© L.Eppelbaum, V.Zheludev, A.Averbuch, 2014

DIFFUSION MAPS AS A POWERFUL TOOL FOR INTEGRATED GEOPHYSICAL ANALYSIS TO DETECTING HIDDEN KARST TERRANES

L.Eppelbaum, V.Zheludev, A.Averbuch

Tel Aviv University
Ramat Aviv 69978, Tel Aviv, Israel

Karst areas occupy about 14% of the world's land and cause serious economic losses, which are estimated at dozens of billions dollars, set aside the physical danger. Among various geophysical methods that can be used to detect these objects at the first stage gravity and magnetic surveys were selected. However, complex host geological media and an unfavorable S/N ratio often do not enable to reveal the geophysical anomalies from the desired targets. Our approach to recognition of separate geophysical field peculiarities and geophysical integration consists in application of modern developments in the wavelet theory and data mining. Wavelet approach is applied for derivation of enhanced (e.g., coherence portraits) and combined images of geophysical fields simulated for the typical physical-geological models of karst development. The methodology based on the matching pursuit with wavelet packet dictionaries enables to extract desired signals even from strongly noised data. The recently developed technique of diffusion clustering combined with the abovementioned wavelet methods is utilized in to integrate the geophysical data and to detect existing irregularities in the subsurface structure. The most important factor is that these obtained results may be applied for revealing karst targets (on the basis of "learning" approach) from the field geophysical observations. The combination of the above approaches enables to create a new methodology, which enhances reliability and confidence of application of any individual geophysical method and geophysical method integration. This methodology could be an effective tool to recognizing not only hidden karst terranes, but also any typical geological (economic minerals, geological mapping, etc.), archaeological, ecological and other targets.

Introduction

Karst areas occupy about 14% of the world land (Price, 2009). Karst terranes of different origin have caused difficult conditions for building, industrial activity and tourism, and are the source of heightened danger for environment. Mapping of karst (sinkhole) hazards, obviously, will be one of the most significant problems of engineering geophysics in the XXI century (Beck, 2003; Fookes et al., 2005). Annual damage of karst hazards consists of about 10 billion $ only in the USA. Early delineation of karst (sinkhole) terranes (when these objects are generating a depth) can be implemented only on the basis of geophysical method application. However, in many cases a low S/N ratio does not allow to recognize the karst terranes. Therefore, we propose to apply the newly developed powerful wavelet methods for retrieving the useful component from the total observed signal for delineation of the desired targets.

Taking into account the complexity of geological media, some unfavourable environments and known ambiguity of geophysical data analysis (Telford et al., 1990), a single geophysical method examination might be insufficient (Khesin and Eppelbaum, 1997). At this stage we select for hidden karst recognition two widely applied geophysical methods: gravity and magnetic fields.

Design of a new modern gravimetric and magnetometric equipment (which enable registering earlier inaccessible small anomalies and improves the observation methodology) and development of new methodologies for gravity-magnetic data processing and interpretation triggered multiple applications of the mentioned fields to karst areas (e.g., Waltham et al., 2004; Butler, 2005; Debeglia et al., 2006; Eppelbaum, 2007; Eppelbaum et al., 2008; Mochales et al., 2008; Jacob et al., 2010; Chalikakis et al., 2011; Kaufmann et al., 2011; Ezersky et al., 2013).

Wavelet methodology as whole has a significant impact on cardinal problems of geophysical signal processing such as: denoising of signals, enhancement of signals and distinguishing of signals with closely related characteristics (Kumar and Foufoula-Georgiou, 1997; Zheludev and Alperovich, 1999; Alperovich et al., 2003; Grinsted et al., 2004; Wapenaar et al., 2005; Yates et al., 2007; Fedi et al., 2010; Eppelbaum et al., 2011; Averbuch and Zheludev, 2014).

Based on the all abovementioned we developed a ***three-phase approach*** to the integrated Gravity–Magnetic (GM) localization of subsurface karsts (the same approach could be used for following monitoring of karst dynamics). The ***first phase*** consists of 3D computer modeling of geophysical

effects from the typical karst terranes. We consider in this article high-precision gravity and magnetic effects (we plan to consider an additional combination of GPR and ERT in a separate publication) from different types of hidden karst terranes. The ***second phase*** determines development of the signal processing approaches to analyzing of profile (in this paper we present modeling of profile results, but computed as 3D) or areal GM observations. Finally, at the ***third phase*** a new methodology of different geophysical data combined interpretation (on the basis of the modern developments in the wavelet technique of signal and image processing) is presented.

Thus, we propose that the wavelet integrated methodology of geophysical field examination will enable to recognizing the karst terranes in complex geological environments even by a small S/N ratio.

Short description of the applied methodology

On the basis of review of numerous literature sources (e.g., Beck, 2003; Waltham et al., 2004; Fookes et al., 2005; Chalikakis et al., 2011; Eppelbaum, 2007; Legchenko et al., 2008; Price, 2009; Ezersky et al., 2010, 2013) and the authors' personal experience, 30 physical-geological models (*PGM*) – typical geological sections containing karst cavities – were developed. The same sections were also computed without the karst cavity – altogether sixty working *PGM* were utilized. Figures 1a, 1b and 1c show gravity and magnetic curves computed (for these computations 3D GSFC-M software (Eppelbaum, 2007; Eppelbaum et al., 2008) was used) from the *PGM* 1a (with cavity), *PGM* 1b (the same, but without cavity), and *PGM* 1c (from cavity only) respectively. *PGM* 1c (as well as similar models with the GM effects computed from cavities only for other classes of models) was utilized for the GM "useful signal learning". Figures 2 and 3 display gravity and magnetic curves computed from *PGM*s 19a and 28a (both with cavity), respectively. Unfortunately, we cannot present in this paper all sixty (and with only cavities effects – ninthly) developed models.

It should be noted that anomalous gravity and magnetic effects from the karst cavities (negative anomalies) were specially modeled as significantly small ones. Physical characteristics presented in the figures demonstrate not only density-magnetization properties but indicate that the next stage of such analysis will include computing GPR and ERT images for the same sections.

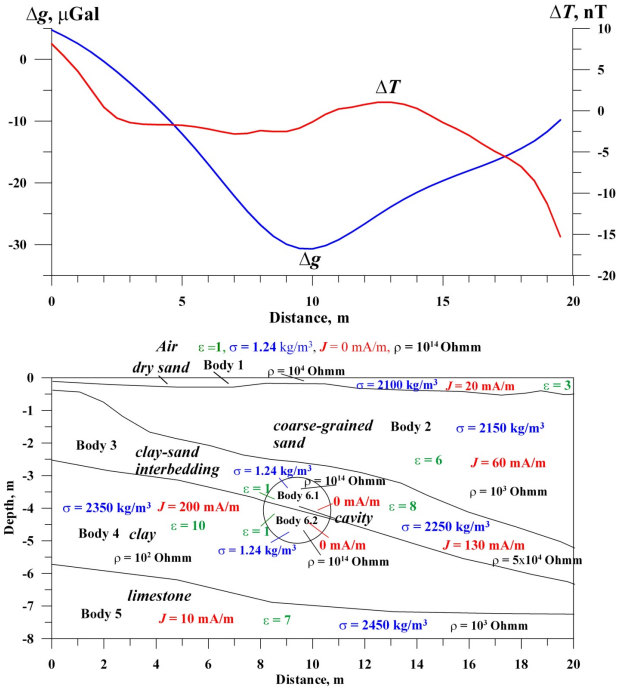

Figure 1a. 3D gravity-magnetic effects from the *PGM* with karst cavity (**Model 1a**). Here σ is the density, J is the magnetization, ρ is the electric resistivity, and ε is the dielectric permittivity (last two parameters are intended for the further GPR and ERT modeling). Gravity curve is presented in microGal (1 μGal = 10^{-8} m/s^2) and magnetic curve – in nanoTesla (nT).

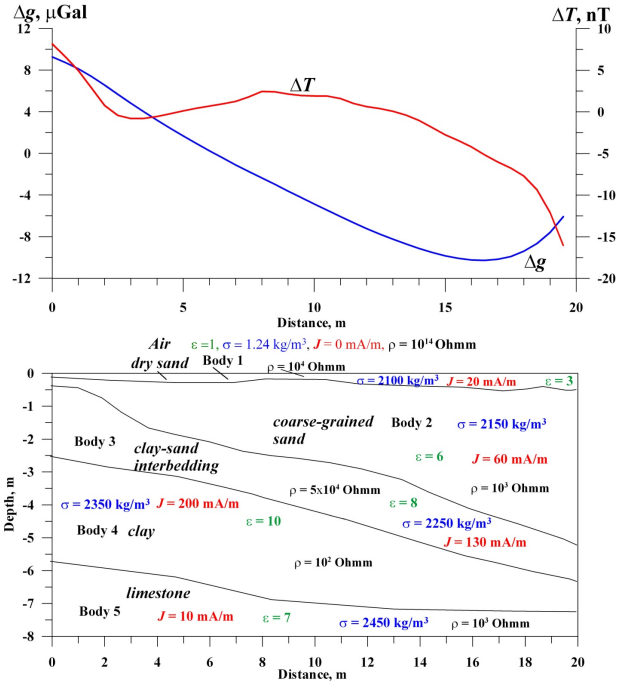

Figure 1b. 3D gravity-magnetic effects from the *PGM* without karst cavity (**Model 1b**).

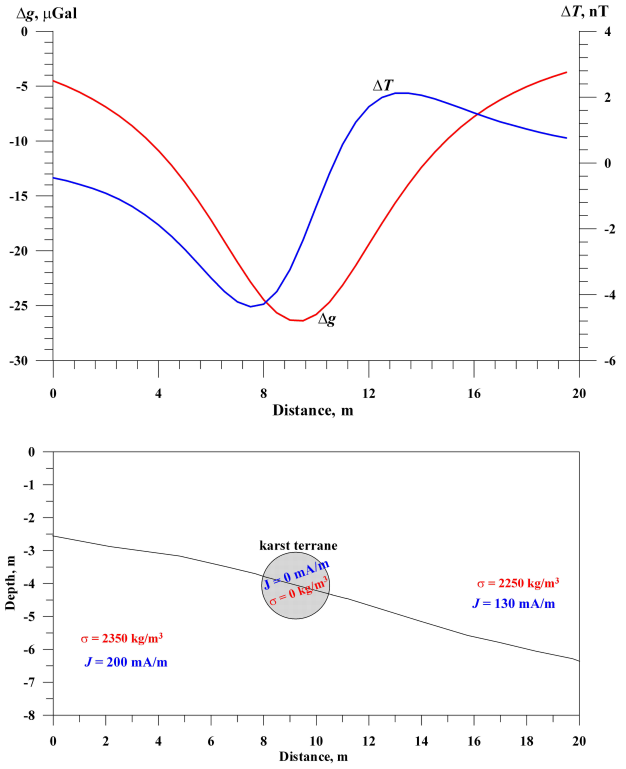

Figure 1c. 3D gravity-magnetic effects from the *PGM* with karst cavity only (**Model 1c**).

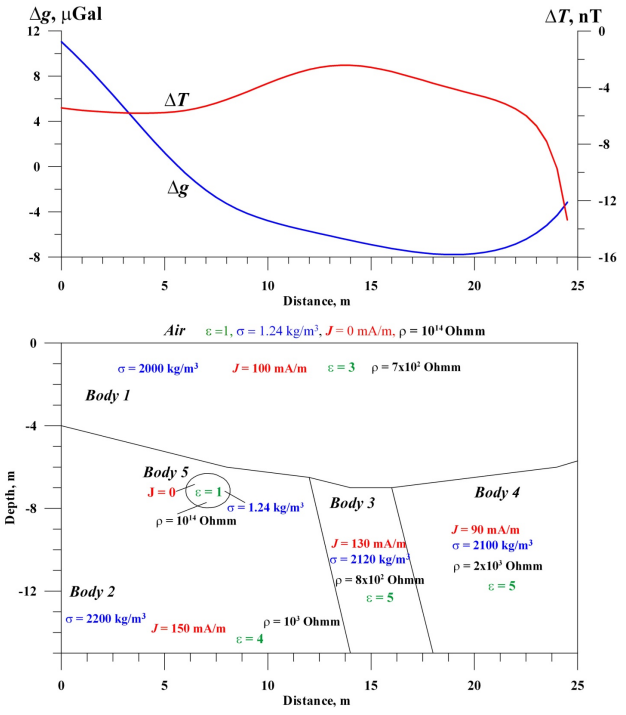

Figure 2. 3D gravity-magnetic effects from the *PGM* with karst cavity (**Model 19a**). Description of physical properties is given in Figure 1a.

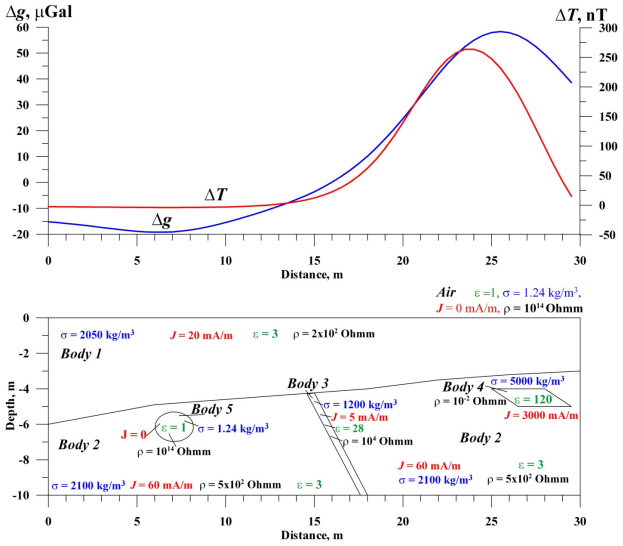

Figure 3. 3D gravity-magnetic effects from the *PGM* with karst cavity (**Model 28a**). Description of physical properties is given in Figure 1a.

We must underline that several last *PGM* (27-30) were specially complicated by gravity-magnetic effects from fault and some metallic target (bodies 3 and 4, respectively in Figure 3). Introducing such complicated *PGM*s, undoubtedly, reflects some real class of karst terranes occurring in such unfavorable for geophysical searching conditions. Anomalous effects from the two mentioned targets significantly disturb the integrated gravity-magnetic anomalies and impede the wavelet methodology employment for underground karst terrane delineation.

The proposed wavelet approach to recognition of karst terranes by the examination of geophysical method integration consists of advanced processing of each geophysical method and non-conventional integration of different geophysical methods between themselves. Modern developments in the wavelet theory and data mining are utilized for the analysis of the integrated data. Wavelet approach is applied for derivation of enhanced (e.g., coherence portraits) and combined images of geophysical fields computed for typical karst models. The methodology based on the matching pursuit with wavelet packet dictionaries enables to extract desired signals even from strongly noised data.

Researchers usually met the problem of extraction of essential features from available data contaminated by a random noise and by a non-relevant background (Averbuch and Zheludev, 2014). If the essential structure of a signal con-

sists of several sine waves then we may represent it via trigonometric basis (Fourier analysis). In this case one can compare the signal with a set of sinusoids and extract consistent ones. An indicator of presence a wave in a signal $f(t)$ is the Fourier coefficient

$$\int f(t)\sin wt\ dt\,. \qquad (1)$$

Wavelet analysis provides a rich library of waveforms available and fast, computationally efficient procedures of representation of signals and of selection of relevant waveforms. The basic assumption justifying an application of wavelet analysis is that the essential structure of a signal analyzed consists of not a large number of various waveforms. The best way to reveal this structure is representation of the signal by a set of basic elements containing waveforms coherent to the signal. For structures of the signal coherent to the basis, large coefficients are attributed to a few basic waveforms, whereas we expect small coefficients for the noise and structures incoherent to all basic waveforms.

Wavelets are a family of functions ranging from functions of arbitrary smoothness to fractal ones. Wavelet procedure involves two aspects. The first one is a *decomposition*, i.e. breaking up a signal to obtain the wavelet coefficients and the 2nd one is a *reconstruction*, which consists of a reassembling the signal from coefficients

There are many modifications of the WA. Note, first of all, so-called *Continuous WA* in which signal $f(t)$ is tested for presence of waveforms $\psi\left(\frac{t-b}{a}\right)$.

Here, a is scaling parameter (dilation), b determines location of the wavelet $\psi\left(\frac{t-b}{a}\right)$ in a signal $f(t)$. The integral

$$(W_\psi f)(b,a)=\int f(t)\psi\left(\frac{t-b}{a}\right)dt \qquad (2)$$

is the Continuous Wavelet Transform.

When parameters a, b in $\psi\left(\frac{t-b}{a}\right)$ take some discrete values, we have the Discrete Wavelet Transform. A general scheme of the Wavelet Decomposition Tree is shown in Figure 4.

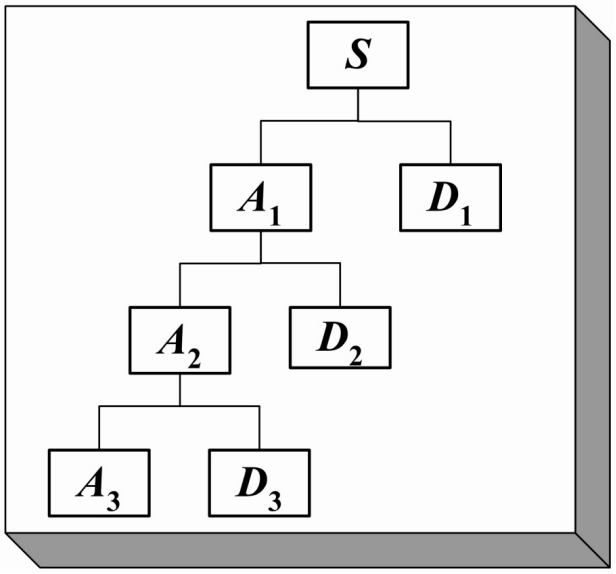

Figure 4. Block diagram of wavelet analysis. A_i = approximation, D_i = details

The signal is compared with the testing signal on each scale. It is estimated wavelet coefficients which enable to reconstruct the 1st approximation of the signal and details. On the next level, wavelet transform is applied to the approximation. Then, we can present A_1 as $A_2 + D_2$, etc. So, if S – Signal, A – Approximation, D – Details, then

$$S = A_1 + D_1 = A_2 + D_2 + D_1 = A_3 + D_3 + D_2 + D_1.$$

Wavelet packet transform is applied to both low pass results (approximations) and high pass results (Details) (Figure 5).

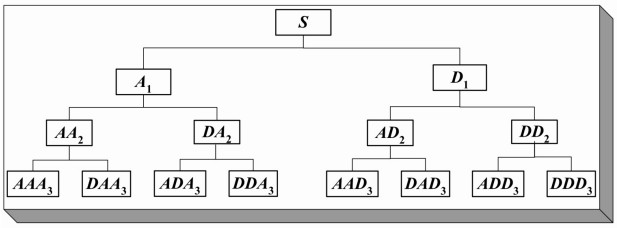

Figure 5. Wavelet packet tree

The goal of the modern geophysical data examination is to detect the geophysical signatures of buried targets at noisy areas via the analysis of some physical parameters with a minimal number of false alarms and miss-detections (Eppelbaum et al., 2011). The recently developed technique of diffusion clustering combined with the abovementioned wavelet methods was utilized to integrate the

GM data and detect existing irregularities. The approach is based on the wavelet packet techniques applied as to the GM graphs (it can be applied also for images) versus coordinates.

For analyzing the GM data, we used a technique based on the algorithm to characterize a GM image by a limited number of parameters (Eppelbaum et al., 2012). This set of parameters serves as a signature of the image and is utilized for discrimination of images containing karst cavity from the images non-containing karst. The constructed algorithm consists of the following main phases: (a) collection of the database, (b) characterization of GM images, (c) and dimensionality reduction. Then, each image is characterized by the histogram of the coherency directions (Alperovich et al., 2013a). As a result of the previous steps we obtain two sets **K** and **N** of the signatures vectors for images from sections containing karst cavity and non-karst subsurface, respectively. The obtained 3D set of the data representatives can be used as a reference set for the classification of newly arriving data.

Gravity and magnetic data processing

As a result of our numerical simulations, we found 30 pairs of data vectors $\mathbf{g}_k$ and $\mathbf{m}_k$, $k = 1,\ldots,30$ which are derived from the gravity and magnetic computations, respectively, from the subsurface without a cavity (**N** set) and the same number of pairs retrieved from the subsurface with a karst cavity (**K** set).

The interim goal is to extract characteristic features of these two sets in a way that the maps of these sets in the features space to be well separated. Once this goal is achieved, these maps can be used as reference sets for the construction of classifiers, which automatically decide, whether a newly arrived data indicate a cavity presence or not.

The goal is achieved via a number of steps. Either of the gravity and magnetic data vectors $\mathbf{g}_k$ and $\mathbf{m}_k$, $k = 1,\ldots,30$ comprises 40÷60 computation points (step of computation was 0.5 m), which, being plotted look like smooth curves. Then, we couple these two vectors into one combined data vector. For further processing we need the data vectors, whose lengths are a power of 2. Thus, we apply to each pair of vectors the following operations. Each vector is symmetrically extended via its left and right ends to the length 64. The window is applied to the extended vector in order to avoid discontinuities when merging (Figure 6).

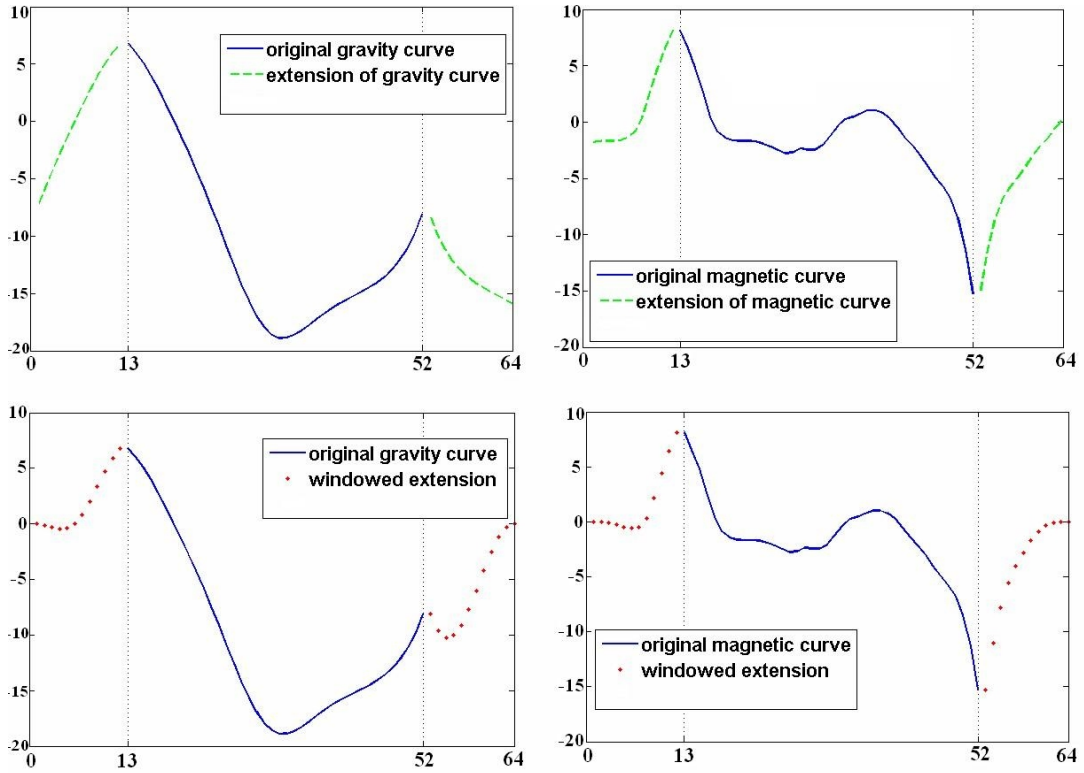

Figure 6. Preprocessing a gravity-magnetic pair of data vectors. Upper panel respects to the case of "no karst", and the lower one shows the case of "karst" presence

And, at last, the extended and mirrored gravity and magnetic data vectors are merged into one combined vector of length 128. The window is applied to the extended vector in order to avoid jumps when merging. Figure 6 demonstrates the main steps of our wavelet preprocessing: extension of the initial gravity and magnetic observational curves to 2^n – point curves, then their smoothing by a window. Figure 7A displays merged gravity and magnetic data curves for a case of *PGMs* without karst cavity, and Figure 7B – merged gravity and magnetic data curves for the *PGMs* with a simulated karst cavity.

We observe that the curves differ by subtle geometrical details. To catch these differences and use them for discrimination of **N** versus **K** sets we use wavelet packet transforms and dimensionality reduction by Diffusion Maps (Appendix A).

Extraction of characteristic features using wavelet packet transforms

We choose to characterize the geometric events in the merged data curves via the representation of the curves in the Fourier domain. The assumption is that each of **K** and **N** data sets is characterized by a number of dominating frequency bands. A perfect tool to reveal these characteristic frequency bands is provided by the wavelet packet (WP) analysis (Averbuch et al., 2001; 2010a). Once implemented, the wavelet packet transform of a signal yields highly redundant partitions of the frequency domain.

The transform consists in iterative application to a signal (curve) a pair (lowpass / highpass) of quadrature mirror filters. To be specific, let **L** and **H** be lowpass and highpass quadrature mirror filters. Their magnitude responses divide the frequency domain into two halves. The one-scale wavelet packet transform of a signal **f** of length $N = 2^j$ consists in application to the signal the filters **L** and **H**, which is followed by down sampling by factor of 2. Thus, we get two coefficient arrays **Lf** and **Hf** of length **N/2** each. The transform is invertible. The array **Lf** contains all the information needed to restore low-frequency component of the signal **f** while the high-frequency component can be restored from **Hf**. Since **L** and **H** are quadrature mirror filters, the energy (sum of squares) in the coefficient arrays **Lf** and **Hf** are equal to the energies in the low- and high-frequency components of the signal **f**, respectively.

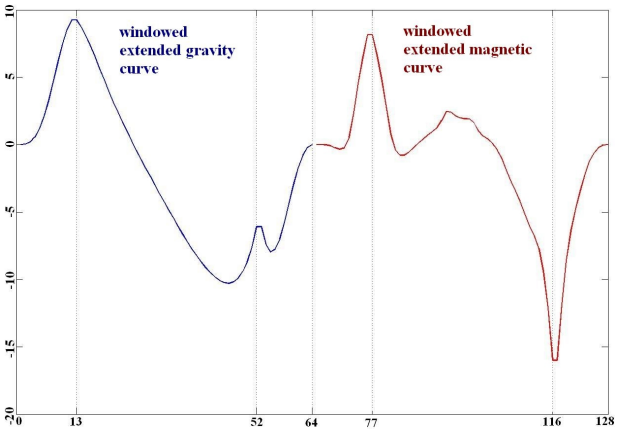

Figure 7A. Merged combined gravity and magnetic curves from the *PGMs* without underground karst

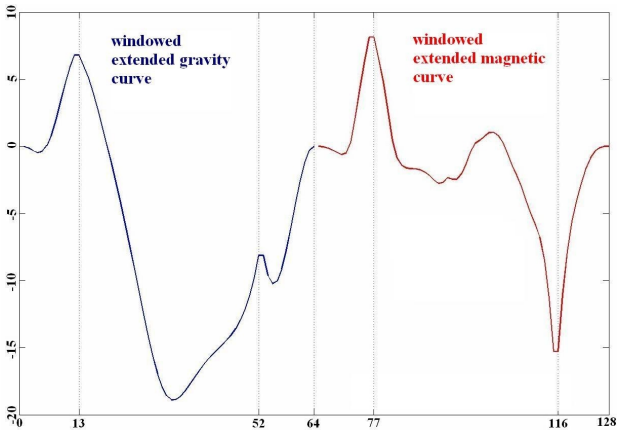

Figure 7B. Merged combined gravity and magnetic curves from the *PGMs* with presence of underground karst

The second-scale transform consists in application of the filters **L** and **H** to either of the arrays **Lf** and **Hf** and subsequent down sampling. As a result, we get four coefficient arrays **LLf**, **LHf** **HLf** and **HHf** of length $N/4$ each, which are related to quarters 3 of the frequency domain. Similarly, to pass to the next scale, the filters **L** and **H** are applied to either of the four second-scale arrays. After down sampling, eight subarrays are produced and so on. Figure 8 displays the diagram of the WP transform. Each block of the wavelet packet coefficients is related to a certain frequency band. Blocks from different scales are related to frequency bands of different widths. Set of energies within the coefficients blocks produces a "spectrum" of frequency bands of a signal rather than the spectrum of frequencies.

Characteristic features of the **N** and **K** sets are extracted in a following way. The data curves from **N** ("no cavity") and **K** ("karst cavity") are

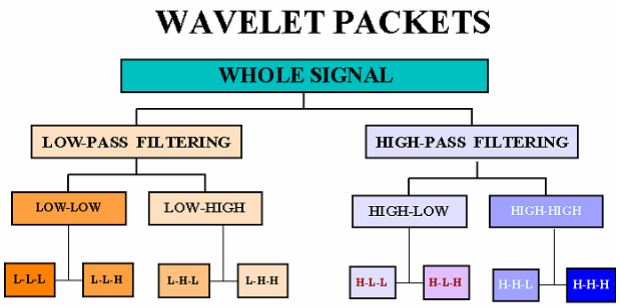

Figure 8. Diagram of WP transform

gathered into two matrices $\mathbf{M}_N$ and $\mathbf{M}_K$, respectively. Both matrices are of size 30 x 128. Then we applied the wavelet packet transform down to 4-th decomposition scale on each row of the matrices $\mathbf{M}_N$ and $\mathbf{M}_K$. Energies (sums of squares) of each out of 31 blocks of the transform coefficients are calculated. Gather these energies into matrices $\mathbf{E}_N$ and $\mathbf{E}_K$ of size 30 x 31, respectively.

Average the energy matrices $\mathbf{E}_N$ and $\mathbf{E}_K$ along the columns are computed. Two vectors $\overline{\mathbf{E}}_N = \text{mean}(\mathbf{E}_N)$ and $\overline{\mathbf{E}}_K = \text{mean}(\mathbf{E}_K)$ of length 31 are regarded as signatures of the classes **N** and **K**, respectively. We calculate the difference $D_{KN} = \overline{\mathbf{E}}_K - \overline{\mathbf{E}}_N$ of the signature vectors. The difference indicates how different blocks of the transform coefficients discriminate between two data sets. Next step is to select C (8-12) blocks, which best discriminate between two data sets. Let $\mathbf{I} = i_1, \ldots i_k$ be a list of indices of the best discriminating blocks. And finally, we reduce the energy matrices $\mathbf{E}_N$ and $\mathbf{E}_K$ to the feature matrices $\mathbf{F}_N$ and $\mathbf{F}_K$ of size 30 x C by discarding all the columns, whose indices do not participate in the list **I**. Thus, each merged data curve is represented by C features.

Figure 9A displays the signature vectors $\overline{\mathbf{E}}_K$ and $\overline{\mathbf{E}}_N$, and Figure 9B – their difference $\mathbf{D}_{KN}$. This difference, taking into consideration low values of anomalous gravity-magnetic effects is sufficiently small but detectable.

Reduction of dimensionality and obtained results

We further reduce the data sets by embedding features vectors into the 3D Euclidean space using the so-called diffusion map. This map enables to reveal the internal structure of the datasets **K** and **N** and to distinctly separate them. For this, a matrix of the diffusion distances for the combined feature matrix $\mathbf{F} = \mathbf{F}_N \cup \mathbf{F}_C$ of size 60 x C is constructed (Coifman and Lafon, 2006; Averbuch et al., 2010b; Appendix A). Then, each row of the matrices $\mathbf{F}_N$ and $\mathbf{F}_C$ is projectted onto three first eigenvectors of the matrix $\mathbf{D}(\mathbf{F})$. As a result, each data curve is represented by a 3D point in the Euclidean space formed by eigenvectors of $\mathbf{D}(\mathbf{F})$.

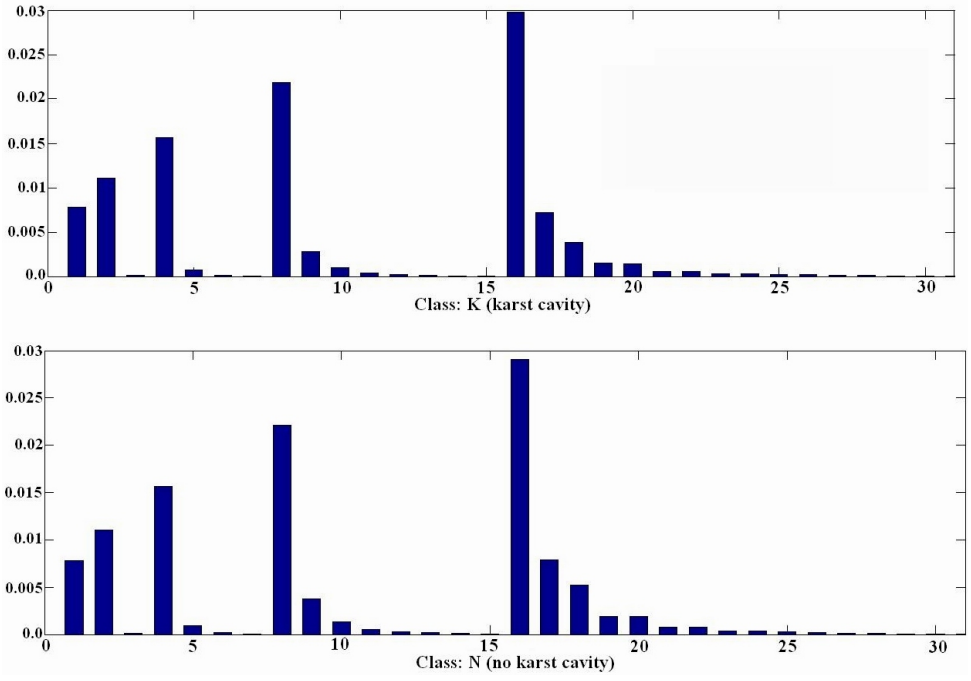

Figure 9A. Signature vectors $\overline{\mathbf{E}}_N$ and $\overline{\mathbf{E}}_K$ of the classes **N** and **K**

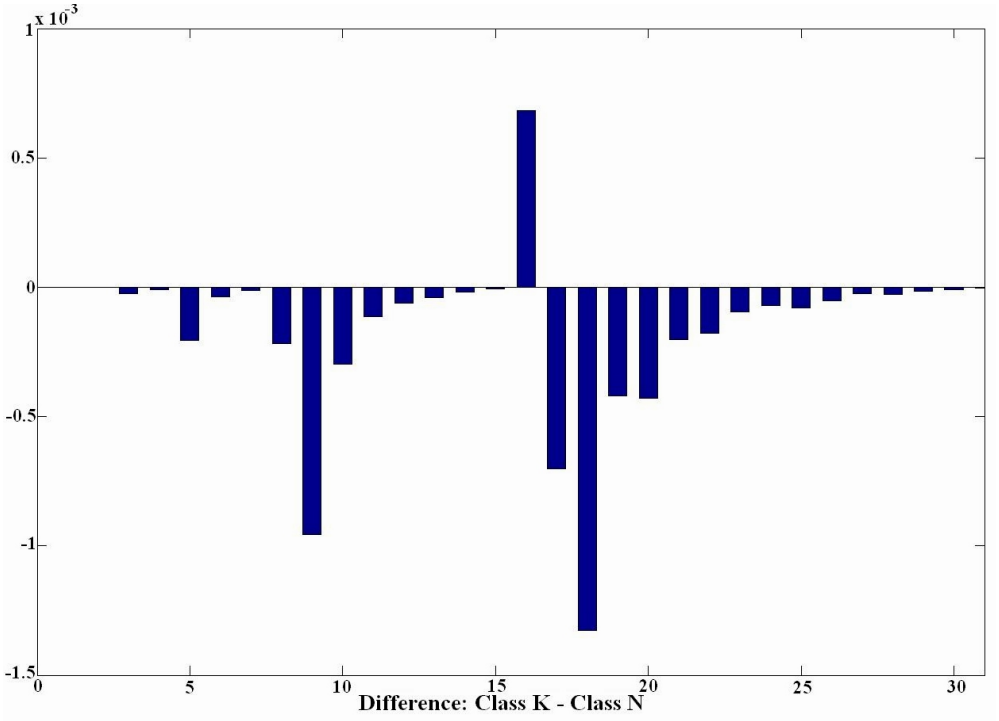

Figure 9B. Difference $\mathbf{D}_{KN}$ for the classes **N** and **K**

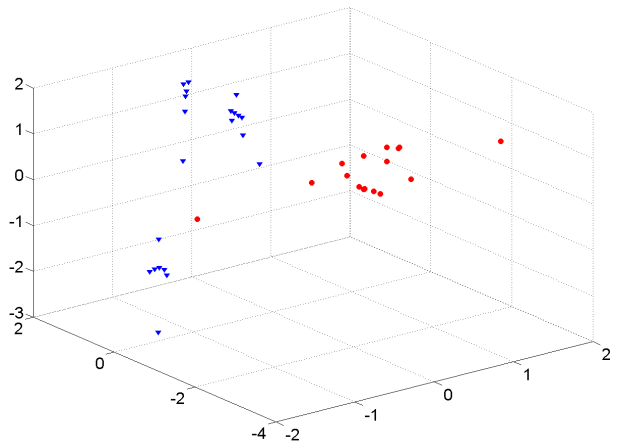

Figure 10. Scattered projections of the data curves onto the diffusion eigenvectors. Blue triangles represent the **N** data curves (non-karst areas) while red circles represent **K** (karst areas)

The Euclidean distances between these 3D points reflect the similarity of the data curves. Figure 10 displays scattered projections of the data curves onto the diffusion eigenvectors. We observe that as a result of the above operations we embedded the original data into 3-dimensional space where data related to the **K** (karst cavity) subsurface are well separated from the **N** (non-cavity) data. This 3D set of the data representatives can be used as a reference set for the classification of newly arriving data. It is clear score of blue triangles (absence of karst) is unfailingly separated from score of red circles (presence of karst). Geophysically it means a reliable division of the studied areas for the karst- containing and free from the karst terranes.

Thus, we demonstrated the real possibilities for classification of geophysical anomalies (on example of series of gravity-magnetic *PGMs*) to detecting hidden karst terranes in complex physical-geological environments. Obviously, such computations in karst areas could be performed for other geophysical method integrations. Some advanced wavelet methodologies were recently applied for the combined GPR – ERT comprehensive examination in the Montagnole testing site, Alps (Alperovich et al., 2013b). Besides this, such an approach may be employed not only for the buried karst terrane delineation, but also for solving the variety of geological-geophysical problems: searching useful minerals, geological mapping, localization of archaeological remains, ecological and engineering examination, etc.

Conclusions

An effectiveness of advanced wavelet methodology to recognizing subsurface objects is demonstrated on example of processing gravity-magnetic fields computed from thirty typical models containing karst cavities and thirty the same models but without the cavities. As a result of ad-

vanced wavelet processing we obtain two sets **K** and **N** of the signatures vectors for images from the karst and non-karst subsurface, respectively. We embed these sets into a low-dimensional Euclidean space using Diffusion Maps. This embedding enables to reveal similarity between the members of the same set and to discriminate between different sets. This 3D set of the data representatives can be used as a reference set for the classification of newly arriving data. It should be noted that this wavelet methodology may be applied to localization of any geological objects by application of arbitrary geophysical (geochemical, geotechnical, etc.) tools. For this aim is necessary to perform preliminary wavelet "learning" on a large assembly of typical physical-geological (physical-archaeological, geochemical, engineering, etc.) models.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under Grant Agreement No. 225663.

Appendix. *A brief description of diffusion framework*

This section describes the diffusion framework that was introduced in (Perona and Malik, 1990; Coifman and Lafon, 2006). Diffusion maps and diffusion distances provide a method for finding meaningful geometric structures in datasets. In most cases, the dataset contains high dimensional data points in $\mathbf{R}^n$. The diffusion maps construct coordinates that parameterize the dataset and the diffusion distance provides a local preserving metric for this data. A non-linear dimensionality reduction, which reveals global geometric information, is constructed by local overlapping structures. Let $\Gamma = \{x_1,\ldots,x_m\}$ be a set of points in $\mathbf{R}^n$. We construct the graph on Γ in order to study the intrinsic geometry of this set. A weight function $w_\varepsilon(x_i,x_j)$ is introduced. It measures the pairwise similarity between the points. For all $x_i,x_j \in \Gamma_i$, the weight function has the following properties:

symmetry: $w_\varepsilon(x_i,x_j) = w_\varepsilon(x_j,x_i)$

non-negativity: $w_\varepsilon(x_i,x_j) \ge 0$

positive semi-definiteness: for all real-valued bounded function f defined on Γ,

$$\sum_{x_i\in\Gamma}\sum_{x_j\in\Gamma} w_\varepsilon(x_i,x_j) f(x_i) f(x_j) \ge 0 . \qquad \text{(A1)}$$

A common choice for w_ε is

$$w_\varepsilon(x_i,x_j) = e^{-\frac{\|x_i-x_j\|}{\varepsilon}} . \qquad \text{(A2)}$$

The non-negativity of w_ε allows to normalize it into a Markov transition matrix *P* where the states of the corresponding Markov process are the data points. This enables to analyze Γ as a random walk. The construction of *P* is known as the normalized graph Laplacian (Chung, 1997). Formally, $P = \{p(x_i,x_j)\}_{i,j=1,\ldots,m}$ is constructed as

$$p(x_i,x_j) = \frac{w_\varepsilon(x_i,x_j)}{d(x_i)} , \qquad \text{(A3)}$$

where

$$d(x_i) = \sum_{x_j\in\Gamma} w_\varepsilon(x_i,x_j) \qquad \text{(A4)}$$

is the degree of the node x_i. *P* is a Markov matrix since the sum of each row in *P* is 1 and $p(x_i,x_j) \ge 0$.

Thus, $p(x_i,x_j)$ can be viewed as the probability to move from x_i to x_j in *one* time step. Consequently, the matrix $P^t = \{p_t(x_i,x_j)\}_{i,j=1,\ldots,m}$ consists of the probabilities for this move in *t* time steps. These probabilities measure the connectivity among the points within the graph. The parameter *t* controls the scale of the neighborhood in addition to the scale control provided by ε.

The transition matrix *P* is conjugate to the matrix $\widetilde{P} = \{\widetilde{p}(x_i,x_j)\}_{i,j=1,\ldots,m}$, where

$$\widetilde{p}(x_i,x_j) = \frac{\sqrt{d(x_i)}}{\sqrt{d(x_j)}} p(x_i,x_j) , \qquad \text{(A5)}$$

which is a symmetric and positive semi-definite. The matrix $\widetilde{P}$ has *m* real eigenvalues $\{\lambda_k\}$ and *m* orthonormal eigenvectors $\{\nu_k\}$. This leads to the following eigen-decomposition of the matrix $\widetilde{P}$:

$$\widetilde{p}(x_i,x_j) = \sum_{k=0}^{m-1} \lambda_k \nu_k(x_i) \nu_k(x_j) . \qquad \text{(A6)}$$

A similar eigen-decomposition is obtained for t-step matrix P^t

$$\widetilde{p}_t(x_i,x_j) = \sum_{k=0}^{m-1} \lambda_k^t \nu_k(x_i) \nu_k(x_j) \qquad \text{(A7)}$$

after advancing *t* times on the graph.

A fast decay of $\{\lambda_k\}$ is achieved by an appropriate choice of ε. Thus, only a few terms are required in the sum in Eq. (A7).

A family of diffusion maps $\Phi_t(x)_{m\in N}$ was introduced in (Coifman and Lafon, 2006):

$$\Phi_t(x) = \begin{pmatrix} \lambda_0^t \nu_0(x) \\ \lambda_1^t \nu_1(x) \\ \vdots \\ \lambda_{m-1}^t \nu_{m-1}(x) \end{pmatrix} .$$

The map Φ_t embeds the dataset into a Euclidian space. The distance between points in this space, the so-called *diffusion distance* is

$$D_t^2(x_i, x_j) = \sum_{k=0}^{m-1} \left(\widetilde{p}_t(x_i, x_k) - \widetilde{p}_t(x_{ik}, x_j)\right)^2.$$

The diffusion distance can be expressed in terms of the eigenvectors of $\widetilde{P}$:

$$D_t^2(x_i, x_j) = \sum_{k=0}^{m-1} \lambda_k^{2t} \left(v_k(x_i) - v_k(x_j)\right)^2. \qquad \text{(A8)}$$

It follows that in order to compute the diffusion distance, one can simply use the eigenvectors of $\widetilde{P}$. Due to the fast decay of the eigenvalues λ_k (and even faster decay of $\lambda_k^t, t > 1,$) only ν terms in the sum (A8) are essential. Thus, actually we implement embedding of the original points into a Euclidean space $\mathbf{R}^\eta$ by:

$$\Xi_t : x_i \to \left(\lambda_0^t v_0(x_i), \lambda_1^t v_1(x_i), \lambda_2^t v_2(x_i), \ldots, \lambda_{\eta-1}^t v_{\eta-1}(x_i)\right). \qquad \text{(A9)}$$

This also provides coordinates on the set Γ . Essentially, $\eta << n$, where n is the original dimensionality of the data points $\{x_i\}$ due to the fast spectral decay of the eigenvalues λ_k of $\widetilde{P}$. Furthermore, η depends only on the primary intrinsic variability of the data as captured by the random walk and not on the original dimensionality of the data. This data-driven method enables the parameterization of any set of points provided the similarity matrix of the points W_ε is available. This parameterization essentially reduces the dimensionality of the data while revealing the internal structure of the dataset.

BIBLIOGRAPHY

ALPEROVICH, L., AVERBUCH, A., EPPELBAUM, L., ZHELUDEV, V. 2013a. Delineation of karst terranes in complex environments: Application of modern developments in the wavelet theory and data mining. Trans. of the 9th EUG Meet., *Geophysical Research Abstracts*, 15, EGU2013-11063, Vienna, Austria, 1-2.

ALPEROVICH, L., EPPELBAUM, L., ZHELUDEV, V., DUMOULIN, J., SOLDOVIERI, F., PROTO, M., BAVUSI, M., LOPERTE, A. 2013b. A new combined wavelet methodology applied to GPR and ERT data in the Montagnole experiment (French Alps). *Journal of Geophysics and Engineering*, 10, 2, 025017, 1-17.

ALPEROVICH, L., ZHELUDEV, V., HAYAKAWA, M. 2003. Use of wavelet analysis for detection of seismogenic ULF emissions. *Radio Science*, 38, 6, 1093, doi: 10.1029/2002RS002687, 1-12.

AVERBUCH, A.Z., HULATA, E., ZHELUDEV, V.A., KOZLOV, I. 2001. A wavelet packet algorithm for classification and detection of moving vehicles. *Multidimensional Systems and Signal Processing*, 12, 1, 9-31.

AVERBUCH, A.Z., ZHELUDEV, V.A. 2014. Image compression using spline based wavelet transforms. Vol. XII in Series: *Wavelets in Signal and Image Analysis*, Springer.

AVERBUCH, A., ZHELUDEV, V., NEITTAANM, P., KOREN, J. 2010a. Block based deconvolution algorithm using spline wavelet packets. *Journal of Mathematical Imaging and Vision*, 38, 3, 197-225.

AVERBUCH, A., HOCHMAN, K., RABIN, N., SCHCLAR, A., ZHELUDEV, V. 2010b. A diffusion frame-work for detection of moving vehicles. *Digital Signal Processing*, 20, 1, 111-122.

BECK, B.F. (ed.). 2003. Sinkholes and the engineering and environmental impacts of karst, geotechnical special publication **122**. American Soc. of Civil Engineers. RestonVA. 730p.

BUTLER, D.K. (ed.). 2005. Near-surface geophysics (investigations in geophysics, No. 13). *Society of Exploration Geophysicists*, 571p.

CHALIKAKIS C., PLAGNES, V., GUERIN, R. VALOIS, R., BOSCH, F.R. 2011. Contribution of geophysical methods to karst-system exploration: an overview. *Hydrogeology Journal*, doi: 10.1007/s10040-011-0746-x, 1-12.

CHUNG, F.R.K. 1997. Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions. *AMS Conf. Ser. Math.*, 92, 209p.

COIFMAN, R.R., LAFON, S. 2006. Diffusion maps, applied and computational harmonic analysis. *Special issue on Diffusion Maps and Wavelets*, 21, 7, 5-30.

DEBEGLIA, N., BITRI, A., THIERRY, P. 2006. Karst investigations using microgravity and MASW: application to Orléans, France. *Near Surface Geophysics*, 4, 215-225.

EPPELBAUM, L.V. 2007. Revealing of subterranean karst using modern analysis of potential and quasi-potential fields. *Coll. of Papers of the 2007 SAGEEP Conf.*, Denver, USA, 797-810.

EPPELBAUM, L.V., EZERSKY, M.G., AL-ZOUBI, A.S., GOLDSHMIDT, V.I., LEGCHENKO, A. 2008. Study of the factors affecting the karst volume assessment in the Dead Sea sinkhole problem using microgravity field analysis and 3D modeling. *Advances in GeoSciences*, 19, 97-115.

EPPELBAUM, L.V., ALPEROVICH, L.S., ZHELUDEV, V., EZERSKY, M., AL-ZOUBI, A., LEVI, E. 2012. Integration of potential and quasipotential geophysical fields and GPR data for delineation of buried karst terranes in complex environments. *Trans. of the 8th EUG Meet., Geophysical Research Abstracts,* 14, EGU2012-3418, Vienna, Austria, 1-3.

EPPELBAUM, L.V., ALPEROVICH, L., ZHELUDEV, V., PECHERSKY, A. 2011. Application of informational and wavelet approaches for integrated processing of geophysical data in complex environments. *Proceed. of the 2011 SAGEEP Conference*, Charleston, South Carolina, USA, 24, 24-60.

EZERSKY, M., EPPELBAUM, L., AL-ZOUBI, A., KEYDAR, S., MEDVIDIEV, B., AKKAWI, E., ABUELADAS, A.-R., AL-RUZOUQ, R. 2013. Geophysical prediction and following development sinkholes in two Dead Sea areas, Israel and Jordan. *Environmental Earth Sciences*, doi: 10.1007/s12665-013-2233-2, 1-16.

EZERSKY, M., LEGCHENKO, A., CAMERLYNCK, C., AL-ZOUBI, A., EPPELBAUM, L., KEYDAR, S., BOUCHER, M., CHALIKAKIS, K. 2010. The Dead Sea sinkhole hazard – new findings based on a multidisciplinary geophysical study. *Zeitschrift Fur Geomorphologie*, 54, 2, 69-90.

FEDI, M., CELLA, F., QUARTA, T., VILLANI, A.V. 2010. 2D Continuous wavelet transform of potential fields due to extended source distributions. *Applied and Computational Harmonic Analysis,* 28, 320-337.

FOOKES, P.G., LEE, E.M., MILLIGAN, G. (eds.). 2005. Geomorphology for engineers, Whittles Publisher. Dunbeath, UK. 851p.

GRINSTED, A., MOORE, J.C., JEVREJEVA, S. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11, 561-566.

JACOB, T., BAYER, R., CHERY, J., LE MOIGNE, N. 2010. Time‐lapse microgravity surveys reveal water storage heterogeneity of a karst aquifer. *Journal of Geophysical Research*, 115, B06402, doi: 10.1029/2009JB006616, 1-18.

KAUFMANN, G., ROMANOV, D., NIELBOCK, R. 2011. Cave detection using multiple geophysical methods: Unicorn cave, Harz Mountains, Germany. *Geophysics,* 76, 3, B71-B77.

KHESIN, B.E., EPPELBAUM, L.V. 1997. The number of geophysical methods required for target classification: quantitative estimation. *Geoinformatics*, 8, 1, 31-39.

KUMAR, P., FOUFOULA-GEORGIOU, E. 1997. Wavelet analysis for geophysical applications. *Review of Geophysics*, 35, 4, 385-412.

LEGCHENKO, A., EZERSKY, M., BOUCHER, M., CAMERLYNCK, C., AL-ZOUBI, A., CHALIKAKIS, K. 2008. Pre-existing caverns in salt formations could be the major cause of sinkhole hazards along the coast of the Dead Sea. *Geophysical Research Letters*, 35, L19404, doi: 10.1029/2008GL035510, 1-5.

MOCHALES, T., CASAS, A.M., PUEYO, E.L., PUEYO, O., ROMAN, M.T., POCOVI, A., SORIANO, M.A., ANSON, D. 2008. Detection of underground cavities by combining gravity, magnetic and ground penetrating radar surveys: a case study from the Zaragoza area, NE Spain. *Environmental Geology*, 53, 1067-1077.

PERONA, P., MALIK, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE trans. on pattern analysis and machine intelligence*, 12, 7, 629-639.

PRICE, D.G. 2009. Engineering Geology. Principles and Practice. Springer. 450p.

TELFORD, W.M., GELDART, L.R., SHERIFF, R.E. 1990. Applied Geophysics. 2nd ed. Cambridge Univ. Press. 792p.

WALTHAM, T., BELL, F.G., CULSHOW, M.G. 2004. Sinkholes and subsidence: karst and cavernous rocks in engineering and construction. Springer, 382p.

WAPENAAR, K., GHOSE, R., TOXOPEUS, G., FOKKEMA, J. 2005. The wavelet transform as a tool for geophysical data integration. *Integrated Computer-Aided Engineering*, 12, 1, 5-23.

YATES, T.T., SI, B.C., FARRELL, R.E., PENNOCK, D.J. 2007. Time, location, and scale dependence of soil nitrous oxide emissions, soil water, and temperature using wavelets, cross-wavelets, and wavelet coherency analysis. *Journal of Geophysical Research*, 112, D09104, doi: 10.1029/2006JD007662, 1-15.

ZHELUDEV, V., ALPEROVICH, L. 1999. Wavelet analysis as a tool for revealing geomagnetic precursors of an earthquake. In: *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes*. Terra Scientific Publishing Company. Tokyo, Japan. 255-269.

Reviewer: Corresponding member of ANAS P.Z.Mammadov