The methodology of choosing the distribution model of echoes for ground penetrating radar with broadband signals

S V Borisov and Yu K Vyboldin
Saint-Petersburg Mining University, 2, 21 line of Vasilievsky Island, Saint-Petersburg, 199106, Russia
E-mail: box26@mail.ru

Abstract. The paper presents an approach to the solution of the problem of assessing the probabilistic characteristics of echo signals of ground penetration radar (GPR). This allows performing the synthesis of decisive statistical rules in the problems of detection and recognition of subsurface objects.

1. Introduction
One of the drawbacks of studying subsurface spaces using GPR is inaccurate a priori information on the composition and characteristics of soils and other elements that make up this space, which leads to distortion of sounding results or a significant decrease in volume of explored space. To overcome a priori uncertainty, laboratory studies of soil samples are carried out, as well as indirect, including remote, measurements of its characteristics, which are summarized in tables or analytical dependencies. In the future, this information will be used in analytical calculations or in programs for constructing and decoding GPR’s range profiles of the underground space [1-6].

One way to overcome a priori uncertainty is to use statistical models of reflected GPR signals to study the subsurface space, which is developed on the basis of an analysis of the results of the primary processing of information from underground radar stations. The quality of decoding GPR data depends on the available a priori data on the distribution of moisture, salinity, the dielectric constant of the soil, the nature and composition of inclusions. These data may also be obtained by laboratory studies of samples.

2. Results and discussions
One of the important tasks solved by GPR methods is the search, determination of the position and shape of subsurface objects and heterogeneities, including an assessment of their depth and linear size. The development of optimal and quasi-optimal methods for detecting and recognizing subsurface objects when processing the results of primary sensing is possible using distribution models characterizing the propagation paths of electromagnetic waves in underground space.

The experimental method for obtaining probabilistic distribution models consists in coordinated processing of reflected pulsed broadband signals from GPR, digitizing the obtained realizations, and selecting the type and parameters of the probabilistic distribution characteristic of particular sounding conditions and studied rocks. In this case, statistics are accumulated for each studied point on the surface and for several points within a radius of 3-5 m from it [7-10].
The reflected GPR signals are represented as a set of random amplitude-phase variations of reflected pulses from flat facies-lithological boundaries, as well as diffuse-diffracted pulses. They are the results of scattering and reflection of electromagnetic waves from structural and petrophysical inhomogeneities of rock masses and soils. The stochastic nature of the variability of these signals cannot be seen on GPR’s echoes profiles. As a result, the reflected signals in GPR are considered deterministic. At the same time, when analyzing the results of repeated soundings, the stochasticity of the GPR signals becomes obvious. This can no longer be neglected when solving petrophysical problems related to the assessment of properties and the state of the subsurface medium. Thus, if the adopted interpretative decisions are made using one-time records of GPR signals, then they will be unstable and the results of studying the properties and state of the subsurface medium will not satisfy the reliability criterion.

The propagation medium of subsurface radar signals are soils. Soils are very complex natural multiphase formations (especially frozen and permafrost), consisting of various components about their properties. These components are in different phase states (solid, perfectly plastic, liquid, gaseous) and are interconnected. The medium can be considered as a single-component continuous only under certain conditions - for example, in a given volume of frozen soil there is no redistribution of individual phases of the soil in time. The contrast of the dielectric constant in the layers determines the reflectivity of the boundaries and the ability of buried objects to form diffracted waves. As the materials of numerous laboratory and field experiments show in the meter wavelength range the real part of the complex dielectric constant, and, consequently, the propagation velocity of electromagnetic waves weakly depend on the frequency and mineral part of the soil, but very much depend on their moisture content.

The tasks of subsurface GPR sounding and remote sensing are interrelated. In particular, the results of studies of the dielectric and radiative properties of the soil can be used as a priori information for processing contact field data of GPR. Remote determination of the hydrophysical parameters of the soils is based on the dependence of radio-emitting and dispersing characteristics of soils in the microwave range on the amount of moisture contained in it. In this case, for soils of different composition at the same moisture content, noticeable variations in the emissivity are observed, depending on the granulometric composition of the soil and, therefore, on the content of bound water. Due to the wide variety of soil types, it is necessary to take into account the regional and even local nature of soil properties. This can be done in two ways. Firstly, interpolation dependences of electrophysical parameters on humidity for each specific territory can be constructed. The second way is to create basic models that would basically contain a certain limited number of parameters that take into account the specifics of the soil. Variations of these parameters provide the opportunity for flexible adaptation of basic models to this particular type of soil. Often the first version of the translation ratios is used in the form of interpolation dependencies [11-14,17].

A related task is remote sensing of frozen soils. In the middle geographic latitudes, in winter, a seasonally frozen soil layer is formed over a large area in which the state of moisture and the nature of its interaction with radio emission substantially change. An applied task in this case can be the remote evaluation of particle size distribution, freezing depth, soil salinity. For the remote determination of soil salinity, it is proposed to use the relationship between the radio brightness temperature of the soil and the salinity established experimentally; there are also methods for remote determination of the thermodynamic temperature and depth of freezing of the soil cover. All this stimulates the conduct of further studies of the moisture and temperature behavior of the dielectric and radio-emission characteristics of soil of various compositions.

Thus, the results of remote sensing can be used to clarify the characteristics of the soil when performing radar subsurface sensing in the analysis of the results already during the primary processing of the received signals. Important is the ability to use remote sensing information about soil properties in real time.

A powerful tool for studying the properties of a subsurface medium is a physical experiment in a laboratory environment, which includes measuring the complex permittivity of soils with different
particle sizes and with varying humidity, salinity, and temperature. Naturally, all these characteristics of the subsurface medium affect the statistical characteristics of the reflected signal. Statistics of received signals are to study carefully to obtain a priori data for the synthesis of algorithms for processing primary GPR information using probabilistic models for various types of soil.

By measuring the statistical characteristics of the reflected GPR signals, it is possible to successfully investigate the dielectric properties of materials, namely, the relative permittivity, which mainly depends on humidity and electrical conductivity. In [15], a GPR was used to assess the quality of concrete. Concrete blocks were tested in which cement, sand, and gravel were mixed in different proportions. Measurements were made for cases of wet and dry concrete. The speed of the signal, the amplitude of the reflected signal, the other properties of the direct signal, which passes along the shortest path from the transmitter to the receiver through the air and material, were measured. The GPR signal speed and other characteristics of the direct waves closely correlated with the water content in concrete. An analysis of the data also shows that the amplitude from peak to peak of the direct signal is sensitive to porosity. These results offer new opportunities for assessing the quality of materials using GPR in situ.

Also in [16], by using the statistical characteristics of the reflected radar signal, the characteristics of the aggregate distribution, which affect the mechanical strength properties of a dense homogeneous medium - concrete, were investigated. A physical model of the distribution of impurities in concrete was developed, and the corresponding statistical characteristics of the propagation of the electromagnetic wave were analyzed. A relationship was found between the GPR echo and the impurity distribution. The adequacy of the proposed method for estimating the distribution of impurities according to the statistical characteristics of the reflected signal was verified using direct radar modeling and a field experiment. The result showed that the method can automatically detect a coarse aggregate distribution and is suitable for detecting aggregate distribution in concrete.

Since the measurement is a procedure for finding amount empirically by means of special technical means, implementing an algorithm including a comparison operation with a known quantity, in static measurements a measure should be applied, reproducing a known quantity. As the algorithm for the analyzer of probabilistic characteristics, the typical procedures for measuring the random processes of probabilistic characteristics can be matched. They may be presented in the following form

\[ \theta^* [X(t)] = S_d \cdot g (X(t)) \]  \hspace{1cm} (1)

\[ \theta^* [X(t)] = S_d \cdot K \cdot g (X(t)) \]  \hspace{1cm} (2)

\[ \theta^* [X(t)] = S_d \cdot g (X(t)) \]  \hspace{1cm} (3)

where \( S_d \) - averaging operator; \( K \) - reference measure comparison operator; \( g \) – expression to define a probabilistic characteristic \( \theta \); \( \theta^* [X(t)] \) – characteristic’s \( X(t) \) measurement result.

The averaging parameter \( d \) determines the principle of averaging over time (\( d = T \)) or over the number of realizations (\( d = N \)). These algorithms differ in the order of operations. The comparison operation with the exemplary measure (\( K \)) can be final - equation (3), performed after the implementation of the operator \( g \), but before averaging - equation (2) and, finally, be the initial one - equation (1).

In these figures, to denote the blocks that implement the operators, entering into expressions, the same notation is used. So, \( g \) is the device that performs the transformation underlying the determination of the probabilistic characteristic \( \theta \); \( S_d \) - averaging device (adder or integrator); \( K \) is the comparator (comparison operator or comparison device), and \( M \) is the measure by which a known quantity \( (\theta_0, g_0, x_0) \) is formed.
Figure 1. Generalized block diagrams of tools for probability characteristics measurements

Presented in figure 1 (a), the measuring instrument implements the following procedure: a set of implementations \( \{x_i(t)\} \) enters the input (when using time averaging - one implementation of \( x_i(t) \) - at the output of node \( g \) we have a set of transformed implementations \( \{g\{x_i(t)\}\} \); after averaging, we obtain the quantity \( S_d[\{g\{x_i(t)\}\}] \), which goes to the comparator, which compares with the known value \( \theta_0 \). as a result of which we obtain the value of the measured probabilistic characteristics \( \theta^*X_t \).

The difference between the procedure implemented by the measuring tool presented in figure 1 (a) and (b) is that after forming the population \( \{g\{x_i(t)\}\} \) it does not go to the averager, but to the comparator, which performs a comparison with the known value \( g_0 \). Thus, a numerical array \( \{g^*\{x_i(t)\}\} \) is formed at the output of the comparator, and averaging is performed in numerical form. At the output of the averager \( S_d \), we have the measurement result \( \theta^*[X(t)] \).

The measuring tool (figure 1, c) is based on the formation of an array of numerical equivalents of instantaneous realizations of random process \( X(t) \), after which the transformation \( g \) and averaging are performed in numerical form. This device is equivalent to a serial connection of the ADC and the computing device (processor). At the exit the ADC forms an array of instantaneous values, and the processor, according to a certain program, provides the implementation of the \( g \) and \( S_d \) operators.

As a result of processing the GPR data statistical estimates of the probability indicators of the distributions based on their histograms and then the approximation of the integral distribution laws can be obtained. The basis for choosing two- and three-parameter probabilistic models can be obtained by two- and three parametric sample characteristics of the reflected subsurface signal. One way to present information about distributions is to use quantile estimates of distributions. Consider the possibilities
of selecting distributions by their quantiles using the Pearson family of distributions. Each family of distributions in the Pearson diagram can be obtained as a solution of the differential equation

$$\frac{df}{dx} = \frac{(x-\mu) f(x)}{\phi_0 + \phi_1 x + \phi_2 x^2} \quad (4)$$

for a random variable $x$ with a distribution density $f(x)$ by appropriate selection of the parameters $\theta_0$, $\theta_1$ and $\theta_2$. The solution of this equation leads to a large number of distribution families, including normal, beta distribution and gamma distribution. The Pearson diagram for distributions in the $(\beta_1, \beta_2)$ plane shows points, lines and regions for different distributions - normal, beta distribution (special case - uniform distribution), gamma distribution (special case - exponential distribution) and log-normal.

The coordinates of the diagram are selected as follows: $\beta_1$ is the square of the skewness, $\beta_2$ is the kurtosis+3. In particular, for the normal distribution, $\beta_1 = 0$ and $\beta_2 = 3$. In the diagram, this distribution is represented by a single point, as well as the exponential and uniform distribution. The gamma distribution, the lognormal distribution, and the t distribution can be selected for all values of $\beta_1$ and $\beta_2$ lying near the average curve. Note that the curve for the gamma distribution is near the curve for the log-normal distribution. This helps to explain the fact that empirical data can be described by both a gamma distribution and a log-normal distribution. The beta distribution, which has two form parameters, occupies a certain area, hence its generalizing character follows. However, a large range of values of $\beta_1$ and $\beta_2$ is not covered by more than one of the distributions considered earlier. The most common distribution for approximating experimental data is the use of a normal distribution. It follows from the central limit theorem that the normal distribution gives an acceptable description of many real phenomena. Similarly, the gamma distribution and the lognormal distribution are used to describe random variables bounded on one side. Although these models lead to distributions of the most various forms, nevertheless they do not give the degree of generalization that is necessary.

To use the graphs shown in figure 2, it is necessary to know the estimated values of $\beta_1$ and $\beta_2$, corresponding to the accepted experimental implementations. These estimates are used to describe the experimental data by one of the presented distributions by calculating sample estimates $\beta_1$ and $\beta_2$ and drawing points with the corresponding coordinates on the distribution diagram. If the locations of the experimental points are located close enough to the point, curve, or region corresponding to one of the types of distribution, it follows that this distribution can be used to describe empirical data. Then they proceed to the adjusted choice of distribution parameters. The graphs on the diagram indicate a wide variety of forms of Pearson distributions and, thus, can be directly used to select the appropriate
approximating distribution. In some cases, to select the approximating Pearson distribution, it suffices to use the estimates of the first four moments of the histograms of the experimental samples.

3. Conclusion

The development of optimal methods for the detection and recognition of subsurface objects in the processing of primary sensing results is possible only when using distribution models characterizing the propagation paths of electromagnetic waves in the underground space. A method of experimental obtaining probabilistic distribution models is coordinated processing of reflected pulsed broadband GPR signals, digitizing them, and selecting the type and parameters of the probabilistic distribution characteristic of particular sounding conditions and studied rocks.

The probabilistic characteristics of the reflected signals of the subsurface location are satisfactorily approximated by non-Gaussian distributions, which can serve as the basis for the development of algorithms for the coordinated processing of the GPR reflected signals when detecting or recognizing subsurface objects for given conditions.

References

[1] Danilev S M and Danileva N A 2018 Engineering and Mining Geophysics, 1 pp. 1-4
[2] Limanskiy D V and Kuskildin R B 2019 Journal of Physics Conference Series 1384 012026 DOI: 10.1088/1742-6596/1384/1/012026
[3] Shevniin V A, Kvon D A, Ryzhov A A 2017 Journal of Mining Institute 226 pp. 397-404 DOI: 10.25515/pmi.2017.4.397
[4] Lalomov D A and Glazunov V V 2018 Journal of Mining Institute 229 pp. 3-12 DOI: 10.25515/pmi.2018.1.3
[5] Rudianov G V, Krapivskii E I and Danilev S M 2018 Journal of Mining Institute 231 pp. 245 DOI: 10.25515/pmi.2018.3.245
[6] Vyboldin Yu K and Borisov S V 2018 IOP Conf. Ser.: Earth Environ. Sci. 194 062036
[7] Neradovsky L G 2009 Methodical Guide to the Study of Permafrost by Dynamic Georadiolocation Method (Moscow: Russian Academy of Science) p 337
[8] Ermakov A P and Starovoitov A V 2010 Moscow University Geology Bulletin 65 pp 422–7
[9] Neradovsky L G 2019 Ice and Snow 59(1) pp 81-92 (In Russian).
[10] Shlyakhin V N 1987 Radiotechnika i elektronika 32 p 9
[11] Shirman Ya D, Leschenko S P and Orlenko V N 2004 Ultrawideband and Ultrashort Impulse Signals Proceedings 1 pp 71-6
[12] Sopelnik Yu V, Borisov S V and Volodin A V 1997 Radiotechnika 5 pp 27-31
[13] Sudakova M S, Vladov M L and Sadurtdinov M R 2018 Moscow University Geology Bulletin 73 pp 206–12
[14] Denisova O V, Belitskiy A A and Shelkovenikova I V 2018 Proc. of the 2018 IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering. EIConRus 1203-4 DOI: 10.1109/EIConRus.2018.8317307
[15] van Dam R L da Silva P R, de Almeida I M, Lima J S, Fonseca R V, Damas P V, de Souza M N, Ludvig P and Aranha P R 2019 Anais do 61 Congr. Brasileiro Concreto CBC 61C CBC 4191 v2 (1)
[16] Yang Y, Zhao W and Li R 2019 Structural concrete Journ. of the fib 21 (2) pp 772-80
[17] Bhattacharyya M et al. 2008 European Journal of Operational Research 191 pp 386–97