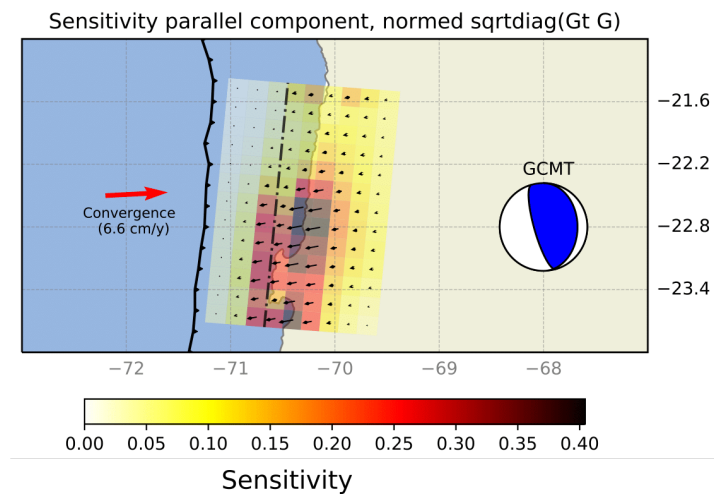
(b) GPS
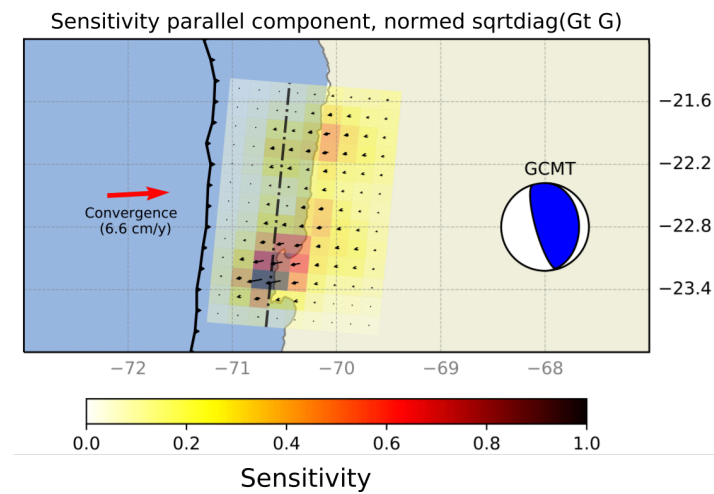
(c) InSAR track 96

Figure 4.36: Sensitivity of parallel to rake component of slip when considering a) all datasets b) GPS and c) InSAR only.
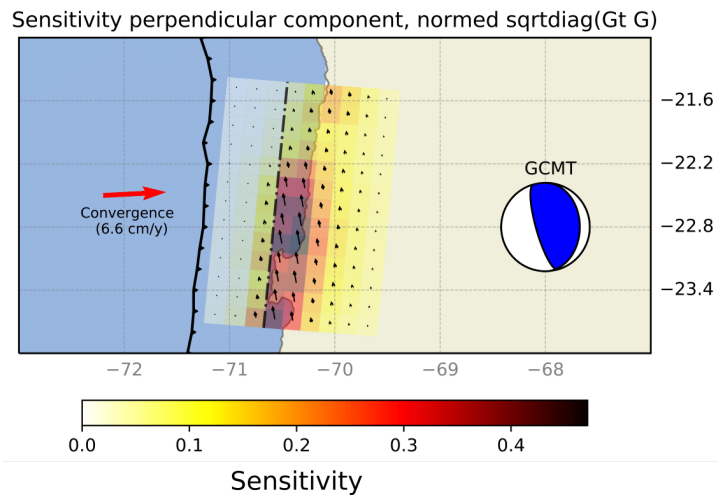
(a) All data
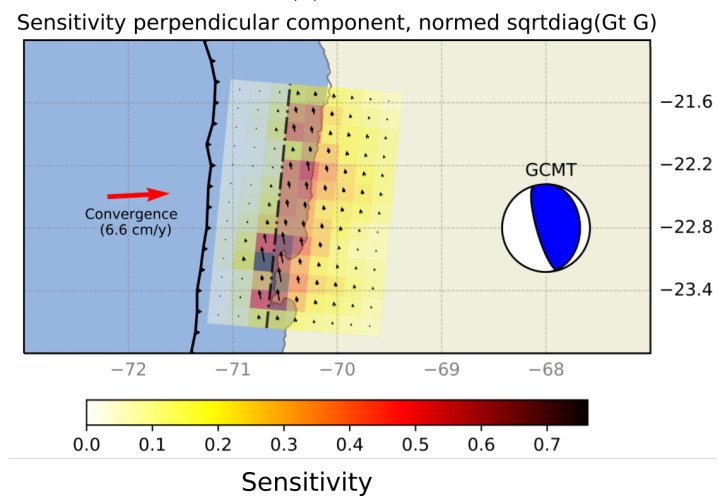
(b) GPS
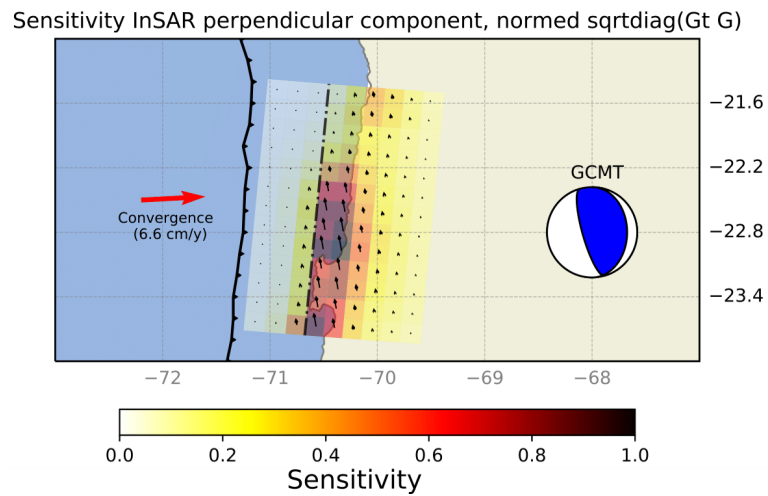
(c) InSAR track 96

Figure 4.37: Sensitivity of the perpendicular to rake component of slip when considering a) all datasets b) GPS and c) InSAR only.

4.8.3   Comparison with other studies

The median solution obtained using GPS and InSAR data (using one descending track and incorporating the epistemic error) shows a zone with high slip in a smaller region offshore respect to previously obtained solutions (fig 4.38.a, 4.38.b). The zone with the highest slip spans an area beyond the coastal edge close to the southern asperity. Nevertheless, to analyze this result it is necessary to have in mind the role of the error in the result. If we consider this error, the solution covers a range of likely solutions, probably containing a solution similar to the result obtained by *Béjar-Pizarro et al.* (2010) using Regularized Least Squares, since the slip values in this solution reside in the interval defined by the standard deviation (Figure 4.38.d). This is probably the case too for the solution obtained by *Minson* (2010), taking into account the finer discretization. On the other hand, to fully characterize the error it is necessary to consider the correlation between patches. This is shown in figures 4.29, 4.30 and 4.31, where we can notice the almost null correlation between patches in the perpendicular to rake component and the anticorrelation between adjacent patches for the parallel to rake component. The latter means that larger slip in a certain patch makes more probable a smaller slip in the neighbors, making less likely solutions where the entire region has larger slip than the mean.

Slip estimates obtained with optimization methodologies (e.g., Least Squares) usually use smoothing as prior information. Thus biasing slip estimates towards smooth models, so it is expected that the zone with higher slip covers a broader region on the fault geometry, compared with the "true" region that slipped during the earthquake. Then, it makes sense that the representative solution obtained using this Bayesian methodology presents a more compact slip area, due to lack of smoothing during the inversion. It should be noted that the results of the Bayesian inversion do not exclude the smoothed solutions obtained using Regularized Least Squares, in fact they are present in the ensemble of models characterizing the solution, but the likelihood associated to those models is lower than the one obtained for the representative models. The completeness of the solution is an advantage of the Bayesian methodology over more typically used methods, not requiring for it the incorporation of additional biases, more than those physically justifiable by the setting of the inverse problem.

The result obtained is in agreement with those obtained by *Duputel et al.* (2015) studying the 2014 Pisagua earthquake, which is a region more compact compared with other studies of the same earthquake.

A first difference that can be noticed from the modeling made by *Minson* (2010) using Bayesian methods is in the fault geometry employed. We set a fault geometry following slab feautures from *Contreras-Reyes et al.* (2012) that covers the region reaching the trench, then, we can show that the rupture did not reach the shallower part of the subduction contact.

Additionally, we can see that the portion of higher slip lies downdip from the limit indicating the change in dip in the subducting plate, which reinforce the idea of this limit acting as a seismic barrier for the updip propagation of the rupture, as we can see in Figure 4.38.c

The updip limit of the rupture area is close to 30km depth, deeper that the kink in the slab identified at 20 km depth. The updip limit seems to be well constrained almost in the entire rupture zone, as we can see from the uncertainties close to this limit. By contrast, the downdip limit seems to be not so constrained, especially in the region close to the northern

asperity. The downdip limit is between 45 km, for the southern asperity and 50 km for the northern.

This result can interpreted as a tendency of the rupture to propagate close to this singular feature in the fault geometry in this region, or a coupling context that favored the propagation close to this singularity. Both cases has different implications for tsunami hazard, and should be studied in more detail.

Solutions shown by other authors suggest a trenchward propagation of the rupture in the zone close to the Mejillones peninsula, reaching very shallow depths, which is likely produced by smoothing constraints used for least squares inversions. On the contrary, as we do not use prior smoothing constraints, estimated slip in this study presents an abrupt decay in the region where there is an expected change in dip interpreted as a barrier to updip seismic rupture propagation.
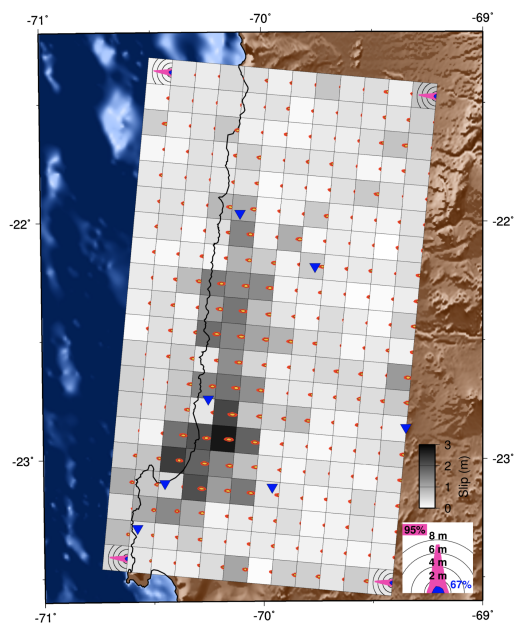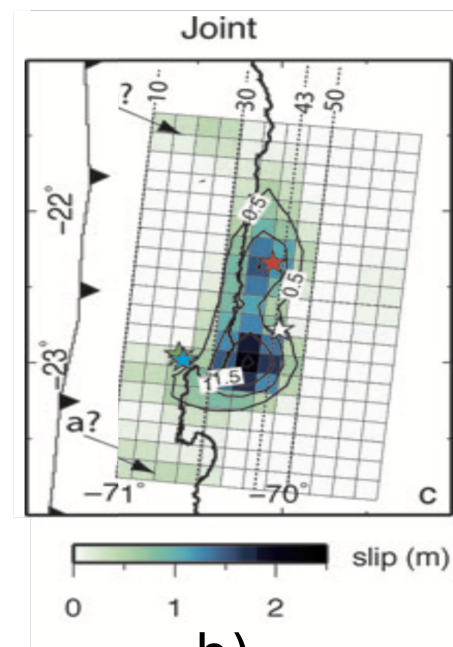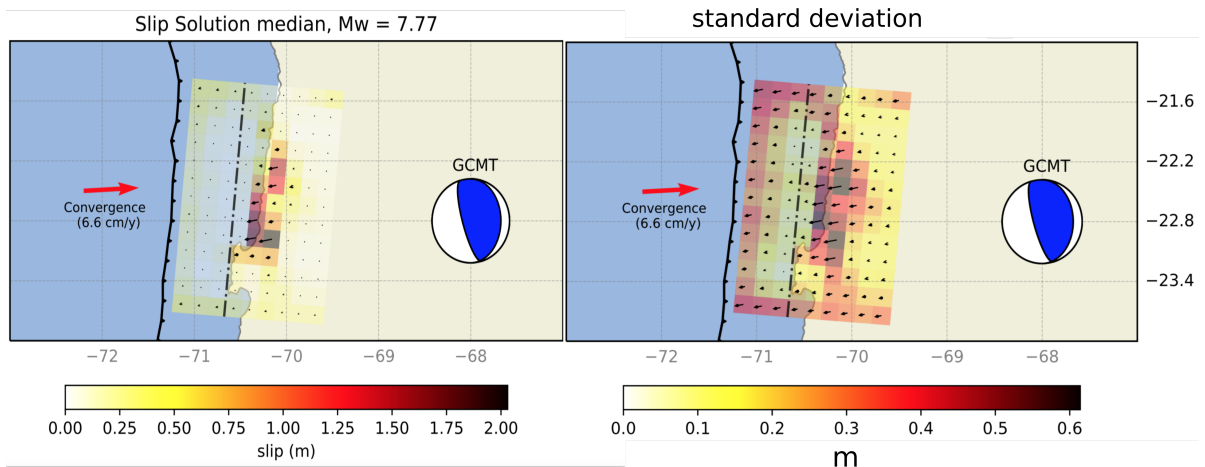
Figure 4.38: a) and b) show the solutions for slip obtained by a)*Minson* (2010) and b)*Béjar-Pizarro et al.* (2010). c) shows the median d) show the median solution obtained in this thesis using Bayesian inversion and d) presents the map indicating the standard deviation.

Chapter 5

Conclusions

The Bayesian methodology employed to solve the coseismic static slip generated by the 2007 M_w 7.7 Tocopilla earthquake allows to find an ensemble of plausible solutions, solving the inverse problem in a more complete and robust form without imposing non-physical restrictions, such as artificial regularization terms (e.g., smoothing) that produce a bias on slip estimates. These solutions can be represented as marginal distributions of model solutions and can be visualized pointing out the pdfs obtained, from which aspects like the number of maxima, and the possible tradeoffs between parameters can be seen. This method allows to incorporate additional sources of error which could be difficult to deal with when more traditional methods are used (e.g., least squares). Besides, it allows better characterization of uncertainties in the resulting parameters. Additionally, the solutions can be represented as an statistical estimator, as the mean, median or MAP solution (maximum a posteriori).

Also, this method allows to visualize aspects about the resolution over the fault area from the shape of the obtained distributions. Zones with distributions similar to the pdf of a Beta distribution can be associated to a lower resolution and consequently the size of the fault discretization in those regions should be larger. Because of this behavior, sometimes a posteriori spatial smoothing is recommended in fault regions poorly constrained by the data due to excessive refinement of the discretization (too small elements). The resolution length or size of the region to perform the local spatial averages can be estimated from the shape of the beta distribution, and after the application of the local average the resulting distributions are smoothed. In a word of caution, we note that this procedure differs in great manner from smoothing used to find regularized least squares solutions, in which the latter is chosen arbitrarily and homogeneously across the fault, whereas with the Bayesian posterior local averages, smoothing is applied only at some fault portions and in an amount justified by the statistics of the ensemble of sampled models.

The Bayesian methodology is costly in computational terms, making difficult to use this technique to solve efficiently an inverse problem with limited computational resources, especially when the number of parameters is big. The number of models to be evaluated by these methods usually is high, as in this case, where we evaluated $\sim 10^{11}$ models, taking the order of 1-3 days to perform the inversion using GPS and 1 InSAR track.

The main results obtained using the Bayesian methodology applied to the rupture process of the Tocopilla earthquake, and incorporating the epistemic source of error can be

summarized as follows:

- The rupture of the Tocopilla earthquake occurred in the deepest part of the seismogenic contact, with largest slip located at two main asperities. Our results agree with previous studies, but differ in the details of the estimated slip and that we produce a full characterization of errors for slip estimates.

- The representative slip model solutions obtained for the 2007 Tocopilla earthquake using the Bayesian methodology present more compact features compared with previous studies, which can be interpreted as a consequence of the absence of smoothing during the modeling. This allows to identify in a better way, without smoothing, regions on the contact area where the rupture occurred, information valuable to better understand rheology features about the fault contact.

- Slip connecting the two main asperities seems to be offshore respect to the previously published solutions.

- The maximum slip obtained is close to 2.0 ± 0.6 m located in the southern patch. This slip maintains similar values towards the coast, downdip the limit of dip change, along which the slip values decay to the north.

- Moment magnitudes obtained using representative models as the average of all ensemble models, median and the maximum likelihood model are closer to M_w 7.8 and M_w 7.9.

- As inferred by this study and previous authors, the downdip kink in the plate interface geometry (*Contreras-Reyes et al.*, 2012) might act as a barrier to rupture propagation, but this claim needs to be further tested.

- The updip limit obtained is close to 30 km, better constrained than the downdip limit, which is more spread, especially in the northern patch.

- Neglecting the epistemic error associated with layered earth model conducts to solutions where the main zones of maximum slip are more localized in the two patches characterizing the solution. When we incorporate this source of error, the solution is more spread, in the form of a more notorious bridge between the main asperities.

- The model representing the ensemble of solutions for the 2007 Tocopilla earthquake agree with previously hypothesized tectonic features, for example the southern barrier in the Mejillones peninsula and the updip barrier, which can be related with the change in geometry properties. This study suggest that theses changes in rheology could be more abrupt than in the previous studies. In the case of the southern edge of the rupture propagation it is even more noticeable that slip occurred below the geometric /rheologic limit.

  For example a possible hypothesis is that the rupture tends to propagate towards singularities like the kink in the slab, this acting as a seismic barrier. This could have implications in the management of seismic hazard, specially for tsunami hazard. Data from future events needs to be processed to confirm (or discard) this hypothesis.

Appendix A

Inversion Using Regularized Least Squares

A.1 Formulation of the Regularized Least Squares Problem

Previous to run the Bayesian inversion of the quasistatic slip of the 2007 Mw 7.7 Tocopilla earthquake a series of inversions were performed using regularized least squares with a restriction of positivity. The aim of these previous tests was to check the results obtained in published studies using the least squares method as well as to check the calculation of the Green's functions.

The positivity constraint requires the use of coordinate axes to insure that the rake is located in the positive quadrant of this system. The axes have been chosen so that the rake is approximately at a 45° angle between them. The azimuthal angle used is 5° while the angles of the coordinate system are 60° and 150°.

The formulation of the inverse problem in this case is in terms of the minimization of the function:

$$\min_{\boldsymbol{\theta}} ||\mathbf{W_x}(\mathbf{G_e}\boldsymbol{\theta} - \mathbf{d})||_2^2 + \epsilon^2||[\mathbf{W_h}(\mathbf{H}\boldsymbol{\theta} - \mathbf{h_0})]||_2^2 + ||\gamma[\mathbf{W}_\delta\boldsymbol{\theta}]||_2^2$$

$$S.A\ m_i \geq 0\ \forall i \quad \boldsymbol{\theta} = \begin{pmatrix} \mathbf{m_{slip-\|}} \\ \mathbf{m_{slip-\perp}} \\ A_1 \\ B_1 \\ C_1 \end{pmatrix} \tag{A.1}$$

Where $\boldsymbol{\theta}$ is the set of slip parameters incorporating additional parameters in the case when InSAR data are considered, $\mathbf{W_x}$ is a weighting matrix accounting the fitting errors, This matrix comes from the Cholesky decomposition of the fitting covariance matrix C_χ^{-1}, where $\mathbf{C}_\chi = \mathbf{C_d} + \mathbf{C_p}$, with $\mathbf{C_p} = 0.05 \times \mathbf{C_d}$. This weighting matrix is multiplied by the residual $\mathbf{G_e}\boldsymbol{\theta} - \mathbf{d}$. The second term corresponds to the Tickonov regularization (order 2), where $\mathbf{H}$ is the Laplacian operator, $\mathbf{h_0}$ is a zero vector and $\mathbf{W_h}$ a weighting matrix

accounting the sensitivity regularization (*Ortega*, 2013). γ is a coefficient associated to a second regularization, this one associated to the difference between components of slip in two directions. $\mathbf{W}_\delta$ is a matrix applied to $\boldsymbol{\theta}$ to obtain a vector measuring the differences between perpendicular components of slip. The tests carried out consider the previous inverse formulation using different geometries, such as a rectangular fault, slab 1.0 and the geometry obtained by *Contreras-Reyes et al.* (2012), assuming a half space and the 1D velocity model *Husen et al.* (1999). Here we will show the case of the bent geometry used in this thesis *Contreras-Reyes et al.* (2012) and the stratified model of the seismogenic zone of northern Chile from *Husen et al.* (1999), which is an elastic structure consistent in 15 layers. Figure A.1 shows the slip obtained using regularized least squares with a positivity constraint when we use only GPS data. Solutions depends on the chosen regularization factor ϵ, and the slip pattern smooths as this factor increases its value. The γ factor is chosen as 1. Figure A.2 shows the fitting between observed and synthetics.

When we incorporate InSAR data from track 96d to the GPS data, the obtained solutions are similar to those obtained using only GPS data as we see in Figure A.3. The main difference is that the patches of maximum slip are separated by a region of very low slip values for a low regularization factor. Figures A.2 and A.5 show the comparison between observed and synthetic GPS displacements and the residuals for InSAR dataset.
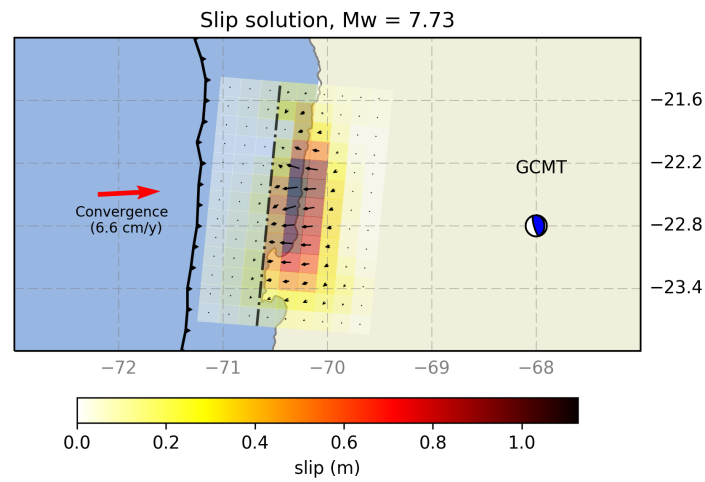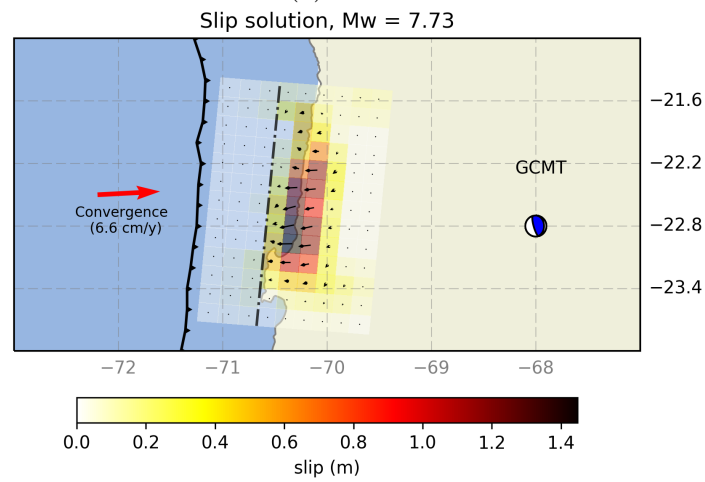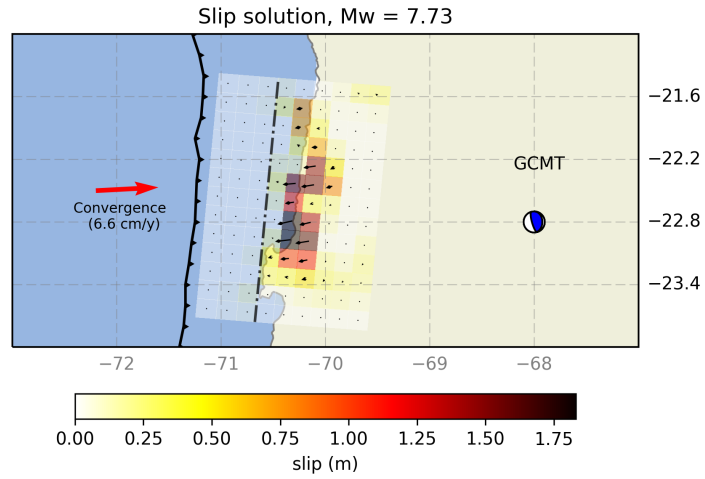
Figure A.1: Slip obtained for the 2007 Tocopilla earthquake using GPS data and a regularized least squares with a positivity constraint method for different regularization factor ϵ.

(a) $\epsilon = 0.03$

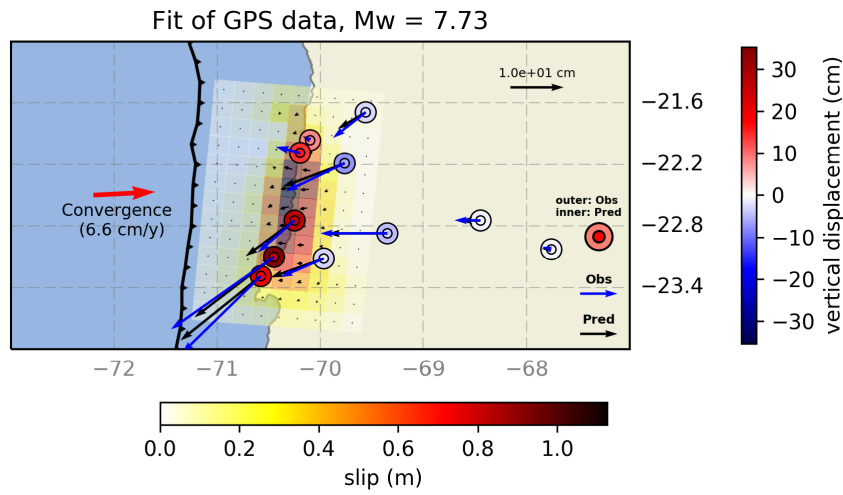(b) $\epsilon = 0.11$

(c) $\epsilon = 0.36$

Figure A.2: Comparison between observed and predicted horizontal and vertical static displacements as a result of the inversion of the 2007 Tocopilla earthquake using GPS data and the regularized least squares method with a positivity constrain for different regularization factor ϵ

(a)

(b)

(c)
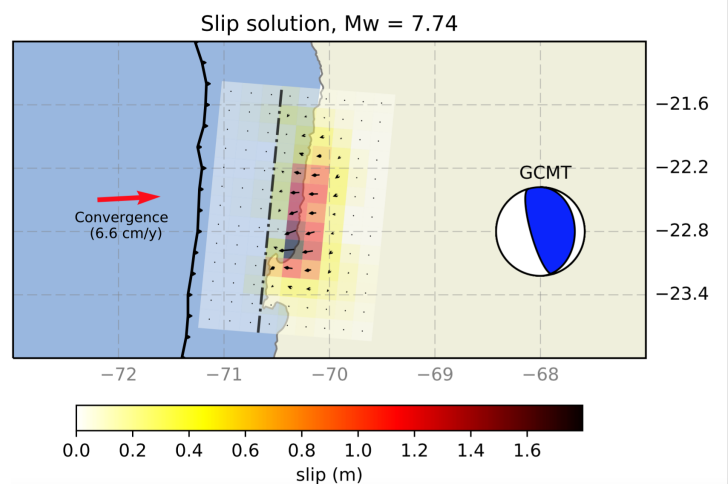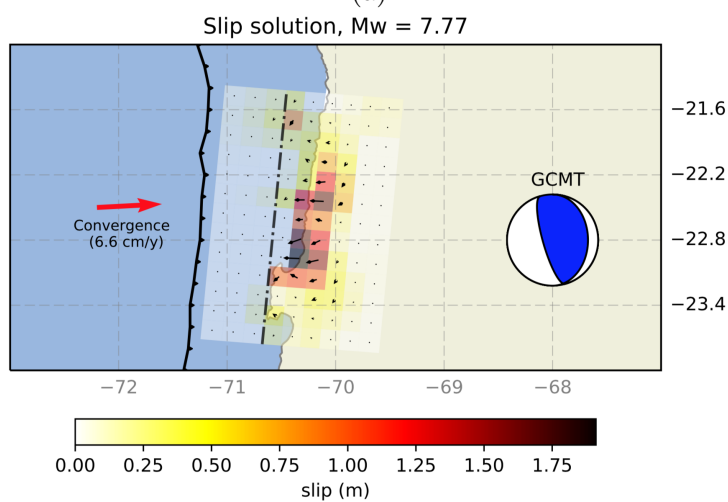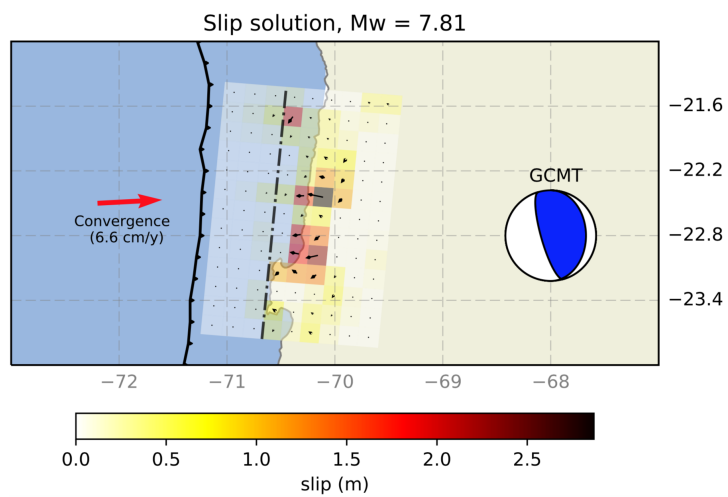
Figure A.3: Slip obtained from the inversion the 2007 M_w 7.7 Tocopilla earthquake using GPS and InSAR data and a regularized least squares with a positivity constraint method for different regularization factor ϵ.
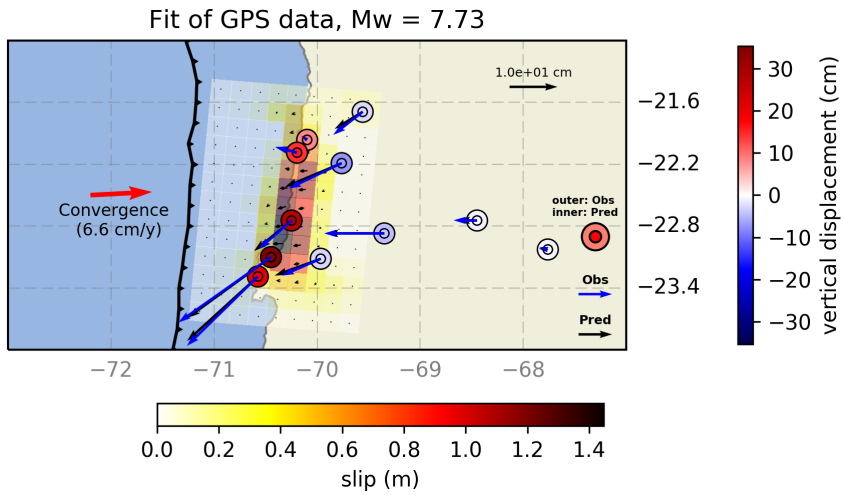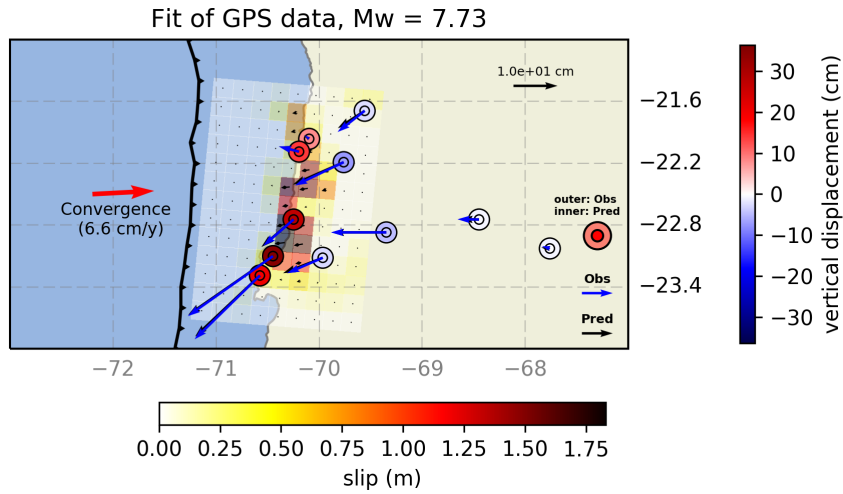
(a)

(b)

(c)
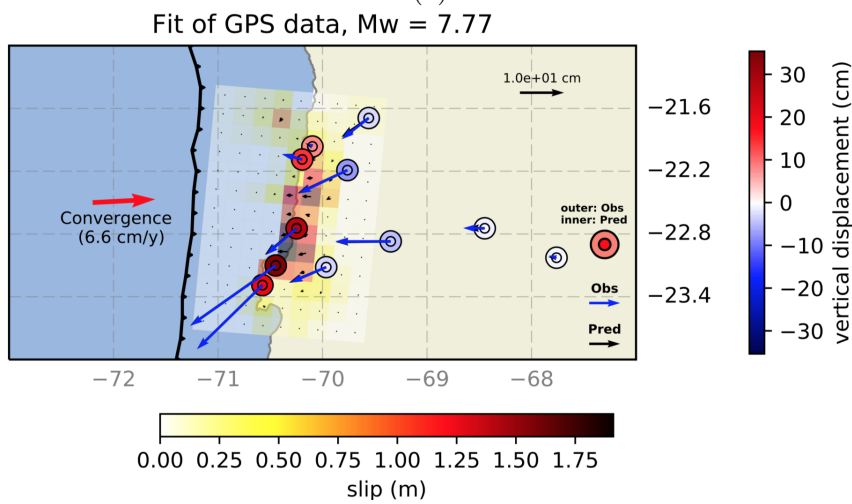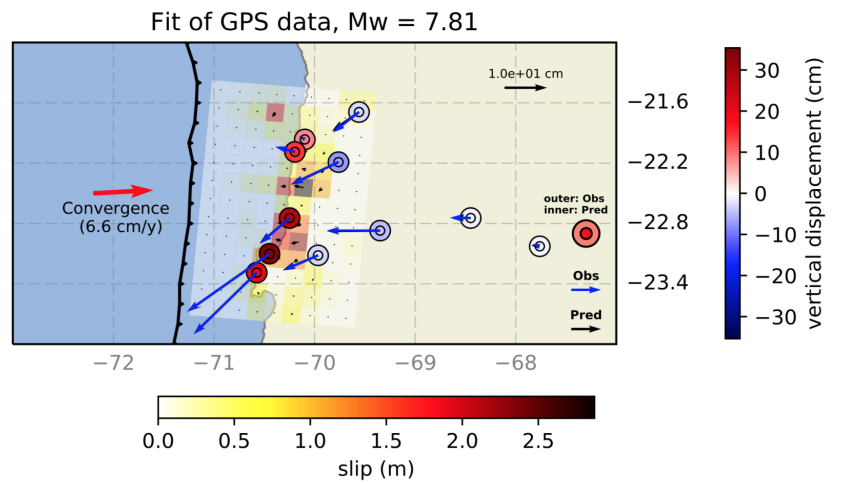
Figure A.4: Fit of GPS data set from the inversion the 2007 Tocopilla earthquake using GPS and InSAR data (track 96) and a regularized least squares with a positivity constraint method for different regularization factor ϵ.
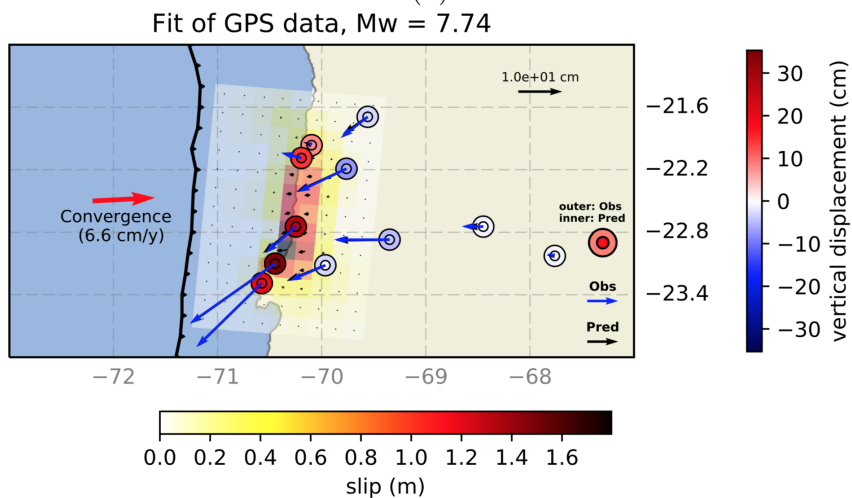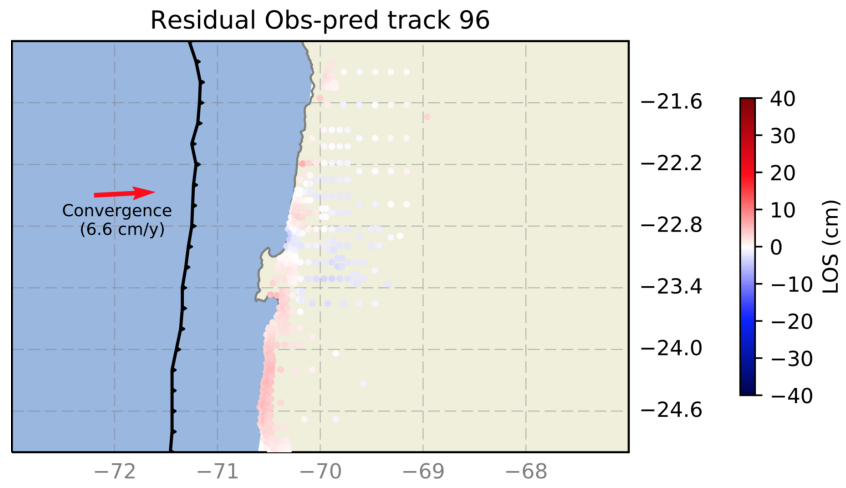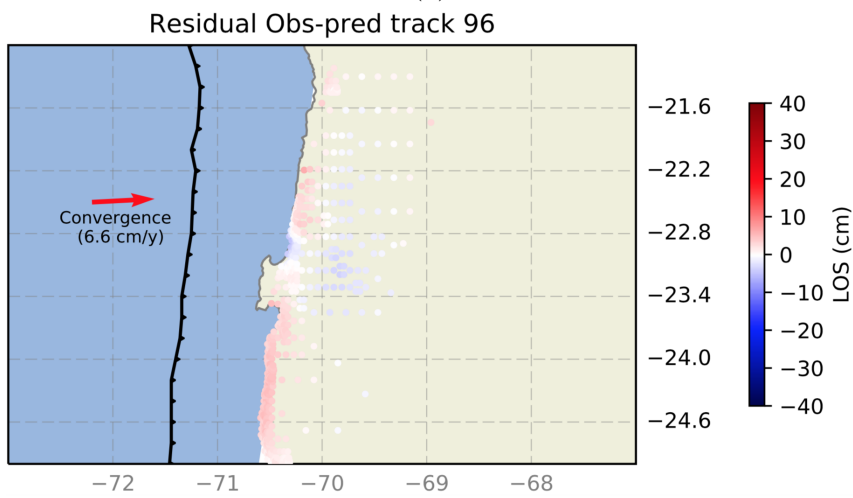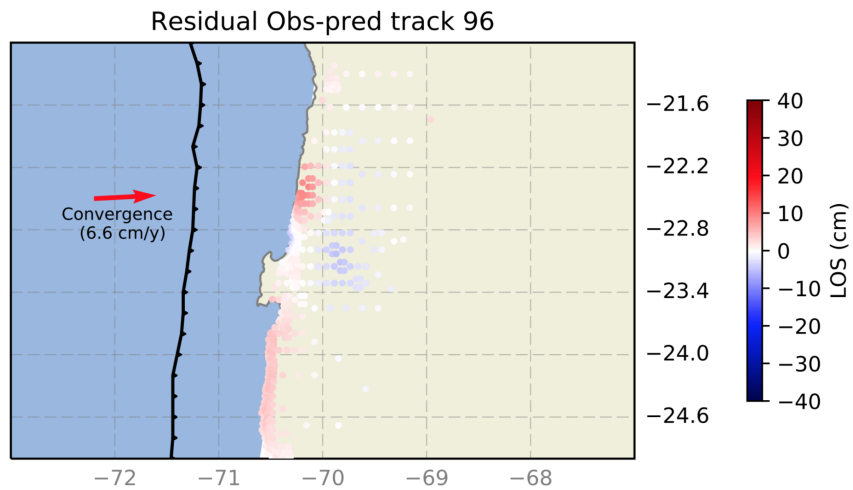
Figure A.5: Residual for InSAR ALOS displacements obtained from the inversion of the 2007 Tocopilla earthquake using GPS and InSAR data (track 96) for the InSAR prediction.

When we use the regularized least squares method it is necessary to define a criterion for the selection of the regularization factor ϵ. Usually this criterion is based on the use of the L-Curve, which compares $\|\mathbf{W_x}(\mathbf{G_e}\theta - \mathbf{d})\|$ with $\|\mathbf{W_h}(\mathbf{H}\theta - \mathbf{h_0})\|$, choosing the model which generates an inflexion point in the L-curve (Figure A.6), balancing the residual with the regularization term. Nevertheless, this criterion does not obey any physical constraint, but an heuristical condition, adding a bias to the solution.
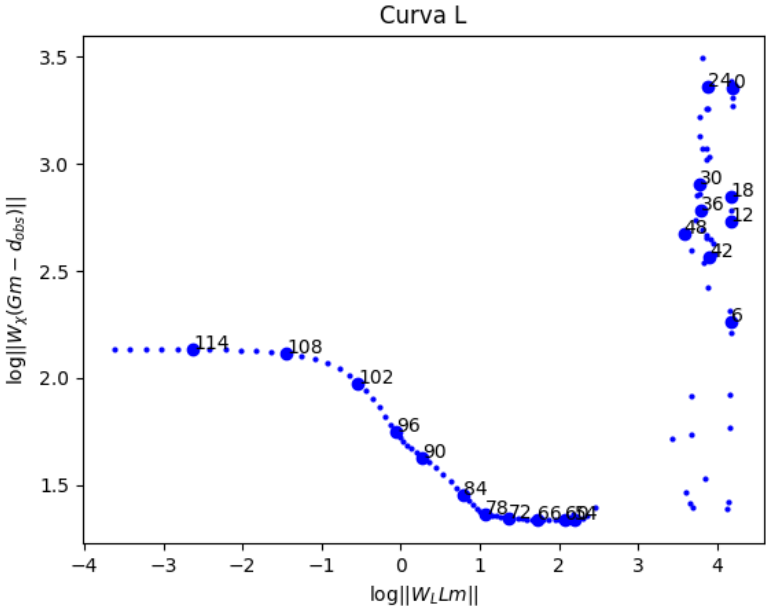
Figure A.6: L curve for the least squares problem with a positivity constraint when GPS and InSAR data from track 96 are used to invert the slip of the 2007 Tocopilla earthquake. Numbers identify order index for the regularization factor ϵ.

Appendix B

Dirichlet Distribution

B.1 Dirichlet Distribution

Dirichlet distribution is used as initial prior distribution for the parallel to rake component of slip in the inversion using AlTar, to take into account the previous knowledge about the magnitude of the 2007 Tocopilla earthquake (Mw 7.7). These initial samples are a starting point for the exploration of the model parameter space. Choosing a set of samples that incorporates previous information about the expected slip distribution makes more likely to accept the samples in posterior stages, but do not restrict the later sampling, as the samples are constrained by the run priors of the processing. In this case, the Dirichlet distribution ensures that initial samples have a magnitude close to $M_w7.7$.

In this section we will explain briefly the Dirichlet distribution and why we use it as the prior distribution for initial samples of the slip parallel to rake.

The Dirichlet distribution is a continuous multivariate probability distribution defined by the pdf

$$f(x_1, x_2, ..., x_k, \alpha_1, \alpha_2, ..., \alpha_k) = \prod_{i=1}^{K} \frac{1}{B(\boldsymbol{\alpha})} x_i^{(\alpha_i-1)} \qquad (\text{B.1})$$

Where $\boldsymbol{\alpha}$ is a vector parametrizing the pdf, $B(\boldsymbol{\alpha})$ a normalization and x_i is a set of random numbers in which the sum of components is 1. (α) is defined as,

$$B(\boldsymbol{\alpha}) = \prod_{i=1}^{K} \frac{\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)} \qquad (\text{B.2})$$

The marginal distribution of the Dirichlet is the Beta function, which is illustrated in Figure B.1.

Total slip

Slip on each patch

(a)
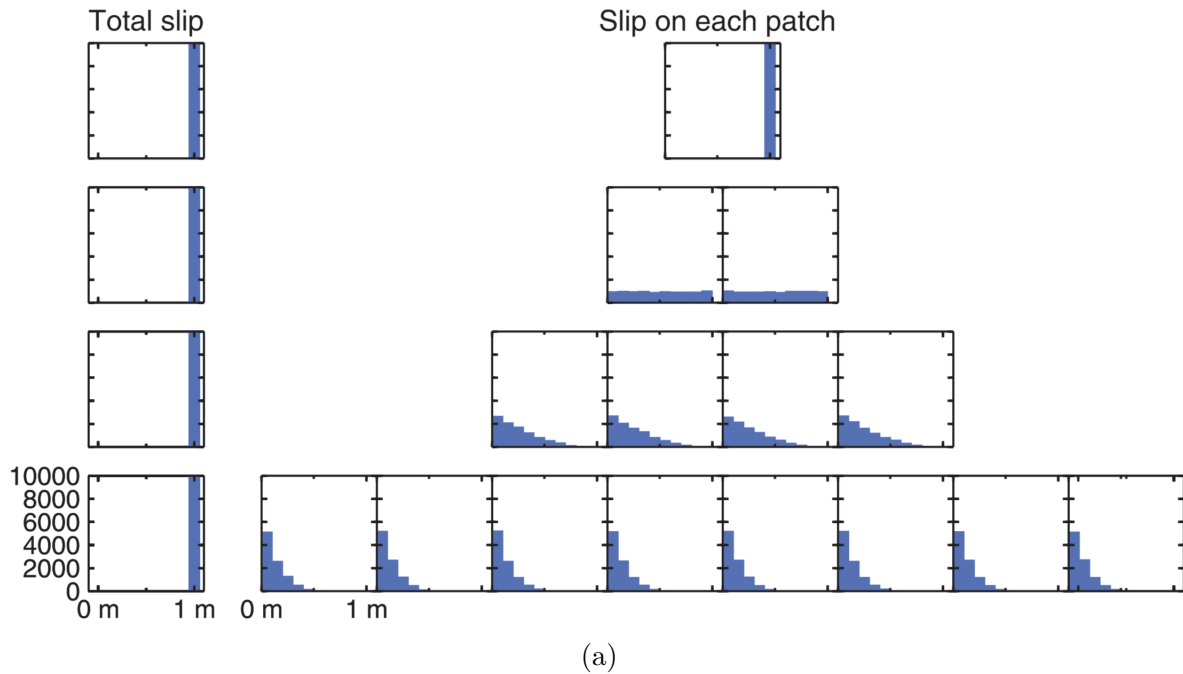
Figure B.1: Figure from *Minson et al.* (2013) illustrating the Dirichlet distribution for different amount of subfaults (1, 2, 4 and 8). The total amount of slip is constrained to 1 m, and the number of samples is 10000.

Applied to the inversion of the 2007 Tocopilla earthquake the condition of a magnitude close to a certain value implies a Dirichlet distribution respect to the seismic moment of the initial samples. The slip in that case is as we can see in Figure B.2. The slip decrease its values in depth as the shear modulus increase to ensure that the seismic moment will be close to a constant value. Figure B.2.c shows the distribution of magnitudes for the initial samples, where we can notice they are around Mw 7.7.
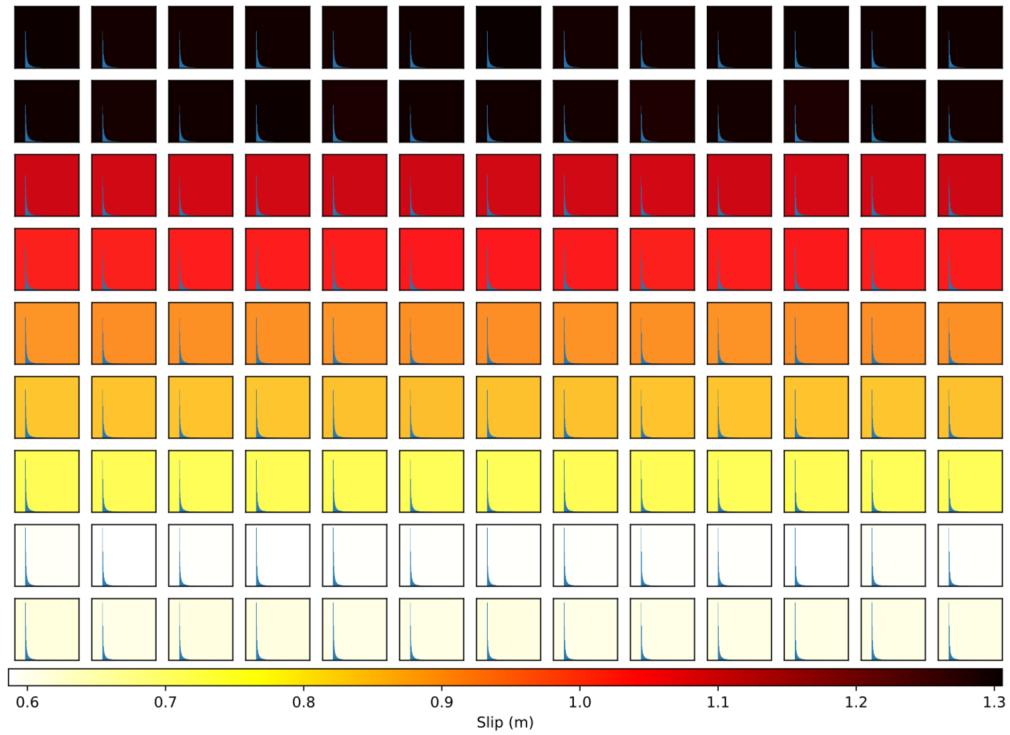
(a) Marginal histograms slip distribution
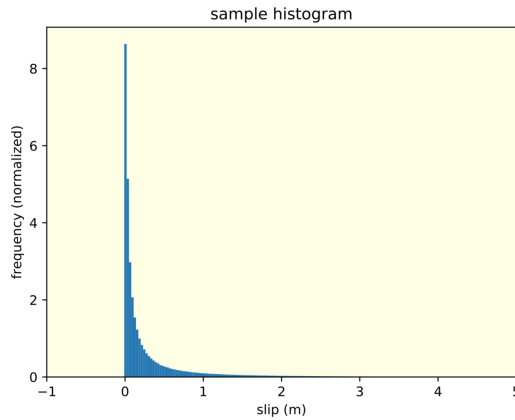
(b) Scale of marginal histograms
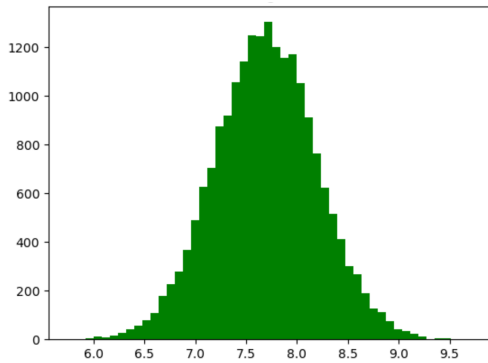
(c) Histograms magnitude

Figure B.2: a) Marginal distributions of initial slip samples employed in the inversion of the Tocopilla earthquake using AlTar. The number of samples is 10^6 (color indicates the mean value of the samples) b) Scale of the histogram located in the lower right corner c) Histogram of magnitudes for initial samples.

Bibliography

Bayes, T. (1763), An Essay towards Solving a Problem in the Doctrine of Chances., *By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S*, *1*(53), 370–418.

Beck, J., & K. Zuev (2013), Asymptotically Independent Markov Sampling: a new MCMC scheme for Bayesian Inference, *International Journal for Uncertainty Quantification*, *3*(5), 445–474.

Béjar-Pizarro, M., et al. (2010), Asperities and barriers on the seismogenic zone in North Chile: state-of-the-art after the 2007 Mw 7.7 Tocopilla earthquake inferred by GPS and InSAR data, *Geophysical Journal International*, *183*(1), 390–406.

Blewitt, G. (2015), 3.11 - gps and space-based geodetic methods, in *Treatise on Geophysics (Second Edition)*, edited by G. Schubert, second edition ed., pp. 307 – 338, Elsevier, Oxford, doi:https://doi.org/10.1016/B978-0-444-53802-4.00060-9.

Bouchon, M. (1981), A simple method to calculate Green's functions for elastic layered media, *Bulletin of the Seismological Society of America*, *71*(4), 959–971.

Ching, J., & Y.-C. Chen (2007), Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, *Journal of engineering mechanics*, *133*(7), 816–832.

Comte, D., & M. Pardo (1991), Reappraisal of Great Historical Earthquakes in the Northern Chile and Southern Peru Seismic Gaps, *Natural Hazards*, *4*(1), 23–44.

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

Delouis, B., M. Pardo, D. Legrand, & T. Monfret (2009), The Mw 7.7 Tocopilla Earthquake of 14 November 2007 at the Southern Edge of the Northern Chile Seismic Gap: Rupture in the Deep Part of the Coupled Plate Interface, *Bulletin of the Seismological Society of America*, *99*(1), 87–94.

Duputel, Z., P. S. Agram, M. Simons, S. E. Minson, & J. L. Beck (2014), Accounting for prediction uncertainty when inferring subsurface fault slip, *Geophysical Journal International*, *197*(1), 464–482.

Duputel, Z., et al. (2015), The Iquique earthquake sequence of April 2014: Bayesian modeling accounting for prediction uncertainty, *Geophysical Research Letters*, *42*(19), 7949–7957.

Gusman, A. R., S. Murotani, K. Satake, M. Heidarzadeh, E. Gunawan, S. Watada, & B. Schurr (2015), Time series analysis of InSAR data: Methods and trends., *Geophysical Research Letters*, *42*(4), 90–102.

Hayes, G. P., D. J. Wald, & R. L. Johnson (2012), Slab1.0: A three-dimensional model of global subduction zone geometries, *Journal of Geophysical Research: Solid Earth*, *117*(B1), doi:10.1029/2011JB008524.

Herrmann, R. B. (2013), Computer programs in seismology: An evolving tool for instruction and research, *S Research Letter*, *84*(1), 1081–1088.

Husen, S., E. Kissling, & G. Flueh, E. R and. Asch (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*(2), 687–701.

Ide, S. (2013), The proportionality between relative plate velocity and seismicity in subduction zones, *Nature Geoscience*, *6*(9), 780–784.

Jara, J., et al. (2018), Kinematic study of Iquique 2014 Mw 8.1 earthquake: Understanding the segmentation of the seismogenic zone., *Earth and Planetary Science Letters*, *503*(0012-821X), 131–143.

Jaynes, E. T. (1957), Information Theory and Statistical Mechanics., *Phys. Rev.*, *106*(4), 620–630.

Kendrick, E., M. Bevis, B. Smalley, B. Brooks, R. Barriga Vargas, E. Lauria, & L. Paulo Souto Fortes (2003), The Nazca–South America Euler vector and its rate of change, *Journal of South American Earth Sciences*, *16*(1), 125–131.

Leon Rios, S., S. Ruiz, A. Maksymowicz, F. Leyton, A. Fuenzalida, & R. Madariaga (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.

Lohman, R. B., & M. Simons (2005), Some thoughts on the use of insar data to constrain models of surface deformation: Noise structure and data downsampling, *Geochemistry, Geophysics, Geosystems*, *6*(1), 1525–2027.

Maksymowiczk, A. (2015), The geometry of the Chilean continental wedge: Tectonic segmentation of subduction processes off Chile, *Tectonophysics*, *659*(659), 183 – 196.

McGrayne, S. B. (2012), *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged Triumphant from Two Centuries of Controversy*, Yale University Press, first published May 14th 2011.

Metropolis, A. R. M. R. A. T., N., & E. Teller (1953), Equation of state calculations by fast computing machines., *Journal of Chemical Physics*, *21*(6), 1087–1092.

Minson, S. (2010), A Bayesian approach to earthquake source studies, Ph.D. thesis, California Institute of Technology.

Minson, S., M. Simons, & J. Beck (2013), Bayesian inversion for finite fault earthquake source models I theory and algorithm, *Geophysical Journal International*, *194*(3), 1701–1726.

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

Ortega, F. (2013), Aseismic Deformation in Subduction Megathrusts: Central Andes and North-East Japan, Ph.D. thesis, California Institute of Technology.

Osmanoglu, B., F. Sunar, S. Wdowinski, & E. Cabral-Cano (2016), Time series analysis of InSAR data: Methods and trends., *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*(1), 90–102.

Pasten-Araya, F., et al. (2018), Fluids Along the Plate Interface Influencing the Frictional Regime of the Chilean Subduction Zone, Northern Chile, *Geophysical Research Letters*, *45*(19), 10,378–10,388.

Peyrat, S., R. Madariaga, E. Buforn, J. Campos, G. Asch, & J. P. Vilotte (2010), Kinematic rupture process of the 2007 Tocopilla earthquake and its main aftershocks from teleseismic and strong-motion data, *Geophysical Journal Internationa*, *182*(3), 1411–1430.

Poli, P., A. Maksymowicz, & S. Ruiz (2017), The Mw 8.3 Illapel earthquake (Chile): Preseismic and postseismic activity associated with hydrated slab structures, *Geology, 45*.

Pritchard, M. E., & M. Simons (2006), An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behavior, *Journal of Geophysical Research: Solid Earth*, *111*(B08405), 1.

Pritchard, M. E., M. Simons, P. A. Rosen, S. Hensley, & F. H. Webb (2002), Co-seismic slip from the 1995 July 30 Mw= 8.1 Antofagasta, Chile, earthquake as constrained by InSAR and GPS observations, *Geophysical Journal International*, *150*(2), 362–376.

Ruiz, S., & R. Madariaga (2018), Historical and recent large megathrust earthquakes in Chile, *Tectonophysics*, *733*(1), 37 – 56.

Ruiz, S., M. Métois, A. Fuenzalida, J. Ruiz, F. Leyton, R. Grandin, C. Vigny, R. Madariaga, & J. Campos (2014), Intense foreshocks and a slow slip event preceded the 2014 Iquique Mw 8.1 earthquake, *Science*.

Schurr, B., G. Asch, M. Rosenau, R. Wang, O. Oncken, S. Barrientos, P. Salazar, & J. Vilo (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*, *117*(B05305), doi:10.1029/2011JB009,030.

Schurr, B., et al. (2014), Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake, *Nature, 512*.

Simons, M., & P. Rosen (2015a), 3.12 - interferometric synthetic aperture radar geodesy, in *Treatise on Geophysics (Second Edition)*, edited by G. Schubert, second edition ed., pp. 339 – 385, Elsevier, Oxford, doi:https://doi.org/10.1016/B978-0-444-53802-4.00061-0.

Simons, M., & P. Rosen (2015b), *Interferometric Synthetic Aperture Radar Geodesy. In: Gerald Schubert (editor-in-chief) Treatise on Geophysics, 2nd edition, Vol 3.*, Society for industrial and applied mathematics, Philadelphia (PA), oxford: Elsevier; 2015. p. 339-385.

Simons, M., et al. (2011), The 2011 Magnitude 9.0 Tohoku-Oki Earthquake: Mosaicking the Megathrust from Seconds to Centuries, *Science*, *332*, 1421–1425.

Song, T.-R. A., & M. Simons (2003), Large Trench-Parallel Gravity Variations Predict Seismogenic Behavior in Subduction Zones, *Science*, *301*(5633), 630–633.

Tarantola, A. (2005), *Inverse problem theory : and methods for model parameter estimation*, Society for industrial and applied mathematics, Philadelphia (PA), 1987 version of the book completely rewritten.

Tong, X., et al. (2010), The 2010 Maule, Chile earthquake: Downdip rupture limit revealed by space geodesy., *Geophysical Research Letters*, *37*(24).

Victor, P., M. Sobiesiak, J. Glodny, S. Nielsen, & O. Oncken (2011), Long-term persistence of subduction earthquake segment boundaries: Evidence from Mejillones Peninsula, northern Chile, *Journal of Geophysical Research*, *116*(2), 362–376.

Wyss, M., & S. Wiemer (1999), How Can One Test the Seismic Gap Hypothesis? The Case of Repeated Ruptures in the Aleutians, *pure and applied geophysics*, *155*(2), 259–278.

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