Geological Quarterly, 2022, 66: 16
DOI: http://dx.doi.org/10.7306/gq.1648

Investigation of lithofacies predictability using the Shannon (information) entropy theorem: the Upper Eocene "Górka Lubartowska" amber deposit of the Siemień Formation

Krzysztof CZURYŁOWICZ[1, *]

[1] University of Warsaw, Faculty of Geology, Żwirki i Wigury 93, 02-089 Warszawa

Czuryłowicz, K., 2022. Investigation of lithofacies predictability using the Shannon (information) entropy theorem: the Upper Eocene "Górka Lubartowska" amber deposit of the Siemień Formation. Geological Quarterly, 2022, 66: 16, doi: 10.7306/gq.1648

Associate Editor: Joanna Rotnicka-Dłużewska

A generalized workflow of scientific process requires data to be obtained, reprocessed, integrated, optionally transformed, modelled and finally interpreted in order to understand the underlying process. This procedure is affected by both objective and subjective uncertainties. In parallel with the development of geostatistics, the role of uncertainty has been widely investigated in geosciences. This has led to the introduction of new concepts, taken for example from thermodynamics, such as entropy. Predicting the subsurface is an especially thankless effort, as data are driven from spatially highly limited direct sources. The following paper provides an review of various applications of the Shannon entropy theorem in geoscience. Information entropy, initially proposed by Shannon (1948) provides an objective measure of overall system uncertainty. Significant concern has been focused on the application of Shannon entropy to provide an objective measure of joint system uncertainty and visualization of its spatial distribution. The area of extensively drilled Eocene amber-bearing deposits located in the Lubelskie voivodeship was selected as a case study to investigate the quality of prediction stochastic lithofacies models. The importance of adding secondary variables to a stochastic model is also reviewed here. Adding new data and rerunning the simulation allows assessment of its impact on the predictability of a stochastic model. The most important conclusion from the study is that the deposition of amber-bearing lithofacies occurred mostly in the northern part of the area investigated, as shown also by ongoing exploitation of the deposit.

Key words: Shannon entropy, uncertainty, facies probability, Multiple-Point Statistics.

INTRODUCTION

One of the key objectives in geological exploration is to map the most prospective areas. In many conceptual mineral system models, the host rock presence is the key factor that controls the spatial distribution of resources. Data collected during ongoing exploration of amber-bearing deposits has shown that the probability of finding amber raw material quadruples when a specific lithofacies is drilled (Czuryłowicz, 2013). The reliable prediction of the spatial architecture of this lithofacies within the depositional sequence investigated is therefore critical. This issue has attracted attention (Deutsch, 1998; Wellmann and Regenauer-Lieb, 2012) in seeking to establish a reliable measure of the prediction quality of discrete variable stochastic models. Introduction of Shannon entropy as a joint measure of overall model uncertainty allows assessment of the impact of knowledge- and data-driven sources of information. The general workflow involves (1) selection of a validation subset of lithofacies borehole profiles, (2) deriving data-driven input parameters of spatial continuity, (3) providing knowledge-based data by the construction of training images and 3D probability cubes and (4) simulation of lithofacies models using various stochastic algorithms supported by different sets of input knowledge- and data-driven information, followed by (5) calculation of Shannon entropy from derived probability models.

* E-mail: kczurylowicz@uw.du.pl

Received: September 30, 2021; accepted: May 20, 2022; first published online: August 26, 2022

BACKGROUND

The term entropy originates from thermodynamics and describes the state of maximum disorder of energy and its distribution in an equilibrium state. The term entropy in information theory, defined by Shannon (1948), is designated to determine the degree of uncertainty of the variable investigated (Wędrowska, 2010). Originally, the entropy theorem was developed to separate a transmitted radio signal with certain statistical characteristics from noise. As geologists, we also face similar problems in analytical work, when certain sediment successions that are indicative of particular sedimentary environments must be extracted from a sometimes seemingly disordered profile. This tool has the basic advantage of combining computational simplicity with intuitiveness and efficiency of interpretation (Doktor et al., 2010).

The first application of the information (Shannon) theorem in sedimentological analyses aimed at the interpretation of cyclicity type and an indication of the randomness of the individual lithologies analyzed by Markov chains (Schwarzacher, 1969; Krawczyk, 1979; Mastej, 2002). Then, for each lithology, the pre- and post-depositional entropy was calculated, and so it was possible to determine the possible asymmetry of the lithological succession. The results were then plotted on Hattori (1976) model nomograms to determine the possible sedimentation environment. For example, a zero value of predepositional entropy for a particular lithofacies indicates directly that a lithofacies investigated is always underlain by a another particular lithofacies. An increase in predepositional entropy implies a situation where there is greater diversity of lithofacies that may have underlain a particular lithofacies within the sedimentological profile investigated (Krawczyk, 1980; Doktor and Krawczyk, 2010).

A Shannon entropy-based measure was proposed by Chiogna et al. (2012) to quantify the dilution of conservative solutes either in a given volume (dilution index) or in a given water flux. Bianchi and Pedretti (2017) introduced a new tool called an entrogram, a logical extension of the Shannon entropy theorem, to investigate fluid dynamics and solute transport in a porous aquifer. The entrogram is a promising concept that can be used to measure the overall persistency of patterns of spatial association in a distributed field. This proposed measure allows robust comparisons between different spatial structures (Bianchi and Pedretti, 2017).

Another application of Shannon entropy is to illustrate the uncertainty of structural interpretation, by stochastic perturbation of stratigraphic boundaries and structural discontinuities (Caers, 2011). It has been shown that an information theorem provides not only a quantitative insight into the overall uncertainty space of a structural and stratigraphic interpretation, but also provides information about the internal connection of the interpretive model with the input data (Wellmann and Regenauer-Lieb, 2012). The possibilities of using entropy go beyond the visualization of uncertainty. For example, by interpreting structural features on a seismic image in the so-called “optimization loop”, it is possible to evaluate the contribution of an additional interpretation portion to the total entropy of the system analyzed. It is then necessary to calculate the differential entropy for the structural-stratigraphic models (Wellmann and Regenauer-Lieb, 2012). Under such conditions, Shannon entropy can be seen as an alternative, objective measure to describe the space of interpretation uncertainty that can be used in parallel with interpretation restoration tools.

Shannon entropy provides an objective measure that also allows visualization of the spatial distribution of uncertainty associated with the stochastic model of any discrete-type variable. The main advantage of information entropy is the possibility of visualizing the spatial variability of uncertainty of all lithofacies simultaneously, by communicating their probability distribution.

This publication addresses the application of the Shannon entropy theorem to assess the uncertainty of stochastic lithofacies models constructed by variogram-based and training image-based algorithms. This issue was raised first by Deutsch (1998), in the context of comparing different cross-validation methods and their application to specific simulation algorithms of a discrete variable.

THEORETICAL BASICS

The variable that was used for the comparative analysis of different realizations of the lithofacies model was the total entropy of information *H* [1], which describes the lack of knowledge in relation to the total space of the depositional system under study (Shannon, 1948; Deutsch, 1998; Wellmann and Regenauer-Lieb, 2012). It takes the following mathematical form:

$$H = -\sum_{k}^{K} p_k \cdot \log(p_k) \qquad [1]$$

where: p_k is the probability of the kth elementary event (k = 1, …, K)

The studies conducted so far has shown that it is possible to determine whether the introduction of new (a posteriori) information or a change in the hypothesis describing lithofacies relations affects the optimization of the model (Deutsch, 1998; Wellmann and Regenauer-Lieb, 2012). As in thermodynamics, adding energy into the system increases its entropy, so the setting of an additional portion of information to the stochastic model developed that incorrectly describe the nature of the spatial organization of lithofacies determines similar behaviour of Shannon entropy (Wędrowska, 2010). The concept of entropy was developed on the assumption that knowing the probability of each letter of the alphabet appearing in the words we speak, it is possible to define a measure describing the missing information in order to recreate the full text of an incomplete or encrypted message. At the same time, this information should not be linked with the meaning of the message. Thus, information entropy is a measure of missing information required for a comprehensive description of the system or process under study. For a depositional system consisting of two lithofacies, the value of entropy approaches the value of 0 when the probability of one of them to occur approaches 100%. In the case where at the unsampled location of the stochastic model there is an equal probability of each lithofacies, then the entropy is maximum and reaches the value 1 (Fig. 1).

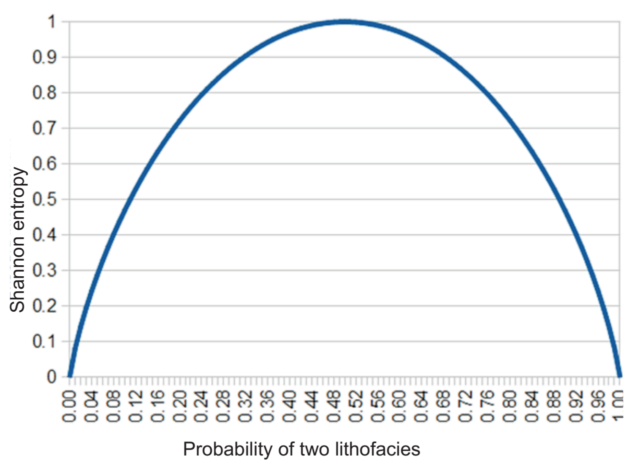

Fig. 1. Shannon entropy for a 1 bit system (two possible outcomes) as a function of the probability of the first outcome (following Wellman and Regenauer-Lieb, 2012)

The parameter C_k describing the "closeness" to the true facies should be interpreted as the average value (E) of the probability of facies k at location u_α in the validation subset of the profile investigated, which has not been used for the simulation stage [2]. Its value should be 1 when we have a complete set of data to develop a reliable lithofacies model. If the variable takes a value <1/lithofacies number (as in the case of the Górka Lubartowska deposit – 0.33), it should be assumed that the reliability of the lithofacies model obtained is minimal and comparable to the completely random model (Deutsch, 1998).

$$C_k = E\,[p(u_\alpha;\ k)\ |\,\text{true} = k],\ k = 1, ..., K \qquad [2]$$

However, in order to compare the results of the validation analysis, the "closeness" in relation to global lithofacies proportions (p_k) was introduced.

$$C_k^{rel} = \frac{C_k - p_k}{p_k},\ k = 1,\ldots K \qquad [3]$$

Thus, the measure of the quality of the model is a variable which, in terms of probability, describes the closeness to reality (Deutsch, 1998).

CASE STUDY AREA

The subject of the analysis is the Upper Eocene Siemień Formation of the "Górka Lubartowska" (Fig. 2) amber deposit (Strzelczyk and Danielewicz, 1990). Deposition of these Upper Eocene strata was related to the transition from transgression, which determined the maximum extent of the Late Eocene sea, to marine regression that initially took place under conditions of enhanced sediment supply (Czuryłowicz et al., 2014). The "Górka Lubartowska" deposit is located within the prospective amber area of the northern Lublin region (Fig. 2A, B). In this area, in the Middle and Late Eocene, on the southern shore of the epicontinental sea, favourable palaeogeographic and hydrodynamic conditions prevailed along the shoreline to form significant amber resources (Kramarska and Kasiński, 2008). The Upper Eocene Siemień Formation deposits were initially deposited within the prodelta zone, then with falling sea level, deposition changed to reflect progradation of submarine distributary channels (Table 1). The thickness of this Upper Eocene formation is 7 m on average. The top of the deposit is at a depth of 10 to 24.5 m, and its base at 13.4 to 29.2 m. In total, 158 boreholes were drilled, with a total length of 4.093.1 m and 150 x 200 m borehole spacing (Fig. 2C).

RESEARCH METHODOLOGY

The uncertainty analysis procedure was performed by separating the borehole database into test and validation subsets (including boreholes 79, 87, 104, 109, 112, 142, 144A, 166A and 169; location: see Fig. 2C). The analysis was conducted on multiple equiprobable realizations of lithofacies architecture (Table 1) generated with the use of stochastic simulation algorithms based on variograms, i.e. Sequential Indicator Simulation – SIS and Truncated Gaussian Simulation – TGS (Deutsch and Journel, 1992; Deutsch, 2002; Kelkar and Perez, 2002; Liu et al., 2005; Armstrong et al., 2011). The Multiple Point Statistics (MPS) algorithm based on reference images was also used in the comparative analysis (Caers and Zhang, 2002; Strebelle, 2002; Harding et al., 2004; Hu and Chugundova, 2008; Boucher, 2011). Probability models and the final most frequent occurring lithofacies model were computed from 500 realizations (Fig. 3). The simulation algorithms of discrete variables were applied to choose the most reliable workflow of lithofacies reconstruction, that stays in accordance with fundamental laws of deposition processes. SIS honours basic parameters describing the spatial structure of each individual lithofacies, as inferred from semivariogram analysis. However, this method does not respect the ordering of lithofacies resulting from the nature of the deposition. TGS is rather used for simulation of ordered sequences. The simulation algorithm itself truncates single gaussian random field into domains, that are determined by lithofacies rules. The Multi-Point Statistics algorithm is a reflection of the intuitive actions of a sedimentologist aimed at lithological correlation. In both cases, the key to inferring the occurrence of a lithofacies at an unsampled location is the spatial relationship of corresponding lithofacies occurring in an analogous geometrical position as in the reference image representing an analogous depositional sequence.

To guarantee comparability of results, the same seed number was used during the simulation. This ensures that the order in which cells of the stratigraphic model are visited is constant. This is important from the point of view of the comparability of building local probability distributions at the stage of the simulation process. Additionally, different chronostratigraphic patterns (Mallet, 2002, 2004) within the stratigraphic model were introduced (Fig. 4). The incorporation of sedimentological interpretations into the stochastic model of depositional architecture took place using 3D probability cubes (Gotway and Young, 2002; Strebelle et al., 2006; Levy et al., 2008a, b; Labourdette et al., 2008). These are computed by integrating the lithofacies depocentre map and vertical proportion curves (Volpi et al., 1997; Ravenne, 2002; Ravenne et al., 2002; Falivene et al., 2006; Purkis et al., 2012; Yarus et al., 2012) using probability aggregation methods (Journel, 2002; Hong et al., 2007; Caers, 2011; Allard et al., 2011; Comunian et al., 2012). Construction of lithofacies depocentre maps was supported by the matrix of vertical proportion curves that reveal the non-stationary nature of lithofacies occurrence and help to delineate thick zones of lithofacies ST (Fig. 5).

Calibration of various data sources into the joint combined conditional probability $P(A/B_1B_2)$ requires good quality data that cover the study area. In this study, partial conditional probabilities are calculated using borehole data in the form of a vertical proportion curve (B_1) and lithofacies depocentre maps (B_2) translated to probability space by simple normalization. Both types of data serve as a source of information on the non-stationarity of lithofacies occurrence. Combination of these two data sources is completed using the Tau model which takes into account data redundancy (Fig. 6). This means that the influence of a secondary variable (B_1 or B_2) on the state investigated (A) depends on how much is "coincidence" between a priori and a posteriori knowledge. Data redundancy measures the "information overlap" between the sources of information used to predict a state (A). This ap-

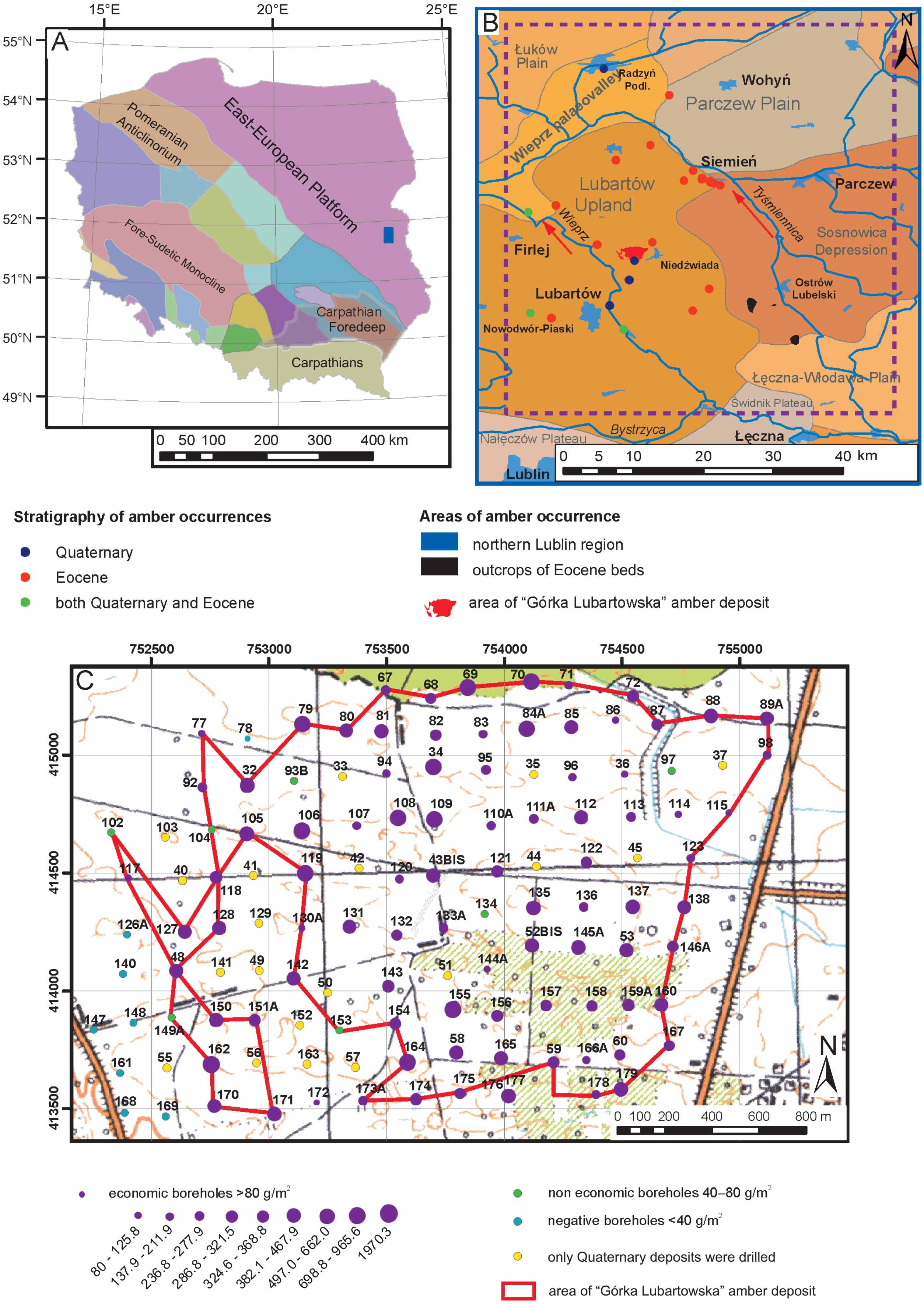

Fig. 2. Location of the case study

A – map of tectonic units on the sub-Cenozoic surface; **B** – map of physico-geographical mesoregions; **C** – map of the area of the "Górka Lubartowska" amber deposit

Table 1

Lithofacies distinguished in the Eocene sequence of the "Górka Lubartowska" amber deposit with their brief description and interpretation

Lithofacies	Description	Sedimentary environment
FD	silt and silty sand with minor amber occurrence	delta plain deposits
ST	silty to fine glauconite sand with significant amount of amber grains	distributary channel deposits
FSm	sandy silt and silt with significant content of glauconite and amber grains	prodelta deposits

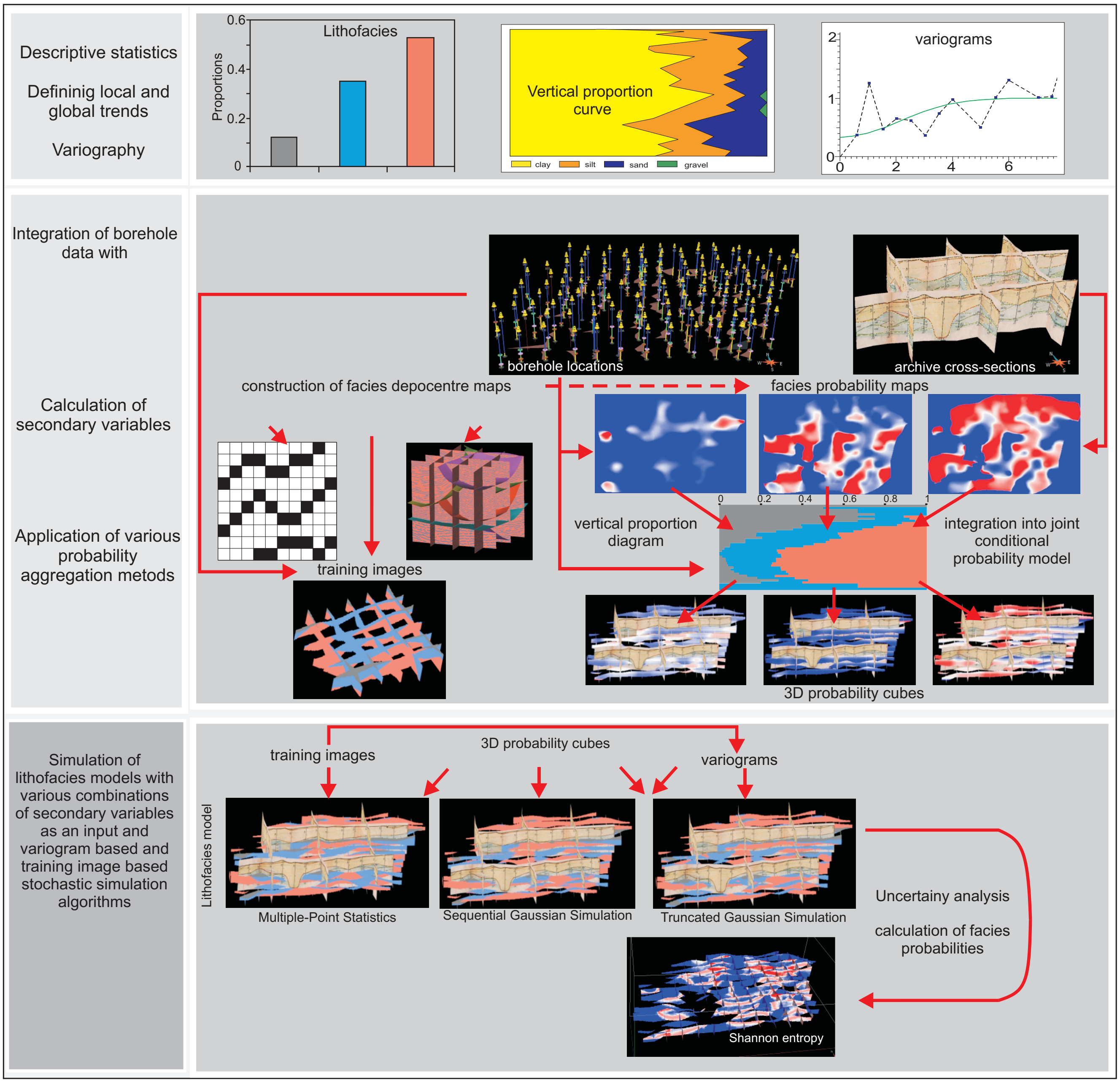

Fig. 3. Methodology of uncertainty analysis in order to calculate Shannon entropy

The workflow shown includes: variogram analysis, construction of secondary variables such as 3D probability cubes, and training images

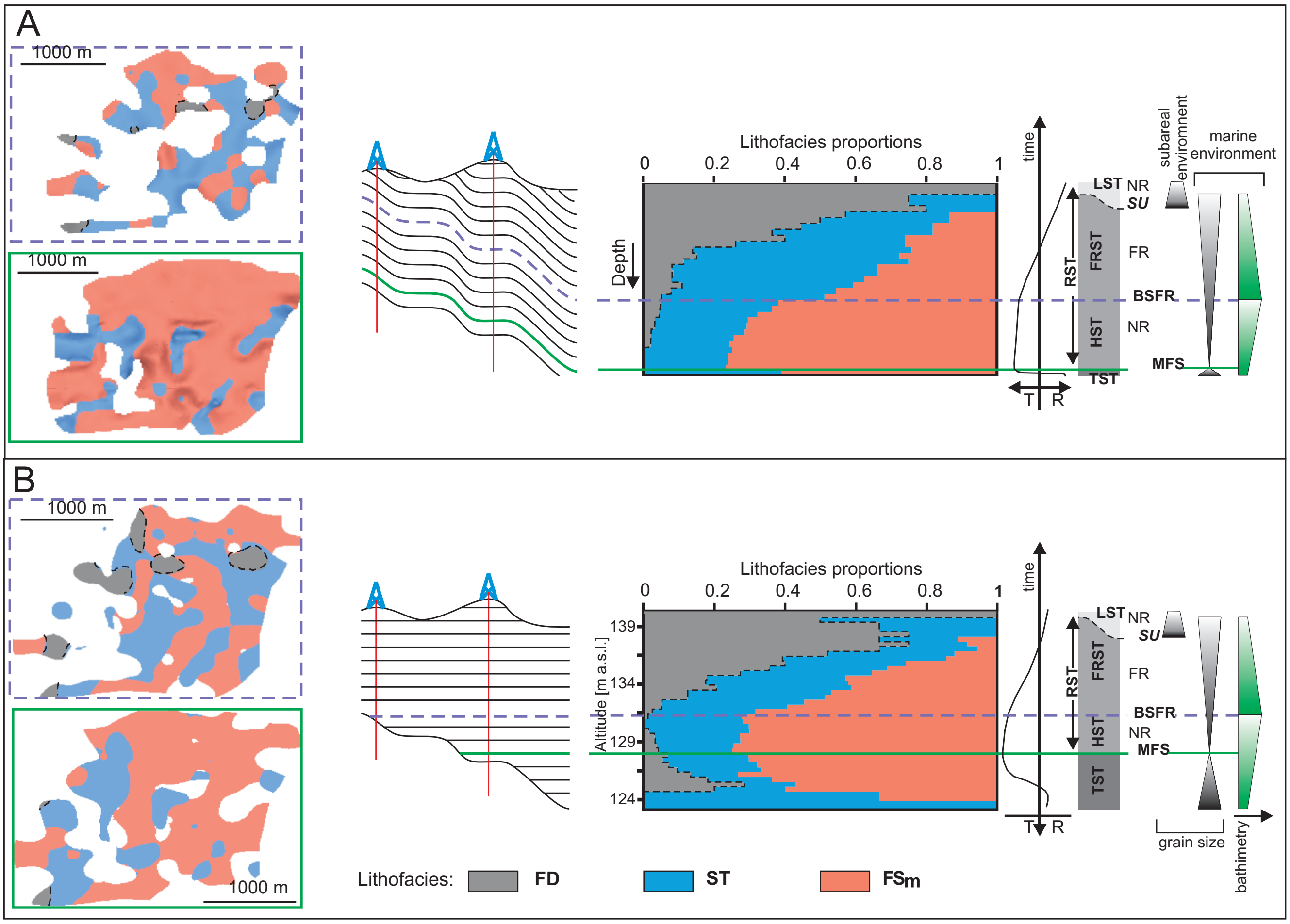

Fig. 4. Interpretation of system tracks of the Eocene depositional sequence based on vertical proportion curves built using: A – onlap chronostratigraphic framework; B – baselap chronostratigraphic framework

BSFR – basal surface of forced regression, FR – forced regression, FRST – forced regression systems tract, HST – highstand systems tract, LST – lowstand systems tract, MFS – maximum flooding surface, NR – normal regression, RST – regressive systems tract, SU – subaerial unconformity, ST – transgressive systems tract

proach assumes that the relative contribution of source B_1 to predict state A is the same regardless of the fact that you have other source of information B_1 (Caers, 2011). Computed 3D probability cubes are the representation of probability of each lithofacies to occur in each cell of the stratigraphic model, based on the delineation of depositional zones and depositional trends within the sequence investigated.

$$b_1 = \frac{1-P(A/B_1)}{P(A/B_1)}, b_2 = \frac{1-P(A/B_2)}{P(A/B_2)}, a = \frac{1-P(A)}{P(A)} \quad [4]$$

$$P(A/B_1B_2) = \frac{a}{a+b_1b_2} \quad [5]$$

where: b_1 quantifies how much is not known about state A knowing the information B_1, b_2 quantifies how much is not known about state A knowing the information B_2 and a quantifies how much is not known about state A before any data is available (Caers, 2011).

The computation of the probability of occurrence of each lithofacies in the location where it has been empirically determined, e.g. on the basis of the borehole core description, can be carried out by adopting two strategies for the separation of the validation subset:

- by removal of the entire lithofacies profile from the borehole,
- by removal of selected or random intervals of the lithofacies profile from the borehole.

The first method was used in this study, as it allows the assessment of the impact of additional variables introducing geological concepts (e.g., 3D probability cubes) to the stochastic model to a much better extent, at the cost of generally underestimated results of the uncertainty analysis (Deutsch, 1998).

In the final stage of the analysis, the results of Shannon entropy and "closeness" to true lithofacies for selected variants of the lithofacies model were compared in a scatter plot (Fig. 7). This diagram was used to compare the scenarios provided and select a lithofacies model characterized by the value of entropy closest to zero and close to one for the measure of "closeness" to true lithofacies.

The basic premise for the selection of a stochastic model retaining the expected vertical ordering of lithofacies was quantitative, i.e. Shannon entropy and "closeness" to the true lithofacies (Deutsch, 1998). These studies constituted the last stage of reconstruction of the lithofacies architecture of the Upper Eocene deposits. The inputs for calculating the parameters

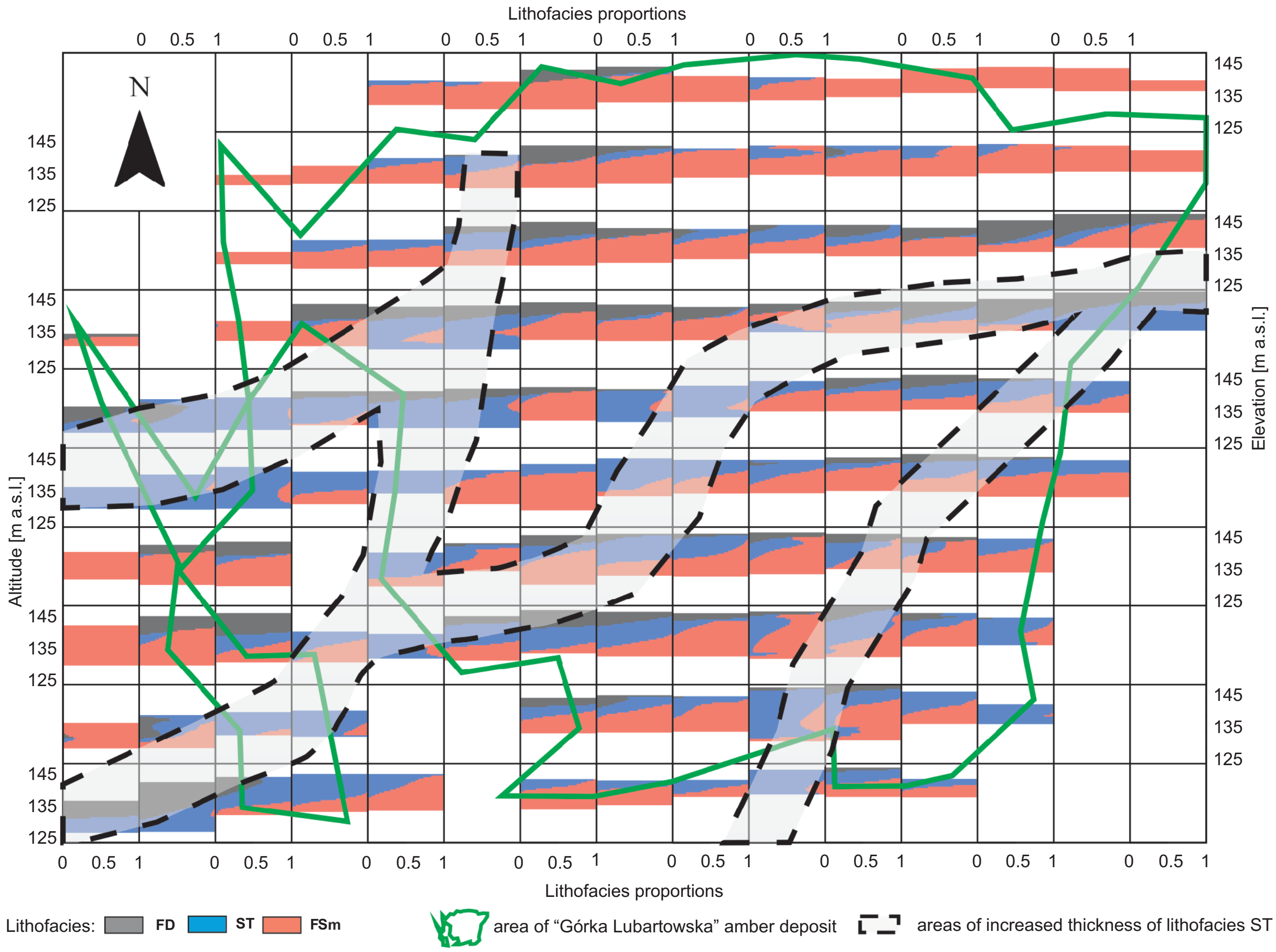

Fig. 5. Matrix of vertical proportion curves computed from lithofacies borehole profiles

of the average entropy (*H*), and “closeness” standardized by the global proportions tested lithofacies (were probability values of lithofacies computed within validation subsets of borehole profiles. These were calculated with various configurations of borehole data and secondary variables. A particular advantage of the application of the theory of entropy was its ability to visualize total model uncertainty simultaneously for all lithofacies of the deposit (Fig. 8).

RESULTS

Analysis of the spatial organization of lithofacies of amber-bearing deposits near Górka Lubartowska show that the probability models obtained solely on the basis of borehole data and variograms fitted to lithofacies profiles are characterized generally by higher entropy compared to those which were updated with secondary probability models i.e. 3D probability cubes (Fig. 3).

Each point within Figure 7 corresponds to a single simulation procedure of 500 realizations of a lithofacies stochastic model. Each procedure uses as input different concepts describing the spatial organization of lithofacies implemented through various types of secondary data, such as 3D probability cubes, variograms, training images, facies depocentre maps and vertical proportion curves. The entropy distribution of stochastic models computed using SIS shows that the presence of zones of equal probability is limited to a maximum of two lithofacies. In turn, lithofacies models based on the TGS algorithm show a significant increase in entropy greater than 1, which is also reflected in geologically unrealistic models (Fig. 7). At the same time, for each of the methods (SIS, TGS and MPS), a clear reduction in entropy is observed as secondary variables are added, successively as vertical proportion curves and 3D probability cubes (Fig. 7). The most promising results in cross-validation analysis are obtained when using 3D probability cubes as secondary data obtained by the aggregation of conditional probabilities of a vertical proportion curve $P(A/B_1)$ and depocentre maps $P(A/B_2)$ using a Tau model (Caers , 2011; Allard et al., 2012).

The comparative studies performed with the assumption of two chronostratigraphic reference patterns (Fig. 4) showed much better-fitting results for the models computed using onlap layering (Fig. 4B). The validation analysis performed illustrates the contribution of the sedimentological concepts of the non-stationary nature of spatial organization of lithofacies towards the understanding of the depositional environment of

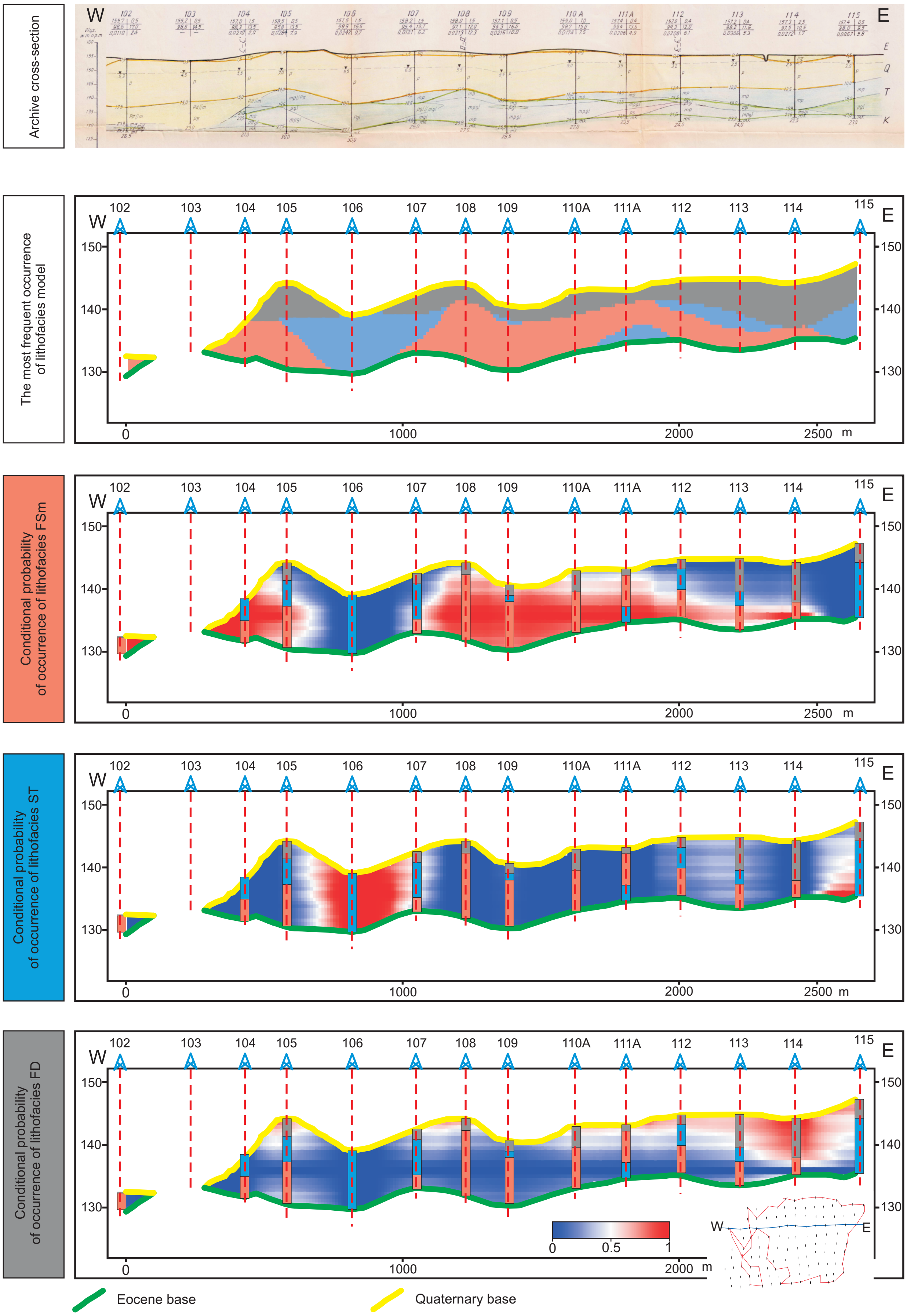

Fig. 6. Summary of the probability aggregation process using the Tau model with reference to the archival lithological correlation (Strzelczyk and Danielewicz, 1990)

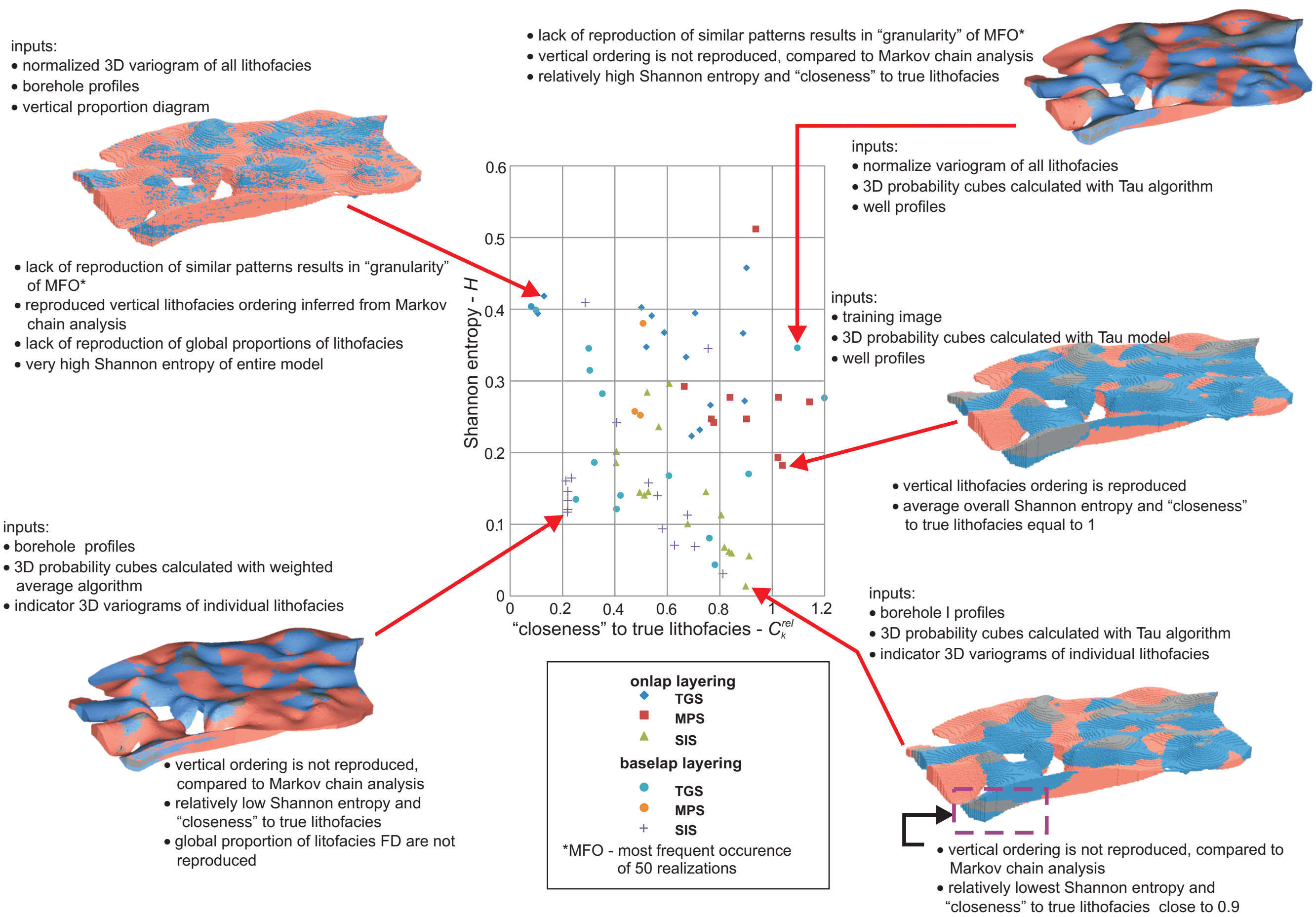

Fig. 7. Scatter plot of Shannon entropy *versus* "closeness" to true lithofacies derived from a validation subset of lithofacies borehole profiles

these Upper Eocene deposits. Nevertheless, when choosing the most optimal set of input parameters, such as variograms, vertical proportion curves and 3D probability cubes, it was not only the realizations meeting the criteria of minimum entropy and "closeness" to true lithofacies equal to one that were taken (Fig. 7). Another crucial criterion that had to be met was the vertical lithofacies ordering inferred from Markov chain analysis (Table 2). Among all the simulation methods used, these criteria were not met by the realizations computed using the algorithm of SIS. The analysis also showed that the realizations based on theoretical models of variograms fitted only to borehole profiles have a higher entropy compared to the realizations, which were based on the variograms built on the basis of the reference image used in the Multiple-Point Simulation (Figs. 5 and 6). In this case, the key role was the relatively high nugget effect compared to total variance. Its presence resulted from the short-range spatial variation of lithofacies architecture that was not able to be inferred from exploration drilling borehole spacing (Olea, 1995; Namysłowska-Wilczyńska, 2006). This effect is magnified when using the TGS algorithm, which is based on a single 3D variogram of normalized data.

Fitted theoretical models of variograms in the vertical and horizontal directions were complex and contained two components – spherical and nugget (*sensu* Stach, 2009). The share of the nugget effect, reaching 45% of the total normalized semivariance, is much higher than in the case of the indicator variograms of individual lithofacies (Tables 3 and 4), used during the sequential indicator simulation. This is the result of the completely different spatial structure and geometry of the bodies that build lithofacies of this Upper Eocene sequence. This effect is also enhanced by post-depositional Neogene and Quaternary erosion, which probably contributed to the blurring of the main directions of their continuity. Modelling the spatial structure of lithofacies simultaneously using only a single and linear combination of acceptable mathematical functions makes it impossible to take into account the individual features of the structure of variability of lithofacies resulting from specific hydrodynamic conditions during their deposition (Gringarten and Deutsch, 2001).

These considerations concerning the impact of the nugget effect of indicator and normalized semivariograms on the simulation process (Tables 3 and 4) also affect the results of Shannon entropy. The "granularity" visible in the realizations is the result of the lack of reproduction of the same lithofacies in nearby cells and in individual stochastic realizations, making them unique and independent (Fig. 7). Consequently, the high

Table 2

Results of Markov chain analysis

A – transition count matrix

A	FD	ST	FSm	Sum →
FD	–	0	0	0
ST	23	–	6	29
FSm	10	50	–	60
Sum ↓	33	55	6	–

B – transition probability matrix

B	FD	ST	FSm
FD	–	0.00	0.00
ST	0.79	–	0.21
FSm	0.16	0.83	–

C – significant facies transitions matrix

C	FD	ST	FSm
FD	–	–	–
ST	**39.77**	–	–
FSm	–	**43.11**	–

Regular facies transitions in percentage scale at degree of freedom α = 0.5 are shown in bold (Staňová et al., 2009)

Table 3

Semivariogram parameters of fitted indicator variogram models constructed with the assumption of an onlap chronostratigraphic framework

	Partial sill variance (C")	Structure type	Maximum horizontal range (a_{max})/azimuth	Minimum range (a_{min})	Vertical range (a_{vert})
FD	0.005 0.045 0.025 0.023	nugget (C_0) spherical spherical spherical	– 384/90 2860/90 4000/90	– 300 2444 4000	– 8.71 7.17 ∞
ST	0.01 0.23	nugget (C_0) spherical	– 622/60	– 395	– 6.7
FSm	0.23	spherical	652/45	400	7.2

Table 4

Semivariogram parameters of fitted variogram models constructed with the assumption of a baselap chronostratigraphic framework

	Partial sill variance (C")	Structure type	Maximum horizontal range (a_{max})/azimuth	Minimum range (a_{min})	Vertical range (a_{vert})
FD	0.033 0.002	spherical Gaussian	555/90 1350/90	384 900	13.8 ∞
ST	0.226	spherical	622/60	458	11.7
FSm	0.226	spherical	660/45	456	16.25

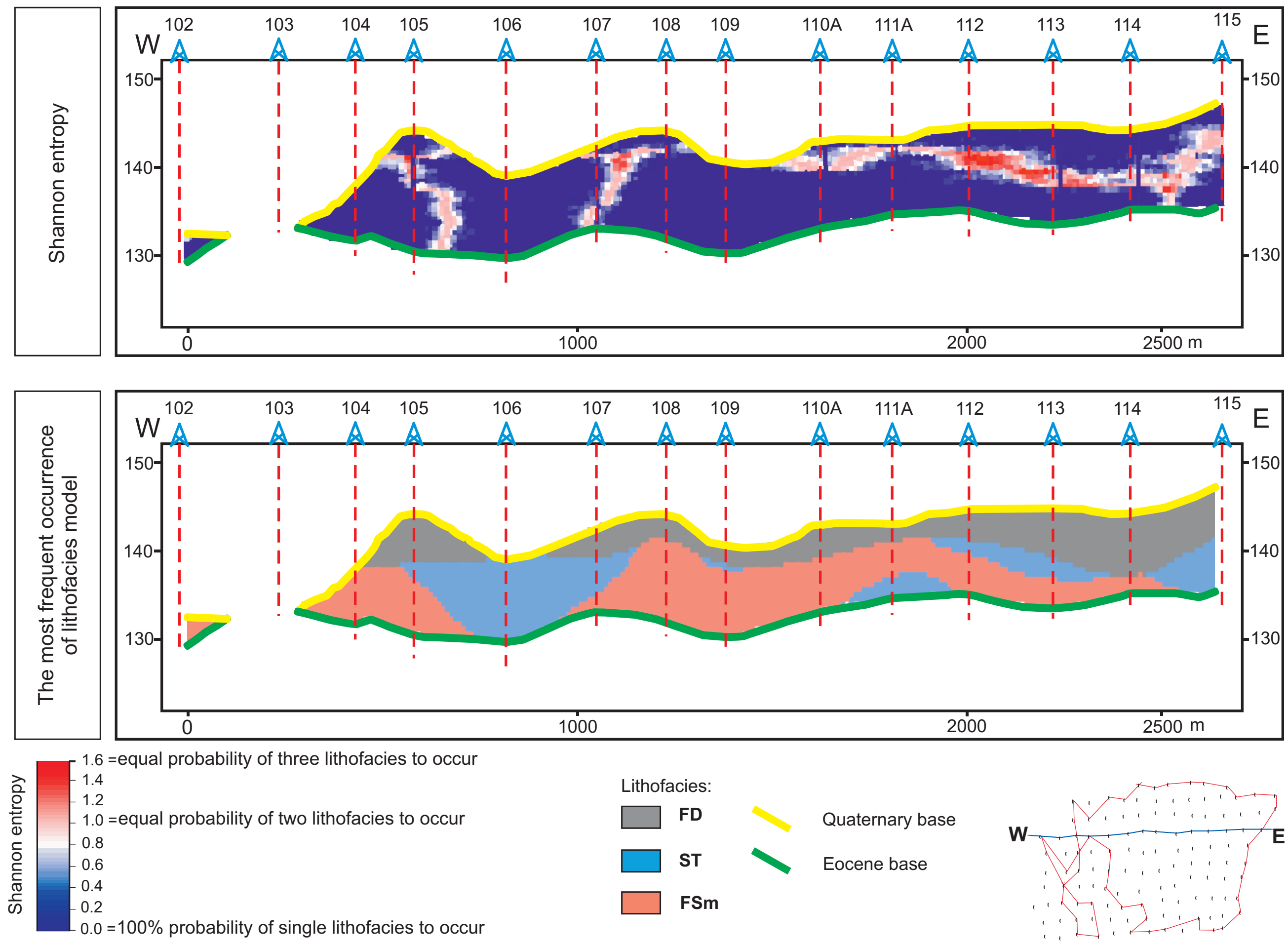

Fig. 8. Cross-section showing coincidence of occurrence of maximal Shannon entropy values with zones of lithofacies overlap and thin intercalations; most of the lithofacies geometry is characterized by a Shannon entropy close to 0

entropy values obtained indicate an almost equal probability of the occurrence of each lithofacies.

At the final stage of the of the uncertainty analysis of the lithofacies model, a reconstruction was selected that met both parametric (i.e. a mean Shannon entropy close to 0 and "closeness" to true lithofacies approaching to 1) and geological criteria (i.e. a vertical ordering of lithofacies consistent with the results of the Markov chain analysis – Table 2). From among multiple modelling scenarios (Fig. 7), a stratigraphic model computed with onlap layering was selected using the Multiple Point Statistics algorithm (Strebelle, 2002; Strebelle et al., 2006; Strebelle and Levy, 2008) based on the reference image (Fig. 6) and 3D probability cube (aggregated with Tau model) (Fig. 5). The spatial distribution of Shannon entropy indicates that the presence of areas with increased entropy coincides almost exclusively with zones of interfingering lithofacies (Fig. 8). It is therefore assumed that addition of additional auxiliary variables, in the form of 3D probability cubes, which incorporate the highest order of information on the spatial non-stationarity of lithofacies, significantly reduces the overall uncertainty of the reconstruction of depositional system architecture.

Comparing simulation algorithms based solely on variograms, it can be seen that the models developed using the Truncated Gaussian Simulation algorithm, regardless of the auxiliary data as an input, show a relatively higher total Shannon entropy (Fig. 9).

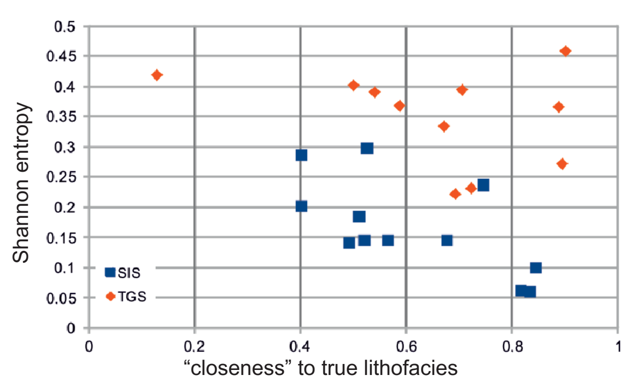

Fig. 9. Scatter plot of Shannon entropy *versus* "closeness" to true lithofacies for variogram-based stochastic simulation algorithms of discrete variables

CONCLUSIONS

Integration of auxiliary data into a joint conditional probability model, describing the conceptual deposition architecture with the Tau model, gives the best results. The advantage of this model is the ability to take into account the contribution of the auxiliary variable information to the understanding of the state investigated. This feature is known in the literature as redundancy (Caers, 2011). This method gives much better results than the classic additive methods, which largely approach the aggregate averaged value of conditional probability and lose an important attribute which stands as the basis of Bayesian inference.

The Shannon entropy theorem has found to be a useful measure, both global and local, describing the amount of information needed to exhaustively describe the system investigated. The goal of the minimum entropy state can be achieved by application of iteration in a simulation procedure and linking sets of auxiliary variables introducing conceptually developed probability models or reference images. Consequently, this approach can contribute to the reduction of investment risk assessment at the stage of planning exploration fieldwork (Wellman and Regenauer-Lieb, 2012). Adding the conceptual ideas of lithofacies relations using 3D probability cubes significantly improved the results of the analysis of Shannon entropy which indicates the relatively greatest uncertainty only for the zone of direct lithofacies contacts and thin intercalations. Application of a looping procedure in simulation runs showed that consecutive probability models computed from multiple realizations are characterized by reduced Shannon entropy.

The comparison of many discrete variable simulation methods, effective in the reconstruction of specific depositional systems with different inputs of secondary data, allowed for the verification of hypotheses describing the spatial organization of lithofacies, and to some extent for the description of the depositional processes of amber accumulations. The amount of information contained in the source data depends on many factors, i.e. spatial configuration, measurement error, nature of the process analysed, measurement methodology and the scale of the area covered by the analysis. The influence of these elements can be quantitatively evaluated by Shannon entropy.

Acknowledgements. I would like to thank an anonymous reviewers for many helpful suggestions and efficient editorial handing, which improved the quality of manuscript.

REFERENCES

Allard, D., Comunian, A., Renard, P., 2011. Probability aggregation methods in geoscience. Research Report No. 48. Unite Biostatistique et Processus Spatiaux Institut National de la Recherche Agronomique Avignon, France.

Armstrong, M., Galli, A., Beucher, H., Le Loc'h, G., Renard, D., Doligez, B., Eschard, R., Geffroy, F., 2011. Plurigaussian Simulations in Geosciences. Springer, Berlin, Heidelberg.

Bianchi, M., Pedretti, D., 2018. An entrogram-based approach to describe spatial heterogeneity with applications to solute transport in porous media. Water Resources Research, **54**: 4432–4448.

Boucher, A., 2011. Strategies for modeling with multiple-point simulation algorithms. Closing the Gap. Gussow Geoscience Conference.

Caers, J., 2011. Modeling Uncertainty in the Earth Sciences. Wiley Blackwell.

Caers, J., Zhang, T., 2002. Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. AAPG Memoir, **80**: 383–394.

Chiogna, G., Hochstetler, D.L., Bellin, A., Kitanidis, P.K., Rolle, M., 2012. Mixing, entropy and reactive solute transport. Geophysical Research Letters, **39**: L20405.

Comunian, A., Allard, D., Renard, P., 2012. A review of probability aggregation methods in Earth sciences. The International Geostatistics Congress, Oslo, Norway, June 11–15.

Czuryłowicz, K., 2013. Geologiczno-złożowe uwarunkowania nagromadzeń bursztynu w utworach paleogeńskich w rejonie Parczewa i Lubartowa (in Polish). Ph.D. Thesis, University of Warsaw.

Czuryłowicz, K., Lejzerowicz, A., Kowalczyk, S., Wysocka, A., 2014. The origin and depositional architecture of Paleogene quartz-glauconite sands in the Lubartów area, eastern Poland. Geological Quarterly, **58** (1): 125–144.

Deutsch, C.V., 1998. A short note on cross validation of facies sumilation methods. Centre for Computational Geostatistics. Annual Report no. 1.

Deutsch, C.V., 2002. Geostatistical Reservoir Modeling. Oxford University Press.

Deutsch, C.V., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and User's Guide. New York, Oxford University Press.

Doktor, M., Krawczyk, A., Mastej, W., 2010. Testing the randomness of lithostratigraphic successions with the Markov Chain methods. Annales Societatis Geologorum Poloniae, **80**: 163–166.

Doktor, M., Krawczyk, A.J., 2010. Model entropii dla węglonośnych utworów karbonu górnego Zagłębia Górnośląskiego i próba jego geologicznej interpretacji (in Polish). Geologia. Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, **36**: 37–47.

Falivene, O., Arbues, P., Gardiner, A., Pickup, G., Munoz, J.A., Cabrera, L., 2006. Best practice stochastic facies modeling from a channel-fill turbidite sandstone analog (the Quarry outcrop, Eocene Ainsa basin, northeast Spain). AAPG Bulletin, **90**: 1003–1029.

Gotway, C.A., Young, L.J., 2002. Combining incompatibile spatial data. Journal of the American Statistical Association, **97**: 632–648.

Gringarten, E., Deutsch C.V., 2001b. Teacher's aide: variogram interpretation and modelling. Mathematical Geology, **33**: 507–534.

Hattori, I., 1976. Entropy in Markov chains and discrimination of cyclic patterns in lithologic successions. Mathematical Geology, **8**: 477–497.

Harding, A., Strebelle, S., Levy, M., Thorne, J., Xie, D., Leigh, S., Preece, R., Scamman, R., 2004. Reservoir facies modelling: new advances in MPS. In. Geostatistics Banff (eds. O. Leuangthong and C.V. Deutsch) **1**: 559–568. Springer, Netherlands.

Hong, S., Ortiz, J.M., Deutsch, C.V., 2007. Integration of Disparate Data with Logratios of Conditional Probabilities. Petroleum Geostatistics 2007. Cascais, Portugal: 10–14.

Hu, L.Y., Chugundova, T., 2008. Multiple-point statistics for modeling subsurface heterogeneity: a comprehensive review. Water Resources Research, **44**: 14.

Journel, A., 2002. Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. Mathematical Geology, **34**: 573–596.

Kelkar, M., Perez, G., 2002. Applied Geostatistics for Reservoir Characterization. Society of Petroleum Engineers Inc. Richardson, Texas.

Kramarska, R., Kasiński, J.R., 2008. Prognozy zasobowe bursztynu w Polsce (in Polish). National Geological Archive. no 38/2009. Państwowy Instytut Geologiczny, Warszawa.

Krawczyk, A.J., 1979. Markov chain as stochastic model of flysch sedimentation. Annales Societatis Geologorum Poloniae, **49**: 155–164.

Krawczyk, A.J., 1980. Some characteristics of the sedimentation process of the Podhale Flysch. Annales Societatis Geologorum Poloniae, **50**: 55–98.

Labourdette, R., Hegre, J., Imbert, P., Insalaco, E., 2008. Reservoir-scale 3D sedimentary modelling: approaches to integrate sedimentology into a reservoir characterization workflow. Geological Society Special Publications, **309**: 75–85.

Levy, M, Harris, P.M., Strebelle, S., 2008a. Carbonate reservoir modeling using Multiple-Point Statistics (MPS)/Facies Distribution Modeling (FDM). Search and Discovery Article #40293

Levy, M, Harris, P. M., Strebelle, S., Rankey, C., 2008b. Geomorphology of carbonate systems and reservoir modeling: Carbonate Training Images, FDM Cubes, and MPS Simulations. Search and Discovery Article #40289

Liu, Y., Harding, A., Gilbert, R., Journel, A., 2005. A workflow for multiple-point geostatistical simulation. In: Geostatistics Banff 2004 (eds. O. Leuangthong and C.V. Deutsch) **1**: 245–254. Springer, Netherlands.

Mallet, J.L. ed., 2002. Geomodeling. Applied Geostatistics Series. Oxford University Press.

Mallet, J.L., 2004. Space-time mathematical framework for sedimentary geology. Mathematical Geology, **36**: 1–32.

Mastej, W., 2002. An application of Markov Chain analysis to studies of lithofacies sequences in the alluvial fans from the "Bełchatów" lignite deposit (Poland). Annales Societais Geologorum Poloniae, **72**: 271–282.

Namysłowska-Wilczyńska, B., 2006. Geostatystyka: teoria i zastosowania (in Polish). Oficyna Wydawnicza Politechniki Wrocławskiej.

Olea, R.A., 1995. Fundamentals of semivariogram estimation, modeling, and usage. AAPG Computer Applications in Geology, **3**: 27–36.

Purkis, S., Vlaswinkel, B., Gracias, N., 2012. Vertical-to-lateral transitions among Cretaceous carbonate facies – a means to 3D framework construction via Markov analysis. Journal of Sedimentary Research, **82**: 232–243.

Ravenne, C., 2002. Sequence stratigraphy evolution since 1970. Comptes Rendus Palevol, **1**: 415–438.

Ravenne, C., Galli, A., Doligez, B., Beucher, H., Eschard, R., 2002. Quantification of facies relationships via proportion curves. In: Geostatistics Rio 2000 (eds. A. Armstrong, C. Bettini, N. Champigny, A. Galli and A. Remacre): 19–39. Springer.

Schwarzacher, W., 1969. The use of Markov Chains in the study of sedimentary structures. Mathematical Geology, **1**: 17–39.

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

Stach A., 2009. Analiza struktury przestrzennej i czasoprzestrzennej maksymalnych opadów dobowych w Polsce w latach 1956–1980 (in Polish). Adam Mickiewicz University Press, Poznań. Seria Geografia: 85.

Staňová, S., Soták, J., Hudec, N., 2009. Markov chain analysis of turbiditic facies and flow dynamics (Magura Zone, Outer Western Carpathians, NW Slovakia). Geologica Carpathica, **60**: 295–305.

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology, **34**: 1–22.

Strebelle, S., Levy M., 2008. Using multiple-point statistics to build geologically realistic reservoir models: the MPS/FDM workfow. Geological Society Special Publications, **309**: 67–74.

Strebelle, S., Thorne, J.A., Harding, A.W., Levy, M., Xie, D., 2006. Method for making reservoir facies model utilizing a training image and geologically interpreted facies probability cube. United States Patent Application Publication. Pub. No.: US 2006/0041409A1.

Strzelczyk, G., Danielewicz, B., 1990. Dokumentacja geologiczna w kategorii C2 złoża kruszywa naturalnego (piaski budowlane czwartorzędowe) wraz z określeniem występowania bursztynu w utworach trzeciorzędowych w rejonie Górka Lubartowska (in Polish). National Geological Archive. nr 484/92. Państwowy Instytut Geologiczny, Warszawa.

Volpi, B., Galli, A., Ravenne, C., 1997. Vertical proportion curves: a qualitative and quantitative tool for reservoir characterization. In Memorias del I Congreso Latinoamericano de Sedimentologia, Sociedad. Venezolana de Geólogos: 351–358.

Wędrowska, E., 2010. Wykorzystanie entropii Shannona i jej uogólnień do badania rozkładu prawdopodobieństwa zmiennej losowej dyskretnej (in Polish). Przegląd Statystyczny, **57**: 39–53.

Wellmann, J.F., Regenauer-Lieb, K., 2012. Uncertainties have a meaning: information entropy as a quality measure for 3-D geological models. Tectonophysics, **526–529**: 207–216.

Yarus, J.M., Chambers, R.L., Mauces, M., Shi, G., 2012. Facies simulation in practice: lithotype proportion mapping and plurigaussian simulation, a powerful combination. Ninth International Geostatistics Congress, Oslo, Norway.