



Extracting the subsurface stratigraphy by means of a novel automated reflection strength tracking procedure: Theory and application to ground-penetrating radar data

Matteo Dossi^{a,b}, Emanuele Forte^{c,*}, Barbara Cosciotti^b, Sebastian Emanuel Lauro^b, Elisabetta Mattei^b, Michele Pipan^c, Elena Pettinelli^b

^a Institute for Electromagnetic Sensing of the Environment (IREA), National Research Council (CNR), via Diocleziano, 328, 80124 Naples, Italy

^b Department of Mathematics and Physics, Roma Tre University, via della Vasca Navale, 84, 00146 Rome, Italy

^c Department of Mathematics, Informatics, and Geosciences, University of Trieste, via Edoardo Weiss, 1, 34128 Trieste, Italy

ARTICLE INFO

Keywords:

Auto-tracking
GPR
Reflection dip
Reflection strength
Stratigraphy

ABSTRACT

We developed an algorithm designed to automatically detect and accurately track reflections within geophysical data sets, whether seismic or georadar, based on their reflection strength. Each signal trace is divided into a series of energy packets containing either reflected, diffracted, interfering, or noise-related signals. Acceptable pairs across nearby traces are then connected using the locally-estimated reflection dips, thus constructing a network of energy packets covering the entire data set. Starting from automatically determined initial seeds, the algorithm moves along the network, selecting at each node the path that most likely follows the reflection that is supposedly being tracked. After constructing all possible horizons, the procedure disregards the expected noise-related false positives, which randomly connect unrelated energy packets in areas lacking any recognizable structure. Among its advantages, the proposed algorithm can be applied with minimal signal preprocessing and it requires only limited input from the interpreter, thus leading to more objective results. Moreover, the procedure is able to track a reflection even when its lateral continuity is locally disrupted by either low signal-to-noise ratios, interfering events, or splitting horizons, which typically hinder auto-picking algorithms that are based on waveform-related attributes. The presented algorithm is tested on a georadar profile acquired on the Boulder Clay Glacier, located in Victoria Land, Antarctica, showing an overall good performance across the data set, which includes a well-stratified internal section with wide-ranging reflection dips, as well as more complex structures along the bedrock and the moraine.

1. Introduction

Ground-penetrating radar (GPR; a.k.a. georadar) is a non-invasive geophysical technique that probes the near-surface with movable antennas, which transmit electromagnetic (EM) wavelets (typically within the 10 MHz – 3 GHz frequency range) into the ground and record the returning signals (Jol, 2009). In the simplest acquisition geometry, a common offset (CO) GPR system involves a single transmitter-receiver pair with a fixed separation distance (a.k.a. antenna offset), that can quickly cover large areas at high spatial acquisition rates (Jol, 2009). As a result of this high data density, CO GPR surveys effectively produce images of the subsurface, or rather of its response to the propagation of

EM signals, with varying degrees of resolution, penetration, and data quality, depending on both systemic and environmental factors (Jol, 2009). Detailed imaging and subsurface models can subsequently be obtained by applying qualitative and quantitative analyses to the recorded data sets. In this regard, a particularly important process is the accurate tracking of the recorded reflections, which are marked with horizons whose arrival times are then combined with the reconstructed EM velocity distribution to recover the subsurface stratigraphy (Forte et al., 2014; Dossi et al., 2018).

Over the years, several methods have been developed to assist in the interpretation of seismic data, particularly for the identification of stratigraphic surfaces (Dorn, 1998; Moqadam and Eisen, 2025).

* Corresponding author.

E-mail addresses: dossi.m@irea.cnr.it (M. Dossi), eforte@units.it (E. Forte), barbara.cosciotti@uniroma3.it (B. Cosciotti), sebastian.lauro@uniroma3.it (S.E. Lauro), elisabetta.mattei@uniroma3.it (E. Mattei), pipan@units.it (M. Pipan), elena.pettinelli@uniroma3.it (E. Pettinelli).

<https://doi.org/10.1016/j.jappgeo.2026.106168>

Received 30 July 2025; Received in revised form 27 December 2025; Accepted 13 February 2026

Available online 20 February 2026

0926-9851/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

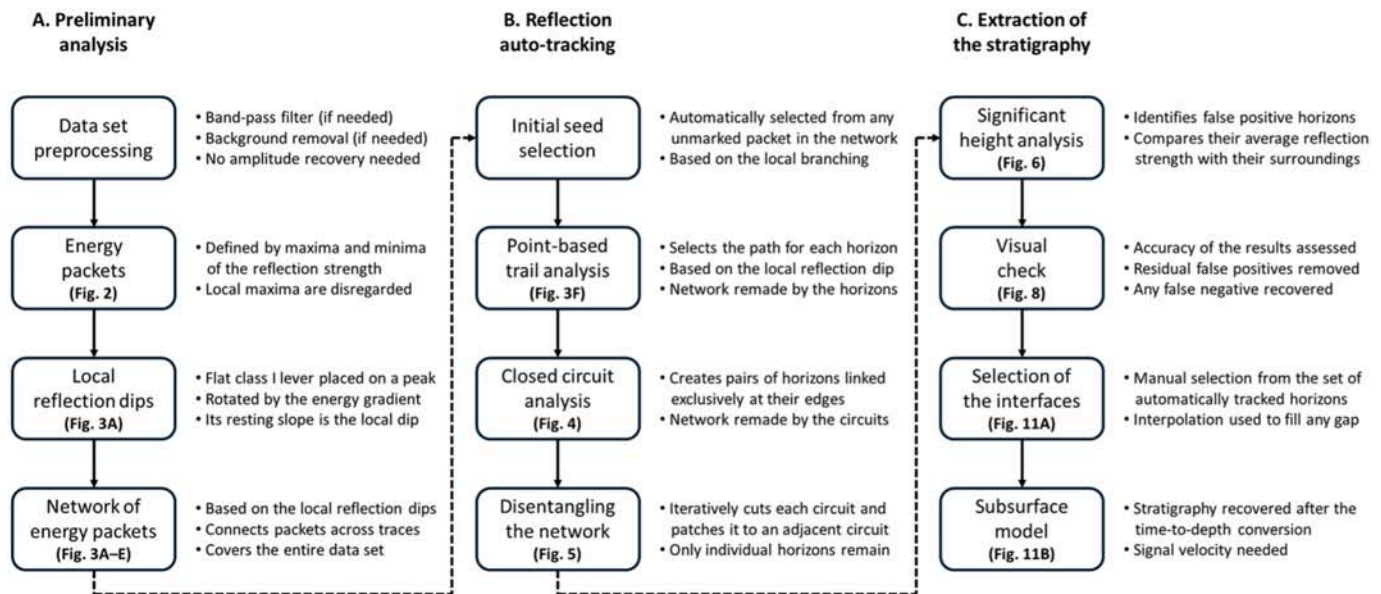


Fig. 1. Flowchart of the proposed automated reflection strength tracking algorithm. The procedure is divided into three main sections, namely (A) a preliminary analysis of the geophysical (i.e., seismic or georadar) profile, to construct a network of reflection strength peaks that covers the entire data set; (B) the actual auto-tracking process, which tracks every laterally-coherent structure through the network; and (C) the post-tracking analysis, to assess the accuracy of the results and construct a model of the subsurface.

Straightforwardly applicable to GPR surveys, these techniques historically included (from least efficient to most efficient) manual picking, interpolation, automated picking, surface slicing, and voxel tracking (Dorn, 1998). All these methods greatly differ in terms of the implemented reflection-tracking criteria, the assumptions made regarding the data quality, the accuracy of the reconstructed stratigraphy, and the independence of the results from the interpreter (Moqadam and Eisen, 2025). For instance, manual picking relies entirely on the expertise of the interpreter, who personally identifies and tracks each reflection by visually assessing the lateral continuity of the amplitude and phase of the reflected wavelets. As a result, manual picking is considered the least effective reflection tracking method, in terms of both the objectivity of the results and the time and effort required for the analysis, especially when dealing with large amounts of data (Dorn, 1998). Similarly to manual picking, interpolation algorithms require the manual selection of initial control points (a.k.a. seeds), which are then automatically connected under the assumption that the event being tracked is locally smooth, or even linear (Dorn, 1998). Although they are generally more efficient than manual picking, interpolation procedures can still lead to poor performances when such assumption is violated, for instance, due to the presence of faults or interfering events causing discontinuities between control points (Dorn, 1998).

The development of computer-assisted methods for the identification of seismic reflections dates back to the 1950s, when Wadsworth et al. (1953) used linear operators as statistical tools in the analysis of time series. Subsequently, statistical signal detection methods were further used successfully in radar studies, with an emphasis on hypothesis testing, and to a lesser extent in nuclear detection seismology and in exploration seismology (Paulson and Merdler, 1968). To give just a few examples of the large variety of reflection-tracking methods developed since then, the algorithm proposed by Paulson and Merdler (1968) marks the time positions of either all peaks or all troughs in a given data set, after which a lateral continuity search correlates these peaks or troughs across traces to define reflection segments. Other techniques apply texture analysis to cluster the various seismic horizons into classes according to their specific features, and then use these horizons as seeds for the segmentation of the geophysical image (Pitas and Kotropoulos, 1989; Pitas and Kotropoulos, 1992). The method presented in Hoetz and Watters (1992) instead combines automatic horizon tracking and

subsequent horizon processing, resulting in maps of various horizon attributes (i.e., reflection time, reflection dip, azimuth, shaded relief, and amplitude). Conversely, Maroni et al. (2001) combine edge following algorithms with multiresolution analysis using the wavelet transform; while Faraklioti and Petrou (2004) present a surface detection technique that identifies the various horizon fragments within 3-D data sets, and then combines them based on their proximity and orientation. Another method uses the Canny edge detection to highlight the points in the data set that most likely belong to horizons, which are then used as input for the Kalman filter to track and recognize horizons trace by trace (Deng and Zhang, 2017; Deng and Hu, 2019). The semi-automatic horizon extraction algorithm proposed by Alves et al. (2023) is instead based on the absolute phase of seismic data obtained through phase unwrapping techniques. Attribute analysis is also implemented in Zhao et al. (2020), who combine envelope, phase attributes, and texture analyses; and in Lou et al. (2020), whose algorithm first generates horizon patches using the reflector dip, and then improves their quality using the instantaneous phase. In addition to the more traditional methods, various machine learning and neural network techniques are more commonly being implemented in recent years (Di et al., 2020; Yang and Sun, 2020; Ball and O'Connor, 2021; Roncoroni et al., 2022), even though some of the earliest examples date back to the 1990s (Roberto et al., 1990; Harrigan et al., 1992).

1.1. Automated picking

The algorithm presented in this paper belongs to the auto-picking (a.k.a. auto-tracking) methods, whose fundamental purpose is to “quickly and accurately perform straightforward event correlations” (Herron, 2000). As a result, these methods allow the interpreter to focus on more complex sections and manually intervene where the algorithm might fail, such as in areas affected by either considerable interference or significantly low signal-to-noise ratios (S/N). A generic reflection auto-tracking algorithm starts from a manually selected initial seed placed on a clearly-identifiable reflected wavelet. From this control point, it then searches the adjacent traces for comparable features (e.g., similar signal amplitudes, wavelet correlation, or lateral phase continuity) within specified time windows. In particular, Dorn (1998) identifies two main classes of auto-picking algorithms, namely (A) feature trackers, which

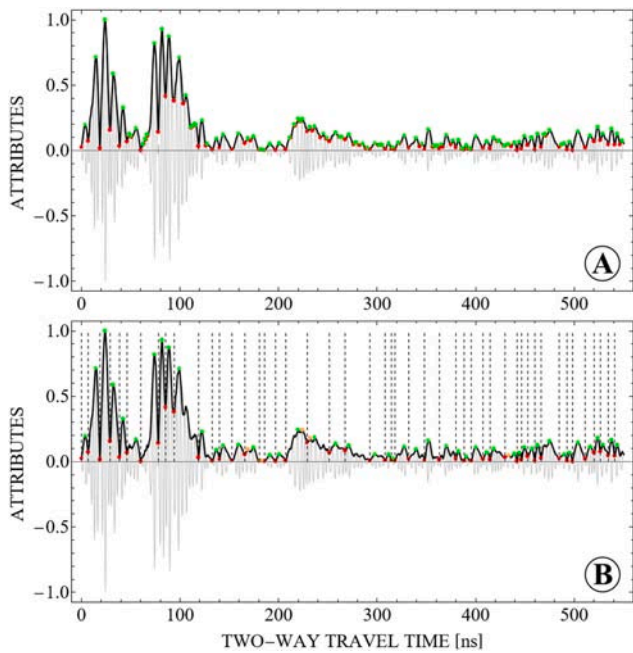


Fig. 2. Partition of a generic signal trace into a series of energy packets. The figure shows the normalized trace before (A) and after (B) the local maxima and minima of the reflection strength are removed. The respective graphs present the signal amplitude (thin line), the reflection strength (thick line), the boundaries of the energy packets (dashed lines), and the reflection strength maxima (green dots), minima (red dots), and secondary maxima (orange dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

search for particular configurations of samples from trace to trace (e.g., peaks, troughs, or zero-crossings), within the specified time window; and (B) correlation trackers, which compare specific sections of adjacent traces through amplitude correlation, while looking for a match within an acceptable lag time. Correlation trackers generally tend to be more robust than feature trackers, albeit also more computationally intensive (Dorn, 1998). Among these two classes, feature trackers are the ones closest to the reflection auto-tracking algorithm proposed in this paper.

In any case, a feature identified in an adjacent trace is recognized as part of the reflection being tracked when the relevant signal attributes and parameters, as defined by the implemented auto-picking criteria, fall within their respective acceptable ranges (Dorn, 1998). The algorithm then moves on to the new position and reiterates the process, searching for similarly acceptable features on traces further away until the auto-tracking criteria are no longer met. At the end of the process, the last acceptable features in their respective moving directions constitute the edges of the constructed horizon, which ideally should mark the automatically tracked reflection in its entirety. However, auto-tracking algorithms can be sensitive to the presence of significant noise and interference, since they generally assume that the analyzed data set is locally continuous, smooth, and consistent (Dorn, 1998).

In particular, Herron (2015) identifies four common pitfalls for horizon auto-picking, namely (A) improperly selecting the grid of initial seeds, which could potentially lead to either grossly erroneous and meaningless results (i.e., a too coarse grid), or needlessly increased computation times due to redundant analyses (i.e., a too fine grid); (B) not accounting for the directional behavior of the algorithm when specifying the initial control points, which for instance could affect peculiar situations where the signal being tracked effectively splits (e.g., horizon convergence, divergence, or signal interference); (C) tracking reflections characterized by poor lateral continuity or affected by significantly low S/N, for which manual tracking and interpolation between control points could be preferable to fully automated procedures;

and (D) failing to recognize geological structures that are not suitable for auto-tracking, as they are not defined by laterally coherent or continuous reflections but are still distinguishable from noisy data, and thus require a more careful interpretation. All these pitfalls can significantly affect the quality of the reconstructed stratigraphic model and they may be chronic throughout a project, albeit with very subtle effects that makes their detection more difficult; or they may be acute, and thus require a significant interpretive effort to correct the issues they create (Herron, 2015). In any case, auto-picking algorithms are primarily meant to aid in the interpretation of a profile and can be expected to produce a certain number of false positives and false negatives in the more complex areas. As a result, they should not be expected to accurately track every single recorded event, nor should the interpretation be exclusively focused on those events that can be accurately tracked by auto-picking algorithms (Herron, 2000; Herron, 2015).

The main advantage of the presented algorithm is the general avoidance of the first three pitfalls identified by Herron (2015), as a result of its innovative auto-tracking criteria, which are summarized in Fig. 1 through a flowchart. The proposed procedure is designed to effectively follow the energy of the reflected signals across a network that covers the entire data set, locally connecting the energy peaks identified in each trace with other peaks in other traces nearby. While constructing an individual horizon along the network, the reflection strength auto-tracking algorithm (A) automatically selects the initial seed from the network based on its likelihood of belonging to an actual reflection, without the need of a subjectively pre-defined grid; (B) individually analyzes the nodes encountered along the network to account for the convergence, divergence, and interference of the auto-tracked events; and (C) is less sensitive to local noise-related disruptions in the lateral continuity of a reflection, since it follows the signal energy rather than more traditional waveform-related attributes. In addition, after constructing all possible horizons, the algorithm separates the horizons marking actual reflections from the expected noise-related false positives (Fig. 1), as a further aid to the interpreter during the final check of the auto-tracking results. The interpreter can then select the various interfaces from the constructed set of horizons, while also assessing the accuracy of the results, and manually intervene in those areas where the algorithm might fail (i.e., the fourth pitfall; Herron, 2015).

2. Preliminary analysis of the reflection strength profile

2.1. Energy packets

The preliminary analysis of the geophysical (i.e., seismic or georadar) profile divides each recorded trace into a series of energy packets (a.k.a. energy peaks), which identify the successive peaks of the reflection strength (a.k.a. trace envelope or instantaneous amplitude; Taner et al., 1979). First described in Forte et al. (2016) in its earlier version, the procedure marks all the maxima and minima of the reflection strength within a given trace (Fig. 2A), and then disregards all the local maxima and minima so that only actual maxima remain, separated by minima that define the boundaries of each energy packet (Fig. 2B). A minimum in the reflection strength S_{ij} (either actual or local) is defined by the following condition:

$$S_{i-1,j} \geq S_{ij} \wedge S_{ij} < S_{i+1,j}, \quad (1)$$

where S_{ij} is the reflection strength in the i -th sample in the j -th trace of the data set. The first and last samples of the trace are automatically considered boundaries of the first and last packets, respectively; while the various maxima (either actual or local) are the samples with the highest reflection strength within each packet (Fig. 2A).

The algorithm analyzes each minimum in a given trace, from the strongest to the weakest, and compares it with the two adjacent maxima. If the minimum is at most equal to a set percentage (e.g., 50%) of the

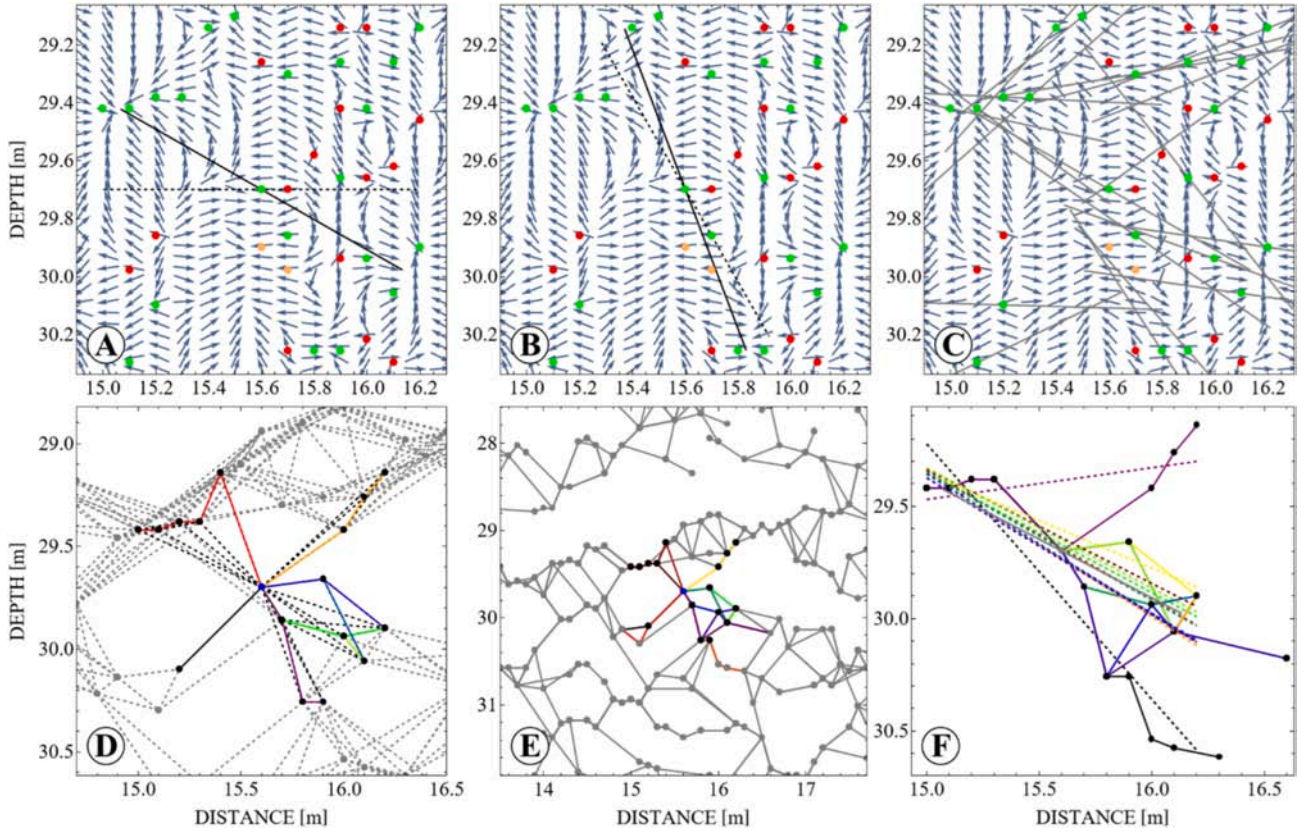


Fig. 3. Different stages of the proposed reflection strength tracking procedure. The figure shows (A) the reflection dip analysis, with the virtual lever in its initial (dashed line) and resting (solid line) positions; (B) the same analysis, but with a different initial position; and (C) the reflection dips of all the nearby energy packets. These results are superimposed to the vector plot (blue arrows) of the gradient of the reflection strength (Eq. (3)), with the reflection strength maxima (green dots), minima (red dots), and secondary maxima (orange dots) also highlighted (Fig. 2B). The figure further shows (D) the original network (gray dashed lines) connecting the reflection strength maxima (gray dots), from which the identified obtuse triangles (black dots and dashed lines) are used to construct eight horizons (colored lines) branching out of the analyzed packet (i.e., two leftward and six rightward). From the simplified network (gray lines) shown in (E), seventeen different paths (colored lines) are subsequently identified (i.e., four leftward and thirteen rightward). Assuming a rightward tracking direction, the trail analysis shown in (F) combines the already travelled path (i.e., one of the four) with all the possible paths forward, to then compare the corresponding linear fits (colored dashed lines) with the local reflection dip (gray solid line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

weaker maximum, the two maxima are marking different peaks separated by the analyzed minimum (Fig. 2). Conversely, if the minimum is above such threshold, the two maxima are marking the same peak, therefore the weaker maximum and the minimum are disregarded as local (Fig. 2). In some cases, however, constructive interference can raise a minimum above the threshold, thus causing the algorithm to erroneously group close but distinct wavelets into the same packet. In order to prevent this, two energy packets can only be joined if (A) the minimum in-between is above the set threshold; (B) either one has a width smaller than a specified time interval (e.g., the time-length $3T_R$ of a reference three-phase wavelet for GPR data sets); and (C) the two maxima are closer in time than such interval. The reference time interval T_R is the average interval of the recorded signal phases, which is roughly estimated from the central frequency ν_C of the signal spectrum using the following equation:

$$T_R = \frac{1}{2\nu_C}. \quad (2)$$

Moreover, given the potentially long energy tails and possible signal distortions, the width mentioned in condition (B) is not the time difference between the two minima at the boundaries of either packet. Instead, it is the temporal distance from the analyzed minimum in which the reflection strength returns either at or below the level of such minimum, in either direction along the trace. Furthermore, also due to possible signal distortions, condition (C) only applies when the weaker

maximum would constitute a secondary maximum within the joined energy packet, that is when its absolute reflection strength difference with the stronger maximum is smaller than its difference with the minimum in-between (Fig. 2B).

For the purpose of reflection auto-tracking, local maxima in general, and secondary maxima in particular, can indicate signal distortions from either noise, interference, or reflection wedges, which could cause individual horizons to split into multiple branches marking different events. On a similar note, the presented auto-tracking algorithm focuses on the main peak of each energy packet, defined as the section within its full width at half maximum, to prevent possible deviations caused by long energy tails.

2.2. Local reflection dips

The algorithm estimates the reflection dip associated with each energy packet within the data set based on the local gradient of the reflection strength ∇S_{ij} , which is obtained from the discrete partial derivatives along the horizontal and vertical axes:

$$\nabla S_{ij} = \left(\frac{S_{ij+1} - S_{ij-1}}{2\Delta x}, \frac{S_{i+1,j} - S_{i-1,j}}{2\Delta z} \right), \quad (3)$$

with

$$\Delta z = \frac{\Delta t}{2v_c}, \quad (4)$$

where Δx is the trace interval and Δz is the depth interval, as calculated from the sampling interval Δt using a reference signal propagation velocity v_c set by the interpreter (e.g., 20 cm/ns for the glaciological GPR data set studied in this paper).

At any given position, this gradient is a vector pointing toward the direction of fastest increase in the reflection strength, although not necessarily toward the closest maximum. Since the partition into energy packets is applied to each trace separately (Fig. 2), some maxima and minima may in reality be on lateral slopes or saddle points in what are actually 2-D energy peaks. Notwithstanding these peculiarities, Eq. (3) defines a vector field that (in general) diverges from the minima and converges toward the maxima of the reflection strength (Fig. 3A). The algorithm thus estimates the reflection dip at a given energy peak by having the gradient field act on a virtual class I lever of half-length L centered on the reflection strength maximum (Fig. 3A). Behaving as a force field, the gradient aligns said lever with other reflection strength maxima nearby by creating a virtual torque τ , given by the sum of the vector products of discrete positions $\vec{r}_{ij}$ along the lever (relative to its fulcrum) with the corresponding gradient values ∇S_{ij} :

$$\tau = \sum_{n=-N_L}^{N_L} \vec{r}_{i+\Delta i, j+\Delta j} \times \nabla S_{i+\Delta i, j+\Delta j}, \quad (5)$$

with

$$N_L = R \left[\frac{L}{\sqrt{\Delta x^2 + \Delta z^2}} \right], \quad (6)$$

$$\vec{r}_{i+\Delta i, j+\Delta j} = (\Delta j \Delta x, \Delta i \Delta z), \quad (7)$$

$$\Delta i = R \left[\frac{n}{\Delta z} \sqrt{\Delta x^2 + \Delta z^2} \sin[\phi] \right], \quad (8)$$

$$\Delta j = R \left[\frac{n}{\Delta x} \sqrt{\Delta x^2 + \Delta z^2} \cos[\phi] \right], \quad (9)$$

$$L = R \left[\frac{3T_R}{\Delta t} \right] \Delta z, \quad (10)$$

where $R[x]$ is a function that rounds x to the nearest integer, and ϕ is the dip of the lever as it rotates. The fulcrum used in Eq. (5) is located on the i -th sample of the j -th trace, whereas the half-length L of the lever (Eq. (10)) is derived from the time interval $3T_R$ of the reference wavelet (Eq. (2)).

At the start of each analysis, the virtual lever is set flat so to spread across several traces in the vicinity of the given energy packet (Fig. 3A), all of which will influence its rotation at least initially. The algorithm then iteratively rotates the lever by a discrete angular step (e.g., 0.5°) in accordance with the sign of the virtual torque (Eq. (5)). The rotation continues until either the torque becomes null, at which point the lever is stopped; or it changes sign, in which case the lever settles on the position with the weaker torque, among the ones immediately before and after the change. At the resting position, the lever is aligned with the nearby reflection strength maxima (Fig. 3A), and the final angle ϕ is the local reflection dip associated with the analyzed energy packet.

2.3. Network of energy packets

As a foundation for the auto-tracking process, the algorithm connects each energy packet with other packets in nearby traces, based on its local reflection dip. A nearby packet is connected to the analyzed one if (A) its maximum is within a $2L \times 2L$ square centered on the latter's maximum (Fig. 3A); (B) its main peak is crossed by the latter's virtual lever at rest (or by its linear extension); and (C) it crosses back the latter's

main peak with its own resting lever (or its linear extension) although, in this case, the initial position of the lever is aligned with the two maxima (Fig. 3B). The last condition accounts for both diverging stratigraphic interfaces and signal interference, which can give different results for different initial positions in the torque analysis, thus producing multiple acceptable paths for the auto-tracking algorithm (Fig. 3C).

After establishing all possible connections, the algorithm analyzes the two sides of each packet separately and identifies all the pairs that form triangles with a vertex on the given packet (Fig. 3D). Moreover, these triangles must be obtuse and the analyzed packet must not be on their obtuse-angled vertex, in order to preserve diverging horizons, as opposed to erratically connecting unrelated packets with a single horizon on either side. The sides connecting the analyzed packet with the most distant vertex (horizontally) of each triangle are then iteratively eliminated, and the remaining sides are joined together at the same time, to construct all possible horizons branching out of the given packet (Fig. 3D). After analyzing each packet in the data set, those horizons that are completely overlapped by other horizons are removed to avoid redundancies.

The remaining horizons form a network of energy packets (Fig. 3E) that covers the entire profile, with some horizons marking actual reflections (or at least sections of them), while others are connecting unrelated packets based on interfering or noise-related signals. The algorithm automatically tracks the recorded reflections along this network, with most packets constituting nodes from which the correct paths must be selected. In order to facilitate this process, the algorithm constructs horizons branching out from each energy packet, covering all possible paths on each side (Fig. 3E). The maximum horizontal length of these horizons is the half-length L of the virtual lever (Eq. (10)) although, if a gap is found at such distance, the considered horizon is extended to the next packet within the network. As a result, the original set of horizons branching out of each peak is expanded to include the horizons that can now reach the same peak from further away. In addition, since these horizons are separately used during the analysis of the energy packets from which they are branching out, those horizons that are completely overlapped by other horizons must not be removed, unless the compared horizons are identical.

3. Automated reflection strength tracking

3.1. Initial seed selection

As with other auto-tracking algorithms (Dorn, 1998), the presented procedure starts from initial seeds within the constructed network of energy packets, although such seeds are not manually selected by the interpreter. The algorithm instead tracks reflections from every single packet within the network, unless they already belong to an auto-tracked horizon. The order in which these seeds are analyzed is based on the assumption that clear and undistorted reflections produce long lines along the network (i.e., without nodes). Conversely, a more intricate network is produced in areas that either lack any recognizable structure, show significant interference, or present particularly low S/N. Therefore, the first seeds to be analyzed are those packets in which the paths branching out from them separate the furthest away. More specifically, the order in which the energy packets are analyzed is defined by (A) a decreasing splitting distance; and (B) a decreasing reflection strength maximum, when the splitting distance is the same.

Considering the energy packets common to all the horizons branching out of a given packet on each side, the splitting distance is the horizontal distance (i.e., difference in the trace positions) between the most distant packets from the two sets. However, this distance is capped on each side at the aforementioned threshold L (Eq. (10)), to avoid giving precedence to horizons that are longer simply because of a gap at such distance.

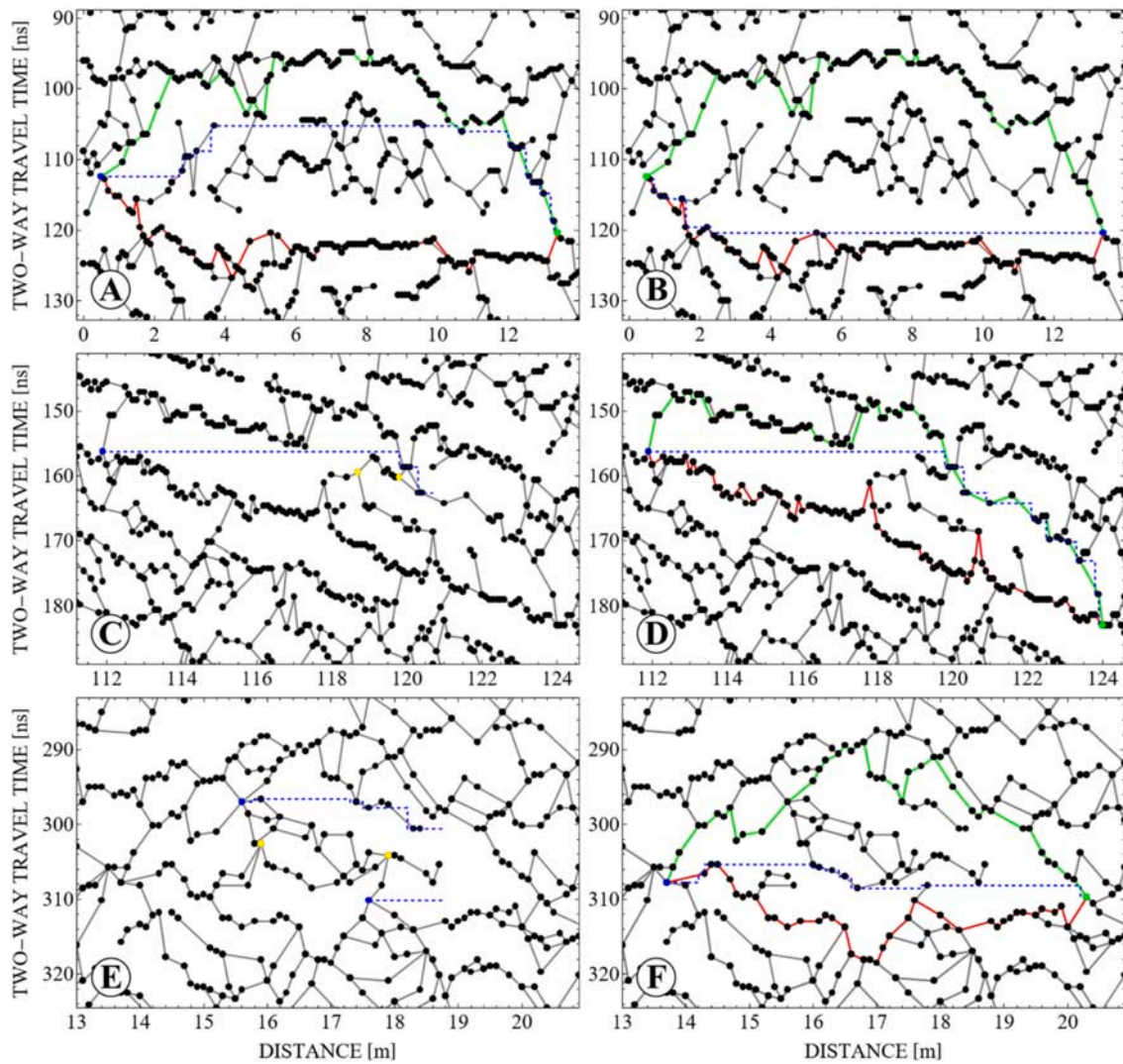


Fig. 4. Closed circuit analysis applied to different scenarios within the network of energy packets. The figure shows (A) the construction of a generic circuit; (B) the construction of the same circuit, but in the opposite direction (i.e., from right to left); (C) the identification and (D) removal of a rightward barrier; and (E) the identification and (F) removal of a leftward barrier. The respective graphs show the trajectory (blue dashed line) followed by the algorithm from one edge of the circuit (blue dot) to the other (green dot), as well as the nodes (yellow dots) that are severed to remove the encountered barriers. The constructed circuits are marked along the network (gray lines) by both a stronger (green line) and a weaker (red line) path, based on the average reflection strength of their energy peaks (black dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Point-based trail analysis

Starting from an initial seed, the algorithm repeats the same analysis at each energy packet encountered along the network (in both directions) to assess the most acceptable path forward, whose next packet is then added to the horizon under construction. The implemented point-based analysis is applied to both branching and non-branching packets, as the latter could still influence subsequent analyses further down the line. At a given packet, the algorithm combines all the possible paths forward with either (A) the already travelled path (Fig. 3F); or (B) all the possible paths backward, in absence of the former (i.e., around the initial seed, at the very start of the process). As previously discussed, the maximum horizontal length of both forward and backward paths is equal to L (Eq. (10)), except for paths with a gap at such distance, which are instead elongated to the next packet along the network. After constructing M trails from all possible combinations of forward and backward paths, the algorithm linearly fits the reflection strength maxima marked by such trails (Fig. 3F), with the resulting dips ϕ_k given by:

$$\phi_k = \arctan[a_k], \quad (11)$$

with $k = 1, 2, \dots, M$, and

$$a_k = \frac{N \sum_i x_i z_i - \sum_i x_i \sum_i z_i}{N \sum_i x_i^2 - (\sum_i x_i)^2}, \quad (12)$$

where a_k is the slope of the N linearly fitted packets of the k -th trail (Cantrell, 2008), whose maxima have spatial coordinates x_i and z_i in the depth-converted profile (Eq. (4)).

A given trail is acceptable if its fitted dip ϕ_k differs from the local reflection dip of the analyzed packet (Fig. 3F) by at most a threshold $\Delta \phi_{max}$ set by the interpreter (e.g., 20°), in which case its forward packets receive one point each. Since interference can distort the local reflection dips, in addition to the acceptable trails, the algorithm also considers all the unacceptable ones with at least one forward packet that retains at least one point from a previous analysis, for the same horizon being constructed. However, the procedure automatically disregards any path whose potential step forward crosses an already constructed horizon, even though crossing segments can be found within the network (Fig. 3E). Among all the considered paths forward, the procedure thus selects either (A) the one with the highest score, averaged among its

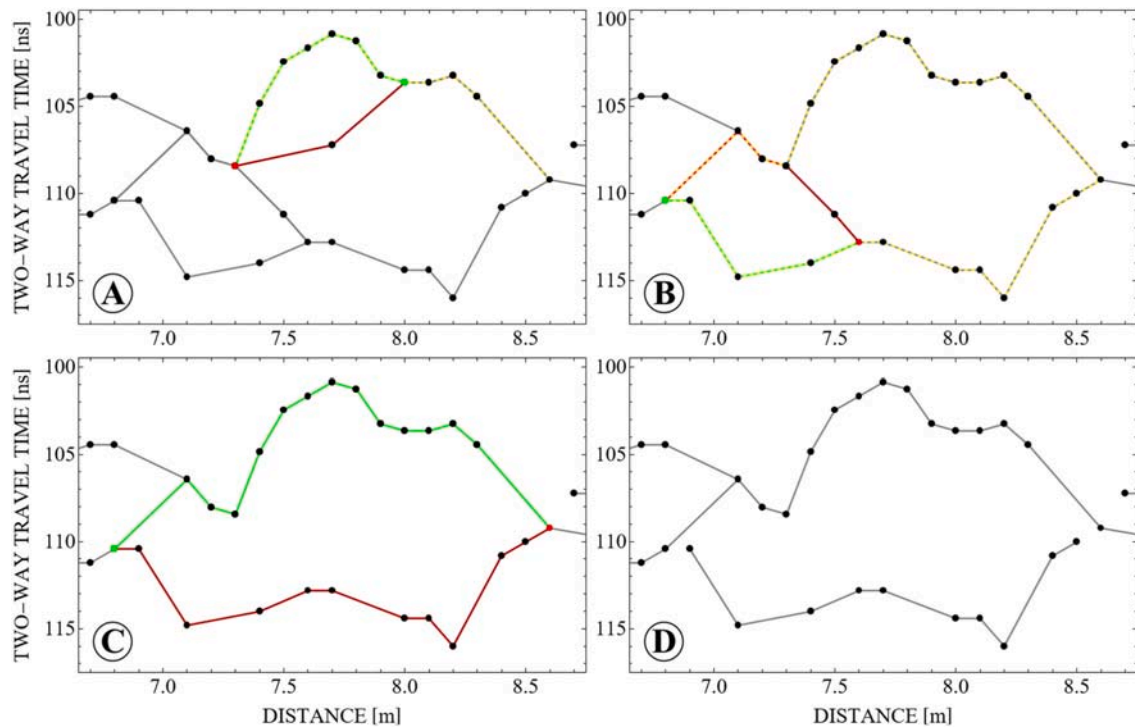


Fig. 5. Local disentanglement of the network of energy packets at different stages. The respective graphs show the cutting of the analyzed closed circuit at a specific section (dark red line) of its weaker horizon (red line), and its subsequent patching (dashed yellow line) with the adjacent circuit (when present). The weaker (red dot) and stronger (green dot) edges of the analyzed circuit, as well as its stronger horizon (green line), are also highlighted. After the cutting and patching analysis, the only remaining nodes are related to non-circuitual horizons (D). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

marked energy packets; or (B) the one whose trail has the highest average reflection strength, among the paths that share the highest averaged score. The algorithm thus moves forward to the next packet along the selected path, which is added to the horizon under construction, to then repeat the same point-based analysis on this new packet. Moreover, if all the considered paths forward share at least one energy packet, the selected path forward is fixed up to the closest common packet irrespective of other possible nodes in-between, to reduce the possibility of deviations into interfering events. Nevertheless, the procedure is applied to each intermediate packet, since they can still influence subsequent analyses, as previously mentioned. Conversely, in order to prevent the trails that deviate from the selected step forward from influencing subsequent analyses, their forward packets have one point removed for each deviating trail that is marking them.

As the algorithm moves along the network, energy packets are iteratively added to the horizon under construction until either no path forward can be selected, or the selected step forward reaches an already constructed horizon, in which case the two horizons are connected at their common packet, and the auto-tracking process is interrupted. After a given horizon is constructed, the algorithm is newly applied to the next initial seed (within the previously compiled list) that is still unmarked. After all possible horizons have been constructed, all the unused paths within the network are removed and a new simplified network is thus created.

3.3. Closed circuit analysis

The final step of the auto-tracking process is to disentangle this simplified network (so that only individual horizons remain), starting from all closed circuits, which are defined as pairs of horizons connected exclusively at their edges, with no other horizon between them connecting said edges (Fig. 4A). Even though the broken lines forming said horizons can present vertical spikes (i.e., upward and downward

concavities), they cannot have lateral ones (i.e., leftward or rightward concavities), which are therefore removed during the construction of a circuit by cutting the network at specific nodes (Figs. 4C–F). The algorithm identifies all closed circuits within the network by analyzing each trace of the profile, from left to right, and treating all energy packets with multiple branches on their right side as the left edges of potential circuits. Only rightward-branching packets are considered in order to save computation time, since no actual circuit would be identified by moving in one direction (Fig. 4A) that would not be identified by moving in the other (Fig. 4B). If more than two horizons are branching out rightward, each pair is analyzed separately, to potentially form different closed circuits sharing at least their left edge.

After defining the initial pair of rightward-branching horizons, the algorithm moves horizontally from the analyzed packet while keeping count of two different sets of horizons, respectively located either above (i.e., the upper set) or below (i.e., the lower set) the followed trajectory, and branching out of either the initial packet or other nodes within the sets. As the algorithm moves forward, more rightward-branching horizons are iteratively added to their respective sets at such nodes, and the energy packet in which the two sets first rejoin constitutes the right edge of the identified circuit (Fig. 4A). Until such packet is encountered, the algorithm keeps moving horizontally unless it is about to cross a horizon from either set, at which point it moves either upward or downward (depending on the set) until the next horizontal step is cleared.

Since the two sets only contain forward-branching horizons, any horizon branching backward (i.e., leftward) from either set constitutes a loose end that can be freely crossed (Fig. 4A). However, if backward-branching horizons from the two sets join to form a barrier with a rightward concavity (in violation of the aforementioned condition), the procedure is interrupted and the barrier is severed at specific nodes (Fig. 4C), after which the analysis is restarted from the same initial packet (Fig. 4D). If the algorithm was moving from right to left (Fig. 4B), barriers with leftward concavities would be identified in a similar

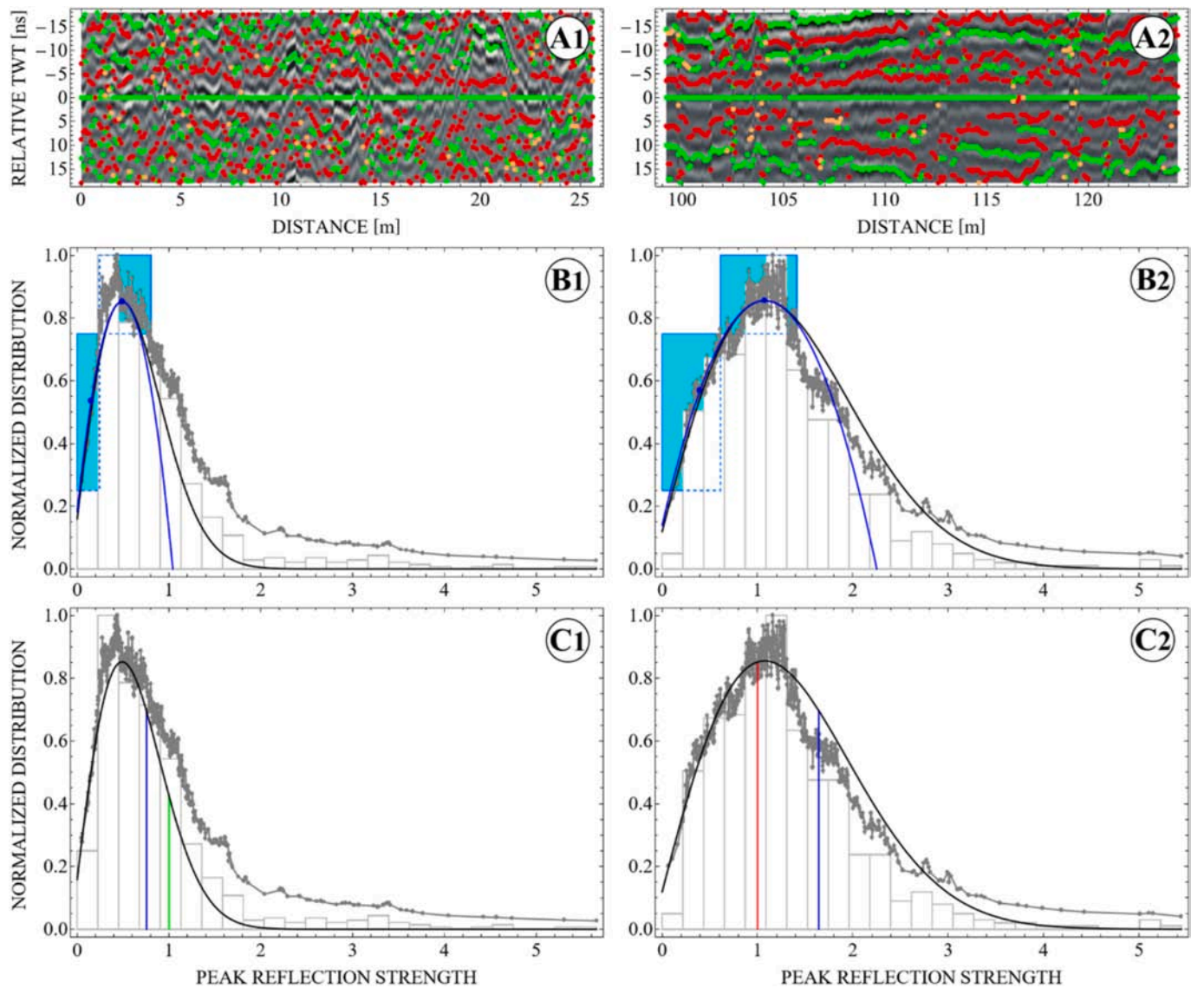


Fig. 6. Significant height analysis applied to two different horizons. In each column, the figure shows (A) the area around the analyzed horizon (flattened for visual clarity), with the local reflection strength maxima (green dots), minima (red dots), and secondary maxima (orange dots) highlighted (Fig. 2B); and (B) the adaptation of the Rayleigh function (black line; Eq. (13)) to the potential function (dark gray dots and line; Eq. (14)) used to approximate the distribution of reflection strength maxima (histogram). As per Appendix B, the areas (blue rectangles) used to define the points (R_M, V_M) and (R_Q, V_Q) (blue dots) crossed by the auxiliary quadratic function (blue line; Eq. (B.7)) are highlighted in (B). The figure also shows (C) the comparison of the average reflection strength (normalized to 1) of the analyzed horizon with the lower margin $R_{1/3}$ (blue line; Eq. (21)) of the upper third of the approximated Rayleigh distribution, which determines whether the analyzed horizon is acceptable (green line) or not (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

manner. In the presented procedure, such barriers are instead detected in specific situations where either the upper or lower set has no more branches propagating rightward (Fig. 4E and F). The identification and removal of barriers during the closed circuit analysis is discussed in detail in Appendix A.

3.4. Disentangling the network

Following the close circuit analysis, the algorithm replaces the network of energy packets with the constructed pairs of horizons from each circuit, while keeping all non-circuitual parts (e.g., isolated, unpaired, or loose-end horizons) so that every connection is maintained. The network is then disentangled from the inside out, by iteratively cutting the smallest circuit and then patching it to a larger adjacent circuit (when present), until there are no closed circuits left. In particular, the procedure analyzes each circuitual horizon from the shortest to

the longest, thus starting with the innermost circuits and then moving outward (Fig. 5), with the length of a horizon defined by the sum of the Euclidean lengths (in the depth-converted profile; Eq. (4)) of its constituting segments (which connect its marked energy peaks; Fig. 4). The cutting and patching procedure assumes that, if the two packets at the edges of the analyzed circuit belong to the same reflection, the stronger circuitual horizon (based on its average reflection strength; Fig. 4) is marking the reflection, while the weaker one is either randomly connecting unrelated energy packets or marking a separate parallel event. The weaker horizon is therefore cut at either (A) a specific section shared with the aforementioned adjacent circuit (Fig. 5B); or (B) its edges, if the external circuit either completely overlaps the horizon (Fig. 5A) or does not exist (Fig. 5C). In order to determine the section to be cut, the procedure first analyzes the stronger edge (based on its reflection strength maximum; Fig. 5), and then moves to the other edge if no acceptable adjacent circuit can be found. If this second attempt also

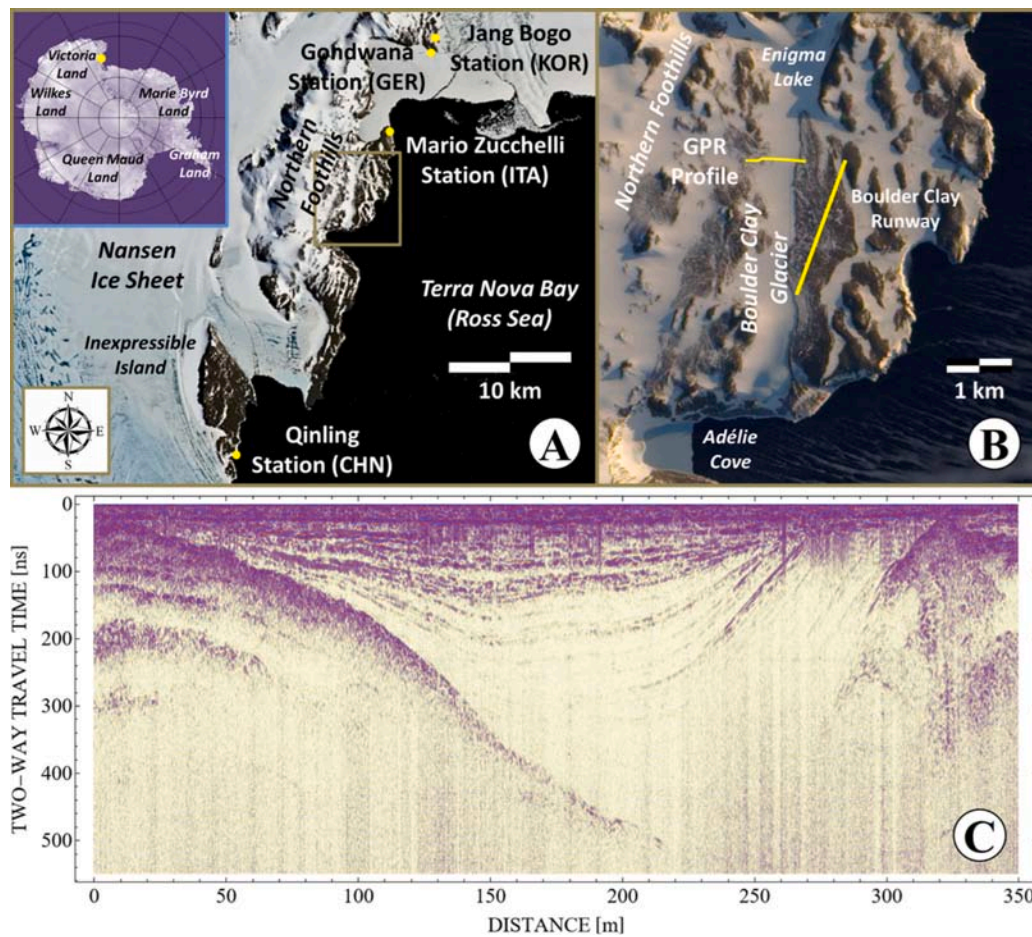


Fig. 7. Location map of the survey area and the analyzed glaciological GPR profile. The figure shows (A) the Northern Foothills, along the western coast of Terra Nova Bay in Victoria Land, Antarctica; and (B) the Boulder Clay Glacier, with superimposed the entire path of the GPR survey, the analyzed section of which is shown in (C). The inset satellite image of Antarctica in (A) is courtesy of professor David Pape, who created it using NASA's Blue Marble data set; while the other two images are courtesy of the NASA Earth Observatory, with data provided by NASA and USGS. In particular, the picture of the Northern Foothills in (A) was created by Joshua Stevens using data acquired in 2020 by the Operational Land Imager onboard the Landsat 8 satellite; while the picture in (B) was created by Jesse Allen and Robert Simmon using data acquired in 2009 by the Advanced Land Imager onboard the Earth Observing-1 satellite.

fails then the entire weaker horizon is to be cut.

The algorithm searches for an adjacent circuit that (A) marks the analyzed edge with either one of its horizons; (B) has one of its own edges marked by the weaker horizon of the analyzed circuit; and (C) has the closest such edge to the analyzed edge (the two cannot coincide), among all the external circuits that satisfy the first two conditions. The weaker horizon is then cut at the two considered edges, creating a new horizon (if long enough) that may still mark a weaker event (Fig. 5D), while the analyzed and adjacent circuits (when the latter is present) are patched together to form a larger circuit (Fig. 5B). Conversely, if the adjacent circuit covers the entire weaker horizon of the analyzed circuit, the overlapping section is simply moved to the stronger horizon (Fig. 5A), which is thus made redundant. Afterwards, the algorithm updates the previously compiled length-based list of circuit horizons and moves on to the next one, repeating the process until all the closed circuits in the network have been cut open.

Following the cutting and patching process, the algorithm removes the last remaining nodes (belonging to the aforementioned non-circuit horizons) through a procedure similar to the trail analysis applied during the auto-tracking process (Fig. 3F). The residual nodes are analyzed from the strongest to the weakest, and the local reflection dip of each one is compared with the slopes (Eq. (11)) of the linear fits of all possible trails branching out of their node. Similarly to the auto-tracking process, the maximum horizontal length of the paths constituting these trails is equal to L (Eq. (10)), although those paths with a gap at such distance on

either side are elongated to the next connected energy packet. The trail with the closest linearly fitted slope to the local reflection dip is preserved by the algorithm, while the remaining paths are severed from the node to form independent horizons (if long enough).

4. Significant height analysis

After the network is completely disentangled, the final horizons are marking either (A) actual reflections, although some may still deviate into interfering or noise-related events; or (B) unrelated energy packets in areas lacking any laterally coherent structure or affected by particularly low S/N. In order to identify the expected false positives, the algorithm compares the average reflection strength of each horizon with the distribution of the reflection strength maxima in their vicinity. The considered area covers horizontally the entire trace interval of the analyzed horizon, and vertically a time interval centered on its marked reflection strength maxima in their respective traces (Fig. 6A). This time interval has a width equal to $12T_R$ (Eq. (2)), so that the area ideally covers all the energy packets just above and below the horizon. If the analyzed horizon contains gaps, in those traces the time interval is centered on the straight lines connecting the edges of such gaps. Furthermore, to account for possible lateral changes in the subsurface reflectivity, the considered reflection strength maxima are normalized (for the sole purpose of this analysis) with respect to either (A) the maxima marked by the horizon in their respective traces; or (B) the

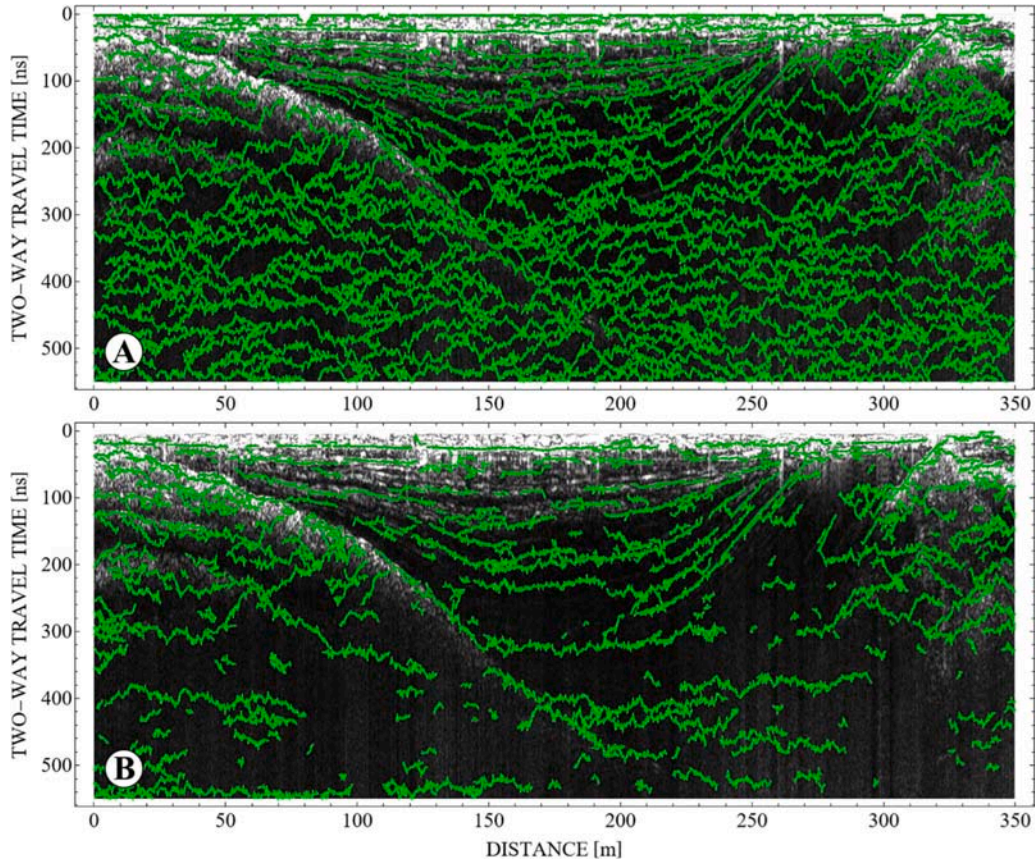


Fig. 8. Reflection auto-tracking results superimposed to the reflection strength profile. The figure shows (A) the automatically constructed horizons marking at least 50 energy peaks (i.e., at least 4.9 m long without gaps); and (B) the horizons remaining after the significant height analysis and marking at least 10 energy peaks (i.e., at least 0.9 m long without gaps).

average reflection strength of the horizon, if such traces contain a gap. Following this normalization, the maxima marked by the horizon are removed from the constructed set, to prevent their identical values (i.e., equal to 1 by construction) from significantly distorting the resulting distribution.

The general principle of the significant height analysis is that the average of the marked reflection strength maxima (which is equal to 1 after the normalization) would stand out in the constructed distribution for genuine reflection horizons. Conversely, noise-related false positives would have similar values to their surroundings (in areas with low S/N), or even lower values (near actual reflections). Inspired by the study of rogue waves in oceanography, the analysis assumes that the reflection strength maxima within a generic section follow the Rayleigh distribution (Stansell, 2004; Muraleedharan et al., 2007):

$$V(r) = \frac{r}{R_M^2} \exp \left[-\frac{r^2}{2R_M^2} \right], \quad (13)$$

where R_M is the mode of the distribution, in which Eq. (13) reaches its maximum.

In the case of a discrete distribution, the mode indicates the value that appears most often in the analyzed set. However, since the considered samples are unlikely to repeat themselves without rounding, the algorithm defines the mode as the position with the highest concentration of samples within the covered reflection strength interval. Such position can be determined using a histogram, which divides the covered interval into segments, to select the one with the most samples (Fig. 6B). Nevertheless, such method can be sensitive to the selected width of the segments, with longer segments increasing the uncertainty of the mode, and shorter segments potentially creating misleading

spikes. A similar problem is encountered when determining the mode from the sample density in the vicinity of each one, again due to the subjective width of the utilized window.

In order to objectively estimate the mode of the considered reflection strength maxima, the algorithm calculates the *potential function* V_i at each sample within the set. Inspired by the gravitational potential, this function is given by:

$$V_i = \sum_{k=1}^{N_p} f(R_k, R_i), \quad (14)$$

with $N_p > 2$, $i = 1, 2, \dots, N_p$, and

$$f(R_k, R_i) = \begin{cases} \frac{1}{|R_k - R_i|} & \text{if } |R_k - R_i| \geq \Delta R \\ 0 & \text{otherwise} \end{cases}, \quad (15)$$

$$\Delta R = \frac{1}{N_p - 2} \sum_{k=2}^{N_p-1} \frac{R_{k+1} - R_{k-1}}{2}, \quad (16)$$

$$R_k \leq R_{k+1}, \quad (17)$$

where N_p is the number of samples, R_k is the k -th maximum in the set, and ΔR is the average difference in the reflection strength between adjacent samples when arranged in an ascending order (Eq. (17)).

Similarly to gravitational wells, the potential function reaches its peaks around the densest clusters of samples (Fig. 6B), with the minimum distance ΔR (Eq. (16)) between interacting samples reducing the number and intensity of misleading spikes (from small clusters of closely spaced samples) and preventing singularities (from matching samples).

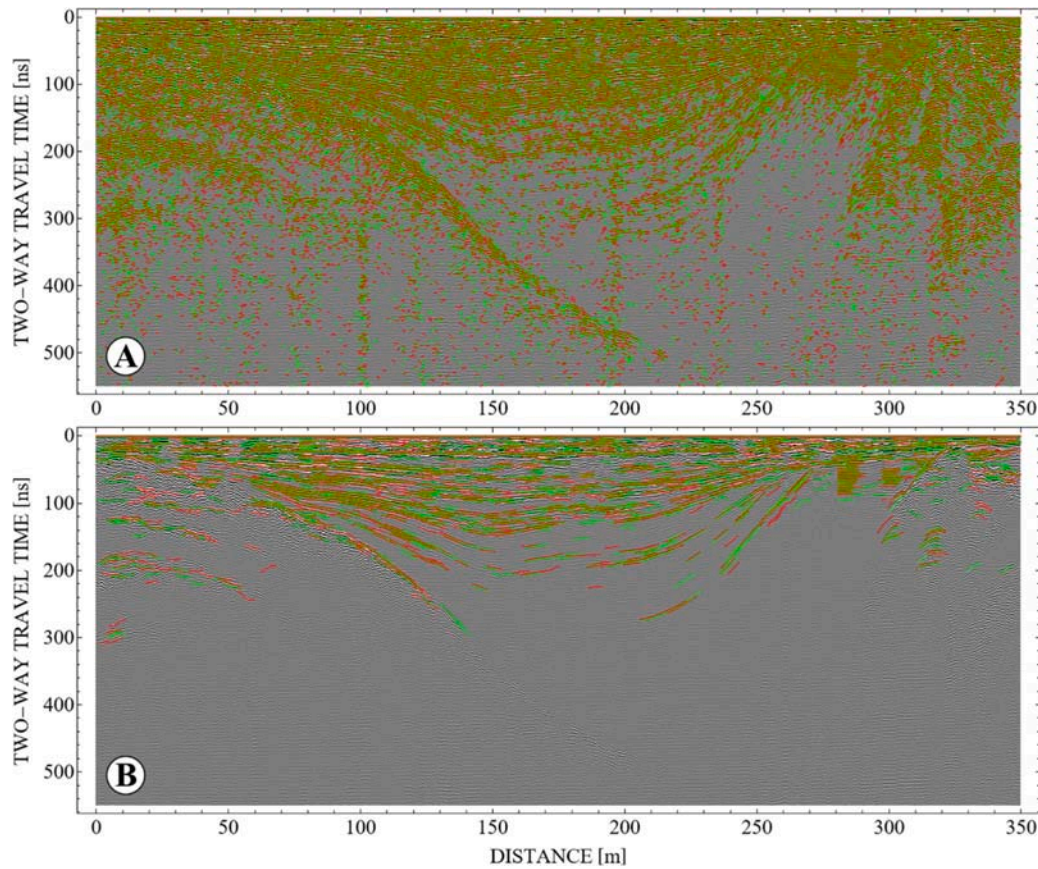


Fig. 9. Reflection phase auto-tracking results superimposed to the signal amplitude profile. The figure shows (A) the automatically constructed horizons marking at least 10 signal phases (i.e., at least 0.9 m long without gaps); (B) the horizons remaining after increasing the minimum length to at least 50 marked signal phases (i.e., at least 4.9 m long without gaps). Positive amplitudes are marked in green, negative amplitudes in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Therefore, if a cluster of samples R_k covers an interval that is shorter than ΔR , Eq. (15) prevents such samples from interacting with each other, while also causing them to act on other samples as a single entity with a cumulative (metaphorical) mass equal to the sum of all the unit masses within the cluster. In any case, after estimating the potential V_i of each sample (Eq. (14)), the algorithm adapts the Rayleigh function (Eq. (13)) to such values, while also accounting for the expected deviations (Fig. 6B), which can include (A) possible minor spikes in the vicinity of the actual mode R_M ; and (B) a horizontal shift Δr of the discrete distribution. The calculation of both the mode and the horizontal shift of the discrete distribution is discussed in detail in Appendix B.

The estimated shift is subtracted from the calculated mode R_M in order to apply the continuous function (which crosses the origin; Eq. (13)), to be then added back after calculating the minimum threshold for the average reflection strength of the analyzed horizon (which is equal to 1 after the previously discussed normalization; Fig. 6C). In particular, after integrating Eq. (13) over the interval $[R_A, \infty]$, the lower margin R_A is given by the area A between the Rayleigh function and the abscissa axis within said interval:

$$R_A = R_M \sqrt{-2 \ln[A]}, \quad (18)$$

with

$$0 < A \leq 1, \quad (19)$$

$$\Delta r = 0. \quad (20)$$

In order for a horizon to be acceptable, the average of its marked reflection strength maxima must be within the upper third of the reconstructed Rayleigh distribution (Fig. 6C). After accounting for the

horizontal shift Δr (Eq. (B.6)) of the distribution in Eq. (18), the lower margin $R_{1/3}$ of such area is given by:

$$R_{1/3} = (R_M - \Delta r) \sqrt{2 \ln[3]} + \Delta r. \quad (21)$$

Therefore, the analyzed horizon is marking an actual reflection if its average reflection strength is at least equal to its estimated minimum threshold $R_{1/3}$ (Fig. 6C-1), otherwise it is disregarded as false positives (Fig. 6C-2).

5. Application to field GPR data

5.1. Data acquisition and processing

The proposed automated reflection strength tracking algorithm is tested on a glaciological GPR profile that was acquired in November 2017 as part of a wider study (i.e., combining boreholes, geophysical, geomorphological, and chemical data) of the Boulder Clay Glacier (Forte et al., 2024; Forte et al., 2025). The survey area is ~5.5 km southwest of the “Mario Zucchelli” Italian Antarctic Station on the Northern Foothills, along the western coast of Terra Nova Bay in Victoria Land, Antarctica (Fig. 7A). The Boulder Clay Glacier is oriented parallel to the coast on a north-south axis, with a total length of ~6 km, a width of ~1.5 km, and a maximum thickness in the 100–150 m range (Kannan et al., 2020). Bordering the perennially frozen Enigma Lake to the north and reaching the Ross Sea at Adélie Cove to the south (Fig. 7B), the cold-based glacier is characterized by a slow eastward movement caused by ice deformation, although it is virtually unable to erode the bedrock underneath (Urbini et al., 2019).

The survey area is partially covered by rocky debris from a large,

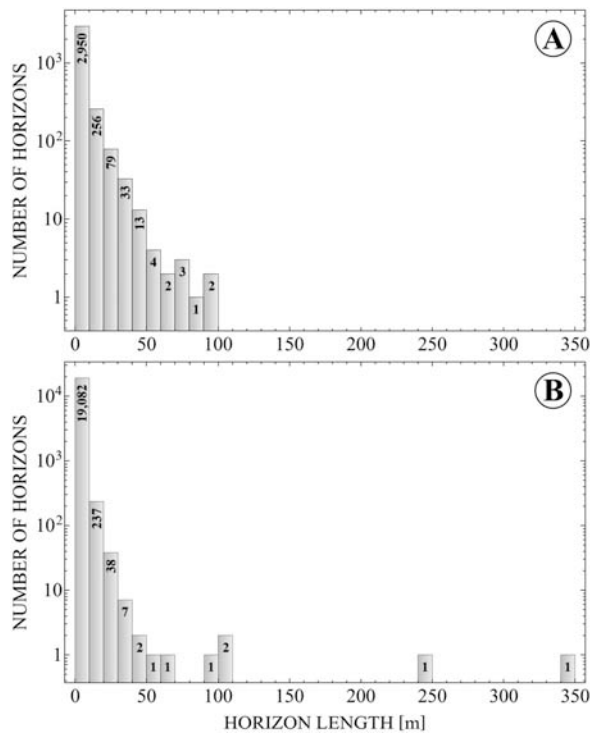


Fig. 10. Comparison of the reflection auto-tracking results based on the reflection strength and the signal phase. The figure shows the histogram of the horizon lengths obtained using (A) the proposed reflection strength auto-tracking algorithm (Fig. 8); and (B) a phase-based auto-tracking algorithm (Fig. 9). The minimum threshold for the horizons in both graphs is set equal to 10 marked traces (i.e., at least 0.9 m long without gaps).

heterogeneous glacial moraine to the east, which occupies more than half the glacier surface (Urbini et al., 2019; Fig. 7B). Transversally crossing the Boulder Clay Glacier from west to east (Fig. 7B), the analyzed 350 m long profile (Fig. 7C) was acquired using a ProEx GPR system equipped with 250 MHz ground-coupled bistatic Malå Geoscience shielded antennas, with a transmitter-receiver offset equal to 31 cm (Forte et al., 2024; Forte et al., 2025). An odometer was used as trigger for the transmitter, resulting in an average trace interval equal to 10 cm. The GPR trace positioning was performed using a Thales Magellan Promark 3 RTK GPS device, with an accuracy of the order of centimeters for the latitude and longitude coordinates and of the order of decimeters for the elevation.

The studied GPR profile is particularly useful in testing the presented auto-tracking procedure given its diverse subsurface stratigraphy (Fig. 7C), which includes the well-layered Boulder Clay Glacier, a steep and more irregular basal interface, other clear reflections within the bedrock, and more complex structures within the moraine. The applied signal processing includes (A) DC removal; (B) drift (a.k.a. time-zero or zero timing) correction; (C) a Butterworth band-pass filter; and (D) a background removal that applies a 201 traces (i.e., 2 m) wide window to signals recorded after 3 ns; while no amplitude recovery is applied to the data set. Indeed, in regards to amplitude decay, the general and gradual decrease of S/N with depth (Fig. 7C) constitutes a further valuable test for the algorithm.

5.2. Results and discussion

The final set of constructed horizons is shown in Fig. 8A, with a minimum threshold for the horizons' length equal to 50 marked energy packets (i.e., at least 4.9 m without gaps), applied for visual clarity. After the significant height analysis, the horizons automatically identified as marking genuine reflections are shown in Fig. 8B, with a minimum

length equal to 10 marked packets (i.e., at least 0.9 m without gaps). Notwithstanding a few cases of interference and noise-related distortions, as well as resolution issues caused by closely spaced interfaces with partly overlapping energy packets, the automatically tracked horizons provide a largely accurate stratigraphic outline of the subsurface. An overall good performance can also be noticed in the more complex sections of the data set, particularly along the ice-bedrock interface, as well as a few deeper structures within both the bedrock and the moraine. As a further indication of the general applicability and versatility of the implemented procedure, the same level of accuracy is found in Dossi et al. (2022), where the reflection auto-tracking algorithm is successfully applied to GPR profiles acquired over well-stratified coastal sand dunes; and in Dossi et al. (2024), where the procedure accurately tracks the basal reflections of a debris-covered glacieret from a pseudo 3-D GPR survey that recorded a particularly challenging ice-bedrock interface.

As expected, a certain number of false positives are found in the deeper sections of the analyzed profile, characterized by a lower S/N, as well as in other areas lacking laterally coherent structures (Fig. 8B). In these areas, the predictably chaotic local reflection dips would produce a more intricate network of energy packets (Fig. 3), and thus cause the algorithm to move more erratically, mostly producing long and flat horizons with quasi-random local deviations (Fig. 8A). For example, the significantly low S/N in the bottom-right section of the profile (i.e., within 220–290 m and 250–550 ns; Fig. 7C) prevents the accurate tracking of the highly dipping interfaces expected both within the glacier and the lower section of the moraine (Fig. 8). More specifically, the local reflection dip analysis (Fig. 3A) is not able to highlight these particularly steep (and considerably weaker) reflections, which are thus lost within the subsequently more intricate local network (Fig. 8A).

Nevertheless, most false positives are removed after the significant height analysis, which produces an overall clearer image of the subsurface stratigraphy (Fig. 8B). It is important to point out, however, that such analysis can be counter-productive when applied to close, parallel, and clear reflections such as the ones in the shallower section of the presented data set (Fig. 8A). In particular, as the purpose of the analysis is to make actual reflection horizons stand out among the weaker false positives in their vicinity, closely spaced reflections would interfere with each other's analyses and result in a certain number of false negatives (Fig. 8B). Notwithstanding these issues, the significant height analysis constitutes a valid starting point for the final stratigraphic interpretation, while any false negative can be manually recovered from the larger set of constructed horizons (Fig. 8A).

In regards to the previously discussed common pitfalls identified by Herron (2015) for auto-picking algorithms, the presented procedure (A) altogether removes the need to define a grid of initial seeds, which instead are automatically and iteratively selected from the constructed network of energy packets; (B) takes into account the directionality of the tracked horizons, whose convergence, divergence, and interferences produce the nodes that are individually analyzed along the network; and (C) shows overall robustness even in areas where the lateral continuity of a reflection is locally disrupted by low S/N. In order to further highlight the advantages of the proposed reflection strength tracking algorithm, when compared to other feature tracking procedures, we apply to the same GPR data set an auto-picking algorithm designed to track any recorded event characterized by lateral phase continuity (Dossi et al., 2015; Forte et al., 2016), with the results shown in Fig. 9. As it can be noticed by comparing Figs. 8A and 9A, one clear advantage of the proposed procedure is the construction of a single horizon for each reflection. In particular, while a correlation tracker would similarly produce a single horizon by matching entire wavelets (Dorn, 1998), a feature tracker like the one applied in Fig. 9A constructs multiple horizons marking the individual phases of the same reflection, with alternating polarities. In the presented case, with a minimum threshold for the horizons' length equal to 10 marked traces (i.e., at least 0.9 m without gaps), the reflection strength auto-tracking algorithm constructed 3343 horizons (Fig. 8), while the phase-based algorithm

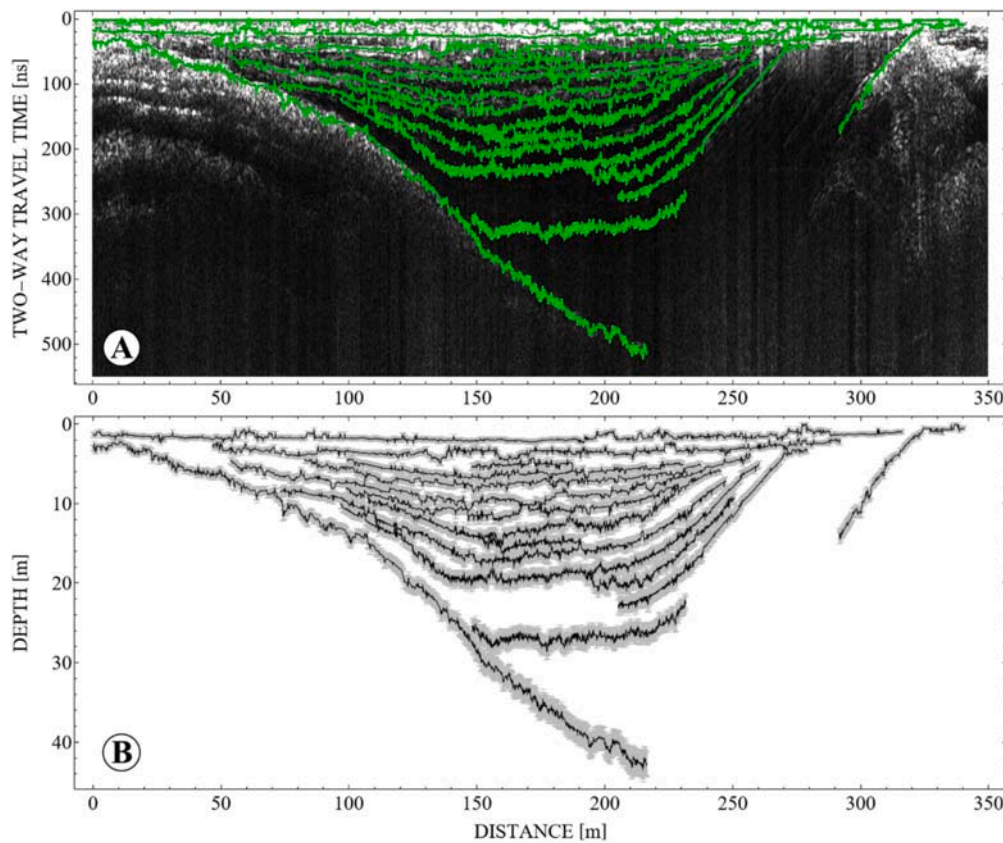


Fig. 11. Reconstruction of the internal stratigraphy of the Boulder Clay Glacier. The figure shows (A) the manually selected interfaces, as inferred from Fig. 8, after the various gaps between and within horizons are filled through an automated interpolation process (up to a specified maximum width); and (B) the resulting subsurface stratigraphy (black lines), with the associated uncertainties (gray bars), after using the EM velocity of pure ice (i.e., 16.8 cm/ns) for the time-to-depth conversion.

constructed 19,373 horizons (Fig. 9).

A more significant feature of the presented algorithm is its ability to accurately track the reflection strength even in areas where the lateral continuity of a reflection is disrupted by either noise or interference-related distortions, which would normally hinder both feature and correlation trackers (Dorn, 1998). For the considered phase-tracking algorithm, this can be observed in the lower half of the GPR profile in Fig. 9A, where the ice-bedrock interface produces a large number of exclusively small horizons, each marking less than 50 signal phases (Fig. 9B). The robustness of the presented procedure (Fig. 8) is evident when comparing the horizon length distributions resulting from the two methods (Fig. 10), which show the energy-based algorithm producing a smaller set of longer horizons compared to the phase-based one. Another notable advantage of the proposed procedure is that it can be applied with minimal signal preprocessing, which is mostly needed in case of significant systemic noise. For example, while band-pass filter and background removal were applied in the presented case study, no amplitude recovery was needed for the analysis. Finally, the presented algorithm requires no pre-training (as opposed to machine learning techniques) and needs only limited input from the interpreter (i.e., a few thresholds), thus producing more objective and easily reproducible results.

Based on the results in Fig. 8, the horizons marking the different reflections are manually selected, and then automatically joined together to form the final interfaces shown in Fig. 11A. Several different post-tracking adjustment methods can be applied to the auto-picked horizons (e.g., manual corrections, interpolations) to counteract possible deviations of genuine reflection horizons into interfering events. Such methods can also be used to fill potential gaps (up to a specified maximum width) between consecutive horizons, as well as

gaps within the horizons themselves. In the presented analysis, the algorithm automatically determines the path to follow across a given gap, potentially including various unmarked and secondary maxima within the gap, as well as possible small horizons in-between. More specifically, the algorithm fills each gap by taking into account the lateral trends of the two consecutive horizons, possible overlaps in their respective trace intervals, as well as the length and average reflection strength along the different possible paths across the gap.

Once the various interfaces are selected (Fig. 11A), the subsurface stratigraphy is reconstructed through a time-to-depth conversion, which requires as input the signal propagation velocity distribution across the data set. In the case of the Boulder Clay Glacier, we use the EM velocity of pure ice (i.e., 16.8 cm/ns) as a reference velocity for the conversion, with an associated uncertainty equal to ± 0.5 cm/ns. In regards to the arrival times of the various reflections, since the selected interfaces are marking their reflection strength maxima rather than their first breaks, a drift correction is applied that subtracts the average arrival time of the shallowest interface (which marks the airwave-groundwave interference; Jol, 2009) from the various interfaces, and then adds the arrival time of the airwave (i.e., 1.03 ns for a 31 cm antenna offset). In theory, the arrival time of the various reflections could be recovered from the upper reflection strength minima of the various marked energy peaks, which would ideally correspond to their first breaks. Compared to the reflection strength maxima, however, these minima are particularly sensitive to both (A) background noise levels, potentially creating long energy tails that move the minima toward much earlier arrival times; and (B) signal interference, where the partly overlapping tails of successive energy peaks lead to highly irregular arrival times for the minima. In any case, the uncertainty associated with the arrival times of the various interfaces is set equal to $\pm 1.5T_R$ (i.e., half the time interval

$3T_R$ of the reference wavelet; Eq. (2)). The reconstructed stratigraphy of the Boulder Clay Glacier along the studied GPR data set is shown in Fig. 11B, reaching a maximum depth of 43.4 ± 1.5 m.

6. Conclusion

We presented a novel auto-picking method that is able to accurately and objectively detect and track reflections within geophysical data sets based on their reflection strength. We detailed the theory and mathematical background of the proposed procedure, thoroughly discussed its advantages and limitations, and tested its robustness by applying it to a glaciological GPR profile that contains both well-stratified sections and more complex structures. The applied algorithm was able to accurately track the vast majority of the recorded reflections, which were then used to reconstruct the subsurface stratigraphy, while also automatically removing most of the expected noise-related false positives. The versatility of the procedure is highlighted by the minimal signal preprocessing requirements, the absence of pre-training (as opposed to machine learning techniques), and the limited input from the interpreter needed for the analysis, which results in a more objective stratigraphic model of the subsurface.

Similarly to other auto-picking techniques, the presented algorithm can be sensitive to the presence of significant noise and interference, which can either cause horizons marking actual reflections to deviate into interfering events or produce a certain number of erratic false positives in areas that lack any laterally coherent structure. In order to greatly improve the visualization of the automatically tracked subsurface stratigraphy, the algorithm filters out these false positives by comparing their signal strength to their surroundings, from which they are less likely to stand out as opposed to actual reflection horizons. However, a certain number of recoverable false negatives can be expected in well-stratified areas with close, parallel, and clear reflections, for which such analysis would effectively be pointless.

In any case, the automated reflection strength tracking algorithm was able to recover the internal stratigraphy of the surveyed Antarctic glacier in a sound and objective way. Most importantly, the accurate and time-saving procedure ultimately allows the interpreter to focus on, and manually intervene in, the more challenging areas of the studied data set, where any auto-picking algorithm can be expected to fail. Future research will apply the presented algorithm to both seismic and GPR data acquired in vastly different contexts, to further demonstrate the general applicability, robustness, and versatility of the procedure

beyond the case study presented in this paper. The studied data sets vary greatly in terms of the surveyed environment, the quality of the data, and the scientific field of interest; and among these are marine reflection seismics data, surface rover GPR data from the Moon, and orbital GPR data from Mars.

CRediT authorship contribution statement

Matteo Dossi: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Emanuele Forte:** Writing – review & editing, Investigation, Funding acquisition, Data curation. **Barbara Cosciotti:** Writing – review & editing. **Sebastian Emanuel Lauro:** Writing – review & editing. **Elisabetta Mattei:** Writing – review & editing. **Michele Pipan:** Writing – review & editing, Supervision. **Elena Pettinelli:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The studied GPR data set was collected during the XXXIII Italian Antarctic Expedition in 2017, as part of the National Antarctic Research Program (PNRA), research project PNRA 2018_00186-E. We wish to thank professor Mauro Guglielmin, from the University of Insubria, for assisting in the field operations. The presented data set is stored in the open-access repository Zenodo, and is available at <https://zenodo.org/records/6090268>.

We also gratefully acknowledge the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) for allowing the publication of their satellite images under the terms of the Image Use Policy of NASA's Earth Observatory online catalogue; and professor David Pape, from the State University of New York at Buffalo, for releasing his satellite image into the Public Domain. The utilized location maps are available at https://commons.wikimedia.org/wiki/File:Antarctica_6400px_from_Blue_Marble.jpg, <https://science.nasa.gov/earth/earth-observatory/curious-clouds-in-the-transantarctic-mountains-147772/>, and <https://science.nasa.gov/earth/earth-observatory/katabatic-winds-rake-terra-nova-bay-41161/>.

Appendix A. Removing barriers during the closed circuit analysis

During the closed circuit analysis, a rightward barrier is encountered when backward-branching horizons from the upper and lower sets join together (Fig. 4C), prompting the algorithm to sever the point of junction and restart the analysis. In the case of multiple junctions, these are iteratively severed starting from the furthestmost (i.e., rightmost) one and then moving backward (i.e., leftward), until an opening is created (Fig. 4D). More specifically, the algorithm identifies all the rightward-branching nodes within the barrier, up to the first two points of connection with the upper and lower sets, and then iteratively cuts the one closest (from the left side) to the leftmost connection (Fig. 4C). In each case, the horizons branching out of the node on the right side are severed, with the exception of either (A) the one for which the node is internal (i.e., the horizon continues leftward); or (B) the one with the highest average reflection strength, if all considered horizons start at the node.

In the case of a leftward barrier, these cause either the upper or lower set to have no more branches propagating rightward as the algorithm progresses. Therefore, among the analyzed rightward-branching energy packets that have such interruptions, the algorithm isolates those pairs in which the upper set of one and the lower set of the other end with the same packet (Fig. 4E). In particular, each pair constitutes the possible points of connection between a potential leftward barrier and the upper and lower sets of a larger circuit. In other words, if the algorithm was moving from right to left within said circuit, the same barrier would be encountered similarly to the previous case, formed by backward (i.e., rightward) branching horizons originating from these points (Fig. 4E).

In order to identify the left edge of the supposed larger circuit, the algorithm uses the already constructed circuits on the left side of the potential barrier to form independent chains. The initial circuit of each chain marks either the upper or lower packet in the considered pair with its upper or lower horizon, respectively, although not at its edges. Following the marking horizon leftward in search for a rightward-branching node, a new circuit is added to the chain if it marks such node the same way the initial circuit marks the initial packet (i.e., both with either their upper or lower horizon, but not at their edges), after which the analysis is repeated on the new marking horizon. As a result, these marking horizons are potentially delineating one of the horizons constituting the aforementioned larger circuit. In particular, if such larger circuit exists, then at least two chains from the two

initial packets will converge toward a circuit (or alternatively to a pair of circuits) that shares the left edge with the larger circuit.

Once this left edge is found, the leftward barrier is cut in a way that mirrors the rightward barrier case. More specifically, among the leftward-branching nodes within the barrier, the algorithm iteratively severs the one closest (from the right side) to the rightmost point of connection (i.e., among the initial pair; Fig. 4E). The process is repeated until it clears the path from the identified left edge, which is then analyzed once again to construct the supposed larger circuit (Fig. 4F). The closed circuit analysis is also reapplied to the left edges of each of the circuits that have been cut during the removal of the leftward barrier.

Appendix B. Adapting the Rayleigh function to the potential distribution

When adapting the Rayleigh function (Eq. (13)) to the potential distribution (Eq. (14)), the mode R_M is given by an average of the reflection strength values R_i , weighted by their respective potentials V_i ; and the corresponding maximum V_M is a simple average of such potentials. However, the two averages are limited to the values V_i that are at least equal to a minimum percentage (e.g., 75%) of the highest potential V_P within the set (Fig. 6B):

$$R_M = \frac{\sum_{i=1}^{N_p} f(V_i, R_i)}{\sum_{i=1}^{N_p} g(V_i)}, \quad (\text{B.1})$$

$$V_M = \frac{\sum_{i=1}^{N_p} g(V_i)}{\sum_{i=1}^{N_p} n(V_i)}, \quad (\text{B.2})$$

with

$$f(V_i, R_i) = \begin{cases} V_i R_i & \text{if } V_i \geq 0.75 V_P \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B.3})$$

$$g(V_i) = \begin{cases} V_i & \text{if } V_i \geq 0.75 V_P \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B.4})$$

$$n(V_i) = \begin{cases} 1 & \text{if } V_i \geq 0.75 V_P \\ 0 & \text{otherwise} \end{cases}. \quad (\text{B.5})$$

The horizontal shift Δr of the discrete distribution V_i with respect to the origin is subsequently estimated as the smaller root of an auxiliary quadratic function, whose vertex coincides with (R_M, V_M) . This function approximates the left side of the Rayleigh distribution (Eq. (13)) by crossing a second point (R_Q, V_Q) calculated from specific samples (R_i, V_i) on the left side of the set (Fig. 6B), with the resulting shift given by:

$$\Delta r = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad (\text{B.6})$$

with

$$f(r) = ar^2 + br + c, \quad (\text{B.7})$$

$$a = \frac{V_Q - V_M}{(R_Q - R_M)^2}, \quad (\text{B.8})$$

$$b = \frac{2R_M(V_M - V_Q)}{(R_M - R_Q)^2}, \quad (\text{B.9})$$

$$c = V_M + \frac{R_M^2(V_Q - V_M)}{(R_Q - R_M)^2}, \quad (\text{B.10})$$

where the square root in Eq. (B.6) is set positive due to the negative value of a (Eq. (B.8); Fig. 6B).

The second point (R_Q, V_Q) is calculated similarly to the vertex (R_M, V_M) ; however, the averages are limited to the central potential values V_i (e.g., 25–75% of the highest potential V_P) of the reflection strength samples R_i below the lowest value R_L used in the calculation of the mode (i.e., the leftmost point (R_i, V_i) whose potential V_i is at least 75% of V_P ; Fig. 6B). In addition, the weights used to calculate R_Q are relative to highest potential V_P , so that they are larger for the samples closer to the origin:

$$R_Q = \frac{\sum_{i=1}^L f(V_i, R_i)}{\sum_{i=1}^L g(V_i)}, \quad (\text{B.11})$$

$$V_Q = \frac{\sum_{i=1}^L h(V_i)}{\sum_{i=1}^L n(V_i)}, \quad (\text{B.12})$$

with

$$f(V_i, R_i) = \begin{cases} (V_P - V_i)R_i & \text{if } 0.25V_P \leq V_i \leq 0.75V_P \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B.13})$$

$$g(V_i) = \begin{cases} V_P - V_i & \text{if } 0.25V_P \leq V_i \leq 0.75V_P \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B.14})$$

$$h(V_i) = \begin{cases} V_i & \text{if } 0.25V_P \leq V_i \leq 0.75V_P \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B.15})$$

$$n(V_i) = \begin{cases} 1 & \text{if } 0.25V_P \leq V_i \leq 0.75V_P \\ 0 & \text{otherwise} \end{cases}. \quad (\text{B.16})$$

Data availability

Data will be made available on request.

References

- Alves, C.D.S., Silva, B.S., Barbosa, C.H.S., Rochinha, F.A., Coutinho, A.L.G.A., Filho, D.M. S., 2023. Horizon extraction using absolute seismic phase obtained by unwrapping techniques. *J. Appl. Geophys.* 218, 1–17. <https://doi.org/10.1016/j.jappgeo.2023.105198>.
- Ball, A., O'Connor, L., 2021. Geologist in the loop: a hybrid intelligence model for identifying geological boundaries from augmented Ground Penetrating Radar. *Geosciences* 11 (7), 1–15. <https://doi.org/10.3390/geosciences11070284>.
- Cantrell, C.A., 2008. Technical note: review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems. *Atmos. Chem. Phys.* 8 (17), 5477–5487. <https://doi.org/10.5194/acp-8-5477-2008>.
- Deng, X., Hu, X., 2019. A new seismic horizon picking method based on the edge detection and the Kalman filter. In: *Proc. IEEE 12th Int. Cong. Image Signal Process. Biomed. Eng. Inform.*, Suzhou, China, October 19–21, 2019, pp. 1–4. <https://doi.org/10.1109/CISP-BMEI48845.2019.8965858>.
- Deng, X., Zhang, Z., 2017. Automatic multihorizons recognition for seismic data based on Kalman filter tracker. *IEEE Geosci. Remote Sens. Lett.* 14 (3), 319–323. <https://doi.org/10.1109/LGRS.2016.2637006>.
- Di, H., Li, Z., Maniar, H., Abubakar, A., 2020. Seismic stratigraphy interpretation by deep convolutional neural networks: a supervised workflow. *Geophysics* 85 (4), WA77–WA86. <https://doi.org/10.1190/geo2019-0433.1>.
- Dorn, G.A., 1998. Modern 3-D seismic interpretation. *Lead. Edge* 17 (9), 1262–1269. <https://doi.org/10.1190/1.1438121>.
- Dossi, M., Forte, E., Pipan, M., 2015. Automated reflection picking and polarity assessment through attribute analysis: theory and application to synthetic and real ground-penetrating radar data. *Geophysics* 80 (5), H23–H35. <https://doi.org/10.1190/GEO2015-0098.1>.
- Dossi, M., Forte, E., Pipan, M., Colucci, R.R., 2018. Quantitative 3-D GPR analysis to estimate the total volume and water content of a glacier. In: *Proc. IEEE 17th Int. Conf. GPR, Rapperswil, Switzerland*, June 18–21, 2018, pp. 1–6. <https://doi.org/10.1109/ICGPR.2018.8441636>.
- Dossi, M., Forte, E., Cosciotti, B., Lauro, S.E., Mattei, E., Pettinelli, E., Pipan, M., 2022. Automated reflection strength tracking for improved stratigraphic interpretation of GPR data sets. In: *Proc. SEG 19th Int. Conf. GPR*, Golden, Colorado, USA, June 12–17, 2022, pp. 123–126. <https://doi.org/10.1190/gpr2022-016.1>.
- Dossi, M., Mattei, E., Pecci, M., Cosciotti, B., Lauro, S.E., Monaco, A., Scozzafava, M., Pettinelli, E., 2024. The last remnants of the Italian Apennine glaciers: 3-D modeling of the Lower Calderone Glacieret by the IDW spatial interpolation of GPR data. *J. Glaciol.* 70 (e32), 1–13. <https://doi.org/10.1017/jog.2024.36>.
- Farakliti, M., Petrou, M., 2004. Horizon picking in 3D seismic data volumes. *Mach. Vis. Appl.* 15, 216–219. <https://doi.org/10.1007/s00138-004-0151-8>.
- 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. *Geophys. J. Int.* 197 (3), 1471–1483. <https://doi.org/10.1093/gji/ggu103>.
- Forte, E., Dossi, M., Pipan, M., Del Ben, A., 2016. Automated phase attribute-based picking applied to reflection seismics. *Geophysics* 81 (2), V141–V150. <https://doi.org/10.1190/geo2015-0333.1>.
- Forte, E., Santin, I., Azzaro, M., Battistel, D., Guglielmin, M., 2024. Evidence of brines interconnections and different flow patterns within the Boulder Clay Glacier and its moraine (Victoria Land, East Antarctica). *Sci. Total Environ.* 957, 1–10. <https://doi.org/10.1016/j.scitotenv.2024.177469>.
- Forte, E., Azzaro, M., Cannone, N., Longhi, A., Santin, I., Guglielmin, M., 2025. A warming pulse in the Antarctic continent changed the landscape during the Middle Ages. *Commun. Earth Environ.* 6 (281), 1–9. <https://doi.org/10.1038/s43247-025-02259-4>.
- Harrigan, E., Kroh, J.R., Sandham, W.A., Durrani, T.S., 1992. Seismic horizon picking using an artificial neural network. In: *IEEE Int. Conf. Acoust. Speech Signal Process.*, San Francisco, California, USA, March 23–26, 1992, pp. 105–108. <https://doi.org/10.1109/ICASSP.1992.226265>.
- Herron, D.A., 2000. Horizon autopicking. *Lead. Edge* 19 (5), 491–492. <https://doi.org/10.1190/1.1438636>.
- Herron, D.A., 2015. Pitfalls in horizon autopicking. *Interpretation* 3 (1), SB1–SB4. <https://doi.org/10.1190/INT-2014-0062.1>.
- Hoetz, H.L.J.G., Watters, D.G., 1992. Seismic horizon attribute mapping for the Annerven Gasfield, the Netherlands. *First Break* 10 (2), 41–51. <https://doi.org/10.3997/1365-2397.1992003>.
- Jol, H.M., 2009. *Ground Penetrating Radar: Theory and Applications*. Elsevier Sci. Ltd.
- Kannan, K., Mansutti, D., Rajagopal, K.R., Urbini, S., 2020. Mathematical modeling of rock glacier flow with temperature effect. In: Cannarsa, P., Mansutti, D., Provenzale, A. (Eds.), *Mathematical Approach to Climate Change and Its Impacts*, 38. Springer INdAM Series, pp. 149–161. https://doi.org/10.1007/978-3-030-38669-6_4.
- Lou, Y., Zhang, B., Lin, T., Cao, D., 2020. Seismic horizon picking by integrating reflector dip and instantaneous phase attributes. *Geophysics* 85 (2), O37–O45. <https://doi.org/10.1190/GEO2018-0303.1>.
- Maroni, C.S., Quinquis, A., Vinson, S., 2001. Horizon picking on subbottom profiles using multiresolution analysis. *Digit. Signal Process.* 11 (4), 269–287. <https://doi.org/10.1006/dspr.2001.0396>.
- Moqadam, H., Eisen, O., 2025. Review article: Feature tracing in radio-echo sounding products of terrestrial ice sheets and planetary bodies. *Cryosphere* 19 (6), 2159–2196. <https://doi.org/10.5194/tc-19-2159-2025>.
- Muraleedharan, G., Rao, A.D., Kurup, P.G., Nair, N.U., Sinha, M., 2007. Modified Weibull distribution for maximum and significant wave height simulation and prediction. *Coast. Eng.* 54 (8), 630–638. <https://doi.org/10.1016/j.coastaleng.2007.05.001>.
- Paulson, K.V., Merdler, S.C., 1968. Automatic seismic reflection picking. *Geophysics* 33 (3), 431–440. <https://doi.org/10.1190/1.1439941>.
- Pitas, I., Kotropoulos, C., 1989. Texture analysis and segmentation of seismic images. In: *IEEE Int. Conf. Acoust. Speech Signal Process.*, Glasgow, Scotland, UK, May 23–26, 1989, pp. 1437–1440. <https://doi.org/10.1109/ICASSP.1989.266709>.
- Pitas, I., Kotropoulos, C., 1992. A texture-based approach to the segmentation of seismic images. *Pattern Recogn.* 25 (9), 929–945. [https://doi.org/10.1016/0031-3203\(92\)90059-R](https://doi.org/10.1016/0031-3203(92)90059-R).
- Roberto, V., Gargiulo, L., Peron, A., Chiaruttini, C., 1990. A knowledge-based system for geophysical interpretation. In: *IEEE Int. Conf. Acoust. Speech Signal Process.*, Albuquerque, New Mexico, USA, April 3–6, 1990, pp. 1945–1948. <https://doi.org/10.1109/ICASSP.1990.115884>.
- Roncoroni, G., Forte, E., Bortolussi, L., Pipan, M., 2022. Efficient extraction of seismic reflection with deep learning. *Comput. Geosci.* 166, 1–9. <https://doi.org/10.1016/j.cageo.2022.105190>.
- Stansell, P., 2004. Distributions of freak wave heights measured in the North Sea. *Appl. Ocean Res.* 26 (1–2), 35–48. <https://doi.org/10.1016/j.apor.2004.01.004>.
- Taner, M.T., Koehler, F., Sheriff, R.E., 1979. Complex seismic trace analysis. *Geophysics* 44 (6), 1041–1063. <https://doi.org/10.1190/1.1440994>.
- Urbini, S., Bianchi-Fasani, G., Mazzanti, P., Rocca, A., Vittuari, L., Zanutta, A., Girelli, V. A., Serafini, M., Zirizzotti, A., Frezzotti, M., 2019. Multi-temporal investigation of the Boulder Clay Glacier and Northern Foothills (Victoria Land, Antarctica) by integrated surveying techniques. *Remote Sens.* 11 (12), 1–23. <https://doi.org/10.3390/rs11121501>.

- Wadsworth, G.P., Robinson, E.A., Bryan, J.G., Hurley, P.M., 1953. Detection of reflections on seismic records by linear operators. *Geophysics* 18 (3), 539–586. <https://doi.org/10.1190/1.1437911>.
- Yang, L., Sun, S.Z., 2020. Seismic horizon tracking using a deep convolutional neural network. *J. Petrol. Sci. Eng.* 187, 1–15. <https://doi.org/10.1016/j.petrol.2019.106709>.
- Zhao, J., Li, S., Zhao, X., Feng, J., 2020. A comprehensive horizon picking method on subbottom profiles by combining envelope, phase attributes, and texture analysis. *Earth Space Sci.* 7 (3), 1–20. <https://doi.org/10.1029/2019EA000680>.