PROCEEDINGS OF SCIENCE

Comparision of Fractal Analysis Methods for Fractal and Random Data Sets

T.G. Dedovich*
Joint Institute for Nuclear Research, Dubna, Russia
E-mail: tdedovich@yandex.ru

M.V. Tokarev
Joint Institute for Nuclear research, Dubna, Russia
E-mail: tokarev@jinr.ru

Properties of self-similarity and fractality in processes of interactions of hadrons and nuclei at high energies are discussed. Different methods of Fractal Analysis (the Box Counting (*BC*), P-adic Coverage (*PaC*), System of the Equations of P-adic Coverage (*SePaC*) methods) are described. The two-step procedure for *PaC* and *SePaC* methods of fractal reconstruction is justified. The search procedure of optimal values of parameters for *BC*, *PaC* and *SePaC* methods is developed. Comparison of the fractal analysis methods for fractal and random data sets is performed. The two-step procedure of fractal analysis for *SePaC* method is shown to has advantages before other ones.

XXII International Baldin Seminar on High Energy Physics Problems,
September 15-20, 2014
JINR, Dubna, Russia

*Speaker.

PoS(Baldin ISHEPP XXII)038

1. Introduction

The study of experimental data which demonstrate a property of self-similarity of hadron interactions at high energy is traditionally of considerable interest. One of such methods is known as z-scaling approach [1]. It was developed in [2, 3, 4, 5]. The method was applied for analysis of numerous experimental data on inclusive spectra of hadrons, direct photons and jets produced in $pp, p\bar{p}, pA$ and AA collisions over a wide kinematic range. Authors have concluded that the structure of hadrons, mechanism of constituent interactions and hadronization process reveal self-similar properties over a wide scale range. In this approach an inclusive particle spectrum is described in terms of dimensionless function ψ depending on single variable z. Dependence of scaling function ψ on z was found to be independent of the collision energy, and angle of produced particle. It is also described by power law $\psi(z) \sim z^{-\beta}$ at high z. The value of slope parameter β is independent of kinematic variables over a wide z-range. It confirms self-similarity of hadron production at various scales. This power law characterizing self-similarity of particle production is typical for fractals.

A fractal is usually determined as an object having a self-similar structure whose fractal dimension D_F is larger than topological one D_T. Topological dimensions of a point, line and plane are equal to zero, one and two, respectively. The value of D_F which provides a finite limit of the following expression:

$$\lim_{\delta \to 0} \sum_{i=1}^{N} l_i^{D_F} = const \tag{1.1}$$

is the fractal dimension. Here $N(\delta)$ is the number of probes covering the object with size l and δ is a maximal size of probes.

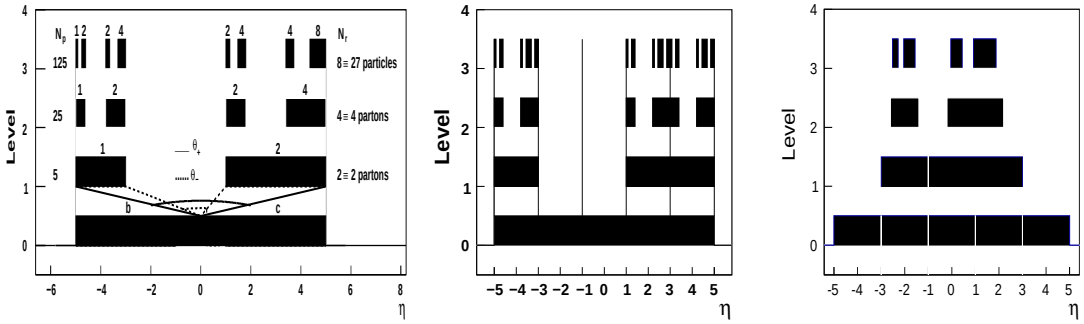
Fractal dimensions δ_1, δ_2 and ε_a, ε_b are quantitative characteristics of fractal structures of colliding objects and fragmentation processes in z-scaling approach [6, 7]. Discontinuity of these dimensions is assumed to be an indication of phase transition which is related to the change of physical subprocesses underlying inclusive particles production. Fractal dimensions have been determined from data analysis of inclusive cross sections and then verified in analysis of other data sets. We plan in future to determine dimensions by using fractal analysis. In present paper we compare different methods of fractal analysis and give estimations of their efficiency and impurity.

2. A parton shower and hadronization as fractal

The hadrons produced in inelastic interactions are usually considered as a set of points of the three-dimensional phase-space (p_T, η, ϕ). Here p_T is a transverse momentum, η and ϕ are the pseudorapidity and azimuthal angle, respectively. The distribution of these points in the phase-space is non-uniformed. It is determined by the processes of particle production. The distribution is assumed to be a fractal and characterized by fractal dimensions, which depend on interaction dynamics. Thus, we conclude that determination of fractal dimensions is an important task to understand hadron and nucleus interactions at a constituent level.

Below we describe the scenario of process of parton shower and hadronization which leads to formation of a fractal. Figure 1(a) shows a schematic diagram of the process. The parton a outgoing from a hard process branches into partons b and c. The admissible opening angles Θ_+, Θ_- are shown by the solid and dashed lines, respectively. The permissible ranges of the pseudorapidity

space are shown by black rectangles. The pseudorapidity η is defined by $\eta = -0.5ln(tg(\Theta/2))$. The number of parts ($N_p$) in the general partition of space is shown on the left side of Fig.1(a). As an example, we show the space at the first level divided into five parts. The number of the permissible ranges (N_r) and their conformity to the number of partons or particles is shown on the right side of Fig.1(a). We see that two ranges at the first level are permissible. The left consists of one part and the right - of two parts. It is assumed that the partons can be to each of these ranges. The right range is considered as the uniformed object consisting of dependent parts. We assume that branching and hadronization keep a specified spatial structure at different levels. The described fractal is named a fractal with a dependent partition.



(a) (b) (c)

Figure 1: The parton shower and hadronization as a fractal process. The permissible ranges of η-space at different levels. Fractals with dependent (a), independent (b) and combine (c) partition.

The fractal dimension is defined as a value D_F which provides the finite limit (1.1). The equation for the fractal described above is written in the form of $1/5^{D_F} + 2/5^{D_F} = 1$. The solution is equal to $D_F \approx 0.5639...$ The box dimension D_b is determined as a solution of the following equation:

$$D_b = \lim_{\delta \to 0} \frac{lnN(\delta)}{ln(\delta)}. \tag{2.1}$$

It is equal to $D_b = ln3/ln5 \approx 0.6828...$ Note that the fractal D_F and box D_b dimensions for this type of fractals are not equal to each other.

In the present analysis we use three types of fractals(see Fig. 1a). The first one includes fractals with dependent partition. The permissible range consists of arbitrary number of parts which are not contactable. It is divided as a uniformed object. The second type includes fractals with independent partition (see Fig. 1b). In this case the permissible ranges consist of one part. The ranges are independently divided. The third type are fractals with combined partition (see Fig. 1c). The permissible ranges consist of arbitrary number of parts which are contactable. The parts belonging to the same range are divided dependently. The different ranges are divided independently. We would like to note that for fractals with independent partition the fractal and box dimensions are equal to each other. These dimensions are different for fractals with dependent and combine partition.

3. Methods of Fractal Analysis

In the section we consider methods of definition of fractal dimension. One of them is the Box counting (*BC*) method [12]. It is widely used for fractal analysis. The other ones are the new methods - *P*-adic Coverages (*PaC*) and System of the Equations of *P*-adic Coverages (*SePaC*) methods which are recently proposed in [13] and [14], respectively.

3.1 Box Counting (*BC*) and *P*-adic Coverages (*PaC*) Methods

The *BC* and *PaC* methods are based on the definition of the box dimension D_b. Below the general features and differences of *BC* and *PaC* methods are described. Both ones include the following steps:
1. Read out data ($\{X = \eta, ...\}$ of particles in event).
2. Construction of *P*-adic coverages: Each coverage is a set of distributions of variable X. The number of bins M_i in distributions of the set is changed as a degree of base P $(M_i = (P)^i)$.
BC: $P = 2$ (as usually), *PaC*: $P = 2, ..., P_{Max}$.
3. The counting number of non-zero bins $N(lev,P)$: saturation condition $N(lev,P) = N(lev+1,P)$ defines the number of levels $N_{lev} = lev$.
4. Basic *PaC* method includes and modified *PaC* method does not include the condition $N(lev,P) = N(1,P)^{lev}$.
5. Finding the parameter D_F of the power function $N \sim M^{D_F}$ and corresponding χ^2 of the linear approximation of this function in a double-log scale for each *P*-adic coverage.
6. Accuracy condition: if $\chi^2 < \chi^2_{lim}$ then the set of particles is considered to be a fractal (P and $D_F(P)$).

The *BC* and *PaC* methods determine the box dimension D_b. They have two parameters. The first one is the base of coverage $P = 2$ for *BC* methods and maximal base P_{Max} for *PaC* method. The second parameter is χ^2_{lim} for both methods. It determines whether the set of points is a fractal or not. In present analysis, we used the basic and modified *PaC* methods.

3.2 System of Equations of *P*-adic Coverages (*SePaC*) Method

Here we briefly describe *SePaC* method [14]. It consists of the following steps:
1. Read out data ($\{X = \eta, ...\}$ of particles in event).
2. Construction of *P*-adic coverages: $P = 3, ..P_{Max}$.
3. Counting number of non-zero bins $N(lev,P)$: saturation condition $N(lev,P) = N(lev+1,P)$ defines the number of levels $N_{lev} = lev$.
4. Basic *SePaC* method includes and modified *SePaC* method does not include the condition $N(lev,P) = N(1,P)^{lev}$.
5. Analysis of the system of the equations to verify the hypothesis on independent or dependent partition:
a) Construction of the system of the equations for all levels:

$$\sum_{i=1}^{N_{lev}} (d_{lev})^{D_F^{lev}} = 1. \tag{3.1}$$

N_{lev} is the number of levels, and d_{lev} is the length of permissible ranges for each level.
b) Finding the solution D_F^{lev} of equations for each level. The dichotomy method is used.
c) Determination of the averaged value $< D_F^{lev} >$ and deviation ΔD_F^{lev}

$$< D_F^{lev} >= \sum_{lev=1}^{N_{lev}} (D_F^{lev})/N_{lev} \qquad \Delta D_F^{lev} = | < D_F^{lev} > -D_F^{lev}|. \tag{3.2}$$

d) Accuracy condition: if $\Delta D_F^{lev} < Dev$, then the set of particles is considered to be a fractal (P is the base and $D_F = D_F^{lev}$ is the fractal dimension.

We would like to note that *SePaC* method determines the fractal dimension D_F. It has two parameters. The first one is maximal base P_{Max} and second one is deviation Dev. The last parameter determines whether the set of points is a fractal or not. In the analysis, we used the basic and modified *SePaC* methods.

3.3 Two-step Procedure of Fractal Analysis

The results of our analysis presented in [14] showed that it is preferable to use different versions of *PaC* and *SePaC* methods for analysis of different types of fractals. Therefore we suggest a two-step procedure of fractal analysis. It takes into account peculiarities of the proposed methods. The procedure includes the following steps:
1. Analysis of the data set by the basic method
- determination of the optimal values of parameters *Par*
- analysis of the data set with the found parameters
- determination of characteristics of reconstructed fractals
- selection of unreconstructed fractals.
2. Analysis of unreconstructed fractals by the modified method.

3.4 Search Procedure of Parameters

The *BC*, *PaC* and *SePaC* methods have parameters $Par = \{\chi^2_{lim}, P_{Max}, Dev\}$. They should be determined. We have developed the search procedure of optimal values of these parameters. It allows us to determine fractal dimension D_F, number of levels N_{lev}, and base P with highest efficiency. The search procedure of parameters *Par* consists of the following steps:
1. Construction of distribution on variable $V = \{D_F, N_{lev}, P\}$ for the different *Par*.
2. Calculation of the distribution ΔD_V for different *Par*:

$$\Delta D_V = \sum_{i=1}^{N_{bin}} |a_i - b_i|. \tag{3.3}$$

Here a_i and b_i are bin contents for the adjacent distribution (Par_j and Par_{j+1}).
3. Calculation of the extended function ΔD_{Ext} for various values of *Par*

$$\Delta D_{Ext} = \Delta D_{D_F} + \Delta D_{N_{lev}} + \Delta D_P. \tag{3.4}$$

4. Choice of the value *Par* based on analysis of function $\Delta D_{Ext}(Par)$

Tables 1 and 2 show the correspondence of number and value of parameters χ^2_{lim} and Dev used in the present analysis.

Table 1: Correspondence of $N_{\chi^2_{lim}}$ and χ^2_{lim}.

$N_{\chi^2_{lim}}$	1	2	3	4	5	6	7	8	9
χ^2_{lim}	10^{-13}	10^{-12}	10^{-11}	10^{-10}	10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}
$N_{\chi^2_{lim}}$	10	11	12	13	14	15	16	17	18
χ^2_{lim}	10^{-4}	10^{-3}	0.01	0.02	0.03	0.04	0.05	0.06	0.07
$N_{\chi^2_{lim}}$	19	20	21	22	23	24	25	26	27
χ^2_{lim}	0.08	0.09	0.1	0.2	0.3	0.4	0.5	0.6	0.7
$N_{\chi^2_{lim}}$	28	29	30	31	32	33	34	35	36
χ^2_{lim}	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6
$N_{\chi^2_{lim}}$	37	38	39	40	41	42	43	44	45
χ^2_{lim}	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5

Table 2: The correspondence of N_{Dev} and Dev.

N_{Dev}	1	2	3	4	5	6	7	8
Dev	10^{-6}	10^{-5}	10^{-4}	$2\cdot10^{-4}$	$5\cdot10^{-4}$	10^{-3}	$2\cdot10^{-3}$	$5\cdot10^{-3}$
N_{Dev}	9	10	11	12	13	14	15	16
Dev	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
N_{Dev}	17	18	19	20	21	22	23	24
Dev	0.09	0.1	0.2	0.3	0.4	0.5	0.6	0.7

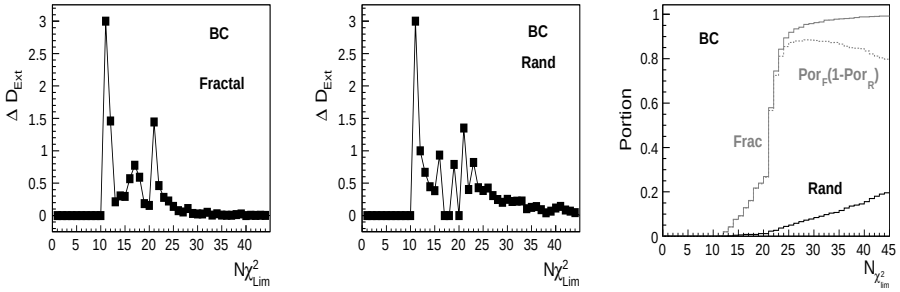
4. Comparison of Fractal Analysis Methods

The used data set consists of 1857 fractals and the same number of random sets. Fractals are generated by independent, dependent and combined partition. The multiplicity distributions for random and fractal data sets are taken to be equal each other. In the analysis we compare the efficiency and impurities of the methods used. Efficiency is define as a portion of the reconstruction fractal Por_{Frac} and impurities - as a portion of the random set reconstructed as a fractal Por_{Rand}.

4.1 *BC* Method

Here we analyze data sets by *BC* method. The fig.2(a,b) shows the dependence of extended function ΔD_{Ext} on $N_{\chi^2_{lim}}$ for the fractal and random data sets. We see that the shapes for both functions are similar. The dependence of the portion of reconstructed fractals Por_{Frac} and random data sets found to be as a fractals Por_{Rand} and function $Por_{Frac}(1-Por_{Rand})$ on $N_{\chi^2_{lim}}$ are presented on fig.2(c).

The choice of $N_{\chi^2_{lim}}$ as being the minimal value on the second plateau of ΔD_{Ext} for fractals (fig.2(a)) corresponds to the maximal portion of reconstructed fractals (fig.2(c)). Note, that shapes of distributions Por_{Frac} and Por_{Rand} are different (fig.2(c)). The shape of $Por_{Frac}(1-Por_{Rand})$

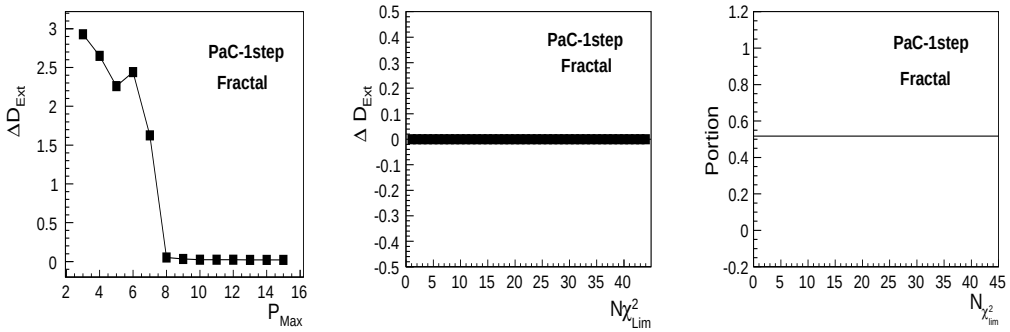


(a) (b) (c)

Figure 2: The dependence of extended function ΔD_{Ext} on $N_{\chi^2_{lim}}$ for fractal (a) and random (b) data sets. The dependences of portions of reconstructed fractals Por_{Frac}, random data sets found to be as fractals Por_{Rand} and function $Por_{Frac}(1-Por_{Rand})$ (c) on $N_{\chi^2_{lim}}$.

(fig.2(c)) allows us to determine the acceptable range of $N_{\chi^2_{lim}}$ at which efficiency is maximal and impurity is minimal. The found value of $N_{\chi^2_{lim}}$ for fractals corresponds to the acceptable range. This is confirmation of the proposed procedure of choice of $N_{\chi^2_{lim}}$.

4.2 Two-step Procedure of *PaC* Method



(a) (b) (c)

Figure 3: First step of *PaC* method. The dependence of extended function ΔD_{Ext} on P_{Max} (a) and $N_{\chi^2_{lim}}$ (b) for fractals. The dependence of the portion of reconstructed fractals Por_{Frac} on $N_{\chi^2_{lim}}$ (c).

In the section we analyze the first step of *PaC* method. The fig.3(a,b) shows the dependence of extended function ΔD_{Ext} on P_{Max} (a) and $N_{\chi^2_{lim}}$ (b) for fractals and the dependence of the portion of reconstructed fractals Por_{Frac} on $N_{\chi^2_{lim}}$ (c). We found that the choice of P_{Max} as being the minimal value on the plateau of extended function ΔD_{Ext} corresponds to the maximal portion of reconstructed fractals Por_{Frac}. As seen from fig.3(b,c) the shapes of extended function ΔD_{Ext} and the portion of fractals Por_{Frac} are independent of $N_{\chi^2_{lim}}$. Therefore the choice of $N_{\chi^2_{lim}}$ is arbitrary. In the analysis we take it to be 10^{-2}. We found that the first step restores 52% of fractals and does not restore the random set as a fractals.

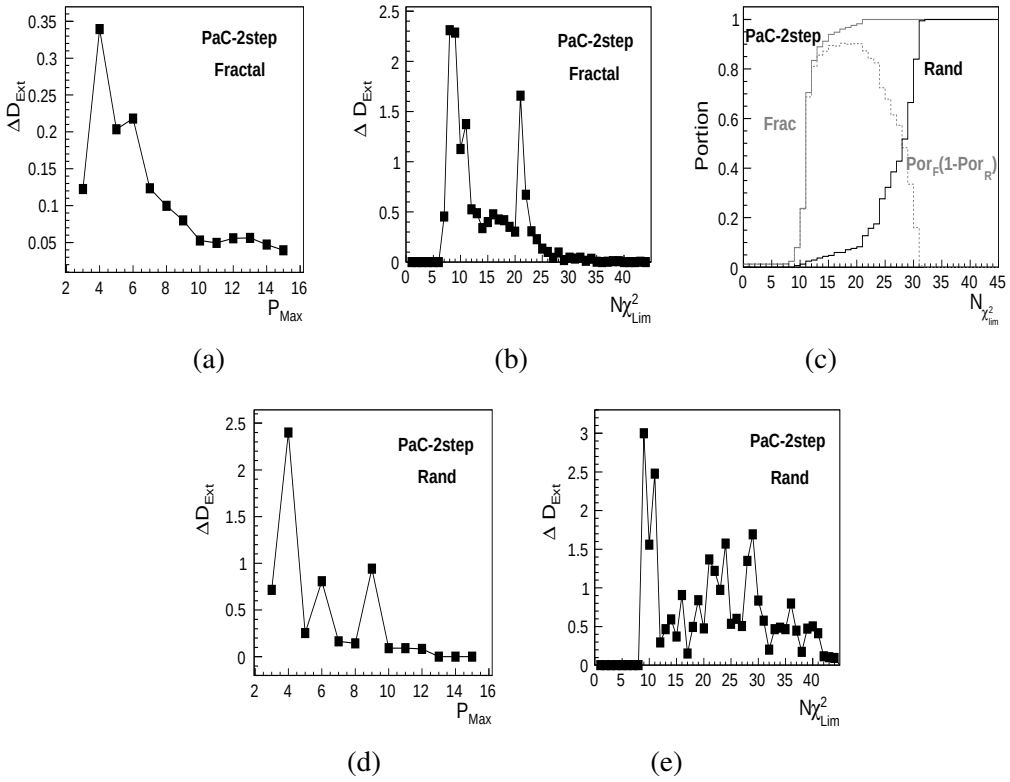
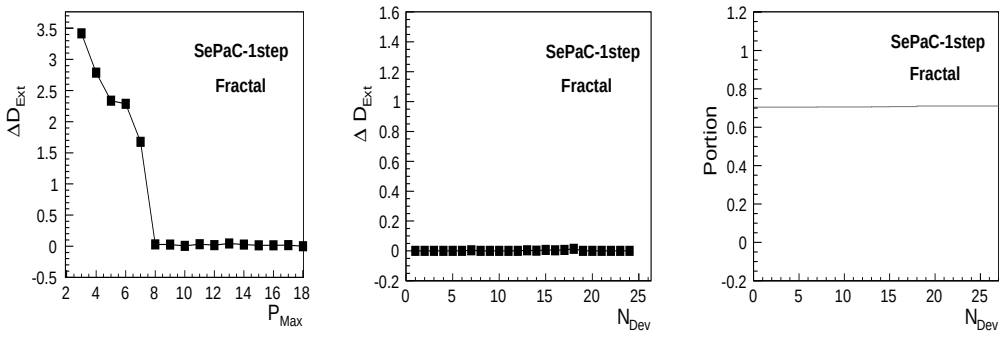


Figure 4: Second step of *PaC* method. The dependence of extended function ΔD_{Ext} on P_{Max} and $N_{\chi^2_{lim}}$ for fractals (a,b) and random data sets (d,e). The dependences of portions of reconstructed fractals Por_{Frac}, random data sets found to be as fractals Por_{Rand} and function $Por_{Frac}(1-Por_{Rand})$ on $N_{\chi^2_{lim}}$ (c).

Here we analyze fractal and random data sets by the second step. The fig.4(a,b) shows the dependence of extended function ΔD_{Ext} on P_{Max} and $N_{\chi^2_{lim}}$ for fractals. The dependences of portions of reconstructed fractals Por_{Frac}, random data sets found to be as fractals Por_{Rand} and function $Por_{Frac}(1-Por_{Rand})$ on $N_{\chi^2_{lim}}$ are plotted on fig. 4(c). We found that the choice of P_{Max} taken as the minimal value on the plateau of extended function ΔD_{Ext} (fig.4(a)) and $N_{\chi^2_{lim}}$ taken as the value of second peak on extended function ΔD_{Ext} fig.4(b) correspond to the maximal portion of reconstructed fractals Por_{Frac} (fig.4(c)). We choose $P_{Max} = 10$ and $N_{\chi^2_{lim}} = 22$. The dependences of extended function ΔD_{Ext} on P_{Max} and $N_{\chi^2_{lim}}$ for random data sets is shown on fig.4(d,e). The function $\Delta D_{Ext}(P_{Max})$ for fractal and random sets has a plateau (fig.4(a,d). The shapes of $\Delta D_{Ext}(N_{\chi^2_{lim}})$ for fractal and random sets are different. We found, that absence of two peaks of function $\Delta D_{Ext}(N_{\chi^2_{lim}})$ indicates that the data sets are not fractals. Note, that shapes of distributions Por_{Frac} and Por_{Rand} are different (fig.4(c)). The shape of $Por_{Frac}(1-Por_{Rand})$ (fig.4(c)) allows us to determine the acceptable range of $N_{\chi^2_{lim}}$ at which efficiency is maximal and impurity is minimal. The found value of $N_{\chi^2_{lim}}$ for fractals corresponds to the acceptable range. Thus the proposed procedure confirms the choice of $N_{\chi^2_{lim}}$.



(a) (b) (c)

Figure 5: First step of *SePaC* method. The dependence of extended function ΔD_{Ext} on P_{Max} (a) and N_{Dev} (b) for fractals. The dependence of the portion of reconstructed fractals Por_{Frac} on N_{Dev} (c).

4.3 Two-Step Procedure of *SePaC* Method

In the section we analyze the first step of *SePaC* method. The fig.5(a,b) shows the dependence of extended function ΔD_{Ext} on P_{Max} (a) and N_{Dev} (b) for fractals and the dependence of the portion of reconstructed fractals Por_{Frac} on N_{Dev} (c). We found that the choice of P_{Max} as the minimal value on the plateau of extended function ΔD_{Ext} corresponds to the maximal portion of reconstructed fractals Por_{Frac}. As seen from fig.5(b,c) the shapes of extended function ΔD_{Ext} and the portion of fractals Por_{Frac} are independent of N_{Dev}. Therefore the choice of N_{Dev} is arbitrary. In the analysis we take it to be 10^{-2}. We found that the first step restores 70% of fractals and does not restore the random set as a fractals.

Here we analyze fractal and random data sets by the second step. The fig.6(a,b) shows the dependence of extended function ΔD_{Ext} on P_{Max} and N_{Dev} for fractals. The dependences of portions of reconstructed fractals Por_{Frac}, random data sets found to be as fractals Por_{Rand} and function $Por_{Frac}(1-Por_{Rand})$ on N_{Dev} are plotted on fig. 6(c). We found that the choice of P_{Max} taken as the minimal value on the plateau of extended function ΔD_{Ext} (fig.6(a)) and N_{Dev} taken as the value of second peak on extended function ΔD_{Ext} fig.6(b) corresponds to the maximal portion of reconstructed fractals Por_{Frac} (fig.6(c)). We choose $P_{Max} = 7$ and $N_{Dev} = 18$. The dependences of extended function ΔD_{Ext} on P_{Max} and N_{Dev} for random data sets are shown on fig.6(d,e). The function $\Delta D_{Ext}(P_{Max})$ for fractal and random sets has a plateau (fig.6(a,d). The shapes of $\Delta D_{Ext}(N_{Dev})$ for fractal and random sets are different. We found, that if first peak of $\Delta D_{Ext}(N_{Dev})$ is smeared then data sets are fractals. Note, that shapes of distributions Por_{Frac} and Por_{Rand} are different (fig.6(c)). The shape of $Por_{Frac}(1-Por_{Rand})$ (fig.6(c)) allows us to determine the acceptable range of N_{Dev} at which efficiency is maximal and impurity is minimal. The found value of N_{Dev} for fractals corresponds to the acceptable range. Thus the proposed procedure confirms the of choice of N_{Dev}.

4.4 Comparison of Methods

Here we compare the results of analysis of fractal and random data sets obtained by *BC*, two-step *PaC* and two-step *SePaC* methods. The efficiency (portion of reconstructed fractals Por_{Frac},

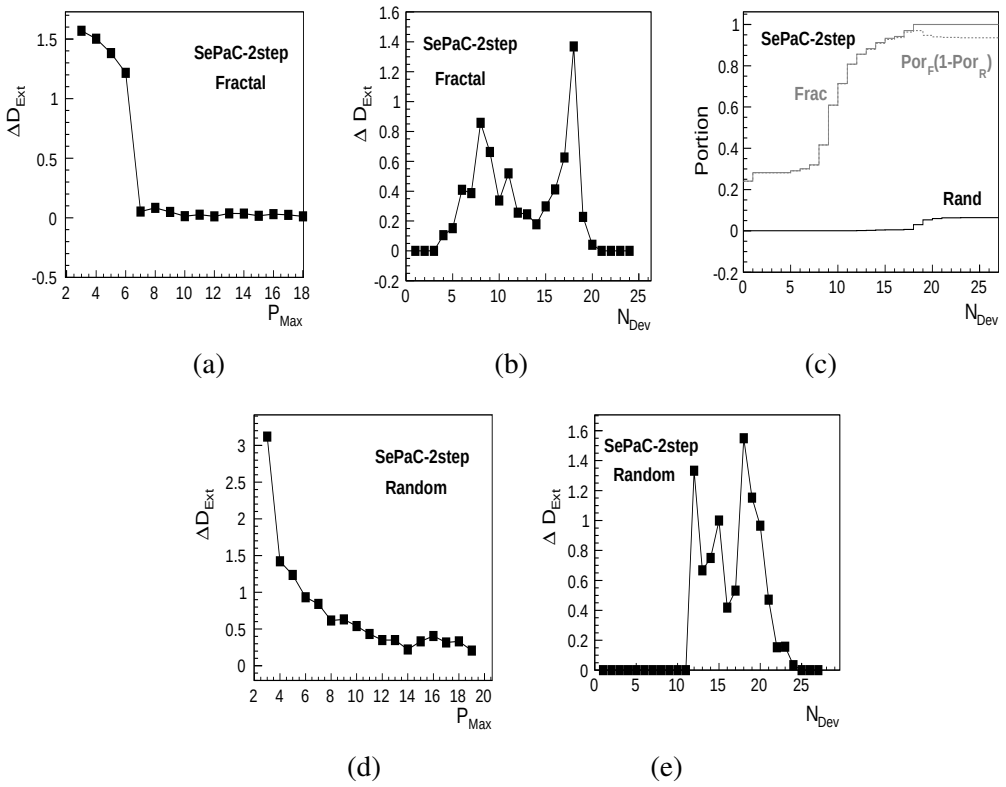


Figure 6: Second step of *SePaC* method. The dependence of extended function ΔD_{Ext} on P_{Max} and N_{Dev} for fractals (a,b) and random data sets (d,e). The dependences of portions of reconstructed fractals Por_{Frac}, random data sets found to be as fractals Por_{Rand} and function $Por_{Frac}(1-Por_{Rand})$ on N_{Dev} (c).

impurity (portion of random set reconstruction as a fractal Por_{Rand}) and function $Por_{Frac}(1-Por_{Rand})$ are presented in Table 3. We see that the two-step procedure of fractal analysis for *SePaC* method has advantages before other ones (100% event reconstruction and lowest impurity).

Table 3: Results of analysis of fractals and random sets obtained by *BC*, *PaC* and *SePaC* methods.

Method	Por_{Frac}	Por_{Rand}	$Por_{Frac}(1-Por_{Rand})$
BC	0.95	0.08	0.87
$two-step\ PaC$	1.0	0.12	0.88
$two-step\ SePaC$	1.0	0.03	0.97

5. Conclusions

The two-step procedure of fractal analysis for *PaC* and *SePaC* methods was suggested. The search procedure of optimal values of parameters for *BC*, *PaC* and *SePaC* methods was developed.

Comparison of the fractal analysis methods for fractals and random data sets were carried out. We found that the two-step procedure for *SePaC* method has advantage before other ones.

References

[1] I. Zborovský, Yu.A. Panebratsev, M.V. Tokarev, G. Škoro, *Phys. Rev.* **D 54**, 5548 (1996).

[2] M.V. Tokarev, T.G. Dedovich, *Int. J. Mod. Phys.* **A 15**, 3495 (2000).

[3] M.V. Tokarev, T.G. Dedovich, *Int. J. Mod. Phys. A* **15**, 3495 (2000).

[4] M.V. Tokarev, O.V. Rogachevski, T.G. Dedovich, *Nucl. Part. Phys.* **26**, 1671 (2000).

[5] M.V. Tokarev, T.G. Dedovich, and I.Zborovský, *Int. J. Mod. Phys. A* **27**, 1250115 (2012).

[6] I. Zborovský, M.V. Tokarev, *Phys. Rev.* **D 75**, 094008 (2007).

[7] I. Zborovský, M.V. Tokarev, *Int. J. Mod. Phys. A* **24**, 1417 (2009).

[8] M. Adamus et. al., *Phys. Lett.* **B 185**, 200 (1987).

[9] A. Bjalas, *Nucl. Phys.* **B 273**, 703 (1986).

[10] R. Hwa, *Phys. Rev.* **D 41**, 1456 (1990).

[11] T. Sjostrand et al., *Computer Physics Commun.* **135**, 238 (2001).

[12] B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).

[13] T.G. Dedovich, M.V. Tokarev, *Phys. Part. Nucl. Lett.* **8**, 521 (2011).

[14] T.G. Dedovich, M.V. Tokarev, *Phys. Part. Nucl. Lett.* **9**, 552 (2011).