Spatial self-arrangement of expanding structures.
3. Novel modeling routine

Jelena Nesterova [a]
Department of Geophysical Sciences, Faculty of Mathematics and Natural Sciences,
Klaipėda University, H. Manto st. 84, LT-92294 Klaipėda, Lithuania

Received 14 April 2010, in revised form 9 December 2010, accepted 10 December 2010

Abstract. This work is devoted to the search of correlations - similarities and differences - between indexes of metric and information organization in the complicated cartographic structures. The indicators give the required quality of landscape and represent the quality of residential environment indirectly. Research data are taken from geographical maps and analysed using *geographic information system* (GIS). Analysis of data is based on methods of mathematical statistics. Results of comparative research of two measures are reported in this study.

Citations: Jelena Nesterova. Spatial self-arrangement of expanding structures. 3. Novel modeling routine – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(9)** 2010 – Pp. 30-36.

Keywords: Spatial organization; Expanding Structures; Organized structures; Indices of organization; Informational organization; Self-arrangement; Metrical organization; Cartographic structures; GIS data.
PACS: 91.10.Da
Short title: Relations in cartographic structures - 3.

Introduction

Forests of definite size are analyzed in this work so their areal is constant as well. Previous two our publications represent an overview of paradigms [1] and novel model description [2] based on information indicators which are useful for such complicated modelling. The indicators give the required quality of landscape and represent the quality of residential environment indirectly. Research data is taken from geographical maps and analysed using *geographic information system* (GIS).

Purpose of this work is an analysis of data based on methods of mathematical statistics.

1. Real forests: methodology of description

Research data is taken from geographical maps and analysed using *geographic information system* (GIS). The analyzed segment of forests is shown in Fig. 1. Area of observation is selected randomly. Then we assign a binome to the functional area according to a model described in Ref. [2]. Fig. 2 represents dividing of layer into binary components.

In order to calculate the indicator of metric organization, we need to take the area of the binome in consideration. To find the area of it, we have to know whole perimeter that consists of outer perimeter of binome and inner perimeter of forest. Fractal dimension D_z of binome $f(z)$ is approaching the value of 2, and for area calculation we use the simplest form:

$$D_z \rightarrow 2 \tag{1}$$

$$f(z) = \left(\frac{P_z}{4}\right)^{D_z} \rightarrow \left(\frac{P_z}{4}\right)^2 \tag{2}$$

$$f(z) = \left(\frac{P_m + P_p}{4}\right)^2 \tag{3}$$

where P_m - perimeter of the binome - area of forest,
P_p – outer perimeter of binome.
It is shown graphically in Fig. 3. This way, by analyzing data from GIS, we may calculate indicators of organization.

2. Evaluation of organization in map layers

2.1. Metric organization

Index of metric organization M_i was described in Refs. [2-3]. We use the following routine for current modeling.

Index of metric organization. $f(x)$ – is a function describing a component of a forest layer (x_i – metrics of forests); $\overline{f}(x_1, ..., x_k)$ - function describing average value of a forest layer; k – amount of components in a layer;

[a]email: *jelena.nesterova@gmail.com*

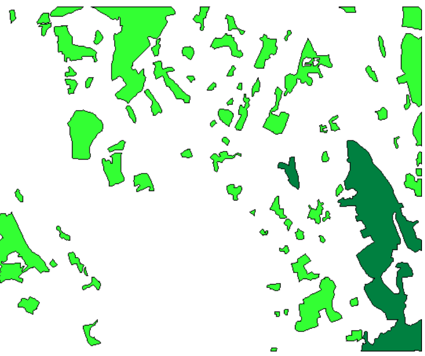

Fig. 1. Layer of digital map containing forest arrays.

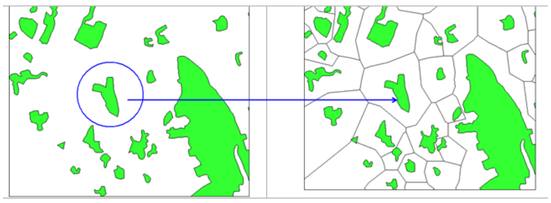

Fig. 2. Dividing of layer (see Fig. 1, center-right) into binary components.

$f(\overline{x_1}, ..., \overline{x_k})$ - function of component average values of a layer;
P_m - perimeter of a component;
P_p - outer perimeter of binome;
D_m - fractal dimension of forests;
S_m - area of a component of forest layer.

$$M_i = 1 - \left[\frac{\overline{f}(x_1, ..., x_k)}{f(\overline{x_1}, ..., \overline{x_k})}\right] \tag{4}$$

$$f(x) = \left\{\left(\frac{P_m}{4}\right)^{D_m} + \left(\frac{P_z}{4}\right)^2\right\} \tag{5}$$

$$P_z = P_m + P_p \tag{6}$$

$$D_m = \frac{\ln S_m}{\ln P_m - \ln 4} \tag{7}$$

$$\overline{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{P_m}{4}\right)^{D_m} + \left(\frac{P_z}{4}\right)^2\right\}_i \tag{8}$$

$$f(\overline{x}) = \left\{\left(\frac{\overline{P_m}}{4}\right)^{\overline{D_m}} + \left(\frac{\overline{P_z}}{4}\right)^2\right\} \tag{9}$$

To determine the difference of a structure over the time of evolution, geometrically correct forms of units should be used. Function describing one element of a layer when describing it by metrics x and y would look like:

$$f(x) = \overline{x} \cdot \overline{y} + r \cdot s_x \cdot s_y. \tag{10}$$

where $\overline{x}$ and $\overline{y}$ are averages of metrics, r - coefficient of

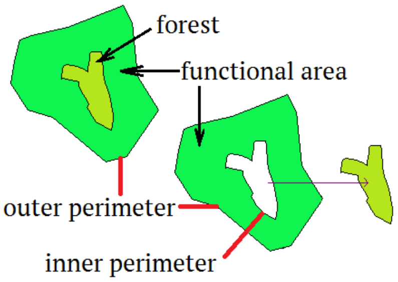

Fig. 3. In order to measure the area of binome, perimeters are calculated, where the perimeter of binome consists of outer and inner perimeters. Inner perimeter is the perimeter of a forest.

correlation of x and y values, s_x and s_y - standard deviations of metrics.

2.2. Informational organization

Three indexes of informational organization were described in Ref. [2]. For current modeling we use the following routine, described in Refs. [4-6].

First index of informational organization.
$H(X)$ - Shanon entropy;
$H_{\max}$ - value of entropy when all probabilities are the same.
N_k - amount of classes.
p_i - probability of a specific result of experiment.
$n(P_m)$ – amount of elements in a class sorted by a perimeter of forests.
$n(d_m)$ – amount of elements in a class sorted by fractal dimension of forests.
N – total amount of elements.
$S(P_m)$ - area of a class sorted by a perimeter of forests.
S - total area of forests.
$S(d_m)$ - area of a class, sorted by fractal dimension of forests.

$$R = 1 - \frac{H(X)}{H_{\max}} \tag{11}$$

$$H(X) = -\sum_{i=1}^{N_k} p_i \cdot \log(p_i) \tag{12}$$

$$H_{\max} = \ln N_k \tag{13}$$

$$p_i = \frac{n(d_m)}{N} \tag{14}$$

$$p_i = \frac{n(P_m)}{N} \tag{15}$$

$$p_i = \frac{S(d_m)}{S} \tag{16}$$

$$p_i = \frac{S(P_m)}{S} \tag{17}$$

Second index of informational organization.
$T(n(P_m), n(d_m))$ - negentropy of experiment;
$H(n(P_m))$ - entropy of experiment with parameter P_m;
$H(n(d_m))$ - entropy of experiment with parameter d_m;
n_i - number of elements in a class, sorted by perimeter;
N - total number of elements;
n_j - number of elements in a class, sorted by fractal dimension;
$H(n(P_m), n(d_m))$ - entropy of matrix $n(P_m) \times n(d_m)$;
$n_{i,j}$ - elements of matrix $n(P_m) \times n(d_m)$.

$$R = \frac{T(n(P_m), n(d_m))}{H(n(d_m))} \times 100\% \tag{18}$$

$$R = \frac{T(n(P_m), n(d_m))}{H(n(P_m))} \times 100\% \tag{19}$$

$$T(P_m, d_m) = H(n(P_m)) + H(n(d_m)) - H(n(P_m), n(d_m)) \tag{20}$$

$$H(n(P_m)) = -\sum_i \frac{n_i}{N} \log \frac{n_i}{N} \tag{21}$$

$$H(n(d_m)) = -\sum_j \frac{n_j}{N} \log \frac{n_j}{N} \tag{22}$$

$$H(n(P_m), n(d_m)) = -\sum_i \sum_j \frac{n_{i,j}}{N} \log \frac{n_{i,j}}{N} \tag{23}$$

Third index of informational organization.
$T(P_m, d_m)$ - negentropy of experiment;
$H(P_m)$ - entropy of experiment with parameter P_m;
$H(d_m)$ - entropy of experiment with parameter d_m;
n_i - area of a class, sorted by perimeter;
N - total area;
n_j - area of a class, sorted by fractal dimension;
$H(P_m, d_m)$ -entropy of a matrix $P_m \times d_m$;
n_{ij} - elements of matrix $P_m \times d_m$.

$$R = \frac{T(P_m, d_m)}{H(d_m)} \times 100\% \tag{24}$$

$$R = \frac{T(P_m, d_m)}{H(P_m)} \times 100\% \tag{25}$$

$$T(P_m, d_m) = H(P_m) + H(d_m) - H(P_m, d_m) \tag{26}$$

$$H(P_m) = -\sum_i \frac{n_i}{N} \log \frac{n_i}{N} \tag{27}$$

$$H(d_m) = -\sum_j \frac{n_j}{N} \log \frac{n_j}{N} \tag{28}$$

$$H(P_m, d_m) = -\sum_i \sum_j \frac{n_{i,j}}{N} \log \frac{n_{i,j}}{N} \tag{29}$$

2.3. Correlations

It is very important to know the level of dependence. Various correlations and other measurements of dependence are used according to Ref. [7].

In order to estimate correlation bonds, the coefficient of these bonds (unit of strength in statistical connections between variables) gains significance. It may be applied only when there is a linear dependence in range [-1÷1]. The connection is very strong if it equals to -1 and very weak when it equals to 1. It is marked as ρ in theory of probability.

Coefficient of correlation $\rho_{X,Y}$ of two random values X and Y when their averages are μ_X and μ_Y, and stardard deviations σ_X and σ_Y is defined as follows:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E\left((X - \mu_X)(Y - \mu_Y)\right)}{\sigma_X \sigma_Y} \tag{30}$$

Standard deviations σ_X and σ_Y need to be finite and unequal zero for coefficient of correlation to have a defined meaning. In the first part of the equation, we see the covariation of variables X and Y that is the average of multiplication of their deviation from average values.

Correlations may be positive or negative, depending on their direction. For example, when there is negative correlation, values of one variable decrease while the values of other variable increase. Pearson's coefficient of correlation is used to measure strength of quantitative connections between variables. Big values of this coefficient, either positive or negative, reflect strong correlation, while small values reflect weak correlation. If correlation is insignificant, coefficient is close to zero. Pearson's coefficient of correlation is calculated by multiplication of pairs of values from two sets, after subtracting the average. The difference is divided by multiplication of standard deviations.

$$r_{xy} = \frac{1}{n-1} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \tag{31}$$

where $\bar{x}$ and $\bar{y}$ are average values of observations x and y, s_x and s_y - standard deviations of x and y. Coefficient of correlation has these properties: when r = 1, all points (x_i, y_i) are in a line that has positive coefficient of direction. When r = -1, all points (x_i, y_i) are in a line that has negative coefficient of direction. When r = 0, all variables are linearly independent. Table 1 represents the interpretation of strength of correlation.

In order to prevent very strong correlation between unrelated variables, there is a check of significance in coefficient of correlation:

i) zero hypothesis H_0 states that coefficient of correlation equals zero;
ii) alternative, H_1 hypothesis states that coefficient is not equal to zero.

Table 1. Interpretation of strength of correlation.

Negative	Positive	Estimation of correlation
-1.0 ÷ -0.9	0.9 ÷ 1.0	very strong
-0.9 ÷ -0.7	0.7 ÷ 0.9	strong
-0.7 ÷ -0.5	0.5 ÷ 0.7	average
-0.5 ÷ -0.3	0.3 ÷ 0.5	weak
-0.3 ÷ 0.0	0.0 ÷ 0.3	insignificant

In order to prove the H_0 hypothesis, criteria t is introduced:

$$t = r_S \sqrt{\frac{n-2}{1-r_S^2}} \tag{32}$$

where r - calculated value of correlation. n - amount of calculated values. Level of significance is free to choose. Let's assume that the level of significance α = 0.05. Hypothesis H_0 is rejected if absolute value exceeds the critical value of Student's distribution with n - 2 degrees of freedom on 0.5·α level.

3. Data sources

Digital map of Lithuania was chosen as a data source. Layer of forests was chosen as an object of observation. Areas of a map were selected randomly.

Required amount of data. Using another form of Student's criteria and selecting the required accuracy, we may determine the required amount of data:

$$t = \frac{(\bar{x} - \hat{x})}{\sigma_n} \tag{33}$$

where $\bar{x}$ is theorical average of data, $\hat{x}$ - mathematical average of data, σ_n - dispersion of data. Switching it for sample parameter, we get:

$$\sigma_n^2 = \frac{\sigma^2}{n} \approx \frac{1}{n(n-1)} \sum_{i=1}^{n} (\Delta x_i)^2 = s_{\bar{x}}^2 \tag{34}$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \tag{35}$$

If there is a need of accuracy of 0.182, we may calculate required amount of data from 2-43:

$$s_{\bar{x}} = 0.182 s_x, \tag{36}$$

$$0.182 s_x = \frac{s_x}{\sqrt{n}}, \tag{37}$$

$$n = \left(\frac{1}{0.182}\right)^2 \approx 30. \tag{38}$$

In order to get accuracy of 0.182, set of amount of 30 values must be used.

4. Results and discussion

Starting data represent the primary as well as derived parameters. Several primary parameters such as areas and perimeters of forest elements, areas and perimeters of binomes, amount of elements in classes sorted by perimeter were used. Nine indicators of organization were calculated - see Table 2.

Also derived parameters such as fractal dimensions, values of model function, average values of model function, values of model function of average values; fractal dimension of forests; areas of classes sorted by perimeter and fractal dimension, indicators of probability and informational organization, matrices for calculation of negentropy, values of negentropy were included.

Main task could be formulated as follows: to find the strong correlations as relations between indicators of metrical (M) and informational organization (R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8). Such type relations could be most interesting. Values of correlations for analyzed indicators are presented in Table 3.

Looking at Table 3, we may conclude that the connection between organizational indicators is weak or insignificant, because correlation does not exceed 0.3.

It is necessary to point out that r displays direct connection. If correlation r = 0, it does not mean that there is no connection at all. Non-linear connection may appear in this case. We may conclude that because r is close to zero, there is no suitable line by the data.

To check H_0 hypothesis, values of t were checked (see Table 4).

When α = 0.4, all values of t exceeding 0.855 are considered significant correlations and they are worth attention. Values lower than 0.855 are considered insignificant. That means that it is useful to analyse connections between following indicators:

i) M and R_3 (1.080), $R_3 = f(M)$; see Fig. 4;
ii) M and R_4 (0.913), $R_4 = f(M)$; see Fig. 5;
iii) M and R_5 (1.308), $R_5 = f(M)$; see Fig. 6;
iv) M and R_6 (1.308), $R_6 = f(M)$; see Fig. 7.

Choice of variables was done according the following schema: X - indicator of metrical organization M, Y - indicator of informational organization R.

Fig. 4 and Fig. 5 represent the dependences $R_3 = f(M)$ and $R_4 = f(M)$ respectively. Following linear equations were established from mentioned Fig. 4 and Fig. 5 respectively (red line):

$$Y = 0.27678 \cdot X + 0.11983 \tag{39}$$

$$Y = 0.19999 \cdot X + 0.10764 \tag{40}$$

We set Student's coefficient to 5% and see that the reliability of correlation is 95%, but 67% of data is unreliable for R_3 and 27% of data is reliable for R_4. Such data allow us to conclude that connection between metrical and informational organization is insignificant or there are no indicators when

Table 2. Dependencies between indicators of metrical (M) and informational organization (R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8).

Nr	M	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
1.	0.53	0.05	0.45	0.10	0.005	0.08	0.15	1.30	1.18
2.	0.58	0.21	0.06	0.21	0.32	1.43	1.20	1.08	1.25
3.	0.60	0.16	0.01	0.12	0.30	1.79	1.52	1.52	1.90
4.	0.38	0.21	0.35	0.35	0.40	1.27	1.53	1.02	1.10
5.	0.42	0.10	0.003	0.25	0.53	1.71	1.55	1.38	2.19
6.	0.47	0.01	0.30	0.77	0.38	1.11	1.55	3.15	1.16
7.	0.50	0.09	0.14	0.30	0.28	1.38	1.46	1.46	1.43
8.	0.68	0.06	0.02	0.06	0.23	0.98	0.94	1.14	1.40
9.	0.64	0.01	0.12	0.16	0.11	0.94	1.05	1.32	1.23
10.	0.56	0.04	0.01	0.14	0.43	1.07	1.02	1.03	1.55
11.	0.64	0.05	0.13	0.04	0.10	1.35	1.46	1.42	1.50
12.	0.59	0.13	0.42	0.74	0.24	1.28	1.93	3.58	1.22
13.	0.63	0.004	0.07	0.11	0.11	0.85	0.92	1.08	1.08
14.	0.32	0.01	0.34	0.31	0.11	0.46	0.69	0.78	0.61
15.	0.51	0.11	0.36	0.10	0.06	1.18	1.64	1.42	1.36
16.	0.60	0.08	0.35	0.14	0.06	1.23	1.72	1.67	1.52
17.	0.54	0.01	0.34	0.20	0.10	0.82	1.22	1.31	1.16
18.	0.48	0.03	0.17	0.15	0.09	1.29	1.51	1.60	1.49
19.	0.60	0.07	0.27	0.22	0.22	1.15	1.48	1.67	1.66
20.	0.63	0.06	0.30	0.18	0.13	1.16	1.55	1.23	1.17
21.	0.66	0.05	0.28	0.32	0.25	0.98	1.30	1.31	1.19
22.	0.51	0.04	0.47	0.17	0.01	0.83	1.52	1.34	1.12
23.	-0.09	0.02	0.36	0.07	0.09	0.43	0.66	1.23	1.27
24.	0.63	0.01	0.53	0.41	0.14	0.45	0.95	1.38	0.95
25.	0.65	0.05	0.47	0.29	0.08	0.05	0.29	1.33	1.03
26.	0.63	0.004	0.17	0.57	0.69	0.21	0.25	0.76	1.06
27.	0.71	0.02	0.12	0.58	0.36	1.70	1.88	1.66	1.09
28.	0.61	0.40	0.64	0.06	0.02	1.51	2.48	1.92	1.85
29.	0.54	0.12	0.53	0.33	0.08	1.24	2.32	2.00	1.44
30.	0.78	0.10	0.42	0.72	0.61	1.55	2.41	2.21	1.59

probabilities are calculated by using areas of forest groups that are classified by fractal dimension of forests (R_3) or by perimeter of forests (R_4).

Fig. 6 and Fig. 7 represent dependences $R_5 = f(M)$ and $R_6 = f(M)$ respectively. Following linear equations were established from Fig. 6 and Fig. 7 (red line):

$$Y = 0.66599 \cdot X + 0.66504 \tag{41}$$
$$Y = 0.89736 \cdot X + 0.84422 \tag{42}$$

We set Student's coefficient to 5% and see that the reliability of correlation is 95%, but 37% of data is reliable for R_5 and 36% - for R_6. Such data allow us to conclude that connection between metrical and informational organization is insignificant but direct when values of informational organization are calculated by using expression of negentropy, where probabilities are calculated by using amount of elements and divided by entropy, which is calculated:

i) by using element number in class, classified by fractal dimension of forests (R_5) or
ii) by numbers of elements in a class, classified by fractal dimension of forests (R_6).

Although all indicators suggest that observed structures are more or less organized, no connection between indicators was noticed after reviewing all charts, because neither way of examination showed strong direct dependence. Analytically, both methods are similar - they both depend on connection between the variables. Unfortunately analysis of correlation did not confirm this assumption. Why there is no connection between them? Organization of spatial structures is a term that has multiple meanings. There is a possibility that this property may not be represented by any single indicator and needs multivariable characteristics. If, for example, one indicator shows homogeneity, another may show anisotropy. In

Table 3. Correlations for indicators.

Indexes	Correlation coefficients	Pearson T-criteria
M and R_1	0.07	0.371
M and R_2	-0.10	0.532
M and R_3	0.20	1.080
M and R_4	0.17	0.913
M and R_5	0.24	1.308
M and R_6	0.24	1.308
M and R_7	0.13	0.694
M and R_8	0.09	0.478

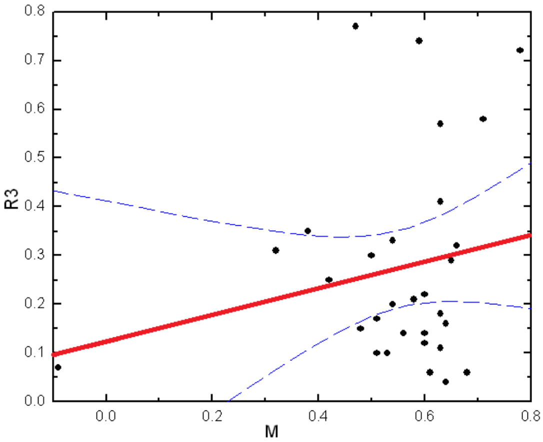

Fig. 4. $R_3 = f(M)$. Dependency of metric organization indicator from information organization indicator when probabilities are calculated based on the areas of forest groups that are classified by fractal dimension of forests.

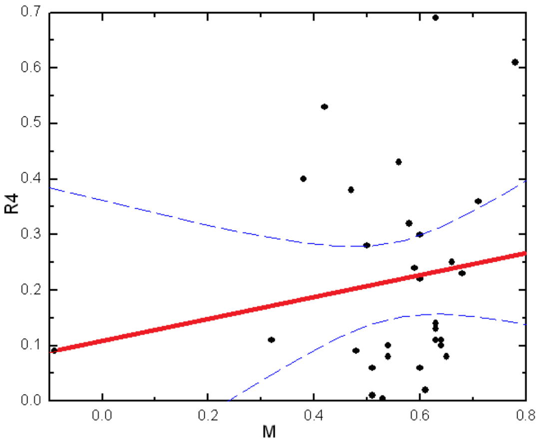

Fig. 5. $R_4 = f(M)$. Dependency of metric organization indicator from information organization indicator when probabilities are calculated based on the areas of forest groups that are classified by perimeters of forests.

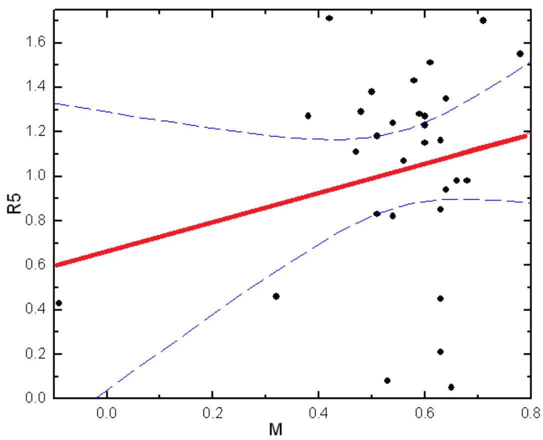

Fig. 6. $R_5 = f(M)$. Dependency of metric organization indicator from information organization indicator when values of information organization are calculated by using negentropy where probabilities are calculated by the amount of elements and divided by entropy which is calculated by the amount of elements in a class, classified by the fractal dimension of forests.

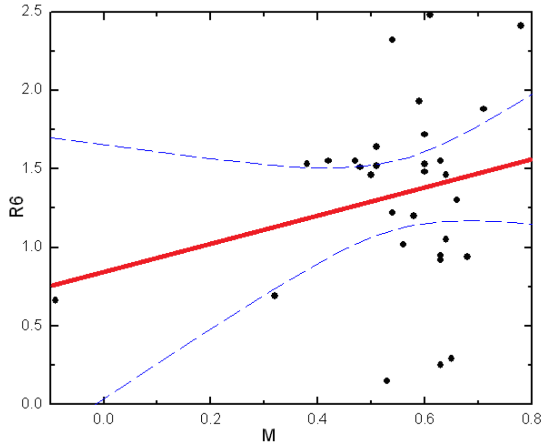

Fig. 7. $R_6 = f(M)$. Dependency of metric organization indicator from information organization indicator when values of information organization are calculated by using negentropy where probabilities are calculated by the amount of elements and divided by entropy which is calculated by the amount of elements in a class, classified by the perimeter of forests.

this case, there may be no correlation between them and only future research will perhaps will solve this problem.

Conclusions

1. Research of cartographic layers in two different ways (19 000 km^2 territory was observed in total) shows that there are no statistically significant correlations between informational and metrical organization indicators. Two conclusions come of this:
   i) these indicators show different aspects of spatial organization;
   ii) research done (variety of layers and numbers) is insufficient to measure the correlation.
2. Variation of informational indicator values is bigger than the characteristics of metrical indicator ($C_u > C_v$). This difference is the significance of a structure.
3. Observation of real layers is not sufficient to determine the sensitivity of indicators. Research should be done by changing properties of layers by applying Moran and Getis indicators in order to compare results. That would help to determine the susceptibility of objects for spatial heterogeneity, anisotropy and of metrical and topological properties of layers.
4. In order to improve the research it is preferable to apply both: GIS and statistical computation routines for estimation of spatial organization in cartography.

Acknowledgments

This work was prepared as a part of the Master Thesis. Author would like to thank Assoc. Prof. Algimantas Tiknius (Klaipėda university) for suggestions and comments.

References

1. Jelena Nesterova. Spatial self-arrangement of expanding structures. 1. Overview of assessment concepts – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(9)** 2010 – Pp. 17-22.
2. Jelena Nesterova. Spatial self-arrangement of expanding structures. 2. Structure model construction – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(9)** 2010 – Pp. 23-29.
3. Algimantas Tiknius, Kęstutis Dučinskas, Benediktas Jankauskas. Deriving the spatial traits of organized land structures. – In: I. Petrosillo et al. Use of landscape sciences for the assessment of environment security. – Springer, 2008. – 121-130. – <http://www.springerlink.com/content/q6u8m2m545582162/>, accessed 2010 08 05.
4. G. Boffetta, M. Cencini, M. Falcioni, A. Vulpiani. Predictability: a way to characterize complexity. – <http://arxiv.org/abs/nlin.CD/0101029>, accessed 2010 04 09.
5. J. Bogaert, A. Farina, R. Ceulemans. Entropy increase of fragmented habitats: A sign of human impact? – <http://www.springerlink.com/content/hw2784516085402x/fulltext.pdf>, accessed 2010 04 09.
6. Ambedkar Dukkipati. On Generalized Measures of Information with Maximum and Minimum Entropy Prescriptions. – <http://etd.ncsi.iisc.ernet.in/bitstream/2005/353/1/G20346.pdf>, accessed 2010 04 09.
7. Bronislovas Martinėnas. Eksperimento duomenų statistinė analizė (in lith.) – Vilnius: Technika, 2004.