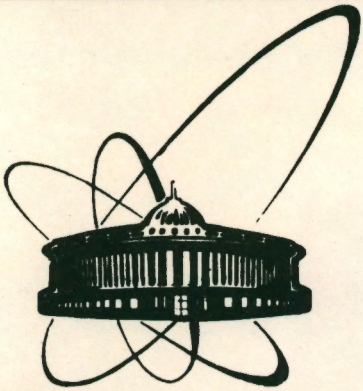


объединенный
институт
ядерных
исследований
дубна

E10-90-323

Yu.N.Pepyolyshev, W.Dzwinel*

CLUSTERING METHODS AND VISUALIZATION ALGORITHMS TO AID NUCLEAR REACTOR OPERATIVE DIAGNOSTICS

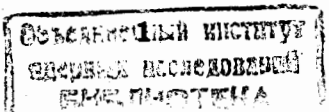
Submitted to International Symposium
on Balancing Automation and Human Action in Nuclear
Power Plants, Münich, FRG, 9-13 July

*Academy of Mining and Metallurgy,Institute
of Computer Science, Cracow, Poland

1990

1. Introduction

Noise and vibration analysis appears to be a valuable tool for non-destructive nuclear reactor condition monitoring [1,2]. It enables early fault detection and gives the opportunity to avoid abnormal operation and accidental situations. However, three dimensional (frequency, amplitude, time) noise patterns analysis gets very complicated. On the one hand, the bulk of information is useful for deep investigations, but on the other, the vast of them is unessential, disturbing and delaying the proper reaction of an operator to the reactor operation changes. To make the efficient and convenient tool for an operator the software system should enable the proper signal processing speed, its handling must be as easy as possible and the maximum information should be monitoring in the form of comprehensive patterns, relatively simple to analyze to perform suitable action. Pattern recognition techniques for noise analysis gives such an opportunity.



The pioneering proposition of clustering application for a reactor control are presented in [3,4] and still have been used and developed (e.g. [5,6]). The reliability of such an approach depends mainly on the proper choice of clustering algorithm.

In this paper some of pattern recognition methods for reactor operative diagnostics are recommended and the software system based on them is presented. In the last but not the least chapter, the information obtained by an operator is exemplified. The conclusions are reported at the end of the paper.

2.Diagnostics system description

The IBR-2 diagnostics system consists of two tightly coupled parts. The first one, fully automated, is responsible for data acquisition and preprocessing. It is represented by two minicomputers SM 1300.01 equipped with CAMAC-standard and IBM PC/386, connected by the RS-232, and software responsible for statistics computations and FFT transformation, which transforms the fluctuations of the power pulsed energy and positions vibrations of moving reflectors, i.e. the main (OPO) and the additional ones (DPO), into the corresponding spectral densities PSD and RMSs, respectively. As earlier investigations of random processes [7,8] taking place in the reactor core show that the character of spectral densities of energy pulse fluctuations is tightly coupled with the current reactor state, their combination may be considered as a noise image

of the reactor. The subsequent noise images, where an instant noise image is added to a host of images registered well before, determine the dynamic changes of the reactor state.

The second part of diagnostics system comprises software for data processing, which is based on clustering techniques and may be fully controlled by the user. Each of PSDs and RMSs, corresponding to the instant measurement of a respective reactor parameter may be represented as a point in M-dimensional Euclidean space (for our purposes M=257), where each of its subsequent coordinates is equal to the value of the spectral density amplitude for a subsequent discrete value of frequency. Clustering methods allow one to reveal the M-dimensional structures constituted by all the points, which are continually recorded. The tendencies of changes in various characteristics of clusters i.e. an abrupt change in the position of an instant cluster center, a change in its shape or initiation of a new cluster, give information on the reactor anomalous operation.

In Fig.1 the software system for nuclear reactor diagnostics is shown. The backbone of the system i.e. the clustering and mapping of patterns algorithms (the MNN and PRR, respectively) and the way of monitoring results and visualization, are described briefly in the next chapters.

3.Clustering method

A proper selection of the clustering method should be

the first stage to design a reliable tool for reactor diagnostics. The clustering techniques can be divided onto two main groups [9]: the hierarchical and non-hierarchical ones. The clustering algorithm ISODATA was the first applied to the reactor control [3,4] and is exploited up to now (e.g. [5,6]). It belongs to the family of non-hierarchical clustering algorithms. However, the algorithm ISODATA has several deficiencies which obliged one to look for a better method i.e.:

- it gives adequate results rather for static than dynamic systems and needs several parameters, which must be fitted to various operation conditions,
- considering the "rigid" clusters structure the new cluster initiation may be delayed,
- for more sophisticated clusters structures it gives wrong results [6,10].

The hierarchical algorithm, slightly modified in comparison with the mutual nearest neighborhood method (MNN) presented in [10], is recommended here and described briefly.

Let us assume that N spectral densities are considered and each of them can be represented by a point $\mathbf{x}_i$ in M-dimensional Euclidean space ω, i.e.:

$$\mathbf{x}_i \in \omega, \quad \dim \omega = M, \quad \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iM}), \tag{1}$$

and $i = 1,2,\ldots,N$.

Two kinds of distances between these points are defined. The first one:

$$s_{ij} = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \tag{2}$$

is the square of the classical Euclidean distance. Whereas, the second is defined as follows:

$$mnv_{ij} = L_i(j) + L_j(i), \tag{3}$$

where $L_i(j)$ means the position of the $\mathbf{x}_j$ point in the list of k_i neighbors of the $\mathbf{x}_i$ point, which is constructed in accordance with the increasing s_{im} distance, where m is the neighboring point number provided that:

$$s_{im} \leq R_{cut}. \tag{4}$$

As the number of the nearest neighbors k_i performing condition (4) is different for each point i, the maximum mnv distance is less or equal to:

$$mnv_{MAX} = \max_i\{k_i\} + \max_i\{\{k_i\} - \{\max_i\{k_i\}\}\} \tag{5}$$

and we can conclude that:

$$\forall i,j;\ 2 \leq mnv_{ij} \leq mnv_{MAX}. \tag{6}$$

The agglomerative procedure commences with the situation, where each point belongs to the one out of N separate and one-element clusters. The agglomeration of them is performed sequentially in accordance with the increasing mnv distance. For the $\{i,j\}$ pairs of points with equal mnv distances, an agglomerative procedure is performed, but now in accordance with the increasing s_{ij} distance. The process is continued until all the pairs with the maximum mnv values will be considered. The final clusters number K depends on R_{cut} (cut-off radius) only. To obtain the reasonable number of clusters, R_{cut} must be fitted once e.g. at the beginning stage of a diagnostics process, and may be corrected, if necessary. In the original MNN method

[10] the number of neighbors k for each point are equivalent and must be given *a priori*. Agglomerative process is finished, when all the points form the single cluster or a "rigid" number of clusters specified earlier. The procedure recommended here gives the "soft" stop condition and the number of clusters may vary in time. Moreover, R_{cut} determines the minimal separation between clusters.

In comparison with the computational efficiency order $O(N^2)$ for the algorithm presented in [10], the proposed version is the $O(N)$ order, assuming that for both of them the distances table $\mathbb{D}=[s_{ij}]$ is computed and stored only once, at the start of the system and is supplemented continually during operation. Using the *linked-lists* concept [11] additional memory savings are obtained. These properties enable the system implementation to use non-sophisticated, standard, table- standing personal computers, that alleviate the system handling and contributes to costs savings. Additionally, the method presented here enables one to recognize the sophisticated clusters structure e.g. branched, spherical, "bridged" and others [10], and lets one to follow the clusters formation process.

4.Clusters analysis and visualization algorithms

The information on the points distribution among clusters is very valuable. The single cluster comprises the points, which are close together or form a specific shape,

due to some "similarity measure" (the Euclidean distance in this paper). It may be assumed that all the points which belong to the same cluster refer to the same physical state of a reactor. They determine the global characteristics of a single cluster. If the subsequent point (which represents the subsequent measurement) falls to some of the existing clusters, the actual operational characteristics can be determined immediately. Else, i.e. a new cluster is formed, additional studies are needed. However, each new point is carrying a new information, which changes more or less a cluster shape and its characteristics. Such a structural change may appear for anomalous operation as well. So that, not only the cluster occupation but the information on the clusters structures and their characteristics appear to be valid. As the great dimension of ω space, their direct observation is impossible, that involves other methods for cluster analysis and visualization.

To determine the most informative frequencies, several principal components analysis algorithms are widely used [4]. Two of them can be invoked optionally in the diagnostics system presented here. They are based on a minimal entropy method and apart from frequencies analysis, they can be used for the decreasing of pattern dimension. An operator can obtain the information, which group of frequencies is responsible for clustering mainly i.e. which group reflects the reactor operation changes. The appearance of the new informative frequencies may signal the anomalous reactor operation and must be carefully examined. For each cluster the information concerning its representation and compactness is available. The distances

between the clusters centers, an average distance between spectra and a cluster center and average standard deviation of distances between spectra and a cluster center, can be computed. The time characteristic shows the cluster number each subsequent spectral density belongs to.

The cluster analysis entities presented above, gives rather poor information about the clusters structural changes. To observe them directly the mapping of patterns methods seem to be attractive. The PRR program, which constitutes the second part of the software system comprises three algorithms transforming the M-dimensional Euclidean space into two-dimensional. The first one represents the linear Karhunen -Loeve transformation. As is shown in [12], it rather poorly preserves a real clusters structure, especially for the sophisticated clusters shapes and the great space dimension. The non-linear mapping of patterns technique presented by Niemann in [12] is a better one.

Let us assume that ω' is an Euclidean space and:

$$\dim \omega' = M', \quad M' < M . \tag{7}$$

The points in M'-dimensional space can be represented as follows:

$$x'_i = (x'_{i1}, x'_{i2}, \ldots, x'_{iM'}). \tag{8}$$

To preserve approximately the M-dimensional clusters structure, (especially the interpoint distances) in the M'-dimensional space, the $\omega'_0 = \{x'_i; (i=1,2,\ldots,N)\}$ must perform the following condition [12]:

$$E(\omega,\omega'_0) = \min_{\omega'} E(\omega,\omega'), \tag{9}$$

where:

$$E(\omega,\omega') = \sum_{k<j} (s_{jk} - s'_{jk})^2 / \sum_{k<j} s^2_{jk} \tag{10}$$

and

$$\sum_{k<j} \equiv \sum_{j=2}^{N} \sum_{k=1}^{j-1} . \tag{11}$$

Niemann in [12] presents the iterative and deterministic technique for a minimum (9) finding. This method- slightly improved- is implemented in the diagnostics system. However, the Niemann algorithm is not able to find the global minimum of (10), but rather a local one. The simulated annealing method [13] which may be optionally invoked should enable one to find global minimum not only for (10) but for other criterion function as well. As this method is very time consuming, authors recommend to apply it rather for long time investigations. Up to now, the improved Niemann algorithm seems to be sufficient for IBR-2 operation monitoring.

5.Monitoring and visualization

The programs mentioned above are written in the FORTRAN 77 and OLYMPUS standard. The system has been developed on IBM PC/386 (387, 20MHz, 1MB) clone. Two work regimes are available. The first one- "suppressed"- is intended to monitoring reactor operation. It uses a set of

the fastest procedures of the software system, is fully automated and needs the limited operator action only. On the EGA display the main information is shown, which may be immediately analyzed by an operator. The second one -"extended"- is destined for long term investigations. Then, all the system possibilities can be used. Some of them are scheduled below.

- The row spectral densities and all the statistical information stored on the data base may be displayed.
- Two clustering methods may be optionally invoked i.e. the ISODATA algorithm and MNN method.
- The clusters agglomeration may be traced and interrupted at the optionally chosen stage.
- The optional "window" of frequencies and optional data sample may be considered.
- Two principal component analysis methods are available.
- Coordinates clustering is possible if necessary.
- Three mapping of patterns methods may be optionally invoked and optional combination of them is possible.
- Three dimensional patterns may be obtained.
- Apart from EGA display, the line printer and plotter may be used as graphical output.

The system has been developed at the IBR-2 reactor and its use in "suppressed" version is exemplified below. The subsequent PSDs are considered as images of the reactor states (with a new moving reflector) from the beginning of December 1987 to January 1990. Instant measurements are added to the host of images registered well before. If they are correct, the programs MNN and then PRR are invoked. The results of computations are stored on the hard disk. After

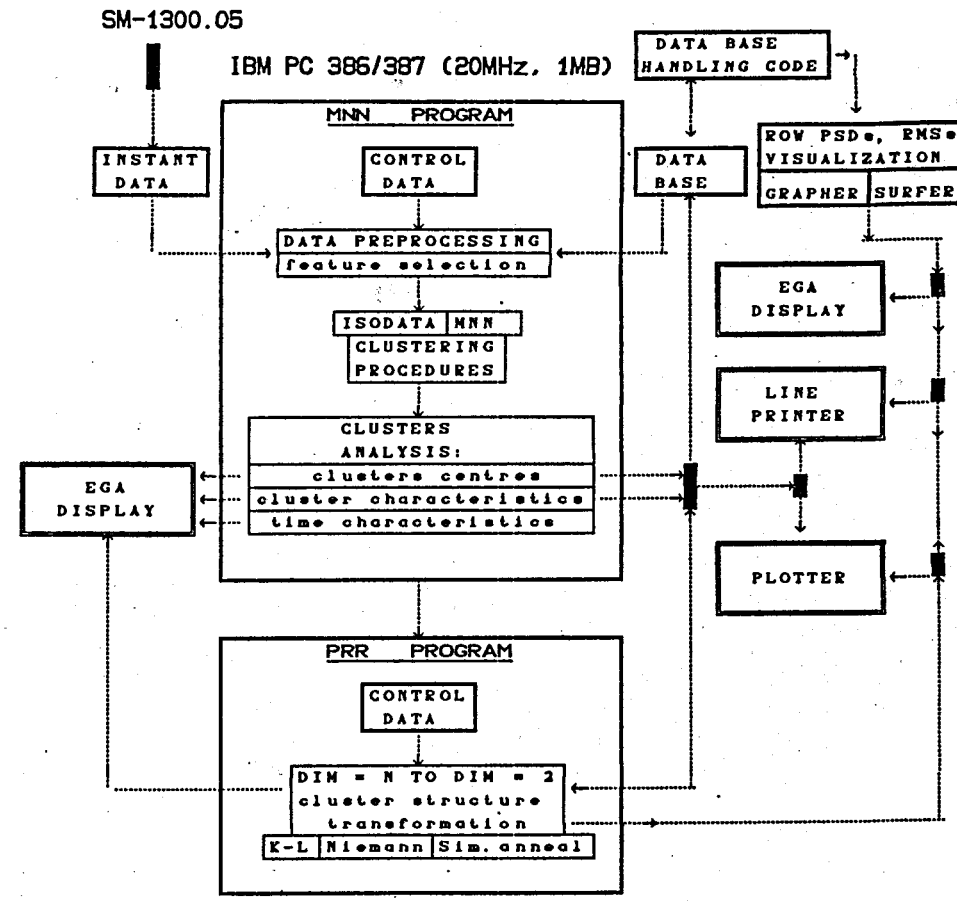


Fig.1 The diagnostics software system

the response on the first prompt, the global characteristics of the clusters structure are displayed (see Fig.2). The positive response on the second prompt gives the possibility to show the internal clusters characteristics. In Fig.3 the information on e.g. the fifth cluster is gathered. An operator can observe the clusters centers and detect their changes. In Fig.4 the result of clusters visualization using Niemann technique is

H I E R C L U S T - M N N
Version 1.2 July 1989

GLOBAL CHARACTERISTICS OF THE CLUSTERS STRUCTURE

Numbers of the most informative features

11 13 14 15 16 17 83 112 116 117 120 121 122 127 128 129 130 131 132 133
134 135 136 137 138 141 145 148 149 150 151 156 159 161 162 168 169 171 172
173 179 180

Current clusters number = 6
Average std.dev.between spectra and clusters centers = 0.106E+01
Average distance between spectra and clusters centers = 0.589E+01

THE TABLE OF DISTANCES BETWEEN CLUSTERS

cluster number	1	2	3	4	5	6
1	0.00E+00	0.15E+02	0.22E+02	0.22E+02	0.24E+02	0.22E+02
2	0.15E+02	0.00E+00	0.99E+01	0.93E+01	0.13E+02	0.92E+01
3	0.22E+02	0.99E+01	0.00E+00	0.95E+01	0.13E+02	0.99E+01
4	0.22E+02	0.93E+01	0.95E+01	0.00E+00	0.11E+02	0.88E+01
5	0.24E+02	0.13E+02	0.13E+02	0.11E+02	0.00E+00	0.86E+01
6	0.22E+02	0.92E+01	0.99E+01	0.88E+01	0.86E+01	0.00E+00

TIME CHARACTERISTIC

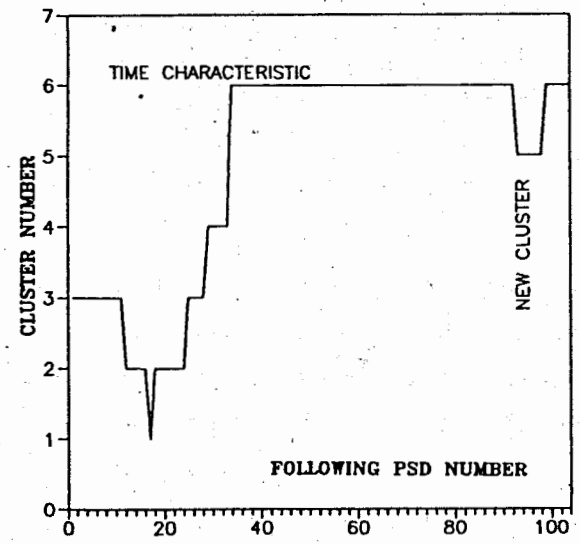


Fig.2 The information on global clusters structure obtained by an operator.

H I E R C L U S T - M N N
Version 1.2 July 1989

CHARACTERISTICS OF THE CLUSTER NUMBER: 5

Representation

Number of elements in cluster no. 5 is equal to: 6

Subseq. Number	Kind of process	Spectra name	Date	Sigma	Delta max.
98:	PSD	st.139	22 12 89	0.770E+01	0.573E+02
97:	PSD	st.138	22 12 89	0.769E+01	0.574E+02
95:	PSD	st.136	22 12 89	0.783E+01	0.596E+02
96:	PSD	st.137	22 12 89	0.765E+01	0.525E+02
94:	PSD	st.135	22 12 89	0.801E+01	0.609E+02
93:	PSD	st.134	22 12 89	0.803E+01	0.571E+02

CLUSTER AVERAGES: sigma.avg = 0.782E+01
deltamax.avg . = 0.575E+02
Average interspectra distance = 0.709E+01
The furthest spectra from cluster center: st.134(distance = 0.841E+01)
Average distance between spectra and cluster center = 0.495E+01
Stnd.Dev. of distances between spectra and cluster center = 0.188E+01

Distances between center of cluster no.5 and other ones

	1	2	3	4	5	6
5	0.24E+02	0.13E+02	0.13E+02	0.11E+02	0.00E+00	0.86E+01

CLUSTER CENTER

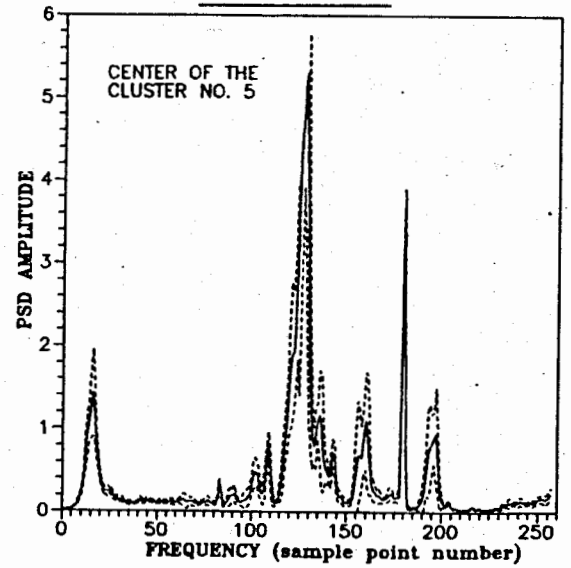


Fig.3 The information on internal characteristics of the fifth cluster.

P R E P R O C E S - P R R
Version 1.0 August 1989

CLUSTERS Visualization (Niemann preprocessing)

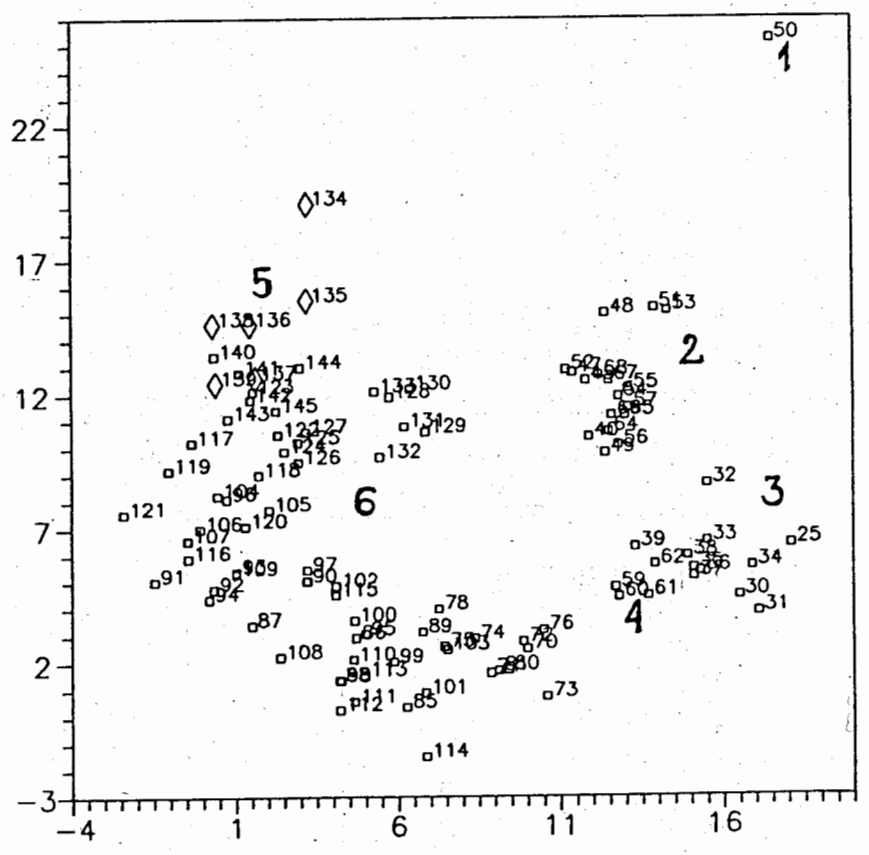


Fig.4 The clusters structure visualization. The two-dimensional pattern obtained on operator display. The points of fifth cluster are signed by rectangles. The substructure of the sixth cluster is visible. The cluster numbers are marked by hand.

presented. It reflects the clusters structure correctly, and gives additional information on the substructure of the sixth cluster, which is not detected by the clustering procedure. The tendencies of reactor operation changes are visualized and may be easily traced. The clusters structure obtained for the IBR-2 reactor refers to its normal operation. Each cluster represents various conditions of reactor operation, which depends on the actual position of moving reflectors and the whole construction vibration. The changes actually observed are fully controlled.

6.Conclusions

In this paper two techniques for nuclear reactor operative diagnostics and operation monitoring are recommended: the hierarchical clustering algorithm MNN and mapping of patterns methods. The software system, which is constructed on the base of these methods, performs the role of friendly human- machine interface. It is designed on the standard computer hardware and computer system independent. The great deal of attention was paid to enable the maximum time and memory computational efficiency. The maximal number of PSDs which can be considered, are limited by 1MB computer memory to 450. The time necessary for their processing on IBM PC/386 (387, 20MHz, 1MB) is equal to several minutes. The on-line, short-term surveillance similar to that presented in [2] would be possible for more sophisticated hardware available. For a reactor diagnostics, the anomalous effects, initiating accidental reactivity disturbances on a 10^{-7}-10^{-6} $\Delta K/K$ level are apparent.

In the near future the data acquisition will be oriented to IBM PC/386 hardware with the LABCARD system, and diagnostics software will be enriched by procedures for the automated operation control, enabling one to process greater number of measured parameters simultaneously, and making the system more user friendly.

Acknowledgements

The authors wish to thank Mrs. T.A. Filimonycheva for her assistance in manuscript preparation.

References

[1] Runkel, J., Stegemann, D., "Non-Destructive Condition Monitoring of Nuclear Power Plants by Use of Vibration Analysis", Non-Destructive Testing (Proc. 12th World Conference)
(Boogaard J., van Dijk G.M.,Ed.), Elsevier Sci. Publishers, Amsterdam, (1989) 628-633.

[2] Thie, J.A.,Damiano, B., Campbell, L.R., Inference of Physical Phenomena From FFTF Noise Analysis, Power Plant Dynamics Control and Testing Symposium, Knoxville, (15-17 May 1989).

[3] Gonzalez, R.C., Fry D.N., Kryter, R.C., Results in The Application of Pattern Recognition Methods to Nuclear Reactor
Core Component Surveillance, IEEE Trans. Nucl. Sci., 21 (1974) 750.

[4] Tou, J., Gonzalez, R.C., Pattern Recognition Principles, Addison-Wesley Publishing Company, (1974).

[5] Bargiel, M., Kitowski, J., Pepyolyshev, Yu.N., Commun. JINR Dubna, 13-88-345, (1988) (in russian).

[6] Weiss, F.P., Anwendung einiger Verfahren der Mustererkennung fur die Rauschdiagnostische Schadensfrherkennung an Kernkraftwerken, PhD thesis, Commun. ZfK Rossendorf, Dresden (1985) (in german)

[7] Kitowski, J., Pepyolyshev, Yu.N., Commun. JINR Dubna, P13-85-657, (1985), (in russian).

[8] Pepyolyshev, Yu.N., Commun. JINR Dubna, 13-85-795, (1985) (in russian).

[9] M.R.Anderberg, Cluster Analysis for Applications, Academic Press, (1973).

[10] Chidananda, K. Gowda, Krishna G., Agglomerative Clustering Using The Concept of Mutual Nearest Neighbourhood, Pattern Recognition, 10 2 (1978) 105.

[11] R.W.Hockney and J.W.Eastwood, Computer Simulations Using Particles, Mc Graw Hill, (1981).

[12] Niemann, H., Linear and Nonlinear Mapping of Patterns, Pattern Recognition, 12 2 (1980) 83.

[13] Kirkpatrick, S., Gelatt, C.D. Jr., Vecchi, M.P., Science, 220 (1983) 671.

Received by Publishing Department
on May 10, 1990.

Пепелышев Ю.Н., Дзвинель В. E10-90-323
Методы группирования и алгоритмы визуализации для оперативной диагностики ядерных реакторов

Описана программная реализация системы, выполняющей функции помощника оператора в диагностировании состояния ядерного реактора. Основу системы составляет анализ шумов основных реакторных параметров, например, мощности, расхода теплоносителя и т.д. Совокупность процессов накопления данных, их предварительной обработки, группирования и визуализации в сочетании с эвристической техникой анализа полученных результатов определяет способ применения системы. При разработке подобных систем рекомендуется использовать метод "ближайшего соседства" в технике распознавания образов и два нелинейных алгоритма визуализации образов. В дополнение к широко известным методам кластеризации ISODATA и преобразования Карунена-Лоэва указанные методы составляют основу программной реализации предлагаемой системы. Особое внимание уделено визуальному представлению данных и возможности легкого управления системой. Существует два режима ее работы: "расширенный" - для подробных и долговременных исследований и "сжатый" - для мониторирования работы реактора.Система создана и проверена на импульсном реакторе ОИЯИ ИБР-2.

Работа выполнена в Лаборатории нейтронной физики ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна 1990

Pepyolyshev Yu.N., Dzwinel W. E10-90-323
Clustering Methods and Visualization Algorithms to Aid Nuclear Reactor Operative Diagnostics

The software system developed plays the role of the aid to an operator for nuclear reactor diagnostics. The noise analysis of the reactor parameters such as power, temperature and coolant flow rate constitutes the basis of the system. Combination of data acquisition, data preprocessing, clustering and cluster visualization algorithms with heuristic techniques of results analysis, determine the way of its implementation. The pattern recognition methods, namely, the mutual nearest neighborhood clustering technique and two nonlinear algorithms for mapping the patterns are recommended here. They supplement each other, and together with the wide known ISODATA clustering method and Karhunen-Loeve transformation constitute the "backbone" of the diagnostics software system. The attention is paid to the proper visualization of results and easy system handling. Two regimes are available. The first one - "extended" - is recommended for a long term investigations and the second - "suppressed" - for the aid to the reactor operation monitoring. The system has been tested and developed at the JINR IBR-2 pulsed reactor.

The investigation has been performed at the Laboratory of Neutron Physics, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna 1990