Title

Development of Specialized Nonlinear Inversion Algorithms, Basis Functions, Eikonal Solvers, and Their Integration for Use in Joint Seismic and Gravitational Tomographic Inversion

Author

Abatchev, Zagid

Publication Date

2019

UNIVERSITY OF CALIFORNIA

Los Angeles

Development of Specialized Nonlinear Inversion Algorithms, Basis Functions, Eikonal Solvers, and Their Integration for Use in Joint Seismic and Gravitational Tomographic Inversion

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Geophysics and Space Physics

by

Zagid Abatchev

2019

ABSTRACT OF THE DISSERTATION

Development of Specialized Nonlinear Inversion Algorithms, Basis Functions, Eikonal Solvers, and Their Integration for Use in Joint Seismic and Gravitational Tomographic Inversion

by

Zagid Abatchev
Doctor of Philosophy in Geophysics and Space Physics
University of California, Los Angeles, 2019
Professor Paul M. Davis, Chair

1 Abstract

Many approaches and packages have been developed for tomographic inversion of seismic data and have consistently remained in high demand in both academic and industrial geophysics for subsurface imaging. However, our survey of published implementations has shown them to suffer from shortcomings. They include: low computational efficiency leading to overreliance on gradient-based inversion methods, non-robust forward models, and overparametrization of invertible fields leading to overconfidence in parameter resolvability and susceptibility to local minima. Additionally, many suffer from poor algorithmic documentation, lacking source code availability, and complicated dependencies of many published packages, resulting in difficulty with their reproducibility, testing, and adaptation. Finally, most available packages do not incorporate different measurement sets, such as gravitational anomaly data, to better constrain seismic inversion.

To address these issues, we have developed our own implementation and integration of an improved set of algorithms for seismic and gravitational tomography and applied it to

a seismic data set from the Peruvian Andes obtained in a joint experiment with Caltech between 2008 and 2013. A summary of the work completed includes the development and/or demonstration of:

An efficiently parametrized scalar field basis function incorporating 2D and 3D Natural Neighbor Interpolation specialized for mantle tomography and mapping of irregular discontinuities such as a perturbed Moho. An emphasis on perturbation continuity and smoothness makes it suitable for use with appropriately damped gradient based inversion methods.

A hybrid Eikonal solver first arrival forward model, based primarily on the Fast Marching Method (FMM)(Sethian, 1997) and utilizing key elements of Vidale Finite Difference Method (VFD)(Vidale, 1991) in order to reduce model error. We demonstrate a simple and common velocity configuration in which VFD experiences a catastrophic failure, while our variant (designated FMM-VFD) performs with well constrained and small model error.

A hybrid inversion algorithm incorporating iterating sequential application of a custom Damped Gauss-Newton (DGN) algorithm and a Markov Chain Monte Carlo (MCMC) method, which allows for problem specific balancing of inversion robustness versus efficiency. Integration of the above into a joint gravitational, local seismic, and teleseismic master inversion algorithm.

In a tangential development with potential for future integration, we developed an experimental $O(n^3 log(n))$ quasi-parallel and sparse implementation of an Acoustic Wave Equation solver wavefront arrival model using FMM-VFD field pre-initialization.

Full disclosure and publication of key source code, written in a maximally simplified manner with no dependencies outside the C++ standard template library, included in a fully

commented form to promote its verification, adaptation, and free use by others.

Local and teleseismic arrivals, as well as the published gravitational anomaly field over Peru have been used to invert jointly for local event locations, 3D mapping of the Moho, and mantle velocity perturbation structure, showing features consistent with the subducting Nazca slab. From the obtained results, we believe to have potentially identified the region in which the subduction of the Nazca plate undergoes flattening.

The dissertation of Zagid Abatchev is approved.

David D. Jackson

Troy A. Carter

Gilles F. Peltzer

Paul M. Davis, Committee Chair

University of California, Los Angeles

2019

Contents

List of Figures

List of Tables

Acknowledgments

Mirzafer and Luiza Abatchev - For raising me, giving me an education, guidance, and always providing support and encouragement.

Polina Abatcheva - For unfailing support and company. For proofreading this thesis and typographical editing.

Nicholas Scianmarello - For valuable feedback, teaching me C++, and embedded programming.

Gary Binder - For valuable feedback, teaching me to program in Python, Markov Chain Monte Carlo methods, and helpful ideas for future work, as well as proofreading this thesis.

Constantine Sideris - For teaching me to program in C, optimization methods, and helpful programming feedback.

Kolton Drake - For teaching me to program in C++ and helpful programming feedback.

David Michels - For thoughtful feedback and suggestions both in research and writing.

Igor Stubailo - For teaching me to use Linux, seismic acquisition, and field deployment.

David Jackson - For providing me with extensive wisdom in inverse theory and for his time spent serving on my committee.

Troy Carter - For his dedication and time spent serving on my committee and for feedback from a physics point of view.

Gilles Peltzer - For his time spent serving on my committee and for feedback from a geochemical point of view. For being an excellent professor for which to teach.

Paul Davis - For exceeding all realistic expectations of a PhD advisor. For teaching me to be independent, helping guide my research direction, and providing important ideas and suggestions to overcome problems. For teaching me how to approach research, field work, and scientific rigor. For being an excellent professor for which to teach.

NSF - Funding for PeruSE data acquisition and analysis.

SCEC - Funding for LASSIE data acquisition and analysis.

Vita

Education

Bachelor of Science in Physics at California Institute of Technology. 2007-2011

Master of Science in Geophysics and Space Physics at University of California, Los Angeles. 2011-2014

Research experience

Undergraduate Researcher in Physics under Professor Buyung Kim. Boise State University. Summer 2008

Undergraduate Researcher in Chemistry under Professor Nate Lewis and Dr. Shannon Boettcher, California Institute of Technology. 2009-2011

Graduate Researcher in Geophysics under Professor Paul Davis. University of California, Los Angeles, 2011-2019

Teaching experience

Teaching Assistant/Apprentice in lower and upper division departmental courses at department of Earth, Planetary, and Space Sciences, University of California, Los Angeles. 2013-2018

2 Introduction

Our initial need for a new tomographic inversion approach was motivated by the need to determine upper mantle structure under the Peruvian Andes, which are characterized by a convergent plate boundary involving the oceanic Nazca and continental South American lithospheric plates. This type of boundary produces what is known as subduction, a process in which the denser of two converging plates sinks toward the lower mantle, with compression stresses causing a crustal thickening and associated mountain building processes (Stacey, Davis 2008). Together with subduction induced volcanism (Stacey, Davis 2008), this is believed to be responsible for the creation of the Andes mountain range, which span a significant majority of the Western end of the South American continent.

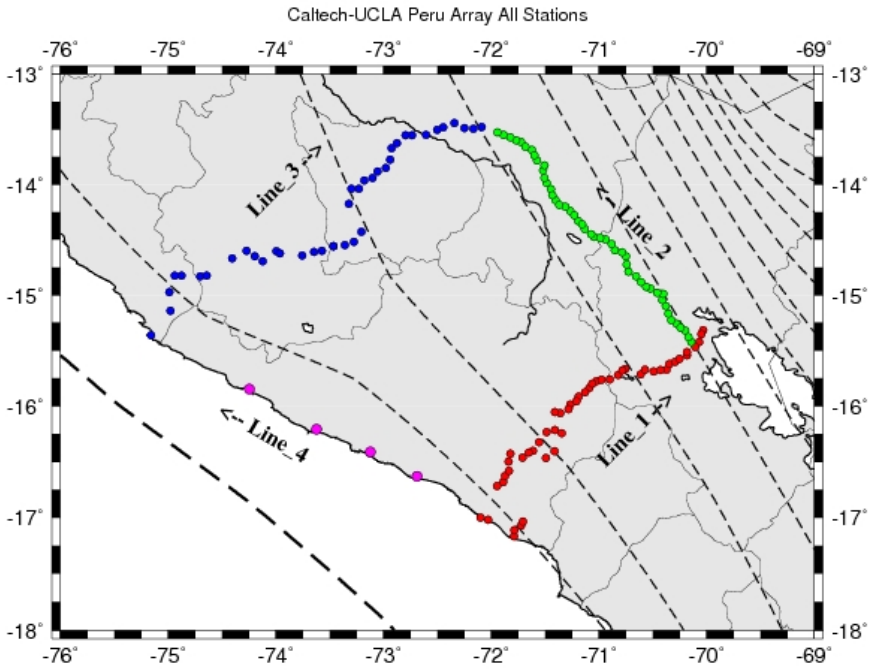


Figure 1: Peru Subduction Experiment (Source: Robert Clayton, Caltech). Colored circles indicate locations of seismic stations deployed by Caltech and UCLA by their designated line. Data collected between 2008 and 2013.

The region under Southern Peru is believed to contain anomalously deep seismicity, cat-

aloged by USGS PDEs (Preliminary Determination of Epicenters catalog) at depths of up to 600km. Additionally, the pattern of seismicity appears to indicate a variable plate subduction angle, the profile of which has not been well constrained. In order to gain a better understanding of the structures involved, the natural approach is through the use of seismic tomography, which has historically been considered the most effective means for the study of both deep and shallow Earth structure. One of the primary types of seismic tomography, known as body wave travel time tomography, uses measured times of arrival of seismic body waves (compressional p-waves, and/or shear s-waves) in order to construct a velocity field that would produce travel times consistent with those measured. This is an example of what is broadly known as an inverse problem; the determination of inputs that produce observed outputs. Other methods for determining deep earth structure include measurement of Earth's natural oscillation frequencies (known as "normal modes"), and use of surface waves. In contrast to normal mode and surface wave inversion, travel time tomography can be applied for imaging both small and large scale anomalies to depths of interest to our problem.

In posing the problem, we must determine the types of hypothetical features, which we would want to resolve using our data. In the case described, one significant feature that we expect to observe is a nonuniform crust-mantle boundary depth, known as the Mohorovic discontinuity, or the Moho. In a standard radially symmetric earth reference model such as PREM (Dziewonski, Anderson 1981) or IASP91 (Kennet et al. 1991), the Moho is defined to be at the depth of 7km under ocean and 35km under continents (Stacey, Davis 2008), though in the physical earth, these values are subject to variability.

Much like in the way icebergs float in water, the crust floats on the mantle material below. When the large scale terrain altitude is significantly deviated from sea level, the crust along with the uppermost brittle layer of the mantle, called the "lithosphere", rises or sinks in

order to be in isostatic equilibrium with the mantle to a state called Airy isostasy. Because the elevation in Peru is regionally variable, the Moho structure underneath is expected to be nontrivial. The Moho is characterized by a significant mineralogical discontinuity from basaltic/granitic material to peridotite, with an accompanying p-wave velocity discontinuity from approximately 6.5km/s above the boundary, to approximately 8km/s below. Theoretically this can be resolved by body wave tomography.

Additionally, in the process of oceanic plate subduction, a cold lithospheric plate of a thickness on the scale of 100-120km descends into much hotter surrounding mantle. Surrounding heat diffuses in to the plate, however the diffusion rate is slow enough to maintain a significant temperature differential. This manifests as a region of anomalously high body wave velocity, which can also theoretically be resolved by body wave tomography.

An early study of the data used in our experiment published by Philips et al. 2014 used data from seismic line 1 (Figure 1) to carry out receiver function and 2D first arrival tomographic inversion under the line. Results in Figures 2 and 3 indicated the presence of a subducting slab under line 1. We aimed to expand on this by using the full PeruSE network, as well as published gravitational anomaly data, to carry out a 3D tomographic inversion of the area under and near the full seismic network.

Having identified the two key features to be resolved, we established a need for an inversion routine suitable for our problem, goals, and constraints. In general terms, an inverse problem is any problem, where one seeks to solve for the inputs or description of a system from the outputs it produces. Mathematically, it involves taking an observed output data set y, a model $f(\beta)$ which describes the system in question as a function of a parameter set β and seeks the values of β such that $f(\beta) = y$. In other words, one wants to describe the parameter set of a model that will give outputs matching the observed data. Using simpli-

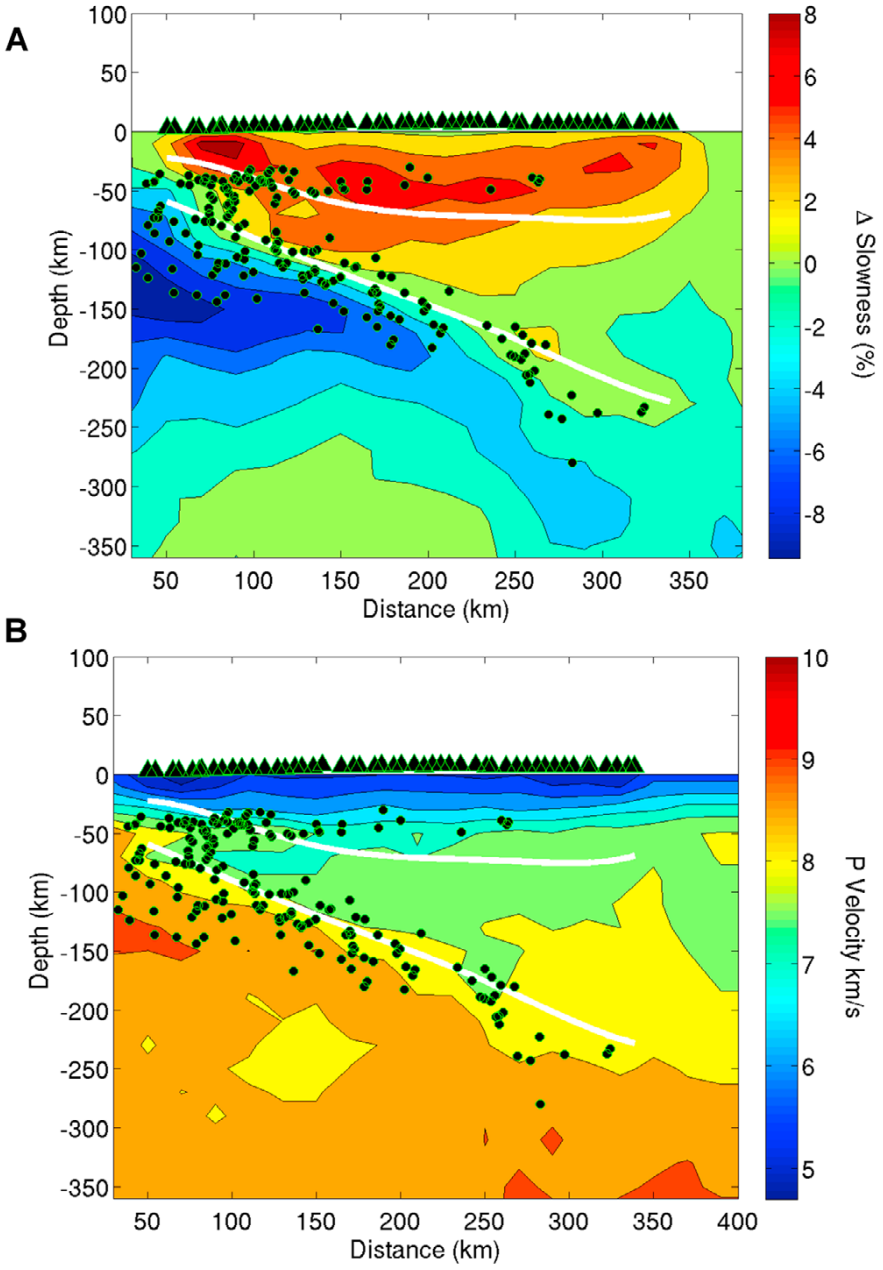


Figure 2: Philips et al. 2014: 2D first arrival tomographic inversion using selected data from PeruSE

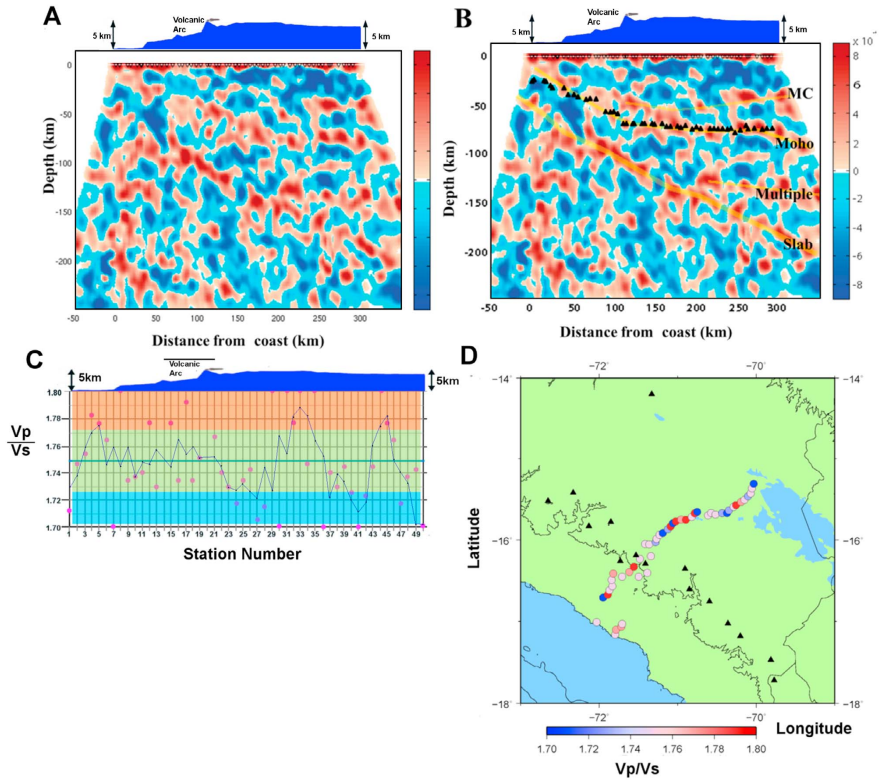


Figure 3: Philips et al. 2014: 2D receiver function inversion using selected data from PeruSE

fied models of physical systems and observables with measurement noise, this solution rarely exists, and real world inversions instead attempt to minimize the misfit between a model output and observables with respect to β. Additionally, there may exist many sets β which produce the same misfit or even model outputs. This is known as nonuniqueness of solutions and is also something that must be addressed by a practical inversion algorithm, both through application of stability tests, as well as through incorporating prior known data to select against physically implausible parameter sets. One fundamental measure to describe the sensitivity of a model's outputs with respect to changes to inputs is known as robustness, and is a feature that is sought in both inverse and forward models.

In order to create a first arrival forward model $f(\beta)$ that satisfies the specific requirements of our inverse problem, we examined 4 model classes, including ray tracers, Eikonal solvers, Hyugens wavelet solvers, and wave equation solvers. Ray tracers work either by exploiting

Fermat's principle, which states that a ray path between two points will be minimized (or maximized) with respect to time. This is done by either parametrizing a path and minimizing point to point path integral time (global application of Fermat's principle), or by integration of segment times while solving for the path iteratively through application of Snell's law (local application of Fermat's principle). An Eikonal solver is a model which numerically solves the Eikonal equation throughout a grid:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = \frac{1}{c(x,y,z)^2} \tag{1}$$

Hyugens and shortest path solvers work based on minimizing point to point time by applying the Huygens principle, which states that every point on a wavefront is itself a source of waves. Finally, first arrival wave equation solvers implement a numeric solution scheme to solve the 3D acoustic wave equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{c(x,y,z)^2}\frac{\partial^2 u}{\partial t^2} \tag{2}$$

The fastest known models for seismic forward modeling are variations of ray tracers. However, in certain physical configurations, those forward models are fragile to inputs and prone to unpredictable errors that can be difficult to constrain. This can include local minimum solutions (multipathing), and nonexistent solutions for direct arrivals in shadow zones. Therefore, ray tracers are generally considered nonrobust in many heterogenous media, especially those containing discontinuities. Conversely, many existing robust nonlinear first arrival forward models, such as the Fast Marching Method(FMM)(Sethian, 1997) and Shortest Path/Dijkstra algorithm Methods(SPM)(Moser 1991, Kvasnicka 1994), are highly robust, but suffer from low computational efficiency (to varying degrees), require use of greedy (gradient based) inversion approaches for acceptable run times, exposing them to increased risk of converging on solutions at local minima for any computationally bound inversion.

One exception to this is an Eikonal solver class algorithm published by John Vidale in 1990, which we know as the "Vidale Finite Difference" model. It is significantly faster than all major known robust first arrival models, while maintaining robustness over a larger range of conditions. However, it too has limitations, including the inability to find certain diffracted arrivals known as head waves in scenarios with large velocity contrasts. To address this, we have created a variant of the Fast Marching Method, incorporating elements of Vidale's Finite Difference Method implemented in a way to overcome the robustness limitations of the original Vidale Finite Difference model, while being significantly more accurate and thus faster than Fast Marching Method alone. Our testing has shown it to be robust in physically anticipated scenarios, while being more computationally efficient than existing robust first arrival methods. Additionally, we prototyped a parallelized GPU implementation of a first arrival time field initialized acoustic wave equation solver for computation of wavefront arrival amplitudes and wave shape at $O(n^3 log(n))$ time, compared to $O(n^4)$ time required by traditional wave equation solvers.

As the primary goal of our inversions was a physical description of the velocity structure under the Peruvian Andes, the parametrization of this field is one of the central problems in our inversion. One of the fundamental trade-offs in the parametrization of a parameter set is between resolution and resolvability, the limits of which are determined by the amount and distribution of observed data, as well as how the data is used. A survey of available methods and published inversions can lead one to note that many current geophysical inversion approaches disregard fundamental limitations on inversion parameter resolvability set by the Shannon-Hartley Theorem, which defines the information carrying capacity of a finite frequency waveform with noise:

$$C_c = B_c * log_2(1 + snr) \tag{3}$$

Where C_c is the channel capacity in bits per second, B_c is channel bandwidth, and snr is the

signal to noise ratio. For a waveform inverting up to n_c cycles of the assumed upper cutoff frequency, this gives a theoretical waveform bit capacity W_b:

$$W_b = n_c * log_2(1 + snr) \tag{4}$$

As a result, disregarding these limitations may significantly overparametrize invertible basis functions without properly describing the null space or parameter resolvability, leading to statistically questionable inversion results. Specifically, overparametrization is often a natural consequence of using regularized basis functions while attempting to locally resolve features at scale sizes that should be resolvable at the receiver spacing and wavelengths used. It is also a shortcoming in many applications of full waveform inversions (FWI), which must necessarily invert a significantly larger parameter set in comparison to a point data inversion of the same spacial resolution and assume a significant majority of poorly resolvable parameters implicitly, in order to converge on a statistically questionable solution. This is due to extreme growth in required parametrization from the multiple frequency dependent relations that govern wave propagation (such as Q and scattering relations), which are well known but often disregarded for convenience. These become especially significant at the higher frequencies and smaller resolution scales, where finite frequency tomography is often credited with improved performance. A set of landmark papers on the study of parameter resolvability was published by Backus and Gilbert in 1967-1968, creating a framework for determining resolvability of invertible parameters by defining the concept of a resolution matrix. In the context of a linear inversion, the resolution matrix can be defined as:

$$R = [J^T J + D]^{-1} J^T J \tag{5}$$

However, increasingly large parameter sets have resulted in a tendency for many to fail to provide resolution matrices or proper error estimations with their results.

Another indication of overparametrization and overconfidence in the parameter space is the existence of a large null space in inverted parameter space. This, by extension, would manifest in those parameter subsets being poorly resolved, if a resolution matrix or a parameter co-variance matrix was to be computed. In 2005, Trampert demonstrated that in the case of finite frequency surface wave tomography, the claimed improvement of invertible resolution of available data over an infinite frequency inversion is in fact lost in the null space.

In order to invert for a field with heterogeneous feature scales and nonuniform model output sensitivity to input parameters while using a sensibly resolved parameter set with a small null space, multi scale parametrization must be used together with robust first arrival forward models. Additionally, the field parametrization should ideally have a smooth sensitivity matrix (Jacobian) in its working parameter space.

One well known and widely used multiscale parametrization used in inversions is Nearest Neighbor Interpolation. However, it suffers from strong discontinuities, which can cause difficulties for gradient based approaches, necessitating significantly higher forward model computational costs by requiring larger grids to perform within acceptable error bounds. Additionally, nearest neighbor interpolation scheme discontinuities have inherently higher forward model error than smoothly parametrized velocity fields, which requires finer grid sizes to be used at great computational expense. Recent advances in computer vision have yielded a set of methods known as Natural Neighbor Interpolation (NNI), which address these issues by allowing for basis functions that generate smooth and continuous fields. This class of interpolation was first implemented in geophysics by Sambridge et al. 1995. We have created our own parametrically efficient 3D multiscale basis function incorporating an NNI implementation known as Discrete Sibson Interpolation (DSI)(Park et al. 2010) for simultaneously resolving 3D structure of strong discontinuities such as the mantle crust boundary, and the more smooth and continuous 3D scalar field body perturbations that would be ex-

pected in the mid and upper mantle.

We do this by rendering an interpolation of a perturbed 2D surface (from hereon referred to as manifold, in order to avoid confusion with the Earth's surface, which is also referenced frequently) in 3 dimensional space to represent a boundary between a fixed velocity upper layer representing the earth's crust, and a 3D Natural Neighbor Interpolation of the perturbed mantle velocity below. The boundary layer around the manifold between the nonuniform bottom layer and uniform upper layer is blended linearly. The same approach is also used to parametrize an invertible density perturbation field, describing a density perturbation due to an offset discontinuity, which in our case is the Moho, as well as the smooth density perturbations in the mantle due to p-wave velocity perturbations relative to a uniform background.

Recognizing the stability and convergence problems posed by relying on gradient based inversion methods, the computational efficiency constraints of unguided Monte Carlo methods, and the nature of our basis function and model parametrization, we designed and implemented a hierarchy of inversion algorithms that incorporate elements of a guided Monte Carlo method in parameter and meta-parameter inversion initialization, together with a staged iterative approach using a customized gradient based inversion, and a Markov Chain Monte Carlo (MCMC) method. It is intended to have usability in cases ranging from well conditioned problems relying more on gradient based computationally efficient inversion, to poorly converging ill conditioned problems requiring heavier reliance on Monte Carlo methods for more robust outcomes. Current results show the Markov Chain evolution to perform reasonably well in eliminating artifacts on large scales, but in inversions that do not enforce smoothness, it can give questionable resolution of small-scale features.

Additionally, we recognized a need for the existence of an automated waveform timing and

phase measurement algorithm for application to nonlinear first arrival tomography. There is also a need for it to be integrated into the inversion routine and to allow for recursive re-measurement and re-selection of traces using converging confidence intervals. We have developed routines for the automated processing of teleseismic and local arrivals; however, the data set used is sufficiently small that in the interest of error reduction, the work presented was done using manually picked first arrivals for local and regional events, and phase offsets for teleseismic events.

Furthermore, we incorporated a simple gravitational anomaly forward model to map gravity perturbations on a set of points at a surface as a function of a perturbation density grid rendered using the aforementioned density parametrization. This generated surface gravity perturbation field is fitted to the corrected Bouguer Anomaly Field in the area of study, using published data from WGM2012. Our gravitational tomographic approach can be contrasted with a study published by Syracuse et. al, in that our goal is far field resolution rather than near surface resolution. Therefore the filtering, and range considerations in model rendering, as well as the implementation of the gravitational integration scheme used are dissimilar.

We then intergated the developed algorithms into a tomographic inversion routine and primarily tested on a data set obtained in the course of a seismic experiment carried out between 2008 and 2013 collaboratively with Caltech named PeruSE (Peru Subduction Experiment). Local and teleseismic arrivals, as well as the gravitational anomaly field, have been used to invert jointly for local event locations, 3D mapping of the Moho, and mantle velocity perturbation structure. The obtained inverted velocity model delineates the thickening of the Moho and locations of slab-like features. It provides evidence that the Peruvian Andes are isostatically compensated by thickened crust and shows where the slab makes a transition from steep subduction in the South to what is likely flat subduction in the North. For many regions the location of slabs are assumed to be given by the locations of the earth-

quakes (Hayes et al. 2012). In a highly heterogenous region such as Peru where the Moho may thicken by 30 km, we find (USGS catalog) earthquake locations based on standard Earth models can be mislocated by tens of kilometers.

This thesis will attempt to present an abridged, algorithmic, and quantitative summary of only the most currently relevant work, current results, and publication objectives. It should be emphasized that more than 95 percent of the time spent in the course of the author's thesis work was spent in tomographic algorithm development, which continued until the time of thesis submission and defense. For this reason, while due diligence was done to test for obvious software bugs and errors, a complete audit was not possible, and the results should be considered preliminary and subject to a more thorough verification. More complete and reliable inversions are expected to be completed for pending peer reviewed publication submissions. Some branching research performed, as well as detailed derivations were omitted in favor of a concise description of the most currently relevant work.

Additionally, in the course of writing of this thesis and confirmatory literature searches, we became aware of published work that precedes ours and is substantially similar to subelements of our body of work, some of which we developed independently and without awareness. Refraining from claims of original authorship on individual subroutines used, our primary claim of original research lies in the full implementation of all sub-units of this work, as well as its application to a poorly studied but geologically interesting area.

Finally, particular effort was given to making the developed code simple, easy to understand, dependency-free, and easily adaptable. The source code of all core algorithms presented is provided in the supplementary materials, and the reader is encouraged to use and adapt it as needed.

3 Methods

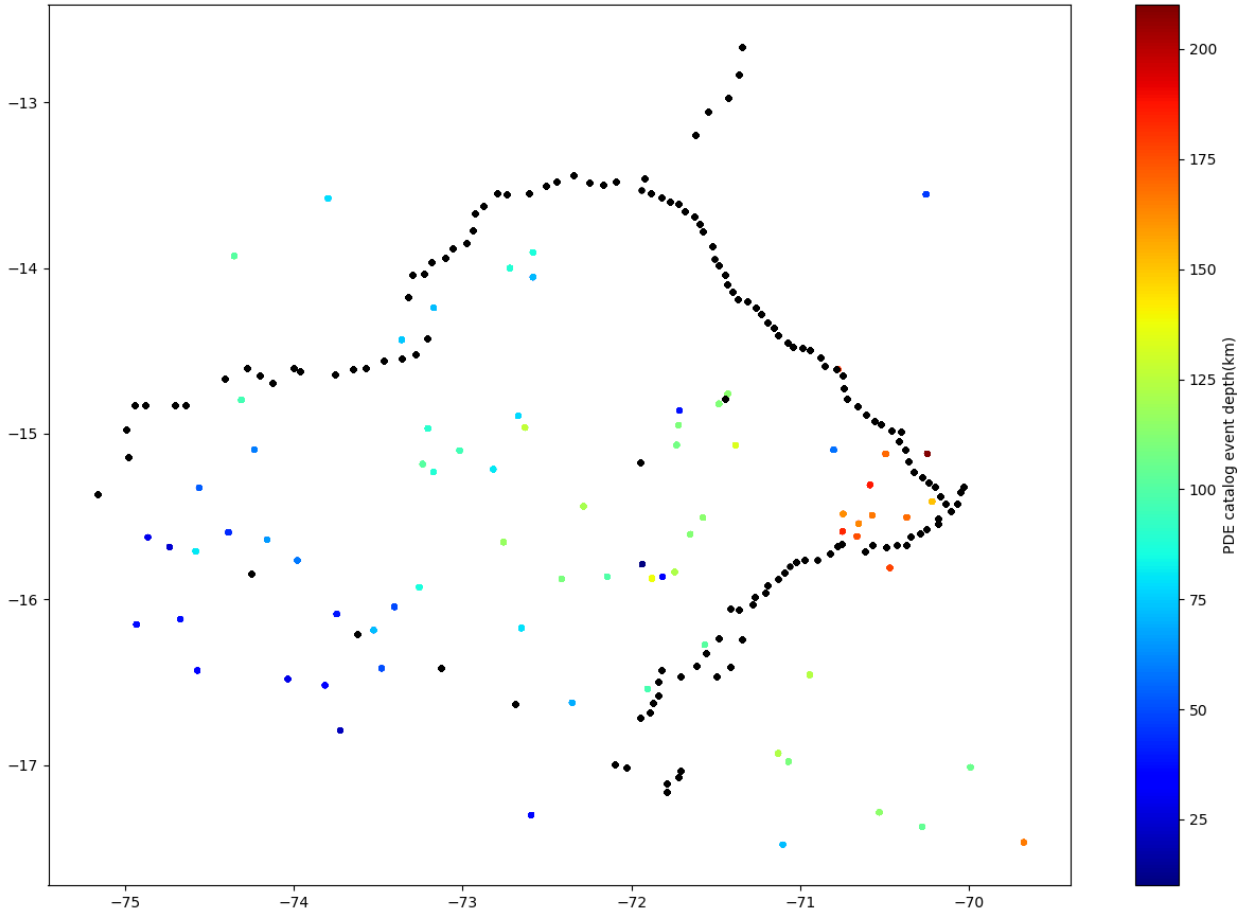


Figure 4: Peru reporting stations(black), USGS PDE catalog event coordinates and starting depths(color).

The primary data set used in the work presented was produced by PeruSE, collected from approximately 100 broadband Guralp CMG-3T seismometers with Quanterra Q330 digitizers placed in a quasi-rectangular pattern in Southern Peruvian Andes between 2007 and 2013 (Figure 1). The experiment was carried out as a collaboration between Dr. Paul Davis et al. at UCLA and Dr. Robert Clayton et al. at Caltech. Recently, some additional data from 8 stations in the ZG and ZD networks in the IRIS data repository was incorporated into our inversions. Data was stored in ".sac" format (Seismic Analysis Code) on a Caltech STP server. The ".sac" file format was developed by IRIS and is shorthand for "Seismic Analysis Code". Python seismic analysis package "Obspy" was used for data import and

pre-processing.

The work presented was developed and evaluated using a custom-built setup utilizing an Intel i7-8700k 6 core CPU at 4.7GHz with 32GB of 3200mhz RAM. CPU inversions were run in a multi-threaded configuration. Massively parallel algorithms were implemented for CPU execution, as well as GPU execution on an 8GB Nvidia GTX-1080 with 2560 physical cores at 1850 MHZ. A second setup utilizing a 32 core AMD TR-2990WX at 4.0 GHZ, 16 GB of ram at 3200 MHZ, and a 1280 core Nvidia GTX1060 at 1850 MHZ was also used. All results incorporating gravitational inversion were run on GPUs concurrently with seismic models on CPUs. Model prototyping, data I/O and plotting was done in Python 3.5 using numpy, scipy, and matplotlib. All performance-critical CPU implementations were done in C++14 compiled on GCC using -O3 optimization. All GPU implementations were in CUDA C and compiled on GCC host side and NVCC device side. No library dependencies outside of STL(Standard Template Library) C++14 were used for CPU implementations. No library dependencies outside the CUDA C API were used in GPU implementations. All matrix and vector operations were done using our own linear algebra and math library.

A short overview of the core software libraries created from scratch and provided in the supplementary materials includes the following libraries:

- zmath - Linear algebra and math library, includes most common math functions and operator overloading for handling vector and matrix types. Generally slower than major distribution but free of non C++ STL dependencies and usable by inclusion of reference in the main source file.

- NNI - Discrete Sibson Interpolation (discrete implementation of a Natural Neighbor Interpolation) library containing functions for 2D and 3D sets of Voronoi nodes, including multi threaded implementations.

- FMM - Fast Marching Method first arrival forward model class and its variants. Includes 1st and 2nd order FMM, as well as our variant utilizing VFD. Can be used with internally initialized point sources, as well as boundary condition initialized sources.

- TFM - User friendly forward model I/O wrapper class, data I/O and top level inversion algorithms.

- SET - Standard Earth library, including standard Earth travel time functions analogous to TauP. Builds and uses SET.dat file to maintain a lookup table for standard Earth travel times.

4 Waveform Pre-processing and Timing

While automated approaches for first arrival timing have been developed, we have refrained from implementing them for use in our inversions, due to the sensitivity of inversions to non-Gaussian pick errors, which are well known to be especially common in phase measurement cross correlation approaches due to cycle skipping. It also allowed us to refrain from picking non-physical arrival times due to instrument clock error on otherwise clean wave forms, which would be picked by a phase measurement or envelope detection algorithm running without a complex weighting meta-algorithm. We found GPS error to be generally non-Gaussian at the larger end, with some waveforms being mistimed by 5-15 seconds. The non-Gaussian distribution of GPS errors, and the difficulty in their detection when they are not significant outliers can be especially detrimental to a least squares (L2) weighted inversion, as the one used, and must be a consideration in the error propagation and analysis of the results.

In the course of inversion algorithm development, it was found that the PeruSE network data set suffered from a systematic offset between two subsets of reporting stations, with a difference of 1s between station clocks. Upon discovery, corrections were applied in a manner maximally consistent with station deployment timing and ownership. The application of corrections improved fit convergence and is believed to have been successful to the author's satisfaction. However, this cannot be guaranteed with a complete certainty.

Initially, pkbaer(), a picking function in Obspy (Baer, 1987) was applied to the PeruSE data set, resulting in arrival time picks for approximately 40 local and 40 teleseismic events. However, the reliability of the picks was found questionable. Subsequently, 83 usable local events were manually timed, of which 77 events were found to be usable for inversion. Automatic timing of teleseismic arrivals was deemed to be prone to error, and at this time, between 37 and 51 manually picked teleseismic arrivals are used in inversions, depending on

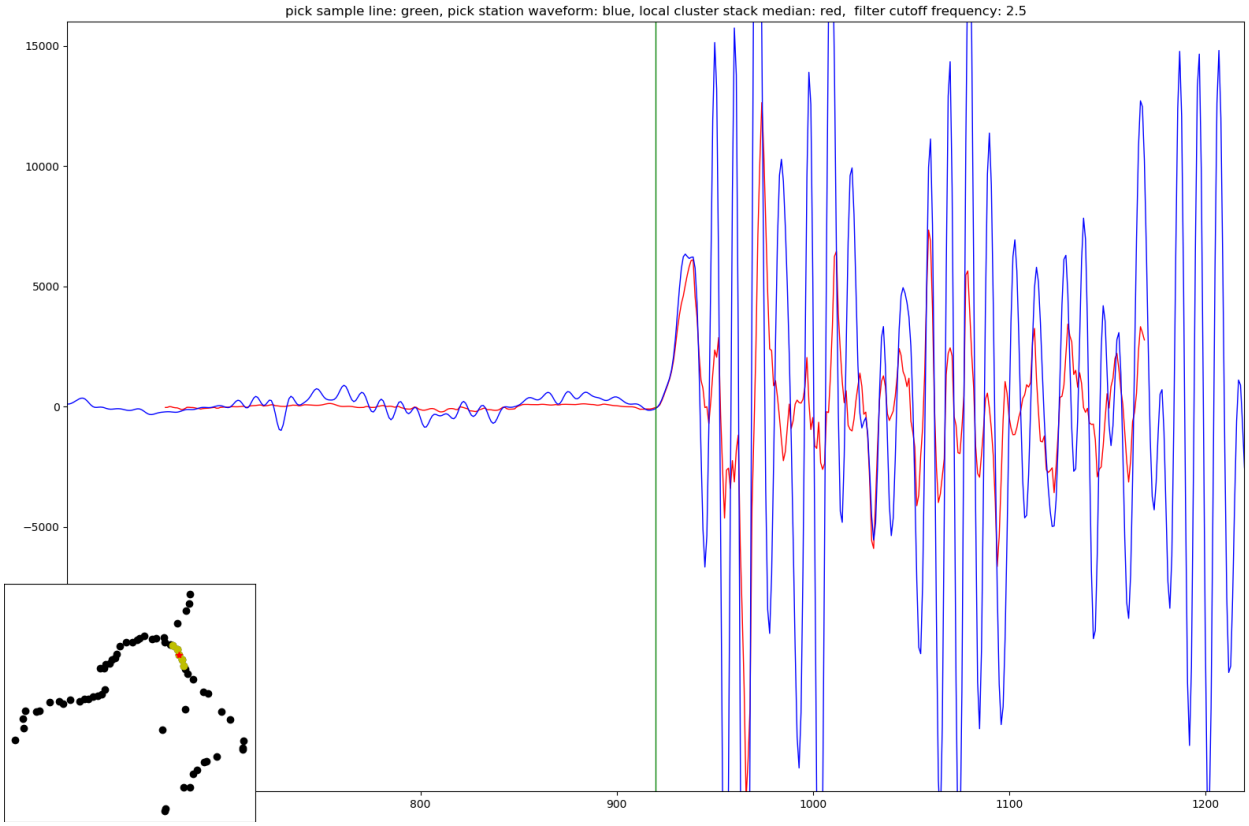


Figure 5: Single station local arrival pick aligned with local cluster wave stack median (20 samples/second).

whether spatial feature emergence restrictions are used or not.

First arrival picks were done by using 20 Hz sample rate wave forms centered displayed between 0 and 100 seconds after catalog event time using the USGS PDE catalog of regional events in Peru, selecting events with preliminary locations within a 300km E-W and 300km N-S box centered at 72.5 W and 15.25 S (Figure 4). Wave forms were plotted and first departures from noise were selected with and without filtering applied (Figure 5). Filter frequencies and pole count varied by event. However, all local arrival picks were made on raw unfiltered wave forms to avoid artifacts.

Teleseismic arrival picks used in the following inversions were made using 20 Hz sample rate wave forms, identifying and picking a waveform feature ruled to be close to the first

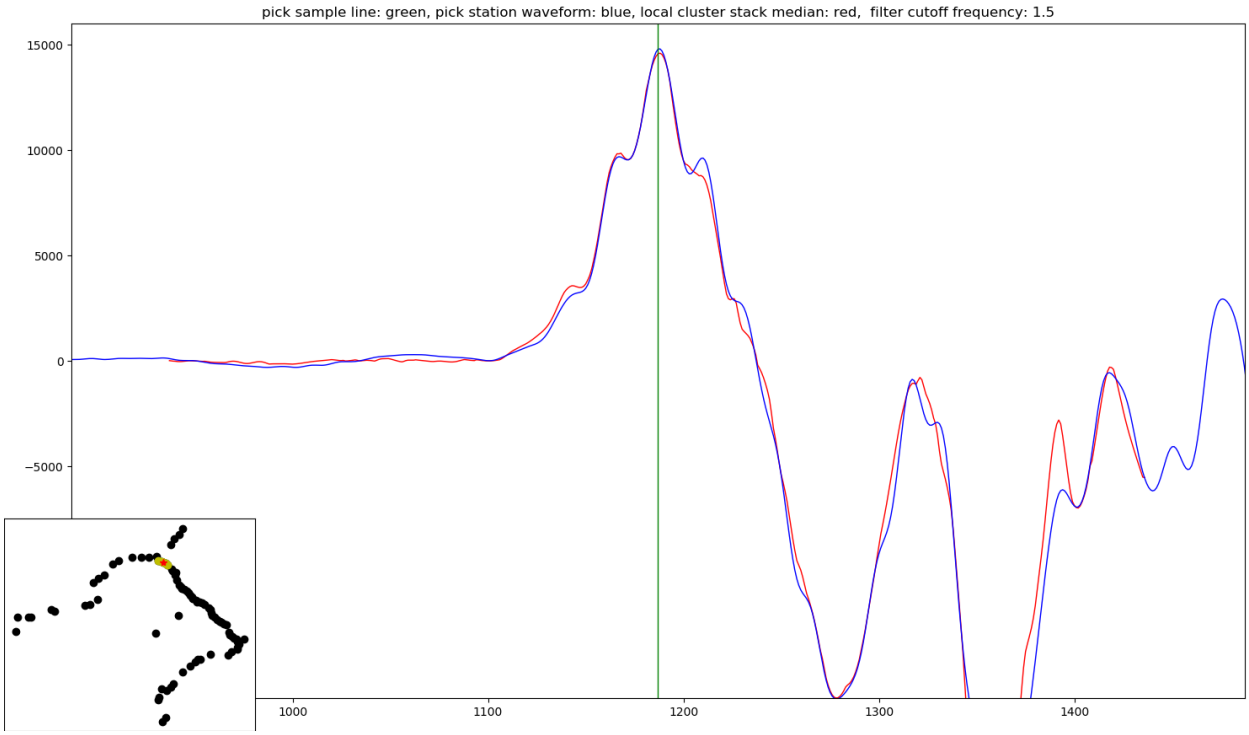


Figure 6: Single station teleseismic arrival pick aligned with local cluster wave stack median (20 samples/second).

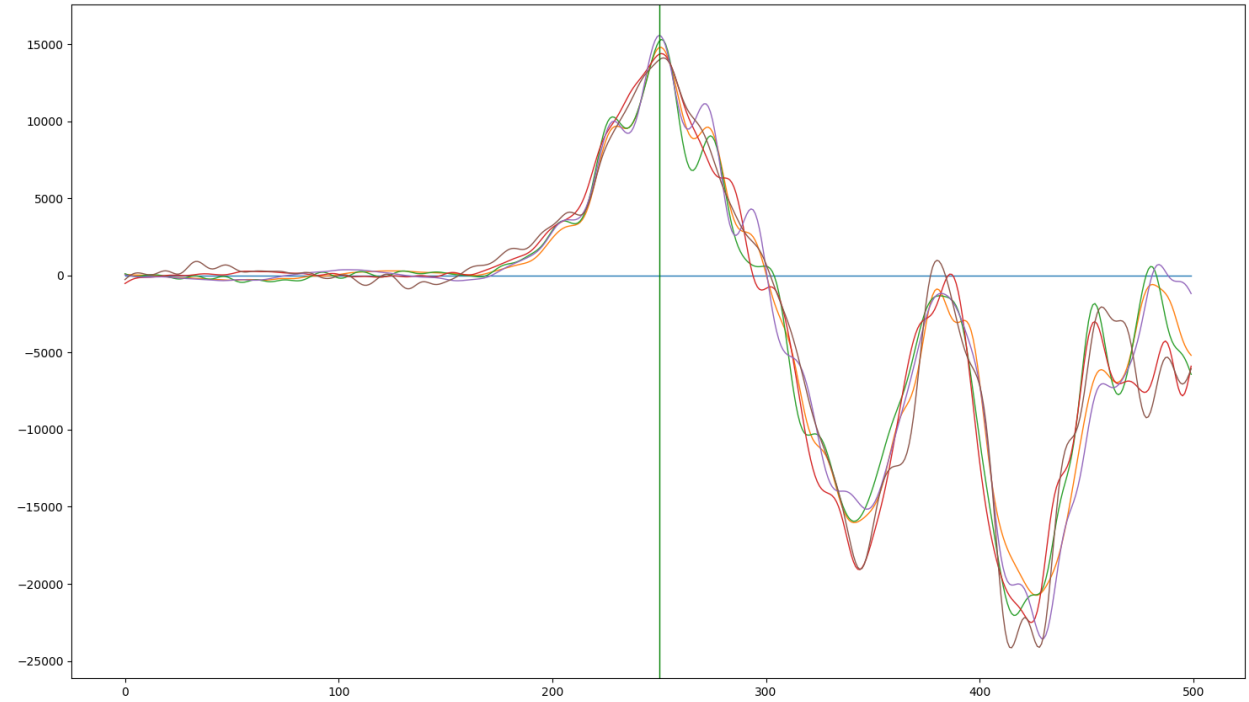


Figure 7: Five station local cluster of aligned teleseismic arrival wave forms (20 samples/second).

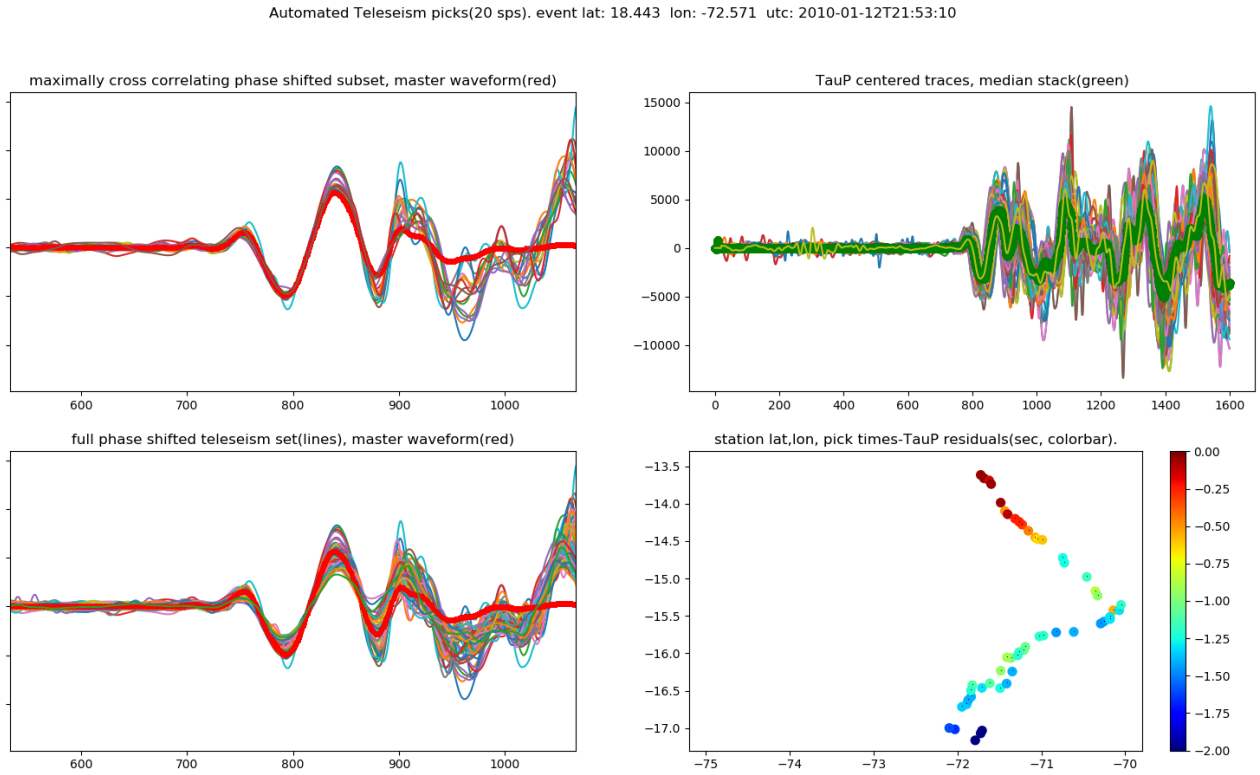


Figure 8: Top left: master waveform and aligned component wave forms. Top right: all traces phase centered at TauP arrival time. Bottom left: Full phase shifted cross correlated trace set. Bottom right: Picked-TauP residuals across early stage PeruSE network

departure from noise, and identifiable across the entire network (not spatially emergent). Waveforms were then stacked in groups of 5, and the relative phase was further adjusted to bring each station in phase with the median waveform constructed from a local station cluster defined by itself and 4 of its nearest neighbors (Figures 6 and 7). Arrivals cataloged outside of the 600km x 600km inversion grid but less than 20 degrees arc to the center of the network were rejected in later inversions due to spatially emergent phases and waveform features, the phase of which could not be reliably measured across all reporting stations.

As our inversion approach is based on least squares minimization, it has L2 norm (quadratic response to error instead of linear) and its performance strongly depends on noise being Gaussian. One persistent source of systematic error in our experiment was mistimed GPS clocks, which as expected, created spatially coherent anomalies in addition to appearing to be dis-

tributed in a manner more consistent with a uniform sampling. Because of this, we placed a priority on careful error rejection. After using spatially smooth location and tomographic inversion, we reviewed each set of waveforms for local monotonicity. When points were found to be more than 1.5 seconds away from monotonicity with neighbors, they were removed. The total loss of data was on the order of 2 percent.

As an automated first arrival measurement technique has not yet been finalized, waveform processing description is limited. The current preliminary automatic waveform picking routine works by event timing, using arrival windows centered around standard earth travel times from the USGS global PDE catalog. Teleseismic picking is then performed by proxy, through measurement of the relative phase of early waveform features across the network using the following algorithm (Figure 8):

1. Pick crude first arrivals using an STA/LTA trigger of the analytic envelope of the waveform (the absolute values of a waveform's Hilbert transform). Among the different approaches tried, using an analytic envelope has been found to be least prone to false positives for single station arrival detection, while reliably showing a signal when one is expected, at the expense of precision, which is lost due to envelope phase dependence on the first arrival signal to noise ratio.

2. Assign Gaussian weight to the waveform time series centered at the pick time to be used for subsequent cross correlation.

3. Cross correlate every event trace with all other traces. Sum the maximal cross correlations with each trace. Alternatively, cross correlate traces across spatially localized clusters.

4. Sort traces by peak cross correlation score sums, select the top quartile to use for creation of a master template waveform.

5. Starting with the highest scoring trace, stack maximally cross correlated phase shifted waveforms in the master waveform subset, reconstructing an averaged clean representation of the wavefront across the network.

6. Cross correlate each reporting trace with master waveform to calculate the phase shift of maximum cross correlation.

7. Convert relative phase to arrival times.

8. Filter out outliers using either peak cross correlation criteria, or using neighborhood median and distributions.

5 Basis Function

Our implementation of the Discrete Sibson Interpolation variant of the Natural Neighbor Interpolation algorithm, which serves as the core of our field parametrization and basis functions, is done in a straightforward manner, following Park et al.'s algorithm described below:

5.1 NNI implementation algorithm

1. Initialize floating sum grid U, contribution counter grid c
2. Iterate through every element $u_n = U[i, j, k]$
3. Determine nearest Voronoi node scalar field value v (Voronoi, 1908), distance to node d
4. Sweep through all u_m in the neighborhood of u_n where $dist(u_n, u_m) < d$. Add v to each neighborhood u_m: $u_m = u_m + v$; increment $c_m = c_m + 1$
5. Divide each element u_m by c_m, return normalized interpolation U

Our 3D NNI implementation has been implemented on both CPU and GPU, with significantly reduced computation time rendering in GPU. Actual render time is highly specific to not only the grid and node count, but also to the node positioning, with fastest interpolations occurring in uniformly distributed Voronoi node sets. If velocity rendering is determined to be a significant contribution to model run time in full 3D inversions, an NNI perturbation scheme under development may be used for Jacobian and incrementally perturbed forward model computations.

Velocity field parametrization was chosen with consideration of the type of perturbations and features expected in mantle tomography. We have thus far prioritized resolution of at least one strong discontinuity, such as a basin boundary or the Moho discontinuity, and

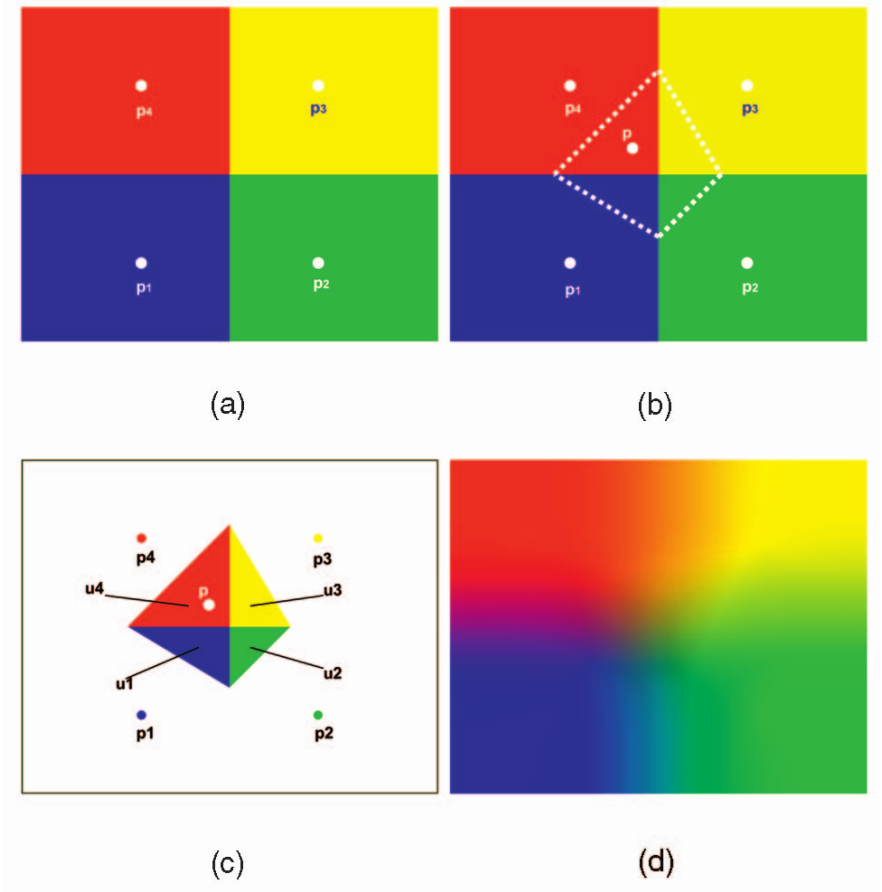


Figure 9: (a) Define Voronoi diagram of field nodes. (b) Define sample p, compute Voronoi cell (c) Compute area fractions of u[i] in u. (d) Weighted avg=field value at p. (Park,2006)

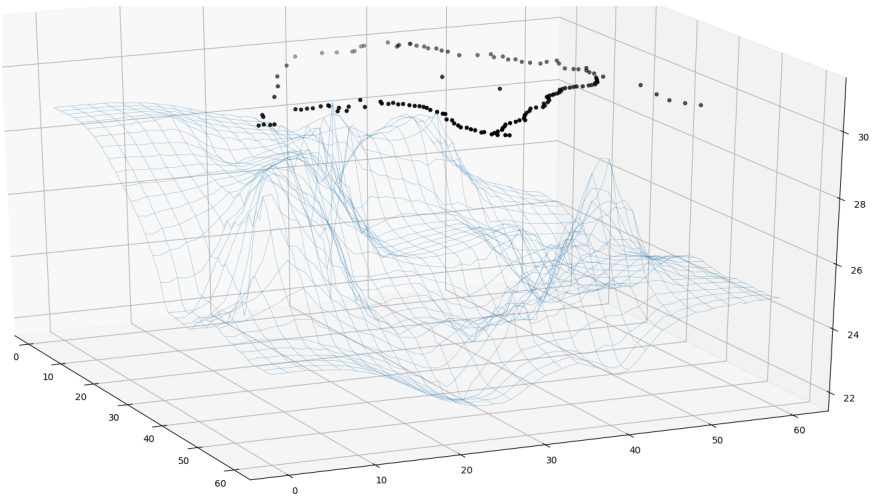


Figure 10: Sample 2D NNI wire-mesh plot using 79 floating Voronoi nodes mapping the Moho surface/manifold (PeruSE network stations in black for reference)

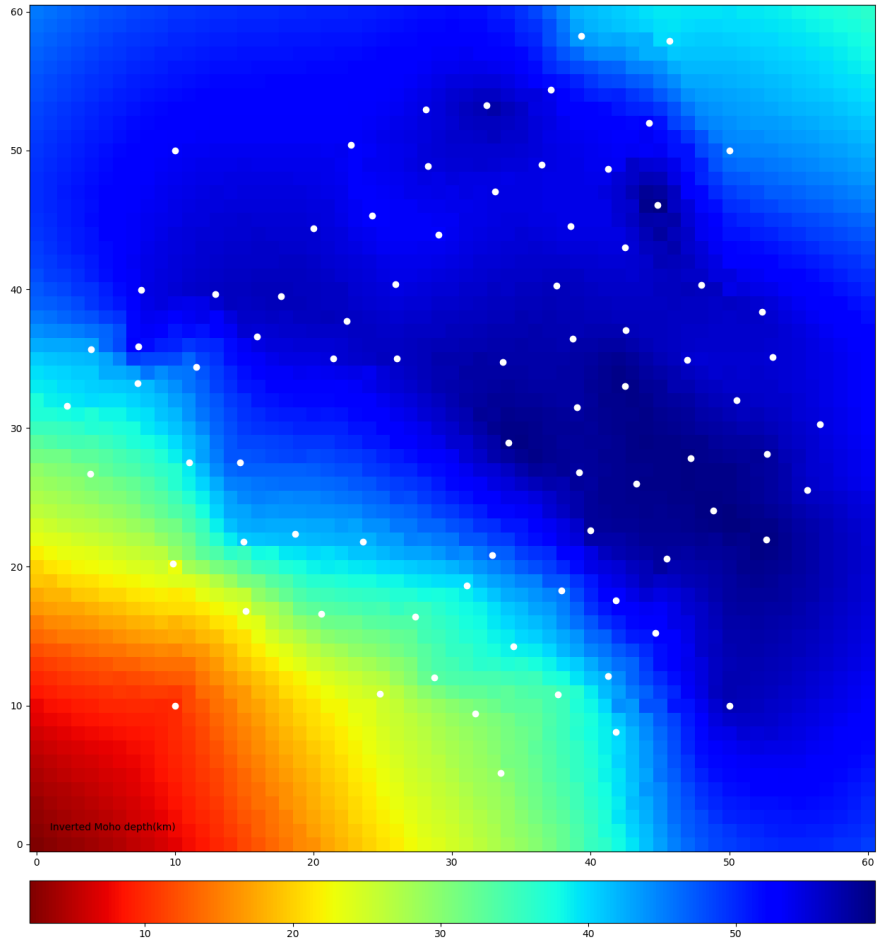


Figure 11: Sample 2D NNI color-map plot using 79 floating Voronoi nodes representing the Moho surface. Node locations in white

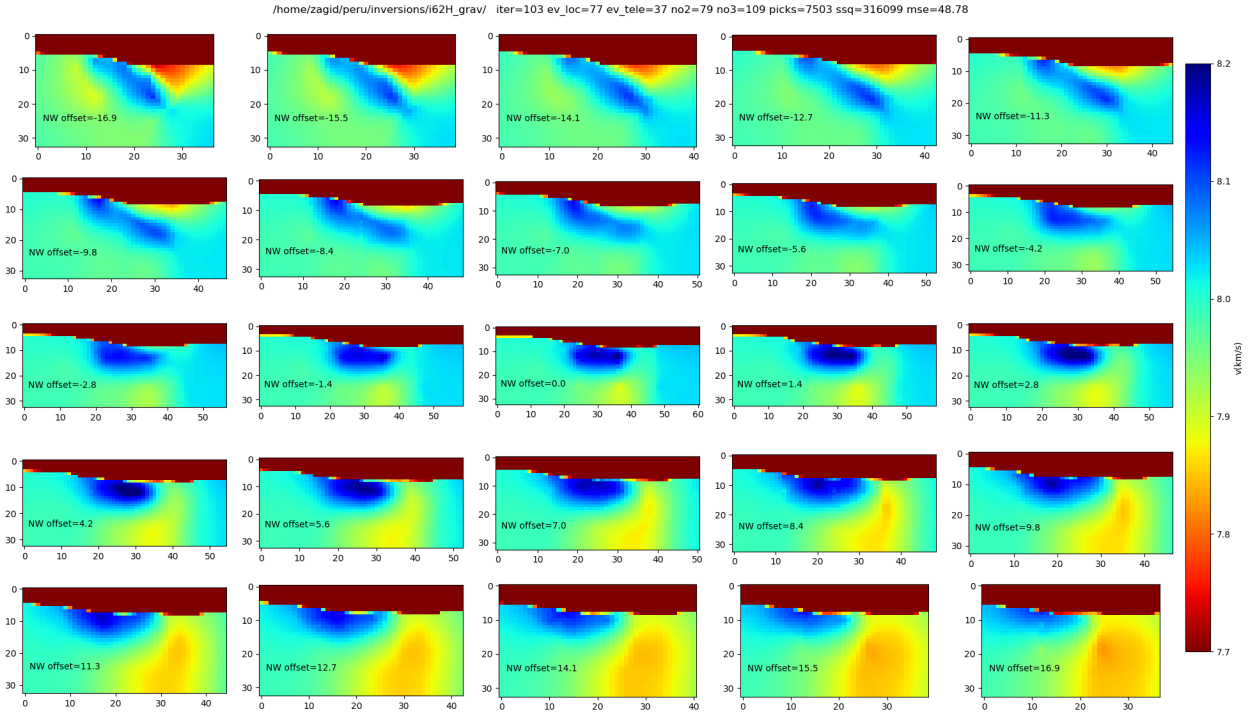


Figure 12: Side progression through sample 2 layer rendering of 79 floating 2D NNI Voronoi nodes representing the Moho manifold, and 109 floating 3D NNI Voronoi nodes representing a sample mantle p-wave velocity perturbation(spatial units are in grid scale, with 10km point spacing).

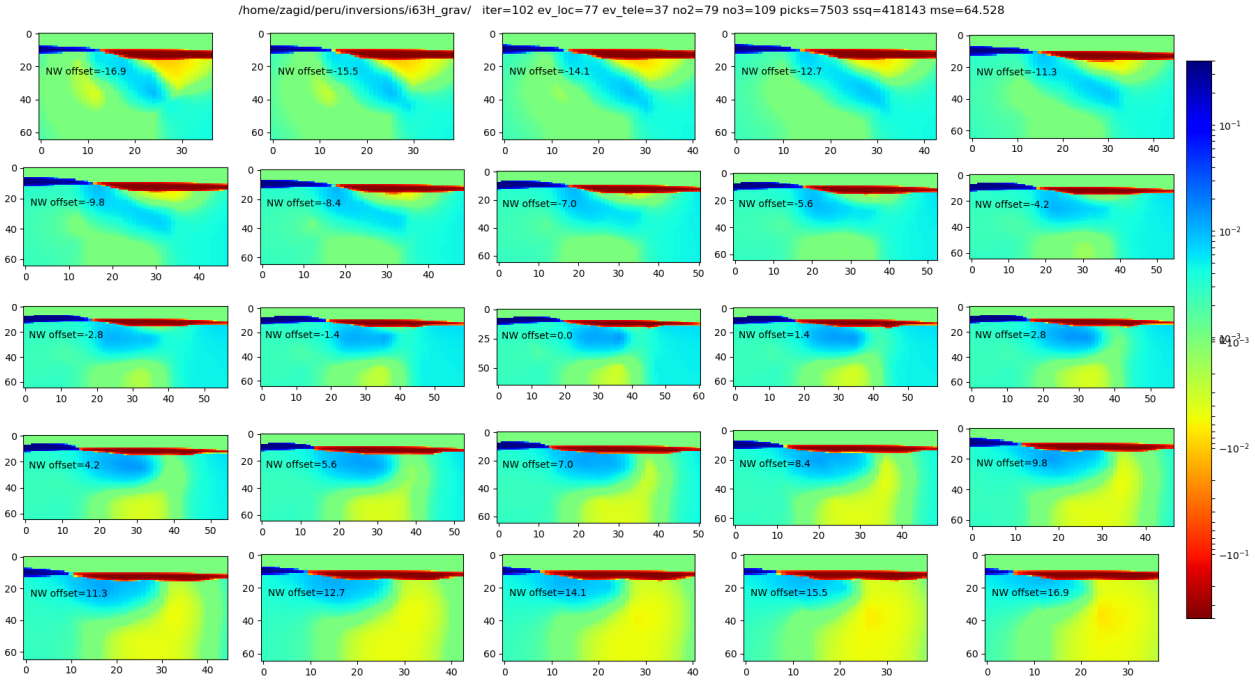


Figure 13: Side progression through sample 2 layer rendering of 79 floating 2D NNI Voronoi nodes representing the Moho manifold, and 109 floating 3D NNI Voronoi nodes representing a sample full density perturbation field versus standard earth density distribution(spatial units are in grid scale, with 10km point spacing, color bar scale is in grams/cc.)

smoother continuous 3D perturbations. To resolve these features, we chose a basis function to consist of a blended 2 layer field model, with a uniform upper layer model that anticipates poor lateral resolvability and an invertible heterogeneous lower layer described using a 3D Discrete Sibson/Natural Neighbor Interpolation (from hereon referred to interchangeably as DSI or NNI). The interface between the layers is mapped using a 2D DSI manifold. 2D DSI Voronoi nodes are each parametrized using 2 floating point spatial coordinates and 1 floating point scalar field coordinate (which in our application maps a 3rd floating point spatial coordinate). 2D DSI of a set of 4 Voronoi nodes is illustrated in (Figure 9).

The 3D DSI implementation Voronoi nodes are parametrized using 3 floating point spatial coordinates and 1 floating point scalar field strength, which in our implementations represents either a p-wave velocity field (dimensionality km/s), or density perturbation field (dimensionality g/cm^3). An important trait of DSI (discussed in more detail) below is its inherent discretization, or roughness, which must be considered when computing Jacobian elements corresponding to coordinate parameters(but not field parameters), as well as inversion damping and positional resolvability. Our spatial coordinates are represented in ENU (East North Up) Cartesian coordinates indexed as $\langle i, j, k \rangle$ We have also implemented a single layer heterogeneous velocity field parametrization using only 3D DSI; however ignoring the prior information about the existence of a discontinuous Moho significantly degrades the model resolvability of the discontinuity. The blended 2 layer basis function can be expressed as:

$$v(i,j,k) = \begin{cases} \beta_{crust} & I_2(\beta_{o2}, i, j) < k - h \\ I_3(\beta_{o3}, i, j, k) + (k - I_2(\beta_{o2}, i, j) + h)\frac{\beta_{crust} - I_3(\beta_{o3}, i, j, k)}{2h} & |I_2(\beta_{o2}, i, j) - k| < h \\ I_3(\beta_{o3}, i, j, k) & k + h < I_2(\beta_{o2}, i, j) \end{cases} \quad (6)$$

Where $I_{on}()$ is the floating linear interpolation of the DSI of a Voronoi node set in n dimensions, described by β_{o2} in 2D and by by β_{o3} in 3D. Each 2D and 3D Voronoi node is described by: $[i, j, f]$ and $[i, j, k, f]$, respectively, with i, j, k defining floating coordinates, and f defining a floating scalar field amplitude. In seismic and gravitational reference models, an unperturbed Moho depth is mapped conformally to the curved (in Cartesian coordinates) Earth surface at the kth coordinate over a grid in i, j to correspond to $d_0 = 35km$ as per iasp91. A sample 2D NNI manifold perturbation field formed by 79 floating Voronoi nodes mapped by an inversion is presented in Figures 10 and 11. A sample full velocity field formed by 79 floating β_{o2} Voronoi nodes defining the Moho and 109 floating β_{o3} is presented in a series of progressing cross section slices in Figure 12.

Density field parametrization is done using the same spatial parameter set as P-wave velocity field parametrization, differing principally by being rendered as a difference field using a background and perturbation model. The background model is a radially symmetric density field with an invertible parameter describing the density discontinuity -ρ_μ across a standard Earth Moho, generally understood to be in the range between 0.35 - 0.5 g/cm^3 with the density field being generally uniform otherwise. The perturbation model has an assigned invertible background mantle reference velocity v_m, used only for calculating density perturbation field, and a uniform crustal density always set equal to the reference. The mantle perturbation field is calculated by using a scaled difference between the perturbed mantle P velocity field and background velocity field, with an invertible scale factor λ_ρ which depending on the source can be calculated to be in a range between 0.05 and 0.37 $\frac{g/cm^3}{km/s}$ (Brocher, 2005)(Sato, 1989) centered at 8 km/s. The explicit piece-wise continuous representation of the density field can be described by taking the difference field:
$\delta\rho = \rho_p - \rho_0$ between a perturbed field ρ_p and background field ρ_0 computed using the following piecewise functions:

$$\rho_0(i,j,k) = \begin{cases} \rho_\mu & k_{M0}(i,j) < k - h \\ \rho_\mu + (k - k_{M0}(i,j) + h)\frac{\rho_\mu}{2h} & |k_{M0}(i,j,k) - k| < h \\ 0 & k + h < k_{M0}(i,j) \end{cases} \tag{7}$$

$$\rho_p(i,j,k) = \begin{cases} \rho_\mu & I_2(\beta_{o2},i,j) < k - h \\ \rho_\mu + (k - I_2(\beta_{o2},i,j) + h)\frac{\lambda_\rho(I_3(\beta_{o3},i,j,k) - v_m) + \rho_\mu}{2h} & |I_2(\beta_{o2},i,j) - k| < h \\ \lambda\rho(I_3(\beta_{o3},i,j,k) - v_m) & k + h < I_2(\beta_{o2},i,j) \end{cases} \tag{8}$$

The relaxation parameters that are allowed to be free floating in joint inversions allow compensation for offset errors in the gravity perturbation data, as well as trading off Moho originated signal against mantle perturbation originated signal. A sample density perturbation field from an inversion used can be seen in Figure 13. In cases where the mass grid points are not indistinguishable from velocity grid points, linear interpolation is used to compute their field position.

6 Forward Models

6.1 Model Error Constraints

A plurality of time spent on research has focused on finding and implementing a fast and robust invertible first arrival travel time forward model. All forward models discussed in this work (with the exception of Standard Earth Traveltime (SET) model and Obspy's TauP) are implemented in three dimensions. As previously mentioned, the two key metrics in an invertible forward model are model robustness and model computational efficiency. We define computational efficiency as a measure of a model's measured run time in relation to the defined acceptable model error bounds. Robustness will be defined in this context as the smooth and continuous response of model outputs and model error to perturbations of its inputs. A basic analysis of error propagation of measurement uncertainty and model uncertainty yields an output uncertainty relation:

$$\sigma^2 = \sqrt{\sigma^2_{measurement} + \sigma^2_{model}} \tag{9}$$

Given this propagation relation, as either measurement or model uncertainty approaches zero, the propagated uncertainty asymptotically approaches the value of the non-zero contribution. An obvious consequence of this is the marginally diminishing utility of reduction of an uncertainty that is significantly smaller than the larger of the pair. In the case of seismic tomography, a fundamental limitation exists on minimum uncertainty bounds of a measurement that is determined by the signal quality. In the types of signals available for mantle tomography and at the dominant frequencies and SNR seen in the types of events used (3-10 Hz incoherent wave forms for local events, 0.1-1 Hz coherent wave forms for teleseismic events), a commonly accepted measurement uncertainty ranges from 0.1 to 0.2 seconds. As computational cost will be demonstrated to grow inversely exponentially in relation to model error tolerance, a reasonable target model uncertainty should be on the order of, or slightly

smaller than the measurement uncertainty. Model travel times for the area imaged under the PeruSE network anticipates up to 60-80 seconds of travel through perturbed velocity fields. Given these constraints, a target fractional uncertainty for mean standard error of the first arrival forward model can be reasonably set to be:

$$\sigma_{model} < \frac{\sigma_{measurement}}{max(T_{model})} \tag{10}$$

Which guided our decision to set a mean seismic model error target no greater than 0.1-0.2%, translating to an absolute mean error of no more than 100ms.

Class	Model(Abbreviation)	Description
Ray	Ray Shot Inversion (RSI)	Inversion of ray shooting angles
Ray	Isotropic Ray Shooting (IRS)	Isotropic ray shooting, receiver interpolation
Ray	Ray Bending Optimization (RBO)	Spline travel time minimization
Huygens	Shortest Path Methods (SPM)	Grid search of shortest point to point travel times
Huygens	Particle Shower Cascade (PSC)	Non-regularized iterative particle cascade mapping
Wave Eq	Acoustic Wave Equation (AWE)	Finite difference solution of $\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$
Wave Eq +Eikonal	Time Initialized Acoustic Wave Equation(TI-AWE)	FMM initialized sparse solution of AWE
Eikonal	Vidale Finite Difference (VFD)	Downwind plane sorted Eikonal solver
Eikonal	Fast Marching Method (FMM)	Upwind level set method Eikonal solver
Eikonal	Eikonal Hybrid (FMM-VFD)	Upwind level set method Fast Eikonal solver

Table 1: Forward model types implemented and tested.

6.2 Seismic Forward Model Survey

Outlined in Table 1 is a summary of seismic forward models that we implemented from scratch in STL C++ and/or CUDA C and evaluated for performance at parameters meeting target model error of 0.1%.

6.3 Seismic Model Summary

- Ray class models (RSI, IRS, RBO) were highly computationally efficient in homogenous media and run well in parallel (CPU and GPU), but are prone to major and unpredictable failures, including but not limited to failure to find diffracted phase first arrivals (RSI, IRS), and selection of non-minimal travel time arrivals due to multipathing (RSI, IRS, RBO). As the heterogeneity of the velocity field grew, instabilities

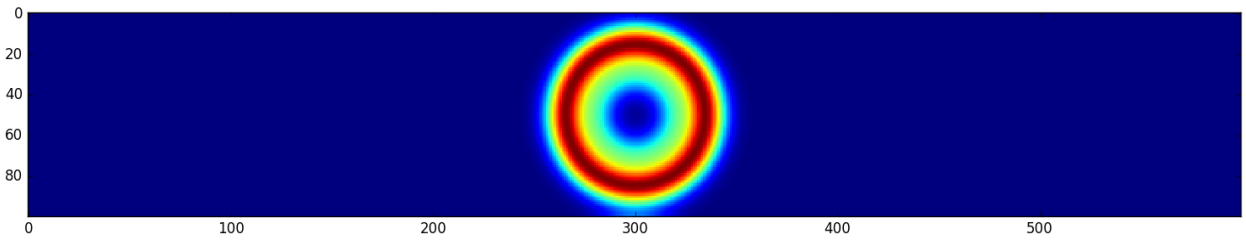


Figure 14: GPU implemented AWE prototype propagation.

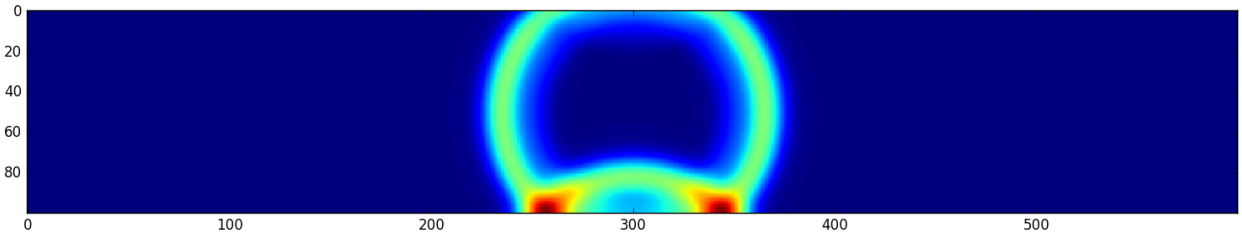


Figure 15: GPU implemented AWE prototype propagation.

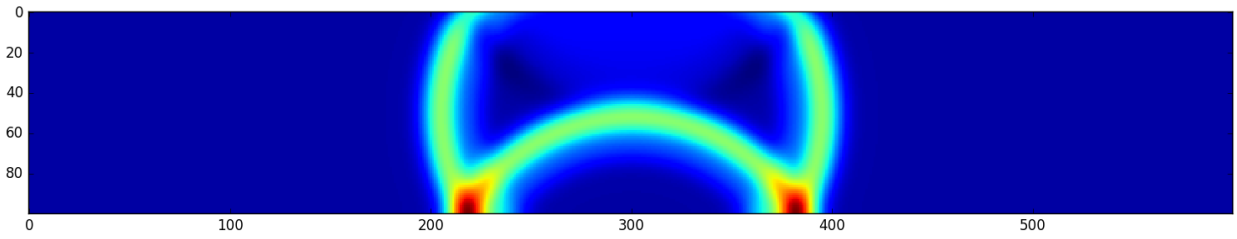


Figure 16: GPU implemented AWE prototype propagation.

became more pronounced.

Computational complexity: RSI=IRS=RBO=$O(n^3)$.

- Hyugens class models (SPM,PSC) were found to be inherently robust as expected, and also easily run in parallel(CPU,GPU). They were found to be among the slowest performing of first arrival models when configured to perform within target model error. Computational complexity: SPM=PSC=$O(n^3)$.

- Wave Equation Solver class models (AWE,TI-AWE) were found to be robust in smooth and continuous media, but broke down in regions of discontinuities, requiring fine grid sizes in order to perform to specification. AWE models run well in parallel execution and were implemented in both CPU and GPU versions. TI-AWE is a sparse implementation of the acoustic wave equation solver for modeling first arrival wavefront amplitudes and with the rate limiting computation running at $O(n^3 log(n))$. This allows amplitude modeling to be done significantly faster than through the use of a dense full wave equation solver. TI-AWE Algorithm description (implementation may vary):

  1. Compute grid first arrival times using a first arrival model such as FMM
  2. Sort each grid point by computed first arrival time
  3. Generate 4D grid of wave potentials u (scalar wave potential defined in 3 spatial dimensions + 1 time dimension) for the wave equation solution, designate as "wave grid"), running in uniform time increments
  4. Iterate along time axis of wave grid, where each wave grid point's time is t_w, first arrival time is t_p, and source cutoff frequency is ν, deleting all grid points where $t_w < t_p$ or $t_w > t_p + \frac{2}{\nu}$
  5. Solve the acoustic wave equation: $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{c(x,y,z)^2}\frac{\partial^2 u}{\partial t^2}$ at each remaining point in u, progressing along the time axis

  This approach is suited for parallel/GPU execution and can be sped up by sorting the

first arrival grid by time and constructing a sparse u from scratch by setting each time iteration of u to the subset of first arrival grid points satisfying the conditions in 4. Computational complexity: AWE=$O(n^4)$, TI-AWE=$O(n^3 log(n))$.

- Eikonal class models (VFD, FMM) had varied characteristics. While VFD was found to be efficient enough to be competitive with ray class models, surpassing them in increasingly heterogenous media, it has known failure modes in media with velocity contrast $\frac{v_{max}}{v_{min}} > 2$, due to imaginary terms in square roots. In contrast, FMM runtime has been consistently measured to be approximately 700-800% of VFD at the same grid size, with significantly higher model errors, to be discussed below. However, FMM was proven by Sethian to be resistant to failure modes seen in VFD and is one of the most robust published first arrival algorithms. Both models are algorithmically serial and cannot be structured to run on large numbers of parallel threads. To attempt to address this while still using the higher theoretical computing power of a GPU, Jacobians utilizing large numbers of VFD instances were run in parallel with each complete instance of VFD running on a separate cuda thread. This solution proved to be unsuccessful, with the large independent memory spaces of each thread causing what can best be described as memory crashes, overwhelming the memory management system of the GPU and either slowing down the rate of execution to a rate significantly slower than that of a CPU, or performing a device memory and process dump. Therefore, it has been concluded that there is no straightforward GPU implementation of VFD or FMM. Gilbert et al. have published Eikonal methods vaguely related to FMM, which have some compatibility with parallel execution, but they do not appear to run significantly faster than CPU implementations of FMM.

  Computational complexity: VFD=FMM=$O(n^3 log(n))$.

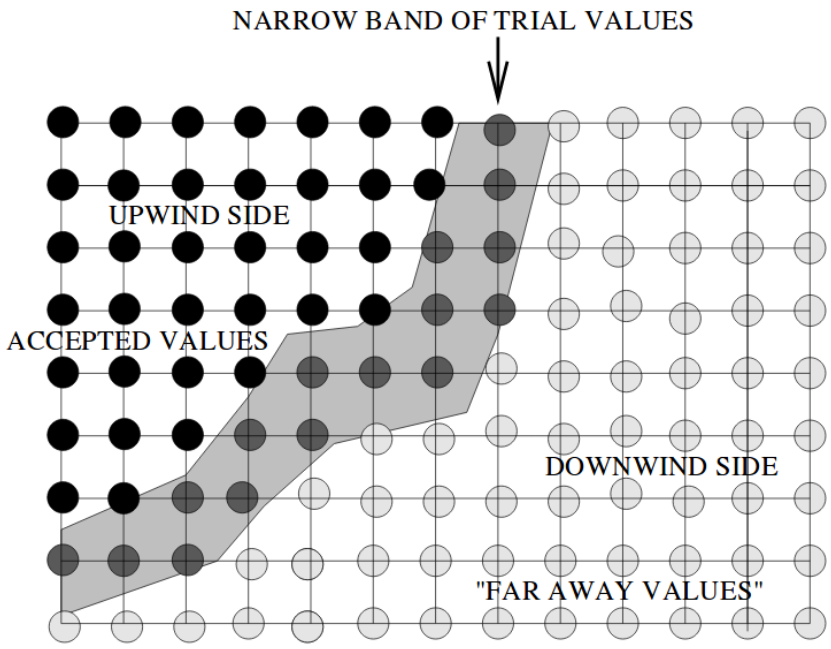


Figure 17: Grid structure in FMM: Sethian, 1997.

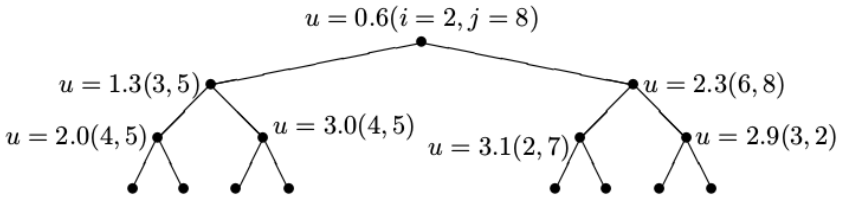
6.4 Seismic Model Selection

Upon consideration of the advantages and disadvantages of each class and model, it was decided to use the Fast Marching Method as the core forward model in our inversion algorithm, for its combination of acceptable efficiency and guaranteed robustness. During the course of its implementation, a variant of FMM was made to incorporate a key component of VFD, significantly reducing model error of FMM to levels comparable to VFD, without a measurable increase in runtime. The following is a pseudocode description of our implementation of FMM-VFD for bounded point sources.

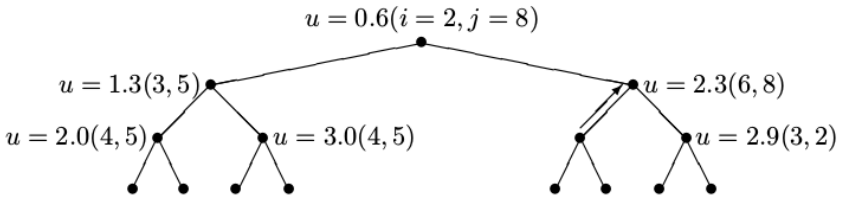
6.5 Forward Model Algorithm Implementation

Fast Marching Method algorithm:

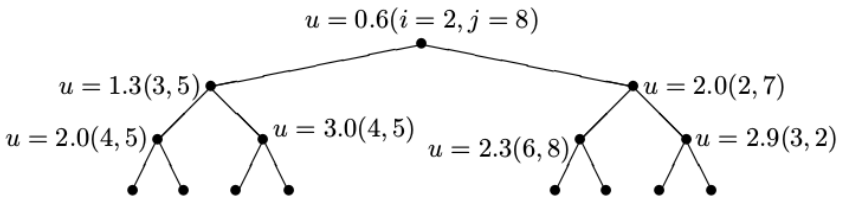
1. Generate a regularized Cartesian floating point grid U for holding first arrival times 't' at each grid point and a corresponding state grid containing the state of each



Step 1: Change u value at (2,7)



Step 2: New value at (2,7); UpHeap



Heap property restored

Figure 18: Update procedure in a binary minheap: Sethian, 1997.

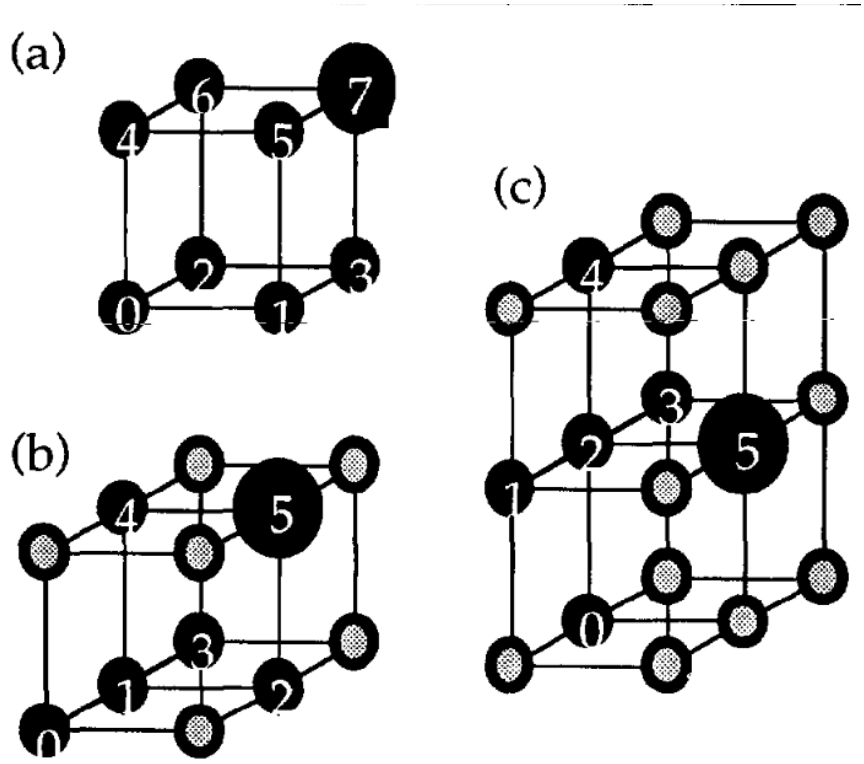


Figure 19: Travel time computation schemes: Vidale, 1990.

grid point. The possible states are: ['far'(unvisited),'close'(visited but not accepted), and 'accepted'(finalized and not subject to revision)]. At time of generation, all non-bounding points are set to state 'far' (Figure 17). Bounding points are initialized at state 'accepted' and assigned a time value much larger than largest possible model time $t_{bound} > 2\frac{Dim(axis_{max})}{c_{max}}$ This is done to prevent the need for bounds checking, and allow faster execution of the marching.

2. Initialize a binary minheap. Binary minheaps are binary trees satisfying a requirement, where the value indexed by each node is either greater than or equal to the values in that node's children(Figure 18). The exact implementation of heap indexing and update methods used is described in Sedgewick, Wayne p315-319. We have also implemented a 2 way linked list in order to maintain 2 way indexing of grid point position/state and heap position/state.

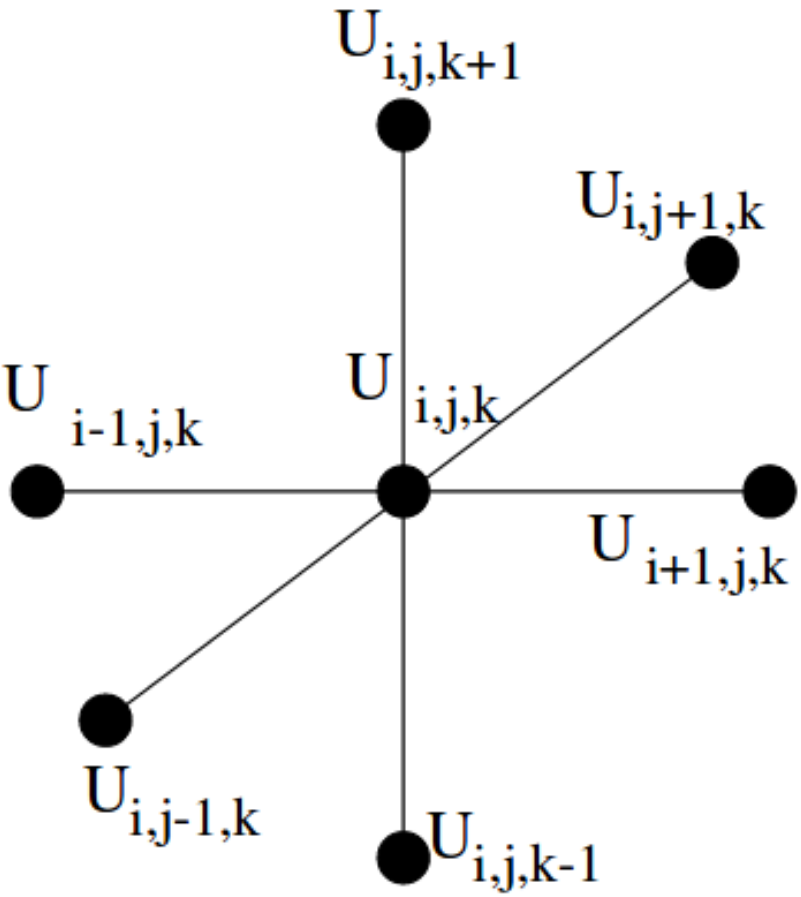


Figure 20: Upwind Eikonal solver scheme: Sethian, 1997.

3. Initialize points within 3-5 grid points of the source coordinates, computing direct line travel time using averaged end point grid slowness (1/v), setting their state to 'close', insert in to bottom of heap and bubble it up until the heap condition is restored.

4. Launch marching algorithm, looping until all points in grid have been converted to "accepted" status:

   (a) Designate top point in heap ("close" u with min time) as "min"

   (b) Apply Vidale Diagonal Recomputation (VDR) to min

   (c) Change status of min to "accepted"; remove from heap

   (d) Iterate through each neighbor of min:

   - If status is "far", set to "close" and add to heap
   - If status is "close", apply Upwind Eikonal Scheme (UES); update heap if u is modified

5. Resultant values in U are the first arrival travel times at each point.

Vidale Diagonal Recomputation (VDR) procedure:

1. On each axis, select neighboring node with the lower time.

2. Sort selected neighbor times as a,b,c where $a < b < c$.

3. Compute mean unit time step using configuration enumerated in Figure 5, scheme a: $f = h * \frac{u_0+u_1+u_2+u_3+u_4+u_5+u_6+u_7}{8}$ where $u = \frac{1}{velocity}$ and h=grid spacing in distance units of u.

4. If all 7 neighbors on the upwind axes and diagonals, as enumerated in "scheme a" in Figure 5 are "accepted", with test point $t = t_7$:

   - Set $t = min(t, t_0 + \frac{\sqrt{6h^2s^2-(t_1-t_2)^2-(t_2-t_3)^2-(t_4-t_1)^2-(t_3-t_5)^2-(t_5-t_6)^2-(t_6-t_3)^2}}{\sqrt{2}})$

5. Else, if 5 neighbors on the upwind axes and diagonals, as enumerated in "scheme b" in Figure 5 are "accepted", with test point $t = t_5$:

   - Set $t = min(t, t_1 + 2\sqrt{2h^2s^2 - \frac{1}{2}(t_0 - t_3)^2 - (t_2 - t_4)^2})$

Upwind first order Eikonal Scheme:

1. On each axis (Figure 6), select neighboring node with the lower time.

2. Sort selected neighbor times as a,b,c where $a < b < c$

3. Compute mean unit time $f = h * \frac{u+u_a+u_b+u_c}{4}$ where $u = \frac{1}{velocity}$ and h=grid spacing in distance units of u.

4. If $a + f > c$:

   - Set $t = min(t, \frac{1}{3}(a + b + c + \sqrt{3f^2 - 2(a^2 + b^2 + c^2 - ab - bc - ca)}))$

5. Else if $a + f > b$:

   - Set $t = min(t, \frac{1}{2}(a + b + \sqrt{2f^2 - 2(a - b)^2}))$

6. Else:

   - Set $t = min(t, a + f)$

6.6 Seismic Forward Model Testing

The described forward model was implemented and tested against a pure FMM implementation using 1st and 2nd order Eikonal solvers through a uniform medium, and % versus analytic travel times was calculated and plotted in Figures 21 - 24. Figure 25 shows % error of VFD through a 2 layer medium with headwave first arrivals, demonstrating model failure, whereas this is contrasted in Figure 26, with FMM-VFD yielding a low and manageable error through the same medium.

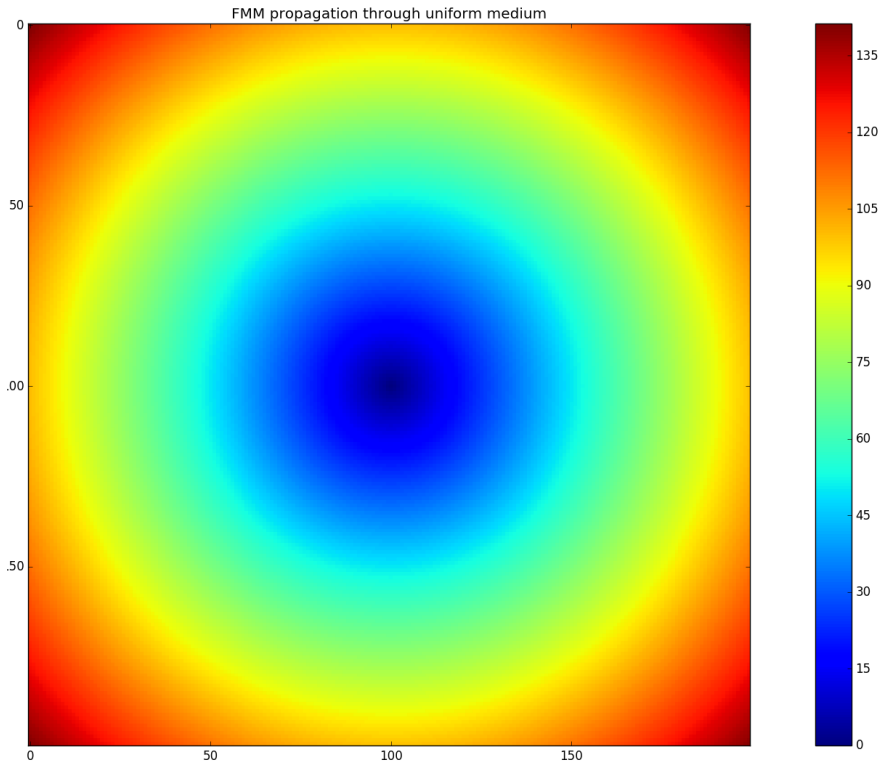


Figure 21: FMM travel times through uniform medium.

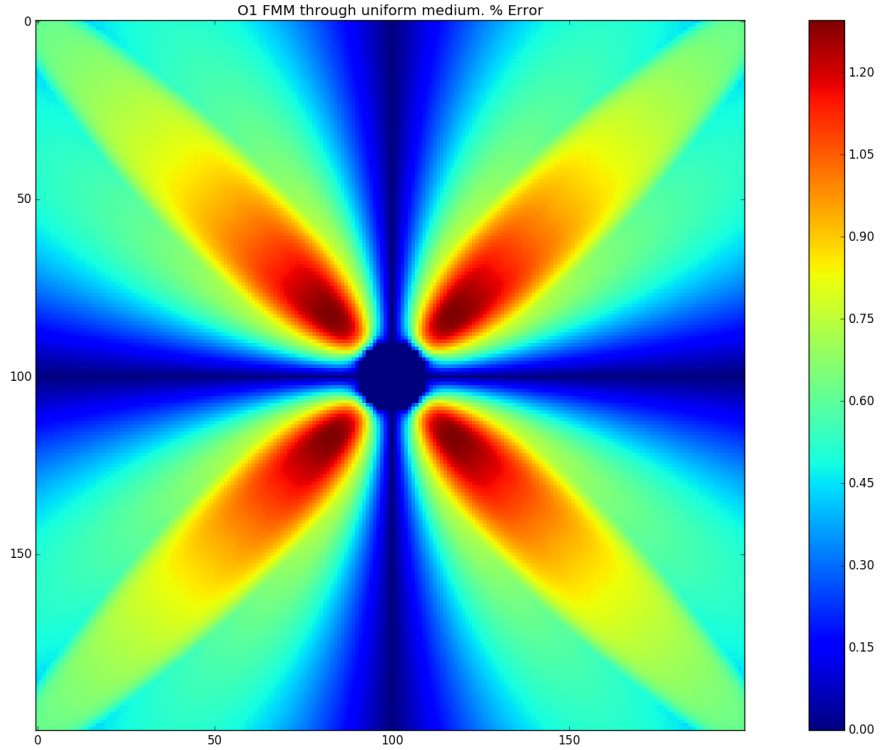


Figure 22: 1st order Eikonal FMM % error through uniform medium.

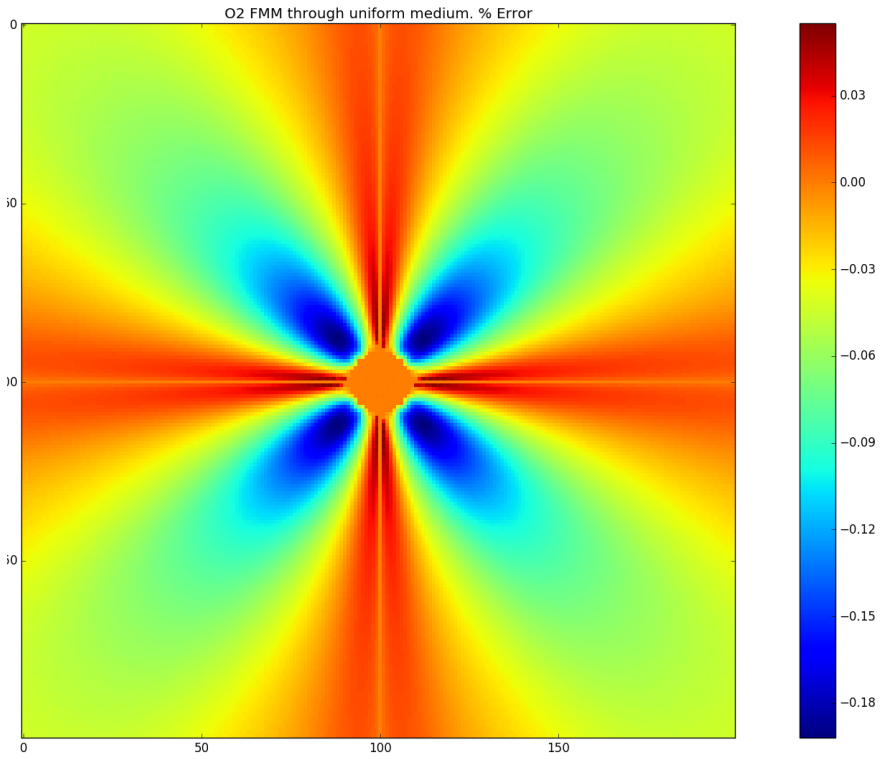


Figure 23: 2nd order Eikonal FMM % error through uniform medium.

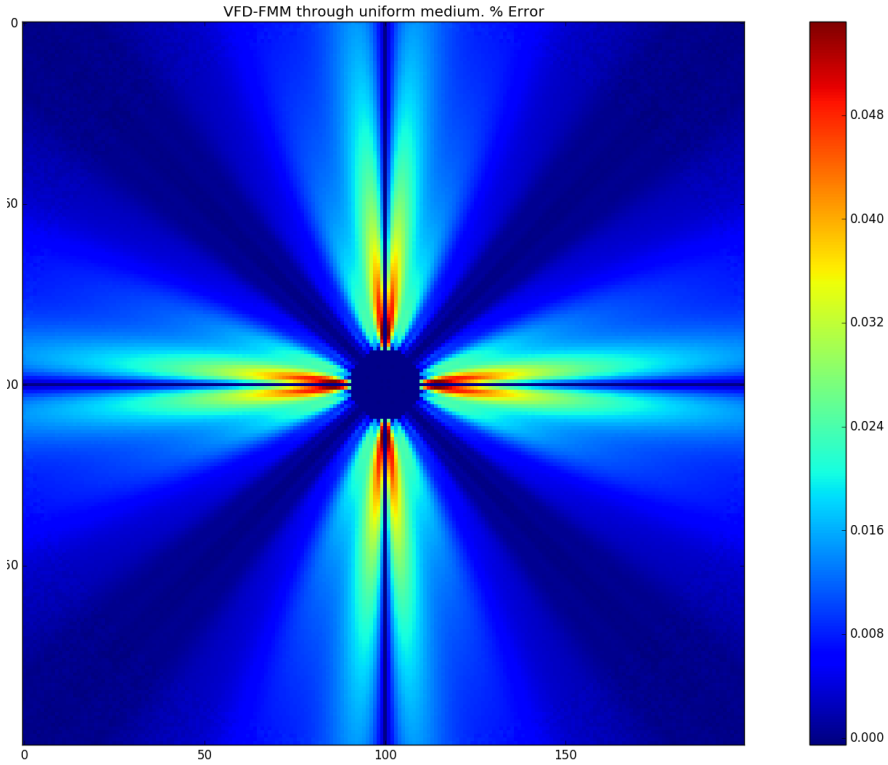


Figure 24: FMM-VFD hybrid forward model % error through uniform medium.

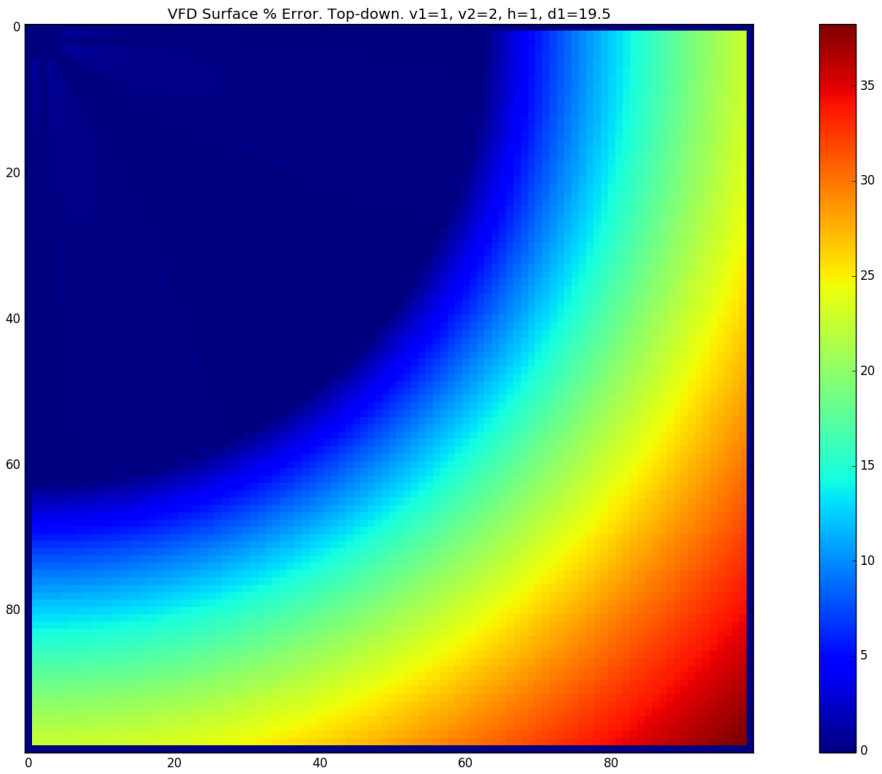


Figure 25: VFD forward model % error through 2 layer medium.

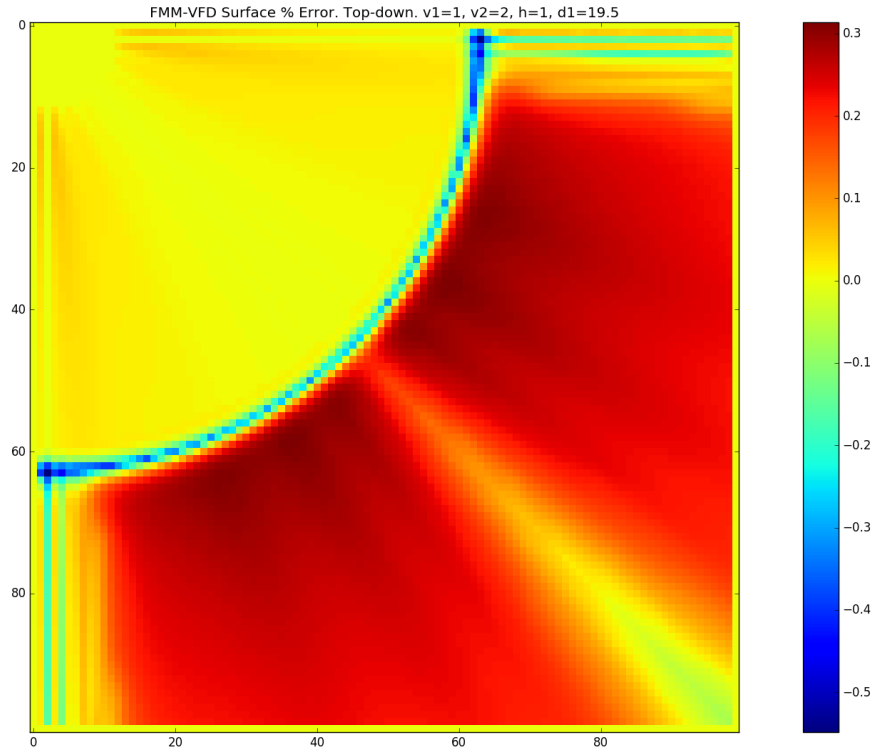


Figure 26: FMM-VFD hybrid forward model % error through 2 layer medium.

6.7 Gravity Model

Another model class that was developed and used for inversion dealt with computing gravity anomaly grids using the density basis function described in the "Basis Functions" chapter. The implementation is a simple grid integrator, which sums vertically projected point mass representation contributions from each point n in a (normally 3 dimensional) density grid to each point m in the (normally 2 dimensional) field measurement grid.

$$\delta g_m = \sum_{n=1}^{N} \frac{G_u \rho(i_n, j_n, k_n) h_s (k_n - k_m)}{((i_n - i_m)^2 + (j_n - j_m)^2 + (k_n - k_m)^2)^{3/2}} \tag{11}$$

Our mass point grid is a scaled analog of the velocity grid, with options for integral partitioning of the vertical and/or horizontal axes. h_s represents the grid spacing in meters, G_u is the universal gravitational constant, and $\rho(i, j, k)$ is an indexed density field. The measurement grid coordinates are described by a 2 dimensional surface mapping earth's topography at a 10km grid spacing, using ETOPO5 global elevation data. The fitted residual was computed using the difference field between model gravitational anomaly, and the WGM2012 Bouguer corrected surface gravity map. The 2 arc minute WGM data was mapped at 10km grid point spacing by averaging WGM data points within 0.2 degrees arc of the field grid point. Figure 27 shows a plot of the averaged WGM2012 Bouguer corrected anomaly field used in inversion, while Figure 28 shows a sample calculated surface Bouguer anomaly for a test parameter set.

One major limitation in using gravitational data in inversions was found to be its run time. Even after some optimizations were applied to the code including multiple restructurings, gravity rendering, when executed on a CPU, takes up to 50 percent more time than the rest of the forward model including NNI rendering and first arrival FMM rendering for 128 events. The performance penalty is major if the velocity grid is further partitioned, with run time growing theoretically at $O(N * M)$, where N is total mass grid points, and M is total measurement grid points. This translates to a theoretical compute bound run time increasing at $O(h_p^2 * v_p)$, where h_p represents the horizontal partitioning factor, and v_p represents

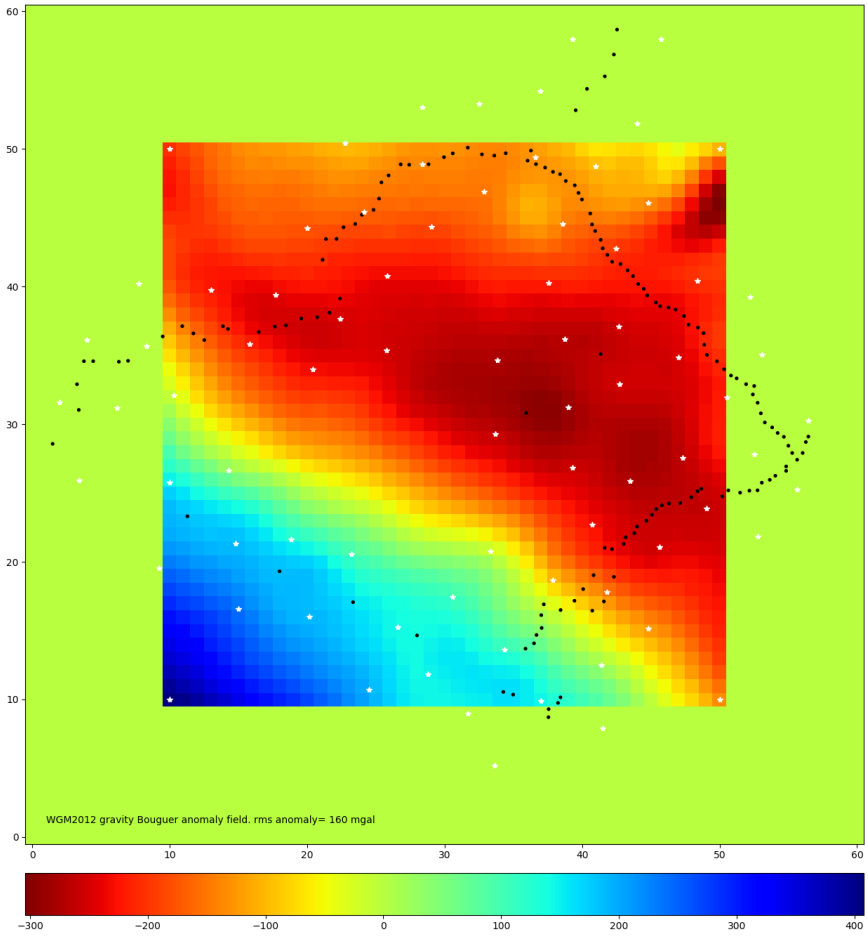


Figure 27: WGM2012 Bouguer anomaly mapping of Peru used for inversion.

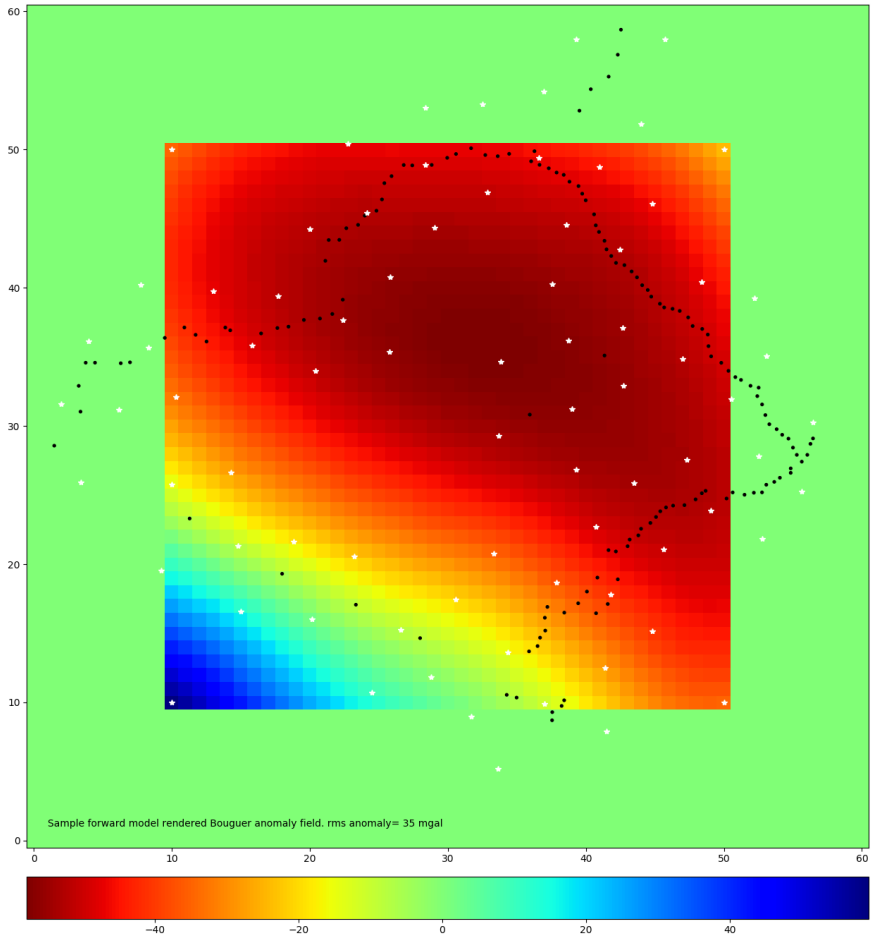


Figure 28: Sample gravity forward model anomaly using test data.

the vertical partitioning factor. Actual run times grow faster due to large grids becoming memory bound while being easily implemented in parallel, we determined this model to be a good candidate for GPU parallalization. When implemented in CUDA C and executed on a GTX1080, the performance increase was found to range between a factor of 5 and 10 relative to a server grade CPU depending on grid size used, with larger grids benefiting more from GPU execution. Importantly, it enabled the CPU to run concurrently during a gravity field render. A single gravity render with the grid dimensions used in our inversions ($61 \times 61 \times 65$ gravity point mass grid, 41×41 gravity surface measurement grid), run times on a GTX1080 ranged between 250m and 350ms versus 1.2+ seconds on a CPU, with roughly 120ms of the quoted GPU run time being due to startup runtime overhead, which would easily be eliminated if a need arose. However, because a forward model execution time for seismic arrival time rendering ranged at between 0.6 and 0.8 seconds, it was determined that no further optimization was necessary for gravity modeling. Furthermore, it enabled us to carry out inversions on a separate build utilizing a 32 core liquid cooled AMD TR-2990WX @ 4.0 GHz with a downgraded GTX1060 GPU with no performance penalty, with the limiting execution dependency still being seismic arrival time computation.

One potential shortcoming of a GPU implementation is the requirement for all gravity model computations to be done in fp32(single floating point precision). Use of fp32 on CPU based models, for the purpose of reducing memory footprint, had previously been found to be a cause of major numeric instabilities when used with our seismic forward model implementation, especially in teleseismic propagation where grid times exceeded about 300 seconds. To verify the validity of GPU based fp32 gravity rendering, we compared outputs from a fp64 CPU model with an fp32 GPU model, with results shown in Figure 29. This would correspond to an rms signal error significantly less than 1 milligal. The measurement uncertainty estimations we use in inversions generally range from 10 to 30 milligal, for which the fp32 error can be considered acceptable with a wide margin.

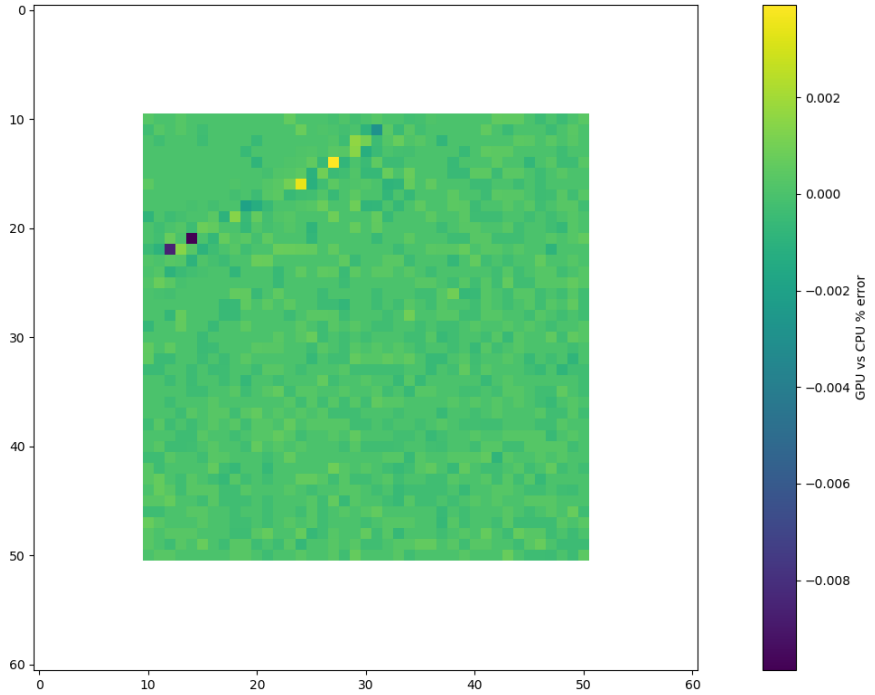


Figure 29: Fractional error for sample density field render using fp32 GPU versus fp64 CPU implementations.

7 Inversion Algorithms

Deferring more conclusive results for potential future work and publication, the following section is a brief introduction of the specific problem faced, an algorithmic overview of the current form of the components integrated in to the Tomographic Inversion Algorithm (TIA) hierarchy described in this chapter. Our core developed components include a Damped Gauss Newton (DGN) and Markov Chain Monte Carlo (MCMC), as well as a top level parameter selecting Monte Carlo algorithm. This complex structure was employed at the expense of simplicity in order to allow to trade off robustness against speed, depending on the computing resources available, the parameter set size, the parameter space volume, and the degree to which the problem is ill conditioned. These trade offs can be made with only configuration option changes, requiring no other modification, giving the user versatility necessary for efficient algorithm development and testing. Many serious bodies of work have been published on inverse theory, and the goal of this work was not to surpass or supersede them. Rather,

our intention was the development of a custom inversion scheme specialized and optimized for the specific type of problem that we are solving, while utilizing well established existing approaches as elements and subprocesses.

For clarity, the parameters in our problem can be distinguished by invertible and non-invertible parameters. Unless specified otherwise, this work assumes that references to parametrization imply that they are invertible parameters. Inversion algorithm input parameters and meta-parameters such as damping, assumed apriori information, or iteration conditions can be considered invertible or non-invertible, depending on whether the top level Monte Carlo algorithm is run enough times to create performance distributions that would qualify it as an effective inversion of those parameters. Below we focus on the theoretical description of the inversion subroutines employed.

7.1 Damped Gauss Newton Algorithm

The Gauss-Newton Algorithm (GNA) is a well established and widely used approach in iterative gradient based least squares inversion, serving as the underlying basis for many modifications. It can be summarized as:

$$\beta_{n+1} = \beta_n + \left[J^T J\right]^{-1} J^T R \tag{12}$$

Where:

$f(\alpha, \beta)$ = model being inverted; generalized form may be referred to in shorthand as f

β = invertible parameter set for the model f

α = all non invertible arguments of f. For conciseness, α will be omitted from notation unless its explicit inclusion is necessary

R = misfit between output of the model and data $R = f(\beta) - y$

J = Jacobian $J = \partial f / \partial \beta$

One common reason for why modifications of GNA are used instead can be explained by the breakdown of the assumed linearization of the response of f to parameter step $\delta\beta = \beta_{n+1} - \beta_n$, implicit in the first order Taylor expansion used to describe perturbations to f around β_n serving as the basis for the inversion. In strongly non-linear scenarios and parameter sets far from their stationary points, inverted steps may not reduce the misfit at all. To address this, most modifications of GNA add damping terms (also known as regularization) to the invertible matrix, trading off reduction in convergence rate against risk of overstepping, which may either fail to reduce the misfit or lead in to a region in the parameter space that may lead to another stationary point entirely (also known as a local minimum).

Broadly speaking, there are two types of approaches to damping: those that attempt to damp based on curvature, which fall into a category of Levenberg-like modifications, and those that attempt to damp based on probabilistic apriori knowledge of data covariance, as well as the apriori knowledge of roughly where the invertible parameter set lies, with a Gaussian probabilistic description of the expected apriori parameter set error, which fall in to the category of Baysean style inversions.

We have combined both of these approaches, along with a 2D parameter grid search in an attempt to maximize convergence rates, designating it as Damped Gauss-Newton, or DGN, which is closely related to the Levenberg-Marquardt algorithm, but accounts for estimated data covariance $C_{d[i][i]} = \frac{1}{\sigma^2_{[i]data}}$, and has a user definable grid search for optimum damping parameters by constructing a damping matrix D using a scaled linear combination $D = d_1 D_{lm} + d_2 D_b$, where:
$D_{lm[i][i]} = [J^T J]_{[i][i]}$ is the damping term specified in the Levenberg-Marquardt algorithm.
$D_{b[i][i]} = \frac{1}{\tau^2_{[i]}}$ is a diagonalized square inverse of the apriori estimated parameter uncertainties.

D_β is also referred to as the apriori parameter covariance matrix inverse C_m^{-1} and used in many Baysean type inversion approaches (Tarantola, 1987). In our approach, we scale the linear combination using a grid search of the permutation of 2 sets of damping scale factors d_1 and d_2. The full diagonal damping matrix can be described by:

$$D_{[i][i]} = d_1[J^T J]_{[i][i]} + \frac{d_2}{\tau^2_{[i]}} \tag{13}$$

The optimal multiplicity of test conditions depends on the size of the gradient invertible parameter vector β, but as a general guideline, in systems where the dominant execution dependency is Jacobian calculation rather than matrix inversion, testing more damping conditions tends to improve convergence rates. We found a multiplicity between 12 and 36 to yield optimal convergence rates, with damping scale factors set to ranges described in the integration sub-chapter. Data covariance is accounted for in our published source code implicitly by scaling data and model outputs by the quotient of the corresponding index data uncertainty.

During a follow-up literature search, it was found that work highly similar to our development of the damping approach in DGN had been published (Rawlinson et al. 2010) where the sole difference between our D_{lm} term and their D term in Eq7, p117 is that in our case, we explicitly set $D_{lm} = J^T J$, while Rawlinson et al. define it vaguely as a "second derivative smoothing operator". Additionally, our approach differs in that DGN explicitly calls for damping and smoothness parameter selection via grid search. Finally, the inversion hierarchy in which DGN is presented differs from that which is used by Rawlinson et al. in relation to its combination to Monte Carlo and Markov Chain Monte Carlo methods.

DGN Algorithm:

- Evaluate model $f(\beta)$ at initial parameter vector β, containing all gradient invertible parameters. Defining $ssq(\beta) = [f(\beta) - y]C_d^{-1}[f(\beta) - y]$ where ssq is the data measurement

uncertainty weighted sum of squares of residuals at β.

- Enter loop, iterating until a break condition or max iteration count is reached.

    1. Compute finite difference Jacobian J using parameter steps $ds_{[i]} = s_r \tau_{[i]}$ or minimum step size, where model noise is significantly smaller than step output range. When using uncertainty ratio for step, we found a step fraction of 0.2 to 0.25 to produce reasonable results.
    2. Define scaled damping matrix $D_{[i][i]} = d_1 J^T J_{[i][i]} + \frac{d_2}{\tau^2_{[i]}}$
    3. Generate 2 column array d with n damping factor permutations for D_{lm}, D_b
    4. Solve inversion for each damping condition, storing each test β as rows in step matrix B: $B_n = \beta + (J^T J + d_{[n][0]} D_{lm} + d_{[n][1])} D_b)^{-1} J^T (y - f)$
    5. Find the test vector in B with the lowest ssq and set as $\beta_{conditional}$.
    6. If $\frac{ssq(\beta_{conditional}))}{ssq(\beta)} > R_{trig}$: exit loop.
    7. Else: set $\beta = \beta_{conditional}$

7.2 Markov Chain Monte Carlo Algorithm

The MCMC class algorithm presented is implemented as two operators: creation and annihilation. The primary optimization metric used by both of the operators is the estimated error of the solution E, which is defined using the estimated curvature of the misfit with respect to the parameter set and can be written in two forms (Tarantola, 1987):

$$E_i(\beta) = \sqrt{[J^T C_d^{-1} J + C_m^{-1}]^{-1}_{[i][i]}} \tag{14}$$

which is derived from Tarantola's estimated posterior parameter covariance matrix when computed using the apriori parameter covariance matrix in both effective damping, and misfit weighting. In an unconstrained well conditioned inversion, the limit of damping required to reduce misfit goes to zero with the distance to a stationary point approaching zero. For

this reason, uncertainty of an unconstrained problem can be the estimated using only the estimated curvature of misfit relative to the parameters:

$$E_i(\beta) = \sqrt{[J^T C_d^{-1} J]^{-1}_{[i][i]}} \tag{15}$$

We calculate uncertainty for both MCMC operators assuming either no damping or minimal damping to make $[J^T C_d^{-1} J]$ invertible.

The annihilation operator is designed to identify parameters weakly sensitive to perturbations in data and/or highly covariant with other parameters, which as a result, carry a high estimated error. After computing estimated error of β, the indices with the highest estimated uncertainties among node field strength parameters are removed from the invertible set.

Annihilation Operator:

- For n in $N_{annihilate}$:

    1. Try to calculate estimated error vector E using $E_i(\beta) = \sqrt{[J^T C_d^{-1} J + d * C_m^{-1}]^{-1}_{[i][i]}}$
    2. While $[J^T C_d^{-1} J + d * C_m^{-1}]$ is singular, do: $d = 10(d + 10^{-6})$ (update d recursively)
    3. Delete β indices of node with highest computed E_i among β_{o3} node field strength parameters
    4. Update Jacobian

Creation Operator:

- For n in N_{create}:

    1. Initialize trial beta, with new initialized 3D Voronoi node
    2. Loop for t iterations:
        (a) Assign random spacial coordinates to test node indices in β_t
        (b) Check distance from test node to nearest neighbor d; if $d < d_{min}$, break;
        (c) Calculate new node's estimated field error $E_i(\beta) = \sqrt{[J^T C_d^{-1} J + d * C_m^{-1}]^{-1}_{[i][i]}}$
        (d) If $E_{trial}[i] < E_{conditional}[i]$ set: $\beta_{conditional} = \beta_{trial}$
    3. If $E(\beta_{trial}[n]) < E(\beta, n)$: accept new node, setting $\beta = \beta_{conditional}$
    4. Update Jacobian

The creation operator is a guided random walk grid search through the Voronoi node parameter space (stepping only to coordinates more than a threshold minimum distance to the nearest node), seeking to lower its field parameter's estimated error E_{test} to below the previously conditionally accepted value. If it meets that requirement, it becomes the new conditionally accepted value.

A notable property of this MCMC implementation is that it is not explicitly designed to weigh the absolute *ssq* as a standalone performance metric, but instead the measure of a

node's estimated parameter uncertainty, which is a relation dependent on that node's scalar field parameter's impact on the model output. A consequence of this is that MCMC executions done in series with DGN actually increase the *ssq* of a DGN run which has converged to a stationary point. However, because of the redistribution of the weakly coupled and highly uncertain nodes in to regions of high sensitivity and lower covariance, a subsequent execution of DGN on the newly perturbed set generally converges to a stationary point at a lowered *ssq* relative to its prior stationary point, which after enough MCMC-DGN loops eventually converges to a noisy bottom around a horizontal asymptote.

By consistently observing a re-convergence of DGN toward lower *ssq* after alternating DGN/MCMC calls, we believe it demonstrates that the addition of an MCMC model focused strictly on reassignment of parameters into high contrast areas is able to move converged DGN solutions out of local minima. We believe that this is an important feature, and if applied correctly, it may be able to improve on the robustness of a Gauss-Newton class inversion approach, while allowing for a user selectable computational cost penalty that is still significantly smaller than that of completely random Monte Carlo approaches. One qualification, however, is that without a proper roughness penalty, such methods may produce results with potentially nonphysical spatial features, which we will discuss briefly in results.

8 Top Level Integration and Tomographic Inversion

Model integration and tomographic inversion is done by the "Traveltime Forward Model" (TFM) and "Tomographic Inversion Algorithm" (TIA), which we developed in parallel with the component basis function and seismic forward model (FMM-VFD).

In order to properly characterize our inverse problem and choose an appropriate inversion

Parameter Type	Dimensionality	Multiplicity
event coordinates	geodetic coordinates	array
event offset times	time	array
Moho NNI node horizontal position parameters	2D grid coordinates	array
Moho NNI node depth parameters	1D grid coordinate	array
mantle NNI node position parameters	3D grid coordinate	array
mantle NNI node field parameters	velocity	array
initial Airy isostatic compensation ratio	dimensionless	scalar
crust velocity	velocity	scalar
crust mantle density contrast	density	scalar
background mantle velocity for gravity	velocity	scalar
density perturbation scaling factor	density/velocity	scalar

Table 2: Parameter dimensionality table.

approach, we must first identify the parameter space and dimensionality with which our problem is dealing with. With the goal of our work being the identification and mapping of deep seismic events, along with what is widely believed to be an accompanying subducting slab structure, while using as much of the available data as possible, we can describe the physical parametrization of our problem as: relocation of event hypocenters to the best fitting solution and physical mapping of velocity structure underneath our seismic network (PeruSE) to the maximum practically resolvable velocity field depth. In order to do this using what we claim to be an efficient parametrization, we must solve for: local event locations as physical coordinates, offset times for local and teleseismic events, upper mantle p-wave velocity field V_p, density field directly corresponding to V_p, crustal velocity, and Moho depth variation.

Already this is an unorthodox problem, due to the existence of at least 4 physical parameter dimensions, 3 of which (event coordinates, velocity field, density field) range from being quasi linear to highly nonlinear, depending on where they are located in the parameter space. Our parameter space ideally includes all realistically plausible physical values and a horizontal spacial extent, roughly bounding the horizontal extent of our station network. Our chosen representation of these physical parameters includes array and scalar parameters with parameter types and dimensionality described in Table 2.

Parameter Type	Range
grid dimensions (East (di), North (dj), Up (dk))	[61,61,33]
grid spacing (km)	10
LM-Baysean damping range permutations	[0,.3,1,3,10,100],[10000,100000]
MCMC node annihilations per cycle	3-10
MCMC node creations per cycle	2-4
source coordinate apriori uncertainty (deg, deg, km, s)	[.4,.4,40,100]
DGN iteration count per top level inversion loop	5-20
MCMC trial count	20-30
outlier trim threshold residual (s)	1.2-1.5

Table 3: Non-invertible meta parameter ranges.

Parameter Type	Range
Airy isostatic terrain-crust compensation	0-6
crust velocity range km/s	5.9-6.2
crust-mantle density contrast range g/cc	.3-.5
initial and background mantle velocity (km/s)	7.7-8.5
density/velocity perturbation scaling factor ($\frac{g/cm^3}{km/s}$)	.04-.4
β_{o2} node apriori uncertainty (km/h,km/h,km/h)	[1-5,1-5,1-2]
β_{o3} node apriori uncertainty (km/h,km/h,km/h,km/s)	[4-5,4-5,4-5,.3-1]

Table 4: Invertible parameter initialization ranges.

8.1 Parameter Initialization

Fixed non invertible parameters are given in Table 3. Model parameters set at physical values considered apriori knowledge use ranges given in Table 4. Other parameters are set using the following mixed random scheme:

1. Extract velocity field parameters from β, render velocity using selected basis function.
2. For each loaded local event data set:
   (a) Initialize a local cluster of gridpoints around source coordinates in FMM-VFD.
   (b) March FMM-VFD to compute first arrival time grid.
   (c) Linearly interpolate for computed travel times at each reporting station.
3. For each loaded teleseismic event data set:

(a) Initialize SET computed travel times at 5 of the grid boundaries, excluding the top surface ajacent to terrain inside FMM-VFD. Add 2nd sheet from bottom of the grid to the FMM-VFD minheap, setting their status to "visited" and setting their time using SET.

(b) March FMM-VFD to compute first arrival time grid.

(c) Linearly interpolate for computed travel times at each reporting station.

4. If gravity inversion is being performed, compute the gravity anomaly residual field using gravity model perturbation field basis function described above, subtract from WGM2012 data, using only points more than "buffer"(set to 10) points away from an edge to minimize edge effects.

5. Return model computed travel times, or times+gravity anomaly field.

We note that in our TFM implementation, teleseismic arrival incorporation is done in a single pass, in contrast with a related tomographic package FMTOMO, published by Sambridge and Rawlinson, which initializes a surface with standard Earth travel times, back propagates to the grid boundaries, assuming an unperturbed velocity model, then re-propagates to the surface. The methods are analytically, but not numerically equivalent. In addition to being more computationally efficient, our approach has a lower model error in principle, due to the absence of model error introduced in the course of an unnecessary back propagation of the waveform to grid boundaries.

Tomographic Inversion Algorithm:

1. Load user specified space delimited .txt files containing earthquake coordinates, station coordinates, picked station arrival times, and optionally a clip of the raw waveform.

2. Convert station coordinates to Earth Centered Earth Fixed(ECEF) Cartesian system and apply a rotation and translation to convert to local East North Up Coordinates

using a centrally defined lat/lon coordinate set and a depth which places the top of the grid above the terrain enclosed horizontally by the grid.

3. If station coordinates are within the defined Cartesian grid, accept them into the inversion data set.

4. Categorize imported event as a local event or teleseism. Mark event type as a flag on each data point.

5. Generate a loose regular coordinate grid and denser station guided coordinate grid of 2D Voronoi nodes with field values set using iasp91 Moho depth perturbed by the Airy factor scaled local ETOPO terrain altitude: $d = 35 + ETOPO5(i, j) * airy$, converting d to ENU coordinate k.

6. Generate a loose regular coordinate grid and denser initial event location biased coordinate grid of 3D Voronoi nodes with amplitude set to initialization parameter v_{mantle}

7. Launch sequential inversion loop with user defined loop multiplicity:

   (a) Execute DGN until convergence or model smoothness breaks down. During DGN application, in configurations with a distinct crust, we enforce velocity limits between 7.5 and 9.5 for every stepping node, limiting nodes to the parameter space in between.

   (b) Apply residual outlier rejection filter to data using model outputs at inverted β

   (c) Apply annihilation operator to remove poorly resolved nodes

   (d) Execute 1-2 DGN iterations

   (e) Apply creation operator to inject node into sparsely populated regions with favorable estimated uncertainty, executing a subloop with creation and 1-2 DGN iterations per loop

The algorithm hierarchy used in an inversion routine is represented in Figure 30.

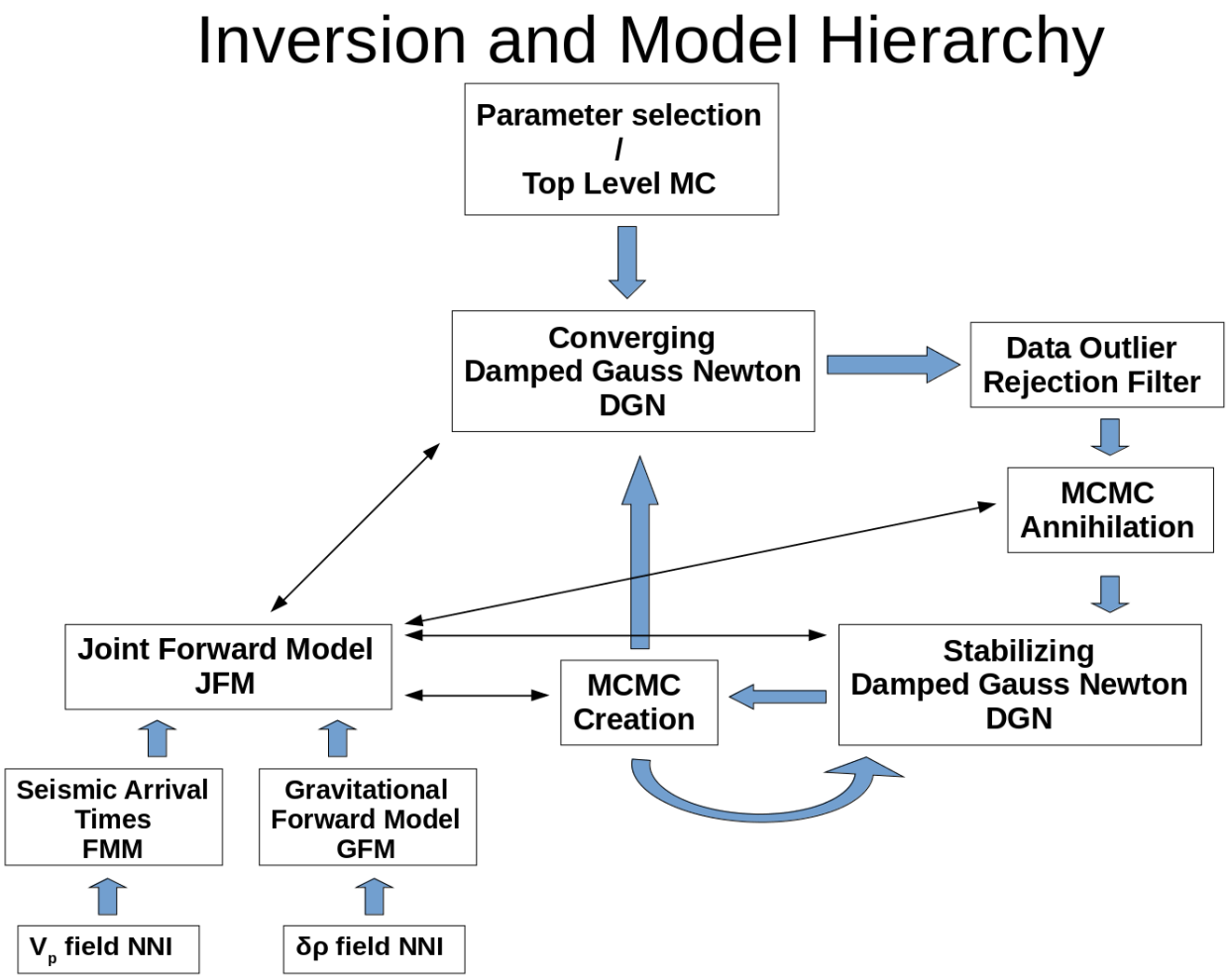


Figure 30: Block diagram of subroutines and models used in our inversion hierarchy.

9 Synthetic Testing

One of the ultimate performance metrics employed in seismic tomography is the recoverability of a physically realistic synthetic parameter set when inverting a null hypothesis parameter set using synthetic parameter set forward model output as fit data. While this test does not guarantee the uniqueness of solutions, it can be a helpful tool in characterizing the resolvable parameter space of an inversion. One widespread synthetic test is the inversion of a checkerboard perturbation of the parameter set. We do not use this test because it is nonphysical, being an unrealistic and unanticipated perturbation geometry in mantle tomography. Instead, we use a simple Gaussian packet synthetic perturbation model, which, while generic, is a more physically realistic perturbation field. The tests presented are a preliminary evaluation, and should not be considered a comprehensive synthetic recovery trial, which should include other structures such as perturbation planes, cylinders, and combinations thereof.

The synthetic models tested included halfspace and 2 layer(mantle/crust) background models and a positive Gaussian perturbation with 1km/s amplitude, centered at $[i, j, k] = [30, 30, 18]$ and Gaussian scale length of 60km. Figures 31 - 32 show synthetic and recovered velocity field cross sections in a perturbed halfspace model, respectively. Figures 33 - 34 show synthetic and recovered velocity field cross sections in a perturbed 2 layer model, respectively. Generating synthetics were made using catalog event locations with same reporting stations as data inversions.

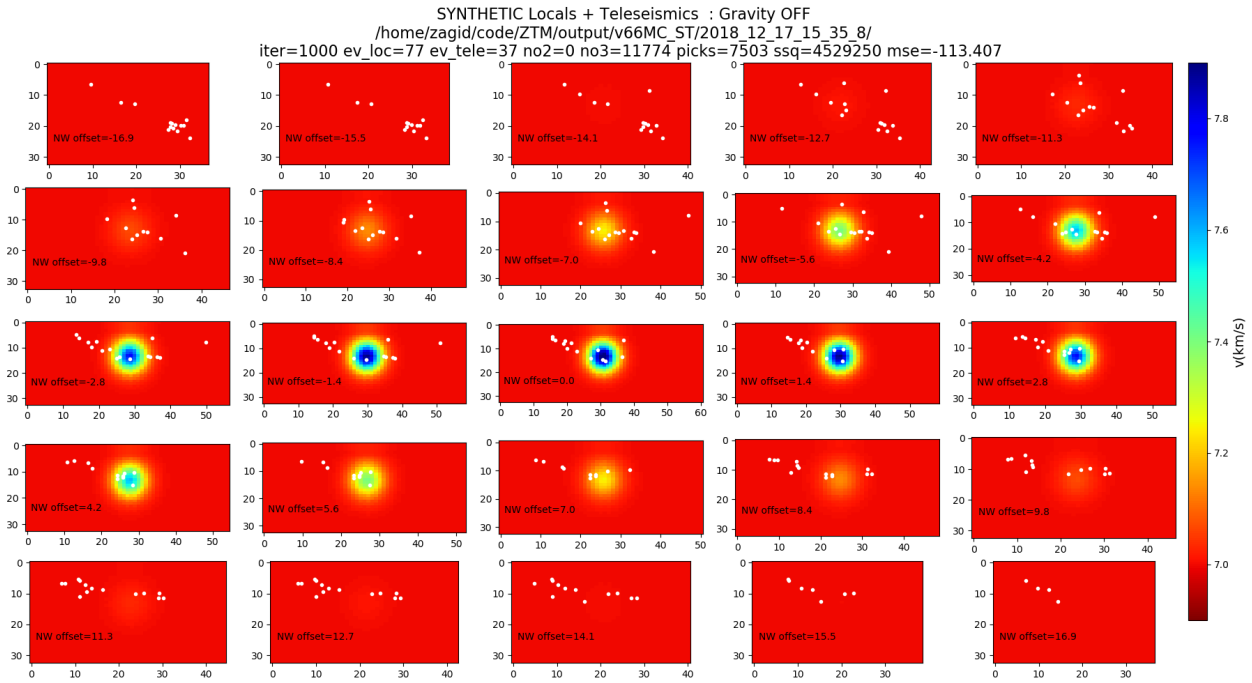


Figure 31: Synthetic data generating field using Gaussian perturbed halfspace model velocity parametrization. Gaussian perturbation amplitude: 1km/s; characteristic length: 60km.

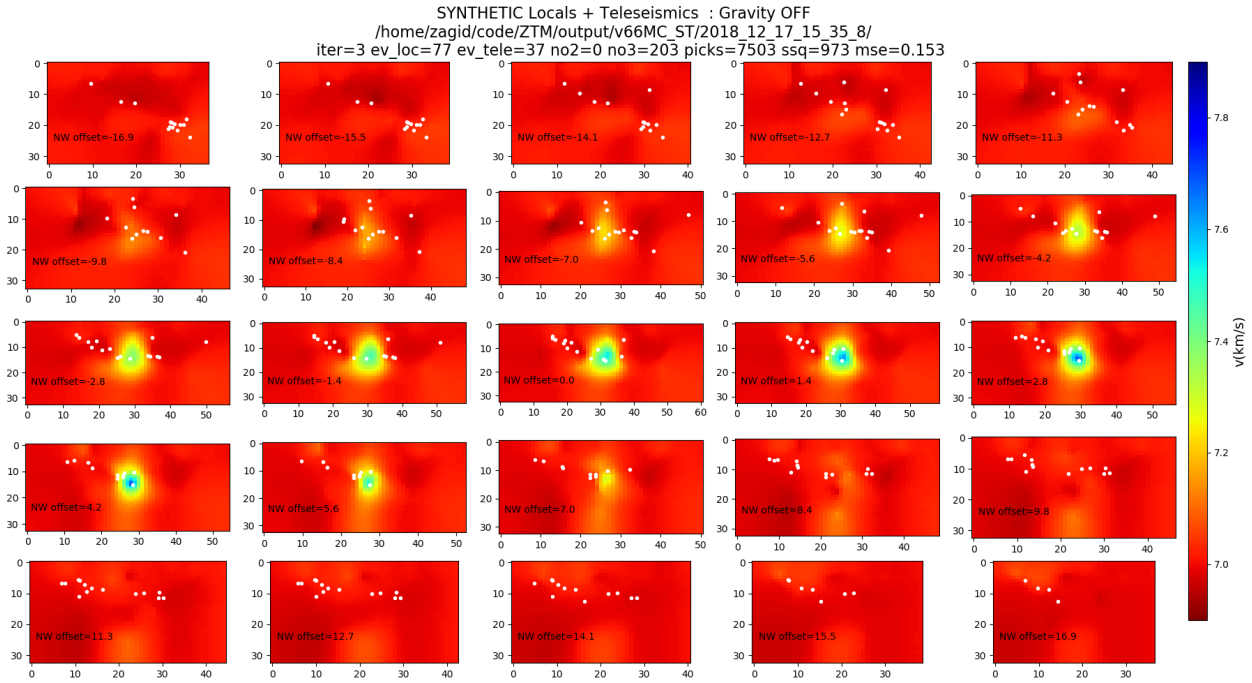


Figure 32: 3rd iteration of damped inversion using synthetic time data produced using 31. Gaussian perturbation amplitude: 1km/s; characteristic length: 60km.

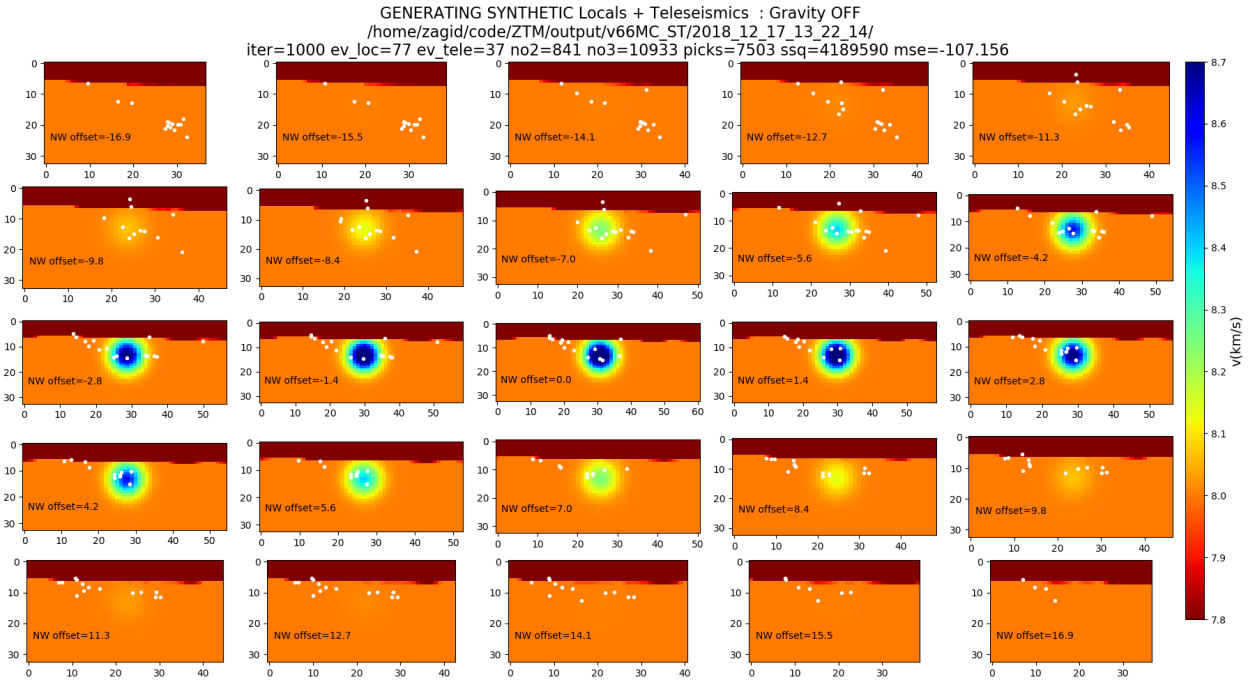


Figure 33: Synthetic data generating field using Gaussian perturbed 2 layer model velocity parametrization. Gaussian perturbation amplitude: 1km/s; characteristic length: 60km.

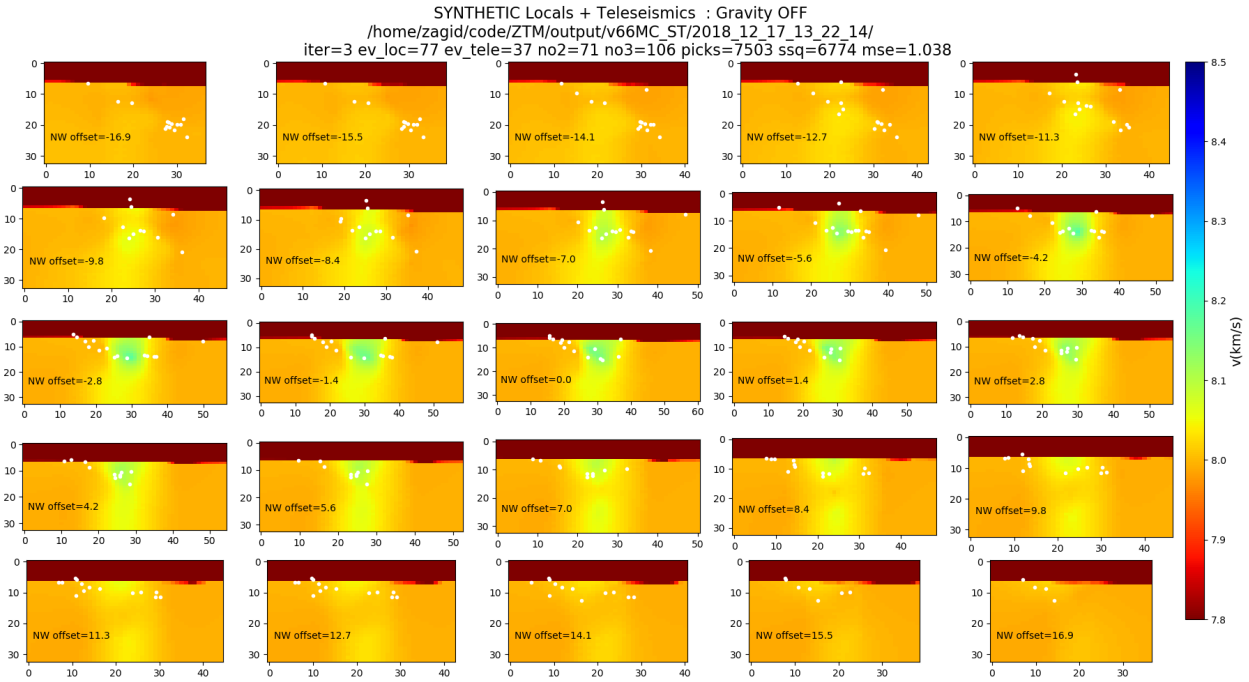


Figure 34: 3rd iteration of damped inversion using synthetic data field and perturbed 2 layer model velocity parametrization. Gaussian perturbation amplitude: 1km/s; characteristic length: 60km.

parameter	value
Airy isostatic compensation	5km thickening/1km terrain.
initial crust velocity	6 km/s
initial mantle velocity	8 km/s
Moho density contrast	0.4 g/cc
drho/dv scaling factor	0.3 g/cc*s/km
time uncertainty scaling	0.3s
gravity uncertainty scaling	20mgal

Table 5: Standard inversion initialization values.

10 Inversion Results and Error Estimation

The hierarchy of algorithms described above was tested using data described in "Methods" and "Waveform Pre-processing and Timing" sections. Parameter initialization in most single parameter set runs was selected with an effort to utilize the information believed to be reasonably well constrained in regards to upper mantle and crust characteristics, while a few selected runs were run as stress tests in order to obtain an indication of solution stability. We present results from inversions that include configurations with/without a Moho/distinct crust, with/without gravity, locals only/teleseisms only, as well as complete inversions. The standard initial parameters used are given in Table 5. The specific iteration conditions used for produced results are:

- 5 DGN executions
- 1.3s residual outlier rejection filter
- 15 applications of annihilation operator
- 2 DGN executions
- 2 loops with 1 creation operator, 1 DGN iteration

However, all figures below, except inversion fields at the end associated with the waveform summary include only selected representative inversion configurations up to 5th DGN

iteration. Each run configuration is summarized in caption, title and color bar information. Care was taken to select and present data in a way to reduce bias, but as with any inversion result, the reader is encouraged to be mindful of potential sources of bias in terms of model, initial parameter, and meta-parameter selection.

10.1 Figure Types

- Unless otherwise noted, figures drawn in 3 dimensions and/or with wire mesh are color mapped by estimated relative uncertainties of velocity node field values. The wire mesh represents inverted Moho.

- Figures with slice progressions and a density color axis are inverted density fields, done only in configurations that invert gravity. These are the only figures with a symmetric logarithmic color mapping.

- Figures with slice progressions and a velocity color axis are the rendered velocity field represented in vertical slices oriented NE toward the right. The progression runs roughly from the SE end of the PeruSE network to its NW end. These figures also contain relocated event hypocenters represented as white dots.

- Twin figure color maps represent the gravity misfit on the left when available (gravity option set to true) and Moho depth in km on the right, with the Moho Voronoi node coordinates plotted as white dots.

Variable	Definition
iter	inversion iteration, where 0 is initial conditions; may also begin at 100
ev_loc	number of local events used in presented inversion
ev_tele	number of teleseismic events used in presented inversion
no2	number of 2 dimensional Voronoi nodes used to describe Moho
no3	Number of 3 dimensional Voronoi nodes used to describe mantle velocity field
picks	number of trace first arrival picks used in inversion
ssq	sum of squares of residuals of all points in inversion
mse	mean square error of residuals normalized to their uncertainty

Table 6: Inversion figure header key.

10.2 Inversion Plots

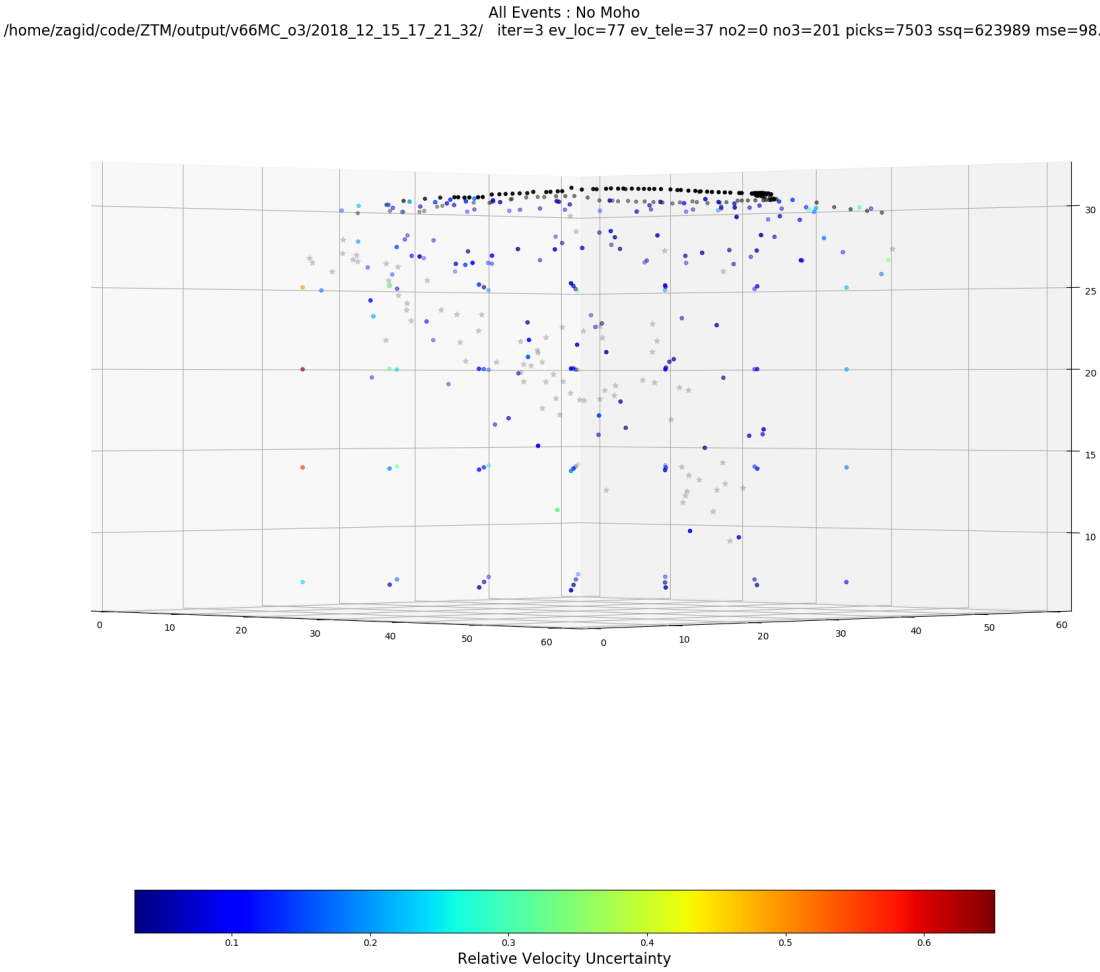


Figure 35: Single layer 3D view of relative node field uncertainty. 3rd iteration. Using locals, teleseisms.

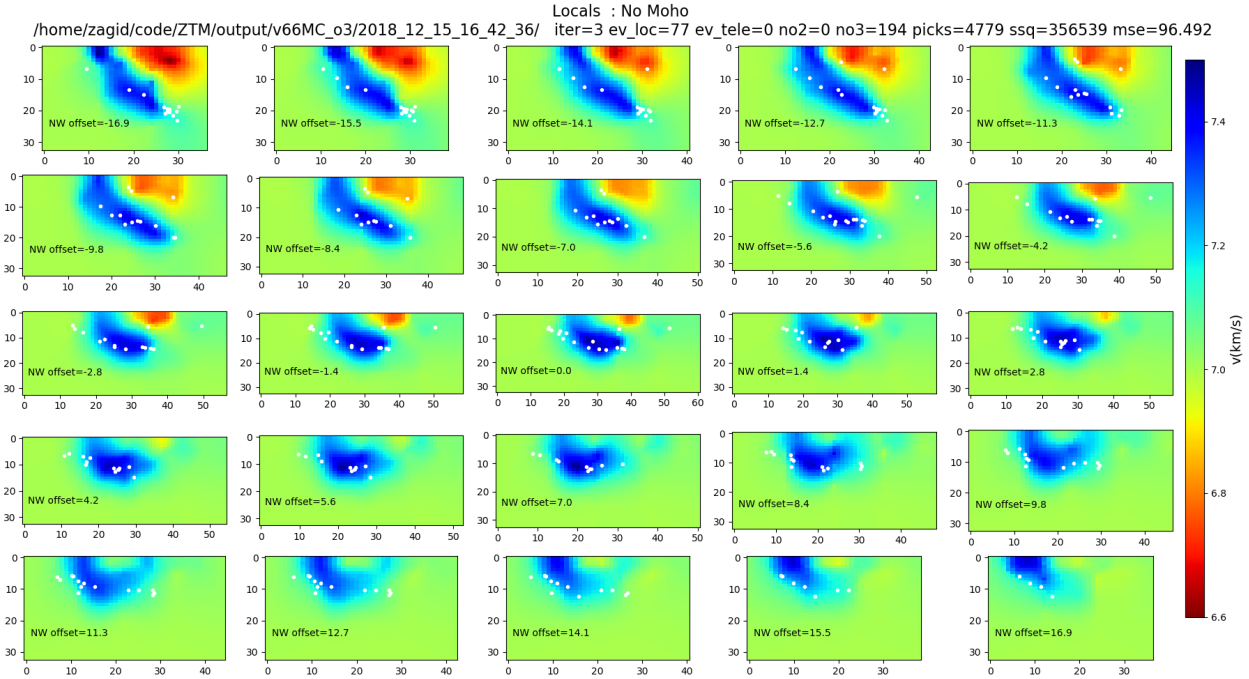


Figure 36: Single layer V_p cross sections. 3rd iteration. Using locals.

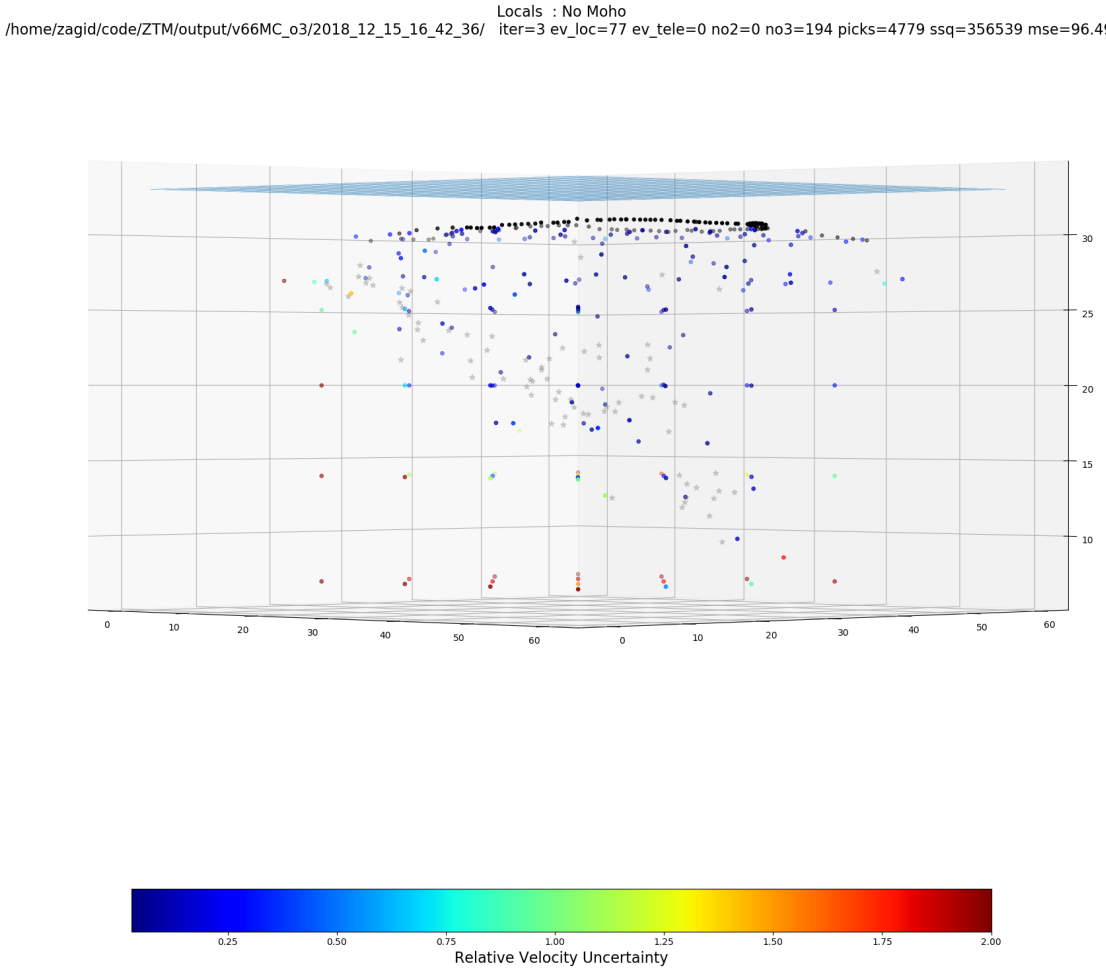


Figure 37: Single layer 3D view of relative node field uncertainty. 3rd iteration. Using locals.

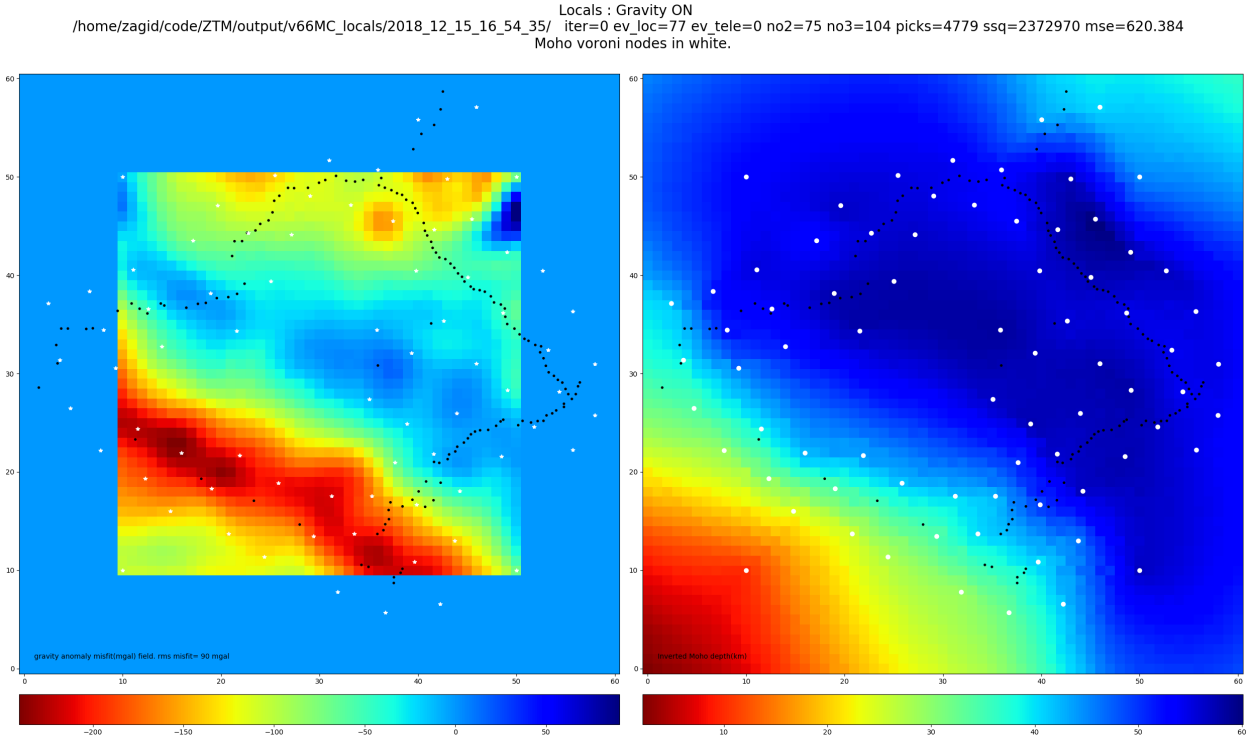


Figure 38: Top down gravity misfit(left), Moho depth(right) map. 0th iteration (initial conditions). Using locals, gravity.

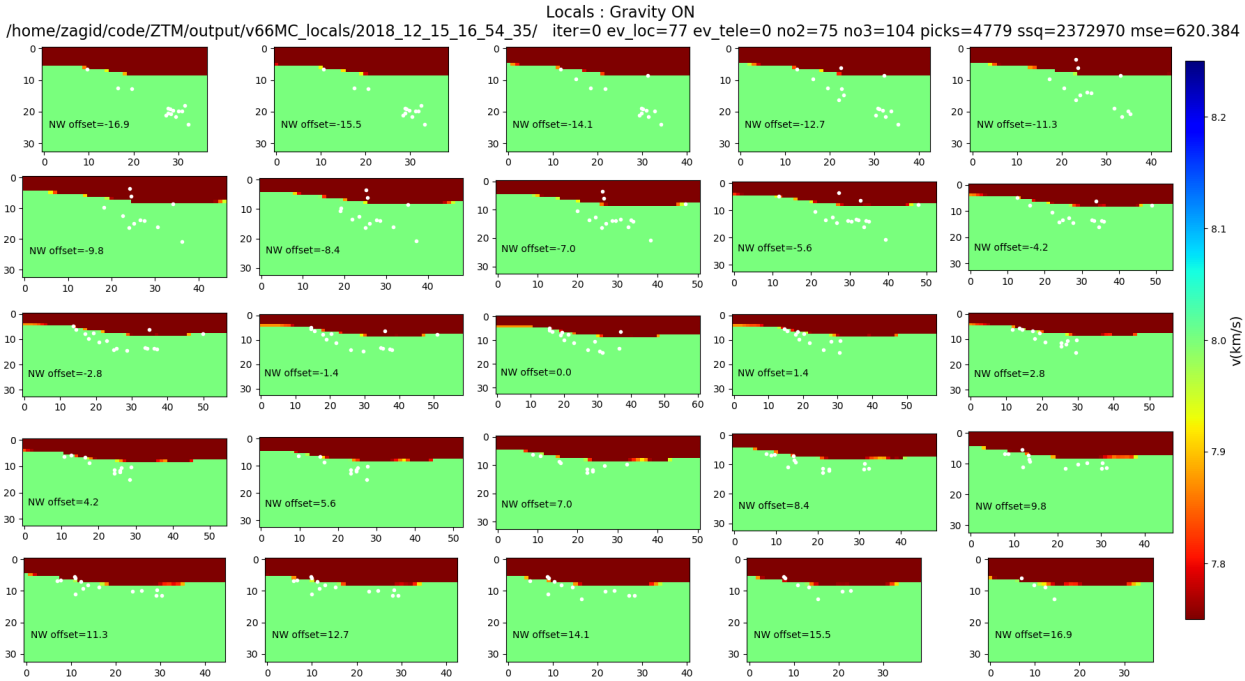


Figure 39: V_p cross sections. 0th iteration (initial conditions). Using locals, gravity.

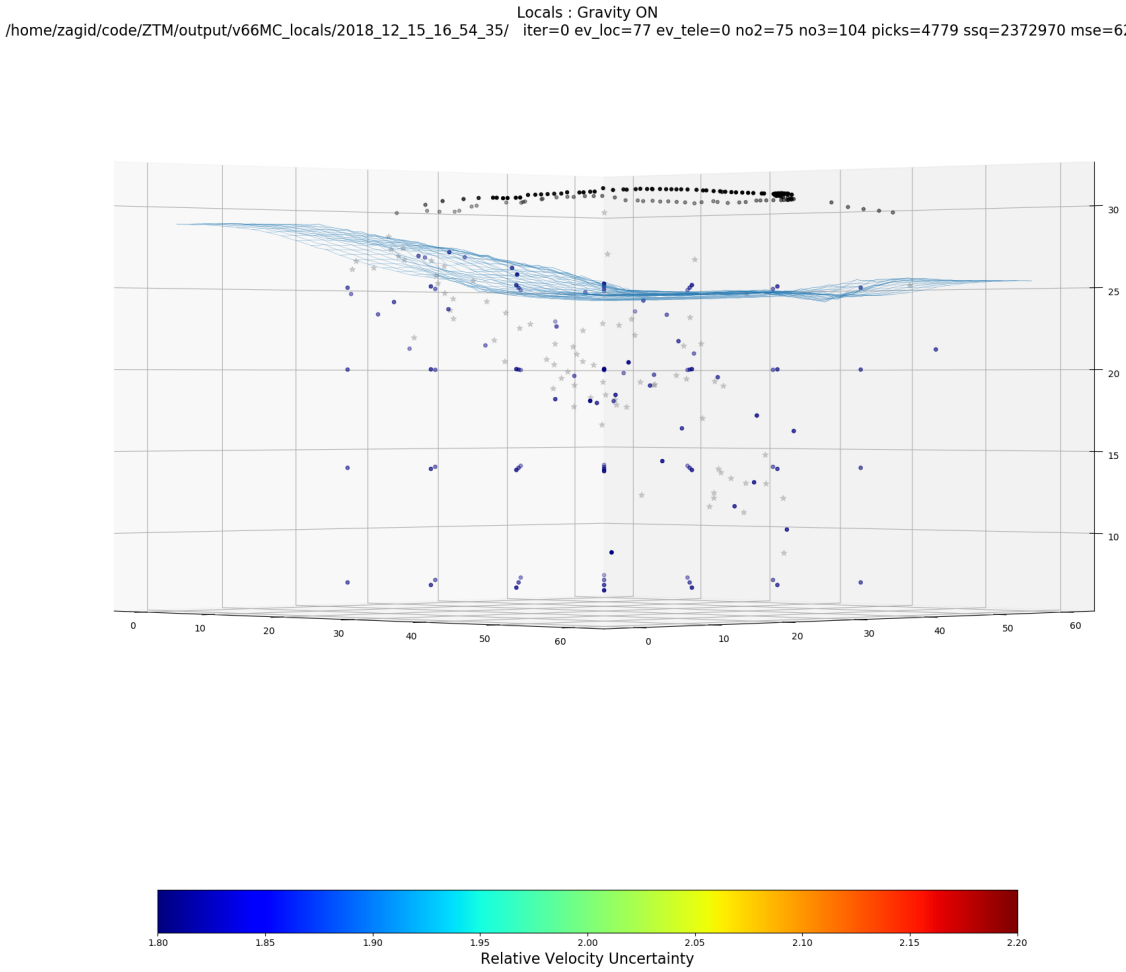


Figure 40: 3D view of relative node field uncertainty. 0th iteration (initial conditions). Using locals, gravity.

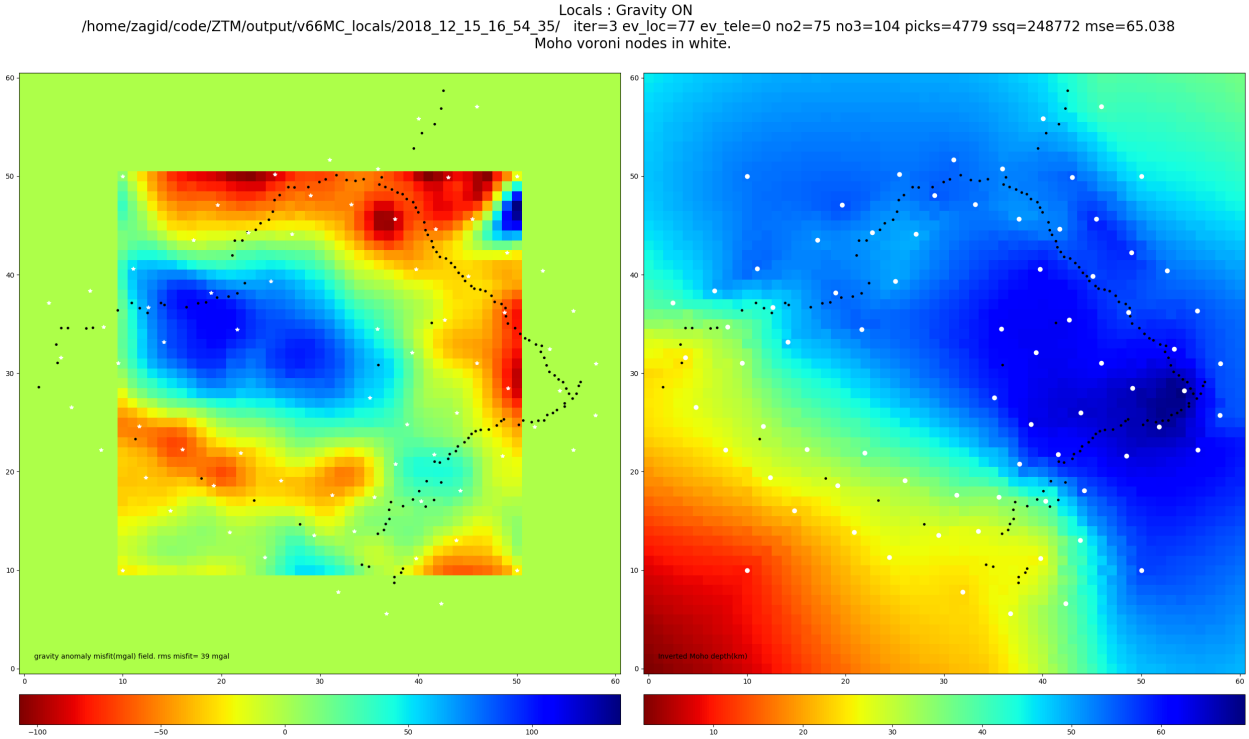


Figure 41: Top down gravity misfit(left), Moho depth(right) map. 3rd iteration. Using locals, gravity.

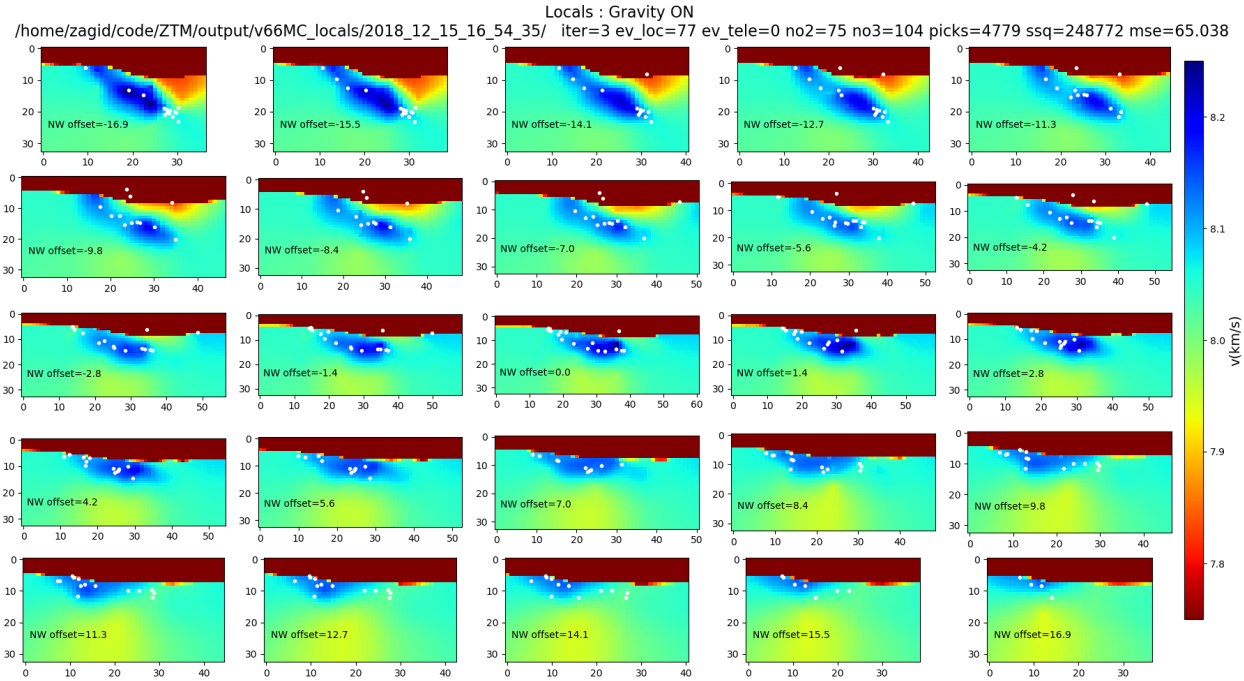


Figure 42: V_p cross sections. 3rd iteration. Using locals, gravity.

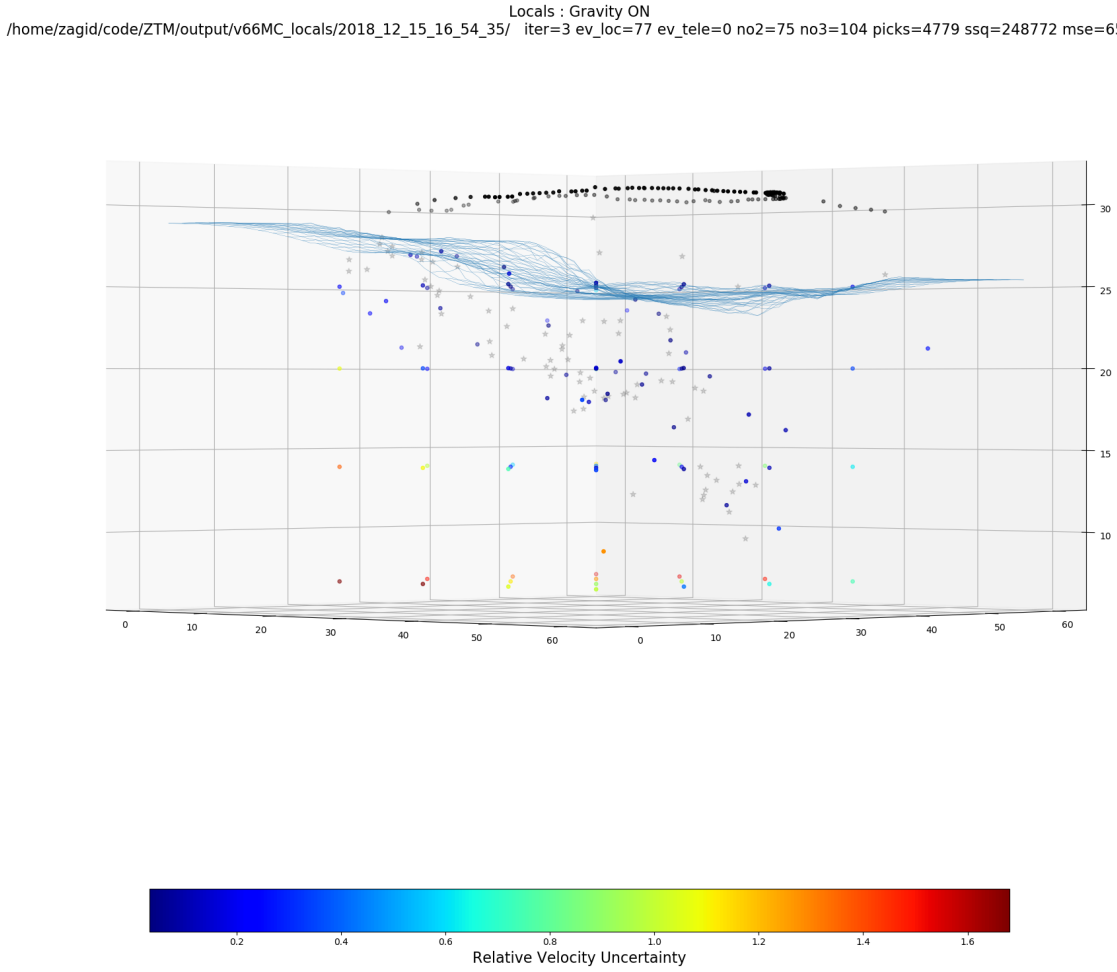


Figure 43: 3D view of relative node field uncertainty. 3rd iteration. Using locals, gravity.

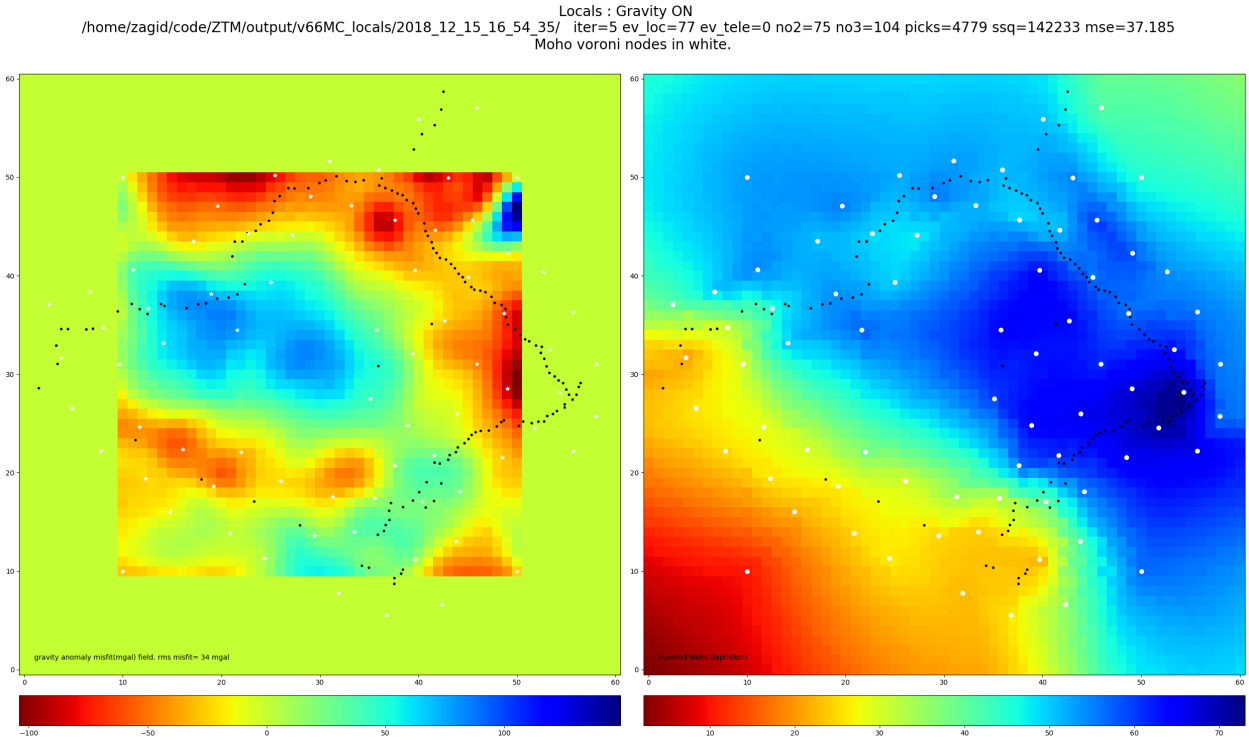


Figure 44: Top down gravity misfit(left), Moho depth(right) map. 5th iteration. Using locals, gravity.

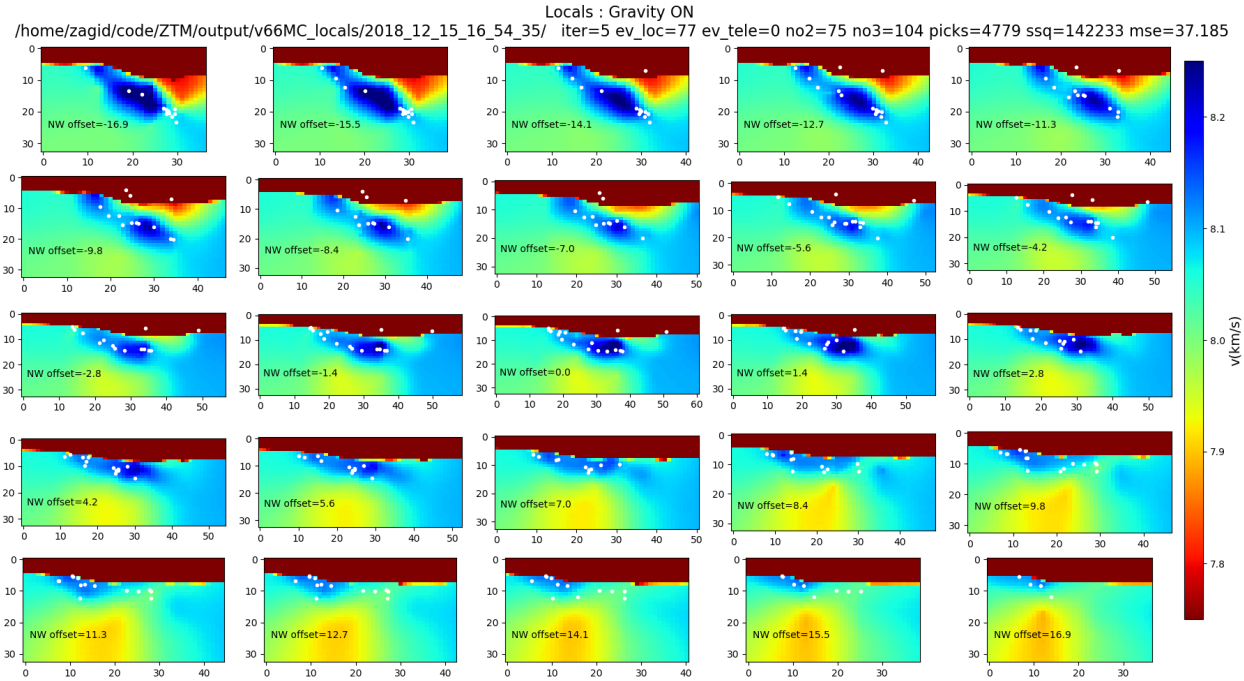


Figure 45: V_p cross sections. 5th iteration. Using locals, gravity.

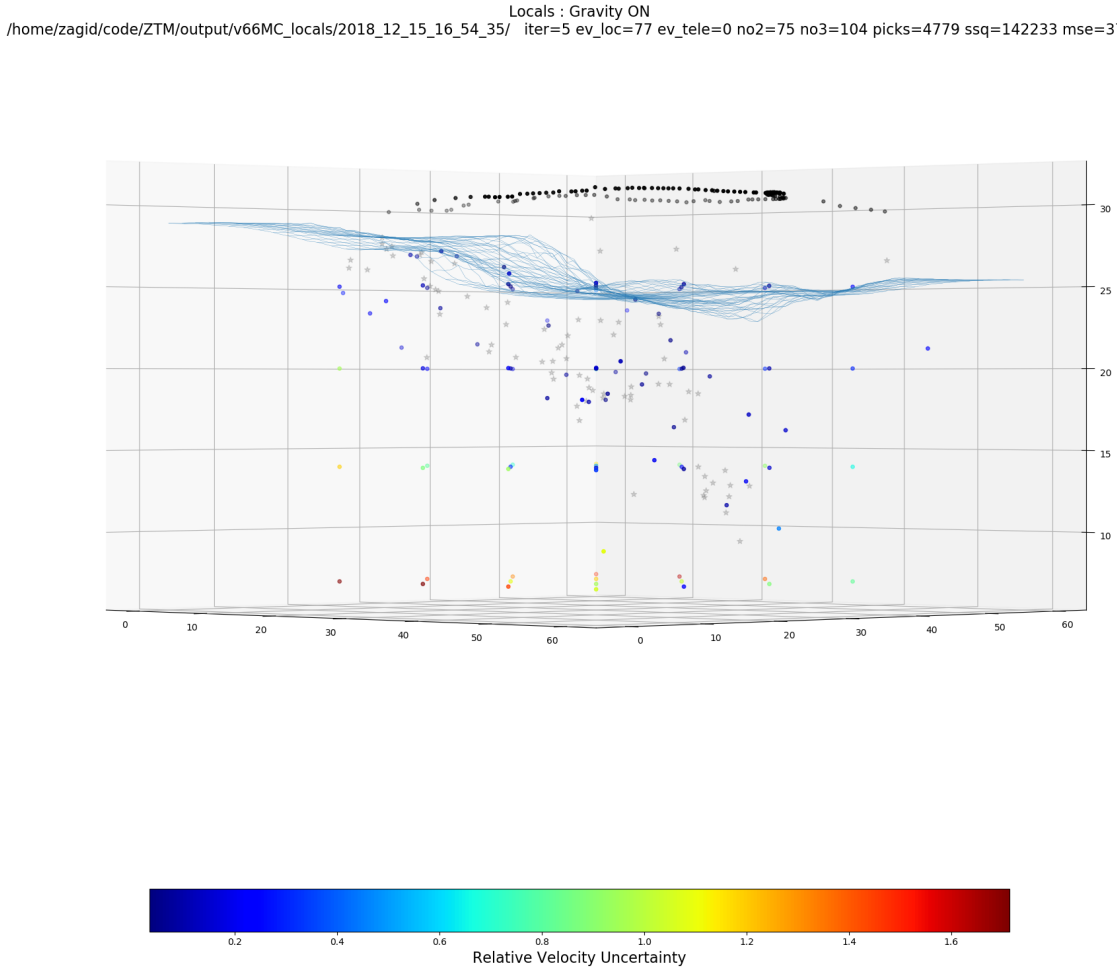


Figure 46: 3D view of relative node field uncertainty. 5th iteration. Using locals, gravity.

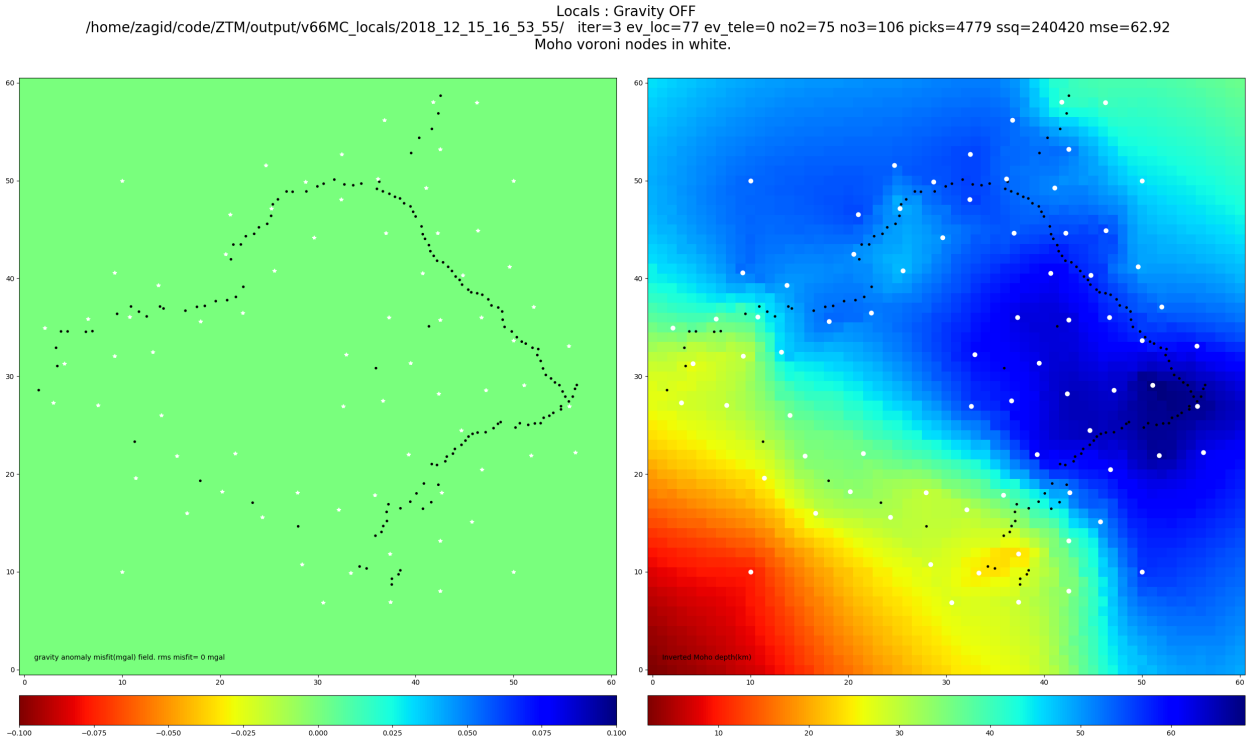


Figure 47: Top down gravity misfit(left), Moho depth(right) map. 3rd iteration. Using locals.

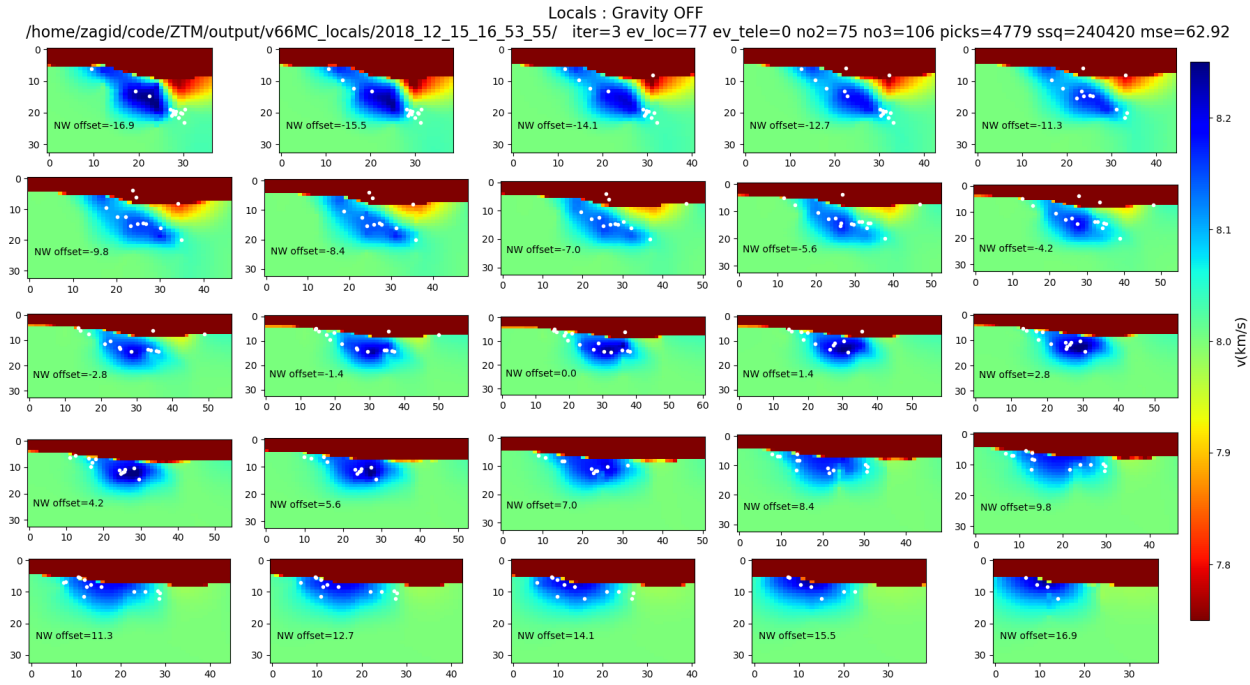


Figure 48: V_p cross sections. 3rd iteration. Using locals.

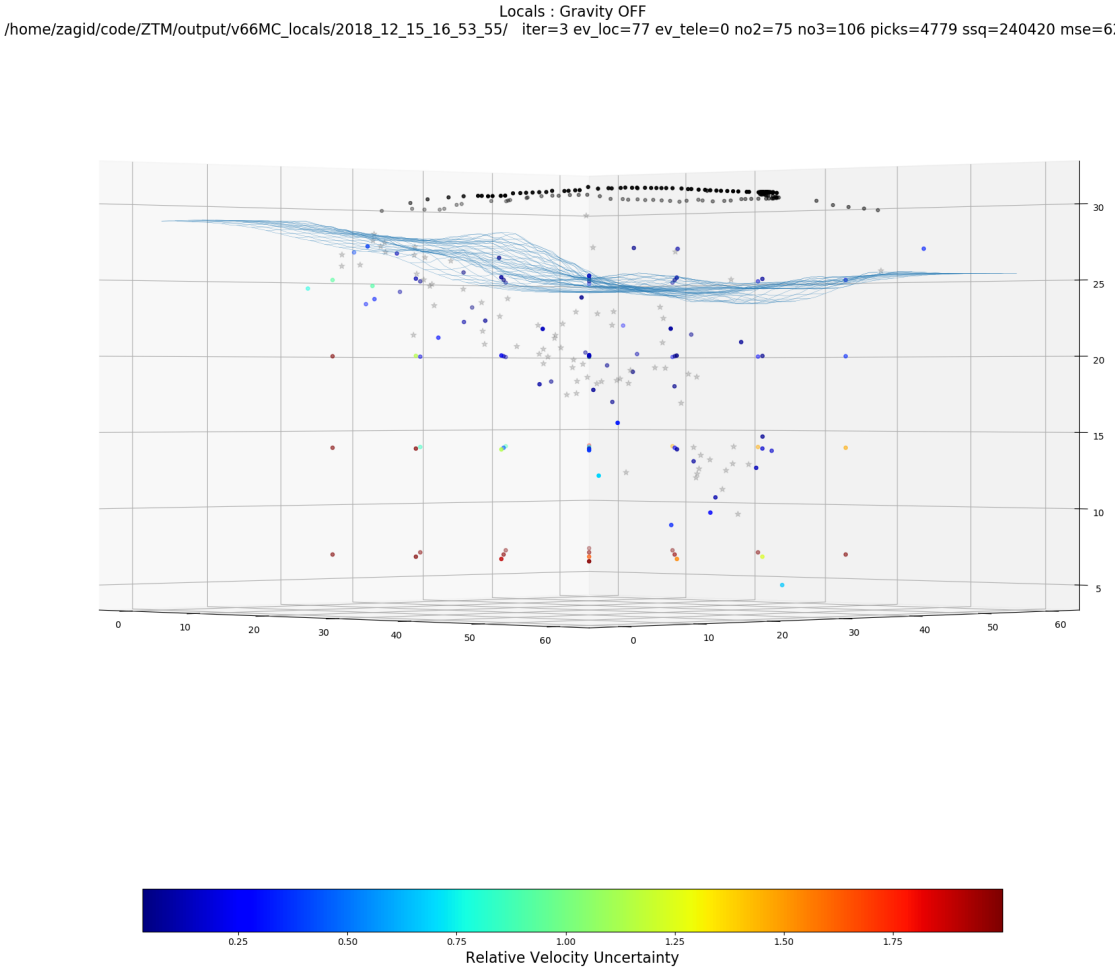


Figure 49: 3D view of relative node field uncertainty. 3rd iteration. Using locals.

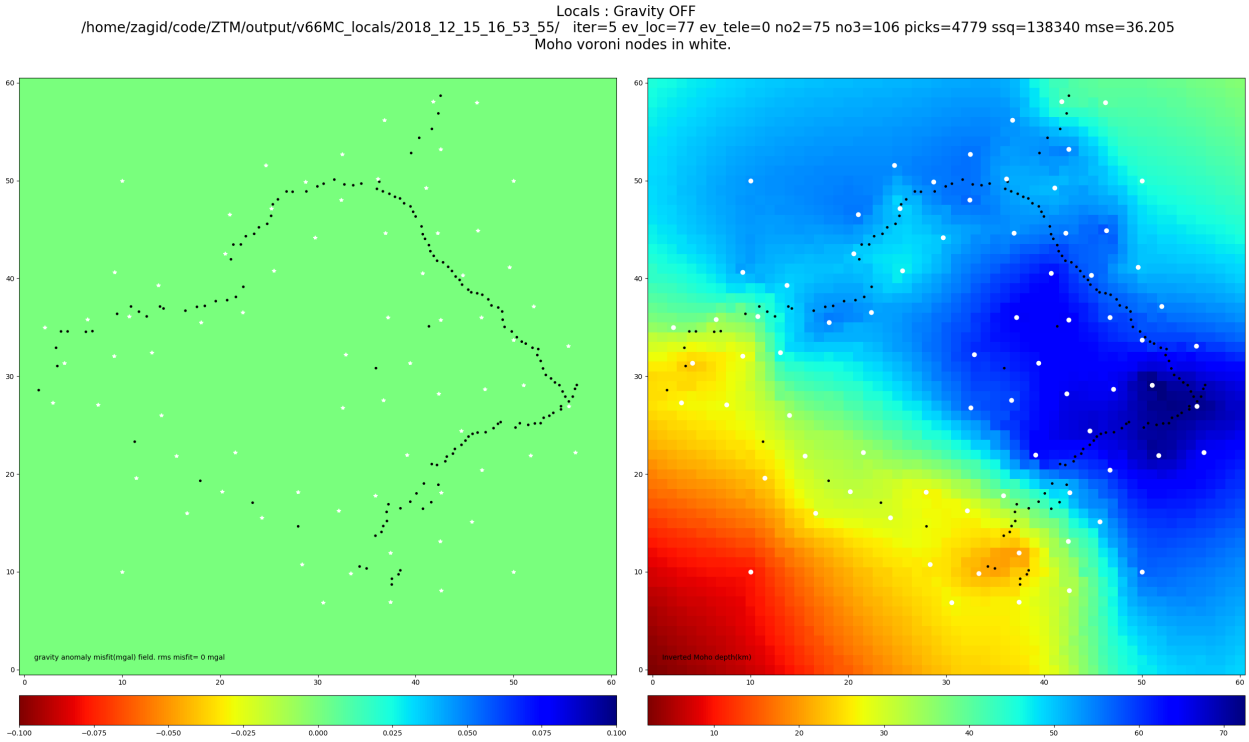


Figure 50: Top down gravity misfit(left), Moho depth(right) map. 5th iteration. Using locals.

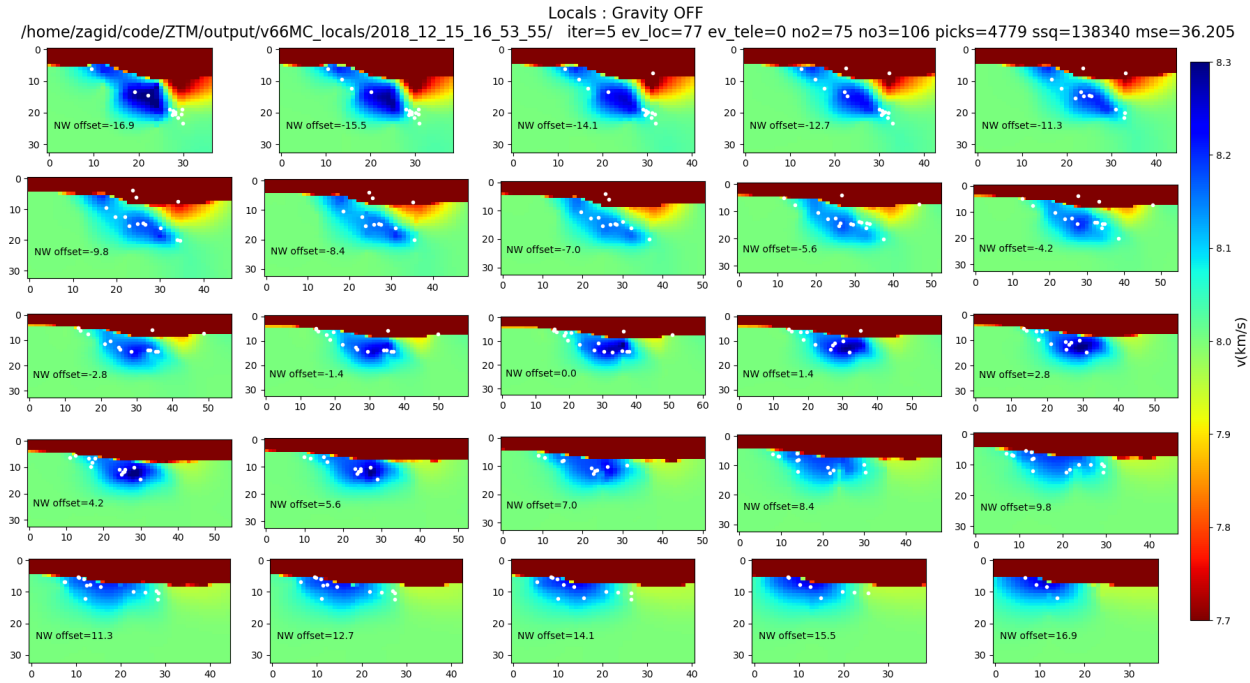


Figure 51: V_p cross sections. 5th iteration. Using locals.

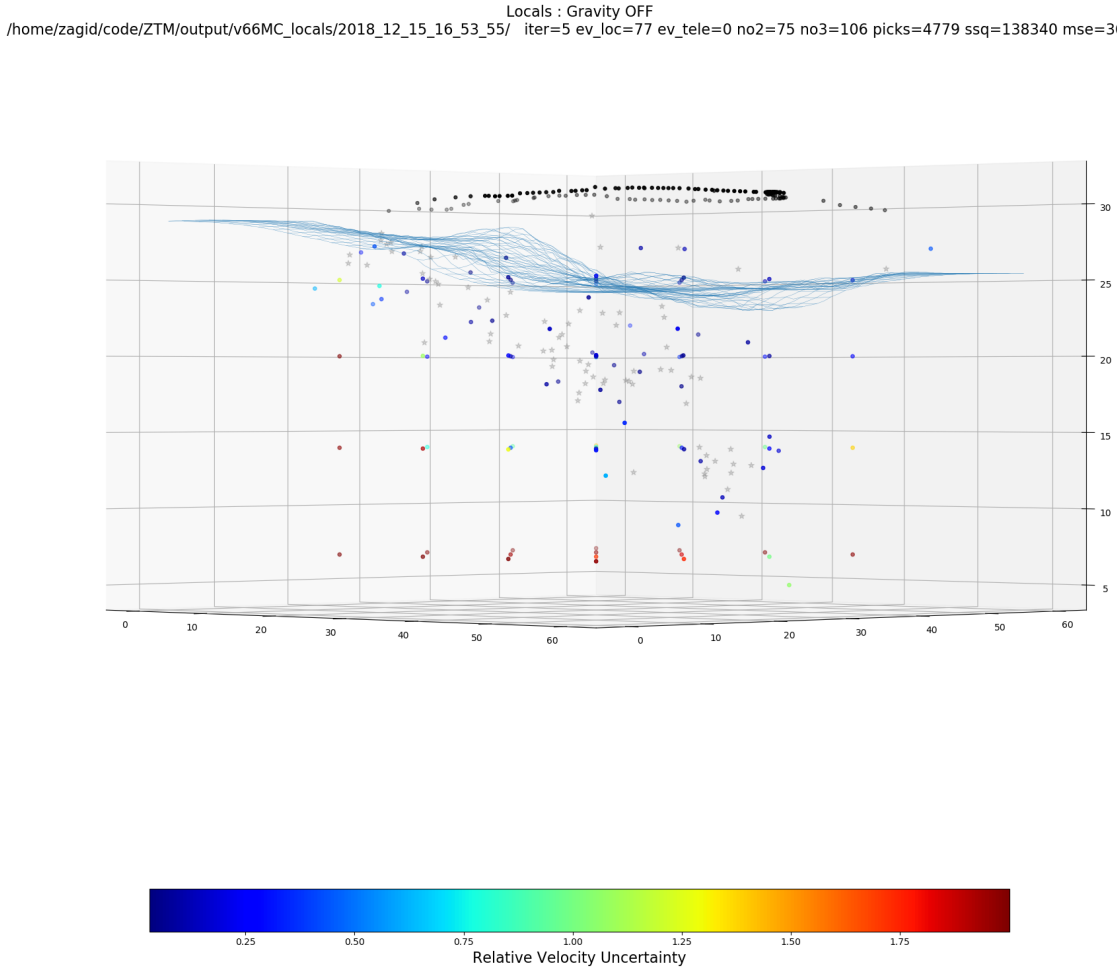


Figure 52: 3D view of relative node field uncertainty. 5th iteration. Using locals.

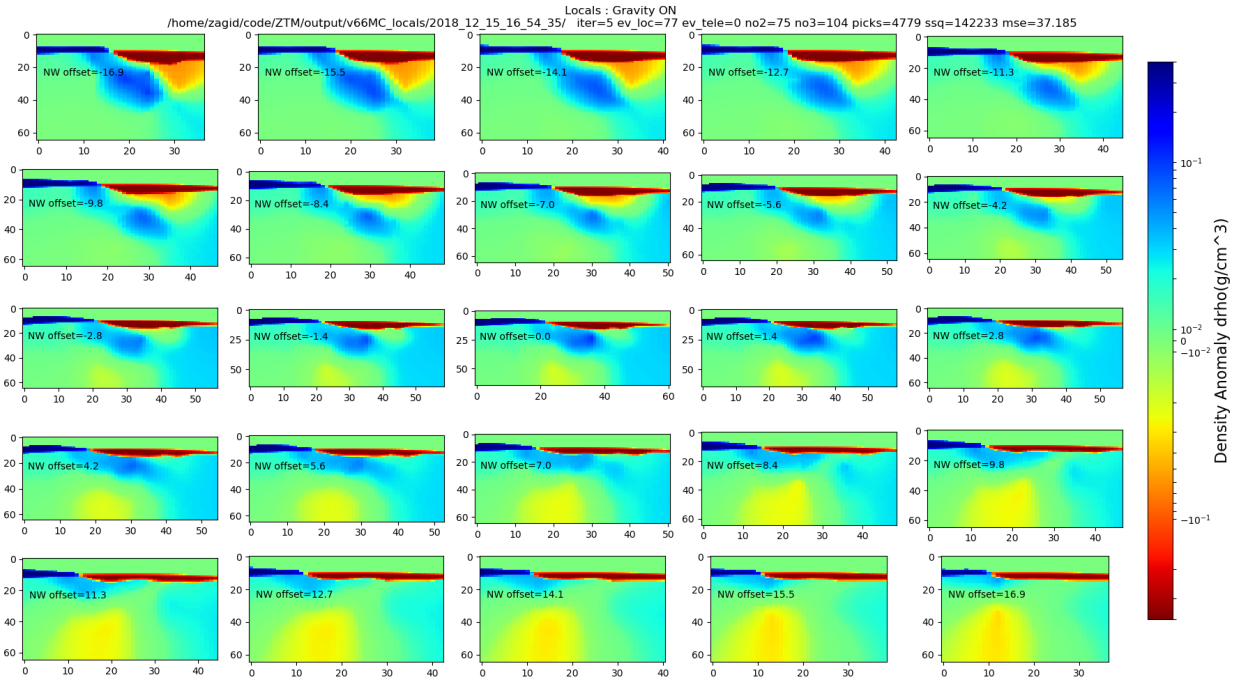


Figure 53: Density anomaly cross sections. 5th iteration. Using locals, gravity.

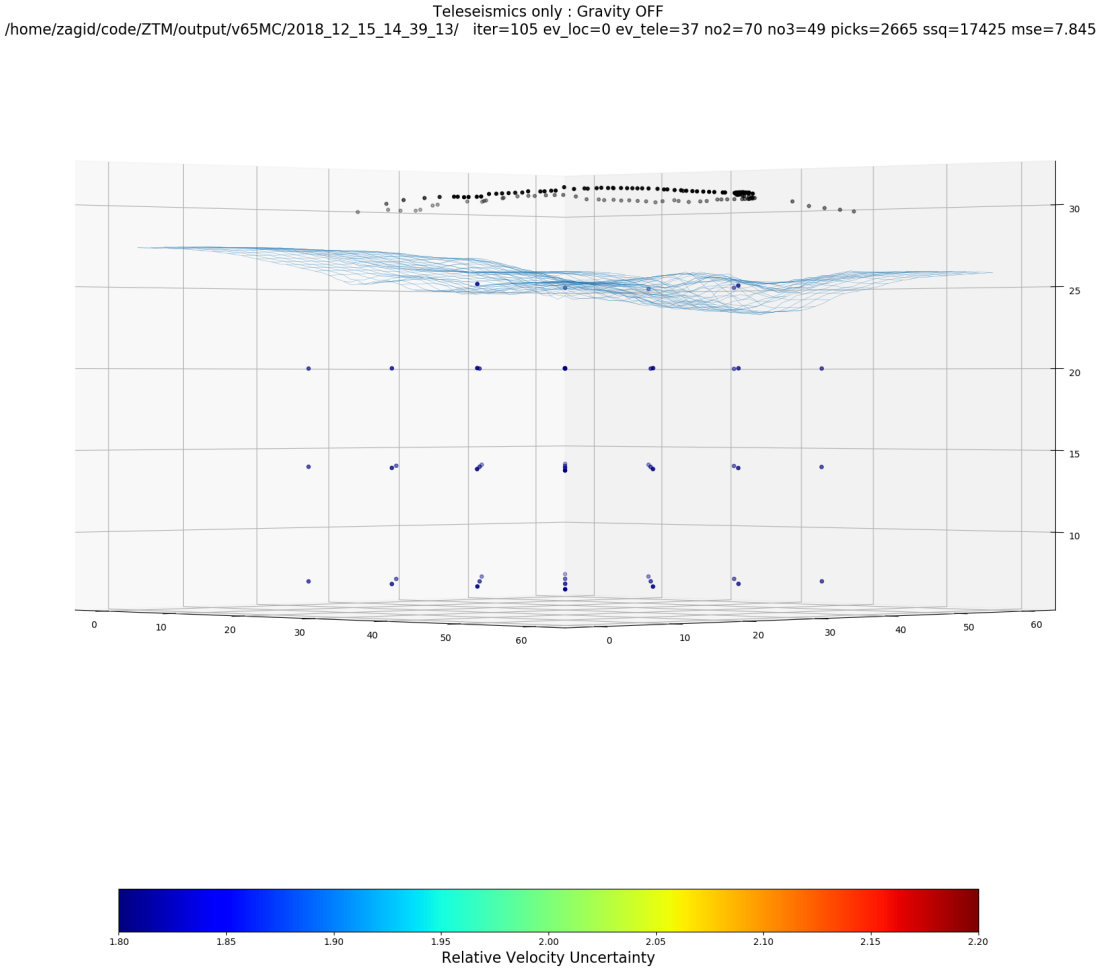


Figure 54: 3D view of relative node field uncertainty. 5th iteration. Using teleseisms.

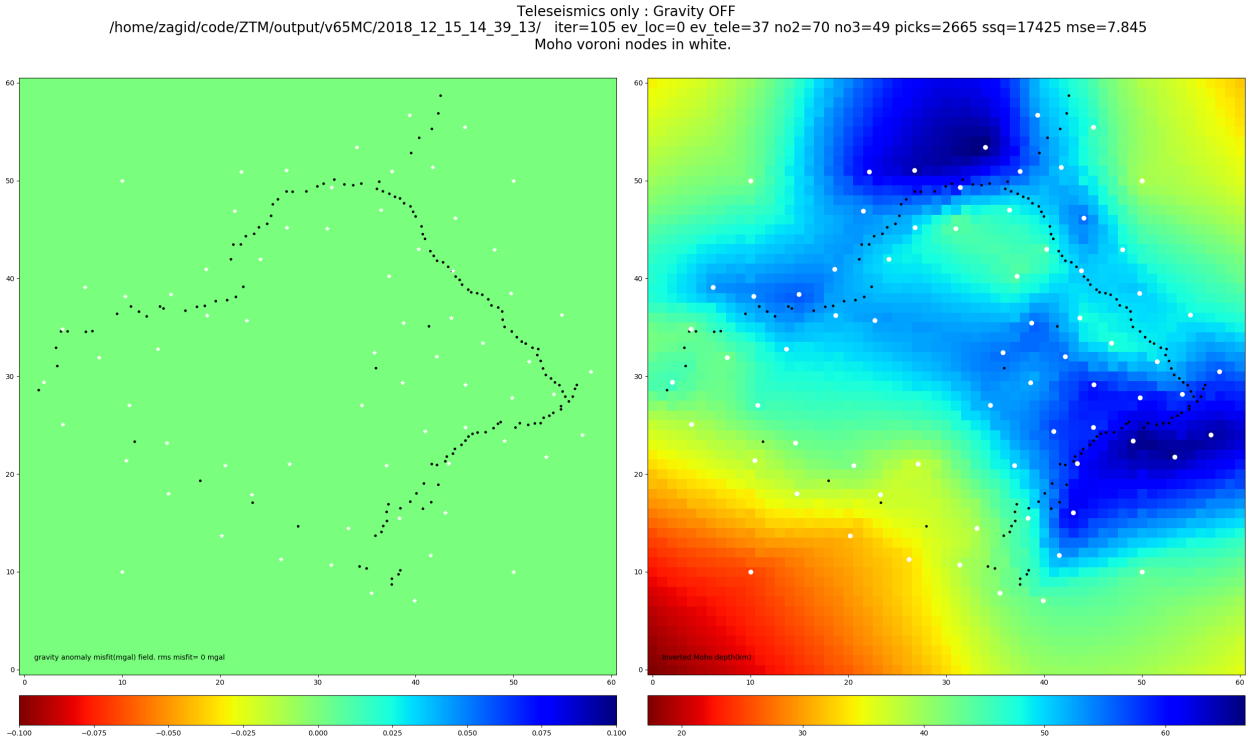


Figure 55: Top down gravity misfit(left), Moho depth(right) map. 5th iteration. Using teleseisms.

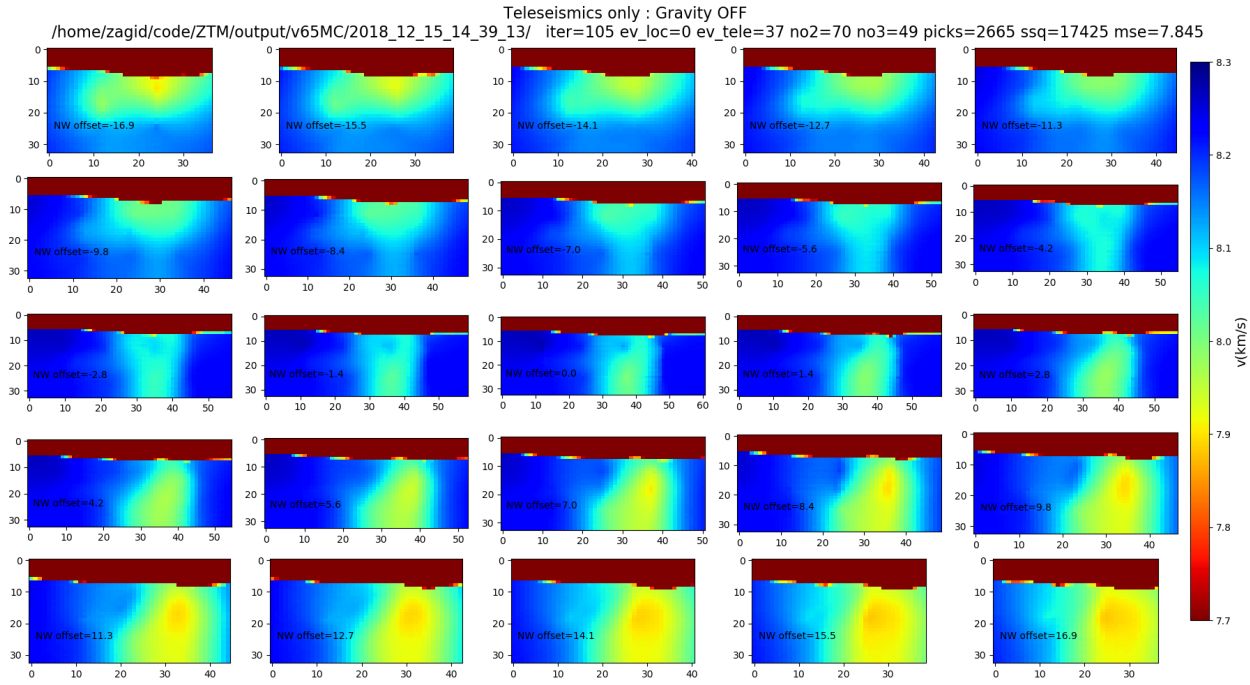


Figure 56: V_p cross sections. 5th iteration. Includes teleseisms.

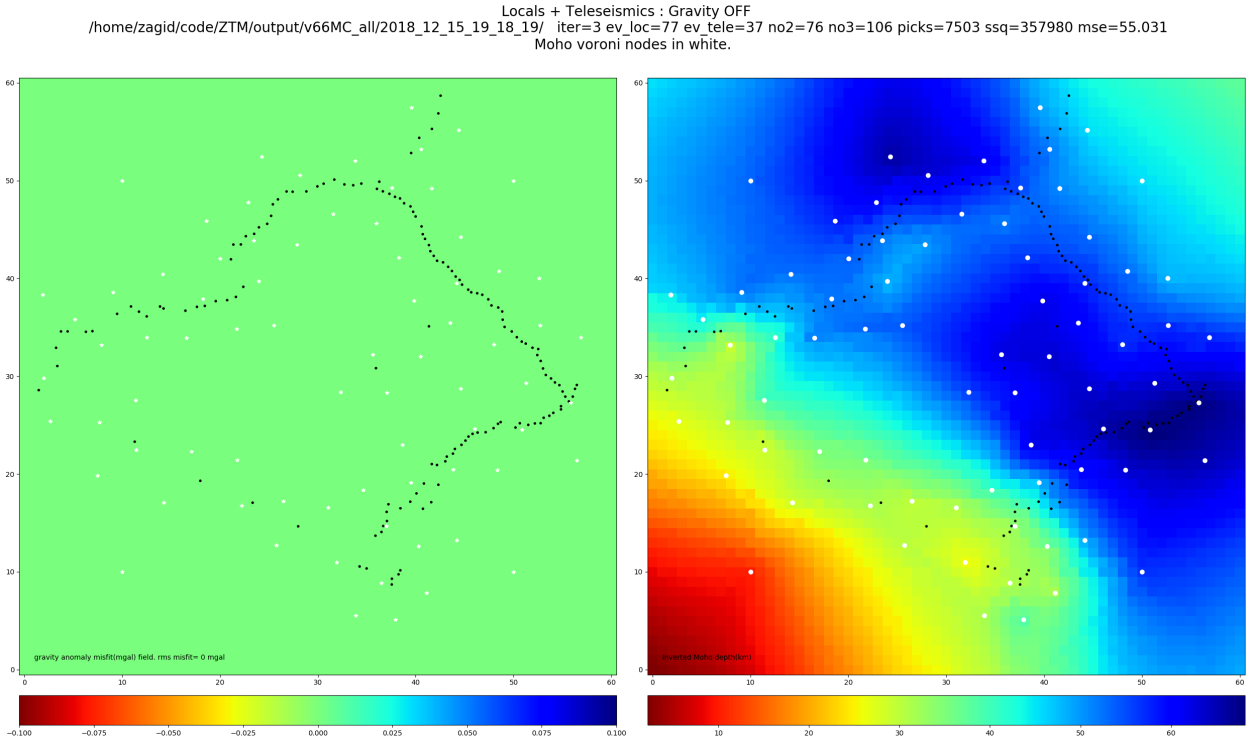


Figure 57: Top down gravity misfit(left), Moho depth(right) map. 3rd iteration. Using locals, teleseisms.

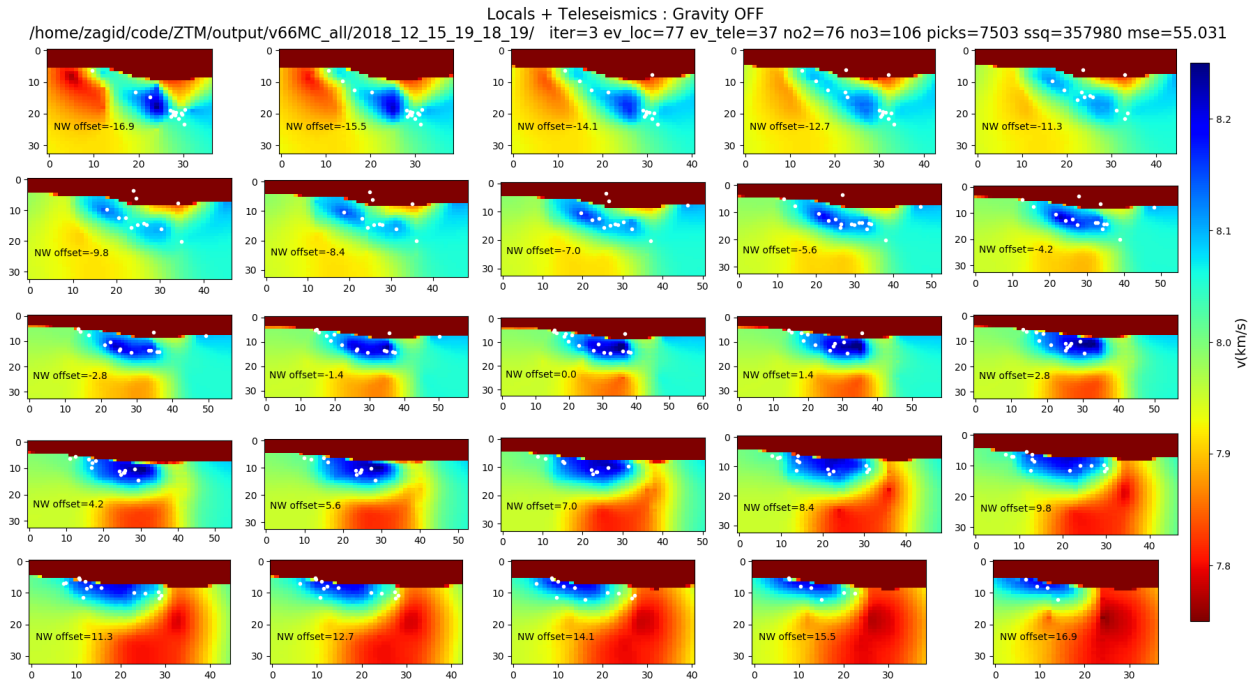


Figure 58: V_p cross sections. 3rd iteration. Using locals, teleseisms.

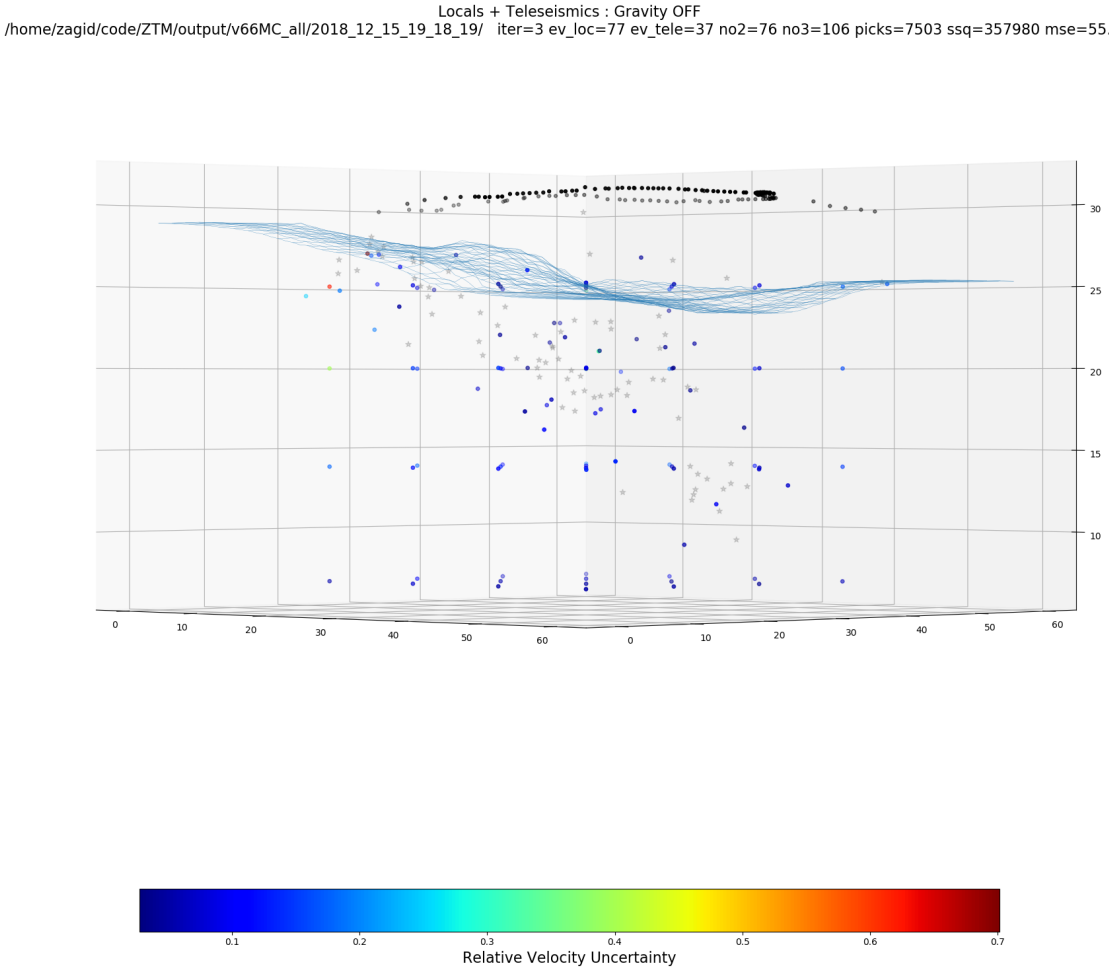


Figure 59: 3D view of relative node field uncertainty. 3rd iteration. Using locals, teleseisms.

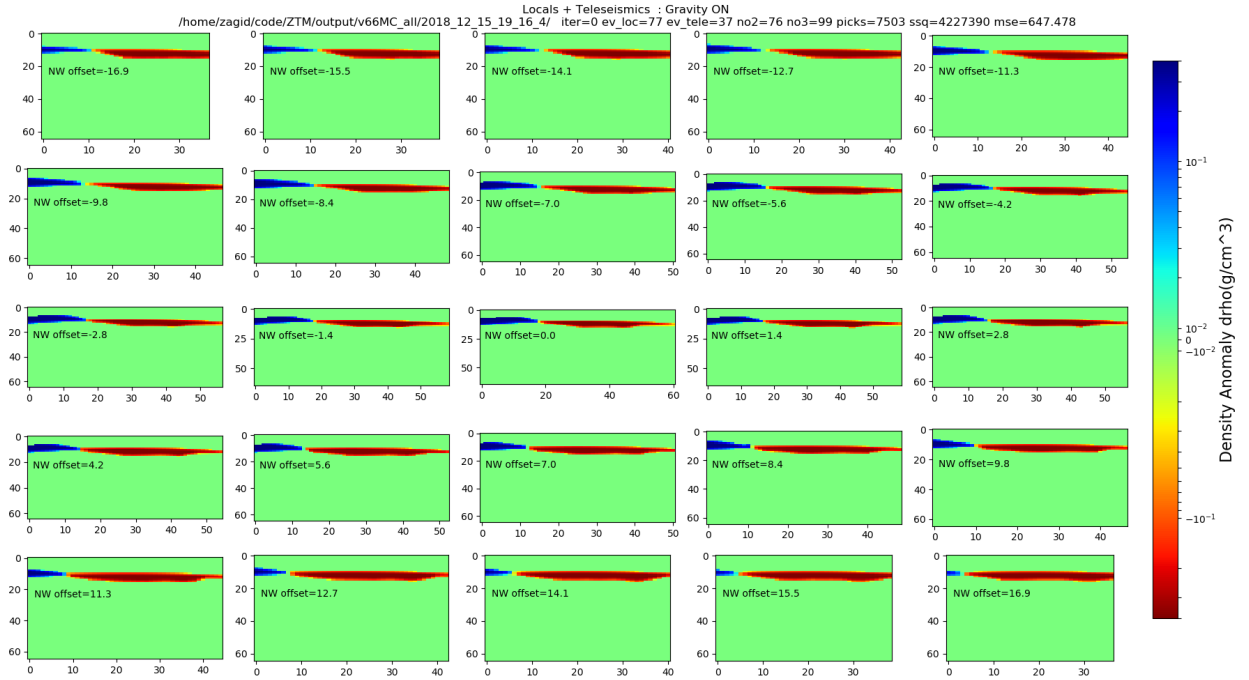


Figure 60: Density anomaly cross sections. 0th iteration(initial conditions). Using locals, teleseisms, gravity.

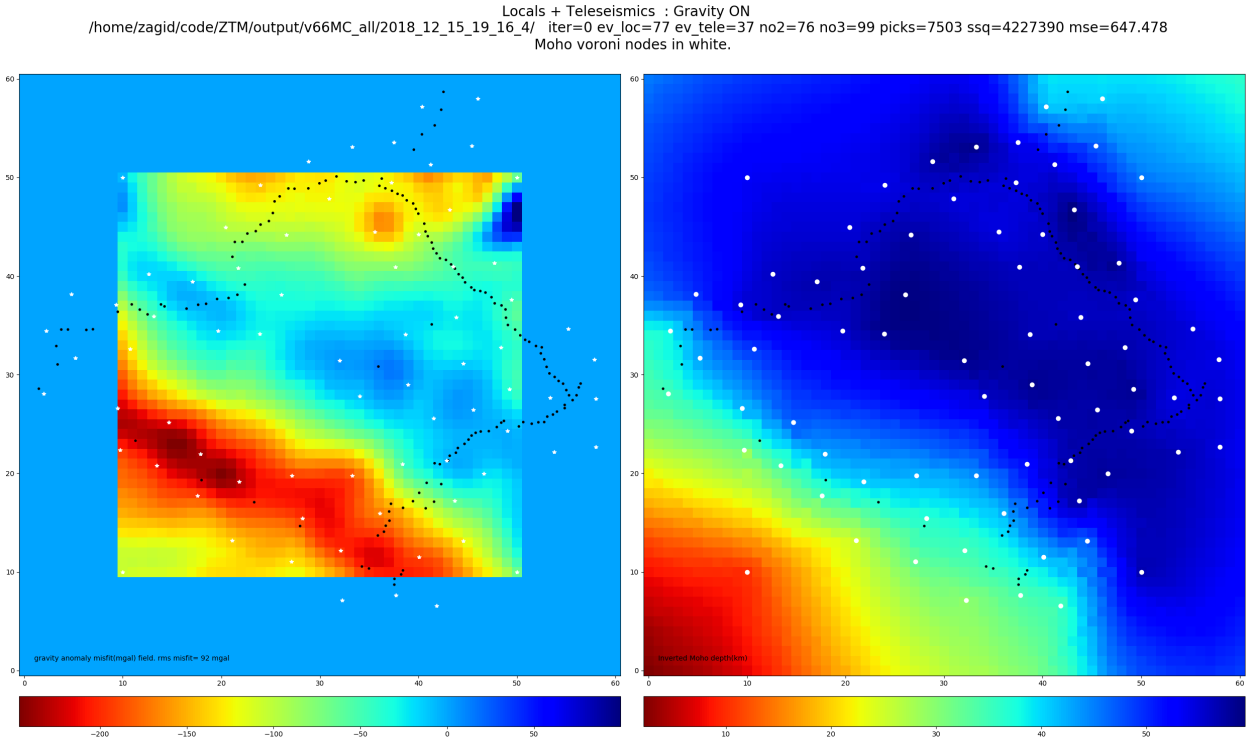


Figure 61: Top down gravity misfit(left), Moho depth(right) map. 0th iteration(initial conditions). Using locals, teleseisms, gravity.

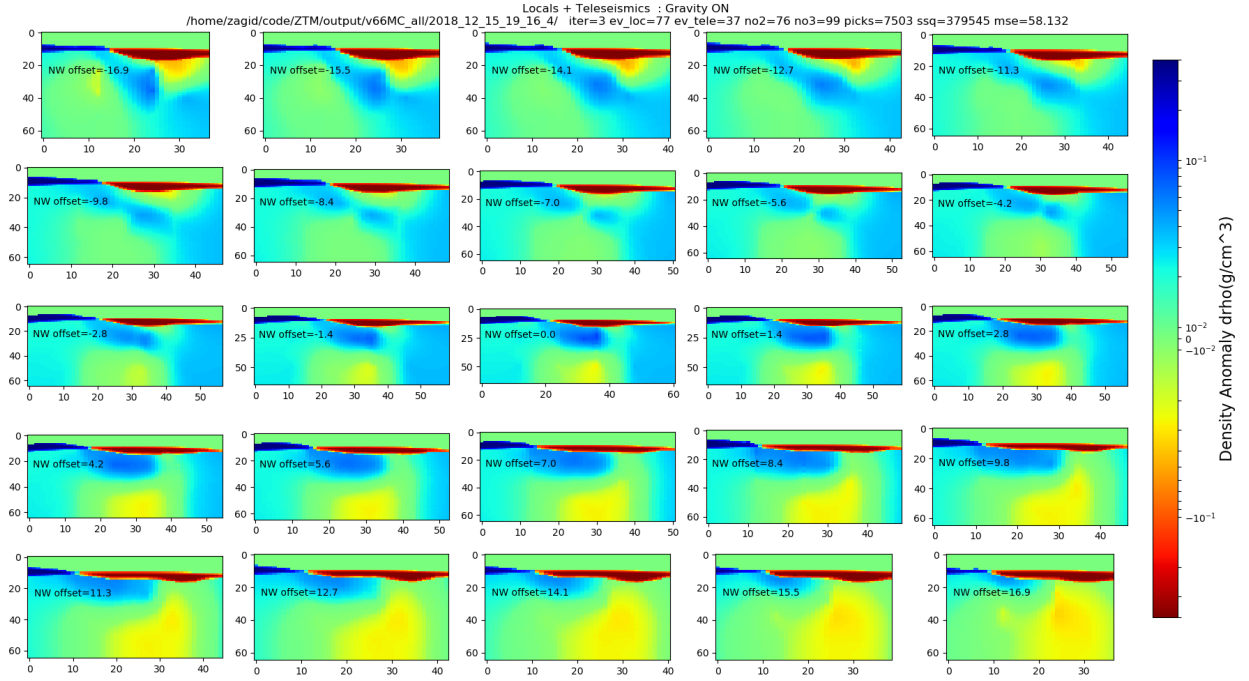


Figure 62: Density anomaly cross sections. 3rd iteration. Using locals, teleseisms, gravity.

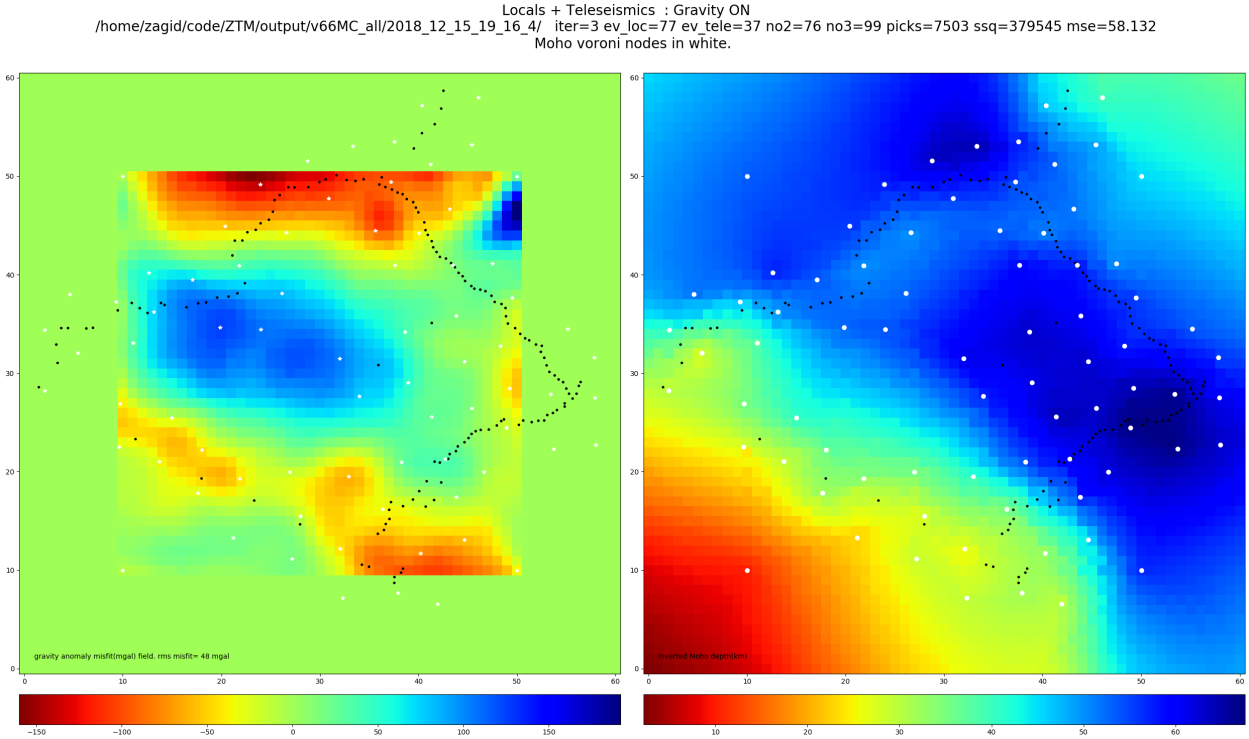


Figure 63: Top down gravity misfit(left), Moho depth(right) map. 3rd iteration. Using locals, teleseisms, gravity.

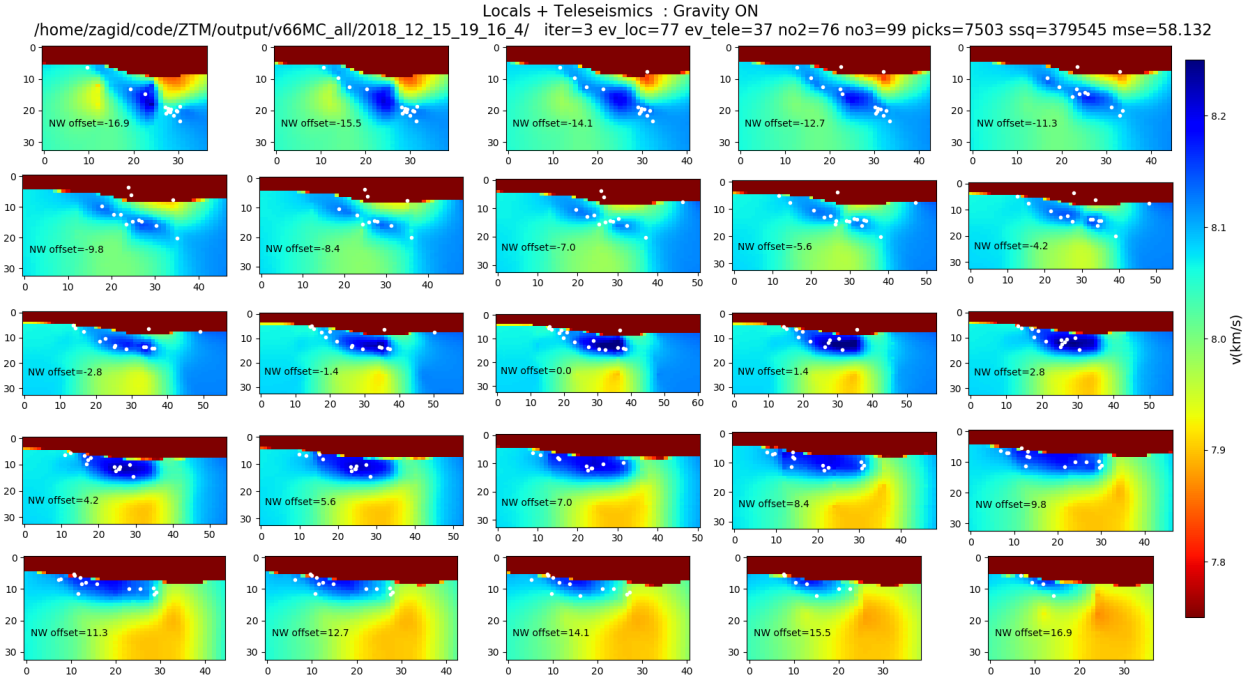


Figure 64: V_p cross sections. 3rd iteration. Using locals, teleseisms, gravity.

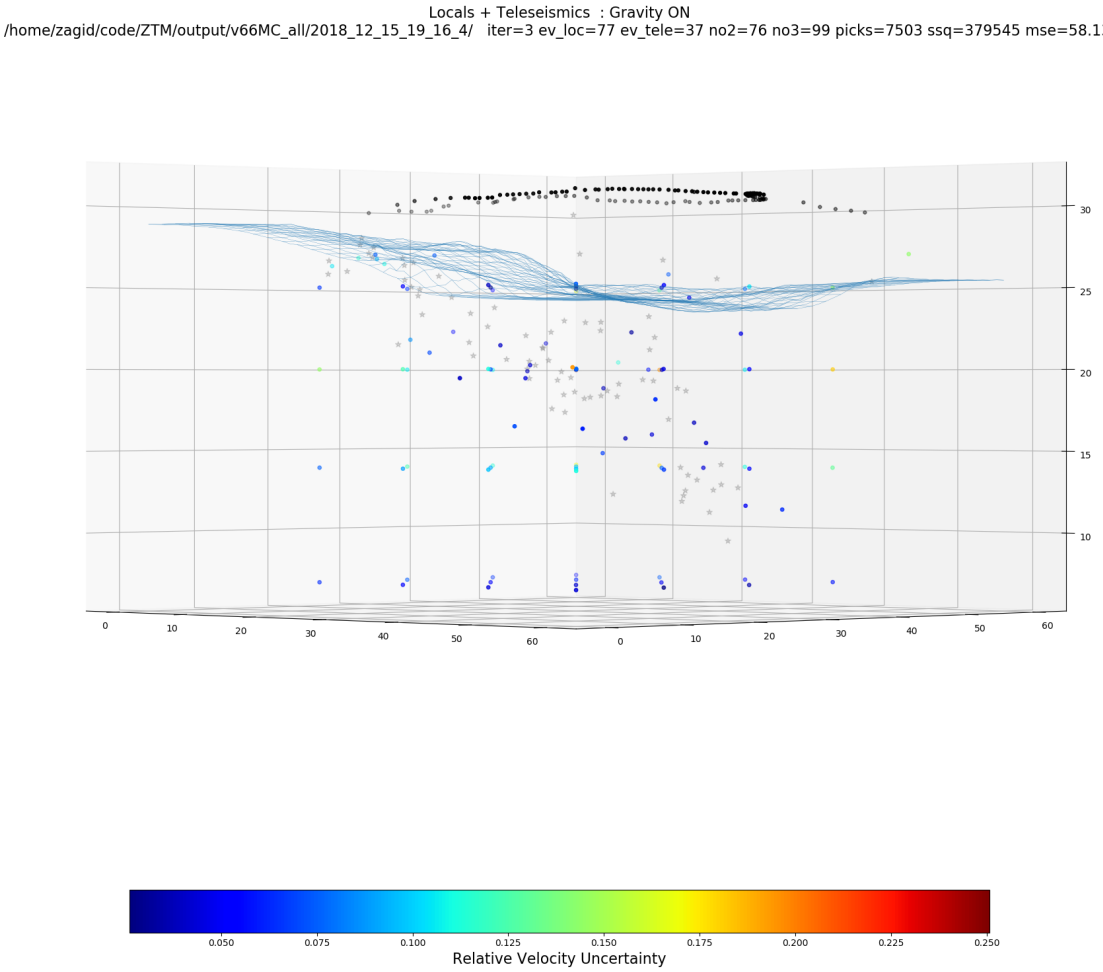


Figure 65: Flat 3D view of relative node field uncertainty. 3rd iteration. Using locals, teleseisms, gravity.

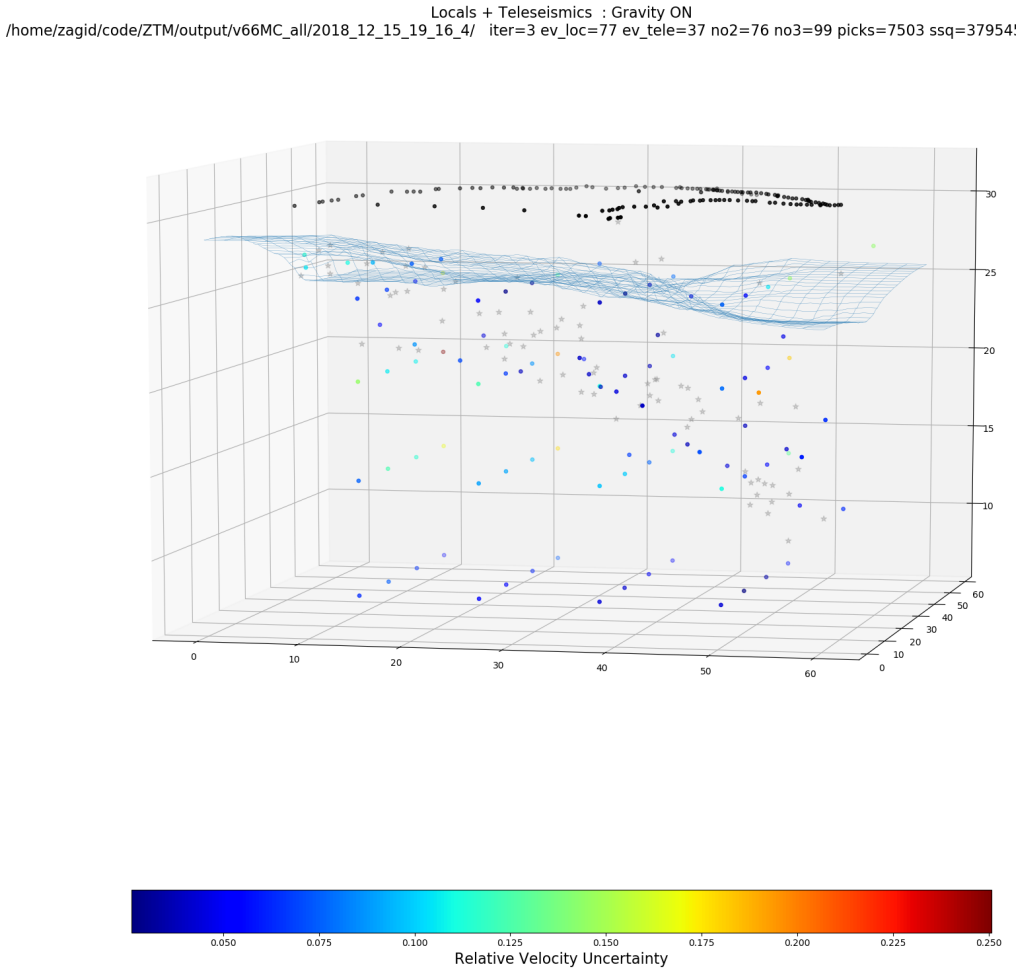


Figure 66: Tilted 3D view of relative node field uncertainty. 3rd iteration. Using locals, teleseisms, gravity.

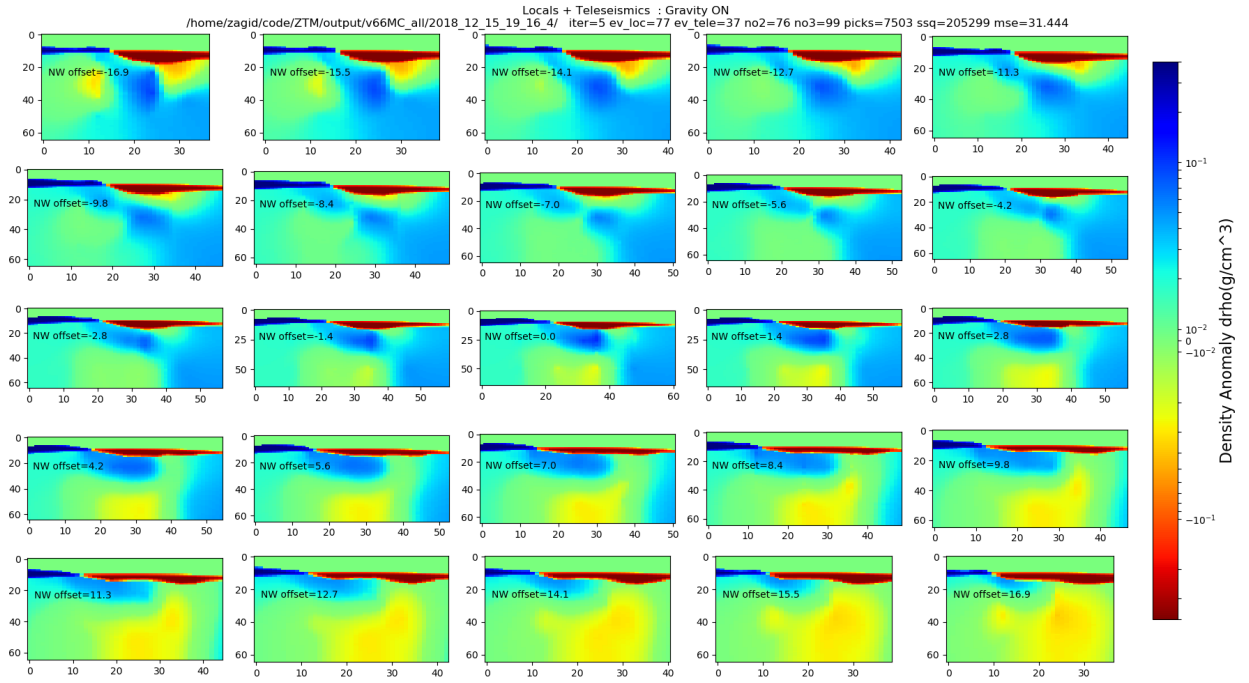


Figure 67: Density anomaly cross sections. 5th iteration. Using locals, teleseisms, gravity.

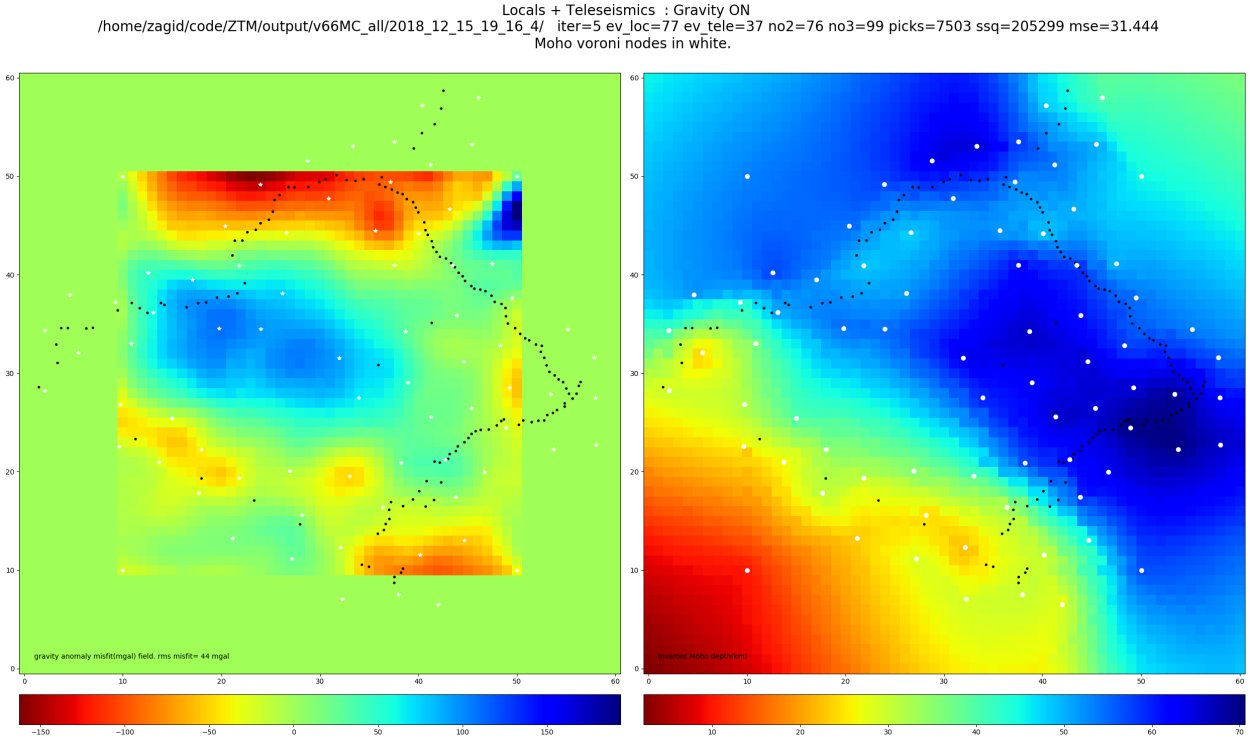


Figure 68: Top down gravity misfit(left), Moho depth(right) map. 5th iteration. Using locals, teleseisms, gravity.

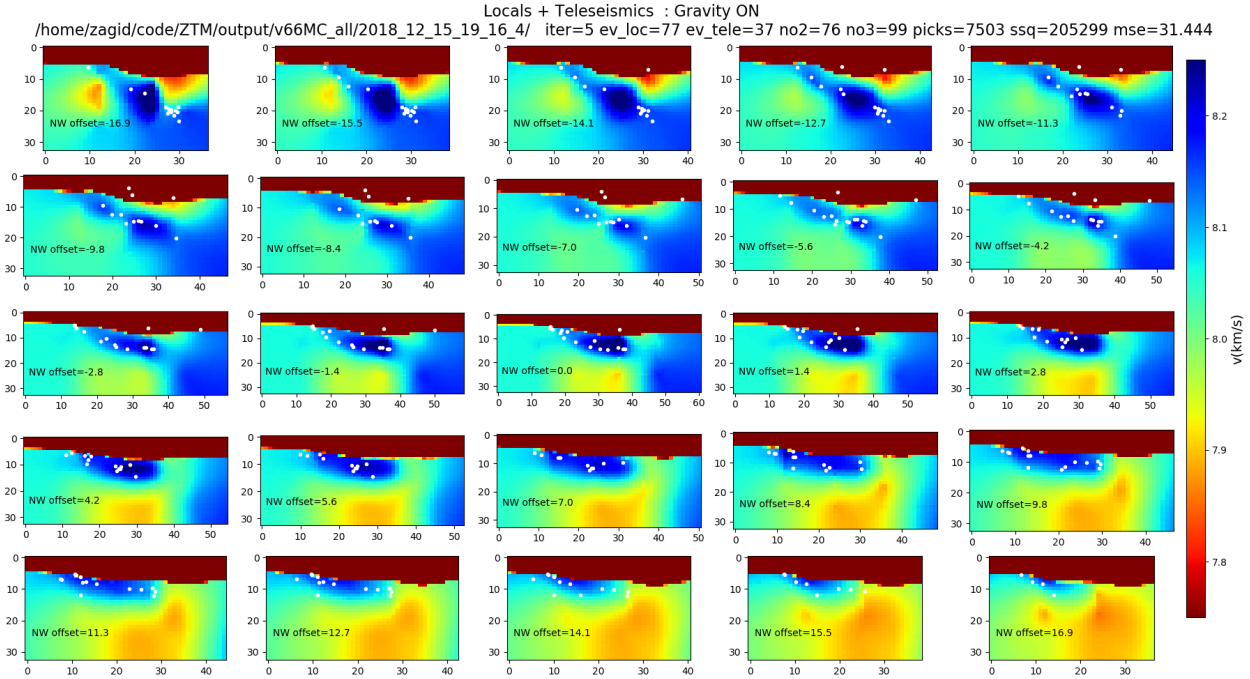


Figure 69: V_p cross sections. 5th iteration. Using locals, teleseisms, gravity.

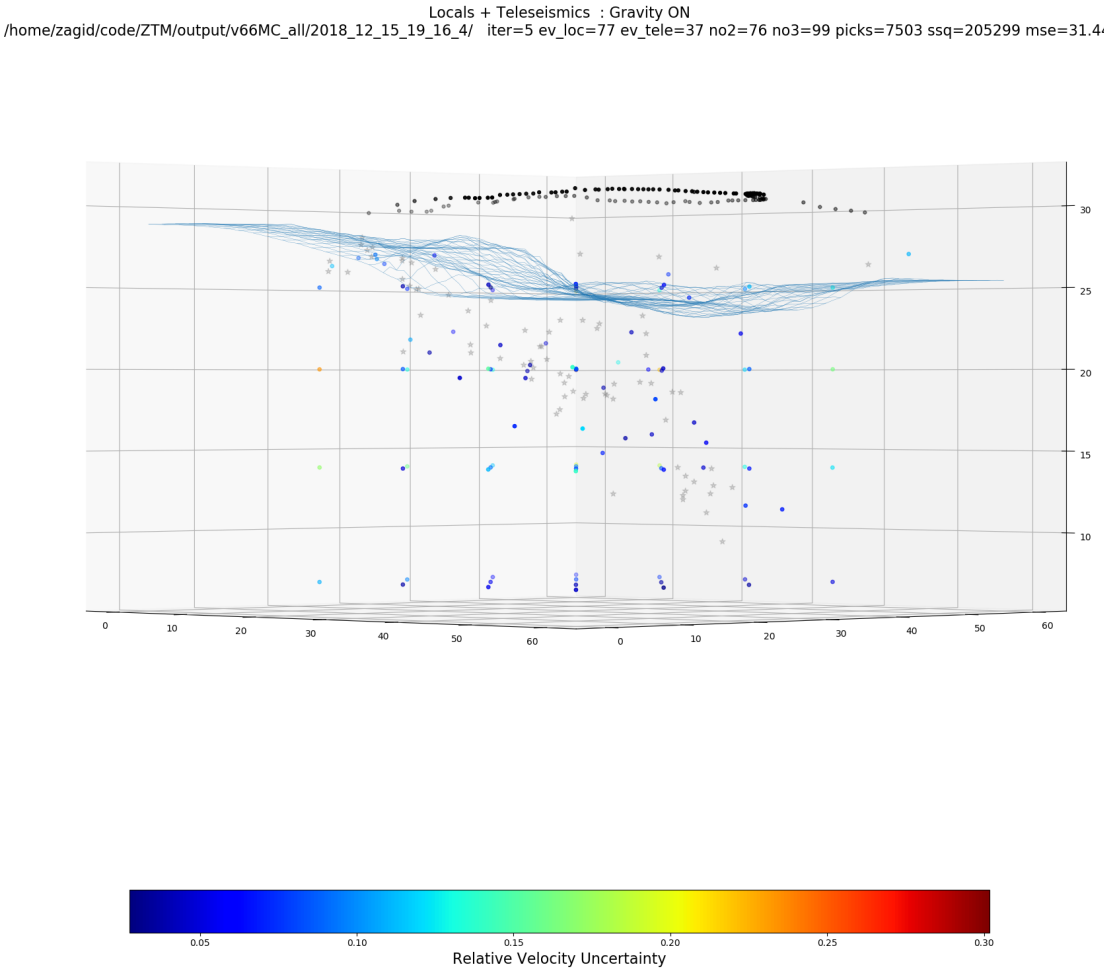


Figure 70: 3D view of relative node field uncertainty. 5th iteration. Using locals, teleseisms, gravity.

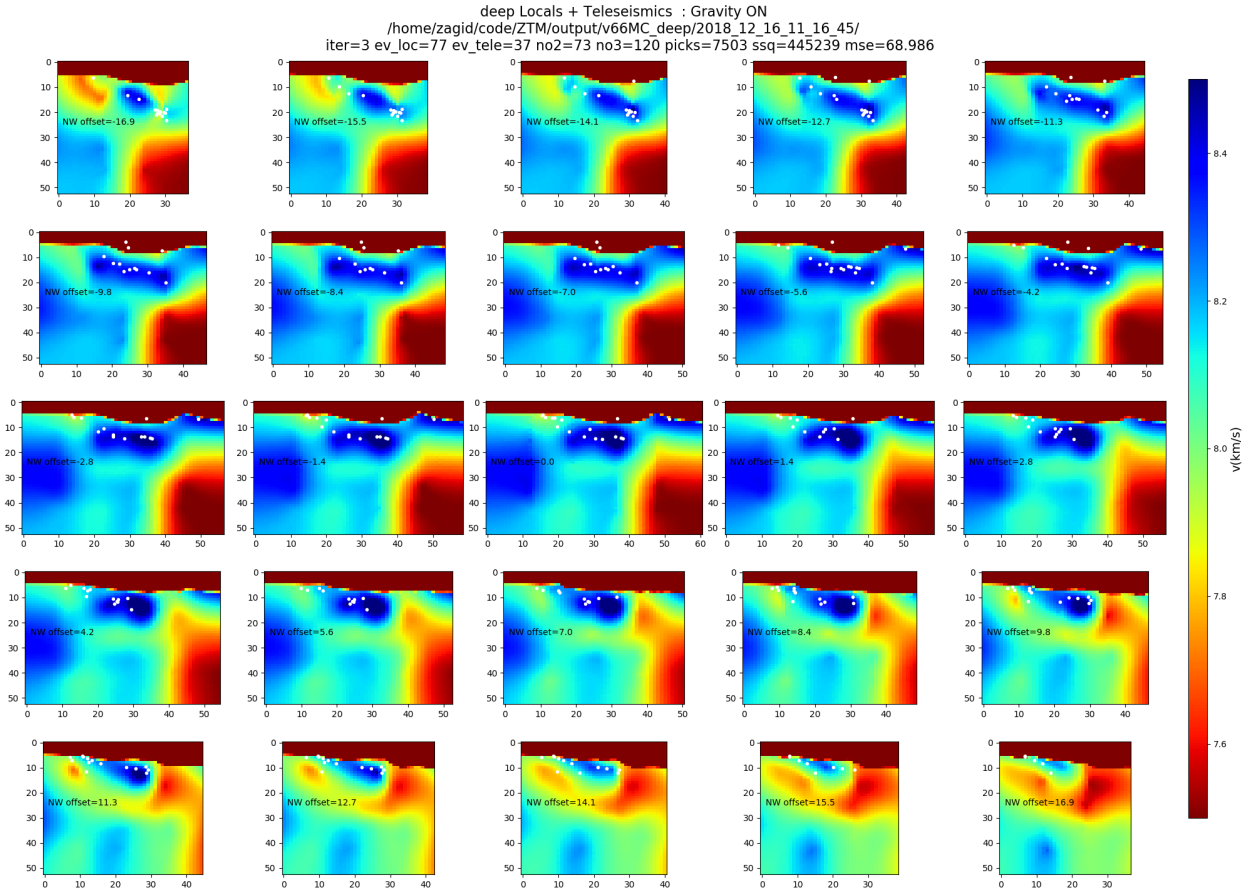


Figure 71: Deep grid V_p cross sections. 3rd iteration. Using locals, teleseisms, gravity.

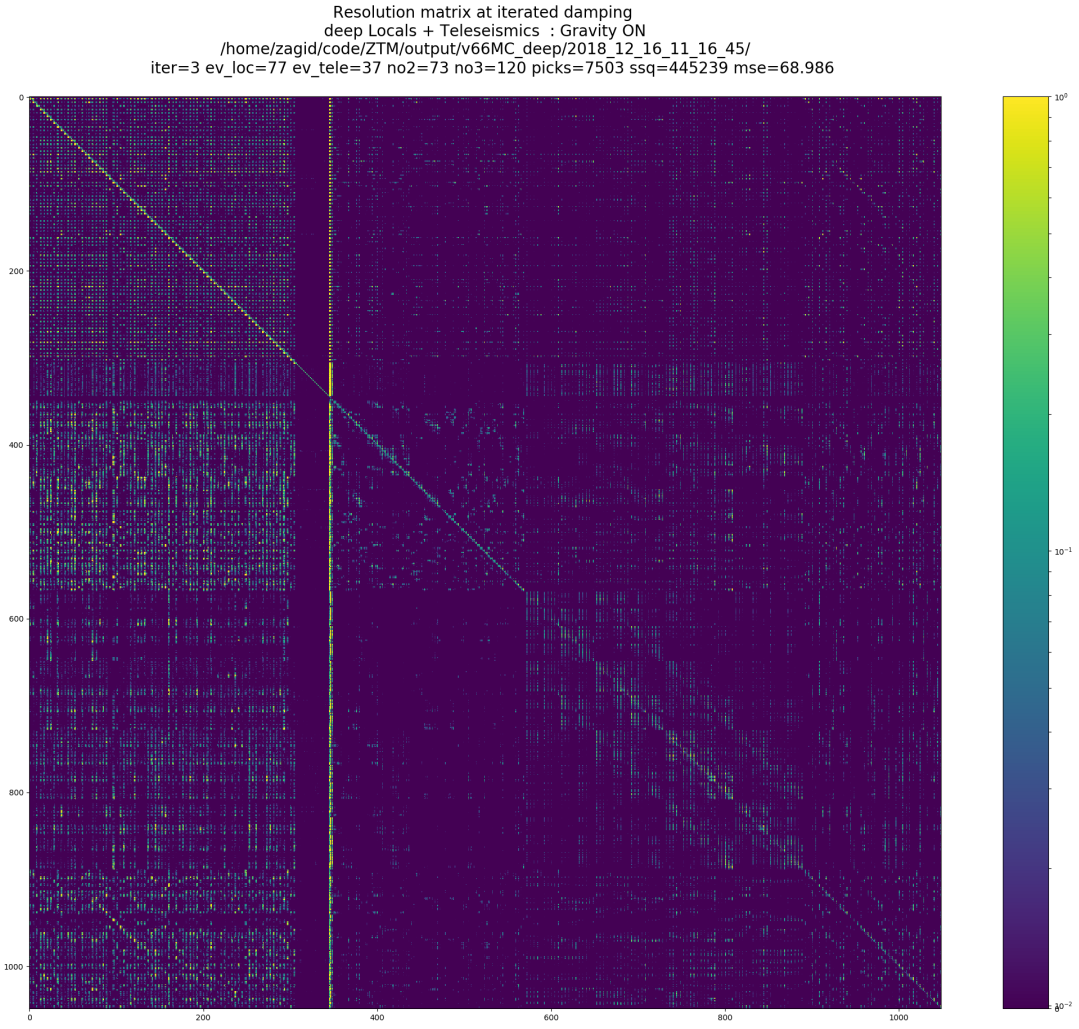


Figure 72: Sample resolution matrix of a complete inversion, assumes iterated damping.

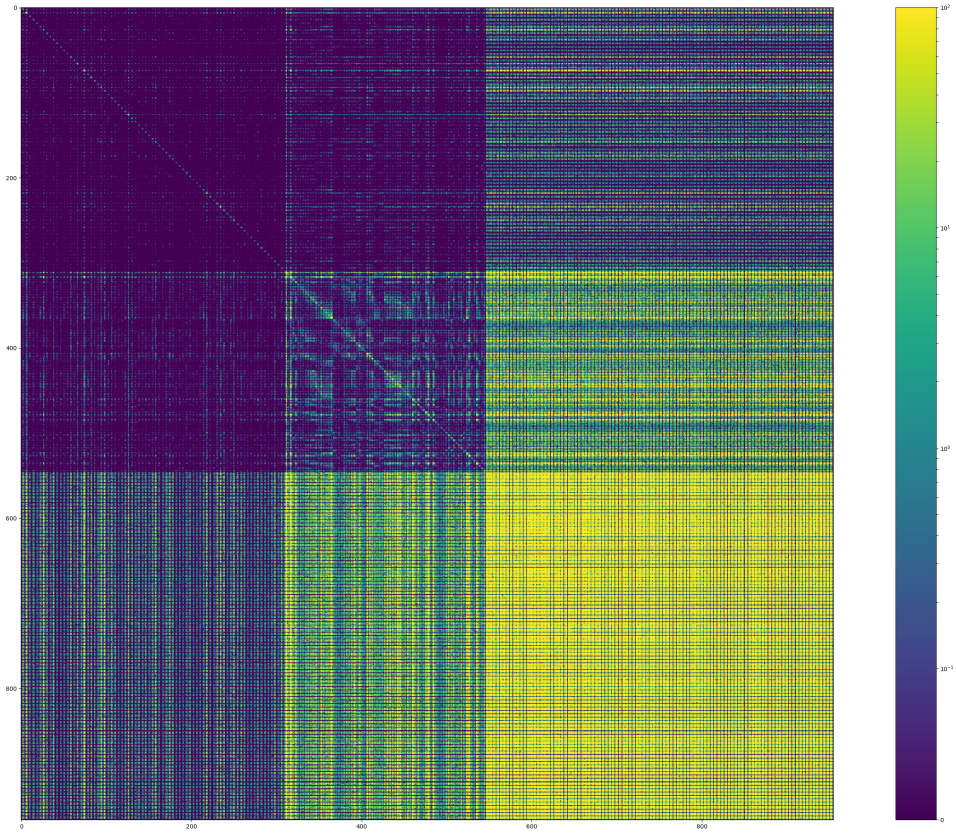


Figure 73: Sample covariance matrix (not normalized) of a local event inversion, assumes damping $D = C_m^{-1}$.

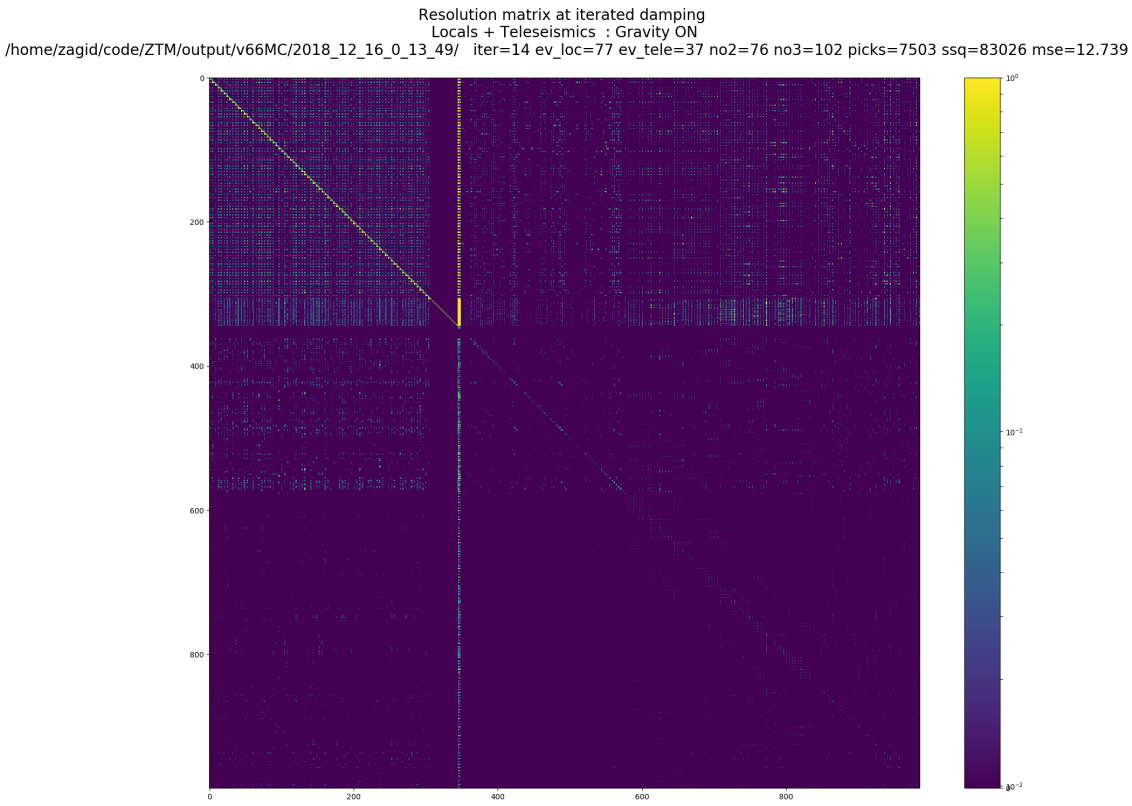


Figure 74: Sample resolution matrix of a complete inversion, assumes iterated damping.

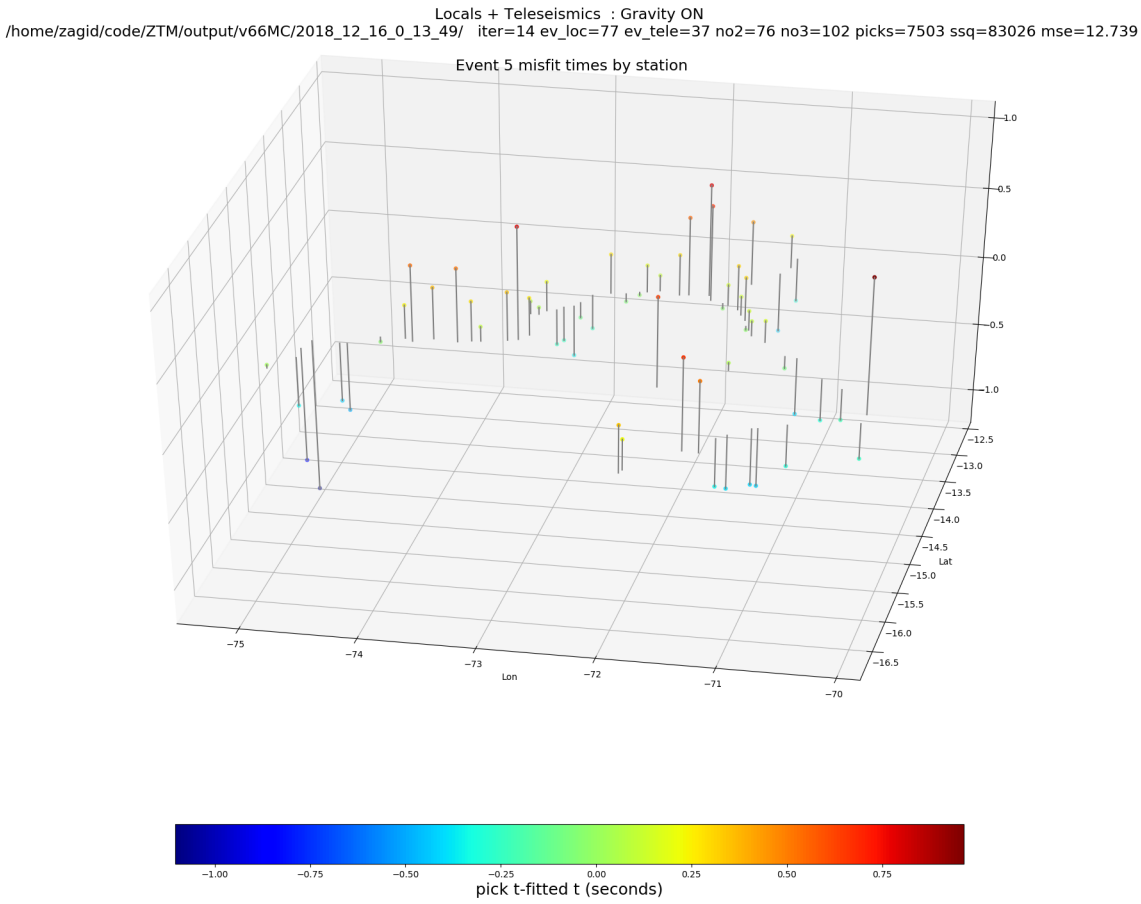


Figure 75: Pick-fitted time residuals by station for an inverted local event.

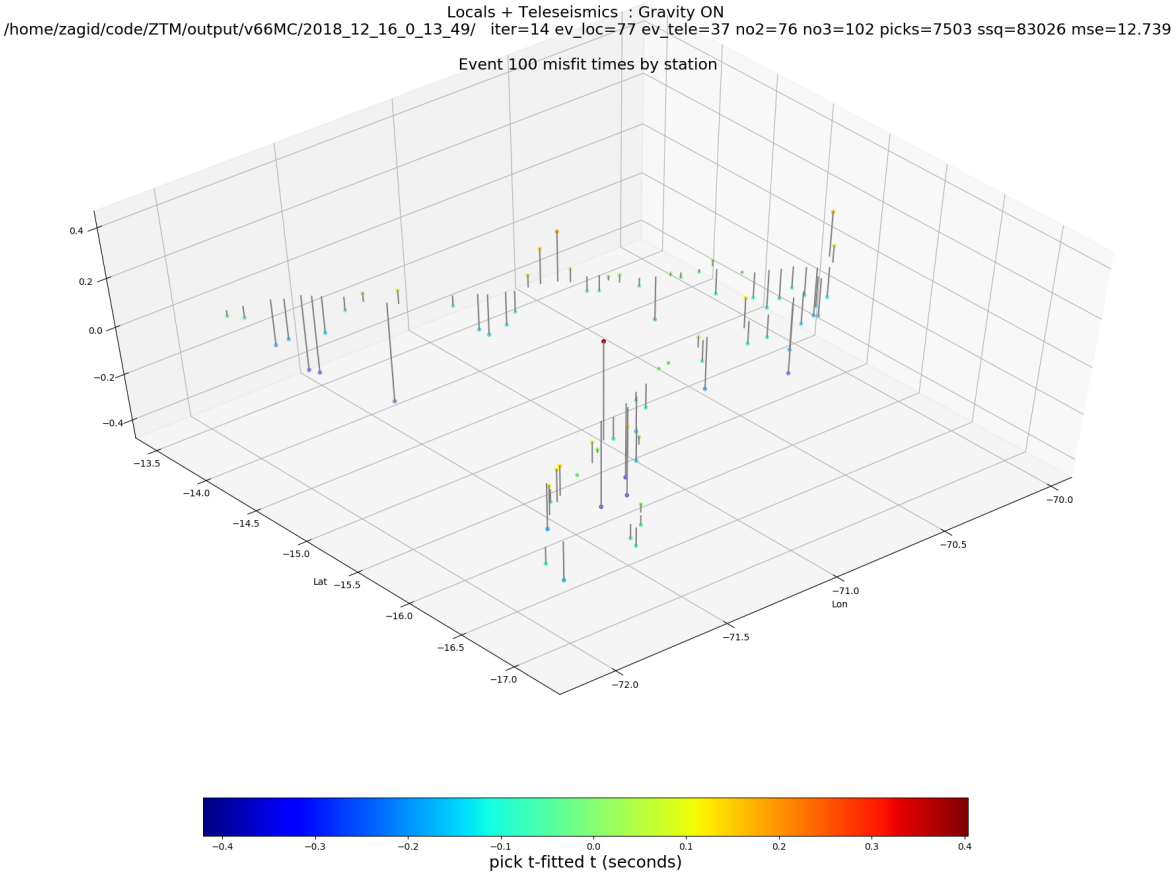


Figure 76: Pick-fitted time residuals by station for an inverted teleseismic event.

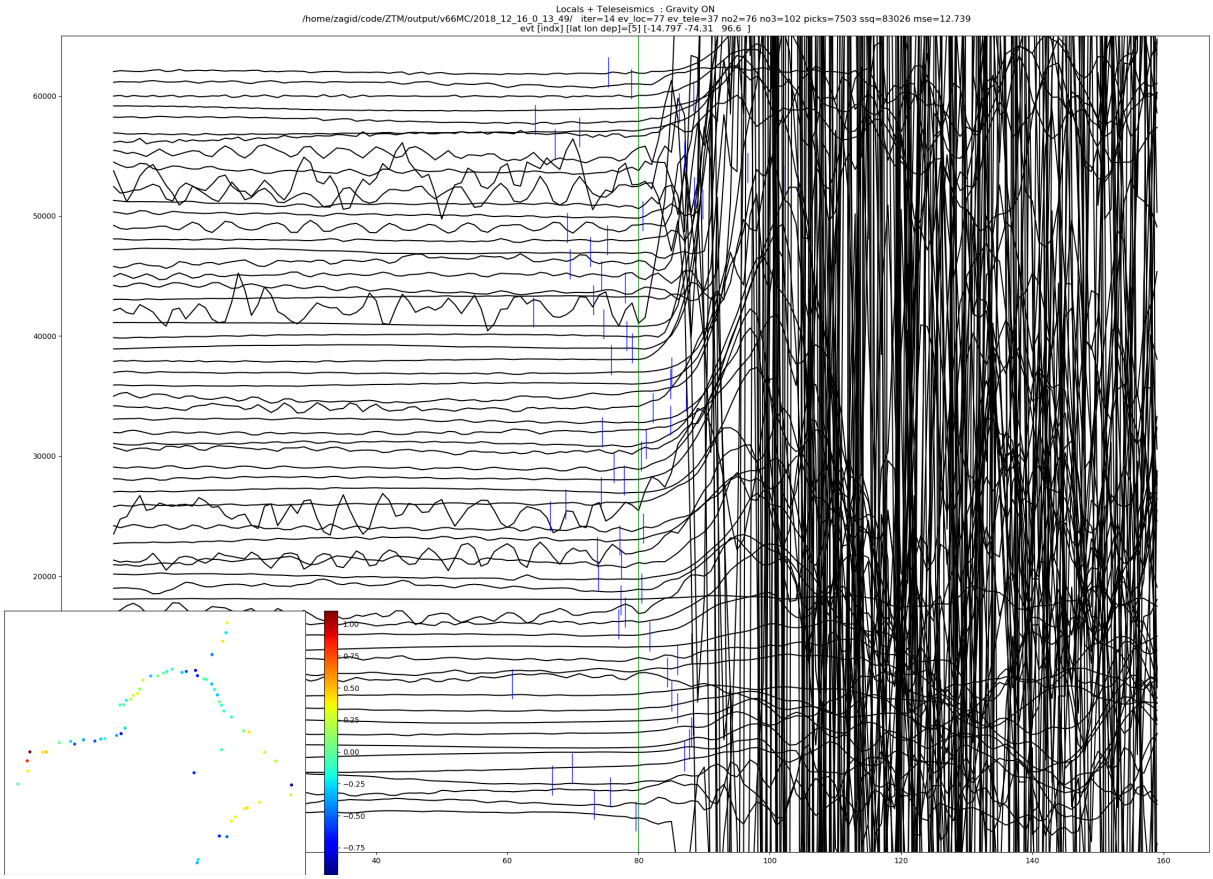


Figure 77: Traces picked for local event aligned by pick sample(green); inverted model time in blue.

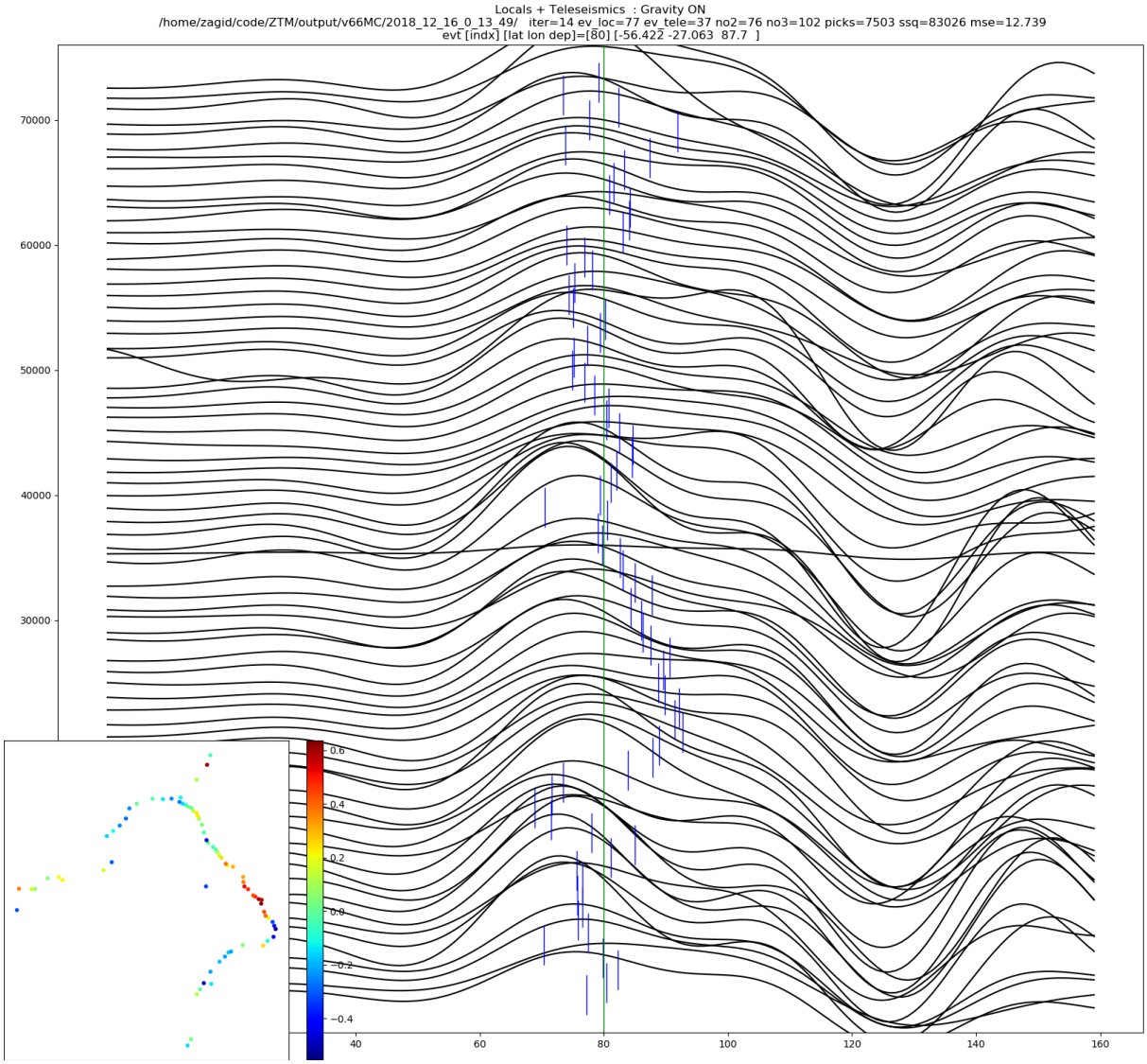


Figure 78: Traces picked for teleseismic event aligned by pick sample(green); inverted model time in blue.

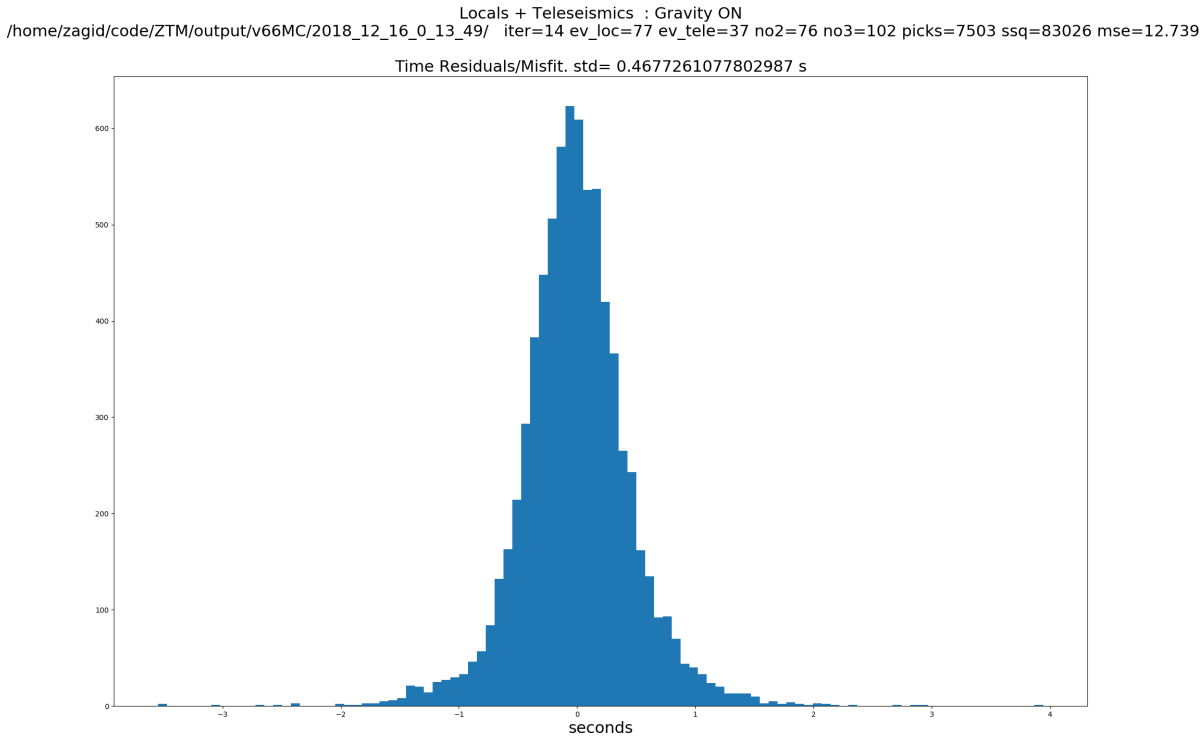


Figure 79: Residual/misfit histogram for sample inverted data set using heavy damping.

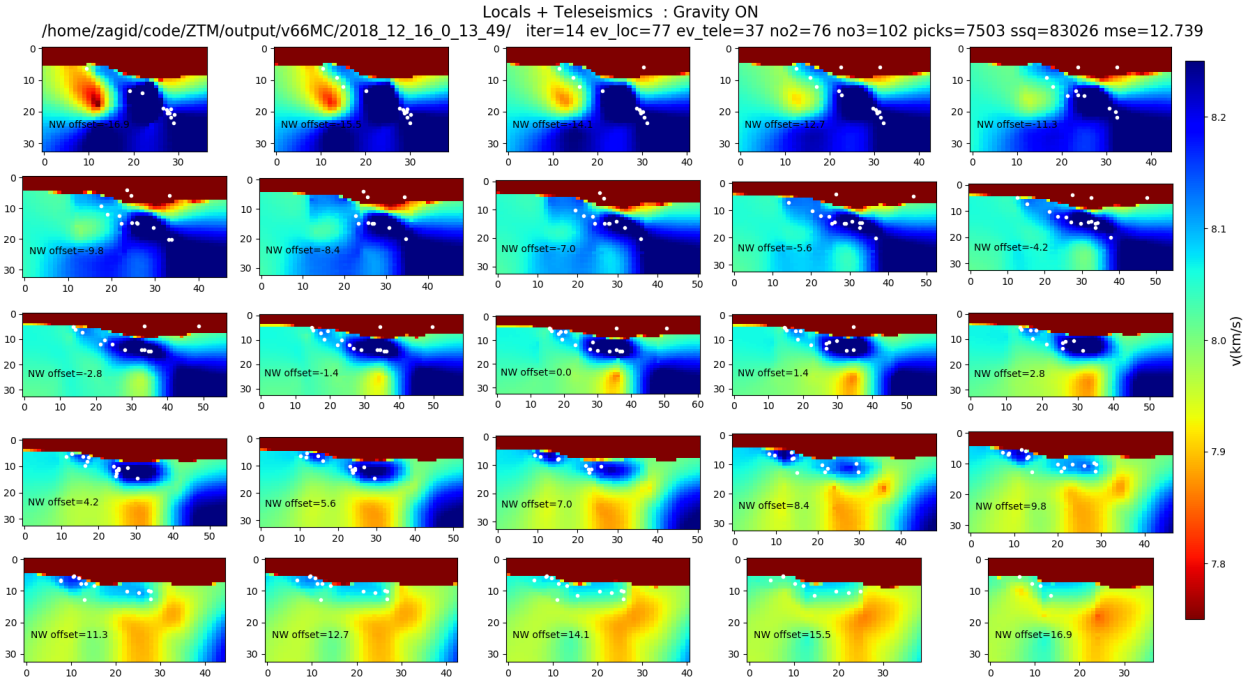


Figure 80: Waveform reference sample inversion velocity field.

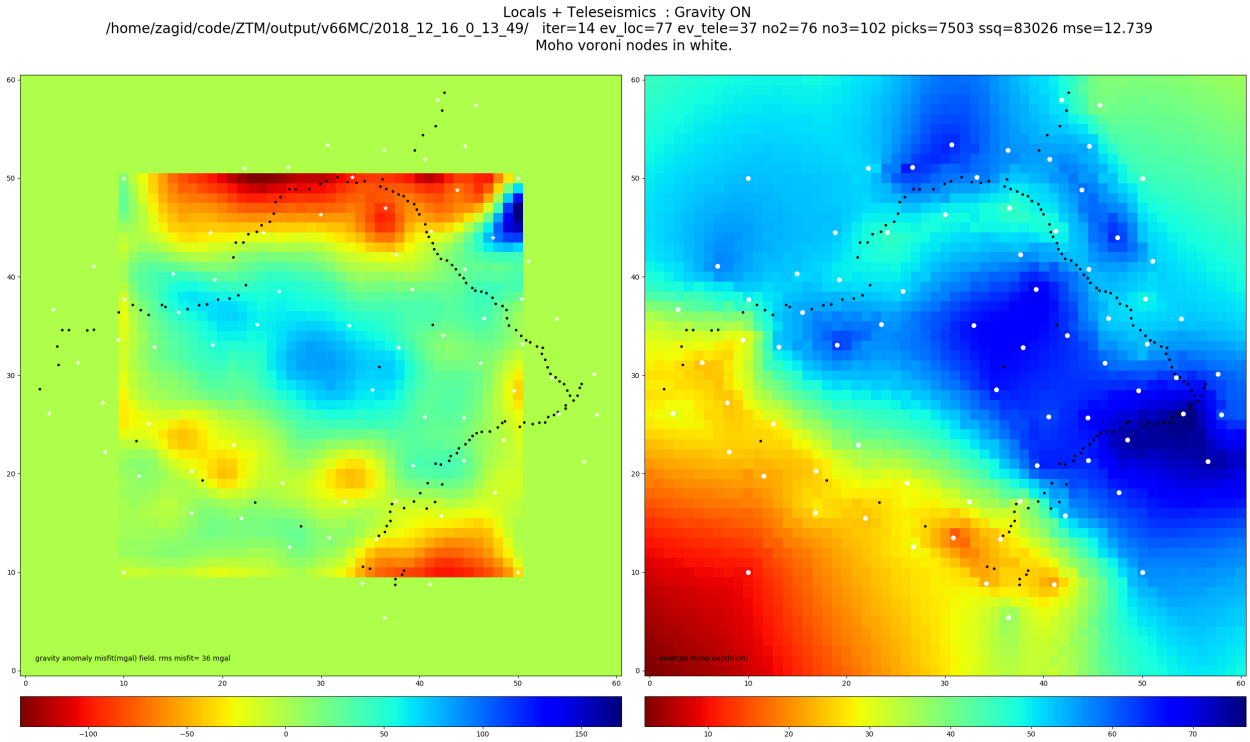


Figure 81: Inversion gravity misfit and Moho depth mapping corresponding to traces in figures 77-78.

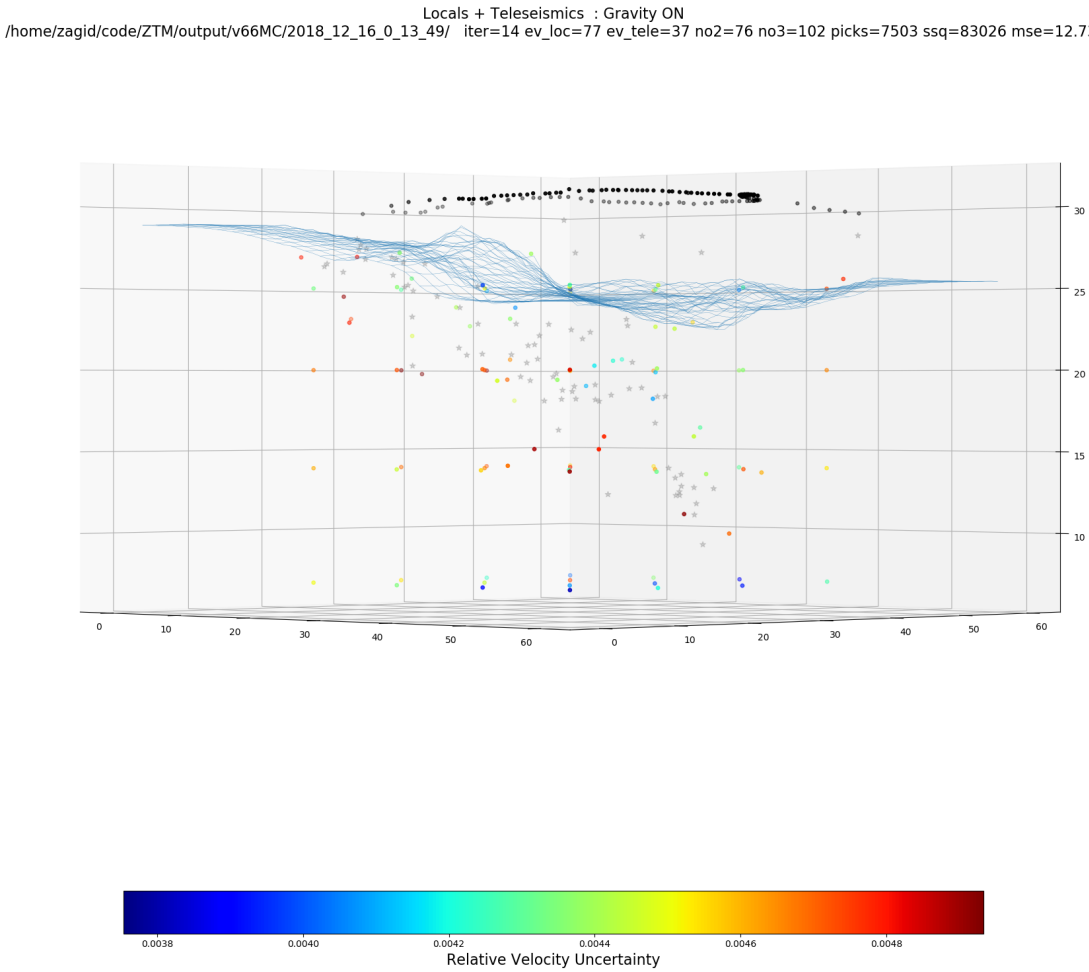


Figure 82: Inversion estimated velocity uncertainty field corresponding to traces in figures 77-78. Note: Iterated damping was used for this particular parameter covariance calculation, while apriori covariance matrix was used for previous uncertainty calculations.

Figures 35 - 73 are results using different combinations of available data types, including local source seismic, teleseismic, and gravity data, presented at different iterations. They are presented individually to demonstrate general tendencies with respect to data types and inversion iterations used. The general interpretation of results is presented in "Conclusions".

Figures 74 - 82 are all associated with one particular complete inversion that was allowed to run for 14 iterations at the time of its analysis. Figures 75 and 76 are simple misfit point plots showing a geographical distribution of misfit. Figures 77 and 78 show all picked traces from one representative local event and one representative teleseismic event lined up by polar angle around network center reference with wave forms centered at their pick sample. Inverted pick samples for each trace are marked in blue.

Figures 73 and 74 are estimated parameter covariance, and resolution matrix respectively. It should be noted that the damping terms used to obtain each were selected to reduce bias toward misleadingly good results. In the case of C_m, it involved using the apriori C_m^{-1} as the damping term, whereas for Figure 74, we used the damping vector used in its latest step. We discuss our reasoning behind this decision in Conclusions.

11 Conclusions

In summary, we believe that we have created the first approach of its kind to incorporate local and global seismicity, gravitational data, and local hypocenter relocation in order to efficiently image mid to upper mantle and Moho structure anomalies using a robust forward model and a parametrically efficient velocity field parametrization characteristic of an irregular Moho with a subducting slab.

In the course of conducting and presenting this work, we have concluded that our developed inversion package appears to perform in a manner consistent with the way it was designed to operate, with no apparent failure modes seen in the physically representative parameter space we tested and operated it in. Additionally, we found the package to demonstrate robustness of some significance in its tendency to converge toward roughly similar features across a varying range of not only starting parameters, but also model types and parametrizations. We found that our field parametrization basis functions performed in a manner that allowed for their parameters to be invertible, despite Discrete Sibson Interpolation having known quantized coordinate response with respect to its parameters.

While typically discontinuous, or quantized, response to inputs would make a model unsuitable for gradient inversion, we were able to demonstrate reasonable convergence in our tests by accounting the scale of quantization noise in evaluation of Jacobian elements. We also believe that this may be one of the central contributing factors to the consistent tendency of converging inversions run to require increasing damping in order to keep reducing misfit with iterations, which would be amplified even more in scenarios using one-sided Jacobians, as in our case.

We have reaffirmed that there is no simple solution for regularization, and that optimization of damping conditions remains a highly parametrization and input dependent problem.

Additionally, we have recognized the need to weigh model smoothness in a performance evaluation of inversions. In other terms, we found emergence of local small scale instabilities and features in many of our inversions that are not consistent with large scale geophysics. The use of a mixed damping scheme, using a linear combination of scaled estimated parametric misfit curvature as in Levenberg Marquardt, and a scaled inverse of the apriori parameter covariance matrix performed reasonably in trading off acceptable convergence rates against seeking to preserve parametric smoothness of inverted solutions for up to roughly five inversion iterations.

We believe that an additional potential reason for the need to use excessively high damping to reduce misfit even at asymptotic convergence, may be due to a very small subset of ill conditioned parameters that have high covariance with a significant portion of the invertible parameters, as visible in some of the rows and column elements of the computed covariance matrix represented in Figure 73. They are represented in our visualization of parameter field uncertainties. Additionally, we found that the covariance and non-linearity ranges for Voronoi node positional parameters are much larger than those of the nodes' field parameters, which have a quasilinear response to first order perturbations, and are generally well behaved and have either continuous or nearly continuous input response. For this reason, it may be worthwhile to introduce parameter type specific regularization scaling factors, or even reclassifying inverted Voronoi node positions to be fixed after a set degree of convergence or iteration count.

Examining misfit statistics in Figures 75 - 79, while the full time misfit set distribution is Gaussian or nearly Gaussian, we see clear patterns of localized clustering and monotonicity in misfit, indicating an underfitted model. While in many cases, one aims to solve a critically or near critically constrained problem, in practical inversions, we find under-fitting to be favorable to over-fitting, as under-fitting generally has a smaller effect on loss of model smoothness and fewer opportunities for the emergence of small scale local minima.

Additionally, by employing a gravitational inversion, we found it to be useful in constraining Moho depth to physical ranges, which is especially helpful for resolving Moho depth in areas where it is not proximal to reporting stations. However, gravitational inversion introduces its own potential errors, especially in areas where Bouguer corrections may have been made with wrong assumptions, such as sedimentary basins or areas with extremely rough terrain. By controlling the relative weight assigned to each type of data, we can try to find a ratio where the sum of back propagated errors from the data sets is minimized, excluding non-physical or solutions deemed too far from prior knowledge to be realistic.

Having initially set a goal of carrying out seismic imaging to the maximal practical depth, we found that the practical imaging depth limit for the PeruSE appears to be somewhat lower than anticipated, with most features we believe to have resolved washing out at depths greater than 200km (Figure 71). This is due in large part to the sparseness of our sensor array and the relatively small volume of invertible data, with less than 10^4 usable first arrivals with certain regions of our grid traversed by very few of the first arrivals used. Despite this, our study has led us to conclude with a reasonably high degree of confidence, that we have imaged the subducting profile of the Nazca plate from approximately the southernmost end of our seismic network, with the slab appearing to be very well correlated to local source hypocentral coordinates, as expected. Indication of a dipping slab is manifested in the consistent emergence of a tilted thin rectangular feature seen in nearly all inversions attempted, including ones initialized at halfspace v_p with no parametrization for a Moho (and thus layering), which is one of the less biased parametrizations and initial conditions. Some of the strongest evidence of a subducting slab came from local only inversions, which appear to resolve it cleanly. However, it is worth conceding that the initial conditions of the local hypocenters, as well as the concentration of Voronoi nodes near the initial hypocenters, could in theory create a bias, even in a scenario with a half space initialization for the complete velocity

field, though such a bias would be smaller than that of most other common configurations, including ones we carried out.

Notably, examination of teleseismic only inversions (Figures 54 - 56) converged on a different geometry and field strength than local only inversions. While appearing to only barely or not at all resolving the region of the presumed subducting lab in the South-East, they exhibit a strong tendency to converge to fields with existence of strong and stable slow anomalies in the northern region of our seismic network. The smoothness and extent of the anomaly are indicative of an error in boundary condition assumptions, which generate bounding grid point times assuming an iasp91 standard Earth everywhere outside the invertible grid.

While our initial top level Monte Carlo run results are still too sparse to describe their effective parameter space with useful confidence, we believe that array inversions with independently initialized parallel gradient based inversions using reasonably randomized key parameters justify the computational cost of such a run, as this approach gives the viewer a less subjective ensemble of start and end configurations. This allows for probabilistic interpretations of results requiring neither the linearization assumptions one must make when relying purely on a gradient based inversion, nor computationally prohibitive dense grid search inversion.

Geologically, we believe our inversion results are strongly indicative of a Moho thickening that is largely consistent with Airy isostasy corresponding to large scale topography, with a maximum Moho depth bounded by our grid to be between 55km and 75km below the surface. We have also observed a consistent and robust emergence of a fast planar high velocity zone of a thickness on the order of 50-100km consistent with a tilted subducting slab roughly coplanar with the seismicity distribution located South and East of our network's projected centroid, which is located approximately at 15.S, 72.5W. North and especially West of this

point, the resolvability of a slab-like feature is significantly degraded, and there is some indication that the subduction is significantly flattened, also consistent with a flattening of the seismicity profile. We have also relocated local event hypocenters, the locations of which are indicative of subducting plate boundaries.

12 Prospective Research

12.1 Prior Knowledge Constraints

One inversion paradigm which was not investigated in detail is iterated use of prior data in an inversion scheme (Tarantola, 1987) and its adaptation for use in our scheme. Current and upcoming research is focused on a new regularization approach to a prior weighted iterating inversion. Preliminary inversion runs were done using a step solution very similar to Tarantola's:

$$\beta_{n+1} = \beta_n + \epsilon\big[J^T J + \Lambda\big]^{-1}(J^T r + \Lambda(\beta_n - \beta_0)) \tag{16}$$

where $\Lambda_{ii} = \frac{\sqrt{m}}{\tau_i^2}$ and m is the number of data points used, and ϵ is a scaling factor that results in the lowest misfit found using a grid search.

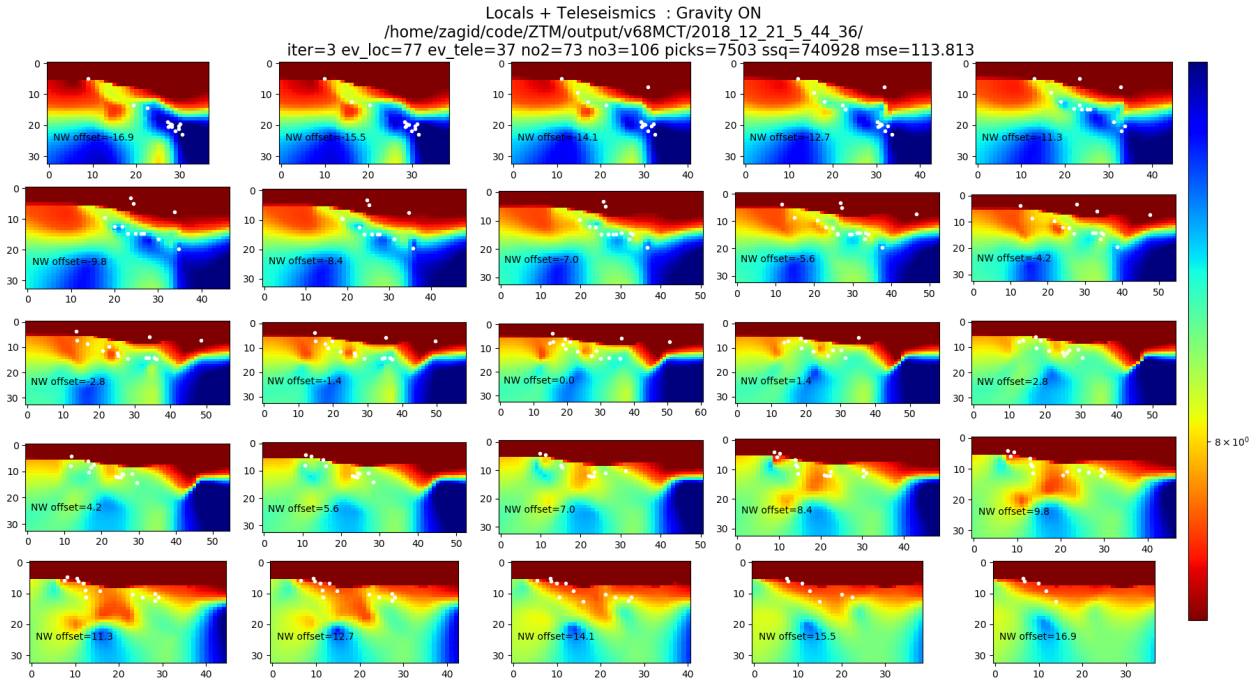


Figure 83: Prior weighted iteration 3.

Preliminary results shown in figures 83 - 84 suggest that while prior weighted inversions reduce large scale instabilities and regional drift, their effect on small scale instabilities is much less pronounced. Covariance induced localized solution instabilities appear to be emergent as the iteration count grows, despite the quadratic penalty imposed for deviation from the initial parameter set.

Another form of prior knowledge inversion that we tested involved placement of a high velocity planar tilted slab coplanar with high velocity regions resolved by our null hypothesis inversions and distribution of seismicity. This initialization was used as the starting parameter set for an iterating inversion series, shown in Figure 85, and the rendered field after 2 inversion iterations in Figure 86. This approach can be useful in identifying potential features that would be lost in the inversion's null space.

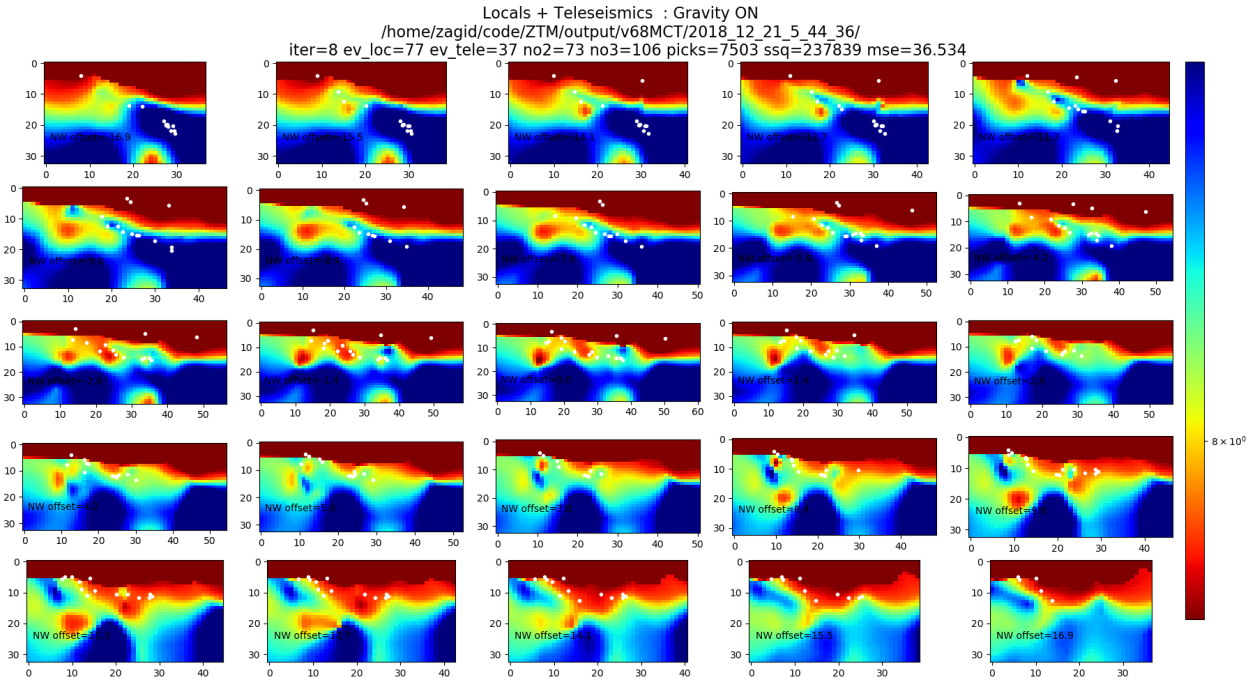


Figure 84: Prior weighted iteration 8.

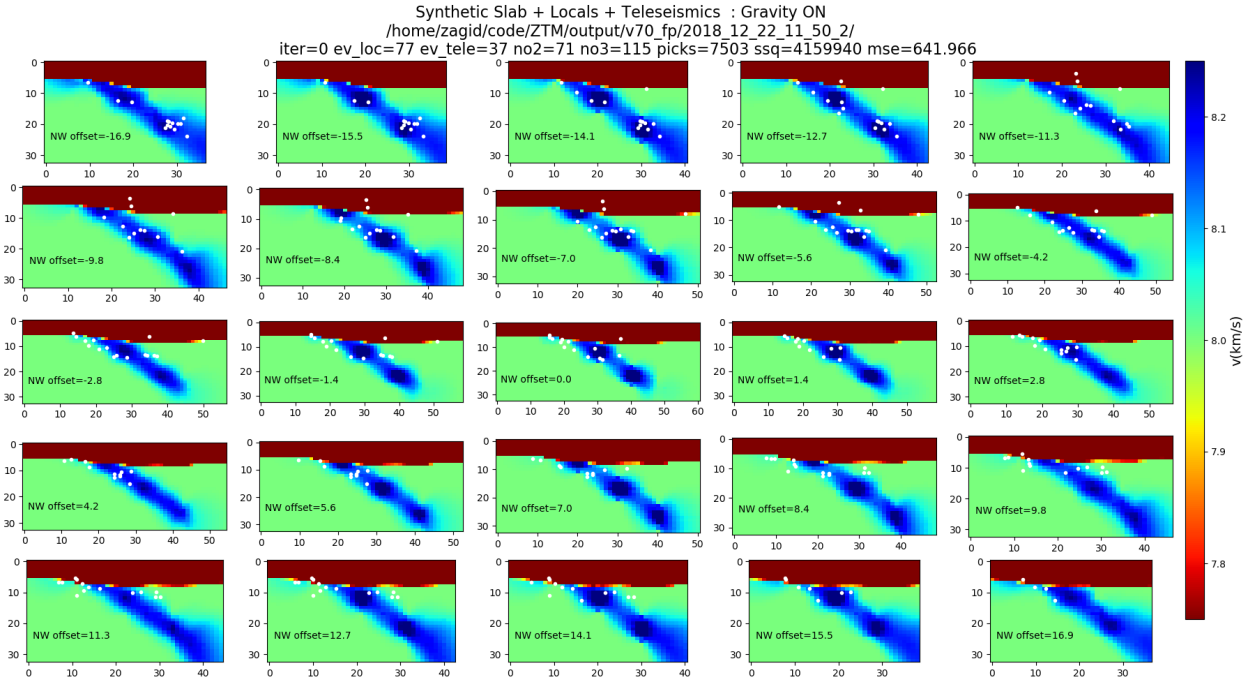


Figure 85: Slab initialized iteration 0.

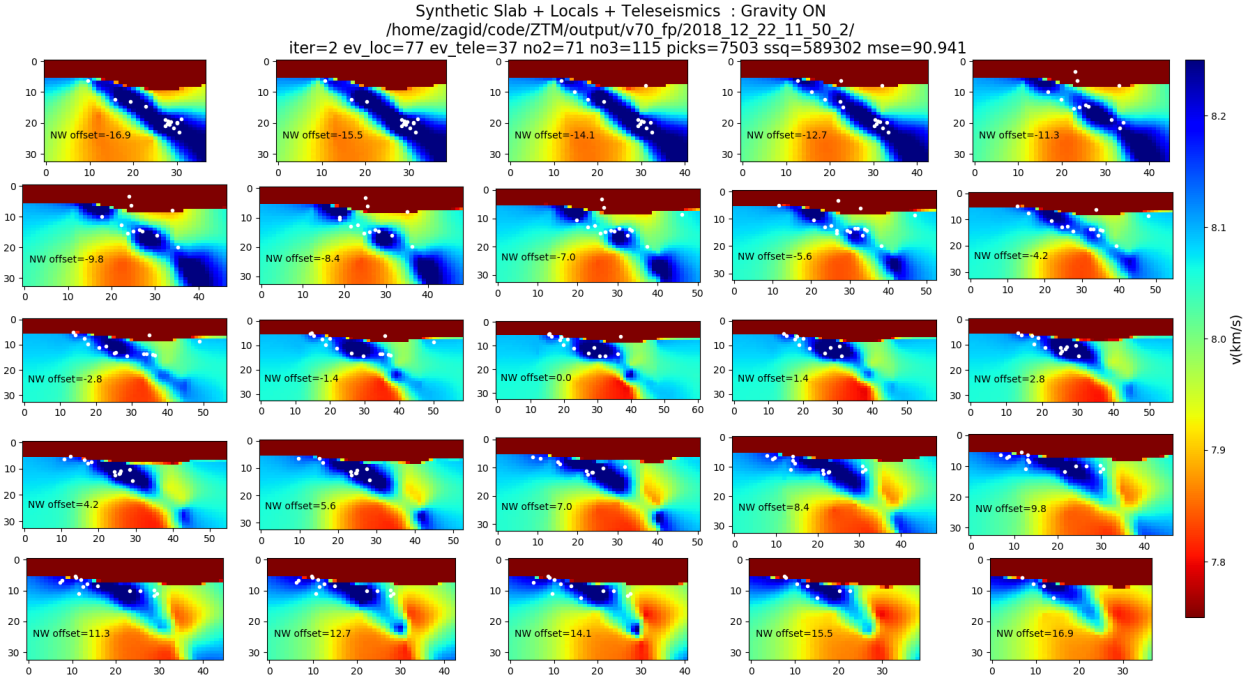


Figure 86: Slab initialized iteration 2.

12.2 Other Objectives

Additional work that needs to be done prior to a potential submission for peer reviewed publication consists primarily of detailed testing and verification of the algorithms and the source code presented. Initially, simple synthetics should be generated using a geologically anticipated velocity model, and after noise injection, the synthetic data needs to be inverted using our inversion routine. Comparison with the generating model will give a more reliable indication of parameter resolvability. A more rigorous test would involve generating synthetic waveforms using a wave equation solver, picking the arrivals, and running inversions on the picked waveforms. This approach would be advantageous, because it would produce synthetics vulnerable to mistiming due to poor signal to noise ratio of diffracted first arrival phases, as observed inside the physical Earth.

Additional characterization of our variant of FMM-VFD is also a priority for bringing its claims to full publication quality. This includes testing over a larger set of grid dimensions and velocity fields.

Another prospective area of research is an implementation of an iterative waveform re-timing routine to be incorporated into our Tomographic Inversion Algorithm. All necessary work on waveform I/O and management in our environment is complete, and an implementation of recursive waveform re-timing is a question of prioritization relative to other objectives.

One significant unresolved matter can broadly be described as error estimation and resolvability of iterated nonlinear inversions. Traditionally, linear inversion has been employed in well conditioned (linear or weakly nonlinear with wide intervals of convergence) problems, and a subset of the methods developed to estimate their solutions' error and resolvability were applied to our problem. In Figures 72 - 74, we present covariance and resolution matrices of sample inversions. However, we note that different damping schemes were applied

to the covariance matrix relative to the resolution matrices. Depending on the assumptions used, three approaches can be described in computation of covariance and resolution matrices. Without delving into details, the differences can be understood as how one applies the previously iterated damping and initial estimates of parameter uncertainty to the calculation of R, previously defined as: $R = [J^T J + D]^{-1} J^T J$. The primary constraint of these error and resolvability estimation schemes is that they require the problem described to be well posed, continuous, and linearized at and near the solution. They also require Gaussian parameter and data error distributions. In scenarios where these conditions are not met, interpreting these metrics will not give accurate characterization. As previously described, the inverse problem we present and solve is not necessarily well conditioned at its stationary points, and because their Jacobians must be calculated using minimal safe step sizes, they can converge only to solutions within the scale of their Jacobian step size, for which standard linearized resolution operators may not be appropriate.

Finally, continued development and testing of GPU implemented AWE and TI-AWE (Figure 14 - 16) is potentially promising, allowing for amplitude inversion to be done jointly with travel times. One qualification to acoustic waveform modeling is that it accounts only for longitudinal stresses, and for this reason, the first arrival amplitudes will not account for S-wave conversion at discontinuities, which may result in an overestimation of amplitude in certain situations. However, this could be addressed by solving the full wave equation in a manner analogous to the acoustic wave equation solver described. The transformation of either acoustic or full wave equation solutions to a sparse problem in application to first arrival wave fronts has the potential to significantly aid in inversion of highly contrasted areas with major anticipated localized amplification, such as in the scenario hypothesized at the Los Angeles basin edges. A full completion of TI-AWE may allow for a travel time and amplitude inversion of sedimentary basin data and could aid in the study of potential focusing mechanisms observed.

13 Bibliography

13.1 Publications

Alkhalifah T, Fomel S; "Implementing the fast marching eikonal solver: Spherical versus Cartesian coordinates", Geophysical Prospecting 49(2), p 165-178; 2001.

Backus G, J F Gilbert; "Numerical applications of a formalism for geophysical inverse problems", Geophys. J. R. Astron. Soc. 13, p 247-276; 1967.

Backus G, J F Gilbert; "The resolving power of gross earth data", Geophysical Journal International Vol 16 I2; 1968.

Backus G, J F Gilbert; "Uniqueness in the inversion of inaccurate method based on iterated function systems", Water Resour. Res., gross earth data, Philos. Trans. R. Soc. Londom Ser. A, 266, 123-192, 30 (6), p 1721-1745; 1970.

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

Brocher T; "Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust", Bulletin of The Seismological Society of America - BULL SEISMOL SOC AMER. 95. 2081-2092. 10.1785/0120050077; 2005.

Dziewonski A, Anderson D; "Preliminary reference Earth model". Physics of the Earth and Planetary Interiors 25(4): p 297-356; 1981.

Hayes G P, D J Wald, R L Johnson; "Slab1.0: A three-dimensional model of global sub-

duction zone geometries", J. Geophys. Res., 117, B01302; 2012.

Kennet B, Engdahl E; "Traveltimes for global earthquake location and phase identification", Geophysical Journal International 105; p 429-465.

Klimes L, Kvasnicka M; "3-D network ray tracing", Geophysical Journal International, 116: p 726-738; 1994.

Moser T J; "Shortest path calculation of seismic rays", Geophysics, Vol. 56, No. 1, p 59-67; 1991.

Park S W et al.; "Discrete Sibson Interpolation", IEEE Transactions on visualization and computer graphics, Vol.12, No. 2; March/April 2006.

Phillips K, R Clayton, P Davis, H R Guy, S Skinner, I Stubailo, L Audin, V Aguilar; "Structure of the Subduction System in Southern Peru From Seismic Array Data", J. Geophys. Res, 117, B11306; 2012.

Popovici A M, J Sethian; "Three-dimensional traveltime computation using the fast marching method", 67th Ann. Internat. Mtg, Soc. of Expl. Geophys., p 1778-1781; 1997.

Rawlinson N, Sambridge M; "Seismic travel time tomography of the crust and lithosphere", Advances in Geophysics,46, 81-197.1, 2003.

Rawlinson N, Pozgay S, & Fishwick S; "Seismic tomography: a window into deep Earth. Physics of the Earth and Planetary Interiors", 178, p 101-135; 2010.

Rawlinson N, Fichtner A, Sambridge M, Young M; "Chapter One - Seismic Tomography and the Assessment of Uncertainty", Advances in Geophysics, Elsevier, Volume 55, p 1-76, ISBN 9780128002728; 2014.

Sambridge M, Braun J, McQueen H; "Geophysical parametrization and interpolation of irregular data using natural neighbors", Geophysical Journal International, p 122: 837-857; 1995.

Sato H, I S Sacks, T Murase; "The use of laboratory velocity data for estimating temperature and partial melt fraction in the low velocity zone: Comparison with heat flow and electrical conductivity studies", J. Geophys. Res., 94(B5), p 5689-5704; 1989.

Sedgewick, Wayne: Algorithms, 4th ed. p 315-319. Addison Wesley; 2011.

Stacey, Davis: Physics of the Earth, 4th ed. Cambridge University Press; 2008.

E M Syracuse, H Zhang, M Maceira; "Joint inversion of seismic and gravity data for imaging seismic velocity structure of the crust and upper mantle beneath Utah, United States", Tectonophysics, Volume 718, p 105-117; 2017.

Tarantola A; "Inverse problem theory: Method for data fitting and model parameter estimation", Elsevier; 1987.

T Gillberg, M Sourouri, X Cai; "A new parallel 3D front propagation algorithm for fast simulation of geological folds", Procedia Computer Science; 2012.

J Trampert, J Spetzler; "Surface wave tomography: finite-frequency effects lost in the null

space", Geophysical Journal International, Volume 164, Issue 2, p 394-400; 1 Feb, 2006.

Vidale J; "Finite-difference calculation of traveltimes in three dimensions", Geophysics, Vol. 55, No. 5, p 521-526; 1990.

Voronoi G; "Nouvelles applications des paramètres continus à la théorie des formes quadratiques", Journal für die Reine und Angewandte Mathematik (133): 97?178; 1908.

Withers M et al.; "A comparison of select trigger algorithms for automated global seismic phase and event detection", Bulletin of the Seismological Society of America, 88 (1), p 95-106; 1998.

13.2 Online Data and Software Sources

Accompanying source code and data set: https://github.com/abatchev/ZTM

USGS Earthquake catalog: https://earthquake.usgs.gov/earthquakes/search/

Obspy: https://github.com/obspy/obspy/wiki

WGM2012: http://bgi.omp.obs-mip.fr/data-products/Grids-and-models/wgm2012

Peruvian seismic ZG, ZD line data: https://www.iris-database.org/

FMTOMO: http://rses.anu.edu.au/ nick/fmtomo.html