1. RESEARCH METHODOLOGY

The research methodology, in addition to the specific methods for investigating the factors of geological processes, has a unified logistics framework for all research fields within Earth Sciences:

- Establishing the research ***objective*** of the geological process
- Identifying the ***factors*** that influence the development of the process, factors that shape the ***conceptual model*** of the process with its three components:
  - ***The space*** and ***time interval*** in which the geological process takes place
  - ***The parametric characteristics of the space*** in which the geological process occurs
  - ***The energy*** that sustains the development of the geological process
- ***Modeling the correlations*** between the research object and the factors that influence it.

The research is based on an ***investigation network*** whose configuration aims to capture, with minimal errors, the spatial and temporal variability of the regionalized variables of geological processes with:

- a minimum number of observation points
- a minimum frequency of measurements at the investigation points

The configuration of the network depends on the number of regionalized variables and the permissible error for evaluating the spatial distribution of all the monitored regionalized variables.

The minimum data required for configuring an investigation network based on a preliminary survey are:

- ***The coordinates*** of the initial observation points (N) of the investigation network ($i = 1,2,..., N$):
  - (X_i, Y_i)
- ***The values of the monitored variable(s)*** at the initial moment in all observation points:
  - $V(T_o)$
- ***Time series of values*** at each observation point:
  - $V(X_1, Y_1, T_1), V(X_1, Y_1, T_2), ..., V(X_1, Y_1, T_{50})$
  - .....
  - $V(X_N, Y_N, T_1), V(X_N, Y_N, T_2), ..., V(X_N, Y_N, T_{50})$
- **Maximum accepted standard deviation*(KSD)*** ***(*** KSD : Kriging Standard Deviation) imposed by the maximum allowable error in evaluating the spatial and temporal distribution in the investigated area, corresponding to an assumed risk.

1.1. Entropy of investigation networks

The use of Shannon entropy (***Fig.1***) in the design of monitoring networks has the following objectives:

- Evaluation of the ***2D average uncertainty*** regarding the spatial variability of regionalized ***random variables***, as a preliminary stage in applying the topo-probabilistic ***method of the fictitious point*** (D. Scrădeanu et al., 2001, 2003), to improve the efficiency of monitoring networks.
- Evaluation of the ***1D entropy of Markov chains***, used to determine the ***sampling interval of time series*** for the monitored variables.

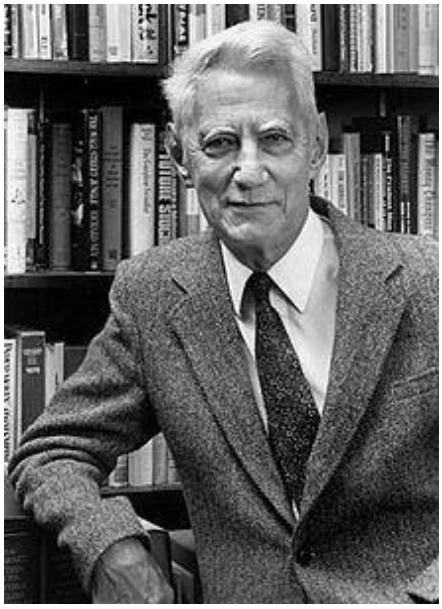

Fig.1.*Claude Shannon, părintele teoriei informaţiei*

The calculation methodology and applications from paragraphs §1.2 and §2.2 are supplemented with two Excel files posted on the website dedicated to the design of monitoring networks:

1. The file **ENTROPIE_2D.xls** allows the calculation of 2D entropy for a series of 55 values that must be placed on the **GREEN ZONE** of column V(To) following a similar procedure, which is described in detail for 1D entropy.
2. The file **ENTROPIE_1D.xls** allows the calculation of 1D entropy for a series of 50 values separated into three value groups (A, B, C), following the procedure below:

a. The **ENTROPIE_1D.xls** file is opened, applying the methodology for a series of 50 values placed in column V(Ti) in the **GREEN ZONE**.
b. Delete the values of the test series.
c. Place the series of values for which the 1D entropy calculation is desired in the **GREEN ZONE** column.

NOTE:

a. For the two excel files go to: http://www.ahgr.ro/specialisti/daniel-scradeanu/conducere-doctorat/rms/conf_network.aspx
b. The three value groups (A, B, C) are established by dividing the range of variation of the studied variable into three equal intervals (other working variants can be chosen by appropriately modifying the limits of the intervals for the three states: A, B, C).
c. The calculation algorithm is explained in paragraphs **§1.2** and **§2.2.**
d. If the series of values exceeds 50/55 values, the calculation formulas from columns **MNOPQRS/ M, N, …. , AB** should be extended up to the row corresponding to the last value of the processed value series.

1.1.1. The 2D entropy of monitoring networks.

1.1.1.1. *Basic concepts.*

Shannon's entropy, inspired by thermodynamics, in a ***probabilistic*** context, is a measure of the information contained in the distribution of the ***regionalized variable*** (V), what is to be monitored at the stations of the monitoring network ($i = 1,2,..., N$; ***Fig.1.1***):

$$V: \begin{pmatrix} V_1 & \dots & V_N \\ p_1 & \dots & p_1 \end{pmatrix} \qquad H(V) = -\sum_{i=1}^{i=N} p_i \cdot log_2(p_i) \tag{1.1}$$

p_i it is the probability of achieving the value from the station V_i .

For the calculation of entropy, the ***logarithm*** is used to allow the summation of the uncertainties of independent variables (U, V): $H(U, V) = H(U) + H(V)$. The unit of measurement for entropy is the ***shannon/bit***, if the base of the logarithm used in the calculation is 2. The informational content of an event with a probability p=1/2 is 1 shannon:

$$1sh = -\sum_{i=1}^{i=2} \frac{1}{2} \cdot log_2\left(\frac{1}{2}\right) = 1$$

To illustrate the calculation of Shannon entropy (H(V)) in a probabilistic context, applied to a monitoring network, we consider ***two extreme situations*** (**Fig. 1.1**) in ***N=12 observation points*** of a monitoring network that identify:

a. ***N=1 possible state/value*** of the variable V: A (**Fig. 1.1.a**));

b. ***N=12 equally possible states/values*** of the variable V: A, B, C, D, E, F, G, H, I, J, K, L (**Fig. 1.1.a**)).

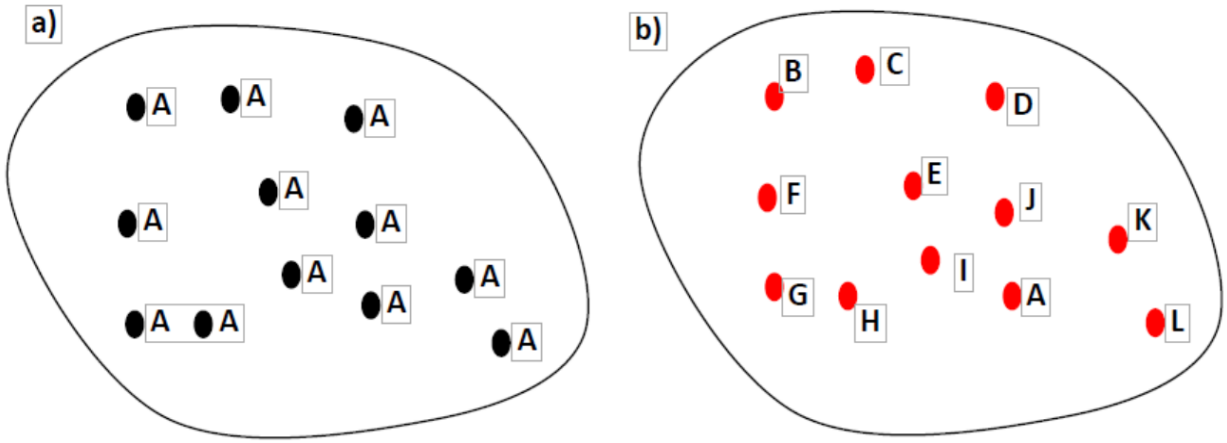

Fig. 1.1*. Two extreme situations for illustrating the calculation of Shannon entropy in a probabilistic context, applied to monitoring networks*.

The tables of the regionalized variable (V) from the 12 observation points, at ***a given moment*** (t_j), corresponding to the two extreme situations are:

a) **Complete knowledge of the variable**, when ***only one state is present*** (state A) with unit probability (uniform distribution):

$$V:\begin{pmatrix} A & A & A & A & A & A & A & A & A & A & A & A \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix};\ p_A = \frac{12}{12} = 1$$

a situation in which the entropy is MINIMAL and equal to ZERO:

$$H(V) = -\sum_{i=1}^{i=12} p_i \cdot \log_2 p_i = -\sum_{i=1}^{i=12} 1 \cdot \log_1 (1) = 0$$

b) **maximum uncertainty** about the state of the variable, when ***all 12 possible states*** (A, B, C, D, E, F, G, H, I, J, K, L) have equal probabilities:

$$V:\begin{pmatrix} A & B & C & D & E & F & G & H & I & J & K & L \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \end{pmatrix};\ p_A = p_B = p_C = \ldots p_L = \frac{1}{12}$$

a situation in which the entropy is MAXIIMAL and equal to $\log_2 12$:

$$H(V) = -\sum_{i=1}^{i=12} p_i \cdot \log_2 p_i = -\sum_{i=1}^{i=12} \frac{1}{12} \cdot \log_2 (\frac{1}{12}) = \log_2 12 = 3.58$$

The variation of entropy between the two extremes is similar to the variation of the function y=x·logx.

For calculations (in Excel), according to the graph (***Fig. 1.2***), **we will consider**:

$$0 \cdot log_2(0) = 0$$

Where N is the number of distinct states/values of the variable V . For details, see Keth Konrad, Probability Distribution and Maximum Entropy, [4].

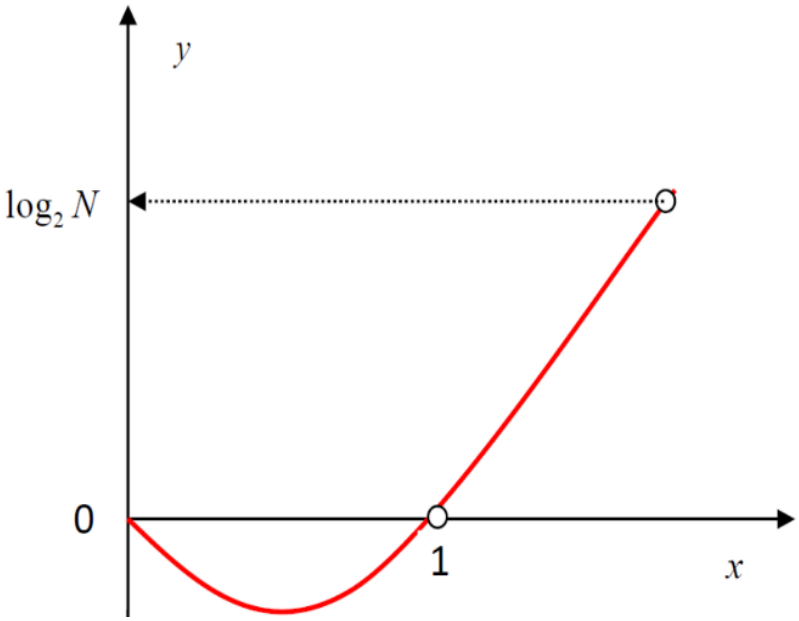

Fig. 1.2. *The graph of the function y = x log(x)*
(after K. Conrad with additions)

1.1.1.2. Application 2D

Calculate the ***2D entropy*** of a monitoring network consisting of ***55 observation points*** based on the values V(T0); (**Table 1.1**).

Tabele 1.1. *The values of the variable V at moment T_0 in the 55 monitoring points.*

NR	X	Y	V(To)	NR	X	Y	V(To)
1	79.23	535.13	24.85	29	862.42	41.88	68.30
2	125.20	577.56	30.38	30	867.72	149.72	73.61
3	54.48	570.49	27.34	31	973.80	20.66	86.13
4	43.87	480.32	19.35	32	873.03	317.67	85.17
5	134.04	404.30	16.41	33	770.49	365.41	72.73
6	171.16	538.66	28.73	34	706.84	421.98	68.34
7	96.91	434.36	17.29	35	625.52	453.80	60.38
8	40.34	312.37	8.39	36	530.05	489.16	52.37
9	123.43	303.53	9.93	37	473.48	551.04	52.46
10	58.02	365.41	11.69	38	650.27	565.18	74.83
11	174.70	215.13	7.64	39	657.34	499.77	68.91
12	54.48	172.70	2.97	40	706.84	545.74	80.57
13	248.95	77.24	6.68	41	756.34	566.95	90.28
14	252.49	255.80	12.99	42	966.73	178.01	91.98
15	319.67	114.36	11.40	43	970.26	338.89	103.98
16	266.63	317.67	17.11	44	922.53	370.71	98.05
17	400.99	41.88	15.18	45	970.26	271.71	98.73
18	132.27	34.81	1.83	46	835.90	598.77	106.81
19	522.98	33.04	25.30	47	869.49	572.25	109.40
20	565.41	98.45	31.32	48	906.62	593.47	118.53
21	662.64	10.06	39.82	49	982.64	529.82	125.22
22	660.88	75.47	41.37	50	966.73	490.93	117.86
23	423.97	475.02	40.42	51	940.21	467.95	110.50
24	56.25	36.57	0.46	52	869.49	444.96	95.69
25	22.66	110.83	1.12	53	957.89	406.07	107.65
26	512.37	381.32	41.26	54	956.12	554.58	123.03
27	0.00	0.00	0.00	55	1000.00	600.00	137.05
28	765.18	71.93	54.88				

Solution 2D:(ENTROPIE_2D.xls)

For the calculation of ***2D entropy***, three value groups are defined by dividing the range of variation into three equal intervals (***Fig. 1.3***):

- A $\in [0;45,69)$
- B $\in [45,69;91,37)$
- C $\in [91,37;137,06]$

The number of value groups is defined based on the level of detail desired for understanding the global average variation of the variable over the monitored area. Comparing the entropy of the same variable for a different number of value groups is used to detect the average continuity in the monitored space.

Spatial zoning of the monitored area can be performed to identify regions with different uncertainties for the same number of groups and the same variable. The entropy of the variable V(T0), calculated for the three groups (***Fig. 1.3***) using Shannon's formula (1.1), is:

$$H(V) = -\sum_{i=1}^{i=3} p_i \cdot \log_2 p_i = -\left[\frac{27}{55}\cdot\log_2\left(\frac{27}{55}\right)+\frac{14}{55}\cdot\log_2\left(\frac{14}{55}\right)+\frac{14}{55}\cdot\log_2\left(\frac{14}{55}\right)\right] = 1{,}51$$

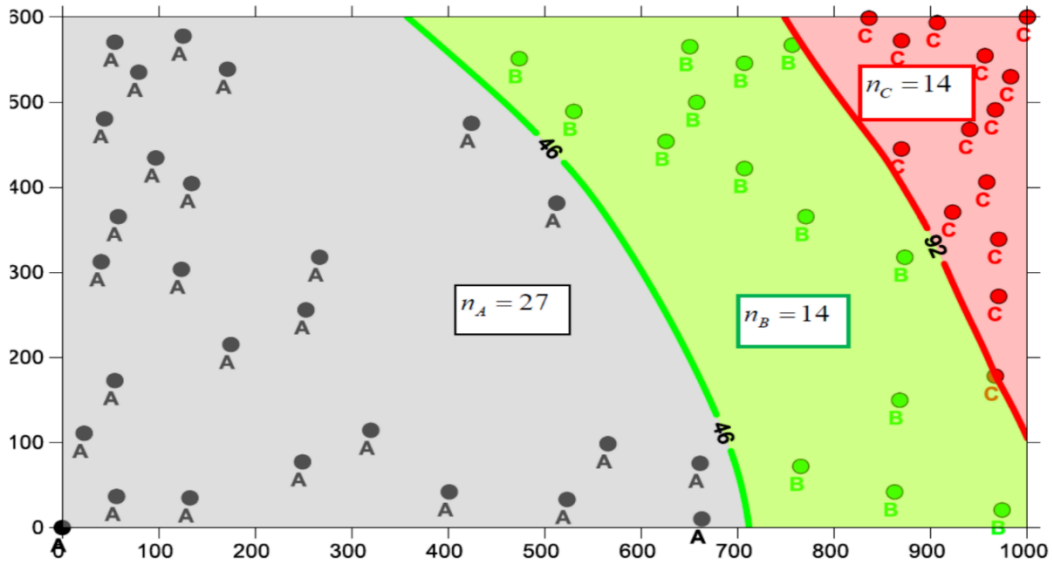

Fig. 1.3. *Position of the 55 observation points in groups A, B,*

1.1.1.3. Conclusions 2D

The "high" value of the monitoring network's entropy (relative to HMAX=$\log_2(3)$= 1.59) indicates a **HIGH UNCERTAINTY** regarding the spatial variability of the variable V(T_0), suggesting the need to **supplement the network with additional observation points**. An efficient way to complete the monitoring network can be achieved using the **fictitious point method**.

1.1.2. 1D entropy of monitoring networks

The variation over time of the value of a variable at a station in the monitoring network is the result of the interference of a large number of factors, with the number of factors being greater the more complex the process the measured variable reflects.

The piezometric level of a groundwater aquifer, for example, is the result of the interaction of a large number of factors (with groundwater flow being a complex process):

- Precipitation in the aquifer recharge area
- Air temperature
- Vegetation cover/land use type
- Land slope
- Lithological composition of formations in the vadose zone
- Thickness of permeable deposits
- Moisture in the vadose zone
- Hydraulic conductivity of the aquifer
- etc.

Building a ***functional model*** to simulate the variation of the piezometric level of the aquifer in relation to all the factors that condition it is burdensome. To study the variation over time of variables that result from a complex process, statistical models are used, where the time variation of the resulting variable values is modeled, and the factorial variables are enclosed in a *"**black box***". ***Markov chains*** are a suitable statistical model for the analysis of complex time series, a model that involves the evaluation of two quantities:

- The ***probability of transitioning*** from one state to another
- The ***entropy*** associated with each state change

The evaluation of the entropy of Markov chains is used to select the sampling interval for the time series of monitored variables. The sampling interval (Δt) of the time series for the monitored variables is:

- ***Inversely proportional*** to the value of ***the entropy*** of the transition probability matrix
- ***Directly proportional*** to ***the admissible error*** (ε(α)) at a given ***assumed risk*** (α).

1.1.2.1. *Basic concepts*

A **Markov process/Markov chain** (***Fig. 2.1***) is a stochastic process in which the current state/value retains all the information about the entire evolution of the process. The **stochastic/random process** is a model that quantifies the uncertainty of the time evolution of a complex natural process, conditioned by a large number of factorial variables. This means that although the initial state of a process is known, there are multiple possible continuations of the process, with some paths being more likely than others.

***Fig. 2.1**. Andrei Markov, a Russian mathematician known for the study of stochastic processes.*

The ***transition matrix*** of a Markov process is the operational tool that allows the identification of the "correlational" component of the stochastic process (Daniel Scrădeanu, 1995, Geological Informatics, Univ. Buc. Press).

A Markov process with only two distinct values/states (**A**, **B**: Ns=2Ns = 2) of the main variable, described by a sequence of 10 values (***Fig. 2.2***), is characterized by:

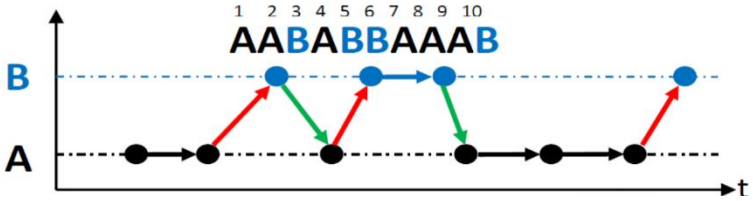

***Fig. 2.2**. Markov chain with two states (A, B).*

- The total number of available values (**A** + **B**):
  - N=10
- The absolute frequency of the two states (**A** and **B**: Ns=2):
  - **A**: nA=6
  - **B**: nB=4
- The average probability for each state:
  - **A**: $p_A = \frac{nA}{N} = \frac{6}{10}$
  - **B**: $p_B = \frac{nB}{N} = \frac{4}{10}$
- The types of transitions:
  - **Tranziţie A->A**
  - **Tranziţie:A->B**
  - **Tranziţie:B->B**
  - **Tranziţie:B->A**
- Total number of transitions (Ntr = 3 + 3 + 1 + 2 = 9) (***Table 2.1***):

- nAA=3
- nAB=3
- nBB=1
- nBA=2

- The transition probability matrix (TPM) (***Table 2.2***)
- The diagram of the transition probability matrix (***Fig. 2.3***)

Tabel 2.1. *Number of transitions*

		A	B	Număr total pe rând (NTR)
NT:	**A**	3	3	6
	B	1	2	3

Tabel 2.2. *Transition probability matrix*

		A	B	NTR
MT:	**A**	3/6	3/6	1
	B	1/3	2/3	1

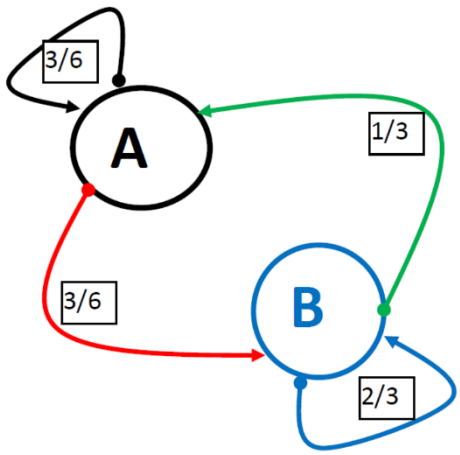

Fig. 2.3. *Diagram of the transition probability matrix (TPM(A,B))*

The uncertainty associated with the transition probability matrix is measured by the ***entropy of the transition probability matrix*** (H(TPM)) calculated using the following relation:

$$H(MT) = -\left(p_A \cdot \left(p_{AA} \cdot \log_2\left(p_{AA}\right) + p_{AB} \cdot \log_2\left(p_{AB}\right)\right) + p_B \cdot \left(p_{BA} \cdot \log_2\left(p_{BA}\right) + p_{BB} \cdot \log_2\left(p_{BB}\right)\right)\right)$$

The entropy applied to the 4 transitions (***Fig. 2.2*** and ***Table 2.2***) is:

$$H(MT) = -\left(\frac{6}{10} \cdot \left(\frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right) + \frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right)\right) + \frac{4}{10} \cdot \left(\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) + \frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right)\right)\right) = 0{,}97$$

or in the general form for Ns states, it is:

$$H(MT) = -\sum_{i=1}^{i=Ns} p_i \cdot \sum_{j=1}^{j=Ns} p_{ij} \cdot \log_2\left(p_{ij}\right)$$

with the extreme values:

- H(TPM)=0: When the process is in a single state, and ***the knowledge of the process's evolution is complete***.
- H(TPM)=1: When all probabilities in the transition matrix are equal, and ***the uncertainty about which state the process will transition to is maximal***.

1.1.2.2. Application 1D

Calculate the ***entropy of the transition matrix*** for an observation point in a monitoring network, based on a series of 50 values measured at ***equal time intervals*** (Δt=7 days) (***Table 2.3***).

Tabel 2.3. *Series of measured values at an observation point in the monitoring network.*

Nr	V(T)	Nr	V(T)	Nr	V(T)	Nr	V(T)
1	17.07	14	17.5	27	25.29	40	20.48
2	19.06	15	28.04	28	28.12	41	25.02
3	17.75	16	20.48	29	31.94	42	26.56
4	25.34	17	28.55	30	19.89	43	26.91
5	19.96	18	17.32	31	28.38	44	17.06
6	30.23	19	19.18	32	18.11	45	26.99
7	20.88	20	30.49	33	23.51	46	32.91
8	27.37	21	18.05	34	17.16	47	28.85
9	31.37	22	27.25	35	25.14	48	24.98
10	17.3	23	32.64	36	28.64	49	26.52
11	24.23	24	26.16	37	32.84	50	30.28
12	26.46	25	26.02	38	31.25		
13	29.91	26	32.24	39	18.29		

1D Solution: (ENTROPIE_1D.xls)

The steps for processing to calculate the entropy of the transition matrix are:

- **Defining the value groups**, dividing the range of variation into equal intervals (***Fig. 2.4***). For the N = 50 available values (***Table 2.3***), the selection range is divided into three value groups (A, B, C; Δ=5.28\Delta = 5.28):
  - $A \in [V\min = 17{,}07; V\min + \Delta = 22{,}35)$
  - $B \in [22{,}35; 27{,}63)$
  - $C \in [27{,}63; 32{,}91]$

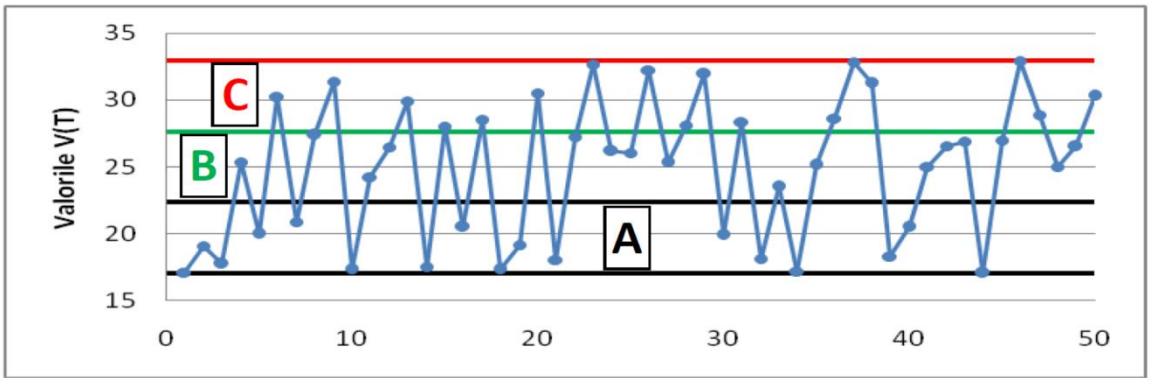

Fig. 2.4. *The series of 50 values of the variable V (measured at an interval: Δt=7 days)*

NOTE: The number of value groups is chosen based on the level of detail desired for understanding the variation of the variable over time during the monitored period.

- **Graphical representation for verifying data grouping (Fig. 2.4):**
  - The variation of the variable over time;
  - The separation limits for the defined value groups.
- Calculation of absolute frequencies for the three value groups:
  - nA=17
  - nB=16
  - nC=17

- Calculation of relative frequencies (average probabilities) for the three value groups:
  - **A**: $p_A = \frac{nA}{N} = \frac{17}{50} = 0{,}34$
  - **B**: $p_B = \frac{nB}{N} = \frac{16}{50} = 0{,}32$
  - **C**: $p_C = \frac{nC}{N} = \frac{17}{50} = 0{,}34$
- Establishing the sequence of the 50 coded values according to their classification into the three value groups (***Table 2.4***)

Tabel 2.4. *The sequence of the 50 values, coded as A, B, C.*

Nr	V(T)	Nr	V(T)	Nr	V(T)	Nr	V(T)
1	A	14	A	27	B	40	A
2	A	15	C	28	C	41	B
3	A	16	A	29	C	42	B
4	B	17	C	30	A	43	B
5	A	18	A	31	C	44	A
6	C	19	A	32	A	45	B
7	A	20	C	33	B	46	C
8	B	21	A	34	A	47	C
9	C	22	B	35	B	48	B
10	A	23	C	36	C	49	B
11	B	24	B	37	C	50	C
12	B	25	B	38	C		
13	C	26	C	39	A		

- Calculation of the absolute frequency of the nine possible types of transitions (***Table 2.5***).

Table 2.5. *Number of Transition*

		A	**B**	**C**	**Total number per row**
NT:	**A**	4	8	5	17
	B	3	5	8	16
	C	9	3	4	16

- Calculation of the transition probability matrix (***Table 2.6***).

Table 2.6. *Transition Probability Matrix*

		A	**B**	**C**	**Sum of probabilities per row**
NT:	**A**	0.24	0.47	0.29	1

	B	0.19	0.31	0.50	1
	C	0.56	0.19	0.25	1

- Creating the diagram of the transition probability matrix (***Fig. 2.5***).

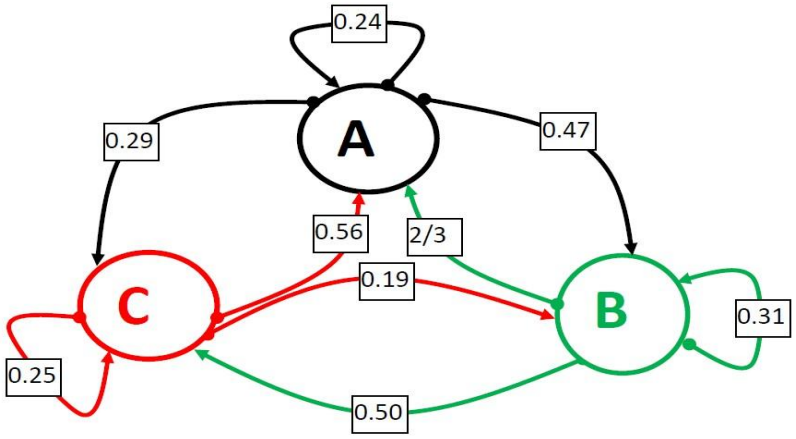

Fig. 2.5. *Diagram of the transition probability matrix for three*

- Calculation of ***the entropy of the transition probability matrix*** (***Table 2.6***) for the sequence of the 50 values, coded as A, B, C (Ns = 3;***Tabel 2.4***).

$$H(MT) = -\sum_{i=1}^{i=Ns} p_i \cdot \sum_{j=1}^{j=Ns} p_{ij} \cdot \log_2(p_{ij}) = 1.47$$

1.1.2.3. Conclusions 1D

In relation to the maximum entropy of the transition probability matrix for a Markov chain with 3 distinct states (H_max = $\log_2 3$ = 1.58), the entropy H(MT) = 1.47 indicates ***a high degree of uncertainty*** regarding the state the process will transition to from a known state. Such a situation recommends **reducing the time interval** (Δt) between two consecutive measurements for the studied variable.

It is recommended to calculate the transition matrix for all observation points of the monitoring network and identify areas where it is necessary to adjust the time interval between consecutive measurements, in order to increase the understanding of the temporal evolution of the monitored variable.

1.2. The method of the fictitious point

The fictitious point method must establish additional observation points that ensure ***the reduction of the standard deviation*** of estimation through **kriging** (Kriging Standard Deviation) across the entire investigated surface, below the maximum allowed value, specified as the validity criterion for the investigation results.[The fictitious point method: "Scrădeanu, D, Popa, R.- APPLIED GEOSTATISTICS, 2001, EUB]

1.2.1. Elementary concepts

The reduction of the zonal estimation error can only be achieved by increasing the density of observation points. The evaluation of the effect of placing a new observation point is possible through the variance of the estimation error, which depends only on the variogram model and the distance between the point where the estimation is made and the observation points (it does not depend on the measured values!!!).The precision gain (CP(po)) associated with an estimation point (po) by introducing a fictitious point in its area of influence is estimated using the following relationship:

$$CP(p_o) = \frac{\left(\tilde{\sigma}_R^2\right)_f - \left(\tilde{\sigma}_R^2\right)_i}{\left(\tilde{\sigma}_R^2\right)_f}$$

in which:

$\left(\tilde{\sigma}_R^2\right)_f$ is the variance of the estimation error after introducing the fictitious point;

$\left(\tilde{\sigma}_R^2\right)_i$ is the variance of the initial estimation error.

The efficiency of placing new observation points is assessed based on the precision gain obtained. The fictitious point method operates in this stage in two directions:

- Elimination of inefficient locations from the monitoring network. For each observation point in the monitoring network, the precision gain is evaluated, and those that do not significantly contribute to reducing estimation errors are eliminated.
- Completion of the monitoring network. In areas with estimation errors greater than the allowed value, fictitious locations are placed, and their efficiency is calculated based on the precision gain they determine.

It is common to use maps with precision gain isolines, created by densifying the existing network with a regular grid of fictitious locations. The selection of locations for additional stations is made

based on a minimum precision gain, which is chosen depending on the level of detail required in the spatial estimates.

1.2.2. The steps for applying the fictitious point method.

The main steps of the fictitious point method are:

- Variographic analysis of the variable at the initial moment:V(To)
- Calculation of the surface variogram
- Estimation of spatial structure characteristics:
  - Variogram model
  - Anisotropy parameters:
    - Anisotropy ratio (R/r)
    - Orientation of the maximum continuity direction (θ)
- Estimation of the standard deviation distribution of interpolation (KSD)
- Identification of areas with interpolation standard deviations greater than the maximum allowed KSDmaxim.
- Determining the positions of the "fictitious" points required to reduce KSD below the maximum allowed value.
- Recalculation of KSD after the introduction of the "fictitious" points.
- Finalization of the monitoring network.
- Calculation and graphical representation of the estimation error distribution for the finalized investigation network, for a significance level of α = 10%, N, and KSD using the function in Excel:
  - $$\varepsilon(x_i, y_j, \alpha) = CONFIDENCE\ (\alpha, KSD(x_i, y_j), N)$$

in which (x, y) are the coordinates of the interpolation network nodes for KSD.

The finalization of the application of the ***fictitious point method*** is illustrated with:

- A map showing the position of the initial investigation points;
- A map of the initial KSD distribution highlighting areas where the maximum allowed KSD is exceeded;
- A map showing ***the position of the fictitious points*** that reduce the KSD to below the maximum allowed value;
- A map of the distribution of the variable values $V(x_i,y_i)$ and the estimation errors $\varepsilon(x_i,y_i,\alpha)$ for the final investigation network, completed with the fictitious points.

1.2.3. 2D Application

Optimizing the Hydrogeological Exploration Network of the Confined Aquifer in the Lignite Layer V in the Motru-Jiu Interfluve

Solution:

For the research of the distribution of hydrogeological parameters of the aquifers in the Motru-Jiu interfluve area (***Fig. 3.14***), 129 hydrogeological boreholes were made, exploring different aquifer horizons, in which selective hydrodynamic tests were conducted. Since the spatial distribution of hydrogeological parameters varies from one parameter to another, the optimization of the research network is done considering one of these parameters. Usually, the

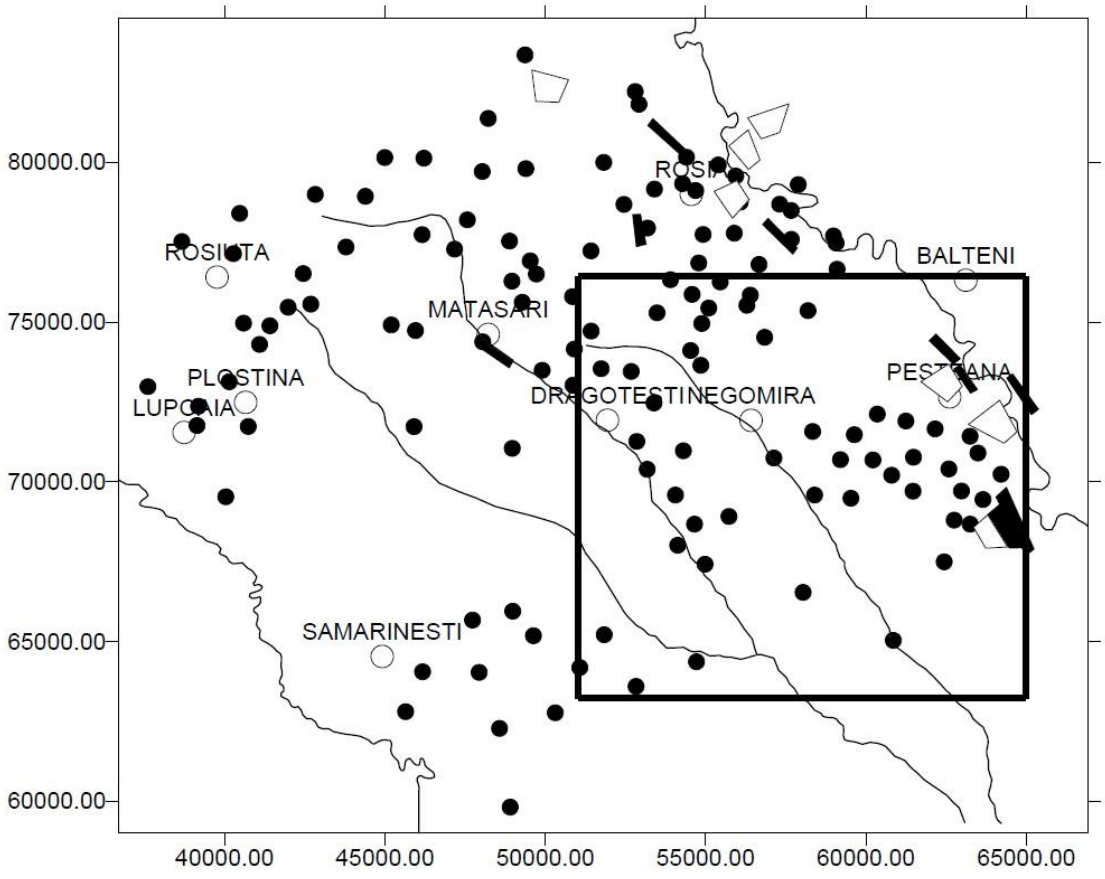

Fig. 3.14. *Location of hydrogeological boreholes for aquifer horizon research*

□ *- the selected sector for the application.*

most important parameter for the research is selected. In the chosen area, the transmissivity of the aquifer in the Lignite Layer V is an important parameter for determining the discharge potential of this aquifer, which has a regional extension.

For this aquifer, 39 transmissivity values are available, ranging from 10 to 130 m²/day, with a lognormal distribution and a skewness coefficient of 1.48. After eliminating 7 values considered unrepresentative for the available data selection, the distribution of the remaining 32 values was normalized through logarithmic transformation.

The surface variogram calculated for the logarithmic values indicates a weak anisotropy, which, given the determination errors of the spatial distribution of transmissivities (***Fig. 3.15***), is negligible. The omnidirectional variogram model used for estimating the spatial distribution of transmissivity is of the spherical type with:

- ***Pebble effect***: c0=0.56
- ***Sill***: c=3.00
- ***Influence radius***: r=7000 m

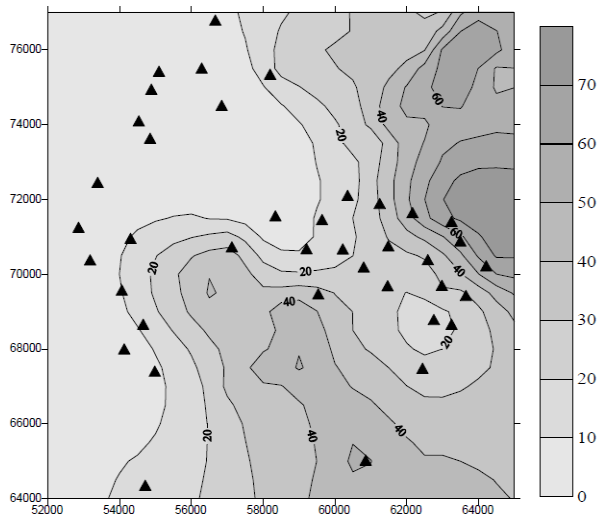

Fig. 3.15. *Distribution of transmissivity for the aquifer in the Lignite Layer V, calculated with normalized values.*

The maximum difference between the two estimates (due to overestimations) is 38 m²/day, which is 50% of the maximum estimated value.

The evaluation of the interpolation error distribution for estimating the transmissivity

Normalization of the transmissivity values distribution helps avoid overestimation. To illustrate the overestimation effect caused by the skewness of the original data distribution, the transmissivity distribution was estimated using the original, non-logarithmic values (***Fig. 3.16***).

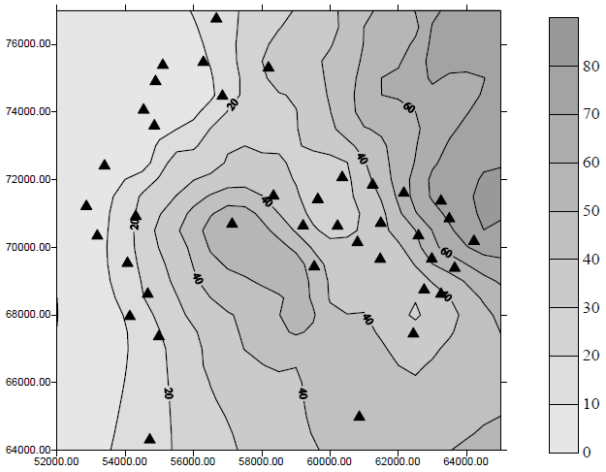

Fig. 3.16. *Distribution of transmissivity for the aquifer in the Lignite Layer V, calculated with unnormalized values.*

distribution with logarithmic values (the correct one, from ***Fig. 3.15***) was performed through zonal kriging on rectangular blocks of 200 x 300 m with 16 discretization points.

The contour map of the standard deviation calculated through kriging (***Fig. 3.18***) indicates maximum values located in the northwestern area of the interfluve and where the density of hydrogeological boreholes is lower. In the areas where boreholes were located and transmissivity was determined, the estimation standard deviation is less than 0.7 m²/day, while in the peripheral areas of the researched perimeter, it reaches maximum values of 2 m²/day.

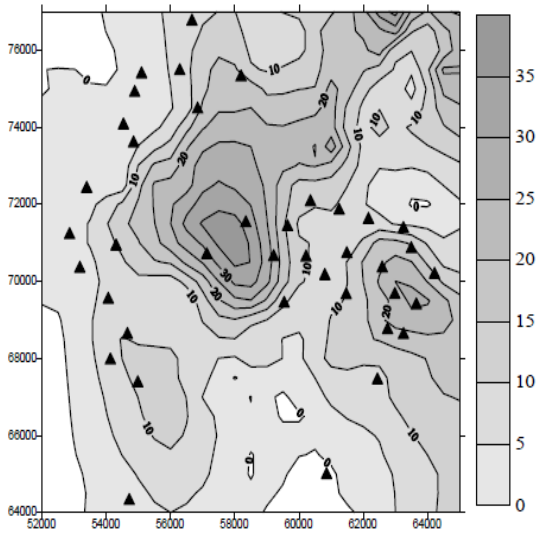

Fig. 3.17. *Distribution of the overestimations of transmissivity for the aquifer in the Lignite Layer V.*

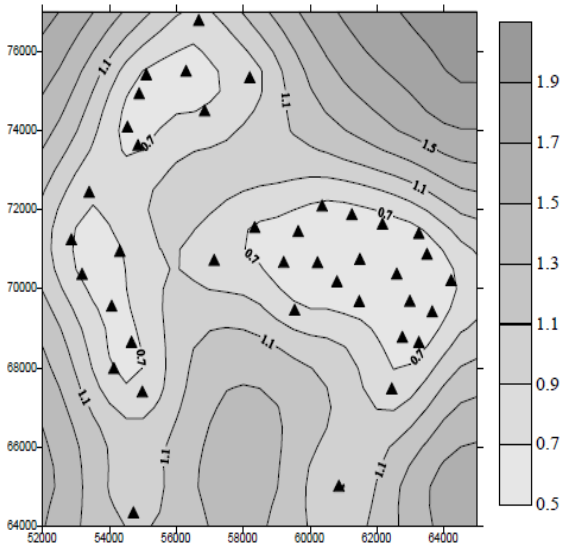

Fig. 3.18. *Distribution of the standard deviation of the transmissivity estimation for the aquifer in the Lignite Layer V.*

The evaluation of the effect of placing a regular network of fictitious observation points in the researched area (***Fig. 3.19***) is carried out using the fictitious point method and is expressed through the precision gain.

To quantitatively illustrate the effect, 380 fictitious observation points were placed in a square grid (19 columns and 20 rows, with a grid parameter of 650 m).

Using the same variogram model (spherical, with a pebble effect of 0.56, plateau of 3, and influence radius of 7000 m), the following was obtained:

- A maximum precision gain of 380% in the peripheral areas of the researched zone;
- A minimum precision gain of 20% in the areas of maximum borehole density (***Fig. 3.20***).

The map with the distribution of the precision gain (***Fig. 3.20***) is used to select the areas where the fictitious points are considered effective, meaning they provide a significant precision gain in estimating the transmissivity distribution.

Thus, if the interest is in the transmissivity distribution in the northeastern area, it is clear that placing observation points will result in a precision increase of over 300%, making their placement efficient. On the other hand, placing additional points in the area of borehole concentration in the central part will result in only a 20% precision gain, and their efficiency is questionable.

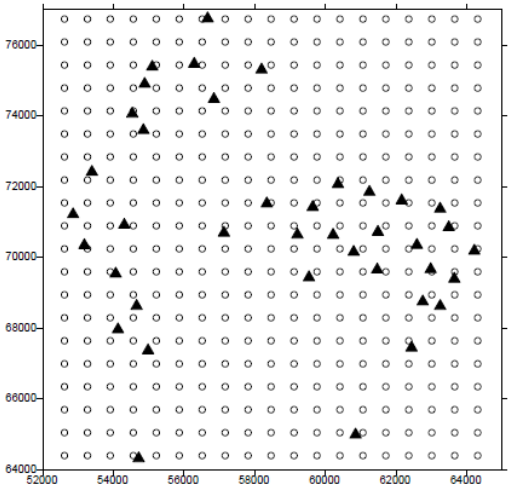

Fig. 3.19. *Distribution of fictitious points (○) and hydrogeological boreholes (▲) within the perimeter.*

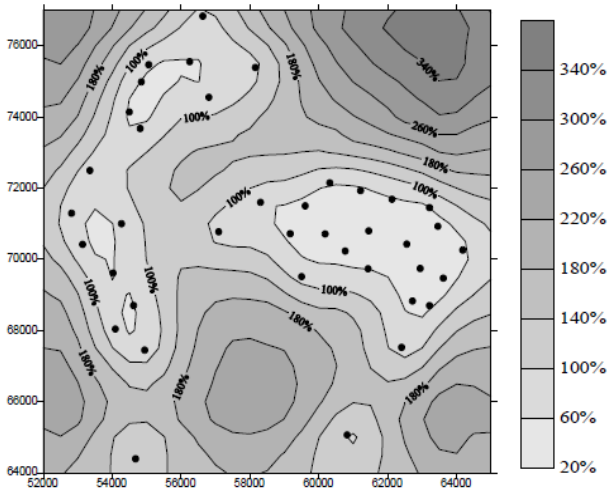

Fig. 3.20. *Distribution of the precision gain determined by the introduction of the 380 fictitious points.*

Of course, these percentage gains can be transformed through confidence intervals into standard deviation values in absolute terms (3.35), which may have clearer significance for those working with transmissivity values in numerical models for simulating aquifer dynamics.

COMMENTARY

Optimizing the research network is the operation that utilizes all the results of geostatistical modeling of spatial structures.

After completing all the processing steps, assuming that all have been correctly performed, we can express our agreement or disagreement with the result based on a complete quantitative foundation.

The main elements that must be considered are:

- ***Representativeness*** of the data for the spatial distribution of the studied variable.
- ***Certainty*** of the estimates, reflected in the estimation error values.
- ***Possibility*** of improving the precision of the estimates based on optimizing the research network.

And here we are, with the structure laid out before us.If we don't like how it looks, we can start over—but not just any way. There is control over imagination through the data used. Every estimate is characterized by a certain degree of model adequacy (the variogram model) and estimation precision (standard deviation of estimation).

If the result obtained does not align with expectations, under the condition that the data are considered representative, it is time to suspect the validity of the hypotheses we are trying to test.

The days when imagination and the intuition of a geologist solved most geological research problems are long gone.Correctly measured data and rigorously applied quantitative methods are the only means by which hypotheses can be confirmed or disproven. These are the foundations upon which the exploitation of oil fields is planned, the location for a water supply system for a town is decided, or whether a gold deposit is profitable or not.

Do not attempt to build a geological map without a well-founded quantitative methodology! Do not paint maps or geological sections with a brush, even if you're wielding the "brush" with a mouse!

Minimal Selective Bibliography

1. Keith Conrad, Probability distributions and maximum entropy (http://www.math.uconn.edu/~kconrad/blurbs/analysis/entropypost.pdf)
2. Isaaks, E.H., Srivasrava,M.R., Un introduction to Applied Geostatistics, New York, Oxford University Press, 1989.
3. Scrădeanu Daniel, Popa Roxana, [2001, 2003], Geostatistică aplicată, Editura Universităţii din Bucureşti
4. Scrădeanu Daniel, [1995], Informatrică geologică, Editura Universităţii din Bucureşti