CrossMark

Quantitative Analysis of GPR Signals: Transmitted Wavelet, Amplitude Decay, and Sampling-Related Amplitude Distortions

M. Dossi,[1] Emanuele Forte,[1] and M. Pipan[1]

Abstract—We study the importance of accurately recording signal amplitudes for the quantitative analysis of GPR data sets. Specifically, we measure the peak amplitudes of signals emitted by GPR antennas with different central frequencies and study their amplitude decay with distance, in order to extrapolate the peak amplitude of the wavelet initially transmitted by each antenna. The purpose is to compare the reference and reflected amplitudes in order to accurately estimate the subsurface EM impedance contrasts. Moreover, we study how sampling-related amplitude distortions can affect the quantitative analysis, and subsequently the resulting subsurface models, even in the absence of aliasing effects. The well-known Nyquist–Shannon theorem gives practical lower limits for the sampling rate in order to preserve the spectral content of a digitized signal; however, we show that it does not prevent possible amplitude distortions. In particular, we demonstrate that significant and unrecoverable loss of amplitude information occurs even at sampling rates well above the Nyquist–Shannon threshold. Interpolation may theoretically reduce such amplitude distortions; however, its accuracy would depend on the implemented algorithm and it is not verifiable in real data sets, since the actual amplitude information is limited to the sampled values. Moreover, re-sampling the interpolated signal simply reintroduces the initial problem, when a new sampling rate is selected. Our analysis suggests that, in order to limit the maximum peak amplitude error within 5%, the sampling rate selected during data acquisition must be at least 12 times the signal central frequency, which is higher than the commonly adopted standards.

Key words: Sampling, Nyquist–Shannon theorem, peak amplitude error, amplitude inversion, GPR, reference amplitude.

1. Introduction

Ground Penetrating Radar (GPR) is a non-invasive near-surface geophysical technique which uses EM waves to study the subsurface (Jol 2009). From a GPR data set, it is possible to extract both qualitative and quantitative information about the subsurface stratigraphy and the materials encountered by the recorded wavelets. The validity of the reconstructed subsurface model mainly depends on the particular algorithm used for the calculations, the assumptions made by the procedure with regards to the EM properties of the subsurface, as well as the accuracy of the measured input quantities. In this paper, we mainly focus on the latter issue, specifically on how inaccurate signal amplitudes can affect the reconstructed subsurface EM impedance contrasts and subsequently all the other properties that can be derived from them. The inversion procedure applied in this paper requires as input the peak amplitude of the recorded reflections, as well as the peak amplitude of the wavelet initially emitted by the transmitter, also referred to in the following as the *reference amplitude* (Forte et al. 2014; Dossi et al. 2016).

In this paper, we estimate the reference amplitude by analyzing the signal reflected from a metallic surface (Saarenketo and Scullion 2000; Al-Qadi and Lahouar 2005), recorded at various distances with different GPR antennas, assuming a total reflection and taking into account amplitude decay with distance, as well as possible noise and interference. This allows us to analyze the amplitude decay caused by spreading losses, while also making the calculation of the reference amplitude for each antenna statistically sound. However, before comparing the reference and reflected amplitudes in the calculation of the subsurface EM impedance contrasts, it is necessary to take into account possible differences in the radiation patterns (Jiao et al. 2000). Specifically, if the reference amplitude is measured from an air-coupled GPR system, while the recorded reflections are from a ground-coupled one, we have to take into account the

[1] Department of Mathematics and Geosciences (DMG), University of Trieste, Trieste, Italy. E-mail: eforte@units.it

refractive focusing associated with the air–ground interface (Jol 2009).

Besides correctly measuring the reference amplitude, the applied inversion algorithm requires that reflected amplitudes depend exclusively on the subsurface EM impedance contrasts (Forte et al. 2014; Dossi et al. 2016). Therefore, the signal processing must counteract the amplitude decay caused by wavefront expansion and possible intrinsic attenuation, while also removing potential signal distortions caused by either noise or interference. Another important issue is that the reflected amplitudes must be accurately sampled in the first place; therefore, possible sampling-related amplitude distortions of the analog signal must be taken into account.

Sampling is an irreversible process used during data acquisition in order to reduce analog signals, which contain an infinite number of values, into smaller and numerically manageable discrete series (Proakis and Manolakis 2006). The procedure causes an inevitable and unrecoverable loss of information between sampled values, which prevents the exact reconstruction of the input analog signal from the recorded discrete series and can cause significant signal distortions.

The general field of studies related to sampling and distortions is known as the *Rate–Distortion Theory*, and it was introduced in 1948 by Claude E. Shannon with his landmark paper (Shannon 1948). There are two different fields of application for the signal quantization theory, namely the *rate–distortion optimized quantization* and the *rounding quantization* (Gersho and Gray 1991). The former is used in source coding for "lossy" data compression algorithms, where the purpose is to manage distortions within the limits of a certain instrument or algorithm. The rounding quantization is instead used to enable a simple approximate representation of a certain quantity that we want to measure. This second quantization class includes the analog-to-digital conversion of a signal. The main objective of such conversion is to retain an adequate signal fidelity after the digitizing process, while at the same time avoiding unnecessary precision or data redundancy, and keeping the dynamic range of the signal within practical limits. All the examples and discussions in this paper are related to the rounding quantization category.

We can identify two different kinds of sampling error, which can cause significant distortions in the recorded data set with respect to the analog signal. The *quantization error* (or round-off error) is the difference between the actual (theoretical) analog value and the resulting digital one at a particular sample point (e.g. Proakis and Manolakis 2006), while the second kind of sampling error is related to the data loss between sampled values and the subsequent erroneous reconstruction of the original analog signal. An example is given by aliasing effects caused by the use of excessively low sampling rates, creating distortions in the spectral content of the recorded signal. In this paper, we are instead interested in *peak amplitude errors*, which also belong to this second class and can significantly affect the quantitative analysis of any recorded data set, even in the absence of aliasing effects. Considering a signal peak, we define as peak amplitude error the difference between the theoretical analog peak amplitude and the actual digitized peak amplitude, which can be shifted from the former by at most half the sampling interval.

In many geophysical applications dealing with the measurement of time-varying phenomena, the most important parameter is the discretization interval used during data acquisition, commonly referred to as the *sampling interval*. A digitized wave field is very similar to any other discrete data set, with the peculiarity of being a function of both space and time. Such data can be analyzed using appropriate signal processing (Yilmaz 2001), image processing (e.g. Chopra and Marfurt 2007), or even techniques originally developed for digital music recording (Dell'Aversana 2013, 2014). The sampling interval (Δt) has a crucial role in signal reconstruction, processing, and analysis, with the attention commonly focused on the spectral content of the recorded signal. The well-known Nyquist–Shannon theorem states that a continuous signal having a band-limited frequency spectrum, with a maximum frequency f_{max}, can be fully specified in a time interval of duration T by a set of $2Tf_{max}$ discrete values (Nyquist 1928; Shannon 1949). This means that such signal can be accurately reconstructed, at least in terms of its

spectral content, by using a *sampling rate* f_S equal to at least $2 \cdot f_{max}$:

$$f_S = \frac{1}{\Delta t}. \qquad (1)$$

In other words, in order to avoid aliasing effects, f_{max} has to be lower than the *Nyquist frequency* f_{Ny}, which is equal to half the sampling rate. A band-limited signal would extend infinitely in time; therefore, the Nyquist–Shannon theorem cannot be strictly applied to a recorded data set, since such signal has a finite time interval and, therefore, it is characterized by an infinite number of frequency components. Nevertheless, it is always possible to define a maximum frequency f_{max} above which the signal energy is negligible. The theorem can thus be considered a useful guide during data acquisition when selecting f_S, which has to be higher than $2 \cdot f_{max}$ in order to preserve the spectrum of the analog signal. For example, when dealing with signals characterized by a central frequency f_C, Jol (2009) recommends using a sampling rate equal to at least $6 \cdot f_C$ for GPR surveys, while Drijkoningen (2003) suggests using a sampling rate equal to at least $8 \cdot f_C$ as the rule of thumb for reflection seismics. The $2 \cdot f_{max}$ threshold for the sampling rate f_S is also known as the *Nyquist rate*, but it is referred to in this paper as the *Nyquist–Shannon threshold*, in order to avoid confusion with the Nyquist frequency f_{Ny}, which is half the sampling rate. The Nyquist frequency is more useful during quantitative analysis, since it can be easily compared to the highest frequency component f_{max} of the recorded signal, which constitutes a minimum threshold for f_{Ny}.

It is interesting to note that Nyquist describes in detail the frequency behavior of a digitized signal in its pioneering 1928 paper, but he disregards the amplitude behavior in the time domain. This is in part due to the fact that he worked on practical applications related to telegraph transmission, where the essential point is that a signal must be properly recorded in terms of frequencies (and phases). In fact, he explicitly wrote that "a criterion of perfect transmission is selected; and a discussion is given of the characteristics which the received wave must have to be nondistorting with the requirement that the frequency range should not be greater than necessary"

(Nyquist 1928). In other words, the main objective was avoiding any frequency (and phase) distortion, while accurate amplitude recording was less important for a telegraph system, as long as the signal-to-noise ratio of the recorded signal was acceptable. Similar observations can be made with regard to the fundamental classical works of Gabor (1946) and Shannon (1949) where, again, the focus is to assure a proper spectral representation or transformation of a digitized signal. Furthermore, it is interesting to note that Shannon (1949) pointed out the approximate formulation of the original Nyquist sampling theorem, while also providing its rigorous mathematical proof. Linvill (1949) further advanced the sampling theory, proving that sampling could be regarded as an amplitude modulation of a periodic "impulse carrier" by the signal being sampled. He also described interpolation as filtering in the frequency domain by means of an ideal low-pass filter (Widrow and Kollar 2008).

However, the effects that the selected sampling rate has on the sampled amplitudes deserve further analysis. In fact, the recorded amplitudes are essential input quantities for several geophysical signal analysis and processing algorithms, as well as for interpretation, and inversion procedures. Inaccurate sampling of amplitudes can affect, for instance, the reconstructed subsurface impedance contrasts, which can be estimated using various procedures, such as amplitude versus offset (AVO) analysis (Castagna and Backus 1994) or amplitude inversion algorithms (Tarantola 1984). While the aliasing problem is very well known and it has been thoroughly discussed in literature, to our knowledge the amplitude behavior as a function of the sampling rate has seldom been quantitatively analyzed in aliasing-free data sets.

In this paper, we analyze the effects of the sampling rate on the recorded amplitudes, not considering rounding or truncation errors, which are both well-known (e.g. Proakis and Manolakis 2006) and less important in this context. We show that the recorded amplitudes, and consequently all the other parameters derived from them, are strongly affected by the sampling rate. In particular, we give some examples of the maximum peak amplitude error as a function of the sampling rate for different signals, as well as examples of common mid-point (CMP) gather

analysis and amplitude inversion of synthetic common offset (CO) data sets. We conclude that the commonly used Nyquist–Shannon threshold does not necessarily guarantee an accurate amplitude representation of either basic or more complex digitized time series, with significant repercussions for their quantitative analyses. Our discussion focuses on GPR data, which are represented in terms of amplitude variations as a function of travel time (i.e. traces); nonetheless our conclusions are valid for any digitized signal.

2. Reference Amplitude Analysis

In order to obtain quantitative information about the subsurface from a GPR data set, it is essential to accurately characterize the EM wavelet emitted by the transmitting antenna, as well as its propagation through different media. A radar signal propagating through the ground is subjected to several processes which can affect its energy, shape, and travel path, and which depend on the EM properties of the subsurface (Jol 2009). These processes affect the recorded amplitudes, which are then compared with the reference amplitude in order to recover the subsurface impedance contrasts and subsequently all the related EM properties.

The two main mechanisms responsible for amplitude decay in a GPR data set are the intrinsic attenuation and the wavefront expansion, with variable contributions from partial reflections and scattering that, respectively, depend on the background-target contrasts and the coarseness of the medium. Intrinsic attenuation is mainly caused by subsurface EM properties like the electric conductivity and dielectric relaxation. Under favorable conditions such effects can be disregarded, for example air–ice mixtures can be considered low-loss materials with relaxation frequencies well outside the 10 MHz–2 GHz frequency range for GPR (Jol 2009). However, in the general case, the decay is not easily corrected through amplitude recovery, since the applied gain function depends on the assumptions made by the interpreter with regard to the subsurface EM properties, which are usually not accurately known in advance. On the other hand, spreading

losses caused by wavefront expansion can be corrected, since the signal can be approximated as having an asymptotic $1/r$ amplitude decay with distance. Nevertheless, assuming that the recorded data set has been corrected for intrinsic attenuation and spreading losses, so that the amplitude decay mainly depends on partial reflections, the reflected amplitudes can be compared with the reference amplitude in order to recover the subsurface EM impedance contrasts (Forte et al. 2014; Dossi et al. 2016).

We estimated the reference amplitude of the signals emitted by three different GPR antennas, by analyzing a series of single reflections from an air-metal interface. The reflections were recorded at various distances from a vertical 3 × 3 m metallic surface, and with different antenna orientations. We used a Malå Geoscience ProEx GPR system equipped with three bistatic shielded antenna pairs, with nominal central frequencies equal to 250, 500, and 800 MHz, and with offsets parallel to the metallic surface, respectively, equal to 31, 18, and 14 cm. For each GPR antenna, four different data sets were recorded, each one with the antennas being rotated by 90°, in order to take into account possible changes in the radiation pattern, systematic artifacts, or possible external interference. In each series, the GPR system was moved along the axis normal to the metallic surface, with a maximum distance of about 9.0 m, and with the transmitter being automatically triggered every 0.5 s. The applied signal processing was limited to DC correction, drift removal, and background removal. The peak amplitudes of the recorded reflections are plotted for the three GPR antenna pairs in Fig. 1, showing their decay with the distance traveled by the recorded signal, which is calculated from the arrival times of the reflected signals and the EM velocity in air (i.e. about 30 cm/ns).

Assuming a total reflection from the metallic surface, the graphs in Fig. 1 can be used to study the amplitude decay caused by spreading losses with the signal propagating in air, and also to extrapolate the peak amplitude of the wavelet initially transmitted by the antennas. The recorded amplitudes are fitted in Fig. 1 with the following amplitude decay function, which assumes an asymptotic $1/r$ amplitude decay with distance:
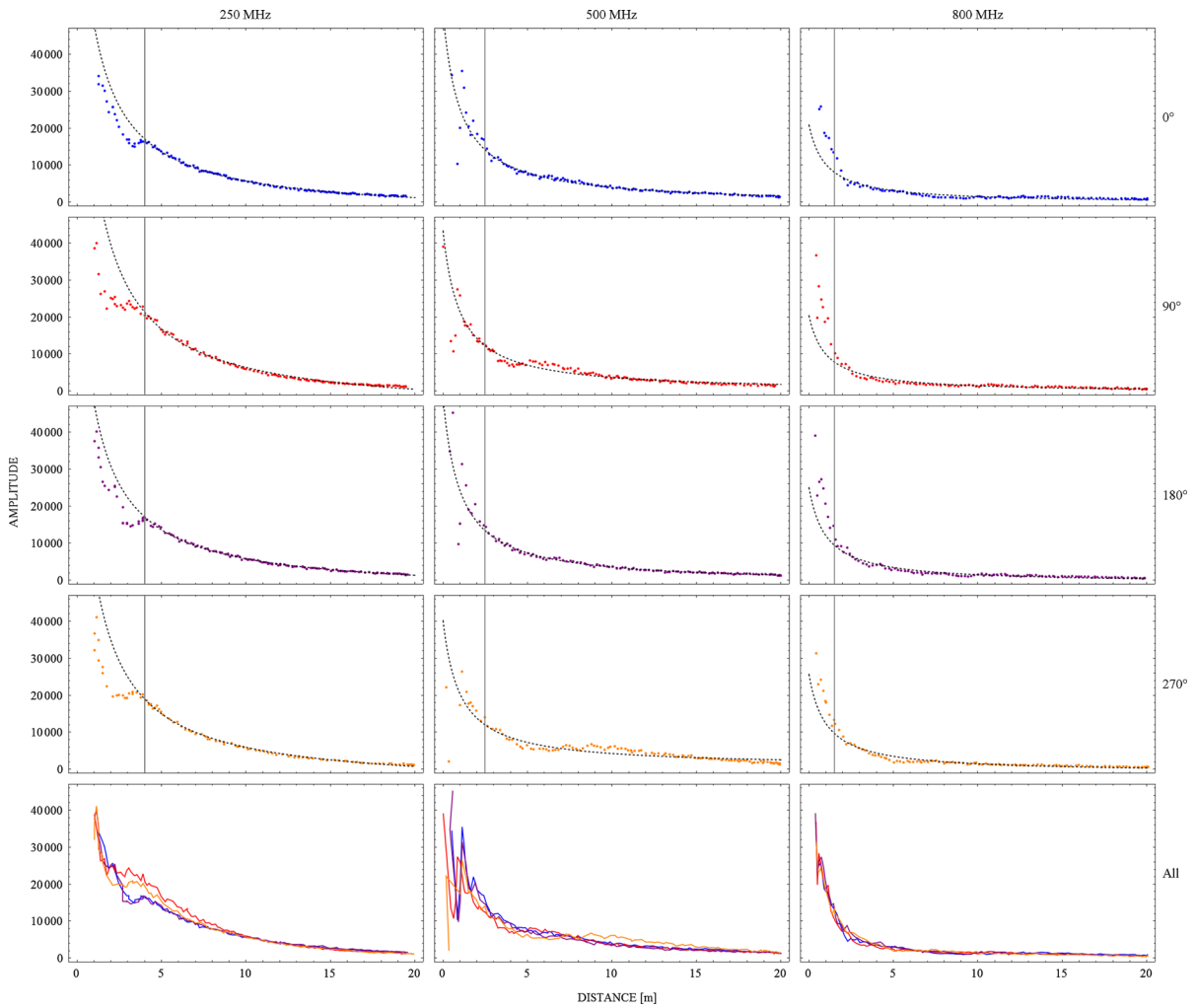
Figure 1
Analysis of the amplitude decay caused by spreading losses of GPR signals reflected from an air–metal interface. The antennas have a nominal central frequency, respectively, equal to 250 (first column), 500 (second column), and 800 (third column) MHz. Four data sets were recorded for each GPR antenna, each one with the antennas being rotated by 90°. The recorded data sets do not show significant deviations with respect to the assumed $1/r$ amplitude decay, nor do they show major changes when rotating the antennas, as it can be also noticed from the fitting parameters given in Table 1. Amplitudes recorded at distances lower than the defined thresholds (vertical gray lines) were not used for data fitting, due to interference with the direct air-wave and possible deviations from the far field approximation

$$A(r) = \frac{A_0}{1 + r} + B. \qquad (2)$$

In the ideal case, the value of the fitting parameter B in Eq. 2 should be either null or negligible so that the theoretical amplitude $A(0)$ of the signal initially transmitted by the antennas is indeed equal to A_0. The fitted curves in Fig. 1 show an overall good correspondence with the recorded amplitudes, with the values of B being consistently lower than 5% of the

corresponding values of A_0 (Table 1). The erratic behavior of the reflected amplitudes at small distances from the metal plate in Fig. 1 is caused by the interference of the reflected signal with the direct air-wave, which can alternatively be constructive or destructive, depending on the arrival time of the reflection in each recorded trace. Deviations from the far field approximation at smaller distances could also contribute to this effect. Such distorted

Table 1

Parameters obtained by fitting the function given in Eq. 2 to the amplitude decays shown in Fig. 1

Central frequency (MHz)	Fitting parameters	Series				
		1	2	3	4	All
250	A_0	103,946	136,518	102,973	119,352	115,374
	B	− 3735	− 6017	− 3540	− 4929	− 4531
	$|B/A_0|$ (%)	3.6	4.4	3.4	4.1	3.9
	$|\Delta A_0/A_{0All}|$ (%)	9.9	18.3	10.7	3.4	0
500	A_0	52,023	43,572	50,387	39,516	46,484
	B	− 756	− 257	− 1005	− 659	− 349
	$|B/A_0|$ (%)	1.5	0.6	2.0	1.7	0.8
	$|\Delta A_0/A_{0All}|$ (%)	11.9	6.3	8.4	15.0	0
800	A_0	21,301	20,860	25,803	26,619	23,702
	B	− 422	− 464	− 828	− 949	− 670
	$|B/A_0|$ (%)	2.0	2.2	3.2	3.7	2.8
	$|\Delta A_0/A_{0All}|$ (%)	10.1	12.0	8.9	12.3	0

The functions show an overall good fit to the data, with the calculated values of B being consistently lower than 5% the values of A_0 in all cases. The parameters also show an average 10% variability (ΔA_0) of the values of A_0 estimated in each series, with respect to the values A_{0All} obtained when combining all the data sets of each GPR system

amplitudes were not used in the analysis, and the minimum distance for fitting was set equal to 4.0, 2.5, and 1.5 m for the 250, 500, and 800 MHz data sets, respectively. These distances were calculated by analyzing the arrival times at which the direct and reflected signals interfere in each GPR data set.

Besides amplitude distortions at small distances, the graphs in Fig. 1 are all consistent with an 1/r amplitude decay, independently from the orientation of the GPR antennas. A not negligible effect on the reflected amplitude can arise at larger distances due to the limited size of the metal plate. Such behavior would be related to the zone of influence of the antenna radiation pattern (Annan 2005); however, it is not noticeable in the analyzed data set. The few cases in which the recorded amplitudes noticeably differ from the fitting function, such as the 270° series in the 500-MHz data set (Fig. 1), are most likely caused by residual coherent noise. Nevertheless, the recorded data sets can show significant variations when analyzing the reference amplitudes estimated for each GPR antenna (i.e. A_0 in Eq. 2). In Table 1, the value of A_0 calculated in each series is compared with the value A_{0All} obtained when combining all the data sets of a particular GPR antenna, showing an average variation of about 10%. The main factors contributing to the uncertainty of the calculated

reference amplitudes are the lack of viable data at small distances, due to the aforementioned interference between the direct and reflected signals, as well as residual noise and possible minor changes in the radiation pattern between each series. The uncertainty of the reference amplitude has to be taken into account when performing any quantitative analysis. Nevertheless, the analyzed data sets do not show significant deviations with respect to the approximated 1/r amplitude decay, nor do they show major changes when the antennas are rotated. It is also interesting to compare the reference amplitudes A_0 obtained for the three analyzed antennas (Table 1), whose significant differences are mainly caused by the specific signatures of each antenna pair. In previous works (Forte et al. 2014; Dossi et al. 2016), the difference between the reflected amplitudes in GPR profiles acquired with various antennas along the same paths was mainly attributed to an increased intrinsic attenuation for higher frequency signals. While such factor could have had an impact, the main factor was most likely a lower reference amplitude for higher frequency antennas, as reported in Table 1.

It is also important to point out that, if the reference amplitude is estimated from an air-coupled data set while the reflected amplitudes are measured from a ground-coupled survey, we have to take into

account the difference in radiation pattern between the two signals, which, respectively, travel through the air and the ground, before being able to compare them in the calculation of the subsurface EM impedance contrasts. The change in directivity between an air-coupled and a ground-coupled GPR system is caused by the refractive focusing associated with the air–ground interface (Jol 2009). The radiation pattern of a ground-coupled GPR system is divided into a main central lobe and several minor lateral lobes. Most of the transmitted energy is contained within a cone defined by the critical angle θ_C of the radiation pattern (Annan and Cosway 1992; Jiao et al. 2000), and the propagating wavefront can be approximated with an expanding spherical cap with surface area S:

$$\theta_C = \arcsin\left[\frac{1}{\sqrt{\varepsilon_r}}\right] \qquad (3)$$

$$S = 2\pi r^2 (1 - \cos[\theta_C]). \qquad (4)$$

The critical angle depends on the relative permittivity of the subsurface for ground-coupled GPR systems, while it is equal to 90° in the case of air-coupled ones (Eq. 3), with the value of ε_r being equal to 1 in air. Therefore, in the air-coupled GPR surveys analyzed in Fig. 1, the wavefront can be approximated with an expanding hemisphere having a surface area equal to $2\pi r^2$. If we assume that the same amount of energy within this hemisphere is concentrated within the spherical cap obtained in the case of ground-based surveys, the estimated values of A_0 have to be corrected by a geometrical factor $1/\sqrt{1 - \cos[\theta_C]}$, before they can be compared with the reflected amplitudes in the calculation of the subsurface EM impedance contrasts. Although approximate, this correction is essential in order to avoid overcompensating for the spreading losses of the reflected amplitudes during amplitude recovery, which would result in an overestimation of the subsurface EM impedance contrasts (Dossi et al. 2016).

3. Sampling-Related Amplitude Distortion Analysis

As previously discussed, the accurate estimation of the subsurface EM impedance contrasts mostly depends on the correct measurement of the reference and reflected amplitudes. This means that the digital data set obtained during data acquisition must accurately sample the peak amplitudes of the original analog signal. Consider a signal defined by an analytic function $A(t)$, with a local maximum located at the time instant t_{max}, in which we want to correctly sample the peak amplitude $A(t_{max})$, or at least to keep the maximum peak amplitude error as low as possible. Given a sampling interval Δt, the largest possible peak amplitude error is obtained when t_{max} is located between two adjacent samples, respectively, at time instants t_k and $t_k + \Delta t$, such that $A(t_k)$ is equal to $A(t_k + \Delta t)$ (Kuffel et al. 2000; Schon 2013). In any other case, one of the two sampled amplitudes would be lower than such value, while the other would be higher, and, therefore, the latter would be considered the sampled peak amplitude. The maximum time shift of the sampled peak amplitude with respect to the actual peak amplitude is equal to $\Delta t/2$, in which case t_{max} is located exactly in the midpoint between t_k and $t_k + \Delta t$, as in the case of a symmetric peak, while in the generic case it is closer to one of the two adjacent samples.

An example of sampling-related signal distortion analysis is given in Fig. 2, which shows a generic analog wavelet (or equivalently a greatly oversampled signal, Fig. 2a), that can be used to simulate both GPR and seismic sources. As previously discussed, we focus on the signal distortions related to the data loss between samples, while disregarding possible round-off errors of the sampled amplitudes. The analog signal is digitized in Fig. 2b, showing a significant amplitude data loss, despite using a sampling rate equal to 2.0 GHz, which is well above the previously defined Nyquist–Shannon threshold. The frequency spectrum of the sampled series (gray line in Fig. 2c) is virtually coincident with the one obtained when sampling the signal with a sampling rate equal, for instance, to 10.0 GHz (black line), in accordance with the Nyquist–Shannon theorem. It is important to point out that the frequency spectrum of the digitized signal is calculated exclusively from the sampled values (black dots in Fig. 2b), which contain the actual amplitude information. Moreover, in order to increase the number of frequency components to be compared in Fig. 2c, resulting from the Fourier analyses of the 2.0 and 10.0 GHz digitized series,
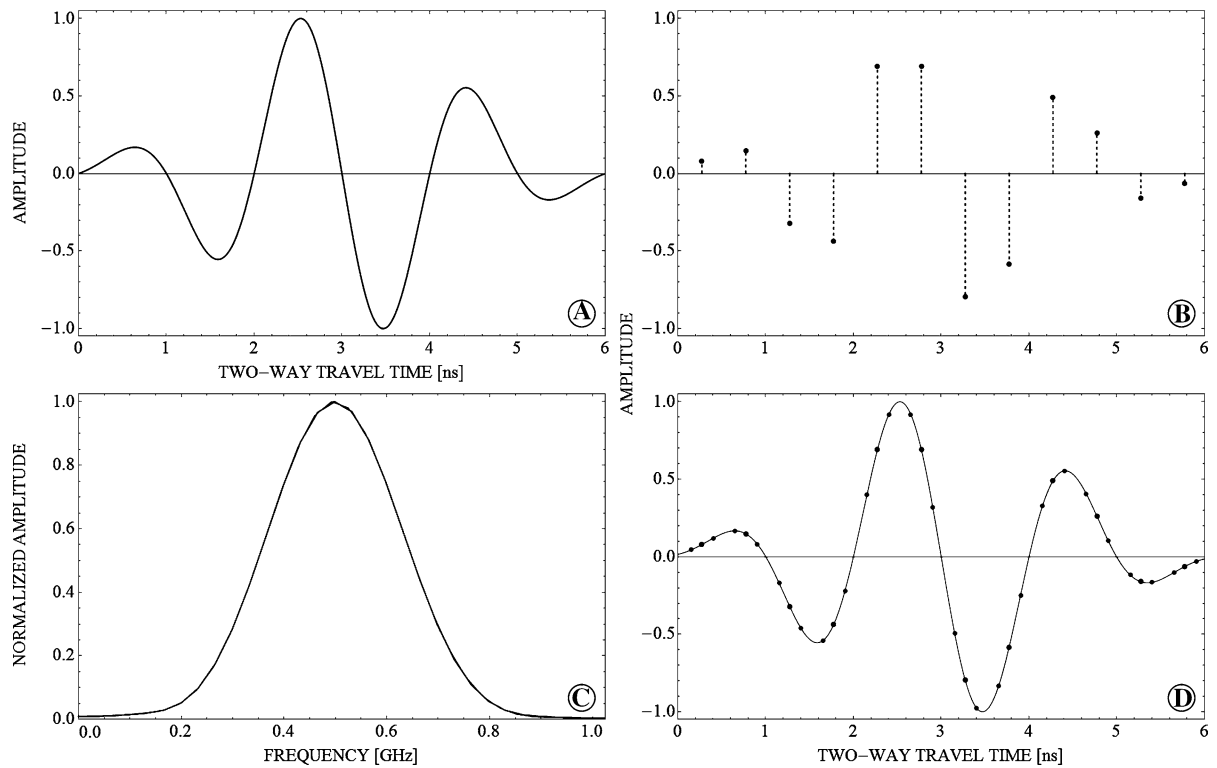
Figure 2
Example of peak amplitude error analysis applied to a generic wavelet. The analytic signal (**a**) is sampled (black dots) in **b** using a sampling rate equal to 2.0 GHz, which is well above the Nyquist–Shannon threshold (i.e. twice the maximum signal frequency, about 0.9 GHz in **c**). The frequency spectrum of the digitized signal (gray line) is compared in **c** to the one obtained with a sampling rate equal to 10.0 GHz (black line), chosen as an example, and they are virtually coincident, in accordance with the Nyquist–Shannon theorem. The analog signal is reconstructed in **d** through a sinc interpolation (gray line), and it is re-sampled using a sampling rate equal to 4.0 GHz (gray dots)

while also satisfying the hypotheses of the afore-mentioned Nyquist–Shannon theorem, the two signals were both sampled in the larger [0, 30] ns time interval, with the amplitude being set identically null outside the analyzed [0, 6] ns time interval (Fig. 2a). Nevertheless, despite the absence of alias-ing effects, the digitized series in Fig. 2b shows significant peak amplitude errors in the various signal phases. Specifically, the analog signal is sampled in Fig. 2b in the worst-case scenario with respect to the third peak in Fig. 2a, with the two samples $A(t_k)$ and $A(t_k + \Delta t)$ around the actual peak having the same value, and the resulting peak amplitude error is equal to 30.9%. The peak amplitude error in Fig. 2b varies significantly between the signal phases, most of which are not sampled in the worst-case scenario, and it depends on the arrival time and shape of the peak, as well as the selected sampling interval. For

example, the peak amplitude error in the fourth peak, which is symmetric to the third one in Fig. 2a, is equal to 20.3%.

The peak amplitude error in a recorded data set can theoretically be reduced by re-sampling the data at higher rates, using polynomial or predictive inter-polating functions in order to reconstruct the original analog signal (Eldar 2015). However, the accuracy of the reconstructed signal would depend on the applied interpolation method, which may require systematic assumptions on the signal behavior between samples. In fact, the actual data are limited to the sampled values, while any information between samples is unrecoverably lost. Strange (2013) analyzes interpo-lation and re-sampling as a possible method to increase temporal resolution, thus focusing on reducing the temporal uncertainty $\Delta t/2$ of a signal peak, rather than the peak amplitude error as we are

here focusing on. He concludes that interpolating GPR traces to increase temporal resolution is less accurate than directly sampling the original analog signal at higher sampling rates, while the time interpolation of data acquired with low sampling rates is always preferable with respect to using the raw sampled data.

From a practical point of view, the interpolation process can be computationally intensive and time consuming, especially for large data sets. On the other hand, limiting the interpolation to a moving window may affect the reconstructed signal and introduce a certain degree of subjectivity or artifacts. In other words, computationally effective interpolation techniques tend to be less accurate, while accurate interpolation techniques tend to be more computationally intensive (Strange 2013). Moreover, while the analog signal may be accurately reconstructed in synthetic data sets, in which the transmitted wavelet and the subsurface model are known, the accuracy of the reconstructed signals is neither necessarily acceptable nor verifiable for real data sets. In geophysical surveys, the signal behavior between samples is not predictable due to several factors, such as spectral variation of the transmitted wavelet with time, statistically random distributions of the subsurface reflectivity, interference phenomena, and time-varying random noise (Yilmaz 2001). In other words, digitized data sets contain amplitude information only at the sampling points and data interpolation (and re-sampling) does not necessarily increase the information content, or decrease the sampling error. More importantly, even if the analog signal is accurately reconstructed through interpolation, the subsequent re-sampling process simply reintroduces the initial sampling-related amplitude distortion problem, since we have to select a new sampling rate with which to re-sample the reconstructed signal. As an example, the analog signal $A(t)$ is reconstructed in Fig. 2d by applying a sinc interpolation to the amplitudes $A(t_k)$ sampled in Fig. 2b, given by the following equation:

$$A(t) = \sum_{k=-\infty}^{+\infty} A(t_k) \frac{\sin[\pi(t - t_k)/\Delta t]}{\pi(t - t_k)/\Delta t}. \qquad (5)$$

This is the reconstruction formula for a band-limited signal in the absence of aliasing effects (Proakis and Manolakis 2006), and it consists in a sum of various time-shifted sinc functions, each centered in one of the sampled time instants t_k and weighted by the corresponding sampled amplitude $A(t_k)$. It can be noticed that the reconstructed signal correctly gives the sampled amplitudes $A(t_k)$ at each time instant t_k, since each time-shifted sinc function in Eq. 5 is equal to 1 at their central time instant t_k and null in all the others. Similarly to the Nyquist–Shannon theorem, the sinc interpolation assumes a band-limited signal, and Eq. 5 requires an infinite number of samples. However, in practice, the recorded signal contains a finite number of samples and the interpolation formula has to be truncated.

In our sampling analysis, Eq. 5 is simplified by the fact that the samples $A(t_k)$ outside the analyzed [0, 6] ns time interval (Fig. 2b) are set identically null. The reconstructed signal is re-sampled in Fig. 2d by increasing the sampling rate to 4.0 GHz, which satisfies the Nyquist–Shannon theorem as well as the threshold of six times the central frequency (i.e. six times 0.5 GHz, Fig. 2c) recommended by Jol (2009) for GPR surveys. The peak amplitude error is visibly reduced in Fig. 2d with respect to Fig. 2b; however, the previously described sampling-related distortion problem remains. Specifically, while the fourth signal peak can be considered accurately sampled, with the error being reduced to 2.4%, the maximum peak amplitude error in the re-sampled data set is still significant, with the error in the third peak being equal to 8.1%. Therefore, even if we re-sample the data set in order to accurately measure the peak amplitude of a specific signal phase, other recorded peaks may still show considerable peak amplitude errors, depending on their shapes and arrival times, as well as the newly selected sampling rate. It is also important to point out that the signal reconstruction in Fig. 2d, through the interpolation of the discrete signal, is favored by the fact that the original analog signal (Fig. 2a) has a particularly regular shape. More irregular analog signals, possibly affected by noise or interference, would be more difficult to accurately reconstruct from the digitized series, especially with relatively low sampling rates. Moreover, the reason we are able to calculate the

peak amplitude errors in Fig. 2b, d is because we know the shape of the original analog signal (Fig. 2a); however, in real data sets it is impossible to determine the recorded sampling errors or the accuracy of the reconstructed analog signal. These considerations are not limited to a specific kind of sampled signal, rather they can be applied to any digitized series. In fact, when only discrete sampled data are available, as it is the case of any digitally recorded data set, it is impossible to evaluate which interpolation performs better in terms of signal reconstruction, simply because no reference analog signal is available. This is definitely a crucial point, which is often disregarded and misunderstood.

Nevertheless, it is possible to limit the peak amplitude error during data acquisition by setting a minimum threshold for the sampling rate, which can be linked to the central frequency of the recorded signal. In the following sections, we provide an estimation of this threshold by analyzing the dependence of the maximum amplitude error on the sampling rate for various signals characterized by different spectral contents.

3.1. Wavelet Analysis

In this section, we analyze three different signals, namely a sine function, a Ricker function, and the aforementioned generic wavelet, respectively, given by the following equations:
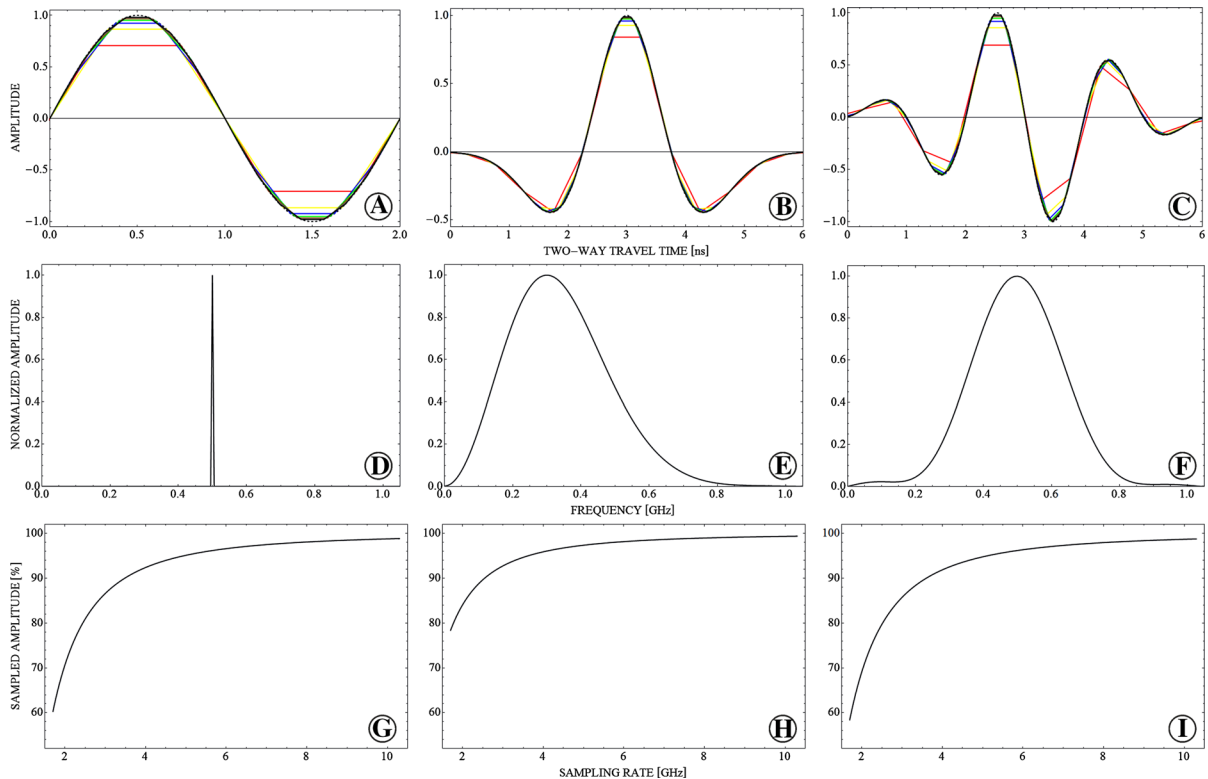

Figure 3

Analysis of the peak amplitude error for a sine function (**a**), a Ricker function (**b**), and a generic wavelet (**c**). The figure shows the analog signals (dashed lines) with superimposed seven different digital series (**a–c**) sampled using sampling rates, respectively, equal to 2.0 (red), 3.0 (yellow), 4.0 (blue), 5.0 (green), 6.0 (brown), 7.0 (gray), and 8.0 (black) GHz; the normalized amplitude spectra (**d–f**); and the resulting peak amplitudes for sampling rates in the [2, 10] GHz frequency interval (**g–i**). In all cases, the amplitudes are sampled in the worst-case scenario with respect to the main peak in each graph (**a–c**), and the sampled amplitudes in **g–i** are shown as a percentage of the actual peak amplitudes. The sampling rates considered are all well above the Nyquist–Shannon thresholds of the respective signals (i.e. twice the maximum signal frequencies in **d–f**), thus avoiding aliasing effects

$$H(t) = \sin[2\pi f_H t] \qquad (6)$$

$$R(t) = \left(1 - \frac{1}{2}(2\pi f_R(t - t_R))^2\right)e^{-\frac{1}{4}(2\pi f_R(t-t_R))^2} \quad (7)$$

$$G(t) = -A\sin[2\pi f_G(t - t_G)]e^{-0.2\pi f_G(t-t_G)^2}. \qquad (8)$$

where A is a normalizing amplitude, the central frequencies f_H, f_R, and f_G are, respectively, equal to 0.5, 0.3, and 0.5 GHz, and the temporal parameters t_R and t_G are both equal to 3 ns.

The analyzed analog signals are shown in Fig. 3a–c, with superimposed seven different digital series, sampled in the worst-case scenario with respect to the main peak in each signal, using sampling rates, respectively, equal to 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, and 8.0 GHz. In Eqs. 6 and 7, the main peak is symmetric with respect to $t_{\max}$; therefore, the worst sampled amplitudes W_H and W_R can be easily calculated from the selected sampling rate f_S using the following equations:

$$W_H(f_S) = \cos\left(\pi\frac{f_H}{f_S}\right) \qquad (9)$$

$$W_R(f_S) = \left(1 - \frac{1}{8}\left(2\pi\frac{f_R}{f_S}\right)^2\right)e^{-\frac{1}{16}\left(2\pi\frac{f_R}{f_S}\right)^2}. \qquad (10)$$

The sampling rates used for the discrete series in Fig. 3a–c are all well above the Nyquist–Shannon threshold of the respective signals (i.e. twice the maximum signal frequencies in Fig. 3d–f), thus avoiding aliasing effects. However, as previously discussed, while the Nyquist–Shannon theorem can be used to accurately reconstruct the frequency spectrum of a band-limited signal, it does not guarantee an accurate sampling of its peak amplitudes. This is evident when analyzing Eq. 9, where the frequency spectrum of the original sinusoidal signal $H(t)$ has a single component with frequency f_H (Eq. 6). If we use a sampling rate exactly equal to $2 \cdot f_H$, in the worst-case scenario we can have a sampled peak equal to 0, since all the sampled amplitudes would be null. If instead the sampling rate is equal to $3 \cdot f_H$, in the worst case the sampled peak is equal to 0.5, that is half the actual peak in $H(t)$. In order to have a sampled peak which at worst is equal to 0.9 (i.e. 90% of the actual peak), we have to use a sampling rate around $7 \cdot f_H$. A similar behavior can be noticed in all three analyzed signals, with Table 2 listing the peak amplitudes sampled by the discrete series shown in Fig. 3a–c. The sampled amplitudes are given as a percentage of the actual peak amplitudes of the original analog signals, and we can notice that the higher the sampling rate is, the lower the maximum peak amplitude error becomes. This behavior can also be observed in Fig. 3g–i, which shows the peak amplitudes sampled in the worst-case scenario for sampling rates in the [2, 10] GHz frequency interval, which is well above the Nyquist–Shannon thresholds of the respective signals (Fig. 3d–f). More specifically, we can notice

Table 2

Analysis of the peak amplitudes sampled in Fig. 3a–c for the sine function (Eq. 6), the Ricker function (Eq. 7), and the generic wavelet (Eq. 8), respectively

Sampling rate	Sampling interval	Sine function		Ricker function		Generic wavelet	
f_S (GHz)	Δt (ns)	f_S/f_C	A (%)	f_S/f_C	A (%)	f_S/f_C	A (%)
2.0	0.50	4.0	70.7	6.7	84.1	4.0	69.1
3.0	0.33	6.0	86.6	10.0	92.7	6.0	85.7
4.0	0.25	8.0	92.4	13.3	95.9	8.0	91.9
5.0	0.20	10.0	95.1	16.7	97.4	10.0	94.8
6.0	0.17	12.0	96.9	20.0	98.2	12.0	96.3
7.0	0.14	14.0	97.5	23.3	98.6	14.0	97.3
8.0	0.13	16.0	98.1	26.7	99.0	16.0	97.9

The table lists the selected sampling rates f_S, the corresponding sampling intervals Δt and sampling-to-central frequency ratios f_S/f_C, as well as the peak amplitudes A of the main signal peak in each graph, sampled in the worst-case scenario and shown as a percentage of the actual peak amplitudes of the original analog signals
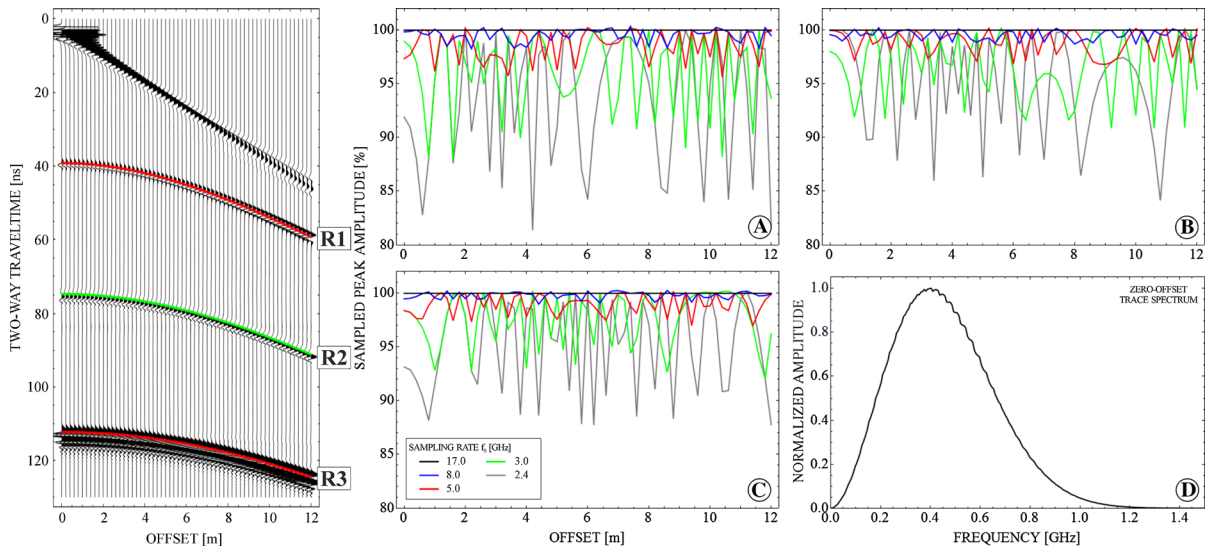
Figure 4
Analysis of the peak amplitudes sampled along three reflections in a synthetic CMP GPR gather. Positive amplitudes in the GPR profile are marked in green, negative amplitudes in red. The graphs show the variations, at different sampling rates, of the peak amplitudes along the reflections $R1$ (**a**), $R2$ (**b**), and $R3$ (**c**), given as a percentage of the amplitudes picked with the highest sampling rate in each trace. The normalized frequency spectrum of the zero-offset trace, sampled with the highest sampling rate, is also shown (**d**). The sampling rates used for the analysis are equal to 2.4 (gray), 3.0 (green), 5.0 (red), 8.0 (blue), and 17.0 (black, used as reference) GHz, all well above the Nyquist–Shannon threshold of the simulated signal (i.e. twice the maximum signal frequency in **d**), thus avoiding aliasing effects

that the higher the sampling-to-central frequency ratio f_S/f_C is, the lower the maximum peak amplitude error becomes. As an example, the central frequencies of the sine function and the generic wavelet are both equal to 0.5 GHz (Fig. 3d, f), and the corresponding sampling errors are very similar for equal sampling rates (Fig. 3g, i, Table 2). On the other hand, the Ricker function has a lower central frequency (i.e. 0.3 GHz, Fig. 3e), and the higher sampling-to-central frequency ratio results in lower sampling errors for the same sampling rates (Fig. 3h, Table 2). The sampling analysis shows that the maximum peak amplitude error is lower than 5% for sampling rates above 4.95 GHz for the sine function, which is about 10 times the signal central frequency (i.e. 0.5 GHz, Fig. 3d); above 3.62 GHz for the Ricker function, which is around 12 times the signal central frequency (i.e. 0.3 GHz, Fig. 3e); and above 5.12 GHz for the generic wavelet, which is more than 10 times the signal central frequency (i.e. 0.5 GHz, Fig. 3f).

3.2. Synthetic CMP Gather Analysis

Quantitative geophysical interpretation methods have been integrating and gradually substituting qualitative analyses, particularly in reflection seismics and GPR applications. Such methods require measurements of the analyzed physical phenomena with a sufficient degree of accuracy to obtain valid quantitative information. In this section, we analyze the behavior of the peak amplitude error in the synthetic CMP GPR gather shown in Fig. 4, which was constructed using the GPRMax 2-D software (Giannopoulos 2005). The input 1-D model consists of three layers, respectively, 5, 3, and 4 m thick, above a semi-infinite base layer. Each layer is homogeneous, lossless, non-magnetic, and non-dispersive, with EM velocities, respectively, equal to 27.9, 17.0, and 21.3 cm/ns in the three layers, and 12.2 cm/ns in the base layer. The simulated survey consists of a ground-coupled 400 MHz bistatic GPR system, transmitting Ricker wavelets with an offset which varies from 0 (i.e. coincident transmitter and receiver) to 12 m, increasing by 0.2 m at each step.

Table 3

Analysis of the maximum peak amplitude error for the sampled amplitudes shown in Fig. 4

Sampling rate	Sampling interval	Frequency ratio	Reflection		
			R1	R2	R3
f_S (GHz)	Δt (ns)	f_S/f_C	A_{min} (%)	A_{min} (%)	A_{min} (%)
2.4	0.42	6.0	81.4	84.1	87.7
3.0	0.33	7.5	87.9	90.8	92.1
5.0	0.20	12.5	95.6	96.8	97.0
8.0	0.13	20.0	98.3	98.7	99.0
17.0	0.06	42.5	100.0	100.0	100.0

The table lists the selected sampling rates f_S, the corresponding sampling intervals Δt and sampling-to-central frequency ratios f_S/f_C, as well as the minimum sampled peak amplitudes A_{min}, given as a percentage of the corresponding peak amplitudes sampled with the highest sampling rate, which are used as reference

We analyzed the peak amplitudes of the main phases of the three recorded reflections (marked as R1, R2, and R3 in Fig. 4), using sampling rates equal to 2.4, 3.0, 5.0, 8.0, and 17.0 GHz. These sampling rates are all well above the Nyquist–Shannon threshold of the simulated signals (i.e. twice the maximum signal frequency in Fig. 4d), and they all satisfy the Nyquist–Shannon theorem as well as the threshold of six times the central frequency (i.e. six times 0.4 GHz) recommended by Jol (2009) for GPR surveys. The sampled peak amplitudes are shown in Fig. 4a–c as a percentage in each trace of the amplitudes obtained with the 17.0 GHz sampling rate, which are used as reference. We can notice a general decrease in the variability of the sampled peak amplitudes, and, therefore, of the maximum peak amplitude error, with increasing sampling rates. This behavior is also highlighted in Table 3, which shows the minimum sampled amplitudes in Fig. 4a–c for each sampling rate, given as a percentage of the amplitudes picked with the highest sampling rate. The results in Table 3 show maximum peak amplitude errors as high as 18.6% (2.4 GHz, gray line), 12.1% (3.0 GHz, green line), 4.4% (5.0 GHz, red line), and 1.7% (8.0 GHz, blue line). In Fig. 4a–c, we can notice a few instances in which the sampled peak amplitude is actually higher at lower sampling rates, showing that even the highest sampling rate does not necessarily guarantee the recovery of the actual peak amplitude, although the resulting errors are negligible. The sampling analysis in Fig. 4 shows that the peak amplitude error is mostly random in nature, and it depends on the selected sampling interval, as well

as the shape and arrival time of the analyzed signal peak. For example, it is interesting to notice the smaller peak amplitude errors in Fig. 4b, c with respect to Fig. 4a, which are most likely caused by a slight elongation of the simulated wavelet with travel time. Specifically, a larger time interval corresponds to a lower central frequency for the analyzed signal phase, and consequently the higher sampling-to-central frequency ratio results in a lower maximum peak amplitude error, as previously discussed.

As it is shown in the following section, a planar reflection, combined with a regular sampling grid and a laterally constant wavelet, would result in a quasi-periodic behavior of the peak amplitude error. In the extreme case of a perfectly flat reflection, the peak amplitude error would be laterally constant. However, in the general case a reflection rarely has a perfectly planar geometry or a laterally constant wavelet, if ever, due to factors like the subsurface stratigraphy, signal attenuation, EM velocity distribution, and impedance contrasts, as well as noise and interference. The resulting quasi-random behavior of the peak amplitude error prevents any possible prediction or estimation of sampling-related signal distortions. However, the results in Table 3 show that the maximum peak amplitude error is lower than 5% in all the reflections for sampling rates equal to at least 5.0 GHz, which is more than 12 times the signal central frequency (i.e. 0.4 GHz, Fig. 4d). Therefore, we can limit the peak amplitude error during data acquisition by increasing the sampling rate above this easily identifiable threshold, which is linked to the central frequency of the transmitted signal.
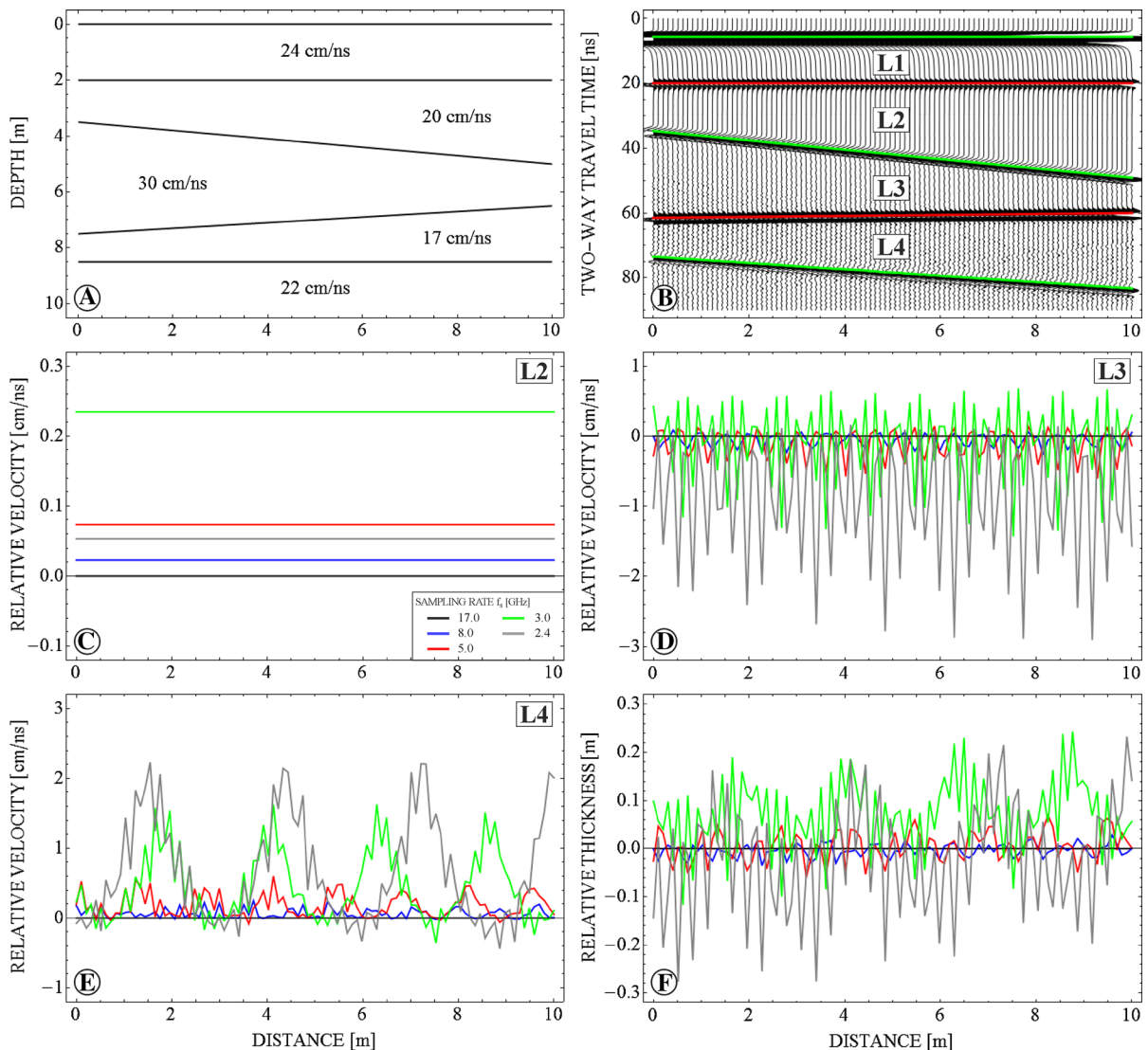
Figure 5

Amplitude inversion of a synthetic CO GPR profile. The figure shows the synthetic model with the simulated EM velocity distribution (**a**); and the resulting GPR profile (**b**) with superimposed the peak amplitudes selected for the inversion. Positive amplitudes in **b** are marked in green, negative amplitudes in red. The figure also shows the EM velocities estimated at different sampling rates for the layers $L2$ (**c**), $L3$ (**d**), and $L4$ (**e**), as well as the reconstructed total thickness of the original model (**f**). These quantities are shown as variations from the values obtained with the highest sampling rate. The sampling rates used for the analysis are equal to 2.4 (gray), 3.0 (green), 5.0 (red), 8.0 (blue), and 17.0 (black, used as reference) GHz, all well above the Nyquist–Shannon threshold of the simulated signal (i.e. twice the maximum signal frequency in Fig. 4d), thus avoiding aliasing effects

3.3. Amplitude Inversion Analysis

In this section, we study the effects of the sampling rate on the quantitative analysis of a GPR data set. Specifically, we apply an amplitude inversion algorithm (Forte et al. 2014; Dossi et al. 2016) to a synthetic CO GPR data set, which was constructed using the GPRMax 2-D software (Giannopoulos 2005), and analyze the changes in the reconstructed thickness and EM velocity distribution with different sampling rates. The input 2-D model contains four layers (marked as $L1$–$L4$ in Fig. 5), with EM velocities, respectively, equal to 24.0, 20.0, 30.0,

and 17.0 cm/ns, above a semi-infinite base layer with an EM velocity of 22.0 cm/ns (Fig. 5a). Similarly to the previous model, the simulated survey consists of a ground-coupled 400 MHz GPR system, transmitting Ricker wavelets with a constant 0.7 m offset and a 0.1 m trace interval. The resulting synthetic data set is shown in Fig. 5b, with superimposed the picked amplitudes used in the inversion. The high-frequency noise visible in the GPR profile below L2 is caused by the geometrical discretization of the dipping reflectors (cell size equal to 2.5×2.5 cm), which can affect the reflected amplitudes, resulting in slight random lateral changes in the reconstructed EM impedance contrasts.

Similarly to the previous CMP analysis, we use sampling rates equal to 2.4, 3.0, 5.0, 8.0, and 17.0 GHz, all well above the Nyquist–Shannon threshold of the simulated signals (i.e. twice the maximum signal frequency in Fig. 4d), thus avoiding aliasing effects. In this analysis, we study how the reconstructed EM velocities change with the sampling rate in the layers L2 (Fig. 5c), L3 (Fig. 5d), and L4 (Fig. 5e). The EM velocity in layer L1 is required as an input of the applied inversion algorithm (Forte et al. 2014; Dossi et al. 2016), and, therefore, it is not included in the analysis. Moreover, we also study the variations in the estimated total thickness of the model (Fig. 5f), which combines the changing EM velocities in each layer with the recorded arrival times, which can also slightly vary with different sampling rates. The estimated quantities in Fig. 5c–f are shown as variations from the results obtained in

each trace with the 17.0 GHz sampling rate, which are used as reference. Similarly to the amplitude analysis in the previous CMP model, we can notice a general decrease in the variability of the estimated quantities, with increasing sampling rates. This behavior is also highlighted in Table 4, which shows the maximum changes in the estimated quantities in Fig. 5c–f for each sampling rate, given as variations from the results obtained with the highest sampling rate. The analysis shows maximum EM velocity errors as high as 0.26 (8.0 GHz, blue line), 0.60 (5.0 GHz, red line), 1.63 (3.0 GHz, green line), and 2.91 (2.4 GHz, gray line) cm/ns. As previously discussed, the quasi-periodic lateral trends of the EM velocity errors (Fig. 5c–e) are caused by the combination of the simulated planar reflections (Fig. 5b), characterized by laterally constant wavelets, with the regular sampling grid. In the case of the shallowest reflection, which is perfectly flat (Fig. 5b), the laterally constant peak amplitude errors produce the laterally constant EM velocity errors in the second layer (Fig. 5c). In the previous CMP analysis (Fig. 4), the hyperbolic reflections instead caused more irregular peak amplitude errors.

The sampling-related EM velocity errors, combined with the arrival times of each reflection, result in the total thickness errors shown in Table 4. The largest error is equal to 0.28 m, which corresponds to 3.3% of the total thickness of the model (i.e. 8.5 m, Fig. 5a), and it is obtained with the 2.4 GHz sampling rate. However, we have to take into account the fact that this result combines both positive and negatives

Table 4

Analysis of the maximum changes in the estimated quantities shown in Fig. 5c–f

Sampling rate	Sampling interval	Frequency ratio	Layer			Total thickness
			L2	L3	L4	
f_S (GHz)	Δt (ns)	f_S/f_C	$|\Delta v_{max}|$ (cm/ns)	$|\Delta v_{max}|$ (cm/ns)	$|\Delta v_{max}|$ (cm/ns)	$|\Delta Z_{max}|$ (m)
2.4	0.42	6.0	0.05	2.91	2.23	0.28
3.0	0.33	7.5	0.23	1.43	1.63	0.24
5.0	0.20	12.5	0.07	0.60	0.60	0.07
8.0	0.13	20.0	0.02	0.26	0.25	0.04
17.0	0.06	42.5	0	0	0	0

The table lists the selected sampling rates f_S, the corresponding sampling intervals Δt and sampling-to-central frequency ratios f_S/f_C, as well as the maximum changes in modulus in the reconstructed EM velocities Δv_{max} and the total thickness ΔZ_{max}, with respect to the values estimated with the highest sampling rate, which are used as reference

EM velocity errors (Fig. 5c–e). Therefore, the resulting overestimations and underestimations of the various layer thicknesses tend to compensate each other in the calculation of the total thickness of the model. As an example, if we combine the maximum velocity errors in each layer in Table 4 with the arrival times of the central trace of the synthetic GPR profile (Fig. 5b), assuming that they are all negative (i.e. we underestimate the EM velocity in all layers), the total model thickness will be underestimated by 0.5 m (i.e. a 5.9% error).

It is important to point out that other possible uncertainty factors can influence the quantitative analysis of a GPR data set, which would then be added to the effects of sampling-related peak amplitude errors. For example, an erroneous input EM velocity of the shallowest layer can greatly affect the inversion results in the deeper layers due to error propagation, while distortions in the input amplitudes can also be caused by noise and interference, as well as possibly inaccurate amplitude recovery functions (Forte et al. 2014; Dossi et al. 2016). However, sampling-related peak amplitude errors in aliasing-free data sets can still be significant, and they can be greatly reduced during data acquisition by simply increasing the sampling rate above a clearly defined threshold. In this particular case, the results in Table 4 show that the maximum EM velocity error is lower than 0.5 cm/ns, which is the typical accuracy limit even for multi fold GPR acquisitions (Tillard and Dubois 1995), in all layers for sampling rates higher than 5.0 GHz, which is more than 12 times the signal central frequency (i.e. 0.4 GHz, Fig. 4d).

4. Discussion

The previous analyses show that significant peak amplitude errors may result from the sampling process even in absence of aliasing effects, and such amplitude distortions can affect any quantitative analysis of the recorded data sets. For example, in Fig. 4 the peak amplitude errors reach values as high as 18.6% of the actual peak amplitudes, while in Fig. 5 the EM velocity changes resulting from peak amplitude errors can potentially lead to a 5.9% error

in the reconstructed thickness of the analyzed synthetic model.

Peak amplitude errors mainly depend on the sampling rate selected during data acquisition, as well as the shape and arrival time of the analog wavelet to be digitized. Given such factors, these sampling-related distortions can be irregularly distributed within the recorded data set, potentially showing a quasi-random behavior even along the same reflection (Fig. 4), thus preventing any possible prediction or estimation. This is particularly true for geophysical wave field data, where the spectral content of the signal is influenced by several factors. For example, different frequency components of the signal spectrum can be affected by selective filtering during propagation, as well as by dynamic and kinematic dispersion. As a general trend, for a fixed sampling rate, such effects result in overall larger peak amplitude errors in the shallowest reflections, which are usually characterized by the highest frequency components. This behavior was highlighted in Fig. 4, where the smaller variability of the peak amplitude error in the deeper reflection was attributed to a slight elongation of the simulated wavelet, resulting in a higher signal-to-central frequency ratio and consequently in a lower maximum peak amplitude error.

As previously discussed, interpolation and re-sampling of the digitized signals offer only limited solutions to the amplitude distortion problem. In fact, the original analog signal to be digitized is unknown, since it depends on the subsurface EM properties at the particular time and location of the GPR survey, and the only objective data available are limited to the sampled values, while any information between samples is unrecoverably lost. Therefore, the accuracy of the analog signal reconstructed through interpolation cannot be verified in real data sets, while its re-sampling simply reintroduces the sampling-related amplitude distortion problem.

It is important to point out that several factors, other than sampling, can distort the recorded peak amplitudes, including both coherent and random noise, signal interference, as well as inaccurate amplitude recovery functions, resulting in erroneous estimations of the subsurface EM impedance contrasts. However, the uncertainty introduced by sampling-related peak amplitude errors in aliasing-
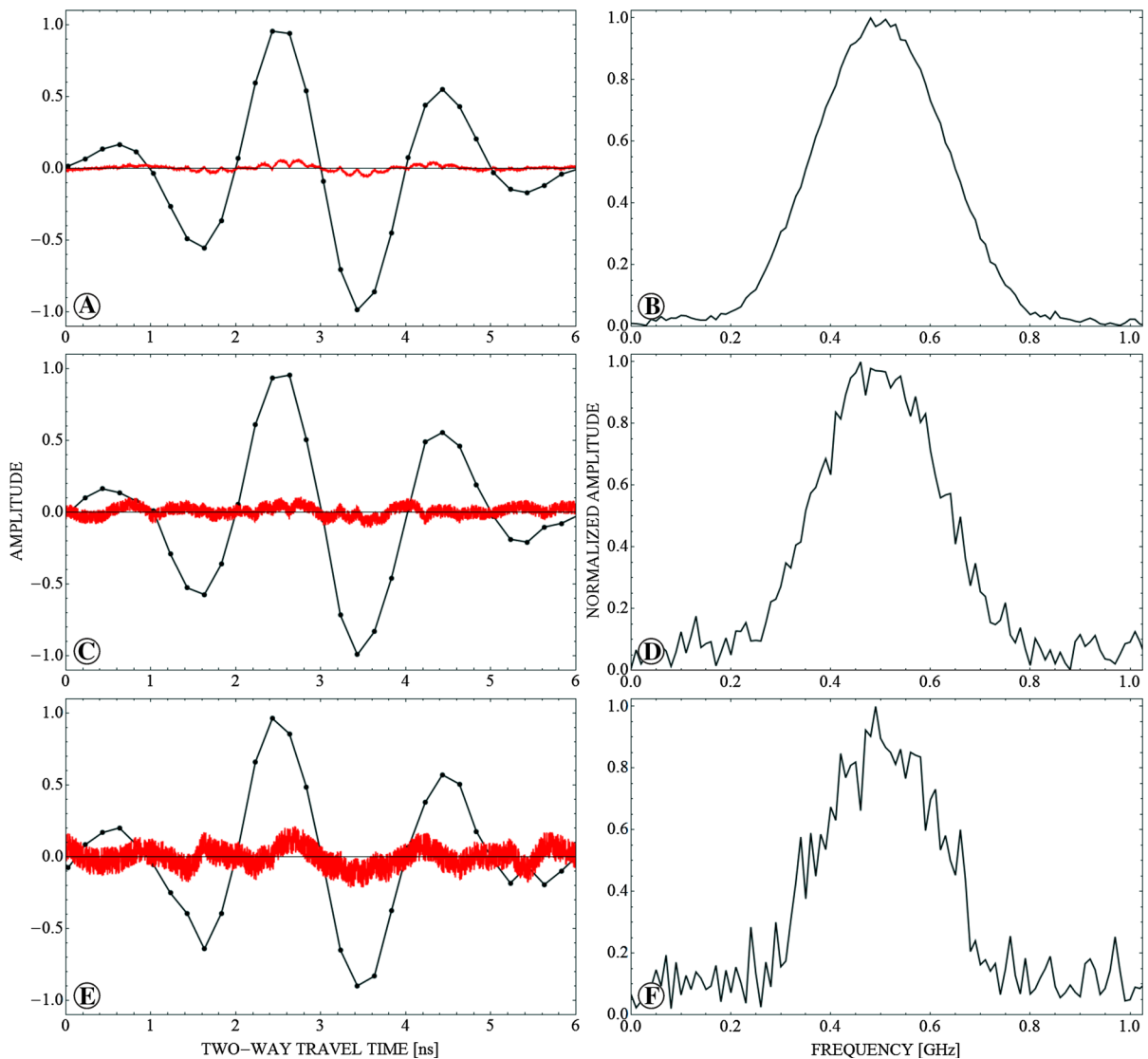
Figure 6

Analysis of the effects of white noise on the synthetic wavelet shown in Fig. 2a, with a sampling rate equal to 5.0 GHz in all three cases, which is ten times the signal central frequency (i.e. 0.5 GHz, Fig. 2c). The figure shows the linearly interpolated digitized series (black line), and the resulting difference (red line) with respect to the original analog signal, affected by white noise equal to 1% (**a**), 5% (**c**), and 10% (**e**) of the signal peak amplitude. The figure also shows the corresponding normalized amplitude spectra in **b–d**, which are also distorted by the presence of the noise

free data sets can still be significant. Furthermore, these errors can be greatly reduced during data acquisition by simply increasing the sampling rate above a clearly identifiable threshold. As an example, white noise is added in Fig. 6 to the previously analyzed generic wavelet (Fig. 2a), while also increasing the sampling rate. Specifically, the signal is sampled

with a 5.0 GHz sampling rate, which is 10 times the signal central frequency (i.e. 0.5 GHz, Fig. 2c), and the amplitudes are sampled at the same time instants in all three cases presented in Fig. 6. The maximum amplitudes of the added white noise are, respectively, equal to 1% (Fig. 6a), 5% (Fig. 6c), and 10% (Fig. 6e) of the peak amplitude of the original analog

wavelet. The added noise causes visible distortions in the frequency spectrum (Fig. 6b, d, f), nevertheless the main energy band of the signal is still dominant. As expected, the higher sampling rate causes the maximum peak amplitude error to be consistently lower in Fig. 6 (i.e. 5.2% without noise, Table 2) with respect to Fig. 2b (i.e. 30.9%, with the sampling rate equal to 2.0 GHz). The white noise does add to the amplitude uncertainty; however, the maximum peak amplitude error is still present, being dominant in Fig. 6a with respect to the noise (i.e. 5.2% against 1%), comparable in Fig. 6c (i.e. 5.2% against 5%), and lower only in Fig. 6e (i.e. 5.2% against 10%). Moreover, if we sampled the signals in Fig. 6 at a lower rate (e.g. 2.0 GHz), the maximum peak amplitude error would be dominant in all three cases (i.e. 30.9% against 1, 5, and 10%).

As previously discussed when analyzing Fig. 2, not all the signal peaks are sampled in the worst-case scenario, and different events may show different frequency contents, which would change the local peak amplitude errors. Moreover, in the smaller peaks the added white noise can be consistently dominant due to the lower signal-to-noise ratio. For example, if we sample the first signal peak in Fig. 2a in the worst-case scenario with a sampling rate equal to 2.0 GHz, the different frequency content results in a maximum amplitude error equal to 35.7%, while the added white noise in Fig. 6e would be equal to 59.1% of the local peak, due to its lower amplitude. In other words, as expected, lower signal-to-noise ratios produce larger signal distortions.

The issues discussed in this paper have a general validity and can have a remarkable impact on real geophysical data sets, especially when reflections have to be quantitatively analyzed, inverted, and recovered using a gain function. Therefore, quantitative analyses based on the recorded amplitudes, like in the cases of AVO (e.g. Ostrander 1984), attributes (e.g. Chopra and Marfurt 2007), and inversion calculations (e.g. Sen 2006), require the application of sampling rates well above the Nyquist–Shannon threshold. The synthetic data analyses presented in this paper indicate that a sampling rate equal to at least 12 times the central frequency of the recorded signal will consistently limit the peak amplitude error within 5%. We showed in the CO analysis (Fig. 5)

that the variability of the inversion results rapidly increases with sampling rates lower than such threshold. Similarly, the amplitude analysis performed on the CMP synthetic data set (Fig. 4) indicates that a sampling rate of about 12 times the central frequency can be considered a good practical choice. Different sampling rates may be selected, based on the specific objectives of the survey and on the effective importance of the resulting peak amplitude errors for the subsequent data analysis.

5. Conclusion

The presented analysis of the amplitude behavior of GPR signals, with specific reference to source wavelet characterization, amplitude decay with distance, and sampling errors, leads the following conclusions:

1. it is possible to estimate the signal emitted by a transmitting antenna (disregarding system instabilities or jitter) by analyzing its reflection from an air-metal interface. After correcting for spreading losses, we can estimate the reference amplitude of a specific antenna and compare it to the recorded reflections in the calculation of the subsurface EM impedance contrasts, taking into account possible changes in the radiation pattern due to the refractive focusing associated with the air–ground interface;
2. before being able to compare the reference and reflected amplitudes, we have to minimize possible sampling-related amplitude errors. Such errors affect the digitized signal in an unpredictable way and re-sampling the data through interpolation offers only limited and unverifiable solutions to the problem;
3. the uncertainty introduced by peak amplitude errors is comparable to other factors, such as noise and interference, and it can significantly impact the quantitative amplitude analysis and inversion of any GPR data set. Nevertheless, while noise and interference are generally uncontrollable and intrinsic to any recorded data set, sampling-related signal distortions can be greatly reduced by

increasing the minimum threshold for the sampling rate;

4. based on our analysis, we recommend using during data acquisition a sampling rate at least 12 times the signal central frequency for most quantitative applications, which is higher than the commonly adopted standards. This rate not only ensures the avoidance of aliasing effects, in compliance with the Nyquist–Shannon theorem, but it also allows to limit the peak amplitude error within 5%. From the operational point of view, the available sampling and storage technologies do not impose severe limitations to the choice of higher sampling rates in most GPR applications, even though it could still be a problem for large 3-D reflection seismic data sets.

The aforementioned problems and the subsequent conclusions have a general validity, and they are true for the sampling of any physical phenomenon as a function of either time, space, or any other parameter. They furthermore apply to data sets with variable sampling intervals.

Acknowledgements

REFERENCES

Al-Qadi, I. L., & Lahouar, S. (2005). Measuring layer thickness with GPR—Theory to practice. *Construction and Building Materials, 19*(10), 763–772.

Annan, A. P. (2005). Ground-penetrating radar: In near-surface geophysics. In D. K. Butler (Ed.), *Investigations in geophysics* (Vol. 13, pp. 357–438). Tulsa: Society of Exploration Geophysicists. ISBN 978-1-56080-130-1.

Annan, A. P., & Cosway, S. W. (1992). Simplified GPR beam model for survey design. In *Proceedings of the 62nd SEG annual meeting, New Orleans, Louisiana, USA, 25–29 October 1992* (pp. 356–359).

Castagna, J., & Backus, M. M. (Eds.) (1994). Offset dependent reflectivity: Theory and practice of AVO analysis. In *Investigations in geophysics* (Vol. 8). Tulsa: Society of Exploration Geophysicists. ISBN 978-1-56080-059-0.

Chopra, S., & Marfurt, K. J. (2007). *Seismic attributes for prospect identification and reservoir characterization.* Tulsa: SEG/EAGE. ISBN 978-1-56080-141-2.

Dell'Aversana, P. (2013). Listening to geophysics: Audio processing tools for geophysical data analysis and interpretation. *The Leading Edge, 32*(8), 980–987.

Dell'Aversana, P. (2014). A bridge between geophysics and digital music—Applications to hydrocarbon exploration. *First Break, 32,* 51–56.

Dossi, M., Forte, E., Pipan, M., Colucci, R. R., & Bortoletto, A. (2016). Automated reflection picking and inversion: Application to ground and airborne GPR surveys. In *IEEE proceedings of the 16th international conference on ground penetrating radar—GPR2016, Hong Kong, 13–16 June 2016* (pp. 1–6). ISBN 978-1-5090-5181-6.

Drijkoningen, G. G. (2003). *Seismic data acquisition.* TA3600, Delft University of Technology, Section Applied Geophysics and Petrophysics.

Eldar, Y. C. (2015). *Sampling theory: Beyond bandlimited systems.* Cambridge: Cambridge University Press. ISBN 978-1-10700-339-2.

Forte, E., Dossi, M., Pipan, M., & Colucci, R. R. (2014). Velocity analysis from common offset GPR data inversion: Theory and application to synthetic and real data. *Geophysical Journal International, 197*(3), 1471–1483.

Gabor, D. (1946). Theory of communications. *Journal of the Institute of Electrical Engineers, 93*(26), 429–457.

Gersho, A., & Gray, R. M. (1991). *Vector quantization and signal compression.* Berlin: Springer. ISBN 978-0-7923-9181-4.

Giannopoulos, A. (2005). Modelling ground penetrating radar by GprMax. *Construction and Building Materials, 19*(10), 755–762.

Jiao, Y., McMechan, G. A., & Pettinelli, E. (2000). In situ 2-D and 3-D measurements of radiation patterns of half-wave dipole GPR antennas. *Journal of Applied Geophysics, 43*(1), 69–89.

Jol, H. M. (2009). *Ground penetrating radar: Theory and applications.* Amsterdam: Elsevier Science Ltd. ISBN 978-0-444-53348-7.

Kuffel, E., Zaengl, W. S., & Kuffel, J. (2000). *High voltage engineering: Fundamentals* (II ed.). Oxford: Newnes. ISBN 0-7506-3634-3.

Linvill, W. K. (1949). Analysis and design of sampled data control systems. *DSc. Thesis, MIT Project Whirlwind Report R-170.*

Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the AIEE, 47,* 617–644.

Ostrander, W. J. (1984). Plane wave reflection coefficients for gas sands at non normal angles of incidence. *Geophysics, 49,* 1637–1648.

Proakis, J. G., & Manolakis, D. G. (2006). *Digital signal processing: Principles, algorithms, and applications* (4th ed.). New Jersey: Prentice Hall Inc. ISBN 978-0131873742.

Saarenketo, T., & Scullion, T. (2000). Road evaluation with ground penetrating radar. *Journal of Applied Geophysics, 43*(2–4), 119–138.

Schon, K. (2013). *High impulse voltage and current measurement techniques: Fundamentals—Measuring instruments—Measuring methods.* Berlin: Springer. ISBN 978-3-319-00377-1.

Sen, M. (2006). *Seismic inversion: An interdisciplinary approach to topics in petroleum engineering and geosciences.* Richardson: Society of Petroleum Engineers. ISBN 978-1-55563-110-9.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 379–423.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE, 37*(1), 10–21.

Strange, A. D. (2013). Analysis of time interpolation for enhanced resolution GPR data. In *IEEE proceedings of the 7th international workshop on advanced ground penetrating radar—IWAGPR2013, Nantes, France, 2–5 July 2013* (pp. 1–5). ISBN 978-1-4799-0940-7.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics, 49,* 1259–1266.

Tillard, S., & Dubois, J. C. (1995). Analysis of GPR data: Wave propagation velocity determination. *Journal of Applied Geophysics, 33,* 77–91.

Widrow, B., & Kollar, I. (2008). *Quantization noise: Roundoff error in digital computation, signal processing, control, and communications*. Cambridge: Cambridge University Press. ISBN 978-0-52188-671-0.

Yilmaz, Ö. (2001). *Seismic data analysis: Processing, inversion and interpretation of seismic data* (2nd ed., Vol. 1). Tulsa: SEG. ISBN 1-56080-098-4.