Ground Motion Residuals, Path Effects, and Crustal Properties: A Pilot Study in Southern California

V. J. Sahakian1,2, A. Baltay1, T. C. Hanks1, J. Buehler2, F. L. Vernon1, D. Kilb3, and N. A. Abrahamson4

1Earthquake Science Center, United States Geological Survey, Menlo Park, CA, USA, 2Now at Department of Earth Sciences, University of Oregon, Eugene, OR, USA, 3Institute for Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA, 4Civil and Environmental Engineering, University of California, Berkeley, CA, USA

Abstract To improve models of ground motion estimation and probabilistic seismic hazard analyses, the engineering seismology field is moving toward developing fully nonergodic ground motion models, models specific for individual source-to-site paths. Previous work on this topic has examined systematic variations in ground-motion along particular paths (from either recorded or simulated earthquake data) and has not included physical properties of the path. We present here a framework to include physical path properties, by seeking correlations between ground motion amplitudes along specific paths and crustal properties, specifically seismic velocity and anelastic attenuation, along that path. Using a large data set of small-magnitude earthquakes recorded in Southern California, we find a correlation between the gradient of seismic $S$ wave velocity and the path term residual, after accounting for an average geometric spreading and anelastic attenuation, indicating that heterogeneity in crustal velocity primarily controls the path-specific attenuation. Even in aseismic regions, details of path-specific ground motion prediction equations can be developed from crustal structure and property data.

1. Introduction

Large uncertainties in ground motion estimation have long afflicted seismic hazard analysis and can lead to the overprediction of ground motion intensity measures at low probabilities of exceedance. This low probability of exceedance is of concern in particular to critical facilities—as each path must be described separately, but the variability is also reduced, leading to reduced hazard levels at low probabilities of exceedance. The source-to-site path, in particular, should affect ground motion amplitudes based on fundamental seismological theory.

An ergodic model is one where all data are considered as representative of the same physical phenomenon. The Next Generation Attenuation West (NGA-West2) ground motion prediction equations (GMPEs), for example, consists of ground motion data from shallow, active crustal earthquakes from California, Taiwan, Japan, and New Zealand (among others) regressed together. The resulting median GMPEs are then assumed to be representative of the average source, site, and path effects in the western United States (e.g., Bozorgnia et al., 2014). A limitation of these ergodic models is that they often do not perform on a regional or subregional scale, as their associated standard deviations are larger than should be for individual source-to-site paths. In contrast, a nonergodic model considers that each region, or more specifically, each path, is unique and potentially different from the median model. Nonergodic models are necessarily more complex, as each path must be described separately, but the variability is also reduced, leading to reduced hazard levels at low probabilities of exceedance. The source-to-site path, in particular, should affect ground motion amplitudes based on fundamental seismological theory.

In most GMPEs, the region of interest is modeled with a single geometric spreading function and anelastic attenuation function, although the observed attenuation along any one path or group of paths may be much different from the median model. Current nonergodic studies have used purely empirical and statistical methods to account for these differences (Abrahamson et al., 2017; Dawood & Rodriguez-Marek, 2013; Landwehr et al., 2016; Villani & Abrahamson, 2015). The results have been successful in reducing the overall model uncertainties; however, they do not include any physical information or understanding of the
material properties. Without any underlying physical mechanisms, these models do not necessarily extrapolate well (say to different magnitude or distance ranges) and are biased when applied to other regions (such as those with low-seismicity rates where statistical methods may fail).

The properties of the upper crust, through which the seismic waves travel, are measurable physical quantities and will remain knowable and repeatable for many earthquakes. This is unlike properties of the source, because stress drop or directivity from any one earthquake does not predict the source physics of the next event. These crustal material properties are therefore obtainable a priori and could be incorporated into nonergodic GMPEs as additional predictive information.

In this paper, we propose a physical framework to move toward path-specific, nonergodic models. In our approach, we integrate known, crustal physical properties into source-to-site, path-specific ground motion models. This is similar in nature to approaches using physical parameters or constraints in ground motion models from source properties, such as stress drop (Ameri et al., 2017; Baltay et al., 2017; Bindi et al., 2017; Kotha et al., 2016; Oth et al., 2017), or incorporation of kappa as a site parameter (e.g., Laurendeau et al., 2013). We use a dense data set from Southern California of over 123,000 peak ground acceleration (PGA) values from earthquakes of magnitudes $1 < M < 4.5$ and a region-specific GMPE to find path-specific residuals or deviations away from an average ground-motion per recording (Sahakian et al., 2018). We compare these path-specific PGA residuals to crustal properties such as seismic velocity, velocity heterogeneity, and the seismic quality factor ($Q$). We find that heterogeneity in seismic $S$ wave velocity, represented by the path integral of the gradient of velocity, correlates with the empirical path residuals, especially at long distances. This implies that local, path-specific heterogeneities (length scales of $>3$ km) in the velocity field are indicative of attenuation (likely from scattering or unmodeled intrinsic attenuation). Incorporating these correlations into a GMPE increases the accuracy and reduces the variability. The fact that the correlations are also physical allows them to be extrapolated in future applications, although testing of the magnitude and distance scale-up is necessary, along with testing this framework in other locations or with other seismic velocity models.

2. Methods

With our proposed methodology, we empirically seek relationships between ground motion amplitudes and crustal properties. Any recording’s ground-motion can be considered a superposition of the earthquake source effects, site or station effects, and effects from the path between the event and the station. In light of this, including physical information of these effects into a GMPE is useful for more accurately estimating ground-motion. One method to explore how various physical source, path, and site parameters may control ground-motion is to study relationships between these parameters, and GMPE residuals. A GMPE represents average properties of ground-motion in a region; therefore, the residual (observed – predicted) ground-motion represents any unmodeled variations in regional parameters. The residual furthermore can be decomposed into event, path, and site residuals for each recording and used to understand physical processes. For example, earthquake stress drop and its variability have shown to be correlated with high-frequency (e.g., PGA) event residuals on an event-by-event basis (e.g., Baltay et al., 2017; Trugman & Shearer, 2018), but application in a forward manner is tricky given that it is nearly impossible to say what the stress drop of any future event may be. Site effects are often represented by physical values such as $V_{S30}$ (the time-averaged shear wave velocity in the top 30 m of the Earth’s crust) and $x$ (the slope or decay of the high-frequency portion of a recorded spectrum; Anderson & Anderson & Hough, 1984; Anderson & Humphrey, 1991; Laurendeau et al., 2013; Yong et al., 2016). Considerations of path-specific effects, however, have heretofore not been studied in relation to physical properties, only in a purely statistical sense. This physics-based, but still empirical, approach is useful for GMPEs as it allows the information to extrapolate to a range of distances, magnitudes, and regions.

Therefore, to understand what physical properties affect path-specific ground-motions, we seek to correlate two main observables: (1) path residuals from a GMPE, describing the amount of uncertainty or scatter in a model attributable to differences in that path from the median model, and (2) informational metrics that describe the material properties along a raypath, such as seismic velocity or anelastic attenuation (Figure 1). If a metric shows a relationship with the recording’s path residual, then we can feasibly explain some component of the unmodeled ground-motion with this obtainable crustal property, and a path-specific
adjustment to the GMPE prediction can be made with this information. This would produce a new nonergodic model with reduced overall scatter and thus standard deviation.

To implement this method, we need the path residuals and material properties for a very large data set in order to have a statistically robust result. Southern California region has a number of dense seismic networks, one of which has been operational for decades (Analysis of Natural Zeismicity at Anza - ANZA; Berger et al., 1984; Vernon, 1989), and a very high level of seismic activity (Hardebeck & Hauksson, 2001). We use the peak ground acceleration (PGA) data and the nonbiased, regional GMPE of Sahakian et al. (2018) to obtain observable (1), the ground motion path residuals. Because the GMPE is developed specifically on this PGA data set, it represents median values of the region’s parameters (such as geometrical spreading and intrinsic attenuation). The path residuals from the median GMPE are therefore representative of any variations away from the average attenuation or elastic structure for that specific path.

In addition, there are a large variety of crustal models (tomographic, seismic attenuation, or quality factor Q) and detailed fault and geologic mapping in Southern California (Allam et al., 2014; Bennett et al., 1996; Fang et al., 2016; Hauksson & Shearer, 2006; Plesch et al., 2007; Shaw et al., 2015). If a relationship between crustal properties and ground motion amplitudes exists here, it can be incorporated into GMPEs and applied in another region where seismic data are absent but crustal structure models are available (e.g., from noise or active-source studies), to improve the ground motion estimations.

2.1. GMPE Path Residuals

Sahakian et al. (2018) used observed PGA values from ~10,000 earthquakes, largely in the range of $1 < M < 3$, all from the year 2013 in Southern California (Figure 2) to develop their regional GMPE. They use only data recorded on a minimum of five stations; this database includes a total of 78 seismic stations and more than 123,000 earthquake recordings. We use the distance metric $R_{rup}$, closest rupture distance, equivalent to hypocentral distance for these small magnitudes; all events are recorded at distances of less than 180 km. A disadvantage of the small-magnitude nature of this data set is that these results cannot define with certainty how any path effects would scale to larger magnitudes; however, the benefit of these small-magnitude events is that they are plentiful and also can be considered primarily point sources, eliminating source complexity in the results. While finite-fault effects for these small events may contribute to the high-

---

**Figure 1.** Schematic showing the methodology employed in this paper. (left) Schematic relationship between earthquake magnitude and ground-motion (circles), and GMPE prediction curve (black line). For each recording, a total residual is obtained and decomposed into source, path, and site contributions. We first determine the path residual, as discussed in section 2. (right) Raypath for an individual recording, from the event (star) to site (cartoon structure). The material properties along the raypath (irregular shape in the cartoon) is a metric of crustal properties for every path. In this work we search for a correlation between this metric of crustal properties (seismic velocity or anelastic attenuation) and the raypath ground motion residual.
frequency energy present in many of our recordings, we do not have a sufficiently high-resolution velocity model, sufficient sampling rate, the necessary station density, or borehole stations to resolve such effects. It would, however, be very useful for future studies.

Path residuals were also computed by Sahakian et al. (2018) by comparing the observed PGA ground-motions from the small-to-moderate magnitude earthquakes in the Anza region to the GMPE constructed in that paper. This is a maximum likelihood model with event and site “terms” (residuals) included as random effects in the model. The functional form has a magnitude-dependent source component (magnitude and magnitude squared), and path function (one value of geometric spreading and one of intrinsic attenuation for the entire region). A $V_{S30}$-based relationship for the site was not included in this model, as $V_{S30}$ was not found to show a relationship to site amplification of our small-magnitude PGA values; the event residual takes up this slack (Sahakian et al., 2018). This GMPE is therefore a function of magnitude and distance for each earthquake $i$ and station $j$. The full representation of ground-motion is written as

$$\ln y_{ij} = f(M, R_{rup})_{ij} + \delta E_i + \delta S_j + \delta WS_{ij}$$

where $\ln y_{ij}$ is the intensity measure for earthquake $i$ and station $j$ (here taken to be PGA from the aforementioned data set), and $f(M, R_{rup})_{ij}$ is the GMPE functional form described above. $\delta E_i$ and $\delta S_j$ are the random effects, event and site terms for earthquake $i$ and site $j$, respectively, as defined in Baltay et al. (2017). These are solved for as a part of the mixed effects regression (Sahakian et al., 2018). The final term, $\delta WS_{ij}$, is the remaining PGA residual. In theory, it consists of both path residual and random or aleatory residual:

$$\delta WS_{ij} = \delta P_{ij} + \delta W^0_{ij}$$

where $\delta P_{ij}$ is the residual for the path between earthquake $i$ and station $j$, and $\delta W^0_{ij}$ is the random residual for that same path. We cannot separate out the path residual and aleatory component at this time, and so we take the entire $\delta WS_{ij}$ to be the “path term” for this recording that we use in the analysis (Figure 2). For further details on the data set, GMPE, or residual decomposition, see Sahakian et al. (2018).

2.2. Seismic Velocity and Quality ($Q$) Models

To compute values that describe crustal properties along each raypath (labeled 2 in Figure 1), we compute raypaths for every recording, including a variety of metrics. We raytraced through the seismic velocity model of Fang et al. (2016) to obtain three-dimensional raypaths using the FM3D software (de Kool et al., 2006). FM3D is based on the fast-marching method, which uses a finite-difference eikonal solver. Thus, we find the fastest-arrival raypaths for a given velocity model and compute the $P$ wave and $S$ wave raypaths for the Fang et al. (2016) $V_P$ and $V_S$ models, respectively (though in this paper we focus on the rays for the $V_S$.
models only as they are the only values that show robust relationships). The vast majority of rays travel through the crust and are upgoing from the event location. Very few, if any at all, travel through the mantle. The metrics we devise and compute for each recording aim to describe several features, including the effects of bulk properties along the raypath—crustal velocity (i.e., shear wave velocity speed, \( V_S \)) and seismic attenuation (related to quality factor, \( Q_S \))—and the heterogeneity of these properties as well. We have found the most effective metric to be our gradient metric, \( dVI \):

\[
dVI = \int \frac{|dV|}{dL} dL
\]

the path integral of the absolute value of the gradient of \( V_S \) in the direction of the ray at every point along the raypath (\( L \)). In a physical sense, this is representative of the degree of heterogeneity in velocity along a raypath (i.e., larger \( dVI \) represents greater heterogeneity). We apply this metric to the velocity model of Fang et al. (2016), which features a grid spacing of 0.03° in both latitude and longitude (~3 km horizontal spacing).

Another metric we devise and investigate is the normalized path integral metric for attenuation, \( QI_{\text{norm}} \):

\[
QI_{\text{norm}} = \frac{\int Q \, dL}{\int Q_{\text{max}} \, dL}
\]

where \( Q_{\text{max}} \) is the maximum value of \( Q \) found in the crustal model; the numerator is the path integral of seismic quality factor (here denoted \( Q_S \)), and the denominator is the path integral of seismic quality, but assuming that the value of \( Q \) along the entire ray is \( Q_{\text{max}} \). We use this formulation to remove the effects of distance, without explicitly dividing by distance; instead, \( QI_{\text{norm}} \) represents a unitless fraction of the total amount of \( Q \) that the ray could encounter, given the defined model.

For every recording \( ij \) in the data set, we first interpolate the raypath obtained when raytracing through the Fang et al. (2016) crustal property model, to obtain the property value at every discretized point along the ray. Then, along each ray, we compute these metrics for the values of the gradient of velocity for the model of Fang et al. (2016) and seismic quality of Hauksson and Shearer (2006). These results are then compared to the path residuals for recording \( ij \). We use a single-value metric (instead of a formulaic metric) so that it can easily be compared to the single-value path residual per recording.

Finally, we computed the gradient metric (equation (3)) for seismic quality (\( Q_P \) and \( Q_S \)) and the normalized path integral metric (equation (4)) for seismic velocity (\( V_P \) and \( V_S \)), but we do not present them here, as they do not show promise toward our goal. Of course, this methodology could be extended to other crustal properties as well with these same formulations (such as density and shear modulus).

3. Results

3.1. Correlation of Path Residuals With Physical Properties

Although we computed \( P \) and \( S \) wave raypaths for every recording in this data set, we focus here on the \( S \) wave path metrics, as the PGA data set we are using is computed using the \( S \) wave arrivals (available from Sahakian et al., 2018). Before exploring any metrics, we look qualitatively at regional variations in the path terms. We color the 3-D raypaths by their path terms (Figure 3). Although each path bends in three dimensions and has many points along its path, it is assigned only one path term. We find some consistent high or low path terms in particular regions, although the panel with all rays is too dense to easily visualize systematic trends. The insets of all rays arriving at a single station (ANZA station RDM and CI station ERR, respectively) are easier to interpret. These examples show clear regional variations. For example, the recordings to the NW of ERR have much greater path residuals than those arriving from the WSW. Likewise, the azimuthal variation is apparent for RDM, such as the rays arriving from the north show lower path residuals.

3.2. Relationships With Crustal Properties

The metric showing the strongest relationship with our path residuals is \( dVI \), or the gradient of velocity along each raypath. This can be seen in Figure 4, where the path residuals (within ±5 \( \Phi_{\text{res}} \) of the mean path residual) versus \( dVI \) are binned by distances. Examining recordings of all distances and all magnitudes together,
there is no apparent relationship between the path residual and $dVI$. However, when binned by distance (Figure 4), some features begin to stand out. At short rupture distances, $R_{rup} < 20$ km, there is a small positive correlation between the path residual, $\delta P_{ij}$, and $dVI$ ($r = 0.12$). This relationship is no longer present between ~20 and 60 km rupture distance, and at longer distances, $>60$ km, the relationship starts to become negative. At much longer distances, $>100$ km, there is a weak to moderate negative relationship between the path residual and $dVI$ ($r = -0.48$). Each of these bins has a very low $p$ value and exceptionally high statistical power with a significance level of 0.05, suggesting the relationship/correlation between the path residuals and $dVI$ is robust. We complete a similar analysis using magnitude bins and find no difference between magnitude bins, suggesting the correlations are distance-dependent and not magnitude-dependent, at least within the magnitude range of our data set.

As $Q$ would be expected to yield a relationship, we present the results of the metric $QI_{\text{norm}}$ (Figure 5). In each distance bin, the statistical power is not as high as what we find for the gradient metric ($dVI$); however, given our large data set (>123,000 values) the results remain strong. These results show lower path residuals with larger values of $QI_{\text{norm}}$ at greater distances, but with a much weaker correlation than for $dVI$. We similarly bin the $QI_{\text{norm}}$ results by magnitude and find no relationship, so we do not present them here.

As mentioned above, we apply equations (3) and (4) to $V_S$, $Q_P$, and $Q_S$ models but only show the results of $dVI$ and $QI_{\text{norm}}$. None of these other metrics show relationships with the path residual, either with magnitude or distance dependence. This suggests that bulk velocity is not related to path-specific ground-motions for the Fang et al. (2016) model and that heterogeneity in seismic attenuation is also not related.

4. Discussion
4.1. Crustal Heterogeneity
The most promising and useful path effect observed in this work is related to structural heterogeneity, or the gradient metric, $dVI$. At longer distances, we observe a relationship between lower path residuals and greater
Figure 4. Path residual (path term in ln residual) versus the path gradient of seismic velocity, $dVI$, along each raypath for various station-to-source ($R_{rup}$) distance bins in kilometers, as labeled. Here, we only include data points that fall within ±5 standard deviations ($\phi_{SS}$) of the mean path term (0 ln residual). The results are shaded by earthquake magnitude, and within each distance bin subplot we give the correlation coefficient ($r$), $p$ value ($p$), and statistical power for the given relationship/effect size assuming a statistical significance (alpha) of 0.05. In each subplot, the distance-binned means of the residuals are plotted with the 95% confidence limit. Thin dash-dotted line: 0 ln residual. Dashed lines: slope of the residuals in each distance bin, as a visual aid to see how the correlation varies with distance.
Figure 5. Path residual (path term in ln residual) versus the normalized path integral of the quality factor $Q$, $Q_{I_{\text{norm}}}$, along each raypath for various $R_{\text{rup}}$ distance bins as labeled. The data are shaded by earthquake magnitude, and for each distance bin we also report the correlation coefficient ($r$), $p$ value ($p$), and statistical power for the given relationship/effect size assuming a statistical significance (alpha) of 0.05. In each subplot, the distance-binned means of the residuals are plotted with the 95% confidence limit.
structural heterogeneity, or larger \( dV_I \) (Figure 4). Our interpretation is that with increasing distance, more or
greater velocity contrasts along a ray's path yields more energy lost to scattering (refractions or reflections),
and less energy transmitted. Recent numerical modeling studies (Olsen et al., 2018) have found similar
results. At closer distances, we observe a weak positive correlation between path residuals and \( dV_I \), which
is perhaps unsurprising. Within the frequency range that our PGA measurements are likely selected from
(5–15 Hz), some numerical simulations that investigated the effects of media heterogeneity on ground-
motion found that increased roughness or heterogeneity may lead to amplified high-frequency ground-
motions. This is, however, still a topic of research, and not all studies arrive at the same conclusion
(Bydlon & Dunham, 2015; Mai et al., 2010; Mena et al., 2010). Additionally, these studies primarily focused
on the near-field \( (R_{rup} < 60 \text{ km}) \), and so we cannot make any inferences regarding our longer
distance observations.

Another explanation why at shorter distances we do not see a relationship between path residuals and struc-
tural heterogeneity is that the velocity model we use to compute the metrics features a grid spacing of
\(~3 \times 3 \text{ km} \) horizontally. With this grid spacing, we cannot resolve the shorter wavelength structural features
that likely produce the high-frequency ground-motions affecting PGA, at the shorter source-to-station dis-
tances. At longer distances, as higher-frequency energy is attenuated faster than lower-frequency energy,
PGA may be coming from a lower range of frequencies, which would be affected by the slightly larger scale
heterogeneities resolvable by this velocity model.

Our observation that there is a positive correlation between lower ground-motions and increased velocity
heterogeneity at only the longer distances could also be attributed to scattering \( Q \). Scattering \( Q \) is different
from intrinsic \( Q \) and has been shown to be the predominant process responsible for generating seismic coda
at longer distances (Aki & Chouet, 1975; Dainty & Toksoz, 1981). If true, this would imply that the PGA and
thus path residual should be lower at longer distances because of greater scattering. This agrees with what
we observe. We therefore propose that the gradient metric could be acting as a proxy for scattering phenom-
ena, which is a useful finding for future studies of path effects for GMPEs.

### 4.2. Null Relationships

Intuitively, we would expect the strongest relationship between ground-motions and crustal properties to be
with seismic quality, or \( Q_{I_{norm}} \). Seismic waves traveling through regions of higher \( Q \) should experience less
seismic attenuation, yield larger values of \( Q_{I_{norm}} \), and lead to less ground-motion than predicted by a GMPE,
attributable to the path. This would produce a positive relationship between the path residual, \( \delta P_{ij} \), and
normalized seismic quality metric, \( Q_{I_{norm}} \), for any path from earthquake \( i \) to station \( j \). This is not what we
observe. Instead, we find a negative relationship between path residuals and \( Q_{I_{norm}} \), which become stronger
at greater distances (Figure 5). We interpret the negative relationship between path residual and seismic
quality to be an artifact of the data set, for a couple of reasons.

We think that the majority of the negative trend between path residuals and \( Q_{I_{norm}} \) at longer distances is
from a data bias, having to do with the relationship between \( M, R_{rup} \), and \( Q_{I_{norm}} \). Although this metric
should be removing distance dependence, there remains some distance dependence inherent within
\( Q_{I_{norm}} \). This is because events traveling longer distances to a station will travel on paths that propagate deeper
through the crust (as our raypaths take the fastest traveltime paths, and seismic velocity increases with
depth), and as such will experience larger values of \( Q_{S} \) since this also tends to increase with depth. Within
each distance bin, the larger magnitudes generally occur at greater distances from the stations (which tend
to be in the center of the region; Figure S1), since larger events can be recorded at farther distances (a typical
\( M \) vs. \( R_{rup} \) distribution). This explains why the larger-magnitude earthquakes produce larger values of
\( Q_{I_{norm}} \), as can be seen in Figure 5. Also, at the longer distances, there is a slight magnitude dependence to the
path residuals (Figure S2), due to the instrument noise floor, such that larger-magnitude events produce
slightly negative path residuals. This magnitude dependence combined with the above magnitude/distance/
\( Q_{I_{norm}} \) relations leads to larger values of \( Q_{I_{norm}} \) at longer distances, associated with negative path residuals.
This produces a slightly negative relationship between the path residuals and \( Q_{I_{norm}} \). We therefore propose
that this weak negative correlation is an artifact of the data set and not a real feature.

This now begs the question as to why there is effectively no relationship between the path residuals and seis-
mic attenuation in this region. We suggest that this may be due to the resolution of the Hauksson and
Shearer (2006) attenuation model, which in comparison to the velocity model has lower resolution. In Figure 6, we show the minimum number of cells that a recording with distance $R_{rup}$ can travel through, in both the velocity model (Fang et al., 2016) and attenuation model (Hauksson & Shearer, 2006). Both models have varying resolution with depth (and additionally one model is oriented directly north, whereas the other is oblique/San Andreas parallel), so for every recording we compute the maximum distance that it could travel in a model node, given the earthquakes depth and rupture distance, by assuming that the event and station are located at extreme corners of the node’s volume. This places a lower bound on the number of cells the wavepath encounters. We find that for a given distance, recordings travel through 4 times as many nodes of the velocity model than attenuation model. With the gradient metric, $dVI$, we begin to see some path effects at ~60 km, where recordings travel through a minimum of ~12 nodes of the velocity model. In contrast, at this distance rays will encounter a minimum of only ~3 nodes of the attenuation model. We suggest that the attenuation model is too coarse to resolve regional features that may control path effects. The anelastic attenuation term in the regional GMPE may be serving as an adequate estimate of the average attenuation in the region; regional variations presented by this $Q$ model are not great enough to be extracted from the path-residual data.

We find the $VI$ and $VI_{norm}$ metrics do not appear to control path-specific ground motion amplitudes in this region, suggesting that bulk velocity properties impose less of an effect on ground-motions than do heterogeneities or scattering effects. Although there are known relationships between seismic velocity and attenuation or quality factor (Barton, 2007; Brocher, 2008), these may not account for the complicated nature of many other parameters affecting seismic quality or attenuation, such as the presence of fluids (Eberhart-Phillips & Chadwick, 2002).

4.3. Implications for Uncertainty in Ground-Motion Estimation

Ultimately, the primary application of any path effects we observe is to move toward path-specific ground motion models which include, for any path of interest, individual adjustments to both predictions of median ground-motion as well as their uncertainty. While for these Southern California data the path effect we observe from the gradient of velocity displays a weak to moderate correlation and is strongest at greater distances, it is still effective for our goal of improving ground motion estimation. In fact, the gradient...
metric produces a larger correlation than many of the proxies already incorporated into ground motion models (such as basin term proxies, i.e., Campbell & Bozorgnia, 2014).

To demonstrate the effectiveness of velocity heterogeneity, or \( dVI \), we show how it would improve the median values of ground-motion and path residual standard deviations (\( \Phi_{SS} \), after Baltay et al., 2017 and Sahakian et al., 2018) if it were included as a path-specific adjustment to every recording in the PGA data set. As there is a distance dependence to this relationship, we fit a quadric surface to the path residual (\( \delta P_i \)), gradient metric (\( dVI \)), and rupture distance (\( R_{\text{rup}} \)), to find a relationship that can be used to predict an adjustment to the median ground-motion, for any given path:

\[
\delta P_i = a \cdot dVI^2 + b \cdot R_{\text{rup}}^2 + c \cdot dVI \cdot R_{\text{rup}} + d \cdot dVI + e \cdot R_{\text{rup}} + f
\]

where the coefficients \( a \) to \( f \) can be used for any path in this region. In a forward sense, all that is necessary is a velocity model and three-dimensional raypaths to compute \( dVI \), for any path of interest to the application; these properties can then be input into equation (5) as the knowable part of the path effects we observe here. Finally, this may be included as an added term in a GMPE in predicting PGA or high-frequency ground-motion.

Given this relationship, we compute the adjusted path residuals for every recording and adjusted standard deviations within each distance bin (Figure 7 and Table 1). To illustrate where this structural heterogeneity has the largest effect on adjusted ground-motions, we show histograms of the median adjustments in each bin (Figure 8 and Table 1). When applied to the entire data set, the reduction in uncertainty is small (0.011 ln units), and the adjustment in the median ground-motion is very low (\( \sim 0.002 \) ln PGA). However, with increasing distance, the adjustment becomes more significant. In the longest distance bin, applying the gradient metric as a path-specific adjustment reduces the uncertainty by nearly 0.06 ln units, and the adjustments in median path terms can be as large as \(-0.7\) or 1.6 ln residual units, with a mean adjustment of 0.427.

From our results (Figures 7 and 8 and Table 1), we find that the largest adjustments to median ground-motion arise at long distances. The largest increases in ground-motion from this path effect occur at long distances, for paths traveling through very homogeneous crust. The largest decrease in ground-motion occurs for paths travelling through very heterogeneous regions. This implies that even though this data set (as well as many others) is dominated by shorter rupture distance recordings, if the raypath traverses >100 km distance through either a very homogeneous or very heterogeneous area, this relationship can have a significant effect on the ground motion estimation (both median and standard deviation), as well as resulting seismic hazard.

From a seismological perspective, it is initially surprising that seismic quality (or anelastic attenuation) is not the dominant property to control path effects, but from an engineering perspective, the gradient of velocity is much more practical. For site-specific applications (such as for critical facilities), it is much easier to create a model of velocity with an acceptable resolution to compute the gradient metric than it is to create a model of seismic attenuation or \( Q \). Computing a model of \( Q \) requires many reliable \( S \) wave spectra in a region (Eberhart-Phillips & Chadwick, 2002; Hauksson & Shearer, 2006), which is difficult if the region is not very seismically active or if the region does not have dense and well-maintained seismic networks. The \( Q \) model we use here may have a much lower resolution than many velocity models available in the region, but its grid spacing and resolution is on par with other \( Q \) models around the world (Eberhart-Phillips & Chadwick, 2002; Stachnik et al., 2004). Our observation of scattering controlling path effects in Southern California is therefore favorable in a practical sense, and it is useful for future applications.

**4.4. A Path Forward**

Studies of path effects for empirical ground motion estimation have so far relied solely on methods requiring large amounts of seismic data, or a purely statistical approach. The novel relationships we have found are encouraging, in that there may be knowable, repeatable properties in a region that can be obtained to update estimations of ground-motion based on path. The methodology we have proposed and tested is a promising and fruitful way forward toward constraining ground motion path effects. There are, however, still some limitations of our findings which will require more research moving forward.
Figure 7. Comparison of original path terms (circles; same as in Figure 4) for each distance, and adjusted path terms (triangles), given the solution for a best-fitting quadric surface (to the path terms, dVI, and $R_{nup}$). The original and adjusted path term standard deviation in each bin is labeled on the top right of each subplot, as well as the reduction in standard deviation or uncertainty. The distance-binned means and 95% confidence intervals are plotted for the adjusted path terms in each subplot.
There are some requirements for seismic hazard assessment that our data set cannot explain or provide. For example, the path effect we observe is only valid for the small-to-moderate magnitude earthquakes in our Southern California data set. It is not clear how our results using small events will scale to larger magnitudes. Using PGA as an intensity measure to investigate path effects may be defensible at smaller magnitudes, but without understanding how our observed path effects are manifested in the frequency domain, it is difficult to make assumptions about how this relationship can be used for larger earthquakes in the region. Furthermore, Q is known to be frequency dependent, so without Fourier amplitude spectra, these effects may be lost or hidden in our study. To move forward in this regard, we believe that it is important to test the gradient metric on these data in the frequency domain, or even with peak ground velocity values. Furthermore, obtaining these frequency-dependent intensity measures would allow us to understand if the velocity model resolution is one of the controlling factors on why we only observe a relationship between PGA and structural heterogeneity at longer distances.

Larger-magnitude data are also important to understand how the path effects we see could be applied to probabilistic seismic hazard analysis (PSHA). It is not clear how \( dV I \) should be calculated for finite source ruptures (i.e., not the point source ruptures that exist in this data set). For example, is the “path effect” represented by a distribution of \( dV I \) for all paths from the finite fault rupture, or is there a particular path that dominates path effects? Using \( R_{rup} \) as a distance metric could also further complicate this analysis. We propose that future studies should focus on magnitude scaling for finite fault applications and include various distance metrics as well such as the mean rupture distance (Thompson & Baltay, 2018). In addition to finite fault effects, studies with a larger range of magnitudes should consider magnitude-dependent geometric spreading. Although of engineering interest, these effects cannot be assessed with this data set.

Another question is if there are other metrics or combinations of metrics that might be better and how a material model’s resolution affects the results. If what we propose is true, and scattering is in fact the driving process in path effects, then studies that look at an assortment of velocity models with varying resolutions will be important to understand on what scale we must resolve heterogeneities for a model to be useful. Another option would be to randomize or perturb the velocity model we have here to impose smaller-scale heterogeneities, or to look at a data set simulated or synthetic earthquakes, since this numerical method can provide an in-depth examination of how the material model affects the results.

Furthermore, if scattering is a controlling process here, then is there a better way to represent, in a physical and less empirical sense, how it affects ground-motions? Scattering Q is often represented as a diffusive process. Can our results be modeled instead as diffusion? If so, this would be a more elegant and practical way to include scattering into path-specific ground motion models, because then one might only need to understand regional variations in diffusivity with respect to scatterers, instead of specifically velocity models.

It may be useful to consider \( QI_{norm} \) in this context as well, as it may relate to scatterers.

Finally, it is possible that our observations are only applicable to the Southern California region. Testing this method in other regions will be important in understanding if our results are a pervasive feature or region-specific. For example, the crustal structure in Northern California is sufficiently different from Southern California that it is not clear if the results from this method is immediately applicable. Furthermore, other

Table 1

<table>
<thead>
<tr>
<th>Distance bin center (km)</th>
<th>Initial ( \gamma_{SS} )</th>
<th>Quadric ( \gamma_{SS} )</th>
<th>Quadric reduction</th>
<th>Minimum adjustment (ln)</th>
<th>Maximum adjustment (ln)</th>
<th>Mean adjustment (ln)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.442</td>
<td>0.431</td>
<td>0.011</td>
<td>−0.712</td>
<td>1.564</td>
<td>−0.002</td>
</tr>
<tr>
<td>10.0</td>
<td>0.435</td>
<td>0.429</td>
<td>0.006</td>
<td>−0.038</td>
<td>0.199</td>
<td>0.046</td>
</tr>
<tr>
<td>30.0</td>
<td>0.434</td>
<td>0.43</td>
<td>0.004</td>
<td>−0.106</td>
<td>0.172</td>
<td>−0.009</td>
</tr>
<tr>
<td>50.0</td>
<td>0.414</td>
<td>0.413</td>
<td>0.0</td>
<td>−0.357</td>
<td>0.032</td>
<td>−0.09</td>
</tr>
<tr>
<td>70.0</td>
<td>0.406</td>
<td>0.398</td>
<td>0.008</td>
<td>−0.491</td>
<td>0.17</td>
<td>−0.06</td>
</tr>
<tr>
<td>90.0</td>
<td>0.471</td>
<td>0.443</td>
<td>0.028</td>
<td>−0.633</td>
<td>0.345</td>
<td>0.007</td>
</tr>
<tr>
<td>110.0</td>
<td>0.571</td>
<td>0.52</td>
<td>0.05</td>
<td>−0.678</td>
<td>0.612</td>
<td>0.201</td>
</tr>
<tr>
<td>150.0</td>
<td>0.654</td>
<td>0.596</td>
<td>0.058</td>
<td>−0.712</td>
<td>1.564</td>
<td>0.427</td>
</tr>
</tbody>
</table>

Note: Initial values for the standard deviations, standard deviations after the quadric surface adjustment, and the reduction in standard deviation after the adjustment given in the second to fourth columns; minimum, maximum, and mean adjustments (same as shown in Figure 8) in each bin given in the fifth to seventh columns. Again, the largest adjustments and greatest reduction in standard deviation is seen in the farthest distance bin.
regions outside of the United States with different geologic and crustal velocity structures, such as Japan (Koketsu et al., 2012), would present very different impedance contrasts. It is unknown whether these regions would present similar observations and findings with our methodology. Understanding these regional variations is crucial in moving forward to include more physics in ground motion models and creating path-specific models.

Figure 8. Histograms of the natural logarithm adjustments to path residuals in each distance bin from Figure 7. The minimum, maximum, and mean adjustment is labeled in each subplot. Because the strongest correlation between the path term and the gradient of velocity is found for the greatest distance bin, the adjustments are also the largest for that distance bin.
5. Conclusions

We develop and test a methodology to constrain path effects and move toward path-specific ground motion models with observable physical properties or processes. We find that structural heterogeneity is the dominant path effect for PGA in the Southern California region. For longer paths, increased heterogeneity in seismic velocity along the path correlates with lower ground motion amplitudes, indicating that energy is lost to refracted or reflected rays likely due to scattering effects and scattering Q. Our results suggest that, for this region and data set, bulk or intrinsic properties of velocity and seismic quality factor Q are not successful for modeling path effects, though in the case of Q this could be due to low model resolution. The relationship we observe between ground-motions and structural heterogeneity can be used to reduce model uncertainty and improve ground motion estimation, in particular for paths with long (>100 km) rupture distances, where paths through very homogeneous structure tend to increase ground-motions and paths through very heterogeneous structure decrease ground-motions. At these longer distances, the reduction in uncertainty is nearly 0.06 (natural logarithm units). We propose that the methodology we have used shows much promise, though there are many factors to consider to fully understand and model path effects, including frequency, magnitude, and region dependence.

References


