

Ирина Чучуева

математик, [chuchueva@mbureau.ru](mailto:chuchueva@mbureau.ru)

## **«Модель экстраполяции временных рядов по выборке максимального подобия»**

### **Аннотация**

В статье предложена новая модель экстраполяции временных рядов по выборке максимального подобия. Приводится описание модели, а также алгоритм решения задачи ее идентификации. Оценка эффективности модели экстраполяции проведена на временных рядах энергопотребления и цен на электроэнергию.

### **Ключевые слова**

прогнозирование, экстраполяция, потребление, цена РСВ, временные ряды, оптовый рынок электроэнергии и мощности

## Введение

Задача прогнозирования будущих значений временного ряда на основании его текущих и прошлых значений является основой для 1) планирования в экономике и торговле, 2) планирования выпуска продукции, 3) складского контроля, 4) управления и оптимизации промышленных процессов и т.д. [1]. В настоящее время методы прогнозирования продолжают интенсивно развиваться параллельно с развитием систем хранения данных и вычислительных мощностей.

Методы прогнозирования можно разделить на интуитивные и формализованные [2]. Согласно обзору [3] формализованные методы делятся на две категории – статистические и методы искусственного интеллекта. В статистических методах для вычисления прогнозируемого значения временного ряда определяется некоторое уравнение, полученное на основании изучения истории временного ряда и отражающее зависимость будущих значений ряда от его истории и/или внешних факторов. Методы искусственного интеллекта имитируют процесс мышления человека в изучении истории и предсказании будущих значений ряда. В работе рассматривается статистический метод прогнозирования временных рядов.

Математическая зависимость значений временного ряда за пределами его фактических значений от этих фактических значений называется *моделью экстраполяции* [4]. В настоящей работе предлагается новая модель экстраполяции временных рядов на основе *выборки максимального подобия*.

Исследование эффективности предложенной модели экстраполяции выполнено на временных рядах цен рынка на сутки вперед оптового рынка электроэнергии и мощности (далее цен РСВ) и энергопотребления того же рынка (далее энергопотребления). Экстраполированные значения энергопотребления и цен РСВ необходимы организациям, регулирующим

энергосистему, для управления ее техническими и экономическими параметрами. Кроме того, экстраполированные значения энергопотребления и цен РСВ помогают участникам энергорынка повышать свои финансовые результаты [3,5-7].

Заметим, что задача прогнозирования цен РСВ для России является новой. Особенность ее состоит в том, что по мере реформирования энергорынка России алгоритм расчета цен РСВ подвергается изменениям. Цены РСВ до 01.01.2008 года рассчитывались по одному алгоритму, затем алгоритм был изменен. Важность прогнозирования временного ряда цен РСВ для участников рынков электроэнергии отмечается, например, в работе [3].

Задача прогнозирования энергопотребления имеет множество специализированных моделей, созданных за последние годы [8-10], однако исследования в данной области продолжаются.

Статья содержит в себе описание предлагаемой модели экстраполяции (раздел 1), а также алгоритм решения задачи идентификации модели (раздел 2). Далее в статье приведены результаты, подтверждающие эффективность предложенной модели (раздел 3). В заключении указаны особенности предложенной модели.

## 1. Модель экстраполяции

В работе речь идет о дискретных временных рядах, значения которых получены в моменты времени  $t_1, t_2, t_3, \dots, t_N$ . Указанные моменты времени могут быть неравноотстоящими.

Временной ряд  $Z = z(t_1), z(t_2), z(t_3), \dots, z(t_N)$  обозначим  $Z_1^N = z_1, z_2, z_3, \dots, z_N$ . Набор последовательных значений  $Z_t^M = z_t, z_{t+1}, z_{t+2}, \dots, z_{t+M-1}$ , лежащих внутри временного ряда  $Z_1^N$ , назовем *выборкой* из этого ряда длины  $M$  с моментом

начала отсчета  $t$ ;  $M \in [1, N-1]$ ;  $t \in [1, N-M]$ . Разность начал отсчетов выборок  $Z_t^M$ ,  $Z_{t-k}^M$  назовем *задержкой*  $k$ ,  $k \in [1, t-1]$ .

Коэффициент корреляции выборок  $Z_t^M$ ,  $Z_{t-k}^M$  определим как

$$\rho_k = \frac{\text{cov}(Z_t^M, Z_{t-k}^M)}{\sqrt{D[Z_t^M]} \cdot \sqrt{D[Z_{t-k}^M]}}, \quad (1)$$

где  $\text{cov}(Z_t^M, Z_{t-k}^M)$  – ковариация исходных выборок, а  $D[Z_t^M]$  и  $D[Z_{t-k}^M]$  – их дисперсии [1].

На основе коэффициента корреляции (1) введем *меру подобия* выборок  $Z_t^M$ ,  $Z_{t-k}^M$

$$\rho_k^M = \frac{|\text{cov}(Z_t^M, Z_{t-k}^M)|}{\sqrt{D[Z_t^M]} \cdot \sqrt{D[Z_{t-k}^M]}} \in [0,1].$$

Величина  $\rho_k^M$  зависит от длины  $M$  выборок  $Z_t^M$ ,  $Z_{t-k}^M$ , а также от задержки  $k$ . Мера подобия  $\rho_k^M$  характеризует степень линейной зависимости выборок  $Z_t^M$  и  $Z_{t-k}^M$ : чем ближе к единице значение  $\rho_k^M$ , тем выше их линейная зависимость.

Определим для выборки  $Z_t^M$  величины  $\rho_1^M, \rho_2^M, \dots, \rho_{t-1}^M$  и найдем их максимальное значение

$$\rho_{k \max}^M = \max(\rho_1^M, \rho_2^M, \dots, \rho_{t-1}^M).$$

Выборку, соответствующую задержке  $k_{\max}$  обозначим  $Z_{t-k \max}^M$  и назовем *выборкой максимального подобия* для выборки  $Z_t^M$ . Легко видеть, что для выборки  $Z_{t-k \max}^M$  справедливо равенство

$$\rho_{k \max}^M = \frac{|\text{cov}(Z_t^M, Z_{t-k \max}^M)|}{\sqrt{D[Z_t^M]} \cdot \sqrt{D[Z_{t-k \max}^M]}}.$$

**Гипотеза подобия.** Если выборки  $Z_t^M$ ,  $Z_{t-k \max}^M$  имеют значение величины  $\rho_{k \max}^M$  близкое к единице, то для некоторых значений  $P$  и выборок  $Z_t^{M+P}$ ,  $Z_{t-k \max}^{M+P}$  значение величины  $\rho_{k \max}^{M+P}$  также близко к единице.

Представленные в разделе 3 результаты прогнозирования подтверждают справедливость гипотезы для временных рядов энергопотребления и цен РСВ. Для временных рядов из других предметных областей справедливость гипотезы необходимо проверять.

Перейдем к векторному обозначению выборки  $\mathbf{Z}_t^M = (z_t, z_{t+1}, \dots, z_{t+M-1})^T$  и временного ряда  $\mathbf{Z}_1^N = (z_1, z_2, \dots, z_N)^T$ .

Аппроксимируем выборку  $\mathbf{Z}_t^M$  с помощью выборки  $\mathbf{Z}_{t-k \max}^M$ :

$$\begin{aligned}\mathbf{Z}_t^M &= a_1 \cdot \mathbf{Z}_{t-k \max}^M + a_0 + \mathbf{E}^M; \\ \check{\mathbf{Z}}_t^M &= a_1 \cdot \check{\mathbf{Z}}_{t-k \max}^M + a_0.\end{aligned}\quad (2)$$

Здесь  $a_1$  и  $a_0$  – вещественные константы,  $\mathbf{E}^M$  –  $M$ -мерный вектор значений ошибок аппроксимации, а  $\check{\mathbf{Z}}_t^M$  аппроксимированные значения  $\mathbf{Z}_t^M$ . Тогда выборку  $\mathbf{Z}_{N+1}^P$  можно выразить через некоторую выборку  $\mathbf{Z}_\tau^P$ , лежащую внутри исходного временного ряда  $\mathbf{Z}_1^N$ , в виде

$$\check{\mathbf{Z}}_{N+1}^P = a_1 \cdot \mathbf{Z}_\tau^P + a_0. \quad (3)$$

Алгоритм определения выборки  $\mathbf{Z}_\tau^P$  имеет следующий вид.

- Для выборки  $\mathbf{Z}_{N-M+1}^M$  находим выборку максимального подобия  $\mathbf{Z}_{k \max^*}^M$ , где  $k \max^* = N - M + 1 - k \max$ .
- Согласно (2), аппроксимируем выборку  $\mathbf{Z}_{N-M+1}^M$  с помощью выборки  $\mathbf{Z}_{k \max^*}^M$ :

$$\check{\mathbf{Z}}_{N-M+1}^M = a_1 \cdot \mathbf{Z}_{k \max^*}^M + a_0. \quad (4)$$

- В соответствии с гипотезой подобия, в качестве выборки  $Z_{\tau}^P$  используем выборку  $Z_{\tau}^P = Z_{k_{\max}^* + M}^P$ , то есть выборку, расположенную на оси времени сразу за выборкой максимального подобия (рис. 1).

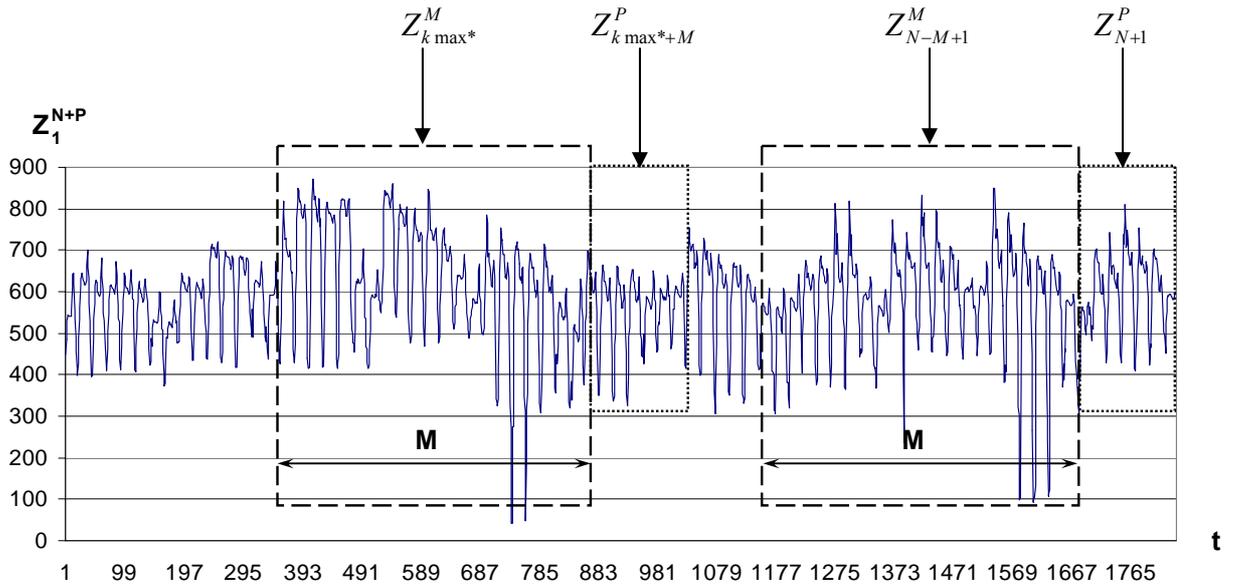


Рис. 1. Положения выборок  $Z_{k_{max}^*}^M$ ,  $Z_{k_{max}^*+M}^P$ ,  $Z_{N-M+1}^M$  и  $Z_{N+1}^P$  на оси времени

В формуле (4) вещественные константы  $a_1$  и  $a_0$  определяются путем решения методом наименьших квадратов задачи

$$S = \sum_{i=1}^M (\check{Z}_{N-M+1-i} - Z_{N-M+1-i})^2 \rightarrow \min,$$

подробно рассмотренной в работе [11].

Таким образом, экстраполированные значения  $\check{Z}_{N+1}^P$  временного ряда  $Z_1^N$  определяются формулой

$$\check{Z}_{N+1}^P = a_1 \cdot Z_{k_{max}^*+M}^P + a_0 = \text{EMMSP}(M), \quad (5)$$

которая представляет собой модель экстраполяции временных рядов по выборке максимального подобия (extrapolation model on most similar pattern, EMMSP).

## 2. Идентификация модели экстраполяции

Задачу идентификации модели (5), то есть задачу нахождения параметра  $M$ , предлагается решать по следующей схеме.

- Исходный временной ряд  $Z_1^N$  разделим на три части в пропорции 2:2:1. Полученные части назовем *базовый период* (40%), *тестовый период* (40%) и *контрольный период* (20%) временного ряда соответственно.
- Исходя из решаемой задачи прогнозирования, определяем значение  $P$ , а также диапазон возможных значений параметра  $M$ . Первоначально рекомендуется брать широкий диапазон возможных значений  $M$ , например,  $M \in [P, 30 \cdot P]$ , а затем его последовательно уточнять.
- Для каждого значения параметра  $M$  из установленного диапазона, прогнозируем значения временного ряда внутри *тестового периода*.
- По результатам прогнозирования определяем среднюю абсолютную ошибку прогноза (mean absolute error,  $MAE$ )

$$MAE = \frac{1}{K} \cdot \sum_{i=1}^K |\tilde{z}_i - z_i|, \quad (6)$$

где  $K$  – количество значений временного ряда, попавших в *тестовый период*.

- Строим график зависимости  $MAE$  от  $M$  для *тестового периода* и определяем диапазон значений  $M$ , соответствующий устойчивому минимуму  $MAE$ .
- На последнем шаге экспертом выбирается параметр модели  $M$  внутри диапазона устойчивого минимума.

Рассмотрим в качестве примера зависимость  $MAE$  от  $M$  для временного ряда энергопотребления европейской территории РФ, представленного на рис. 2. Первоначально был выбран диапазон  $M \in [24, 360]$ , внутри которого выделены значения  $M \in [168, 240]$ . На последнем шаге экспертно было выбрано значение  $M = 216$ .

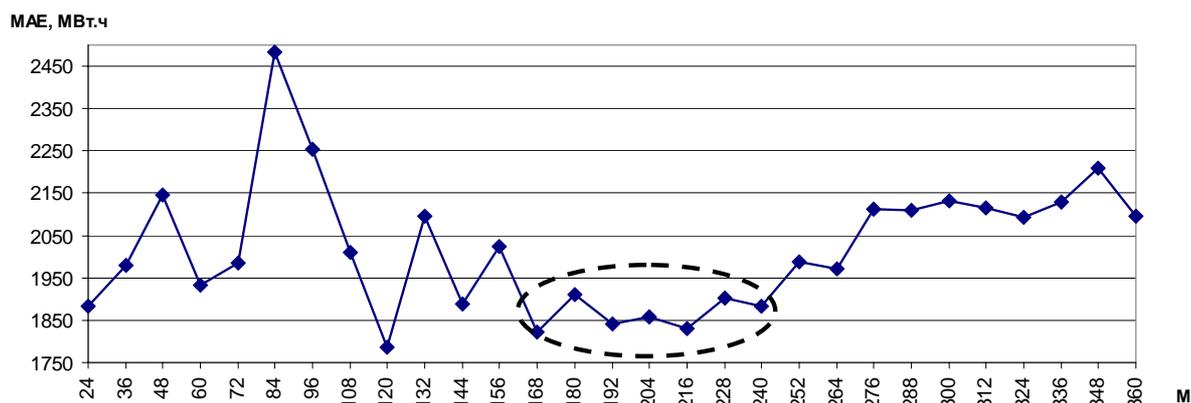


Рис. 2. Зависимость  $MAE$  от параметра модели  $M$ ,  $M \in [24,360]$

### 3. Эффективность модели экстраполяции

Исследования эффективности модели экстраполяции (4) выполнено на следующих временных рядах, предоставленных ОАО «АТЭС» [12]:

- энергопотребление европейской ценовой зоны (ЕЦЗ);
- энергопотребление сибирской ценовой зоны (СЦЗ);
- цена РСВ европейской ценовой зоны (ЕЦЗ);
- цена РСВ сибирской ценовой зоны (СЦЗ).

Каждый временной ряд содержит почасовые равноотстоящие значения за период с 01.09.2006 по 30.09.2009. Кроме экстраполяции почасовых значений указанных временных рядов на следующие сутки, ставится задача экстраполяции агрегированных суточных значений – суммарного энергопотребления за сутки и среднего значения цены РСВ за сутки. Параметры временных рядов представлены в таблице 1.

**Таблица 1. Параметры временных рядов**

Разре- шение	Временной ряд	Длина ряда	Среднее значение	Стандартное отклонение	Минимум	Максимум
Почасовое	Энергопотребление ЕЦЗ (МВт.ч)	27 023	81 441	10 640	57 847	110 586
	Энергопотребление СЦЗ (МВт.ч)	27 023	22 338	3 023	15 328	30 666
	Цена РСВ ЕЦЗ (руб./ МВт.ч)	27 023	610	194	0	1 559
	Цена РСВ СЦЗ (руб./ МВт.ч)	27 023	368	185	0	1 029
Суточное	Энергопотребление ЕЦЗ (МВт.ч)	1 126	1 954 502	218 255	1 519 197	2 438 008
	Энергопотребление СЦЗ (МВт.ч)	1 126	536 101	68 930	408 408	695 606
	Цена РСВ ЕЦЗ (руб./ МВт.ч)	1 126	610	144	221	1 204
	Цена РСВ СЦЗ (руб./ МВт.ч)	1 126	368	176	0	734

Задача экстраполяции временных рядов цен РСВ и энергопотребления, имеющих почасовое разрешение, заключается в нахождении 24 почасовых значений временного ряда на будущие сутки. Для рядов, имеющих суточное разрешение, задача экстраполяции состоит в нахождении одного значения на будущие сутки. Согласно работе [3] обе задачи экстраполяции относятся к категории краткосрочного прогнозирования.

В качестве *контрольного периода* прогнозирования использован период с 01.03.2009 по 30.09.2009 (семь месяцев, около 5000 значений для временных рядов, имеющих часовое разрешение; около 200 значений для рядов, имеющих суточное разрешение).

В таблицах 2, 3 приведены численные результаты точности прогнозирования, а также указаны параметры моделей  $M$  каждого временного ряда, полученные путем решения задачи идентификации по схеме раздела 2. Приведенные значения параметра  $M$  могут быть взяты за основу при работе с временными рядами энергопотребления и цен РСВ. Для временных рядов другой предметной области необходимо решать задачу идентификации заново.

Для сравнения точности экстраполяции кроме величины  $MAE$  оценивалась также средняя абсолютная ошибка в процентах (mean absolute percentage error, MAPE)

$$MAPE = \frac{1}{Q} \cdot \sum_{i=1}^Q \frac{|\tilde{z}_i - z_i|}{z_i} \cdot 100\% .$$

Здесь Q – количество значений временного ряда, попавших в *контрольный период*.

**3.1. Прогнозирование энергопотребления.** Результаты прогнозирования энергопотребления представлены в таблице 2.

Таблица 2. Результаты прогнозирования энергопотребления

Временной ряд	Разрешение	M	MAE, МВт.ч	MAPE, %	Время экстраполяции, час
Энергопотребление европейской ценовой зоны	Почасовое	216	1 347	<b>1.04</b>	1.80
	Суточное	6	19 736	<b>1.10</b>	0.15
Энергопотребление сибирской ценовой зоны	Почасовое	24	378	<b>1.86</b>	4.20
	Суточное	8	7 727	<b>1.44</b>	0.05

Полученные значения величины *MAPE* краткосрочного прогнозирования лежат в диапазоне от 1.04% до 1.86%.

Проведем сравнение достигнутой точности с точностью прогнозирования энергопотребления, указанной в других публикациях. В работе [3] описаны результаты краткосрочного прогнозирования энергопотребления Шанхайской энергосистемы (Shanghai Power Grid) и приведены значения *MAPE* различных моделей экстраполяции в диапазоне от 2.8% до 3.4% в зависимости от их особенностей. Разработчики специализированного программного пакета «Энергостат» в статье [7] приводят значения *MAPE* краткосрочного прогнозирования энергопотребления для регионов РФ за 2000 год, которые лежат в диапазоне от 1.21% до 3.32% (в зависимости от региона). Австрийская компания iRM (<http://www.irm.at>), производитель программного продукта iOPT PRO для прогнозирования энергопотребления,

указывает значения  $MAPE$  краткосрочного прогнозирования энергопотребления 1 – 2%.

Приведенные результаты краткосрочного прогнозирования энергопотребления показывают эффективность применения предложенной в работе модели экстраполяции временных рядов по выборке максимального подобия для решения задачи прогнозирования временных рядов энергопотребления.

**3.2. Прогнозирование цен РСВ.** Результаты прогнозирования цены РСВ представлены в таблице 3.

Из таблицы видно, что значения  $MAPE$  для цен РСВ сибирской ценовой зоны велики в связи с тем, что в *контрольном периоде* имеются нулевые значения цен РСВ, которые при анализе ошибок замещались на 0.01 (т.е. 1 коп./МВт.ч).

Таблица 3. Результаты прогнозирования цен РСВ

Временной ряд	Разрешение	M	MAE, руб./МВт.ч	MAPE, %	Время экстраполяции, час
Цена РСВ европейской ценовой зоны	Почасовое	360	49.54	<b>7.00</b>	2.10
	Суточное	16	34.50	<b>4.81</b>	0.04
Цена РСВ сибирской ценовой зоны	Почасовое	84	65.90	<b>39.78</b>	4.93
	Суточное	30	60.39	<b>32.93</b>	0.04

Выше отмечалось, что задача экстраполяции цен РСВ для энергорынка России является новой, потому нет возможности выполнить сравнение точности краткосрочного прогнозирования. В единственно доступном источнике [13] приведено значение  $MAE$  для краткосрочного прогнозирования цен РСВ, равное 55.09 руб./МВт.ч.

Для аналогичной задачи прогнозирования цен энергорынков Испании и Калифорнии в работе [5] указаны значения  $MAPE$ , равные 5 – 10% (в зависимости от месяца). Кроме того, в работе [5] указано, что данный

уровень точности прогнозирования является достаточным для эффективного финансового планирования на указанных рынках. Предложенная в данной работе модель экстраполяции позволила получить для европейской ценовой зоны значения *MAPE*, равные 4.8% и 7.0% (таблица 3). На этом основании можно утверждать, что разработанная модель экстраполяции эффективна для краткосрочного прогнозирования цен РСВ европейской ценовой зоны. В тоже время, из таблицы 3 видно, что задача экстраполяции цен РСВ сибирской ценовой зоны требует дополнительных исследований и доработок модели.

## Заключение

В работе предложена модель экстраполяции временных рядов по выборке максимального подобия. Основные особенности модели экстраполяции заключаются в следующем.

- Модель экстраполяции по выборке максимального подобия решает задачи краткосрочного прогнозирования (от 1 до 50 значений вперед).
- Модель не накладывает ограничений ни на непрерывность, ни на стационарность временного ряда.
- Для работы с моделью желательно, чтобы временной ряд содержал не менее 1000 отсчетов.
- Модель может быть расширена с целью решения задачи экстраполяции временного ряда с учетом влияния внешних факторов, представленных в виде других временных рядов (например, учет температуры при прогнозировании энергопотребления).

Представленные в работе результаты подтверждают эффективность использования модели для прогнозирования исследуемых временных рядов цен РСВ и энергопотребления. В развитие работы планируется оценить

эффективность модели экстраполяции для задач прогнозирования временных рядов из других предметных областей.

Автор выражает благодарность своему научному руководителю д.т.н., профессору Ю.Н. Павлову за помощь в обсуждении модели, постановке задачи для экспериментов, а также подготовке статьи.

## Литература

1. Бокс Дж., Дженкинс Г.М. Анализ временных рядов, прогноз и управление. – М.: Мир, 1974. – С. 406.
2. Тихонов Э.Е. Прогнозирование в условиях рынка. – Невинномысск, 2006. – С. 221.
3. M. Sc. Jingfei Yang. Power System Short-term Load Forecasting – Elektrotechnik und Informationstechnik der Technischen Universität Darmstadt. – 2006. – С. 81 – 84.
4. Extrapolation [Электронный ресурс] // The free encyclopedia «Wikipedia» 2001. (<http://en.wikipedia.org/wiki/Extrapolation>).
5. Reinaldo C. Garcia A. GARCH Forecasting Model to Predict Day-Ahead Electricity Prices. – German Institute of Economic Research, Germany. – 2003. – С. 9 – 12.
6. Макоклюев Б.И., Ёч В. Ф. Взаимосвязь точности прогнозирования и неравномерности графиков электропотребления // Энергетик. – 2005 г. – № 5. – С .20.
7. Макоклюев Б.И., Владимиров А.И., Фефелова Г.И. Прогнозирование потребления электроэнергии АО Мосэнерго // Энергетик. – 2001 г. – № 4. – С .12.
8. Badran Saeed M. Short term Electrical Load Forecasting // Australian Journal of Basic and Applied Sciences. – 2009. – №1. – С. 9.

9. Kumar M. Short-term load forecasting using artificial neural network techniques. – Department of Electrical Engineering National Institute of Technology Rourkela, 2009. – С. 48.
10. Al-Shareef A. J., Mohamed E. A., Al-Judaibi E. One hour ahead load forecasting using artificial neural network for the Western Area of Saudi Arabia. – World Academy of Science, Engineering and Technology, 2008. – С. 6.
11. Draper N. R., Smith H. Applied regression analysis. – New York: Wiley, In press, 1981. – С. 693.
12. Открытое акционерное общество «Администратор торговой системы оптового рынка электроэнергии и мощности». (<http://www.atsenergo.ru/>).
13. BI EnergoPrice. «Прогнозирование цен на электроэнергию». ([http://www.bi-grouplabs.ru/Rech/electricity/BI\\_EnergoPrice.html](http://www.bi-grouplabs.ru/Rech/electricity/BI_EnergoPrice.html)).