Statistical and geostatistical study of Rn and hydrocarbon components of a soil gas monitoring system: an application to surface hydrocarbon exploration

Janos Geiger[1], Noémi Jakab[1], Bálint Csökmei[1], Zsolt Horváth[2] and Balázs Gellért[2]

[1] University of Szeged, Department of Geology and Paleontology; (corresponding author: matska@geo.u-szeged.hu, j.noemi@geo.u-szeged.hu, balint.cs@gmail.com)
[2] MOL Hungary; (ZsoHorvath@mol.hu, BGellert@mol.hu)

Crossref
Similarity Check
Powered by iThenticate

Abstract

This work addresses three topics: (1) the study of the joint areal distribution of the Rn and hydrocarbon components of soil gases over a large region overlying some known hydrocarbon reservoirs in the southern part of Hungary; (2) the relationships between the positive anomalies of Rn and hydrocarbon components of soil gases to the existing reservoirs; (3) suggestions for new targets for surface hydrocarbon exploration based on the results. Given the very low correlation coefficients between the Rn and hydrocarbon components of the soil gases, factor analysis was used to reveal a background process controlling the common migration of hydrocarbon and Rn components. The lateral distribution of the factor scores were studied using sequential Gaussian distribution. The E-type grid generated from 100 realizations indicated several positive anomalies at the surface. Indications with a larger than 0.7 probability were kept for further analysis. Seismic sections of a 3D survey support the comparison of the surface locations of these anomalies and the surface projections of the known reservoirs. The results proved the connection between the known reservoirs and the Rn and HC components of soil gases. From the positive verification, regions with a high probability positive anomaly of factor scores, but without any reservoir counterparts may be suggested as targets for further surface hydrocarbon exploration.

1. INTRODUCTION

Surface geochemical techniques have been used to explore for hydrocarbons ever since the late 1920s. These techniques look for the presence and effects of minute levels of hydrocarbons that migrate through the imperfect seals that cover every reservoir and migrate either as macroseepage via faults or as microseepage vertically through the reservoir overburden (COLEMAN et al., 1977, KLUSMAN, 1993, WRIGLEY et al., 2012). Direct and indirect hydrocarbon indicators can be measured using these methods. Hydrocarbons measured on the surface are regarded as direct indicators, meanwhile the results of alternations induced by presence of hydrocarbons can be considered as indirect indicators. Soil gases pertain to direct indicators, and radon measurements, which show anomalously high values at the surface due to enrichment in uranium induced by the presence of migrating hydrocarbons, belong to indirect indicators.

Radiometric mapping, as a petroleum exploration tool, began at the end of 1920s (MORSE et al., 1982; MORSE & ZINKE, 1995). PIRSON found „halos" of high radiation in a petroleum basin in Texas, noting that each halo countered definable regions of lower gamma flux (PIRSON, 1969). Since then, several papers have reported on successful HC exploration strategies using radiometric mapping. PALACIOS et al. (2013) demonstrated positive radon anomalies around abandoned gas wells, LI et al. (2006), DYCK & JONASSON (2000), GHAGREMANI (1985) used Rn-anomalies as one of the indicators of hydrocarbon leaks from the subsoil. LANDRUM et al. (1989) and PATRICK et al. (2011) highlighted more emissive zones that could be related either to main faults or to secondary fractures acting as migration pathways.

However, most of the existing applications rely only on detecting either direct or indirect indicators. On the ground, not surprisingly, the geostatistical analysis deals with these sources of information separately. CIOTOLI et al. (2007) studied the spatial character of a soil gas monitoring system; OLIVIER & KHARYAT (1999), BUTTAFUOCO et al. (2007) dealt with the geostatistics of Rn-monitoring systems; MALDONATO & CAMPBELL (1992) or, recently, CASTILLO et al. (2015) used geostatistical tools in the spatial analysis of the hydrocarbon components of soil gases. Only a few papers have dealt with a geostatistically supported integrated interpretation of the Rn and hydrocarbon components of soil gases.

The aims of this study were to (1) determine the common areal distribution of the Rn and hydrocarbon components of soil gases in a large region above some known hydrocarbon reservoirs in the southern part of Hungary (Fig.1); (2) relate positive anomalies of Rn and hydrocarbon components of soil gases to the existing reservoirs, and (3) to suggest new targets for the surface hydrocarbon exploration on the basis of the results.

The studied area is located in the Southwestern part of Hungary, close to the Croatian border. In this region, six known hydrocarbon reservoirs occur, including from North to South: Berzence, Somogyudvarhely, Vízvár-Nord, Vízvár, Heresznye and Görgeteg-Babócsa (Fig. 1). They were developed in the Szolnok, Algyő and Újfalu Formations of the Miocene sequences. The studied area was also surveyed by 3D seismic survey, the results of which were used in the latter part of this paper to compare and validate the findings of the geostatistical analyses.
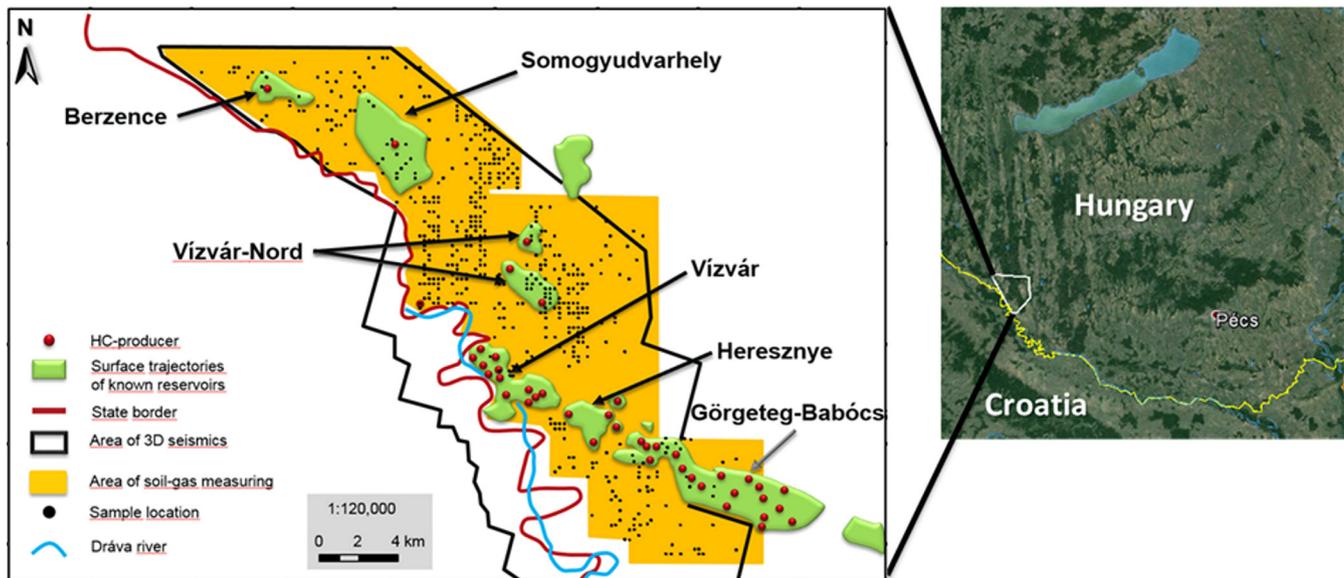
Figure 1. Study area.

The basic geochemical theory

The parent-elements of radon are supposed to migrate from the subsurface HC-reservoir. Essentially, they are ^{238}U, ^{230}Th and ^{226}Ra. Radon (Rn) can appear in two different decay series:

(1) ^{238}U $\rightarrow$ ^{226}Ra $\rightarrow$ ^{222}Rn

(2) ^{232}Th $\rightarrow$ ^{224}Ra $\rightarrow$ ^{220}Rn

Although radium (Ra) appears in both decay series, the two isotopes (^{222}Ra and ^{220}Ra) are not identical: their difference lies in their half-lives. The half live of ^{222}Ra (from ^{238}U) is 38 days, while it is only 55 seconds for ^{220}Ra (from ^{232}Th). When equal quantities of each isotope are available under the same physical environment, the average travel distance of ^{222}Ra is around 100 cm, whereas that of ^{220}Ra is only 1 cm. That is why ^{222}Ra is expected to make much more of a contribution to the gamma flux.

The literature explains that the geochemical background of radiometric signals measured above hydrocarbon reservoirs is very wide. Some authors (e.g. ARMSTRONG & HEEMSTRA, 1973, SAUNDERS et al., 1993) proposed that deep seated hydrocarbon pools provided the sources of radionuclides and their progeny. They thought that radionuclides were transported by formation fluids from the deep pool, along some hypothetical pathways, up to the surface. In their view, relatively high gamma emissions form the „halo", while low emissions exist within the boundary of the halo.

MORSE & ZINKE (1995) provided an elegant interpretation for the formation of the radiological „halo". Their theory was based on Laubmeyer's concept of vertical gas migration. In this process, with the passage of time, low molecular weight hydrocarbons migrating upward induce a chemically reducing environment in zones overlying the hydrocarbon (HC) accumulations. When oxidized uranium enters into this zone, its upward movement stops. Consequently, the uranium concentration becomes lower than it would have been with no hydrocarbon present. The uranium attempting to transit into the reducing zone will precipitate at the zone boundary. This precipitation appears as a halo as reported by PIRSON (1969).

The theories say that the radiation anomalies are not directly connected to the deep HC reservoirs. Rather, "geochemical cells", developed above the HC reservoirs, initialize such geochemical changes which cause the adsorption of the uranium content of the formation fluids. Radon, as the most mobile member of uranium-decay, can give effective information about these anomalies.

KRISTIANSSON & MALMQVIST (1982) explained the non-diffusive transport of Rn by carrier-gas transport. One of the most complete theories for the relationship between the subsurface reservoirs and the surface radiological measurements are given by TEDESCO (1995). In Tedesco's theory, a geochemical cell with the cloud of inverse electric charge develops between the (subsurface) reservoir and the surface. It initializes the migration of ionic geochemical constituents. It is very important to note that only the long-lived parent elements of Rn take part in this process. Radon is only a tracing element of this process, which can migrate via microseepage in a water-saturated permeable agent.

VÁRHEGYI et al. (2008) supposed that vertical migrations of formation fluid and gas can be connected to the subsurface HC accumulations which carry radionuclides from the formations between the reservoirs and the surface. In this case the source of radioactive elements is not the HC reservoir, but the rock bodies lying above the HC layers. The reservoir only initializes the following simplified transport mechanism: HC reservoir $\rightarrow$ geochemical cell $\rightarrow$ change of redox conditions near the surface $\rightarrow$ selective adsorption of the dissolved uranium at the redox-fronts $\rightarrow$ uranium-radium and radium-radon transformation $\rightarrow$ upward migration of radon. Figure 2. shows this combination of the hypothesized geochemical cell and the microseepage transport model of Rn migration (VÁRHEGYI et al., 2008). In this implementation the geochemical cell and the connected microseepage transport model of Rn assume high water saturation. The migration processes slows down above the surface of underground water. There is no chance of detecting Rn indication if the water surface is below 10 m (VÁRHEGYI et al., 2008).

In summary, it can be concluded that the radiation anomalies are not directly connected to the deep HC reservoirs. Rather, "geochemical cells" developed above the HC reservoirs initialize such geochemical changes which cause the adsorption of the uranium content onto the formation fluids. Radon, as the most mobile member of uranium-decay, can give effective information about these anomalies.
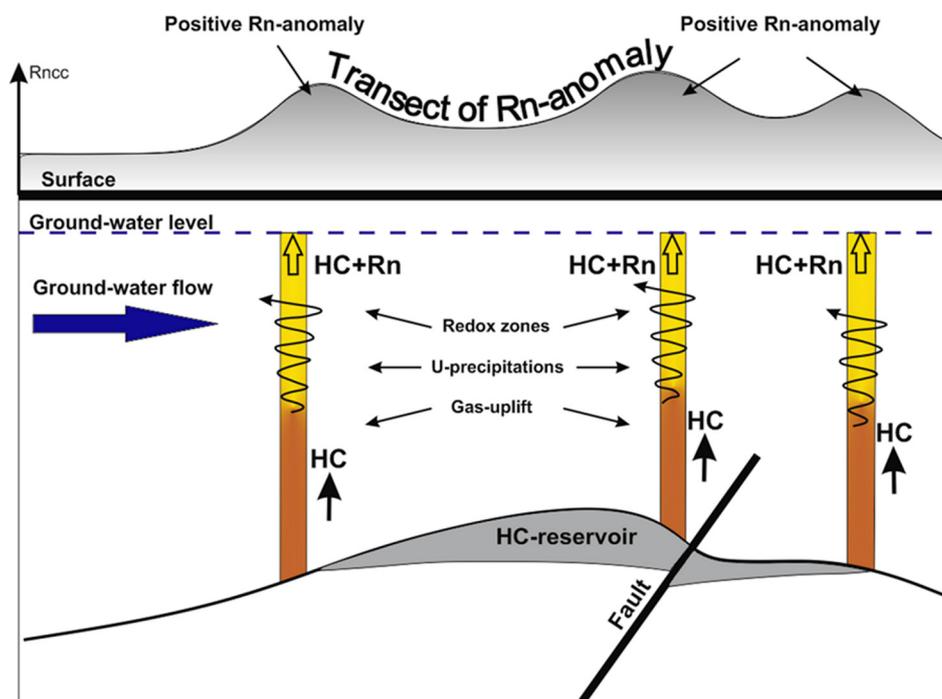
Geiger et al.: Statistical and geostatistical study of Rn and hydrocarbon components of a soil gas monitoring system ... **257**

Geologia Croatica

Figure 2. A geochemical model of the radiation anomaly above hydrocarbon reservoirs VÁRHEGYI et al., 2008.

The critical point in the applicability of the research-method based on radon detection is the differentiation of the deep origin and the background radiation. VÁRHEGYI et al. (2008) have developed a method for measuring this.

2. SOIL GAS AND RADIOLOGICAL MEASUREMENTS

2.1. Soil gas measurements

During the radon-monitoring field measurements, a specified mass (12 g on average) of soil was taken for the headspace gas method of soil gas analysis. The samples were placed into a glass container of 20 ml volume, and sealed by a rubber membrane. This glassware was then heated at 85 °C for 20 minutes prior to gas-chromatographic measurement. The headspace of the glass container was sampled by syringe through the membrane and the gas sample in the syringe was injected into a gas-chromatograph with an FID detector. With this apparatus, simple hydrocarbons C1, C2, C3, nC4, iC4, nC5 and iC5 (methane, ethane, propane, normal butane and isobutane, normal pentane and isopentane) were measured. To assess the reliability of the measurements, how the concentrations of hydrocarbon components decreased in "natural" and "artificial" soil samples were analyzed in the laboratory. From the measured hydrocarbon components (HC), only C2, C3, nC4 and nC5 are involved in the further statistical and geostatistical analysis.

2.2. Analysis of the ^{222}Rn concentration of soil gas

The samples were taken in a regular surface grid where the nodes were 250 m apart. The change of background Ra radiation was followed by the so called basic registrations process. In the monitoring locations, three other properties were measured and calculated besides the ^{222}Rn concentration. They were the uranium equivalent of the ^{222}Rn concentration (UeRn), the difference between ^{222}Rn and ^{226}Ra (Rn-Ra), and the so called soil-corrected ^{222}Rn (Rn-Ai) content.

The uranium equivalent of the measured Rn concentration (UeRn) was defined as the concentration of the parent element of Rn in the soil. It was directly measured using gamma-spectrometry. Under general conditions, the ^{226}Ra concentration of the soil environment is the origin of the ^{222}Rn concentration. However, migration processes can increase or decrease this "regular" ^{222}Rn concentration. When the migration adds Rn to the "regular" concentration, this Rn-Ra difference is positive, otherwise it is zero or negative. The local characteristics of soil geochemistry and its physical properties can significantly affect the Rn concentration. The effects of these features on the local Rn concentration can even be much larger than those of the migration processes. That is why correction of the measured concentrations to the soil-variability was of primary importance.

The samples were grouped in K sets with homogeneous soil characteristics, $S_k, k = 1, 2, \ldots, K$. Then everages and variances were calculated for each set, $S_k, k = 1, 2, \ldots, K$. Finally, $Rn - Ai_j^{(k)}$ was determined for each sample as a "standardized" value:

$$Rn - Ai_j^{(k)} = \frac{Rn_j^{(k)} - Mean^{(k)}}{\sigma^{(k)}} \qquad (1)$$

where Rn_j^k is the Rn content of the sample, $j^{th} \left(j = 1, 2, \ldots, J \right)$ in group $k^{th} \left(k = 1, 2 \ldots, K \right)$, $Mean^{(k)}$ is the average, and $\sigma^{(k)}$ is the variance of the Rn content in $k^{th} \left(k = 1, 2 \ldots, K \right)$ group. Note that $Rn - Ai_j^{(k)}$ can have positive and negative values.

3. FIELD DATA COLLECTION AND ITS DESCRIPTIVE STATISTICS

The geochemistry suggested that only the C2 (ethane), C3 (propane), nC4 (normal butane) and nC5 (normal pentane) components of the HC gases were worth keeping in the statistical-geostatistical modeling, since they could be expected to be of deep origin. All these components were expressed in ppb. In contrast, all the parameters measured or calculated in the radiological monitoring were involved in the analyses. They were as follows: Rn (radon content of soil gas, in kBq/m^3), UeRn (uranium equivalent

of the Rn content, in ppm), Rn-Ra (the difference between Rn and Ra components, in kBq/m^3), and Rn-Ai (Rn measurements corrected by the soil environments, dimensionless).

In the sampling process of soil gas and radon-monitoring, 2441 samples altogether were analyzed. However, among them there were only 538 in which all the parameters could have been measured (Fig. 1). They were the target of the geomathematical analyses. Table 1. shows their general statistical characteristics. Since the hydrocarbon components had very skewed probability distributions, their original values were transformed by a log10 function before calculation of the statistics in Table 1.

Table 1. Descriptive statistics of the variables analyzed.

Variables	Valid N	Mean	Median	Minimum	Maximum	Standard Deviation
Rn (kBq/m^3)	538	35.39	30.38	1.00	176.50	24.95
Rn-Ra (kBq/m^3)	538	−0.88	−5.34	−81.76	145.30	23.13
Rn-Ai (dimensionless)	538	0.02	0.11	−3.86	2.89	1.00
Log(C2) (ppb)	538	0.90	0.90	−0.67	2.68	0.41
Log(C3) (ppb)	538	0.26	0.24	−0.74	2.56	0.35
Log(nC4) (ppb)	538	−0.21	−0.24	−1.04	2.78	0.31
Log(nC5) (ppb)	538	0.07	−0.15	−0.95	1.98	0.37

The minimum value of Rn-Ra, in Table 1, has a negative sign indicating that there were sampling locations from where the Rn migrated into the surroundings. However, the high positive value of the maximum Rn-Ra property indicated that there were sampling sites where the migrated Rn content were added to those generated "in situ" (Table 1).

The correlation coefficients between the radon and hydrocarbon components in Table 2 (coloured light grey) were significant, but the deterministic coefficients (r^2) of the highest correlations between the Rn and HC components were, at best, 0.04.

Table 2. Correlation coefficients between the variables.

Variables	Correlations (N=538)						
	Rn	Rn-Ra	Rn-Ai	Log(C2)	Log(C3)	Log(nC4)	Log(nC5)
Rn	1.00	0.77	0.78	−0.17	−0.12	0.02	−0.01
Rn-Ra		1.00	0.73	−0.01	−0.03	0.04	0.02
Rn-Ai			1.00	−0.02	−0.06	0.01	0.05
Log(C2)				1.00	0.75	0.47	0.46
Log(C3)					1.00	0.63	0.55
Log(nC4)						1.00	0.49
Log(nC5)							1.00

These results were the consequences of very heterogeneous soil types, and the temperature and humidity differences between the sampling periods. This fact (among others) had three important consequences: (1) it was not possible to make a direct comparison of Rn and HC components; (2) instead of forcing the analysis of their pairwise relations, the pattern of their inter-correlations was targeted in the analysis; (3) quite high uncertainty could be expected in the measured data.

Point (2) meant that a factor analysis as a data integration tool could be expected to reveal the connections between the HC and Rn components (JÖRESKOG et al., 1976; OLIVIER & KHARYAT, 1999). Point (3) suggested the use of a stochastic

simulation in the analysis of the spatial distribution. This approach does give the best estimation, but tends to honour the statistical characteristics while emphasizing the small scale heterogeneities (MUCSI et al., 2013).

3.1. Methods used in the statistical and geostatistical modeling

The workflow of statistical and geostatistical analysis shows the steps in which field and surface seismic data are integrated and their spatial connections revealed (Fig. 3).

The statistical and geostatistical analyses were based on three principles. (1) The factors were regarded as compact summarizations of the background processes affecting the Rn-and HC-components. Because factors represented continuous spatial processes in the joint migration of Rn and hydrocarbon gases, they could be regarded as continuous spatial variables and analyzed by geostatistical tools. (2) Because of the scattered sampling, the spatial uncertainty of the geostatistical model was also important to study. (3) The final goal was to indicate the probability of those regions where the co-variability of Rn-and HC components could suggest HC structures at depth.

3.2. Factor analysis

Factor analyses are very effective tools of linear data integration. The main applications of factor analytical techniques are to reduce the number of variables and to detect structure in the relationships between variables, that is, to reduce data dimensionality. Here, we used this technique in the latter sense, as a structure detection method. In this so called Exploratory Factor Analysis (EFA) we tried to uncover complex patterns by exploring the da-
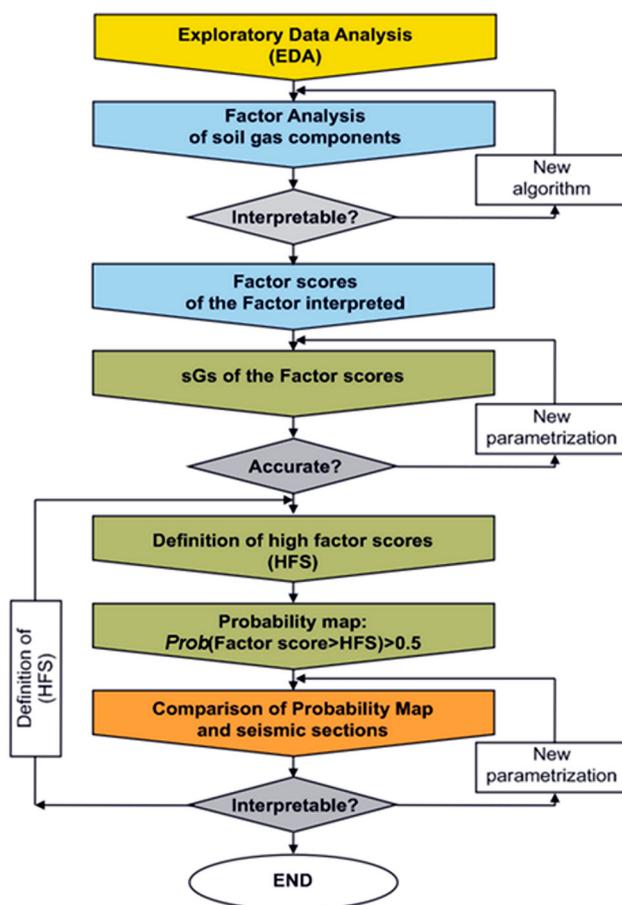

Figure 3. Workflow of analysis strategy.

Geiger et al.: Statistical and geostatistical study of Rn and hydrocarbon components of a soil gas monitoring system ... **259**

Geologia Croatica

taset and testing predictions (CHILD, 2006). EFA is used when the aim is to discover the number of factors influencing variables and to analyze which variables 'go together'. A basic hypothesis of EFA is that there are m common 'latent' factors $(F_1, F_2, \ldots, F_m)$ to be discovered in the dataset, and the goal is to find the smallest number of common factors, p, that will account for the correlations of the variables, $X_1, X_2, \ldots, X_p$ (MCDONALD, 1985).

The model assumes that each observed variable is a linear function of these factors together with a residual variance. This model intends to reproduce the maximum correlations.

$$X_j = \sum_{i=1}^{m} a_{ji} \cdot F_i + e_j \qquad (2)$$

where $a_{ji} (i = 1, 2, \ldots, m)$ are the factor loadings, and e_j is the specific or unique factor.

Generally, a factor analysis decomposes the modified correlation matrix. The modification is that the original diagonal elements (they equal to 1) of the correlation matrix are replaced by the prior communality estimates. It means that this technique does not decompose the whole variance of each variable since it may also have error variance. The estimated proportion of variance of the variable (communality) is assumed to be free of error variance and is shared with other variables in the matrix. It is the variance of a variable in common with all others together (communalities are smaller than 1). In this way factor analysis gives enough room to take measurement error and measuring uncertainty into consideration.

One of the most convenient ways to interpret the results of factor analysis is the interpretation of factor loadings. The factor loadings are the correlation coefficients between the variables and factors. They represent to what extent a variable is explained by a factor. Analogous to Pearson's r, the squared factor loading is the percent of variance in the particular indicator variable explained by the factor. Loadings can range from (-1) to 1. Loadings close to (-1) or 1 indicate that the factor strongly affects the variable. In the interpretation of factor loadings the first important decision is to identify the high loadings. There is a rule of thumb saying that loadings should be 0.7 or higher. The rationale of this statement is that the 0.7 level corresponds to about one half $(0.7^2 = 0.49)$ of the variance in the indicator being explained by the factor. However, some researchers, particularly for exploratory purposes, use a lower level such as 0.4 for the central factor and 0.25 for other factors. In any event, factor loadings must be interpreted in the light of theory, not by arbitrary cutoff levels (STEVENS, 2002). Factor analysis can also be viewed as a transformation of the original sample space (defined by the original variables) into an orthogonal space defined by the factors. The coordinates of this factor space are called factor scores. Computing factor scores facilitates the search for factor outliers. Also, factor scores may be used as variables in subsequent modeling.

Here the main goal was to reveal such a factor in which both the Rn and all the hydrocarbon components appear with high positive factor loadings. After several trials a special factoring algorithm was discovered in which these requirements have been realized. In this algorithm the communality was approached by the multiple correlation coefficients, and we used the centroid method (THURSTONE, 1931) to extract the factors. The eigen values of the extracted factors were larger than 1. The rationale of this criterion was that the variance of the defined factor should be at least as large as that of each standardized variable. The cutoff for the definition of a high factor score was 0.5. For each sample the factor scores were also calculated. Factor scores represent the amount of the effect of the actual factor exercised in each object (sample). Consequently, the areal distribution of factor scores could be expected to show the areal variability of the 'hidden' process controlling the joint migration of Rn and the hydrocarbon components of soil gases.

3.3. Geostatistical tools

3.3.1. Variography

In the geostatistical literature, spatial patterns are usually described in terms of the dissimilarity (instead of similarity) between observations as a function of the separation distance. The average dissimilarity between data separated by a vector h is measured by the experimental semivariogram $\hat{\gamma}(h)$, which is computed as half the average of the squared difference between the components of every data pair:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(u_i) - z(u_i + h) \right]^2 \qquad (3)$$

where $N(h)$ is the number of data pairs for a given distance.

The main purpose of the variogram calculation is partly to analyze the directional heterogeneity, and partly to transfer this information to kriging or simulation procedures. Therefore, a continuous function must be fitted to the values of the experimental variograms so as to deduce semivariogram values for any possible lag h required by prediction algorithms and also to smooth out sample fluctuations. The difficulty is that only conditionally negative definite functions can be considered as semivariogram models, in order to ensure the non-negativity of the variance of the prediction error. In practice, only a few models are known to be permissible (e.g. DEUTSCH & JOURNEL, 1998, DEUTSCH 2002 or GOOVAERTS, 1997).

Typically, two or more permissible models must be combined to fit the shape of the experimental semivariogram. Combinations of permissible models are permissible as long as the contribution of each basic model is positive, that is the nested model is written as:

$$\gamma(h) = \sum_{i=0}^{L} b^i g_i(h) \quad with \quad b^i \geq 0 \qquad (4)$$

where b^i is the positive sill or slope of the corresponding basic semivariogram model $g_i(h)$.

The variogram map (sometimes called variogram surface) takes the idea of calculating the variogram in a number of directions. It can be created by posting the variogram values on a map, where the center of the map is the lag distance of zero (e.g. ISAAKS & SRIVASTAVA, 1989 or DEUTSCH, 2002).

The way in which the permissible models are chosen and their parameters (range, sill) are estimated is still controversial (MCBRATNEY & WEBSTER, 1986; GOOVAERTS, 1997). There are several ways to check the goodness-of-fit of the permissible model. It can be done per view, or using some criteria based on minimizing the least square differences between the experimental and fitted model. However, the tool of variogram map can also be applied for this purpose. By comparison of the variogram map coming from original data and that of variogram models, the goodness of the variogram models can be evaluated.

3.3.2. Cell declustering

The sample points in this study show neither regular nor random spatial distribution over the region (Fig. 1). This kind of 'sampling' is said to be preferential. This situation means that the available samples are quite far from statistical representativity. Consequently, neither the summary statistics of the data set nor

the shape of the histograms of the measured attributes may be representative for the entire volume of interest.

Although most contouring or mapping algorithms adjust this preferential clustering by default, simulation algorithms fail in the process. These methods rely on the "intrinsic" assumption of the sample distribution being representative of the entire volume of interest (DEUTSCH, 2002). Since the "very-leaky" spatial arrangements needed the application of stochastic simulations, the situation amounts for data-"declustering". There are two commonly used solutions: the polygonal method (ISAAKS & SRIVASTAVA, 1989; GOOVAERTS, 1997; GEIGER, 2012) and the so called cell-declustering approach (JOURNEL, 1983; DEUTSCH, 1989; GOOVAERTS, 1997; GEIGER, 2012). In this study we used the latter method.

The cell-declustering approach calls for definition of a grid system over the A area of interest and counting the number B of cells that contain at least one datum and the number n_b of data falling within each cell b. Then the weights are defined as:

$$w_j = \frac{1}{B \cdot n_b}, \; j = 1, 2, \cdots, n \qquad (5)$$

These weights are inversely proportional to the number of data in each cell. Hence, these weights give more importance to isolated locations. A weight greater than 1 implies that the sample space is being over weighted, and a weight less than 1 shows that the corresponding location is being under weighted. In the latter case it is clustered with other locations.

In the implementations the "best" cell size for declustering the sample points can be recognized by plotting the declustered mean versus cell size for a range of cell sizes. From this plot the size with the lowest declustered mean should be chosen when data are clustered in high-valued areas or the size with largest declustered mean in the inverse situation (DEUTSCH, 1989).

3.3.3. Sequential Gaussian Simulation

Consider the simulation of the continuous attribute z at N grid nodes u_i' conditional to the data set $\{z(u_i), i = 1, 2, \ldots, n\}$. Sequential simulations (JOURNEL & ALABERT 1988; ISAAKS, 1990; GÓMEZ-HERNÁNDEZ & SRIVASTAVA, 1990; MALVIC, 2008) amount to modeling the conditional cumulative distribution function (ccdf):

$$F\left(u_i; z|n(u_i)\right) = Prob\left\{Z(u_i) \le z|n(u_i)\right\}, \qquad (6)$$

then sampling it at each of the grid nodes visited along a random sequence. To ensure reproduction of the z-semivariogram model, each ccdf is made conditional not only to the original n data but also to all values simulated at previously visited locations. Other realizations are obtained by repeating the entire sequential drawing process. Sequential Gaussian simulation (sGs) typically involves a prior transform of the z-data into normal score data. The simulation is then performed in the normal space and the simulated normal scores are back-transformed in order to identify the original z-histogram.

3.3.4. Goodness of the probabilistic model (simulation)

The goodness of a probabilistic model may be checked by its accuracy and precision. In general, accuracy refers to the ultimate excellence of the data or computed results, e.g. conformity to truth or to the standard. Precision refers to the repeatability or refinement (significant figures) of a measurement or computed result.

DEUTSCH (1997) proposed specific definitions for these terms. In his view accuracy and precision were based on the actual fraction of true values falling within symmetric probability

intervals of varying with p. A probability distribution is accurate if the fraction of true values falling in the p interval exceeds p for all p in [0,1]. The precision of an accurate probability distribution is measured by the closeness of the fraction of true values to p for all p in $[0,1]$.

Stochastic simulation leads to L stochastic realizations at each data location. These realizations provide a model for the conditional cumulative distribution function (ccdf):

$$F\left(u_i; z|n(u_i)\right) = Prob\left\{Z(u_i) \le z|n(u_i)\right\} \qquad (7)$$

where $n(u_i)$ is the set of n data at location u_i. These local models may be derived from a set of L realizations, calculated from indicator kriging or defined by the Gaussian mean, variance and transformation.

The probabilities associated with the true values $z(u_i), i = 1, 2, \ldots, n$ are calculated from the previous ccdf as $F\left(u_i; z|n(u_i)\right)$.

Consider a range of symmetric p-probability intervals (*PIs*). The symmetric $p - PI$ is defined by the corresponding lower and upper probability values:

$$p_{low} = \frac{1-p}{2} \; and \; p_{upp} = \frac{1+p}{2} \qquad (8).$$

Next, define an indicator function $\xi(u_i; p)$ at each location u_i as:

$$\xi(u_i; p) = \begin{cases} 1, & if \; F\left(u_i; z|n(u_i)\right) \in \left(p_{low}, p_{up}\right) \\ 0, & otherwise \end{cases} \qquad (9)$$

The average of $\xi(u_i; p)$ over the n location u_i:

$$\overline{\xi(p)} = \frac{1}{n} \sum_{i=1}^{n} \xi(u_i; p) \qquad (10)$$

is the proportion of locations where the true value falls within the symmetric $p - PI$.

According to the earlier definition of accuracy, the simulation algorithm used to generate the ccdfs is accurate when $\overline{\xi(p)} \ge p, \forall p$. A grapcal way to check the assessment of accuracy is to plot $\overline{\xi(p)}$ versus p and to see if all of points fall above or on the 45° line. This plot is referred to as an accuracy plot (DEUTSCH, 1997).

3.3.5. Analyses of uncertainty

The advantage of stochastic simulation over interpolation (kriging) is that it allows the reproduction of statistics (histogram, semivariogram, scattergram) inferred from the data, hence the model or realization looks more "realistic" than a smooth estimated map. Also, one can generate multiple realizations that all reasonably match the same sample statistics and exactly identify the conditioning data. The set of alternative realizations provides a visual and quantitative measure (actually a model) of spatial uncertainty. Spatial features are deemed certain if seen on most of the simulated maps. Conversely, a feature is deemed uncertain if seen only on a few simulated maps. Simulation can be accomplished using a growing variety of techniques, which differ in the underlying random function model (multiGaussian or nonparametric), the amount and type of information that can be accounted for, and the computer requirements (GOTWAY & RUTHERFORD, 1994; SRIVASTAVA, 1996; DEUTSCH & JOURNEL, 1998).

According to SRIVASTAVA (1996), geostatistical realizations can be used for three main purposes: (1) assessing the im-

pact of uncertainty; (2) reproducing the spatial variation; (3) performing Monte-Carlo risk analysis. The concept of space of uncertainty, and the associated issue of equiprobability of realization were discussed in detail in the 1990s (e.g. JOURNEL, 1994; SRIVASTAVA, 1994). Some authors believed that the space of uncertainty must have been theoretically defined outside the use of a particular algorithm. Others stated that the space of uncertainty could only be defined through the algorithm and consists of all possible realizations that could be generated by the algorithm.

There is currently no theory that allows us to determine the set of all possible outcomes if fairly sampled. The characterization of the space of uncertainty is rendered difficult by the fact that only a limited number of realizations is usually generated.

The uncertainty can be quantified by measures such as entropy, interquartile range, or variance. In this paper, the variance was preferred because of its simplicity and wide acceptance as a measure of spread. The uncertainty of a probabilistic model was defined as the average conditional variance of all locations in the area of interest:

$$U = \frac{1}{N} \sum_{i=1}^{N} \sigma^2 (u_i) \qquad (11)$$

where there are N locations of interest u_i, $i = 1, 2, \ldots, N$ with each variance $\sigma^2 (u_i)$ calculated from the local ccdf $F(u_i; z|n(u_i))$.

3.3.6. Probability maps

SGS results in the generation of conditional cumulative probability distributions (ccdf) at each grid node via the calculation of stochastic realizations. Hence, calculation of any probability in the sense of (7) is quite straightforward at grid nodes if we have large enough number of realizations. The 100 equally probable values simulated in grid nodes provided the possibility of this derivation. The accepted geochemical model said that the migration process of interest could be made probable in the region of "high" factor scores. If the cut-off of these high factor scores were known, the probability belonging to this cut-off in each grid node would characterize the possibility of the migration process in grid nodes. This thought needed two practical steps: (1) finding a cut-off to define "high" scores of the factor describing the migration process; (2) contouring the probability of this cut-off.

For the definition of cut-off of "high" scores, $score_{cut-off}$, two approaches were considered. The first was the box and whisker plot technique (TUKEY, 1977), the second being the $(mean + \sigma)$ approach, where σ was the standard deviation. The latter provides an inflection point of the normal distribution. From these two definitions the latter was used in further analysis, since this gave a larger number of regions on the map to be compared with the seismic data. After choosing the proper cut-off, we calculated the

$$Prob\big(score_{cut-off} \leq Scores\ of\ Factor_i\big) \qquad (12)$$

probabilities in each grid node from the 100 simulated values. In this probability map, regions outlined by high probabilities were expected to show the surface projections of the subsurface.

3.3.7. Comparing geostatistical results and 3D seismic

The 3D seismic covers almost the whole sampling area. In the seismic data sets, analysis of average amplitude and the interpretation of some seismic attributes have occurred. As a result, all the known subsurface reservoirs and the most important faults have been identified. Three sections were cut from the 3D seismic data set. Their traces are shown in Fig. 4. Along these poly-
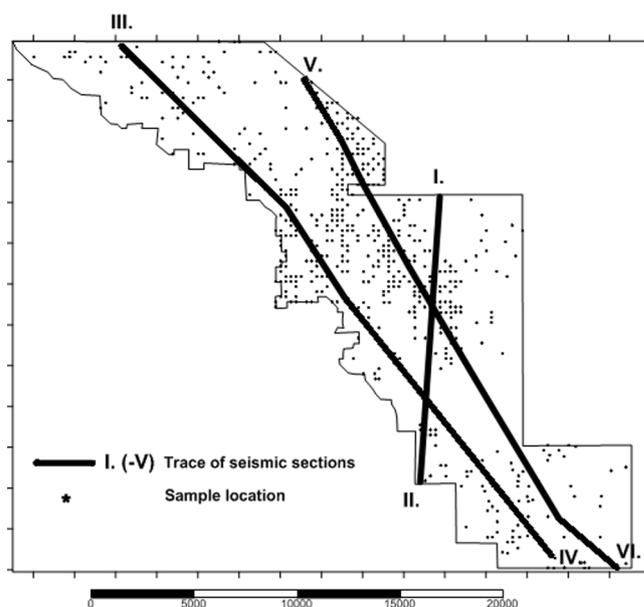

Figure 4. The five traces of seismic sections.

lines three transects were created from the probability map. These transects were compared with the corresponding seismic sections. This facilitated the checking of whether the structures appearing in the probability map coincided with the subsurface reservoirs or that such a fault system could be a potential pathway for the common geochemical migration of hydrocarbon and parent elements of Rn.

4. RESULTS

The result of factor analysis showed only two factors meeting the initial criterion (let the eigenvalue be larger than 1). The two (unrotated) factors shown in Table 3. described 65% of the total variance. They defined almost equal proportions of the total variance: 31.5% and 34.1% respectively.

The first factor positively correlated all the other variables (Table 3, Factor 1). Therefore we supposed it reflected a complex process increasing both the HC and the Rn components of soil gases. Consequently this factor was called a "Factor of migration", which was exactly what we were looking for. Its low contribution to the total variance could be explained by the high dependence of Rn components on the very variable physical environment: humidity, soil type, season of measurements, etc. On the basis of the applied geochemical model this factor was expected to outline regions, surrounding the surface projections

Table 3. The result of factor analysis (high factor loadings are colored by light grey).

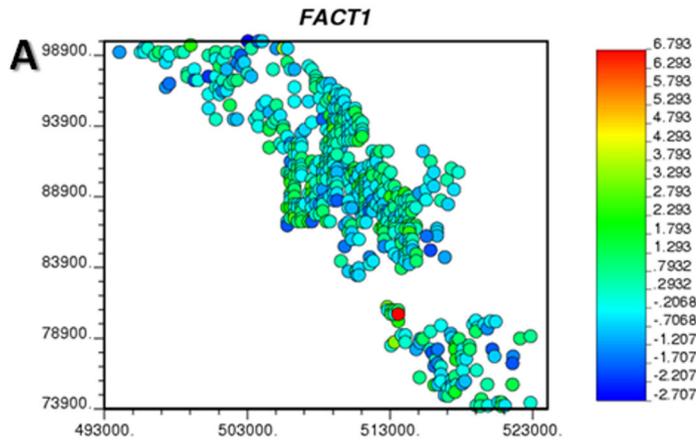| Variables	Unrotated Factor Loadings Extraction: Principal factors (Centroid)	
	FACTOR 1	FACTOR 2
Rn	0.536540	−0.736023
Rn-Ra	0.570151	−0.623170
Rn-Ai	0.571089	−0.643253
Log(C2)	0.526628	0.540717
Log(C3)	0.659843	0.652462
Log(nC4)	0.545612	0.420220
Log(nC5)	0.501242	0.389047
Proportion of Total Variance	0.314396	0.341410

Figure 5. Lateral distribution of Factor 1 scores (A), frequency distribution of declustered Factor 1 scores (B), normal scores of declustered Factor 1 scores (C).
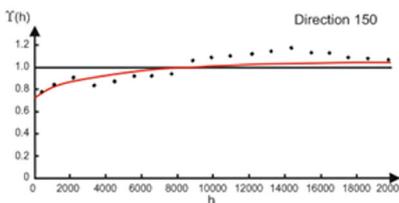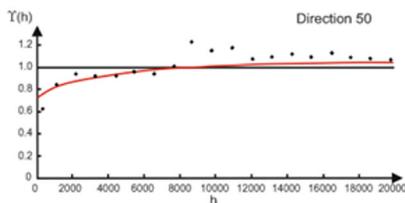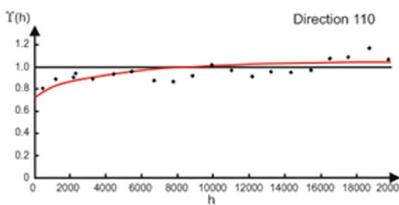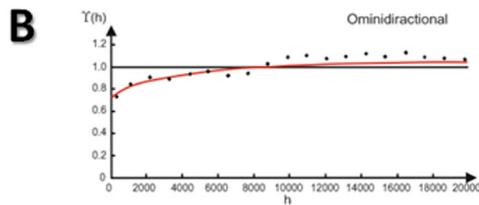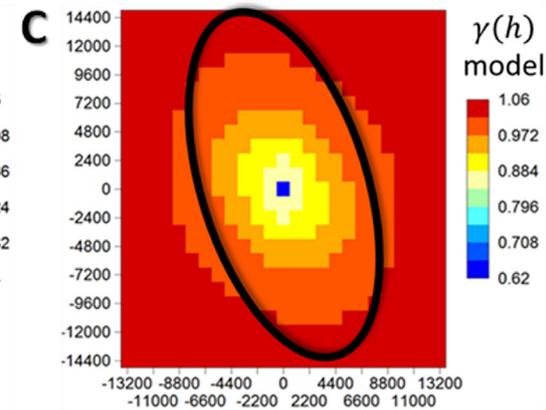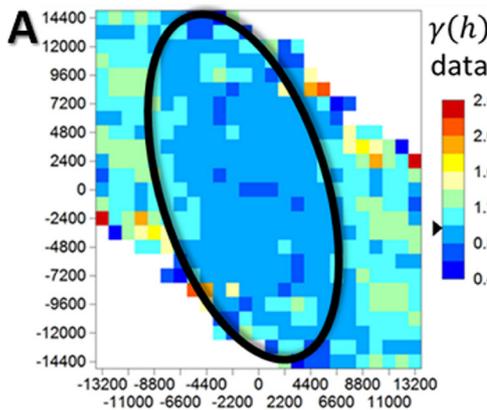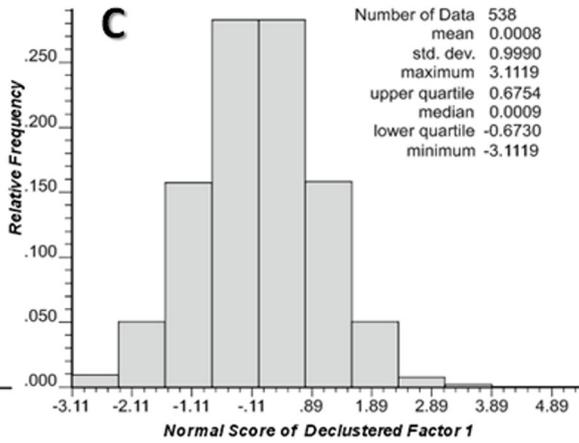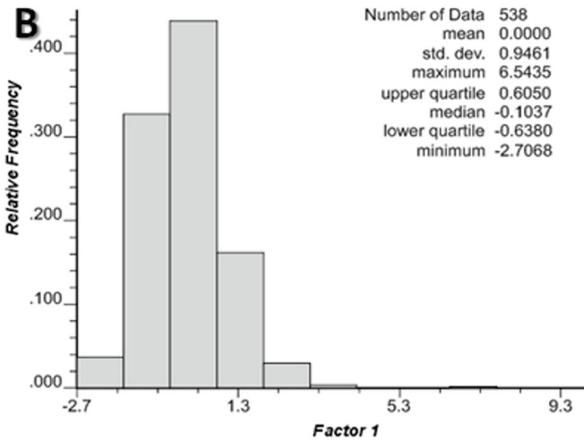

Figure 6. Empirical variogram map of the normal scores of Factor 1 scores (A), Variogram models of the normal scores of Factor 1 (B), Variogram map of the variogram model of (C).

of the existing subsurface reservoirs ("Positive radon anomaly" in Fig. 2).

The second factor showed a high positive correlation with all the Rn and high negative correlations with hydrocarbon components, which was inconsistent with the assumed geochemical process. Even the information content of this factor was very low: it was able to explain only 3% of the total variability by itself (Table 3, lowermost row) and is why Factor 2 was ignored at the next stage of modelling.

The lateral distribution of Factor 1 scores appeared with high positive values in the southern part of the studied area (Fig. 5, A). The clustered appearance of the sample locations required a cell declustering approach to get more representative descriptive statistics of the frequency distribution. The frequency distribution derived from the declustered data is skewed toward the high positive figures (Fig. 5, B). The normal score transformation of original declustered Factor 1 scores is shown in Fig. 5, C.

The variogram maps calculated for the normal score transformations of the Factor 1 scores showed a definite NE-SW continuity direction (Fig. 6 A). The empirical continuity expressed by the variogram map was modeled by a nested theoretical variogram model:

$$\gamma(h) = 0.63 + 0.26 \cdot Exp.2000(h) + 0.14 \cdot Gauss.1400(h) + \\ + 0.03 \cdot Gauss.2400(h) \qquad (13)$$

where the directions and the anisotropy ratios were 118°, 40°, 35°, and 0.76, 40 and 50, respectively.

In this model, the short small scale heterogeneity had 1400 m of range, while 2400 m of range belonged to the large scale heterogeneity. The nugget effect was very large (0.63 in (13)) expressing that about 60% of the total spatial variability cannot be characterized by any linear geostatistical model.

Variogram maps calculated from the models honoured the main anisotropic-characters of Factor 1 (Fig. 6, C). It can also be seen that the models did not cover the whole of the initial heterogeneity. However, the comparison of model and empirical variograms proved that this loss of information was beyond the range (Fig. 6 A, C).

Using the variogram model and declustering weights, 200 equally probable stochastic realizations were generated by sequential Gaussian simulation. The simulated normal scores were back-transformed to the original scale of the Factor 1 score. We
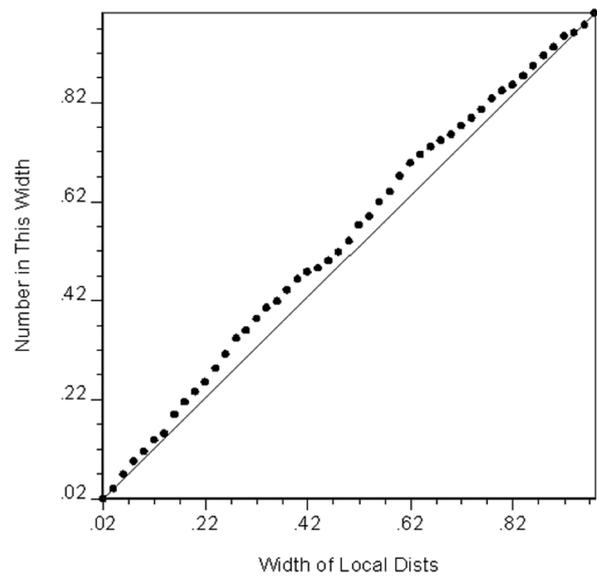

Figure 7. Accuracy plot of the simulation.

checked the simulation results by the accuracy plot (Fig. 7) and a visual comparison of the distributions of original and simulated scores (Fig. 8). Since points in the accuracy plot of the simulation results lie above the 45° line, the result of the simulation was accurate, but not exact (Fig. 7).

Figure 8. shows both the original and simulated Factor 1 scores. The spatial dependence of the Factor 1 scores did not allow the use of any statistical test. By visual examination we could say that the simulation decreased the minimum of the input data set. Fortunately, this did not affect the acceptance of the simulation, since we were looking for the high positive and not for the low negative factor scores.

In grid nodes, the E-type estimation for Factor 1, was calculated from the backtransformed normal scores (Fig. 9).

The grid-average of the Factor 1 scores and their variance were −0.072 and 0.192, respectively. The 'high' scores were defined as those being larger than the expected value plus variance. In fact, this definition produced a larger set than that of the Boxplot (Fig. 10). Consequently it resulted in higher possibilities in the lateral outlining of the area with high Factor 1 scores.

These laterally continuous regions of high Factor 1 scores were the targets which we were looking for.
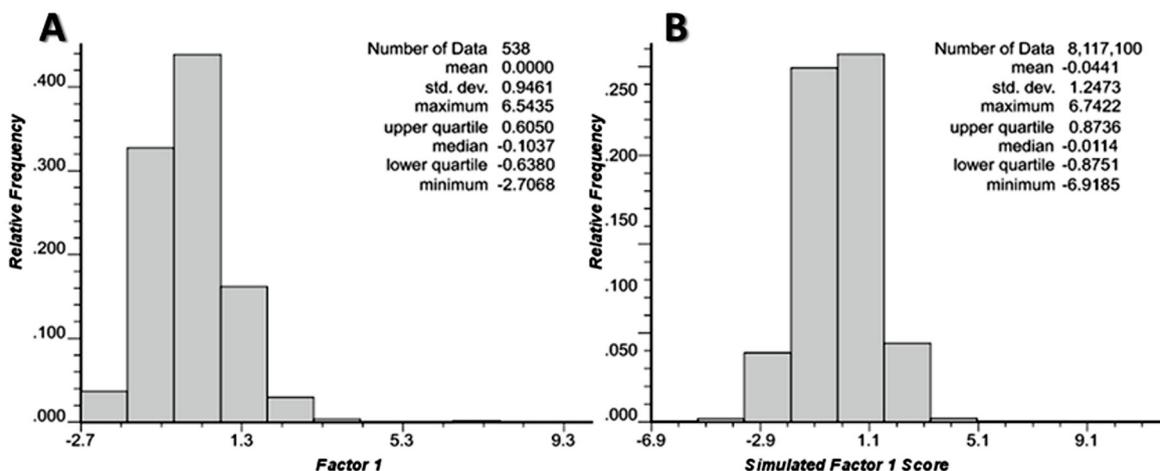

Figure 8. Comparisons of original (A) and simulated (B) Factor 1 scores.
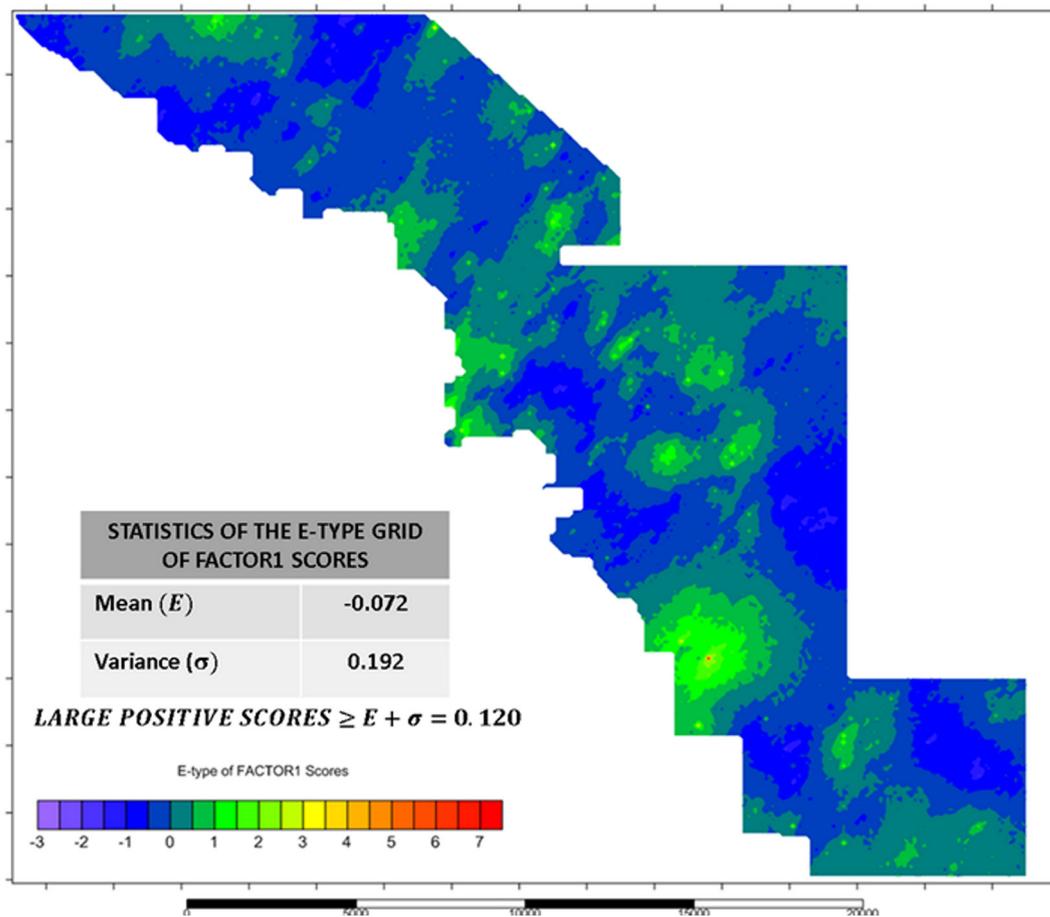
Figure 9. E-type estimation of Factor 1 scores from 200 stochastic realizations.

The map of conditional variance was used to characterize the uncertainty of the stochastic model from which the E-type estimation was derived (Fig. 11, A). This map suggested that although the general uncertainty was low in most of the area, there was a very definite subregion with high uncertainty. Fig. 11B shows the stack map of the E-type estimation (3D surface in Fig. 11B) and the contours of conditional variance (contour map above the 3D surface in Fig. 11B). This processing clearly shows that the highest uncertainty coincided with a subregion of high Factor 1 scores.
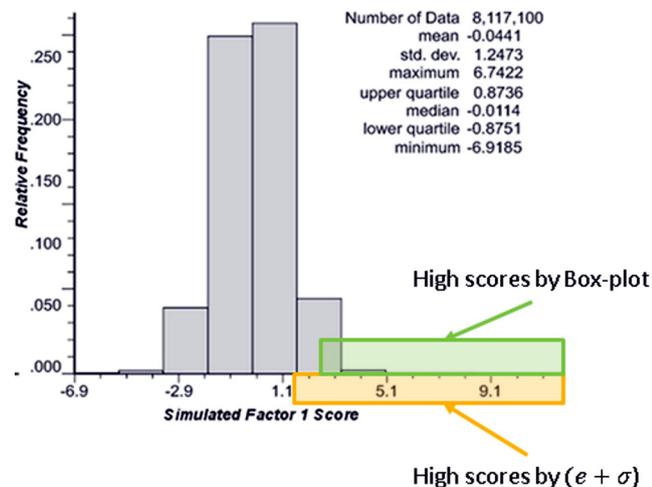

Figure 10. Comparison of the sets of "high Factor 1 scores defined by the Box-plot and $(e + \sigma)$ techniques. e is the expected value of the distribution; σ is the standard deviation of the distribution.

The probability map showing the probability of locations where Factor 1 scores were larger than the lower boundary of the high scores, 0.120 (mean+variance) was derived from the stochastic realizations (Fig. 12, A). On this map, at least six larger continuous subregions of high probabilities could be easily recognized. These were expected to reflect the subsurface hydrocarbon reservoirs. Three transects were cut out from the probability grid along the surface projections of the seismic sections (Fig. 12, B,C and D).

Along the first transect (I-II in Fig. 13) we identified an almost typical halo geometry of high Factor 1 scores (Fig. 13, I). Figure 12, II demonstrates the seismic section along this transect. The lateral extension of Vízvár and Vízvár-North fields could be identified in the probability transect. The southern anomaly of Factor 1 scores (Heresznye-anomaly in Fig. 13) appeared with more than 0.7 probability and matched with the subsurface position of the Vízvár Field. The northern anomaly was a typical halo-anomaly corresponding to the subsurface location of the Vízvár-North Field.

The second transect (III to IV in Fig. 12) went through three characteristic anomalies of Factor 1 scores appearing with more than 0.7 probability. They were, from NW to SE, the Berzence, Somogyudvarhely and Heresznye anomalies (Fig. 14). Among them Somogyudvarhely and Heresznye developed above the Vízvár hydrocarbon field, but there was no known equivalent hydrocarbon accumulation below the Berzence anomaly (Fig. 14).

The third transect also crossed three spots of high Factor 1 scores: Somogyudvarhely, Vízvár-North and the Komlósd anomalies. The probability of their appearance was more than 0.7 (Fig. 15). The subsurface counterparts of Vízvár-North and Komlósd were
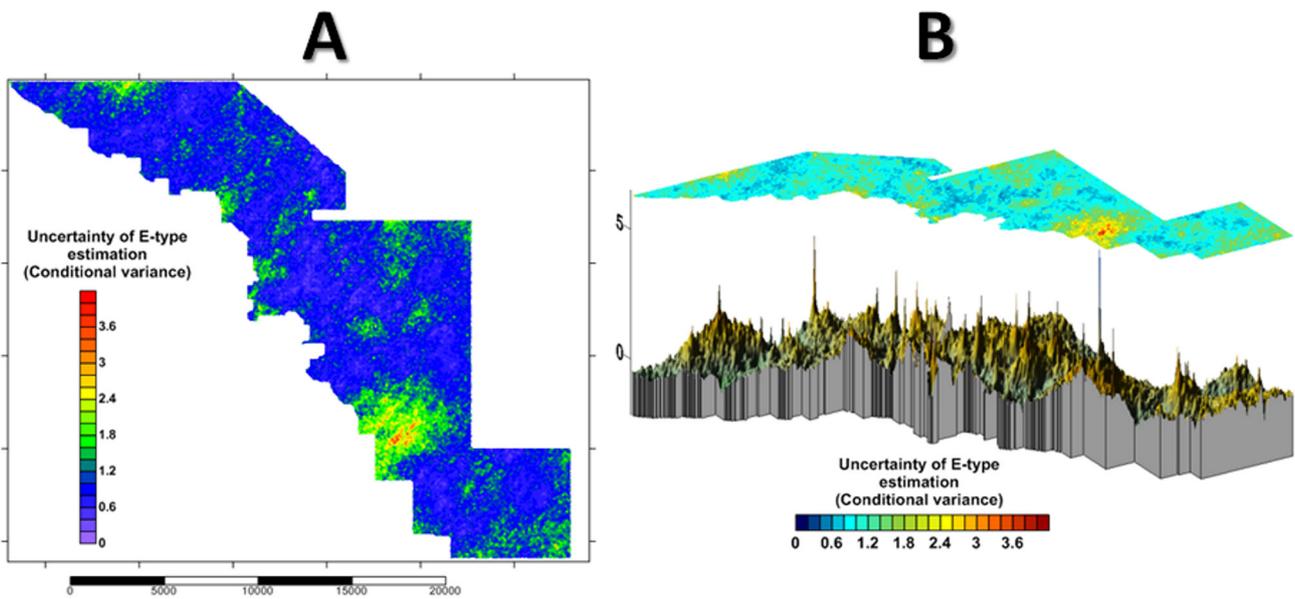
Geiger et al.: Statistical and geostatistical study of Rn and hydrocarbon components of a soil gas monitoring system ... **265**

Geologia Croatica

Figure 11. A: The uncertainty of the stochastic model by which the E-type estimation was derived (A); B: Stacked map of the E-type estimation (3D surface) and the conditional variance (contour map above the 3D surface).
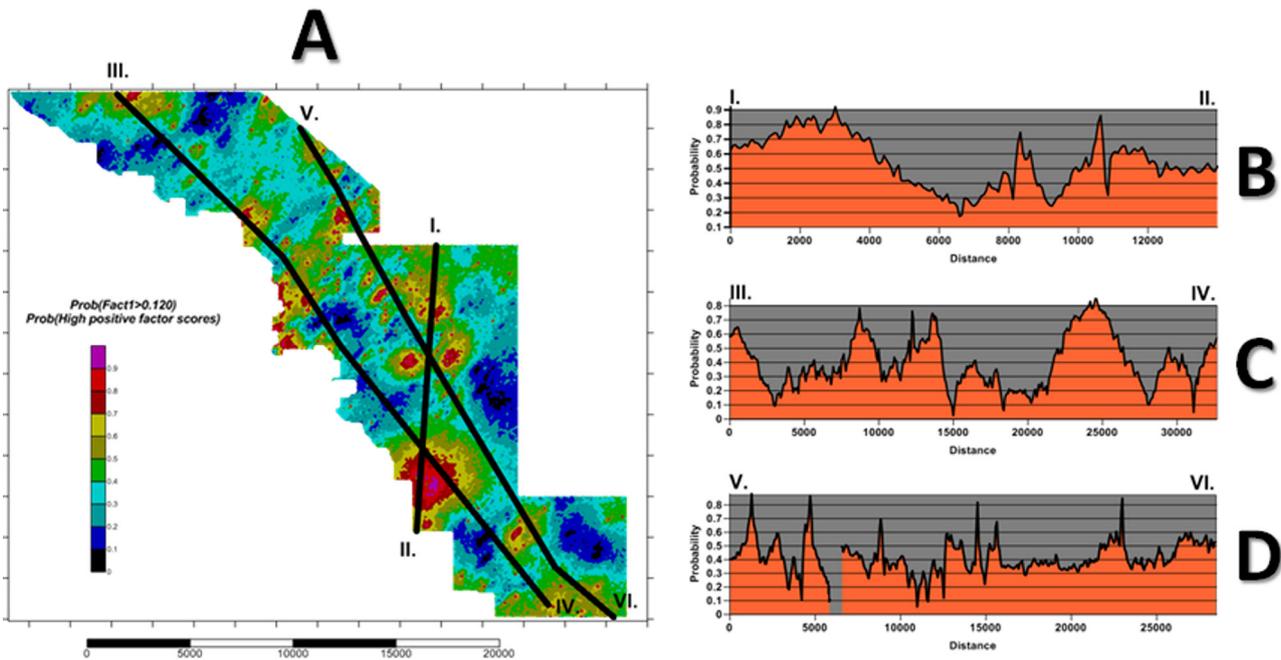

Figure 12. Map showing the lateral distribution of *Prob*(*Factor* 1 > 0.120). A: probability contours with the surface projections of seismic sections (I to VI). B, C and D are the transects created along the surface polygons of the seismic sections.

the Vízvár-North and Görgeteg-Babócsa Fields. The Miocene fault appearing below the Somogyudvarhely anomaly was regarded as being the genetic background of this anomaly.

5. DISCUSSION AND CONCLUSION

The presented statistical and geostatistical analysis of the Rn and HC components of soil gas samples was based on two expectations. Firstly, that the process controlling the common migration of these components could be recognized by a particular factor in the factor matrix of the studied components. The consequence of this hypothesis was that the strength of this process could be characterized by the scores of this factor.

Secondly, that the regional appearance of this migration process could be analysed by the tools of stochastic simulations.

The choice of stochastic simulation was supported by two facts: (1) after filtering out some 'uncertain' results of the HC gas analyses, the remaining samples showed a very scattered pattern; (2) the factor scores were not direct measurements of the migration process, consequently we had to use probabilistic approaches.

The basis of the factor analysis approach was the lack of any strong correlation between the HC and Rn components. In fact, the application of factor analysis did not follow the traditional way, in which the first few factors are used in the interpretation. In this study we searched for "the" factor correlating with both the Rn and CH components. In this strategy we accept, that the radiological and HC properties of soil gases could be affected by several processes. It was also admitted that the targeted phenomenon (factor expressing the joint migration of Rn and HC gases)
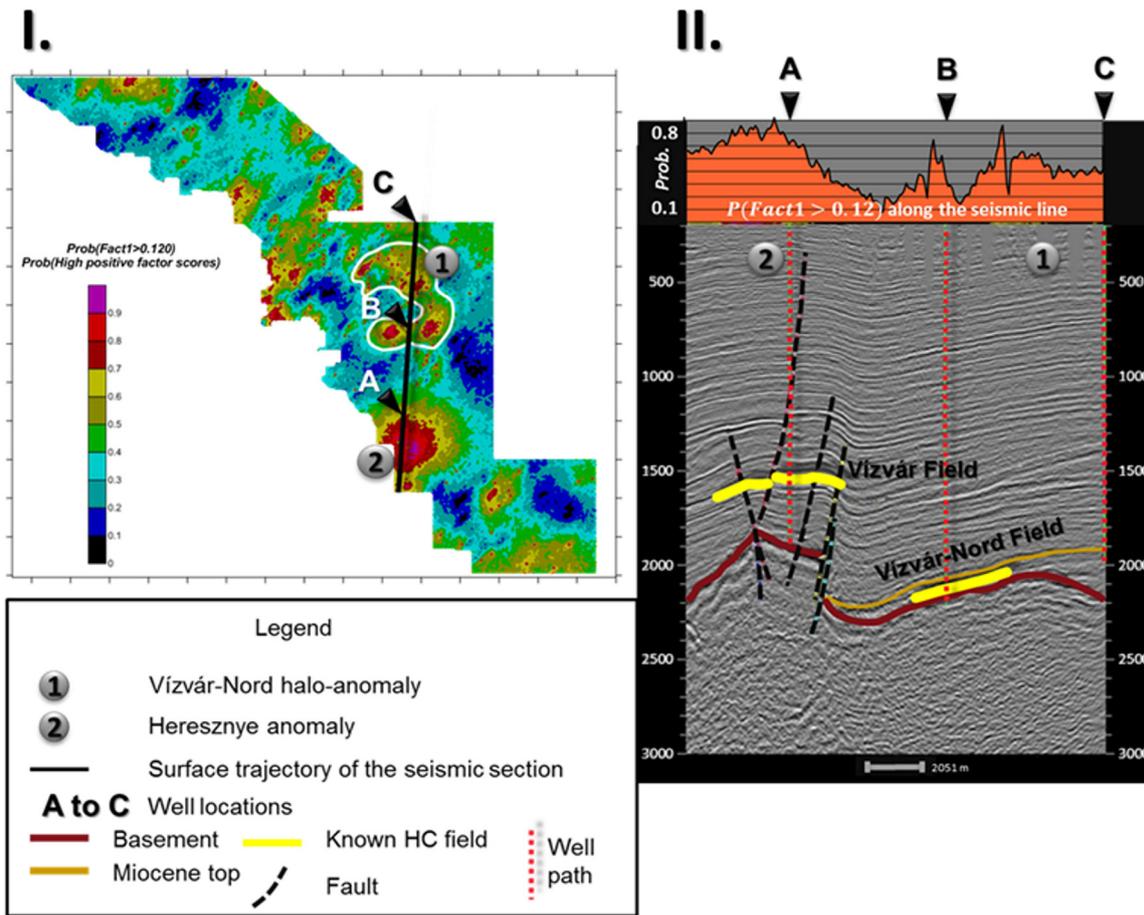
Figure 13. Correlation of the probability map of high Factor 1 scores and the subsurface hydrocarbon fields along the first transect.
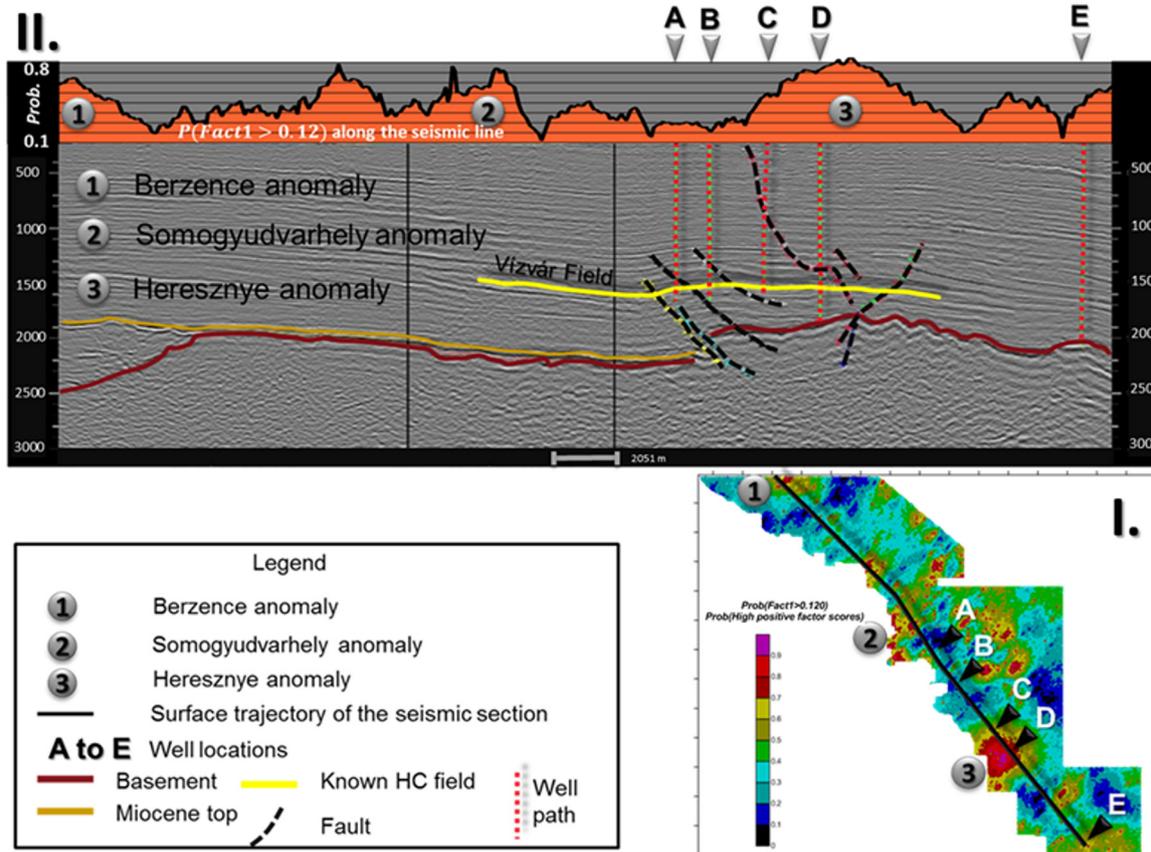

Figure 14. Correlation of the probability map of high Factor 1 scores and the subsurface hydrocarbon fields along the second transect.

Geiger et al.: Statistical and geostatistical study of Rn and hydrocarbon components of a soil gas monitoring system ... **267**
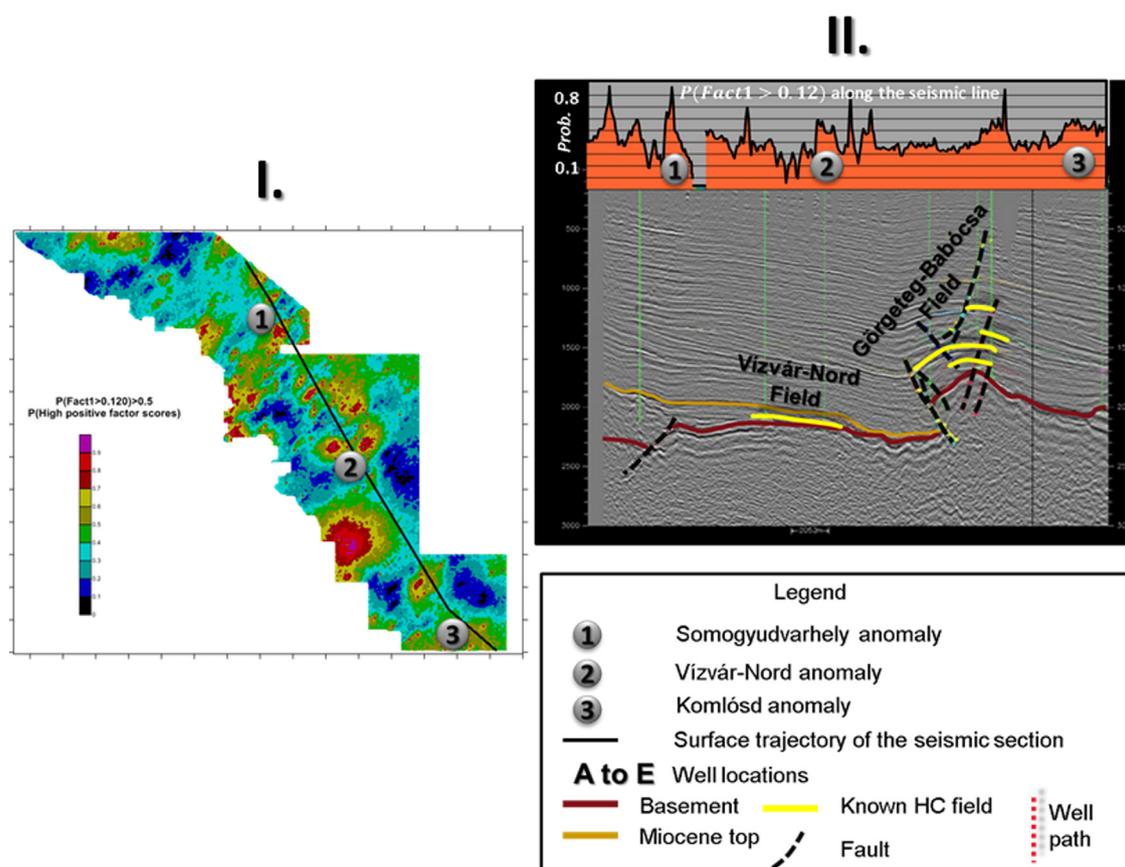
Geologia Croatica

Figure 15. Correlation of the probability map of high Factor 1 scores and the subsurface hydrocarbon fields along the third transect.

might be not necessarily the dominant process. For the purpose of this study, the only important thing was the existence of such a factor. This was the main reason for (1) the application of only the unrotated factor matrix and (2) the spatial analysis of Factor 1 even if it did not have high communality.

The loadings in Factor 1 showed exactly what we were looking for: the higher the Rn content, the higher the CH content is (Table 3). This result was in absolute harmony with the basic geochemical theory (Fig. 2). In Factor 2, the signs of factor loadings belonging to the Rn and HC component were contradictory. This process could be characteristic above such regions, where the vertical migration of HC gases slowed down while the migration of the Rn components did not change significantly. These phenomena may be characteristic above the water phase of the reservoirs. Unfortunately this situation could not be checked on the basis of the available seismic survey.

The spatial analysis of the factor scores highlighted some spatial outliers. For example, in Figure 5A, a high score appeared (red dot on the map) among average scores (green and blue dots). Under regular circumstances, these outliers should be filtered out. We did not follow this rule, since the factor scores were regarded as measures of the strength of the background process. Also, the existence of such outliers calls for the application of a non-parametric simulation approach. Fortunately, in this case the spatial outliers did not form any chain-like pattern which would have necessitated an indicator approach.

The 200 equally probable stochastic realizations honoured well the statistical characters of the input set (Figs. 7 and 8). The simulation result was accurate, although the simulated set showed slightly larger variability than that of the input. Since a Gaussian

algorithm resulted in a unimodal distribution, the E-type grid could be accepted to show the average or general lateral distribution of the results of the migration process. The areal distributions of regions with high Factor 1 scores of the E-type estimation roughly matched the surface projections of the known fields (Figs. 1 and 9). The uncertainty of this rough matching was characterized by the conditional variance of the probability distributions generated at the grid nodes. Unfortunately, the highest uncertainty values are concentrated in the larger area of the high Factor 1 scores (Fig. 11B). Perhaps the heterogeneity of Factor 1 scores in this area and its surrounding were the reason for the less accurate estimation results.

The probability maps showed the possibility of locating such regions, where Factor 1 scores are large in the sense of the definition given above. The results showed that the regions where this event had a large probability (in general the probability was > 0.7) could be clearly correlated with the subsurface hydrocarbon reservoirs identified in seismic profiles (Figs. 13–15). This correlation produced a positive result even in the region where the uncertainty was very high (Heresznye anomaly in Fig. 1, Fig. 13 and Fig. 14). The probability counterparts of the other known reservoirs on the map of the E-type estimation of Factor 1 scores could be outlined with a significantly smaller uncertainty (Fig. 9, Fig. 11 and Fig. 15).

These results suggested that the research achieved its original purpose: the connection between the known reservoirs and the Rn and HC components of soil gases have been proven in the studied area. This verification seems to be more precise than those of the relevant literature, since it used not only the Rn components alone, but also the C2–C5 hydrocarbon components of soil gases. Because of the positive verification, regions with a

high probability of a positive anomaly of Factor 1 scores, but without any reservoir counterparts, may be suggested as targets for surface hydrocarbon exploration.

However, it can also be said that the presented modelling strategy demonstrated a correlation of the measured signal to reservoirs rather than more widespread petroleum system sources.

Acknowledgement

REFERENCES

ARMSTRONG, F.E. & HEEMSTRA, R.J. (1973): Radiation halos and hydrocarbon reservoirs: a review. U.S. Department of the Interior, Bureau of Mines, 52 p.

BUTTAFUOCO, G., TULLARICO, A. & FALCONE, G. (2007): Mapping Soil Gas Radon Concentration: A Comparative Study of Geostatistical Methods. Environmental Monitoring Assesssment, 131, 135–151. doi: 10.1007/s10661-006-9463-7.

CASTILLO, P.I.C, TOWNLEY, B.K., EMERY, X.F., PUIG, Á.F. & DECKART, K. (2015): Soil gas geochemical exploration in covered terrains of northern Chile: data processing techniques and interpretation of contrast anomalies.– Geochemistry: Exploration, Environment, Analysis, May 2015, v. 15, 222–233. doi: 10.1144/geochem2014-283

CHILD, D. (2006): The essentials of factor analysis. (3rd ed.). New York. Continuum International Publishing Group.

CIOTOLI, G., LOMBARDI, S. & ANNUNZIATELLIS, A. (2007): Geostatistical analysis of soil gas data in a high seismic intermontane basin: Fucino Plain, central Italy.– Journal of Geophysical Research, Vol. 112, B05407, doi:10.1029/2005JB004044, 2007

COLEMAN, D.D., MEENTS, W.F., LIU, C.-L. & KEOGH, R.A. (1977): Isotopic identification of leakage gas from underground storage reservoirs- a progress report.– Report 111, Illinois State Geological Survey. 10 p.

DEUTSCH, C.V. (1989): DECLUS: A FORTRAN 77 program for determining optimum spatial declustering weights, Computers & Geosciences, 15/3, 325–332. doi: 10.1016/0098-3004(89)90043-5.

DEUTSCH, C.V. (1997): Direct Assessment of Local Accuracy and Precision.– In: BAAFI, E.Y. & SCHOFIELD, N.A. (eds.): Geostatistics Wollongong. Kluwer Academic Publishers, 115–125.

DEUTSCH, C.V. (2002): Geostatistical reservoir modeling.– New York: Oxford University Press, 376 p.

DEUTSCH, C.V. & JOURNEL, A.G. (1992) GSLIB: Geostatistical software library and user's guide.– New York: Oxford University Press, 340 p.

DYCK, W. & JONASSON, I.R. (2000): In Hale, M. (ed.): Geochamical Remote Sensing of the sub-surface. Chapter 11. Elsevier.

GEIGER, J. (2012): Some thoughts on the pre-and post-processing in sequential Gaussian simulation and their effects on reservoir characterization.– In: GEIGER, J., PÁL-MOLNÁR, E. & MALVIĆ, T. (eds.): New horizons in Central European geomathematics, geostatistics and geoinformatics: Selected studies of the 2011 Croatian-Hungarian Geomathematical Convent, Mórahalom. Szeged: GeoLitera, 17–35.

GHAGREMANI, D.T. (1985): Radon prospecting for hydrocarbon: potential strategy for devonian shale gas in N.E. Ohio; PhD thesis, Department of Geological Sciences, Case Western Reserve University, 259 p.

GÓMEZ-HERNÁNDEZ, J. & MOHAN SRIVASTAVA, R. (1990): ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program, Computers & Geosciences, 16/4, 395–440. doi: 10.1016/0098-3004(90)90010-q.

GOOVAERTS, P. (1997): Geostatistics for natural resources evaluation.– New York: Oxford University Press. 483 p.

GOTWAY, C.A. & RUTHERFORD, B.M. (1994): Stochastic simulation for imaging spatial uncertainty: Comparison and Evaluation of available methods.– In: ARMSTRONG, M. & DOWD, P.A. (eds.): Geostatistical simulations proceedings of the Geostatistical Simulation Workshop, Fontainebleau, France, 27–28 May 1993. Dordrecht: Springer, 1–21.

ISAAKS, E.H. (1990): The application of Monte Carlo methods to the analysis of spatially correlated data.– PhD thesis. Stanford University, 213 p.

ISAAKS, E.H. & SRIVASTAVA, M.R. (1989): An introduction to applied Geostatistics.– New York: Oxford University Press. 561 p.

JOURNEL, A.G. (1983): Nonparametric estimation of spatial distributions.– Journal of the International Association for Mathematical Geology, 15/3, 445–468. doi: 10.1007/bf01031292.

JOURNEL, A.G. (1988): Focusing on spatial connectivity of extreme-valued attributes: stochastic indicator models of reservoir heterogeneities.– Society of Petroleum Engineers Annual Technical Conference and Exhibition: SPE Paper 1832.

JOURNEL, A.G. (1994): Modeling Uncertainty: Some Conceptual Thoughts.– In: DIMITRAKOPOULOUS, R. (ed.): Geostatistics For the Next Century. Kluwer Academic Publishers, 30–43. doi: 10.1007/978-94-011-0824-9_5

JÖRESKOG, K.G., KLOVAN, J.E. & REYMENT, R.A. (1976): Geological factor analysis.– Methods in Geomathematics 1, 180 p. Elsevier Scientific Publishing Co. ISBN 0 444 41367 7

KLUSMAN, R.W. (1993): Soil Gas and Related Methods for Natural Resource Exploration. – John Wiley, New York, 483 p.

KRISTIANSSON. K. & MALMQVIST, L. (1982): Evidence for nondiffusive transport of Rn in the ground and a new physical model for the transport.– Geophysics, 47, 1444–1452.

LANDRUM, J.H., RICERS, D.M., MAXWELL, L.E. & FALLGATTER, W. (1989): A comparative hydrocarbon soil-gas study of the Flathead Valley and Townsend Valley areas, Montana.– Journal of Geochemical Exploration, 34/3, 303–335. doi: 10.1016/0375-6742(89)90119-2-

LI, Y.L., YUAN, G.J. & PENG, D. (2006): Discussion on geochemical exploration predicts the depth of oil reservoir (Chinese).– Wutan Huatan Jisuan Jushu, 28/4, 367–371.

MALDONATO, T.J. & CAMPBELL, G.E. (1992): An Evaluation of Total Petroleum Hydrocarbon Data.– Proceedings of the Focus Conference on Eastern Regional Ground Water Issues. October 13-15, 1992. Boston Marriott Newtown. Newtown, Massachusetts. http://info.ngwa.org/GWOL/pdf/920157297.pdf

MALVIC, T. (2008): Kriging, cokriging or stochastical simulations, and the choice between deterministic or sequential approaches.– Geologia Croatia, 61/1, 37–47. doi: 10.4154/GC.2008.06

MCBRATNEY, A.B. & WEBSTER, R. (1986): Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates.– Journal of Soil Science, 37/4, 617–639. doi: 10.1111/j.1365-2389.1986.tb00392.x.

MCDONALD, R.P. (1985): Factor analysis and related methods.– Hillside, N. Lawrence Erlbaum Associates, Inc. 259 p.

MORSE, J.G., RANA, M.H. & MORSE, L. (1982): Radon mapping as indicators of subsurface oil and gas.– Oil Gas J., 80, 227–246.

MORSE, J.G. & ZINKE, R. (1995): The origin of radiometric anomalies in petroleum basins-a proposed mechanism.– Oil and Gas Journal, 36–38., June 5, 1995.

MUCSI, L., GEIGER, J. & MALVIC, T. (2013): The Advantages of Using Sequential Stochastic Simulations when Mapping Small-Scale Heterogeneities of the Groundwater Level.– Journal of Environmental Geography, 6/3–4, 39–47. doi: 10.2478/jengeo-2013-0005

OLIVIER, M.A. & KHARYAT, A.L. (1999): Investigating the spatial variation of radon in soil geostatistically.– In: DIAZ, J, TYNES, R., CALCWELL, D & EHLEN, J. (eds): GeoComputation CD-ROM. ISBN 0-9533477-1-0. http://www.geocomputation.org/1999/095/gc_095.htm.

PALACIOS, D., FUSELLA, E., AVILA Y., SALAS, J., TEIXEIRA, D., FERNÁNDEZ, G., SALAS, A., SAJO-BOHUS, L., GREAVES, E., BARROS, H., BOLÍVAR, M. & REGALADO, J. (2013): Radon measurements over a natural-gas contaminated aquifer. Radiation Measurements, Vol. 50, 116–120. doi: 10.1016/j.radmeas.2012.10.016.

PATRICK, R., FRÉDÉRIC, P., BHARAT, P.K., FRÉDÉRIC, G., MUKUNDA, B. & SOMA, N.S. (2011): Temporal signatures of advective versus diffusive Radon transport at a geothermal zone in Central Nepal.– Journal of Environmental Radioactivity, 102, 88–102.

PIRSON, S.J. (1969): Geological, geophysical, and geochemical modifications of sediments in the environment of oilfields.– SMU Symposium 1 on Unconventional Methods in Exploration for Petroleum and Natural Gas. 1969, Dallas.

SAUNDERS, D.F., BURSON, K.R., BROWN, J.J., & THOMPSON, C.K. (1993): Combined geological and surface geochemical methods discovered Agaritta and Brady Creek fields, Concho County, Tex.– AAPG Bull., Vol. 77 No. 7 July 1993, 1221–1239.

SRIVASTAVA, M. (1994): The visualization of spatial uncertainty.– In: YARUS, J. & CHAMBERS, R. (eds.): Stochastic Modeling and Geostatistics.– AAPG Computer Applications in Geology, Volume 3, 339–345.

STEVENS, J.P. (2002): Applied multivariate statistics for the social sciences (4th ed.).– Hillsdale, NS: Erlbaum, 641 p.

TEDESCO, S.A. (1995): Surface Geochemistry in Petroleum Exploration.– Springer, 206 p. doi: 10.1007/978-1-615-2660-5. ISBN 9781-4613-6142-8.

THURSTONE, L.L. (1931): The Measurement of Social Attitudes.– Journal of Abnormal and Social Psychology, 27, 249–269. doi: 10.1037/h0070363

VÁRHEGYI, A., GORJÁNÁCZ, Z. & HORVÁTH, ZS. (2008): Komplex radiometriai módszer alkalmazása a hazai szénhidrogén-kutatásban (Application of a complex readimetriy method in hydrocarbon exploration, in Hungarian).– Bányászati és Kohászati Lapok, 141/3, 64–70.

WRIGLEY, M.J., STOLPMANN, H., AL-JENAIBI, M. & VAN LAER, P. (2012): A Direct Hydrocarbon Indicator Method and Results of a Demonstration of Capabilities in Abu Dhabi.– AAPG Search and Discovery Article #40895 (2012). Posted March 31, 2012.

TUKEY, J.W. (1977): Exploratory data analysis (Addison-Wesley series in behavioral science - quantitative methods).– 17th edn. Reading, Addison-Wesley.