AFRL-RV-PS-
TR-2020-0084

AFRL-RV-PS-
TR-2020-0084

FULL ENVELOPE TEMPLATES FOR LOW MAGNITUDE DISCRIMINATION AND YIELD ESTIMATION AT LOCAL AND REGIONAL DISTANCES

Seung-Hoon Yoo

Applied Research Associates
Weston Geophysical Group
801 North Quincy Street, Suite 700
Arlington, VA 22203

19 June 2020

Final Report

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
3550 Aberdeen Ave SE
AIR FORCE MATERIEL COMMAND
KIRTLAND AIR FORCE BASE, NM 87117-5776

DTIC COPY

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by AFMC/PA and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RV-PS-TR-2020-0084 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//	//SIGNED//
Dr. Frederick Schult	Dr. Thomas R. Caudill, Chief
Program Manager/AFRL/RVBN	AFRL Geospace Technologies Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
19-06-2020	Final Report	22 Apr 2016 – 31 May 2020

4. TITLE AND SUBTITLE
Full Envelope Templates For Low Magnitude Discrimination and Yield Estimation at Local and Regional Distances

5a. CONTRACT NUMBER
FA9453-16-C-0022

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62601F

6. AUTHOR(S)
Seung-Hoon Yoo

5d. PROJECT NUMBER
1010

5e. TASK NUMBER
EF129461

5f. WORK UNIT NUMBER
V13W

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Applied Research Associates
Weston Geophysical Group
801 North Quincy Street, Suite 700
Arlington, VA 22203

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Space Vehicles Directorate
3550 Aberdeen Avenue SE
Kirtland AFB, NM 87117-5776

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RVBN

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RV-PS-TR-2020-0084

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited. (AFRL-2020-0153 dtd 29 Sep 2020)

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Monitoring seismologists have successfully used seismic coda for event discrimination and yield estimation for over a decade. In practice seismologists typically analyze long-duration, S-coda signals with high signal-to-noise ratios (SNR) at regional and teleseismic distances, since the single back-scattering model reasonably predicts decay of the late coda. However, seismic monitoring requirements are shifting towards smaller, locally recorded events that exhibit low SNR and short signal lengths. To be successful at characterizing events recorded at local distances, we must utilize the direct-phase arrivals, as well as the earlier part of the coda, which is dominated by multiple forward scattering. To remedy this problem, we have developed a new hybrid method known as full-waveform envelope template matching to improve predicted envelope fits over the entire waveform and account for direct-wave and early coda complexity. We accomplish this by including a multiple forward-scattering approximation in the envelope modeling of the early coda. The new hybrid envelope templates are designed to fit local and regional full waveforms and produce low-variance amplitude estimates, which will improve yield estimation and discrimination between earthquakes and explosions. To demonstrate the new technique, we applied our full-waveform envelope template-matching method to the six known North Korean (DPRK) underground nuclear tests. We successfully discriminated the event types and estimated the yield for all six nuclear tests.

15. SUBJECT TERMS
seismic discrimination, yield estimation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Frederick Schult
Unclassified	Unclassified	Unclassified	Unlimited	106	**19b. TELEPHONE NUMBER** *(include area code)*

This page is intentionally left blank.

TABLE OF CONTENTS

LIST OF FIGURES

1. SUMMARY

To improve amplitude predictions and subsequent yield estimates from low yield explosions in regions of interest, Applied Research Associate's Weston Geophysical Group (Weston) has developed a new full-waveform amplitude processing technique under funding from the Air Force Research Laboratory (AFRL). We have accomplished this by implementing two significant and innovative upgrades to the conventional coda-processing methodology (c.f., Mayeda et al., 2003; Mayeda & Walter, 1996; Yoo et al., 2011): 1) a new hybrid waveform envelope modeling technique; and 2) a full 2-D calibration approach that includes both the envelope shape and amplitude attenuation along the propagation path. The new hybrid envelope modeling technique includes both multiple forward- and wide-angle- scattering effects (e.g., Saito et al., 2003) in full-waveform 'envelope templates' that fit the direct and scattered phases in local- and regional-distance observations of low-yield explosions. In the full 2-D calibration approach, all modeling parameters for the hybrid template are physics-based and calibrated for laterally varying scattering/absorbing structures and it also incorporates rigorous uncertainty estimates. We calibrated full-envelope-templates using the new hybrid envelope modeling method to account for laterally varying attenuation in East Asia. We performed a template matching to the six Democratic People's Republic of Korea (DPRK) underground nuclear tests. We were successfully able to discriminate the event type and estimate the yield. We also applied the same technique to the 2015 Tianjin explosions in China, and the 12 May 2010 event (Zhang & Wen, 2015). Our combined R&D development efforts are producing significantly improved amplitude predictions for P_n, P_g, S_n, and L_g phases, their codas and subsequently superior estimates of explosion yield and event discrimination. The new processing technique is also shifting conventional coda-based analysis from a strictly empirical basis to a physical one. This is a unique approach that will significantly improve the long-standing single-scattering methodology and, more importantly, lower magnitude thresholds and extend coda methodology to low-yield local explosions.

2. INTRODUCTION

Over the last decade coda calibrations have been successful for various seismic monitoring applications, particularly when applied to relatively large-magnitude events recorded at regional to teleseismic distances. The single back-scattering model has been successful in previous applications, primarily because it utilizes long duration S-coda signal with high SNR, which in late coda development is well-described by the single back-scattering process. However, as the monitoring target extends to smaller and locally recorded events, the lower SNR and shorter length signals force us to use the earlier part of the coda, including the direct phase arrival. The single back-scattering model does not correctly predict the direct phases and early coda since multiple forward scattering dominates this part of the signal.

To remedy this problem, we apply a new hybrid method to improve envelope fitting for the entire waveform and account for direct-wave and early coda complexity. The net result is an analytic, hybrid template model that is optimized for short-duration envelopes from tamped local explosions and small yield explosions at regional distances. This is a unique approach that will significantly improve the long-standing single-scattering methodology and, more importantly, lower magnitude thresholds and extend coda methodology to low-yield local explosions. This work not only provides more accurate yield estimation by improving fits to observations with shorter signal length, but also extends coda analysis from an empirical basis to a physical basis, which is of significant interest to monitoring agencies. Our ultimate objective in this project involved developing full-waveform, quasi-empirical model templates that fit local and regional waveform envelopes while maintaining low-variance amplitude estimates, which will then improve yield estimation and discrimination of potential future nuclear tests in the DPRK.

In this report we detail multiple layers of R&D effort to develop the full-waveform envelope template matching technique. We begin with model development of a new hybrid synthesis of the full-waveform envelope by combining an analytic solution to the Radiative Transfer Equation and a Markov Approximation to the parabolic wave equation. We move to describing the techniques used to characterize scattering and absorbing properties and the 2-D calibration to estimate envelope template shape parameters. We also discuss how we estimate the uncertainty in the attenuation model and yield estimates using a hierarchical uncertainty model. Finally, we apply the newly developed full-waveform envelope template method for event discrimination and yield estimation to 4 specific datasets: 1) DPRK nuclear tests; 2) 2015 Tianjin explosions; 3) the low yield event in North Korea on 12 May 2010; and 4) the low SNR events at the Musan mine in North Korea. Our results from these 4 datasets show that the new full-waveform envelope template-matching method significantly improves upon longstanding single-scattering coda prediction techniques. More importantly, the new method allows monitoring seismologists to extend coda-based techniques to lower magnitude thresholds and low-yield local explosions.

3. TECHNICAL APPROACH

This section provides background information and pre-processing on the datasets used in this analysis and details on algorithm development.

3.1. Building the Database

3.1.1. Earthquake Database

The ultimate goal of this project is to fully calibrate the path and site effects for all regional seismic phases (*Pn*, *Pg*, *Sn*, and *Lg*) and their coda for precise yield estimation and discrimination of potential future nuclear tests in the aseismic North Korean region. The use of all available seismic networks and experimental data was maximized to improve path coverage of the phases over the North Korean region.

First, a waveform database was constructed from Korea Meteorological Administration (KMA) and Korea Institute of Geoscience and Mineral (KIGAM) seismic networks. We acquired continuous waveforms recorded at 87 stations in the KMA seismic network (50 Broadband, 37 Short-period stations between Jan 2003 and April 2015), and 23 stations in KIGAM seismic network (18 Broadband and 5 Short-period stations between July 2003 and Sep 2014). Also acquired was a triggered (or event-windowed) dataset recorded at all KMA and KIGAM seismic networks (~1,500 events recorded at 256 broad and short-period sensors between Jan 1999 and May 2016).

Next, all available continuous seismic data near the Punggye-ri Nuclear Test Site were downloaded from the Incorporated Research Institutions for Seismology (IRIS), in order to construct databases using the following 5 seismic networks:

1) North Korea Seismic Deployment (XL; http://www.fdsn.org/networks/detail/XL_1995)
2) NE China Seismic Experiment (XI; http://www.fdsn.org/networks/detail/XI_1998),
3) Natural & Man-Induced Seismicity in the Yanquing-Hualia Basin and the Heicheng Area, China (XG; http://www.fdsn.org/networks/detail/XG_2002, doi:10.7914/SN/XG_2002)
4) North China Interior Structure Project-Experiment 6 (1A; http://www.fdsn.org/networks/detail/1A_2007, doi:10.7914/SN/1A_2007)
5) Northeast China Extended Seismic Array (YP; http://www.fdsn.org/networks/detail/YP_2009, doi:10.7914/SN/YP_2009)

We then developed a unified database by merging all individual databases into a single CSS 3.0 database (Figure 1). This unified database also included waveforms from the Mudanjiang (MDJ) station in the New China Digital Seismograph Network; 12 stations located along the North Korea-China and Russia borders (triggered data from China National Seismic Network, Ussuriysk Array, and Vladivostok); and 20 (out of 127) stations in the YP PASCCAL deployment within 4 degrees from the Punggye-ri Nuclear Test Site.

Event origin information from various contributors was collected to form a unified event origin database. We searched all of the origins for M1.5+ events occurring between 1995 and 2015, within a 6-degree radius from the Ganseong station (KSA), which is the closest station from the

Punggye-ri Nuclear Test Site among the South Korean seismic networks. An initial origin database was formed by combining ~25,000 origins from the IRIS database and ~1,500 origins from the local event catalogs published by KMA (http://web.kma.go.kr) and KIGAM (https://quake.kigam.re.kr).

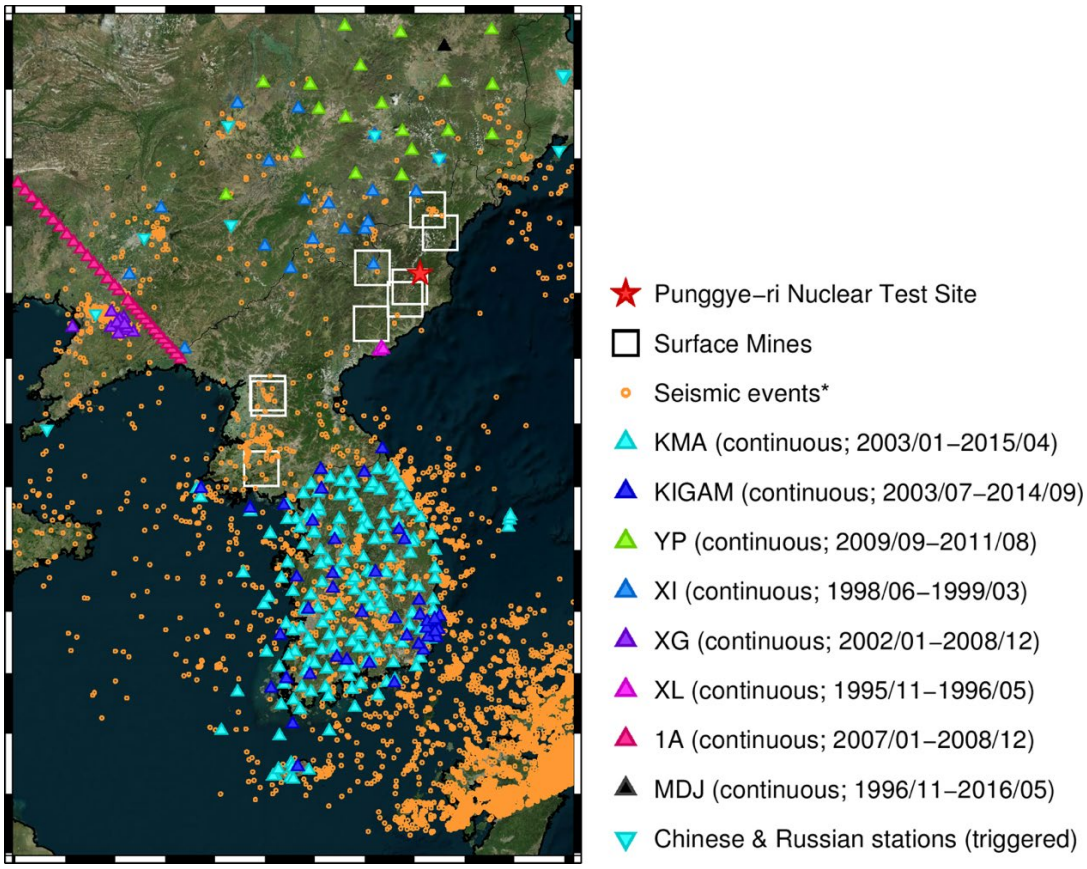

Figure 1. Map showing the location of seismic stations and events in the new database that were used during this study. *Seismic stations in the waveform database are plotted as color-coded triangles (continuous) and inverted triangles (triggered), and events in the initial origin database are plotted as orange circles. The North Korean Punggye-ri Nuclear Test Site is shown with a red star, and the locations of the verified surface mines are plotted as white boxes.*

3.1.2. Mining Explosion Database in North Korea

North Korea is rich in minerals, and mining is important to the country's economy. Mining activities can be found in most areas of North Korea, but especially in the province of Hamgyeong-do where the Punggye-ri Nuclear Test Site is located. It is important to characterize the level of background seismic activity and magnitude of mining events in North Korea for discrimination and event screening purpose, particularly near the Punggye-ri Nuclear Test Site (e.g., Zhang & Wen, 2015).

First, we searched information on mining activities through various sources (Korean journals, reports, news, and websites) to look for mines in North Korea, and then cross-validated the location of some of the surface mines with Google Earth (shown as white boxes in Figure 1). For example, Figure 2a shows a satellite image of one of the surface mines known as the Musan iron mine

(42.23º N, 129.26º E). The Musan mine is located in north-east North Korea in North Hamgyong Province (70 miles north of the test site). There was a M_L 1.8 (m_b 2.6) event in the International Data Centre (IDC) event catalog well recorded on the YP network located near the mine. We ran a cross-correlation detector (e.g., Gibbons & Ringdal, 2006) on the YP network data using the IDC event as a master event and found over 350 events with a high cross-correlation coefficient.

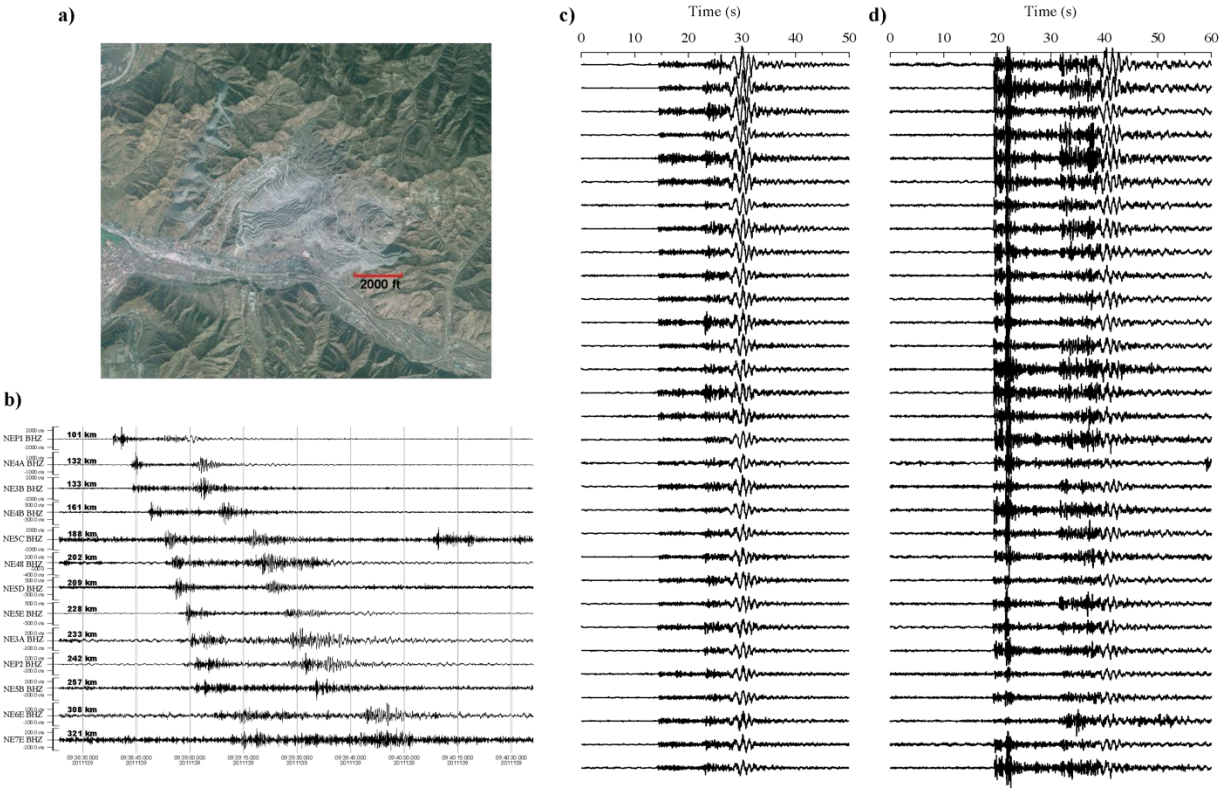

Figure 2. a) Zoomed-in view of the Musan iron mine, visible from Google Earth (42.23º N, 129.26º E), located in the North Hamgyong Province in north-east North Korea. b) A suspected mining explosion near the mine that was well recorded on the YP network. c) and d) Example of detected events using cross-correlation detection at the NE3C and NEP1 YP stations, located 70 and 100 km away from the mine, respectively.

3.1.3. Data Review and Quality Control

Initially we performed quality control on the databases to determine a final origin database for use in in-depth analysis and further processing. Using the initial origin information and available stations for each origin, a theoretical first *P* arrival time was computed for each event-station pair using the IASP91 velocity model (Kennett & Engdahl, 1991), which we used to set a 1-minute window, 30 seconds before and after the *P* arrival. Next we applied a short-term/long-term average (STA/LTA) first-break pick detection to each given window. Then the quality of waveforms was visually inspected, and adjustments were made on the auto-picks event by event for all sub-networks described below.

The conventional coda calibration method is based on the simple exponential coda envelope which doesn't fit the direct arrival and earlier coda well; therefore, the method is designed to avoid the direct and early coda windows for amplitude measurement. This makes the conventional coda calibration method difficult to apply to a small event with a short duration of signal. In contrast, the new hybrid template method simulates the entire waveform envelopes from the direct arrival through the coda and uses the entire signal windows for measurement, so the onset of the phase arrival time is important in the new method.

The first arrivals (*Pn* or *Pg*) comprise most of our arrival database. The auto-pick made by the short-term average to long-term average (STA/LTA) detector is the largest portion in the first arrival picks, followed by the hand-picked and collected arrivals from ISC bulletin. The other later arrivals such as (*Pg*, *Sn* and *Lg*) are very small portion of the database and are mostly RSTT predictions based on given origin information. We manually refined the arrival picks previously made by the STA/LTA detector and the RSTT predictions, and added more high-quality picks for the other later arrival such as *Pg*, *Sn* and *Lg*. For effective processing, we designed a semi-automated pick refinement process that combines Modified Energy Ratio (MER; Wong et al., 2009) and Support Vector Machine (SVM; Cortes & Vapnik, 1995).

Modified energy ratio methods are based on the differences in the energy densities of the noise and signal in the same manner as the STA/LTA method. While the time pick occurs at the maximum of the rising slope of the ratio in the STA/LTA method, the time pick occurs at the peak of the MER attribute in the MER method. For this reason, the MER yields more consistent results than the STA/LTA method for data with various SNR for the minimum-phase signal. Like all other STA/LTA variants, however, the MER method requires an appropriate time window for the non-minimum-phase signal due to the impact of the delayed maximum amplitude signal.

We applied a supervised pattern recognition technique called the Support Vector Machine (SVM) to provide the optimal time window for the MER auto-picker. SVM is one of the most popular supervised machine learning techniques because of its effectiveness with high dimensional features. SVM is easily implemented into automatic processing and gives an unambiguous result to an ambiguous problem.

First, for training data for the SVM classifier, we collected all hand-picked arrivals and corresponding waveforms. We then resampled the waveforms to 20 samples per second, cut out two seconds before and after the pick (four seconds total), and normalized it with the average amplitude of the first two seconds. We have trained the SVM classifier using training data and have applied to the unrefined data. Figure 3 shows an example of how the SVM prediction and MER auto-picker effectively refine the arrival picks. The top plot shows a high-pass filtered root-mean-squared (RMS) envelope calculated by the vector sum of three-component velocity seismogram. The following plots show SVM prediction, MER, and time derivative of MER, respectively. In the process, an analysist only selects which window will be refined by picking around the SVM predictions, and the code automatically searches the maximum MER peak within the selected time window. We found that the peak of the time derivative of the MER show great agreement with the analysist pick, so we designed the code picks to be the arrival pick at the maximum of the time derivative of the MER.

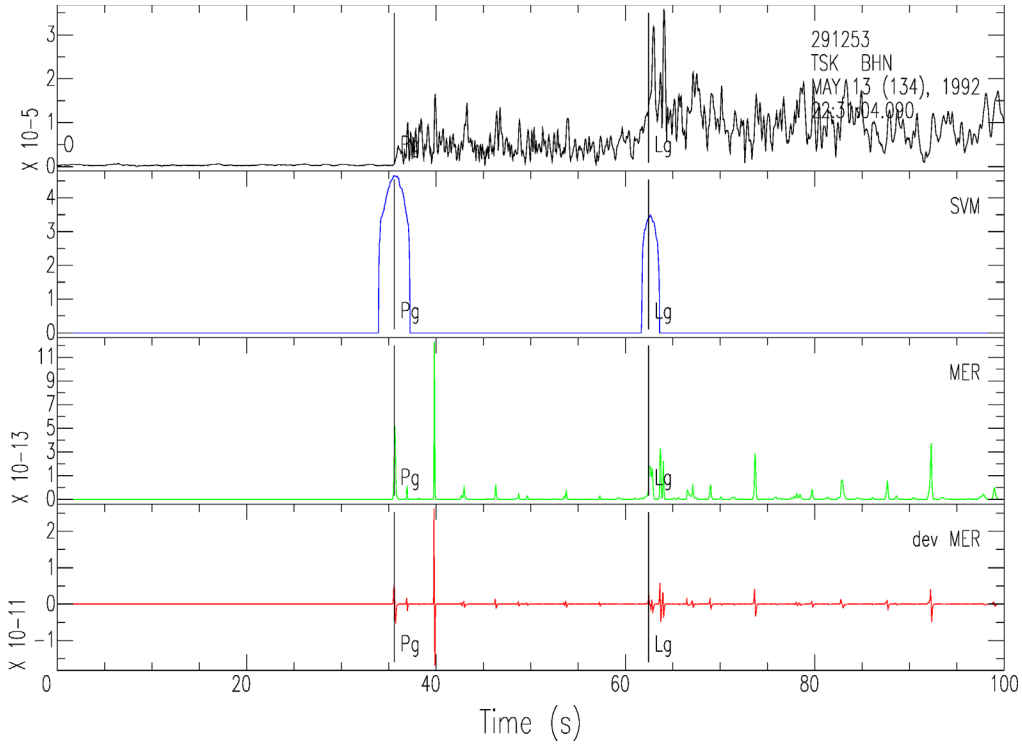

Figure 3. Example of the semi-automated pick refinement. *From top to bottom, RMS envelope calculated by a vector sum of the three-component velocity seismogram, the SVM prediction, the MER and time derivative of the MER.*

3.1.4. Updates to the Database

During the 4-years of this project, we continuously collected and processed additional data to improve the coverage and resolution of the calibration parameters for the Korean Peninsula region, including those areas surrounding the Punggye-ri nuclear test site in North Korea. Although we already collected and analyzed the seismic data for a long period of time at the start of the project (1990-2014), due to low seismicity, lack of available seismic networks in North Korea and the strong attenuation of the East Sea, the current resolution of some sub regions is unsatisfactory. In order to improve the resolution of the calibration, we continued to collect and process additional data and updated the calibration model to include more recent earthquakes in North Korea. The final database at the end of the contract includes over 800,000 of 3-C seismograms for 10,083 events recorded at 1,412 stations.

3.2. Algorithm Development

3.2.1. Hybrid Synthesis of Full-waveform Envelope

A main focus of the project has been on developing a new forward-modeling algorithm and deriving modeling parameters that can simulate an entire waveform envelope from the direct arrival through the coda. The new algorithm is required to correctly describe the direct wavefront, the envelope broadening caused by multiple forward scattering, as well as the late coda caused by multiple wide-angle scattering. The algorithm needs to be fully analytic and parameterized for a

physical model. It must also have a reasonable computational cost for application to a calibration approach using a massive amount of data. However, the algorithm should also have some flexibility in its formulation to allow for extra calibration of unmodeled effects (e.g., site amplification, scattering conversion, and phase blockage effect, etc.).

We compared previously proposed modeling techniques to determine the most suitable algorithm for our objective. To summarize, we chose a hybrid formulation originally proposed by Saito et al. (2003). However, we have replaced the forward-scattering term (from Markov approximation) and multiple scattering term (from Radiative Transfer Theory; RTT) in the Saito et al. (2003) formulation with time-domain approximate solutions (Paasschens, 1997; Sato et al., 2012). By using time-domain solutions of the RTT and Markov approximations in our new hybrid-envelope method, we can keep the computational complexity and cost as low as that used for the single-backscattering model in the current coda calibration method (Mayeda & Walter, 1996).

We note that many studies we reviewed were based on 2-D random or Gaussian media, which are far from a realistic representation of the Earth's inhomogeneity. Thus, we analytically derived all modeling parameters for a 3-D von Karman random media with a non-integer decay power. Our new forward-modeling algorithm is fully analytic and parameterized for a physical model. It allows us to simulate the multiple-phase envelope without needing prior calibration for coda-peak velocity and shape parameters. In the following section, we provide details of our new hybrid formulation.

The ensemble-averaged envelope computed from full-waveform modeling in random media is the most complete solution to the problem we are trying to solve. However, due to its high computational cost, full-waveform based methods have so far been applicable only in 2-D space. RTT developed for describing energy propagation in inhomogeneous media presents a good alternative to a full-waveform modeling approach. In particular, a Monte-Carlo simulation based on RTT is a useful research tool that can model a probabilistic distribution of scattered energy density in time and space. The Monte-Carlo simulation method has been used for the synthesis of high-frequency seismograms in complex structures (e.g., Hoshiba, 1994; Sanborn et al., 2017; Yoshimoto et al., 1997); however, the computational cost of Monte-Carlo simulation is still too high to be applied in our study. We require compact, analytic solutions to the forward and inverse problems that can incorporate large numbers of events and stations.

Radiative Transfer Solution for Coda Excitation

Coda excitation is primarily caused by wide-angle multiple scattering due to short-wavelength spectral components of velocity inhomogeneities. An analytic solution to the conventional radiative transfer equation (RTE), describing the isotropic impulsive radiation in a homogeneous random media, can reasonably explain the coda excitation in the multiple scattering regime (e.g., Sens-Schönfelder & Wegler, 2006; Wegler et al., 2006).

The RTE for a 3-D scattering medium characterized by the isotropic scattering coefficient g_0, and background velocity V_0, where the total energy W is spherically and impulsively radiated from a point source, is given by a convolution integral of the form:

$$E(\mathbb{x},t) = WG(\mathbb{x},t) + V_0 g_0 \iiint \int_{-\infty}^{\infty} G(\mathbb{x}-\mathbb{x}',t-t')E(\mathbb{x}-\mathbb{x}',t-t')\,dt'd\mathbb{x}'.$$

eq. 1

Here, the Green's function is given by:

$$G(\mathbb{x},t) = \frac{1}{4\pi V_0 r^2} H(t)\delta\left(t-\frac{r}{V_0}\right)e^{-(V_0 g_0 t)},$$

eq. 2

where $r = |\mathbb{x}|$.

Using this convolution integral equation (eq. 1), and taking the Fourier transform in space and the Laplace transform in time of the integral equation, we have

$$\tilde{\tilde{E}}(\mathbb{k},s) = \frac{W\tilde{\tilde{G}}(\mathbb{k},s)}{1-V_0 g_0 \tilde{\tilde{G}}(\mathbb{k},s)}.$$

eq. 3

The local energy density $\tilde{\tilde{G}}(\mathbb{k},s)$ can be defined as an analytic form describing the conventional solution for the impulsive radiation (Sato et al., 2012; eq. 8.24). Using the inverse Fourier-Laplace transformation, we can formally derive the space-time distribution of energy density. In practice, we can numerically evaluate the expression by substituting the inverse Fourier transform for the inverse Laplace transform (replacing s with $-i\omega$) and using a 2-D FFT (Sato et al., 2012; Zeng, 1993).

As an alternative to the numerical approach, Paasschens (1997) heuristically proposed an approximate time-domain solution of the radiative transfer equation for a 3-D isotropic scattering medium (eq. 1):

$$E(r,t) \approx \frac{We^{-(V_0 g_0 t)}}{4\pi V_0 r^2}\delta\left(t-\frac{r}{V_0}\right) + W\frac{[1-r^2/(V_0 t)^2]^{1/8}}{[4\pi V_0 t/(3g_0)]^{\frac{3}{2}}}e^{-(V_0 g_0 t)}\mathrm{M}\left[V_0 g_0 t\left(1-\frac{r^2}{V_0^2 t^2}\right)^{\frac{3}{4}}\right]H\left(t-\frac{r}{V_0}\right),$$

eq. 4

where

$$\mathrm{M}(x) \equiv 8(3x)^{-3/2}\sum_{n=1}^{\infty}\frac{\Gamma\left(\frac{3}{4}n+\frac{3}{2}\right)}{\Gamma\left(\frac{3}{4}n\right)}\frac{x^n}{n!} \approx e^x\sqrt{1+2.026/x}$$

eq. 5

Figure 4 (from Fig 2 in Paasschens, 1997) shows that the difference between exact solution of eq. 1 and its approximation is barely visible on the scale from the direct arrival to coda. The relative error of the approximation is on the order of ~2% except for the direct arrival (Paasschens, 1997). Moreover, in our hybrid envelope, we use this time-domain solution to model coda excitation by

replacing the direct term in the first term of eq. 4 with the solution of Markov approximation. Thus, the small inaccuracy issue near the ballistic peak of the approximate solution is not an issue in our hybrid formulation.

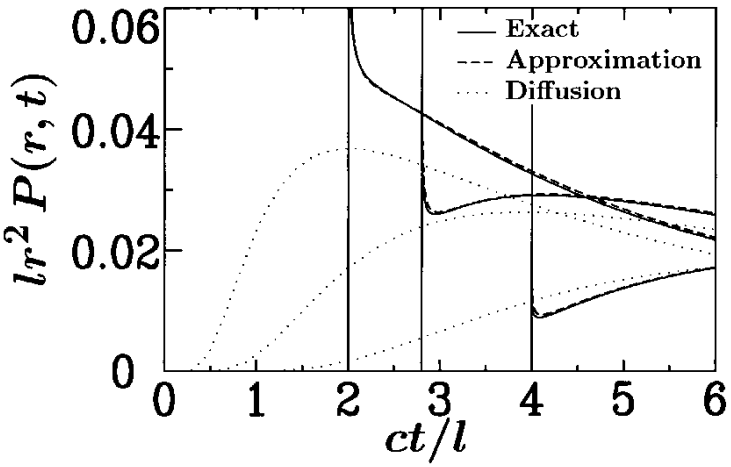

Figure 4. Fig. 2 from Paasschens (1997, Physical Review E). Angular average of the intensity P(r,t) for three dimensions as a function of time t, at distance r = 2.0l, 2.8l, and 4.0l, from left to right. *The solid lines are the exact result, which is very close to the interpolation formula (dashed lines). The dotted lines are the diffusion result. The intensity has a minimum for r greater than some* r_c .

Markov Approximation for the Direct and Early coda

When the wavelength is much smaller than the correlation distance, diffraction and scattering in the forward direction dominates. In this scattering regime, we may use the parabolic wave equation, which describes wave propagation in one direction. Assuming a slowly varying harmonic wave U, propagating in the z direction in an inhomogeneous media with fractional velocity fluctuation (ξ), the parabolic wave equation is given by

$$2ik_0\partial_z U + \Delta_\perp U - 2k_0^2\xi U = 0$$

eq. 6

where $\Delta_\perp \equiv \partial_x^2 + \partial_y^2$ is the Laplacian operator in the transverse plane and $k_0 = \omega/V_0$ is wavenumber. Eq. 6 describes the scattering in the forward direction, neglecting backward scattering.

Next, we define the two-frequency mutual coherence function (TFMCF), Γ_2, which is the correlation of the wave field between two locations on the transverse plane ($\mathbb{x}'_\perp, \mathbb{x}''_\perp$) and two frequencies ($\omega', \omega''$) (a rigorous derivation can be found in Sato et al., 2012, Chapter 9):

$$\Gamma_2(\mathbb{x}'_\perp, \mathbb{x}''_\perp, z, \omega', \omega'') = \langle U(\mathbb{x}'_\perp, z, \omega')U(\mathbb{x}''_\perp, z, \omega'')^*\rangle.$$

eq. 7

Then, the intensity spectral density (ISD) of wavefield can be written using the inverse Fourier transform of Γ_2 with respect to angular frequency difference, $\omega_d \equiv \omega' - \omega''$:

$$\hat{I}(z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Gamma_2(\mathbb{x}_{\perp d} = 0, z, \omega', \omega'') e^{-i\omega_d\left(t-\frac{z}{V_0}\right)} d\omega_d,$$

eq. 8

where $\mathbb{x}_{\perp d} = |\mathbb{x}'_{\perp} - \mathbb{x}''_{\perp}|$. This ISD describes the mean-square (MS) bandpass-filtered envelope centered at angular frequency $\omega_c \equiv (\omega' + \omega'')/2$.

Multiplying eq. 6 by U and taking the ensemble average,

$$2i\partial_z\Gamma_2 + \left(\frac{\Delta'_{\perp}}{k'_0} - \frac{\Delta''_{\perp}}{k''_0}\right)\Gamma_2 - 2\langle(k'_0\xi' - k''_0\xi')U'U''^{*}\rangle = 0$$

eq. 9

and substituting the relation:

$$\langle(k'_0\xi' - k''_0\xi')U'U''^{*}\rangle = -\frac{i}{2}[(k'^2_0 - k''^2_0)A(0) - 2k'_0k''_0A(r_{\perp d})]\Gamma_2,$$

eq. 10

where function *A* is defined by the longitudinal integral of the autocorrelation function of random media, we have the master equation of the TFMCF:

$$2i\partial_z\Gamma_2 + \left(\frac{\Delta'_{\perp}}{k'_0} - \frac{\Delta''_{\perp}}{k''_0}\right)\Gamma_2 + i[(k'^2_0 - k''^2_0)A(0) - 2k'_0k''_0A(r_{\perp d})]\Gamma_2 = 0$$

eq. 11

Eq. 11 can be used to extrapolate the TFMCF recursively in distance away from the source once the TFMCF at a slightly smaller distance is known. The term '*Markov approximation*' has its roots in the concept of a Markov process, in which the probability of future events is dependent only on most recent events (Tatarski et al., 1961). The Markov approximation of the parabolic wave equation is an effective method for synthesizing the multiple forward-scattering effects caused by long-wavelength spectral components of velocity inhomogeneities (2-D Gaussian random media; Fehler et al., 2000; numerical solutions in 3-D von Karman random media; Saito et al., 2002).

The analytic forms for Γ_2 and the Markov envelope G_M without a wandering term, at a distance r, are

$\Gamma_2(r, \theta_d, \omega_d, \omega_c) = \frac{1}{4\pi V_0}\frac{2e^{i\pi/4}\sqrt{t_M\omega_d}}{\sin\left(2e^{i\pi/4}\sqrt{t_M\omega_d}\right)}$; and

eq. 12

$$G_M(r,t;\omega_c) = \frac{1}{2\pi r^2}\int_{-\infty}^{\infty} \Gamma_2(r, \theta_d, \omega_d, \omega_c) e^{-i\omega_d(t-r/V_0)} d\omega_d,$$

eq. 13

where t_M is the characteristic time that controls envelope and peak delay from multiple forward scattering (M. Fehler et al., 2000; Sato et al., 2012). t_M is proportional to the square of the travel distance and is frequency independent for Gaussian random media and frequency dependent for a media having rich short-scale components, such as von Karman random media with $k = 0.1$. Fehler et al. (2000) derived the characteristic time for Gaussian random media as a simple function

of the form, $t_M = \frac{\sqrt{\pi}\varepsilon^2 r^2}{2aV_0}$. Saito et al. (2002) numerically estimated the characteristic time for von Karman random media having non-integer k. Most recently, Sato (2016) derived an analytic form of the characteristic time for von Karman random media, which we use in our hybrid method. A later section will contain details about t_M in Sato (2016).

We can easily evaluate eq. 13 using the Discrete Fourier Transform (DFT) for a given t_M. Instead, we implemented a time-domain solution obtained by Shishov (1974) to minimize computational complexity:

$$G_M(r,t;\omega_c) = \frac{1}{4\pi V_0 r^2}\frac{\pi^2}{2t_M}\sum_{n=1}^{\infty}(-1)^{n+1}n^2 e^{-\frac{\pi^2 n^2}{4t_M}(t-r/V_0)}\ H\left(t-\frac{r}{V_0}\right)$$

eq. 14

using a series expansion:

$$\frac{s_0}{\sin(s_0)} = 1+\sum_{n=1}^{\infty}\frac{(-1)^n 2{s_0}^2}{{s_0}^2-\pi^2 n^2}.$$

eq. 15

eq. 14 can be cast in a simplified form using an elliptic theta function of the fourth kind (see Sato et al., 2012, p 351):

$$\vartheta_4(v,q) = 1+\sum_{n=1}^{\infty}(-1)^n q^{n^2}\cos 2nv,$$

eq. 16

leading to

$$G_M(r,t;\omega_c) = \frac{1}{4\pi V_0 r^2}\frac{\pi^2}{16t_M}\vartheta_4''\left(0, e^{-\frac{\pi^2}{4t_M}(t-r/V_0)}\right) H\left(t-\frac{r}{V_0}\right),$$

eq. 17

where ϑ_4'' is the second derivative of elliptical theta function of the fourth kind with respect to v. The infinite series sum in the elliptical theta function converges very quickly due to exponential decay in time. Thus, eq. 17 has a much lower computational cost compared to eq. 13. Moreover, we only need to compute the Markov envelope once in practice, because the Markov envelope is scalable by t_M. By using time-domain solutions of RTT and Markov approximations in our hybrid envelope, we are able to keep the computational complexity and cost the same as the single-backscattering model in the current coda calibration method (Mayeda & Walter, 1996).

Hybrid Formulation

Figure 5 shows a schematic illustration of energy-packet propagation in a scattering medium based on the concept of RTT. For smooth random media, small angle scattering is dominant around the forward direction. The Markov approximation method, which correctly accounts for multiple-forward scattering and disregards wide-angle scattering, reliably models propagation around the direct path. On the other hand, for rough random media, the short-wavelength components of

inhomogeneity efficiently excite wide-angle scattering. In this situation, scattering distributed far from the direct path contributes to the excitation of coda waves.

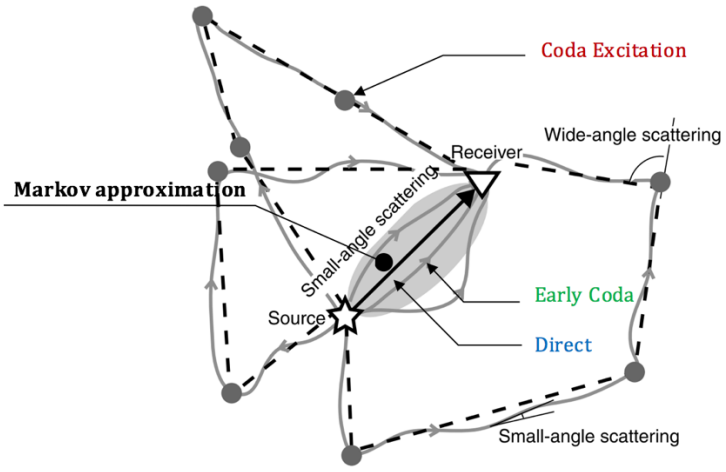

Figure 5. Modified from Figure 7 in Saito et al. (2003). Schematic illustration of energy-packet propagation in scattering media based on the concept of the radiative-transfer theory. *The MS envelope can be considered as the time trace of energy density, which is composed of energy packets radiated from the source. Gray curves show energy-packet paths. The straight solid line between the source and the receiver shows the direct path. The gray curves around the direct path show multiple energy-packet paths with small angle scattering, which form the early part of the envelope, while the paths containing wide-angle scattering form coda excitation. The Markov approximation method can model multiple paths due to small angle scattering around the direct path between a source and a receiver, as shown by the shaded area. The multiple isotropic-scattering model with momentum-transfer scattering coefficients models energy-packet paths as straight trajectories composed of isotropic scattering (dashed lines), in order to model the coda excitation.*

The initial pulse in RTT-derived envelopes are too impulsive due to the use of the delta function propagator in the RTE (see green curves in Figure 11). Ultimately, the RTT solution fails to explain the envelope near the direct arrival because it neglects the narrow angle scattering around the forward direction near the direct arrivals. On the other hand, assumption of the parabolic wave equation neglects the wide-angle scattering and the coda excitation is not modeled in the Markov approximation (see red dotted curves in Figure 11).

In order to model the coda excitation due to wide-angle scattering and envelope broadening due to multiple forward scattering, Saito et al. (2003) first proposed to replace the direct propagation term of the RTE solution with the solution of the Markov approximation envelope for the direct propagation:

$$E(r,t;\omega_c) = WG_M(r,t;\omega_c) + E^1(r,t;\omega_c) + E^M(r,t;\omega_c)$$

eq. 18

where E^1 and E^M are a single- and multiple-scattering energy envelopes, respectively.

The propagator used in the RTE solution (Saito et al., 2003) was still a delta function-type, and thus Sato et al. (2004) proposed an improved hybrid method that utilized the Markov approximation as a propagator in the radiative transfer integral equation:

$$\tilde{\hat{E}}(\mathbb{k}, s) = \frac{W\widetilde{\widehat{G_M}}(\mathbb{k}, s)}{1 - V_0 g_0 \widetilde{\widehat{G}_M}(\mathbb{k}, s)}.$$

eq. 19

We note that, in a similar fashion as with eq. 3, we can solve eq. 19 using the 2-D FFT technique.

Figures 6 and 7 show comparisons between the two hybrid envelopes of Saito et al. (2003) and Sato et al. (2004) in a 2-D von Karman random media ($k = 0.1$, $\varepsilon = 0.05$, $a = 5$), with an envelope synthesized using the finite-difference method. The results show that the envelopes computed with the hybrid methods agree well with envelopes averaged over an ensemble of finite-difference simulations of the wave equation for a suite of random media. More interestingly, the simple hybrid method proposed by Saito et al. (2003) explains both the envelope broadening around the peak and the coda excitation with a far lower computational cost.

For our template modeling, we chose the hybrid approach proposed by Saito et al. (2003) because of its computational simplicity. The hybrid method by proposed Sato et al. (2004) requires computation of the Markov approximation kernel in time and space, as well as a 2-D Fourier-Laplace transform for the convolution integral equation. It therefore has a much higher computational complexity and cost compared to the method in Saito et al. (2003). We can also accelerate the calculation using an approximate analytic time-domain solution (e.g., RT solution for Paasschens, 1997; time-domain solution of the Markov approximation for Sato et al., 2012).

Replacing the first term in the 3-D radiative-transfer solution (eq. 4) with the Markov envelope (eq. 17), and including a scattering energy loss term $e^{-(V_0 g_{S0} t)}$(we will explain further in 3.2.2), we derive a 3-D hybrid envelope:

$$E(r,t) = \frac{1}{4\pi V_0 r^2}\frac{\pi^2}{16 t_M}\vartheta_4''\left(0, e^{-\frac{\pi^2}{4t_M}(t - r/V_0)}\right) e^{-(V_0 g_{S0} t)} H\left(t - \frac{r}{V_0}\right) + W\frac{[1 - r^2/(V_0 t)^2]^{1/8}}{[4\pi V_0 t/(3g_m)]^{\frac{3}{2}}} e^{-(V_0 g_m t)} \mathrm{M}\left[V_0 g_m t\left(1 - \frac{r^2}{V_0^2 t^2}\right)^{\frac{3}{4}}\right] H\left(t - \frac{r}{V_0}\right),$$

eq. 20

Based on eq. 20, the bandpass filtered velocity envelope centered at frequency f at a distance r can be calculated by

$$A(r, t, f; t_M, g_0, b) = S(f)P(f)[2\rho_0^{-1}E(r, t, f; t_M, g_0)e^{-bt}]^{1/2} + N(f)$$

eq. 21

where S, P and N are respectively the source, frequency dependent site effect, and noise spectra, and b is the intrinsic attenuation factor ($b \equiv \frac{2\pi f}{Q_i}$) and ρ_0 is density. Doubling the energy density accounts for the boundary condition for energy transfer in a half-space (Sens-Schönfelder & Wegler, 2006).

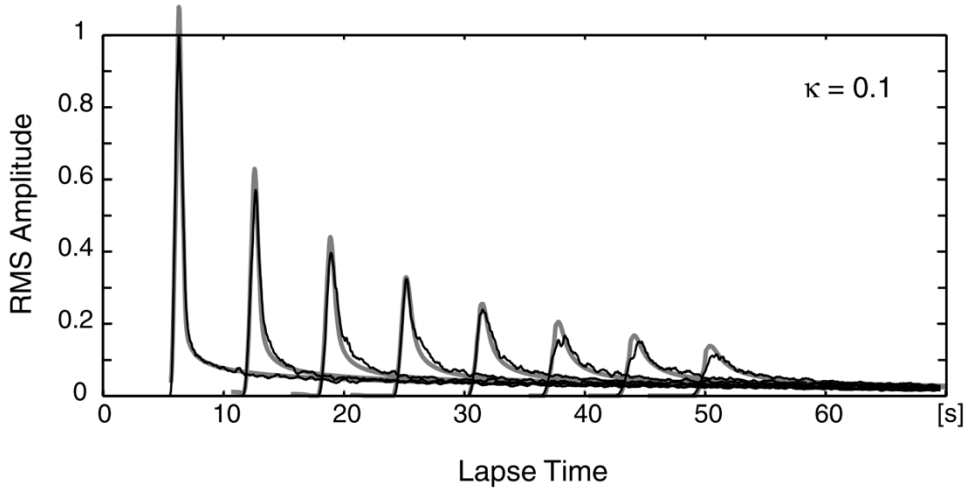

Figure 6. Figure 8 in Saito et al. (2003). A comparison of root-mean-square (RMS) envelopes derived from the new hybrid method (gray curves) with envelopes synthesized by the finite-difference method (solid curves) for k=0.1.

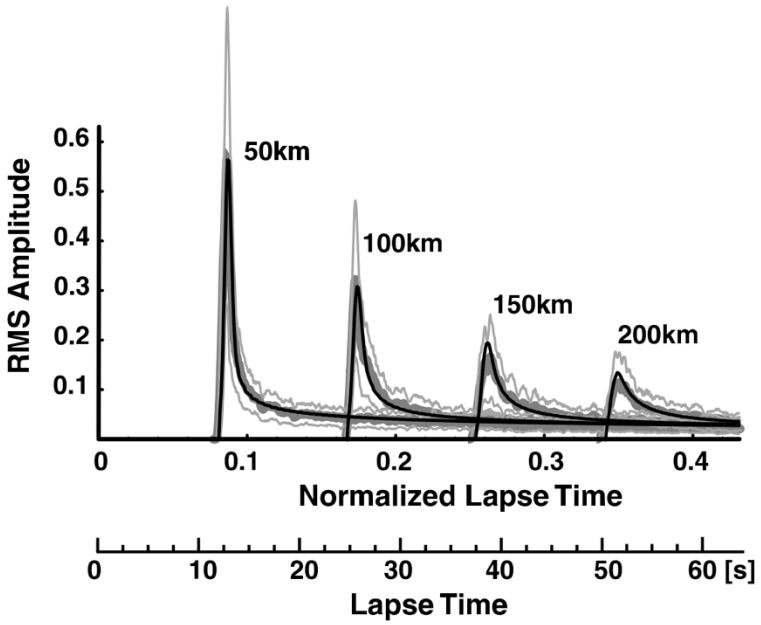

Figure 7. Figure 12 from Sato et al. (2004). Root-mean-square (RMS) envelopes of waves for a 2-Hz Ricker wavelet source radiation in 2-D von Karman–type random media ($\boldsymbol{k = 0.1}$, $\boldsymbol{\varepsilon = 0.05}$, $\boldsymbol{a = 5}$). *The numbers above the curves indicate the distance from the source. Solid curves and thick-shaded curves show RMS envelopes based on the hybrid synthesis versus those found through finite-difference simulations, respectively. Fine-shaded curves indicate ±1 standard deviation of finite-difference envelopes.*

Figure 8 illustrates some results for a 3-D von Karman-type random media. We show envelopes computed using the hybrid method compared to those computed using the analytic solution of the radiative transfer equation and the Markov approximation of parabolic wave equation.

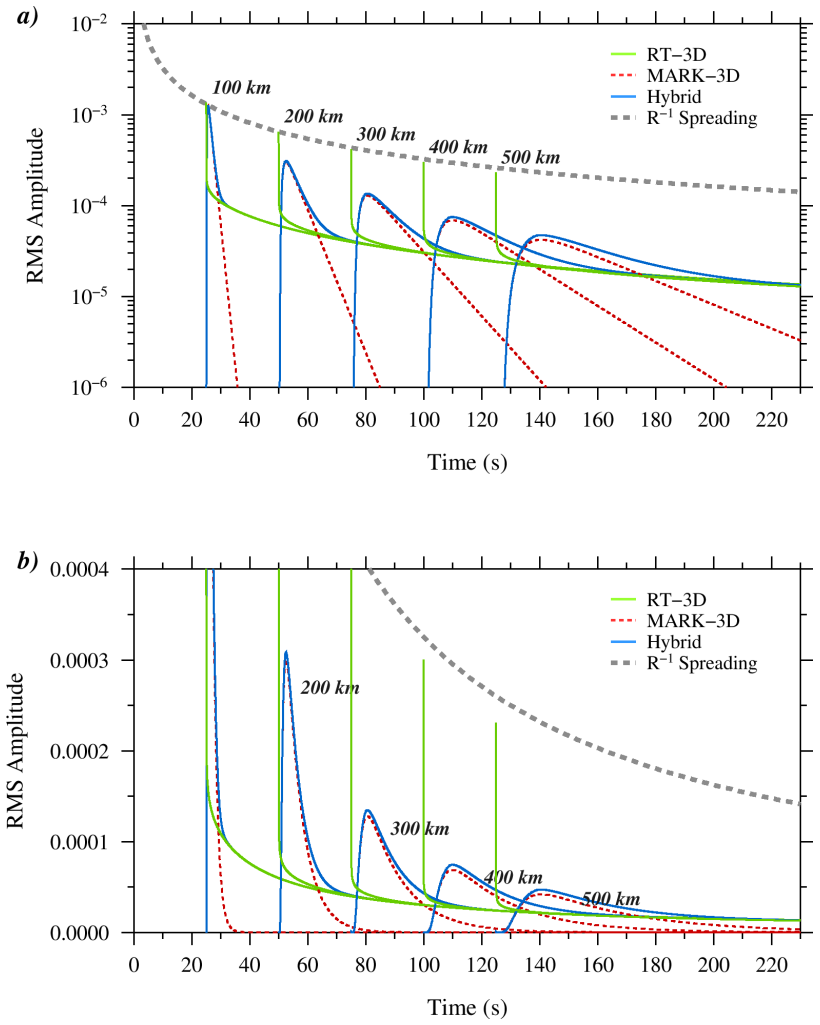

Figure 8. a) semi-log and b) linear plots of the hybrid synthetic envelopes (blue lines) at 2 Hz, for von Karman-type random media ($\boldsymbol{k = 0.5}$, $\boldsymbol{\varepsilon = 0.05}$, $\boldsymbol{a = 5}$), where $\boldsymbol{k}$ is the Hurst number, $\boldsymbol{\varepsilon}$ and $\boldsymbol{a}$ are the fractional fluctuation and characteristic length of inhomogeneity, respectively. *Estimated total scattering coefficient is* $\boldsymbol{8 \times 10^{-4} km^{-1}}$ *for a given random media. Green solid and red dashed lines represent the analytic solutions of the radiative transfer (RT) equation and the Markov approximation of the parabolic wave equation for the random media, respectively. The Markov envelope (red dashed lines) shows the peak delay and envelope broadening with increasing travel distance, but no excitation of coda. On the other hand, the radiative transfer solution for impulsive radiation (green solid lines) models the coda excitation, but the direct arrival in the model is too impulsive, because too little energy has been forward scattered in the portion immediately after the direct arrival. The Markov and hybrid envelopes show more rapid peak amplitude decay than the RT envelope and 1/r geometrical spreading due to energy loss from the multiple forward scattering.*

3.2.2. Envelope Broadening and Scattering in von Karman Random Media

Sonic well-logging data show that velocity inhomogeneities are random and their power spectral density function (PSDF) decreases with some negative power of wavenumber. Von Karman-type random media having power-law spectra at large wavenumber are often used for describing Earth's inhomogeneity. The autocorrelation function (ACF) of a von Karman random media characterized by correlation distance (a) and fractional fluctuation (ε) is given by

$$R(\mathbb{x}) = R(r) = \frac{\varepsilon^2 2^{1-k}}{\Gamma(k)} \left(\frac{r}{a}\right)^k K_k \left(\frac{r}{a}\right) \qquad \text{for } k = 0{\sim}1,$$

eq. 22

where $\Gamma(k)$ is the gamma function and K_k is the modified Bessel function of the second kind of order k. For the 3-D case, the PSDF is

$$P(\mathbb{m}) = P(m) = \frac{8\pi^{3/2}\Gamma(k+3/2)\varepsilon^2 a^3}{\Gamma(k)(1+a^2m^2)^{k+3/2}},$$

eq. 23

where $\mathbb{m}$ is the wavenumber vector and $m = |\mathbb{m}|$. The PSDF obeys a power law for large wavenumbers ($am \gg 1$), where the decay rate is controlled by the Hurst number, k. The von Karman random media is rich in short wavelength components compared with that of Gaussian-type random media. The von Karman random media with $k = 0.5$ is an exponential random media.

Saito et al. (2002) numerically solved the master equation of TFMCF (eq. 11) for a von Karman random media having non-integer k. Most recently, Sato (2016) solved the equation fully analytically by decomposing the random velocity fluctuation into two components: the low-wavenumber spectral component and the high-wavenumber spectral component.

As a simple mathematical model, Sato (2016) proposed a new stochastic synthesis of the scalar wavelet envelope in 3-D von Karman-type random media, for the case when the center wavenumber of the wavelet ($k_c = \omega_c/V_0$) is higher than the corner wavenumber (a^{-1}) of the PSDF. The key idea is to split the random medium spectrum into two components using the center wavenumber as a reference: the long-scale (P_L; low-wavenumber spectral) component produces the peak delay and the envelope broadening by multiple scattering around the forward direction; the short-scale (P_S; high-wavenumber spectral) component attenuates wave amplitude by wide-angle scattering:

$$P(\mathbb{m}) = P_L(\mathbb{m}) + P_S(\mathbb{m}).$$

eq. 24

Sato (2016) took the corner wavenumber of the PSDF of the short-scale component as

$$a_S^{-1} = \zeta k_c,$$

eq. 25

where $k_c = (\omega_c/V_0)$ is the center wavenumber and ζ is a precise tuning parameter of the order of one. Sato (2016) defined the parameter ν in the following equation:

$$\nu \equiv \frac{a_S^{-1}}{a^{-1}} = \zeta a k_c.$$

eq. 26

From eq. 23, eq. 25 and eq. 26, the PSDF of the short scale component of random media (P_S) is

$$P_S(m) \equiv \frac{8\pi^{\frac{3}{2}}\Gamma\left(k+\frac{3}{2}\right)\varepsilon_S^2 a_S^3}{\Gamma(k)(1+a_S^2 m^2)^{k+\frac{3}{2}}} = \frac{8\pi^{\frac{3}{2}}\Gamma\left(k+\frac{3}{2}\right)\varepsilon^2 a^3}{\Gamma(k)(\nu^2+a^2m^2)^{k+\frac{3}{2}}}$$

$$= \frac{8\pi^{3/2}\Gamma(k+3/2)\varepsilon^2 a^3}{\Gamma(k)((\zeta a k_c)^2+a^2m^2)^{k+3/2}}.$$

eq. 27

Envelope Broadening and Peak Delay due to the Long-Scale Component in von Karman Random Media

The time width of an envelope is well characterized by the characteristic time, t_M. By decomposing a random medium into two components, Sato (2016) analytically derived the characteristic time for the long-scale (low-wavenumber spectral) component of 3-D von Karman type random media (P_L) as:

$$t_M(k,\xi,k_c,r_0) = \frac{\varepsilon^2}{2\nu_0 a} C_L(k,\xi,k_c) r_0^2,$$

eq. 28

where

$$C_L(k,\xi,k_c) = \begin{cases} \frac{\sqrt{\pi}\Gamma\left(k+\frac{1}{2}\right)}{(2k-1)\Gamma(k)}\left[1-(\zeta a k_c)^{1-2k}\right] & \text{for } k \neq \frac{1}{2} \\ \ln \zeta k_c & \text{for } k = \frac{1}{2} \end{cases}.$$

eq. 29

As we see from the equations, t_M is proportional to the square of travel distance and frequency dependent as a function of wavenumber. The frequency dependency increases for rougher media with lower k. For 3-D von Karman random media, we found that the delay time between the onset and the peak arrival is $\sim 0.37 t_M$.

Attenuation Correction for the Markov Approximation due to the Short-Scale Component in von Karman Random Media

The Markov envelope with the characteristic time t_M simulates the envelope broadening effect by multiple scattering in a narrow cone around the forward direction, due to the long-scale component of random media. This approximation ignores attenuation due to wide angle scattering that is caused by the short-scale component of random media.

Sato (2016) proposed to use the total scattering coefficient derived using the short-scale component of random media P_S as the lowest correction to the Markov envelope for scattering attenuation by wide angle. We note that Saito et al. (2003) numerically estimated this effect.

The scattering coefficient defined as the scattering power per unit volume is directly related to the PSDF of the velocity fractional fluctuation. For the 3-D case, the scattering coefficient for the angular frequency (ω) and scattering angle (ψ) is given by (Sato et al., 2012, eq. (4.25)):

$$g(\psi;\omega) \equiv 4\pi \frac{1}{L^3}\langle\frac{d\sigma}{d\Omega}\rangle = \frac{k_0^4}{\pi} P\left(2k_0 \sin\frac{\psi}{2}\right),$$

eq. 30

where k_0 is wavenumber (ω/V_0). The total scattering coefficient is given by the solid angle average:

$$\begin{aligned} g_0(\omega) &= \frac{1}{4\pi}\oint g(\psi;\omega) d\Omega(\psi,\omega) \\ &= \frac{1}{2}\int_0^{\pi} g(\psi;\omega)\sin\psi\, d\psi \\ &= \frac{k_0^4}{2\pi}\int_0^{\pi} P\left(2k_0 \sin\frac{\psi}{2}\right)\sin\psi\, d\psi \\ &= \frac{k_0^2}{2\pi}\int_0^{2k_0} P(m) m dm \end{aligned}$$

eq. 31

The total scattering coefficient of the short-scale component is derived from eq. 30 and eq. 31:

$$g_{S0}(k,\zeta,k_0) = \frac{\varepsilon^2}{a}\frac{2\pi^{1/2}\Gamma(k+1/2)}{\Gamma(k)}\left[1-\frac{1}{(1+4\zeta^{-2})^{k+1/2}}\right]\frac{(\zeta a k_0)^{1-2k}}{\zeta^2}$$

eq. 32

Scattering Coefficients for the Coda Excitation

Another important quantity in radiative transfer theory is the transport scattering coefficient, which is also called the momentum transfer scattering coefficient (Morse et al., 1954). Even though the scattering is nonisotropic, the scattering process can be well described by isotropic scattering when the multiple scattering process is dominant. The transport scattering coefficient is defined by adding an additional factor $(1-\cos\psi)$ to exclude forward scattering to the total scattering coefficient eq. 31:

$$\begin{aligned} g_m(\omega) &= \frac{1}{4\pi}\oint g(\psi;\omega)(1-\cos\psi) d\Omega(\psi,\omega) \\ &= \frac{1}{2}\int_0^{\pi} g(\psi;\omega)(1-\cos\psi)\sin\psi\, d\psi \end{aligned}$$

$$= \frac{k_0^4}{2\pi}\int_0^{\pi} P\left(2k_0 \sin\frac{\psi}{2}\right)(1-\cos\psi)\sin\psi\, d\psi$$
$$= \frac{1}{4\pi}\int_0^{2k_0} P(m)m^3 dm$$

eq. 33

The factor $(1-\cos\psi)$ also works as a filter that eliminates the long-scale component of the spectra of the random media (Figure 9). Thus, the transport scattering coefficient can be interpreted as the effective isotropic scattering in the multiple scattering regime, which is the average contribution of wide-angle scattering.

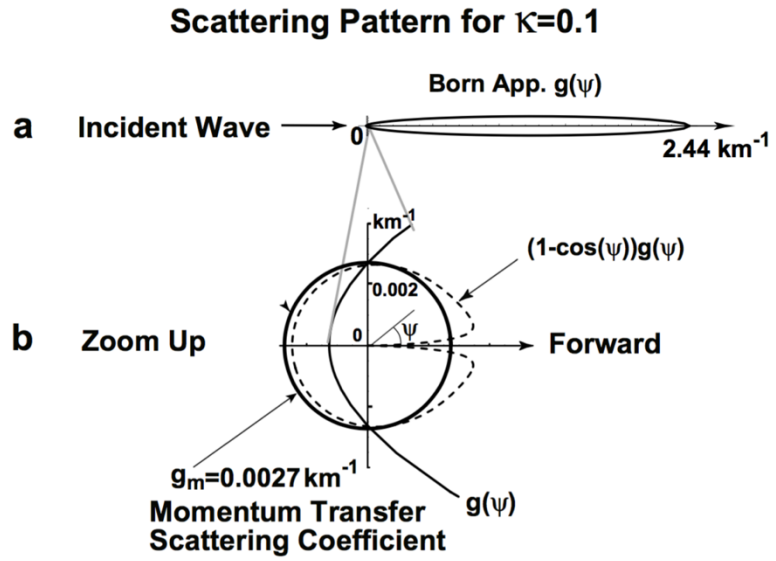

Figure 9. a) Scattering coefficient g(ψ) based on the Born approximation at 2 Hz in a 2-D von Karman random medium with κ =0.1. b) Comparison of the momentum transfer scattering coefficient g_m with the scattering coefficient g(ψ) (Sato et al., 2004).

We derived the analytic form of the moment transfer scattering coefficient for a 3-D von Karman random media having non-integer ***k*** from eq. 24 and eq. 31:

$$g_m(k,\zeta,k_0) = \frac{\varepsilon^2}{a}\frac{2\pi^{1/2}\Gamma(k+1/2)}{(2k-1)\Gamma(k)}\left[1-\frac{1+2a^2k_0^2(2k+1)}{(1+4a^2k_0^2)^{k+\frac{1}{2}}}\right],$$

eq. 34

and for the exponential random media ($k = 0.5$),

$$g_m(k,\zeta,k_0) = \frac{\varepsilon^2}{a}\left[\ln(1+4a^2k_0^2) - \frac{4a^2k_0^2}{1+4a^2k_0^2}\right].$$

eq. 35

For comparison, we also derived the total scattering coefficient from eq. 24 and eq. 30 and a backscattering coefficient ($g_\pi \equiv g(\psi = \pi)$).

$$g_0(k, \zeta, k_0) = \varepsilon^2 a k_0^2 \frac{2\pi^{1/2}\Gamma(k + 1/2)}{\Gamma(k)} \left[1 - \frac{1}{(1 + 4a^2 k_0^2)^{k+1/2}}\right]$$

eq. 36

$$g_\pi(k, \zeta, k_0) = \frac{8\pi^{1/2}\Gamma(k + 3/2)}{\Gamma(k)} \frac{\varepsilon^2 a^3 k_0^4}{(1 + 4a^2 k_0^2)^{k+3/2}}$$

eq. 37

Figure 10 shows wavenumber and frequency dependence of the scattering coefficients for 3-D von Karman random media. The wavenumber and frequency dependence of the transport scattering become stronger for a rough media (e.g., $k < 0.5$). We note that the scattering attenuation effect (g_{S0}) in the Markov solution due to the short scaling component of random media is also important for the rough media.

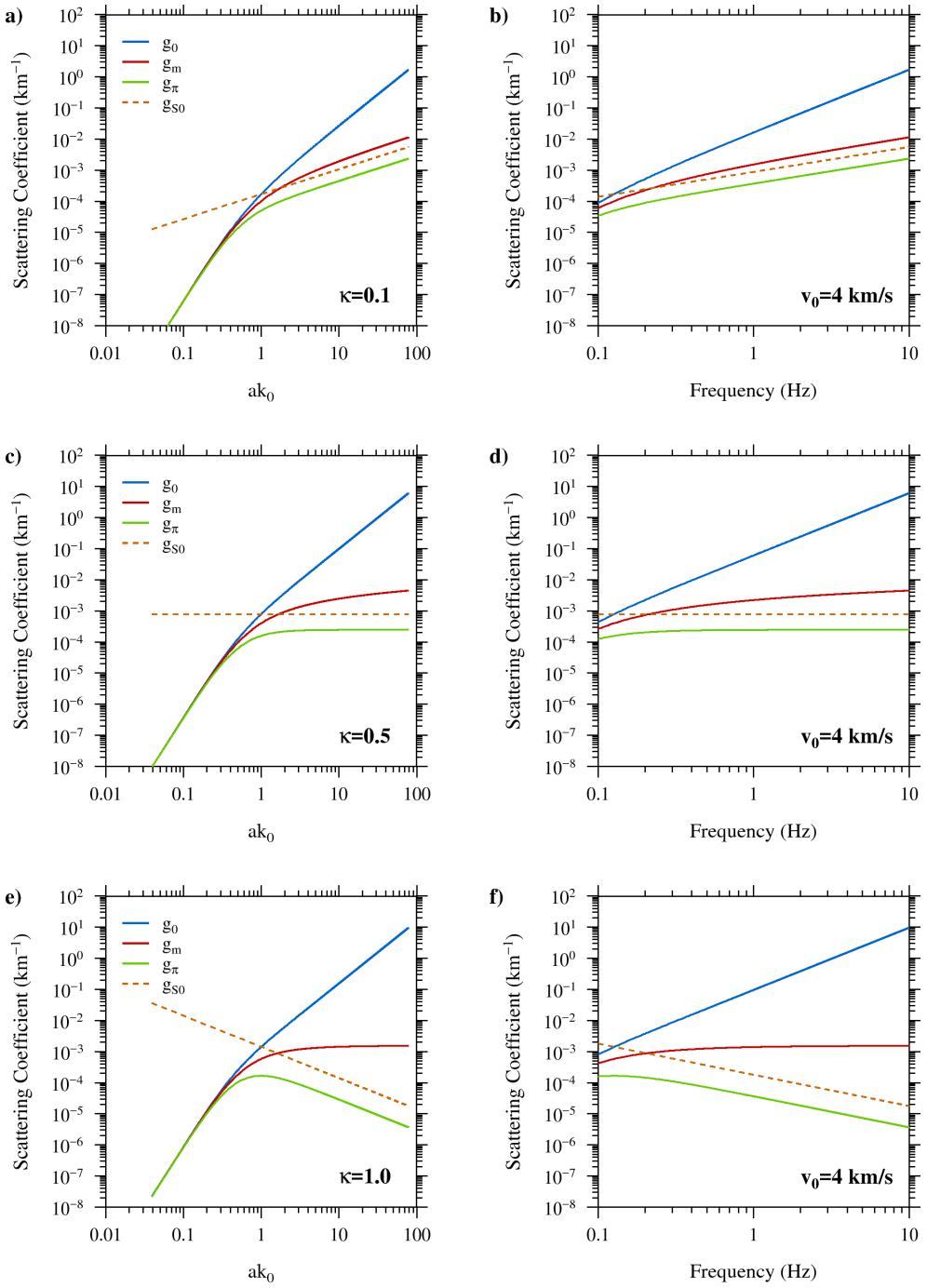

Figure 10. Wavenumber and frequency dependence of scattering coefficients in 3-D von Karman random media characterized by the Hurst number k=0.1 (a and b), k=0.5 (c and d), and k=1.0 (e and f). $\boldsymbol{g_\pi}$, $\boldsymbol{g_0}$, and $\boldsymbol{g_m}$ *are the backward, total, and momentum transfer scattering coefficients, respectively.* $\boldsymbol{g_{S0}}$ *is the scattering due to the short-scale component of the random media (Sato, 2016).*

4. RESULTS AND DISCUSSION

4.1. Hybrid Full-waveform Envelope Modeling and Parameterization

We have developed a new hybrid envelope formulation which can model both the coda excitation due to wide-angle scattering and envelope broadening due to multiple forward scattering. In the new hybrid envelope model, energy density at distance r for a unit source can be given by (for derivations and details of each parameter, refer to 3.2. Algorithm Development):

$$E(r,t) = \frac{W}{4\pi V_0 r^2}\frac{\pi^2}{16 t_M}\vartheta_4''\left(0, e^{-\frac{\pi^2}{4t_M}(t-r/V_0)}\right) e^{-(V_0 g_0 t)} H\left(t-\frac{r}{V_0}\right) + W\frac{[1-r^2/(V_0 t)^2]^{1/8}}{[4\pi V_0 t/(3g_0)]^{\frac{3}{2}}} e^{-(V_0 g_0 t)} \mathrm{M}\left[V_0 g_0 t\left(1-\frac{r^2}{V_0^2 t^2}\right)^{\frac{3}{4}}\right] H\left(t-\frac{r}{V_0}\right),$$

eq. 38

where,

$$M(x) \equiv \frac{8}{(3x)^{3/2}}\sum_{n=1}^{\infty}\frac{\Gamma\left(\frac{3}{4}n+\frac{3}{2}\right)}{\Gamma\left(\frac{3}{4}n\right)}\frac{x^n}{n!}$$

Eq. 39

and ϑ_4'' is second derivative of an elliptical theta function of the fourth kind, t_M is the characteristic time that controls envelope broadening and peak delay due to the multiple forward scattering (M. Fehler et al., 2000; Sato et al., 2012), and g_0 is the total isotropic scattering coefficient. The first and second terms in right hand-side of eq. 38 model envelope broadening due to multiple forward scattering and the coda excitation due to wide-angle scattering, respectively. We note that the Multiple Lapse Time Window Analysis (MLTWA) studies using an analytic solution of the Radiative Transfer equation only use the second term and ignore multiple forward scattering (e.g., Carcolé and Sato, 2010).

Assuming a frequency dependent intrinsic absorption and uncorrelated noise, a Root-Mean-Square (RMS) envelope can be expressed as

$$A(r,t,f;a,\varepsilon,\kappa,\zeta,Q_0,\eta \rightarrow t_M, g_0, b) = S(f)P(f)[2\rho_0^{-1}\boldsymbol{E}(\boldsymbol{r},\boldsymbol{t},\boldsymbol{f};\boldsymbol{t_M},\boldsymbol{g_0})e^{-bt} + N^2(f)]^{1/2}$$

eq. 40

where S, P and N are the source, frequency dependent site effect, and noise spectra, respectively, b is the intrinsic attenuation factor ($b \equiv \frac{2\pi f}{Q_o f^{\eta}}$) and ρ_0 is density.

The characteristic time t_M and total isotropic scattering coefficient g_0 are computed using the spectrum division method (Sato, 2016; Sato & Emoto, 2017, 2018) for given von Karman random media with the correlation length a, standard deviation of fractional velocity fluctuation ε, Hurst exponent κ, and a tuning parameter ζ. Sato (2016) analytically derived the characteristic time for the long-scale (low-wavenumber spectral) component of 3-D von Karman type random media (P_L):

$$t_M(\kappa, \zeta, k_0, r) = \frac{\varepsilon^2}{2v_0 a} C_L(\kappa, \zeta, k_0) r^2,$$

eq. 41

$$\text{where } C_L(\kappa, \zeta, k_0) = \begin{cases} \frac{\sqrt{\pi}\Gamma\left(\kappa+\frac{1}{2}\right)}{(2\kappa-1)\Gamma(\kappa)}\left[1-(\zeta a k_0)^{1-2\kappa}\right] & \text{for } \kappa \neq \frac{1}{2} \\ \ln \zeta k_0 & \text{for } \kappa = \frac{1}{2} \end{cases}.$$

eq. 42

And the total scattering coefficient of the short-scale component is given by (Sato, 2016, eq. 39):

$$g_0(k, \zeta, k_0) = \frac{\varepsilon^2}{a} \frac{2\pi^{1/2}\Gamma(k+1/2)}{\Gamma(k)} \left[1 - \frac{1}{(1+4\zeta^{-2})^{k+1/2}}\right] \frac{(\zeta a k_0)^{1-2k}}{\zeta^2}$$

eq. 43

where $k_0 = \left(\frac{2\pi f}{V_0}\right)$ is wavenumber.

With given envelope modeling parameters ($a, \varepsilon, k, \zeta, Q_0, \eta$) and pre-event noise, we measure the envelope amplitude of each event-station pair by fitting the observed envelope with the synthetic envelope using the Metropolis-Hastings algorithms (Hastings, 1970; Metropolis et al., 1953). We directly estimate uncertainty of the measured amplitudes from *a posterior* distribution of the sampled amplitudes.

4.2. Exploring parameter space using Markov Chain Monte Carlo Sampling Method

Using the new forward modeling code, we explore the model parameter space of the new forward modeling by matching the observed envelope recorded at local and regional distances. This step validates the new forward modeling and examines the sensitivity of each model parameter.

In this new formulation, all modeling parameters that control the forward and wide-angle scatterings were consistently derived for a 3-D von Karman random media. This new parameterization can be numerically very efficient, especially for the multi-frequency band problem. For example, in a forward simulation of each phase envelope, we will need three

parameters, t_M, g_0, b, for each single frequency band. However, we can associate the first two parameters t_M, g_0 for an arbitrary frequency to three characteristic parameters for a von Karman media, which are correlation distance (a), fractional fluctuation (ε) and the Hurst number (k). Thus, we can reduce the modeling parameters from two times the number of frequency bands (N_f) to three for the scattering parameters by assuming the von Karman media. Similarly, we assumed that in the frequency dependent intrinsic attenuation ($Q_i(f) = Q_0 f^{\eta}$), there is a parameter reduction of N_f-to-2.

With the new parametrization assuming von Karman random method, we could reduce the total number of modeling parameters from $3 \times N_f$ to 5. However, 5 parameters are still too many for a conventional grid-search type approach; therefore, we solved the problem using the Markov Chain Monte Carlo (MCMC) sampling method, more specifically the Metropolis–Hastings algorithm (Hastings, 1970; Metropolis et al., 1953).

We explored the modeling parameters control on the shape and attenuation of the crustal phases' envelopes (P_g and L_g). For validation purpose, we chose same dataset we have studied using the L_g-coda calibration method (Yoo et al., 2010). The dataset includes 409 earthquakes recorded at 25 broadband stations in the Korean Peninsula.

In this test, we solved the P_g and L_g templates separately, assuming no scattering conversion. Also, we only fit a single frequency band, 2-4 Hz, and we directly sample the source amplitude rather than sample the seismic moment and corner frequency in this test. We note that we did not sample the site term in the MCMC model search, so the source estimates still include the site amplification in the current test.

First, we assume that the misfit is distributed normally with zero mean and some known variance, σ^2. We then express the log-likelihood function for envelope model,

$$l(a, \varepsilon, k, Q_0, \eta, S) = \log\left[\prod_i^{NT \times NF} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{[\log_{10} A(r,t,f;a,\varepsilon,k,Q_0,\eta,S) - \log_{10} D(r,t,f)]^2}{2\sigma^2}\right)\right]$$

eq. 44

where A is a model prediction proposed by parameters $a, \varepsilon, k, Q_0, \eta, S$ using eq. 21 and D is the observed RMS envelope defined as a vector sum of the three component velocity seismogram ($u_{N,E,Z}$),

$$D(t,f) = [u_N^2 + u_E^2 + u_Z^2]^{\frac{1}{2}}$$

eq. 45

For each observation, we proposed a total of 2 million sets of models (1 million for a burn-in process and another 1 million for a model sampling), sampled 1,000 sets from the 1 million and obtained a distribution of the predicted envelope models and model parameters. From the set of predicted model parameters, we computed the distribution of scattering and intrinsic attenuations,

$$Q_S^{-1} = \frac{v g_0}{2\pi f} \text{ and } Q_i^{-1} = (Q_0 f^{\eta})^{-1}.$$

eq. 46

4.3. Application of New Hybrid Model to Crustal Phases

Using the new hybrid model provides excellent fits to the entire envelope for both crustal phases. Also, the characteristic time, scattering and intrinsic attenuation coefficients are well separated and constrained by the MCMC approach. Figure 11 shows the 2-4 Hz RMS envelopes and predicted models for a Mw 4.6 earthquake that occurred in the mid-eastern Korean Peninsula on 20 January 2007. Our new hybrid model fits the data very well and shows quite a stark contrast to the previous coda model (see Figure 12 from Pasyanos et al., 2012). The simple exponential coda model seems to work reasonably well for the later arrival of S-coda, but their model fails to predict the entire P-coda envelopes and earlier part of S-coda as well. Note that the simple exponential coda model is originally designed for S-coda after roughly twice the travel time of an S-wave.

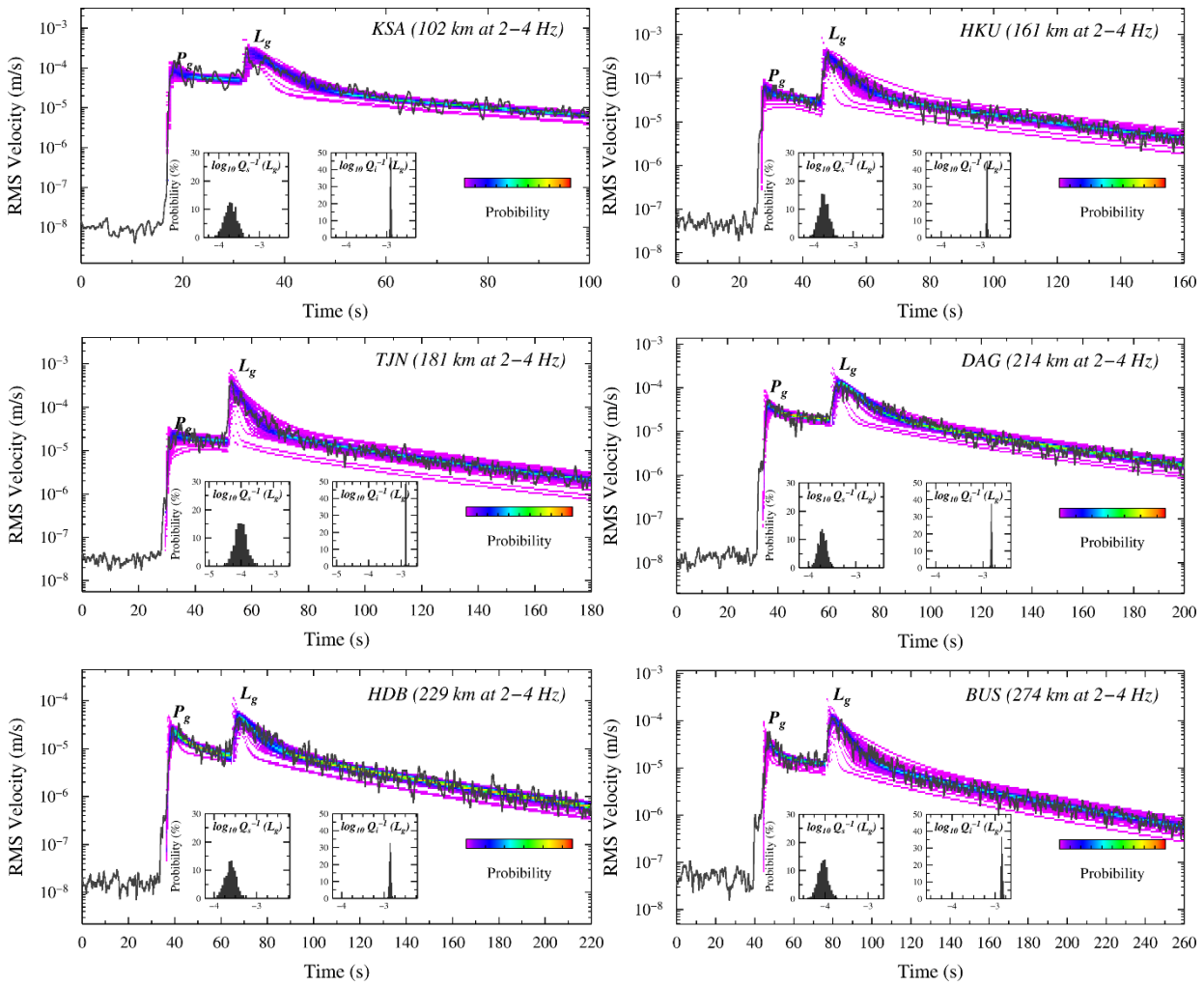

Figure 11. Example of the model-search using the MCMC sampling to obtain the optimal model parameters for a Mw 4.6 earthquake occurred in the mid-eastern Korean Peninsula on 20 January 2007. *Black solid line shows the 2-4 Hz RMS envelope and color-scaled image shows posterior distributions of predicted envelopes. Two inserted plots in each plot show posterior probability distributions of scattering and intrinsic attenuations of L_g, respectively.*

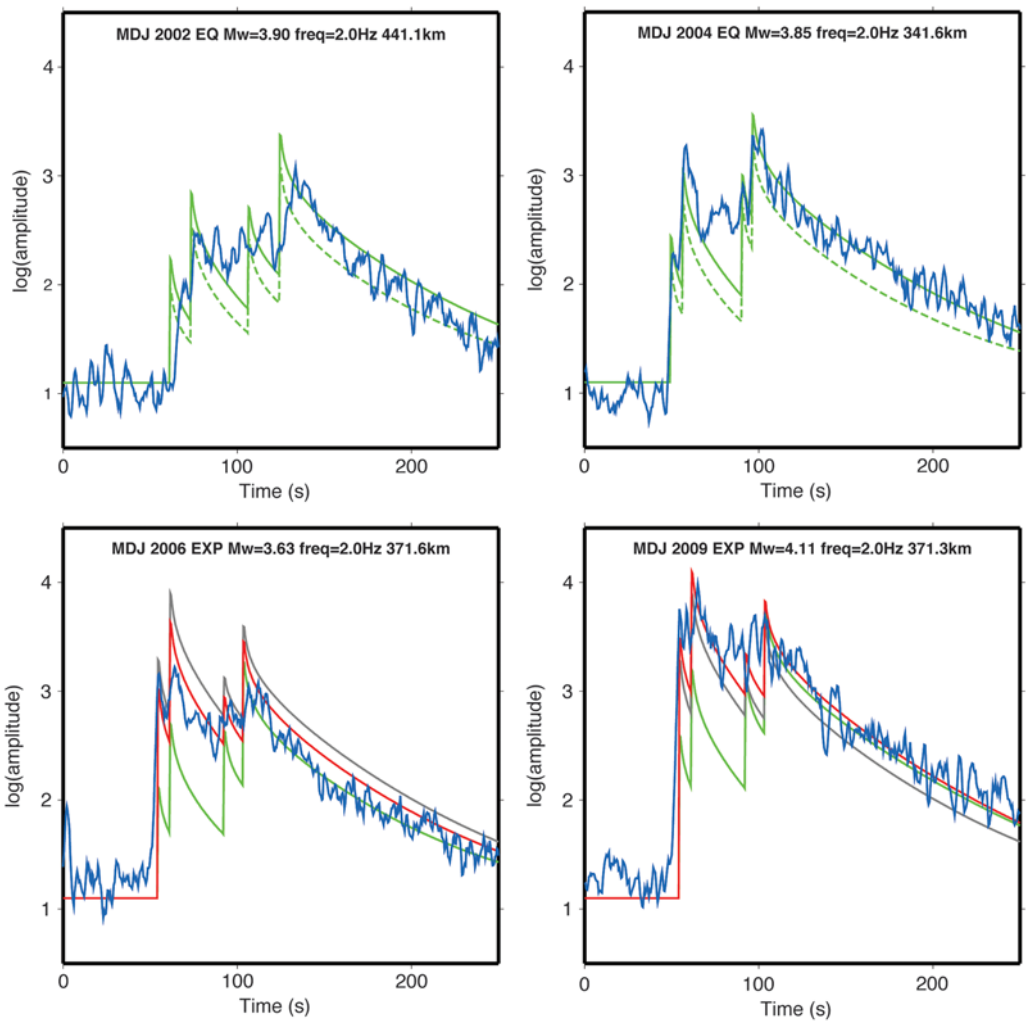

Figure 12. Figure 6 in Pasyanos et al., (2012). Waveform envelopes for the two earthquakes and two explosions recorded at station MDJ in the 2–4 Hz passband. *Synthetic envelopes for earthquakes are shown in green, with dashed and solid lines showing uncorrected and corrected synthetics, respectively. Synthetic envelopes for a 1-kt explosion at 100-m depth are shown in gray, while envelopes for the best-fitting explosions are shown in red.*

Figures 13 and 14 show the estimated scattering and attenuation parameters for L_g and P_g, respectively. For the L_g case, the estimated scattering attenuation is much smaller in magnitude than intrinsic attenuation, indicating that intrinsic absorption is the main cause of the high frequency attenuation process in this area. This observation agrees well with previous studies using the Multi-Lapse Time Windows Analysis based on the Monte Carlo simulation method (Chung et al., 2010).

The characteristic time, t_M, controls envelope broadening and peak delay due to the multiple forward scattering. The maximum peak arrival is delayed by $\sim 0.37 t_M$ seconds for 3-D propagation (spherical wavelet). The estimated characteristic time from L_g is greater than that from P_g, and shows a strong trend for increasing distance which means the peak of L_g is delayed more than that of P_g due to the strong forward scattering. However, the scattering attenuation was twice as low as the MLTWA estimate. It may be difficult to directly compare the scattering coefficients

between our results and MLTWA results since the MLTWA approach assumes only wide-angle scattering (e.g., Chung et al., 2010; isotropic scattering), while the new approach considers both narrow and wide scattering. We will try to devise a rational comparison method in the future.

The characteristics of the P-coda are very different from that of the S-coda. This may make it somewhat difficult to interpret the estimated parameters from P_g envelope as they are. However, the P_g envelope fittings clearly require higher scattering attenuation than intrinsic attenuation which is contrary to the L_g case.

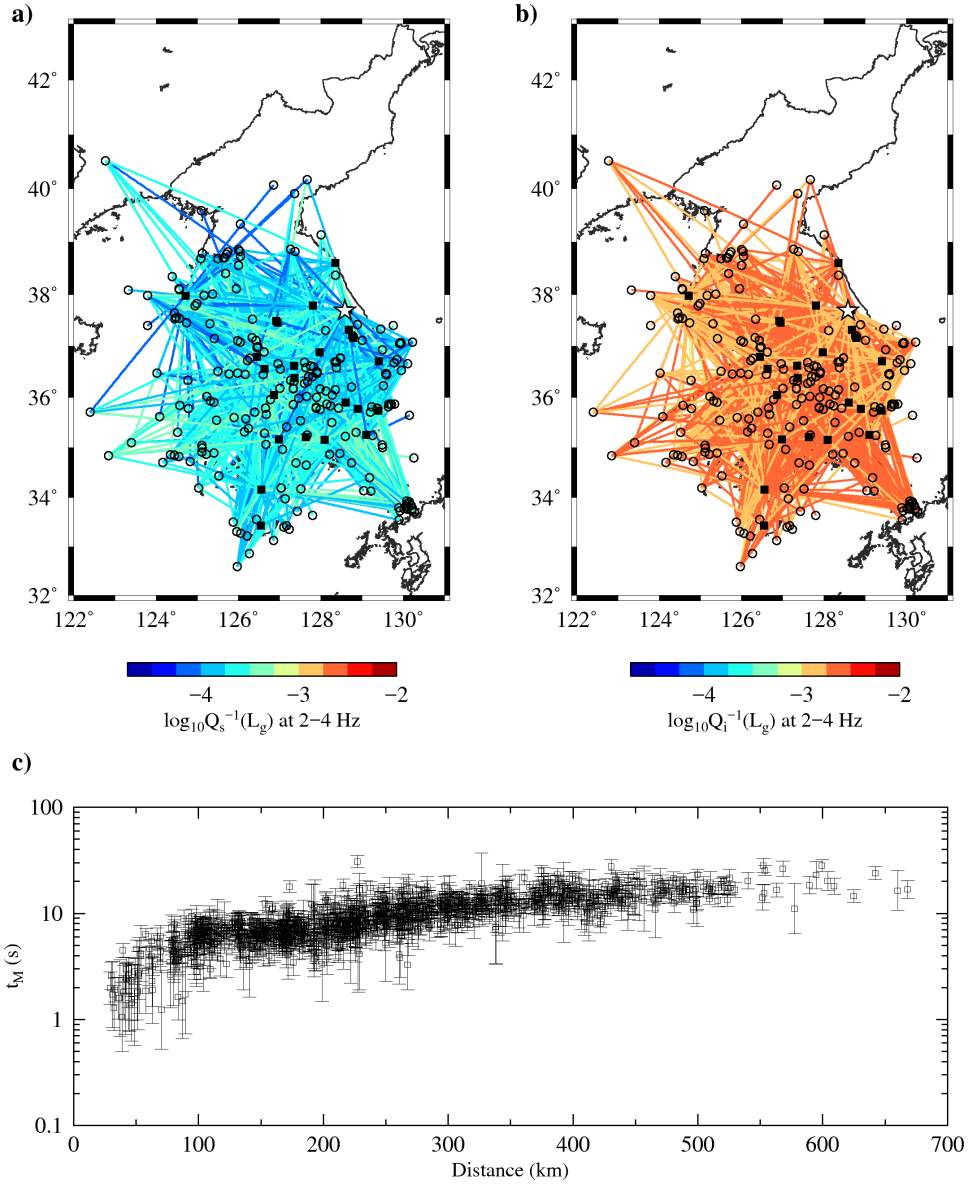

Figure 13. Map shows the a) scattering and b) intrinsic attenuations of L_g estimated at each event-station path. c) Estimated characteristic time $\boldsymbol{t_M}$ against distance. *Solid and dashed lines represent the best fit and ±σ. The Mw 4.6 earthquake on 20 January 2007 is shown as a star symbol.*

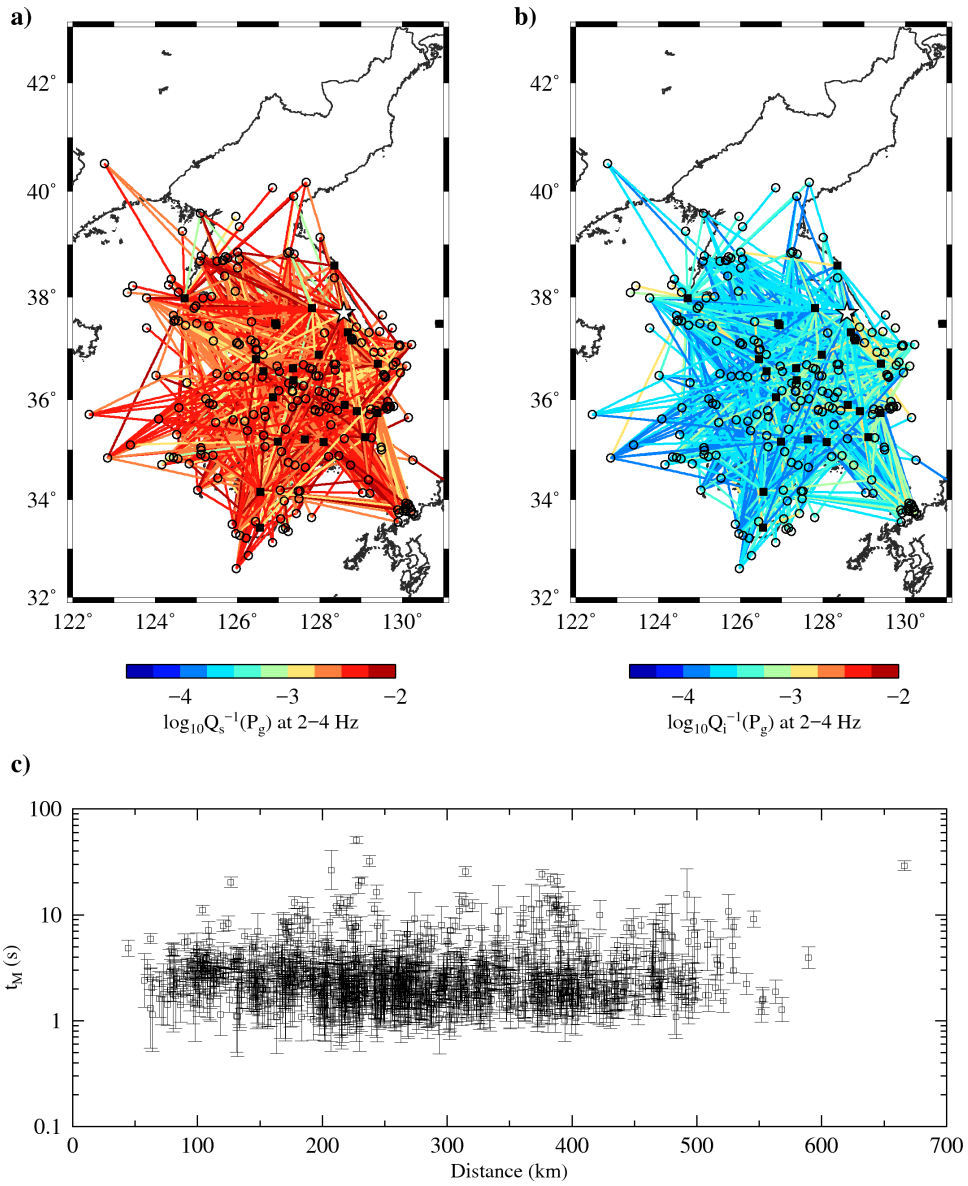

Figure 14. Same to Figure 13, but for P_g. Map shows the a) scattering and b) intrinsic attenuations of P_g estimated at each event-station path. c) Estimated characteristic time $\boldsymbol{t_M}$ against distance. *Solid and dashed lines represent the best fit and ±σ. The Mw 4.6 earthquake on 20 January 2007 is shown as a star symbol.*

4.4. Characterizing Scattering and Absorbing Properties

We characterized scattering and absorbing properties of the Korean Peninsula and the surrounding area for full-waveform envelope template modeling. We have quantified the scattering and absorbing characteristics of the crustal P- and S-waves (P_g and L_g) of the region using a MCMC sampling method with source and site constraints from direct L_g amplitude tomography inversion.

We found that a single set of P_g-template model parameters can model the shape of the observed envelope reasonably well for the whole study area. However, we found that strong regional

variation in scattering and absorbing parameters depends on the tectonic setting for L_g-template. In general, we found strong scattering and intrinsic attenuation in Honshu, Japan and weaker scattering and intrinsic attenuation in Kyushu, Japan, the Korean Peninsula and Northeast China. To properly model the L_g-template shape for both the Korean Peninsula and Japan, we needed at least two sets of template model parameters. The model parameters that are needed for Japan are ~30% longer characteristic time, ~80% higher scattering coefficient, and ~50% lower intrinsic Q on average compared to the model parameters needed for the Korean Peninsula.

The wide-angle scattering coefficients and intrinsic attenuation estimated at local distances (less than ~150km) are well consistent with previous studies using Multi-Lapse Time Windows Analysis (MLTWA) for local distance earthquakes (Carcolé & Sato, 2010; Chung et al., 2010; Michael Fehler et al., 1992). However, the estimates at regional distances are significantly lower than the estimates at local distances and are generally in good agreement with estimates of other regional studies (e.g., Sens-Schönfelder & Wegler, 2006).

Similarly, Wang & Shearer (2017) recently found that the time and range dependence of observed coda waves in Southern California cannot be well explained by half-space models of constant scattering strength. They showed that the observations can be reasonably fit with a two-layered model composed of a shallow crustal layer with strong wide-angle scattering and high intrinsic attenuation and a deeper layer with weaker scattering and lower intrinsic attenuation.

4.4.1. Direct L_g Amplitude Tomography for Source and Site Constraints

It is very important to characterize and quantify the scattering and attenuation characteristics of the study area in order to obtain more stable source parameters through accurate prediction of coda energy. The MLTWA based on the coda-normalization is one most popular ways to separate the scattering and intrinsic attenuation and quantify their contribution in an observed energy envelope (Fehler et al., 1992; Hoshiba, 1994). The coda-normalization is a very easy and convenient way to remove the source and site effect from the envelopes. However, application of the analysis is limited to local distances (~150km) because of the inherent limitations of the coda normalization for the shorter coda.

In this study, we use a more direct approach. We solve the source and site terms independently using direct amplitude tomography (Phillips & Stead, 2008) and use it as prior constraints in our MCMC method to directly sample the scattering/absorbing parameters. By doing this, we can avoid a possible trade-off between scattering/absorbing effects and source/site effects and stabilize the inversion process.

We started the tomography inversion with 20,703 unique event-station paths from 1,937 events and 394 stations. Figure 15 shows triangular tessellation grid used in the tomography inversion. We inverted for a smooth model, estimating the residual spread using robust techniques (MADn: 1.4826 times the Median Absolute Deviation), and trimming residuals outside two standard deviations before re-inverting using lower smoothing constraints (Phillips & Stead, 2008). A total of 18,601 unique event-station-paths were kept after the iterative MADn filtering inversion. The map shows a good agreement with surface topography and known geological features (Figure 16).

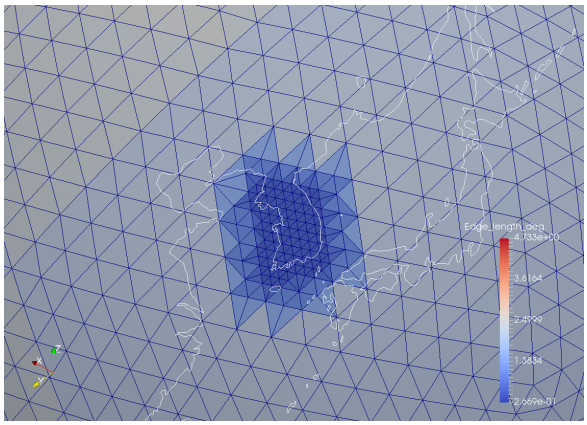

Figure 15. View of grid used in tomography inversion. *We use a uniform triangular tessellation with approximately 2° edge lengths and refine the grid to 0.5° edge lengths near the Korean Peninsula, where there are rich short paths with the highest ray-path density.*

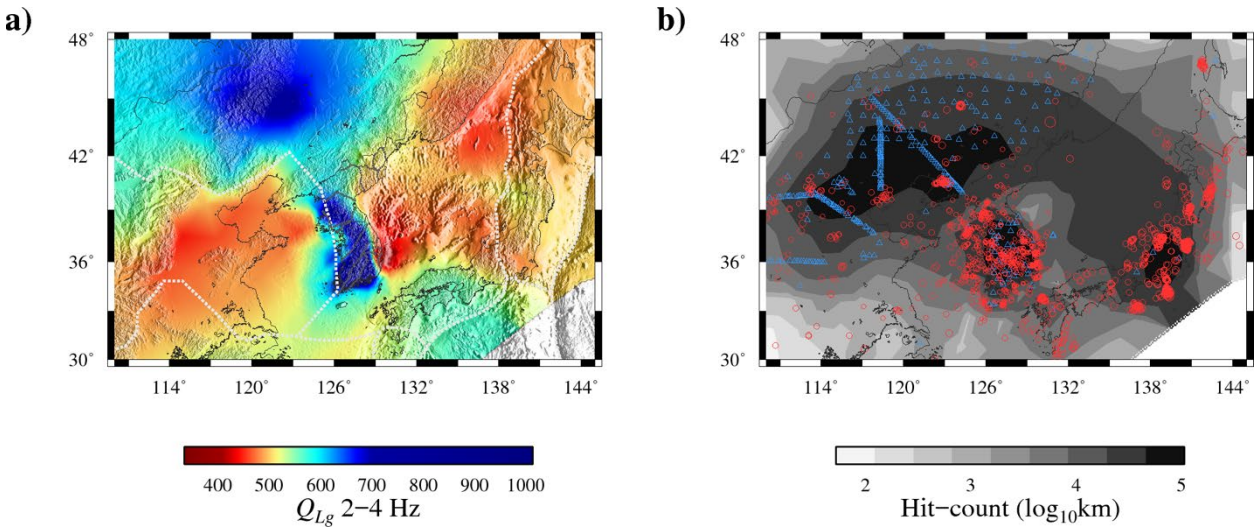

Figure 16. a) Tomographic Q images of direct L_g at 2-4 Hz and b) hit-count map. *The dashed white line represents the plate boundary as defined by Bird (2003).*

4.4.2. Extracting Scattering and Absorbing Parameters

We found 3,650 event-station pairs showing over 200 s of coda window with SNR>2.0. Now that we have independent estimates of the source and site terms from the tomography inversion, we searched for the scattering/absorbing parameters control on the shape and attenuation of the S-coda templates using the MCMC sampling method (Hastings, 1970; Metropolis et al., 1953).

Figures 17 and 18 show ensemble average of the model parameters using the MCMC model search for the 3,650 event-station pairs at 2-4 Hz and 4-8 Hz, respectively. The characteristic time estimated for shorter distance paths less than ~70 km increases according to roughly the square of travel distance, which implies that peak delay and envelope broadening obey a theoretical

asymptote of a spherical wave propagation in a half space. However, the characteristic time estimated for longer distance paths are significantly smaller than a predicted value based on short distance paths, and also largely deviated from the theoretical asymptote. This could be due to both the wave-guide effect in L_g wave and depth-dependent heterogeneity, but further analysis and research will be needed in the future.

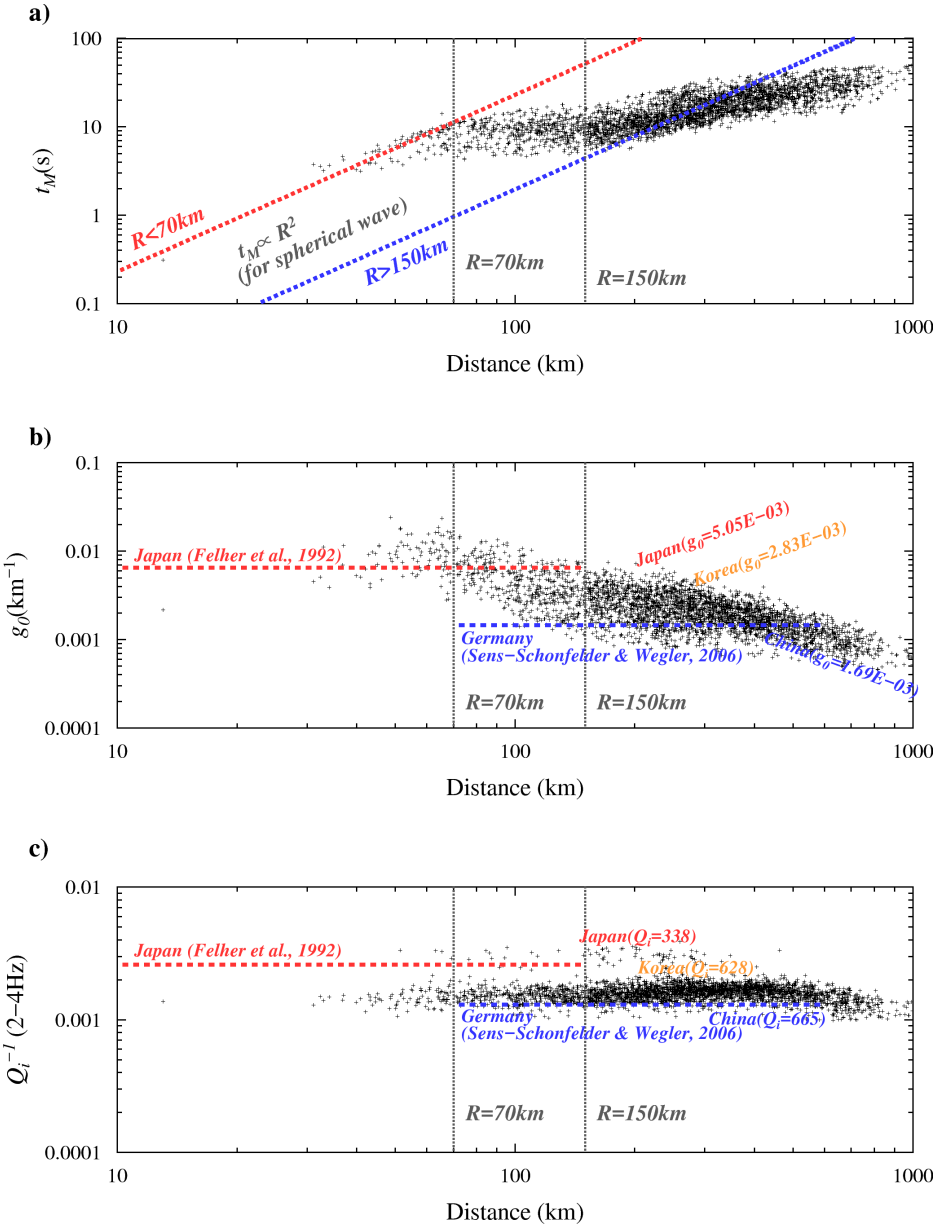

Figure 17. Ensemble average of the model parameters, a) the characteristic time, b) wide-angle scattering coefficient, and c) intrinsic attenuation for S-coda template at 2-4 Hz.

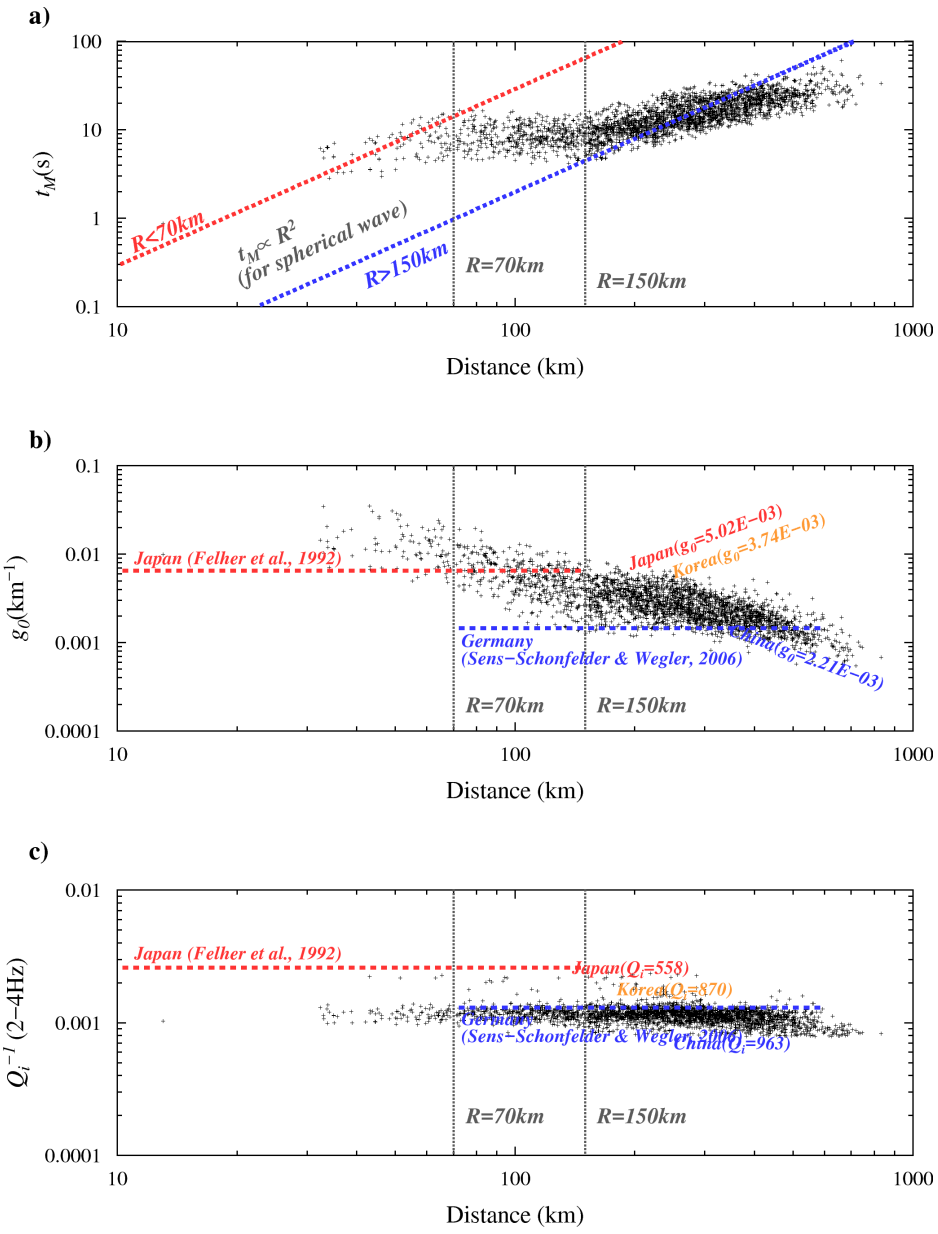

Figure 18. Ensemble average of the model parameters, a) the characteristic time, b) wide-angle scattering coefficient, and c) intrinsic attenuation for S-coda template at 4-8 Hz. *We note that estimates from Fehler et al. (1992) and Sens-Schonfelder & Wegler (2006) are still at 2-4 Hz.*

We observed a strong range dependency in the scattering coefficients. The wide-angle scattering coefficients and intrinsic attenuation estimated at local distances (less than ~150km) are well consistent with previous studies using MLTWA for local distance earthquakes (Fehler et al., 1992; Chung et al., 2010; Carcole and Sato, 2010). However, the estimates at regional distances are significantly lower than the estimates at local distances and they are generally in good agreement with estimates of other regional studies (*e.g.*, Sens-Schönfelder & Wegler, 2006).

We also found regional variation and frequency dependency in wide-angle-scattering coefficients, but it is much weaker than the range dependency. We found strong scattering in Honshu, Japan

and weaker scattering in Kyushu, Japan, the Korean Peninsula and Northeast China (see Figure 17). The frequency dependency seems stronger for the short distance paths than long distance paths.

Figure 19 shows an example of the model-search result using the MCMC sampling to obtain the optimal scattering and absorbing parameters. Red lines are an ensemble average of model predictions using the hybrid modeling and green lines show contribution of the wide-angle-scattering in the model predictions. Thus, the area between the red and green lines represent the contribution of the forward-scattering. At short travel distance, the amount of peak delay and envelope broadening in early coda due to the forward scattering is small. As travel distance increases, the amount of peak delay and envelope broadening is larger. It looks like there is a trade-off between forward-scattering and wide-angle-scattering in the fitting. As we see in Figure 20, however, the forward-scattering effect and wide-angel-scattering are not arbitrarily selected, and they are physically bonded by a von Karman random media. Additionally, our results are well consistent with other corresponding local/regional studies that do not consider forward scattering. The observed range dependency in wide-angle-scattering coefficient does not seem to have a significant effect on inclusion of the forward scattering term.

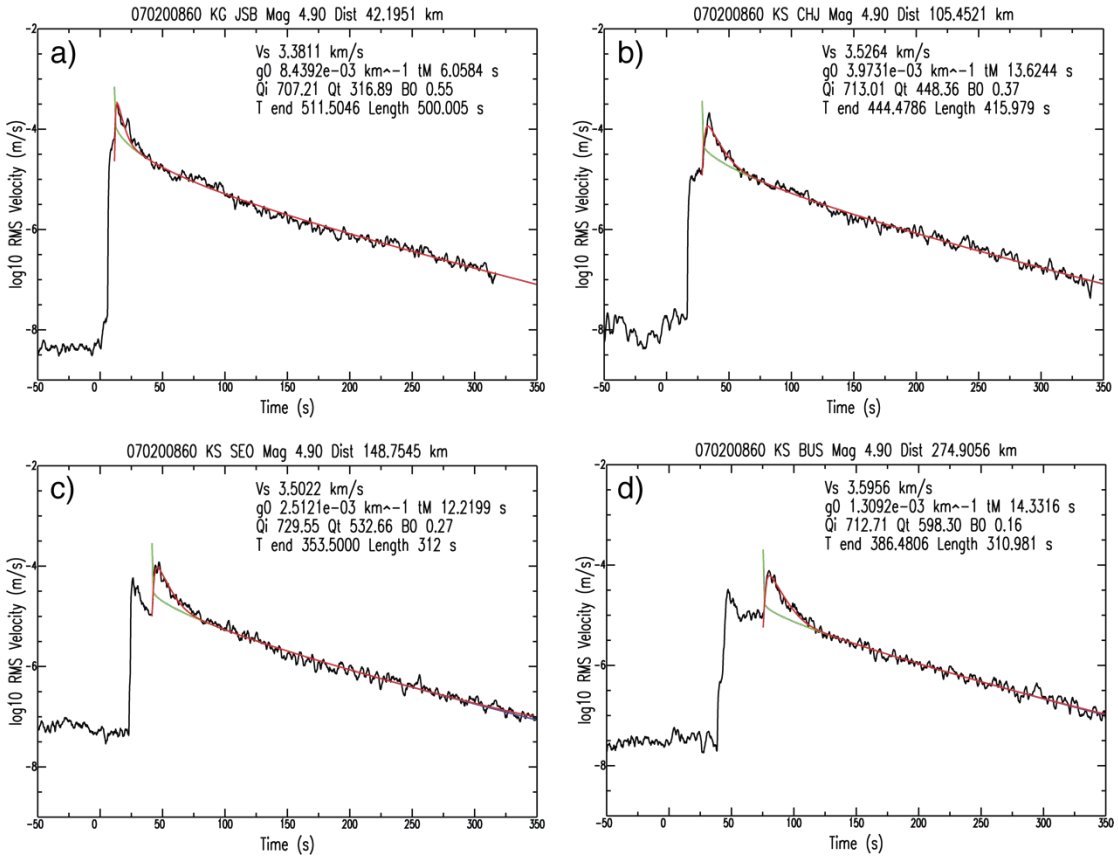

Figure 19. Example of the model-search using the MCMC sampling to obtain the optimal model parameters. *Red lines are ensemble average of model predictions and green lines show the contribution of the wide-angle-scattering in the model predictions.*

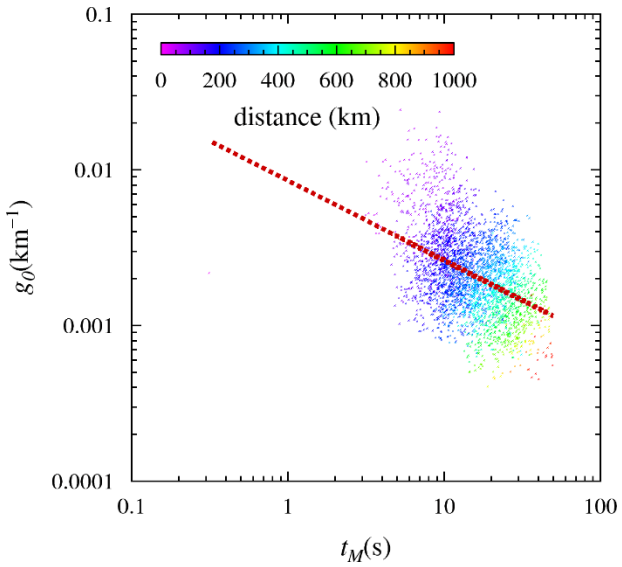

Figure 20. The characteristic time vs wide-angle-scattering coefficients at 2-4 Hz. *Red dotted line shows a least-square regression of data.*

Recently, Wang & Shearer (2017) found that the time and range dependence of observed coda waves in Southern California cannot be well explained by a half-space model of constant scattering strength (Figure 21). They showed the observations can be reasonably fit with a two-layered model composed of a shallow crustal layer with strong wide-angle scattering and high intrinsic attenuation and a deeper layer with weaker scattering and lower intrinsic attenuation.

Similarly, we found that we underpredict the coda level for the short-distance prediction if we use the weak scattering from the long-distance estimates and we overpredict the coda level for the long-distance prediction if we use the strong scattering from the short-distance estimates (Figure 22). To explain the observed range dependence in the wide-angle-scattering, however, we certainly need two scattering models to deal with different distance ranges or the depth-dependent scattering model.

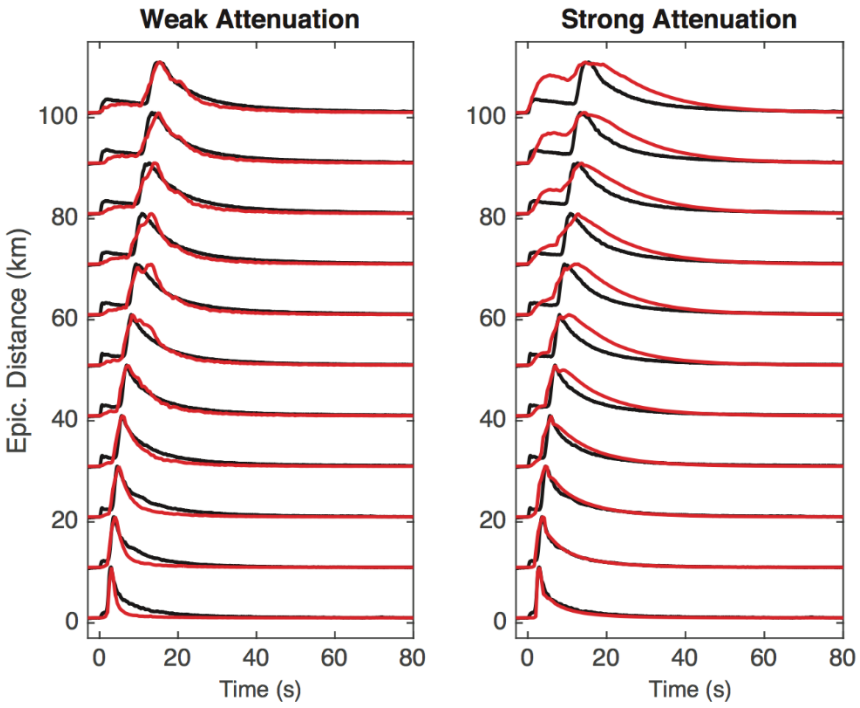

Figure 21. Figure 3 in Wang & Shearer (2017). Examples of uniform half-space intrinsic and scattering attenuation models that can fit the data stacks at long ranges (left) or at short ranges (right). *The black lines show the data stacks and the red lines show synthetics computed using the Monte Carlo method. The left model has relatively weak scattering and attenuation, which is required to fit the data at 50 to 100 km range, but underpredicts the coda amplitudes at close ranges. The right model has relatively strong scattering and attenuation, which is required to fit the data at 0 to 30 km range, but overpredicts the coda amplitudes at long ranges.*

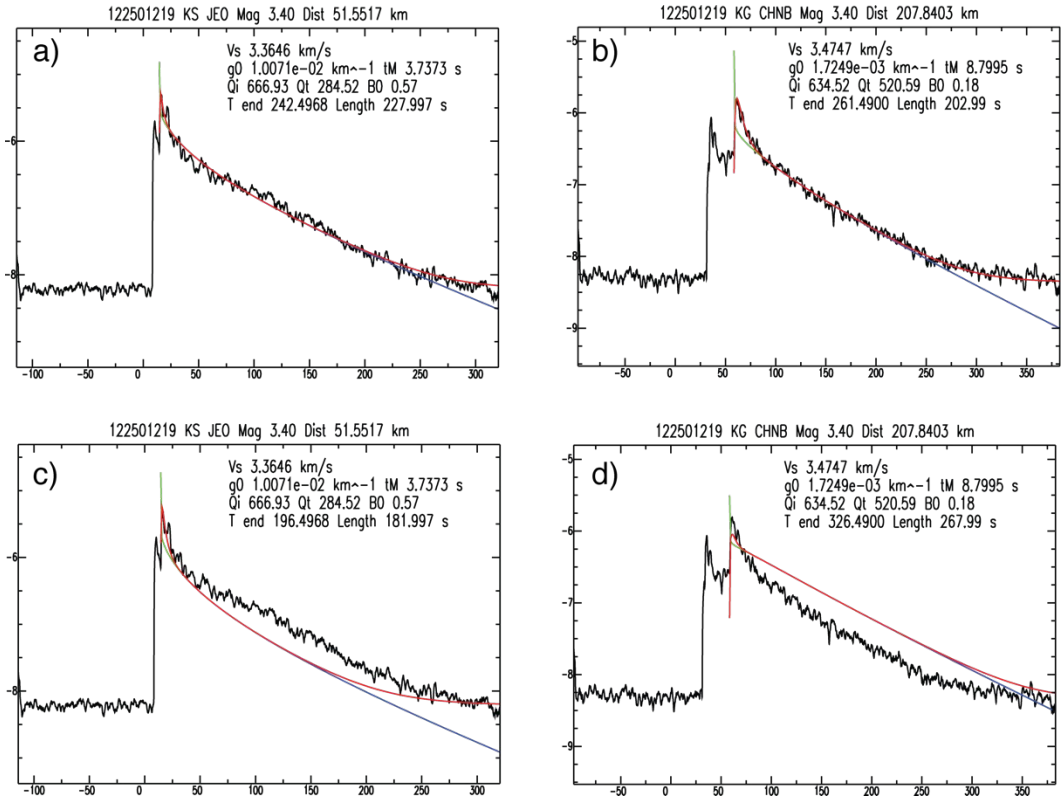

Figure 22. Examples of model-search for a) short- and b) long- distance records. *The red line shows an ensemble average of mode predictions and the blue line is a noise-corrected prediction of the red. The green line shows the wide-angle-scattering contribution of the red line*. If we use the weak wide-angle-scattering from the long-distance estimate for the short-distance prediction (c), we will underestimate the coda level. On the other hand, if we use the strong wide-angle-scattering from the short-distance estimate for the long-distance prediction (d), we will overestimate the coda level.

Figure 23. shows a spatial distribution of the scattering coefficient and intrinsic attenuation at 2-4 Hz. We mapped the parameters at the location of each station and event and computed a mean of values within a 1-degree by 1-degree block. Both scattering coefficients and intrinsic attenuation show strong regional variation depending on the tectonic setting. We found strong scattering and intrinsic attenuation in Honshu, Japan and weaker scattering and intrinsic attenuation in Kyushu, Japan, the Korean Peninsula and Northeast China. We also found some good spatial correlation of the intrinsic attenuation to the direct L_g Q result from direct wave tomography inversion (Figure 16).

We do not claim that our approach finds a physically unique model for the trade-offs between the scattering and intrinsic attenuation. However, our model is a physical model for the radiative transfer theory and can accurately predict both shape and amplitude of entire waveform envelope from early- to late-coda. Our physics-based approach will be of great help in saving considerable effort and time required for future application studies on new target areas.

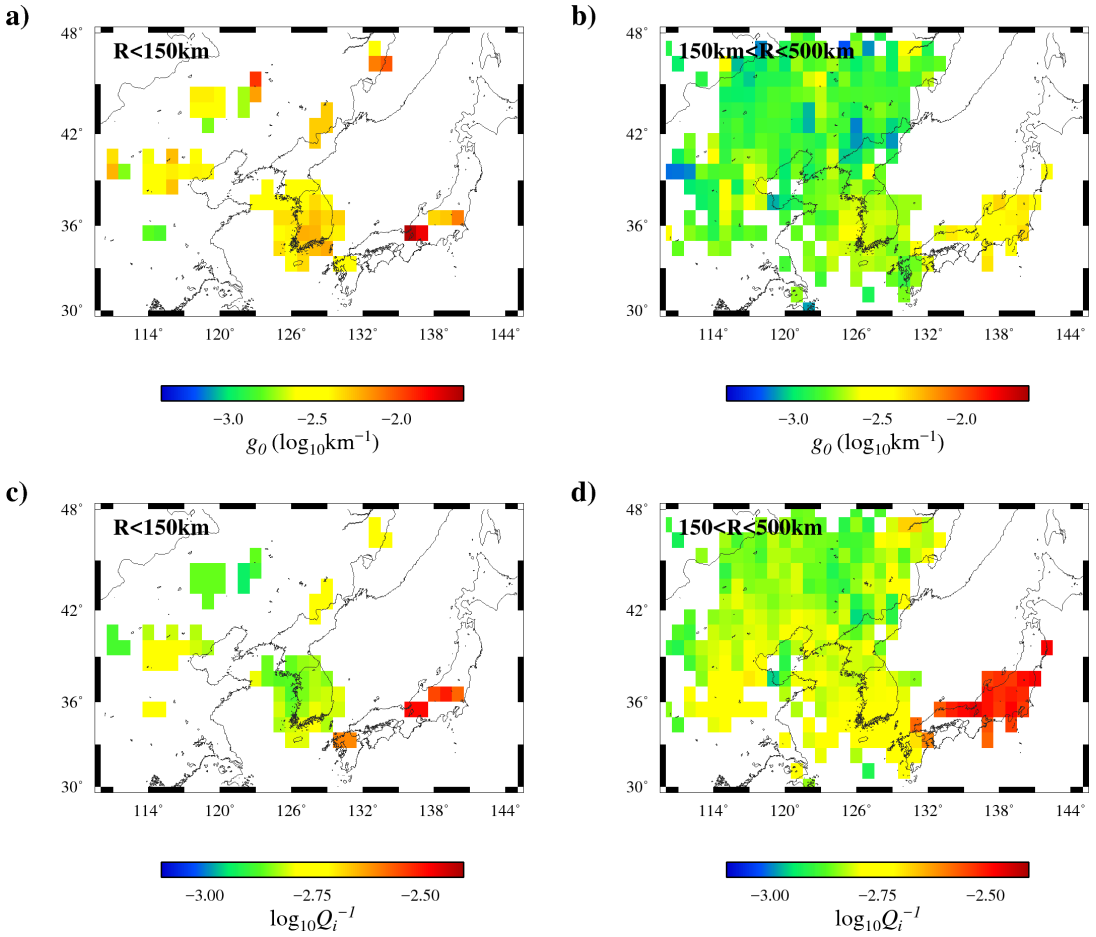

Figure 23. Map shows 1-degree block-mean of scattering (a and b) and intrinsic attenuations (c and d) of L_g coda from the short- (a and c) and long- (b and d) distance paths at 2-4 Hz.

4.5. 2-D Shape Calibration using Trans-dimensional Bayesian Tomography

We have developed a Bayesian tomography code that can map the 2-D distribution of the template modeling parameters through the State Department funded project (DOS Contract No SAQMMA-16-M-2404). This code determines the 2-D distribution of parameters required for hybrid template modeling by applying the latest trans-dimensional hierarchical Bayesian algorithm (Bodin et al., 2012). As the development of this tomography code is completed, unlike the existing method, which is defined as a radially symmetric function and an average value for a wide area, optimized template modeling parameters can be obtained for a given source-station pair. We believe this new approach will enable more sophisticated template modeling for regions where the shape of the template changes horizontally.

4.5.1. Bayesian Mapping Algorithm for Template Shape Parameters

The new hybrid modeling needs three model parameters, 1) characteristic time for the forward-scattering (t_M); 2) scattering coefficient for the coda exception ($g_0 = \frac{2\pi f}{vQ_s}$); and 3) inelastic energy

loss for the later coda decay rate (Q_i^{-1}). We have already implemented an algorithm that can extract posterior distribution of optimal envelope modeling parameters for given waveform data by applying the Bayesian inversion technique to the hybrid envelope model. The Bayesian tomography code uses the posterior distribution of modeling parameters as an input.

Our new hybrid modeling technique is a more advanced method compared to the single back-scattering model widely applied in previous monitoring applications. However, it is still oversimplifying the real Earth. The analytic solutions included in the hybrid formula are derived for a homogeneous random half space. This ignores the stratigraphic characteristics of Earth's crust and the effect from this appears as a distance dependence of the model parameters. 3-D mapping is required to take into account this layered property, but it is not realistic to apply 3-D calibration over a large area. In the previous 1-D coda calibrations, the distance dependence of the model parameters has been functionalized using hyperbolic curve fitting (Mayeda et al., 2003; Mayeda & Walter, 1996; Yoo et al., 2011).

We basically followed the method proposed by Takahashi et al., (2007) to correct the distance dependency of the model parameters, but also added a frequency term to the original method to account for the frequency dependency of the model parameters. We evaluated the significance of the improvement in model misfit Chi-Squared by an F-test and found that a bilinear model can reasonably fit the data. We found a reference 1-D model of the logarithm of modeling parameters against hypocentral distance and central frequency using Metroplis-Hastings MCMC method,

$$\log_{10} t_M = a_{0(t_M)} + a_{1(t_M)} \log_{10} R + a_{2(t_M)} \log_{10} f + a_{3(t_M)} \log_{10} R \log_{10} f$$
$$\log_{10} g_0 = a_{0(g_0)} + a_{1(g_0)} \log_{10} R + a_{2(g_0)} \log_{10} f + a_{3(g_0)} \log_{10} R \log_{10} f$$
$$\log_{10} Q_i^{-1} = a_{0(Q_i^{-1})} + a_{1(Q_i^{-1})} \log_{10} R + a_{2(Q_i^{-1})} \log_{10} f + a_{3(Q_i^{-1})} \log_{10} R \log_{10} f$$

eq. 47

where t_M, g_0 and Q_i^{-1} are the modeling parameters for the hybrid envelope modeling, R is the hypocentral distance, f is the central frequency, and a_0, a_1, a_2, a_3 are regression coefficients for each modeling parameters (see Table 1 and Figure 24).

Table 1. Parameters of the reference shape model

P	a_0	a_1	a_2	a_3
t_M	-0.9512 (±0.0600)	0.9486 (±0.0618)	0.2115 (±0.0677)	-0.1658 (±0.0639)
g_0	-1.1507 (±0.0640)	-0.6283 (±0.0623)	-0.2675 (±0.0795)	0.1071 (±0.0669)
Q_i^{-1}	-2.4726 (±0.0598)	-0.0580 (±0.0616)	-0.8863 (±0.0613)	0.1748 (±0.0638)

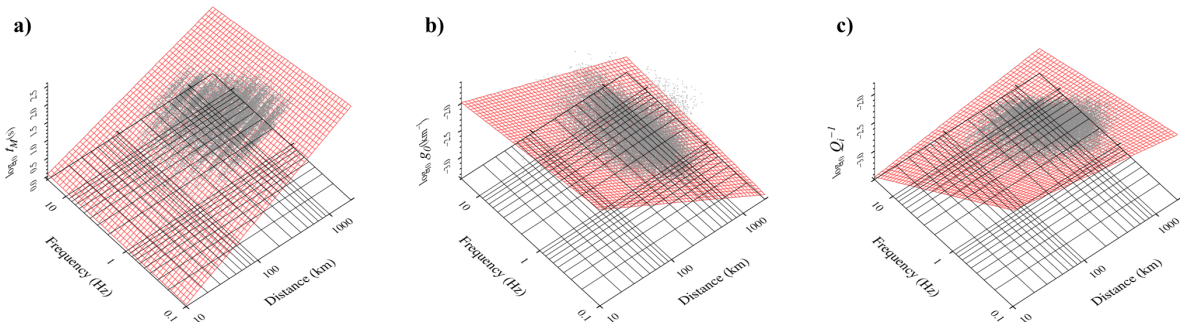

Figure 24. Reference shape models for a) characteristic time ($\boldsymbol{t_M}$), b) total scattering coefficient ($\boldsymbol{g_0}$), and c) inelastic energy loss ($\boldsymbol{Q_i^{-1}}$).

Then, we defined the logarithmic deviation of each model parameter from the reference model as,

$$\Delta\log_{10}t_M = \log_{10}t_M - \log_{10}t_{M(\text{ref})}$$
$$\Delta\log_{10}g_0 = \log_{10}g_0 - \log_{10}g_{0(\text{ref})}$$
$$\Delta\log_{10}Q_i^{-1} = \log_{10}Q_i^{-1} - \log_{10}Q_{i(\text{ref})}^{-1}$$

eq. 48

where $\Delta\log_{10}t_M$, $\Delta\log_{10}g_0$, and $\Delta\log_{10}Q_i^{-1}$ are the values for the Bayesian tomographic mapping. Our current interpretation for these values are the strength of scattering/absorbing due to the heterogeneities along the source-receiver path (e.g., Pezzo et al., 2016). Thus, the values are given by,

$$\Delta\log_{10}t_{M_n} = \frac{1}{R}\sum_m \Delta\log_{10}t_M{}^m \cdot dr^{m,n}$$
$$\Delta\log_{10}g_{0_n} = \frac{1}{R}\sum_m \Delta\log_{10}g_0{}^m \cdot dr^{m,n}$$
$$\Delta\log_{10}Q_i^{-1}{}_n = \frac{1}{R}\sum_m \Delta\log_{10}Q_i^{-1\,m} \cdot dr^{m,n}$$

eq. 49

where $\Delta\log_{10}t_{M_n}$, $\Delta\log_{10}g_{0_n}$, and $\Delta\log_{10}Q_i^{-1}{}_n$ are the values of ray n, and $\Delta\log_{10}t_M{}^m$, $\Delta\log_{10}g_0{}^m$, and $\Delta\log_{10}Q_i^{-1\,m}$ are the model parameters assigned to cell m, and $dr^{m,n}$ is the length of ray n across cell m.

4.5.2. Application to Northeast Asia

We selected about 4800 ray paths for each frequency band between 1-2 Hz, and about 700 ray paths for the 8 Hz band after quality control processes on the pre-sampled posterior of model parameters using the hybrid envelope modeling. Figure 25 shows ray-path coverage at each frequency band for the trans-dimensional Bayesian mapping and checkerboard model used in a synthetic test. For each frequency band and model parameter, we ran the Bayesian tomography code for 2 million MCMC interations (1 million for burn-in and 1 million for sampling) for 48 multiple chains (total 96 million MCMC iterations). Each run took 2-3 days on high-performance Linux workstation using 48 CPUs.

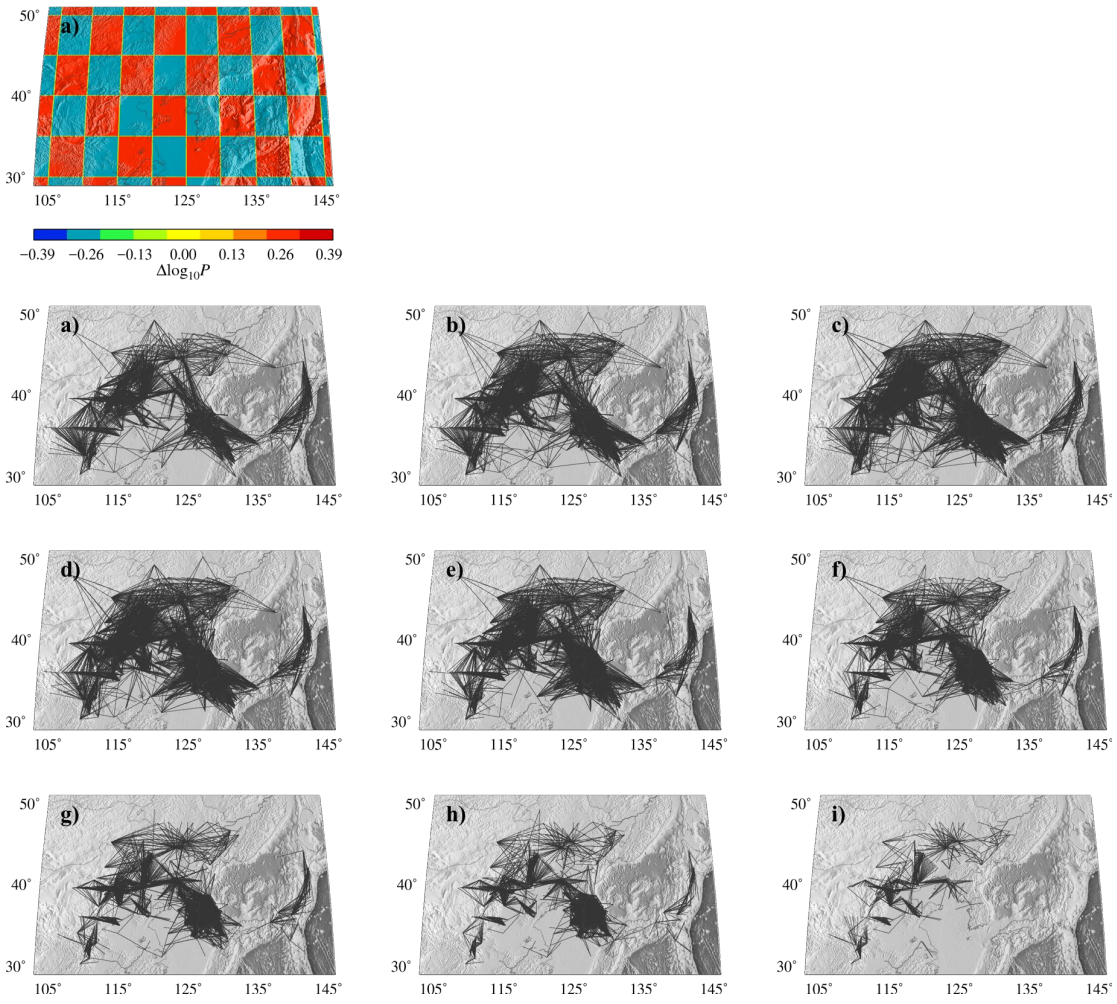

Figure 25. a) Checkerboard and the selected ray-paths at a) 0.5, b) 0.7, c) 1 Hz, d) 1.4, e) 2, f) 2.8, g) 4, h) 5.7, and i) 8 Hz.

Figures 26-34 show the ensemble average and the corresponding variance of the model parameters obtained from the trans-dimensional Bayesian mapping. Since t_M, g_0, and Q_i^{-1} all have identical ray path coverage, the checkerboard recovery test for each will be nearly identical. In Figures 26-34, we show the checkerboard recovery test for tm only. In most cases, except for the highest frequency band which does not have sufficient ray path coverage, we found good checkerboard recovery and low variance of the parameters for the Korean Peninsula, northeast China, and Japan. However, we could not obtain meaningful results for the East Sea (Sea of Japan) region due to low ray path coverage (see Figure 25) for all frequency bands.

Similar to what we have seen in the previous 1-D approach, Japan showed higher forward scattering and absorbing effect on the envelope shape compared to other regions. We found strong scattering/absorbing in Honshu, Japan and weaker in Kyushu, Japan, the Korean Peninsula and northeast China.

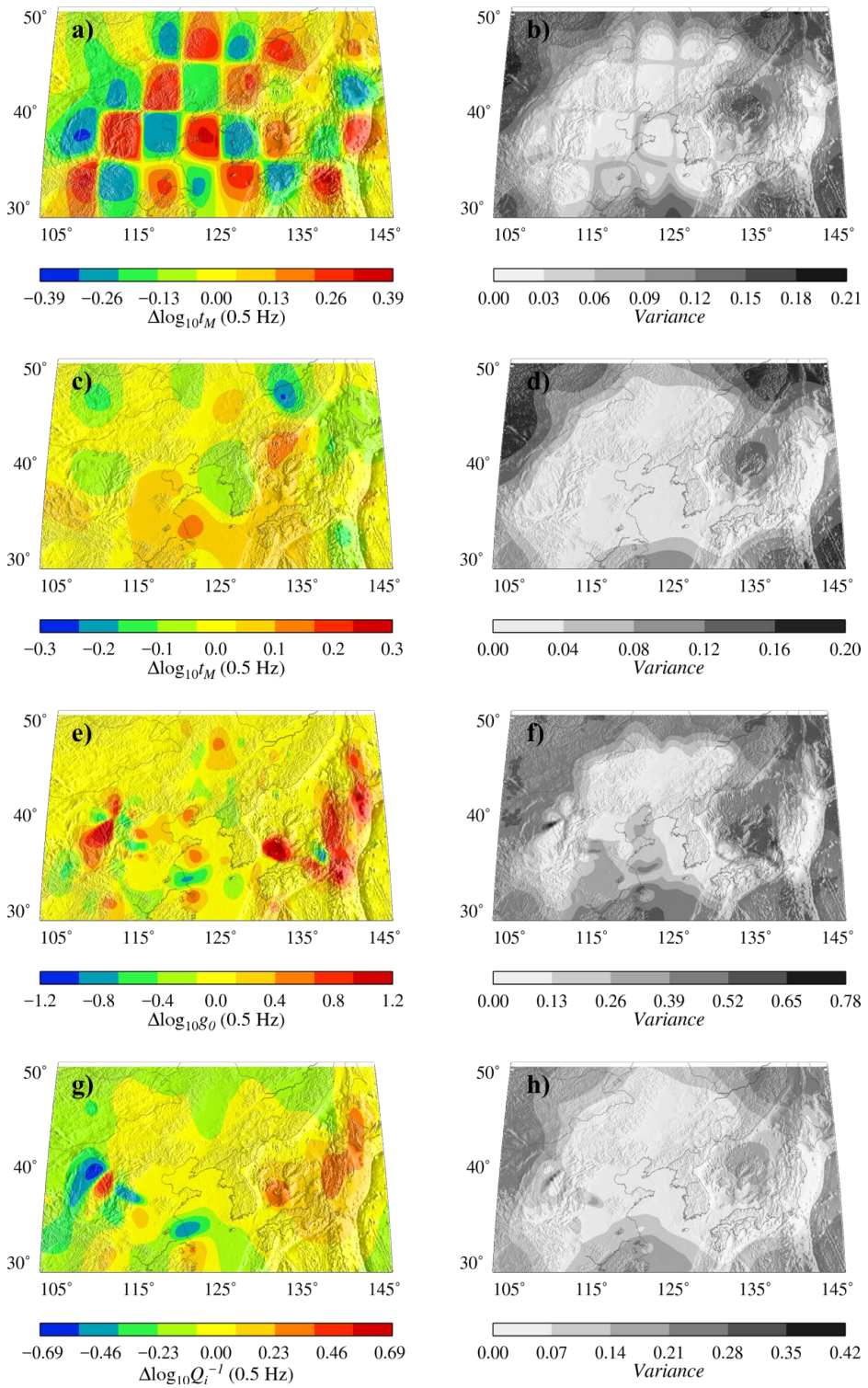

Figure 26. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 0.5 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

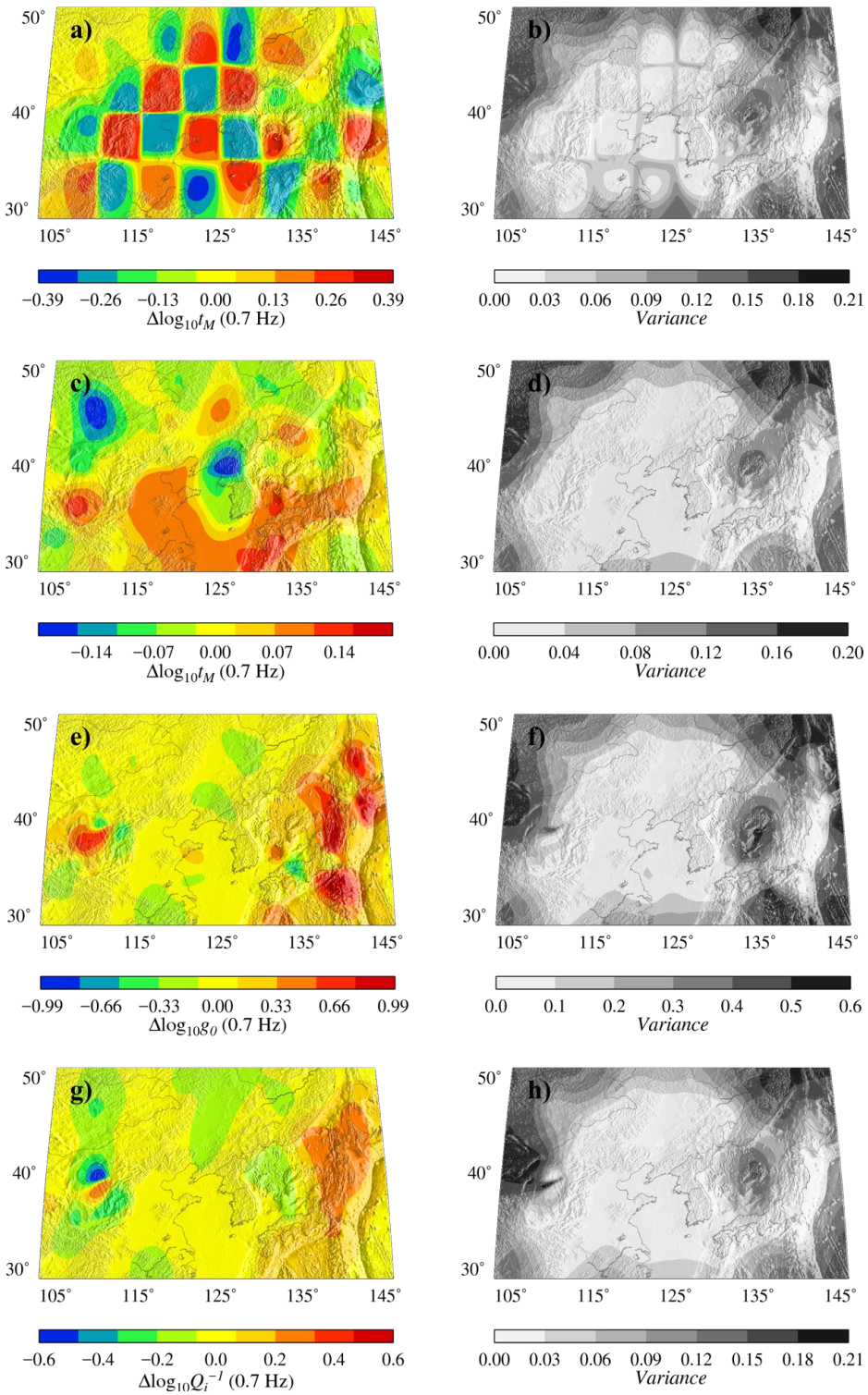

Figure 27. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 0.7 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

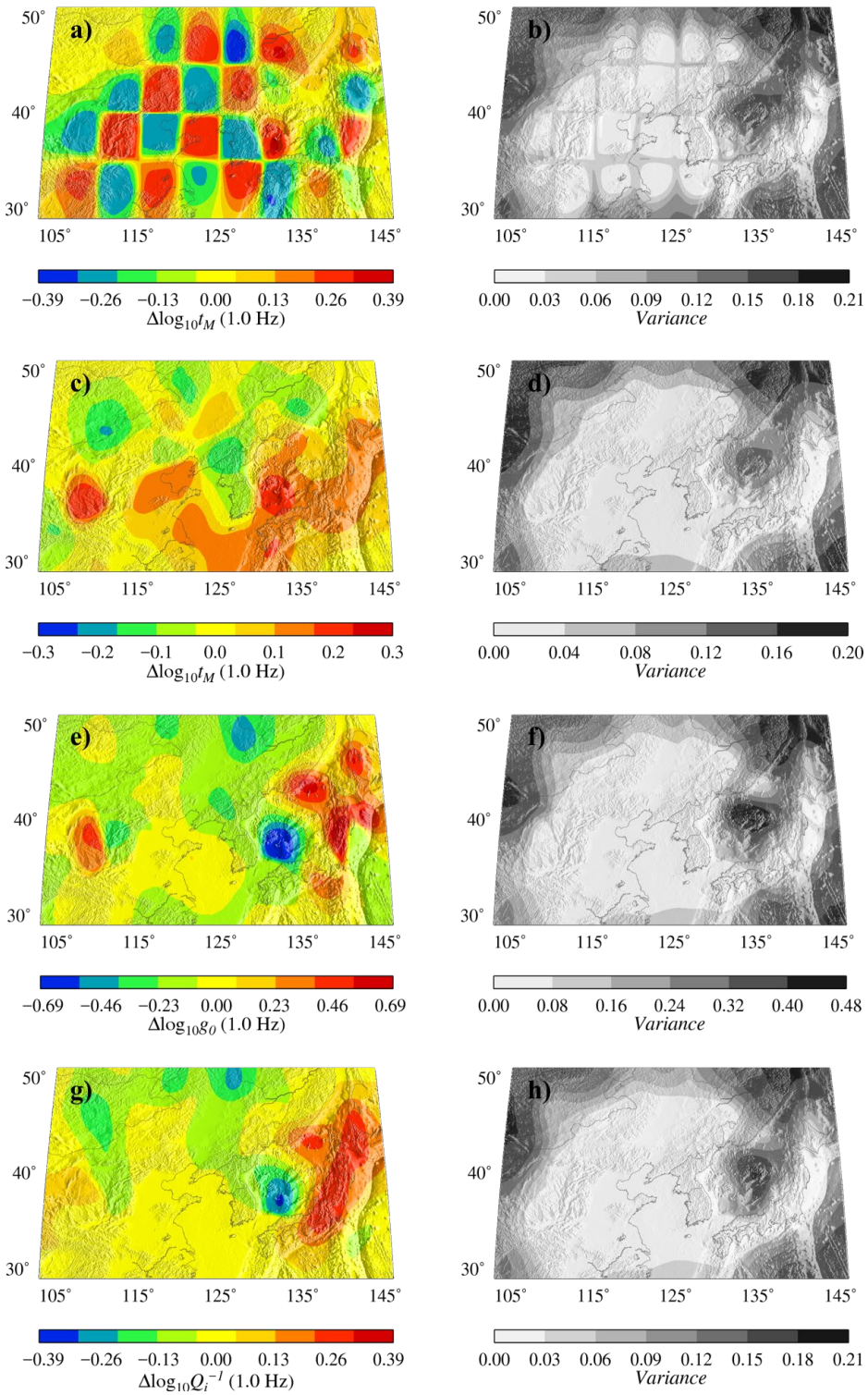

Figure 28. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 1 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

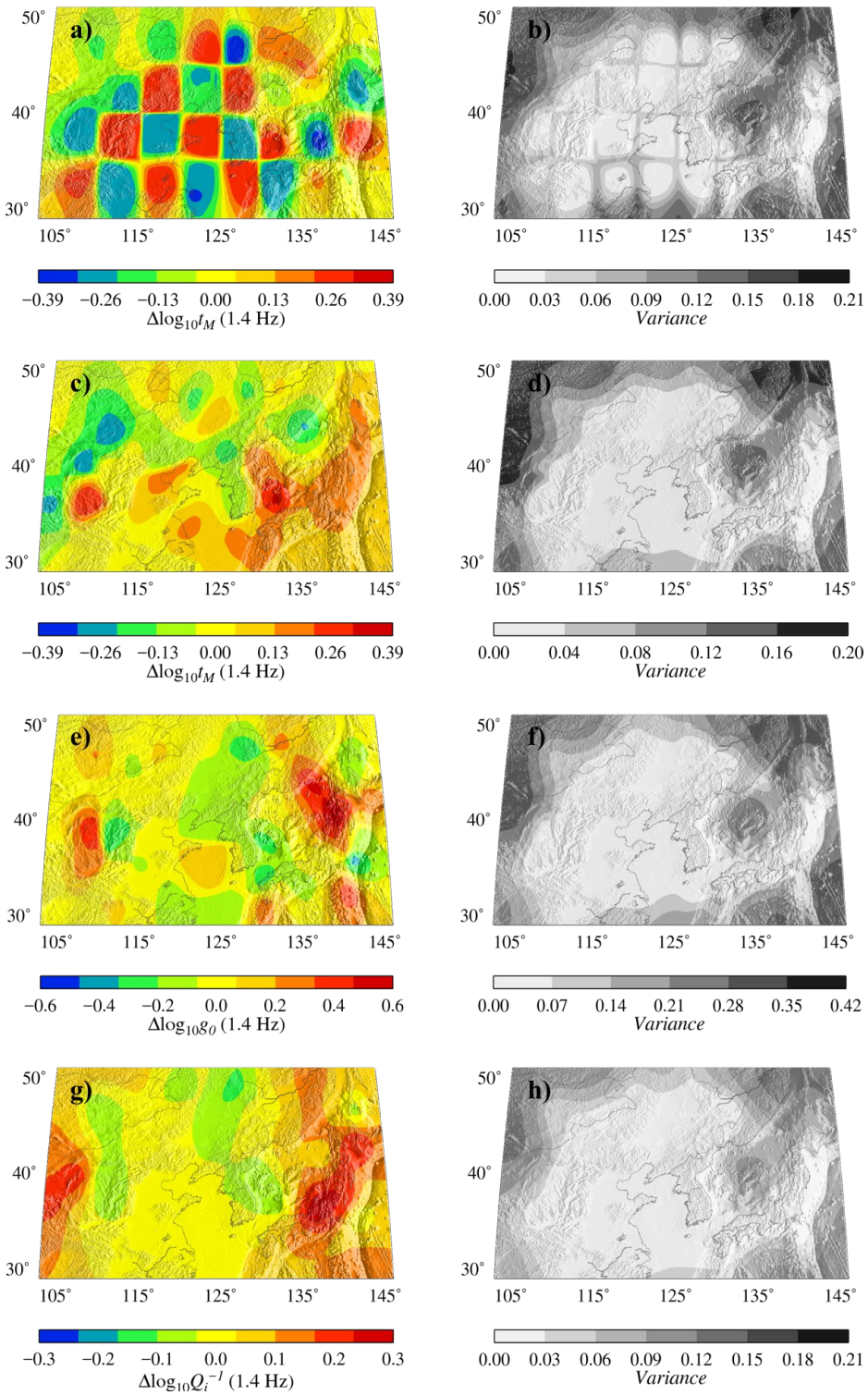

Figure 29. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 1.4 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

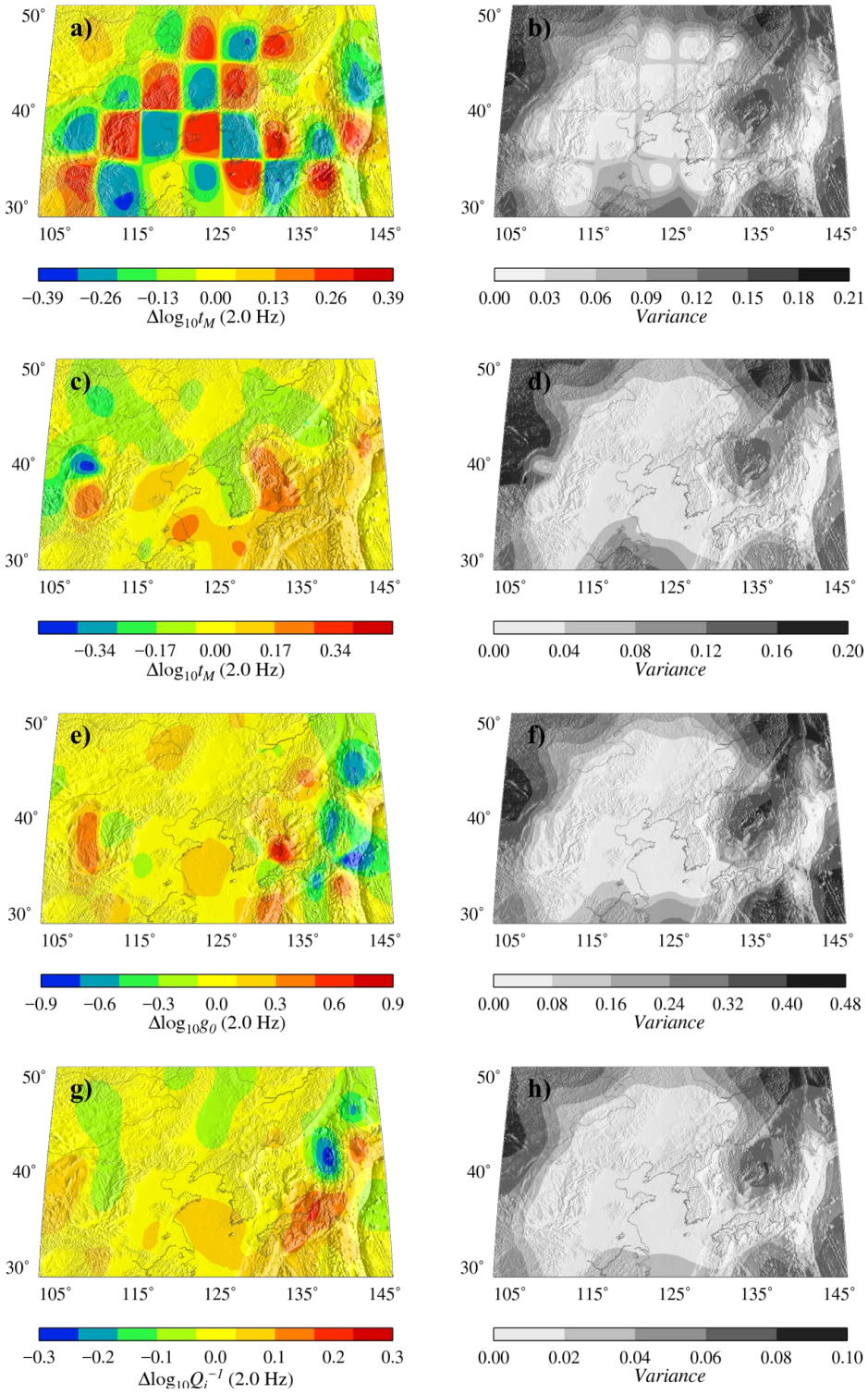

Figure 30. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 2 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration*.

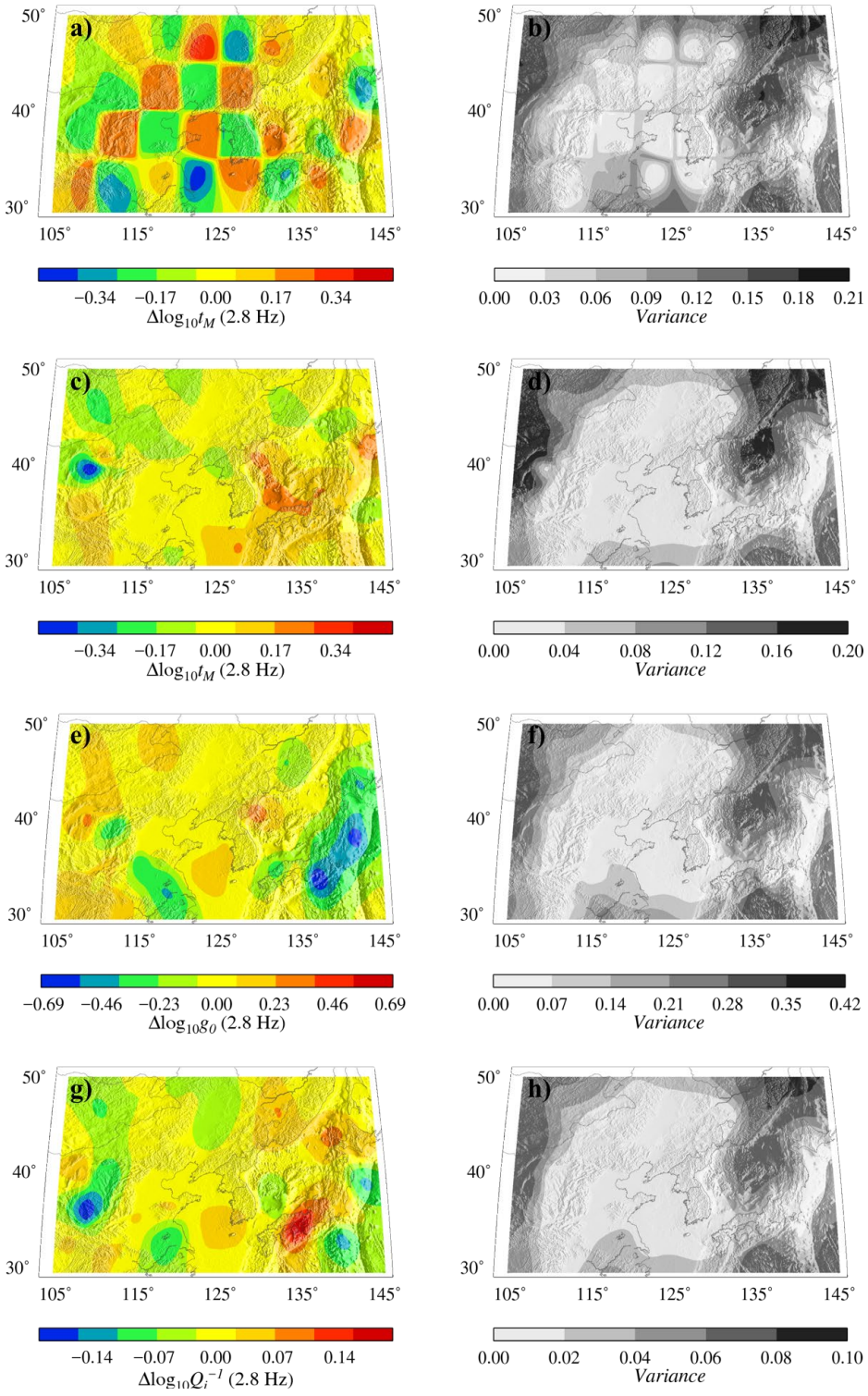

Figure 31. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 2.8 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

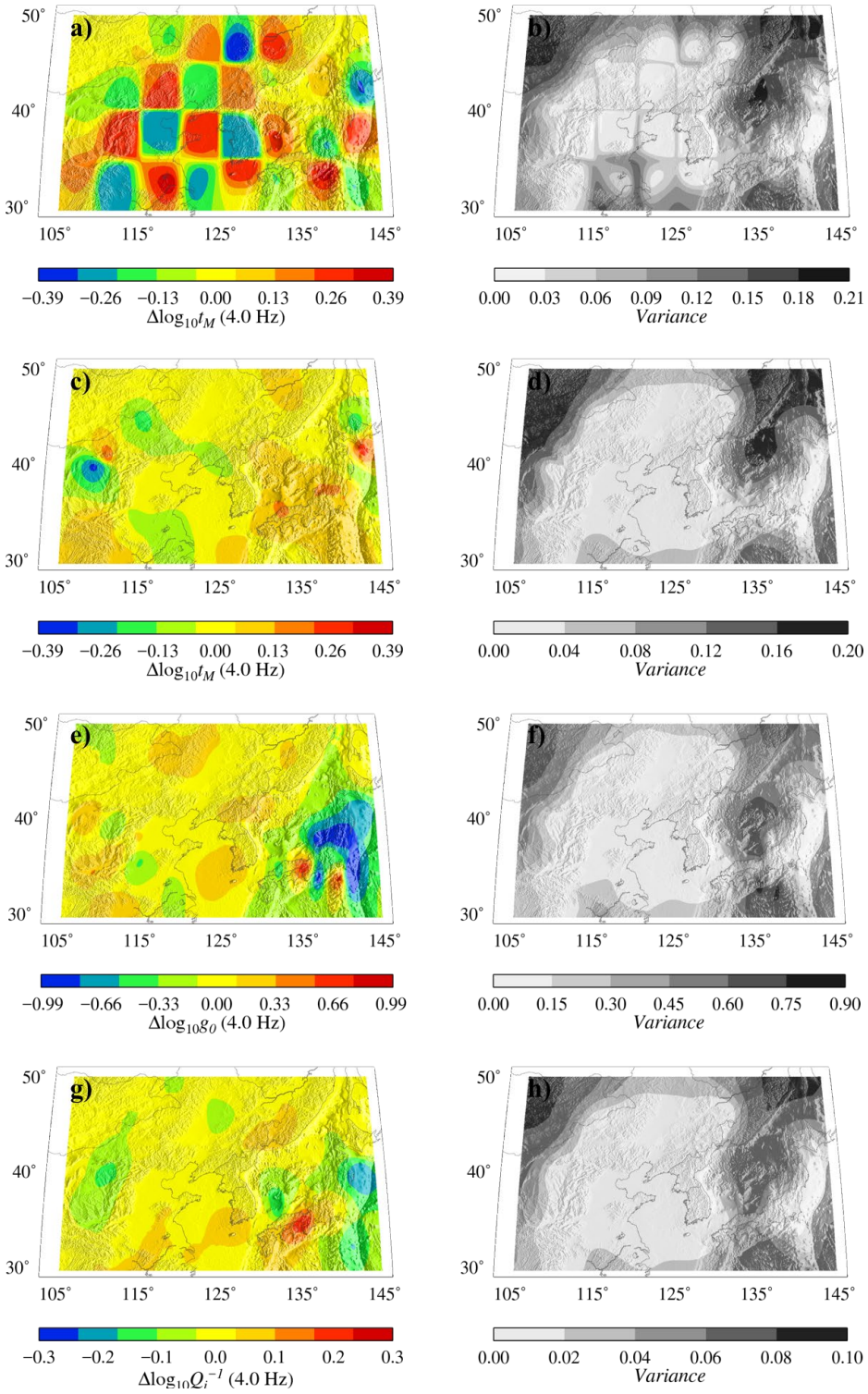

Figure 32. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 4 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

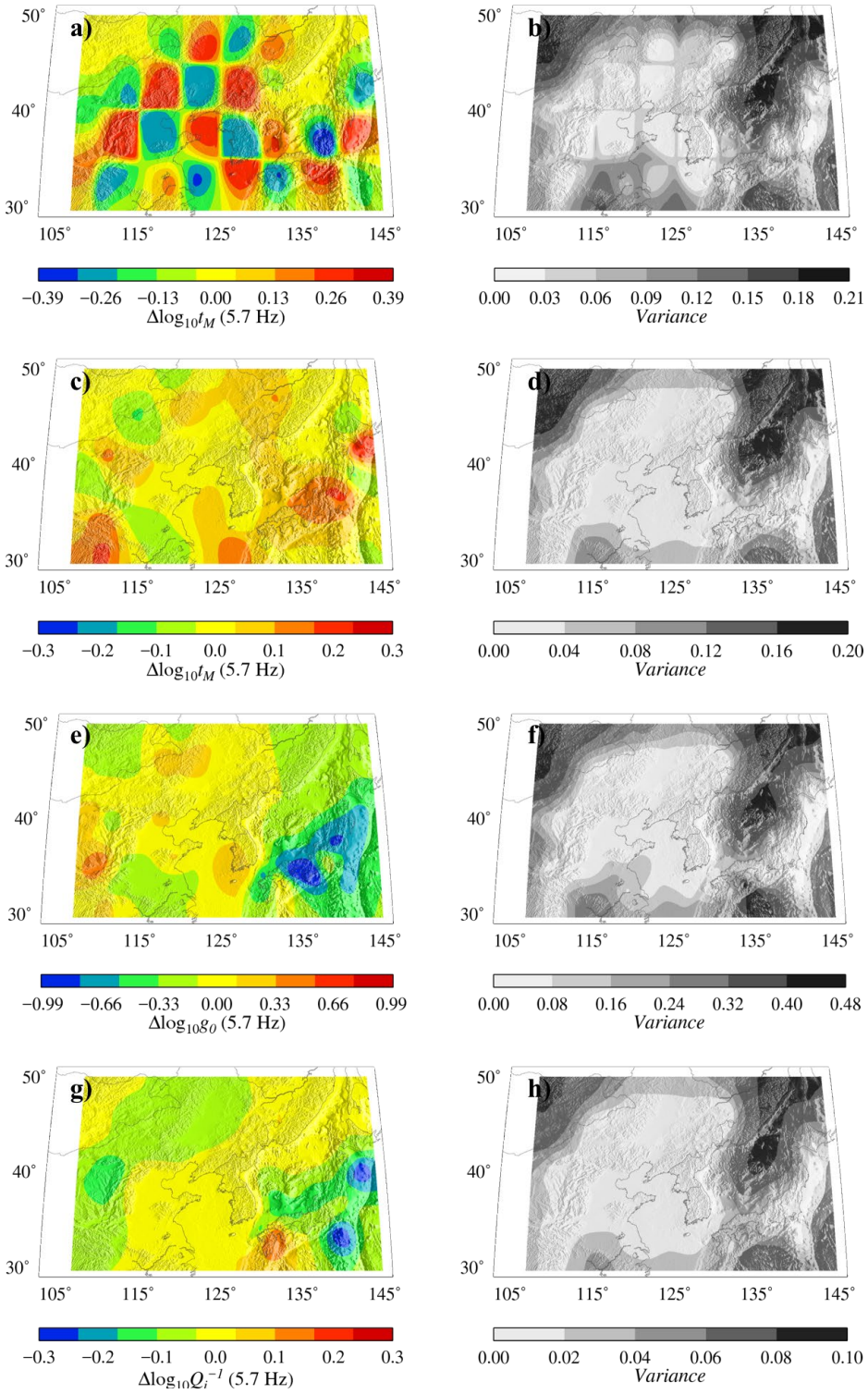

Figure 33. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 5.7 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

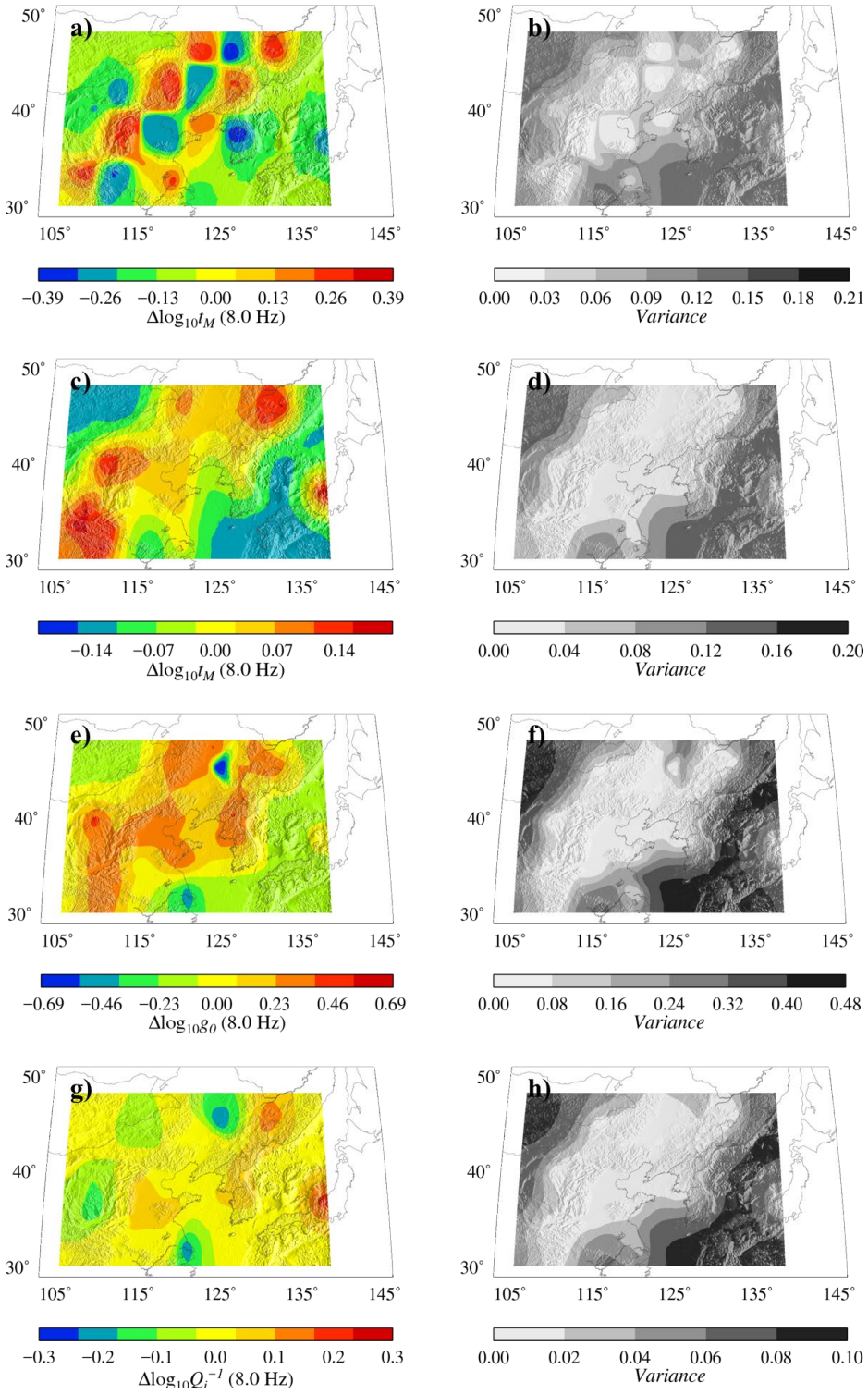

Figure 34. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 8 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

4.6. Full-waveform Template Matching with Full 2-D Calibration

We obtained a 2-D calibration model for the L_g-template shape parameters using the Trans-Dimensional Bayesian tomography technique (e.g., Bodin et al., 2012) newly developed under the support of the U.S. Department of State's V-Fund program (DOS Contract No SAQMMA-16-M-2404). We completed compiling the envelope shape model to a new forward method which uses optimal template modeling parameters for a given event-station pair for template matching. We have confirmed that the new forward method improves template matching compared to the previous approach using the binary model (e.g., Island of Japan).

In Figure 35 we show the estimated characteristic times (t_M) of the L_g-wave template at 1, 2, and 4 Hz, obtained from applying the trans-dimensional Bayesian mapping technique. We estimate the optimal forward scattering parameters from the 2-D calibration model for each propagation path and frequency band and use them in the hybrid envelope template modeling. The results show that the 2-D calibration model more accurately predicts the observed narrowband envelope shapes (Fig. 35(A,B,C)) for three earthquakes at similar distances (~200 km) from the seismic station BUS (Busan, S. Korea). The shape of the observed narrowband envelope varies considerably depending on event location. Notably, the shape of the earlier coda varies significantly due to the differences in the forward-scattering characteristics of each propagation path. For example, the offshore Japan earthquake in Figure 35C shows more broadened and delayed early coda compared to the inland Korean earthquake in Figure 35A.

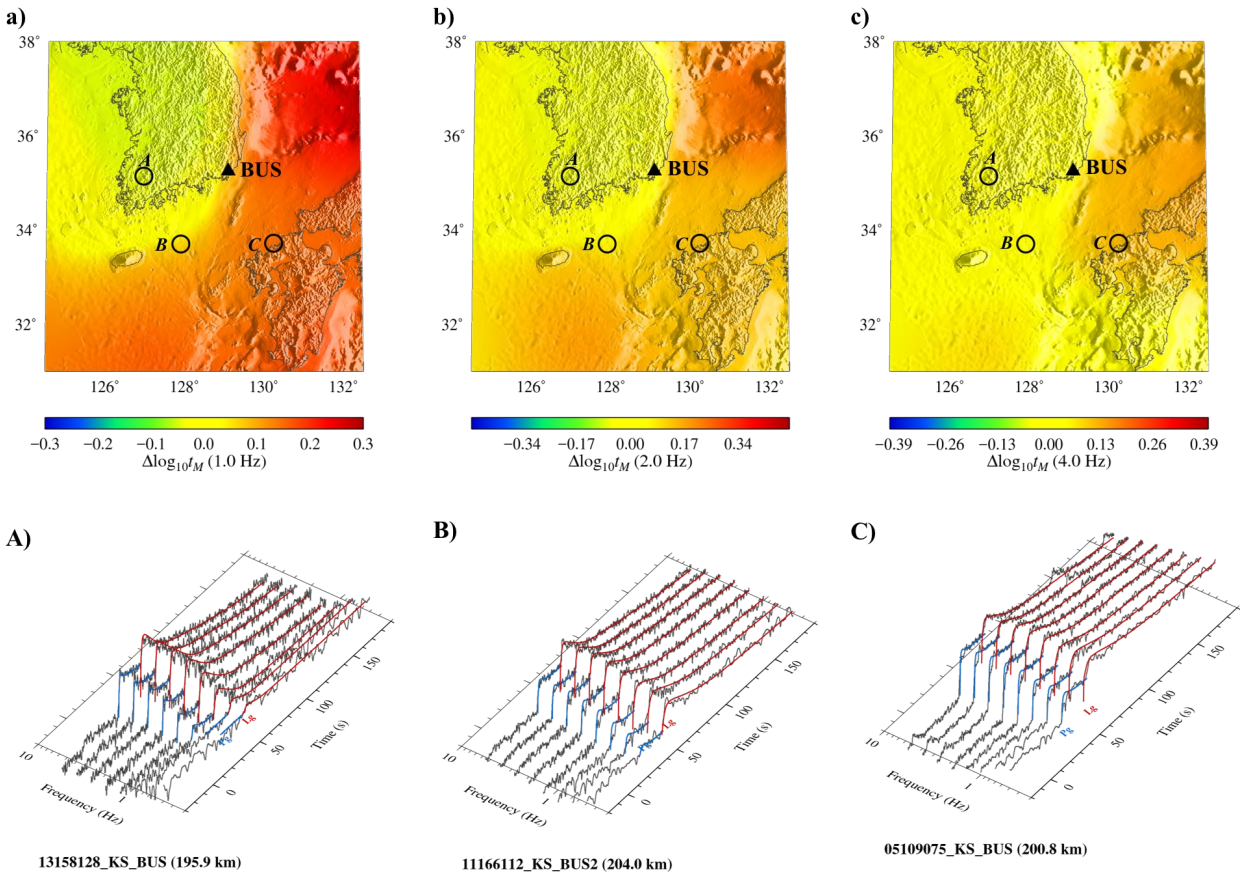

Figure 35. Example illustrating the use of a 2-D calibration model and full-waveform template matching approaches. a), b), and c) show the calibration maps for characteristic time ($\boldsymbol{t_M}$) of the L_g-template at 1, 2, and 4 Hz, respectively. Circles A, B, and C show the location of earthquakes at similar distances (~ 200km) from the Busan station (triangle). Earthquake A: a M3.2 earthquake southeast of the Korean peninsula on 2013/06/07 20:56:59, Earthquake B: a M3.8 earthquake in the Korea Strait on 2011/06/15 05:21:48, and Earthquake C: a M4.7 earthquake off-shore of Fukuoka, Japan on 2005/04/19 21:11:25. In subplots A), B), and C) (bottom row) we show the excellent fits that result from applying full-waveform template modeling for the three earthquakes.

4.7. 2-D Calibration of P_g-template

The P_g phase is one of the most critical components for yield estimation and event discrimination of underground nuclear explosions at local and near-regional distances. Thus, it is essential to characterize and quantify the scattering and attenuation characteristics of the Korean Peninsula to obtain more stable source parameters through accurate prediction of P_g coda energy. The P_g-template model initially developed in the previous year is based on the hybrid formula with a 1-D shape parameterization and produces significantly improved amplitude predictions over the entire envelope compared to the previous methods based on the back-scattering formula (Pasyanos et al., 2012; Phillips et al., 2011).

To further reduce the amplitude measurement error in the P_g coda, we focused on obtaining the 2-D envelope shape calibration model for the P_g template. However, the P_g -wave has a much shorter time window compared to the L_g-wave. This makes it very difficult to confidently constrain P_g envelope modeling parameters using a direct sampling approach of individual seismic records which was a very successful method applied to the long-lasting L_g-coda wave. To characterize the P_g envelope modeling parameters and its uncertainty, we designed a new full-Bayesian inversion method to find an optimal von Karman random media with frequency dependent intrinsic attenuation that explains the given stacked P_g and L_g source-normalized envelopes simultaneously.

We iteratively update the envelope model parameters and tomographic calibration models. Through the iterative inversion process, the envelope model parameters and the calibration model can be sequentially optimized by reducing the bias due to the trade-off between the two model components. Once we have an updated tomographic calibration models, we compute the source-normalized envelopes using independent estimates of the source and site amplification terms obtained through the tomographic inversion. We then stack them into each 50 km distance bins.

Figures 37, 38, and 39 show the stacking of source-normalized *Pg* and *Lg* envelopes at 0.5, 1, 2, 4, and 8 Hz from near and around the Korean Peninsula, northeastern China, and central Japan (see A, B, and C in Figure 36c). The shape of the earlier coda is significantly varied due to the difference in forward and wide-angle scattering characteristics across the regions. For example, strong forward and wide-angle scattering with high attenuation in central Japan (Figure 39 and C region in Figure 36c) cause broadened earlier coda and rapidly decaying late coda. The Korean Peninsula (Figure 37 and A region in Figure 36c) and northeastern China (Figure 38 and B region in Figure 36c) show a very similar level of later coda decay, whereas the envelope broadening from the forward scattering is different between the two regions. We see that the shape and attenuation of the P_g envelope also varies from region to region and seems to correlate with the variations of the L_g envelope.

We explored the envelope modeling parameters and its uncertainty which control the shape and attenuation of both P_g- and L_g-envelopes. We applied a full-Bayesian inversion technique based on a MCMC sample method (Hastings, 1970; Metropolis et al., 1953) to obtain the posterior distributions of the envelope modeling parameters which explain P_g and L_g source-normalized envelopes simultaneously. We follow the spectrum division method (Sato, 2016; Sato & Emoto, 2017, 2018) to estimate the narrow and wide-angle scattering coefficients for a given von Karman random media.

Figures 40, 41, and 42 show inversion results for the Korean Peninsula, northeastern China, and central Japan. The homogeneous random model with frequency-dependent attenuation formula provides a reasonable fit to the data for a broad range of distances and frequencies for both P_g and L_g. Compared to the previous direct sampling methods that use single event station data at a single phase, this new method of using averaged envelopes with a wide range of distances and frequency bands simultaneously tends to provide much more stable behavior of the model parameters. More importantly, the new approach provided fully characterized uncertainty information for individual model parameters. The uncertainty of each modeling parameter will be used to quantify amplitude measurement errors which are directly related to the envelope modeling process.

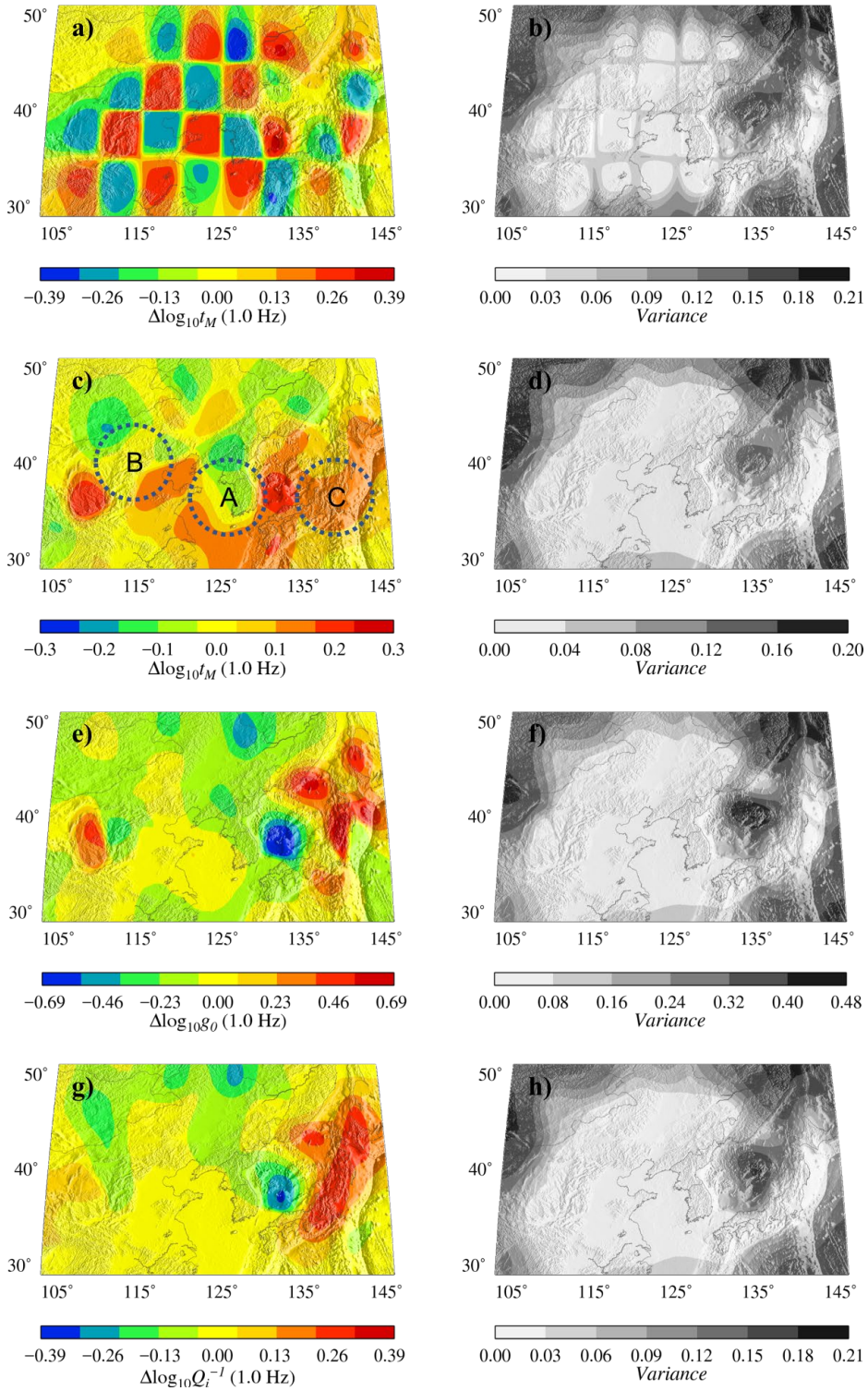

Figure 36. a) Checkerboard recovery test and tomography results of c) $\boldsymbol{t_M}$, e) $\boldsymbol{g_0}$, and g) $\boldsymbol{Q_i^{-1}}$ at 1 Hz from the trans-dimensional hierarchical Bayesian tomography. b), d), f) and h) are the corresponding variance from the checkerboard test and real data. *We note that the* $\boldsymbol{Q_i^{-1}}$ *is for template modeling and is different from the attenuation Q for the amplitude calibration.*

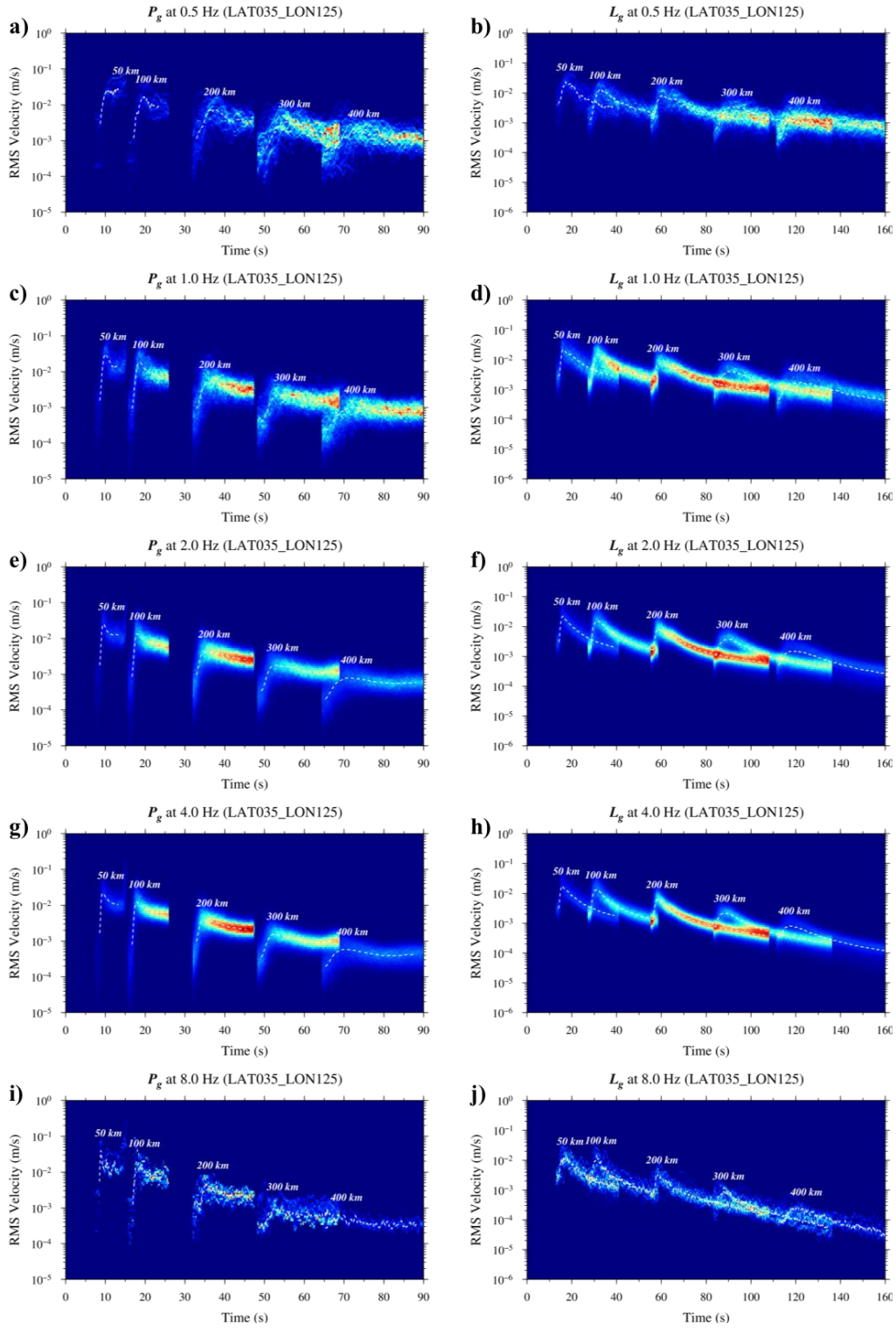

Figure 37. Stacking of source-normalized *Pg* (a, c, e, g, and i) and *Lg* (b, d, f, h, and j) envelopes at 0.5, 1, 2, 4, and 8 Hz from near and around the Korean Peninsula (see A region in Figure 2c).

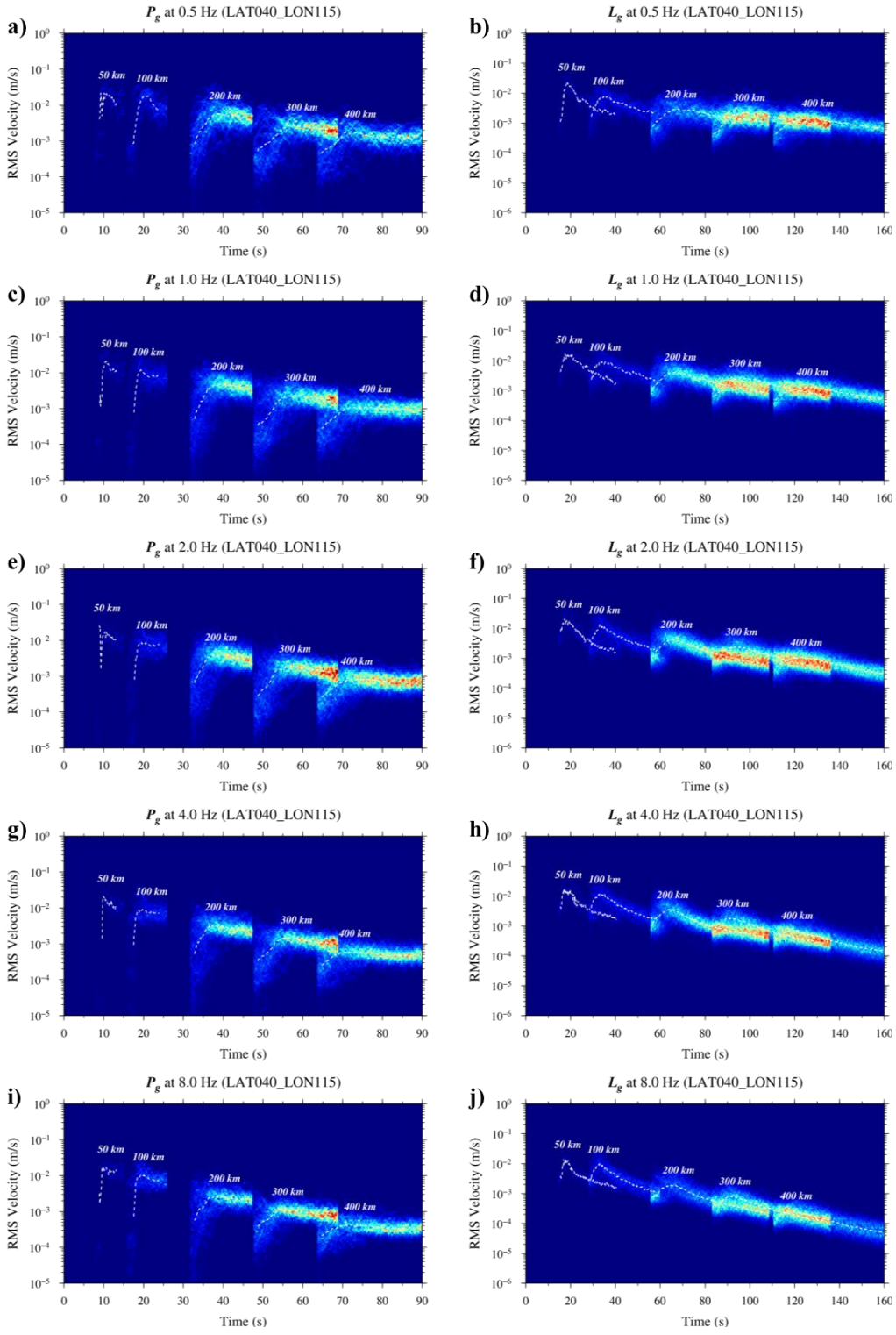

Figure 38. Stacking of source-normalized *Pg* (a, c, e, g, and i) and *Lg* (b, d, f, h, and j) envelopes at 0.5, 1, 2, 4, and 8 Hz from near and around northeastern China (see B region in Figure 2c).

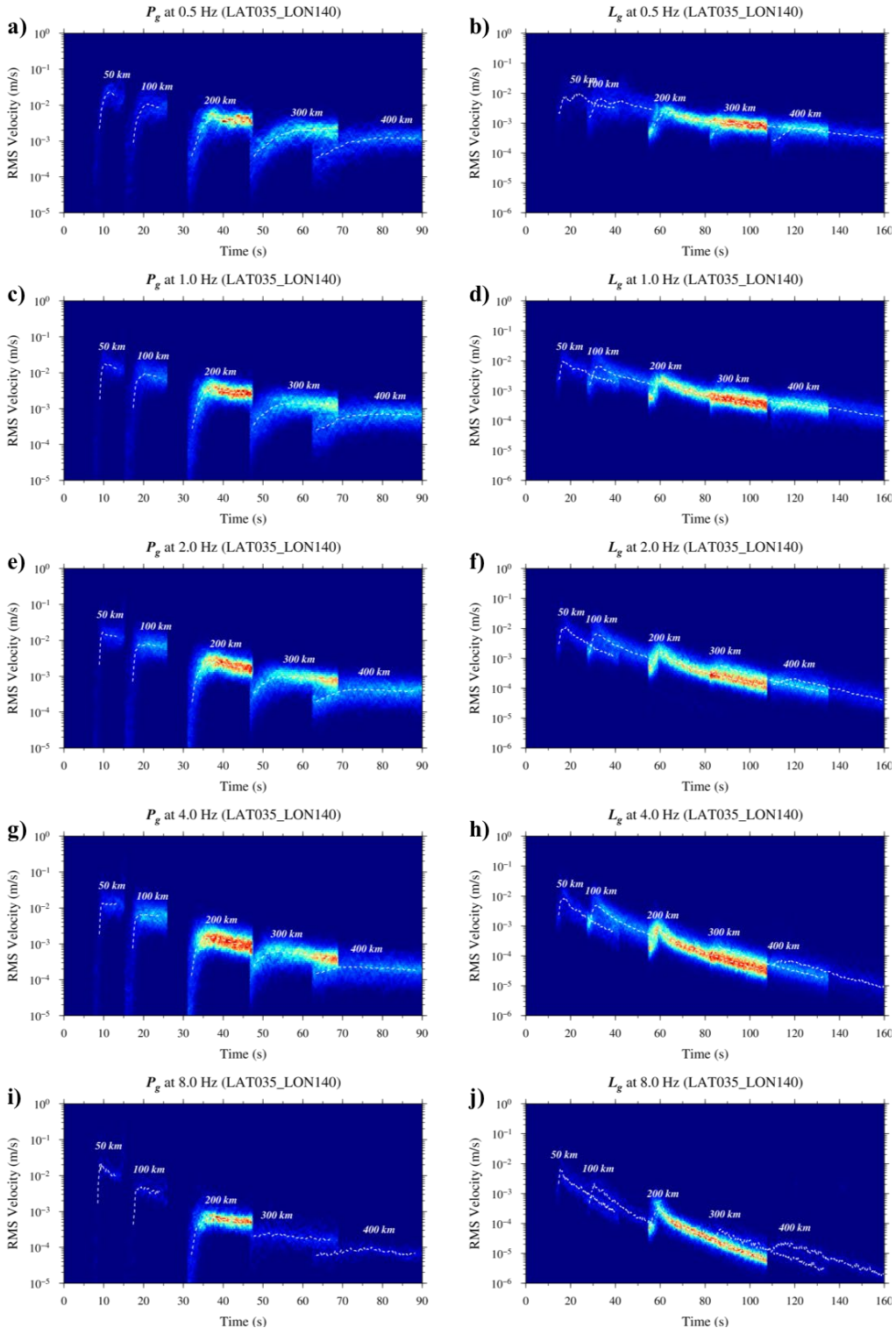

Figure 39. Stacking of source-normalized *Pg* (a, c, e, g, and i) and *Lg* (b, d, f, h, and j) envelopes at 0.5, 1, 2, 4, and 8 Hz from near and around central Japan (see C region in Figure 2c).

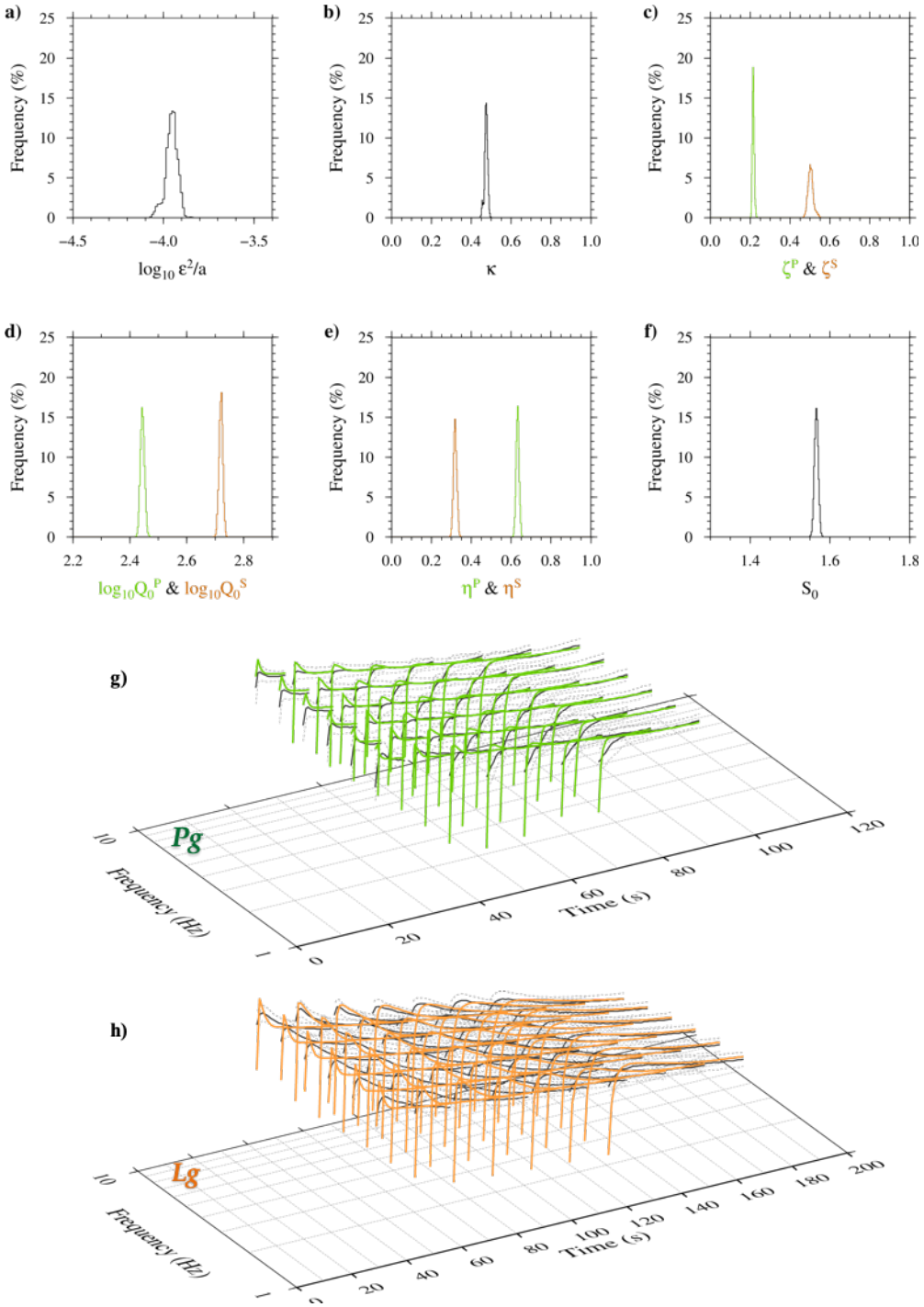

Figure 40. A full Bayesian inversion result for the envelope modeling parameters for the Korean Peninsula, a) $\log_{10} \varepsilon^2/a$, b) the Hust number κ, c) tuning parameters for spectrum division (Sato, 2016; Sato & Emoto, 2017, 2018), d) and e) parameters of frequency-dependent attenuation for P_g (green) and L_g (orange), respectively, and f) constant for source level adjustment. *Black, green, and orange lines in g) and h) are the averaged, predicted envelopes for P_g and L_g, respectively.*

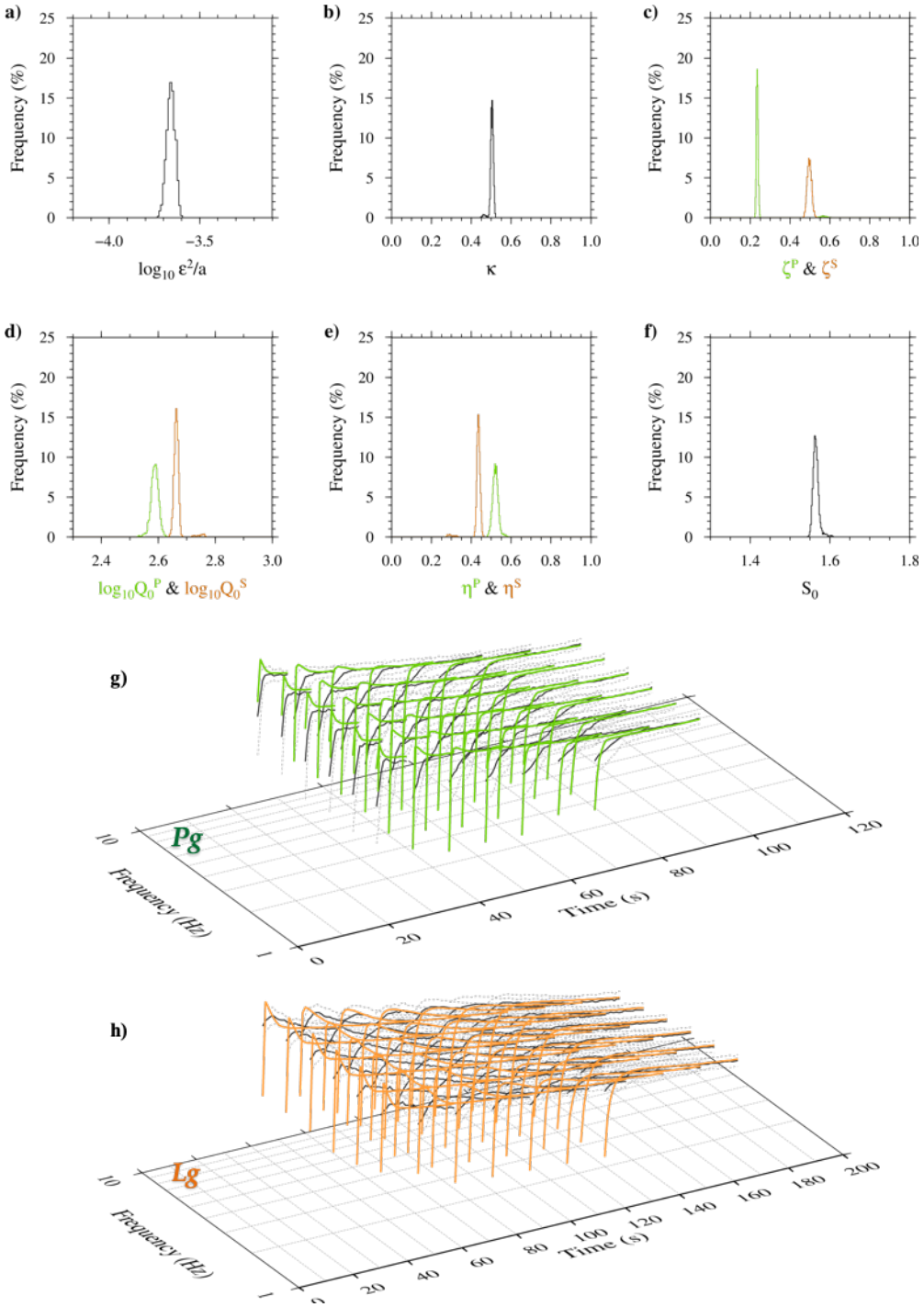

Figure 41. A full Bayesian inversion result for the envelope modeling parameters for northeastern China, a) $\boldsymbol{\log_{10} \varepsilon^2/a}$, b) the Hust number $\boldsymbol{\kappa}$, c) tuning parameters for spectrum division (Sato, 2016; Sato & Emoto, 2017, 2018), d) and e) parameters of frequency-dependent attenuation for P_g (green) and L_g (orange), respectively, and f) constant for source level adjustment. *Black, green, and orange lines in g) and h) are the averaged, predicted envelopes for P_g and L_g, respectively.*

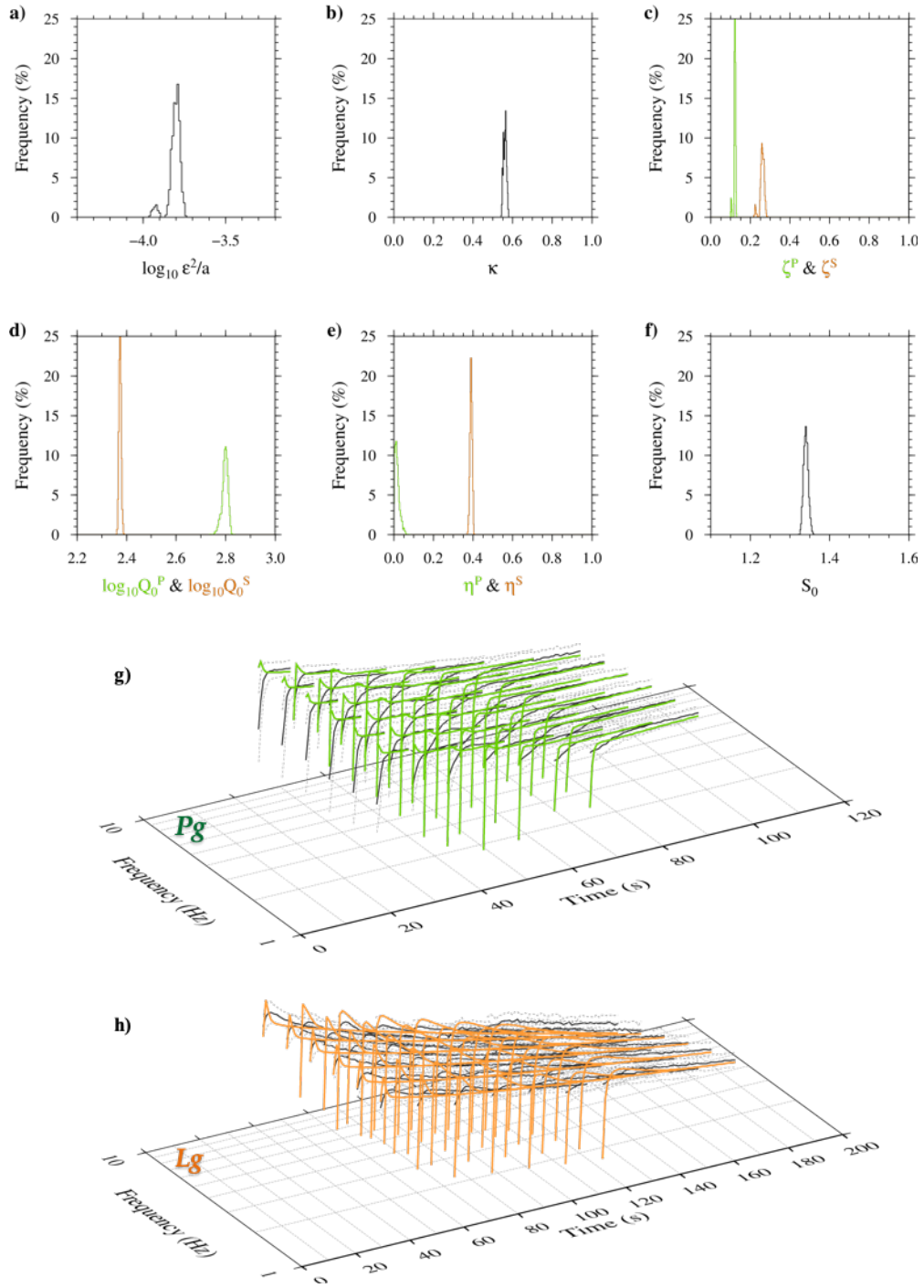

Figure 42. A full Bayesian inversion result for the envelope modeling parameters for central Japan, a) $\log_{10} \varepsilon^2/a$, b) the Hust number κ, c) tuning parameters for spectrum division (Sato, 2016; Sato & Emoto, 2017, 2018), d) and e) parameters of frequency-dependent attenuation for P_g (green) and L_g (orange), respectively, and f) constant for source level adjustment. *Black, green, and orange lines in g) and h) are the averaged, predicted envelopes for P_g and L_g, respectively.*

4.8. Envelope Shape Calibration of Mantle Phases

The shorter time windows and much fewer high SNR records compared to the corresponding crustal phases make it difficult to constrain the envelope shape parameters of the mantle phases. In particular, the S_n phase in our study area is not predominant, making it more difficult to obtain sufficient data for training. We revised and improved the shape calibration of the mantle phases using the source-normalization method which was previously developed to calibrate regional P_g-templates. Compared to the previous direct sampling and regression approach, we obtained a more stable solution by applying the new full-Bayesian method, which uses stacked source-normalized envelopes, to the P_n-template of the Korean Peninsula earthquake. Above all, we were able to fully characterize the uncertainty of each template modeling parameter through the new Bayesian approach.

We applied a full-Bayesian inversion technique based on a MCMC method to explore the template modeling parameters and its uncertainty which control the shape and attenuation of the mantle phases. First, we computed source-normalized envelopes using independent estimates of the source and site amplification terms obtained through the tomographic inversion and stacked them into each 50 km distance bins (Figure 43). Then, we used the MCMC sampling method (Hastings, 1970; Metropolis et al., 1953) to obtain the posterior distributions of the template modeling parameters which explains the given P_n and S_n source-normalized envelopes simultaneously.

Unlike results obtained by applying this new method to crustal phases, we failed to find a model that simultaneously explains P_n and S_n envelopes. The biggest issue is that our hybrid envelope modeling formula does not well explain the shape of the source-normalized S_n. This result is presumably because S_n is not predominate in the study area, resulting in a more distorted appearance by the preceding P_g-coda energy.

Figure 44 shows an inversion result for the P_n envelope parameters of the Korean Peninsula earthquake. The new full-Bayesian method applied to the stacked source-normalized envelopes showed a more stable solution convergence than the previous direct sampling and regression approach. Most importantly, we can understand the uncertainty of template modeling parameters through new methods. The uncertainty of each modeling parameter can be very important in quantifying amplitude measurement errors which are directly related to the envelope modeling process.

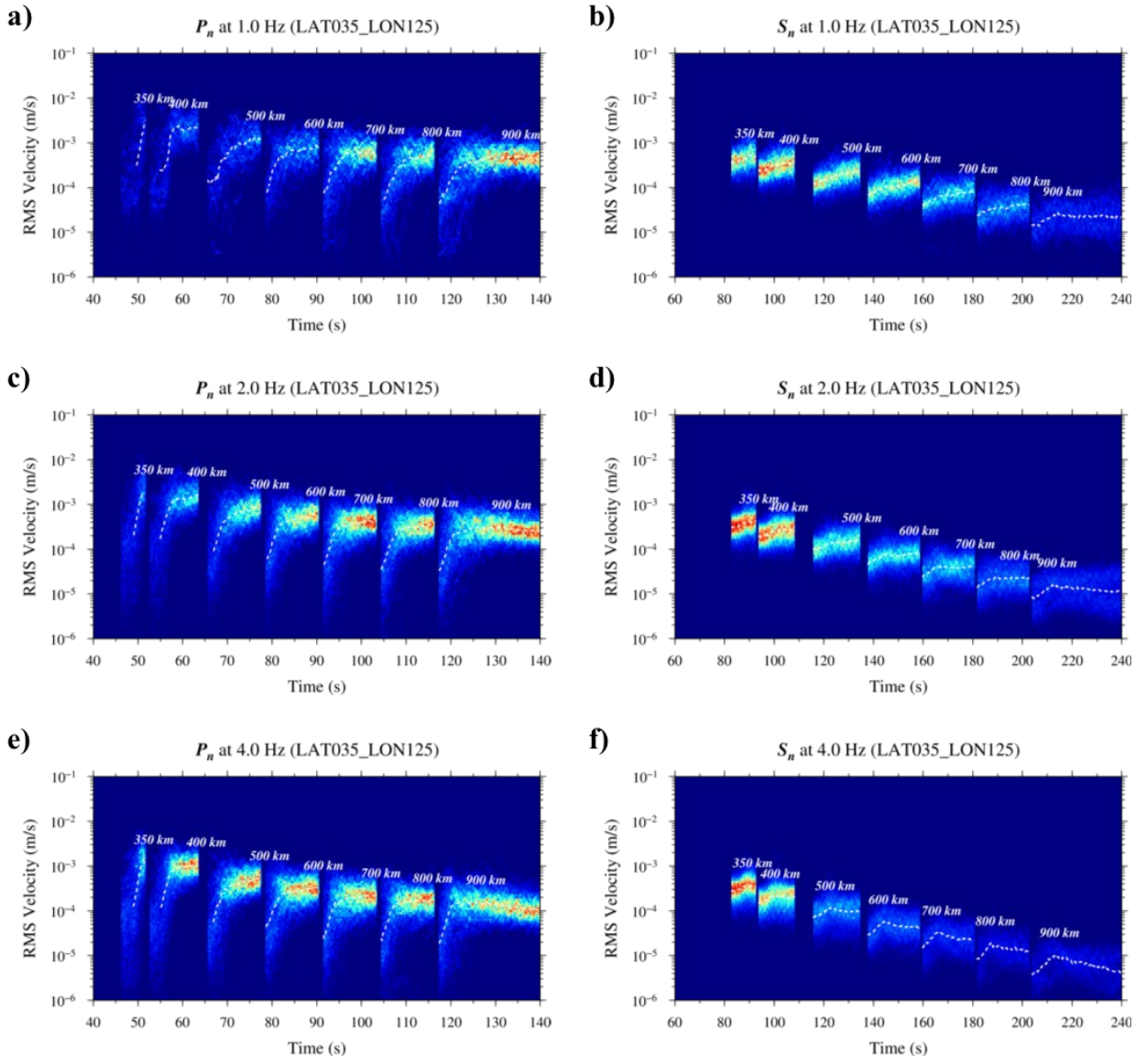

Figure 43. Stacking of source-normalized P_n (a, c, and e) and S_n (b, d, and f) envelopes at 1, 2, and 4 Hz (Korean Peninsula).

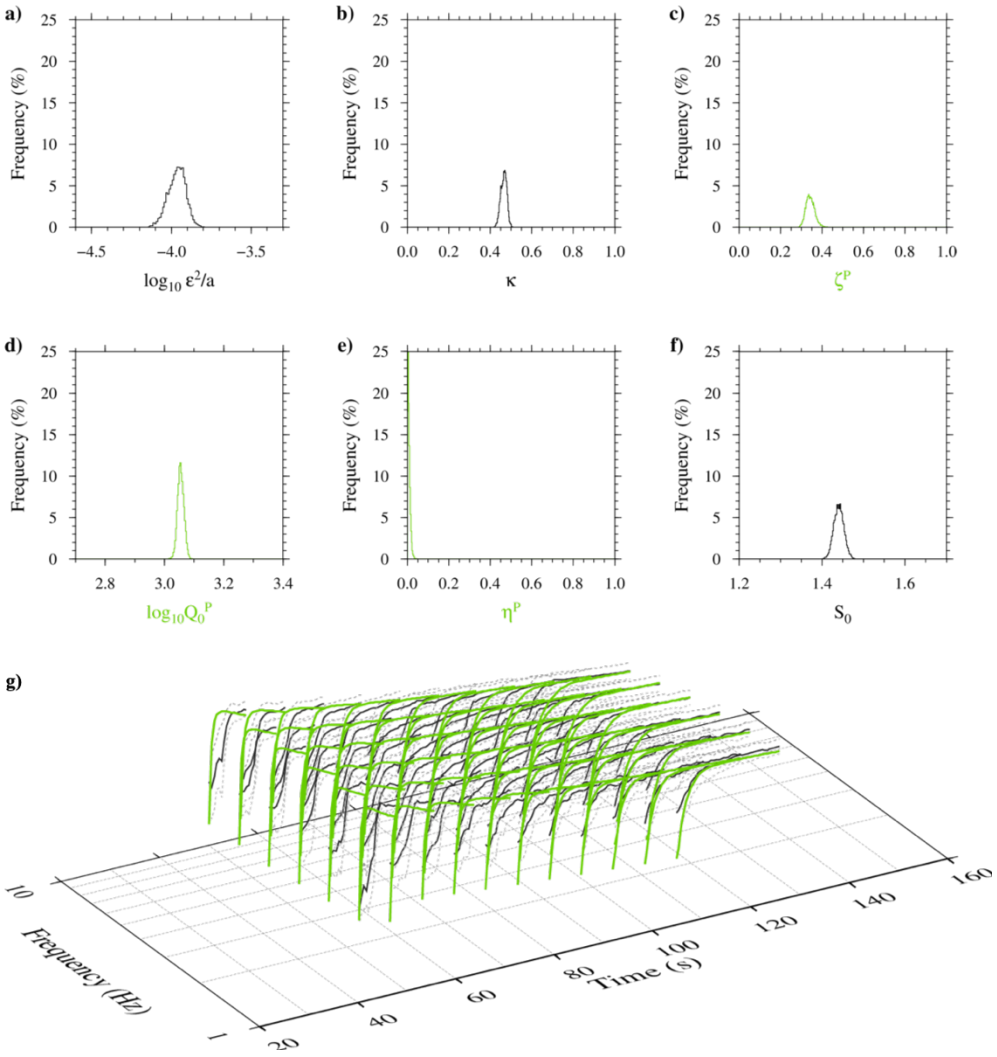

Figure 44. Inversion result for the homogeneous random media, a) $\log_{10}\varepsilon^2/a$, b) the Hust number κ, c) tuning parameters for spectrum division (Sato, 2016; Sato & Emoto, 2017, 2018), d) and e) parameters of frequency-dependent attenuation for P_n (green), and f) constant for source level adjustment. *Black and green lines in g) are the averaged and predicted envelopes for P_n, respectively.*

4.9. 2-D Amplitude Calibration for Full-waveform Envelope Template Matching

Using the derived template model parameters, we processed the template matching for the entire data set and measured the template amplitudes. Then we used the amplitudes to calibrate laterally varying path effects, site amplification, and the source-to-template transfer function with low- and high-frequency source parameters constraints using independent regional moment tensors (Rhie & Kim, 2010; J. Rhie, *pers. comm.*) and GT source spectra (Yoo et al., 2010, 2011), respectively.

4.9.1. Accounting for Noise and Defining Template Amplitudes

As the monitoring target extends to smaller recorded events, we will have to use lower SNR records more often. To lower our template matching analysis to the lower SNR records and not to overestimate source energy by noise, we need to account for the noise in the template matching. Pasyanos et al. (2012) introduced the noise term in their multiple amplitude envelope method. In their method, the waveform envelope was described as the sum of the noise and all of the regional phase envelope amplitudes:

$$A_{total} = A_{Pn} + A_{Pg} + A_{Sn} + A_{Lg} + A_{noise}$$

eq. 50

According to our tests using eq. 50, their method seems to seriously overestimate the noise as the SNR decreases. Mathematically, their formula is comparable to assuming that the signal and noise are perfectly correlated, which is very unrealistic. Assuming signal and noise are independent random variables that are normally distributed, we modified eq 50,

$$A_{total} = \left(A_{Pn}^2 + A_{Pg}^2 + A_{Sn}^2 + A_{Lg}^2 + A_{noise}^2\right)^{1/2}$$

eq. 51

For even very small magnitude earthquakes at distances that were too noisy for the previous study, the hybrid template still provided a stable amplitude measurement by accounting for the noise. In Figure 45, for example, SNR of the M_L 2.3 earthquake at 274 km distance is less than 0.5, which would usually be rejected for a measurement in the previous coda method. However, the new approach with the hybrid template model still shows a reasonable fit.

We estimate the template amplitude from the matched noise-corrected template at a specific point in time. For the current study, we defined this time for P- and S-template amplitudes as:

$$t_{amp}^{P} = \sqrt{3} \times t_{P_g}$$
$$t_{amp}^{S} = d/0.75v_{L_g},$$

eq. 52

in which t_{P_g} is arrival time of P_g, d is the distance in kilometers and v_{L_g} is velocity of observed L_g onset. In Figures 45 and 46, cross symbols represent the noise corrected template amplitudes. We note that we fit the P-template only between the P_g onset to the theoretical S_n arrival to prevent a possible contamination by S phases (Figure 46).

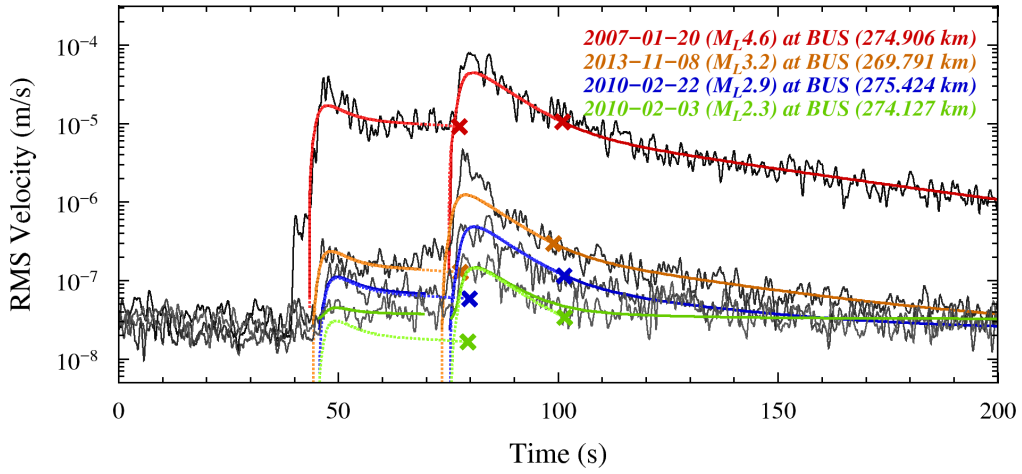

Figure 45. Example of the hybrid template matching for earthquakes recorded at regional distances. *Gray-scaled, color-scaled and dotted lines are data and ensemble average of model prediction and noise-corrected synthetic, respectively. Crosses show the P- and S- template amplitudes. SNR of the M_L 2.3 earthquake at 274 km distance is less than 1, which would usually be rejected for a measurement in the previous coda method. However, the new approach with the hybrid template model still shows a reasonable fit.*

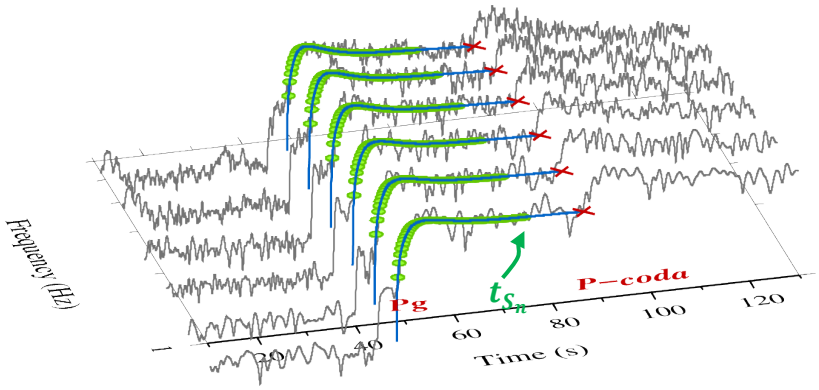

Figure 46. Example of the P-template matching for multi frequency bands. *Observed full-waveform envelopes (thin gray lines) were matched with the new synthetic templates (green circles) for an earthquake at regional distance. Blue lines and red crosses represent the noise-corrected synthetics and template amplitudes, respectively.*

4.9.2. Full-Template Calibration for Northeast Asia

We processed the template matching for entire data set and measured the template amplitudes and used the amplitudes to calibrate laterally varying path effects, site amplification, and the source-to-template transfer function. A total of 49 events were used to constrain the source parameters (Kim and Rhie, 2010; J. Rhie, *pers. comm.*; Yoo et al., 2010 & 2011). After the iterative MADn filtering inversion (Phillips & Stead, 2008), a total of 386,352 template amplitudes for 9 narrowband frequencies were kept from 40,511 unique event-station paths from 3,145 events recorded at 698 stations.

Figure 47 shows the amplitude calibration models and expected signal-to-noise ratio (SNR) of the 2006 DPRK nuclear test from the models. The effective attenuation models for both P- and S-template amplitudes correlate well with surface topography and known geological features (e.g., high attenuation in the East Sea, Bohai Bay and Songliao Basin). We also found that the attenuation derived using crustal earthquakes predicts the amplitudes very well of near-surface explosions. We are also planning to include new seismic data recorded at the seismic stations near the North Korean and Chinese border through academic collaboration with Korean seismologists in the near future.

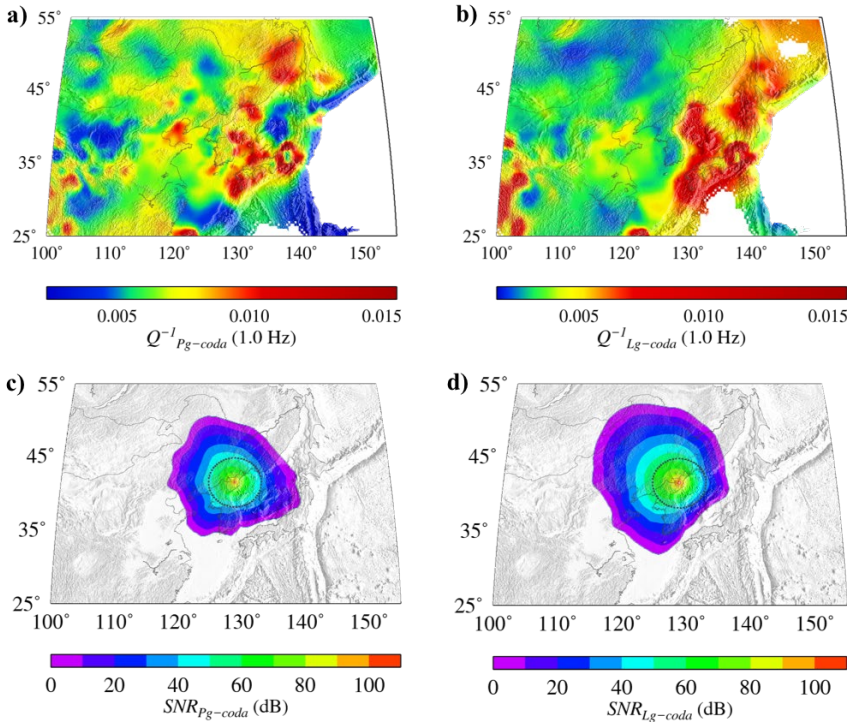

Figure 47. Map shows attenuation of a) P-and b) S- template amplitudes from 2-D tomography inversion (*e.g.*, Phillips and Stead, 2008) and predicted SNR for the 2006 DPRK nuclear test of c) P- and d) S- template amplitudes at 1 Hz. *For the prediction map, we assume 0.78 kt at 113 m depth for the nuclear explosion and average noise and site amplification. We note that the SNR of the direct phase might be higher than the prediction because the template amplitude has decayed some amount of times from the direct phase.*

During the 4 years of this contract, we have continually updated and refined the envelope modeling parameters and tomographic calibrations to more accurately predict the propagation of full-waveform envelopes with additional seismic records and newly available source constraint information. We have been working to calibrate the East Sea region with the possibility of including the records of seismic stations located in Japan in the analysis of the North Korean nuclear tests. However, despite these efforts, the crustal paths through the East Sea region caused a large correction error compared to the paths through the continent regions (Korean Peninsula and China). This is likely due to our simple 2-D attenuation model, which ignores the thickness changes of the crust and probably does not explain the amplitude changes due to the sudden change

in crustal thickness between the East Sea and the surrounding crust. To fix this problem in the new tomography inversion, we excluded the P_g and L_g amplitudes from the paths extending more than 150 km in the dashed area in Figure 12a. As a result, our new model has significantly reduced model roughness for the East Sea region without corrupting the data fit.

Figure 48 shows the revised 2-D attenuation models at 1 Hz for the four regional phases. In the following section, we will explore the non-uniqueness of our calibration model by applying null-space shuttle (Deal & Nolet, 1996; De Wit et al., 2012) to obtain a more quantitative measure of the uncertainty in the models.

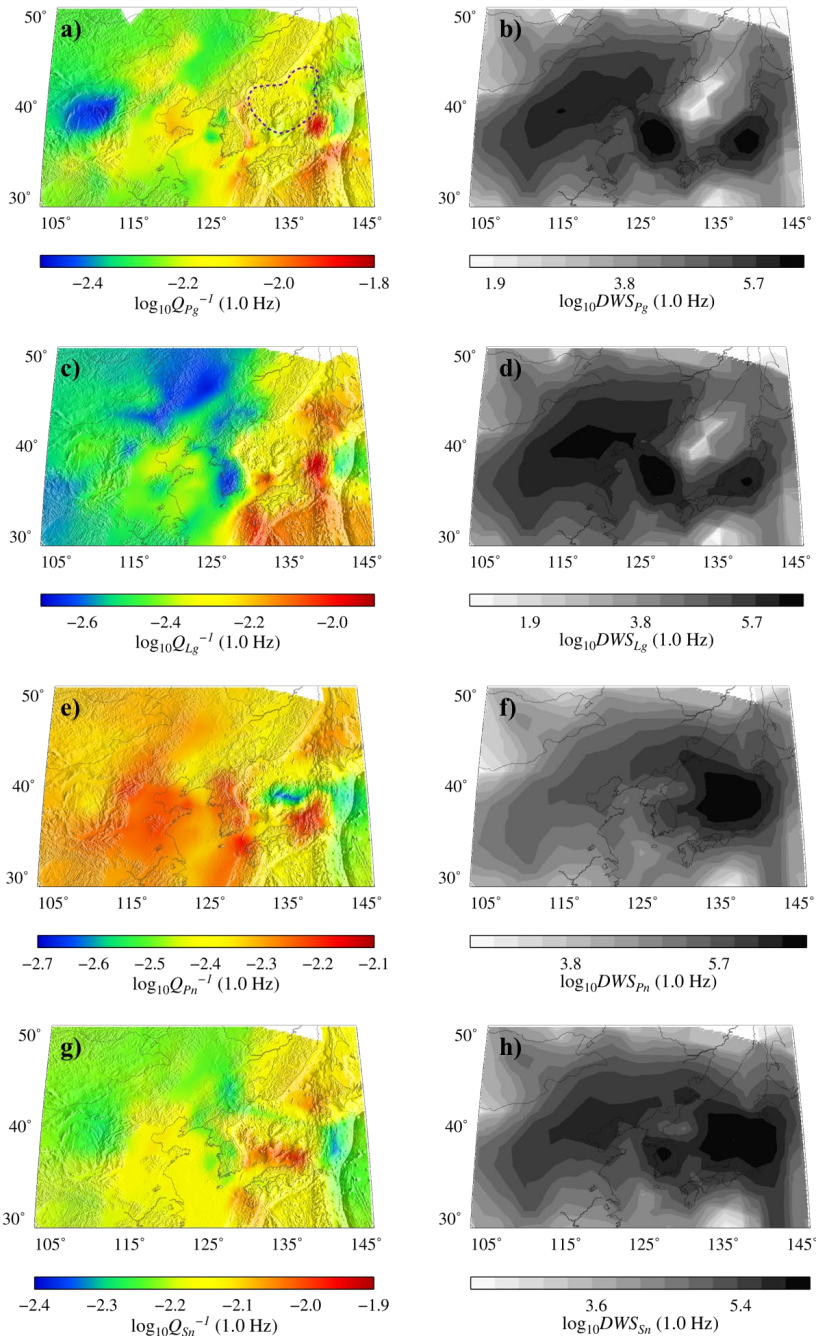

Figure 48. Revised 2-D phase attenuation models for a) P_n, c) P_g, e) S_n and g) L_g waveform template amplitudes at 1 Hz. b), d), f), and h) are corresponding Derivative Weighted Sum (DWS) values which total the weighted ray length in each cell.

4.10. Uncertainty of Attenuation Model

We use a tomography method to derive the 2-D attenuation model. However, seismic tomography is generally an underdetermined problem, and a solution is fundamentally non-unique. Mathematical regularization and smoothness constraints often can lead to misrepresenting model

uncertainty. Therefore, it can be a very challenging study to fully characterize the uncertainty of the 2-D attenuation model obtained using a least square tomography technique.

We explore the non-uniqueness of the tomographic calibration model using a method called 'null-space shuttle' designed by Deal and Nolet (1996) to assess robustness in the tomography solution by examining the null-space of the forward operator. We follow the procedures as described by De Wit et al. (2012), who apply the method to travel time tomography.

An optimum model m_d from inversion of data d can be obtained, that is

$$\boldsymbol{Ld = m_d}$$

eq. 53

using some general inverse operator L. We consider a test model m_t that may be close to m_d but which does not necessarily explain the data. In general, m_t has components both in the range (m_t^{range}) and in the null-space of a forward operator G (m_t^{null}),

$$m_t = m_t^{range} + m_t^{null}$$

eq. 54

We can approximately estimate the model component in the range using inversion of a synthetic data corresponding to the test model ($Gm_t = d_t$),

$$\widetilde{m}_t^{range} = Ld_t$$

eq. 55

Consequently, we can obtain the approximated null-space component of m_t,

$$\widetilde{m}_t^{null} = m_t - \widetilde{m}_t^{range}$$

eq. 56

Using the approximation of the null-space component of m_t, we define a new solution

$$\widetilde{m}_{new} = m_d + \alpha \widetilde{m}_t^{null}$$

eq. 57

with a scaling factor α. Since $G\widetilde{m}_t^{null}$ is not exactly zero, the new solution corresponds to a slightly different data misfit than the original solution. By using the null-space component of the final model, we implicitly asses its robustness and infer a range of acceptable models that explain the data within a chosen data misfit tolerance.

The blue line in Figure 49 shows the data misfit versus model roughness for new solutions by using the optimum model m_d as a test model m_t. The stronger and weaker regularization cases are shown in red and green lines, respectively. By changing the scaling factor α, we can minimize

the model roughness or improve the data misfit. A broad range of solutions exist that explain the data and the best data misfit model is achieved for $\alpha = 1.5$. We allow 1% tolerance of the minimum data misfit and find two extreme models bounded by $\alpha = -1$ and $\alpha = 4$.

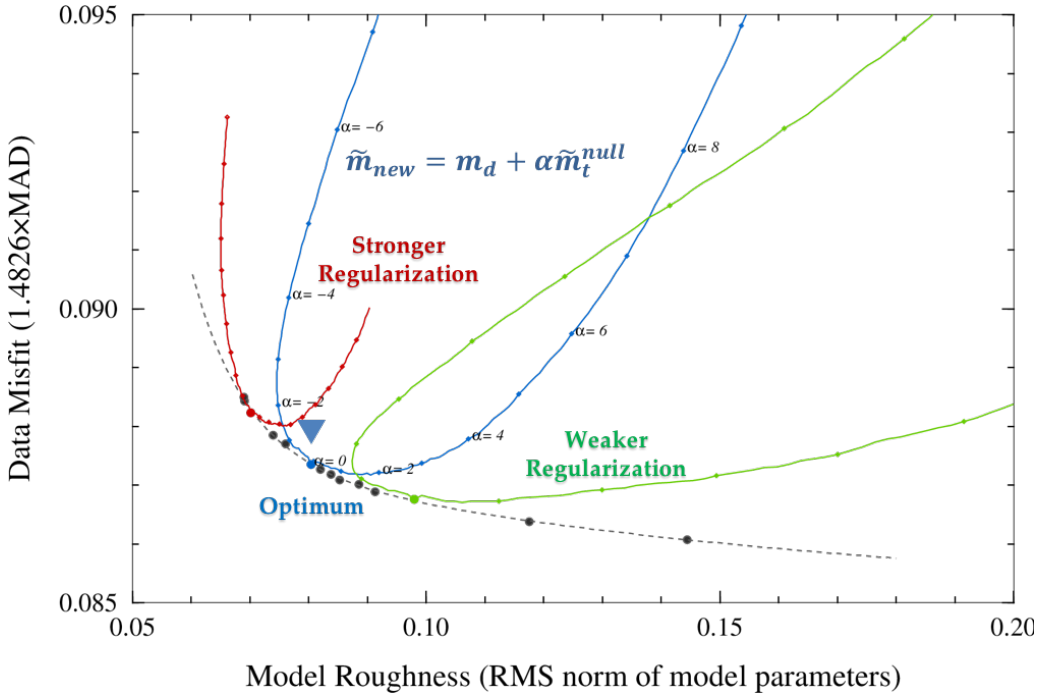

Figure 49. Data misfit versus model roughness of the attenuation parameters in the tomographic solution. *Blue circle denotes the least-square solution given regularization. The blue, red, and green lines represent the null-space shuttle for optimal, stronger, and weaker regularizations, respectively.*

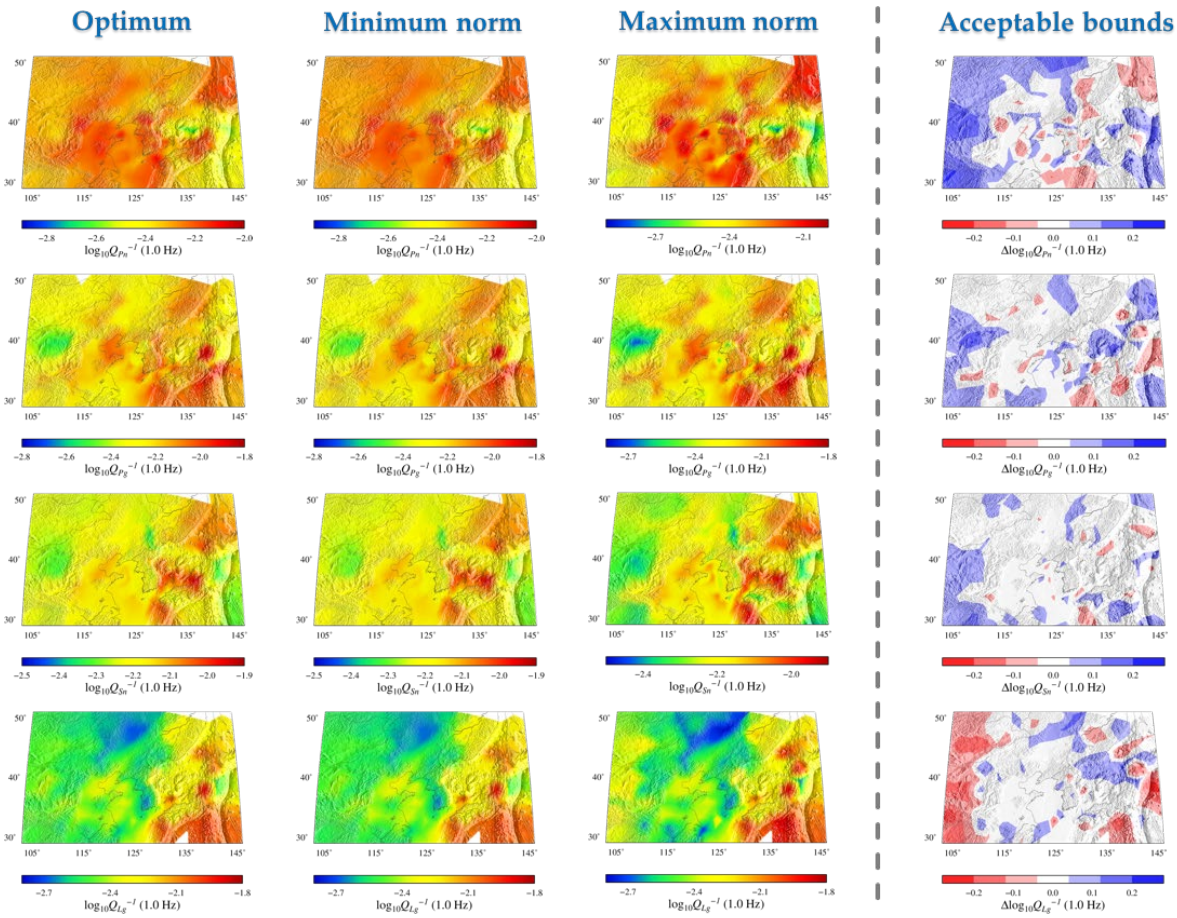

Figure 50. (left) Our final solution, and two bounds at α=-1 and $\alpha = 4$ of acceptable solutions given a tolerance on the data misfit. (right) Difference between the extremes of the range of acceptable solutions.

Figure 50 shows the optimum model, the two extremes of the range of acceptable models given tolerance on data misfit and their difference. As we already find from the path coverage map qualitatively, we can see the highest amplitude of anomalies at the edge of the model which indicates the model has a large variance. However, we still can see a high amplitude anomaly in the East Sea region. This indicates that our model is not robust in that region.

The model difference is sensitive to the chosen data misfit tolerance. Future research is expected to require a variety of tests on the tolerance level and acceptable models bounds. However, it can still be beneficial information in source estimation to adaptively weight the objective function of different propagation paths based on its quantified uncertainty.

4.11. Data Noise and Model Uncertainty in Bayesian Yield Estimation

4.11.1. Data Noise and Model Uncertainty

We use a Laplacian likelihood (L_1 norm) function in our application which is more robust to outliers,

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\prod_{i=1}^{N}(2\boldsymbol{\sigma}_i)} \times \exp\left\{\sum_{i=1}^{N} -\frac{|\mathbf{g}(\mathbf{m}) - \mathbf{d}|}{\boldsymbol{\sigma}_i}\right\}$$

eq. 58

In seismic source estimation processes, we measure the amplitudes and correct for the propagation and site effects, then find optimal source parameters (**m**) that fit the corrected amplitudes (**d**) for given source model (**g**). Uncertainty from measurement and corrections ($\boldsymbol{\sigma}_i$) adaptively weights objective functions and scales uncertainty of the source parameters. As an ideal case, assuming that we can realistically estimate the uncertainty of all underlying processes, we can estimate a reasonable uncertainty for the source parameters. However, there is always a possibility that the uncertainty is underestimated or overestimated if the error is not simply propagated through the multiple steps of data processing or if the correlation between each error is ignored. Figure 51 shows posterior distributions of yield and depth of the 2017 North Korean test properly scaled by estimated amplitude and correction errors (green) comparing to cases with unrealistic error estimates (purple and magenta).

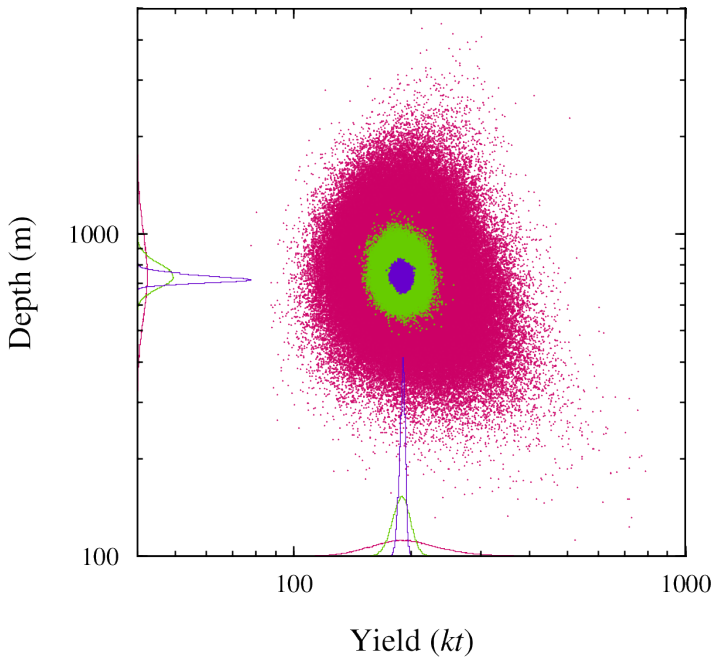

Figure 51. Data noise (e.g., measurement and corrections) and uncertainty in the source parameters. *Green, purple, and magenta color represent posterior distributions of source parameters for proper, underestimated, and overestimated estimates of measurement and correction errors.*

4.11.2. A Hierarchical Uncertainty Model

We applied the null-space shuttle (Deal & Nolet, 1996) to quantify acceptable model bounds of model uncertainty in the 2-D correction models for both path- and site-effects. However, the acceptable bound is still a relative uncertainty from the calibration models because it depends on the given threshold. We use a hierarchical Bayesian formula (e.g., Bodin et al., 2012) to combine the absolute uncertainty in amplitude measurements (σ_i^M) and the relative uncertainty in the calibration models ($\Delta\sigma_i^{T+P}$) with a scale parameter ($\boldsymbol{\lambda}$),

$$\boldsymbol{\sigma}_i = \sigma_i^M + \boldsymbol{\lambda}\Delta\sigma_i^{T+P}$$

eq. 59

In a hierarchical Bayesian algorithm, the $\boldsymbol{\lambda}$ is treated as an unknown using a hyper-parameter and automatically adjusted as a function of the required data fit. The newly designed uncertainty model is incorporated into a Laplacian likelihood function:

$$p(\mathbf{d}|\mathbf{m},\lambda) = \frac{1}{\prod_{i=1}^{N}(2\boldsymbol{\sigma}_i)} \times \exp\left\{\sum_{i=1}^{N} -\frac{\left|\log_{10} A_{syn}(\mathbf{m}) - \log_{10} A\right|}{\boldsymbol{\sigma}_i}\right\}$$

eq. 60

After the path and site-amplification corrections using the tomographic calibration models, we can predict theoretical coda amplitude (A_{syn}) for given source parameters (*i.e.*, explosive yield and depth of burial; $\mathbf{m}$) and source models (*i.e.*, Mueller and Murphy, 1971; Stevens and Day, 1985; Denny and Johnson, 1991; Walter and Ford, 2018). The posterior distribution of the source parameters ($\mathbf{m}$) given corrected coda amplitudes ($\mathbf{d}$) can be obtained from the hierarchical Bayesian inference:

$$p(\mathbf{m},\lambda|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m},\lambda)p(m|\lambda)p(\lambda)$$

eq. 61

4.12. Full-Waveform Envelope Templates for Event Discrimination and Yield Estimation

A new hybrid envelope modeling method with the Bayesian yield estimation method developed under this contract is a major step. We have developed a new discrimination and yield estimation method known as full-waveform envelope template matching to improve predicted envelope fits over the entire waveform and account for direct-wave and early coda complexity. We accomplished this by including a multiple forward-scattering approximation in the envelope modeling of the early coda. The new hybrid envelope templates are designed to fit local and regional full waveforms and produce low-variance amplitude estimates, which improve yield estimation and discrimination between earthquakes and explosions.

To demonstrate the new technique, we applied our full-waveform envelope template-matching method to the six known North Korean (DPRK) underground nuclear tests. We successfully discriminated the event types and estimated the yield for all six nuclear tests. We also applied the same technique to the 2015 Tianjin explosions in China, and another suspected low-yield

explosion at the DPRK test site on 12 May 2010. Our results show that the new full-waveform envelope template-matching method significantly improves upon longstanding single-scattering coda prediction techniques. More importantly, the new method which is possible using the hybrid envelope modeling method with the Bayesian sampling method allows monitoring seismologists to extend coda-based techniques to lower magnitude thresholds (m_b < 4) and low-yield local explosions.

4.12.1. Application to the North Korean Nuclear Tests

To demonstrate the calibration, we applied our full-waveform envelope template-matching method to the six known DPRK underground nuclear tests.

Figure 52 shows waveform envelopes and matched hybrid synthetics for the 3 September 2017 DPRK nuclear test recorded at four seismic stations: (1) the KMA station in Sehwa, South Korea (SEHB), (2) the KMA station in Seoul, South Korea (SEO2), (3) Mudanjiang in China (MDJ), and (4) Inchon in South Korea (INCN). The new hybrid model fits the crustal Pg and its coda very well for broad range of frequency bands. We measured the template amplitudes from the template matching and corrected for path attenuation, site amplification, and the coda transfer function obtained from the 2D calibration and estimated the source spectra. Then, we used the explosion source model (MM71; Mueller and Murphy, 1971) to estimate yield and depth of burial. In the source inversion code, the yield and depth of burial are estimated through MCMC sampling.

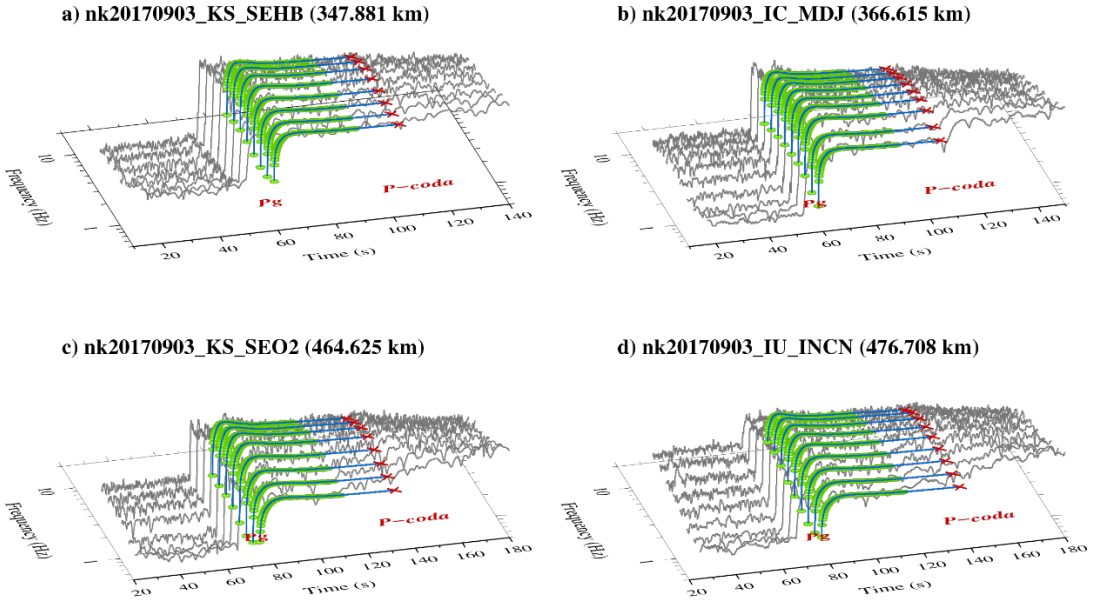

Figure 52. An example of observed full-waveform envelopes (thin gray lines) that match the new synthetic templates (green circles) for 3 September 2017 DPRK nuclear test event. *Blue lines and red crosses represent the noise-corrected synthetics and template amplitudes, respectively.*

Figure 53 shows an example analysis for the 2006 North Korean nuclear explosion using Mueller and Murphy (MM71) explosion source model. Our estimations for the yield and depth are 0.8 *kt* at 129 *m* with 95% credible intervals of 0.71 to 0.85 *kt* and 96 to 154.5 *m*, respectably. Our yield estimate is relatively consistent with the results obtained from other studies which are from 0.7 to

2 *kt*. However, the depth estimate differs from a depth estimated based on elevation of tunnel entrance assuming horizontal-shaft. Our yield estimation for fixed depth at the 310 *m* is 1 *kt* with 95% credible interval of 0.93 to 1.13 *kt*, which is 25 % higher than the estimate from the variable depth inversion.

For comparison with Denny and Johnson (DJ91) and Walter and Ford (WF18) source models, we computed the theoretical source spectrum for the estimated yield and depth from the MM71 source model. For given source parameters of 0.8 *kt* at 129 *m*, DJ91 and WF18 models predict less high frequency energy than the MM71 model and thus, the two models require higher yield and depth to explain the corrected source spectrum from our current calibrations. On average, yield and depth estimates using the WF18 model for 6 North Korean tests are about 2.6 and 2.0 times larger than estimates using MM71, respectively.

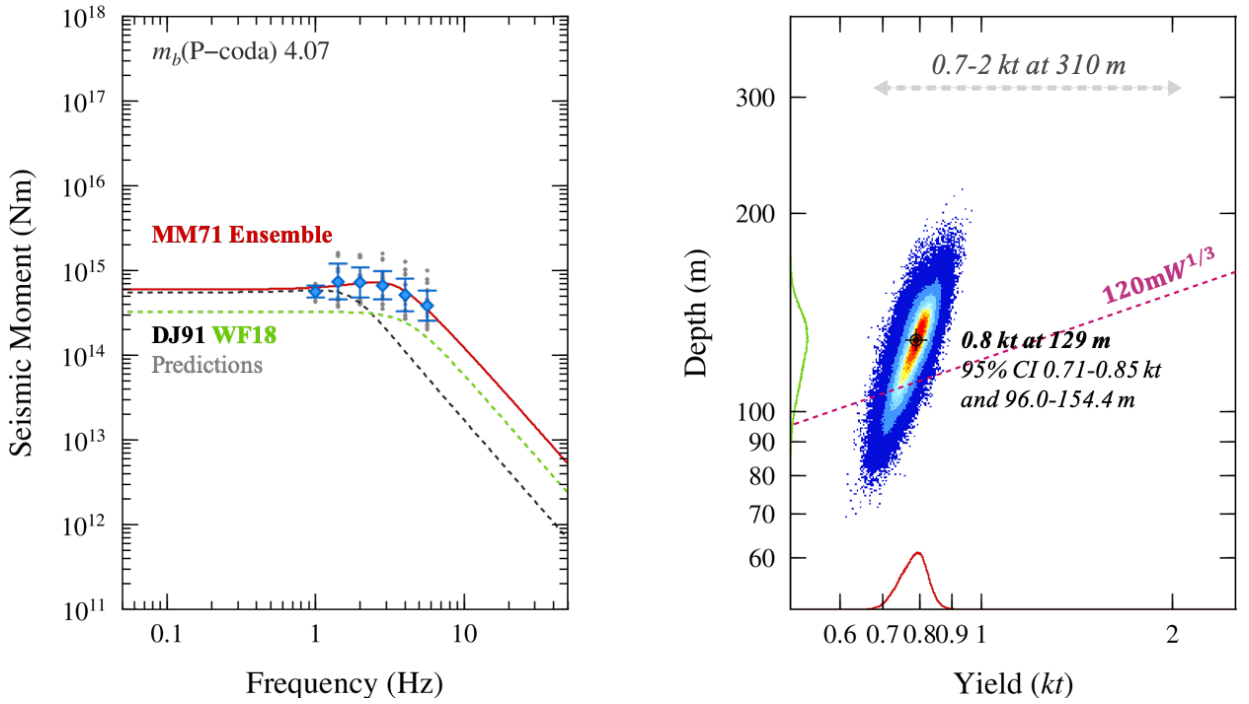

Figure 53. Source spectrum and posterior probability distributions of yield and depth estimates of 9 October 2006 North Korean nuclear explosion using the MM71 explosion source model. *Red, black, and green lines represent ensemble average of the MM71 source spectra, and predicted source spectra using DJ91 and WF18 source models for 0.8 kt at 129 m, respectively.*

Figure 54 shows the posterior distribution of source parameters for 6 North Korean tests using the MM71 explosion source model with our current calibrations for the area. The yield estimates are in good agreement with the estimates of other studies. More importantly, the depth estimates are in good agreement with the estimated depth based on the elevation of Mt. Mantap and portal entrances except the 2006 event, believed to be located at a neighboring mountain. The source spectrum of the 2017 event shows somewhat higher errors than the other five events. The lack of low-frequency measurements using our current frequency bands seems to cause weaker spectral

constraints for the event. Overall, MM71 or SD85 shows better spectral fit than DJ91 and WF18 for the DPRK nuclear explosions.

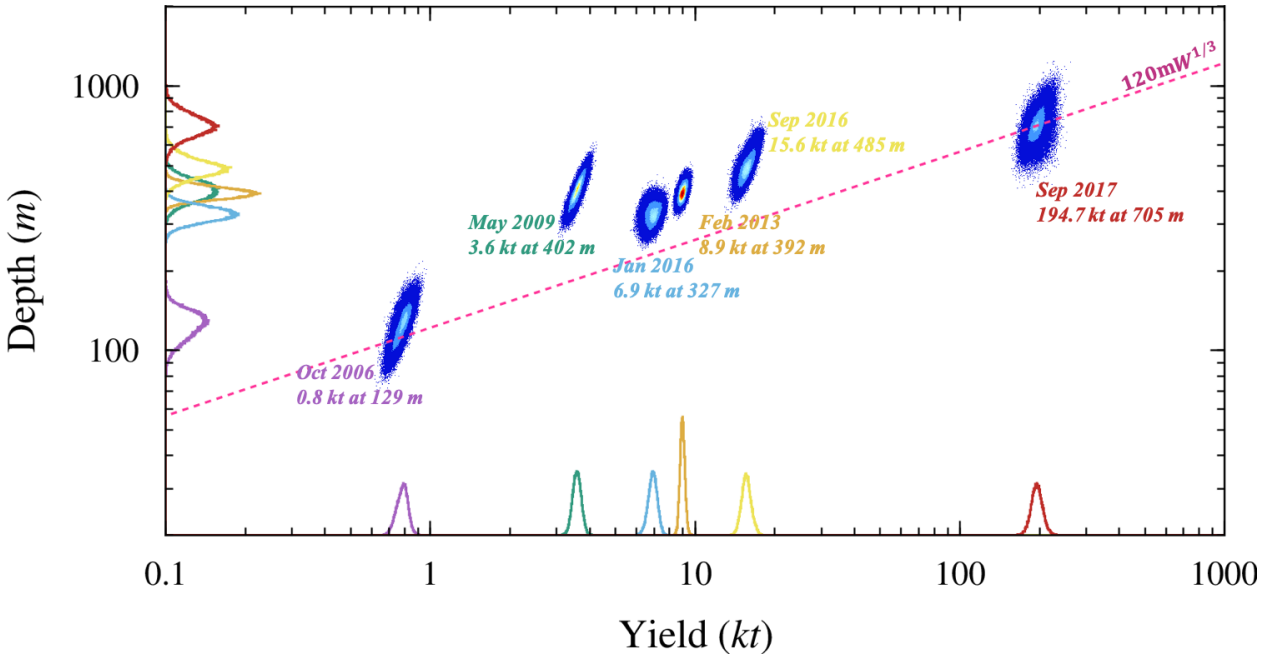

Figure 54. Estimated yield and depth for the six North Korean nuclear explosions using the MM71 explosion source model.

Figure 55 shows posterior distribution for hyperparameters $\boldsymbol{\lambda}$ for three different source models for the six North Korean explosions. For the MM71 model, the estimated scale factor $\boldsymbol{\lambda}$ is about 1.11, which indicates that the path and site corrections have about 2.8 times greater uncertainty than the amplitude measurement error on average. For WF18 and DJ91 models, the uncertainty scale factors were about 16% and 36% larger than that of MM71 (or SD85), respectively. The spectra misfit seems to absorb into the scale factor to compensate for the model error.

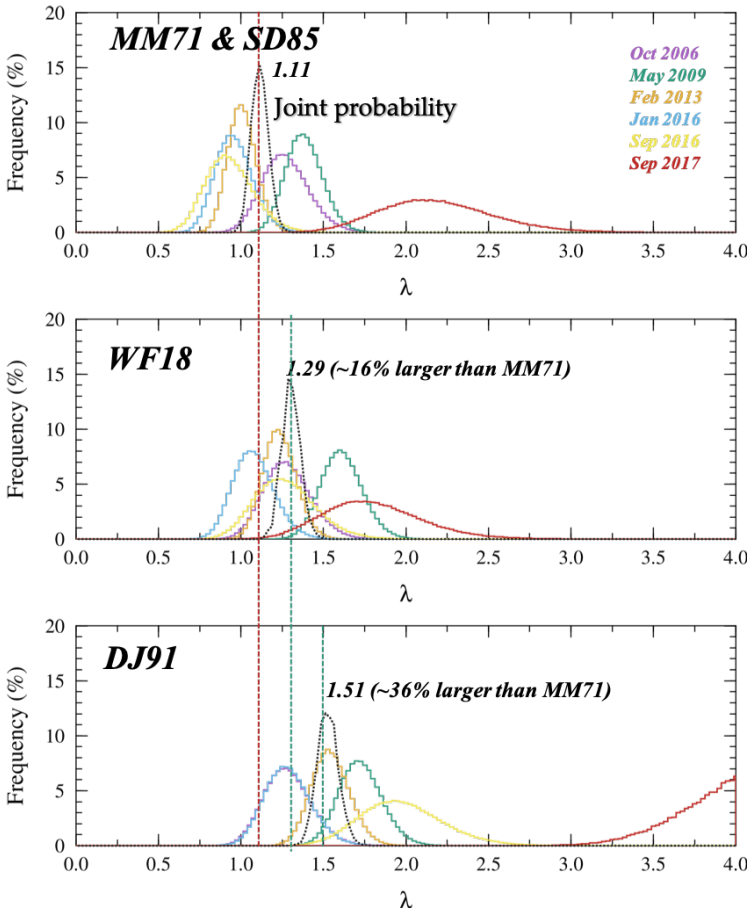

Figure 55. Posterior distribution for hyperparameters $\boldsymbol{\lambda}$ for three different source models. *For each model, colored lines and black dotted line represent posterior distribution from the six North Korean explosions and their joint probability excluding the 2017 test (red).*

Figure 56 shows a comparison between the estimated source spectra of crustal earthquakes and the DPRK nuclear tests. The crustal earthquakes show very comparable source spectra from both P- and S-template matching and we were able to fit the spectra with Brune source model (Brune, 1970). For the nuclear tests, on the other hand, source spectra from the P-template matching are significantly different from the S-template matching source spectra. The explosion source model (MM71; Mueller & Murphy, 1971) provides a better fit to the P-source spectra than the Brune model.

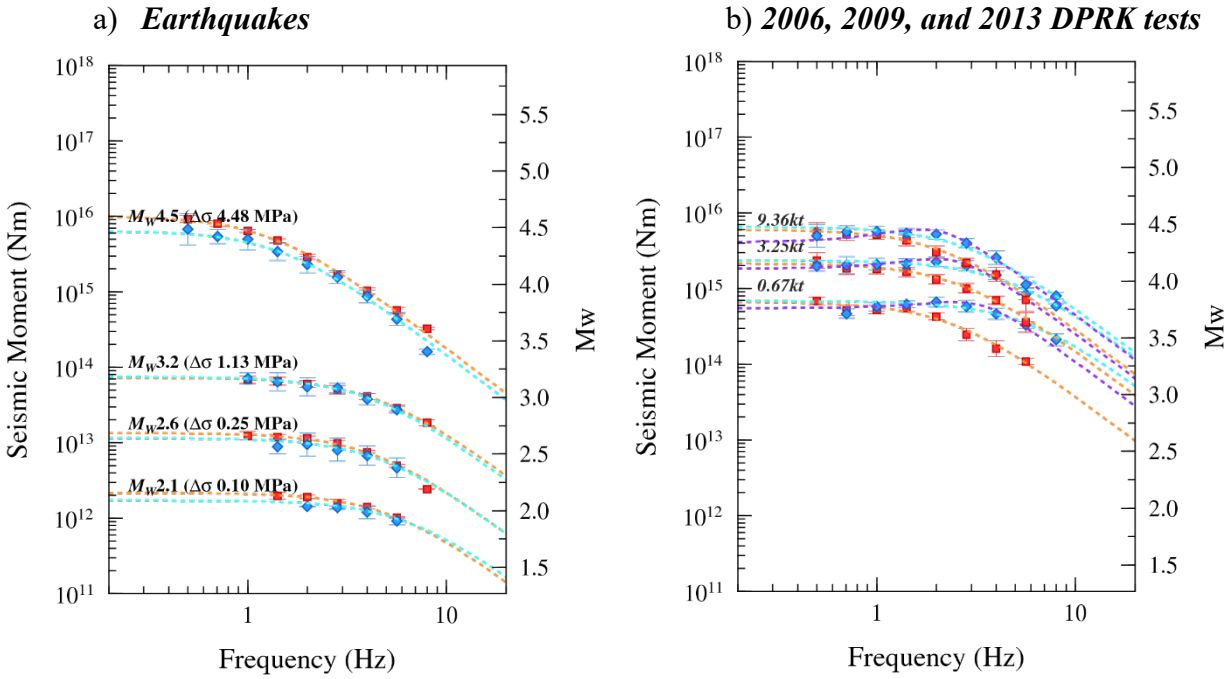

Figure 56. Source spectrum comparison between a) earthquakes, b) 2006, 2009, 2013 DPRK tests. *Blue diamonds and red squares represent estimated source spectra using P- and S- template matching, respectively. Blue and orange dotted lines and purple dotted are fits using Brune source model and MM71, respectively.*

4.12.2. Application to the 2015 Tianjin explosions

We applied our template matching method to two massive chemical explosions in Tianjin, China. On 12 August 2015, a series of explosions killed 173 people and injured hundreds of others at a container storage station at the Port of Tianjin. The first two explosions occurred within 30 seconds of each other at the facility, which is located in the Binhai New Area of Tianjin, China (https://en.wikipedia.org/wiki/2015_Tianjin_explosions). We downloaded the waveforms recorded at the Baijiatuan station (BJT) in Beijing (~174 km) and performed template matching. We note that we couldn't process the S-template matching for the first event at the BJT station because the first explosion's S-coda was contaminated by the second explosion after the first. Figures 57 and 58 show the template matching and estimated source spectra of the explosions.

Similar to the DPRK tests, we found the source spectrum from the P-template matching shows higher spectral levels than that from the S-template matching for the second explosion. However, the scaling of source spectra is significantly different from the DPRK explosions and is more similar to other chemical explosions at the free-surface (e.g., DTRA's Humming Roadrunner series and 2017 Kabul, Afghanistan explosion). We applied a yield formula for a free-surface chemical explosion described in Yoo (2017) and estimated yield of the first and second Tianjin explosions as 90 tons and 960 tons, respectively.

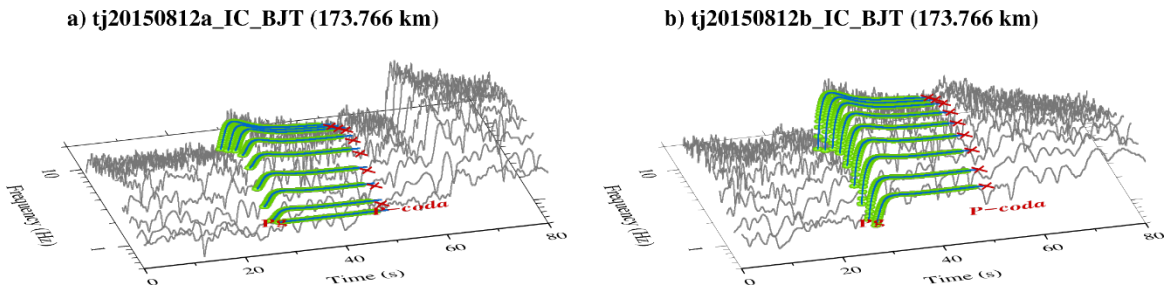

Figure 57. Observed full-waveform envelopes (thin gray lines) that match the new synthetic templates (green circles) for a) the first and b) second explosions during a series of explosions in Tianjin on 12 August 2015. *Blue lines and red crosses represent the noise energy corrected synthetics and template amplitudes, respectively.*

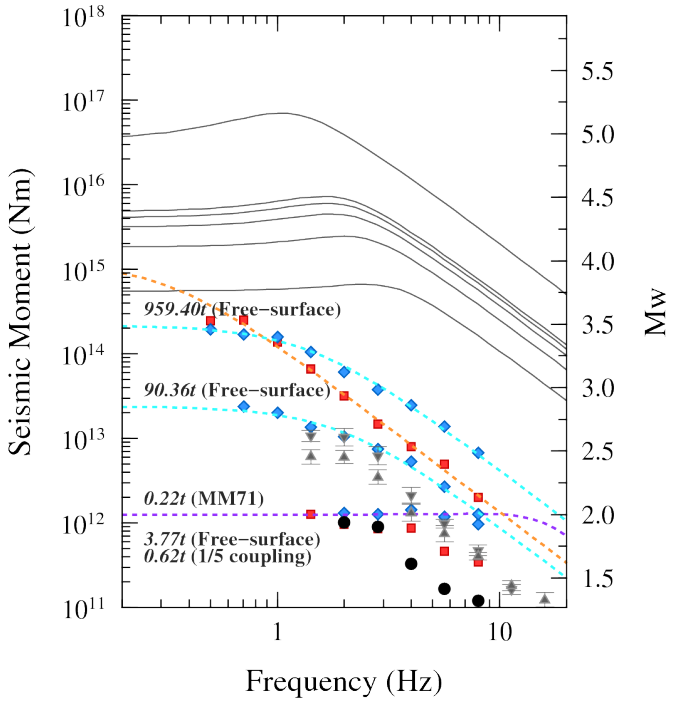

Figure 58. Source spectrum comparison between free-surface chemical explosions and DPRK tests. *Blue diamonds and red squares represent estimated source spectra of the 2015 Tianjin explosions using P- and S- template matching, respectively. And gray and black symbols show source spectra of HRR-1 (18 tons), HRR-5 (45 tons) in DTRA's Humming Roadrunner experiment, and 2017 Kabul bombing event, respectively.*

4.12.3. Application to the suspected low yield event on May 12, 2010

We analyzed the suspected low-yield nuclear event on 12 May 2010 recorded on the Dongbei Broadband Network (DBN; Chun *et al.*, 2011). The data was provided by Professor Paul Richards (Columbia University) and also included the 2006 and 2009 North Korea Nuclear Tests. We note that acquisition and processing of the data are performed under the State Department funded project (DOS Contract No SAQMMA-16-M-2404).

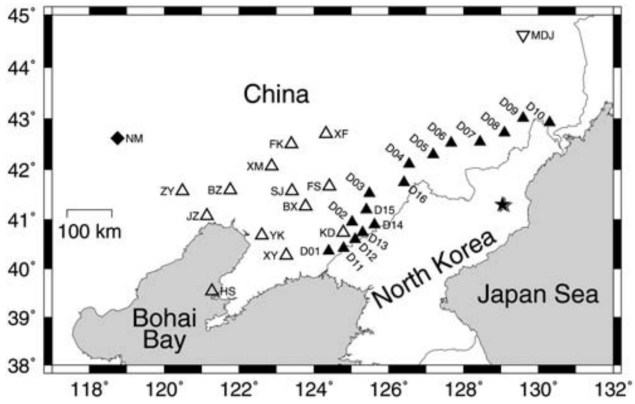

Figure 59. Map showing the North Korean nuclear test site (star) and the seismic stations in the Dongbei Broadband Network (solid triangles) (Chun et al., 2011, copyright by the Seismological Society of America).

The locations of the stations in the network are shown in Figure 59 represented by black triangles. The nearest station to the North Korean nuclear test center is about 1.34 degrees. For now, this is the closest distance we can obtain.

Compared to 2006 and 2009 events, both P and S amplitudes of the 2010 event are very small and most of coda signals are below the noise level (Figure 60). Only direct waves and some portion of early coda signals appear to be available for further analysis.

We applied the measured amplitudes to the preliminarily amplitude attenuation models and reconstructed its source spectrum. Then, we applied the MM71 explosion source model to estimate the yield and depth of burial (Mueller & Murphy, 1971). Due to the very small size compared to other nuclear tests, the corner frequency will be beyond the limit of the calibration band. This lack of high frequency source spectrum made it difficult to predict the yield and depth simultaneously using the explosion model. Therefore, we estimated the yields according to two assumptions listed below to calculate the two depths used in the MCMC inversion (Figure 61).

First, assuming the explosion depth follows the $122m/kt^3$ scaling (Stevens & Day, 1985), the explosion yield is equivalent to about 200kg of TNT at 7m below ground. However, if we assume the depth to 200 m deep, the yield estimate is as high as 700 kg TNT. Even assuming we have similar depths to other DPRK nuclear tests, yield estimates do not appear to be higher than 1 ton.

Our finding is in good agreement with the conclusions reached by Ford and Walter (2015) through analysis of the IMS networks around the North Korean nuclear test site.

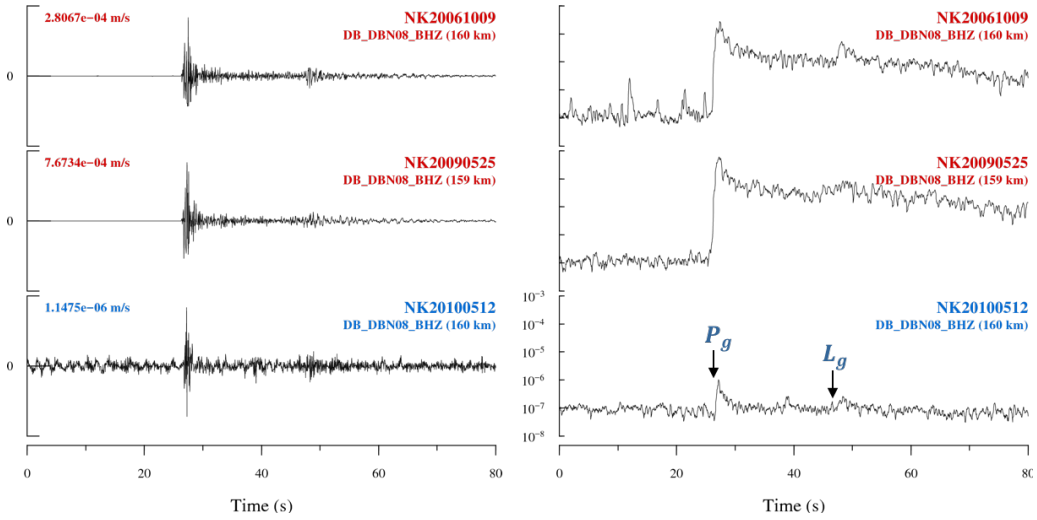

Figure 60. Vertical component waveform (left) and the narrowband envelope (right) of the observed the 2006, 2009, and 2010 events recorded at DBN08.

Our hybrid envelope model not only has superior data representation for such a short coda signal, but also can be applied to data with a very low signal-to-noise ratio by parameterizing the noise in the model. We were able to make template amplitude measurements at four stations in the DBN for the 2010 event.

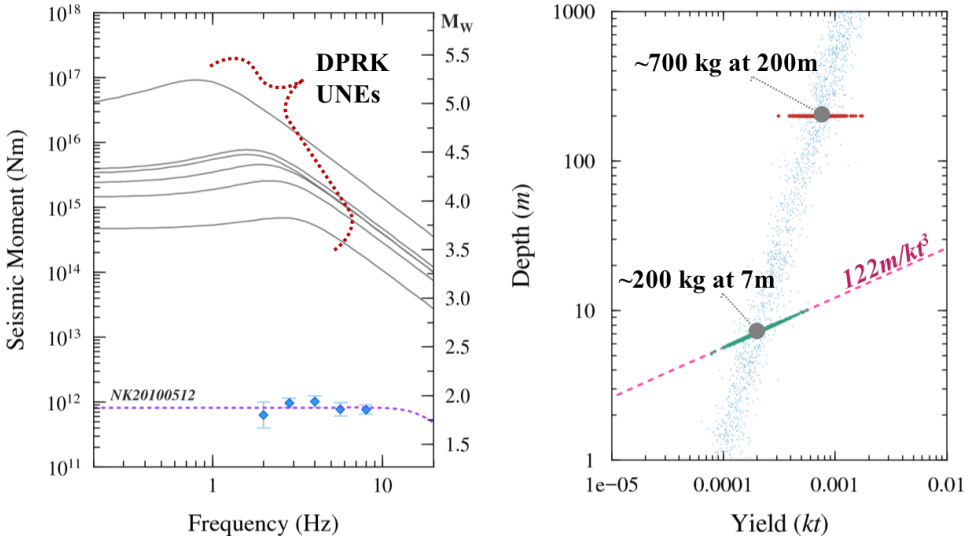

Figure 61. (left) The recovered source spectrum of the 2010 May event (blue diamond) along with the best fit of MM71 source models for the six North Korean nuclear tests (solid line). (right) The posterior distribution of yield and depth estimates from the MCMC inversion (blue dot) using the two assumed depths of 7m and 200m. *Green and red symbols represent the posterior distribution of the estimates assuming the 122m/kt3 scaling and 200m fixed depth, respectively.*

4.12.4. Application to the Low SNR events in the Musan mine, North Korea

We examined the characteristics of the seismic source by applying the newly updated template model to about 200 low SNR seismic events near the Musan mine in North Korea (Figure 62). The mean moment magnitudes of these events estimated using template matching are approximately 2.5 (from 2.2 to 2.8) and the average stress drop is about 0.04 MPa (Figure 63b). The stress drop estimate is significantly lower than 1.8 MPa, which is the average value of the entire crustal earthquake in this area, and 0.3 MPa, which is the average value of similar earthquakes. We suspect these results are due to the fact that these seismic events occurred at very small depths or occurred at the free-surface. Also, we found that the high frequency P wave spectral amplitudes were 2.2 times higher than that of S waves while the mean value of the crustal earthquakes spectral amplitudes is 1 (Figure 63c). Based on these two estimates, we were able to successfully discriminate the seismic activity of the region from the Musan mine events.

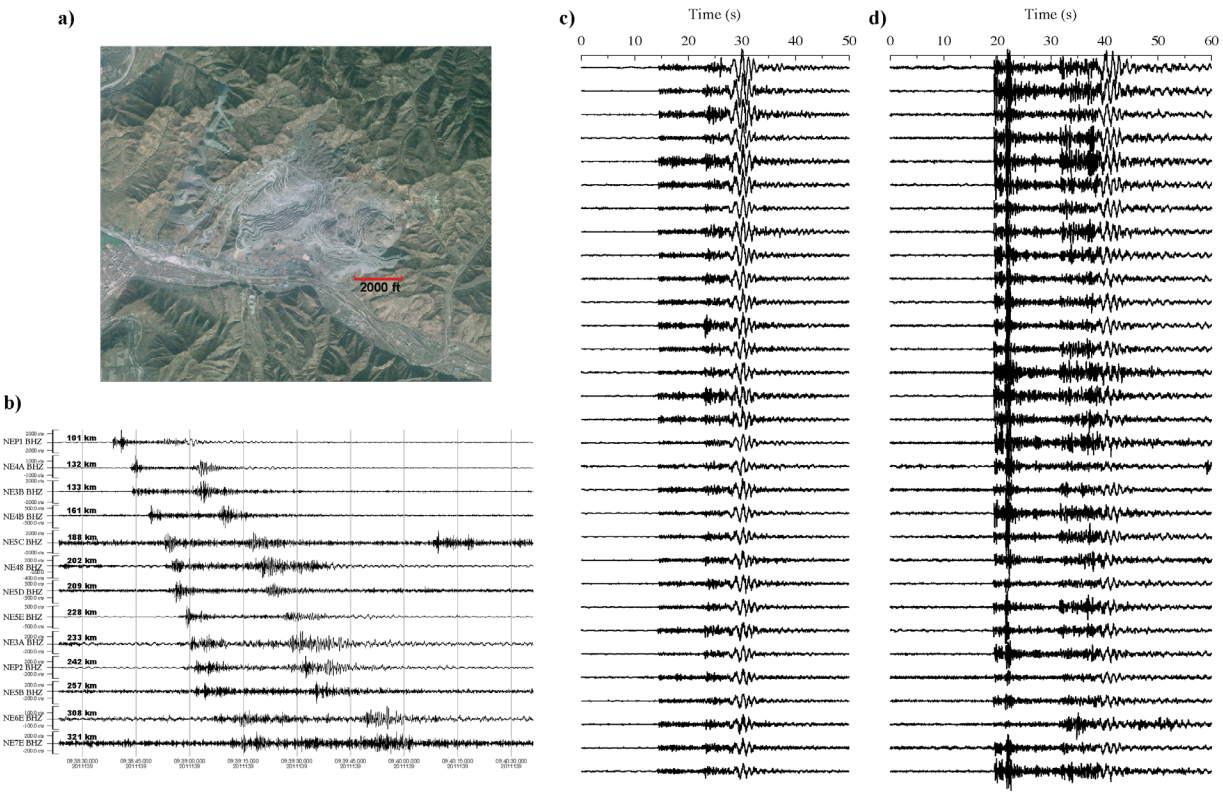

Figure 62. a) Zoomed-in view of the Musan iron mine, visible from Google Earth (42.23º N, 129.26º E), located in the North Hamgyong Province in northeast North Korea. b) A suspected mining explosion near the mine that was well recorded on the YP network. c) and d) Examples of detected events using cross-correlation detection at the NE3C and NEP1 YP stations, located 70 and 100 km away from the mine, respectively.

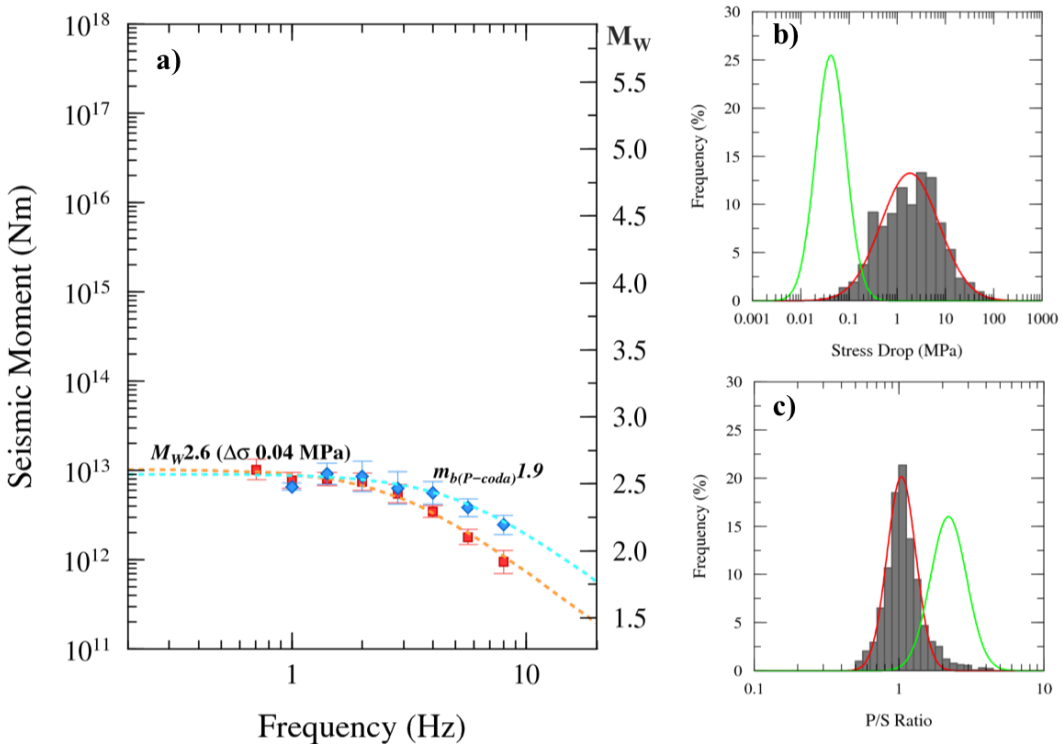

Figure 63. a) Example of estimated source spectrum from P- and S- template matching for one of the low SNR events in the Musan mine. Blue diamonds and red squares represent estimates using P- and S- template matching, respectively. Distribution of estimated stress drop (b) and high frequency P/S ratio (c) show clear separation between crustal earthquakes (red) and the surface mining events in the Musan mine (green).

4.13. Validation Dataset for low SNR events in the North Korean region

To test and validate the calibrated full-envelope templates for low SNR regional events near North Korea, we applied the full-envelope templates processing to new low-SNR events that were not used in the calibration. We found 454 additional small-magnitude seismic events ranging M_L 0.5-3.9 from the NECIS event catalog.

The NECIS origin information for these events has no depth estimate or values after the decimal point for locations and origin times. Therefore, we had to re-determine the origin information for further processing. We first applied a machine-learning-based phase-identification and arrival-picking algorithm, Generalized Phase Detection (GPD), to the database (Ross et al., 2018). Then we applied a Bayesian seismic event location method, Bayesloc (Myers et al., 2007), to relocate the events.

Figure 64a shows an example of the GPD method applied to a three-component seismogram from an M_L 3.9 seismic event that occurred near Songrim in southern Pyongyang, North Korea. This event was recorded at the YC2B station in Yeoncheon, South Korea, about 137 km from the

estimated epicenter. We used probability time series for P- (red) and S-waves (blue) as characteristic functions for the phase detection. Applying the Bayesloc to the detected phase arrivals, we re-determined the origin information of all 454 events in the new database with their associated uncertainties (Figure 64c and Figure 65).

Although the Deep Learning model was trained using only local distance data within 100 kilometers, it showed relatively good performance for the regional data. However, the detection probability tends to be slightly lower than the local data application, so we set the probability threshold lower, which resulted in a higher false-detection rate. We estimated the theoretical arrival time of each regional phase based on the KMA's initial origin information and used it to eliminate these false detections. Re-training our Deep Learning model with regional data sets could improve the detection performance for the regional dataset.

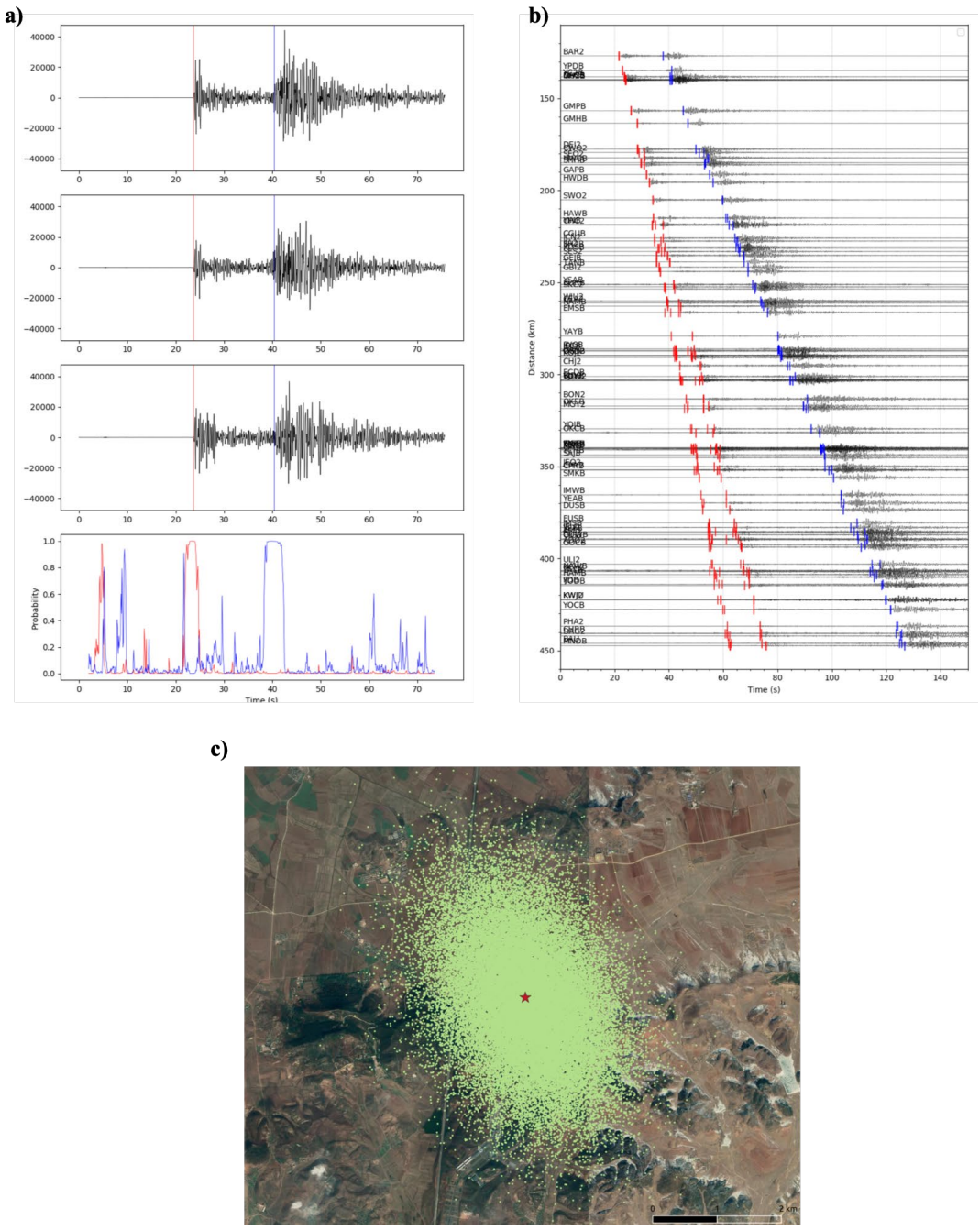

Figure 64. Application example of the Generalized Phase Detection (GRD; Ross et al., 2018) and Bayesloc (Myers et al., 2007) to the M_L 3.9 seismic event in Songrim, southern Pyongyang. (a) and (b) are an example of the GPD applied to a three-component seismogram recorded at YC2B station in Yeoncheon, South Korea and other stations in KMA seismic network, respectively. *Red and blue time series show the probability of detection for P- and S-waves, respectively.* (c) Posterior epicenter distribution of the M_L 3.9 seismic event (EVID 191784168; see Table 2), estimated using the Bayesloc location algorithm.

We found several clustered groups of events from the relocation results and cross-validated their location with some of the surface mines using Google Earth's satellite images, confirming that some of the small-magnitude seismic events were related to mining activity in the region (Figure 65).

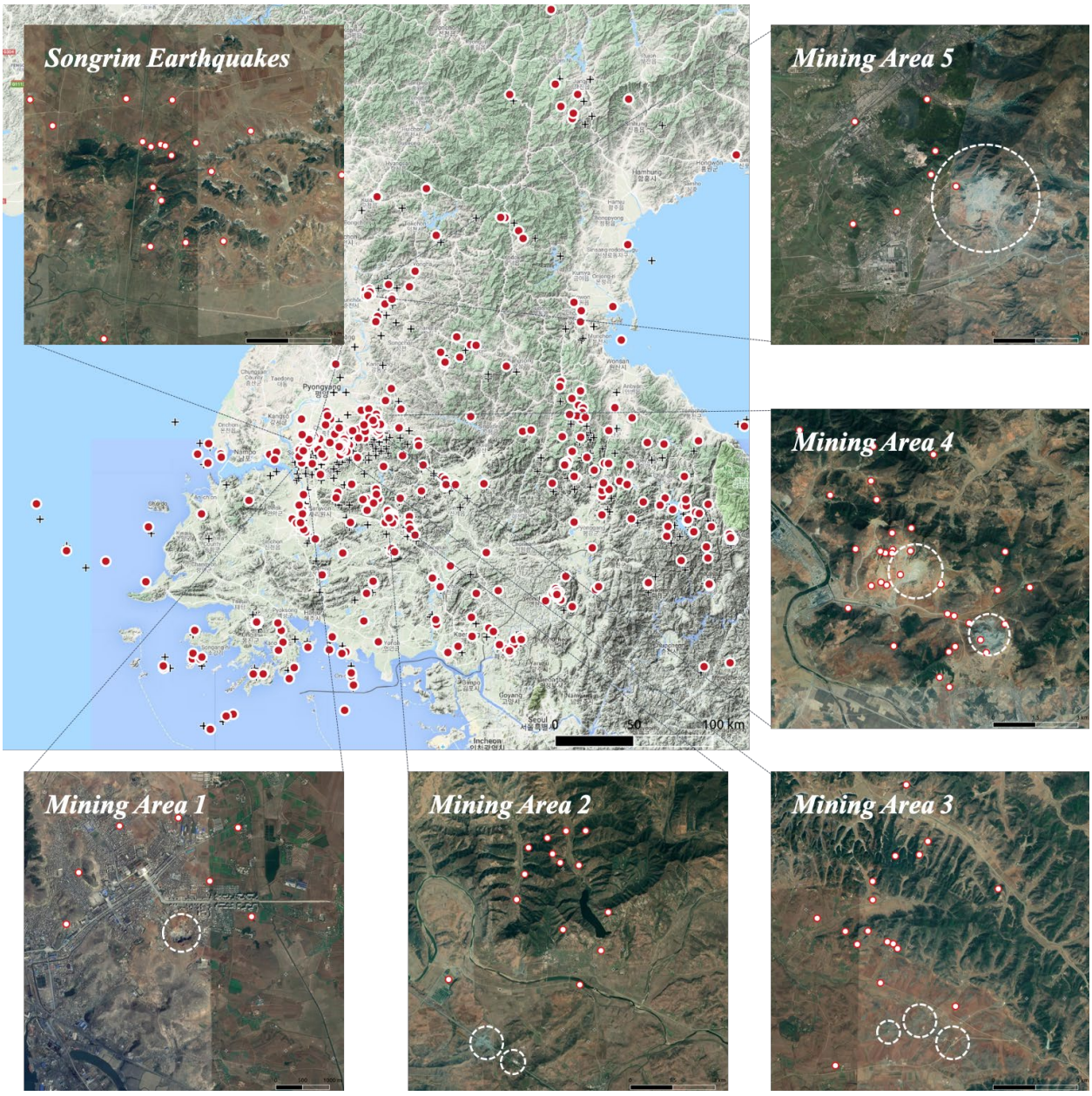

Figure 65. Location map of the 454 events includes in the analysis with zoomed-in satellite images of the Songrim seismic sequence and five verified surface mining areas in the North Korea region. *Crosses and red circles represent the KMA locations and Bayesloc locations made with the GPD picks, respectively. The surface mines are marked as white dotted circles.*

Seismic Activity in Songrim, North Korea

KMA recently reported seismic activity occurring in the Songrim, North Hwanghae Province in North Korea. The sequence started with an M_L 2.5 event on 21 June 2019, followed by an M_L 2.7 and two M_L 3.4 events over the next several days. On 27 June 2019, another M_L 2.7 event occurred and, about six minutes later, an M_L 3.9 event occurred near the same location.

We searched events around the epicenter of the M_L 3.9 event on 27 June 27 2019, using our previous and current event databases and found a total of 15 seismic events (Table 2). We applied the full-waveform template matching to the seismic events in the sequence to estimate moment magnitude and static stress drop. We also compared the P- and S- source spectra to examine whether the event was caused by an artificial explosion. Due to the small size of the events, data recorded at less than 250 km from the estimated epicenters dominated the analysis. Therefore, the P- and S-source spectra were mainly derived from *Pg*- and *Lg*- template amplitudes, respectively.

The estimated moment magnitude of the events ranges from 2.4 to 3.7 (Table 2). The stress drop of the largest earthquake (Mw 3.7) is about 1.3 MPa. The derived source spectra were well described by the Brune source model (Brune, 1970), and the moment magnitude and static stress drop estimated from P- and S- source spectra, respectively (Figure 3), also agreed well. In our preliminary analysis of these events, we could not identify any signs of a man-made explosion.

Table 2. Refined origin information of Songrim seismic sequence, North Korea

Event Id	Date	Time	Lat.	Lon.	Depth (km)	Origin Uncertainty (S.D.)				M_w	σ_S (MPa)
						North (km)	East (km)	Depth (km)	time (s)		
121641196	2012-06-12	00:01:25.384	38.8181	125.6922	17.5	1.3381	0.9455	2.9	0.358	2.51(±0.03)	0.29(±0.18)
143481442	2014-12-14	04:39:40.079	38.8080	125.7120	19.7	1.2685	0.8338	2.4	0.350	2.58(±0.01)	0.19(±0.06)
152251493	2015-08-13	03:36:19.221	38.8352	125.7206	13.2	1.0118	0.7062	2.1	0.309	2.54(±0.01)	0.21(±0.06)
162313575	2016-08-18	13:21:11.506	38.8432	125.6811	22.7	1.0299	0.7547	2.2	0.323	2.70(±0.01)	0.33(±0.04)
162723720	2016-09-28	09:54:22.274	38.8252	125.7081	16.6	1.0935	0.8179	2.3	0.327	2.69(±0.02)	0.09(±0.06)
162723721	2016-09-28	10:00:02.754	38.8429	125.6957	17.2	1.1808	0.8502	2.4	0.346	2.53(±0.02)	0.06(±0.06)
172273886	2017-08-15	12:32:10.474	38.8077	125.7000	10.2	1.1276	0.8276	2.3	0.332	2.62(±0.02)	0.08(±0.05)
172273887	2017-08-15	15:35:23.293	38.8067	125.6888	9.4	1.0531	0.7404	2.2	0.315	2.39(±0.02)	0.07(±0.06)
191724162	2019-06-21	04:20:42.404	38.8319	125.6920	21.5	0.8272	0.6789	1.8	0.290	2.76(±0.01)	0.37(±0.06)
191724163	2019-06-21	20:44:00.072	38.8326	125.6863	24.3	0.7303	0.5768	1.7	0.275	2.81(±0.01)	0.40(±0.04)
191724164	2019-06-21	20:49:36.991	38.8314	125.6889	26.3	0.7073	0.5582	1.6	0.267	3.44(±0.01)	0.82(±0.02)
191734165	2019-06-22	20:26:23.628	38.8323	125.7032	26.7	0.6822	0.5402	1.6	0.269	3.33(±0.01)	0.43(±0.02)
191744166	2019-06-23	04:15:09.172	38.8316	125.6935	19.2	1.0921	0.8355	2.3	0.326	2.49(±0.02)	0.13(±0.07)
191784168	2019-06-27	06:19:52.847	38.8292	125.6955	23.9	0.7240	0.5588	1.6	0.275	3.74(±0.01)	1.29(±0.02)
191784169	2019-06-27	06:37:28.273	38.8213	125.6896	21.8	0.7737	0.6837	1.7	0.284	2.83(±0.01)	0.29(±0.05)

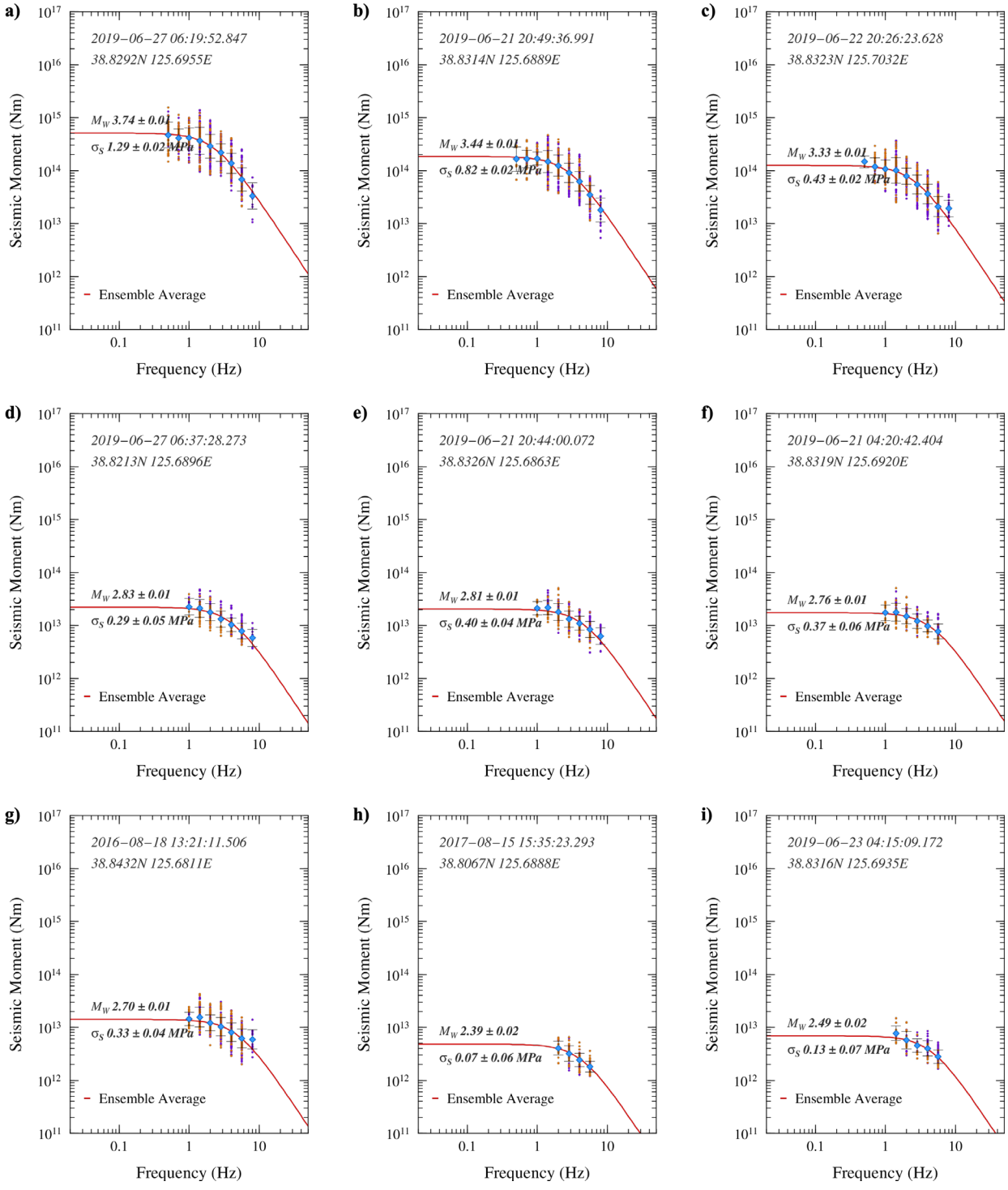

Figure 66. Estimated source spectrum from the full-waveform template matching for the low SNR seismic events in the Songrim sequence. *Purple and orange circles represent the source spectra amplitudes derived from P- and S- template matching, respectively. The red curve shows an ensemble average of an earthquake source model (Brune, 1970).*

5. CONCLUSIONS

We have developed a new hybrid method known as full-waveform envelope template matching to improve predicted envelope fits over the entire waveform and account for direct-wave and early coda complexity. We accomplish this by including a multiple forward-scattering approximation in the envelope modeling of the early coda. The new hybrid envelope templates are designed to fit local and regional full waveforms and produce low-variance amplitude estimates, which will improve yield estimation and discrimination between earthquakes and explosions.

To successfully develop the robust full-waveform envelope template matching technique, we implemented and improved multiple techniques to measure seismic phases, attenuation, and uncertainty. Specifically, to extend capabilities to low SNR events with shorter time windows, we incorporated a full-Bayesian inversion technique based on a MCMC method to explore the template modeling parameters and its uncertainty which control the shape and attenuation of the mantle phases.

To demonstrate the new technique, we applied our full-waveform envelope template-matching method to the six known DPRK underground nuclear tests. We successfully discriminated the event types and estimated the yield for all six nuclear tests. We also applied our new technique to the 2015 Tianjin explosions, the 12 May 2010 low yield North Korean event and a dataset of low SNR events in the Musan mine in North Korea.

Our results show that the new full-waveform envelope template-matching method significantly improves upon longstanding single-scattering coda prediction techniques. More importantly, the new method allows monitoring seismologists to extend coda-based techniques to lower magnitude thresholds and low-yield local explosions.

REFERENCES

Bird, P. (2003), An updated digital model of plate boundaries,,*Geochemistry, Geophysics, Geosystems*, *4*(3). https://doi.org/10.1029/2001GC000252.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P. (2012), Transdimensional tomography with unknown data noise, *Geophysical Journal International*, *189*(3), pp. 1536–1556, https://doi.org/10.1111/j.1365-246X.2012.05414.x.

Brune, J. N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *Journal of Geophysical Research*, *75*(26), pp. 4997–5009, https://doi.org/10.1029/jb075i026p04997.

Carcolé, E. and Sato, H. (2010), Spatial distribution of scattering loss and intrinsic absorption of short-period S waves in the lithosphere of Japan on the basis of the Multiple Lapse Time Window Analysis of Hi-net data, *Geophysical Journal International*, *180*(1), pp. 268–290, https://doi.org/10.1111/j.1365-246X.2009.04394.x.

Chun, K. Y., Wu, Y., and Henderson, G. A. (2011), Magnitude estimation and source discrimination, A close look at the 2006 and 2009 North Korean underground nuclear explosions, *Bulletin of the Seismological Society of America*, *101*(3), pp. 1315–1329, https://doi.org/10.1785/0120100202.

Chung, T. W., Yoshimoto, K., and Yun, S. (2010), The separation of intrinsic and scattering seismic attenuation in South Korea, *Bulletin of the Seismological Society of America*, *100*(6), pp. 3183–3193, https://doi.org/10.1785/0120100054

Cortes, C. and Vapnik, V. (1995), Support-vector networks, *Machine Learning*, *20*(3), pp. 273–297, https://doi.org/10.1007/BF00994018.

Deal, M. M. and Nolet, G. (1996), Nullspace shuttles. *Geophysical Journal International*, *124*(2), pp. 372–380, https://doi.org/10.1111/j.1365-246X.1996.tb07027.x.

Denny, M. D. and Johnson, L. R. (1991), The explosion seismic source function: Models and scaling laws reviewed, In *Explosion Source Phenomenology*, pp. 1–24.

Fehler, M., Sato, H., and Huang, L. J. (2000), Envelope broadening of outgoing waves in 2D random media: A comparison between the Markov approximation and numerical simulations, *Bulletin of the Seismological Society of America*, *90*(4), pp. 914–928, https://doi.org/10.1785/0119990143.

Fehler, Michael, Hoshiba, M., Sato, H., and Obara, K. (1992), Separation of scattering and intrinsic attenuation for the Kanto-Tokai region, Japan, using measurements of S-wave energy versus hypocentral distance, *Geophysical Journal International*, *108*(3), pp. 787–800, https://doi.org/10.1111/j.1365-246X.1992.tb03470.x.

Ford, S. R. and Walter, W. R. (2015), International Monitoring System Correlation Detection at the North Korean Nuclear Test Site at Punggye-ri with Insights from the Source Physics Experiment, *Seismological Research Letters*, *86*(4), pp. 1160–1170, https://doi.org/10.1785/0220150029.

Gibbons, S. J. and Ringdal, F. (2006), The detection of low magnitude seismic events using array-based waveform correlation, *Geophysical Journal International*, *165*(1), pp. 149–166, https://doi.org/10.1111/j.1365-246X.2006.02865.x.

Hastings, W. K. (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, *57*(1), pp. 97–109.

Hoshiba, M. (1994), Simulation of coda wave envelope in depth dependent scattering and absorption structure, *Geophysical Research Letters*, *21*(25), pp. 2853–2856, https://doi.org/10.1029/94GL02718.

Kennett, B. L. N. and Engdahl, E. R. (1991), Traveltimes for global earthquake location and phase identification, *Geophysical Journal International*, *105*(2), 429–465, https://doi.org/10.1111/j.1365-246X.1991.tb06724.x.

Mayeda, K. and Walter, W. R. (1996), Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *Journal of Geophysical Research: Solid Earth*, *101*(B5), pp. 11195–11208, https://doi.org/10.1029/96jb00112.

Mayeda, K., Hofstetter, A., O'Boyle, J. L., and Walter, W. R. (2003), Stable and transportable regional magnitudes based on coda-derived moment-rate spectra, *Bulletin of the Seismological Society of America*, *93*(1), pp. 224–239, https://doi.org/10.1785/0120020020.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), Equation of state calculations by fast computing machines, *The Journal of Chemical Physics*, *21*(6), pp. 1087–1092, https://doi.org/10.1063/1.1699114.

Morse, P. M., Feshbach, H., and Hill, E. L. (1954), Methods of Theoretical Physics, *American Journal of Physics*, *22*(6), pp. 410–413, https://doi.org/10.1119/1.1933765.

Mueller, R. A. and Murphy, J. R. (1971), Seismic Characteristics of Underground Nuclear Detonations Part I. Seismic Spectrum Scaling, *Bulletin of the Seismological Society of America*, *61*(6), pp. 1675–1692.

Myers, S. C., Johannesson, G., and Hanley, W. (2007), A Bayesian hierarchical method for multiple-event seismic location, *Geophysical Journal International*, *171*(3), pp. 1049–1063, https://doi.org/10.1111/j.1365-246X.2007.03555.x.

Paasschens, J. C. J. (1997), Solution of the time-dependent Boltzmann equation, *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *56*(1), pp. 1135–1141, https://doi.org/10.1103/PhysRevE.56.1135.

Pasyanos, M. E., Walter, W. R., and Mayeda, K. M. (2012), Exploiting regional amplitude envelopes: A case study for earthquakes and explosions in the Korean peninsula, *Bulletin of the Seismological Society of America*, *102*(5), pp. 1938–1948, https://doi.org/10.1785/0120120012.

Pezzo, E. Del, Ibañez, J., Prudencio, J., Bianco, F., and Siena, L. De. (2016), Absorption and scattering 2-D volcano images from numerically calculated space-weighting functions, *Geophysical Journal International*, *206*(2), pp. 742–756, https://doi.org/10.1093/gji/ggw171.

Phillips, W. S., and Stead, R. J. (2008), Attenuation of Lg in the western US using the USArray, *Geophysical Research Letters*, *35*(7), https://doi.org/10.1029/2007GL032926.

Phillips, W. S., Hartse, H. E., Randall, G. E., Stead, R. J., Begnaud, M. L., and Anderson, D. N. (2011), Coda Types, *Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, pp. 555–561.

Rhie, J. and Kim, S. (2010), Regional moment tensor determination in the southern Korean Peninsula, *Geosciences Journal*, *14*(4), pp. 329–333, https://doi.org/10.1007/s12303-010-0038-9.

Ross, Z. E., Meier, M. A., Hauksson, E., and Heaton, T. H. (2018), Generalized seismic phase detection with deep learning, *Bulletin of the Seismological Society of America*, *108*(5), pp. 2894–2901, https://doi.org/10.1785/0120180080.

Saito, T., Sato, H., and Ohtake, M. (2002), Correction to "Envelope broadening of spherically outgoing waves in three-dimensional random media having power law spectra" by Tatsuhiko Saito, Haruo Sato, and Masakazu Ohtake, *Journal of Geophysical Research*, *107*(B5), 2089, https://doi.org/10.1029/2002JB002063.

Saito, T., Sato, H., Fehler, M. C., and Ohtake, M. (2003), Simulating the envelope of scalar waves in 2D random media having power-law spectra of velocity fluctuation, *Bulletin of the Seismological Society of America*, *93*(1), pp. 240–252, https://doi.org/10.1785/0120020105.

Sanborn, C. J., Cormier, V. F., and Fitzpatrick, M. (2017), Combined effects of deterministic and statistical structure on high-frequency regional seismograms, *Geophysical Journal International*, *210*(2), pp. 1143–1159, https://doi.org/10.1093/gji/ggx219.

Sato, H. (2016), Envelope broadening and scattering attenuation of a scalar wavelet in random media having power-law spectra, *Geophysical Journal International*, *204*(1), pp. 386–398, https://doi.org/10.1093/gji/ggv442.

Sato, H. and Emoto, K. (2017), Synthesis of a scalar wavelet intensity propagating through von Kármán-type random media: Joint use of the radiative transfer equation with the Born approximation and the Markov approximation, *Geophysical Journal International*, *211*(1), pp. 512–527, https://doi.org/10.1093/GJI/GGX318.

Sato, H. and Emoto, K. (2018), Synthesis of a Scalar Wavelet Intensity Propagating Through von Kármán-type Random Media: Radiative Transfer Theory Using the Born and Phase-Screen Approximations, *Geophysical Journal International,* https://doi.org/10.1093/gji/ggy319.

Sato, H., Fehler, M. M. C., and Saito, T. (2004), Hybrid synthesis of scalar wave envelopes in two-dimensional random media having rich short-wavelength spectra, *Journal of Geophysical Research: Solid Earth*, *109*(6), pp. 1–11, https://doi.org/10.1029/2003JB002673.

Sato, H., Fehler, M. M. C., and Maeda, T. (2012), *Seismic Wave Propagation and Scattering in the Heterogeneous Earth: Second Edition*. *Springer* (Vol. 1), https://doi.org/10.1017/CBO9781107415324.004.

Sens-Schönfelder, C., and Wegler, U. (2006), Radiative transfer theory for estimation of the seismic moment, *Geophysical Journal International*, *167*(3), pp. 1363–1372, https://doi.org/10.1111/j.1365-246X.2006.03139.x.

Shishov, V. L. (1974), Effect of refraction on scintillation characteristics and average pulsars, Sov. Astron., 17, pp. 598–602.

Stevens, J. L. and Day, S. M. (1985), The physical basis of mb:Ms and variable frequency magnitude methods for earthquake/explosion discrimination, *Journal of Geophysical Research*, *90*(B4), pp. 3009–3020, https://doi.org/10.1029/JB090iB04p03009.

Takahashi, T., Sato, H., Nishimura, T., and Obara, K. (2007), Strong inhomogeneity beneath quaternary volcanoes revealed from the peak delay analysis of S-wave seismograms of microearthquakes in northeastern Japan, *Geophysical Journal International*, *168*(1), pp. 90–99, https://doi.org/10.1111/j.1365-246X.2006.03197.x.

Tatarski, V. I., Silverman, R. A., and Chako, N. (1961), Wave Propagation in a Turbulent Medium, *Physics Today*, *14*(12), pp. 46–51, https://doi.org/10.1063/1.3057286.

Walter, W. R. and Ford, S. R. (2018), *A Preliminary Explosion Seismic Spectral Model for Saturated / Hard Rock*, LLNL-TR-754292 940838, Lawrence Livermore National Labaoratory, Livermore, CA.

Wang, W. and Shearer, P. M. (2017), Using direct and coda wave envelopes to resolve the scattering and intrinsic attenuation structure of Southern California, *Journal of Geophysical Research: Solid Earth*, *122*(9), pp. 7236–7251, https://doi.org/10.1002/2016JB013810.

Wegler, U., Korn, M., and Przybilla, J. (2006), Modeling full seismogram envelopes using radiative transfer theory with born scattering coefficients, *Pure and Applied Geophysics*, *163*(2–3), pp. 503–531, https://doi.org/10.1007/s00024-005-0027-5.

De Wit, R. W. L., Trampert, J., and Van Der Hilst, R. D. (2012), Toward quantifying uncertainty in travel time tomography using the null-space shuttle, *Journal of Geophysical Research: Solid Earth*, *117*(3), https://doi.org/10.1029/2011JB008754.

Wong, J., Han, L., Bancroft, J. C., and Stewart, R. R. (2009), Automatic time-picking of first arrivals on noisy microseismic data, *Cseg*, pp. 1–6.

Yoo, S.-H. (2017), Stable and transportable seismic yield estimation from local full-envelope template matching, *Bulletin of the Seismological Society of America*, *107*(2), pp. 660–673, https://doi.org/10.1785/0120150148.

Yoo, S.-H., Rhie, J., Choi, H., and Mayeda, K. (2010), Evidence for non-self-similarity and transitional increment of scaled energy in the 2005 west off Fukuoka seismic sequence, *Journal of Geophysical Research: Solid Earth*, *115*(8), B08308, https://doi.org/10.1029/2009JB007169.

Yoo, S.-H., Rhie, J., Choi, H., and Mayeda, K. (2011), Coda-derived source parameters of earthquakes and their scaling relationships in the Korean peninsula, *Bulletin of the Seismological Society of America*, *101*(5), pp. 2388–2398, https://doi.org/10.1785/0120100318.

Yoshimoto, K., Sato, H., and Ohtake, M. (1997), Three-component seismogram envelope synthesis in randomly inhomogeneous semi-infinite media based on the single scattering approximation, *Physics of the Earth and Planetary Interiors*, *104*(1–3), pp. 37–61, https://doi.org/10.1016/S0031-9201(97)00061-7.

Zeng, Y. (1993), Theory of Scattered P-Wave and S-Wave Energy in a Random Isotropic Scattering Medium, *Bulletin of the Seismological Society of America*, *83*(4), pp. 1264–1276.

Zhang, M. and Wen, L. (2015), Seismological Evidence for a Low-Yield Nuclear Test on 12 May 2010 in North Korea, Seismological Research Letters, 86(1), pp. 138–145. https://doi.org/10.1785/02201401170.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

ACF	Auto-Correlation Function
AFRL	Air Force Research Laboratory
BUS	Busan Station in South Korea
DBN	Dongbei Broadband Network
DJ91	Denny and Johnson Explosion Source Model
DOB	Depth-of-burst
DOS	Department of State
DPRK	Democratic People's Republic of Korea
DTRA	Defense Threat Reduction Agency
FFT	Fast Fourier transform
GT	Ground-Truth
HOB	Height-of-burst
HRR	DTRA's Humming Roadrunner experiment
IDC	International Data Centre
IMS	International Monitoring System
INCN	Inchon Station in South Korea
IRIS	Incorporated Research Institutions for Seismology
ISD	Intensity Spectral Density
KIGAM	Korea Institute of Geoscience and Mineral
KMA	Korea Meteorological Administration
KSA	Ganseong station in South Korea
MAD	Median Absolute Deviation
MCMC	Markov Chain Monte Carlo
MDJ	Mudanjiang station in China
MER	Modified Energy Ratio
MLTWA	Multi-Lapse Time Windows Analysis
MM71	Mueller and Murphy Explosion Source Model
MS	Mean-Square
PSDF	Power Spectral Density Function
RMS	Root-Mean-Square
RT	Radiative Transfer
RTE	Radiative Transfer Equation
RTT	Radiative Transfer Theory
SD85	Stevens and Days Explosion Source Model
SNR	Signal-to-Noise Ratio
STA/LTA	Short-Term/Long-Term Average Ratio
SVM	Support Vector Machine
TFMCF	Two-Frequency Mutual Coherence Function
TNT	Trinitrotoluene
WF18	Walter and Ford Explosion Source Model
Weston	Weston Geophysical Group (Applied Research Associates, Inc.)

DISTRIBUTION LIST

DTIC/OCP
8725 John J. Kingman Rd, Suite 0944
Ft Belvoir, VA 22060-6218 1 cy

AFRL/RVIL
Kirtland AFB, NM 87117-5776 1 cy

Official Record Copy
AFRL/RVB/Dr. Frederick Schult 1 cy