




# DPS Clustering: New Results

Sergey M. Agayan <sup>1</sup>, Shamil R. Bogoutdinov <sup>1,2</sup>, Boris A. Dzeboev <sup>1,\*</sup> , Boris V. Dzeranov <sup>1</sup> , Dmitriy A. Kamaev <sup>3</sup> and Maxim O. Osipov <sup>4</sup> 

<sup>1</sup> Geophysical Center of the Russian Academy of Sciences, 119296 Moscow, Russia

<sup>2</sup> Schmidt Institute of Physics of the Earth of the Russian Academy of Sciences, 123995 Moscow, Russia

<sup>3</sup> Research and Production Association “Typhoon”, 249038 Obninsk, Russia

<sup>4</sup> Obninsk Institute for Nuclear Power Engineering, National Research Nuclear University Moscow Engineering Physics Institute (MEPhI), 249040 Obninsk, Russia

\* Correspondence: b.dzeboev@gcras.ru

**Abstract:** The results presented in this paper are obtained as part of the continued development and research of clustering algorithms based on the discrete mathematical analysis. The article briefly describes the theory of Discrete Perfect Sets (DPS-sets) that is the basis for the construction of DPS-clustering algorithms. The main task of the previously constructed DPS-algorithms is to search for clusters in multidimensional arrays with noise. DPS-algorithms have two stages: the first stage is the recognition of the maximum perfect set of a given density level from the initial array, the second stage is the partitioning of the result of the first stage into connected components, which are considered to be clusters. Study of qualities of DPS-algorithms showed that, in a number of situations in the first stage, the result does not include all clusters which have practical sense. In the second stage, partitioning into connected components can lead to unnecessarily small clusters. Simple variation of parameters in DPS-algorithms does not allow for eliminating these drawbacks. The present paper is devoted to the construction on the basis of DPS-algorithms of their new versions, more free from these drawbacks.



**Citation:** Agayan, S.M.;

Bogoutdinov, S.R.; Dzeboev, B.A.;

Dzeranov, B.V.; Kamaev, D.A.;

Osipov, M.O. DPS Clustering: New

Results. *Appl. Sci.* **2022**, *12*, 9335.

[https://doi.org/10.3390/](https://doi.org/10.3390/app12189335)

[app12189335](https://doi.org/10.3390/app12189335)

Academic Editor: Laurent Daudeville

Received: 29 July 2022

Accepted: 13 September 2022

Published: 17 September 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** discrete mathematical analysis; density; localization radius; fuzzy set; perfect set

## 1. Introduction

One of the most interesting and widely used approaches to the multidimensional data analysis are cluster analysis or clustering methods. Currently, there are many clustering algorithms. Despite significant differences between them, they all rely on the initial postulate of compactness: in the space of objects, all “close” objects must belong to the same cluster, and all different objects, respectively, must be in different clusters. The concepts of “proximity” are interpreted differently in different clustering algorithms.

Within the framework of Discrete Mathematical Analysis (DMA)—an original approach to data analysis that uses fuzzy mathematics and fuzzy logic [1], methods of so-called DPS-clustering are being developed. The present study is devoted to DPS clustering and continues the series of papers on this problem [2–7].

The initial notion of DPS-clustering is a fuzzy model of the fundamental mathematical property “limit”. It is called the density and represents a non-negative function depending on an arbitrary subset and any point in the initial space in which clustering is supposed.

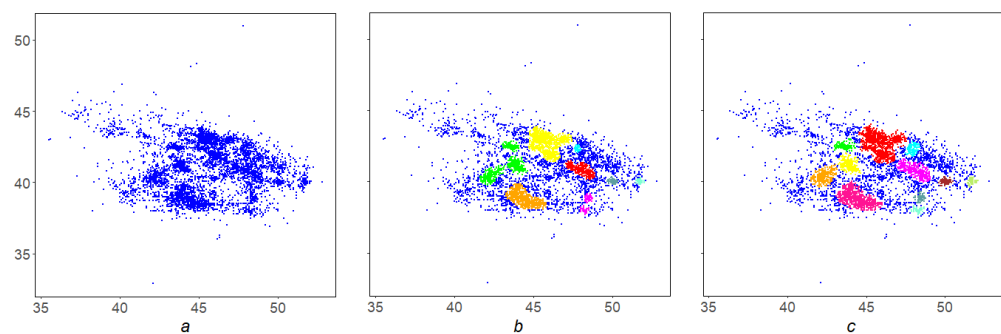
The value of density can be understood as the strength of the connection between a subset and a point, as the degree of influence of a subset on a point, or dually, as the degree of limiting a point to a subset. This view of density automatically requires its monotonicity over a subset: the larger the subset, the stronger its impact on the point, and it is more limiting for it.

Nontrivial densities always exist in a finite metric space (FMS). Each density on the FMS gives a new look at it and a new program of its study, so that the density is a new mathematical concept.

Fixing the density level  $\alpha$  and interpreting it as a limit level, we can introduce the concept of discrete perfection with level  $\alpha$ : a subset is called discretely perfect with level  $\alpha$  ( $\alpha$ -DPS- or simply DPS-set) if it consists of exactly all points of the original space  $\alpha$ -limit to it. A rigorous theory of DPS-sets has been constructed within the framework of DMA. In particular, it has been shown that DPS-sets have the properties of clusters. This, as well as the comparison of DPS-clustering with modern cluster analysis algorithms and its applications, is described in detail in [5].

The DPS-clustering algorithms created to date operate in finite metric spaces, depend on three parameters (density  $P$ , density level  $\alpha$ , and local coverage radius  $r$ ) and have two stages.

At the first stage, topological filtering of the original space is carried out. It is cleaned from noise. The DPS-algorithm iteratively cuts out from the original space (Figure 1a) the maximum  $\alpha$ -perfect subset (Figure 1b).



**Figure 1.** Comparison of algorithms: (a)—initial array; (b)—the result of clustering by SDPS-algorithm; (c)—the result of clustering by the DBSCAN-algorithm.

At the second stage, the DPS-algorithm splits the result of the second stage into  $r$ -connected components for which it considers to be clusters (Figure 1c).

For the array in Figure 1a and for the SDPS-algorithm that is the most known of the DPS-algorithms Section 2.2.2, the clusters are shown in different colors in Figure 1b. Similar clusters obtained as a result of working on a given array of well-known cluster analysis algorithms DBSCAN [8] and OPTICS [9] practically coincide and are shown in Figure 1c. This allows us to conclude that DPS-algorithms, like the DBSCAN and OPTICS algorithms, represent a new stage in cluster analysis, since they not only split the initial space into homogeneous parts, but also preliminarily clear it of noise (filter).

Studies show that there are situations where the result of the first stage does not include all noteworthy clusters. Decreasing the limitation level  $\alpha$  leads to a decrease in the quality of a cluster recognition and is not a way out of the situation.

Furthermore, due to the locality of the radius, the partitioning into  $r$ -connected components of the maximum  $\alpha$ -perfect subset in the second stage is often too shallow, and detailed, and needs to be enlarged as can be seen in Figure 1b.

The present work is devoted to correcting these disadvantages.

## 2. Materials and Methods

This section defines DPS-sets and describes DPS-clustering algorithms.

### 2.1. Discrete Perfect Sets

DMA has a rigorous theory of discrete perfect sets (DPS-sets) in finite spaces, which is summarized in [5]. A complete rigorous justification can be found in [3].

Let  $X$  be a finite set, and let  $A, B, \dots$  and  $x, y, \dots$  be subsets and points in it, respectively.

**Definition 1.** Let us call a density  $P$  on a set  $X$  a product of mapping  $2^X \times X$  to non-negative numbers  $\mathbb{R}^+$ , increasing on the first argument and trivial on the empty:

$$P(A, x) = P_A(x) \quad (1)$$

$$\forall x \in X, A \subset B \Rightarrow P_A(x) \leq P_B(x), P_\emptyset(x) = 0$$

For a fixed  $x$ , the function  $P_A(x)$  is a non-normalized fuzzy measure on  $X$ , so the density  $P$  is a family of such measures parameterized by  $X$  itself.

By fixing a density level  $\alpha$  and interpreting it as a limit level, one can define any topological notions in  $X$ , and discrete perfection with level  $\alpha$  particularly.

**Definition 2.** A set  $A$  consisting of exactly all  $\alpha$ -perfect points of the original space  $X$  is called an  $\alpha$ -discrete perfect (just perfect, DPS-) set in  $X$ :

$$A - \text{DPS-set in } X \Leftrightarrow A = \{x \in X : P_A(x) \geq \alpha\} \quad (2)$$

Numerous studies and examples below show that DPS-sets are clusters in  $X$  and are closely related to clustering in  $X$ .

In the works [2,3,5], a construction is given that allows a subset  $A(\alpha) \subset X$  to be constructed from a subset  $A$  in the space  $X$  and the level  $\alpha$  of the density  $P$ . Under the condition of nontriviality, it will be an  $\alpha$ -DPS-set in  $X$ . It does not have to lie in  $A$ , but if  $A$  itself is an  $\alpha$ -DPS-set in  $X$ , then the set  $A(\alpha)$  constructed for it will coincide with it. Thus, the construction  $A \Rightarrow A(\alpha)$  depends on the space  $X$ , subset  $A$ , density  $P$ , and level  $\alpha$ . To emphasize this fact, we introduce the notation

$$A(\alpha) = A_P(\alpha|X). \quad (3)$$

The properties of the construction  $A \Rightarrow A_P(\alpha|X)$  are formulated in Statement 1, which is proved in [3].

**Statement 1.** The dependences on  $A, P$  and  $X$  are increasing, and the dependence on  $\alpha$  is decreasing:

1. If  $A \subseteq B$ , then  $A(\alpha) \subseteq B(\alpha)$  ;
2. If  $P, Q$  are densities on  $X$  and  $P_A(x) \leq Q_A(x) \forall x \in X$ , then  $A_P(\alpha) \subseteq A_Q(\alpha)$ ;
3. If  $\alpha < \beta$ , then  $A(\beta) \subseteq A(\alpha)$ .
4. If  $A \subseteq X \subset Y$  and density  $P$  is given on  $Y$ , then  $A(\alpha|X) \subseteq A(\alpha|Y)$ .

## 2.2. DPS-Algorithms

In the case of the entire space  $X$ , the transition  $X \Rightarrow X(\alpha)$  is an iterative cutting:

$$X(\alpha) = \bigcap_i X^i(\alpha) \quad (4)$$

$$X^{i+1}(\alpha) = \{x \in X : P_{X^i(\alpha)}(x) \geq \alpha\}, X^0(\alpha) = X$$

If there is a metric on the space  $X$  and the density  $P$  is compatible with it, the property of  $\alpha$ -perfectness is inherited by the connected components of the set  $X(\alpha)$ . Precisely, they most closely correspond in the first approximation to empirical clusters.

In what follows, a metric  $d$  is given on the space  $X$ , so that  $(X, d)$  is a FMS.

### 2.2.1. Topological Digression

Two points  $x$  and  $y$  in  $A$  are called  $r$ -connected if there exists a chain of  $r$ -close to each other points  $x_0, \dots, x_n$  in  $A$  with the beginning  $x_0 = x$ , and the end  $x_n = y$  and distances  $d(x_i, x_{i+1}) \leq r, i = 0, \dots, n - 1$ . The  $r$ -connectivity relation is an equivalence splitting the set  $A$  into  $r$ -connectivity components, which, depending on the context, will

be denoted  $cA(k) = c_r A(k), k = 1, \dots, k^* = k^*(A, r)$ . Let us denote their collection by  $C(A) = C_r(A) = \{cA = c_r A(k) \mid_1^{k^*}\}$ . Thus:

$$A = \vee_{cA \in C(A)} cA = \vee_{k=1}^{k^*} cA(k), \tag{5}$$

where the sign  $\vee$  denotes a disjoint union of sets.

For  $x \in X, A \subset X$ , we denote by  $D_A(x, r)$  the closed ball in  $A$  centered at  $x$  with radius  $r: D_A(x, r) = \{a \in A : d(x, a) \leq r\}$ .

**Definition 3.** Let  $P$  be a density on  $X, r > 0$ . A density  $P$  is called  $r$ -local if

$$\forall x \in X, A \subset X \Rightarrow P_A(x) = P_{D_A(x,r)}(x) \tag{6}$$

**Statement 2.** If the density  $P$   $r$ -local, then every  $r$ -connected component of the set  $X(\alpha)$  is  $\alpha$ -perfect.

**Proof.** According to (2), for any  $k = 1, \dots, k^*$ , it is necessary to establish the equality of the component  $c_r X(\alpha)(k)$  with the set  $\{x \in X : P_{c_r X(\alpha)(k)}(x) \geq \alpha\}$ , which we denote by  $c_r X(\alpha)(k)^1$ .

The inclusion of  $c_r X(\alpha)(k) \subseteq \{x \in X : P_{c_r X(\alpha)(k)}(x) \geq \alpha\}$ .  $x \in c_r X(\alpha)(k) \Rightarrow x \in X(\alpha) \Rightarrow P_{X(\alpha)}(x) \geq \alpha$ . Furthermore,  $D_{X(\alpha)}(x, r) = D_{c_r X(\alpha)(k)}(x, r)$ , since  $d(x, c_r X(\alpha)(\bar{k})) > r$  for  $\bar{k} \neq k$ . Due to the  $r$ -locality of  $P$   $P_{c_r X(\alpha)(k)}(x) = P_{D_{c_r X(\alpha)(k)}(x,r)}(x) = P_{D_{X(\alpha)}(x,r)}(x) = P_{X(\alpha)}(x) \geq \alpha, x \in c_r X(\alpha)(k)^1$ .

The inclusion of  $c_r X(\alpha)(k) \supseteq \{x \in X : P_{c_r X(\alpha)(k)}(x) \geq \alpha\}$ .

First case  $(x \notin c_r X(\alpha)(k)) \wedge (x \notin X(\alpha))$ . Then,  $P_{c_r X(\alpha)(k)}(x) \leq P_X(x) < \alpha \Rightarrow x \notin c_r X(\alpha)(k)^1$ .

Second case  $(x \notin c_r X(\alpha)(k)) \wedge (x \in X(\alpha))$ . Then,  $(x \in c_r X(\alpha)(\bar{k}))$  and  $d(x, c_r X(\alpha)(\bar{k})) > r$ , and therefore  $D_{c_r X(\alpha)(k)}(x, r) = \emptyset$ .

Taking into account the normalization of the density  $P$  and its  $r$ -locality,  $P_{c_r X(\alpha)(k)}(x) = 0$ . Thus, in this case,  $x \notin c_r X(\alpha)(k)^1$ .  $\square$

**Definition 4.** DPS-algorithm depends on three parameters: radius  $r, r$ -local density  $P$  and level  $\alpha$ :  $DPS = DPS(P, \alpha, r)$  and has two stages:

1. The first stage of DPS  $\Leftrightarrow$  the process of constructing for FMS  $(X, d)$  based on the  $r$ -local density  $P$  its  $\alpha$ -envelope  $X(\alpha)$ :

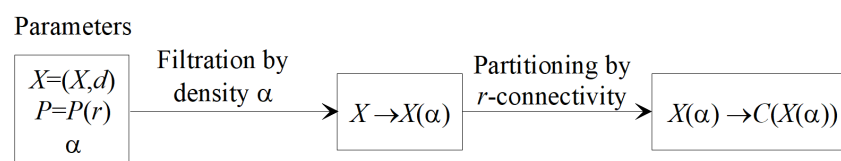
$$DPS(P, \alpha, r) \Leftrightarrow X \Rightarrow X(\alpha)$$

2. The second stage of DPS  $\Leftrightarrow$  the partitioning of the  $\alpha$ -envelope  $X(\alpha)$  into  $r$ -connected components:

$$DPS_c(P, \alpha, r) \Leftrightarrow X \Rightarrow C(X(\alpha)) \subset 2^{2^X}.$$

Further in the text,  $DPS(P, \alpha, r)$  denotes, depending on the context, either the algorithm itself or its first stage.

Let us present a flowchart of the DPS-algorithm (Figure 2).



**Figure 2.** Flowchart of the DPS-algorithm.

Let us summarize the above from the perspective of a cluster analysis: 1st stage  $DPS(P, \alpha, r)$  carves the maximal subset  $X(\alpha)$  dense in the common background from the

original space  $X$ ; 2nd stage  $DPS_c(P, \alpha, r)$  partitions  $X(\alpha)$  into components of  $r$ -connectivity  $cX(\alpha)$ , each of which combines density on background and connectivity, i.e., formally expresses empirical clustering [5].

DPS-clustering has developed methods for choosing the parameters  $\alpha, r$ , which are described in detail in [5] (see also Section 4.3.1). In particular, the density index  $\alpha$  is determined through the level of its extremity  $\beta(\alpha)$ , which answers the question on the scale  $[-1, 1]$ : "To what extent can  $\alpha$  be considered large against the background of all values of the density  $P$  on  $X$ ?"

When constructing  $\beta(\alpha)$ , a fuzzy comparison of  $n$  between non-negative numbers is used. It is expressing on the scale  $[-1, 1]$  the degree of superiority of the larger of them over the smaller one (the left part in (7))

$$\beta(\alpha) = \frac{\sum_{x \in X} n(P_X(x), \alpha)}{|X|}, \quad \frac{\sum_{x \in X} n(P_X(x), \alpha(\beta))}{|X|} = \beta. \tag{7}$$

Due to the properties of  $n$ , the correspondence  $\alpha \Rightarrow \beta(\alpha)$  is unique and the inverse dependence  $\beta \Rightarrow \alpha(\beta)$  is implicitly given by the right side of (7). Thus, the DPS algorithm has a second parameterization, which will be used in the future:

$$\begin{aligned} X(\beta) &\Leftrightarrow X(\alpha(\beta)) \\ DPS(P, \beta, r) &\Leftrightarrow DPS(P, \alpha(\beta), r) \\ DPS_C(P, \beta, r) &\Leftrightarrow DPS_C(P, \alpha(\beta), r) \end{aligned} \tag{8}$$

### 2.2.2. SDPS-Algorithm

Historically, the first in a series of DPS-algorithms was the set-theoretic SDPS. It is based on the density  $S$ , which has the name "Number of points" ("Amount of space") [2,3,6] and conveys the degree of concentration of space  $X$  around each of its points  $x$  (the most natural understanding of the density  $X$  in  $x$ ).

The density  $S_A(x)$  depends on the localization radius  $r$  and the non-negative parameter  $p$ , which takes into account the distance to  $x$  in the ball  $D_A(x, r)$ :

$$S_A(x) = S_A(x|p) = \sum_{y \in D_A(x,r)} \left(1 - \frac{d(x,y)}{r}\right)^p. \tag{9}$$

With  $p = 0$ , we obtain the usual number of points, which explains the name  $S$ :

$$S_A(x|0) = |D_A(x, r)|.$$

The density  $S$  is  $r$ -local, and the SDPS-algorithm is the operation of a DPS-scheme based on  $S$ :  $SDPS = DPS(S, \beta, r)$ .

The result of SDPS-condensing in  $X \Leftrightarrow$  sets locally containing "many  $X$ ". They are formally the most consistent with empirical clusters. By varying the SDPS parameters, one can obtain a fairly complete idea of the hierarchy of clusters in  $X$ .

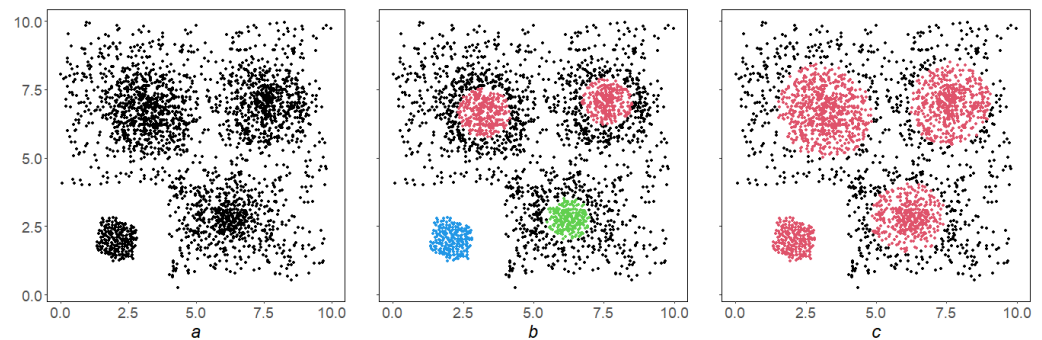
The SDPS-algorithm will serve as a testing ground where new results for the DPS-series algorithms will be tested and shown.

### 3. Results: Iterative DPS

The first DPS-stage of the  $DPS(P, \beta, r)$  algorithm carves a maximal perfect subset  $X(\beta)$ ,  $\beta$ -extremely  $P$ -dense in the background of  $X$  at each of its points, from the space  $X$ .

Let us turn to Figure 3b: it shows the result  $X(\beta)$  of the DPS algorithm on array  $X$  (Figure 3a) in red when  $\beta = 0.02$ . It is easy to see that not all noteworthy condensations from  $X$  were included. The reason for their non-inclusion in the result  $X(\beta)$  is explained by the contradiction between the  $r$ -local character of the view of  $X$  and the global approach to determine the level  $\alpha(\beta)$  of density  $P$  by its extremality level  $\beta$  based on the whole image

$P_X(X)$  (7) and (8): the density level of worthy condensations in the complement  $\overline{X(\beta)}$  appears to be below  $\alpha(\beta)$ .



**Figure 3.** Dependence of the DPS-algorithm result on the level of extremeness: (a)—initial array; (b)—red color shows the result  $X(\beta)$  of the DPS-algorithm, green color shows the result of the algorithm on the complement  $\overline{X(\beta)}$ ,  $\beta = 0.02$ , the blue cluster is the result of one more transition to the complement;  $\beta = 0.02$ ; (c)—red color shows the result at  $\bar{\beta} = -0.25$ .

To include them in the final result, we need to lower the extremum level  $\beta$ :  $\beta \Rightarrow \bar{\beta} < \beta$  and switch from the  $DPS(P, \beta, r)$  to the  $DPS(P, \bar{\beta}, r)$  algorithm. In the control example, the first level of  $\bar{\beta}$ , for which the result  $X(\bar{\beta})$  will include all worthy condensations, will be  $\bar{\beta} = -0.25$ . The result is shown in red in Figure 3c: we see that the result  $X(\bar{\beta})$  of the algorithm  $DPS(P, \bar{\beta}, r)$ , along with  $X(\beta)$  and worthy condensations from  $\overline{X(\beta)}$ , included weak points of the  $r$ -halo of the set  $X(\beta)$ , and it helped them in this.

It can be done otherwise: keep the extremum level  $\beta$ , but change the original space by transition from  $X$  to  $\overline{X(\beta)}$ . The result of  $DPS(P, \beta, r)$  on  $\overline{X(\beta)}$  is shown in green in Figure 3b. If we make another similar transition, a blue cluster appears in Figure 3b.

Such way of using the DPS scheme is called the “iterative DPS” algorithm.

### 3.1. Iterations by Extremality

It will be remembered that the level  $\alpha(\beta)$  of density  $P$  by extremality level  $\beta$  is determined in the  $DPS(P, \beta, r) = DPS(P, \alpha(\beta), r)$  algorithm from Equation (8). Let us introduce the partition  $X = X(\beta) \vee \overline{X(\beta)}$ . According to (7) and (8),

$$\begin{aligned} \beta = n(P_X(X), \alpha(\beta)) &= \frac{\sum_{x \in X(\beta)} n(P_X(x), \alpha(\beta)) + \sum_{x \in \overline{X(\beta)}} n(P_X(x), \alpha(\beta))}{|X|} = \\ &= \frac{|X(\beta)|}{|X|} \frac{\sum_{x \in X(\beta)} n(P_X(x), \alpha(\beta))}{|X(\beta)|} + \frac{|\overline{X(\beta)}|}{|X|} \frac{\sum_{x \in \overline{X(\beta)}} n(P_X(x), \alpha(\beta))}{|\overline{X(\beta)}|}. \end{aligned}$$

From the properties of convexity, hence the inequality

$$\max \left( \frac{\sum_{x \in X(\beta)} n(P_X(x), \alpha(\beta))}{|X(\beta)|}, \frac{\sum_{x \in \overline{X(\beta)}} n(P_X(x), \alpha(\beta))}{|\overline{X(\beta)}|} \right) \geq \beta.$$

However, the left mean  $\leq 0$  because  $\forall x \in X(\beta) \Rightarrow P_X(x) \geq P_{X(\beta)}(x) \geq \alpha(\beta)$ . Hence, if  $\beta \geq 0$ , the right mean must be  $\geq \beta$ :

$$\sum_{x \in \overline{X(\beta)}} \frac{n(P_X(x), \alpha(\beta))}{|\overline{X(\beta)}|} \geq \beta.$$



Then, from the inequalities  $\forall x \in X \Rightarrow P_{\overline{X(\beta)}}(x) \leq P_X(x)$  and the properties of fuzzy comparison  $n$ , we have:

$$\begin{aligned} n\left(P_{\overline{X(\beta)}}\left(\overline{X(\beta)}\right), \alpha(\beta)\right) &= \frac{\sum_{x \in \overline{X(\beta)}} n\left(P_{\overline{X(\beta)}}(x), \alpha(\beta)\right)}{|\overline{X(\beta)}|} \geq \\ &\geq \sum_{x \in \overline{X(\beta)}} \frac{n\left(P_X(x), \alpha(\beta)\right)}{|\overline{X(\beta)}|} \geq \beta. \end{aligned}$$

It follows that the density level  $\alpha_1(\beta)$ , required for the operation of  $DPS(P, \beta, r)$  algorithm on the space  $\overline{X(\beta)}$ :

$$n\left(P_{\overline{X(\beta)}}\left(\overline{X(\beta)}\right), \alpha_1(\beta)\right) = \beta$$

does not exceed the level  $\alpha(\beta)$ , which we will consider as zero:  $\alpha(\beta) = \alpha_0(\beta)$ , and can be strictly less than it. In addition, this, in turn, means the possible nontriviality of such operation, i.e., the set  $\overline{X(\beta)}(\beta)$ . Let us denote it by  $X_1(\beta)$ .

In the situation of non-triviality  $X_1(\beta)$ , one can define the first iteration on the extremity  $X(\beta, 1)$  of the set  $X(\beta)$ , which is naturally considered as its null iteration, assuming  $X(\beta, 0) = X(\beta)$ .

**Definition 5.** Let  $X_1(\beta)$  be non-trivial, and we will call by the first iteration of the space  $X$  with respect to the extremity level  $\beta$  the disjoint union

$$X(\beta, 1) = X(\beta, 0) \vee X_1(\beta).$$

**Note 1.** Given this definition, the above can be understood as the first precondition for the existence of the first iteration on extremality when  $\beta$  is non-negative. In this connection, the zero level of  $\beta$  appears to be the most productive and interesting, at which the second precondition (non-triviality of the result) will be the weakest.

A direct check shows that  $P_{X(\beta, 1)}(x) \geq \alpha_1(\beta) \forall x \in X(\beta, 1)$ , so, by repeating the above deduction with replacement  $\alpha_0(\beta) \Rightarrow \alpha_1(\beta)$ ,  $X(\beta, 0) \Rightarrow X(\beta, 1)$ , we obtain the level  $\alpha_2(\beta) \leq \alpha_1(\beta)$  of the  $DPS(P, \beta, r)$  algorithm on the complement  $\overline{X(\beta, 1)}$  and the possible nontriviality of its result  $X_2(\beta) = \overline{X(\beta, 1)}(\beta)$ .

If all this can be continued up to and including the  $i$ -th step, i.e., the sets  $X_1(\beta), \dots, X_i(\beta)$ , will be nontrivial, then their union with  $X(\beta, 0) = X(\beta)$  will be called the  $i$ -th iteration of  $X$  by extremum level  $\beta$ :

$$X(\beta, i) = X(\beta, 0) \vee X_1(\beta) \vee \dots \vee X_i(\beta). \tag{10}$$

### 3.2. Iterations of the DPS Algorithm

Considering the algorithms  $DPS(P, \beta, r)$  and  $DPS_c(P, \beta, r)$  as their zero iterations:  $DPS^0(P, \beta, r) = DPS(P, \beta, r)$ ,  $DPS_c^0(P, \beta, r) = DPS_c(P, \beta, r)$ , we define their  $i$ -th iterations as processes of building for space  $X$  its  $i$ -th iteration  $X(\beta, i)$  and then breaking it into  $r$ -connectivity components:

$$\begin{aligned} DPS^i(P, \beta, r) &\Leftrightarrow X \Rightarrow X(\beta, i) \\ DPS_c^i(P, \beta, r) &\Leftrightarrow X \Rightarrow C_r(X(\beta, i)) \end{aligned} \tag{11}$$

Due to the disjunctive nature of the decomposition (10), the iteration index  $\text{ind}$  is correctly defined at the  $i$ -th iteration of  $X(\beta, i)$ :

$$\forall x \in X(\beta, i) \Rightarrow \text{ind}(x) = \begin{cases} 0, & \text{if } x \in X(\beta) \\ \bar{i} \in [1, i], & \text{if } x \in X_{\bar{i}}(\beta) \end{cases} .$$

The index  $\text{ind}$  moves from points to connectivity components  $c = cX(\beta, i)$ , thus becoming multivalued:

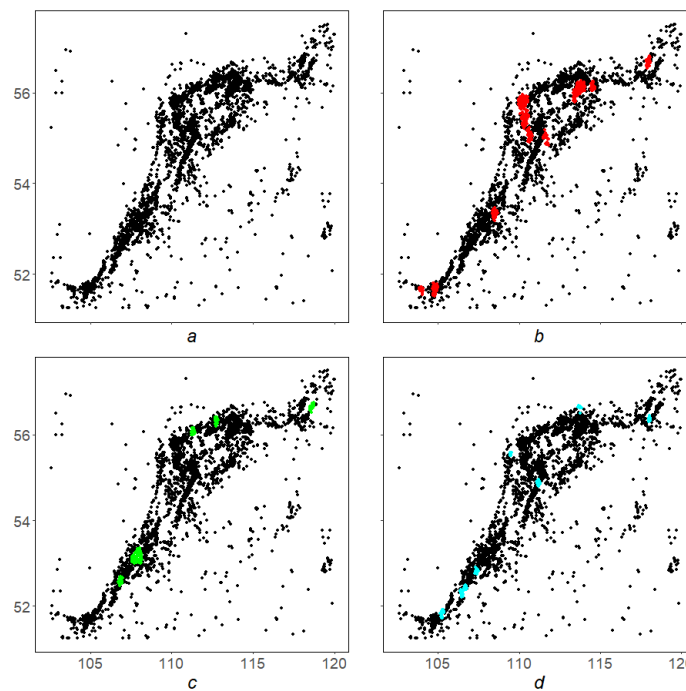
$$\text{ind}(c) = \{\text{ind}(x) : x \in c\}.$$

In addition, this, in turn, makes it possible for any subset  $I \in [0, i]$  to define a conditional version of the  $\text{DPS}^I$  algorithm:

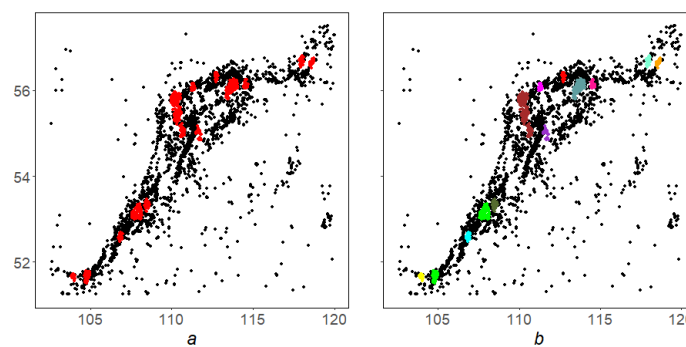
$$\begin{aligned} \text{DPS}^I(P, \beta, r) &= \bigvee_{i \in I} X_i(\beta) \\ \text{DPS}_c^I(P, \beta, r) &= \{c \in C(X(\beta, i)) : \text{ind}(c) \in I\} \end{aligned} \tag{12}$$

**Example 1.** Figure 4b–d show the sets  $X_i(\beta)$ ,  $i = 0, 1, 2$  for the array  $X$  shown in Figure 4a, obtained as a result of the operation of the SDPS-algorithm at  $\beta = -0.09$  based on the fuzzy comparison  $n(t, s) = (s - t) / (s + t)$  of non-negative numbers  $s$  and  $t$  [1].

Figure 5 shows both stages of the  $\text{SDPS}^1$ -algorithm. Its result  $\text{SDPS}_c^1(X)$  indicates that, in the general case, the second stage of DPS cannot be considered final: many components of the  $r$ -connectivity in Figure 5b need further connection. This will be done by the third stage of DPS, the results of which on a given array  $X$  for a given SDPS-algorithm will be shown in Example 4.



**Figure 4.** DPS-algorithm iterations: (a)—initial array; (b)—the set  $X_0(-0.09)$ ; (c)—the set  $X_1(-0.09)$ ; (d)—the set  $X_2(-0.09)$ .



**Figure 5.**  $\text{SDPS}^1$ -algorithm: (a)—the first stage; and (b)—the second stage.



#### 4. Results: Third Stage of DPS

The first stage of the algorithm  $DPS \Leftrightarrow DPS(P, \alpha, r)$  cuts out from the original FMS  $X$  the maximum  $\alpha$ -perfect subset  $X(\alpha) : DPS(X) = X(\alpha)$ . The second stage of the algorithm  $DPS_C \Leftrightarrow DPS_C(P, \alpha, r)$  is considered to be its result and is the set of all components of the  $r$ -connectivity of the set  $X(\alpha) : DPS_C(X) = C_r X(\alpha)$ . Each  $c_r X(\alpha)$  component is considered to be a DPS-cluster because, by virtue of the initial assumption that the initial density is  $r$ -local,  $P$  is a DPS-set in  $X$  and, in particular, is  $\alpha$ -dense at each of its points.

At this stage, the result  $DPS_C(X)$  is taken as a given and is not subject to further transformation. The radius  $r$  is assumed to be infinitely small, and the level  $\alpha$  of  $P$  is infinitely large. As a consequence, each component  $cX(\alpha)$  is considered to be a single and indivisible spot (large point).

Spots  $cX(\alpha)$  are interpreted as fragmentary manifestations (exits) of global anomalous entities in  $X$ . To understand their true scales, if possible, further connection of spots is necessary. This is the third and final stage of the DPS-algorithm.

##### 4.1. Logic and Action Plan

A collection of spots  $C \subset DPS_C(X)$  is considered by the expert as a whole if the degree of advantage in the closeness of internal transitions (paths) between any of its spots over external transitions from  $C$  to  $\bar{C}$ , allows the expert to conclude that  $C$  is non-random.

Thus, the cluster of spots  $C$  must have a tighter internal connection between them in the general background of all spots from  $DPS_C(X)$ . There are no other constraints on  $C$ , in particular on shape.

Because of the  $\alpha$ -perfectness of the  $cX(\alpha)(k)$  spots, it is natural to consider the minimum distance between the closest points in them as the distance between them:

$$d(cX(\alpha)(k), cX(\alpha)(\bar{k})) = \min\{d(y, \bar{y}) : y \in cX(\alpha)(k), \bar{y} \in cX(\alpha)(\bar{k})\}. \tag{13}$$

It will no longer be a metric since there are no triangle inequalities for  $d$  (13), as the example below shows:

**Example 2.** Consider subsets of the real line:  $A = [1, 1]$ ,  $B = [2, 3]$ ,  $C = [4, 4]$ . It is obvious  $d(A, B) + d(C, B) < d(A, C)$ ,  $d(A, B) = |2 - 1|$ ,  $d(B, C) = |4 - 3|$ ,  $d(A, C) = |4 - 1|$ .

Based on the distance  $d$ , the formalization of the expert’s logic for connecting spots is constructed: in each cluster of spots  $C$ , any two spots can be connected by a chain of spots whose distance  $d$  between neighboring links will be smaller than the distance  $d$  from  $C$  to  $\bar{C}$ . This is the  $d$ -clustering condition necessary for  $C$ . Its formal analysis will be the first part of the functioning program to algorithmise the third stage of the DPS. Note that  $d$ -clustering and DPS-clustering are different interpretations of clustering: the first is based on connectivity, and the second on limiting.

The second part of this program consists of constructing, based on the initial distance  $d$  in the original space  $X$ , measures of proximity and distance between spots expressing in this the expert. Based on these, only those spots are selected from the  $d$ -clusters of spots in which the proximity advantage of the inner transitions over the range of the outer ones will be non-random in the expert’s opinion.

##### 4.2. FirstPart: The Theory of $d$ -Clusters

Initial data and designations:  $X$  is a finite set,  $d$  is a quasi-metric in  $X$  ( $(X, d)$  is a quasi-metric space)

$$\begin{aligned} \forall x \in X &\Rightarrow d(x, x) = 0 \\ \forall x, y \in X &\Rightarrow d(x, y) = d(y, x) \neq 0 \end{aligned}$$

**Definition 6.** By the enumeration  $X(x)$  of space  $X$  with origin at the point  $x$ , we call the sequence  $x = x_0, \dots, x_{|X|-1}$  for which for any  $i = 0, \dots, |X| - 2$

$$d(x_{i+1}, X_i(x)) = d\left(X_i(x), \overline{X_i(x)}\right), \tag{14}$$

where  $X_i(x) = \{x_0, \dots, x_i\}$ .

4.2.1. The Notation

$d(x)$  is a numerical sequence of distances  $d_i(x) = d\left(X_{i-1}(x), \overline{X_{i-1}(x)}\right)$ :

$$d(x) = \{d_1(x), \dots, d_{|X|-1}(x)\} \tag{15}$$

**Definition 7.** Let us call an eigen  $d$ -cluster  $C = C_k(x)$ ,  $1 \leq k \leq |X| - 2$  with center (origin) in  $x$  the eigen segment  $X_k(x)$  of the enumeration  $X(x)$ , for which  $\max_{i=1}^k d_i(x) < d_{k+1}(x)$ :

$$C = C_k(x) \Leftrightarrow \begin{matrix} \text{eigen} \\ d\text{-cluster} \end{matrix} \Leftrightarrow (C = C_k(x)) \wedge \left(\max_{i=1}^k d_i(x) < d_{k+1}(x)\right) \tag{16}$$

**Statement 3.** The cluster  $C_k(x)$  is independent of the choice of origin within itself:

$$C_k(x) = C_k(y) \quad \forall y \in C_k(x) \setminus x$$

In other words, in the case of  $d$ -clustering, there is a set-theoretic equality  $X_k(x) = X_k(y)$  that is, for any  $y \in C_k(x)$ , the first  $k + 1$  terms in the enumeration  $X(y)$  will lie in  $C_k(x)$ .

**Proof.** Induction on the number  $i$  of steps  $y_{i-1} \Rightarrow y_i, i = 1 \leq i \leq k$  in the enumeration  $X(y)$ . Let  $y = x_j, j = 1, \dots, k$ .

- $i = 1$ . We have to show that  $y_1 \in C_k(x)$ . By virtue of  $y \in C_k(x)$ , we have

$$d\left(y, \overline{C_k(x)}\right) \geq d\left(C_k(x), \overline{C_k(x)}\right) = d_{k+1}(x).$$

The element  $y = x_j$  in  $C_k(x) = X_k(x)$  has two neighbors  $x_{j-1}$  and  $x_{j+1}$ , if  $1 \leq j < k$ , and one neighbor  $x_{k-1}$ , if  $j = k$  and  $y = x_k$ . To any of the neighbors, the distance from  $y$  will be less than  $d_{k+1}(x)$ , so the next element after  $y = y_0$  in the sequence  $X(y)$  necessarily lies in  $C_k(x)$ .

- Let the assumption that the first  $i$  steps of  $y_i, i \leq k - 1$  in the sequence  $X(y)$  lie in  $C_k(x)$  be satisfied. Let us show that the element  $y_{i+1}$  will also lie in  $C_k(x)$ .

Under our assumptions  $X_i(y) \subset C_k(x)$ , therefore

$$d(X_i(y), \overline{C_k(x)}) > d(C_k(x), \overline{C_k(x)}) = d_{k+1}(x).$$

Assume  $y_m = x_{s(m)}, m = 0, \dots, i$ . If  $s_0 < \dots < s_i$  is the correct ordering of array  $s(m)|_0^i$ , then two cases are possible:

1st: the array  $s_p|_0^i$  has no gaps, that is, it is an eigensegment in  $[0, k]$ . In this case,  $C_k(x)$  necessarily contains either an element  $x_{s_0-1}$ , or an element  $x_{s_i+1}$ , whose distance from  $X_i(y)$  is nontrivial and less than  $d_{k+1}(x)$ . Hence, the term  $y_{i+1}$  in the sequence  $X(y)$  necessarily lies in  $C_k(x)$ :

2nd:  $s_p|_0^i$  array has gaps. Let  $p^*$  be the first number for which  $s_{p^*+1} - s_{p^*} > 1$ . In this case, in the array  $s_p|_0^i$ , it has no number  $s_{p^*} + 1$  and in the set  $X_i(y)$  of an element  $x_{s_{p^*+1}}$ , whose distance from  $x_{s_{p^*}}$  will be less than  $d_{k+1}(x)$ . Hence, in this case too, the term  $y_{i+1}$  in the sequence  $X(y)$  lies in  $C_k(x)$ .

□

**Consequence 1.** *Eigen  $d$ -clusters behave like non-Archimedean balls: either one contains the other or they do not intersect. In particular, two  $d$ -clusters of the same order either coincide or do not intersect.*

**Proof.** Let  $C = C_k(x)$ ,  $\bar{C} = C_{\bar{k}}(\bar{x})$  and  $y \in C \cap \bar{C}$ . According to Statement 3,  $C = X_k(y)$ ,  $\bar{C} = X_{\bar{k}}(y)$ . If  $k \leq \bar{k}$ , then  $C \subseteq \bar{C}$ .  $\square$

The  $d$ -clusters of  $C_k(x)$  are eigennonpoint subsets in  $X$ . Let us remove this restriction on  $k$  in Definition 7 using the notion of a connectivity exponent, which develops and continues the topic of finite connectivity Section 2.2.1.

**Definition 8.** 1. *Connectivity index  $r(A)$  of a subset  $A \Leftrightarrow$  minimum  $r$  for which  $A$  is  $r$ -connected:*

$$r(A) = \min\{r : A \text{ } r\text{-connected}\}.$$

2. *The isolation index  $e(A)$  of a subset  $A \Leftrightarrow$  the distance to its complement  $\bar{A}$ :*

$$e(A) = d(A, \bar{A}).$$

**Statement 4.**

$$r(X_k(x)) = \max_{i=1}^k d_i(x)$$

**Proof.** 1. Proof of the inequality  $r(X_k(x)) \leq \max_{i=1}^k d_i(x)$ . Induction on  $k$ ,  $1 \leq k \leq |X| - 2$

- $k = 1$ :  $X_1(x) = \{x_0, x_1\}$  and  $d(x_0, x_1) = d_1(x)$  according to (14) and (15).
- Suppose that  $r(X_{k-1}(x)) \leq \max_{i=1}^{k-1} d_i(x)$ , that is, any two points in  $X_{k-1}(x)$  can be connected by a path with distance transitions  $\leq \max_{i=1}^{k-1} d_i(x)$ . Let  $x^*$  be the point in  $X_{k-1}(x)$ , closest to  $x_k$ . According to (15),  $d_k(x) = d(x^*, x_k)$ . Through  $x^*$  any point in  $X_{k-1}(x)$  can be connected to  $x_k$  by a path with distance transitions  $\leq \max_{i=1}^k d_i(x)$ .

2. Proof of the inequality  $r(X_k(x)) \geq \max_{i=1}^k d_i(x)$ : let  $d_{i^*}(x) = \max_{i=1}^k d_i(x)$ , then  $d_{i^*}(x) = d(X_{i^*-1}(x), \overline{X_{i^*-1}(x)})$ , and any path with ends in  $X_{i^*-1}(x)$  and  $\overline{X_{i^*-1}(x)}$  has at least one jump between them. Therefore,  $r(X_k(x)) \geq d_{i^*}(x)$ .  $\square$

For the eigen  $d$ -cluster  $C$  (Definition 7), the proved statement means that the necessary condition is fulfilled

$$r(C) < e(C). \tag{17}$$

It is invariant not only from the beginning of the enumeration but also from the enumeration itself, which is obligatory at this point for the definition of  $C$ . Therefore, it is natural to check this condition for sufficiency of  $d$ -clustering.

**Statement 5.** *If (17) is satisfied for an eigensubset of  $A$  in  $X$ , then  $A$  is a  $d$ -cluster and  $A = C_{|A|-1}(x) \forall x \in A$ .*

**Proof.** There exists  $r^1$ :  $r(A) \leq r^1 < d(A, \bar{A})$  for which the set  $A$  is  $r^1$ -connected. Hence, in the enumeration  $X(x)$  of space  $X$  from any point  $x \in A$  the elements from  $\bar{A}$  will appear only at the  $|A|$ -th step.  $\square$

The result of Definition 8, Statement 4 and Statement 5 is a new invariant and more complete definition of a  $d$ -cluster.

**Definition 9.** *Let us call a  $d$ -cluster in  $X$  the subset  $C$  for which there is a true inequation:  $r(C) < d(C, \bar{C})$ .*

To the eigenclusters  $C$  ( $1 < |C| < |X|$ ) will be added all points  $x \in X: x = C_0(x)$ , and all space  $X: X = C_{|X|-1}(x) \forall x \in X$ .

Thus, the restrictions in Definition 7 on  $k$  ( $1 \leq k \leq |X| - 2$ ) are removed:

- $x \Leftrightarrow d$ -cluster, since  $0 = r(x) < e(x) \forall x \in X$ ,
- $X \Leftrightarrow d$ -cluster, since  $r(X) < e(X) = \infty$ : no appearance for  $X$  means that it is at infinity.

#### 4.2.2. Conclusions

The collection of  $d$ -clusters  $C(X, d)$  forms a hierarchy of sets on  $X$  based on the notion of discrete connectivity and different from traditional hierarchical cluster analysis hierarchies, usually binary [10].

$C(X, d)$  is the first part of the formalization of the third stage of DPS.

#### 4.3. Second Part: Final Selection of D-Clusters

Thanks to (14)–(16), the search for  $d$ -clusters in the space of  $r$ -connectivity components  $DPS_C(X)$  with respect to distance (13), that is, the formation of the space  $C(DPS_C(X), d)$  is constructive. Each  $d$ -cluster  $C \in C(DPS_C(X), d)$  has two characteristics: internal  $r(C)$  and external  $e(C)$  (17). Based on this information, expert E must decide: is it necessary to combine the spots from  $C$  into a single whole or not?

Let  $DPS_f(X) = DPS_f(X|E)$  denote the set of maximal  $d$ -clusters in  $DPS_C(X)$ , obtained after joining spots by expert E. According to Consequence 1, they are all disjoint.

Under the condition that  $DPS_f(X)$  is nontrivial in the set  $X(\alpha)$ , a partition appears in the general case larger than the partition (5) into  $r$ -connectivity components:

$$X(\alpha) = \{C : C \in DPS_f(X)\} \vee \{cX(\alpha) : \forall C \in DPS_f(X) \Rightarrow cX(\alpha) \notin C\} \tag{18}$$

Let it also denote by  $DPS_f(X)$  and consider it the third and final stage of the DPS algorithm. Full history is presented in Figure 6.

Thus, for each component  $c_r X(\alpha)$ , one of three things can happen:

- $c_r X(\alpha) \in DPS_f(X) \Leftrightarrow$  component  $c_r X(\alpha)$  is sufficiently isolated from the rest to be of interest to the expert, it is the only way out of the global entity behind it on  $X$ ;
- $c_r X(\alpha) \notin DPS_f(X)$ , but  $c_r X(\alpha) \subset C \in DPS_f(X) \Leftrightarrow$  component  $c_r X(\alpha)$  is part of the  $d$ -cluster  $C$ , which is of interest to the expert and in the team represents a global entity in  $X$ ;
- $c_r X(\alpha) \notin C \forall C \in DPS_f(X) \Leftrightarrow$  any  $d$ -cluster containing  $c_r X(\alpha)$  is of no interest to the expert: either it is not internally dense enough, or it is externally separable. According to the expert, if  $c_r X(\alpha)$  is a fragment of the output of the global entity on  $X$ , then it is not clear enough.

Furthermore, Example 3 will show that all options for  $c_r X(\alpha)$  are possible.

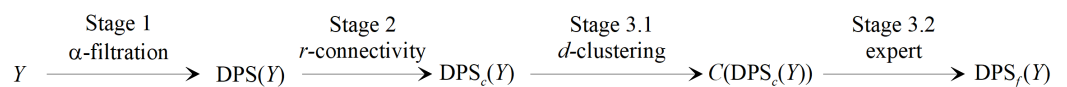


Figure 6. Complete block-diagram of the DPS-algorithm.

At stage 3.2 (Figure 6), the expert can act in different ways. Let us present the simplest Boolean variant of actions for the formation of  $DPS_f(X)$ .

##### 4.3.1. Boolean Variant

Expert E decides whether  $d$ -cluster  $C$  is included in the final result  $DPS_f(X)$  based on comparison of  $r(C)$  and  $e(C)$  with their proximity and distance thresholds  $r_E$  and  $e_E$ :  $C \in DPS_f(X) \Leftrightarrow r(C) \leq r_E$  and  $e(C) \geq e_E$ .

Expert E considers the parameter  $r$  of the DPS-algorithm to be very small (infinitesimal), much less than  $r_E$ , which, in turn, according to E, is much less than  $e_E$ :

$$r \leq r_E \leq e_E.$$

The threshold  $r_E$ , like the radius  $r$  in the DPS-algorithm, is built using the Kolmogorov averaging of non-trivial distances of the FMS  $X$  [5]

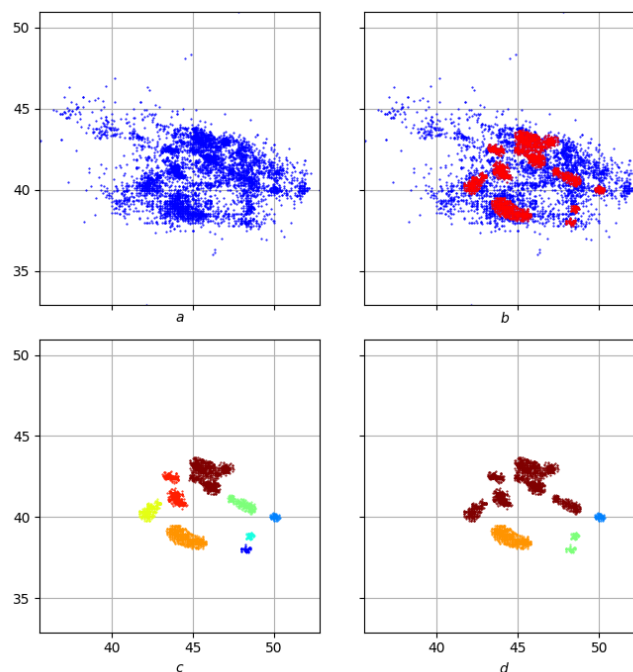
$$r_E = \left( \frac{\sum_{x \neq y \in X} d(x, y)}{|X|(|X| - 1)} \right)^{1/q}, q = q(r_E)$$

For the parameter  $r$ , numerous applications of DPS-series algorithms have established that the choice of  $q(r) \in [-3, -2]$  can be considered optimal. The studies carried out within the framework of present paper show that  $q(r_E) \in [-2.5, -1, 5]$ . The intersection of the areas of parameters  $r, r_E$  is explained both by the fuzzy perception of proximity by the expert and by the diversity in the arrangement of an arbitrary FMS.

The threshold  $e_E$  is obtained from  $r_E$  by formalizing the expert judgment “ $r_E \ll e_E$ ” using fuzzy comparison. For the comparison given in Example 1, this would be the inequality  $e_E/r_E \geq 5/3$ .

**Example 3.** Array  $X$  in Figure 7a has already been seen above: it was a testing ground for DPS, DBSCAN, and OPTICS algorithms. Figure 7b–d show the complete DPS-scenario with parameters  $\beta = -0.25, q(r) = -2.7, q(r_E) = -2.3, e_E/r_E = 5/3$ .

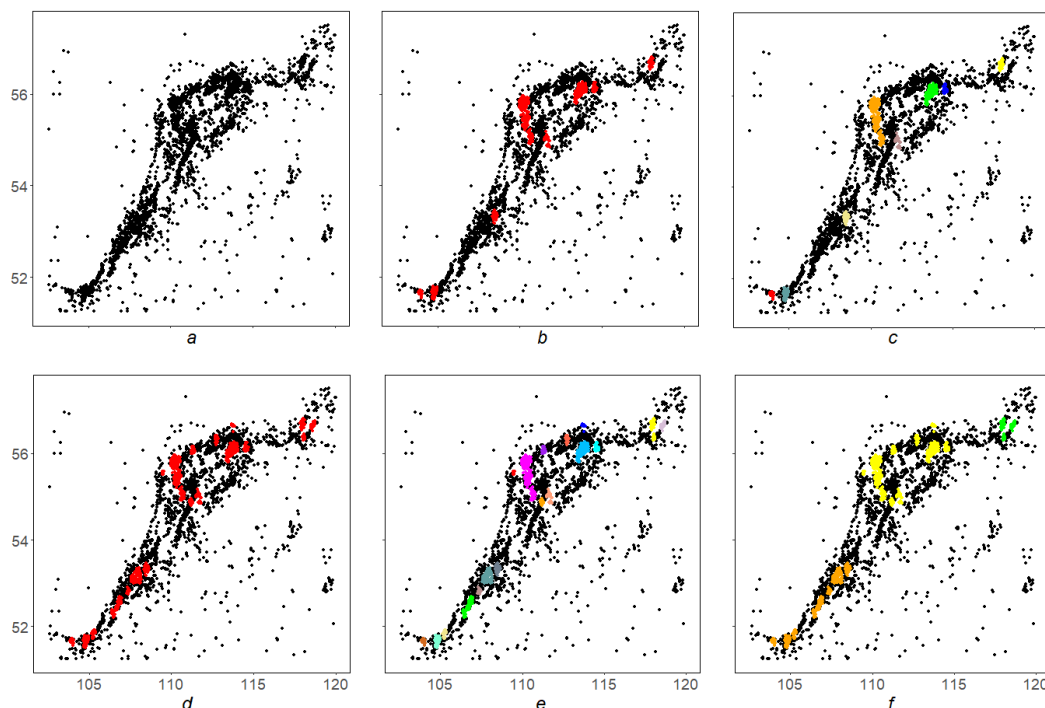
The result of cutting at the first stage (Figure 7b) was divided into eight  $r$ -connectivity components at the second stage (Figure 7c). Two components (blue and light brown) and two more in a green combination passed independently to the third stage (Figure 7d). The remaining four brown components did not pass the third stage.



**Figure 7.** The complete scheme of the SDPS algorithm: (a)—initial array; (b)—the first stage; (c)—the second stage; (d)—the third stage.

**Example 4.** In the conditions, notation and parameters of Example 1 Figure 8a–c show the original array and two stages of the SDPS algorithm on it. Figure 8d–f show the complete scenario for his second iteration of SDPS<sup>2</sup>.

Comparison of Figure 8c,f serves as clear evidence of the work done in the paper: the full version of the second iteration of SDPS<sup>2</sup> with the same parameters performs better than the incomplete zero version (the original SDPS algorithm) that existed before this article.



**Figure 8.** Comparison of SDPS and SDPS<sup>2</sup> algorithms: (a)—initial array; (b)—the first stage of SDPS; (c)—the second stage of SDPS; (d)—the first stage of SDPS<sup>2</sup>; (e)—the second stage of SDPS<sup>2</sup>; (f)—the third stage of SDPS<sup>2</sup>.

### 5. Discussion

The density  $P$  on a finite space  $X$  is a definition in  $X$  in the language of fuzzy mathematics of the property of limiting. From a formal point of view,  $P$  is a fuzzy structure on the direct product  $2^X \times X$ , monotonic in the first argument (1).

Fixing a level  $\alpha \in [0, 1]$  for  $P$ , we can define a closure operator on  $2^X$  and, through it, the usual topology  $\tau_\alpha$  on  $X$ . Thus, an increasing family of topologies  $\{\tau_\alpha, \alpha \in [0, 1]\}$ , arises on  $X$ , starting at zero with a trivially inseparable minimal topology of sticking together points and ending with a trivially separable maximal topology of all subsets.

Each density on the space  $X$  gives its own view of it and its own program of its study. It is a concept that does not lie in classical mathematics.

In addition to the density  $S$  Section 2.2.2, nontrivial densities always exist in a finite metric space. The description of the most important densities and their significance for data analysis are given in [5,11].

The study of the space  $X$  with the help of the density  $P$  defined on it began with the study of perfect sets in the topology  $\tau_\alpha$ . The explanation for this is as follows: setting  $P$  on  $X$  defines a clustering in  $X$ :  $K$  is a cluster ( $P$ -cluster) in  $X$  if it consists of exactly all points that are  $P$ -limit to it. If  $\alpha$  is the limit threshold, then the formal expression of what has been said coincides with  $\alpha$ -perfection:  $K = \{x \in X : P_K(x) \geq \alpha\}$ .

The DPS-algorithm builds all such clusters, but there are too many of them for meaningful cluster analysis in  $X$ : first of all, we need “connected components” of the maximal  $\alpha$ -perfect subset  $X(\alpha)$  in  $X$ .

All this exists when  $X$  is FMS and the density  $P$  is local on it. It is in this situation that the DPS-algorithm operates. It depends on the design of the density  $P$ , its level  $\alpha$  and the localization radius  $r$ :  $DPS = DPS(P, \alpha, r)$ . DPS operates in two stages: first, it cuts out



a subset  $X(\alpha)$  in  $X$  (first stage), and then splits it into  $r$ -connected components, which it considers to be DPS-clusters in  $X$  (second stage).

Despite generally good results and efficient applications [4,5,11–13], the DPS-algorithm in its current version has drawbacks: at the first stage, the cutting  $X \Rightarrow X(\alpha)$  is not always thorough and of high quality; at the second stage, the partitioning of  $X(\alpha)$  into  $r$ -connectivity components is too small and detailed due to the locality of the radius  $r$ . A proper partition of  $X(\alpha)$  must come from the global view of  $X(\alpha)$  induced by the entire space  $X$ .

In the present paper, we have made attempts to correct or mitigate these drawbacks: our answers were at the first stage—the iterative cutting scheme  $DPS^i(X)$ , at the second—its third stage  $DPS_f(X)$  of connecting the components of the  $r$ -connectivity.

The iterations allowed the DPS-algorithm at the first stage to achieve the completeness of its  $DPS(X)$  result, while maintaining a high level of extremality of cutting.

At the moment, the result  $DPS_C(X)$  of the second stage of the DPS-algorithm is considered its final result. The additional connections of components from  $DPS_C(X)$  proposed in the paper occur in two stages. On the first of them, which has the necessary character,  $d$ -clusters are searched in  $DPS_C(X)$ . According to the results of the article, their search is constructive.  $d$ -clustering is a necessary condition for connecting components from  $DPS_C(X)$  into a single whole. The set of  $d$ -clusters  $C(DPS_C(X), d)$  serves as the basis for the second stage, which is already sufficient. The main thing here is the expert: his criteria form the final choice of  $DPS_f(X)$  in  $C(DPS_C(X), d)$  and at the same time the final partition of  $DPS_C(X)$ . There are various formal variants of this choice in DMA. The simplest one of them is given in Section 4.3.1.

DPS-series algorithms are actively applied in many geological and geophysical studies (analysis of seismic catalogs, search for signals on geophysical records, in the problem of radioactive waste disposal, etc.) [1,4,5,7,11–13]. It seems that the new version of DPS developed in the present paper will make it possible to improve this application.

The addition of the third stage to the DPS according to the authors makes its architecture sufficient. Further development of DPS-algorithms should take place through their parameters: new constructions of densities and their connection with fuzzy logic will give (and already give [5]) the possibility of deep study of finite metric spaces. Furthermore, the following circumstance is fundamental: in contrast to the Euclidean space, the FMS  $X$  is local at each of its points  $x$ ; as a rule, it is arranged differently; therefore, the parameters  $\alpha$ ,  $r$  of the DPS-algorithm, which have a local nature, must depend on  $x$ :  $r = r(x)$ ,  $\alpha = \alpha(x)$ . The iterative carving scheme only softens the matter.

**Author Contributions:** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by S.R.B., B.A.D., B.V.D., D.A.K. and M.O.O. The first draft of the manuscript was written by S.M.A., and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was conducted in the framework of budgetary funding of the Geophysical Center of RAS, adopted by the Ministry of Science and Higher Education of the Russian Federation.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Agayan, S.M.; Bogoutdinov, S.R.; Krasnoperov, R.I. Short introduction into DMA. *Russ. J. Earth Sci.* **2018**, *18*, 1–10.
2. Agayan, S.M.; Bogoutdinov, S.R.; Dobrovolsky, M.N. On one algorithm for searching the dense areas and its geophysical applications. In Proceedings of the Materials of 15th Russian National Workshop “Mathematical Methods of Pattern Recognition, MMRO-15”, Petrozavodsk, Russia, 11–17 September 2011; pp. 543–546.
3. Agayan, S.M.; Bogoutdinov, S.R.; Dobrovolsky, M.N. Discrete Perfect Sets and Their Application in Cluster Analysis. *Cybern. Syst. Anal.* **2014**, *50*, 176–190. [[CrossRef](#)]

4. Agayan, S.M.; Tatarinov, V.N.; Gvishiani, A.D.; Bogoutdinov, S.R.; Belov, I.O. FDPS algorithm in stability assessment of the Earth's crust structural tectonic blocks. *Russ. J. Earth Sci.* **2020**, *20*, 1–14. [[CrossRef](#)]
5. Agayan, S.; Bogoutdinov, S.; Kamaev, D.; Kaftan, V.; Osipov, M.; Tatarinov, V. Theoretical Framework for Determination of Linear Structures in Multidimensional Geodynamic Data Arrays. *Appl. Sci.* **2021**, *11*, 11606. [[CrossRef](#)]
6. Dzeboev, B.A.; Gvishiani, A.D.; Agayan, S.M.; Belov, I.O.; Karapetyan, J.K.; Dzeranov, B.V.; Barykina, Y.V. System-Analytical Method of Earthquake-Prone Areas Recognition. *Appl. Sci.* **2021**, *11*, 7972. [[CrossRef](#)]
7. Agayan, S.M.; Losev, I.V.; Belov, I.O.; Tatarinov, V.N.; Manevich, A.I.; Pasishnichenko, M.A. Dynamic Activity Index for Feature Engineering of Geodynamic Data for Safe Underground Isolation of High-Level Radioactive Waste. *Appl. Sci.* **2022**, *12*, 2010. [[CrossRef](#)]
8. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, OR, USA, 2–4 August 1996; AAAI Press: Palo Alto, CA, USA, 1996; pp. 26–231.
9. Ankerst, M.; Breunig, M.; Kriegel, H.-P.; Sander, J. OPTICS: Ordering Points To Identify the Clustering Structure. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Philadelphia PA, USA, 1–3 June 1999; ACM Press: New York, NY, USA, 1999; pp. 49–60.
10. Jambu, M. *Hierarchical Cluster Analysis and Correspondences*; Finansy i Statistika: Moscow, Russia, 1988.
11. Agayan, S.; Bogoutdinov, S.; Soloviev, A.; Sidorov, R. The Study of Time Series Using the DMA Methods and Geophysical Applications. *Data Sci. J.* **2016**, *15*, 16. [[CrossRef](#)]
12. Dzeboev, B.A.; Karapetyan, J.K.; Aronov, G.A.; Dzeranov, B.V.; Kudin, D.V.; Karapetyan, R.K.; Vavilin, E.V. FCAZ-recognition based on declustered earthquake catalogs. *Russ. J. Earth Sci.* **2020**, *20*, 1–9. [[CrossRef](#)]
13. Gvishiani, A.D.; Agayan, S.M.; Losev, I.V.; Tatarinov, V.N. Geodynamic hazard assessment of a structural block holding an underground radioactive waste disposal facility. *Min. Inf. Anal. Bull.* **2021**, *12*, 5–18. [[CrossRef](#)]