Analysis and prediction of hydrological series based on generalized precedents

Vladimir Naumov [1], Elena Nelyubina[1], Vladimir Ryazanov[2] and Alexander Vinogradov[2]

[1]Kaliningrad State Technical University, Sovietsky prospect 1, 236022 Kaliningrad, Russia

[2]Dorodnicyn Computing Centre, Federal Research Centre Computer Science and Control of Russian Academy of Sciences; Vavilova 40, 119333 Moscow, Russia

This work was done under support of RFBR, Russia

1. Generalized precedents

Generalized precedents (GPs) are computational tools that allow using **on unified basis** different a priori, directly observable or preferable for one reason or another, local regularities in data.

A good example is in the field of **Image Processing**, where one of the important tasks is the restoration of images deformed by interference of smear type. The **a priori information** in this case is that the bright points of the original image appear as **smear lines**. This prepares the basis for efficient reconstruction of the smear parameters and for subsequent restoration of the image as a whole.

At the same time, other local regularities can be more complicated. For example, they may represent **typical geometric specialties** of the training sample considered as spatial object. So, in the simplest case such specialties are **subsets of a certain shape** that are rather densely filled with objects of the training sample. We consider such clusters as independent and self-sufficient **objects of higher level – Basic Clusters (BC)**.

2. Typical basic clusters

Typical local patterns repetitive in the data structure act as **multi-dimensional texture** and can determine **additional individual characteristics of classes**. These characteristics could be used in the problem of **joint processing**, when the set of new objects is formed of representative groups from different classes. In this case, multi-dimensional textural features of classes can be added to *N* basic features. Use of BC **yield many other opportunities.** Some of the last are listed below:

- smaller clusters could contain enough share of training objects whatever basic structure of clusters is chosen;
- densely filled smaller clusters provide obtaining detailed decision rule;
- basic clusters represent inherent **local dependencies in data** and can be considered as **precedents of dependencies themselves**;
- when using dependencies as new objects, the feature space can be transformed to **simpler parametric space**, and volume of training data can be reduced.

3. Examples:

- One-dimensional parameter of **space-filling curve** (such as Peano curve) corresponds to vertices of **quad-tree**.
- Parametric approximation of the sample by **normal mixture**

  $$\sum C_i = \sum \mu_i \exp(-0.5(x_i-x)^T \boldsymbol{\sigma}^{-1}(x_i-x))$$

  with constant covariance matrix **σ.** Each component $N(x_i,\boldsymbol{\sigma})$ represents compact spatial cluster.
- Approximation by **Elementary Logical Regularities of the 1st kind (ELR-1)**. Corresponding clusters of hyper-parallelepipeds in R^N are described each by conjunction $L = \&R_i$, $R_i=(A_i<x_i<B_i)$, that is interpreted as joint manifestation of feature values $x=(x_1,x_2,\ldots,x_N)$ on intervals $(A_i<x_i<B_i)$.
- **ELRs of the second kind (ELR-2)** correspond to the linear constraints of general form $L_i = \&_j R_{ji}$, $R_{ji}=(n_{jin}, x_n)<Thr_{ji}$, where $\boldsymbol{n}_{ji}$ is the normal vector to the *j*-th facet of the *i*-th convex hull, Thr_{ji} is the boundary threshold.

In the first three examples just **limited number of parameters** is used for description of cluster and its filling. Dimensions of new parametric spaces are *1+1, N+1* and *2N+1*:

- in positional tree cluster is coded by one integer and one real parameter;
- In normal mixture cluster is represented by the couple (x_i, μ_i);
- ELR-1 is described by *2N* border marks A_i, B_i and the weight of regularity *L*;
- **in the fourth case** one has to deal with parameters of all hull's facets in use, and it's necessary to maximally reduce this value using **coherent ELR-2s**.

4. Parameterization and basic clusters. Hough transform in higher dimensions.

All of mentioned clusters have **simple parameterization** and represent some local or partial dependency in data.

Typicality of a local dependency itself, as well as **repeating values of its parameters**, can be detected through the analysis of the **secondary clustering structure** in corresponding parametric spaces.

This outlined scheme has obvious correlations with methodology and application of **transforms of Hough type** in **IP** and **SA**. But there are **serious differences**.

1. The main difference is that in this case the parameterized model may correspond to a cluster in **abstract feature space of arbitrary dimension**.
2. It is equally important that the role of primary **spatial differentiation** can play variety of procedures used for identification significant clusters, the shape of which is given in advance and changes within controlled limits. In particular, this is right for many well-studied methods of **approximation the empirical distribution** by a set of elements of **certain type**, just as in the case of Gaussian mixture.
3. There is another significant difference from the classical scheme of Hough transform: building the best approximation is essentially **non-local process**, the outcome of which depends on the geometry of the whole sample.

5. Non-locality

Let's analyze the last difference in more detail. At primary glance, non-locality can devalue all the constructions presented above. But it is also obvious that the presence of **local relationships and dependencies** among parameters of objects is not exclusive or rare event.

In fact, some features of data can be known a priori, and this directly affects the choice of the shape of basic clusters. For example, in IP, when working with images damaged by linear smear, lineament is usually chosen as basic cluster.

Thus one uses **a priori knowledge**, but there is **more unbiased way** to select the basic form of clusters, which corresponds to **inherent local dependencies** in data. As criteria we can use a set of functionals assessing the accuracy of the description of the training sample on the basis of particular types of basic clusters:

- Let $\boldsymbol{B}_s$, $s=1,2,\ldots,S,$ is a set of cluster descriptions that could claim to be the basic. Each object $\boldsymbol{B}_s$ contains parameters of cluster shape that may be relevant to the task of detecting differences between classes $\lambda = 1,2, \ldots, l$. Let $\boldsymbol{Q}_z$, $z=1,2,\ldots,Z,$ is a set of quality criteria that are applied in approximation task for representation of class K_λ, $\lambda=1,2,\ldots,l$, using basic clusters of certain shape $\boldsymbol{B}_s$.
- Thus, we keep in denotation just two variables we need for setting the criterion $\boldsymbol{Q}_z=\boldsymbol{Q}_z(s, \lambda)$, $s=1,2,\ldots,S$, $\lambda =1,2,\ldots,l$, with which we establish $S\times Z$-matrix of votes for selection this or that shape of cluster as basic. Applying the shape set $\boldsymbol{B}_s$, $s=1,2,\ldots,S,$ and the list of criteria $\boldsymbol{Q}_z(s, \lambda)$, $z=1,2,\ldots,Z,$ to λ-th class $X_\lambda \subset X$ of the training sample, we obtain a set of matrices $q_{sz}(\lambda)$, $\lambda=1,2,\ldots,l$, containing votes for basic shape $\boldsymbol{B}_s$ for class λ with respect to z-th criterion $\boldsymbol{Q}_z$.
- The set $q_{sz}(\lambda)$, $\lambda=1,2,\ldots,l$, may serve as an objective basis for selection certain shapes of clusters as basic.

Of course, such choice can be made further on the base of different strategies. Notice, that in all ways we've used **minimum of a priori or subjective knowledge** here.

6. GP as typical dependency in data

As result, the description of the **lowest complexity and minimal error** can indicate that chosen basic form is **relevant to intrinsic relationships**, as well as to their **typicality** in available data. All this is consistent with the concept of **GP**, as noted above.

We present below the calculation scheme of **Hough-type transform in higher dimensions**, where set of **ERL-2** is used as set of **basic clusters**. Goal is to find most typical orientations of border hyper-planes represented in parametric space:

a) at first we construct a set $\boldsymbol{L} = \{L\}$ by finding all ELR-2s that form some covering of a class ;

b) it is chosen a limited number of parameters characterizing border hyper-plane of ELR-2 and their position relatively to the main axes of the feature space. In particular, further we consider parametric space $\boldsymbol{C}$ which provides representation of guide angles α_n of the normal vector n to some border hyper-plane;

c) one-to-one mapping $\vartheta: \boldsymbol{L} \rightarrow \boldsymbol{C}$ of the set $\boldsymbol{L}$ into selected parametric space $\boldsymbol{C}$ is constructed, and there some secondary clustering is performed;

d) while clustering we search for the set $\boldsymbol{C}^T$ of expressed compact clusters $\boldsymbol{c}^t$, $t \in T$, in the space $\boldsymbol{C}$. Each cluster $\boldsymbol{c}^t \in \boldsymbol{C}^T$ represents some typical direction of normal vectors to border planes of different ELR-2 revealed at the first step;

Having the set $\boldsymbol{C}^T$ we can deform representations of ELR-2s collected in the set $\boldsymbol{L}=\{L\}$ with aim to arrange more pairs of **coherent** regularities among them. In what follows, new information about presence of such clusters is used to optimize **DR** starting from analysis of derivative distributions to realization **DR** in the original feature space R^N.

Reverse assembly of the DR can be done on the base of detected **GPs**, when typical basic clusters are restored in R^N from points $c^t \in \boldsymbol{c}^t$, $t \in T$, by inversing $\vartheta^{-1}: \boldsymbol{C} \rightarrow \{L\}$. When this, priority may be given to different elements of $\boldsymbol{C}^T = \{\boldsymbol{c}^t\}$, $t \in T$, depending on the nature of data, requirements on the solution, etc.

In **Fig.1** a model example of ELR-2 covering $\{L\}$ constructed for a class in two dimensions $N=2$ is presented. Corresponding one-dimensional parametric subspace is shown in **Fig.2.**

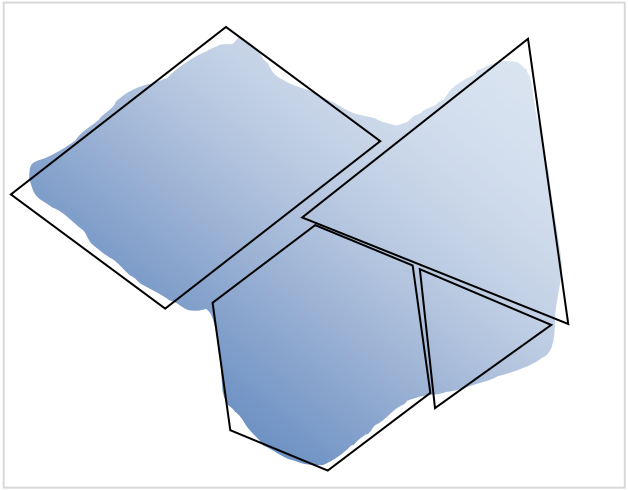

Fig.1. Modeled 4-component ELR-2 covering $\{L\}$ of class K_λ.

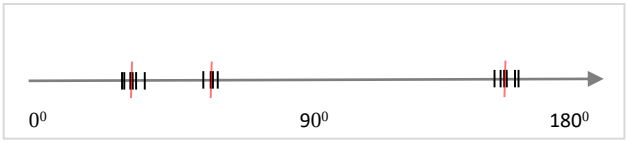

Fig.2. One-dimensional section $\boldsymbol{C}'$ of parametric space $\boldsymbol{C}$ for the covering $\{L\}$. Black dashes show angles between facet normals and the horizontal axis by mod(180^0). Red dashes are unified representatives of the main clusters.

Let $f=1,2,\ldots,F^L$ be the **index of facets in ELR-2 hulls of *{L}***, and $\boldsymbol{n}_f$ is the normal to the f-th facet. We consider clustering in parametric subspace $\boldsymbol{C'} \subset \boldsymbol{C}$ of coordinates of $\boldsymbol{n}_f$. Other parameters of ELRs represented in ***C*** are ignored while clustering in ***C'***, and all facets of all ELRs are **considered simultaneously**. Moreover, to improve the decision rule as whole, we have to mix ELR-2s of the coverings of all classes $K_\lambda, \lambda = 1,2,\ldots,l$, of the training sample and further look for coherent ELR-2 subsets that may unite different classes. Since $|\boldsymbol{n}_f|=1$, the dimension for ***C'*** is chosen *N-1*.

Elements of **improved covering** will have restricted variety of normals to facets. Fig.3 shows a new covering constructed of unified representatives of facets.

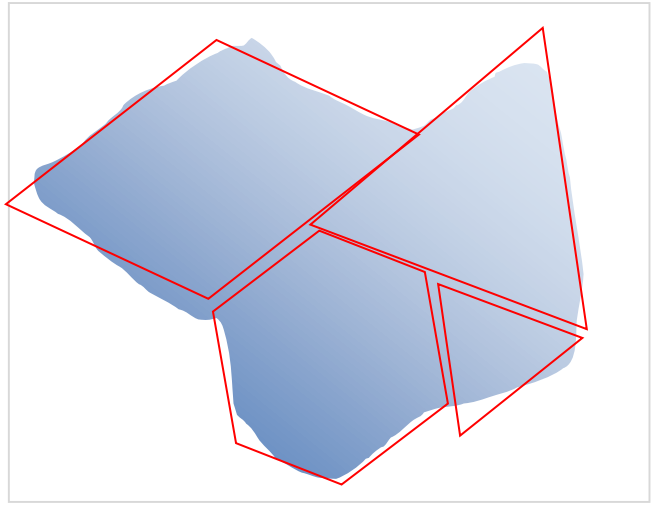

Fig.3. Covering *{L}* improved with unified representatives. Calculating membership in the class K_λ requires 3 convolutions instead of 16 ones.

Notice that normalization of vectors, as calculation of angles, is not an obligation. Here we show **another way to construct the same improved covering** for the class K_λ.

Each boarder hyper-plane corresponding to facet is a **linear manifold of co-dimension 1**. It means that the ideal $\boldsymbol{I}_f$ in the ring $\boldsymbol{R}(x_1,x_{2,\dots},x_N)$ of polynomials on variables x_n, $n=1,2,\dots,N$, that defines the manifold containing f-th facet, is principal ideal produced by a single polynomial of the first order $P_f = (x,\boldsymbol{a}_f)-\boldsymbol{b}_f$, $\boldsymbol{I}_f = (P_f)$, where $\boldsymbol{a}_f$ is a vector orthogonal to f-th facet. If we just look for variety of vectors $\boldsymbol{a}_f$ orthogonal to different facets of all ELR-2s revealed in the training sample and selected at the first step a) of the scheme, we can use directly the **parametric space of coefficients** without normalizing each vector $\boldsymbol{a}_f$.

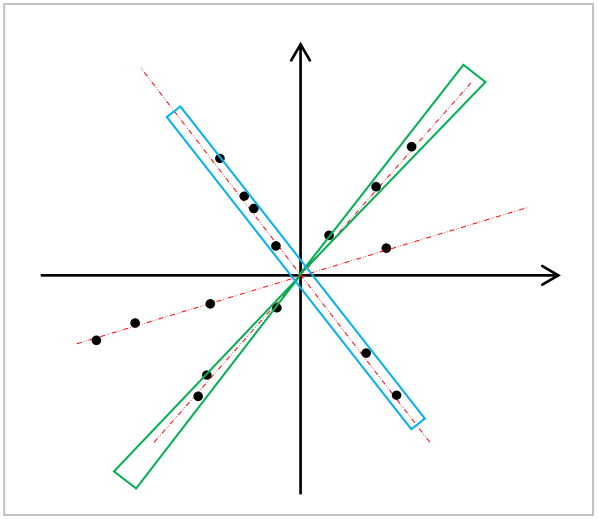

Fig.4 shows 2-dimensional parametric space of such kind for covering $\{L\}$ in Fig.1. As known, each polynomial $(x,\boldsymbol{a}_f)-\boldsymbol{b}_f$ being multiplied by any real value still belongs to the ideal $\boldsymbol{I}_f$. Thus **we obtain a priori knowledge**, that in this case any hyper-plane that's parallel with f-th facet is represented by a point of **one and the same line crossing zero** in the parametric space.

Fig.4. Parametric space of coefficients $\boldsymbol{a}_f$ can be used again in the same scheme in the role of new feature space; centered lineaments (blue) or narrow cones (green) are preferable as basic clusters.

Thus we have **double applying of Hough-type scheme**. Triple of dotted red lines in Fig.4 corresponds to the same choice of representatives.

7. Reconstructing dynamics of hydrological series

Forecasting vs recording actual volume of river runoff is an urgent scientific and economical task. In view of advent of **satellite weather data**, the distribution of precipitation levels and temperatures on weather maps can be reconstructed with **high degree of detail**. Thus, studies on the similar **detailing of river runoff** are also promising. In some cases, the network of gauging stations can be detailed, as in Kaliningrad region:

Name of river	Distance km from		Catchment sq km	Opened
	source	estuary		
Pregola	67.0	56.0	13600	01.04.1869
Angrapa	139.0	30.0	2460	14.03.1894
Instruch	51.0	50.0	587	01.01.1885
Pissa	87.0	11.0	1360	01.08.1894
Pregola	1.00	125	5210	01.05.1886
Pregola	114	8.50	14700	01.01.1811
Lava	271	18.0	7020	01.01.1896
Deyma	32.0	5.00	(distributary)	01.01.1939
Sheshupe	265	43.0	5830	01.09.1955
Zlaya	50.0	12.0	142	31.01.1961
Nelma	26.0	4.00	163	27.09.1963
Mamonovka	45.0	6.20	300	01.10.1959
Neman	878	59.0	91800	01.01.1811
Matrosovka	19.0	24.0	(distributary)	17.12.1968

In less-lived areas, recording actual flow rates for local parts of the river basin is often confronted with fundamental constraints, which makes it difficult to compare precipitation levels and recorded runoff values.

8. Phenomenological model of river flow with precipitation feeding

Our goal is to build **a river basin model** with rain feeding, in which the analysis of actual data and subsequent forecast of runoff behavior rely on the use of GPs as implementations of **local hydrological regularities** that are described by a limited set of parameters. We further show how this approach can reconstruct the flow features in certain regions of the basin, including **regions with complex hydrology**, on the basis of an analysis of only the observed dynamics of **river runoff as a whole**, as well as **detailed meteorological data** for its basin.

We will be interested in the differences in degrees of the **damping (accumulating) effect** of the flow characteristics $Flow_i(t)$ of individual regions R_i on the runoff $Flow(t)$ of river as whole. The main object of further analysis and search will be the dependence F_i of the instantaneous flow rate $Flow_i(t)$ on the current moisture level $Level_i(t)$ at time t in the region R_i:

$$Flow_i(t) = F_i(Level_i(t)). \quad (1)$$

At each moment t, the volume of moisture that enters the region is expressed by the integral value $Input_i(t) = \int^t Prec_i(\tau)d\tau$. The instantaneous rate of flow of the region R_i is determined by the dependence F_i. We assume that all the moisture that has fallen down and been accumulated goes into the runoff of the region, the total volume of the region's runoff to the moment t is also expressed by the integral value $Output_i(t) = \int^t F_i(Level_i(\tau))d\tau$. Since $Level_i(t) = Level_i(0) + Input_i(t) - Output_i(t)$, we obtain integral equation

$$Output_i(t) = \int^t F_i(Level_i(0) + Input_i(\tau) - Output_i(\tau))d\tau, \quad (2)$$

which can be correctly solved only if the initial condition is known $Level_i(0)$, that is the water level for region R_i at the initial moment $t=0$. Solving the equation, we get the instantaneous rate of flow of the region at each moment t: $Flow_i(t) = F_i(Level_i(t))$.

9. Typical flow characteristics as generalized precedents

The main a priori assumption is that each dependence F_i is described by an increasing function, namely, the derivative F'_i is strictly positive, $F'_i>o$, in the function's domain, Fig.5.

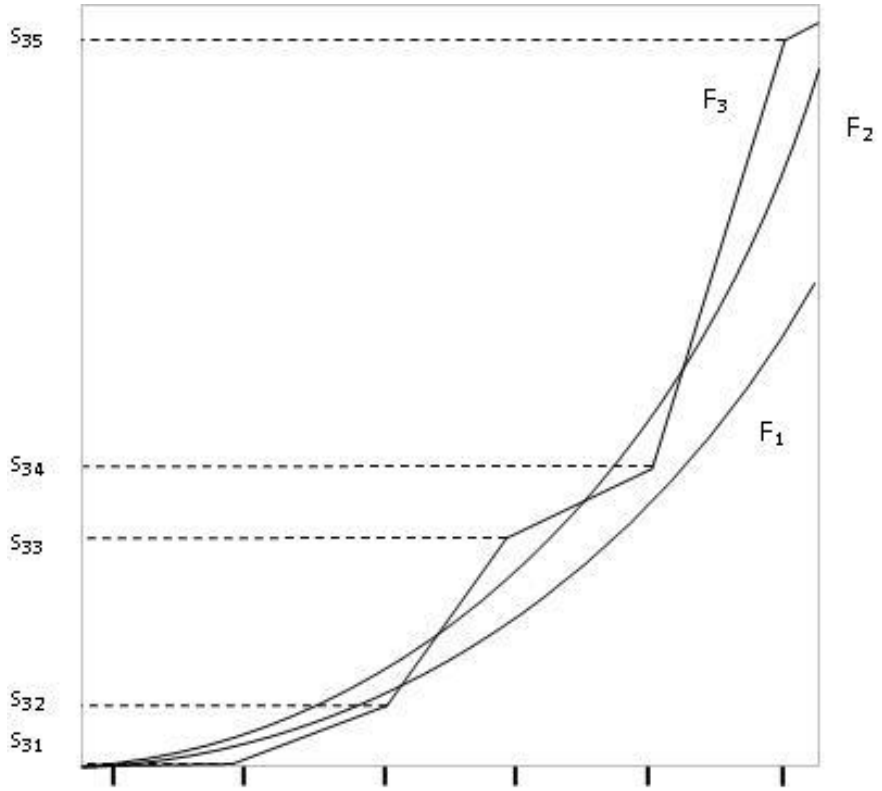

Fig. 5. Variants of natural dependencies F_i

In order to use some dependency F_i as a generalized precedent, we must choose parametric representation for it. Here functions $F_i(x) = s_i x^2$ are used, each of which is uniquely described by scalar parameter $s_i > 0, i = 1,2$. The variant that refers immediately to a more complicated geometric shape s_{ij}, $j=1,2,\ldots,J$, is given for some function F_3 satisfying the restriction $F'_i>o$, where index j, $j = 1, 2, \ldots, 5,$ points particular parameter number.

10. Realizations of generalized precedents in parametric space

Further, we act in accordance with conventional Hough transform procedure.

We refer to $\boldsymbol{s} = [s_{ij}]$ as **structural matrix**, the *i*-th row of this matrix describes some possible character of the flow in the region R_i. In case of usual Hough transform, the appearance of various realizations in the parametric space is controlled by simple criterion - threshold value of the **brightness gradient**. In our case, the role of spatial operation plays not differentiation, but the **solution of equation (2),** which determines evolution of the water level $Level_i(t)$ in regions, using the generalized precedents $\boldsymbol{s}=[s_{ij}]$, $j=1,2,\ldots,J$, $i=1,2,\ldots,I$, and the sums of precipitations, actual at time *t*. In our case, the **criterion for structural matrices *s*** is the following condition

$$\sum \cdot Flow_i(t-\eta_i) = Flow(t). \qquad (3)$$

Here values η_i describe time delays in evolvement the flows of regions R_i into the main watercourse.

Thus, the structural matrices ***s*** satisfying condition (3) appear as points of the discrete secondary distribution on the *IxJ*-dimensional parametric space ***C***. Points corresponding to structural matrices that are most adequate for the available statistics gather in the vicinity of the main mode in the form of expressed cluster ***c****. In this case, the mode vertex $c^* \in \boldsymbol{c^*}$ can be the result of our analysis if the corresponding structural matrix ***s**** better than others fits the statistics used.

Fig. 6 shows the possible shape of cluster ***c**** and the point of the solution matrix ***s**** in 2-D parametric space for pairs (F_1, F_2,). Each point represents one reading for river entire runoff during a month. It reflects the currently relevant meteorological statistics processed in accordance with the structural hypothesis $\boldsymbol{s}=[s_i]$, $i = 1,2$. Deviations from the center mode are caused by measurement errors, truncations, imperfection of the 2-D model $\boldsymbol{s}=[s_i]$, $i = 1,2$, itself, and so on.

Center of the mode is disposed below the diagonal. This means that the inequality $s_1 > s_2$ is most consistent with objective statistics, and thus, the region R_1 has greater damper capacity than the region R_2. In fact, as can be seen from the graph of functions F_1 , F_2 (Fig.3), the rate of flow in region R_1 changes more slowly by variations in the moisture level than in the region R_2.

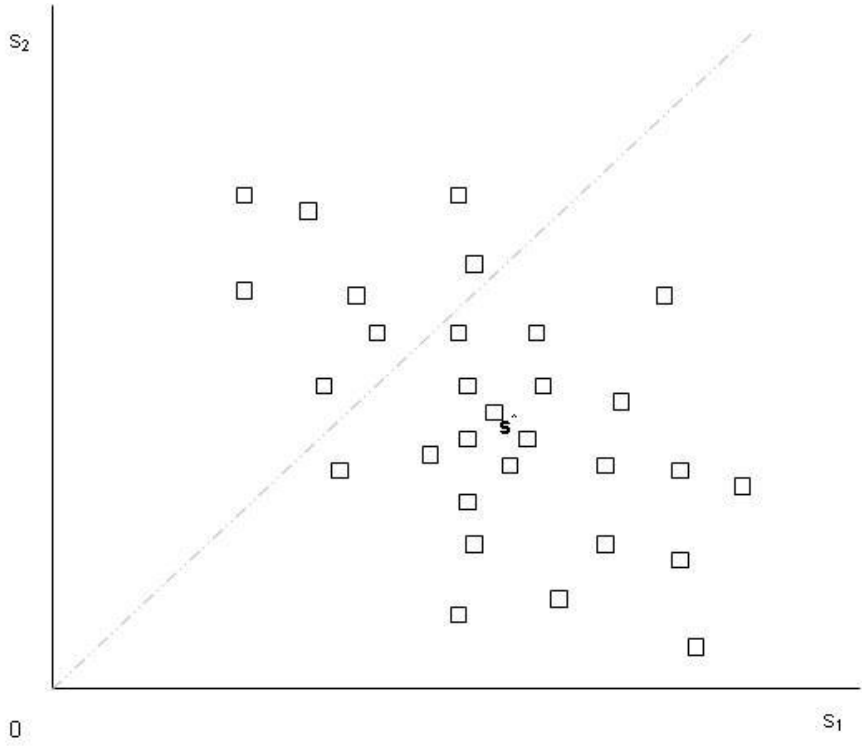

Fig.6. Possible shape of cluster ***c**** with solution point ***s****

1) Having the solution s^* and condition (3), we acquire base to calculate absolute values $Flow_i(t-\lambda_i)$ in accordance with structural hypothesis $[s_{ij}]$, $j=1,2,\ldots,J$, $i=1,2,\ldots,I$. As we pointed out earlier, for some regions direct instantaneous measurement of the values $Flow_i(t-\lambda_i)$ may be complicated or impracticable.

2) Now it's possible to make environmental warnings: if we have big short-time rainfall sum $\int^t Prec_1(\tau)d$ or $\int^t Prec_2(\tau)d\tau$, then for the second region R_2 flood-type situation is more probable.

3) And vice versa, when the sum $\int^t Prec_1(\tau)d$ or $\int^t Prec_2(\tau)d\tau$ is close to zero, fast soil drying is probable for the same region R_2 and not for region R_1 because water leaves the last slower.

We could continue the row of opportunities opened, but it should be said about pitfalls in the presented model, too.

If we have in our example $\eta_1 = \eta_2$, $Prec_1(t) = Prec_2(t)$ for all moments t under consideration, then the secondary distribution on the parametric space $\boldsymbol{C}$ for $s_i>0$, $i=1,2,$ will be symmetric with respect to the diagonal.

In particular, even if the best approximation of the statistical data is achieved with asymmetric choice $s_1 \neq s_2$, nevertheless, in the empirical distribution on $\boldsymbol{C}$ two equally important main clusters symmetrically disposed relative to the diagonal will be present.

This means that the difference in the damping properties of regions R_i, $i=1,2$, is actually found, but it is impossible to tell which of them corresponds to the larger of the parameters s_i, $i=1,2$.

Similar situations are not excluded for greater values of I and J , too. But, such situations arise only in quite exceptional cases and do not have significant effect on the operability of the scheme.

11. CONCLUSIONS

It was considered the concept of generalized precedents as unified computational tool that yields to use local dependencies and regularities in data.

The main stages of applying generalized precedents are presented, and close relationship is shown with the Hough transform.

A computational scheme is presented that is suitable for further optimization of fast linear decision rules from a wide class of them.

Examples of the use of proposed approach are given, in particular, an example of double Hough-type transform in higher dimensions.

On this basis, some possibilities of comparison and joint analysis of meteorological data and actual data on the volume of river flow are investigated. In this case, the generalized precedents are typical nonlinear relationships between certain hydrological parameters.

The goal was to identify differentiation of the regions of the river basin by their accumulating capabilities. We show how this can be done on the basis of an analysis of time-limited contemporary statistics.

Obtained flow characteristics in the regions can be further used for short-term forecasting of river level variations and other hydrological processes and phenomena, including flood and drought situations.

These characteristics can also serve as an important factor in the study of ecosystems, geology of the region and other similar purposes.

Thank You !