УДК 621.391:519.237.8

РАЗРАБОТКА МЕТОДА КЛАСТЕРИЗАЦИИ ВЕЙВЛЕТ-ПРЕОБРАЗОВАННЫХ ГРАФИКОВ ЭЛЕКТРИЧЕСКИХ НАГРУЗОК

***А. В. Волошко, докт. техн. наук, Т. Н. Лутчин, магистр* (*НТУУ «КПИ»*)**

Розглянуто метод кластеризації інформаційних сигналів на основі адаптивного коефіцієнта відхилення відносно загальної вибірки значень. Розроблено спосіб визначення чітких границь між кластерами, які складаються з середніх потужностей вейвлет-коефіцієнтів по субсмугах розкладання.

Ключові слова: вейвлет, метод кластеризації, розпізнавання, сигнал, відхилення, графіки, електричне навантаження.

Рассмотрен метод кластеризации информационных сигналов на основе адаптивного коэффициента отклонения относительно общей выборки значений. Разработан способ определения четких границ между кластерами, состоящими из средних мощностей вейвлет-коэффициентов по субполосам разложения.

Ключевые слова: вейвлет, метод кластеризации, распознавание, сигнал, отклонение, графики, электрическая нагрузка.

The method of clusterization of information signals on the basis of adaptive factor of a deviation concerning the general sample of values is considered. The way to define the precise borders between clusters that consist of average powers wavelet-coefficient on substrips of decomposition is developed.

Key words: wavelet, method of clusterization, recognition, signal, deviation, schedules, electric loading.

Введение. В настоящее время вейвлет-преобразование получило широкое применение для анализа нестационарных информационных сигналов. Целесообразность его использования в электроэнергетике обуславливается тем, что графики электрических нагрузок (ГЭН) носят апериодический и нелинейный характер. Учитывая, что вейвлет-анализ представляет собой особый тип линейного преобразования информационных сигналов и отображаемых этими сигналами физических данных о процессах и физических свойствах природных сред и объектов, данный метод может быть успешно использован для математической обработки необходимых сигналов. С его помощью можно не только выявить характерные частоты сигнала, но и получить сведения об определенных локальных координатах, при которых проявляются эти частоты [1]. Результатом вейвлет-преобразования являются вейвлет-коэффициенты, которые в дальнейшем поддаются кластеризации, то есть автоматическому разбиению на группы по общему признаку [2]. Иными

словами, кластеризация представляет собой рациональный способ разложения данных и формирования представительных выборок.

Существующие методы кластерного анализа ГЭН представляют собой классический математический аппарат [3]. Например, для обработки графиков ГЭН часто используется кластеризация с учетом распределения плотности вероятности по нормальному закону. Однако этот метод, в отличие от предлагаемого, имеет усложненный алгоритм вычислений и требует дополнительного определения пределов между кластерами. Кроме того, он нечетко отображает разные по форме ГЭН. Исходя из того, что электрические сети постоянно изменяются, и при этом часто возрастает количество измеряемых параметров, возникает необходимость усовершенствования методов кластеризации с дальнейшим упрощением алгоритмов и сохраняемых данных.

В данной работе рассматривается метод кластеризации ГЭН на основе адаптивного коэффициента отклонения относительно общей выборки значений. В качестве вектора признаков для кластеризации рассматривается упорядоченная последовательность средних мощностей вейвлет-коэффициентов по субполосам [4].

Цель работы. Создать устойчивый метод кластеризации измеряемых характеристик энергосистемы, результат которого не зависел бы от сезонности и форм ГЭН.

Изложение основных материалов исследования. Вектор признаков кластеризации ГЭН формируется следующим образом (рис. 1). В вейвлет-спектре, сформированном на основе вейвлет-пакетов, усредняется мощность рассчитанных вейвлет-коэффициентов (аппроксимирующих и детализирующих) в пределах каждой субполосы разложения ($k_{m,\,n}(i)$). Усредненные коэффициенты ($\overline{P}_{m,\,n}$) нормируются и, в соответствии с их местами в общей пирамиде вейвлет-пакетов (слева направо и сверху вниз) [5], преобразуются в вектор признаков кластеризации *Y*, *m* – шаг разложения; *N* – общее количество вейвлет-коэффициентов на *m*-м шаге разложения; *n* – порядковый номер вейвлет-коэффициента на определенном шаге разложения; $k_{\,m,n}$ – значение вейвлет-коэффициента на 5-м шаге разложения.

Вначале определяется вид закона распределения нормированных коэффициентов вектора признаков. Предполагается, что члены ряда (вектора признаков) независимы друг от друга и носят случайный характер. Для установления закона распределения рассчитываются основные характеристические параметры ряда – среднее значение (P_{ci}) и среднее отклонение (δ_{ci}). Результаты вычислений приведены на рис. 1. Анализ полученных значений показал, что закон распределения данных вектора признаков является нормальным. Это будет учитываться при проведении дальнейших исследований.

Кластеризация на основе адаптивного коэффициента отклонения относительно общей выборки значений

Часто для оценки степени отклонения относительно среднего значения используется коэффициент вариации, который, в отличие от среднего квадра-

тичного или стандартного отклонения, характеризует не абсолютную, а относительную меру разброса значений признака в статистической зависимости.

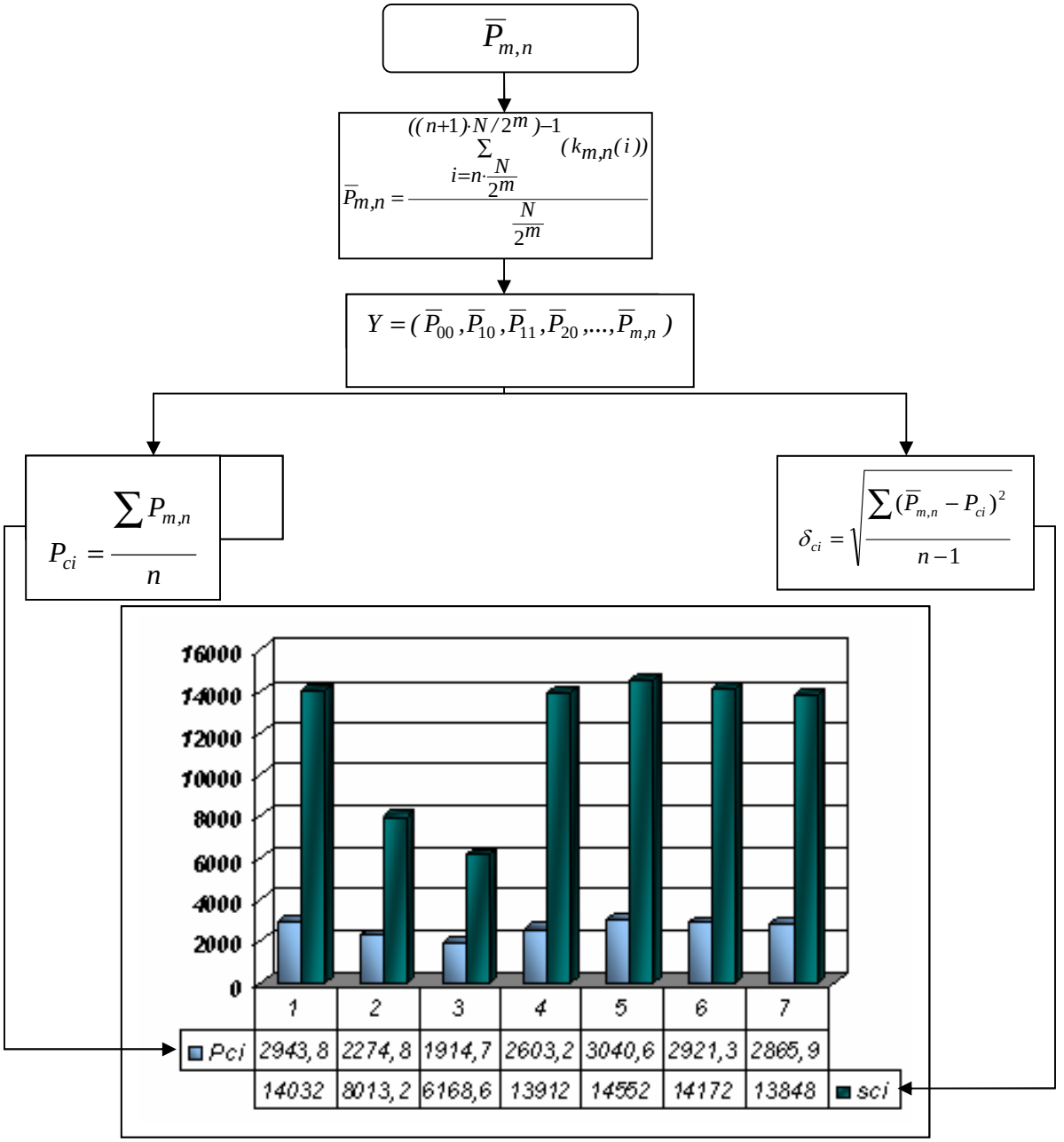

Рис. 1. Расчет базисных параметров преобразованных выборок данных

При этом использование коэффициента вариации позволяет:

сравнить вариацию одного и того же признака разных групп объектов;

выявить степень различия одного и того же признака у одной и той же группы объектов в разное время;

сопоставить вариацию разных признаков у одних и тех же групп объектов.

Коэффициент вариации определяется как

$$\gamma = \frac{\delta_i}{\overline{P}_{ci}}. \quad (1)$$

В связи с тем, что коэффициент вариации не в полной мере отображает классовые различия между ГЭН, как отмечается в [6], с использованием (1) невозможно произвести четкую кластеризацию данных. Более эффективным показателем для кластеризации данных является следующий коэффициент:

$$\beta = \frac{D_i}{\sum_{i,j=1}^{m,n} P_{m,n}}, \quad (2)$$

где D_i – дисперсия, которая определяется по выражению

$$D_i = \sum_{i,j=1}^{m,n} (P_{m,n} - \overline{P}_{ci})^2. \quad (3)$$

Результаты анализа семи ГЭН с применением выражений (2) и (3) представим графически (рис. 2).

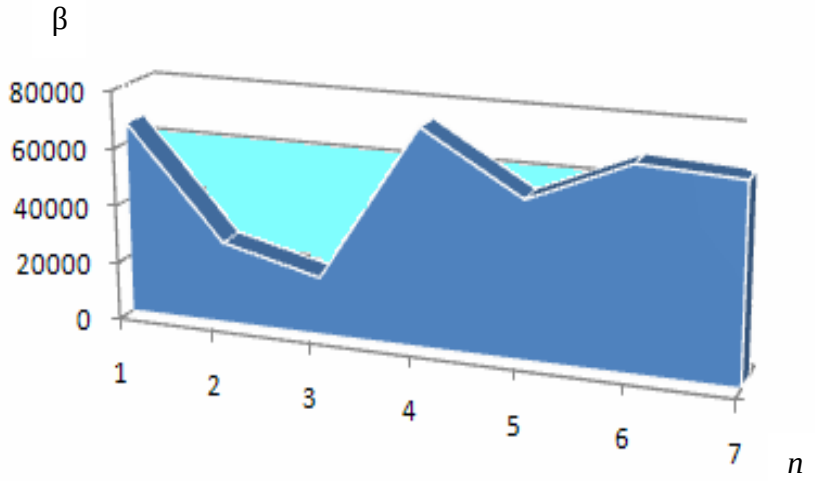

Рис. 2. Кластеризация ГЭН по адаптивному коэффициенту степени отклонения относительно всей выборки значений

Как показали результаты проведения кластеризации с учетом адаптивного коэффициента степени отклонения относительно всей выборки значений, формируются две группы вариационных рядов. Полученные группы соответствуют различным видам графиков потребления, а значит, и есть решением поставленной задачи. Отнесение к определенной группе производится на основе следующего выражения:

$$\beta \le \beta_{\text{пред}} \text{ и } \beta > \beta_{\text{пред}},$$

где $\beta_{\text{пред}}$ – граничное значение, которое определяется по формуле (4).

$$\beta_{\text{пред}} = \frac{1}{2}\left(\frac{D_{\text{вmax}}}{\sum\limits_{i,j=1}^{m,n} P_{m,n\text{в max}}} + \frac{D_{\text{pmin}}}{\sum\limits_{i,j=1}^{m,n} P_{m,n\text{pmin}}} \right), \qquad (4)$$

где $D_{\text{вmax}}$ и D_{pmin} – дисперсии соответственно максимального потребления выходного дня и минимального потребления рабочего дня; $\sum\limits_{i,j=1}^{m,n} P_{m,n\text{в max}}$ и $\sum\limits_{i,j=1}^{m,n} P_{m,n\text{p min}}$ – суммы средних значений мощности соответственно максимального потребления выходного дня и минимального потребления рабочего дня.

Для приведенной выборки исходных данных характерны диапазоны отличия. Формирование диапазонов считается завершенным, если определены граничные значения пределов (рис. 3).

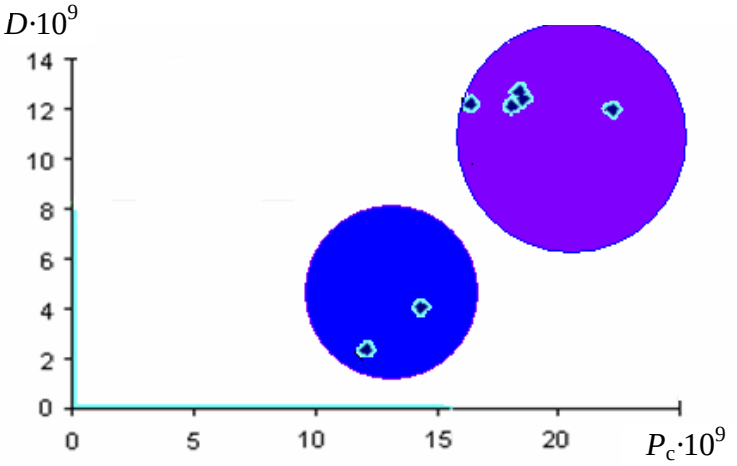

Рис. 3. Количественное отношение между дисперсией D и суммарными значениями ГЭН P_c

Следует отметить, что при определении предельного критерия кластеризации для данного объекта выборка исходных данных должна иметь представительный характер, то есть включать в себя данные для разных периодов потребления (например, зимний и летний период). Таким образом, рассчитанный предел будет определен оптимально и соответствует условиям выбора. В случае изменения объемов потребления электроэнергии полученная точность обеспечит его применимость для дальнейшего использования.

Для определения граничного значения коэффициента $\beta_{\text{пред}}$ объем исходных данных зависит от количественного значения устанавливаемой величины ряда и величины отклонения относительно среднего значения. С учетом того, что объем выборки незначителен для оценки степени корреляции рассматриваемых величин, применяется формула (5):

$$r_{P_c D} = \frac{N\sum P_{ci} D_i - \sum P_{ci} \sum D_i}{\sqrt{N\sum P_{ci}^2 - (\sum P_{ci})^2}} \cdot \frac{1}{\sqrt{N\sum D_i^2 - (\sum D_i)^2}}. \quad (5)$$

В результате математических действий определяется наличие связи между величинами, полученное значение $r_{P_c D}$ = –0,12 соответствует слабой корреляционной связи. Данный результат подтверждает необходимость применения дополнительного показателя, который бы отображал влияние двух рассматриваемых переменных. Проверим обоснованность использования адаптивного коэффициента степени отклонения относительно всей выборки значений.

Изучение аппаратов оценки математической статистики и теории классификации показало, что по результатам кластер-анализа объекты разбиваются на группы [7]. Для осуществления такого анализа необходимо оперировать следующими математическими характеристиками:

центром кластера z_j;

радиусом $\overline{D}_j$.

На первом этапе вычисляется положение центров кластеров:

$$z_{Dj} = \sqrt[N_{Dj}]{\prod_{i=1}^{N_{Dj}} D_i} \text{ и } z_{Pj} = \sqrt[N_{Pj}]{\prod_{i=1}^{N_{Pj}} P_{ci}}, \quad (6)$$

где N_{Dj} и N_{Pj} – соответственно количество значений дисперсий и суммарных средних значений мощности.

В результате центр кластера составляет $z(z_{Pj}; z_{Dj})$.

На втором этапе вычисляется радиус кластера, который соответствует допустимому разбросу значений выборки данных относительно центра кластера таким образом, чтобы выполнялось условие кластерного соответствия рассматриваемого ряда. В математическом виде радиус кластера описывается выражением:

$$\overline{D}_j = \sqrt{\left(\max\left(D_i - z_{Dj}\right)\right)^2 + \left(\max\left(P_{ci} - z_{Pj}\right)\right)^2}. \quad (7)$$

Полученные расчетные значения переменных в выражениях (6) и (7) сведены в табл. 1.

Таблица 1. Математические характеристики кластера

	P_c	$D \cdot 10^9$	$\overline{D}_p \cdot 10^9$
z_p	1,75	8,85	6,46
$z_в$	1,53	7,02	5,17

На основании данных приведенной выше таблицы оценивается соответствие значений проверочной выборки данных установленным классам. Для этого руководствуются критерием

$$\overline{D}_j \geq \sqrt{(D_i - z_{Dj})^2 + (P_{ci} - z_{Pj})^2}\ .$$

Если условие выполняется, рассматриваемая выборка данных принадлежит данному классу. В противоположном случае проверяется ее соответствие соседнему классу. Результаты такого сравнения представлены в табл. 2.

Из них следует, что к первой группе относятся все значения выборки, кроме третьей. При сравнении дисперсий табл. 1 и 2 было установлено классовое соответствие первой и второй группы. Можно сделать вывод, что предложенная модель обеспечивает выполнение условий кластеризации.

Таблица 2. Результаты проверки кластеризации

$\overline{D}_p \cdot 10^9$	$\overline{D} \cdot 10^9$	Проверка
6,5	5,1	+
	5,95	+
	8,43	-
	6,18	+
	5,76	+
	5,82	+

Осуществим еще одну проверку достоверности установленных границ, как это было представлено на рис. 2. Графически результат такой проверки представлен на рис. 4. Определенные значения β формируют кластеры в виде областей, границы которых задаются диапазонами. При установлении классового соответствия выборок данных те, которые принадлежат к одной группе, в действительности относятся к группе ГЭН, подобных по форме.

При использовании информативного количества выборок определяется частота попадания ГЭН в заданный диапазон. При соблюдении сохранения поочередности ввода данных можно также установить повторяемость характера нагрузки (рис. 5).

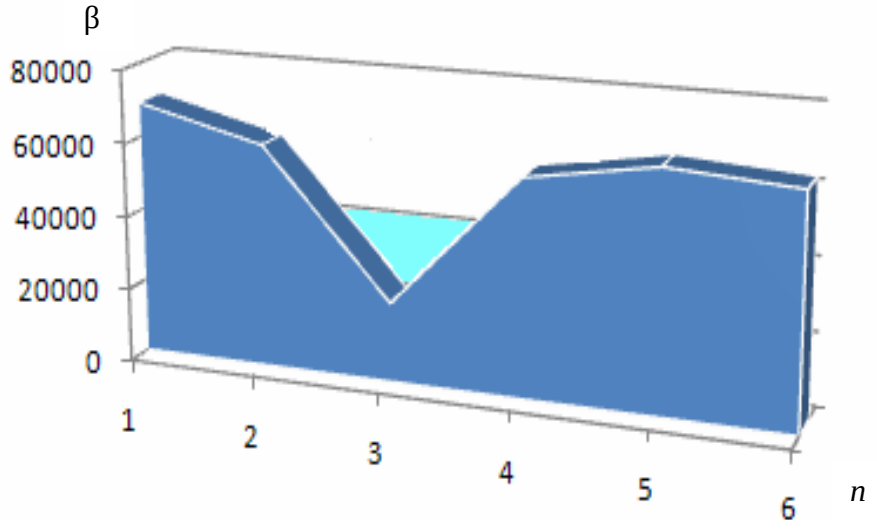

Рис. 4. Проверка кластеризации адаптивного коэффициента степени отклонения относительно всей выборки значений

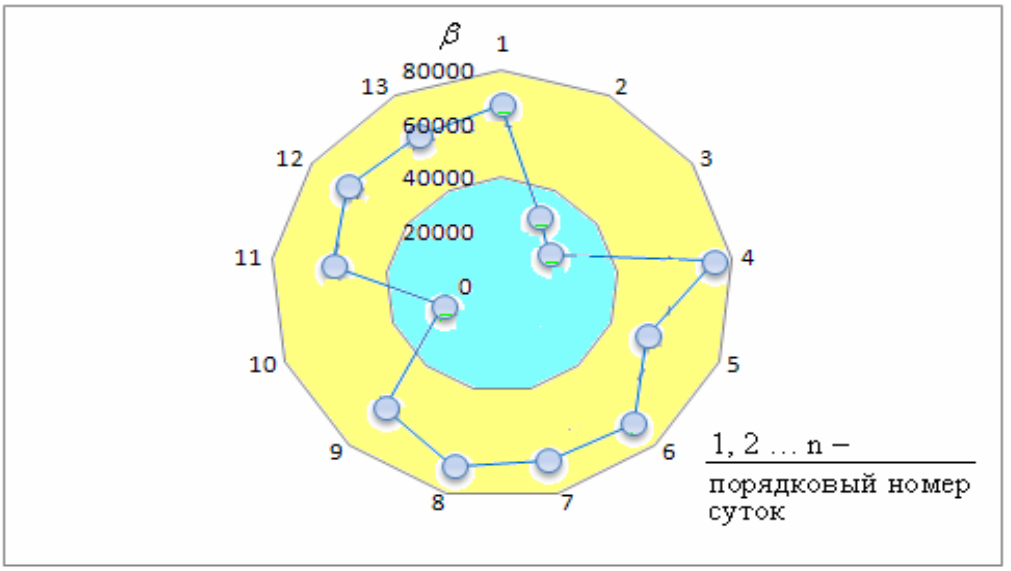

Рис. 5. Общее отображение величин адаптированных коэффициентов степени отклонения относительно всей выборки значений

Используя реальные ГЭН для разных сезонов, можно убедиться, что предложенный метод кластеризации достоверен и применим без учета зимнего и летнего периода.

Выводы

Множество существующих вейвлетов позволяет эффективно обрабатывать сигналы с целью уменьшения объемов хранимых данных, их кодировки и решения других задач. В основном результатом таких преобразований становится нахождение решения только одной проблемы. А полученные вейвлет-коэффициенты в случае дальнейшей обработки требуют применения сложных математических аппаратов вычисления.

Изменение последнего шага обработки вейвлет-коэффициентов привело к формированию такой выборки значений, которая подчиняется существующим статистическим законам и увеличивает точность кластеризации.

Для проведения оптимального анализа вычисленных значений и определения четких границ между группами предложен адаптивный коэффициент степени отклонения относительно всей выборки значений вместо коэффициента вариации.

Преимущество использования разработанного метода заключается в определении адаптивного коэффициента отклонения относительно общей выборки значений, который обеспечил увеличение расстояния между кластерами, что привело к снижению влияния величины потребления и формы ГЭН на точность результатов кластеризации.

1. *Астафьева Н. М.* Вейвлет-анализ: основы теории и примеры применения / Н. М. Астафьева. – Успехи физических наук. – 1996. – Т. 166. – № 11. – С. 1145–1170.

2. *Дремин И. М.* Вейвлеты и их использование / И. М. Дремин, О. В. Иванов, В. А. Нечитайло. – УФН. – 2001. – Т. 171. – № 5. – С. 465–501.

3. *Праховник А. В.* Энергосберегающие режимы электроснабжения горно-добывающих предприятий / А. В. Праховник, В. П. Розен, В. В. Дегтярев. – М.: Недра, 1985. – 232 с.

4. *Волошко А. В.* Метод формирования признаков классификации графиков электрических нагрузок на основе вейвлет-преобразования / А. В. Волошко // Промелектро. – 2009. – № 1. – С. 39–43.

5. *Дворников С. В.* Метод распознавания на основе вейвлет-пакетов / С. В. Дворников, А. М. Сауков // Научное приборостроение, 2004. – Т. 14. – № 1. – С. 57–65.

6. *Козлов П. В.* Вейвлет-преобразование и анализ временных рядов / П. В. Козлов, Б. Б. Чен. – Вестник КРСУ. – 2002. – № 2. – С. 48–56.

7. *Гнатюк В. И.* Методика оптимальным управлением энергопотребления / В. И. Гнатюк. – Энергетика. – 2008. – № 4. – С. 128–136.