Comparison of Methods for Joining Pointwise Geological Data to Interpolation Grids

Aivars Spalvins, *Riga Technical University*, Inta Lace, *Riga Technical University*

Abstract **– Pointwise geological initial data σ_{in} are used for creating hydrogeological models (HM). By processing σ_{in} by interpolation methods, σ - maps are created on (x, y) – grids of HM. The maps represent geometrical and physical features of geological layers. If solutions of boundary field problems are applied as interpolation results (σ - maps) then σ_{in} serve as the boundary conditions of the first kind. They must be joined to nodes of the interpolation grid. In this paper, three methods are compared that may be applied to perform this task when the local date search region is a square, a circle, an area enclosed by hyperbola arcs. Features of these methods are examined and recommendations on their optimal use are given.**

Keywords **- hydrogeological models, interpolation of data**

I. INTRODUCTION

To simplify explanations, the normalized uniform (x, y) - grid is considered where the plane approximation steps h_x=h_x=1.0. For the node p_0, the local normalized search area is the 2×2 square enclosed by the eight neighbouring nodes $p_1, p_2, \dots, p_8$ (see Fig. 1 a)). The node p_0 represents the origin of the local $\bar{x} = x_0 - x$, $\bar{y} = y_0 - y$ local coordinate system where x_0 , y_0 and *x*, *y* are coordinates of the node p_0 and the datum point σ_{in} , accordingly.

Three forms of the data search regions Ls are considered (see Fig. 1 a), 1 b), 1c)): a square, a circle, an area which perimeter is formed by four arcs of hyperbola (the Lh form). By using the parameter $c = 1 - |\bar{x}|_{\max}$ or $c = 1 - |\bar{y}|_{\max}$ (1≥*c*≥0) the regions are defined, as follows:

$$|\bar{x}| \le 1 - c, \quad |\bar{y}| \le 1 - c \;, \tag{1}$$

$$\sqrt{\bar{x}^2 + \bar{y}^2} \le 1 - c \,, \tag{2}$$

$$(1 - |\bar{x}|)(1 - |\bar{y}|) \ge c \;, \tag{3}$$

where the formulas (1), (2), (3) correspond to the square, the circle, the Lh form. The value 2(1-*c*) represents the maximal *x* or *y* dimension of Ls.

The data located within Ls are received by the node p_0 . The area of Ls represents the R_0-region with respect to the node p_0. However, the eight neighbouring nodes $p_1, p_2, \dots p_8$ also have their R_p-regions. Interference of these nine R-regions creates the data “give out” G-regions L*i* (*i*=1, 2, 3, 4). The figure *i* shows the number of nodes that receive data located within L*i*. If Ls is the circle (Fig.1 b)), it contains 21 G-regions: twelve L3 regions, four L2 and L4 regions and one L1 region. Each of them has a different pattern of their data spread out.

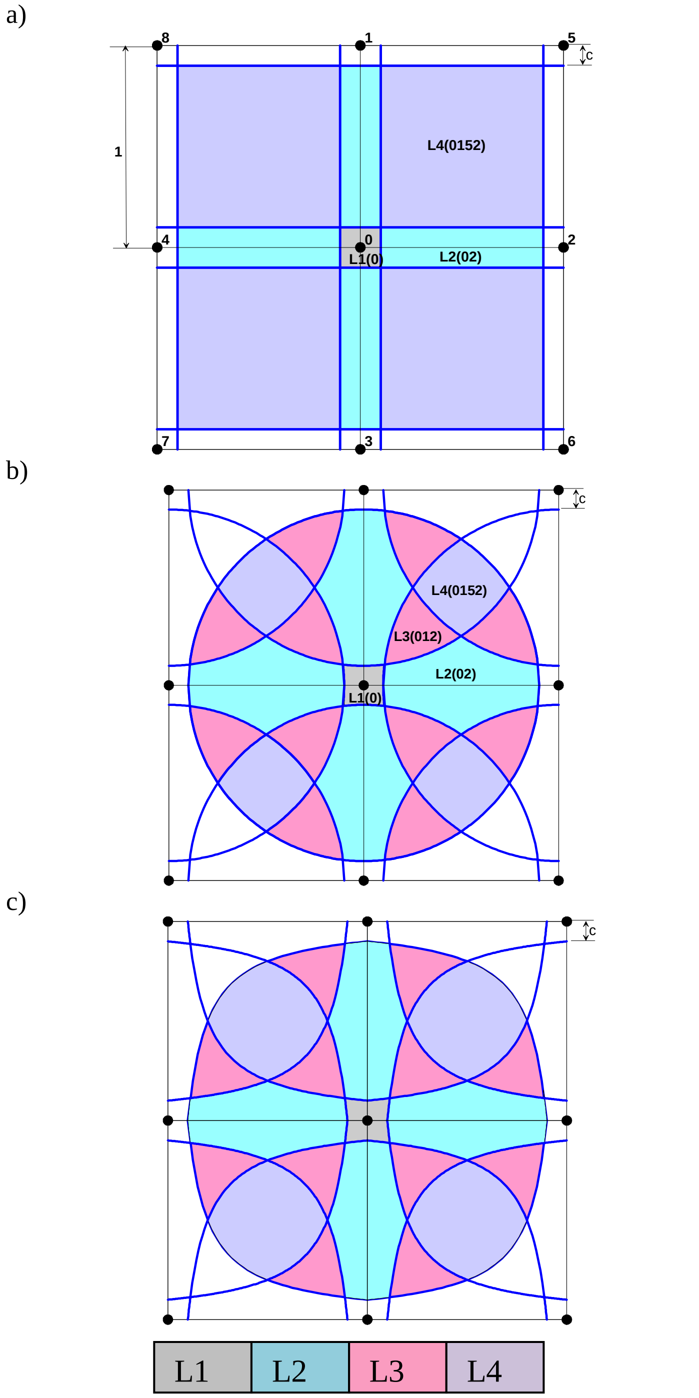

Fig. 1. Data search R-regions within the 2×2 area with the data, “give out” G-regions (L1, L2, L3, L4) included: a) the square; b) the circle; c) the area enclosed by arcs of hyperbola (the Lh form); *c*=0.1

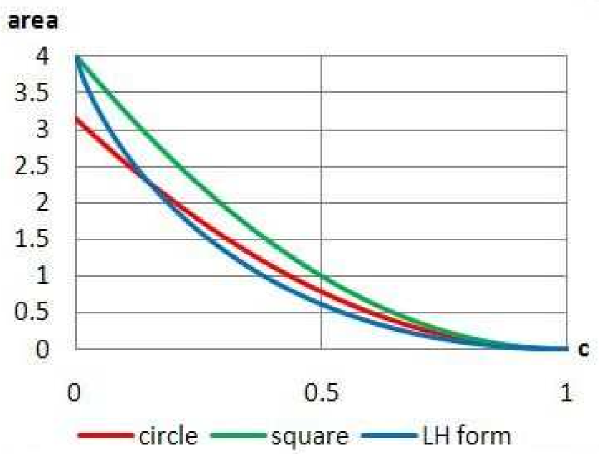

Fig. 2. Area of the R-regions versus the parameter c for three data search methods

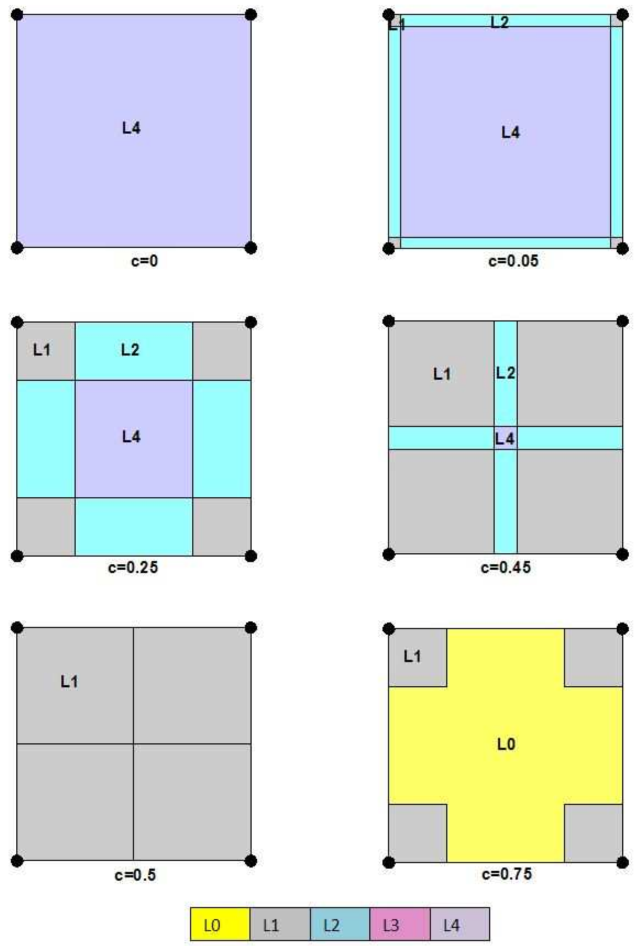

Fig. 3. The R-region is the square. Distributions of the G-regions within the 1×1 area if the parameter *c* has different values

For example, from the regions L4 (0, 1, 5, 2); L3(0,1,2); L2(0,2); L1(0) data are given out to the *p*-nodes 0, 1, 3, 2; 0, 1, 2; 0, 2; 0, correspondingly. Distributions and size of the G-regions depend on the method used for creating the R–regions (the square form contains no L3 regions) and on the value of *c*. Objectives of our investigation are to pick out the best method for joining data to the grid and to find the optimal value of the parameter *c*. This investigation broadens the scope of results reported in [1].

II. FEATURES OF THE R AND G-REGIONS

It follows from the distributions of G-regions (Fig.1) that "productiveness" of their data (value i=1, 2, 3, 4) changes abruptly when the borderlines separating the regions are crossed. The change Δ is minimal (Δ=1.0) if the R-regions are the circle and the Lh form. For these cases, Δ=2 only in 12 points where the regions L1 and L3; L2 and L4 are in touch.

For the square (Fig. 1 a)), Δ=2 with respect to the L2 and L4 regions. In four points, Δ=3 (L1 andL4 regions are in touch there). If *c*=0 then only the L4 region is present, for the square and the Lh form and crossing of the 2×2 region borderline results in unacceptable data jumps to the neighbouring search regions. For the circle (*c*=0), such jumps may take place in four points where circles touch the 2×2 region. To avoid this fault, the value *c*>0 must be used for all three forms of Ls.

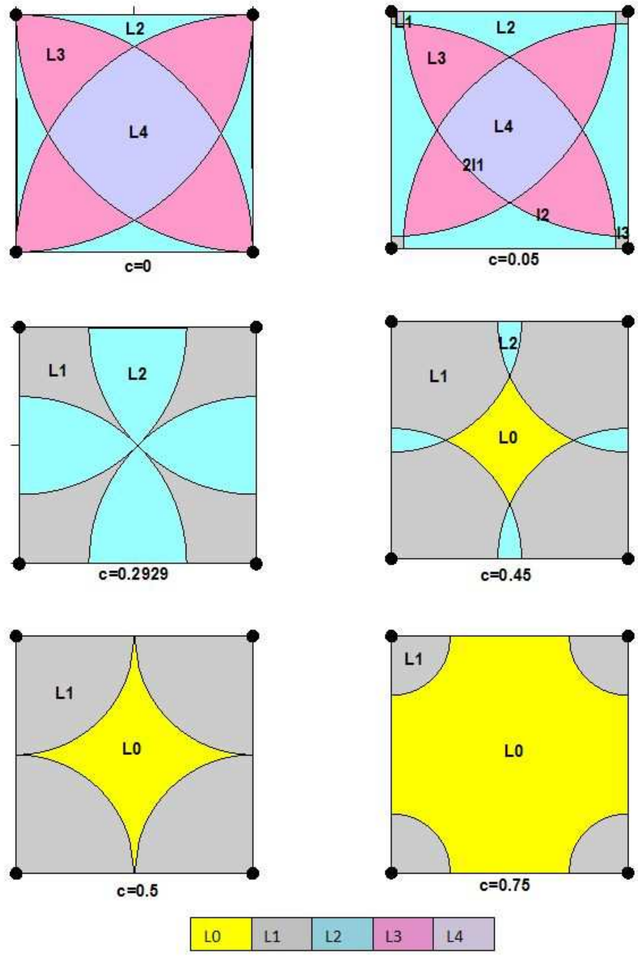

Fig.4. The R-region is the circle. Distributions of the G-regions within the 1×1 area if the parameter *c* has different values

In the sense of the minimal Δ values, the circle and the Lh form are equal and better than the square form.

The area Ls and its perimeter Ps depend on the value *c*. The following formulas are used to compute these parameters:

$$Ls = 4(1-c)^2, \quad Ps = 8(1-c), \qquad (4)$$

$$Ls = \pi(1-c)^2, \quad Ps = 2\pi(1-c), \qquad (5)$$

$$Ls = 4(1-c(1-\ln c)), \quad Ps = 8(1-0.8462\sqrt{c} - A), \qquad (6)$$

$$A = 0.1666c^2(1-0.0768c^2)$$

where the formulas (4), (5), (6) are used for the square, the circle and the Lh form accordingly. In Fig. 2, graphs of Ls versus *c* are shown.

From the 2×2 area, the perimeter Ps cuts of the part $\overline{L}s = 4 - Ls$. Data located within $\overline{L}s$ are not sent to the node *p0*. With respect to the R0-region, the perimeter Ps is the border that separates two different areas (data sent / not sent). If Ps is the circumference, it presents the shortest borderline for the given Ls area.

The following criterions are used to estimate closeness of the form of any region to the circle:

$$q = 4\pi Ls / Ps^2 \le 1.0 \quad \text{or} \quad \overline{q} = \sqrt{q} \le 1 \,. \qquad (7)$$

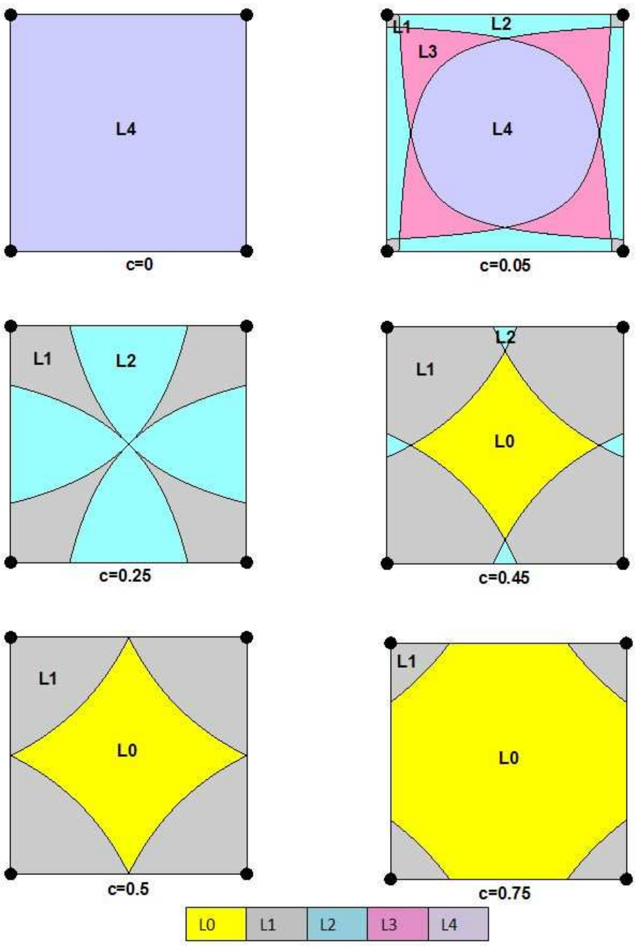

Fig.5. The R-region is the Lh form. Distributions of the G-regions within the 1×1 area if the parameter *c* has different values

The version of $\overline{q} = \sqrt{q}$ is used when some combinations of various Ls must be estimated (see later). For the circle and the square q=1.0 and π/4=0.785, accordingly. For the Lh form, the value of *q* changes if 1≥c≥0. If *c*=0 and 1.0 then *q*=0.785; *q*≅ 1.0 if *c*=0.1716. Therefore, the Lh form should have some features that belong to the square and the circle both.

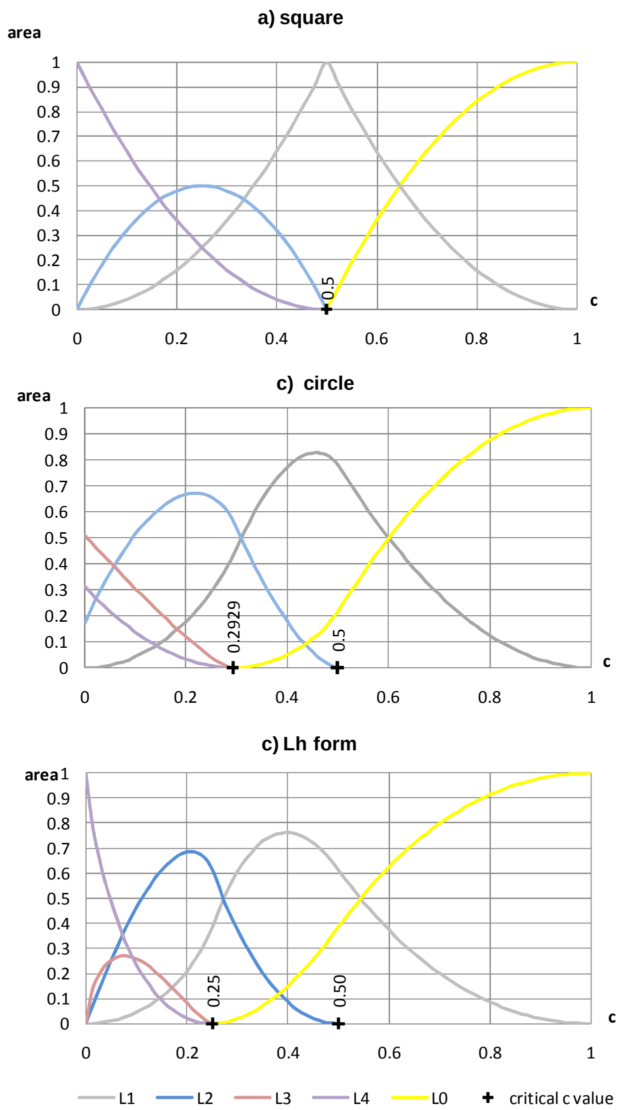

Fig.6. Areas of the G-regions L0, L1, …, L4 versus the parameter *c* for three data search methods

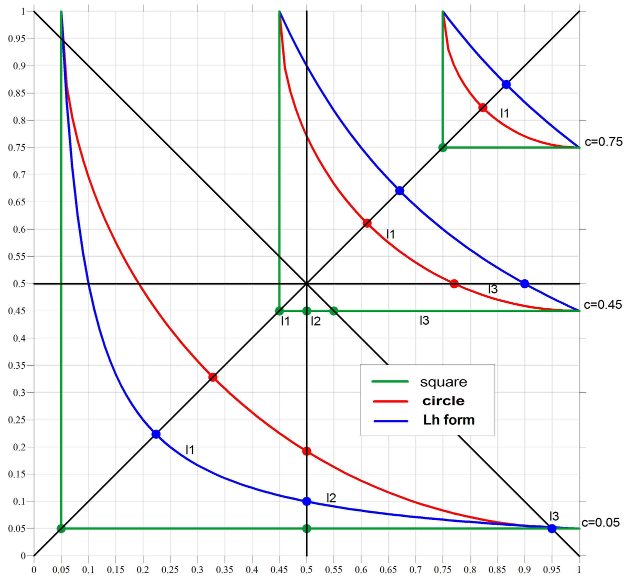

Fig. 7. Appearance of the characteristic line segments l1, l2, l3 for the cases of square, circle, Lh form if *c*=0.05; 0.45; 0.75

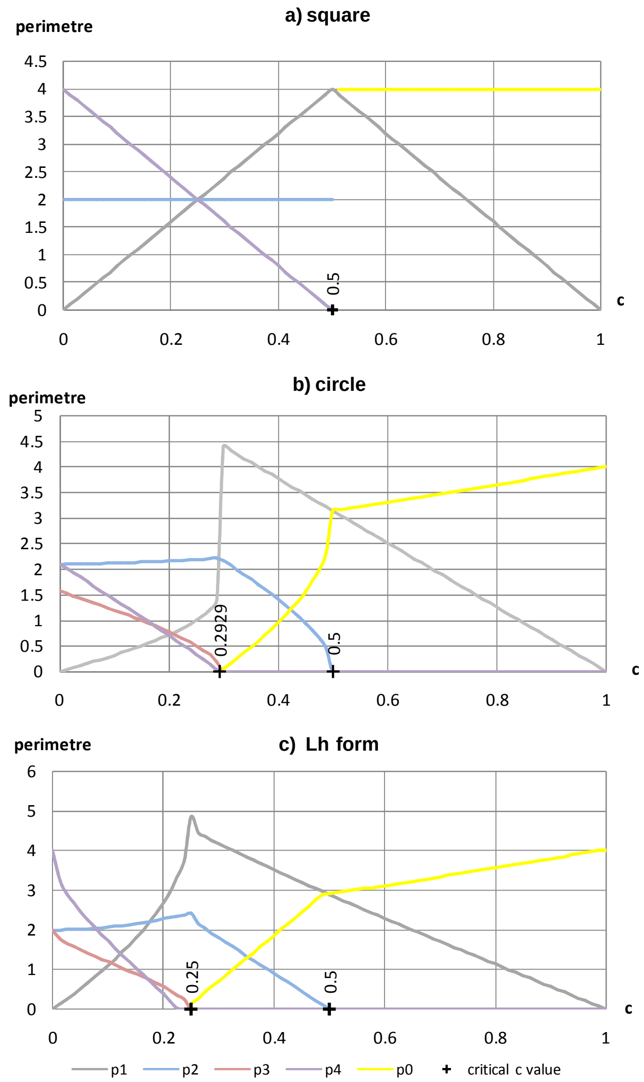

Fig.8. Perimetres of the G-regions L0, L1, …, L4 versus the parameter *c* for three data search methods

III. G-REGIONS WITHIN THE 1×1 AREA

It is simpler to consider only the 1×1 area that is the characteristic part of any R-region. Within this area

$$\sum_{i=0}^{4} Li = 1.0 \ , \qquad (8)$$

where the figure i=0 presents the area L0 unreachable for data searching.

In Fig. 3, the case of the square is presented if the parameter c has six different values 0.75≥c≥0. If c=0 then only the L4 area exists. If c=0.5 (critical c value) then only four L1 areas are present. If c≥0.5 then the L0 area appears.

In Fig. 4, the circle case is examined. There are two critical values of c: 0.2929; 0.5 when the L3, L4 and L2 areas disappear, correspondingly. The L0 area starts if c≥0.2929.

In Fig. 5, the Lh form case is considered. There are two critical values of *c*: 0.25; 0.5 when the L3, L4 and L2 areas disappear, correspondingly. The L0 area starts if *c*≥0.25.

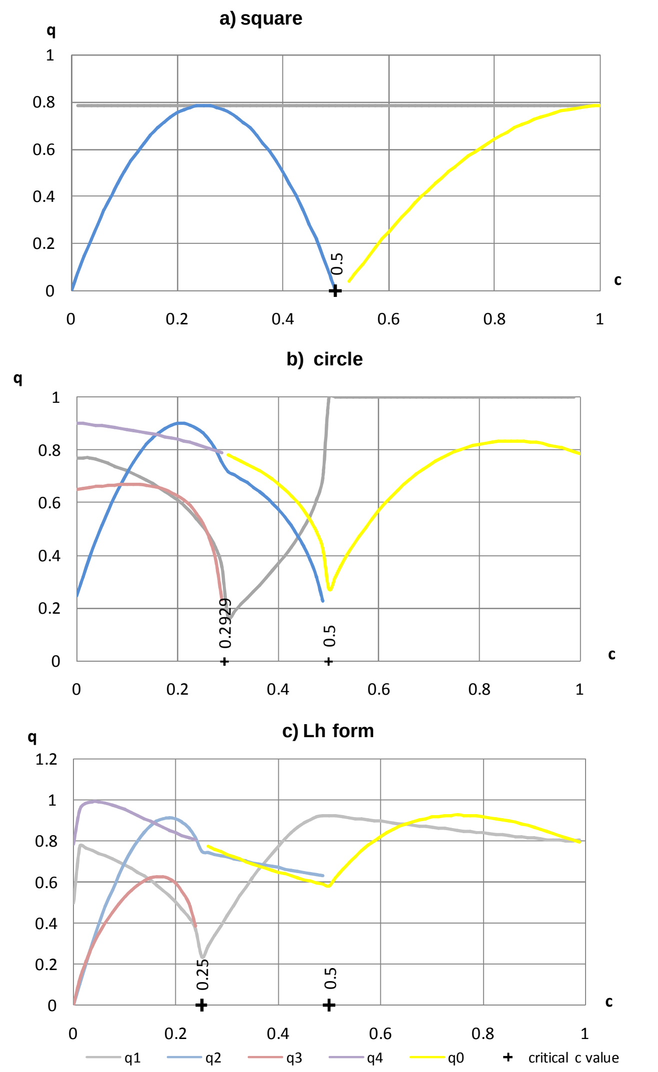

Fig. 9. The criterion q versus the parameter *c* for three data search methods

It follows from the above graphical illustrations that the distributions of Li regions bear drasticall alterations when the parameter c changes (1.0≥*c*≥0). By using formulas of Table 1, areas Li of the G-regions were obtained if they were located within the 1×1 area. By accounting for the Li distributions of Fig.1, this area contains one L1 and L0 region, two L2 regions and four L3 regions (if they exist). This result can also be obtained if the 1×1 area is projected on a toroidal surface (opposite edges of the 1×1 square are connected).

In Fig. 6, graphs of Li versus *c* (1.0≥*c*≥0) are presented for the square, the circle and the Lh form. The relationships L0(*c*), L1(*c*), L2(*c*) are rather similar for the all above cases. The graphs L3(*c*) behave quite differently: for the square L3 does not exist; for the circle, L3(0)=0.5; L3(0.29290)=0; the graph scales down between these *c* values; for the Lh form, L3(0)=L3(0.25)=0; the graph has a maximum located between these two c values. The L4(c) graph shows how large is the area which is the most productive data sender in the four directions uniformly. The area is maximal if c=0. The Lh form has rather large L4 (L4≥0.4) if c is small (c≤0.05). Provisory, the Lh form seems advantageous, if small c values are used (rather large L4(c), L3(c) close to its maximum, L2(c) ≥0.3).

TABLE 1
AREAS OF Li REGIONS

Parameter c	Areas of regions
square	
0≤c≤0.5	$L1 = 4c^2$; $L2 = 4c(1-2c)$; $L4 = (1-2c)^2$
0.5≤c≤1.0	$L1 = (1-c)^2$; $L0 = 1 - L1$
circle	Auxilary terms $r = 1 - c$; $a = 0.5(1 - \sqrt{2r^2 - 1})$; $A_{0.5} = \sqrt{4r^2 - 1} + 4r \ \arcsin(0.5/r)$;
0≤ c≤0.2929	$L1 = 4(a - r^2 \arcsin(a/r)$; $L2 = 4 - 2L1 - A_{0.5}$; $L4 = 1 + \pi r^2 - A_{0.5}$; $L3 = 1 - L1 - L2 - L4$;
0.2929≤c≤0.5	$L2 = 2\pi r^2 - A_{0.5}$; $L1 = 2\pi r^2 - 2L2$; $L1 = \pi r^2 - 2L2$; $L3 = 0$
0.5≤ c≤1.0	$L1 = \pi r^2$; $L2 = L3 = 0$; $L0 = 1 - L1$;
Lh form	Auxilary terms $a = 0.5(1 - \sqrt{1-4c})$
0≤c≤0.25	$L1 = 4(-a + c(+2\ln((1/(1-a))$; $L2 = 8c\ln 2 - 2L1$; $L3 = 1 - (L1 + L2 + L4)$; $L4 = 1 - 4c(1 - \ln 4c)$; $L2 = 4(1 - 2c(1 - \ln 2c)$
025≤ c≤0.5	$L0 = 1 - 4c(1 - \ln 4c)$; $L1 = 1 - L2 - L0$; $L3 = 0$
0≤c≤0.5	$L1 = 4(1 - c(1 - \ln c)$; $L0 = 1 - L1$

To formulate well founded recommendations, the criterion q_i should obtained, for the Li regions. To compute q_i, the perimeters Pi are needed. To obtain them, lengths of characteristic line segments l1, l2, l3 must be known. Their appearance is shown in Fig.7. To obtain the perimeter P0 of L0, an extra segment l0 is needed. In Table 2, formulas are given which are applied to create the graphs Pi of Fig. 8. In Table 3, the perimeters Pi are given as various sums of the line segments l0, l1, l2, l3.

In Fig.9, the graphs q_i versus c are given for the three search methods. The formulas used for computing these q_i are present in Table 4. The criterion q_i have to be applied to the single Li region. For this reason, in formulas (of Table 3), the values of the areas L3 and L2 are divided by 4 and 2, correspondingly. The perimeters Pi of Fig. 8 are also given for the single G- regions.

In Fig. 9 for the Lh form, q_4~1.0 if c=0.043. It means that the L4 area is almost the circle when the q_4 graph reaches its maximal value. The other graphs of Fig. 9 provide no valuable information that can be used for practical purposes. The value c=0.043 has been implemented in real interpolation algorithms [1].

TABLE 2
LENGTHS OF CHARACTERISTIC LINE SEGMENTS

Parameter c	Length of lines
square	
0≤c≤0.5	$l1 + l2 + l3 = 1 - c$; $l1 = l2 = 0.5 - c$; $l3 = c$
0.5≤c≤1.0	$l1 = 1 - c$; $l0 = c - 0.5$; $l3 = c$
circle	Auxilary terms $r = 1 - c$; $a = 0.5(1 - \sqrt{2r^2 - 1})$;
0≤ c≤0.2929	$l1 + l2 + l3 = \pi r/4$; $l1 = r(\pi/4 - \arcsin(0.5/r)$; $l2 = r(\arcsin(0.5/r) - \arcsin(a/r))$; $l3 = r\arcsin(a/r)$
0.2929≤c≤0.5	$l1 + l3 = \pi r/4$; $l1 = r(\pi/4 - \arcsin(\sqrt{4r^2-1}/2r)$; $l1 = r\arcsin(\sqrt{4r^2-1}/2r)$; $l2 = 0$; $l0 = l1$
0.5≤ c≤1.0	$l1 = \pi r/4$; $l2 = 0.5 - r$; $l2 = l3 = 0$
Lh form	Auxilary terms $1 - a = 0.5(1 + \sqrt{1-4c})$; $I_{\sqrt{c}} = 0.8462\sqrt{c}$; $I_{0.5} = 0.5(1 - (16c^2/6)(1 - -0.0768 \times 16c^2))$; $I_{1-a} = (1-a)(1 - (1/6)(\sqrt{c}/(1- -a))^4(1 - 0.0768(\sqrt{c}/(1-a))^4)$; $I_1 = 1 - (c^2/6)(1 - 0.0768c^2)$; $I_{2c} = c(2 - (1/48c^2)(1- -0.0768/(16c^2))$
0≤c≤0.25	$l1 + l2 + l3 = I_1 - I_{\sqrt{c}}$; $l3 = I_1 - I_{1-a}$; $l2 = I_{1-a} - I_{0.5}$; $l1 = I_{0.5} - I_{\sqrt{c}}$
025≤ c≤0.5	$l1 + l3 = I_1 - I_{\sqrt{c}}$; $l1 = I_{2c} - I_{\sqrt{c}}$; $l3 = I_1 - I_{2c}$; $l2 = 0$; $l0 = l1$
0≤c≤0.5	$l1 = I_1 - I_{\sqrt{c}}$; $l0 = c - 0.5$; $l2 = l3 = 0$

More complex criterion Q can be obtained if three regions L1, L3+L4, L4 are involved. They are enclosed by the perimeters 8l3, 8l2, 8l1, accordingly. If these perimetres are used as weight coefficients and the criterion $\overline{q}$ is applied, Q is given the following formula (c≤0.25):

$$Q=(l1\bar{q}_4+l2\bar{q}_{3,4}+l3\bar{q}_1)/(l1+l2+l3)=$$
$$=\frac{\sqrt{\pi}(\sqrt{L1}+\sqrt{L3+L4}+\sqrt{L4}}{4(l1+l2+l3)} \qquad (9)$$

$$L3=0\,,\quad l1+l2+l3=1-c \quad \text{(for square)}$$

$$l1+l2+l3=\pi(1-c)/4 \quad \text{(for circle)}$$

$$l1+l2+l3=I_1-I_{\sqrt{c}} \quad \text{(for Lh form).}$$

TABLE 3
EXPRESSIONS FOR OBTAINING THE CRITERION *q*

Parameter *c*	Perimertres	Criterion *q*
square		
0≤*c*≤0.5	$p1=8l3$, $p2=4(l2+l3)$, $p3=2(l1+l2)$, $p4=8l1$	$q_1=\frac{\pi}{16}L1/l3^2$; $q_2=\frac{\pi}{8}L2/(l2+l3)^2$; $q_3=\frac{\pi}{4}L3/(l1+l2)^2$; $q_4=\frac{\pi}{16}L4/l1^2$
0.5≤*c*≤1.0	$p1=8l1$, $p0=8(l1+l0)$	$q_1=\frac{\pi}{16}L1/l1^2$; $q_0=\frac{\pi}{16}L0/(l1+l0)^2$
circle		
0≤ *c*≤0.2929	$p1=8l3$, $p2=4(l2+l3)$, $p3=2(l1+l2)$, $p4=8l1$	$q_1=\frac{\pi}{16}L1/l3^2$; $q_2=\frac{\pi}{8}L2/(l2+l3)^2$; $q_3=\frac{\pi}{4}L3/(l1+l2)^2$; $q_4=\frac{\pi}{16}L4/l1^2$;
0.2929≤*c*≤0.5	$p1=8(l1+l3)$, $p2=4l3$, $p0=8l1$, $p3=p4=0$	$q_1=\frac{\pi}{16}L1/(l1+l3)^2$; $q_2=\frac{\pi}{8}L2/l3^2$; $q_0=\frac{\pi}{16}L0/l1^2$;
0.5≤ *c*≤1.0	$p1=8l1$, $p0=8(l1+l0)$	$q_1=\frac{\pi}{16}L1/l1^2$; $q_0=\frac{\pi}{16}L0/(l1+l0)^2$;
Lh form		
0≤*c*≤0.25	$p1=8l3$, $p2=4(l2+l3)$, $p3=2(l1+l2)$, $p4=8l1$	$q_1=\frac{\pi}{16}L1/l3^2$; $q_2=\frac{\pi}{8}L2/(l2+l3)^2$; $q_3=\frac{\pi}{4}L3/(l1+l2)^2$; $q_4=\frac{\pi}{16}L4/l1^2$;
025≤ *c*≤0.5	$p1=8(l1+l3)$, $p2=4l3$, $p0=8l1$, $p3=p4=0$	$q_1=\frac{\pi}{16}L1/(l1+l3)^2$; $q_2=\frac{\pi}{8}L2/l3^2$; $q_0=\frac{\pi}{16}L0/l1^2$;
0≤*c*≤0.5	$p1=8l1$, $p0=8(l1+l0)$	$q_1=\frac{\pi}{16}L1/l1^2$; $q_0=\frac{\pi}{16}L0/(l1+l0)^2$;

For the three search methods, the relationship *Q* versus *c* is given in Fig.10. Only for the Lh form there exist a flat maximum of *Q*, within the interval 0.05≥*c*≥0. The value *c*=0.02 gives maximum of *Q(c)*.

Comparison of the *Q(c)* graphs of the search methods leads to conclusion that the Lh form provides higher *Q(c)* values within the interval 0.25≥*c*≥0. For the cases of the square and the circle, maximal value of *Q(c)* corresponds to *c*=0. Considering of the relationship *Q(c)* leads to conclusion, that small values of *c* should be applied for all search methods.

In Table 4, comparison of the data search methods is presented. It follows from these results that the search method using the Lh form is preferable (large relative size of L4, L4 has the circle form, smooth data search).

TABLE 4
COMPARISON OF THE DATA SEARCH METHODS

Parameter	Square	Circle	Lh form
Smoothness of data search (Δi value)	crude Δi =2	good Δi =1	good Δi =1
Relative size of L3, L4	L3=0 L4=1	L3=0.5 L4=0.3	L3=0.25 L4=0.5
Form of L3, L4 regions	no L3 L4 square	L3 triangular L4 circular	L3 triangular L4 circle
Recommended value of *c*	*c*≤0.01	*c*≤0.01	*c*=0.043

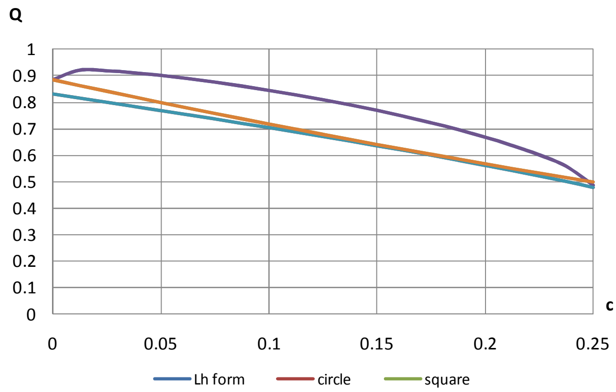

Fig. 10. The criterion *Q* versus the parameter *c* for three data search methods

IV. Conclusions

Thorough examination of main features of the three local data search methods (the square, the circle, the Lh form) has been carried out. Smoothness of data spread out, the size of the R and G – regions, the form (criterion q) of G – regions have been considered. It was found out that the Lh form should be used, because it combines advantages that are possessed by the cases of the square and the circle methods.

References

[1] Spalvins, A., Slangens, J., Local interpolation of geological environment data. *In: proceedings of International Seminar on Environment Modelling*, Jurmala, Latvia, October 9-12, 1995. Riga, 1995, p. 159-174.

Aivars Spalvins was born 1940, Latvia. In 1963, he graduated Riga Polytechnical institute (since 1990, the Riga Technical university) as computer science engineer. In 1967, A. Spalvins received degree of science candidate confirmed by thesis entitled *"Hybrid computers for solving boundary field problems "*. In 1994, this degree was transferred to the one of the doctor of engineering sciences.
Aivars Spalvins has been with the university science 1958 (as a student) until now. His present scientific interests are computer modeling of groundwater flows and migration of contaminants. He is author of about 300 scientific papers. His present position is Director of Environment Modelling centre of Riga Technical University. He is member of International Association of Hydrogeologists, since 1995.
Address: 1/4 Meza str., Riga, LV-1007, Latvia
Phone: +371 67089511
E-mail: emc@cs.rtu.lv

Inta Lace was born in Latvia. In 1971, she graduated Riga Polytechnic institute (since 1990, the Riga Technical university) as computer science engineer. In 1995, I. Lace received degree of M.sc. (applied computer science). Inta Lace has been with the university science 1965 (as a student) until now. Since 1971, she is researcher of Environment Modelling Centre, Faculty of Computer Science and Information Technology, Riga Technical University. Since 1991, she took part in projects of Latvian Science Council on informatics for hydrogeology and other projects. Her present scientific interests are computer modeling of groundwater flows and migration of contaminants. She is author and co-author of about 100 scientific papers on software problems of solving boundary field problems.
Address: 1/4 Meza str., Riga, LV-1007, Latvia
Phone: +371 67089511
E-mail: emc@cs.rtu.lv

Aivars Spalviņš, Inta Lāce. Metožu salīdzināšana ģeoloģisko datu piesaistei interpolācijas režģiem
Punktveida ģeoloģiskos datus σ_{in} izmanto hidroģeoloģisko modeļu (HM) izveidošanai. Apstrādājot σ_{in} ar interpolācijas metodēm, iegūst σ−kartes (x,y) režģim HM. Šīs kartes apraksta ģeoloģisko slāņu fizikālās un ģeometriskās īpašības. Ja kā interpolācijas rezultātu (σ−kartes) izmanto lauku teorijas robežproblēmu atrisinājumu, tad σ_{in} tiek izmantoti kā pirmā veida robežnoteikumi. Tos ir jāpiesaista interpolācijas režģa mezgliem.
Šajā publikācijā tiek salīdzinātas trīs metodes, kuras var izmantot šī uzdevuma veikšanai: ja lokālais datu meklēšanas apgabals ir kvadrāts, aplis un apgabals, kuru ierobežo hiperbolas loki. Šo metožu īpašības tiek pētītas un dotas rekomendācijas metožu optimālai izmantošanai. Rakstā doti datu izmantošanas laukumu grafiskie attēli raksturīgajām meklēšanas parametra *c* vērtībām. Tabulu veidā sakārtotas formulas, kuras lieto datu izmantošanas laukumu perimetru, platības un formas kritēriju aprēķiniem. Šo aprēķinu rezultāti doti grafiku veidā. Konstatēts, ka variantam, kurā datu meklēšanas apgabalu ierobežo hiperbolas loki ir priekšrocības attiecībā pret apļa un kvadrāta variantu. Visām trim metodēm ir jāizmanto mazas meklēšanas parametra c vērtības ($0.04 \geq c > 0.01$).

Айвар Спалвиньш, Инта Лаце. Сравнение методов применяемых для привязки точечных данных к интерполяционным сеткам
Точечные геологические данные σ_{in} применяются для построения гидрогеологических моделей (ГМ). Путем обработки интерполяционными методами, получают σ−карты для (x,y) сеток ГМ. Эти карты характеризуют физические и геометрические свойства геологических горизонтов. Если результат интерполяции (σ−карта) является решением краевой задачи теории поля, тогда используются в качестве граничных условий первого рода, которые необходимо привязать узлам интерполяционной сетки. В этой публикации предлагается сравнение трех методов, которые можно применять для решения этой задачи, если локальная область поиска имеет форму квадрата, окружности и области ограниченной дугами гиперболы. Исследуются свойства этих методов и даются рекомендации по их оптимальному использованию.
В статье приведены графические образы областей использования данных для характерных значений параметра поиска *c*. В виде таблиц приведены формулы, которые используются для расчета периметра, площади и критерия качества формы этих областей. Результаты этих расчетов оформлены в виде графиков. Сделаны выводы, что метод, который применяет область поиска ограниченный дугами гиперболы, имеет преимущества перед методами с применением областей с формой круга или квадрата. Выяснено, что для всех методов рекомендуется применение малых значений параметра *c* ($0.04 \geq c > 0.01$)