

Joining statistics and geophysics for assessment and uncertainty quantification of three-dimensional seismic Earth models

Carène Larmat¹  | Monica Maceira² | David M. Higdon³ | Dale N. Anderson¹

¹Geophysics Group, EES-17, Los Alamos National Laboratory, Los Alamos, New Mexico

²Global Security Directorate, Oak Ridge National Laboratory, Oak Ridge, Tennessee

³Social Decision Analytics Laboratory, Biocomplexity Institute of Virginia Tech, Arlington, Virginia

Correspondence

Carène Larmat, Geophysics Group, EES-17, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM.

Email: carene@lanl.gov

Seismic inversions produce seismic models, which are 3-dimensional (3D) images of wave velocity of the entire planet retrieved by fitting seismic measurements made on records of past earthquakes or other seismic events. Computing power of the TeraFlop era, along with the dataflow from new, very dense, seismic arrays, has led to a new generation of 3D seismic Earth models with an unprecedented level of resolution.

Here we compare two recent models of western United States from the Dynamic North America (DNA) seismic imaging effort. The two models only differ in the wave propagation that was used for their inversion: one is based on ray theory (RT), and the other on finite frequency (FF). We evaluate the two models using an independent numerical method and statistical tests. We show that they differ in how they produce seismic signals from a subset of earthquakes that were used in the original inversion and were recorded on the US array. This is especially true for measurements done in the Yellowstone area which has a large negative seismic anomaly. This result is of importance for seismologists who have been debating on the practical benefit of using FF in ill-posed Earth inversions. Model evaluation, such as the one reported here, represents an opportunity for collaboration between geophysical and statistical communities. More opportunities should arise with the upcoming Exascale era, which will provide enough computational power to explore together several sources of errors in models with thousands of parameters, opening the way of uncertainty quantification of seismic models.

KEYWORDS

permutation test, seismic models, seismology, spectral element method, *t*-test, uncertainty quantification

1 | INTRODUCTION

Seismic waves are unique geophysical observables that allow imaging of buried Earth structures thanks to their long traveling distances and multiple wavelengths. Measurements on seismic waves are used to produce 3-dimensional (3D) seismic images of the Earth in a way similar to how ultrasounds are used for medical imaging. These images, which represent distribution of seismic wave velocities and other elastic

properties within the Earth—and which are thereafter referred to as *seismic models*—are not only used to infer the history and inner workings of our planet but are also used for specific applications related to energy exploration and national security. For example, protocol for the Comprehensive nuclear Test-Ban Treaty (CTBT) calls for the implementation of an International Monitoring System (IMS) and an associated International Data Centre (IDC). IDC's responsibility includes the release of bulletins that report on locations and

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2017 The Authors. *Statistical Analysis and Data Mining: The ASA Data Science Journal* published by Wiley Periodicals, Inc.

analyses of seismic events. These analyses assess whether or not the event has nuclear explosion characteristics (eg [3,30]) and depend on the seismic models they use. The accuracy on seismic event location is critical to any onsite inspection (OSI) should sufficient IDC technical evidence warrant an OSI. Error much smaller than tens of kilometers is required to guide an OSI under the CTBT when such error is acceptable for global Earth studies.

Each generation of seismologists has worked toward a better quantitative estimate of seismic models by advancing the methods used for the forward and inverse problems – solving the wave equation in 3D structures and handling often ill-posed inversions. Pioneering work (eg [1,2,9,48]) was focused on mathematical understanding of the inverse problem and how to reliably estimate elastic properties within the Earth from sparse surface data. The underdetermined nature of geophysical inverse problems led to mathematical innovations that are now also used in other disciplines (eg [4–6,20,60]). First seismic inversion results at the continental scale were soon published by Menke [33], Romanowicz [41,42], and Taylor and Toksöz [52], among others. Today, opportunities of progress in 3D geophysical imaging are very diverse, coming from data collection to methods and tools (eg [43]). Very dense networks of seismic sensors provide huge volumes of high-quality data. Numerical methods and large high-performance computing (HPC) resources allow for the first time reaching accurate wave propagation modeling in complex media. At the same time, our understanding of both wave propagation and inversion techniques has greatly improved (eg [11,18]). The exploitation of the amount of data and the size of the computational resources needed is challenging, but seismologists are now producing 3D seismic models with the best resolution ever. However, the question of reliability of these models remains largely unexplored when it is crucial for applications such as event screening and location for CTBT implementation. Traditionally and for most seismic imaging studies, uncertainty quantification (UQ) is limited to resolution analyses, such as checkerboard and spike tests, that assess how well a particular inversion method reproduces the original checkerboard or spike pattern by inversion of synthetic seismograms created with these artificial seismic models. Areas where the input model is reasonably recovered are potentially resolved thanks to the paths provided by the set of seismic events and stations used for the inversion. Assessing resolution is a good thing, but it is different from conducting UQ. UQ is a problem far vaster than path coverage, and there are many other issues with these tests. For example, they do not convey proper information on the multiscale resolution of the model needed for a complex object like the Earth because they are often limited to specific velocity variations (eg [11]). They also do not provide meaningful quantifications of nonuniqueness [27,34]. A more robust and better approach is needed.

Model assessment is one of the elements needed in uncertainty quantification. Model assessment usually refers to

quantifying the model uncertainty from some metrics of how accurately this model predicts seismic signals when compared to real data. In this study, we will instead compare 2 different seismic models to each other. Our goal is to conduct a statistical analysis on measurements made on synthetic seismograms generated by an independent numerical method. In the following, we first report results of our analysis. We then discuss the opportunity of a handshake between the geophysics and statistics communities, and identify methods that are emerging as means for future UQ analysis of 3D seismic earth models.

2 | ANALYSIS SETUP

In this paper, we assess 2 seismic models that were produced by 2 different inversion methods, ray theory (RT) and finite frequency (FF). For that purpose, we generate synthetic seismograms that are the modeled ground motion generated at an array of stations by a given seismic source. In our case, the seismic sources are a set of 14 earthquakes that were included in the original inversions. Seismic measurements are made on the seismograms such as picking the arrival time and/or the amplitude of a specific phase, or computing cross-correlation with another set of seismograms. These measurements for our 2 models are generally modeled as follows:

$$\begin{aligned} \text{Measurement}_{\text{RT}} &= \text{Measure}(F(\text{model}_{\text{RT}}, \text{source})) + \text{error}, \\ \text{Measurement}_{\text{FF}} &= \text{Measure}(F(\text{model}_{\text{FF}}, \text{source})) + \text{error}, \end{aligned} \quad (1)$$

where $F()$ is a *seismic propagator*, which computationally propagates seismic waves generated by the source through the model, and *Measure* represents measuring a traveltime or cross-correlation coefficient on the synthetic seismograms. In the following, we describe the 2 models, the set of earthquakes and the seismic propagator we chose. Also, we assume that the source location and mechanism are fixed and known.

2.1 | The seismic models: RT and FF DNA09 models

The group under the direction of Pr. R. Allen at the University of California, Berkeley, has developed several versions of high-resolution models of the western United States (see their website <http://seismo.berkeley.edu/~rallen/research/dna/index.php>) using the unprecedented coverage provided by the seismic network USArray [28]. We chose to work with the version released in 2009, DNA09, because 2 specific models were available for that version: one developed using the body-wave FF approach [37,62], and the other using the classical approximation of RT. In this approximation, seismic waves are only sensitive to the structure along the ray path, an infinitesimally narrow ray between the source and the recording station such as in optics (Figure 1). The extensive use of RT in seismic imaging stems from its simplicity, its intuitive

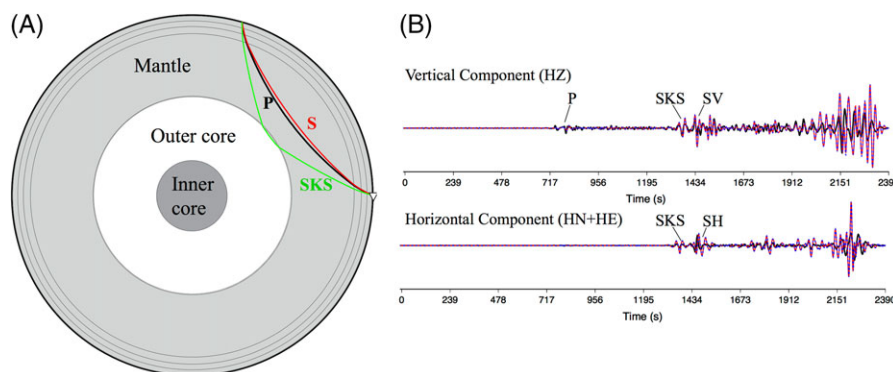


FIGURE 1 (A) Ray paths of the body waves, P, S, and SKS, through the Iasp91 model and between one event and station (inverted triangle) traced with TauP. P waves are fast compressional waves arriving as precursor to the rest of the seismic waveforms. S waves are also body waves traveling along a path but involving much slower shearing motions. SKS have reached the outer core. These types of body waves were used for establishing the DNA09 models. (B) Example of seismograms recorded or computed for one station; actual data that were recorded for the Marianna event are shown in black, and the synthetic data computed with a model called ray theory (RT) and another one called finite frequency (FF) are shown in red and blue, respectively. The two components of the S wave recorded on the vertical and horizontal components are called SV and SH, respectively. HZ, HE, and HN refer to the 3 component channels, vertical (Z), west-east (E), and south-north (N)

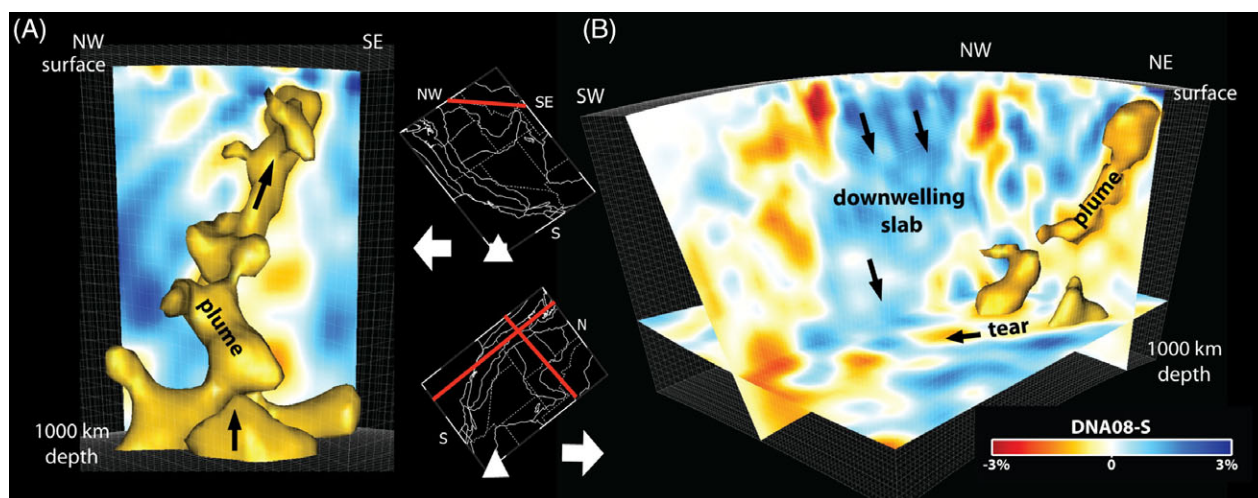


FIGURE 2 Three-dimensional seismic models beneath the western United States. The quantity plotted here is dV_s/V_s from the FF DNA09 model (see text). Red lines on insert maps indicate location of cross sections shown in (A) and (B). Seismic velocities depend on several parameters but can be interpreted at first order as indicator of temperature. Warm colors show low velocity regions interpreted as hot upgoing material (motion indicated by arrows); blue is high velocity interpreted as cold material that has a tendency to sink. (A) Yellowstone plume appearing in the model. (B) Evidence of a subducting slab under the Western part of the United States and of complex interaction with the Yellowstone plume

interpretation, and the fact that algorithms computing ray paths require very low computational resources. Damping and smoothing regularization are used to stabilize the inversion. The FF model is based on the Born-Fréchet kernel theory, taking into account the frequency-dependent sensitivity and wavefront healing [36,59,61]. In that case, the constructive and destructive inferences of scattered waves extend the sensitivity to a few wavelengths off the ray and seismic anomaly, a phenomenon known in optics as Fresnel zone. In consequence, the sensitivity kernels associated with seismic traveltimes and amplitude take the shape of a banana-doughnut, as they follow the banana-shaped ray and there is no sensitivity exactly on the ray (eg [17,56]). Although FF provides a more complete theory to represent wave propagation, and that more and more models are built taking into account some FF effects, the debate continues as to whether its application to 3D modeling produces better models (eg

[8,31,39,49,53]). It is a rare opportunity to be in a position to compare 2 models that only differ on the theory used to build them [7].

The 2 models resolve the mantle down to a depth of ~ 1000 km and with a lateral and vertical resolution of ~ 200 km. They were developed using the very same dataset—arrival times recorded on the Earthscope-US Array from 142 seismic events occurring between 2006 and 2009. A full seismic record contains many different phases, each with its own resolution. The inversion of DNA09 aims to fit the arrival time of the following body waves: S, P, and SKS (see Figure 1). Figure 2 shows the resulting velocity model for the shear (S) wave, hereafter referred to as V_s . Velocity anomalies are shown at each point of the model as

$$\frac{dV_s}{V_s} (\%) = \frac{V_s - V_s^{\text{Iasp91}}}{V_s^{\text{Iasp91}}} \times 100, \quad (2)$$

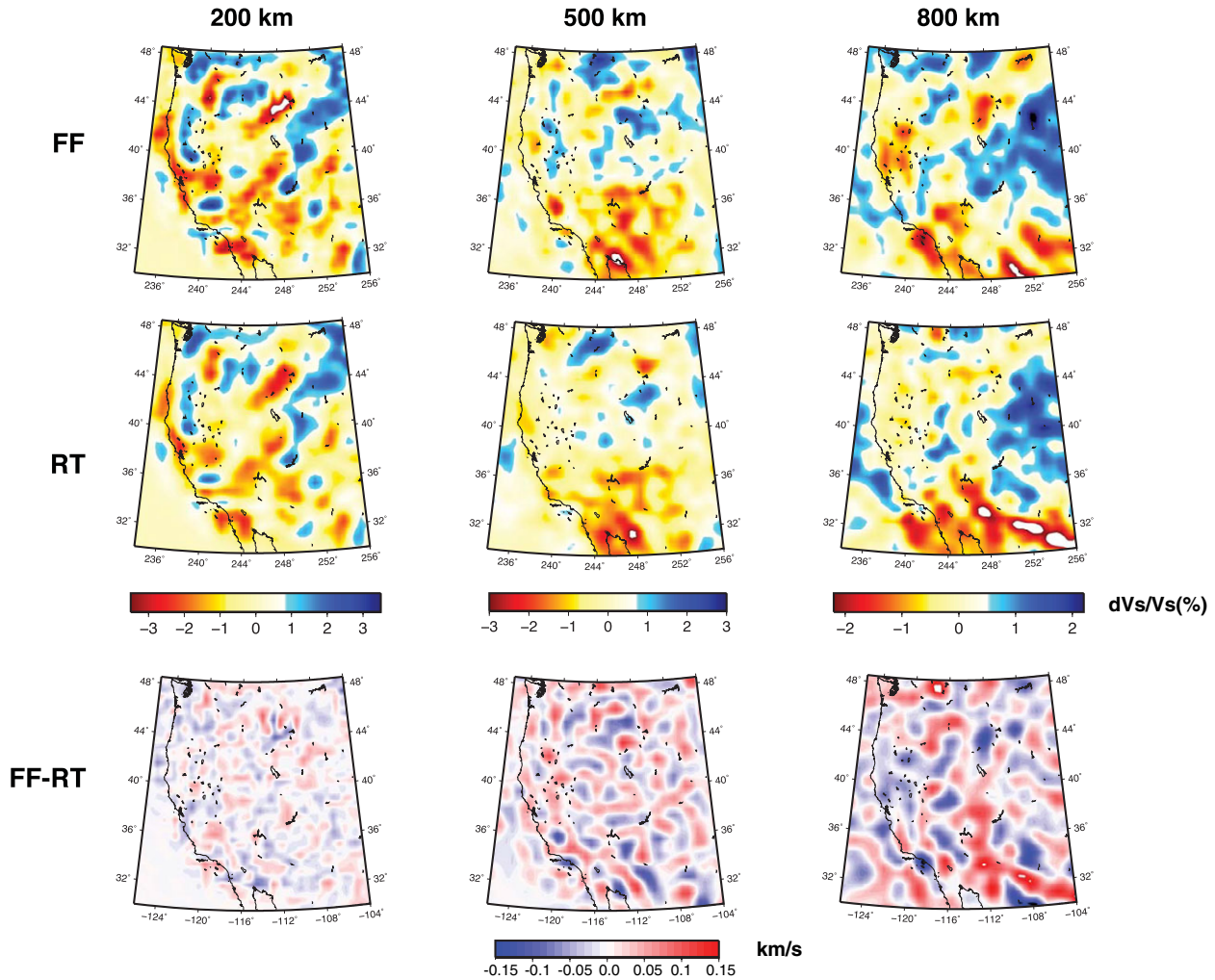


FIGURE 3 Maps of the shear wave anomalies (dV_s/V_s) of the finite frequency (FF) and ray theory (RT) models and their difference in velocity (bottom). The velocity is computed as the difference of dV_s/V_s between the FF and RT models multiplied by the V_s for Iasp91. We display maps for depths 200 km, 500 km (upper mantle), and 800 km (lower mantle)

where V_s^{Iasp91} is the value of the shear wave for a 1D reference model, iasp91 [23]. The 1D model captures the vertical layering of the Earth, especially the general increase of seismic velocities with increasing depths due to higher confinement pressures. Negative and positive values of the quantity calculated in Equation 2 are plotted in warm and cold colors, respectively, as they are interpreted as material of high temperature (slow velocity) and low temperature (high velocity). It must be noted that temperature is only one of the possible processes that affect the seismic velocity. Others possible processes are abnormal chemical composition or water content, but seismic velocities are considered a good indicator of temperature at first order (eg [54]). The left panel, corresponding to a cross-section under Yellowstone, shows a continuous low-velocity conduit extending to a depth of 1000 km, deep into the Earth's mantle and way further than any previous studies. This model gives ground to the hypothesis that the plume responsible for Yellowstone volcanic activity originates from deep mantle processes. The right panel shows a cross-section of the Juan de Fuca plate subducting beneath North America from the Pacific coast. The slab, portion

of the tectonic plate being subducted, is depicted in blue and it appears to get separated in different fragments eastward of the Yellowstone plume, suggesting a break-up of the slab due to interaction with the latter. These features appearing in DNA09 are exemplary of the level of details now resolvable with the new generation of high-resolution 3D models.

Figure 3 shows maps of dV_s/V_s for the RT and FF models at different depths. The 2 models appear to agree on large features showing fast velocities (indicated by blue colors) on the eastern part of the model corresponding to the relatively old material of the stable North America, while the western United States displays slower velocity (shown with warm colors) corresponding to structures such as the Yellowstone plume. However, differences between the 2 models exist, as shown on the difference maps at the bottom of Figure 3. They appear as fluctuations of smaller size than the geophysical features described earlier and become more and more important in depth. The histograms of the seismic velocity anomalies are shown in Figure 4; they are close to a Gaussian shape but shifted to the negative values, with a mean of -0.0762 and

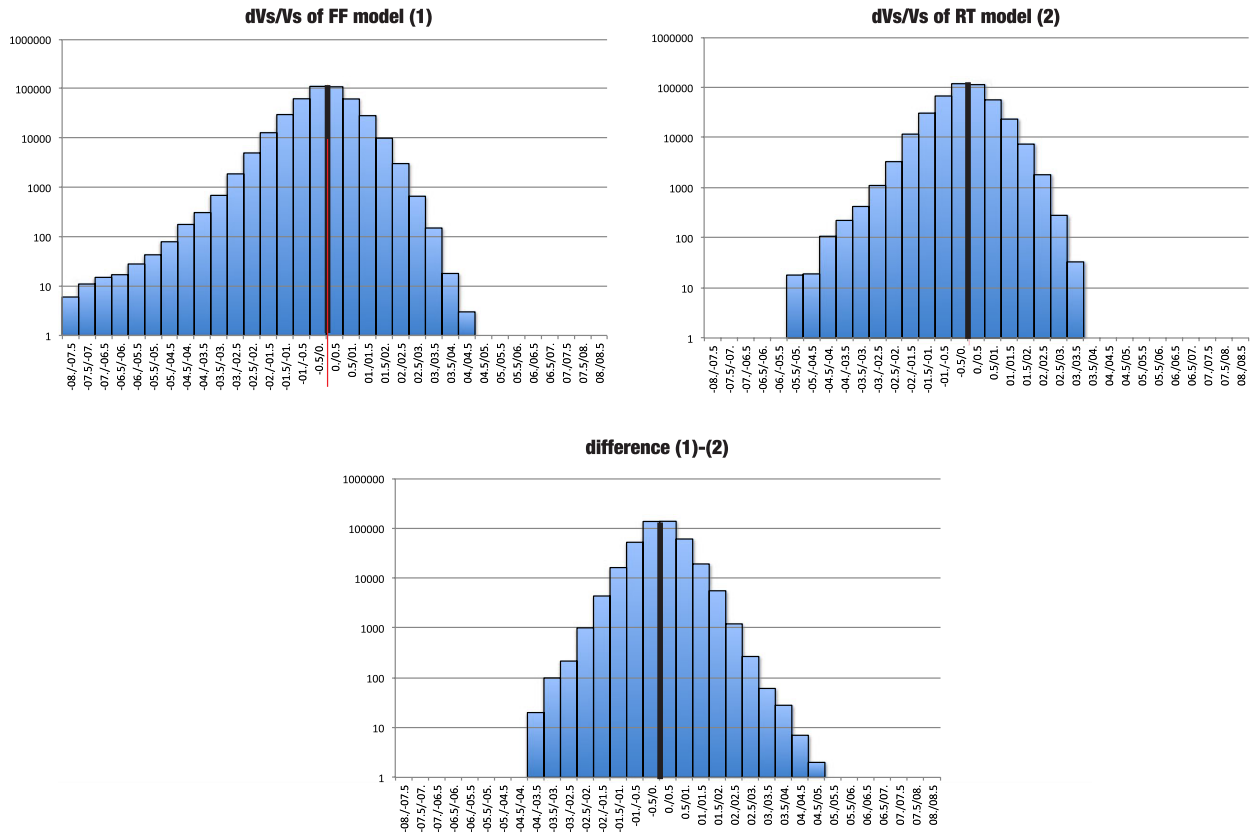


FIGURE 4 Histograms of dV_s/V_s for the seismic models finite frequency (FF) and ray theory (RT) over the 436 896 grid points of each model. Also presented is the histogram of the difference between the two models

−0.0436, respectively, for the RT and the FF model. On the contrary, the difference between the 2 models is noticeably a zero-centered distribution, suggesting that, if present, any bias in the 2 models is similar.

2.2 | Data and synthetic databases

Figure 5 shows the locations of the 1061 seismic stations (blue triangles) and 14 seismic events (red circles in inset) that were used for our study. We chose to work with earthquakes that were included in the original inversion because we know that their paths have been resolved by the inversions. This way, we avoid discussion about how to estimate data coverage for both theories. We select 14 earthquakes from the 100 earthquakes used in the inversion in order to have a good azimuthal coverage and a good representation of all types of source mechanisms (Table 1). These mechanisms are models of the type of motions happening during an earthquake, which depend on the tectonic setting. Each type of earthquake has its own sets of challenges concerning the source model and data. We chose events with magnitude (size) between 6.0 and 8.1, which offers good data quality (signal to noise ratio) and for which good seismic source models exist.

We have 2 main sets of observations or seismograms. On one hand, we have the real seismograms generated by the 14 seismic events and recorded at many of the 1061 stations

(note that not all events are recorded by all the stations). On these real seismograms, we can measure the arrival time of different seismic waves using advanced tools combining hand-picking and multichannel cross-correlation. The arrival times are transformed into time delays to a reference earth model, following

$$\delta t = t - t^{\text{asp}91}, \quad (3)$$

where t is the arrival time in seconds. δt would be 0 for all stations and events if there was no velocity anomaly. Time delay measurements were limited to S waves, SH, and SV, depending on the instrument component (see Figure 1). The number of measurements available varies for each event due to data quality and recording issue, and is indicated in Table 1. Notice that event #14 has no such measurements because of our inability to measure due to contamination of the S-wave on the records.

On the other hand, we have a set of synthetic seismograms that we compute. For each seismic event and each station, we can compute the predicted seismogram using DNA09 FF or DNA09 RT as the propagators. Therefore, we have a set of synthetic seismograms using the FF model and a different set using the RT model (see example Figure 1). To compute these synthetic seismograms, we use the open-source package SPECFEM3D that is based on the spectral element method (SEM). SEM is a direct numerical method (i.e. not based on any mathematical representation but simple discretization), which solves the wave equation in its integral form on

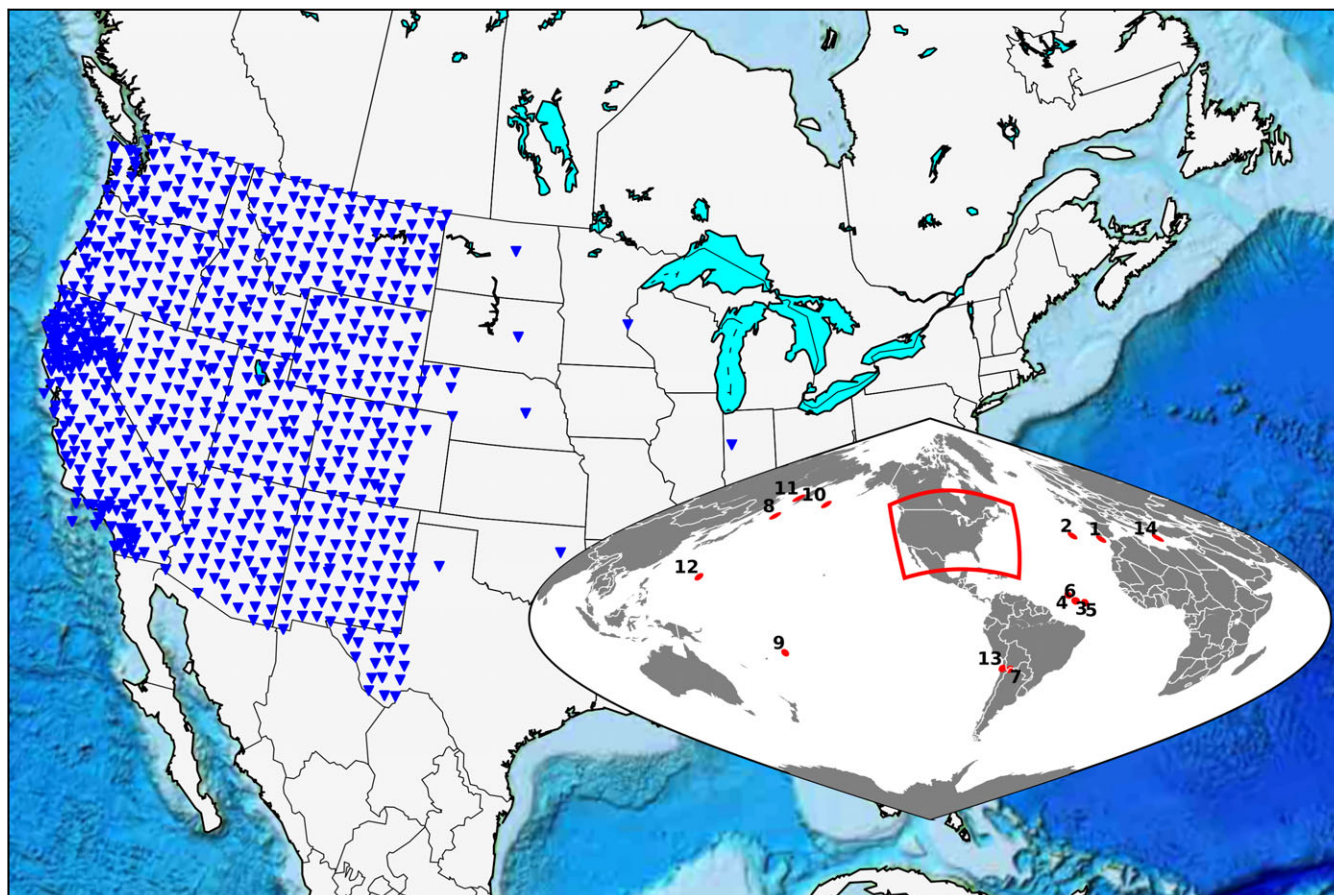


FIGURE 5 Locations of the 1061 seismic stations used in this study are shown as blue triangles. Most of the stations are from the Earthscope-USArray that provides a dense and regular coverage of western United States. We also have some stations from the IRIS and USGS networks. (Inset) Locations of the 14 seismic events used for evaluating the ray theory (RT) and finite frequency (FF) models

TABLE 1 List of seismic events used in assessment of ray theory (RT) and finite frequency (FF) models. This set of seismic events covers different geographical locations providing a good azimuthal coverage and also different tectonic setting (see column type, which refers to the source mechanism). The last 3 columns show the number of measurements of each type on real data or synthetic seismograms

<i>i</i>	Location	Date	Magnitude	Type	Delay SV	Delay SH	cc synth
1	Azores-Cape	February 12, 2007	6.0	Reverse	145	113	3*1061
2	Azores	April 5, 2007	6.3	Normal	128	286	3*1061
3	Central-Mid-Atlantic	August 20, 2007	6.5	Strike-slip	408	219	3*1061
4	Northern Mid-Atlantic	February 8, 2008	6.9	Strike-slip + reverse	252	276	3*1061
5	Central-Mid-Atlantic	May 23, 2008	6.5	Strike-slip	400	63	3*1061
6	Central-Mid-Atlantic	September 10, 2008	6.6	Strike-slip + reverse	461	331	3*1061
7	Jujuy	November 18, 2007	6.0	Reverse	53	265	3*1061
8	Kuril	January 13, 2007	8.1	Normal	278	334	3*1061
9	Samoa	December 13, 2007	6.2	Reverse	275	53	3*1061
10	Aleutians	December 19, 2007	7.1		445	467	3*1061
11	Sea of Ohkost	November 24, 2008	7.3	Reverse	344	543	3*1061
12	Mariana	October 31, 2007	7.2	Reverse	123	122	3*1061
13	Chile	November 14, 2007	7.7	Reverse	361	367	3*1061
14	Greece	February 14, 2008	6.8	Reverse	0	0	3*1061

customized meshes made of hexahedral elements (eg [25,26]). It employs a high-order finite element method with exponential convergence for smooth solutions while maintaining the geometric flexibility of finite elements. The two models DNA09 FF and RT were implemented into

SPECFEM3D and tapered out into the reference model iasp91 for areas of the globe not covered by DNA09. The whole Earth is meshed by 19 million hexahedral elements thanks to a meshing technique called the “cubed sphere” [44]. The smallest element size is 25 km, which is the smallest wavelength

that can be resolved by the mesh according to the sampling condition, which requires a minimum number of grid points per wavelength. In this case, it means that all waves with period above 10 s are correctly modeled. The time step is fixed at 0.1054 s following the classical Courant-Friedrichs-Lewy (CFL) condition. For each seismic event, we computed 40 min of seismic records for each 3 components of the 1061 seismic stations contained in our model (see last column of Table 1). The total computational cost of the whole simulations amounts to 70 K CPU-hours on a machine with peak performance of 38.3 TFlops/s, meaning that studies like ours require access to some computational resources. The choice of SPECFEM3D ensures a fair comparison of the 2 DNA models, as none of the tested models uses that technique as forward modeling and as SEM is a direct numerical method. There is no assumption except for the high-frequency cut-off due to the sampling condition and the form of wave equation implemented, which in this case is quite complete, incorporating the presence of oceans, the ellipticity of the Earth, the topography and bathymetry of its surface, the gravity, and the rotation of the Earth [24].

We then perform the same post-processing and measurements on the synthetic seismograms as the ones performed on the real data. In the following, we present our statistical analysis on the different types of measurements and frequency domains.

3 | STATISTICAL ANALYSIS

3.1 | *t*-Test on S-wave arrival times

In Figure 6, we plot the time delay (δt) measured on the synthetics (y-axis) vs the ones measured on the real data (x-axis) and this for the 13 events with these types of measurements. Note that we have 2 datasets considered independent, the measurement on the SV and SH components that sense differently the Earth, and 2 different sets of synthetics, 1 for the FF model and 1 for the RT model. Therefore, for each event, we have 4 scatter-plots. If the models were fitting the data perfectly, all the data points would be on the line $x=y$ on these plots. We can observe that the fit to the real data varies with each seismic event. Events #3, #4, and #6 display good agreement between data and synthetics; meanwhile, the models correlate poorly with observations for events #8, #10, and #13. We can also see that there are no obvious differences between the synthetics computed with the FF and RT models.

We now focus on the mean misfit to the data of the 2 tested models. It is computed as

$$\text{misfit} = \sum |\delta t_{\text{synthetic}} - \delta t_{\text{data}}|. \quad (4)$$

The open and full circles on top row of Figure 7 show the values obtained for each earthquake when the sum is made over all of the arrival time measurements available for each

seismic event. A value is obtained for each type of synthetics (FF or RT) and for each component (SH or SV). We then take the difference of this mean misfit for FF and RT, which are plotted as blue dots. A negative difference means that FF fits the data better than RT, as the mean absolute misfit with data is smaller. We then compute the average of the difference over these 13 events, which was found negative for both the SV and SH dataset. Taking these 13 seismic events to be a representative collection, we perform a *t*-test to test whether that negative average is statistically significant. A *P*-value of 0.003 was found for the SV component and a *P*-value of 0.4 was found for the component SH. This indicates that there is a significant difference between RT and FF for the SV component. However, we cannot conclude either model produces more accurate arrival times for the SH component. We can attribute this result to the fact that signal-to-noise ratio is known to be poorer for the horizontal components due to the design of traditional seismometers. The SH component also contains more shear signature relative to SV, and the latter is known to become easily complex when propagating through 3D structures, making the picking of the arrival time more challenging.

We then perform a 1-sided *t*-test on the significance of the average of the difference of the misfit for RT and FF for each individual event, meaning that the difference between each pair of open and full circle is tested over the dataset made of the available stations for this event. The logarithm of the obtained *P*-values (y-axis) is plotted vs the difference average (x-axis) on the bottom row of Figure 7. Any data point situated on the left on the plot means that the difference between FF-modeled time delays and actual real data is smaller than for RT, meaning that the FF model fits the data better than RT. It must be noted that each *P*-value on that graph is computed on a different statistical baseline: the number of stations and the average difference are different for each earthquake. Also, the computed *P*-values assume that the measurements are independent within each earthquake, which is not probably true. For these reasons, we have to be cautious about any interpretation on the *P*-values, especially about comparing them between events. However, such a plot allows identifying which event would have passed or failed a *t*-test if it had been tested independently and assuming independent measurements. The same type of subduction earthquakes (Aleutians #10, Mariana #12, and Chile #13) consistently favors FF vs RT, and this for both SH and SV. For SV measurements, subduction events are joined by events of different areas and types that all reach *P*-values below 0.01.

3.2 | Spatial permutation tests and waveform cross-correlation

It is hypothesized that the 2 models treat the Yellowstone plume differently. If this were the case, we would expect syn-

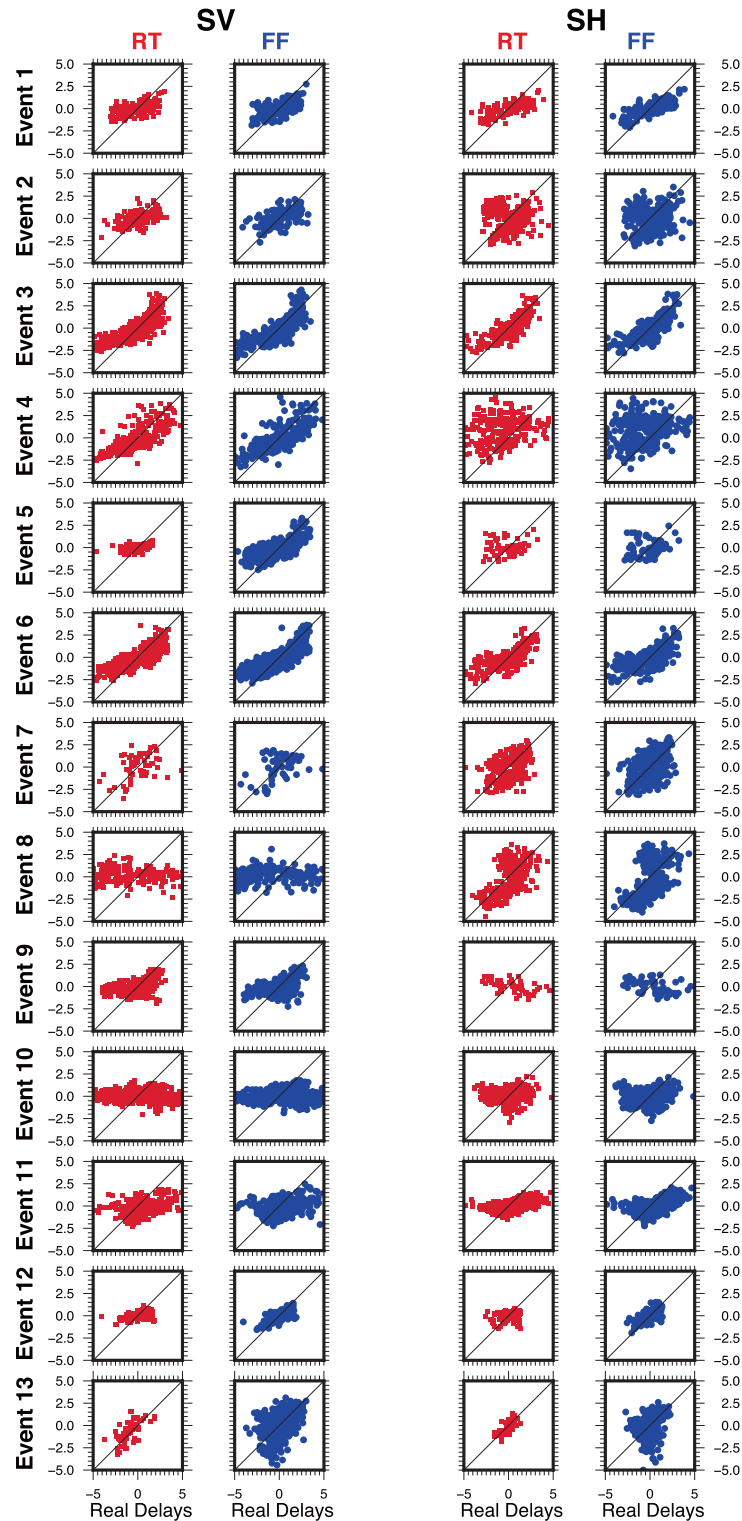


FIGURE 6 Scatter-plots of time delays of S waves (δt_s) measured on synthetic seismograms vs those measured on real seismic records. For each event, we present a scatter-plot for both the SH and SV component, and for the ray theory (RT) and finite frequency (FF) model. The line of slope 1 corresponding to identical measurements is plotted for guidance

thetic seismograms to differ more for sensors located near the Yellowstone region than for sensors located further away.

Hence we now focus on the synthetic seismograms and compute waveform cross-correlation coefficients between RT and FF synthetics for a 55-s time window around the S wave. We first filtered the seismograms to keep the period content between 10 and 50 s (period band used to measure

the arrival time of the S wave) and systematically computed cross-correlation coefficients of 55-s-long time window around the S-wave arrival between RT and FF synthetics. This was carried out for the 3 instrument components: vertical (SV), north-south (SN), and east-west (WE).

The cross-correlation coefficients obtained are plotted at the location of the seismic station in Figure 8. The first obser-

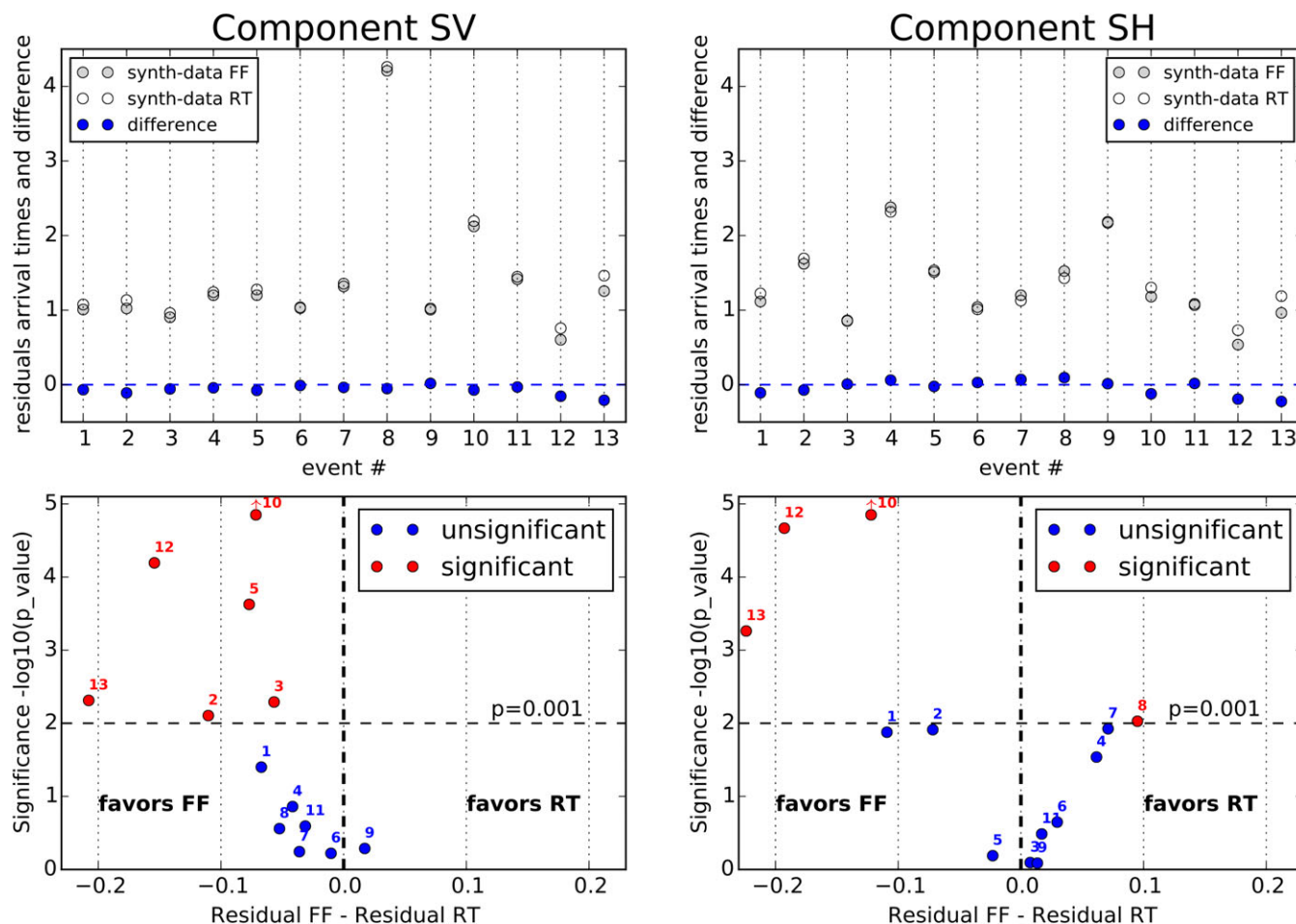
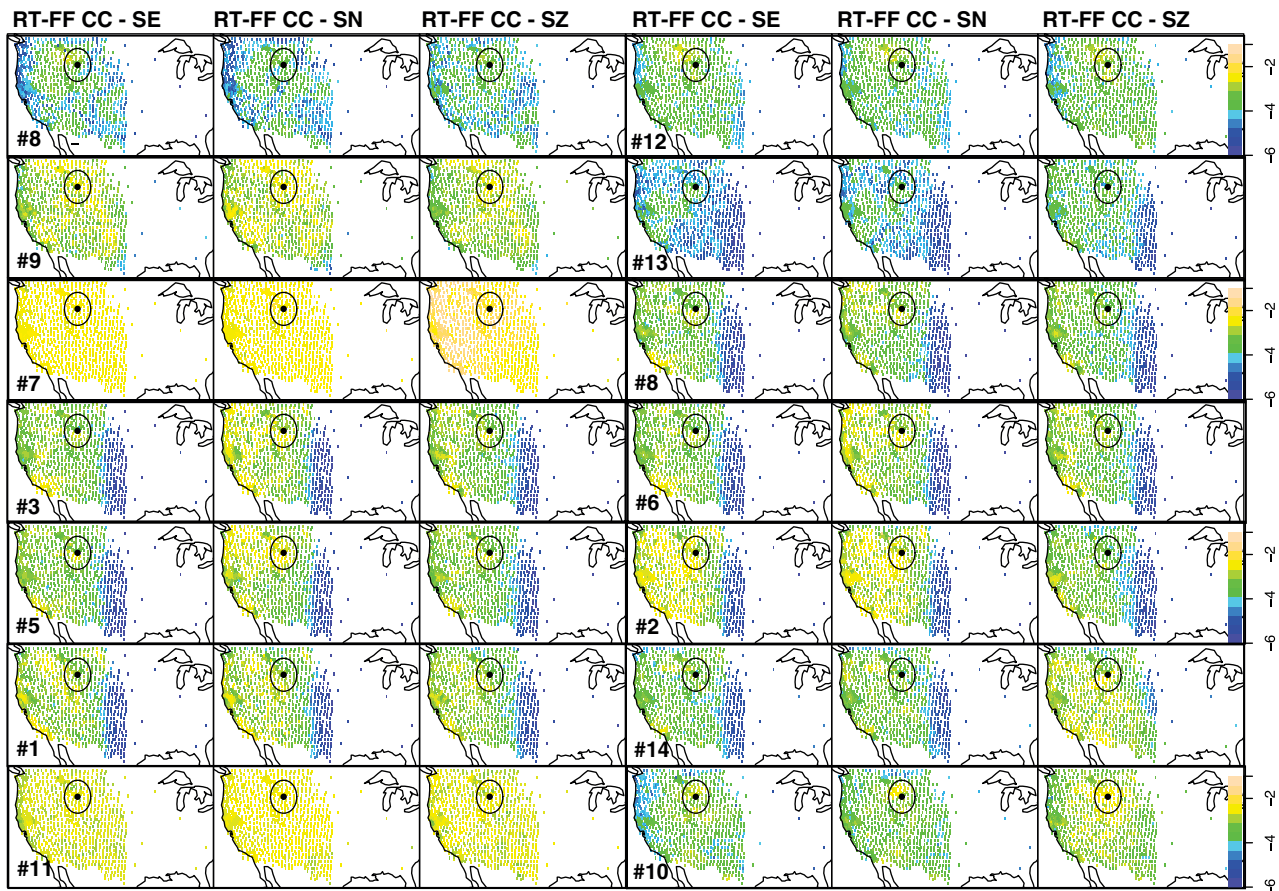


FIGURE 7 In the top row, for each of the 13 events of the database, the mean of the absolute residual between actual and modeled measurements is plotted for both the ray theory (RT; open circle) and finite frequency (FF; full circle). The difference $FF - RT$ is also plotted in blue. In the bottom row, the scatter-plot of the significance of the difference ($FF - RT$) vs its fold-value is presented. Data point passing the P -test (P -value below 0.01) are colored in red. For event 10, significance is above the maximum of the plot for both components

vation is that the coefficient values are very close to 1—we had to use a color code based on the first digit position different from 9. The differences are at the most 1% (digit = 2) for the Jujuy (#7) and Sea of Ohkost (#11) events. As it was the case for the arrival times, the coefficients values seem to mostly depend on the seismic event, meaning the main direction along which the seismic energy has traveled from the seismic event to the seismic stations. The cross-correlation coefficients maps show smaller coefficients in the western part of the United States, suggesting a larger difference between the RT and FF models for this area. In particular, Yellowstone appears as a zone with small coefficient values between the 2 models. This result was confirmed by computing the average cross-correlation coefficient (see fig. 3 of [31]). To test whether this difference is statistically significant between Yellowstone and other regions, we applied a spatial permutation test [32], selecting target locations uniformly at random over the spatial region. As the test statistic, we take the average coefficient value over all stations within a 300-km radius of Yellowstone (black circle in Figure 8), comparing it with cross-correlation coefficient averages computed for randomly situated circles of 300 km radius. We

only considered circle locations that contained at least 10 stations. We randomly generated 100 000 such circles. The rank (from smallest to largest) of the Yellowstone average over 100 001 gives the 1-sided P -value for this test. The permutation test demonstrates that for the Kuril event (event #8), stations within 300 km of Yellowstone lie in the 5th percentile ($P = 0.05$) of these 100 000 realizations, which is a rough cut-off for determining significance for the difference between FF and RT synthetic seismograms. We summarize the P -values obtained for all events in Table 2—note that the significance measure has not been adjusted to account for multiple testing. We can observe an azimuthal dependence: for events #9, #10, #11, and #12, all of which occurred to the west, Yellowstone stands out as a statistically significant anomaly, whereas the remaining events fail to demonstrate if Yellowstone is or is not significantly different from other regions.

Finally, we also computed the cross-correlation coefficient between the real data and RT and FF synthetics for 2 frequency bands for some of the events in this study (1): 10 to 250 s is a period band that encompasses small wavelengths (down to 25 km) that are still discussed in global imaging, (2) 50 to 200 s corresponds to the intermediate wave-



−6: cross cor = .999999, −4: cross cor = .9999XX, −2: cross cor = .99XXXX

FIGURE 8 Maps of the cross-correlation coefficients computed between the synthetics of the ray theory (RT) and finite frequency (FF) models (for each of the 3 components, SE, SN, SZ). The latter were filtered between 10 and 50 s. Each station location is color-coded according to the first digit position different from 9. For example, the yellow color correspond to cross-correlation coefficient than are more than 1% different from 1. The black circle represents the area that was used for Yellowstone in the spatial randomization test

lengths of Earth models for which models are in general good agreement. Maps of the cross-correlation coefficients for the Mariana event are presented in Figure 9. Our analysis confirms that the DNA models predict very well real data for periods over 50 s. The fit is still good but degraded for the period band of 10 to 250 s. The same observations were made on a few other events.

4 | CONCLUSIONS AND DISCUSSION

Accurate image of local structures, especially of the complex Earth crust, are not only crucial to a better understanding of our planet but also essential in technology applications such as oil and gas exploration, as well as national mission areas of nuclear nonproliferation and energy security (geothermal energy, geological carbon sequestration, used-fuel disposition, and repository science). For each of these areas, UQ—understanding implicitly and explicitly the range of possible solutions permitted by the set of observables—is crucial to build confidence on decision making and understand the predictions generated based on these models.

Here we presented different statistical assessment approaches to compare 2 high-resolution seismic models, 1

built with RT and 1 with FF. We paid particular attention to ensure a fair and independent comparison of the 2 imaging methods and models. To that end, we made use of the SEM, which is a direct numerical method that was not used in the inversion for generation of the tested models (DNA09). Aided by very simple statistical tools, we were able to conclude that that FF approaches to seismic imaging exhibit measurable improvement. This conclusion was reached with a limited statistical sampling of 14 seismic events whose location plays a major role in the fit between synthetics and data. The mentioned improvement is more evident for subduction events situated to the west of the model and on components of the seismograms with the best data quality. We hypothesized that the paths for these events are less affected by structures outside the model when optimizing the sampling of the two models under the recording stations. We also demonstrated that the difference between the two models is only significant in regions with pronounced low velocity anomalies, such as Yellowstone. These results suggest how complex the processes of validation and UQ are for seismic models.

Our study also gives some indication of the level of accuracy to the data of the 2 models. We confirm that the fit to the real data is good for long periods but the level of

TABLE 2 *P*-Values (rounded to the second significant digit) obtained when testing the rank of Yellowstone local average cross-correlation coefficient over a 100 000 random realization. *P*-Values are displayed on columns 3 to 5 for the WE, NS, and vertical components, respectively. The event number of the earthquake in our database is shown in column 2. Earthquakes are color-coded according to the success or failure of the test: synthetic seismograms of the events in blue indicate that the ray theory (RT) and finite frequency (FF) models are significantly different around the Yellowstone area; the test is inconclusive for the events displayed in red and yellow. Yellow indicates that at least one component has a low *P*-value

Event name	Event number	SE	SN	SZ
Kuril (January 13, 2007)	8	<i>P</i> = 0.05	<i>P</i> = 0.06	<i>P</i> = 0
Mariana (October 31, 2007)	12	<i>P</i> = 0.09	<i>P</i> = 0.09	<i>P</i> = 0.04
Samoa (December 13, 2007)	9	<i>P</i> = 0.09	<i>P</i> = 0.08	<i>P</i> = 0.04
Aleutians (December 19, 2007)	10	<i>P</i> = 0	<i>P</i> = 0	<i>P</i> = 0.01
Sea-of-Ohkost (November 24, 2008)	11	<i>P</i> = 0.02	<i>P</i> = 0.28	<i>P</i> = 0.11
Central-Mid-Atlantic (May 23, 2008)	5	<i>P</i> = 0.16	<i>P</i> = 0.03	<i>P</i> = 0.21
Chile (November 14, 2007)	13	<i>P</i> = 0.12	<i>P</i> = 0.08	<i>P</i> = 0.26
Jujuy (November 18, 2007)	7	<i>P</i> = 0.18	<i>P</i> = 0.42	<i>P</i> = 0.59
Northern Mid-Atlantic (February 8, 2008)	4	<i>P</i> = 0.18	<i>P</i> = 0.05	<i>P</i> = 0.32
Central Mid-Atlantic (August 20, 2007)	3	<i>P</i> = 0.29	<i>P</i> = 0.19	<i>P</i> = 0.28
Central Mid-Atlantic (September 10, 2008)	6	<i>P</i> = 0.31	<i>P</i> = 0.16	<i>P</i> = 0.3
Azores (April 5, 2007)	2	<i>P</i> = 0.5	<i>P</i> = 0.45	<i>P</i> = 0.47
Azores-Cape (February 12, 2007)	1	<i>P</i> = 0.42	<i>P</i> = 0.49	<i>P</i> = 0.38
Greece (February 14, 2008)	14	<i>P</i> = 0.19	<i>P</i> = 0.5	<i>P</i> = 0.11

accuracy of smaller structures remains largely an open question (eg [43]). These small features may be at the origin of the remaining misfit to data, which amounts to about 1 s for S-wave arrival for most earthquakes in our study. It must be noted that these small features are the core of geophysical discovery, as they are the evidence that the Earth is way more complex than predicted by the theory of plate tectonics [19,35,57].

We feel that the handshake between the geosciences and statistical communities, which was needed for the successful outcome of this study, is the way to a continuous and productive interaction between the 2 communities. There are many sources of errors in 3D seismic models, and the question of UQ could benefit from this interaction between the 2 communities. In fact, we could advance UQ for seismic models by answering the following questions:

(1) *How dependent is a model on its numerical methods?* Despite the simplicity of the statistical methods used, this study illustrates how to address that question. It is now feasible to use direct numerical methods such as SEM that solve the equations of motion with limited assumptions. Analyses on the modeled full waveform seismograms can test the assumptions used for the generation of seismic models. Other numerical methods such as the finite difference method (eg [15,22,58]) and discontinuous Galerkin method [10,21] offer alternatives to SEM.

(2) *How much data is needed for a model to be good enough?* The sheer number of parameters used in 3D Earth imaging renders systematic differentiation unfeasible for direct calculation of sensitivity and Hessian kernels that give indication on the data coverage and resolution [13,40]. A breakthrough was reached in 2005 with the first implementation by Tromp et al. [56] of the adjoint method proposed by Tarantola [50], which makes accurate calculation of Frechet

kernels in 3D Earth models possible. A new approach based on randomization theory has been recently proposed to estimate resolution matrix of a given 3D Earth model [29,55]. The method has been developed further to provide other resolution measures such correlation length and interparameter tradeoffs, and this with a limited computational cost [12,14].

(3) *How well can we uniquely identify sources of error?* (4) *How good is a model at representing true physics?* The most general approach to error estimates for the inverse problem is the Bayesian framework [51]. In this framework, solving the inverse problem is not limited to finding the best model but to estimating the *a posteriori* probability distribution of model parameters given *a priori* probability distribution of error on data and model parameters. In its general form, Bayesian inversion combines all types of uncertainties, and the quality of the knowledge of the 3D structure can be estimated by examining the best model, the average model, and the distribution of the probability around these models. The highly nonlinear character of seismic inversion requires a sampling method that combines efficiency and completeness. The “Monte Carlo” approach has been successfully applied to geophysical imaging problems (eg [16,38]). A recent method that has found some success in global seismology is that of the “neighborhood algorithm” [45,46] in which the data are used to determine the number of parameters.

Each of the methods presented above brings specific information on sources of uncertainty in seismic models. The issue of UQ for seismic models is becoming harder as the inverse problems are getting more complex using less approximations and more information [47]. Statistical and geoscience researchers will have to become knowledgeable of advanced aspects of methods used in the 2 communities in order to answer the question of UQ of seismic models.

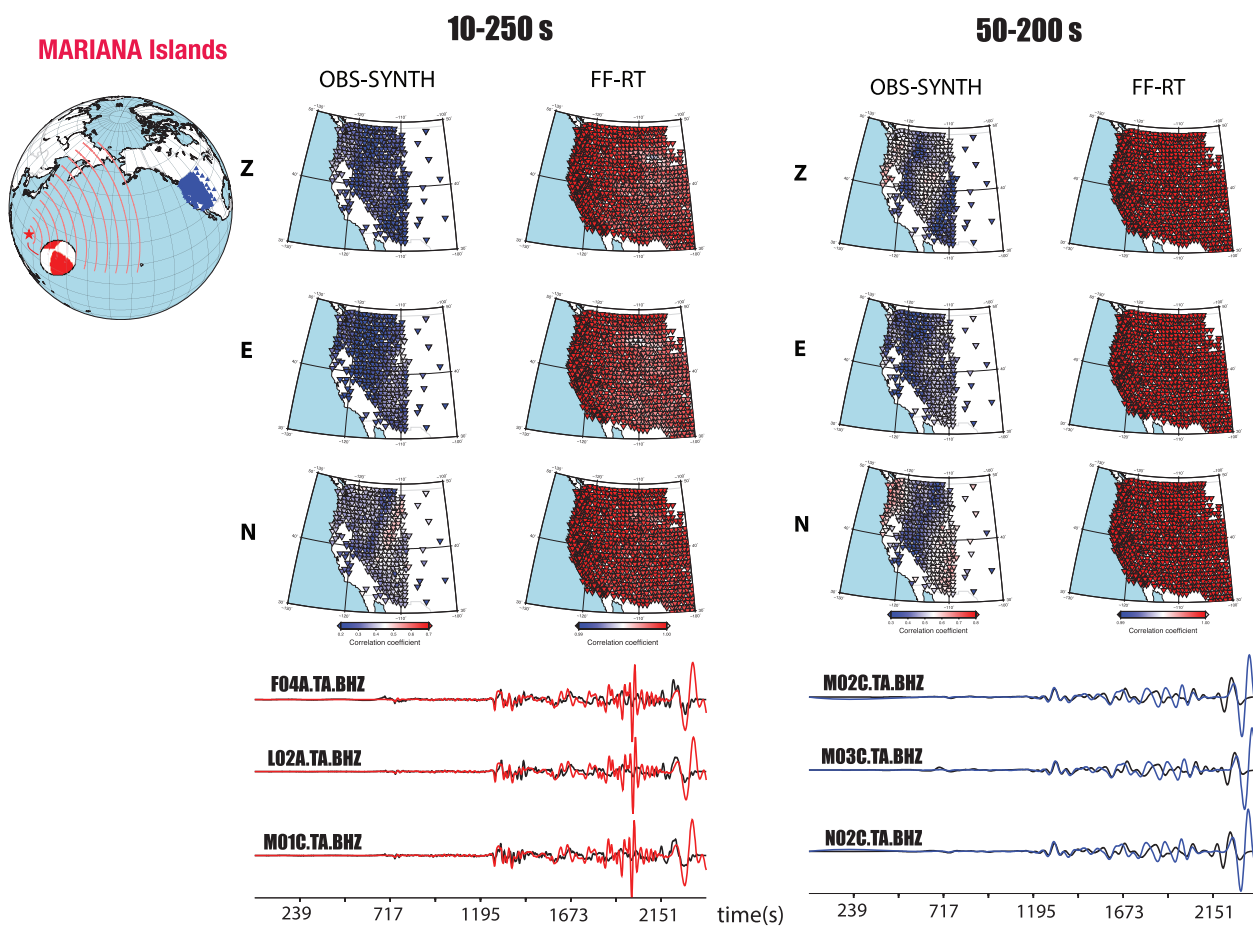


FIGURE 9 Maps of the cross-correlation coefficients between data and synthetics for the ray theory (RT) and finite frequency (FF) models and two different frequency bands (10 to 50 s and 50 to 200 s) for the Marianna earthquake (event #12). The filtered waveforms around the P and S arrivals recorded at one station central to the USArray are plotted at the bottom

ACKNOWLEDGMENTS

We thank R.M. Allen, M. Obrebski, and R. Porritt for kindly providing us the DNA09 model as well as for fruitful discussions. We also thank IRIS for making the data so easily accessible to us (data from II, IU, TA, and US networks were used in this study), the developers of the Generic Mapping Tools software (Wessel and Smith, 1991) and Basemap. We especially thank the organizers of the CoDA conference for the opportunity of cross-disciplinary handshake. We also acknowledge the CIG program and the developers who make available SPEC-FEM3D. This research was supported by U.S. Department of Energy by Los Alamos National Laboratory under contract DE-AC52-06NA25396/LA12-SignalPropagation-NDD2Ab.

ORCID

Carène Larmat  <http://orcid.org/0000-0002-3607-7558>

REFERENCES

1. K. Aki and W. H. K. Lee, *Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model*, J. Geophys. Res. 81 (1976), 4381–4399.
2. K. Aki, A. Christofferson, and E. S. Husebye, *Determination of the three-dimensional seismic structure of the lithosphere*, J. Geophys. Res. 82 (1977), 277–296.
3. D. N. Anderson, W. R. Walter, D. K. Fagan, T. M. Mercier, and S. R. Taylor, *Regional multistation discriminants: magnitude, distance, and amplitude corrections, and sources of error*, Bull. Seismol. Soc. Am. 99 (2009), 794–808.
4. G. E. Backus and J. F. Gilbert, *Numerical applications of a formalism for geophysical inverse problems*, Geophys. J. Int. 13 (1967), 247–276.
5. G. Backus and F. Gilbert, *The resolving power of gross earth data*, Geophys. J. Int. 16 (1968), 169–205.
6. G. Backus and F. Gilbert, *Uniqueness in the inversion of inaccurate gross earth data*, Philos. Trans. R. Soc. Lond. Ser. A 266 (1970), 123–192.
7. T. W. Becker, *On recent seismic tomography for the western United States*, Geochim. Geophys. Geosyst. 13 (2012), Q01W10.
8. L. Boschi, T. W. Becker, G. Soldati, and A. M. Dziewonski, *On the relevance of Born theory in global seismic tomography*, Geophys. Res. Lett. 33 (2006), L06302.
9. A. M. Dziewonski, B. H. Hager, and R. J. O'Connell, *Large-scale heterogeneities in the lower mantle*, J. Geophys. Res. 82 (1977), 239–255.
10. V. Etienne, E. Chaljub, J. Virieux, and N. Glinesky, *An hp-adaptive discontinuous Galerkin finite-element method for 3-D elastic wave modelling*, Geophys. J. Int. 183 (2010), 941–962.
11. A. Fichtner, *Full seismic waveform modelling and inversion*, Advances in geophysical and environmental mechanics and mathematics, Springer, Berlin, Heidelberg, 2011, ISBN 978-3-642-15806-3.
12. A. Fichtner and T. Leeuwen, *Resolution analysis by random probing*, J. Geophys. Res. Solid Earth 120 (2015), 5549–5573.
13. A. Fichtner and J. Trampert, *Hessian kernels of seismic data functionals based upon adjoint techniques*, Geophys. J. Int. 185 (2011), 775–798.

14. A. Fichtner and A. Villaseñor, *Crust and upper mantle of the western Mediterranean—constraints from full-waveform inversion*, *Earth Planet. Sci. Lett.* 428 (2015), 52–62.
15. R. W. Graves, *Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences*, *Bull. Seismol. Soc. Am.* 86 (1996), 1091–1106.
16. J. Hauser, K. M. Dyer, M. E. Pasyanos, et al., *A probabilistic seismic model for the European Arctic*, *J. Geophys. Res. Solid Earth* 116 (2011), B01303.
17. S.-H. Hung, F. A. Dahlen, and G. Nolet, *Wavefront healing: a banana-doughnut perspective*, *Geophys. J. Int.* 146 (2001), 289–312.
18. H. Igel, *Computational seismology: a practical introduction*, Oxford University Press, Oxford, 2017, ISBN 978-0-19-871740-9.
19. B. Isacks, J. Oliver, and L. R. Sykes, *Seismology and new global tectonics*, *J. Geophys. Res.* 73 (1968), 5855.
20. D. D. Jackson, *Interpretation of inaccurate, insufficient and inconsistent data*, *Geophys. J. Roy. Astron. Soc.* 28 (1972), 97.
21. M. Käser and M. Dumbser, *An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—I. The two-dimensional isotropic case with external source terms*, *Geophys. J. Int.* 166 (2006), 855–877.
22. K. R. Kelly, R. W. Ward, S. Treitel, and R. M. Alford, *Synthetic seismograms: a finite-difference approach*, *Geophysics* 41 (1976), 2–27.
23. B. L. N. Kennett and E. R. Engdahl, *Travel times for global earthquake location and phase identification*, *Geophys. J. Int.* 105 (1991), 429–465.
24. D. Komatitsch and J. Tromp, *Spectral-element simulations of global seismic wave propagation—II. Three-dimensional models, oceans, rotation and self-gravitation*, *Geophys. J. Int.* 150 (2002), 303–318.
25. D. Komatitsch and J. P. Vilotte, *The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures*, *Bull. Seismol. Soc. Am.* 88 (1998), 368–392.
26. D. Komatitsch, J. Ritsema, and J. Tromp, *The spectral-element method, Beowulf computing, and global seismology*, *Science* 298 (2002), 1737–1742.
27. J. J. Lévêque, L. Rivera, and G. Wittlinger, *On the use of the checker-board test to assess the resolution of tomographic inversions*, *Geophys. J. Int.* 115 (1993), 313–318.
28. M. D. Long, A. R. Levander, and P. M. Shearer, *An introduction to the special issue of Earth and Planetary Science. Letters on USArray science*, *Earth Planet. Sci. Lett.* 402 (2014), 1–5.
29. J. K. MacCarthy, B. Borchers, and R. C. Aster, *Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross-validation for large geophysical inverse problems*, *J. Geophys. Res. Solid Earth* 116 (2011), B10304.
30. J. K. MacCarthy, D. N. Anderson, and J. L. Bonner, *Combined Rayleigh- and Love-wave magnitudes for seismic event discrimination and screening analysis*, *Bull. Seismol. Soc. Am.* 103 (2013), 3334–3340.
31. M. Maceira, C. Larmat, R. W. Porritt, D. M. Higdon, C. A. Rowe, and R. M. Allen, *On the validation of seismic imaging methods: finite frequency or ray theory?* *Geophys. Res. Lett.* 42 (2015), 323–330.
32. B. F. J. Manly, *Randomization, bootstrap, and Monte Carlo methods in biology*, 3rd ed., Chapman and Hall, London, 2007, MR2257066.
33. W. H. Menke, *Lateral inhomogeneities in P velocity under the Tarbela array of the lesser Himalayas of Pakistan*, *Bull. Seismol. Soc. Am.* 67 (1977), 725–734.
34. W. Menke, *Geophysical data analysis. Discrete inverse theory*, Vol 45, Academic Press, San Diego, CA, 1989, MR3293907.
35. W. J. Morgan, *Rises, trenches, great faults, and crustal blocks*, *J. Geophys. Res.* 73 (1968), 1959–1982.
36. G. Nolet and F. A. Dahlen, *Wave front healing and the evolution of seismic delay times*, *J. Geophys. Res. Solid Earth* 105 (2000), 19043–19054.
37. M. Obrebski, R. M. Allen, M. Xue, and S.-H. Hung, *Slab-plume interaction beneath the Pacific Northwest*, *Geophys. Res. Lett.* 37 (2010), L14305.
38. M. E. Pasyanos, G. A. Franz, and A. L. Ramirez, *Reconciling a geophysical model to data using a Markov chain Monte Carlo algorithm: an application to the Yellow Sea–Korean Peninsula region*, *J. Geophys. Res. Solid Earth* 111 (2006), B03313.
39. D. Peter, L. Boschi, and J. H. Woodhouse, *Tomographic resolution of ray and finite-frequency methods: a membrane-wave investigation*, *Geophys. J. Int.* 177 (2009), 624–638.
40. N. Rawlinson, A. Fichtner, M. Sambridge, and M. K. Young, *Seismic tomography and the assessment of uncertainty*, In *Advances in Geophysics*, Vol 55, Elsevier, 2014, 1–76.
41. B. A. Romanowicz, *Seismic structure of the upper mantle beneath the United States by three-dimensional inversion of body wave arrival times*, *Geophys. J. Int.* 57 (1979), 479–506.
42. B. A. Romanowicz, *A study of large-scale lateral variations of P-velocity in the upper mantle beneath Western-Europe*, *Geophys. J. Roy. Astron. Soc.* 63 (1980), 217–232.
43. B. A. Romanowicz, *Using seismic waves to image Earth's internal structure*, *Nature* 451 (2008), 266–268.
44. C. Ronchi, R. Iacono, and P. S. Paolucci, *The “cubed sphere”: a new method for the solution of partial differential equations in spherical geometry*, *J. Comput. Phys.* 124 (1996), 93–114.
45. M. Sambridge, *Geophysical inversion with a neighbourhood algorithm. I. Searching a parameter space*, *Geophys. J. Int.* 138 (1999), 479–494.
46. M. Sambridge, *Geophysical inversion with a neighbourhood algorithm. II. Appraising the ensemble*, *Geophys. J. Int.* 138 (1999), 727–746.
47. M. Sambridge, C. Beghein, F. J. Simons, and R. K. Snieder, *How do we understand and visualize uncertainty?* *Leading Edge* 25 (2006), 542–546.
48. M. K. Sengupta and M. N. Toksöz, *Three dimensional model of seismic velocity variation in the Earth's mantle*, *Geophys. Res. Lett.* 3 (1976), 84–86.
49. A. Sieminski, J.-J. Lévêque, and E. Debayle, *Can finite-frequency effects be accounted for in ray theory surface wave tomography?* *Geophys. Res. Lett.* 31 (2004), L24614.
50. A. Tarantola, *Inversion of seismic-reflection data in the acoustic approximation*, *Geophysics* 49 (1984), 1259–1266.
51. A. Tarantola and B. Valette, *Generalized nonlinear inverse problems solved using the least squares criterion*, *Rev. Geophys.* 20 (1982), 219–232.
52. S. R. Taylor and M. N. Toksöz, *Three-dimensional crust and upper mantle structure of the northeastern United States*, *J. Geophys. Res. Solid Earth* 84 (1979), 7627–7644.
53. J. Trampert and J. Spetzler, *Surface wave tomography: finite-frequency effects lost in the null space*, *Geophys. J. Int.* 164 (2006), 394–400.
54. J. Trampert, P. Vacher, and N. Vlaar, *Sensitivities of seismic velocities to temperature, pressure and composition in the lower mantle*, *Phys. Earth Planet. Inter.* 124 (2001), 255–267.
55. J. Trampert, A. Fichtner, and J. Ritsema, *Resolution tests revisited: the power of random numbers*, *Geophys. J. Int.* 192 (2013), 676–680.
56. J. Tromp, C. Tape, and Q. Liu, *Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels*, *Geophys. J. Int.* 160 (2005), 195–216.
57. F. J. Vine and D. H. Matthews, *Magnetic anomalies over oceanic ridges*, *Nature* 199 (1963), 947.
58. J. Virieux, *P-Sv-wave propagation in heterogeneous media—velocity-stress finite-difference method*, *Geophysics* 51 (1986), 889–901.
59. Z. Wang and F. A. Dahlen, *Validity of surface-wave ray theory on a laterally heterogeneous earth*, *Geophys. J. Int.* 123 (1995), 757–773.
60. R. A. Wiggins, *The general linear inverse problem: implication of surface waves and free oscillations for Earth structure*, *Rev. Geophys.* 10 (1972), 251–285.
61. P. R. Williamson, *A guide to the limits of resolution imposed by scattering in ray tomography*, *Geophysics* 56 (1991), 202–207.
62. M. Xue and R. M. Allen, *Mantle structure beneath the western United States and its implications for convection processes*, *J. Geophys. Res. Solid Earth* 115 (2010), B07303.

How to cite this article: Larmat C, Maceira M, Higdon DM, Anderson D. *Joining statistics and geophysics for assessment and uncertainty quantification of three-dimensional seismic Earth models*, *Statistical Analysis and Data Mining* 2017;10:277–289. <https://doi.org/10.1002/sam.11353>.