

**Keywords:** renewable and alternative energy sources, comprehensive approach, long-term forecasting, priority technologies of energy generation, risk assessment.

1. Prakhovnik A.V. Distributed generation in the energy supply systems. K.: Osvita Ukraine, 2007. – 464 p.
2. Energy strategy of Ukraine till 2030: Cabinet of Ministers of Ukraine N 145-p от 15.03.2006. – K.: Parlam. izd., 2006. – 129 p.
3. Dictionary of Cybernetics / edited by V.S. Mikhalevich. - 2nd Edition - K.: 1989. - 751 p.
4. Ksenofontov M. Theoretical and applied aspects of the socio-economic forecasting / M. Ksenofontov. - Moscow: ISESP, 2002. - 312.
5. Winters P.R. Forecasting sales by exponentially weighted moving averages/ P.R. Winters // Management Science. - 1960. - Vol. 6. - №3.
6. Kudrya S.A. «Atlas of energy potential of renewable and unconventional energy sources of Ukraine» / Kudrya S.A., Yacenko L.V., Dushina G.P. and others. – NAS of Ukraine, State Commette of Ukraine on Energy Saving, - K., 2001, 41p.
7. Saaty, Thomas L. Decision making in dependence and feedback: The analytic network. Per. from English. / Scientific. Ed. AV Andreychikov, O.N. Andreichikova. - Moscow: Publishing LKI, 2008. - 360 p.

Надійшла 15.10.2013

Received 15.10.2013

УДК 517.4: 519.652

**А. В. Волошко**, канд. техн. наук, доцент; **Т. Н. Лугчин**; **Я. С. Бедрак**  
Национальный технический университет Украины «Киевский политехнический институт»

## **КОМБИНИРОВАННЫЙ МЕТОД ВОССТАНОВЛЕНИЯ УЧЕТНЫХ ДАННЫХ ЭНЕРГОПОТРЕБЛЕНИЯ НА ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЯХ**

*В статье предложен комбинированный метод восстановления учетных данных энергопотребления промышленных предприятий. Алгоритмизированы этапы локальных и глобальных методов восстановления значений. Рассмотрены распространенные глобальные методы описания распределения данных. Представлены оценки качества точности выбора модели для описания исследованных выборок.*

*Разработана процедура выбора оптимального математического метода восстановления учетных данных энергопотребления. Также предложено графическое отображение количественных оценок критериев качества моделей. При этом, расчет зависимостей для выборок потребления энергоресурсов и количества производства продукции предлагается вести с учетом их последовательности на круговой диаграмме.*

**Ключевые слова:** глобальные и локальные методы, восстановление данных, учет энергопотребления.

### **Введение**

С ростом технического обеспечения приборов учета энергоресурсов на промышленных предприятиях приобретают актуальность вопросы, связанные с достоверностью измеренной информации, задачами накопления, хранения и передачи данных. При этом полученные данные поддаются анализу рядом методов, связанных с предварительной их обработкой, группированием, кластеризацией, классификацией, сжатием, восстановлением. Одна из самых проблематичных ситуаций возникает при потере учетных данных (вследствие несанкционированных вмешательств, технических и программных ошибок), так как на их основе формируются расчетные счета, оценивается эффективность работы предприятия, а сбои подобного рода могут привести к ошибочным прогнозам и выводам.

Восстановление данных необходимо начинать с проверки массивов данных на наличие нетипичных выбросов значений. Для выявления случайных значений ряда целесообразно использовать критерий Ирвина [1]. Далее необходимо определится с требуемыми результатами, а также методами и способами их достижения.

© Волошко А. В., Лугчин Т. Н., Бедрак Я. С., 2013

**1 Методы восстановления данных и алгоритмизация их вычислений**

Наиболее распространенные методы восстановления и алгоритмы их расчетов детально рассмотрены в работе [2]. Группирование методов, которые используются для восстановления данных, можно осуществить следующим образом [3]:

Простые алгоритмы — итеративные алгоритмы, основанные на простых арифметических операциях, регрессионном моделировании.

Сложные алгоритмы — итеративные алгоритмы, которые предусматривают оптимизацию некоторого функционала, который отображает точность расчетов, и подставляется на место пропущенного значения. Их можно разделить на глобальные и локальные.

При проведении анализа функции на основе учетных данных с разрывами первого рода с целью ее восстановления предлагается применять локальные методы. При использовании локальных методов (табл. 1) функция разбивается на отдельные сегменты, между которыми находится разрыв и полученные выборки рассматриваются отдельно. Иначе при нахождении одной зависимости для всего интервала возникает дополнительная погрешность в местах разрыва функции.

Таблица 1

Алгоритм восстановления данных локальным методом

№ п/п	Наименование процедуры
1	<i>Проведение сортировки по объему потребления энергии от минимального до максимального значения</i>
2	<i>Разделение выборки на несколько групп в зависимости от потребления электроэнергии</i>
3	<i>Определение отклонений выборок от начального значения в пределах своих границ</i>
4	<i>Построение аппроксимационных однофакторных уравнений</i>
5	<i>Выбор метода, обеспечивающего наилучшее качество модели при помощи критериев качества модели</i>
6	<i>Оценка скачков графика энергопотребления</i>
7	<i>Процедура восстановления данных</i>

Следует отметить, что на результаты исследований существенно влияют: закон распределения данных, объем выборок и наличие пропусков значений. Данные вопросы были представлены в работах [4–6].

Глобальные методы восстановления данных предлагается применять, если исследуемая функция на всем отрезку не имеет очевидных разрывов и имеет относительно равномерный характер. Алгоритм восстановления данных глобальным методом представлен в табл. 2.

Таблица 2

Алгоритм восстановления данных глобальным методом

№ п/п	Наименование процедуры
1	<i>Выявление параметров производства, которые влияют на потребление энергоресурсов и их оценивание</i>
2	<i>Построение зависимостей расхода одного от других энергоресурсов</i>
3	<i>Анализ формы построенных зависимостей</i>
4	<i>Расчет коэффициента корреляции для всех функциональных зависимостей</i>
5	<i>Построение аппроксимационных однофакторных уравнений поочередно для зависимостей одного энергоресурса от другого и наоборот. (Если таких зависимостей несколько, например, 3, то необходимо найти следующие оптимальные аппроксимационные уравнения <math>X_1 = f(X_2)</math>, <math>X_2 = f(X_3)</math>, <math>X_3 = f(X_1)</math>)</i>
6	<i>Проверка достоверности восстановления данных по критериям качества модели</i>
7	<i>Выбор метода с наилучшим качеством модели</i>
8	<i>Процедура восстановления данных</i>

Существующие математические функции позволяют для большинства задач применять глобальные подходы описания функции, то есть описывать полностью функцию на всем диапазоне ее рассмотрения без разбивания на части и рассматривать как одно целое с описанием в виде одной характеристики.

Считаем, что нужно для восстановления данных применить комбинированный метод, сочетающий, как локальные методы, так и глобальные восстановления данных. Такой метод должен учитывать особенности потребления энергоресурсов для каждого предприятия.

Для уменьшения ошибки при восстановлении утерянных данных расхода энергоресурсов предлагается использовать информацию о потреблении или других энергоресурсов, или расхода продукции, которые тесно взаимно зависимы друг от друга за один и тот же период времени.

Встречаются ситуации, когда величина электропотребления зависит от одной величины (например, потребление электроэнергии насосной станции зависит от расхода воды за один и тот же период времени). В таком случае построение аппроксимационных однофакторных уравнений целесообразно производить с помощью программы Matlab в библиотеке Curve Fitting Toolbox [7].

Данная библиотека содержит следующие широко используемые математические методы для аппроксимации табличных функций, а именно:

– Метод полиномов (Polynomial) – подбираются коэффициенты полинома, степень которого может варьироваться от 1 до 9 включительно:

$$y(x) = \sum_{i=0}^n p_i x^{n-i};$$

– Метод экспоненциальный (Exponential) – в качестве аппроксимирующей функции выбирается одна из следующих:

$$y(x) = ae^{bx}; \quad y(x) = ae^{bx} + ce^{dx};$$

– Метод Фурье (Fourier) – данные приближаются отрезками ряда Фурье для  $1 < n < 8$ :

$$y(x) = a_0 + \sum_{i=1}^n (a_i \sin i\omega x + b_i \cos i\omega x);$$

– Метод Гаусса (Gaussian) – модель содержит следующий набор функций ( $1 < n < 8$ ):

$$y(x) = \sum_{i=1}^n \frac{a_i}{e^{\left(\frac{(x-b_i)}{c_i}\right)^2}};$$

– Метод отношения полиномов (Rational) – табличная функция аппроксимируется отношением двух полиномов со степенями не выше 5 (не обязательно равных):

$$y(x) = \frac{\sum_{i=0}^n p_i x^{n-i}}{\sum_{i=0}^m q_i x^{m-i}}, \quad q_0 = 1;$$

– Метод суммы синусоид (Sum of Sin Function) – аппроксимирующая функция является отрезком следующего ряда ( $1 < n < 8$ ):

$$y(x) = \sum_{i=1}^n a_i \sin(b_i x + c_i).$$

После того, как математическая модель построена (значение всех коэффициентов модели вычислены), необходимо проверить ее качество. Для оценки эффективности построенной зависимости чаще всего используются следующие критерии качества модели [8]:

1. Среднеквадратичное отклонение *RMSE* ( $\hat{e}$ ):

$$RMSE(\hat{a}) = \sqrt{\frac{\sum_{i=1}^n (y_m(i) - y_p(i))^2}{n}},$$

где  $y_m(i)$  – измеренное значение энергопотребления;

$y_p(i)$  – прогнозное значение;

$n$  – горизонт прогнозирования (количество временных интервалов в выборке).

2. Средняя абсолютная процентная ошибка *MAPE*:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_m(i) - y_p(i)|}{y_m(i)}.$$

3. Информационный критерий Акаике *AIC*, который определяет степень простоты модели:

$$AIC = 2 \cdot k + n \cdot [\ln(RSS)],$$

где  $k$  – число параметров в статистической модели;

$RSS$  – остаточная сумма квадратов ошибок модели.

4. F-критерий Фишера, который указывает на значимость модели [9].

Для упрощения математических вычислений следует использовать один и тот же вид математической модели (метода из библиотеки Curve Fitting Toolbox).

В случае однофакторной зависимости предлагается выбирать модель, у которой значение *MAPE* и *RMSE* ( $e$ ) одного и того же математического метода восстановления данных для зависимостей  $X_1 = f(X_2)$  и  $X_2 = f(X_1)$  минимальны.

В промышленности часто расход электроэнергии зависит от нескольких величин. Тогда если коэффициент корреляции расхода электроэнергии с каждой из этих величин выше 0,75, то для восстановления данных также предлагается использовать библиотеку Curve Fitting Toolbox программы Matlab.

Если зависимая переменная зависит от нескольких независимых переменных, то довольно сложно определить оптимальную модель даже по одному критерию качества модели. Поэтому предлагается использовать для выбора оптимальной модели геометрический способ при помощи приведенного ниже алгоритма.

## 2 Алгоритм применения геометрического метода для выбора оптимального метода восстановления данных при учете критерия качества *MAPE*

1. Строится круговая диаграмма в полярных координатах с делением на  $n$  секторов, по количеству однофакторных зависимостей. Например, для трёх взаимосвязанных величин используются зависимости  $X_1 = f(X_2)$ ,  $X_2 = f(X_3)$ ,  $X_3 = f(X_1)$ .

2. Разделяется каждый сектор, как показано на рис., на векторы по количеству использованных методов (например, экспоненциальный E, Фурье F, Гаусса G, полиномов P, отношения полиномов R, суммы синусоидальных функций S).

В пределах этих секторов все значения критерия *MAPE* разных методов отображаются последовательно в виде масштабированной по величине средней абсолютной процентной ошибки векторов, началом которых является центр координат.

3. Соединяются между собой вершины векторов одинаковых моделей из каждого сектора, образуя  $n$ -угольники. Число сторон в  $n$ -угольнике равняется количеству взаимозависимых учетных величин.

4. Вычисляются площади  $n$ -угольников путем деления их на треугольники с вершинами в координатах P1 ( $\rho_1; \varphi_1$ ), P2 ( $\rho_2; \varphi_2$ ), P3 ( $\rho_3; \varphi_3$ ), площади которых находятся по формулам для полярных и декартовых координат [10]:

$$S = \frac{1}{2} \cdot [\rho_1 \cdot \rho_2 (\sin(\varphi_2 - \varphi_1)) + \rho_2 \cdot \rho_3 (\sin(\varphi_3 - \varphi_2)) + \rho_3 \cdot \rho_1 (\sin(\varphi_1 - \varphi_3))] \quad (1)$$

или

$$S = \frac{1}{2} \cdot |x_1 y_2 - x_2 y_1 + x_2 y_3 - x_3 y_2 + x_3 y_1 - x_1 y_3|. \quad (2)$$

5. Используется для восстановления данных метод, которому соответствует минимальное значение площади  $n$ -угольника.

Для определения площади многоугольника возможно использовать то, что многоугольник является полигоном. Площадь полигона определяется по формуле [11]:

$$S = \frac{1}{2} \sum_{i=0}^{n-1} |(x_i y_{i+1} - x_{i+1} y_i)|, \quad (3)$$

где  $n$  – номер стороны полигона;

$(x_i; y_i)$ ,  $i \in [1, n]$  – координаты вершин полигона;  $x_{k+1} = x_1$  и  $y_{k+1} = y_1$ .

Пример выбора оптимального метода восстановления данных предприятия представлен на рис. для трех взаимозависимых учетных величин  $X_1$ ,  $X_2$  и  $X_3$ , которые формируют секторы. Далее рассматриваются зависимости  $X_1 = f(X_2)$ ,  $X_2 = f(X_3)$ ,  $X_3 = f(X_1)$ .

В полярных координатах, как представлено на рис., три сектора содержат столько лучей, сколько и методов (применяются следующие методы: экспоненциальный E, Фурье F, Гаусса G, полиномов P, отношения полиномов R, суммы синусоидальных функций S). В масштабе наносятся значения *MAPE* на каждом луче. Точки на лучах, соответствующих одному и тому же методу, соединяются в треугольники. Треугольник с минимальной площадью определяет метод с минимальным значением критерия качества модели *MAPE*.

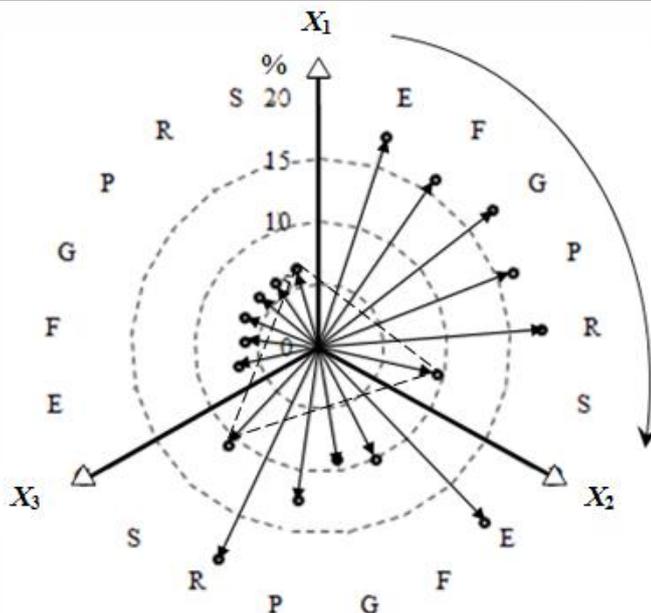


Рис. Выбор оптимального метода восстановления данных при учете одного критерия качества

### Выводы

Предлагается применить для восстановления данных комбинированный метод, который учитывает особенности потребления энергоресурсов каждого предприятия.

Если исследуемая функция имеет разрывы первого рода, то для восстановления данных предлагается применять локальные методы. Глобальные методы более эффективны, если исследуемая функция на всем отрезке без очевидных разрывов и имеет относительно равномерный характер.

В результате исследований доказана рациональность использования геометрического подхода для выбора метода, обеспечивающего наилучшее качество восстановления данных.

### Список литературы

1. Красс М.С. Математика для экономистов / М.С. Красс, Б.П. Чупрынов // СПб.: Питер. – 2005. – 464 с.
2. Снитюк В. Е. Эволюционный метод восстановления пропусков в данных / В. Е. Снитюк // К.: Сб. труд. VI МК «Интеллектуальный анализ информации». – 2006. – С. 262-271.
3. Злоба Е. А. Статистические методы восстановления пропущенных данных / Е. А. Злоба, И. Р. Яцкив // Computer Modeling & New Technologies – Vol. 6 – 2004. – С. 51-61.
4. Волошко А. В. Відновлення втрачених облікових даних / А.В. Волошко, Т.М. Лутчин, Д.К. Міщенко, Я.С. Бедерак // Вісник КНУ. – Вип. 2. – № 73. – 2012. – С. 40–44.
5. Волошко А. В. Влияние сортировки массива данных энергопотребления на точность их восстановления / А.В. Волошко, Т.Н. Лутчин, Я.С. Бедерак // Zbior Raportow Naukowych / Warszawa. – 2012. – Р. 29–33.
6. Волошко А. В. Восстановление учетных данных энергопотребления на промышленных предприятиях / А.В. Волошко, Т.Н. Лутчин, Я.С. Бедерак, А.В. Ткаченко // Материалы МНПК «Техника и технология: Новые перспективы развития», Москва. – 2012. – С.179–188.
7. Основы работы в Curve Fitting Toolbox – [Ел. ресурс] – Режим доступу – URL: <http://matlab.exponenta.ru/curvefitting> (дата обращения 2.10.2012).
8. Лук'яненко І. Г. Економетрика: Підручник / І. Г. Лук'яненко, Л. І. Краснікова // К. : Товариство «Знання», КОО. – 1998. – 494 с.
9. Находов В. Ф. Вибір методу побудови «Стандартів» енергоспоживання на основі кількісних та якісних критеріїв / В. Ф. Находов, Д. О. Іванько, А. В. Головка // Енергетика. Екологія. Людина. – 2013. – С. 386–393.
10. Корн Г. Справочник по математике для научных работников и инженеров / Г. Корн, Т. Корн // М. : Наука. – 1974. – 832 с.
11. Zwillinger D. CRC Standard Mathematical Tables and Formulae / D. Zwillinger // CRC, Boca Raton. – 2003. – 857 p.

УДК 517.4: 519.652

**А. В. Волошко, канд. техн. наук, доцент, Т. М. Лутчин, Я. С. Бедерак**  
**Національний технічний університет України «Київський політехнічний інститут»**  
**КОМБІНОВАНИЙ МЕТОД ВІДНОВЛЕННЯ ОБЛІКОВИХ ДАНИХ ЕНЕРГОСПОЖИВАННЯ НА**  
**ПРОМИСЛОВИХ ПІДПРИЄМСТВАХ**

*У статті запропоновано комбінований метод відновлення облікових даних енергоспоживання промислових підприємств. Алгоритмізовані етапи локальних і глобальних методів відновлення значень. Розглянуто поширені глобальні методи опису розподілу даних. Представлені оцінки якості точності вибору моделі для опису досліджених вибірок.*

*Розроблено процедуру вибору оптимального математичного методу відновлення облікових даних енергоспоживання. Також запропоновано графічне відображення кількісних оцінок критеріїв якості моделей. При цьому, розрахунок залежностей для вибірок споживання енергоресурсів та кількості виробництва продукції пропонується вести з урахуванням їх послідовності на круговій діаграмі.*

**Ключові слова:** глобальні та локальні методи, відновлення даних, облік енергоспоживання.

**A. Voloshko, T. Lutchyn, Ya. Bederak**  
**National Technical University of Ukraine «Kyiv Polytechnic Institute»**  
**COMBINED METHOD OF DATA RECOVERY OF ENERGY CONSUMPTION FOR INDUSTRIAL**  
**ENTERPRISES**

*In this paper a combined method of restoring measurements of energy consumption in enterprises is proposed. Algorithms were created for steps of the local and global recovery methods. Considered common methods for describing the global distribution of data. Estimates for the accuracy quality of the model selection are presented for the description of the investigated data samples.*

*The selecting procedure of the optimal mathematical method for measurements recovery of energy consumption was created. Also a graphical representation of quantitative evaluation of quality models criteria is proposed. In this case, the calculation of consequences for samples of energy consumption and production quantities are proposed to calculate with their sequence in a round chart.*

**Keywords:** global and local methods, data recovery, energy accounting.

1. Krass M.S. Mathematics for economists / M.S. Krass, B.P. Chuprunov // SPb. Peter. – 2005. – 464 p.
2. Snytyuk V.Ye. Evolution method of recovery missing data / V.Ye. Snytyuk // K.: Proc. IC VI "Intelligent analysis of information." – 2006. – P. 262–271.
3. Zloba E.A. Statistical methods of the missing data recovery / E.A. Gall, I.R. Yatskiv // Computer Modeling & New Technologies. – Vol. 6. – 2004. – P. 51–61.
4. Voloshko A.V. Recovery of lost accounting data / A.V. Voloshko, T.N. Lutchyn, D.K. Mishchenko, Ya.C. Bederak // Bulletin of KNU. – Vol. 2. – № 73. – 2012. – P. 40–44.
5. Voloshko A.V. Influence of sorting data sets on the accuracy of energy consumption recovery / A.V. Voloshko, T.N. Lutchyn, Ya.C. Bederak // Zbiór Raportów Naukowych / Warszawa. – 2012. – P. 29–33.
6. Voloshko A.V. Recovery of accounting data of energy consumption in industrial enterprises / A.V. Voloshko, T.N. Lutchyn, Ya.C. Bederak, A.V. Tkachenko // Materials ISPC "Equipment and Technology: New Perspectives for Development", Moscow. – 2012. – P.179–188.
7. Basics of work in Curve Fitting Toolbox – [El. resource] – access mode – URL: <http://matlab.exponenta.ru/curvefitting> (date of access 2.10.2012).
8. Lukyanenko I.G. Econometrics: Tutorial / I.G. Lukyanenko, L.I. Krasnikova // K.: "Knowledge". – 1998. – 494 p.
9. Nahodov V.F. Method choice of construction "Standards" of energy consumption based on quantitative and qualitative criteria / V.F. Nahodov, D.A. Ivanko, A. Golovko // Energy. Ecology. Human. – 2013. – P. 386–393.
10. Korn G. Mathematical Handbook for Scientists and Engineers / G. Korn, T. Korn // M.: Nauka. – 1974. – 832 p.
11. Zwillinger D. CRC Standard Mathematical Tables and Formulae / D. Zwillinger // CRC, Boca Raton. – 2003. – 857 p.

Надійшла 23.09.2013  
Received 23.09.2013