THE SEISMOLOGICAL SOCIETY OF AMERICA
400 Evelyn Ave., Suite 201
Albany, CA 94706-1375
(510) 525-5474; FAX (510) 525-7204
www.seismosoc.org

Bulletin of the Seismological Society of America, Vol. 101, No. 2, pp. 652–666, April 2011, doi: 10.1785/0120100134

Spectral Analysis of K-NET and KiK-net Data in Japan, Part I: Database Compilation and Peculiarities

by Adrien Oth, Stefano Parolai, and Dino Bindi

Abstract The wealth of accelerometric recordings collected by the K-NET and KiK-net networks in Japan since 1996 provides a unique opportunity to improve our understanding of many important seismological research questions. Subsets of these data have been used for many case studies, most of them, however, not focusing specifically on the best practices for data selection and giving relatively little attention to the properties and peculiarities directly observable from the data. Yet for many applications, these steps are an important prerequisite for successful and reliable analysis. For this reason, we devote this article to the extraction of a large data set of surface and borehole recordings from the K-NET and KiK-net databases with strong emphasis on data quality and reliability. The final data set available for subsequent work consists of 78,840 records from 2201 earthquakes covering the Japan Meteorological Agency (JMA) magnitude range 2.7–8, observed at 1681 sites throughout Japan. We explain how this data set has been compiled, including automatic phase picking and relocation of events. We also present an overview of the general features of the data set, providing important information for subsequent analysis. Strong amplification effects at high frequencies are immediately visible on the surface recordings. Furthermore, there is a clear presence of downgoing waves in the borehole records, as deconvolution of borehole/surface recording pairs indicates.

Introduction

Due to their tectonic situation, the Japanese Islands and their surroundings are characterized by extremely high seismic activity. In the northeast, the Pacific Plate subducts beneath the Eurasian Plate, whereas the southwestern part of Japan is sitting over the subducting Philippine Sea Plate (e.g., Iwasaki *et al.*, 2002). Due to this situation, the seismicity in Japan is characterized both by interplate (reaching up to M 8) and intraplate earthquakes, the latter attaining magnitudes of the order of M 7.

Following the devastating Kobe (Hyogo-ken Nanbu) earthquake on 17 January 1995 (see, e.g., Ide *et al.*, 1996; Fukushima *et al.*, 2000), causing a death toll of more than 6500, large efforts were devoted to establishing dense and uniformly spaced strong-motion, high-sensitivity and broadband seismic networks covering the entirety of Japan. These efforts resulted in the creation of the networks K-NET (strong-motion), Hi-net (high-sensitivity), and F-net (broadband) (Okada *et al.*, 2004). These networks have since collected enormous quantities of high-quality digital data, providing the means for investigating important research questions. These include, for instance, the issue of earthquake scaling, improved characterization of site amplification for engineering applications, and the analysis of variability issues in ground-motion prediction equations (GMPEs).

Subsets of the K-NET and KiK-net databases have been used in recent years to address some of these questions, and we can just provide a few examples here. Cotton *et al.* (2008), for instance, used a subset of early KiK-net data to investigate the effect of magnitude-dependent functional forms in the derivation of GMPEs. Moya and Irikura (2003) analyzed the aftershocks of the 2000 Tottori event and derived site amplification functions and $Q(f)$ in the area, while Edwards and Rietbrock (2009) discussed the spectral characteristics of ground motion using two subsets of KiK-net borehole data and a set of small-magnitude Hi-net data in central Japan. Choy and Boatwright (2009) used K-NET recordings of the 1996 Kyushu and 2000 Tottori earthquakes to compare regional and teleseismic estimates of radiated energy. Nonlinearity and temporal changes in site response and crustal seismic velocities were studied by several authors (Sawasaki *et al.*, 2006, 2009; Wu *et al.*, 2009; 2010), and an earthquake early warning feasibility study for the ElarmS system was performed using a subset of the K-NET database (Brown *et al.*, 2009).

However, as a general rule, most studies making use of K-NET and KiK-net data do not fully address the issue of what the best practices are for data selection, nor do they provide a detailed discussion of the features that can directly

be observed from the data without the need of extensive fitting or inversion procedures. Without a clear understanding of these basic characteristics, it can be difficult to properly assess what models and assumptions are appropriate for use with these data sets, and the obtained results can be driven to a large extent by those assumptions. For instance, one important issue that is generally not addressed is the potential influence of downgoing waves on the ground motions recorded by the KiK-net downhole sensors. From deconvolution of surface/downhole recording pairs at the Ataköy vertical array in Istanbul, Parolai *et al.* (2009) showed that downgoing waves are clearly observed down to a depth of 140 m. Furthermore, spectral fitting procedures are usually subjected to potential trade-offs between different parameters of the model. In order to be able to assess these trade-offs, extensive knowledge of the inherent properties of the data set in use is an indispensable precondition.

For these reasons, we dedicate this article to data selection and processing issues and describe the most relevant characteristics directly recognizable on the selected set of K-NET and KiK-net recordings. We apply an automatic picking algorithm on the entire available K-NET and KiK-net databases, from their beginnings in 1996 until February 2009, in order to determine *P*- and *S*-wave onsets. We then relocate more than 3200 earthquakes, select a subset of 2201 events with well-constrained locations, and remove remaining false picks by comparing observed with synthetic travel times. For the KiK-net sensors, differences in horizontal orientation of downhole and surface sensors are determined through correlation analysis of low-pass filtered recordings of large earthquakes. We briefly touch upon the frequency content of the observed ground motions, an issue that we address in detail in Oth *et al.* (2011; hereinafter referred to as companion paper), and finally, through deconvolution of downhole/surface recording pairs, we show that effects of downgoing waves are visible in the borehole recordings throughout the entire KiK-net database, even for sensors located as deep as 2 km.

By combining a variety of powerful seismological techniques available in the literature, we provide the reader with a complete methodology on how to efficiently extract, at the same time, the largest possible and most reliable data subsets from extensive strong-motion databases such as K-NET and KiK-net and to obtain a quick and robust overview of their first-order characteristics, including downgoing wave propagation at borehole stations. These issues deserve particular attention in the case of the K-NET and KiK-net data because these are becoming increasingly popular among researchers, and their general features have not yet been discussed in such detail and on the basis of as many records before. In the companion paper, we then apply a nonparametric generalized inversion approach in order to separate source spectra, attenuation characteristics, and site response functions throughout Japan and use the information collected here for assessing these spectral models.

The K-NET and KiK-net Networks

Immediately after the 1995 Kobe (Hyogo-ken Nanbu) earthquake, the National Research Institute for Earth Science and Disaster Prevention (NIED) started constructing a nationwide strong-motion network, K-NET, as well as a network of borehole high-sensitivity stations with a spacing of 15–20 km, Hi-net. All Hi-net stations were installed in boreholes (most of them at depths between 100 and 200 m, but some deep observation sites with borehole depths up to 3 km also exist) in order to ensure optimal signal-to-noise (SN) conditions. Furthermore, each Hi-net station is additionally equipped with a pair of uphole/downhole strong-motion sensors, which are separately referred to as KiK-net, and a two-component tiltmeter at borehole depth (Fujiwara *et al.*, 2004; Okada *et al.*, 2004).

At the time of the writing of this article, K-NET is composed of 1037 stations, whereas KiK-net consists of 688 sites (observation site lists from the K-NET and KiK-net web sites, see Data and Resources section). The instruments have sampling frequencies of 100 or 200 Hz (100 or 200 Hz for K-NET, 200 Hz for KiK-net sensors). The resolution of the instruments is less than 10^{-4} cm/s^2, the maximum recordable acceleration is 2*g* (respectively, 4*g* for the newer K-NET sensors), and the overall frequency response of the instruments is flat from d.c. to 30 Hz (Fujiwara *et al.*, 2004). Following the information given on the K-NET and KiK-net websites (see Data and Resources), the transfer function amplitude starts, in fact, to slightly decrease for frequencies larger than ~15 Hz, and this decay starts to be significant above 25–30 Hz, such that for the consideration of frequencies above this limit, a full correction for the transfer function would be necessary. Because, following a signal-to-noise ratio (SNR) analysis (see the section Spectral Properties at a Glance), we restrict our analysis to frequencies below 25 Hz, we only correct for the necessary scale factor instead of fully deconvolving the transfer function. For more technical details on the instruments, we refer the reader to Okada *et al.* (2004) and Fujiwara *et al.* (2004). For most KiK-net stations, seismic velocity profiles obtained from downhole *PS* logging are available, reaching down to the borehole sensor depth, while for the K-NET stations, soil conditions have been characterized only down to 20 m depth (Fujiwara *et al.*, 2004).

The extraordinary dense station coverage of the K-NET and KiK-net networks, unrivaled in the world, is shown in Figure 1. For the vast majority of the KiK-net boreholes, the shear-wave velocity determined at the downhole sensor depth is close to or larger than 1 km/s (corresponding either to rock or hard-rock sites following the National Earthquake Hazards Reduction Program [NEHRP] classification, International Code Council, 2006). Thus, no significant amplification effects due to the near-surface geology are expected for the borehole recordings.

For the purpose of our study, we downloaded all available records between May 1996 and February 2009 for K-NET and October 1997 and February 2009 for KiK-net

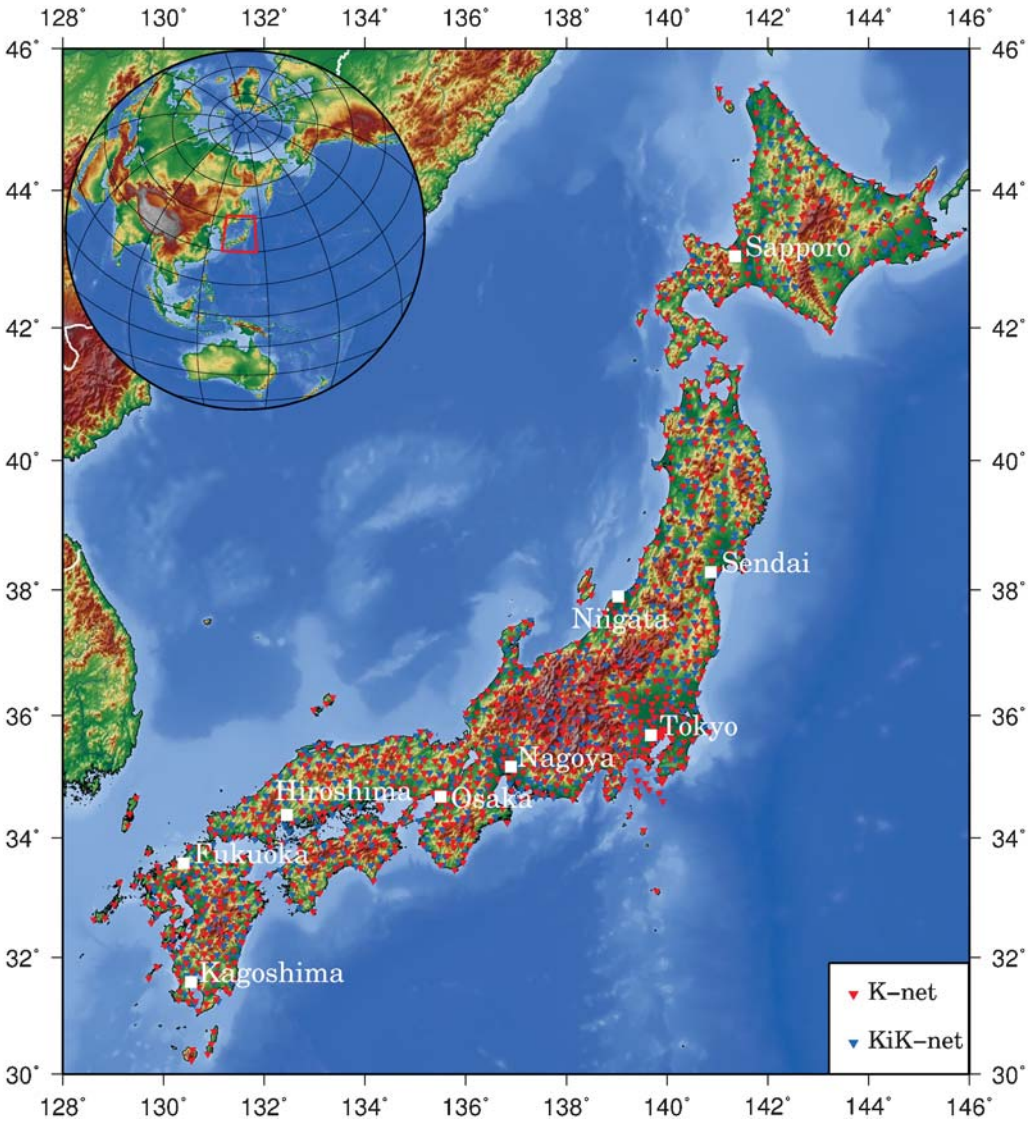

Figure 1. Map of Japan showing locations of K-NET (red inverse triangles) and KiK-net (blue inverse triangles) stations.

from the respective web sites. In total, K-NET data from 6571 and KiK-net data from 5029 earthquakes were obtained this way. For the calculation of ground-motion spectra, we downsampled all recordings with sampling frequency 200 Hz to 100 Hz (filtered low-pass with a corner frequency of 40 Hz prior to downsampling) to obtain a uniform sampling rate.

Automatic Phase Picking and Location

For some applications, it is desirable to be able to use separate *P*- and/or *S*-wave windows instead of the entire recordings, so we first performed automatic phase picking. It should be noted right from the beginning that the usefulness of automatic phase picks depends on their accuracy and the type of study in mind. For the purpose of spectral analysis of *S*-wave windows, as discussed in this paper, as long as the determined *S*-wave onset time is not later than the arrival time of the main *S*-wave energy, less accurate estimates are also acceptable in order to increase the number of available records, which is not the case for other applications such as tomographic studies.

In this study, we used the algorithm of Earle and Shearer (1994), which we only briefly describe here. First, the envelope function of the seismogram under consideration is computed. A short-term average (STA) and long-term average (LTA) moving window filter is then applied to the envelope function, and the obtained STA/LTA ratio function is smoothed with a Hanning filter to avoid rapid fluctuation leading to inaccurate arrival-time estimates. Finally, every time the STA/LTA ratio function exceeds a predefined threshold (termed a trigger point by Earle and Shearer), a phase arrival is declared. For pick confidence, the amplitude of the local maximum in the STA/LTA ratio function is defined. For the local short-period data set that we are dealing with, we set the STA and LTA window lengths to 0.1 and 1.5 s, the

length of the Hanning filter to 0.33 s, and the threshold for defining the trigger point to 2, respectively.

For each three-component record, we ran the autopicker for all three components and retained the pick with the highest confidence on each component. In order to immediately remove inconsistent picks, we only proceeded with the analysis if these highest confidence picks on the three components had a time difference between each other of less than 0.5 s. We then used the following criteria to decide whether to associate the pick with a *P*- or *S*-wave onset (we note at this point that as a general rule, automatic *S*-wave picks are of course significantly less reliable than *P*-wave picks, e.g., Satriano *et al.*, 2008):

1. Because, under normal conditions with reasonable SNR, we expect that the pick with highest confidence corresponds to the *P* wave, we associated the pick with a *P*-wave arrival if either, (1) for all three components, the highest confidence picks were the first ones and their confidences were all larger than 5 or (2) the pick confidence on the vertical component was at least 1.5 times larger than the pick confidence on the horizontal component.
2. If the pick confidence of the vertical component pick was smaller than one of the horizontal components, we used a criterion on the cumulative absolute velocity for the part of the seismogram before the pick (CAV_{BP}) and the entire seismogram (CAV_{ENT}). If $CAV_{BP} \leq 0.02 \cdot CAV_{ENT}$ on all three components, we assumed the pick to be a *P*-wave onset because only a negligible amount of the record's energy had been released up to the time of the pick. If, however, $0.02 \cdot CAV_{ENT} < CAV_{BP} \leq 0.70 \cdot CAV_{ENT}$, we associate the pick with the *S*-wave arrival. Because, for some records with rather low SNR, the algorithm tended to return the (erroneous) highest confidence picks close to the end of the trace, we included this 70% upper-bound on CAV_{ENT}.

For the KiK-net records, automatic picking was performed on the borehole recordings because these are less affected by surface noise. If the analysis described in the above list resulted in the association of the pick with the *P*-wave onset, we applied the autopicker once more to the horizontal components, this time, however, only starting on the *P*-wave onset time as an attempt to also determine the *S*-wave onset, and if the resulting highest confidence picks on the two horizontal components were less than 0.5 s apart from each other, they were associated with the *S*-wave arrival of the record. Visual inspection of the results showed that overall, this picking procedure seemed to work very well. However, some outliers do, of course, remain, and the goals of the following analysis steps are to evaluate the overall consistency of the automatic phase picks in a systematic way and to identify and remove outliers from the final data set.

We relocated all events with picks (*P* and/or *S*) available from at least 6 stations, which led to a subset of 3223 events to be considered. If, for a given earthquake, the automatically determined phase picks are consistent, we expect to obtain a well-constrained location for the event. We use the NonLinLoc software package (NLL) (Lomax *et al.*, 2000, 2001; Lomax, 2005) for this purpose, employing the equal differential travel-time (EDT) formulation for the likelihood function. The EDT formulation has the advantage over the classical L2-norm formulation that it is much more robust in the presence of outliers. Thus, if the majority of the picks for a given event are consistent with one another, the usage of this approach should lead to well-constrained locations and also allow us to eliminate remaining erroneous picks by simple arrival-time residual analysis.

We used the Japan Meteorological Agency (JMA) 2001 velocity model (Ueno *et al.*, 2002) for synthetic travel-time calculations. In order to avoid including P_n and S_n arrivals, we generally only used stations within an epicentral distance of 120 km of the location provided on the K-NET and KiK-net web sites (the latter corresponds to the rapid location determined by JMA, with a horizontal resolution of 0.1°, K. Shiomi, personal comm., 2010). For eight moderate or large events that occurred far offshore, seismograms recorded at epicentral distances up to 200 km were used (and for the Tokachi-Oki $M_{JMA} = 8$ event, up to 300 km were used). In order to assess the quality of a specific earthquake location, we computed the standard horizontal and vertical errors, ERH and ERZ, using the Gaussian estimators calculated by NLL. For small and moderate earthquakes, we considered a location as well-constrained if ERH $\leq$ 2 km and ERZ $\leq$ 2 km. For large offshore events ($M_{JMA} > 5.5$), we manually checked the picks and accepted somewhat larger standard errors because, in such cases, the source-station geometry is less favorable for accurate location, and we did not want to discard too many large earthquakes. This concerns only 26 events, and of these, only 6 have standard errors larger than 5 km.

In total, 2201 earthquakes fulfilled the criteria on location quality, showing that, in general, the automatic phase picks lead to well-constrained locations. Our locations are in good agreement with the rapid JMA hypocenters provided in the K-NET and KiK-net databases, and the obtained epicenters are depicted in Figure 2 (see also Fig. 3a, b). The subcrustal events (depth > 30 km) roughly delineate the eastern coastline of Japan, related to the subduction of the Pacific and Philippine Sea plates. The seismic events retained provide a good coverage of the entire Japanese archipelago, except for Hokkaido, where only very few crustal events have occurred. The M_{JMA} distribution of the selected events is depicted in Figure 3c. As expected, the main contributions to the data set arise from records of small-to-moderate events in the M_{JMA} range 3.5–5. However, it should be noted also that the contribution from large events, 44 earthquakes with $M_{JMA} \geq 6$, is exceptionally large as well.

As a final step of the data selection procedure, we analyzed the residuals between synthetic and picked arrival times for stations with hypocentral distance lower than

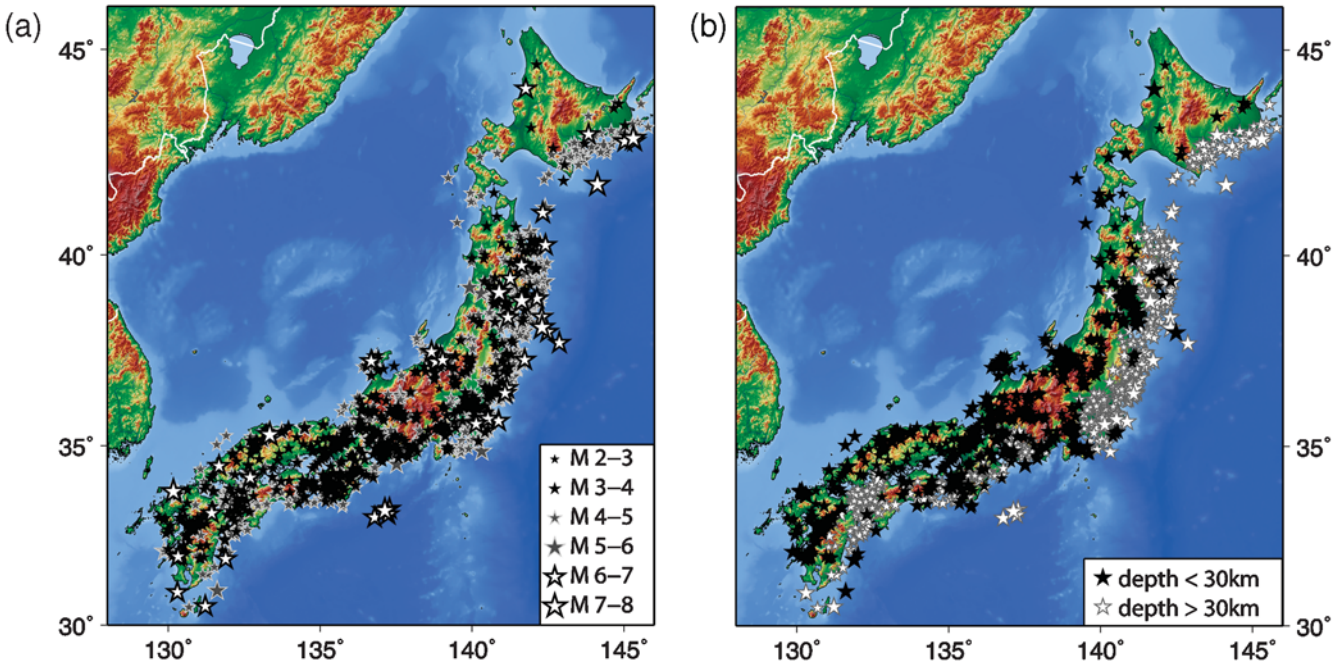

Figure 2. Epicenters of the 2201 events kept in the database. (a) Different symbol sizes denote different M_{JMA} ranges. (b) Same as (a), but highlighting the distribution of crustal (depth ≤ 30 km, black stars) and subcrustal (depth > 30 km, white stars) earthquakes.

250 km, and removed *P*- and *S*-wave picks with residuals larger than 0.5 s. For records with acceptable *P* picks only (i.e., no *S*-wave onset could be picked), we assigned the *S*-wave onset using the calculated synthetic arrival time, which works well enough to be able to cut an *S*-wave window starting around 1 s before the *S*-wave onset for spectral analysis (e.g., Oth *et al.*, 2009). Furthermore, in order to take full advantage of the large number of records from large earthquakes ($M_{JMA} \geq 5.5$) available in the K-NET and KiK-net databases, we also used the synthetic *S*-wave arrival times for records for which no automatic phase picks on the observed time history could be determined (records checked manually).

The final data set resulting from this procedure is composed of 78,840 surface records from 2201 earthquakes recorded at 677 KiK-net and 1004 K-NET sites, and 34,456 borehole records from 1872 earthquakes recorded at 677 KiK-net sites. Three examples of KiK-net recordings and the associated *S*-wave windows for spectral analysis are shown in Figure 4, and the locations of the events from which these records originate are indicated in Figure 5.

Orientation of KiK-net Surface and Borehole Sensors

As a general rule, the orientation of the horizontal components of the K-NET and KiK-net sensors is north–south (NS) and east–west (EW). However, as indicated on the KiK-net website, deviations do occur, and this problem may concern both surface and borehole sensors. Depending on the type of analysis (especially if surface and borehole data are being used together, for instance, for the calculation of surface-to-borehole [S/B] spectral ratios or deconvolution of downhole/surface recording pairs, as we discuss later in this article), these deviations should be accounted for.

Because, at each KiK-net (respectively, Hi-net) site, the three borehole sensors (short-period seismometer, accelerometer, and tiltmeter) share the same casing (Fujiwara *et al.*, 2004), they have identical orientation of the components within the casing. In order to estimate the azimuth deviation of the borehole sensor's NS component from true north, Shiomi *et al.* (2003) calculated cross-correlations of the long-period waveforms (100–200 s period) for several large teleseismic earthquakes observed by the tiltmeters of the Hi-net stations with seismograms recorded at nearby broadband sensors of F-net. This information has been compiled for all Hi-net stations and can be obtained from the Hi-net webpage (see Data and Resources section).

This approach using the tiltmeter data, however, does not provide information on the relative orientation mismatch between borehole and surface sensors. In order to nevertheless estimate this azimuth difference, Maeda *et al.* (2005) followed a similar technique based on correlation of long-period acceleration waveforms (bandpass-filtered between 0.02 and 0.2 Hz) from several strong local earthquakes recorded at the each KiK-net downhole/surface sensor pair on the island of Hokkaido. They performed a systematic rotation of the borehole records through the range of rotation angles ±180°, calculated the transverse component at surface and borehole records, and searched for the relative azimuth of the surface with respect to the borehole sensor (or vice-versa) which led to highest correlation between the waveforms. They found that

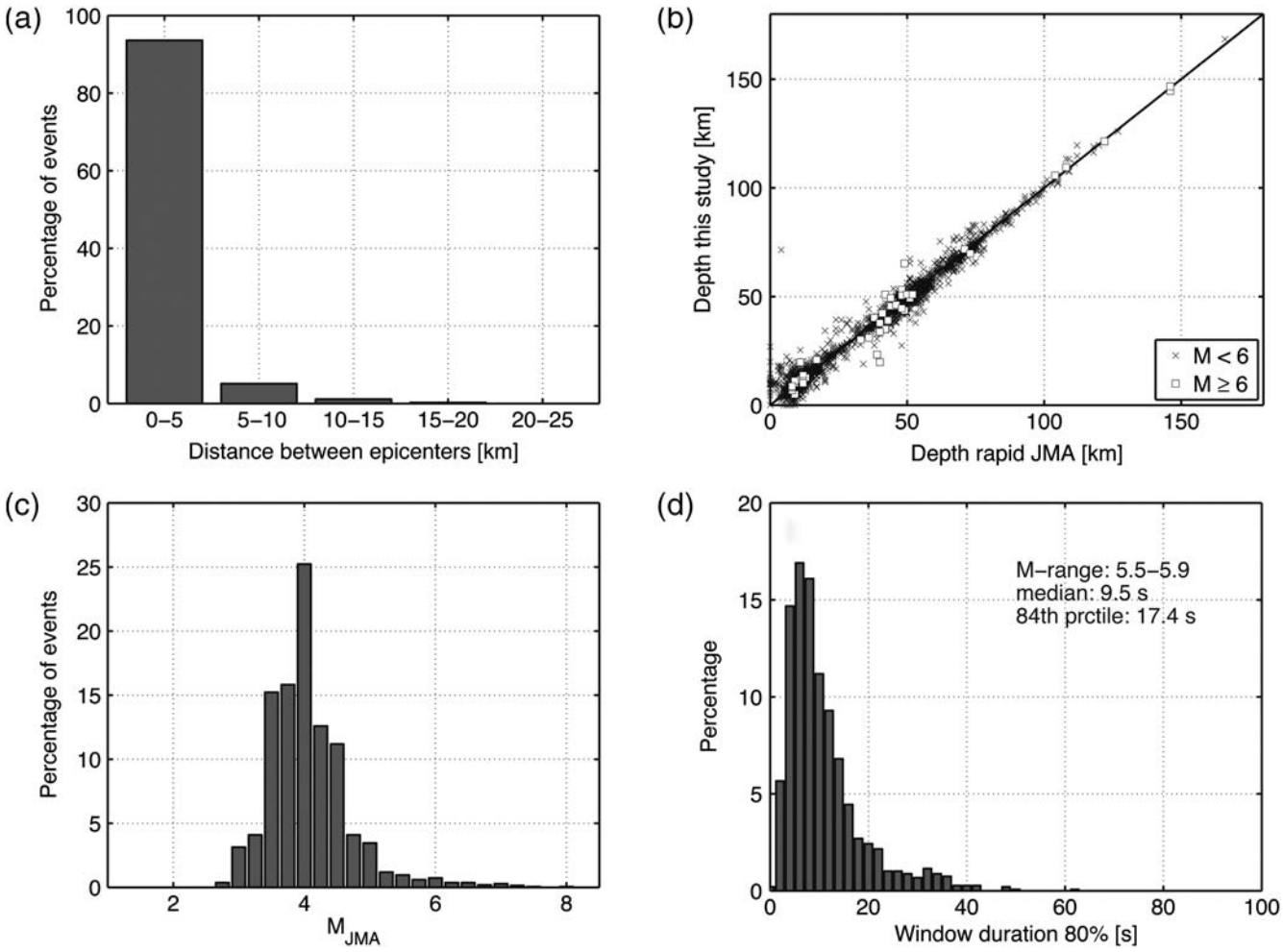

Figure 3. (a) Histograms showing the distance between the epicenters from this study and given in the K-NET and KiK-net database. (b) Comparison of depth determined in this study using the NLL software with the depth of the rapid JMA location provided in the K-NET and KiK-net file headers. (c) Percentage of events with given M_{JMA} in the data set. (d) Example for the analysis of durations of *S*-wave windows (in this case, M_{JMA} range 5.5–5.9 is shown). The given measure of duration is until 80% of the records total energy is released.

on the island of Hokkaido, the relative azimuth errors are generally smaller than 20°, but that 12 stations showed significantly stronger deviations and that the effect of correcting for the relative azimuth between the surface and borehole sensor orientation leads to a lower level of variability when calculating S/B spectral ratios, especially at low frequencies.

We followed the approach of Maeda *et al.* (2005) in order to characterize the relative azimuth mismatch between surface and borehole sensors at the KiK-net stations throughout all of Japan. At each station, we used only the available records from $M_{JMA} \geq 6$ earthquakes in our data set and only considered the determined azimuth to be reliable if the maximum correlation of the long-period acceleration waveforms was greater than 0.95; the results are in excellent agreement with the ones from Maeda *et al.* (2005) for stations on Hokkaido. The relative azimuth mismatch between the surface and borehole sensors is generally also smaller than 20° for the majority of the KiK-net sites throughout all of Japan even though larger deviations do occur and need to be taken into account. At all KiK-net stations where possible, we corrected for the azimuth mismatch between borehole and surface sensors for the remaining analysis.

Spectral Properties at a Glance

In order to get an overview of the spectral characteristics of the K-NET and KiK-net data sets and prepare the necessary input for the generalized inversion presented in the companion paper, we calculated the Fourier amplitude spectra (FAS) of the *S*-wave signals as well as preevent noise windows (when available). We then investigated the SNR and the frequency content of the signals.

Before calculating the FAS of the *S*-wave signals, it is necessary to decide upon the window lengths to be used such that the main *S*-wave energy is included and the amplitude spectra are not too heavily contaminated by the presence of surface waves. Common approaches, for instance, are to use signal windows that start 1 s before the *S*-wave onset time

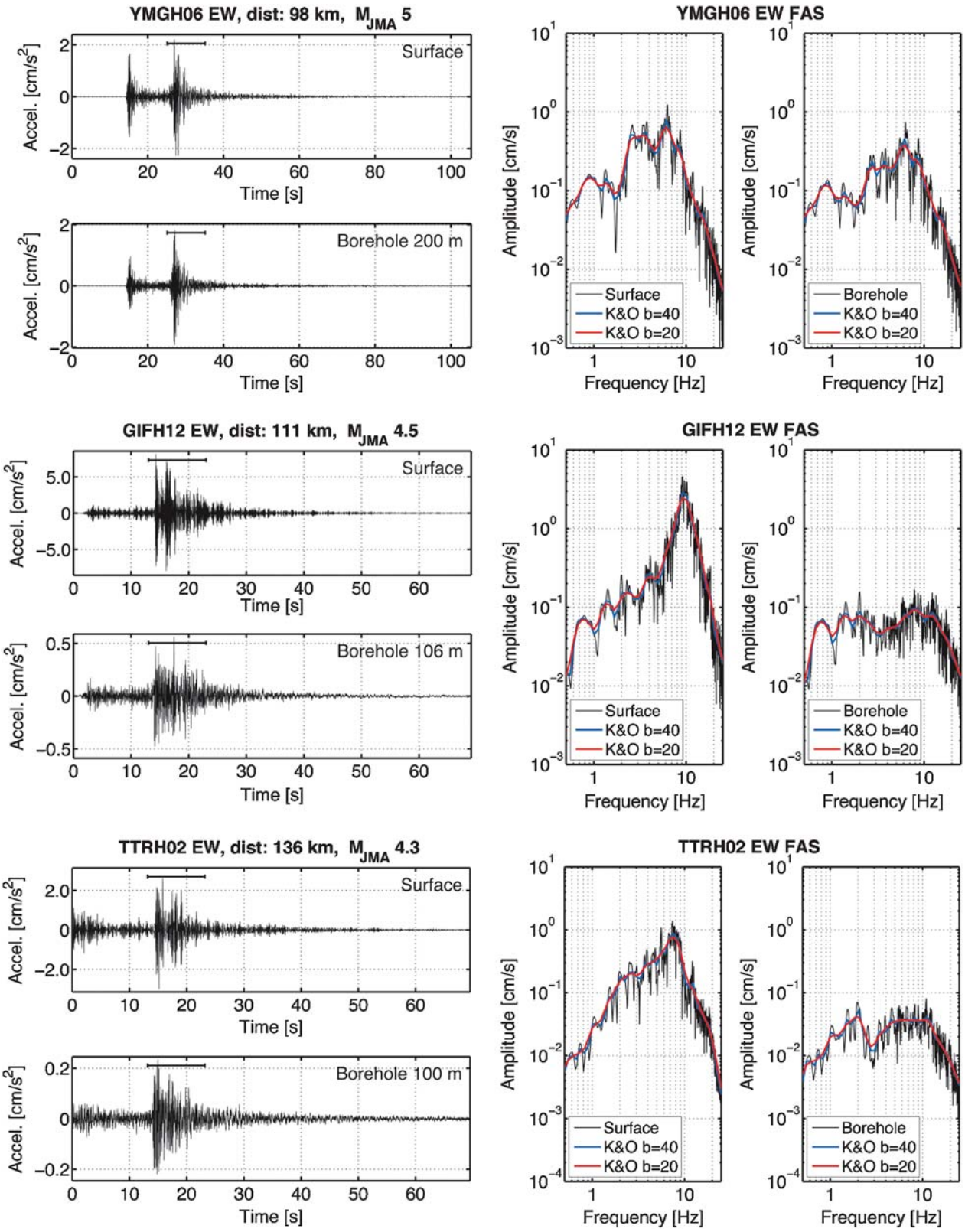

Figure 4. Three examples of the observed EW component accelerograms, showing surface and borehole traces (left panels) and the corresponding FAS (right panels). The smoothed spectra using the Konno–Ohmachi window with $b = 40$ and $b = 20$ are shown in blue and red, respectively, and the *S*-wave window used to calculate the FAS is indicated with the black bar over the time series on the left in each case. Note the trough in the borehole FAS of station TTRH02 between 2 and 5 Hz. The 7–8 Hz peak in the surface spectrum of this station may be related to the vibration characteristics of the observation house (Yoshimura *et al.*, 2003).

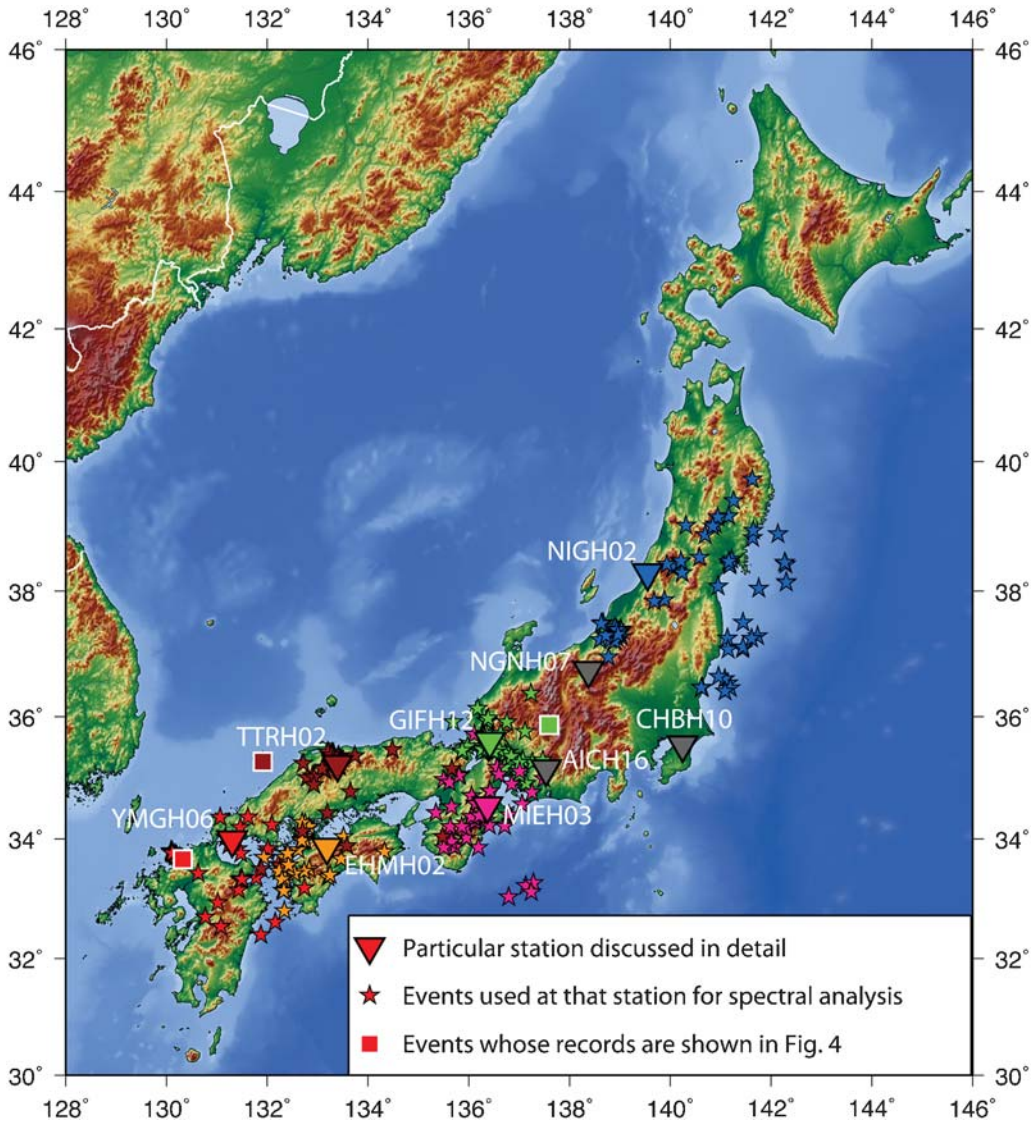

Figure 5. Location of six KiK-net stations (colored inverse triangles), for which the velocity profiles and waveform deconvolution results are exemplarily shown in Figures 8 and 9 (and which are also used as examples when discussing the generalized inversion results in the companion paper). For each of these stations, the respective earthquakes used for spectral analysis are indicated as stars with the respective color matching the station. The events for which the records are shown in Figure 4 are indicated as colored squares. The gray triangles show the deep borehole station CHBH10 (see also Fig. 10) and two additional stations AICH16 and NGNH07 mentioned in the companion paper.

and end when 80% of the record's total energy is reached (e.g., Parolai *et al.*, 2001; Oth *et al.*, 2009) or to determine an estimate of the strong-motion duration from the cumulative acceleration root-mean-square (rms) function (McCann and Shah, 1979).

We estimated the duration of *S*-wave windows following both approaches, and in general, the obtained results are in good agreement even though the estimates obtained from the cumulative rms function are usually somewhat smaller. In Figure 3d, the obtained 80% energy window lengths are shown for all records from events with $5.5 \leq M_{\mathrm{JMA}} \leq 5.9$. For most recordings in that magnitude range, the chosen *S*-wave window lengths are smaller than 20 s, with the eighty-fourth percentile being 17.2 s. Excessively long estimated *S*-wave window durations often result from the presence of secondary events in the considered recordings. Therefore, we set a maximum allowed window length for different M_{JMA} ranges, choosing a value slightly larger than the eighty-fourth percentile of the distribution. On the other hand, we also use a minimum window length of 10 s to ensure acceptable resolution at low frequencies, forcing us to include some degree of surface wave energy in our windows. Maximum allowed window lengths are 15 s for $M_{\mathrm{JMA}} < 5$, 20 s for $5 \leq M_{\mathrm{JMA}} < 6$, 25 s for $6 \leq M_{\mathrm{JMA}} < 7$, 30 s for $7 \leq M_{\mathrm{JMA}} < 7.5$, and 40 s for the Tokachi-Oki event ($M_{\mathrm{JMA}} = 8$).

In the following (and also in the companion paper), we only use records of weak and not-too-strong ground motions

with peak ground accelerations (PGAs) lower than 0.2*g* in order to reduce the risk of potential nonlinear site response that might bias our results. The selected windows were tapered using a 5% cosine taper (in order to ensure that we do not taper any part of the main *S*-wave pulse, we cut an *S*-wave window starting 2 s instead of 1 s before the *S* arrival for records from $M_{JMA} \geq 6$ events). FAS were then calculated for all three components of each recording and smoothed around 40 frequency points equidistant on a logarithmic scale between 0.2 and 40 Hz using the Konno and Ohmachi (1998) windowing function with $b = 40$ (both $b = 20$, stronger smoothing, and $b = 40$, weaker smoothing, were tested, leading to only small differences, see Fig. 4).

For all records where a preevent window of equal length as the signal window was available, we calculated SNRs for the frequency range 0.2–40 Hz. Figure 6 shows, at each frequency, the percentage of the number of recordings for which the SNR is larger than 3, both for *P*- and *S*-wave windows of the horizontal components (the *P*-wave windows were chosen to start 1 s before the *P*-wave onset and end 1 s before the *S*-wave onset). As expected, the *P*-wave windows generally show low SNR at frequencies below 1–2 Hz because their frequency content is usually shifted to higher frequencies compared with *S* waves. Furthermore, more than 75% and 95% of the *S*-wave records have an SNR ≥ 3 at 0.5 Hz and 1 Hz, respectively. Considering these results and the fact that the instruments' amplitude transfer function starts to decrease significantly for frequencies above about 25–30 Hz, the frequency range with optimal data quality is 0.5–25 Hz, and we reapplied the Konno and Ohmachi (1998) smoothing filter for 40 frequency points in the range 0.5–25 Hz. The right-hand panels in Figure 4 show the FAS in the latter frequency range.

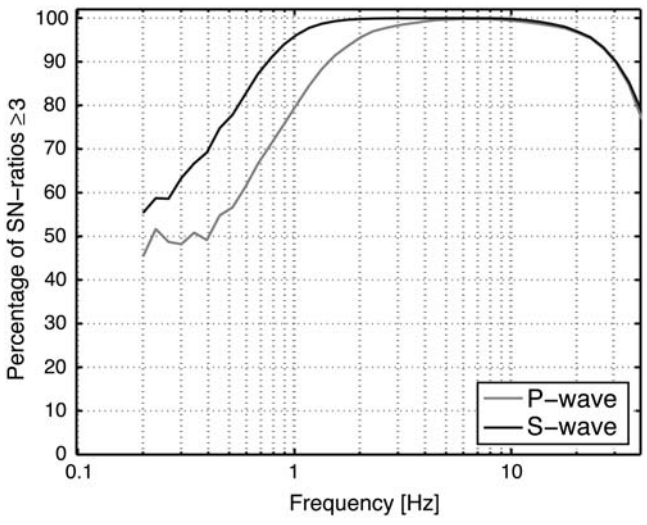

Figure 6. Evaluation of SNRs for the surface recordings of the data set. For each frequency, the percentage of SNRs larger than 3 of all records for which enough preevent noise was available for an evaluation is shown.

Prominent features that can be seen directly by looking at the spectra of the records depicted in Figure 4 are, for instance, the strong amplification peak at station GIFH12 in the frequency range around 9 Hz and the fact that the surface and borehole recordings show practically identical amplitudes at frequencies below 1 Hz. Furthermore, troughs are visible in the borehole spectra (for instance, for GIFH12, between 2 and 7 Hz or, in the case of TTRH02, between 2 and 5 Hz), indicating the potential presence of downgoing waves, which we investigate in more detail in the next section (Borehole Data and Downgoing Wave Propagation).

In order to further highlight the strong high-frequency amplification effects observed for many stations at the surface sensors, we performed a time-frequency analysis of the *S*-wave windows using the *S* transform (Stockwell *et al.*, 1996; Parolai, 2009). Figure 7 shows the results for the EW component record at station GIFH12, already depicted in Figure 4 (middle left panel). The color scales of the *S*-transform coefficients (bottom panels) are normalized to unity, such that relative shifts in frequency content between the borehole and surface sensor are highlighted. In the borehole recordings, the main signal energy is covering the frequency range 6–20 Hz, and some lower frequency components are also recognizable. In contrast, the waveform recorded at the surface is almost monochromatic, with all signal energy shifted into the much narrower frequency band 8–12 Hz (compare with FAS shown in Fig. 4, amplification of a factor 30 as compared with the borehole record). This observation is consistently made at a significant number of stations, showing that especially high-frequency amplification effects are severe in many cases.

Figure 8 shows the velocity profiles of six selected stations that provide a good sample of the general effects observed in the data set and whose locations are depicted in Figure 5. As expected, stations with strong high-frequency peaks in the horizontal component surface spectra often show interfaces with significant velocity contrasts, and depending on the depth of these and the seismic velocities within the layers, the peak resonance frequency of the site is shifted to lower or higher frequencies. However, as shown in the companion paper, amplification effects of up to a factor of 100 are observed, which are difficult to explain with the velocity contrasts in the *PS*-logging profiles alone; other factors contributing to these observations may include high-frequency surface waves induced within a near-surface, thin, and soft layer unresolved in the *PS*-logging data (Maeda, 2004 showed, using microtremor array measurements, that at the KiK-net station TTRH02, the seismic velocities near the surface are likely to be lower than those derived by *PS* logging) and potential soil-structure interaction with the station shed (Yoshimura *et al.*, 2003; Maeda, 2004).

These general first-order observations are corroborated by the calculation of horizontal-to-vertical and S/B spectral ratios for all stations, which will be discussed in detail in

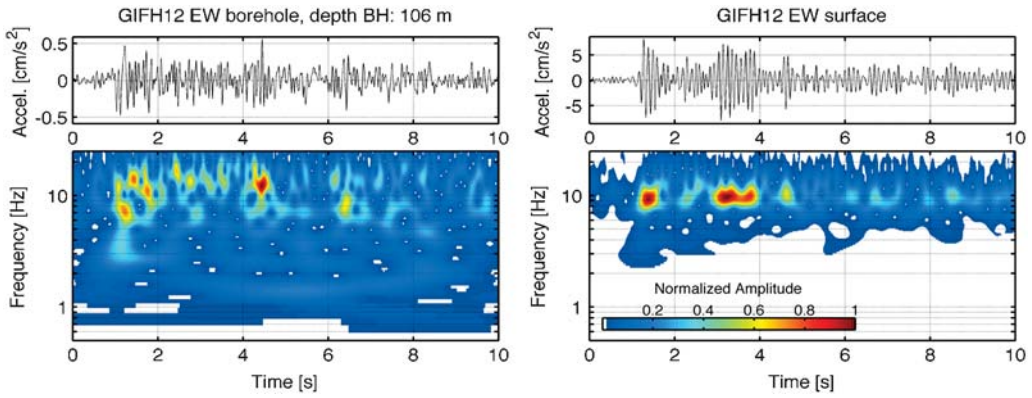

Figure 7. Time-frequency analysis via *S* transform for the record of an M_{JMA} 4.5 earthquake at station GIFH12 (see also Fig. 4, middle panel). (a) Borehole record and its *S* transform. (b) Surface record and its *S* transform. The color scale represents the normalized amplitude of the *S* transform in each case (i.e., amplitudes relative to the maximum present in the trace, see also color bar in the lower right plot). Note the almost monochromatic nature of the surface recording due to strong amplification effects at a frequency of around 9–10 Hz.

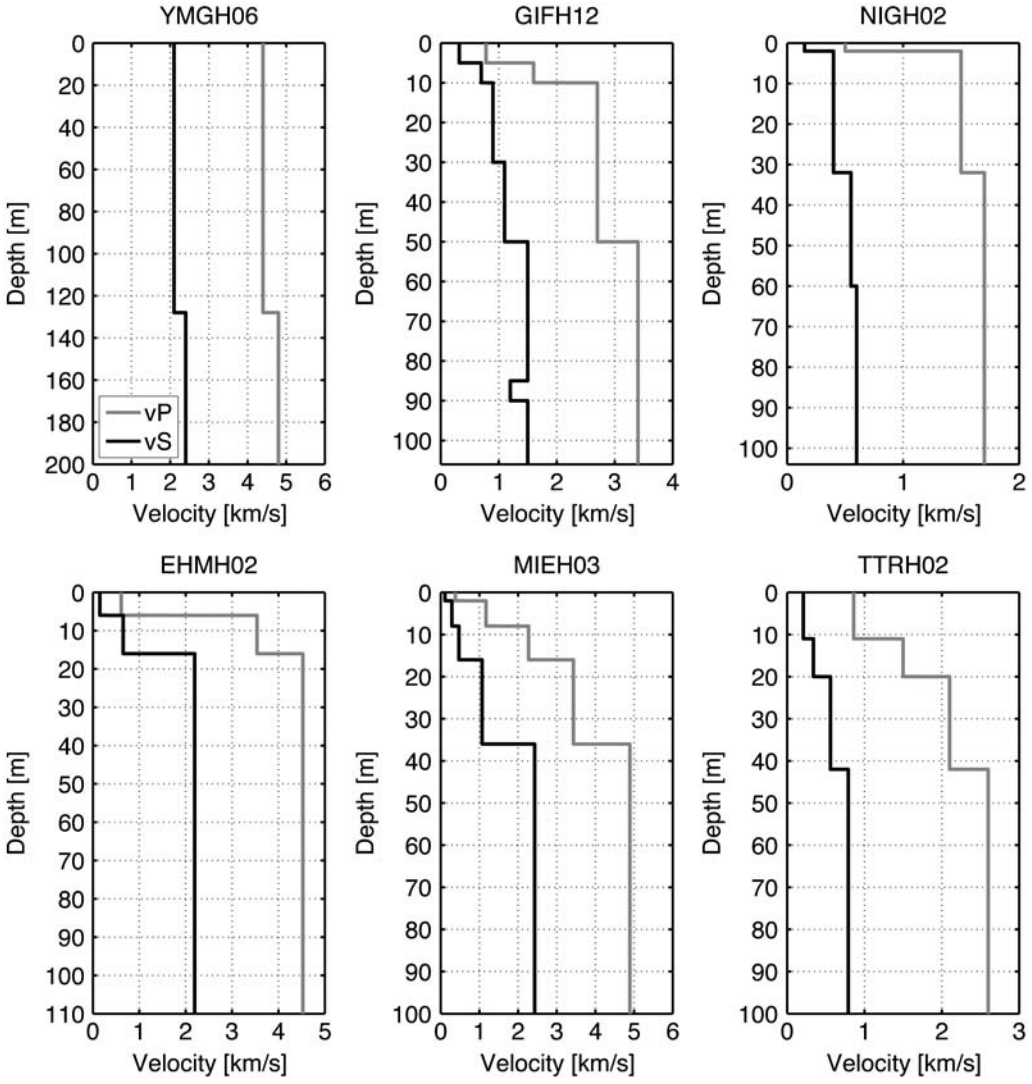

Figure 8. *P*- (gray) and *S*-wave (black) velocity profiles at the six sites discussed in Figure 9.

conjunction with the site response functions derived through generalized spectral inversion in the companion paper.

Borehole Data and Downgoing Wave Propagation

Finally, we address the issue of how far the recordings of the KiK-net borehole stations are affected by downgoing waves. As a general rule, the downhole recordings cannot be considered equivalent to input motion for a rock site because they are, in principle, affected by reflected waves from the free surface or from interfaces with strong impedance contrasts, as are present at many KiK-net sites. Safak (1997) discussed the effect of these downgoing waves on the S/B spectral ratios, and Parolai *et al.* (2009) deconvolved uphole/downhole recording pairs from one another in order to investigate the propagation of up- and downgoing waves at the Ataköy vertical array in Istanbul. They found that downgoing waves could be clearly identified down to the deepest sensor at 140 m depth.

Assimaki *et al.* (2008) studied in depth the seismograms from the Miyaki-Oki aftershock sequence recorded at a subset of 38 KiK-net stations on northern Honshu. They performed an inversion for the velocity and density profiles of the studied sites and analyzed, amongst others, the effects of downgoing wave propagation on the S/B spectral ratios, coming to the conclusion that downgoing-wave effects are reduced at stations with large impedance contrasts at shallow depth, which lead to a trapping of energy in the near-surface layers. With the waveform deconvolution results presented in the following paragraphs, we do not intend to rival the detailed nature with respect to potential engineering applications of the Assimaki *et al.* (2008) study. Rather, we provide the reader with robust first-order information on whether or not, as a general characteristic, severe downgoing wave effects need to be expected at the KiK-net borehole sensors. The influence of these downgoing waves on the S/B and, consequently, on the suitability of the latter as site response estimators is discussed more closely in the companion paper.

We follow the deconvolution approach based on the iterative (unconstrained) Landweber regularization technique presented in Parolai *et al.* (2009) and Bindi *et al.* (2010), and for a detailed discussion of the algorithm and regularization of the solution, we refer the reader to these publications. We tested several values for the iteration number and found that $n = 1000$ represented a good compromise between maintaining acceptable resolution power and guaranteeing stability. For each surface/borehole recording pair at a particular station, we deconvolved the surface *S*-wave recording of the transverse component of ground motion (considering the azimuth difference between surface and borehole sensors as discussed earlier) from the borehole recording and stacked the deconvolved waveforms recorded from several events to obtain an estimate of the propagator function within the borehole.

The results of our analysis are shown for seven example stations (six boreholes with depths between 100 and 200 m, Fig. 9, and one deep borehole of 2-km depth, Fig. 10), which provide a good overview of the typically observed effects. For each propagator, the time origin (i.e., zero time) corresponds to simultaneous arrival times at the borehole and surface sensors. Therefore, the pulse corresponding to the upgoing wave propagation will be visible in the acausal part, while the pulse corresponding to downgoing waves reflected from the surface will be located in the causal part of the deconvolved waveform. Pulses corresponding to downgoing waves generated at interfaces with high impedance contrasts may be located, depending on the up- and downgoing wave travel time to/from that interface, both in the acausal or causal part. In Figure 9, the estimated arrival times of up- and downgoing waves related to reflection from the surface, calculated using the velocity profiles shown in Figure 8, are indicated as red dashed lines. Furthermore, in order to be able to assess pulses related to downgoing waves from interfaces with strong velocity contrasts, we also plotted the estimated arrival times of reflected waves from interfaces with *S*-wave velocity contrasts larger than 0.3 km/s as blue dashed lines.

A first observation that can be made on these propagators is that their structure is generally rather complex, with multiple peaks in many cases. The up- and downgoing waves related to the surface reflection (symmetric with respect to time origin) are clearly visible in most cases (e.g., YMGH06 or NIGH02). For some stations, peaks can be well associated with downgoing waves from interfaces, for instance, at stations EHMH02 or GIFH12, while for others, the nature of the complex propagator remains elusive (e.g., TTRH02). One possible source of uncertainty could exist in the simplification of the velocity structure (Fig. 8) from the original *PS*-logging data, potentially affecting the calculation of expected arrival times of wave propagation pulses in the deconvolved traces, and soil-structure interactions with the observation sheds may play a role as well (Yoshimura *et al.*, 2003; Maeda, 2004). Furthermore, Parolai *et al.* (2009) presented evidence for the potential presence of downgoing mode conversion (*S*-*P* conversions) in the borehole records at the Ataköy vertical array in Istanbul. A significant number of KiK-net stations, such as EHMH02 or GIFH12 in Figure 8, show very pronounced *P*-wave velocity contrasts, which have the potential to facilitate such conversions, and some of the peaks in the deconvolved time series may be due to this effect. Station YMGH06 shows a strong peak exactly at time zero, which cannot be related to any particular interface and could potentially be linked to the presence of surface waves, but further analysis on the polarization is necessary to corroborate this hypothesis.

As the examples in Figure 9 show, nonnegligible downgoing wave effects are generally expected for the KiK-net borehole recordings, and we will investigate their effects on ground-motion spectra using generalized spectral inversion in the companion paper. Even within deep boreholes,

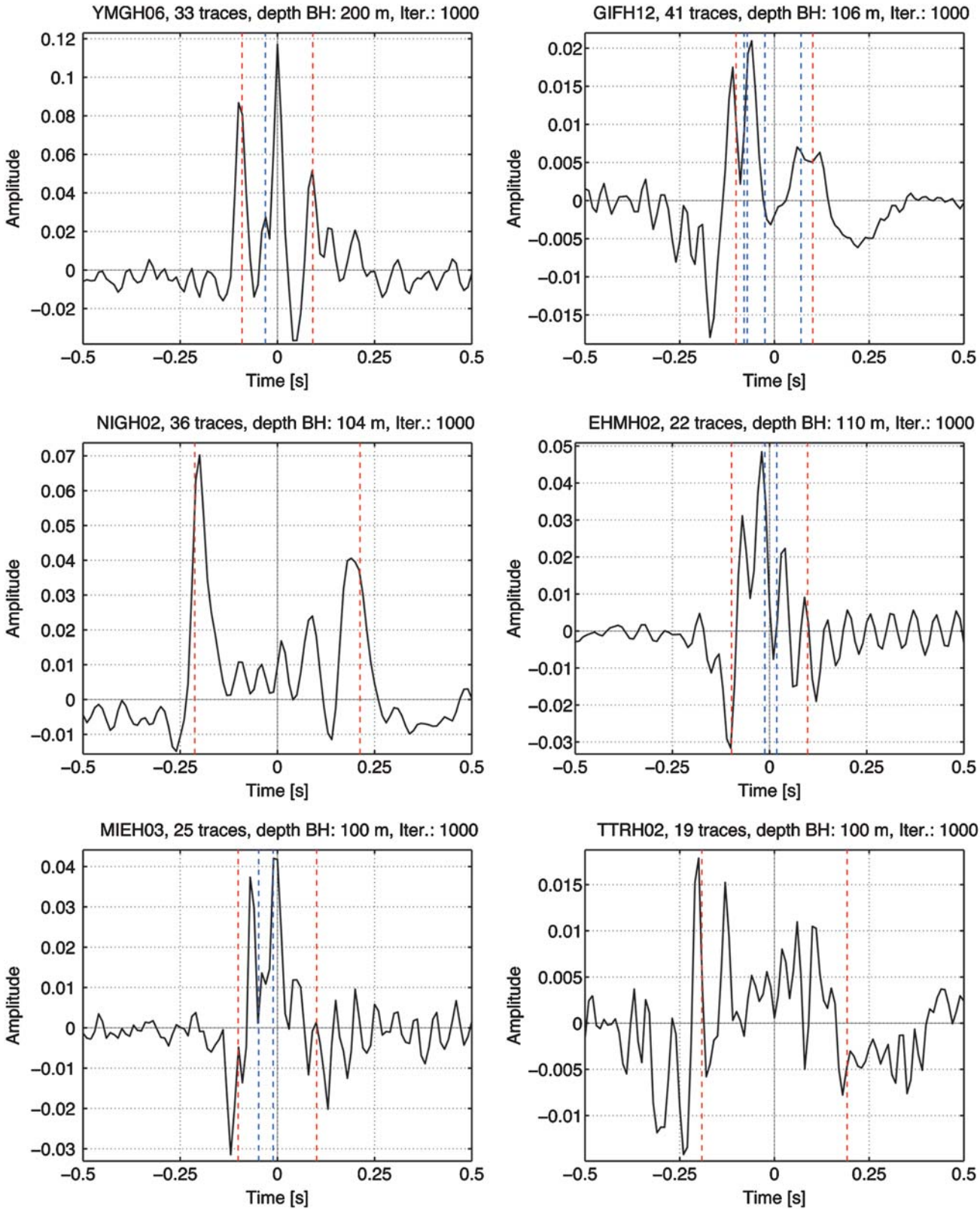

Figure 9. Propagator time functions between borehole and surface sensors obtained by deconvolving the surface transverse *S*-wave recording from the borehole recording at six example stations. Shown are the stacked propagator functions for several events recorded at each station. The expected arrival times for up- and downgoing (due to reflection at the free surface) *S* waves, estimated from the velocity profiles, are indicated as red dashed lines. Estimated arrivals of downgoing waves due to reflections at interfaces with large velocity contrasts (only velocity contrasts larger than 0.3 km/s are considered) are indicated as blue dashed lines.

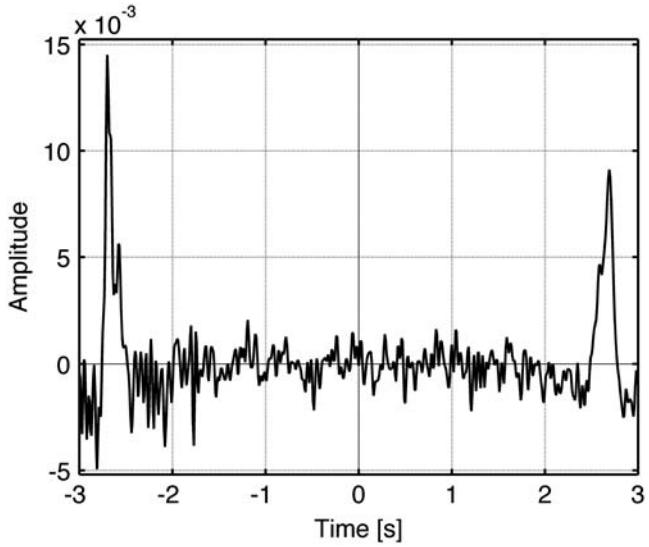

Figure 10. Propagator time function calculated at station CHBH10 (for the station location, see Fig. 5). Shown is the stacked propagator function for 51 events recorded at this station. The borehole depth at CHBH10 is 2 km, and the number of iterations used is 1000.

downgoing waves can play a significant role. Figure 10 shows the propagator function derived at KiK-net station CHBH10, with the borehole sensor located at a depth of 2 km. We do not have velocity profile information, nor was it possible to correct the azimuth difference in sensor orientation between surface and borehole at that site; nevertheless, we clearly observe a downgoing wave pulse with amplitude larger than half of the upgoing one. Even though at other deep boreholes, we did not observe either up- or downgoing wave pulses, hinting at a complete decoupling of the recordings, this result proves that the influence of downgoing waves is not necessarily limited to shallow borehole sites only.

Conclusions

For more than a decade, the extraordinary strong-motion networks in Japan, K-NET and KiK-net, are providing high-quality digital data with an unprecedented density of recording sites, and the availability of these data sets enables scientists to study a large variety of seismological problems. However, for most applications, a detailed analysis of the basic data set properties is indispensable before applying more complex analysis techniques. These properties range from the availability of *P*- and *S*-wave onsets and the investigation of event location reliability to a fundamental understanding of the spectral characteristics of the recordings and, as far as borehole recordings are concerned, the wave-field propagation characteristics within the boreholes.

To this end, we extracted a large data set of more than 78,000 records from 2201 earthquakes recorded at 1681 sites throughout Japan from the K-NET and KiK-net databases (covering the recording period May 1996–February 2009) based on the reliability of the automatic *P*- and *S*-wave picking and location accuracy of the events. We find that, following this approach, a highly robust data set of *P*- and *S*-wave windows can be extracted in an automated way, minimizing the mixing of *P*- and *S*-wave energy for spectral analysis and leading to an immediate removal of records with insufficient SNR because, for excessively noisy recordings, inconsistent picks are usually obtained on the different components of ground motion, and our approach for checking their reliability proved to be very successful in discarding these traces right away.

Furthermore, we have systematically checked and accounted for the relative azimuth mismatch in sensor orientation between surface and borehole sensors, which has not been done in a vast majority of studies based on KiK-net records, and analyzed the general spectral properties of the retained recordings. Many sites show strong amplification effects at frequencies higher than 10 Hz that are already visible directly on the ground motion spectra.

We also confirmed through deconvolution of downhole/surface recording pairs that in general, downgoing waves are clearly visible in the KiK-net borehole data. The obtained propagator functions characterizing the wave field within the borehole show a strong degree of complexity in many cases, with reflections both from the surface and high impedance contrast interfaces and, possibly, downgoing *S*-*P* conversions as well. Even though, as discussed in the companion paper, this downgoing wave contamination turns out to be of minor amplitude when directly compared to the site amplification effects of many surface stations, it may need to be taken into account depending on the application for which the borehole records are being used. For instance, our results show that the latter cannot simply be assumed to be representative for input motions on rock conditions in engineering applications, such as the derivation of GMPEs. Downgoing wave effects have not been taken into account in many past studies using KiK-net borehole data (e.g., Cotton *et al.*, 2008; Cadet *et al.*, 2010), and the recommendation that arises from our Japan-wide study is that authors should verify the potential importance of these downgoing waves with respect to the specific study goals.

The data set presented in this work and the gained insights into the effects inherent in the K-NET and KiK-net recordings provide a robust basis for further investigations, and in the companion paper, we apply a nonparametric generalized inversion approach in order to separate attenuation characteristics, source spectra, and site amplification functions throughout Japan. Other applications, such as the analysis of interevent and interstation variability in GMPEs, are also a matter of ongoing research. We will also continuously update the data set with recent K-NET and KiK-net data using the procedures detailed in the preceding paragraphs.

Data and Resources

The seismograms and geotechnical data used in this study were collected from the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan. Data can be obtained from the K-NET, KiK-net, and Hi-net web sites at www.k-net.bosai.go.jp, www.kik.bosai.go.jp, and www.hinet.bosai.go.jp, respectively, (last accessed April 2010). The NonLinLoc software package used for event location can be obtained at http://alomax.free.fr/nlloc/ (last accessed January 2011). A table with the relative azimuth differences between surface and borehole sensor horizontal component orientation for the KiK-net stations and the time domain propagator functions for all analyzed stations is available from the first author upon request.

Acknowledgments

We wish to thank the National Research Institute for Earth Science and Disaster Prevention (NIED) for making the K-NET and KiK-net data available, Anthony Lomax for sharing the NonLinLoc software, and Bogdan Enescu for making the autopicker routine available to us. We are in debt to Katsuhiko Shiomi and Takashi Furumura for their friendly assistance, providing valuable information on the orientation of borehole/surface sensors and the JMA2001 velocity model as well as helping us to find our way on the Japanese Hi-net website. Thanks are also due to Claudio Satriano for helpful discussions and advice concerning the NonLinLoc software and earthquake location in general. Associate Editor Zhigang Peng and the anonymous reviewers provided constructive comments that helped to improve the manuscript. This publication was supported by the National Research Fund, Luxembourg (FNR/10/AM4/25).

References

Assimaki, D., W. Li, J. H. Steidl, and K. Tsuda (2008). Site amplification and attenuation via downhole array seismogram inversion: A comparative study of the 2003 Miyagi-Oki aftershock sequence, *Bull. Seismol. Soc. Am.* **98,** 301–330.

Bindi, D., S. Parolai, M. Picozzi, and A. Ansal (2010). Seismic input motion determined from a surface-downhole pair of sensors: Aconstrained deconvolution approach, *Bull. Seismol. Soc. Am.* **100,** 1375–380.

Brown, H. M., R. M. Allen, and V. F. Grasso (2009). Testing ElarmS in Japan, *Seismol. Res. Lett.* **80,** 727–739.

Cadet, H., P. Y. Bard, and A. Rodriguez-Marek (2010). Defining a standard rock site: Propositions based on the KiK-net database, *Bull. Seismol. Soc. Am.* **100,** 172–195.

Choy, G. L., and J. Boatwright (2009). Differential energy radiation from two earthquakes in Japan with identical M_w: The Kyushu 1996 and Tottori 2000 earthquakes, *Bull. Seismol. Soc. Am.* **99,** 1815–826.

Cotton, F., G. Pousse, F. Bonilla, and F. Scherbaum (2008). On the discrepancy of recent European ground-motion observations and predictions from empirical models: Analysis of KiK-net accelerometric data and point-sources stochastic simulations, *Bull. Seismol. Soc. Am.* **98,** 2244–261.

Earle, P. S., and P. M. Shearer (1994). Characterization of global seismograms using an automatic-picking algorithm, *Bull. Seismol. Soc. Am.* **84,** 366–376.

Edwards, B., and A. Rietbrock (2009). A comparative study on attenuation and source-scaling relations in the Kanto, Tokai, and Chubu regions of Japan, using data from Hi-net and KiK-net, *Bull. Seismol. Soc. Am.* **99,** 2435–460.

Fujiwara, H., S. Aoi, T. Kunugi, and S. Adachi (2004). Strong-motion observation networks of NIED: K-NET and KIK-net, in *Cosmos Report*, National Research Institute for Earth Science and Disaster Prevention, Japan.

Fukushima, Y., K. Irikura, T. Uetake, and H. Matsumoto (2000). Characteristics of observed peak amplitude for strong ground motion from the 1995 Hyogoken Nanbu (Kobe) earthquake, *Bull. Seismol. Soc. Am.* **90,** 545–565.

Ide, S., M. Takeo, and Y. Yoshida (1996). Source process of the 1995 Kobe earthquake: Determination of spatio-temporal slip distribution by Bayesian modeling, *Bull. Seismol. Soc. Am.* **86,** 547–566.

International Code Council (2006). *International Building Code (IBC)*, International Code Council, Falls Church, Virginia.

Iwasaki, T., T. Yoshii, T. Ito, H. Sato, and N. Hirata (2002). Seismological features of island arc crust as inferred from recent seismic expeditions in Japan, *Tectonophysics* **355,** 53–66.

Konno, K., and T. Ohmachi (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. Seismol. Soc. Am.* **88,** 228–241.

Lomax, A. (2005). A reanalysis of the hypocentral location and related observations for the great 1906 California earthquake, *Bull. Seismol. Soc. Am.* **95,** 861.

Lomax, A., J. Virieux, P. Volant, and C. Berge (2000). Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations, In *Advances in Seismic Event Location*, Amsterdam: Kluwer, 101–134.

Lomax, A., A. Zollo, P. Capuano, and J. Virieux (2001). Precise, absolute earthquake location under Somma–Vesuvius volcano using a new three-dimensional velocity model, *Geophys. J. Int.* **146,** 313–331.

Maeda, T. (2004). Observatory shed effect on strong motion records identified by micro-tremor measurement, in *Proc. of the Thirteenth World Conf. on Earthq. Eng.*, Vancouver, British Columbia, Canada, Paper No. 568.

Maeda, T., T. Sasatani, N. Takai, and G. Shimizu (2005). Azimuth estimation of KiK-net surface seismometers deployed in Hokkaido, Japan, *Geophys. Bull. Hokkaido Univ.* **68,** 141–152 (in Japanese with English abstract).

McCann, M. W., and H. C. Shah (1979). Determining strong-motion duration of earthquakes, *Bull. Seismol. Soc. Am.* **69,** 1253–265.

Moya, A., and K. Irikura (2003). Estimation of site effects and Q factor using a reference event, *Bull. Seismol. Soc. Am.* **93,** 1730–745.

Okada, Y., K. Kasahara, S. Hori, K. Obara, S. Sekiguchi, H. Fujiwara, and A. Yamamoto (2004). Recent progress of seismic observation networks in Japan-Hi-net, F-net, K-NET and KiK-net, *Earth Planets Space* **56,** xv–xxviii.

Oth, A., D. Bindi, S. Parolai, and D. Di Giacomo (2011). Spectral analysis of K-NET and KiK-net data in Japan, Part II: On attenuation characteristics, source spectra, and site response of borehole and surface stations, *Bull. Seismol. Soc. Am.* **101,** 2 667–687.

Oth, A., S. Parolai, D. Bindi, and F. Wenzel (2009). Source spectra and site response from S waves of intermediate-depth Vrancea, Romania earthquakes, *Bull. Seismol. Soc. Am.* **99,** 235–254.

Parolai, S. (2009). Denoising of seismograms using the S transform, *Bull. Seismol. Soc. Am.* **99,** 226–234.

Parolai, S., A. Ansal, A. Kurtulus, A. Strollo, R. Wang, and J. Zschau (2009). The Ataköy vertical array (Turkey): Insights into seismic wave propagation in the shallow-most crustal layers by waveform deconvolution, *Geophys. J. Int.* **178,** 1649–662.

Parolai, S., D. Bindi, and L. Trojani (2001). Site response for the RSM seismic network and source parameters in the Central Apennines (Italy), *Pure Appl. Geophys.* **158,** 695–715.

Safak, E. (1997). Models and methods to characterize site amplification from a pair of records, *Earthq. Spectra* **13,** 97–129.

Satriano, C., A. Lomax, and A. Zollo (2008). Real-time evolutionary earthquake location for seismic early warning, *Bull. Seismol. Soc. Am.* **98,** 1482–494.

Sawazaki, K., H. Sato, H. Nakahara, and T. Nishimura (2006). Temporal change in site response caused by earthquake strong motion as

revealed from coda spectral ratio measurement, *Geophys. Res. Lett.* **33,** L21303, doi 10.1029/2006GL027938.

Sawazaki, K., H. Sato, H. Nakahara, and T. Nishimura (2009). Time-lapse changes of seismic velocity in the shallow ground caused by strong ground motion shock of the 2000 Western-Tottori earthquake, Japan, as revealed from coda deconvolution analysis, *Bull. Seismol. Soc. Am.* **99,** 352–366.

Shiomi, K., K. Obara, S. Aoi, and K. Kasahara (2003). Estimation on the azimuth of the Hi-net and KiK-net borehole seismometers, *Zisin* **56,** 99–110 (in Japanese).

Stockwell, R. G., L. Mansinha, and R. P. Lowe (1996). Localization of the complex spectrum: The *S* transform, *IEEE Trans. Signal Process.* **44,** 998–1001.

Ueno, H., S. Hatakeyama, T. Aketagawa, J. Funasaki, and N. Hamada (2002). Improvement of hypocenter determination procedures in the Japan Meteorological Agency, *Quart. J. Seism.* **65,** 123–134 (in Japanese with English abstract).

Yoshimura, C., H. Hibino, Y. Uchiyama, T. Maeda, N. Kurauchi, and S. Aoi (2003). Vibration characteristics of the observation house at KiK-net Hino, in *2003 Japan Earth and Planetary Science Joint Meeting* (in Japanese with English abstract).

Wu, C., Z. Peng, and D. Assimaki (2009). Temporal changes in site response associated with strong ground motion of 2004 M_w 6.6 Mid-Niigata earthquake sequences in Japan, *Bull. Seismol. Soc. Am.* **99,** 6, 3487–3495, doi 10.1785/0120090108.

Wu, C., Z. Peng, and Y. Ben-Zion (2010). Refined thresholds for nonlinear ground motion and temporal changes of site response associated with medium size earthquakes, *Geophys. J. Int.* **183,** 1567–1576, doi 10.1111/j.1365-246X.2010.04704.x.

European Center for Geodynamics and Seismology
19, rue Josy Welter
L-7256 Walferdange
Grand-Duchy of Luxembourg
adrien.oth@ecgs.lu
(A.O.)

Earthquake Risk and Early Warning Section
Helmholtz Centre Potsdam
GFZ German Research Centre for Geosciences
Telegrafenberg, 14473
Potsdam, Germany
parolai@gfz-potsdam.de
(S.P.)

Centre for Disaster Management and Risk Reduction Technology (CEDIM)
Helmholtz Centre Potsdam
GFZ German Research Centre for Geosciences
Telegrafenberg, 14473
Potsdam, Germany
bindi@gfz-potsdam.de
(D.B.)

Manuscript received 19 May 2010