

UC Berkeley

UC Berkeley Previously Published Works

Title

A Nonergodic Ground-Motion Model for the San Francisco Bay Area for Small-Magnitude Earthquakes

Permalink

<https://escholarship.org/uc/item/3hh3271z>

Journal

Bulletin of the Seismological Society of America, 116(3)

ISSN

0037-1106

Authors

Lacour, Maxime
Abrahamson, Norman
Nakata, Rie
[et al.](#)

Publication Date

2026-06-01

DOI

10.1785/0120250094

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A Non-Ergodic Ground-Motion Model for the San Francisco Bay Area for Small-Magnitude Earthquakes

Maxime Lacour ^{*1}, Norman Abrahamson ^{†1}, Rie Nakata ^{‡2, 3}, Nori Nakata ^{§2, 4},
and Camilo I. Pinilla-Ramos^{2, 5, 6}

¹Department of Civil Engineering, University of California, Berkeley

² Energy Geosciences Division, Lawrence Berkeley National Lab, Berkeley, CA
94720, USA

³ Earthquake Research Institute, University of Tokyo, Tokyo 113-0032, Japan

⁴ Department of Earth, Atmospheric and Planetary Sciences, MIT, Cambridge,
MA 02139, USA

⁵University of Southern California, Los Angeles, California

⁶Statewide California Earthquake Center, SCEC

August 20, 2026

^{*}Corresponding author: maxlacour@berkeley.edu

[†]abrahamson@berkeley.edu

[‡]rnakata@lbl.gov

[§]nnakata@lbl.gov

Abstract

Recently, generative models have become a computationally efficient alternative to physics-based numerical simulations of ground motions. Neural networks can learn from existing ground-motion data to generate unobserved ground-motion data at new source and site locations. A key challenge with generative models is ensuring that predicted ground motions remain within a physically realistic range. For that purpose, we developed an empirical, non-ergodic ground-motion model (GMM) for small-magnitude earthquakes in the San Francisco Bay Area based on about 5,000 recordings per component for $M \leq 4$ earthquakes.

The non-ergodic GMM predicts spatially varying median source, site, and path effects for both the Fourier amplitude spectrum (FAS) and the Fourier phase derivative (a proxy for duration), as well as the corresponding epistemic uncertainty for each term. For FAS, our model shows above-average source and site effects in the western part of the region and below-average effects in the eastern part, with regional effects exhibiting larger spatial correlation lengths with increasing frequency. For duration, the source term is negligible for small-magnitude earthquakes, and the site term leads to site-specific variations up to 5 seconds. Path effects for FAS and duration depend on the source-site pair and are extrapolated spatially using recent methods for path-effect modeling. The aleatory variability of the within-site within-path residuals is similar to the variability found in previous studies for other regions.

The non-ergodic model provides two key contributions: first, median adjustment terms that are transferable to larger magnitude earthquakes, further reducing aleatory variability in probabilistic seismic hazard analysis (PSHA); second, region-specific criteria for validating machine-learning-based ground-motion generators to evaluate whether synthetic ground motions exhibit physically realistic source, site, and path effects.

24 Introduction

25 Numerical simulation of ground motions using physics-based methods with a 3-D velocity structure
26 can be computationally expensive (McCallen *et al.*, 2021a,b) which limits the number of different
27 scenarios that can be modeled.

28 A critical challenge for both physics-based simulations and AI-generated ground motions is the
29 validation of the generated ground motions. Traditional validation approaches compare simulated
30 ground motions against ergodic ground-motion prediction equations (GMPEs); however, ergodic
31 GMPEs gather data from multiple regions and cannot capture systematic spatial variations in
32 source, site, and path effects. Non-ergodic models can significantly reduce the aleatory variability
33 in PSHA and provide region-specific benchmarks for validating AI-based simulations. As an
34 alternative, artificial intelligence (AI) methods provide more efficient methods for generating
35 ground motions for a specific source-site combination.

36 To support validation of AI-based simulations and to enable non-ergodic probabilistic seismic
37 hazard analysis (PSHA) applications, we develop empirically based ground-motion models (GMMs)
38 for the velocity Fourier amplitude spectrum (FAS) and the phase derivative for small-magnitude
39 earthquakes in the San Francisco region. Ground motions from small-magnitude events provide a
40 large empirical dataset for the region that can be used to train and validate AI-based simulation
41 methods for source, path, and site effects. The FAS GMM provides constraints on ground-
42 motion amplitude, and the phase-derivative GMM provides constraints on waveform duration.
43 Simulations consistent with both empirical models demonstrate physically realistic behavior for
44 these fundamental characteristics.

45 In this paper, we first develop an ergodic GMM for FAS. Using the ergodic GMM as a reference,
46 a non-ergodic GMM for FAS is developed that models the systematic source, site, and path effects,

including path effects due to the 3-D velocity structure (Sung *et al.*, 2023). The same approach is used to develop ergodic and non-ergodic GMMs for the phase derivative.

The non-ergodic GMM developed in this study provides two significant contributions. First, for PSHA applications, median non-ergodic terms based on small-magnitude earthquakes can be used to estimate median non-ergodic terms for large-magnitude earthquakes for linear site and path effects, further reducing aleatory variability in hazard calculations. The transferability of non-ergodic site and path corrections from small-magnitude earthquakes to large-magnitude earthquakes is based on the principle that ground motions from large earthquakes can be decomposed as the sum of contributions from multiple smaller subevents. The Southern California Earthquake Center Broadband Platform validation studies (Goulet *et al.* (2015)) explicitly validate this subevent summation principle. Their validation exercise employed kinematic rupture generators that synthesize finite-fault ground motions from subfault contributions, successfully reproducing observed strong-motion recordings across magnitudes M 4.6 to M 7.6 in multiple tectonic regimes. The non-ergodic source effects from small-magnitude earthquakes may be correlated to non-ergodic source effects from large-magnitude earthquakes, but this requires further investigation. Second, for AI-based simulation validation, the non-ergodic GMM establishes region-specific acceptance criteria. AI models can be validated against the empirically derived distributions over the region of study to ensure their predictions match observed systematic regional variations, which provides an objective framework for validating synthetic ground motions.

Data Acquisition and Processing

Seismograms from small-magnitude earthquakes in the San Francisco region were downloaded from the NCEDC database (NCEDC, 2014), where the earthquake data is available as velocity

time series. The stations were selected in a rectangular area of 120km x 120km centered on the Hayward fault, and the earthquakes within 100km of the stations were retained. This led to 626,423 3-component velocity time series from 1990 to 2023. Only the two horizontal components are used in this study. The start of the time window was set at 10 seconds before the origin time of the event, and the total length of the time window was set to 70 sec. The Nyquist frequency is 50 Hz for each recording.

The first 10 seconds before the origin time were used to define the noise of each recording, and the remaining 60 seconds were used to define the earthquake signal, assuming that the noise is stationary over 70 seconds. Using the time window before the origin time ensures that the noise window does not overlap with the earthquake signal.

The FAS of the signal and the noise were smoothed using the [Konno and Ohmachi \(1998\)](#) procedure with a smoothing parameter $be\!xp = 20$. This value of $be\!xp$ corresponds to smoothing over a broader frequency window than typically used for input to random vibration procedures ([Kottke et al., 2021](#)). The broader smoothing reduces the troughs in the FAS which results in fewer records falling below the minimum signal-to-noise ratio threshold.

The ratio of the signal and noise amplitude spectra was then calculated for each recording and only the recordings that had a signal-to-noise ratio (S/N) greater than 3 over the entire frequency range of interest were kept. Because the number of recordings above a given signal-to-noise ratio decreases as we broaden the frequency range of study, we chose the frequency range of 2-15 Hz to retain a reasonable amount of data and a broad enough frequency range for applications.

We selected events with magnitude $M \leq 4$ and recordings with rupture distances less than 100 km. A maximum magnitude of 4 was used so that a finite-fault term is not needed in the GMM. A maximum of 100 km was used to limit the spatial coverage for the model to the Bay Area region.

We also only kept events that were recorded by 5 or more stations and stations that had

93 recorded 5 or more events to improve the statistical constraints on the event and site terms in
94 the mixed-effects regression ([Abrahamson and Youngs, 1992](#)). An additional visual inspection of
95 the seismograms was then performed. This procedure identified recordings that included multiple
96 earthquake recordings in the 70-second time window record or did not look like earthquake time
97 series. Approximately 5,000 recordings for each component passed this visual inspection.

98 Using this subset of reliable data, we again rejected the events and stations that had less than
99 five recordings per station and per event. As a last step, we applied a bandpass Butterworth filter
100 with four poles and corner frequencies of 2.5 Hz and 12 Hz to filter the frequencies outside of
101 the range of interest so that the filtered time series could be used in time-domain validations in
102 subsequent studies.

103 The final data set includes just under 1% of the total number of recordings that were downloaded.
104 This low percentage reflects the noise in seismograms from small-magnitude earthquakes. The final
105 number of selected recordings per component, as well as the corresponding number of stations and
106 events, are shown in [Table 1](#). The number of events per magnitude bin for each component is also
107 given in [Table 1](#). The locations of the selected stations and events and the ray paths between the
108 stations and sources for the H1 component are shown in [Figure 1](#). The magnitude-distance scatter
109 plot, as well as the FAS data for two earthquake scenarios, are shown in [Figure 2](#). Additional
110 figures showing the magnitude, distance and depth distributions in [Figure S1](#), the within-event
111 residuals at 2Hz, 5Hz and 10Hz in [Figure S2](#) and the within-site within-path residuals at 2Hz, 5Hz
112 and 10Hz in [Figure S3](#) are available in the Supplemental Material section.

Ground-Motion Model for the Velocity Fourier Amplitude Spectrum

Ergodic Ground-Motion Model for Velocity FAS

The functional form of the ergodic-base model for velocity FAS is kept relatively simple as more complex features will be added in the non-ergodic model. The functional form for the ergodic velocity FAS of the two horizontal components is given in Equation (1):

$$\ln FAS = c_0 + c_1 M + c_2 \ln(R_{rup}) + c_3 Z_{hyp} + \delta B_e + \delta S2S_s + \frac{R_{rup}}{100} \delta P_e + \delta WSP_{es} \quad (1)$$

in which M is the magnitude, R_{rup} is the rupture distance, and where Z_{hyp} is the hypocentral depth.. The δB_e , $\delta S2S_s$, and δP_e are zero-mean, normally distributed random effects for the source, site, and path, respectively, with standard deviations of τ , ϕ_{S2S} , and τ_P , respectively. The remaining unmodeled part of the ground motion is given by the within-site within-path residual δWSP_{es} with a standard deviation of ϕ_{SP} .

A finite-fault term is not included in the geometrical spreading term because the selected events were limited to magnitudes less than or equal to 4, for which the point-source assumption is reasonable. We also did not include a V_{S30} term in the ergodic model because there are inconsistencies between the V_{S30} estimates from the velocity model used for 3-D simulations and the V_{S30} values estimated for the seismic stations (Pinilla-Ramos *et al.*, 2025). Excluding V_{S30} from the model formulation implicitly incorporates its effect into the $\delta S2S$ site terms and does not reduce the model applicability for the two main uses of the non-ergodic GMM.

131 The δB_e and $\delta S_2 S_s$ random effects are commonly included in GMMs to account for correlations
 132 in the ground-motion data. The third random effect, δP_e , is a new term that accommodates an
 133 event-specific large-distance attenuation. Adding the $\frac{R_{rup}}{100} \delta P_e$ terms helps to reduce the mapping of
 134 source-specific path effects into the source terms that can occur for events with distance attenuation
 135 that is flatter or steeper than average. The normalization by 100 km leads to similar sizes of the
 136 τ and τ_p , which helps the stability of the regression. This type of event-specific linear distance
 137 term was included in the duration model by [Pinilla-Ramos *et al.* \(2023\)](#). Its use here is the first
 138 application for ground-motion amplitudes.

139 The coefficients from the regression were estimated in R using the linear mixed-effects package
 140 LMER ([Bates *et al.*, 2015](#)). The period dependence of the coefficients was smoothed using a simple
 141 polynomial of degree 1 for the magnitude term and a polynomial of degree 2 for the distance and
 142 depth terms. The smoothing of the coefficients leads to a smooth spectrum with frequency. In
 143 addition, the smoothing can be used to impose physical constraints on the coefficients ([Bayless
 144 and Abrahamson, 2019](#)). The smoothed regression coefficients are shown in Figure 3 and are listed
 145 in table 2.

146 The total standard deviation for the ergodic GMM is given by:

$$\sigma_{FAS} = \sqrt{\tau^2 + \phi_{S_2 S}^2 + \phi_{S P}^2 + \left(\frac{R_{rup}^2}{100^2}\right) \tau_p^2} \quad (2)$$

147 For developing non-ergodic GMMs that reflect the full non-ergodic path effects, we combine
 148 the δWSP and $\left(\frac{R_{rup}}{100}\right) \delta P_e$ terms into a within-site residual:

$$\delta W S_{es} = \delta W S P_{es} + \left(\frac{R_{rup}}{100}\right) \delta P_e \quad (3)$$

149 The standard deviation of $\delta W S_{es}$ is denoted ϕ_{SS} . The frequency dependence of τ , $\tau_p(R_{rup}/100)$,

150 ϕ_{S2S} , and ϕ_{SS} is shown in Figure 4 together with the total standard deviation, σ_{FAS} . The four
 151 components of the standard deviation are similar in size, with values generally between 0.35 and
 152 0.5 LN units.

153 **Non-Ergodic Ground-Motion Model for Velocity Fourier Amplitude** 154 **Spectrum**

155 In non-ergodic GMMs, the source-specific, site-specific, and path-specific deviations from the
 156 ergodic model (δB_e , $\delta S2S_s$, and $\delta P2P_{es}$) become spatially dependent and are functions of the
 157 coordinates of the source e and site s . These terms are modeled using Gaussian Processes which
 158 provide smooth maps of their median estimation as well as their epistemic uncertainty. Following
 159 the notation from [Sung *et al.* \(2023\)](#), the non-ergodic GMM is expressed as:

$$\begin{aligned}
 GMM_{NE}(M, R, S, ..., \vec{t}e_e, \vec{t}s_s) = & c_0 + c_1 M + c_2 \ln(R_{rup}) + c_3 Z_{hyp} + \delta L2L(\vec{t}e_e) + \delta S2S(\vec{t}s_s) \\
 & + \delta P2P(\vec{t}e_e, \vec{t}s_s) + \delta B_e^0 + \delta WSP_{es}
 \end{aligned} \tag{4}$$

160 in which $\vec{t}e_e$ and $\vec{t}s_s$ are the coordinates of the earthquake e and site s . The $\delta L2L(\vec{t}e_e)$ is the
 161 median deviation of the source term from the ergodic model for earthquake e , $\delta S2S(\vec{t}s_s)$ is the
 162 median shift in the site term for site s , and $\delta P2P(\vec{t}e_e, \vec{t}s_s)$ is the median shift in the path term for
 163 the ray path between earthquake e and site s .

164 The between-path term, $\delta P2P(\vec{t}e_e, \vec{t}s_s)$, represents how a specific travel path leads to systemat-
 165 ically different motions from the average distance scaling captured from the ergodic GMM, and
 166 the $\delta B^0 + \delta WSP$ term is the remaining aleatory variability after removing the systematic source,
 167 site, and path terms.

The total aleatory variance for the fully non-ergodic GMM, σ_{NE}^2 , is given by:

$$\sigma_{NE}^2 = \tau_0^2 + \phi_{SP}^2 \quad (5)$$

in which the τ_0 and the ϕ_{SP} are the standard deviations of the δB_e^0 and δWSP_{es} , respectively.

To compute the non-ergodic source and site terms, we first estimate the source and site random effects δB_e and $\delta S2S_s$ for the ergodic model. The source and site random effects are removed from the total residuals of the ergodic model to obtain the within-site residuals. Next, we model the δB_e and $\delta S2S_s$ as spatially varying Gaussian Processes for $\delta L2L(\vec{t}e_e)$ and $\delta S2S(\vec{t}s_s)$. The within-site residuals are then used to estimate the path effects between each source-site pair with Gaussian Process $\delta P2P(\vec{t}e_e, \vec{t}s_s)$. The hyperparameters of these Gaussian Processes were estimated via the maximum likelihood approach, and a single deterministic value was selected for the hyperparameters leading to the maximum likelihood for simplicity. The details of the approach used to estimate the non-ergodic terms are described in the following sections.

The non-ergodic model leads to a smaller standard deviation σ_{NE} in the residuals than the standard deviation σ_{FAS} from the ergodic model because it includes additional source, site and paths effects. The standard deviation is reduced by 0.35 ln units at 2Hz and by 0.6 ln units at 15Hz for FAS, as shown in Table 3. These large reductions in the standard deviation will have a significant effect on the seismic hazard, especially at long return periods..

Non-Ergodic Source Term

In an ergodic GMM, the source term, δB_e , represent the average isotropic source effect over all distances for earthquake e . The non-ergodic source term, $\delta L2L$, captures the spatial variation of the δB_e . For small-magnitude earthquakes, the non-ergodic source term is often interpreted to

188 represent spatial variations in the average stress drop, but the event terms can also include some
 189 wave propagation effects.

190 The non-ergodic source term is modeled as a Gaussian Process, following [Landwehr *et al.* \(2016\)](#).
 191 Gaussian Processes are fully described by their mean value and covariance function. The mean
 192 function is imposed to be zero, and the covariance function for the source term is modeled by an
 193 exponential kernel:

$$\kappa_{source-FAS}(\vec{te}, \vec{te}') = \tau_{L2L}^2 \exp\left(-\frac{\|\vec{te} - \vec{te}'\|}{\rho_e}\right) + \tau_0^2 \delta(\vec{te}, \vec{te}') \quad (6)$$

194 in which $\vec{te}$ and $\vec{te}'$ are the coordinates of the hypocenter of events e and e' . The τ_{L2L} term includes
 195 the variability in the source terms that are spatially correlated, with a correlation length of ρ_e . The
 196 τ_0 is the standard deviation of δB^0 , which is the remaining aleatory variability in the source term.
 197 The exponential kernel was chosen over a square-exponential because it gave a larger likelihood.

198 We estimated the parameters τ_{L2L} , τ_0 and ρ_e using maximum likelihood at each available
 199 frequency between $2Hz$ and $15Hz$ for both the East-West and North-South components. The
 200 results are given in Table 3. τ_{L2L} and τ_0 increase slightly with frequency but remain close to
 201 constant, and the correlation length ρ_e increases with frequency, which is likely due to the mapping
 202 of regional site terms into the source terms as discussed by [Abrahamson and Gulerce \(2022\)](#). The
 203 estimated correlation lengths are larger than the source separations, and the resulting Gaussian
 204 Process model effectively averages over many sources which avoids overfitting. The values of the
 205 hyperparameters are shown in Figure 5.

206 The median estimates of the non-ergodic source term are shown in Figure S4 for 2Hz, 5Hz,
 207 10Hz, and 15Hz for the East-West component. The corresponding maps of epistemic uncertainty
 208 in the median source term are shown in Figure S5. The sources located along the San Andreas

209 fault have an above-average median source term compared to sources located along the Hayward
 210 fault. The regions with similar non-ergodic adjustments become broader with frequency due to
 211 the increase in the correlation length with frequency. The epistemic uncertainty in the median
 212 predictions increases at source locations away from the available source locations, and regions of
 213 similar epistemic uncertainty widen with increasing frequency due to the increase in the correlation
 214 length with frequency.

215 We compared our results with the a recent non-ergodic models developed for California. The
 216 median non-ergodic source term at 5Hz are compared to the non-ergodic source terms from
 217 [Lavrentiadis *et al.* \(2023\)](#) in Figure 6,. The [Lavrentiadis *et al.* \(2023\)](#) model is based on the
 218 NGA-West2 database that only includes a few events in the San Francisco region. Consequently,
 219 the non-ergodic source term are poorly constrained in the region, and the median prediction of
 220 the source term from [Lavrentiadis *et al.* \(2023\)](#) is close to zero. On the other hand, the new
 221 model is based on a much larger number of events in the San Francisco region which leads to
 222 better constraints and a wider range of the non-ergodic source terms. This comparisons show the
 223 improvements in the constraints on the non-ergodic source terms for our GMM as compared to a
 224 previous study. They are not used as a validation of our model.

225 Non-Ergodic Site Term

226 Similar to the non-ergodic source term, the non-ergodic site term is modeled as a Gaussian Process,
 227 with a covariance function given by a squared exponential kernel:

$$\kappa_{site-FAS}(\vec{ts}, \vec{ts}') = \phi_{S2S,COR}^2 \exp\left(-\frac{\|\vec{ts} - \vec{ts}'\|^2}{2\rho_s^2}\right) + \phi_{S2S,NCOR}^2 \delta(\vec{ts}, \vec{ts}') \quad (7)$$

228 The $\phi_{S2S,COR}^2$ is the variance of the spatially correlated site term, ρ_s is the correlation length of the

spatially correlated site term, and $\phi_{S2S,NCOR}$ is the remaining variability in the systematic site term that is uncorrelated between sites (or the correlation length is very short). The square-exponential kernel was chosen over the exponential because it gave a larger likelihood.

We estimated the parameters $\phi_{S2S,COR}$, $\phi_{S2S,NCOR}$ and ρ_s using maximum likelihood at each available frequency. As with the source terms, the standard deviations $\phi_{S2S,COR}$ and $\phi_{S2S,NCOR}$ increase slightly with frequency but remain close to constant, and the correlation length increases with frequency, as observed in Figure 7. The correlation lengths for the site term are larger than the site separations, which indicates that there is not an issue of overfitting. The results are listed in Table 3.

The increase in the correlation length of the non-ergodic site term with frequency seems counterintuitive because higher frequencies will have shorter wavelengths and be more sensitive to the variability of the shallow velocity structures; however, the high-frequency FAS is strongly affected by kappa, and this large spatial correlation length likely reflects the spatial correlation of kappa, which shows similar regional patterns in the San Francisco Bay Area (Nye *et al.*, 2022). If the kappa is correlated with the regional surface geology, then the spatial correlation of the high-frequency FAS would reflect the spatial dimension of the surface geology (Eberhart-Phillips, 2016) and not the shallow velocity structure.

Figure S6 shows the forward prediction of the non-ergodic site term at 2Hz, 5Hz, 10Hz, and 15Hz for the East-West component. The corresponding maps of epistemic uncertainty in the prediction are shown in Figure S7. Sites located along the coast west of San Francisco have above-average median site terms, whereas sites located in the East Bay tend to have below-average site terms. The epistemic uncertainty increases at prediction sites located farther from the site locations in the dataset, and regions of similar epistemic uncertainty widen with increasing frequency due to the increase in the correlation length with frequency.

we compare the median non-ergodic site term at 5Hz from the new model to the site terms from [Lavrentiadis *et al.* \(2023\)](#) in In Figure 8. The data set used by includes a significant number of stations in the San Francisco region providing better constraints on the non-ergodic site terms than on the non-ergodic source terms. The hyperparameters for the 5Hz FAS for the two models are similar, with correlation lengths close to 15 km and standard deviations close to 0.3 ln units. The two models give similar spatial variations in the predicted non-ergodic site terms.

Non-Ergodic Path Term

We use the methodology developed by [Lacour *et al.* \(2024\)](#) to model non-ergodic path effects due to the 3-D velocity structure in the region of study. This approach includes an azimuthally dependent site term and can efficiently accommodate the full data set in the forward prediction of the non-ergodic path terms. For the correlation between the path effects for two ray paths, the approach considers the paths as vectors, which can be defined by three parameters: the origin, direction, and length. The distance between the site locations ($|SS'|$), the difference in azimuths ($\Delta\theta$), and the difference in rupture distances (ΔR_{rup}) define the differences in the origins, azimuths, and lengths, respectively. The parameters used to define the difference between two ray paths are illustrated in Figure 9.

The functional form used for the covariance function of path effects is given by:

$$k_{Path-FAS}(P, P') = \phi_{P2P}^2 \exp\left(\frac{-\Delta R_{rup}^2}{2\rho_R^2}\right) \exp\left(\frac{-\Delta\theta^2}{2\rho_\theta^2}\right) \exp\left(\frac{-|SS'|^2}{2\rho_S^2}\right) + \phi_{SP,NE}^2 \delta(P, P') \quad (8)$$

The *NE* subscript is included, to distinguish the standard deviation of the within-path within-site residuals for the non-ergodic GMM from the standard deviation of the within-path within-site

272 residuals for the ergodic GMM with the path random effects (Equation 2).

273 This form of the covariance function was shown to better fit the correlation between paths
274 than other approaches used in non-ergodic GMMs (e.g., [Sung et al., 2023](#)). In particular, this form
275 of the covariance function leads to more accurate estimation of the correlation of path effects for
276 the case in which data are available with moderate correlation (near 0.5) and better extrapolation
277 of path effects into regions where data are not available. Additionally, the new method uses a
278 path correlation function that is separable between its azimuth, rupture distance, and site location
279 terms, which allows the use of efficient numerical methods that drastically reduce computational
280 and memory requirements when dealing with large datasets as shown by [Lacour \(2022\)](#).

281 To avoid overfitting, the correlation lengths ρ_R , ρ_θ and ρ_S were estimated by fitting the semi-
282 variogram of the within-site residuals rather than estimating them as part of the regression. With
283 the correlation lengths held fixed, the variance terms ϕ_{P2P} and $\phi_{SP,NE}$ were estimated using
284 maximum likelihood. The term ϕ_{P2P} represents the variability of path effects that can be explained
285 via the correlation function used to characterize the similarity between path effects, whereas $\phi_{SP,NE}$
286 represents the remaining variability that cannot be explained by the model. The values obtained
287 for ϕ_{P2P} are consistent with the values obtained by other models for various regions around the
288 world for the frequency range of this study [2Hz ; 15Hz], around 0.40, as described in [Sung and](#)
289 [Abrahamson \(2024\)](#). The estimates of the hyperparameters are listed in Table 3 and the values
290 from other regions from [Sung and Abrahamson \(2024\)](#) are shown in Figure S8.

291 To illustrate the results, we calculate the forward prediction of the non-ergodic path term at 2
292 Hz, 5 Hz, 10 Hz, and 15 Hz for the East-West component at two different site locations and sources
293 located at 8 km of depth. The sites (number 55 and number 87 in the dataset) were chosen to be
294 on each side of the bay, and the depth of the source locations for the prediction was chosen as the
295 average depth from all the sources available in the dataset. We show the prediction of these path

effects at 5Hz in Figures 10 and 11 and at 2 Hz, 5 Hz, 10 Hz, and 15 Hz in Figures S9 and S10. The site in the West Bay (Figure 10) shows strong positive path effects from the sources located along the east side of the Hayward fault and shows smaller path effects sources located along the San Andreas fault. The site in the East Bay (Figure 11) shows moderate path effects but only from sources located in the south end of the Hayward fault. The net effect of the non-ergodic median source, site, and path effects and the corresponding epistemic uncertainty at these two sites are shown in Figures S11 and S12.

There are no existing empirical models of the path effects due to 3-D velocity structure ($\delta P2P$ terms) in the San Francisco region that can be used for comparison with our GMM. Previous empirical non-ergodic GMMs for path effects for this regions only use the approach by Dawood and Rodriguez-Marek (2013) which only capture path effects due to the 3-D anelastic attenuation and impacts the model at larger distances ($R_{rup} > 50$ km) and higher frequencies (> 5 Hz).

Ground-Motion Model for the Phase Derivative

The Fourier phase angles describe the shape of the waveform, but because the phase angles are modulo 2π , it is difficult to model the phase angles directly. As an alternative, we use the standard deviation of the distribution of the phase derivatives as a parameter to characterize the phase angles. The standard deviation of the phase derivative is related to the duration of the time series (Baglio *et al.*, 2023). Similar to the approach used for the FAS model, we first develop an ergodic model for the standard deviation of the phase derivative and then develop the non-ergodic model by modeling the spatial variations of the systematic effects of the source, site, and path effects on the phase derivative using Gaussian Processes.

Ergodic Ground-Motion Model for Phase Derivative

The phase derivatives are computed over the frequency range $f \in [2; 15]$ Hz:

$$\phi'(f) = \frac{d\phi(f)}{df} \simeq \frac{\Delta\phi(f)}{\Delta f} = \frac{\phi(f_{i+1}) - \phi(f_i)}{f_{i+1} - f_i}, \quad (9)$$

If the phase difference, $\phi(f_{i+1}) - \phi(f_i)$, is not in the interval of $[-\pi, \pi]$, then 2π is added or subtracted to the phase difference such that the adjusted phase difference is in the interval $[-\pi, \pi]$.

We fit a logistic distribution to the phase-derivative values between 2-15 Hz. The logistic distribution was selected because it has been shown to better fit the tails of the phase-derivative distribution than a normal distribution (Baglio *et al.*, 2023). The logistic distribution is given by:

$$f_{\phi'} = \frac{e^{-(\phi' - \mu_{\phi'})/\sigma_{\phi'}}}{\sigma_{\phi'}(1 + e^{-(\phi' - \mu_{\phi'})/\sigma_{\phi'}})} \quad (10)$$

in which ϕ' is the phase derivative with mean $\mu_{\phi'}$ and standard deviation $\sigma_{\phi'}$.

The standard deviation of the phase derivative is related to the duration of ground motion as shown in Figure 12. For time series with longer duration, the distribution of the phase derivative is wider, and vice versa. A broader distribution of the phase derivative leads to larger variations of the phase angles, which leads to a longer duration. This phenomenon is independent from the region of study. The standard deviation of the phase derivative is shown as a function of the D_{5-95} duration in Figure 13.

Given the strong correlation between the duration and the standard deviation of the phase derivative, we select a functional form for the phase derivative that is similar to the functional form used in duration ground-motion models. In particular, we follow the approach of Pinilla-Ramos *et al.* (2023), which incorporates two main improvements in duration models. The first improvement

335 is the use of a power normal distribution to approximate the distribution of duration. Power
 336 distributions can accommodate a skewness between a normal and a lognormal distribution (for
 337 exponents between 0 and 1). Using a power distribution improves the models for the tails of the
 338 distribution. Another fundamental improvement in the duration model by [Pinilla-Ramos *et al.*](#)
 339 [\(2023\)](#) is that the source, site, and path effects are additive and not multiplicative as used in some
 340 duration models ([Afshari and Stewart, 2016](#)), [Bahrapouri *et al.* \(2021\)](#). This additive formulation
 341 incorporates the source, site, and path effects within the exponent of the power law. The addition
 342 of these terms inside the power makes it more challenging to estimate the random effects. Following
 343 [Pinilla-Ramos *et al.* \(2023\)](#), the formulation of our ergodic model for the phase derivative is given
 344 by Equation (11):

$$\begin{aligned}
 (\sigma_{\phi'})^{0.3} = & c_0 + c_1 \left(10^{0.5(M-4)} + c_2 \Delta S 2S_s + c_3 R_{rup} + c_4 \Delta P_e \frac{R_{rup}}{100} + c_5 (Z_{hyp} - 8) \right)^{0.3} \\
 & + \delta W S P_{es}
 \end{aligned} \tag{11}$$

345 In Equation (11), $10^{0.5(M-4)}$ is related to the source duration, and $c_5(Z_{hyp} - 8)$ is related to
 346 the depth scaling of the duration. The two random-effect terms are denoted with Δ here because
 347 we make the assumption that they are log-normally distributed to avoid negative values inside
 348 the power law (as opposed to the δ notation commonly used for normally distributed random-
 349 effect terms in LN units). The mean values of $\delta S 2S$ and δP_e are zero, leading to mean values
 350 of $\Delta S 2S$ and ΔP_e of unity. As a result of this formulation, the average site term is c_2 , and the
 351 site-specific deviation from the average site effect is $c_2(\Delta S 2S - 1)$. Similarly, the average path
 352 effect is $c_3 R_{rup} + c_4 R_{rup}/100$, and the event-specific deviation from the average path effect is
 353 $(c_4(\Delta P_e - 1)R_{rup}/100)$.

354 The source duration for small-magnitude earthquakes has a negligible effect on the total duration
 355 and cannot be determined as part of the regression. For example, for an M3 event at a distance of
 356 20 km, the source term of $10^{0.5(M-4)}$ is 0.3 and the path term is 1.1. At larger distances, the path
 357 term will become even larger, which will make the fractional contribution of source term to the
 358 total duration even smaller. Therefore, we constrained the coefficient for the magnitude scaling to
 359 equal 0.5 based on the expected magnitude scaling of the source duration for $\log_{10}(Area) = M - 4$
 360 with an aspect ratio of 1. We also imposed the site-specific duration scaling, c_2 , to be equal to 0.7
 361 to ensure similar values as found in the study of [Pinilla-Ramos *et al.* \(2023\)](#).

362 Finally, the exponent for the powernormal distribution was selected to be 0.3 because this
 363 transformation led to a distribution of the δWSP that was closest to a normal distribution, as
 364 seen in the probability plot in Figure 14. The regression was performed using PyMC3 with the
 365 non-linear mixed-effect regression function. The estimated coefficients for both the East-West and
 366 North-South components are listed in Table 4.

367 **Non-Ergodic Ground-Motion Model for Phase Derivative**

368 In this section, we develop the first non-ergodic ground-motion model for duration. Because of
 369 the novelty of the results, there are no alternative models available to compare with results of
 370 this analysis. Non-ergodic source effects on the duration from small-magnitude events are much
 371 smaller than the site and path effects. Because the effects are additive, the source terms cannot be
 372 reliably estimated for these small-magnitude events. Therefore, a non-ergodic source term was
 373 not included in the model, and the non-ergodic GMM for the phase derivative only includes the
 374 spatially varying site and path effects.

Non-Ergodic Site Term

As with the site terms for the FAS, the spatial variation of the site random effects for phase derivative is modeled by a Gaussian process with a squared-exponential covariance function given by:

$$\kappa_{Site-\phi'}(\vec{ts}, \vec{ts}') = \phi_{S2S-PD,COR}^2 \exp\left(-\frac{\|\vec{ts} - \vec{ts}'\|^2}{2\rho_{s-PD}^2}\right) + \phi_{S2S-PD,NCOR}^2 \delta(\vec{ts}, \vec{ts}') \quad (12)$$

The square-exponential kernel was chosen over the exponential because it gave a larger likelihood. The estimated values of the hyperparameters for the East-West and North-South components are listed in Table 5.

Figure 15 shows an example of the forward prediction of the non-ergodic site term for the East-West component and the corresponding epistemic uncertainty in the median prediction. The correlation length of the site term for the standard deviation of the phase derivative is 10-12 km for the two horizontal components, leading to broad regions with systematically longer or shorter duration. The epistemic uncertainty also presents broad regions and increases for stations away from the study region.

Non-Ergodic Path Term

Similar to the FAS model, we model the non-ergodic path effects using the approach by [Lacour et al. \(2024\)](#); however, for the phase-derivative path effects, we use different input parameters to characterize the correlation between ray paths: source separation distance, azimuth difference, and site separation distance. Different combinations of input parameters for the covariance model were considered. Unlike the FAS data, the differences in the lengths of the rays (ΔR_{rup}) did not have a strong effect on the semivariogram. Instead, the difference in the earthquake locations ($|EE'|$)

had a larger effect on the semivariogram than the difference in the R_{rup} values. The covariance function for the path effect for the standard deviation of the phase derivative is modeled by:

$$k_{Path-\phi'}(P, P') = \sigma_{P-PD}^2 \exp\left(\frac{-|EE'|}{\rho_{E-PD}} + \frac{-\Delta\theta}{\rho_{\theta-PD}} + \frac{-|SS'|}{\rho_{S-PD}}\right) + \phi_{SP-PD,NE}^2 \delta(P, P') \quad (13)$$

Following the approach used for the FAS, the correlation lengths ρ_{E-PD} , $\rho_{\theta-PD}$ and ρ_{S-PD} were estimated from the semi-variogram, and then the terms σ_{P-PD} and ϕ_{SP-PD} were estimated using maximum likelihood.

The values of the hyperparameters for the two horizontal components are listed in Table 5. The estimated σ_{P-PD} values (0.2 radians/Hz) are substantially greater than the corresponding $\phi_{SP-PD,NE}$ values of 0.04 to 0.06 radians/Hz. This large difference indicates that most of the variability of the within-site residuals for the phase derivative can be explained by systematic path effects.

It is easier to understand changes in the duration than changes in the standard deviation of the phase derivative. Therefore, we converted the path effects for $\sigma_{\phi'}^{0.3}$ to path effects on duration using the empirical linear relationship shown in Figure 13. With the power transformations in Equation (11), the change in the duration will depend on the earthquake scenario (M and R_{rup}).

Maps of the median path effect for the duration are shown in Figure 16 for a magnitude 3 earthquake at 8 km depth for two site locations. For the station located on the west side of the bay, there are strong path-specific durations from sources located south of the Hayward Fault and below-average path-specific durations from sources located along the southern part of the San Andreas Fault. For the station located east of the Hayward Fault, there are moderately increased durations from sources south of the Hayward Fault and negligible path-specific duration terms

415 from sources located along the San Andreas Fault.

416 The epistemic uncertainty is also shown in Figure 16. The path duration terms grow linearly
417 with source distance in an isotropic way for the two stations due to the linear R_{rup} term used in the
418 functional form for the $\sigma_{\phi'}$ given in Equation (11). The R_{rup} term dominates the spatial variability
419 of the epistemic uncertainty in the path effects because of the short correlation lengths obtained
420 from the regression of Equation (13), which are listed in Table 5.

421 Conclusion

422 Frequent small-magnitude earthquakes (M1-4) sample a large set of ray paths that can be used
423 to constrain linear non-ergodic site and path effects in the San Francisco Bay Area. A limitation
424 of small-magnitude data is the poor signal-to-noise ratio for many recordings. Of the 626,423
425 available recordings in the study region between 1990 and 2023, only about 5000 recordings had
426 an acceptable signal-to-noise ratio. This smaller data set is still adequate to constrain spatially
427 varying source, site, and path effects for the FAS and phase-derivative in the San Francisco Bay
428 Area.

429 An issue with most ground-motion data sets is the tradeoff between path effects and source
430 effects. Including an additional random effect for the event-specific linear distance scaling, $\frac{R_{rup}}{100}\delta P_e$,
431 reduces the tradeoff between path effects and source effects, leading to more reliable source terms.

432 For the velocity FAS, sources near the San Andreas fault have positive non-ergodic source
433 terms, whereas sources near the Hayward fault have negative non-ergodic source terms. The
434 spatial extent of the positive and negative source terms becomes larger with increasing frequencies,
435 reflecting an increase in the correlation length with frequency. A limitation of our approach is
436 that the non-ergodic source terms cannot fully separate stress-drop variations from unmodeled

437 path effects, particularly for small-magnitude earthquakes where the source duration is negligible
438 compared to path effects. Future work could address this by comparing the spatial distribution of
439 source terms with independent stress-drop estimates. [Baltay et al. \(2017\)](#) demonstrated a positive
440 correlation between location-based source terms and spatially averaged stress drops, suggesting
441 that regional variations in stress drop can be identified through such comparisons. Similar analyses
442 with independent stress-drop catalogs or forward simulations using 3D velocity models with varying
443 stress drops could help validate whether observed source-term patterns reflect true stress-drop
444 variations or unmodeled path effects.

445 Similarly, sites located in the western part of the region have positive non-ergodic site terms,
446 whereas sites in the eastern part of the region have negative site terms. The site terms contours
447 are also broader for higher frequencies.

448 The path effects depend on the source and site location and are parameterized by a map of
449 the path effects for each site separately. After removing the path effects, the remaining standard
450 deviation of the within-path residuals is similar to the standard deviations found for other regions
451 around the world.

452 For the phase derivative, there is a high correlation between D5-95 duration and the standard
453 deviation of the phase derivative, confirming that the standard deviation of the phase derivative is
454 a good proxy for duration, in agreement with [Baglio et al. \(2023\)](#). For the non-ergodic model of
455 phase derivative, only site and path-specific random effects were included in the model because
456 the effect of the source on duration is negligible for small-magnitude earthquakes. The spatially
457 varying site effects for the phase derivative lead to variations of the duration of up to 5 seconds.
458 For the path effects, the non-ergodic phase derivative terms lead to up to a 3-second change in the
459 duration.

460 The ergodic and non-ergodic models for the FAS and phase derivative from the small-magnitude

461 events can be used to test AI-based approaches for developing non-ergodic models that may be
462 more efficient numerically than large-scaling 3-D simulations of non-ergodic source, path, and site
463 effects on the ground motions.

464 **Data and Resources**

465 The FAS data used in this study are given in the electronic attachment. We used the LME4 package
466 in R developed [Bates *et al.* \(2015\)](#) and the PyMC3 package in Python developed by [Abril-Pla *et al.*](#)
467 [\(2023\)](#) to perform the linear and non-linear regressions, respectively. We present additional figures
468 of the epistemic uncertainty of the predictions from our non-ergodic model in the Supplemental
469 Material section.

470 **Acknowledgments**

471 This work was supported by the Laboratory Directed Research and Development Program of
472 Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-
473 05CH11231 and partially supported by the PG&E long-term seismic program.

474 We are grateful to two anonymous reviewers for their thorough reviews and insightful comments,
475 which significantly improved the clarity and quality of this manuscript.

476 **Declaration of Competing Interests**

477 The authors acknowledge there are no conflicts of interest recorded.

References

- Abrahamson, N. and R. Youngs (1992). A stable algorithm for regression analyses using the random effects model, *Bulletin of the Seismological Society of America* **82**(1), 505–510.
- Abrahamson, N. A. and Z. Gulerce (2022). Summary of the Abrahamson and Gulerce NGA-SUB ground-motion model for subduction earthquakes, *Earthquake Spectra* **38**(4), 2638–2681.
- Abril-Pla, O., V. Andreani, C. Carroll, L. Dong, C. Fongnesbeck, M. Kochurov, R. Kumar, J. Lao, C. Luhmann, O. Martin, M. Osthege, R. Vieira, T. Wiecki, and R. Zinkov (2023). PyMC: a modern, and comprehensive probabilistic programming framework in Python, *PeerJ Computer Science* **9**, paper e1516.
- Afshari, K. and J. P. Stewart (2016). Physically Parameterized Prediction Equations for Significant Duration in Active Crustal Regions, *Earthquake Spectra* **32**(4), 2057–2081.
- Baglio, M., A. Cardoni, G. P. Cimellaro, and N. Abrahamson (2023). Generating ground motions using the Fourier amplitude spectrum, *Earthquake Engineering & Structural Dynamics* **52**(15), 4884–4899.
- Bahrapouri, M., A. Rodriguez-Marek, and R. A. Green (2021). Ground motion prediction equations for significant duration using the KiK-net database, *Earthquake Spectra* **37**(2), 903–920.
- Baltay, A. S., T. C. Hanks, and N. A. Abrahamson (2017). Uncertainty, Variability, and Earthquake Physics in Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America* **107**(4), 1754–1772.

498 Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting Linear Mixed-Effects Models
 499 Using lme4, *Journal of Statistical Software* **67**(1), 1–48.

500 Bayless, J. and N. A. Abrahamson (2019, 09). Summary of the BA18 Ground-Motion Model for
 501 Fourier Amplitude Spectra for Crustal Earthquakes in California, *Bulletin of the Seismological
 502 Society of America* **109**(5), 2088–2105.

503 Dawood, H. M. and A. Rodriguez-Marek (2013). A Method for Including Path Effects in Ground-
 504 Motion Prediction Equations: An Example Using the Mw 9.0 Tohoku Earthquake Aftershocks,
 505 *Bulletin of the Seismological Society of America* **103**(2B), 1360–1372.

506 Eberhart-Phillips, D. (2016, 10). Northern California Seismic Attenuation: 3D QP and QS Models,
 507 *Bulletin of the Seismological Society of America* **106**(6), 2558–2573.

508 Goulet, C. A., N. A. Abrahamson, P. G. Somerville, and K. E. Wooddell (2015). The SCEC
 509 Broadband Platform Validation Exercise: Methodology for Code Validation in the Context of
 510 Seismic-Hazard Analyses, *Seismological Research Letters* **86**(1), 17–26.

511 Konno, K. and T. Ohmachi (1998, 02). Ground-motion characteristics estimated from spectral
 512 ratio between horizontal and vertical components of microtremor, *Bulletin of the Seismological
 513 Society of America* **88**(1), 228–241.

514 Kottke, A. R., N. A. Abrahamson, D. M. Boore, Y. Bozorgnia, C. A. Goulet, J. Hollenback,
 515 T. Kishida, O.-J. Ktenidou, E. M. Rathje, W. J. Silva, E. M. Thompson, and X. Wang (2021).
 516 Selection of random vibration theory procedures for the NGA-East project and ground-motion
 517 modeling, *Earthquake Spectra* **37**(1_suppl), 1420–1439.

518 Lacour, M. (2022). Efficient non-ergodic ground-motion prediction for large datasets, *Bulletin of
 519 Earthquake Engineering* **21**, 5209–5232.

- 520 Lacour, M., N. Abrahamson, and C. Sung (2024). New Approach to Non-Ergodic Path Effects in
521 Ground-Motion Models, *Bulletin of the Seismological Society of America* (submitted) **X**, X.
- 522 Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016, 12). A Nonergodic Ground-
523 Motion Model for California with Spatially Varying Coefficients, *Bulletin of the Seismological*
524 *Society of America* **106**(6), 2574–2583.
- 525 Lavrentiadis, G., N. A. Abrahamson, and N. M. Kuehn (2023). A non-ergodic effective amplitude
526 ground-motion model for California, *Bulletin of Earthquake Engineering* **21**(11), 5233–5264.
- 527 McCallen, D., A. Petersson, A. Rodgers, A. Pitarka, M. Miah, F. Petrone, B. Sjogreen, N. Abra-
528 hamson, and H. Tang (2021). EQSIM—A multidisciplinary framework for fault-to-structure
529 earthquake simulations on exascale computers part I: Computational models and workflow,
530 *Earthquake Spectra* **37**(2), 707–735.
- 531 McCallen, D., F. Petrone, M. Miah, A. Pitarka, A. Rodgers, and N. Abrahamson (2021). EQSIM—A
532 multidisciplinary framework for fault-to- structure earthquake simulations on exascale computers,
533 part II: Regional simulations of building response, *Earthquake Spectra* **37**(2), 736–761.
- 534 NCEDC (2014). Northern California Earthquake Data Center [Data set].
- 535 Nye, T., V. J. Sahakian, E. King, A. Baltay, and A. Klimasewski (2022, 12). Estimates of κ_0 and
536 Effects on Ground Motions in the San Francisco Bay Area, *Bulletin of the Seismological Society*
537 *of America* **113**(2), 823–842.
- 538 Pinilla-Ramos, C., A. Pitarka, D.B. McCallen and R. Nakata (2025). Performance evaluation of
539 the USGS velocity model for the San Francisco Bay Area, *Earthquake Spectra* **41**(1), 457–494.
- 540 Pinilla-Ramos, C., N. Abrahamson, V. Phung, R. Kayen, and P. Castellanos-Nash (2023, 10).

541 Ground-Motion Model for Significant Duration Constrained by Seismological Simulations, *Bulletin*
542 *of the Seismological Society of America* **114**(2), 1015–1032.

543 Sung, C., N. Abrahamson, and M. Lacour (2023). Methodology for Including Path Effects Due to
544 3-D Velocity Structure in Non-Ergodic Ground-Motion Models, *Bulletin of the Seismological*
545 *Society of America* **113**(5), 2144–2163.

546 Sung, C.-H. and N. Abrahamson (2024). Update to the Non-ergodic GMM for France Including
547 Path Effects Due to the 3-D Velocity Structure. Technical report, Electricite De France.

548 Sung, C.-H. and N. Abrahamson (2024). Comparison of the Spatial Correlation of Non-ergodic
549 Terms in GMMs Utilizing Empirical and Simulation Data from Diverse Regions, *Seismological*
550 *Society of America Annual Meeting*, Abstract

551 **Full Mailing Address for Each Author**

552 Maxime Lacour
553 University of California, Berkeley
554 Department of Civil & Environmental Engineering
555 750 Davis Hall
556 Berkeley, CA 94720

557

558 Norman Abrahamson
559 University of California, Berkeley
560 Department of Civil & Environmental Engineering
561 750 Davis Hall

562 Berkeley, CA 94720

563

564 Rie Nakata

565 Lawrence Berkeley National Laboratory

566 1 Cyclotron Road, 074A-0107

567 Berkeley, CA 94720

568

569 Nori Nakata

570 Lawrence Berkeley National Laboratory

571 1 Cyclotron Road, 074A-0102

572 Berkeley, CA 94720

573

574 Camilo Pinilla-Ramos

575 University of Southern California

576 3651 Trousdale Parkway

577 Los Angeles, CA, 90089

578

Table 1: Summary of the small-magnitude data set.

Component	Recordings	Selected	Stations	Events	$1 < M < 2$	$2 < M < 3$	$3 < M < 4$
East-West	626,423	5,108	266	737	38	496	203
North-South	626,423	5,301	269	725	39	470	216
Both	626,423	3,174	211	577	18	369	290

Table 2: Smoothed model coefficients for the FAS model for both the East-West and North-South components. f is the frequency in Hz.

Coefficient	Smoothed Model
c_1	$3.43 - 0.52 M$
c_2	$-1.51 - 0.09 \ln(R) - 0.03 \ln(R)^2$
c_3	$-0.03 - 0.02 Z - 0.01 Z^2$

Table 3: Hyperparameters for the FAS model. f is the frequency in Hz.

Type	Hyperparameter	East-West Component	North-South Component
Source	ρ_e (km)	$3.2 f - 0.6$	$1.7 f + 4$
	τ_{L2L} (ln units)	$0.005 f + 0.1$	$0.004 f + 0.12$
	τ_0 (ln units)	0.20	0.21
Site	ρ_s (km)	$1.1 f + 5.7$	$1.9 f + 1.2$
	$\phi_{S2S,COR}$ (ln units)	$0.01 f + 0.14$	$0.004 + 0.17$
	$\phi_{S2S,NCOR}$ (ln units)	0.40	0.40
Path	$\rho_{\Delta R}(km)$	$1.3 f + 4.0$	$0.8 f + 6.5$
	$\rho_{\Delta\theta}(degrees)$	15	16
	$\rho_{\Delta S}$ (km)	13	11
	ϕ_{P2P} (ln units)	$0.01 f + 0.40$	$0.01 f + 0.40$
	$\phi_{SP,NE}$ (ln units)	0.40	0.40

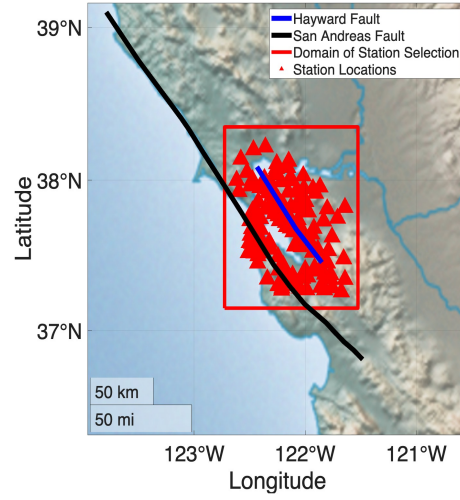
Table 4: Coefficients for the ergodic model for the standard deviation of the phase derivative (combined EW and NS components).

c_0	c_1	c_2	c_3	c_4	c_5
0.33	1.22	0.7	0.054	3.76	0.036

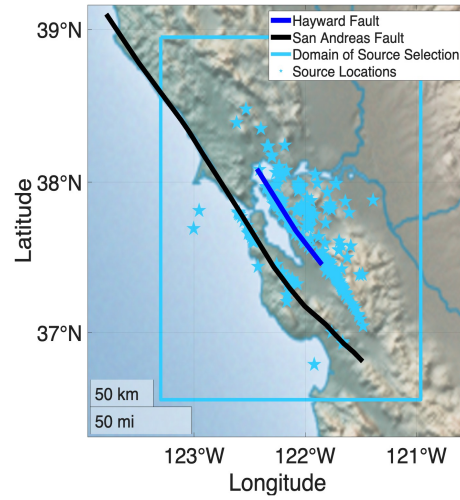
Table 5: Hyperparameters for the phase-derivative model.

Type	Hyperparameter	East-West Component	North-South Component
Source	ρ_{e-PD} (km)		
Site	ρ_{s-PD} (km)	17	12
	$\phi_{S2S-PD,COR}$ (rad/Hz)	0.46	0.32
	$\phi_{S2S-PD,NCOR}$ (rad/Hz)	0.42	0.40
Path	$\rho_{\Delta E-PD}(km)$	10	12
	$\rho_{\Delta\theta-PD}(degrees)$	10	10
	$\rho_{\Delta S-PD}$ (km)	10	12
	σ_{P-PD} (rad/Hz) ^{0.3}	0.20	0.20
	$\sigma_{SP-PD,NE}$ (rad/Hz) ^{0.3}	0.04	0.06

(a)



(b)



(c)

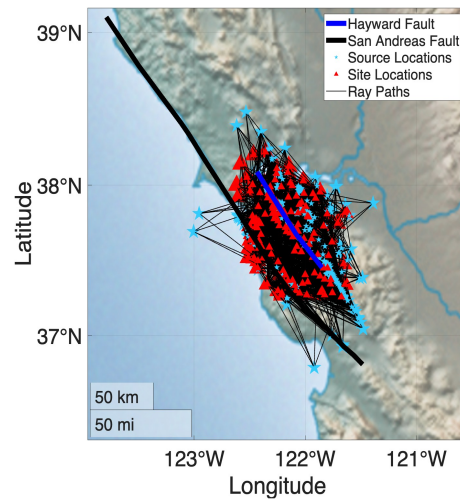
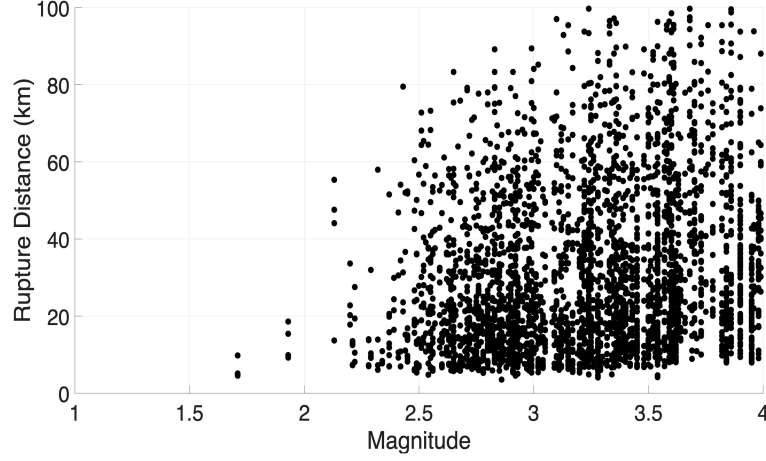


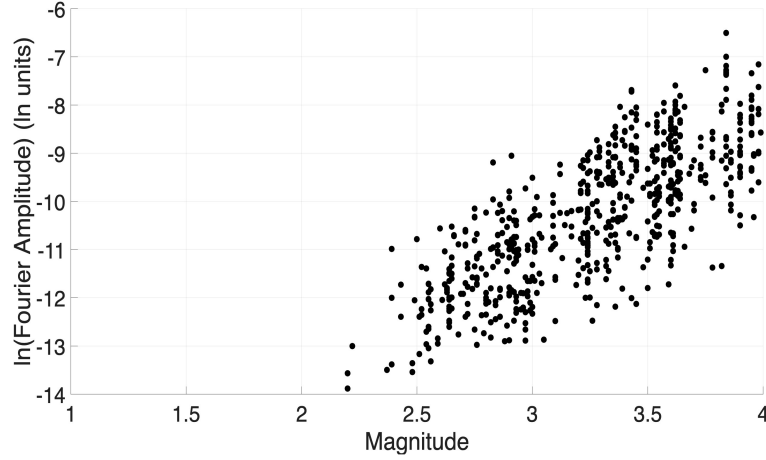
Figure 1: (a) Station locations, (b) source locations and (c) ray paths from the selected sources and sites for the H1 component data set.

Alt-text: Maps of the San Francisco Bay Area. Panel (a) shows the distribution of seismic stations as triangles across the region. Panel (b) shows earthquake source locations as stars along the San Andreas and Hayward faults. Panel (c) shows ray paths as connecting lines between earthquake sources and stations for the H1 component, illustrating the spatial coverage of the dataset.

(a)



(b)



(c)

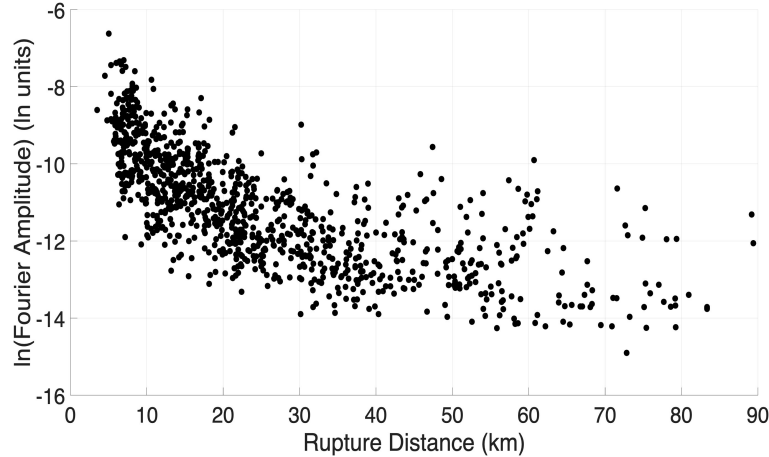
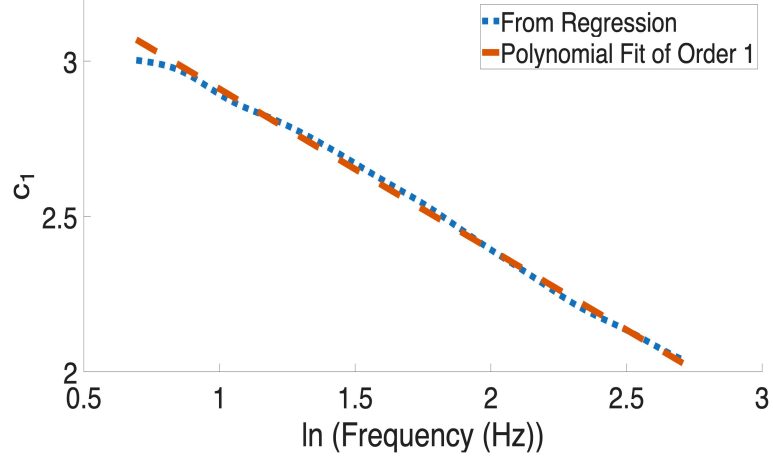


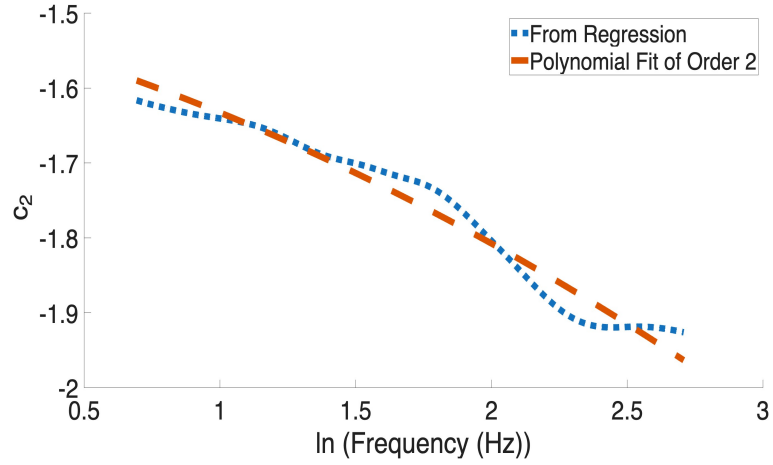
Figure 2: (a) Magnitude-rupture distance scatter plot, (b) Fourier amplitude with magnitude for $R_{rup} \in [20; 30] \text{ km}$ (c) Fourier amplitude with rupture distance for $M \in [2.5; 3]$ at 5Hz for the East-West component.

Alt-text: Three diagnostic plots for the dataset. (a) Scatter plot of earthquake magnitude versus rupture distance, showing coverage of events in the dataset. (b) Fourier amplitude spectrum (FAS) values at 5 Hz plotted against magnitude for recordings at rupture distances between 20–30 km; amplitudes increase systematically with magnitude. (c) FAS values at 5 Hz plotted against rupture distance for M 2.5–3 events; amplitudes decay with distance.

(a)



(b)



(c)

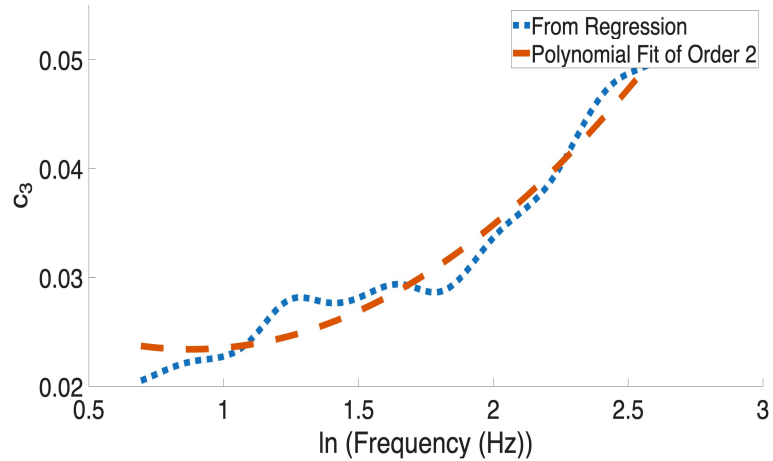


Figure 3: Smoothed coefficients for the ergodic FAS GMM for the east-west and north-south components. (a) Magnitude term, (b) distance term, (c) depth term.

Alt-text: Frequency-dependent regression coefficients for the ergodic FAS model. (a) Magnitude scaling term, (b) Distance scaling term, and (c) Depth scaling term. Each coefficient is shown as a smooth curve across frequency, illustrating consistent spectral trends.

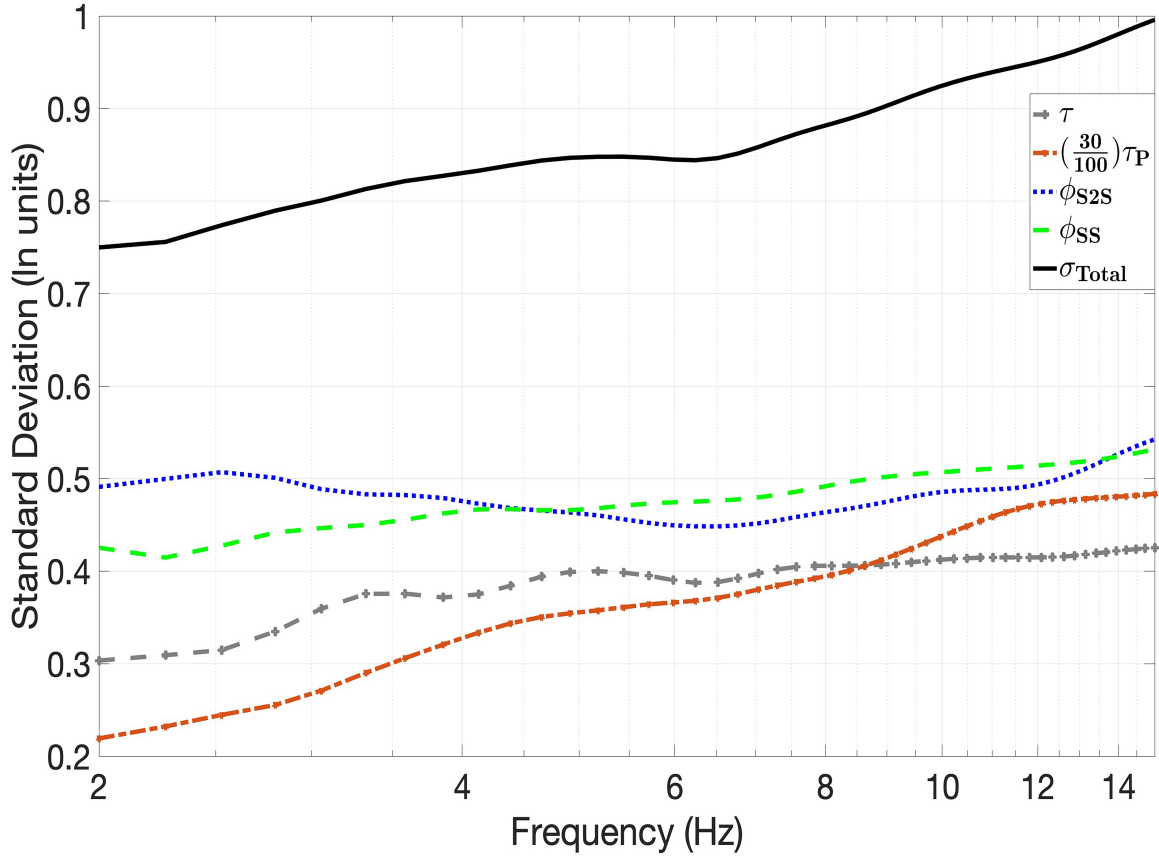


Figure 4: Components of the standard deviation of the FAS for $R_{rup} = 30$ km.

Alt-text: Line plot of the variance components that contribute to the standard deviation of the ergodic FAS prediction at $R_{rup} = 30$ km. Separate curves show between-event, site-to-site, within-site/within-path, and event-specific distance effects, as well as the total standard deviation.

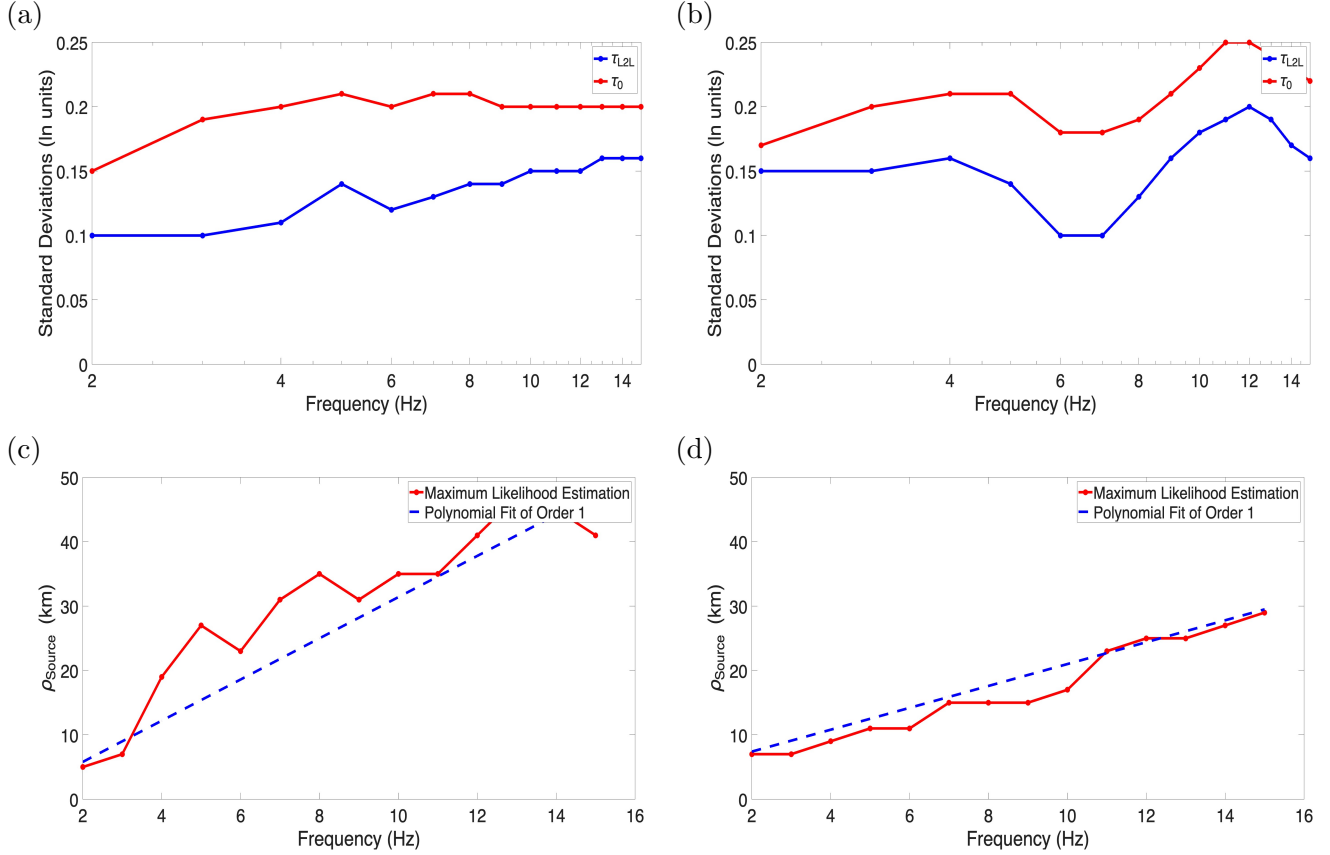


Figure 5: Frequency dependence of the standard deviations and of the correlation lengths for the non-ergodic source term. (a), (c) EW component ; (b), (d) NS component.

Alt-text: (a, c) Frequency dependence of standard deviations and correlation lengths for non-ergodic source terms in the East–West component. (b, d) Same quantities for the North–South component. Plots show correlation length increasing with frequency.

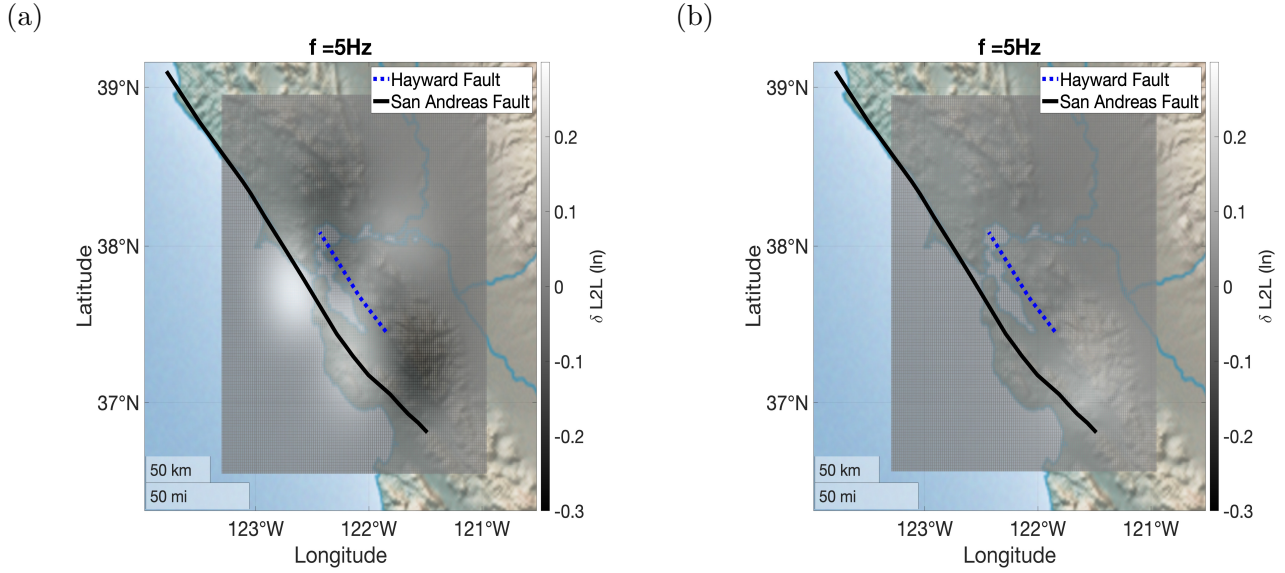


Figure 6: Maps of the median FAS source terms $\delta L2L$ at 5Hz for the East-West component from the presented model (left) and from the EAS model from [Lavrentiadis *et al.* \(2023\)](#) (right).
 Alt-text: Side-by-side comparison maps of median non-ergodic source terms at 5 Hz for the East–West component. The left map (this study) shows positive adjustments (above-average amplitudes) near the San Andreas fault and negative adjustments near the Hayward fault. The right map (Lavrentiadis et al. 2023) shows values close to zero in most of the region due to limited local data coverage.

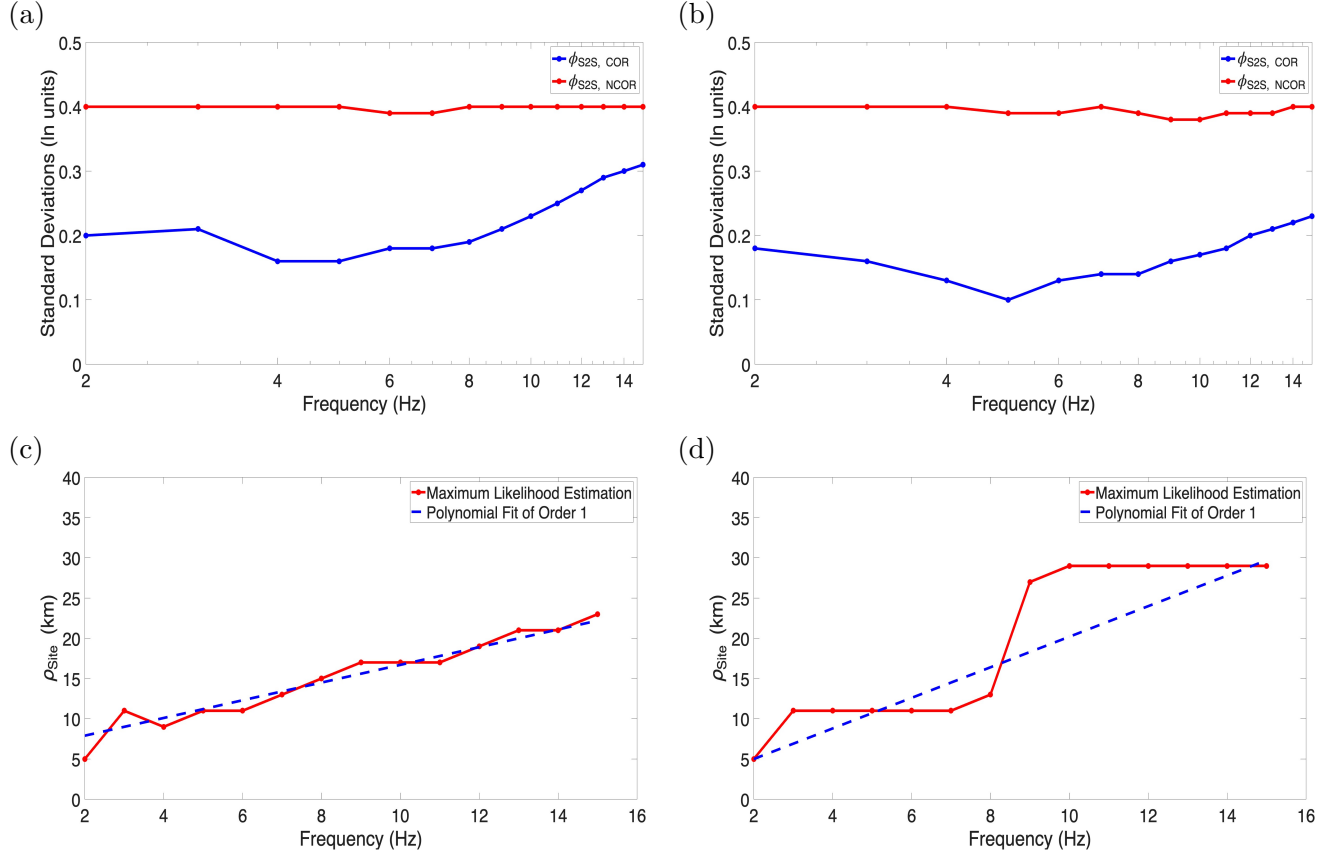


Figure 7: Frequency dependence of the standard deviations and of the correlation lengths for the non-ergodic site term. (a), (c) EW component ; (b), (d) NS component.

Alt-text: (a, c) Frequency dependence of standard deviations and correlation lengths for non-ergodic site terms, East–West component. (b, d) Same for North–South component. Correlation lengths and standard deviations both increase slightly with frequency.

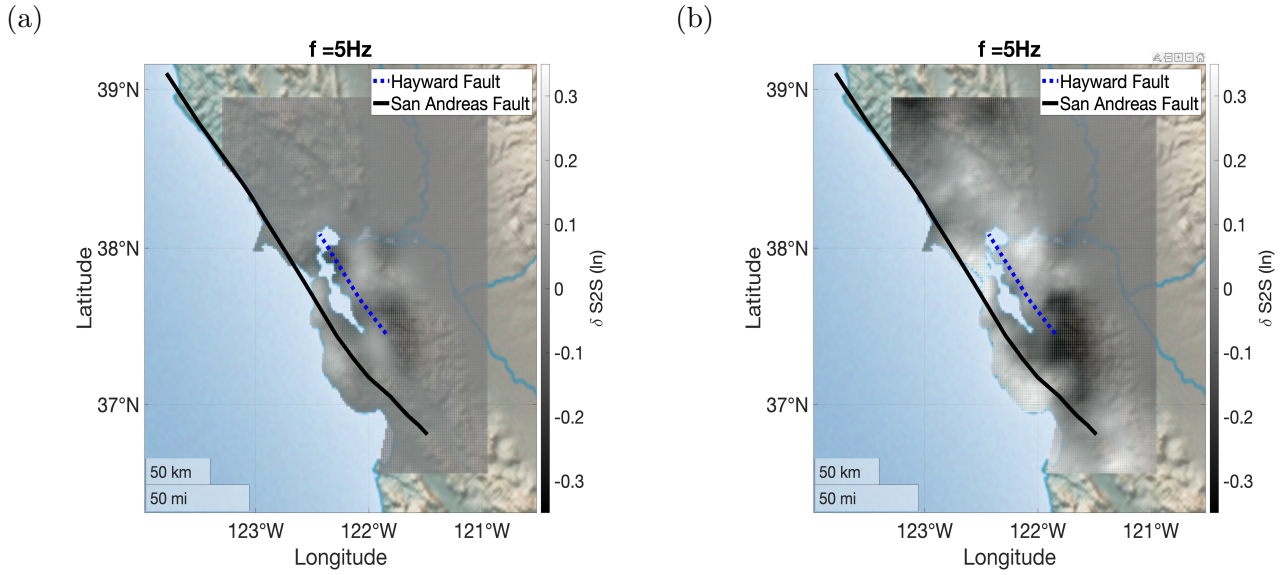


Figure 8: Maps of the FAS site term, $\delta S2S$ at 5Hz for the East-West component using the presented model (left) and from the EAS model from [Lavrentiadis *et al.* \(2023\)](#) (right).

Alt-text: Comparison maps of median non-ergodic site terms at 5 Hz, East–West component. The left map (this study) shows positive site effects along the western coastal margin near San Francisco and negative effects inland in the East Bay. The right map ([Lavrentiadis et al. 2023](#)) shows broadly similar but less spatially detailed site effects.

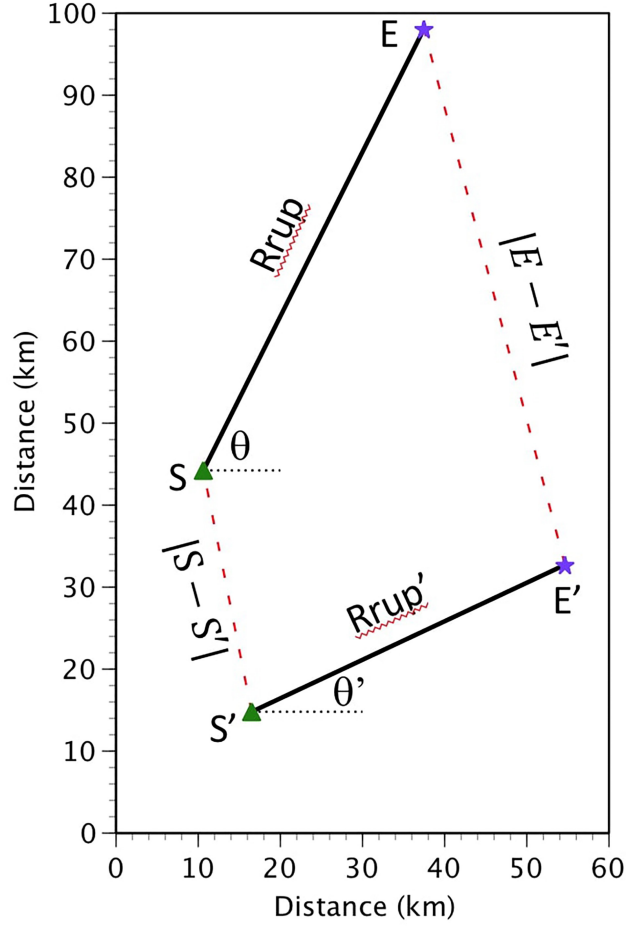


Figure 9: Illustration of the parameterization of the ray paths by their origin (site locations S), azimuth (θ) from the site to the source (earthquake location E), and length of the ray path ($Rrup$). The separation distance between the two paths uses the distance between the two sites, $|S - S'|^2$, the difference between the two azimuths, $(\theta - \theta')^2$, and the difference between the two path lengths, $(Rrup - Rrup')^2$.

Alt-text: Conceptual schematic map illustrating how ray paths are parameterized: each path is defined by its site origin, azimuth angle from site to source, and rupture distance. Differences between paths are quantified by separation in site coordinates, azimuth, and length.

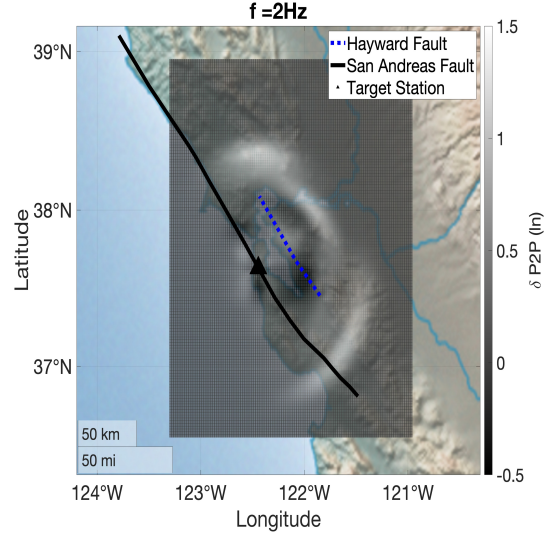


Figure 10: Maps of the median FAS path terms, $\delta P2P$, for site 55 at 2Hz for the East-West component. The site location is shown by the black triangle.

Alt-text: Map of median non-ergodic path terms for site 55 (black triangle) at 2 Hz, East-West component. Warm colors indicate stronger-than-average amplitudes along certain paths, especially from sources on the east side of the Hayward fault, while cooler colors indicate weaker-than-average paths.

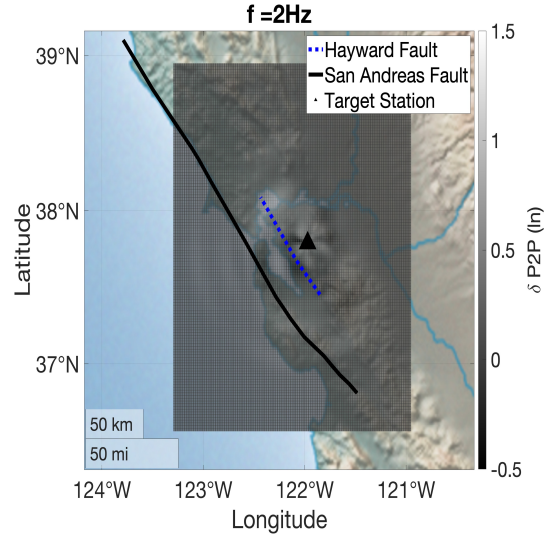
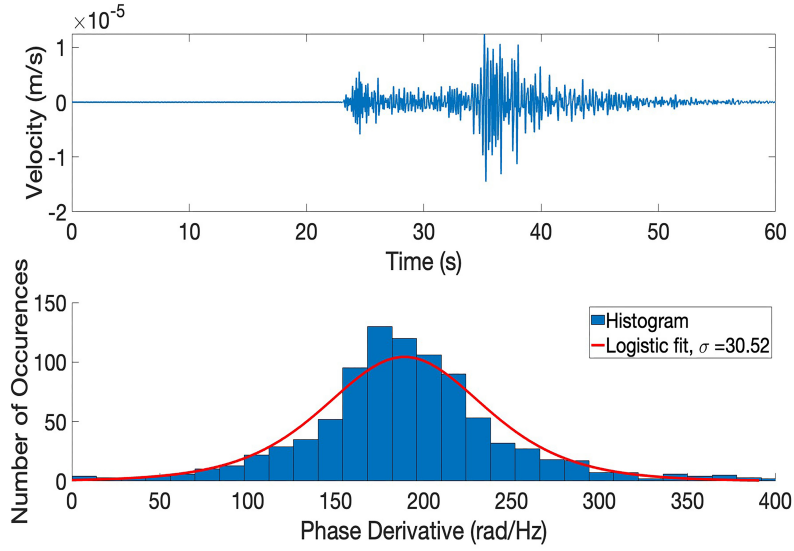


Figure 11: Maps of the median FAS path terms, $\delta P2P$, for site 87 at 2Hz for the East-West component. The site location is shown by the black triangle.

Alt-text: Map of median non-ergodic path terms for site 87 (black triangle) at 2 Hz, East-West component. The strongest positive path effects are concentrated in the southern Hayward fault region, while other paths show smaller or negligible adjustments.

(a)



(b)

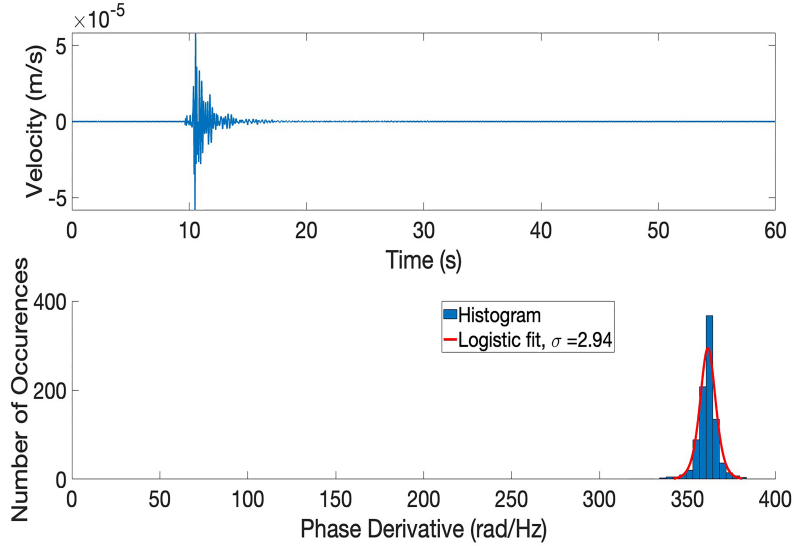


Figure 12: Examples of the velocity time series and the distribution of the phase derivative over 2-15 Hz. The standard deviation of the phase derivative reflects the duration of the time series. (a) Longer duration case due to larger distances ($M=3.3$, $R_{rup}=95$ km). (b) Shorter duration case due to smaller distances ($M=1.7$, $R_{rup}=5$ km).

Alt-text: Two examples of velocity seismograms with associated phase-derivative distributions. (a) Longer-duration time series ($M = 3.3$, $R_{rup} = 95$ km), where the broader distribution of phase derivatives indicates longer duration. (b) Shorter-duration case ($M = 1.7$, $R_{rup} = 5$ km), with a narrower distribution corresponding to shorter duration.

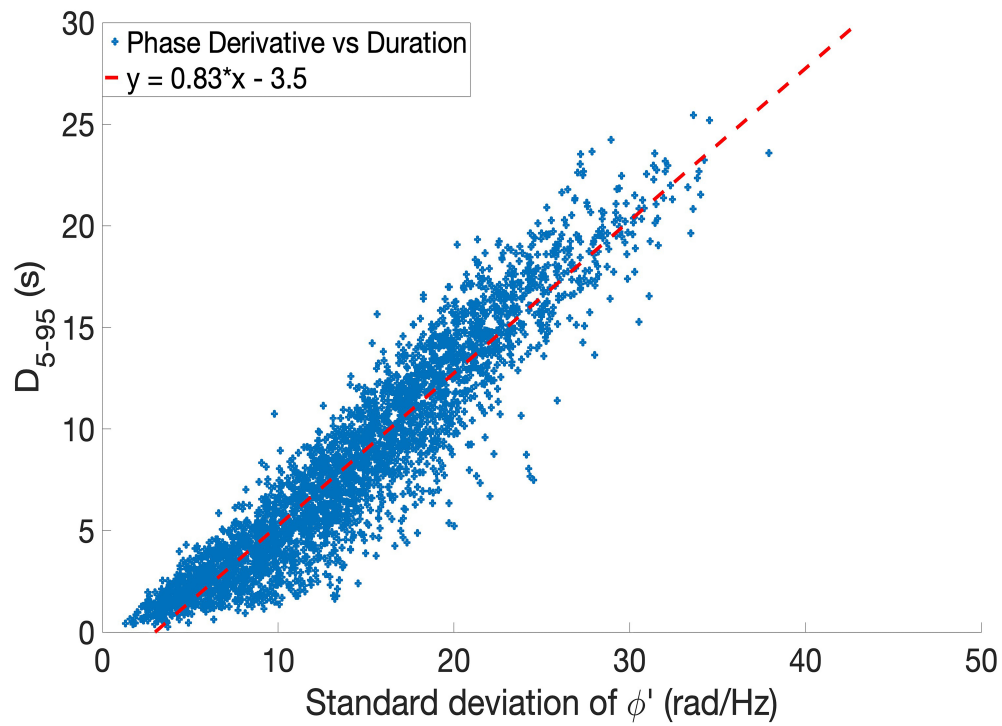


Figure 13: Relation between the standard deviation of the phase derivative and the D_{5-95} Duration for the East-West component.

Alt-text: Scatter plot of standard deviation of the phase derivative versus D5–95 duration for the East–West component. Data points fall along a linear positive trend, confirming that phase-derivative variability is a good proxy for duration.

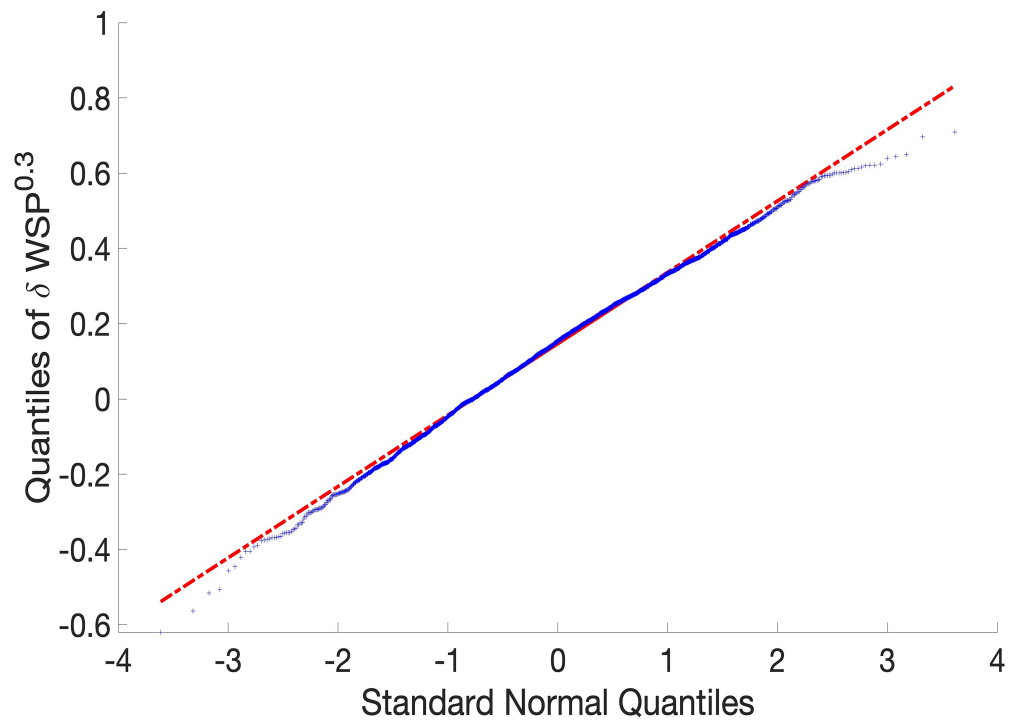


Figure 14: Quantile-quantile plot of the quantiles of δWSP versus the theoretical quantile values from a normal distribution.

Alt-text: Quantile-quantile (Q-Q) plot comparing residuals within-site-path residuals against a theoretical normal distribution. Points closely follow the 1:1 line, indicating that residuals are approximately normally distributed.

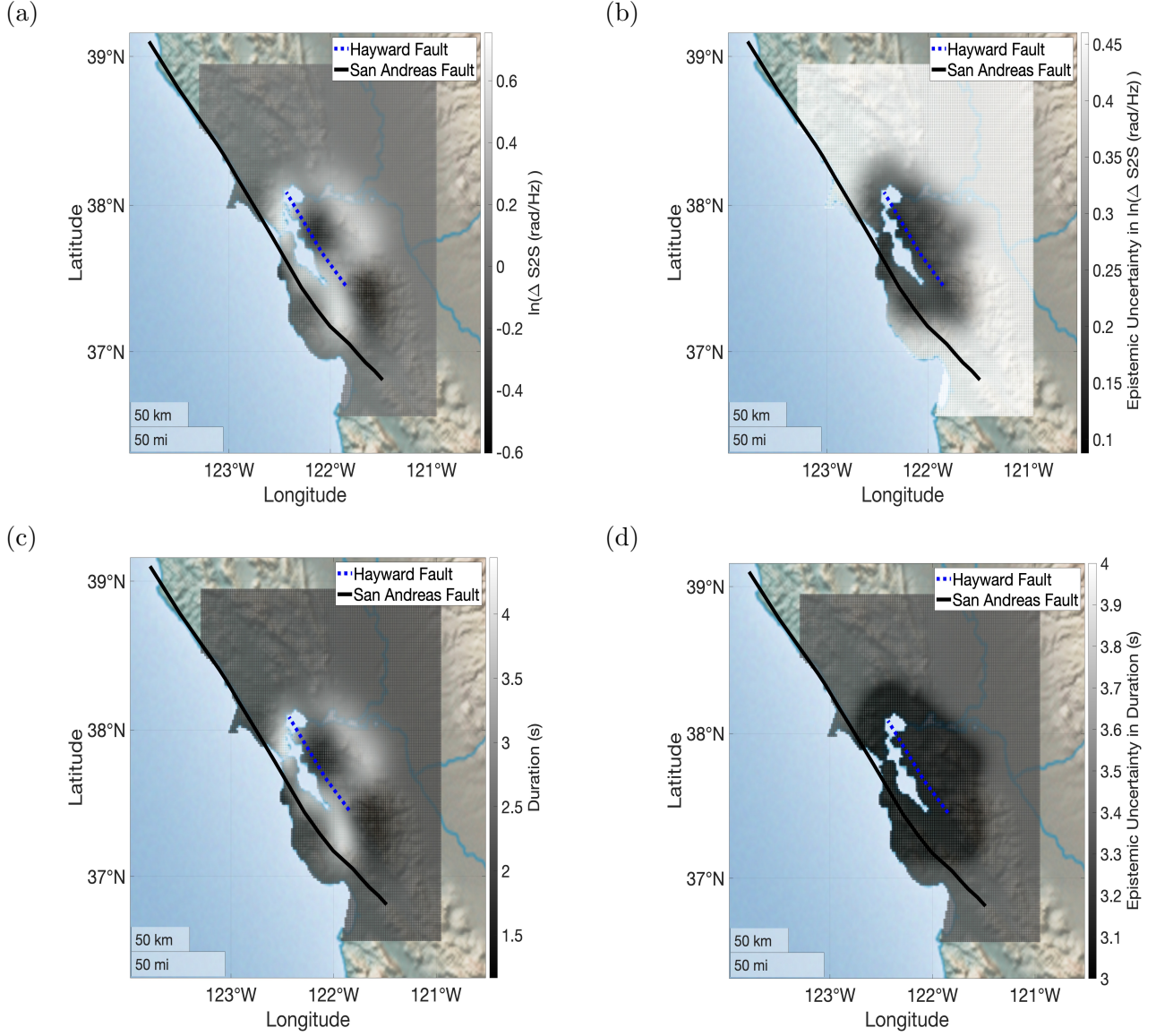


Figure 15: Map of the median site term (a) and the corresponding epistemic uncertainty (b) for the standard deviation of the phase derivative for the East-West component, converted to median duration (c) and epistemic uncertainty in duration (d) using their linear relationship: $D_{5-95} = 0.83\sigma_{\phi'} - 3.5$.

Alt-text: Maps of non-ergodic site effects on duration for the East–West component. Panels (a) and (b) show the median site term for the standard deviation of the phase derivative and its epistemic uncertainty. Panels (c) and (d) convert these values into predicted D5–95 durations, showing longer durations at western coastal sites and shorter durations inland, with uncertainty growing farther from instrumented sites.

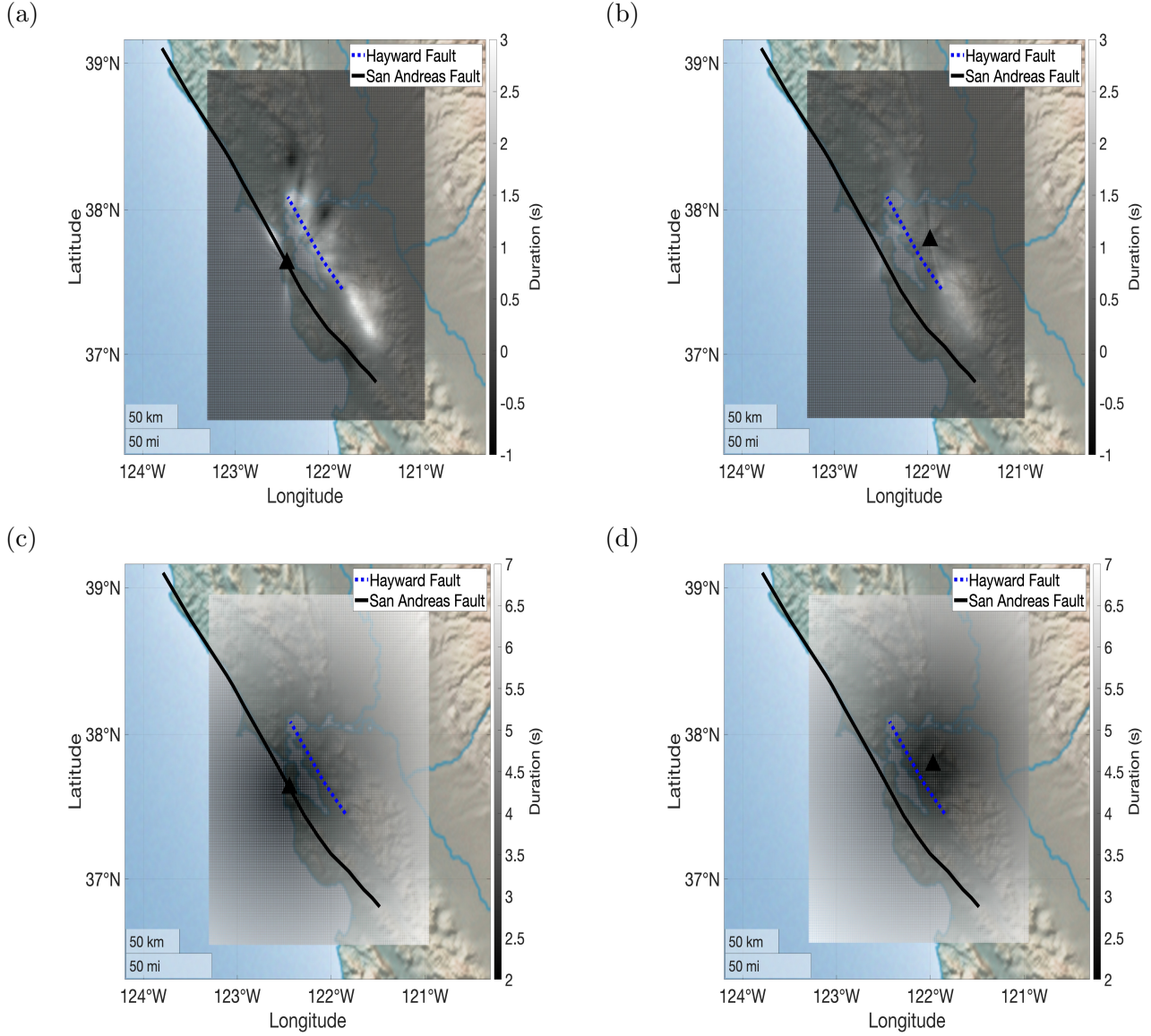


Figure 16: Maps of the median path terms and of the epistemic uncertainty of the phase derivative for the East-West component, converted to duration using their linear relationship. The site locations are shown by the black triangles. (a) Median path terms for Station 55, (b) Median path terms for Station 87, (c) Epistemic uncertainty in median path terms for Station 55, (d) Epistemic uncertainty in median path terms for Station 87.

Alt-text: Maps of non-ergodic path effects on duration for the East–West component, shown for two example stations (black triangles). Panels (a) and (b) display the median path-induced duration adjustments: Site 55 shows longer durations for paths from the south Hayward fault, while Site 87 shows moderate increases in the same region. Panels (c) and (d) show epistemic uncertainty, which increases with source distance due to path correlation limits.

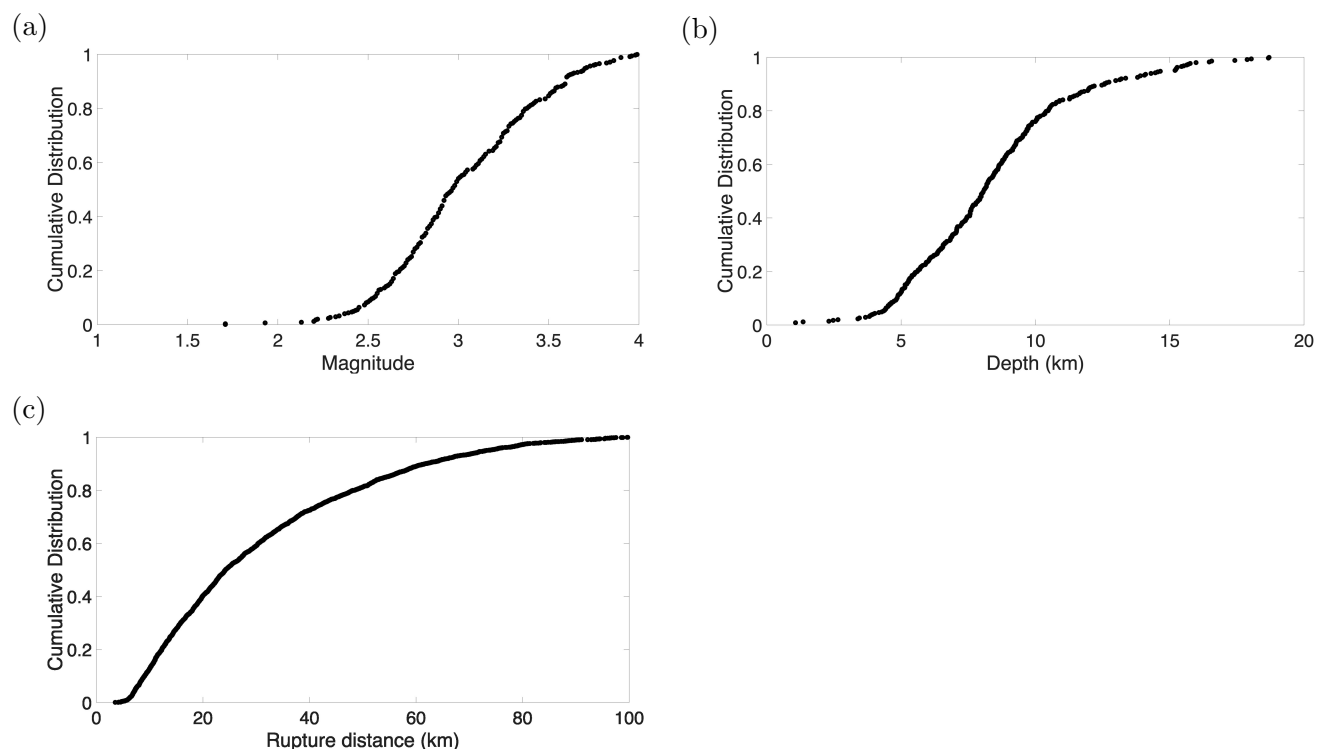


Figure S1: Cumulative distribution of events with magnitude (a), depth (b) and rupture distance (c)

Alt-text: Three cumulative distribution plots showing: (a) Rate of occurrence versus magnitude from 1 to 4, displaying a sigmoid curve rising from near 0 to 1; (b) Rate of occurrence versus depth in kilometers from 0 to 20, showing a similar sigmoid pattern; and (c) Rate of occurrence versus rupture distance in kilometers from 0 to 100, also exhibiting a sigmoid curve. All three plots demonstrate smooth cumulative distribution functions approaching 1.0.

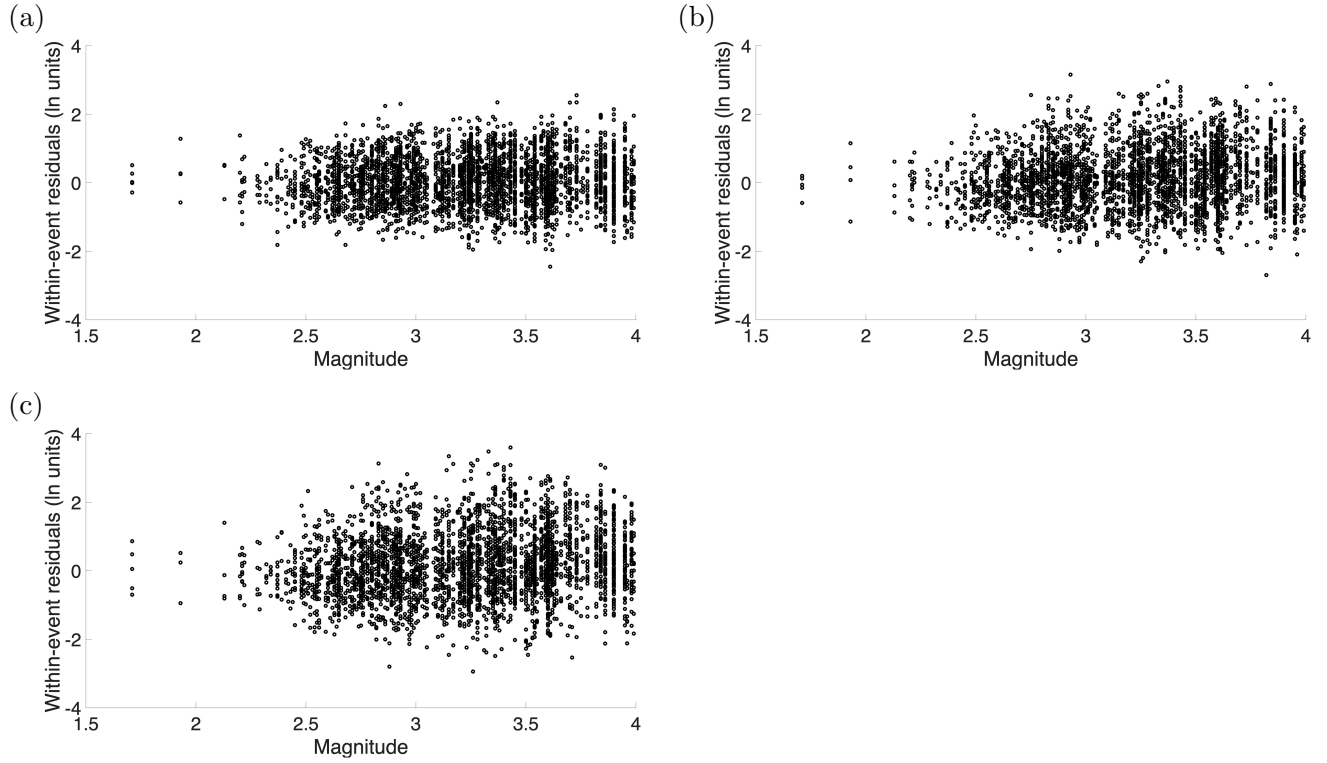


Figure S2: Within-event residuals with magnitude at 2hz (a), 5Hz (b), and 10Hz (c)

Alt-text: Three scatter plots displaying within-event residuals plotted against earthquake magnitude for frequencies of 2Hz (panel a), 5Hz (panel b), and 10Hz (panel c). Each plot shows data points across magnitudes ranging from 1 to 4, with residuals spanning from -4 to 4 ln units. The data exhibit dense concentration with no clear magnitude-dependent trend.

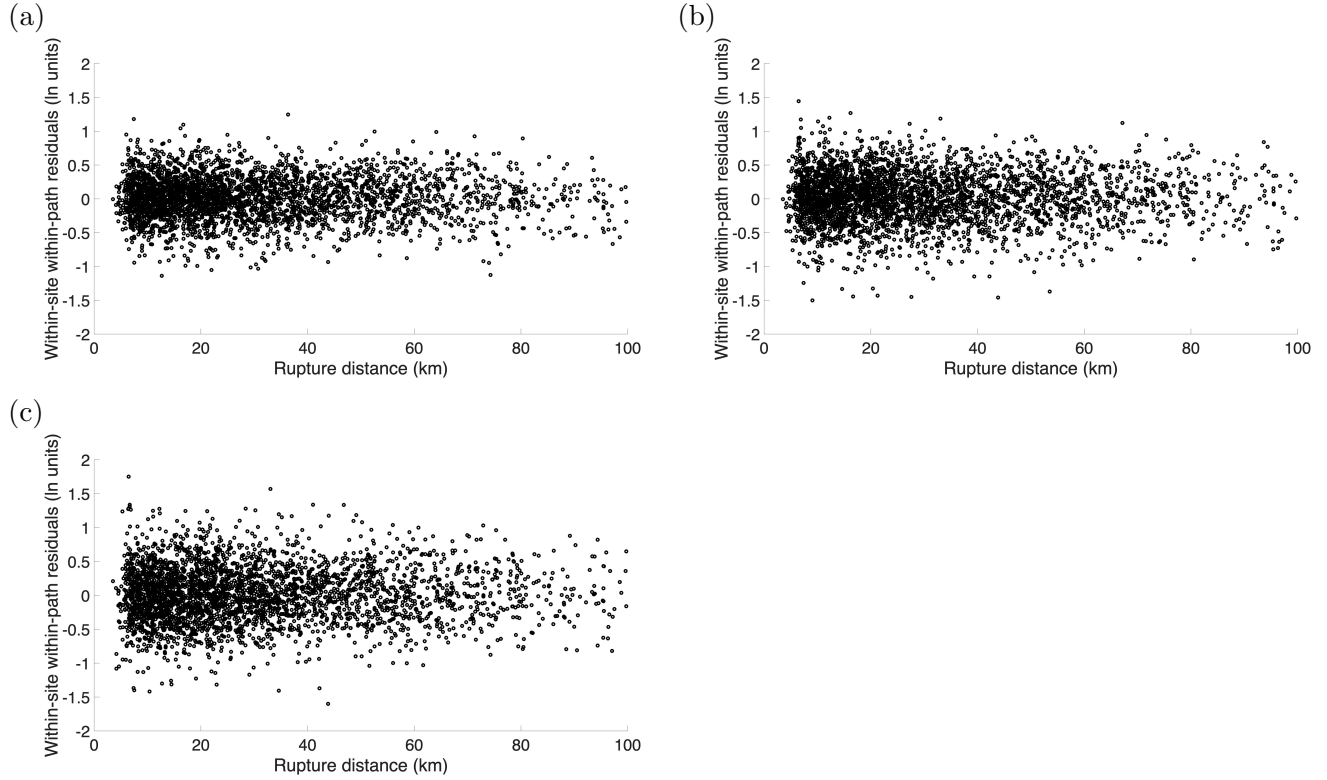


Figure S3: Within-site within-path residuals with rupture distance at 2hz (a), 5Hz (b), and 10Hz (c)

Alt-text: Three scatter plots showing within-site within-path residuals plotted against rupture distance in kilometers for frequencies of 2Hz (panel a), 5Hz (panel b), and 10Hz (panel c). Each plot shows dense clustering of data points near zero residuals at shorter distances (0-30 km) with increasing scatter at greater distances up to 100 km.

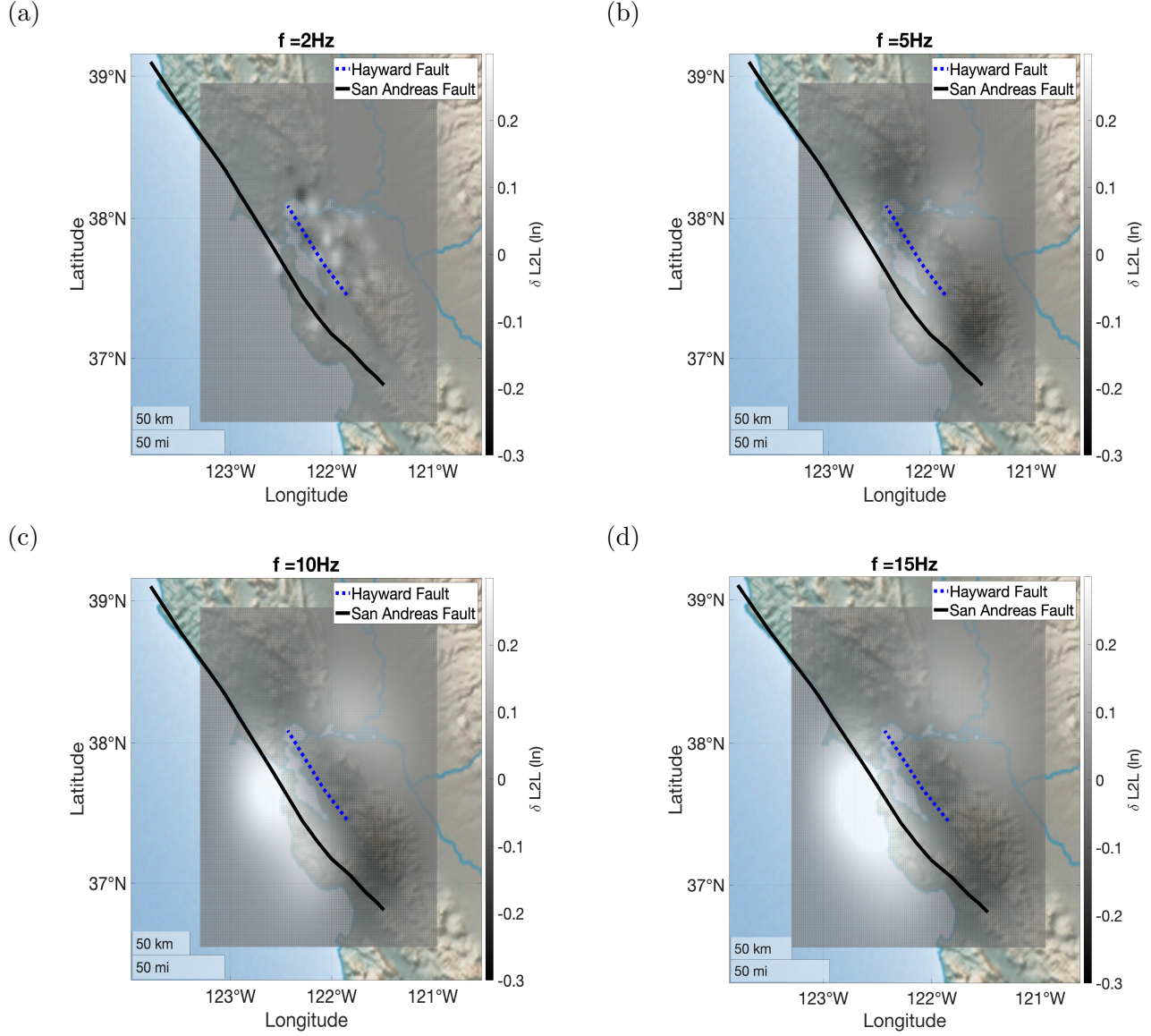


Figure S4: Maps of the median FAS source terms, $\delta L2L$, at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component.

Alt-text: Four maps of median source terms for the East–West component at 2, 5, 10, and 15 Hz. Positive values cluster near the San Andreas fault, while negative values cluster near the Hayward fault. Spatial patterns broaden with frequency due to increasing correlation length.

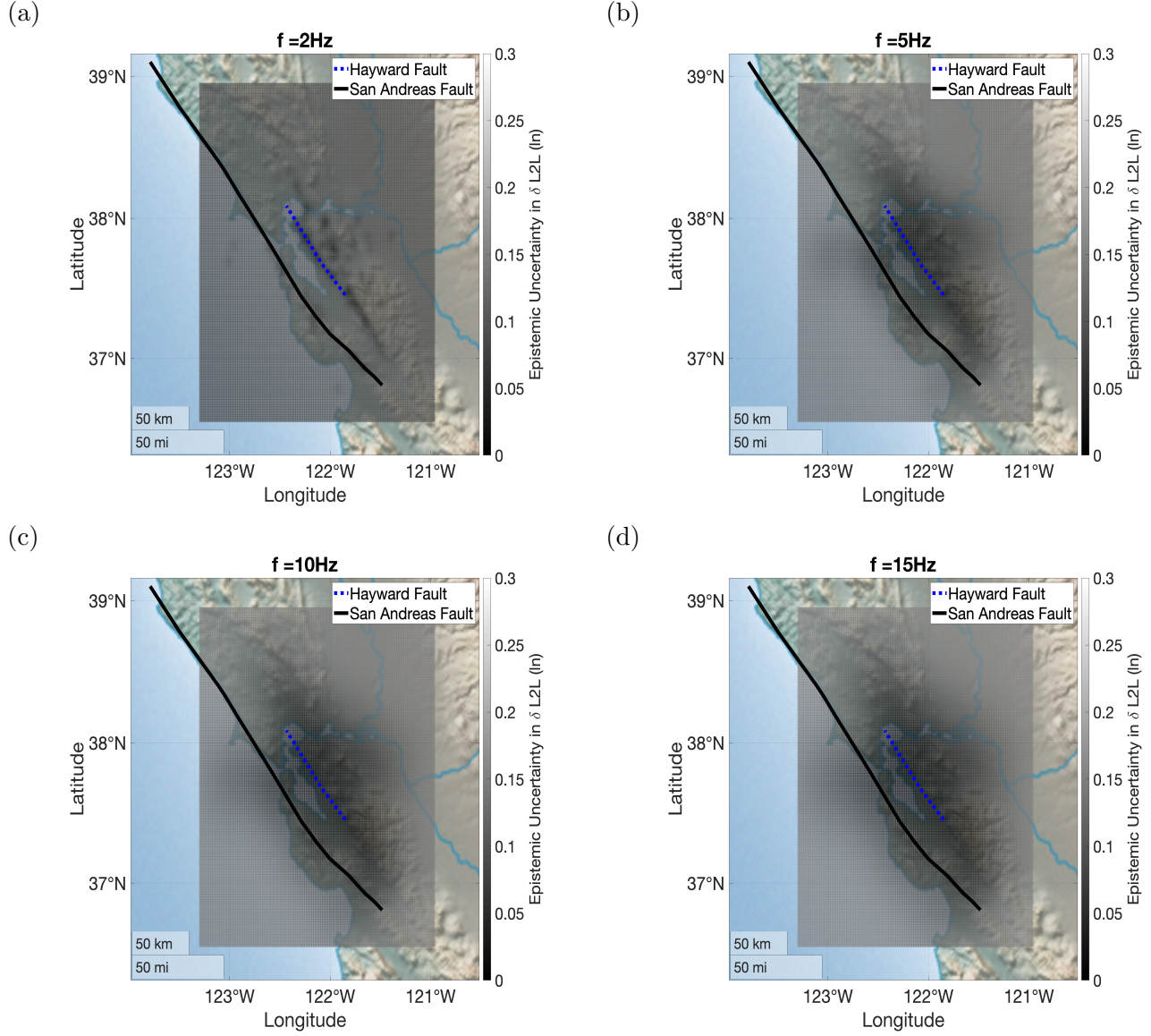


Figure S5: Maps of the epistemic uncertainty in the median FAS source terms, $\delta L2L$, at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component.

Alt-text: Corresponding maps of epistemic uncertainty in the median source terms (East–West component) at 2, 5, 10, and 15 Hz. Uncertainty is lowest near observed events and increases in regions with sparse earthquake coverage.

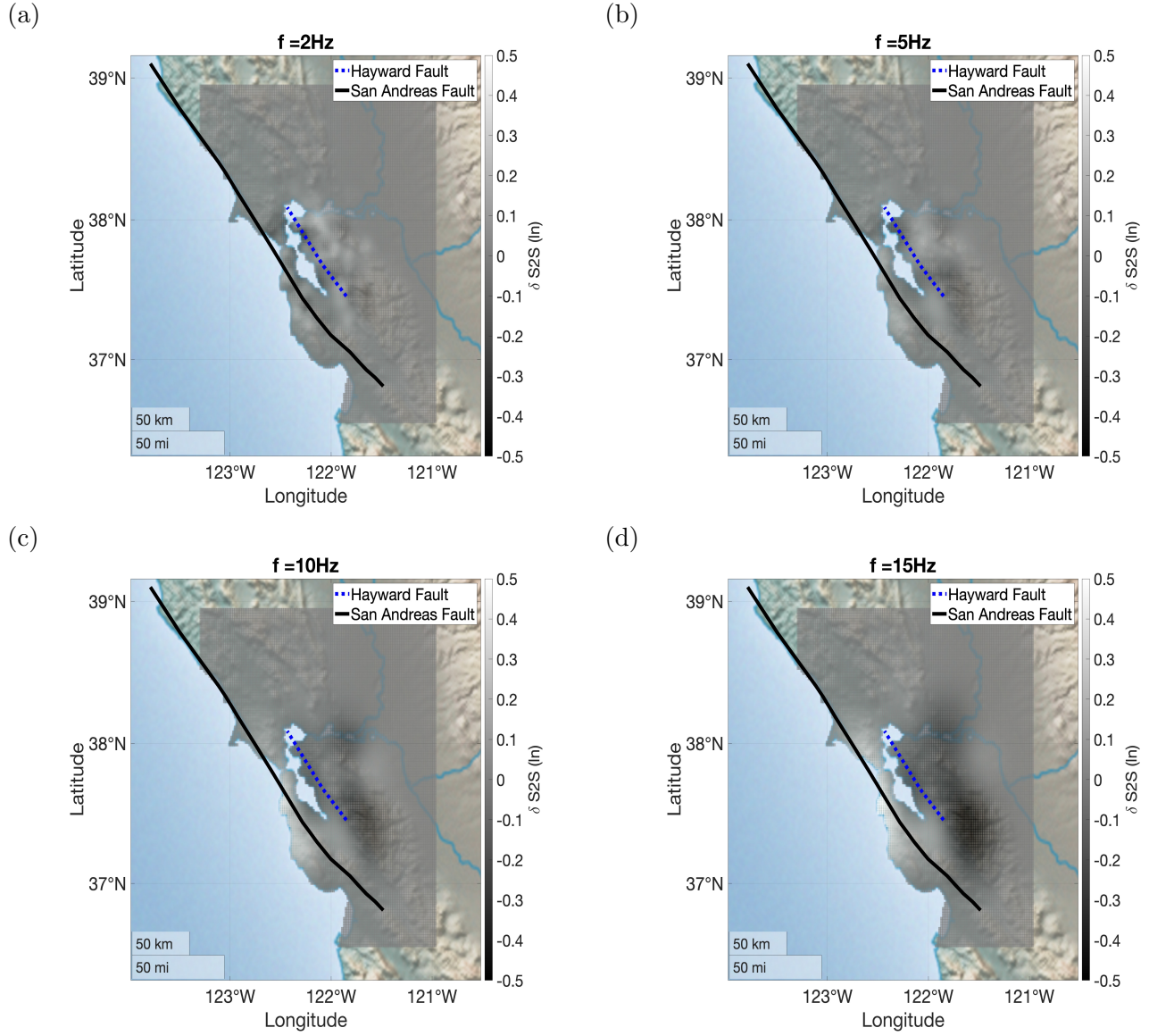


Figure S6: Maps of the FAS site term, $\delta S2S$, at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component.

Alt-text: Four maps of median site terms for the East–West component at 2, 5, 10, and 15 Hz. Coastal sites west of San Francisco exhibit strong positive adjustments, while East Bay sites show negative adjustments.

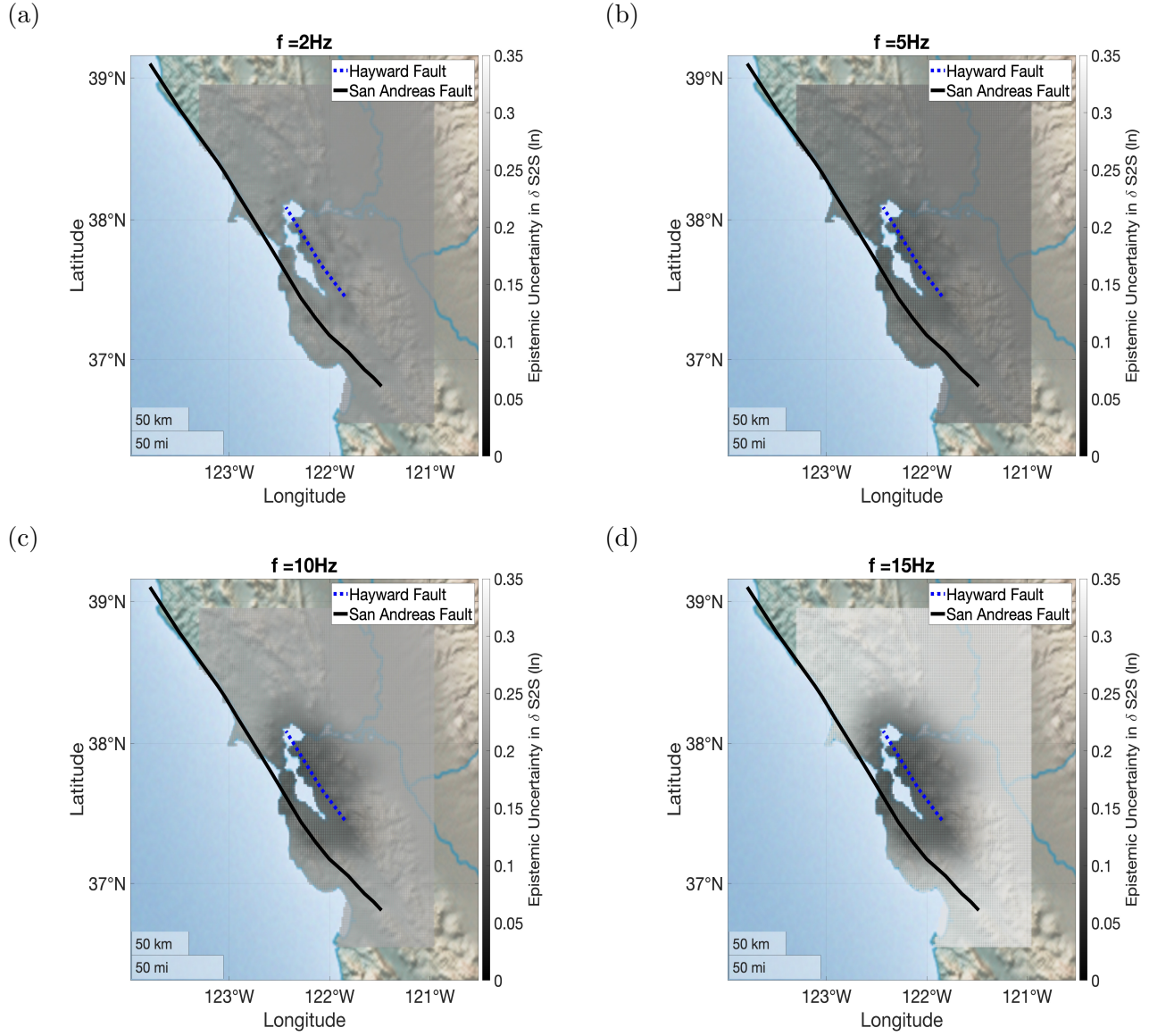


Figure S7: Maps of the epistemic uncertainty in the median FAS site terms, $\delta S2S$, at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component.

Alt-text: Corresponding maps of epistemic uncertainty in the median site term at 2, 5, 10, and 15 Hz, showing lowest uncertainty at recorded sites and highest uncertainty in unrecorded sites in inland or offshore regions.

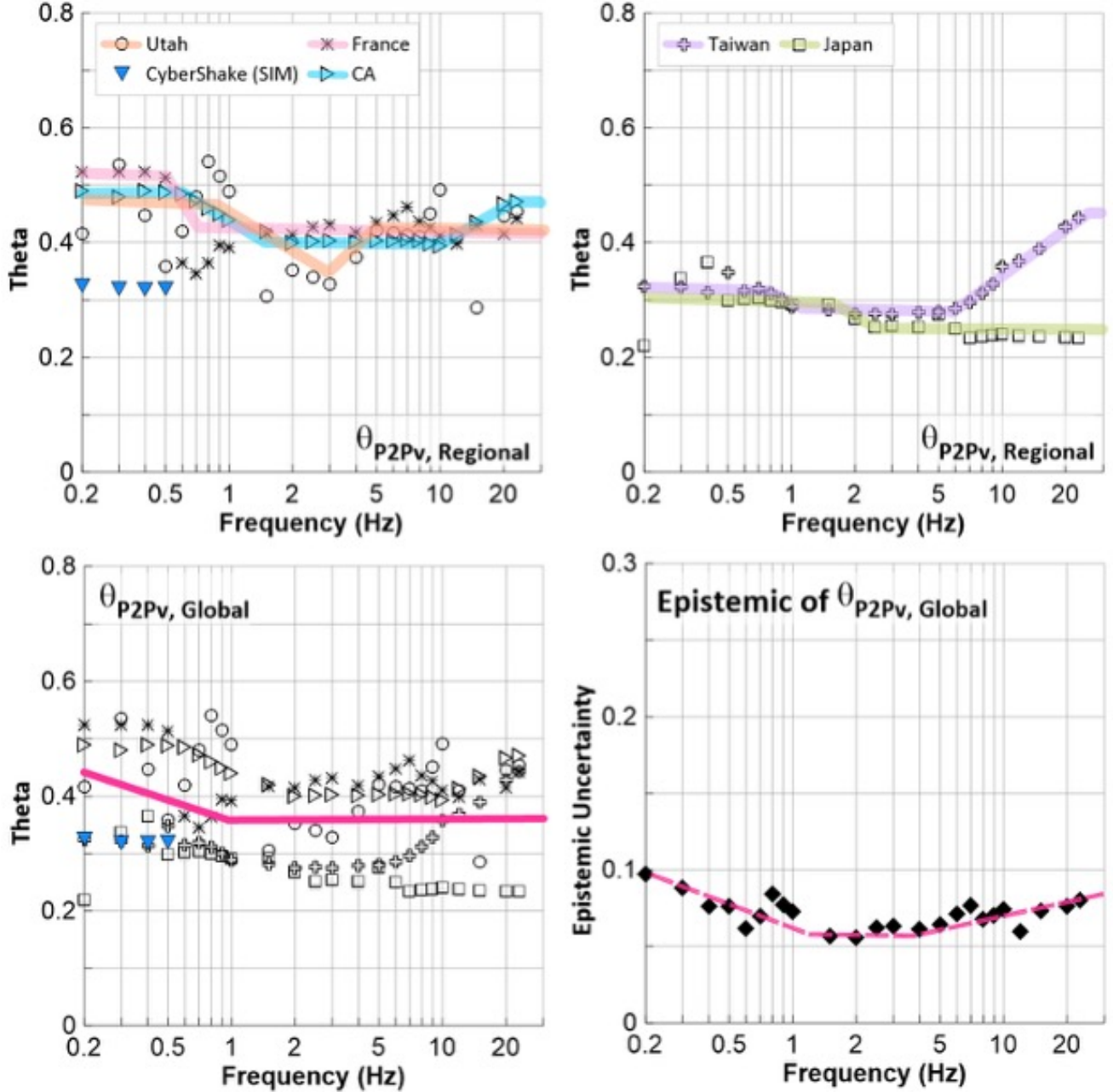


Figure S8: Comparison of non-ergodic path terms ϕ_{P2P} from different regions from [Sung and Abrahamson \(2024\)](#). (ϕ_{P2P} is noted as θ_{P2P} in the notations from [Sung and Abrahamson \(2024\)](#))
 Alt-text: Four-panel comparison of non-ergodic path terms ϕ_{P2P} across frequency ranges from 0.2 to 20 Hz. Top left panel shows theta values for Utah, France, CyberShake simulation, and California datasets with regional corrections, displaying relatively flat trends around 0.4-0.5. Top right panel compares Taiwan and Japan data with regional corrections, showing an increase from approximately 0.25 at low frequencies to 0.45 at high frequencies. Bottom left panel displays the global ϕ_{P2P} correction as a flat magenta line at approximately 0.35 with scattered data points from all regions. Bottom right panel shows epistemic uncertainty of the global correction, with values ranging from 0.05 to 0.1 across frequencies, represented by black diamond markers.

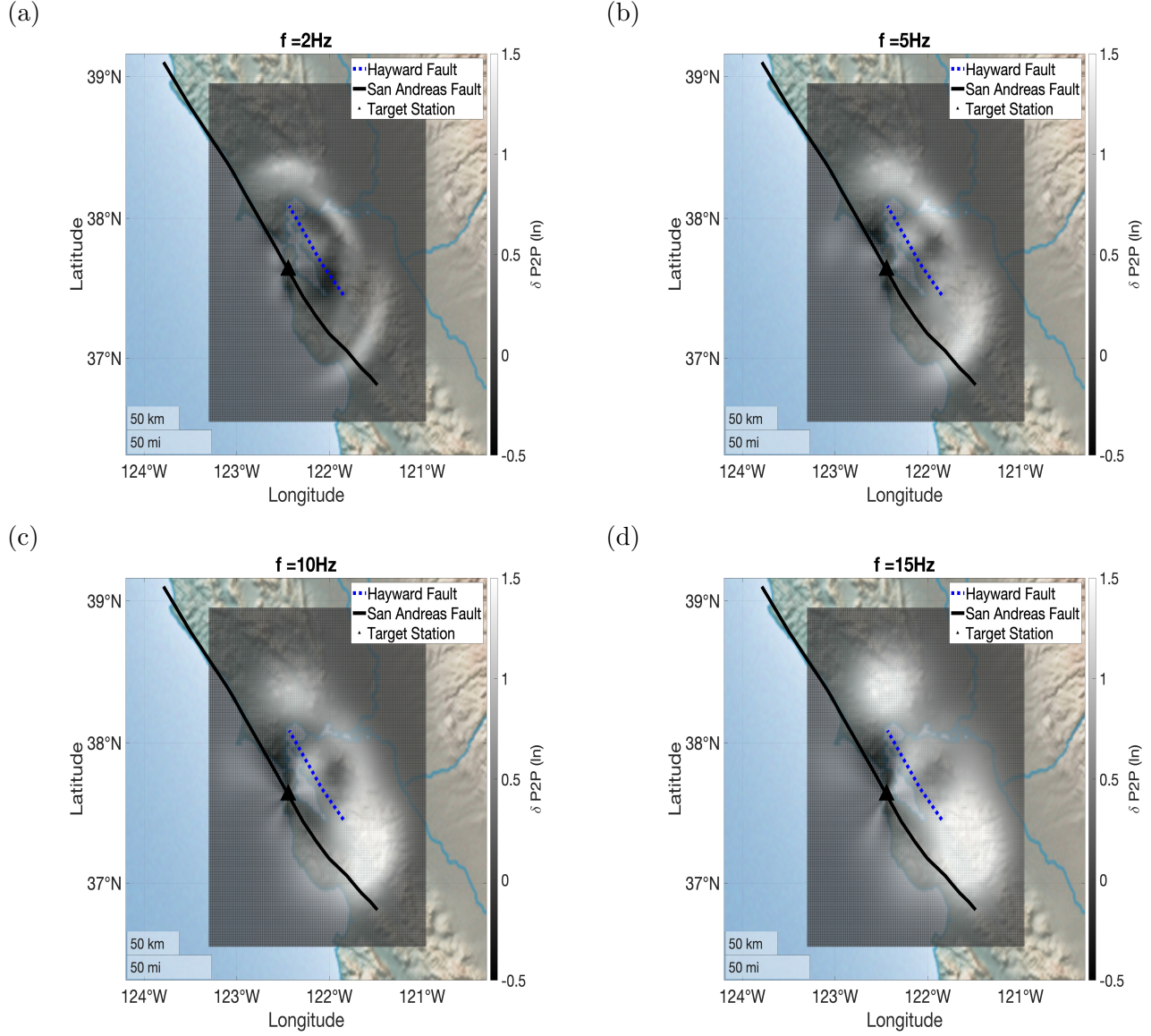


Figure S9: Maps of the median FAS path terms, $\delta P2P$, for site 55 at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component. The site location is shown by the black triangle.
 Alt-text: Four maps of median path terms for site 55 at 2, 5, 10, and 15 Hz (East–West component). The black triangle marks site 55, with positive adjustments for paths from the Hayward fault region and smaller effects for San Andreas fault sources.

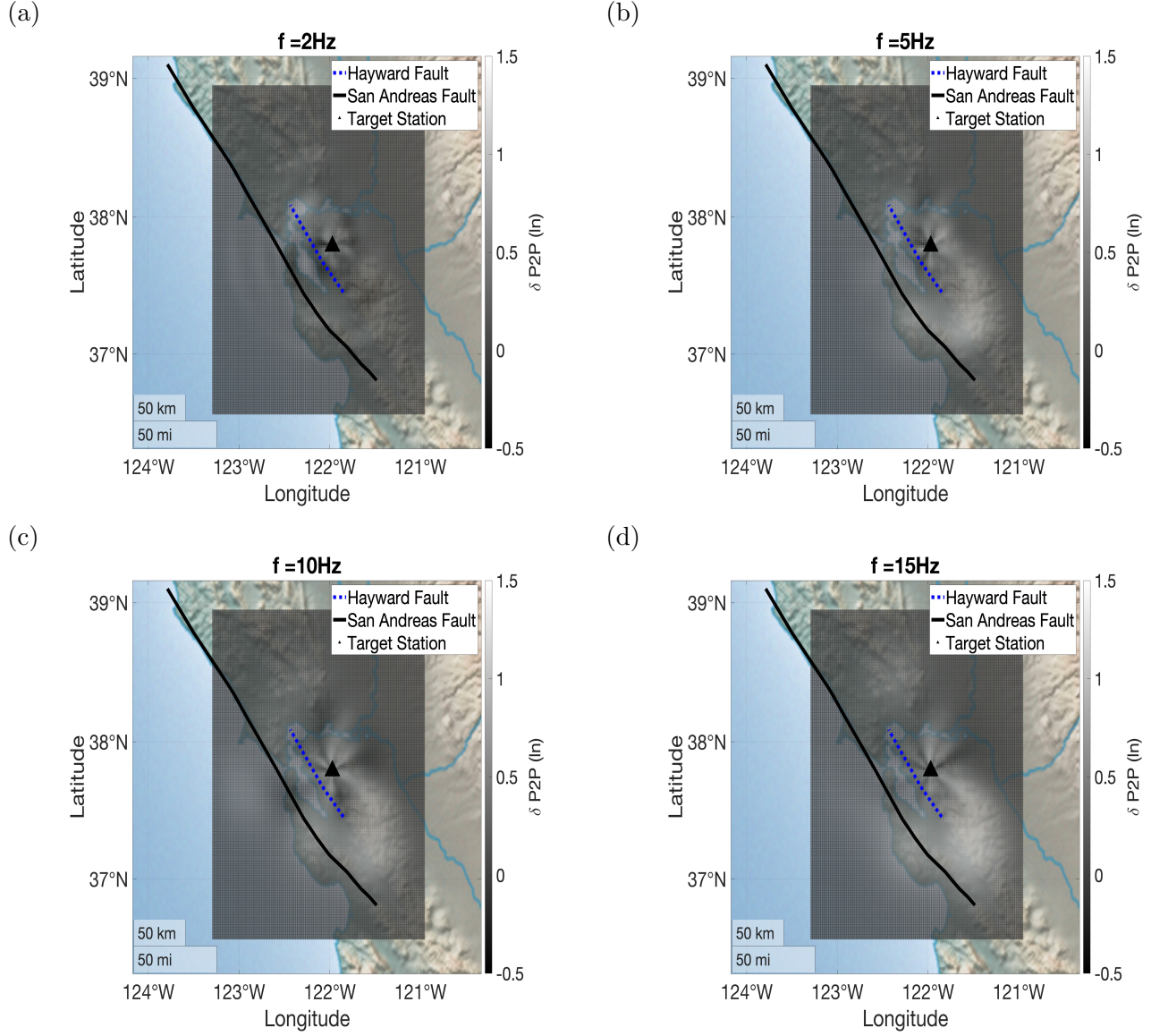


Figure S10: Maps of the median FAS path terms, $\delta P2P$, for site 87 at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component. The site location is shown by the black triangle.
 Alt-text: Four maps of median path terms for site 87 at 2, 5, 10, and 15 Hz (East–West component). The black triangle marks site 87. Maps show site 87 receiving stronger path effects from southern Hayward fault sources compared to other directions.

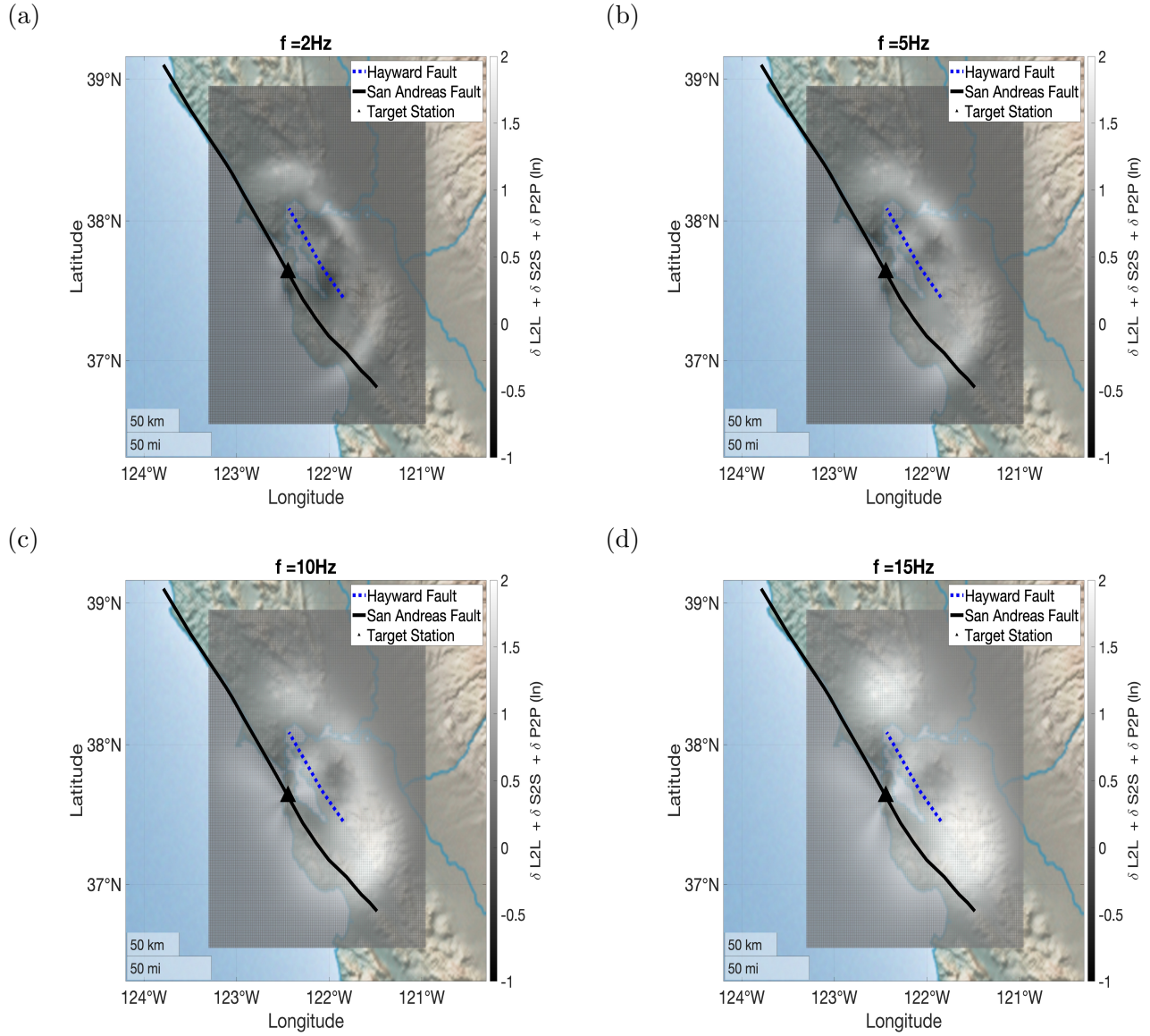


Figure S11: Maps of the net adjustment of the non-ergodic median FAS, $\delta L2L + \delta S2S + \delta P2P$, for site 55 at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component. The site location is shown by the black triangle.

Alt-text: Maps of the net non-ergodic median adjustment (source + site + path) at 2, 5, 10, and 15 Hz for site 55 (East–West component). Patterns combine the positive San Andreas source effects, coastal site effects, and directional path effects.

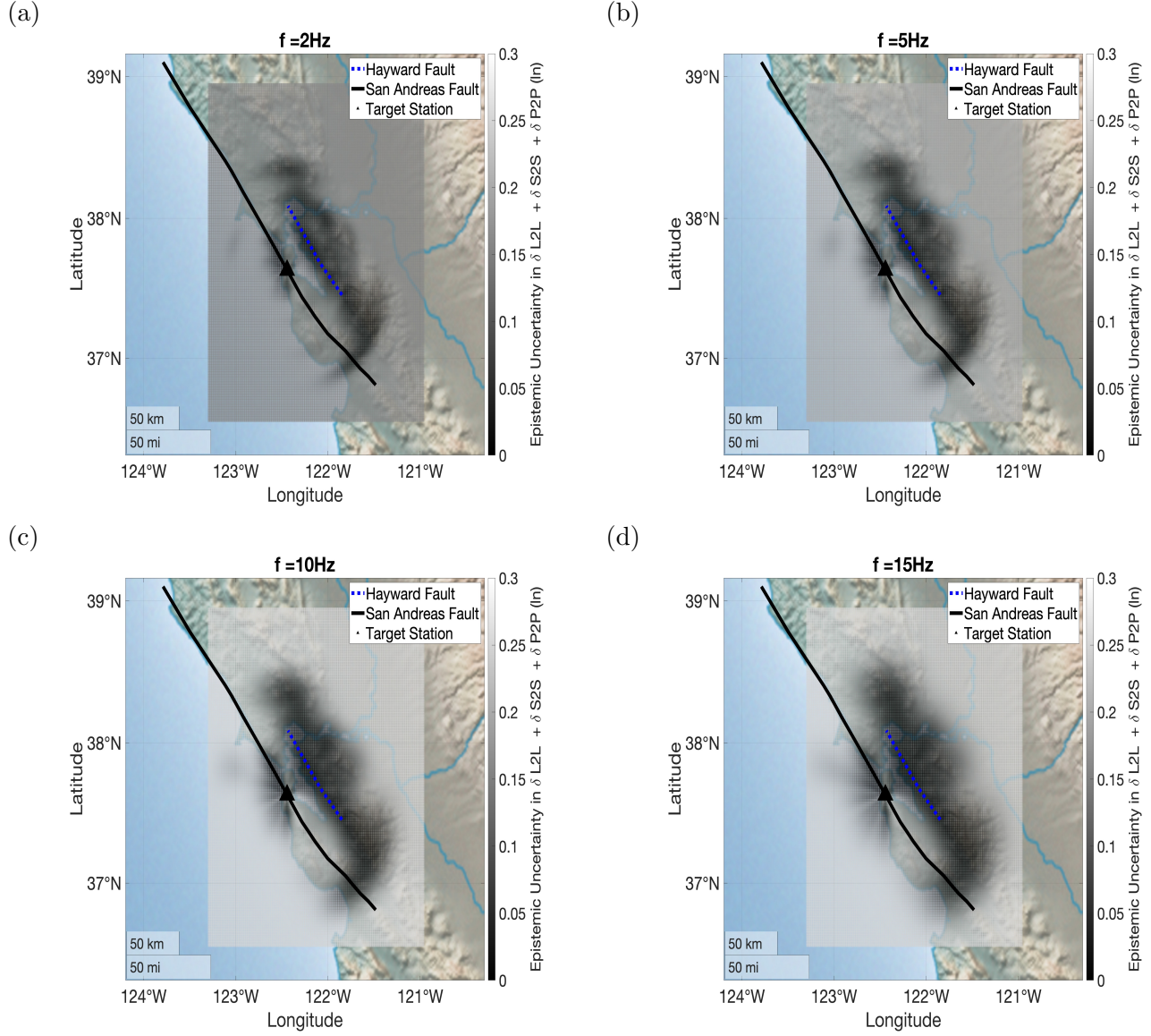


Figure S12: Maps of the epistemic uncertainty in the net adjustment of the non-ergodic median FAS, $\delta L2L + \delta S2S + \delta P2P$, for site 55 at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component. The site location is shown by the black triangle.

Alt-text: Corresponding maps of epistemic uncertainty in the net adjustment for site 55 at 2, 5, 10, and 15 Hz. Uncertainty grows with distance from unobserved source locations.

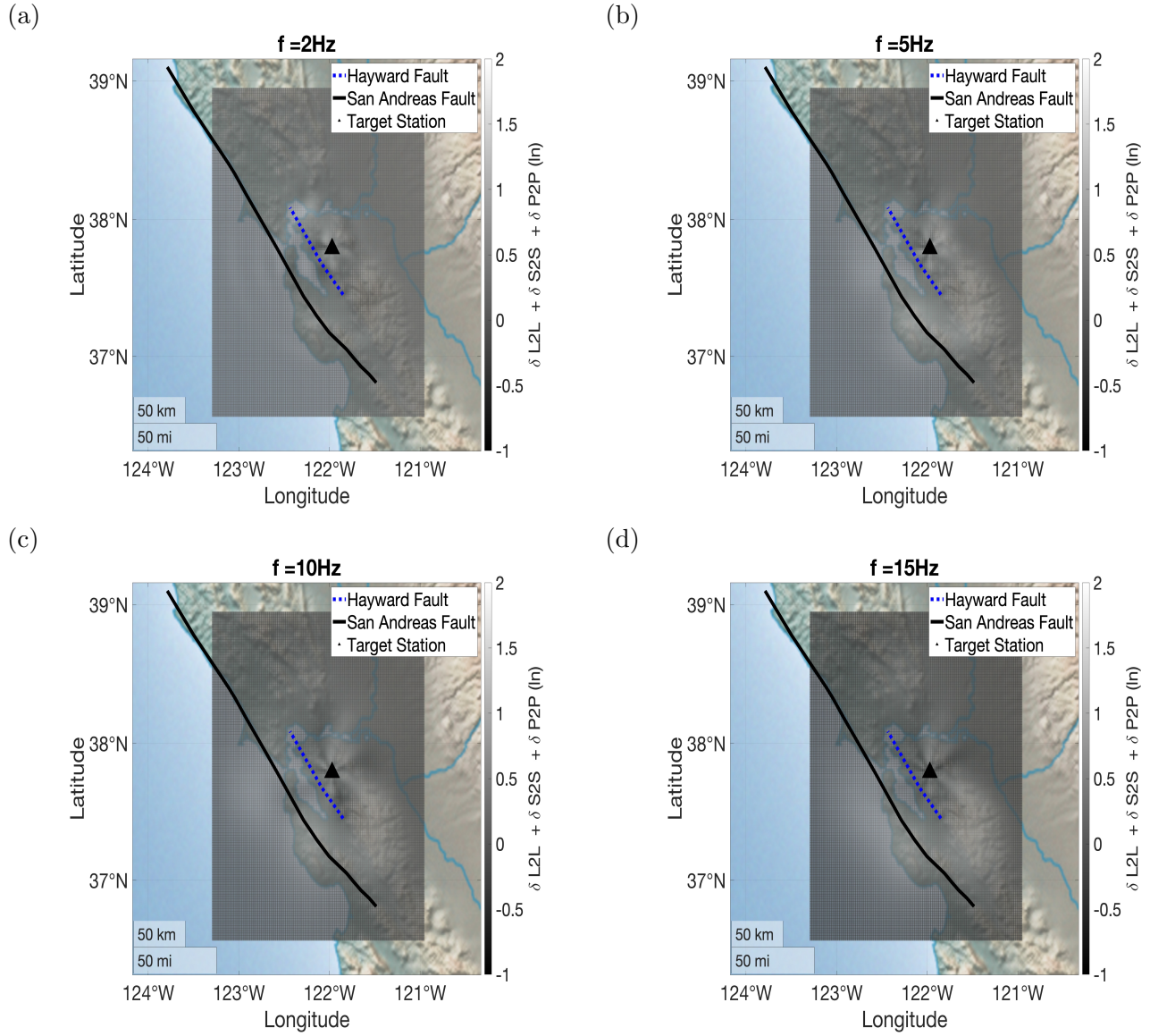


Figure S13: Maps of the net adjustment of the non-ergodic median FAS, $\delta L2L + \delta S2S + \delta P2P$, for site 87 at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component. The site location is shown by the black triangle.

Alt-text: Maps of net non-ergodic median adjustment for site 87 (East–West component) at 2, 5, 10, and 15 Hz, showing moderate positive effects from southern Hayward sources and weaker adjustments elsewhere.

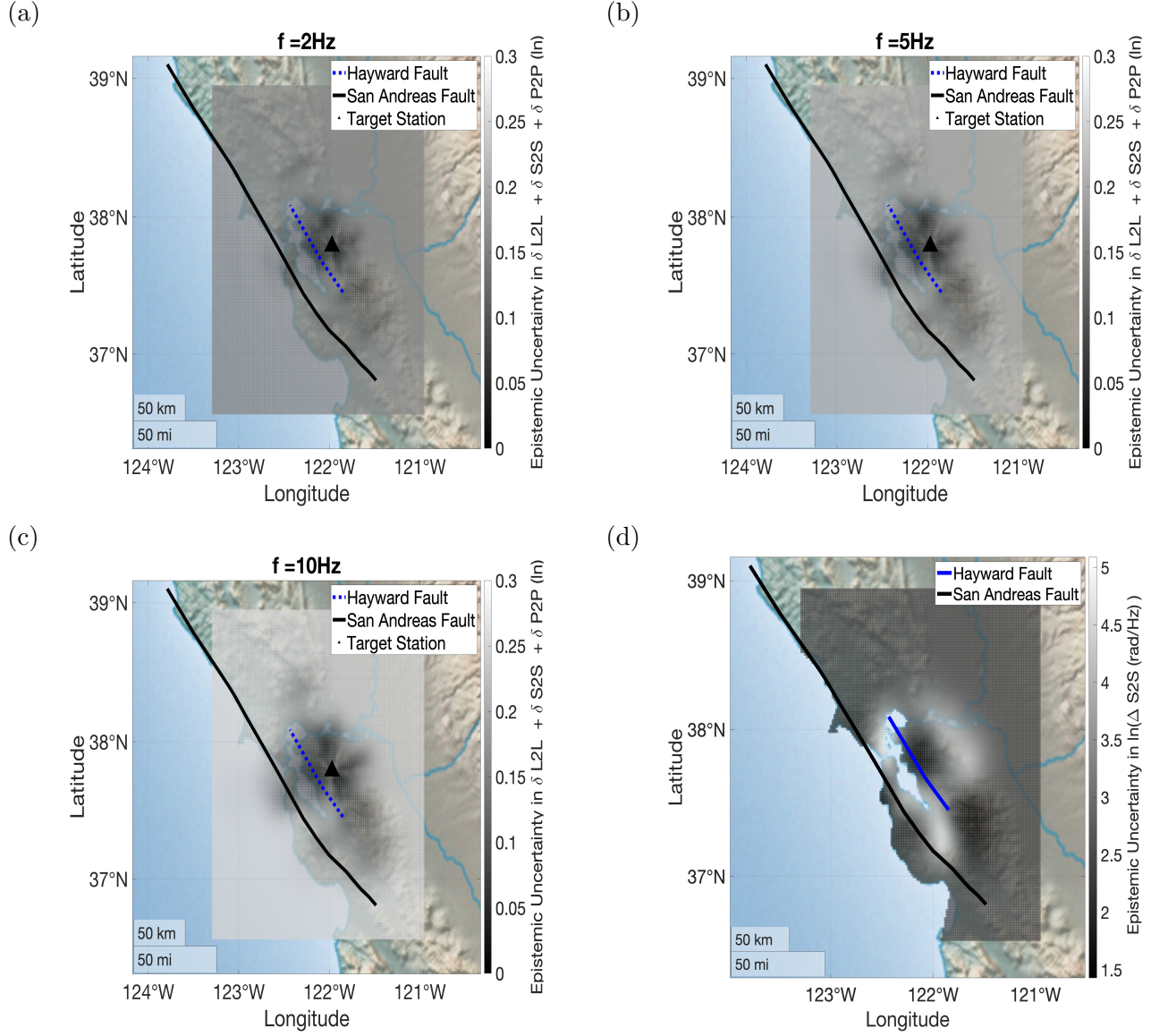


Figure S14: Maps of the epistemic uncertainty in the net adjustment of the non-ergodic median FAS, $\delta L2L + \delta S2S + \delta P2P$, for site 87 at 2Hz, 5Hz, 10Hz and 15Hz for the East-West component. The site location is shown by the black triangle.

Alt-text: Corresponding maps of epistemic uncertainty in net adjustment for site 87, with higher uncertainty away from the dataset's core ray path coverage.