

Оригинальное теоретическое исследование
УДК 620.9; 005.1
DOI: 10.46418/3033-7852_2026_2_2_49-58

МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ РЕГРЕССИОННОГО И ФАКТОРНОГО АНАЛИЗА ПРИ ЭКСПЛУАТАЦИИ ТЕПЛОВЫХ ЭЛЕКТРОСТАНЦИЙ

Александр Дмитриевич Ширяев, Владимир Германович Злобин

Высшая школа технологии и энергетики Санкт-Петербургского государственного университета промышленных технологий
и дизайна, Санкт-Петербург, Россия

© Ширяев А. Д., Злобин В. Г., 2026

Автор-корреспондент: А. Д. Ширяев, e-mail: a. d. shiryayev@rambler.ru

АННОТАЦИЯ

Развитие систем теплоснабжения и электроснабжения промышленных предприятий и потребителей коммунально-бытовой сферы осуществляется путем максимального использования возможностей ТЭС, а также совершенствования тепловых схем паротурбинных установок (ПТУ) и внедрения современных конструкций паровых и газовых турбин различной мощности. В последние годы разработан и освоен целый ряд высокоэкономичных паровых и газовых турбин. Технически грамотная эксплуатация таких тепловых электростанций возможна при постоянном поиске способов повышения эффективности управляемого процесса при использовании их оборудования. Поэтому актуальными становятся вопросы получения математических моделей зависимостей показателей эффективности тепловых электростанций от различных факторов, влияющих на их эксплуатацию [1]. Целью данного исследования является разработка рекомендаций по применению регрессионного и факторного анализа в системе эксплуатации тепловых электростанций на базе паровых и газовых турбин. Предметом исследования были выбраны зависимости показателей эффективности ТЭС и надежности основного оборудования от различных факторов, а также возможности применения методов регрессионного и корреляционного анализа для анализа показателей эффективности эксплуатации ТЭС. Научная новизна работы заключается в обосновании возможности применения корреляционно-регрессионного анализа для получения эмпирических зависимостей и использования полученных зависимостей в цифровых моделях энергетического оборудования. Исследуются вопросы оценки достоверности моделей при влиянии на них различных (в том числе случайных) факторов, и варианты проверки на адекватность по различным критериям. Практическая значимость заключается в предоставлении персоналу ТЭС математического аппарата, позволяющего получать цифровые модели процесса эксплуатации и правильно управлять этим процессом. Полученные выводы: управление такой сложной системой, как ТЭС, будет эффективным только при переходе на цифровое управление при построении цифрового прообраза энергетического оборудования с учетом влияния всех факторов на показатели работы.

Ключевые слова: системный анализ, случайная величина, эффективность управляемого процесса, регрессионный анализ, адекватность уравнений регрессии, коэффициент детерминации, математическая модель, цифровизация управления

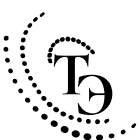
Для цитирования: Ширяев А. Д., Злобин В. Г. Методология применения регрессионного и факторного анализа при эксплуатации тепловых электростанций // Технология и энергетика. 2026. № 2 (2). С. 49–58. DOI: 10.46418/3033-7852_2026_2_2_49-58

METHODOLOGY OF REGRESSION AND FACTOR ANALYSIS APPLICATION FOR THERMAL POWER PLANT OPERATION

Alexander Dmitrievich Shiryayev, Vladimir Germanovich Zlobin

Higher School of Technology and Energy of Saint Petersburg State University of Industrial Technologies and Design, St. Petersburg, Russia

Corresponding author: A. D. Shiryayev, e-mail: a. d. shiryayev@rambler.ru



ABSTRACT

The advancement of heating and power supply systems for industrial facilities and residential utilities is achieved through the maximum utilization of thermal power plant (TPP) potential, along with the optimization of thermal cycle configurations for steam turbine units (STU) and the implementation of state-of-the-art steam and gas turbines across various power ratings. In recent years, a number of highly efficient steam and gas turbines have been developed and introduced into service. Technically competent operation of such thermal power plants requires a continuous search for methods to enhance the efficiency of the controlled process in the course of equipment utilization. Therefore, the development of mathematical models describing the dependence of thermal power plant performance indicators on various factors influencing their operation becomes a relevant task [1]. This study aims to develop recommendations for applying regression and factor analysis to the operation system of steam and gas turbine-based thermal power plants. This study focused on the relationships between TPP performance indicators, main equipment reliability, and various influencing factors, as well as the applicability of regression and correlation analysis methods for evaluating TPP operational efficiency. The scientific novelty of this research consists of validating the applicability of correlation-regression analysis for deriving empirical relationships, as well as the utilization of these relationships in digital models of power generation equipment. This paper addresses the challenges of assessing model reliability when exposed to various influencing factors (including random ones), as well as the options for adequacy checking employing diverse statistical criteria. The practical significance of this work consists of equipping thermal power plant staff with a mathematical framework that allows for obtaining digital models of the operational process and effectively controlling this process. Conclusions: effective management of a complex system like a TPP requires a transition to digital control through the development of a digital twin of the power equipment, accounting for all factors affecting performance.

Keywords: system analysis, random variable, efficiency of the controlled process, regression analysis, adequacy of regression equations, coefficient of determination, mathematical model, digitalization of management

For citation: Shiryayev A. D., Zlobin V. G. (2026) 'Methodology of regression and factor analysis application for thermal power plant operation', *Technology and Energy*, no. 2 (2), pp. 49–58. DOI: 10.46418/3033-7854_2026_2_2_49–58

ВВЕДЕНИЕ

Большинство задач по эксплуатации тепловых электростанций (ТЭС) требуют получения математических моделей, представляющих собой зависимость показателя, который выбран для оценки качества процесса эксплуатации, от других показателей, влияющих на него по физической сущности процесса или по предположению. Особенно это стало важно и актуально, когда в теплоэнергетике начался процесс создания цифровых двойников энергетического оборудования. Например, существует очевидная зависимость показателей безотказности от срока службы, количества пусков, режимов использования, качества технического обслуживания и ремонта и т. п. Но для того чтобы получить количественные оценки, необходимо описать эту связь математически, создать математическую модель [2].

Целью данного исследования является разработка рекомендаций по применению регрессионного и факторного анализа в системе эксплуатации тепловых электростанций на базе паровых турбин для построения математической модели, позволяющей определить влияние различных, в том числе случайных, факторов для достижения заданного уровня эффективной эксплуатации ТЭС.

Связь между различными показателями, характеризующими процесс эксплуатации ТЭС, и возможные

изменения самих показателей зависят от влияния многочисленных общих для них случайных факторов. Поэтому исследовать приходится систему случайных величин, каждая из которых представлена своей статистической выборкой и характеризуется полученными на их основе точечными оценками и законами распределения. При этом строгая функциональная зависимость между отдельными показателями нарушается, а её характеристика по данным отдельных наблюдений носит случайный (стохастический) характер. При стохастической связи одна случайная величина реагирует на изменение другой изменением своего закона распределения [3]. Поэтому предметом исследования стала стохастическая зависимость показателей эффективности ТЭС и основного оборудования от различных, в том числе случайных, факторов, а также возможность применения методов регрессионного и корреляционного анализов для оценки показателей эффективности эксплуатации ТЭС.

В статье последовательно решаются следующие задачи:

- выявляется связь между случайными величинами; исследуются их характеристики и форма связи, которая определяется функцией регрессии;
- определяется теоретическая линия регрессии, которая заключается в подборе и обосновании типа кривой и расчете коэффициентов её уравнения так,



чтобы она наилучшим образом отражала статистические данные;

- определяется вид уравнения регрессии, который заключается, как правило, в подборе известных моделей однофакторной или множественной корреляции;
- определяются коэффициенты уравнения регрессии методом максимального правдоподобия или методом наименьших квадратов;
- после построения регрессионной модели производится проверка её адекватности и значимости коэффициентов уравнения регрессии по критерию *Стьюдента*;

– рассматривается вопрос построения уравнений регрессии при наличии многих факторов в матричном виде.

Научная новизна данной работы заключается в обосновании возможности применения корреляционно-регрессионного анализа для получения эмпирических зависимостей от одного или множества факторов, в том числе случайных, и использовании полученных зависимостей в цифровых моделях энергетического оборудования.

ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Объектами исследования являются массивы данных, полученных в результате эксплуатации энергетического оборудования тепловых электростанций, и факторы, оказывающие непосредственное влияние на параметры данного массива. К параметрам массива относятся, например, мощностные и экономические показатели, показатели безотказности и срока службы, количества пусков и режимов использования, качества технического обслуживания и ремонта и т. п. К факторам, оказывающим влияние на массив данных, можно отнести следующие: температуру, давление и влажность наружного воздуха, качество топлива (основного — природного газа и резервного — мазута), изменение нагрузки потребителей и т. п.

Для оценки влияния факторов на параметры эксплуатации энергетического оборудования ТЭС предлагается применить методы обработки математической статистики, в том числе регрессионный и факторный анализ.

1. СВЯЗЬ МЕЖДУ СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ И ЕЁ ХАРАКТЕРИСТИКА

Для характеристики связи между случайными величинами, к которым в теплоэнергетике можно отнести внешние условия, качество топлива, тепловую и электрическую нагрузку потребителей, скачки нагрузки, суточную и сезонную неравномерность и т. д., используют условное математическое ожидание. Для дискретной случайной величины Y условное математическое ожидание равно сумме произведений возможных её значений на их условные вероятности:

$$M(Y / X = x) = \sum_{i=1}^m y_i P(y_i / x). \quad (1)$$

Для непрерывной случайной величины:

$$M(Y / X = x) = \int_{-\infty}^{\infty} y \varphi(y / x) dy, \quad (2)$$

где $\varphi(y/x)$ — условная плотность случайной величины Y при условии, что независимая случайная величина приняла значение $X=x$.

Условное математическое ожидание одной случайной величины является функцией другой случайной величины [4]:

$$M(Y/x) = f(x); M(X/y) = f(y). \quad (3)$$

Зависимость между одной случайной переменной и условным средним значением другой случайной переменной называется корреляционной зависимостью. А условное математическое ожидание случайной величины Y , рассмотренное как функция другой случайной величины X , дает кривую, которая называется линией регрессии.

Считают, что показатель Y находится в корреляционной зависимости от показателя X , если:

- каждому значению аргумента x соответствует ряд распределения функции y , например, каждому значению температуры внешней среды соответствует мощность энергоблока при сохранении постоянными других факторов;
- теоретическая линия регрессии y по x не параллельна оси абсцисс.

Любая корреляционная зависимость характеризуется формой и теснотой связи. Первой задачей при исследовании корреляционной зависимости, в том числе и в теплоэнергетике, является установление формы связи, которая определяется функцией регрессии. Этот этап и называется собственно регрессионным анализом. Он заключается в аналитическом описании функции регрессии. Второй задачей исследования корреляционной зависимости является измерение тесноты связи между коррелированными случайными величинами. Этот этап называется корреляционным анализом. Иногда в литературе по статистическому анализу оба этапа объединяют и называют либо корреляционным анализом, либо регрессионным анализом, либо корреляционно-регрессионным анализом [5].



2. ОПРЕДЕЛЕНИЕ ФОРМЫ СВЯЗИ МЕЖДУ СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ

Предположение о виде функции регрессии делается на основе расположения точек на поле корреляции, которое показано на рисунке 1. Границы поля корреляции определяются размахом варьирования случайных величин. На оси координат наносят границы разрядов случайных величин. На поле указывают все значения случайной величины Y , полученные при определенных значениях, которые принимает независимая случайная величина X в процессе испытаний.

По полю корреляции строится корреляционная таблица, которой удобно пользоваться при расчетах, необходимых при регрессионном анализе. Корреляционная таблица, соответствующая корреляционному полю рисунка, приведена в таблице 1.

Если вычислить среднее значение случайной величины Y в каждом интервале случайной величины X , нанести эти точки на поле корреляции и соединить их между собой, то вид полученной ломаной линии показывает, как в среднем изменяется Y (например, давление в конденсаторе) в зависимости от изменения X (теплоты отведенной потребителю). Эта ломаная линия называется эмпирической линией регрессии.

Эмпирическую линию регрессии можно аппроксимировать (сгладить) функцией регрессии. В качестве функции регрессии выбирают:

- линейную модель;
- квадратичный полином;
- показательную функцию и т. п.

Таблица 1. Корреляционная таблица

Table 1. Correlation table

Границы и середина разрядов Y	Границы и середина разрядов X								Частота разряд., m_Y
	1–2 1,5	2–3 2,5	3–4 3,5	4–5 4,5	5–6 5,5	6–7 6,5	7–8 7,5	8–9 8,5	
10–20 15	4	5							9
20–30 25	1	3	1						5
30–40 35	2	3	6	5	3	1			20
40–50 45		5	9	19	8	7	2	1	51
50–60 55		1	2	7	16	9	4	2	41
60–70 65			1	5	6	4	2	2	20
70–80 75							1	3	4
Число значений Y в инт. X , m_X	7	17	19	36	33	21	9	8	150

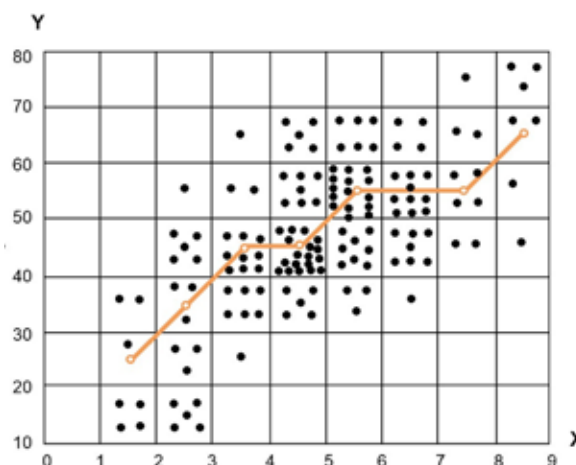


Рисунок 1. Поле корреляции

Figure 1. Correlation field

Функция регрессии представляет собой геометрическое место условных математических ожиданий одной из случайных величин, соответствующих заданным значениям другой случайной величины:

$$M(Y/X=x)=y(x) \vee M(X/r=y)=x(y). \quad (4)$$

Процесс определения теоретической линии регрессии заключается в подборе и обосновании типа кривой и в расчете коэффициентов её уравнения так, чтобы она наилучшим образом отражала статистические данные.

Для определения функции регрессии используют уравнение $y(x)=f(x)$, которое называется уравнением регрессии.

Определение вида уравнения регрессии заключается, как правило, в подборе известных моделей:

$$\hat{y}(x) = f(x) = b_0 + b_1 x \text{ — линейной,} \quad (5)$$

$$\left. \begin{aligned} \hat{y}(x) &= f(x) = b_0 + b_1 x + b_2 x^2 \\ \hat{y}(x) &= f(x) = b_0 + b_1 x + b_2 x^2 + \dots + b_k x^k \end{aligned} \right\} \text{ — нелинейной.} \quad (6)$$

Если исследуется связь между несколькими показателями (например, изменение тепловой и электрической нагрузки потребителей), то корреляцию называют множественной $y(x) = f(x_1, x_2, \dots, x_k)$. Аппроксимация эмпирической линии регрессии при множественной корреляции наиболее часто осуществляется с помощью следующих моделей:

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i; \quad (7)$$

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i \neq j} b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2; \quad (8)$$

$$\hat{y} = b_0 \prod_{i=1}^k b_i^{x_i}; \quad (9)$$

$$\hat{y} = b_0 \prod_{i=1}^k x_i^{b_i}. \quad (10)$$

Для определения коэффициентов уравнения регрессии используется несколько методов. Наиболее распространены метод максимального правдоподобия и метод наименьших квадратов.

Метод максимального правдоподобия заключается в максимизации общей вероятности попадания экспериментальных точек на сглаживающую кривую. Для этого вводят функцию правдоподобия:

$$L = \sum_{i=1}^n \ln f(x_i). \quad (11)$$

Эта функция имеет максимальное значение (экстремум) при тех же параметрах, что и общая вероятность попадания экспериментальных точек на сглаживающую кривую. Параметры сглаживающей кривой определяются из условия максимизации функции правдоподобия:

$$\left\{ \begin{aligned} \frac{\partial L}{\partial b_0} &= 0 \\ \frac{\partial L}{\partial b_i} &= 0, \quad i \in \overline{1, k} \end{aligned} \right.,$$

где b_0, b_i — параметры сглаживающей кривой.

Метод наименьших квадратов (МНК) сводится к обеспечению выполнения требования: сумма квадратов отклонений экспериментальных точек от сглаживающей кривой должна быть минимальной:

$$U = \sum_{i=1}^n [y_i - f(x_i)]^2 = \min. \quad (12)$$

Определим коэффициенты уравнения регрессии по методу наименьших квадратов:

$$\hat{y} = f(x, b_0, b_1, \dots, b_n).$$

Параметры $b_0, b_1, \dots, b_n$ должны быть вычислены таким образом, чтобы выполнялось условие:

$$U = \sum_{i=1}^n [y_i - f(x_i, b_0, b_1, \dots, b_n)]^2 = \min.$$

Для отыскания экстремума этого функционала продифференцируем U по параметрам и приравняем производные к нулю:

$$\left\{ \begin{aligned} \frac{\partial U}{\partial b_0} &= 0 \left[\sum_{i=1}^n [y_i - f(x_i, b_0, b_1, \dots, b_n)] \left(\frac{\partial f(x_i)}{\partial b_0} \right)_i = 0 \right. \\ \frac{\partial U}{\partial b_1} &= 0 \left[\sum_{i=1}^n [y_i - f(x_i, b_0, b_1, \dots, b_n)] \left(\frac{\partial f(x_i)}{\partial b_1} \right)_i = 0 \right. \\ &\dots \dots \dots \\ \frac{\partial U}{\partial b_n} &= 0 \left[\sum_{i=1}^n [y_i - f(x_i, b_0, b_1, \dots, b_n)] \left(\frac{\partial f(x_i)}{\partial b_n} \right)_i = 0 \right] \end{aligned} \right. \quad (13)$$

где $\left(\frac{\partial f(x_i)}{\partial b_0} \right)_i, \left(\frac{\partial f(x_i)}{\partial b_1} \right)_i, \dots, \left(\frac{\partial f(x_i)}{\partial b_n} \right)_i$ — значения частной производной функции $f(x, b_0, b_1, \dots, b_n)$ по соответствующему параметру в точке x_i .

Число уравнений системы (13) равно количеству неизвестных параметров, поэтому она решается известными методами. Рассмотрим для примера её решение, когда уравнение регрессии является линейной функцией $f(x) = b_0 + b_1 x$. В этом случае:

$$\left(\frac{\partial f}{\partial b_0} \right)_i = 1; \quad \left(\frac{\partial f}{\partial b_1} \right)_i = x_i.$$

Подставляя уравнение функции $f(x)$ и значения частных производных в систему (13), получим:

$$\left\{ \begin{aligned} \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] \cdot 1 &= 0 \\ \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] \cdot x_i &= 0 \end{aligned} \right. \quad (14)$$

Решение системы уравнений (14) дает статистические оценки коэффициентов b_0 и b_1 по известным выборкам X и Y :

$$\left\{ \begin{aligned} \sum_{i=1}^n y_i - \sum_{i=1}^n (b_0 + b_1 x_i) &= 0 \\ \left(\sum_{i=1}^n y_i \right) x_i - \sum_{i=1}^n (b_0 + b_1 x_i) \cdot x_i &= 0 \end{aligned} \right. \quad \left\{ \begin{aligned} nb_0 - b_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \end{aligned} \right.$$



$$\text{Из второго уравнения: } b_0 = \frac{\sum_{i=1}^n x_i y_i - b_1 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i}.$$

Подставив это выражение в первое уравнение, получим:

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}. \quad (15)$$

Вычислив b_1 , легко найти b_0 из первого уравнения:

$$b_0 = \frac{1}{n} \sum_{i=1}^n y_i - b_1 \frac{1}{n} \sum_{i=1}^n x_i = \bar{y}^* - b_1 \bar{x}^*. \quad (16)$$

Характерная особенность МНК состоит в том, что для его использования необходим большой объем статистических данных на участке наблюдения в системе эксплуатации тепловой электростанции; без этого МНК дает существенные погрешности [6]. Объем выборки должен быть таким, чтобы на каждый входящий в модель аргумент приходилось несколько точек. В частности, для линейной модели $n_b \geq 3k + 1$.

Коэффициент уравнения регрессии b_i — это мера, которая в среднем показывает влияние изменения объясняющей переменной x_i на зависимую переменную y . Постоянная b_0 в уравнении регрессии является выравнивающей.

После построения регрессионной модели производят проверку её адекватности и значимости отдельных коэффициентов. Статистические выборки из одной и той же генеральной совокупности имеют разное рассеяние. Поэтому может случиться, что построение регрессионной модели одного и того же процесса по данным разных выборок дает разную степень соответствия модели опытным данным, несмотря на то, что тип модели выбирается один и тот же.

Для проверки адекватности регрессионной модели используется несколько статистических критериев, основанных на дисперсионном анализе статистических данных и широко применяемых в оценке реальных показателей работы энергообъектов [8, 9].

Общая дисперсия случайной переменной y , которая характеризует разброс её отдельных наблюдаемых значений около её среднего значения:

$$S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}^*)^2, \quad (17)$$

раскладывается на две её составляющие:

$$S_y^2 = S_y^2 + S_{ост}^2, \quad (18)$$

где S_y^2 — дисперсия, обусловленная регрессией; она представляет ту часть рассеяния переменной y , которая обусловлена влиянием переменных $x_k, k \in \overline{1, m}$:

$$S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}^*)^2, \quad (19)$$

где $S_{ост}^2$ — остаточная дисперсия; она измеряет ту часть рассеяния y , которая возникает из-за влияния прочих неучтенных факторов (l — число коэффициентов уравнения регрессии).

$$S_{ост}^2 = \frac{1}{n-l} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (20)$$

Чем больше $S_{ост}^2$ приближается к нулю, тем меньше эмпирические значения y_i отклоняются от значений регрессии $\hat{y}_i$.

Проверка адекватности регрессионной модели по коэффициенту детерминации

$$B_{yx} = \frac{S_y^2}{S_y^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}^*)^2}{\sum_{i=1}^n (y_i - \bar{y}^*)^2}. \quad (21)$$

Чем больше S_y^2 , тем больше S_y^2 формируется за счет влияния объясняющей переменной x , и, следовательно, связь между переменными y и x более интенсивная.

$$0 \leq B_{yx} \leq 1.$$

Чем большую долю в общей дисперсии составляет S_y^2 , тем лучше регрессионная модель соответствует эмпирическим данным.

$B_{yx} = 1$, если $\hat{y}_i = y_i$; $B_{yx} = 0$, если $\hat{y}_i = \bar{y}^*$, т. е. величина y — детерминированная.

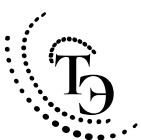
Проверка адекватности регрессионной модели по критерию Фишера

$$F_H = \frac{S_y^2}{S_{ост}^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}^*)^2}{\frac{1}{n-l} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (22)$$

где F_H — наблюдаемое (рассчитанное) значение критерия Фишера.

Критерий Фишера показывает, во сколько раз рассеяние относительно уравнения регрессии меньше рассеяния относительно среднего значения.

F_H сравнивается с табличным значением F_{α, f_1, f_2} , выбранным по таблицам математической статистики



по уровню значимости $\alpha = 0,01 \div 0,05$ и числам степеней свободы $f_1 = n - 1; f_2 = n - l$. Чем больше F_H превышает F_{α, f_1, f_2} , тем эффективнее уравнение регрессии.

Проверка адекватности уравнения регрессии по критерию Фишера — Снедекора

$$F_H = \frac{\bar{y}^*}{S_{ост}^2} = \frac{\frac{1}{n} \sum_{i=1}^n y_i}{\frac{1}{n-l} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (23)$$

F_H сравнивается с табличным значением F_{α, f_1, f_2} , выбранным по таблицам математической статистики по уровню значимости $\alpha = 0,01 \div 0,05$ и числам степеней свободы $f_1 = n; f_2 = n - l$. При $F_H > F_{\alpha, f_1, f_2}$ регрессионную модель считают адекватной.

Проверка значимости коэффициентов уравнения регрессии производится по критерию Стьюдента.

Наблюдаемое значение критерия Стьюдента для каждого коэффициента уравнения:

$$t_j = \frac{|b_j|}{S_{b_j}}, \quad (24)$$

где $S_{b_j}^2 = \frac{S_y^2}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y}^*)^2$ — дисперсия коэффициента,

сравнивается с табличным $t_{\alpha, f}$, выбранным по таблицам математической статистики для уровня значимости $\alpha = 0,05 \div 0,1$ и числа степеней свободы $f = n - 1$.

Если $t_j < t_{\alpha, f}$, то коэффициент b_j незначим.

Иногда определяют величину доверительного интервала для коэффициентов уравнения регрессии:

$$\Delta b_j = \pm t_{\alpha, f} S_{b_j} = \pm t_{\alpha, f} \frac{S_y}{\sqrt{n}}. \quad (25)$$

Значимым считается коэффициент, для которого $|b_j| \geq \Delta b_j$.

При исследованиях множественной корреляции (при числе факторов более двух) регрессионный анализ выполняют в матричной форме.

3. РЕГРЕССИОННЫЙ АНАЛИЗ В МАТРИЧНОЙ ФОРМЕ

Пусть уравнение регрессии:

$$\hat{y} = b_0 x_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k,$$

где x_0 — фиктивная независимая переменная, равная 1.

Исходные данные для получения коэффициентов уравнения регрессии можно представить матрицами:

$$X = \begin{pmatrix} x_{01} & x_{11} & \dots & x_{k1} \\ x_{02} & x_{12} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ x_{0n} & x_{1n} & \dots & x_{kn} \end{pmatrix}; Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Требуется определить матрицу:

$$B = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_k \end{pmatrix}.$$

Коэффициенты уравнения регрессии определяют методом наименьших квадратов из условия:

$$U = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min.$$

Система уравнений для определения коэффициентов имеет вид:

$$b_0 \sum_{i=1}^n x_{0i}^2 + b_1 \sum_{i=1}^n x_{0i} x_{1i} + \dots + b_k \sum_{i=1}^n x_{0i} x_{ki} = \sum_{i=1}^n x_{0i} y_i$$

$$b_0 \sum_{i=1}^n x_{1i} x_{0i} + b_1 \sum_{i=1}^n x_{1i}^2 + \dots + b_k \sum_{i=1}^n x_{1i} x_{ki} = \sum_{i=1}^n x_{1i} y_i$$

.....

$$b_0 \sum_{i=1}^n x_{ki} x_{0i} + b_1 \sum_{i=1}^n x_{ki} x_{1i} + \dots + b_k \sum_{i=1}^n x_{ki}^2 = \sum_{i=1}^n x_{ki} y_i.$$

В матричной форме эта система запишется:

$$X^T X B = X^T Y,$$

$$\text{где } X^T = \begin{pmatrix} x_{01} & x_{02} & \dots & x_{0n} \\ x_{11} & x_{12} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{kn} \end{pmatrix},$$

Тогда

$$B = (X^T X)^{-1} \cdot X^T Y. \quad (26)$$

Аналогично можно определять коэффициенты более сложных регрессионных моделей, учитывающих эффекты взаимодействия факторов и квадратичные эффекты. Для этого надо расширить матрицу X , включив в неё столбцы взаимодействия и квадратичные эффекты. Такое расширение матрицы X называется линеаризацией. Она возможна только тогда, когда все коэффициенты входят в уравнение линейно, т. е. уравнение является линейным по параметрам.

Соответственно, в векторе B при этом появятся элементы, соответствующие эффектам взаимодействия и квадратичным. Для реализации матричного регрессионного анализа необходимо, чтобы матрица $(X^T X)$ была невырожденной, т. е. её определитель не должен быть



равен нулю. В противном случае она не будет иметь обратной матрицы. Для этого необходимо, чтобы переменные $x_1, x_2, \dots, x_k$ были линейно независимы.

Матрица $(X^T X)^{-1}$ называется ковариационной матрицей. Она недиагональна, следовательно, все коэффициенты регрессии взаимосвязаны, и нельзя проверить значимость каждого из них в отдельности. Поэтому

отношение $t_j = \frac{|b_j|}{S_{b_j}}$ можно рассматривать только как

средство ранжировки факторов. Можно исключать наименее значимые факторы последовательно до тех пор, пока уменьшается остаточная дисперсия.

Таким образом, при решении задач исследования зависимости любой характеристики, по которой оценивается эффективность управляемого процесса, от воздействующих на неё факторов всегда возникает проблема отбора факторов. Основными требованиями к ним являются независимость и значимость. В рассмотрение должны приниматься только независимые и значимые факторы, влияющие на исследуемую характеристику.

Так, при исследовании экономичности ТЭС в процессе эксплуатации обычно учитывают влияние факторов внешней среды, режимов использования, техническое состояние. При оценках трудоемкости технического обслуживания и ремонта ТЭС учитывают наработку оборудования, его техническое состояние, поломки и аварии, продолжительность периода от момента последнего технического осмотра и планового ремонта, срок службы оборудования ТЭС. Решению таких задач всегда предшествует предварительный анализ, основным содержанием которого является выбор и обоснование факторов, влияющих на процесс. Очень часто здесь руководствуются соображениями здравого смысла, опыта, сложившихся подходов в прошлом. Однако при этом всегда есть опасность упустить значимые факторы, и тогда модель будет неадекватной процессу, или получить громоздкую, сложную для расчетов модель, в которой большинство факторов оказывают незначительное влияние на исследуемую характеристику.

Примерами применения регрессионного и корреляционного анализа при эксплуатации тепловых электростанций могут служить аналитические характеристики

Таблица 2. Энергетические характеристики современных теплофикационных паровых турбин

Table 2. Energy performance of modern heating steam turbines

Типы турбин	Энергетические характеристики
T-105/120-130	$Q_{т.г} = 122,11 \cdot p_t + 2,326 \cdot N_s - 1,314 \cdot N_t + Q_i; N_m = \frac{0,546 \cdot (Q_m - 15,12)}{10 \cdot p_m}.$ При работе со встроенным пучком $N_m = \frac{0,602 \cdot (Q'_m - 134,9)}{10 \cdot p_m^{0,14}}, \text{ где } Q'_m = Q_m + Q_{пуч}$
T-175/210-130*	$Q_{т.г} = 226,8 \cdot p_t + 2,326 \cdot N_s - 1,302 \cdot N_t + Q_i; N_m = \frac{0,602 \cdot (Q_m - 24,42)}{10 \cdot p_m^{0,14}}$
T-250/300-240	При работе по электрическому графику $Q_{м.г} = 290,75 \cdot p_m + 1,977 \cdot N_s + 1,05 \cdot Q_m - \frac{0,65(Q_m - 40,7)}{(10 \cdot p_m^{0,14})} - 9,304 + \left[\frac{0,65}{(10 \cdot p_m^{0,07})} + 0,48 \right] \cdot Q_{с.н}.$ При работе по тепловому графику $Q_{м.г} = 290,75 \cdot p_m + Q_m + \frac{0,72 \cdot (Q_m - 40,7)}{(10 \cdot p_m^{0,07})} + \left[1,63 - \frac{0,72}{(10 \cdot p_m^{0,07})} \right] \cdot Q_{с.н};$ $N_m = \frac{0,709(Q_m - 140,7)}{10 \cdot p_m^{0,07}} - N_{м.н} + \left(0,619 - \frac{0,709}{10 \cdot p_m^{0,07}} \right),$ где $N_{м.н} = 0,0232 \cdot Q_{м.г} - 4,7$
ПТ-60/75-130/13*	$Q_{м.г} = 16,3 + 2,33 \cdot N_s - 1,314 \cdot N_m + Q_m + Q_n;$ $N_m = 0,529 \cdot Q_m \cdot (0,12/p_m)^{0,14} + 0,305 \cdot (1,3/p_n)^{0,34} \cdot Q_n - (9,9 - 0,048 \cdot Q_m)$
ПТ-80/100-130/13*	$Q_{м.г} = 16,3 + 1,98 \cdot N_s - 0,965 \cdot N_m + Q_m + Q_n;$ $N_m = 0,542 \cdot Q_m \cdot (10/p_m)^{0,14} + 0,301 \cdot (1,3/p_n)^{0,34} \cdot Q_n - (11,6 - 0,217 \cdot Q_m)$
ПТ-135/165-130/15	$Q_{м.г} = 20,9 + 1,95 \cdot N_s - 0,942 \cdot N_m + Q_m + Q_n;$ $N_m = 0,542 \cdot Q_m \cdot (0,12/p_m)^{0,14} + 0,284 \cdot (1,5/p_n)^{0,34} \cdot Q_n - (21 - 0,0553 \cdot Q_m)$
P-50-130/13	$Q_{м.г} = 1,16 + 1,01 \cdot N_m + Q_n; N_m = 0,301 \cdot (Q_n - 48,3) \cdot (1,3/p_n)^{0,55}$
P-100-130/15	$Q_{м.г} = 6,4 + 1,01 \cdot N_m + Q_n; N_m = 0,301 \cdot (Q_n - 98,9) \cdot (1,5/p_n)^{0,55}$



теплофикационных паровых турбин, представленных в таблице 2 [12]. Что позволяет, наряду с термодинамическими параметрами, применять в теплоэнергетике математические модели тепловых схем ТЭС.

ЗАКЛЮЧЕНИЕ

По результатам проведенного исследования можно сформулировать следующие выводы.

Во-первых, процесс эксплуатации ТЭС с позиций системного анализа включает в себя ряд составляющих его этапов: подготовка к использованию, использование по назначению, техническое обслуживание различных видов, плановые ремонты оборудования, консервация, утилизация.

Во-вторых, поскольку целью деятельности инженерной службы ТЭС является обеспечение их технической готовности, то необходимо управлять всеми этими этапами. Это непростая задача, так как различные стадии процесса эксплуатации могут иметь место одновременно на разных агрегатах станции.

В-третьих, понимая сложность процесса управления энергетическим оборудованием ТЭС, приходится отказываться от привычного понимания эксплуатации ТЭС в рамках единичного объекта и отыскивать закономерности и тенденции процесса эксплуатации в рамках большой организационно-технической системы, которая его реализует. При этом под системой эксплуатации ТЭС понимается как организационно-техническая структура, предназначенная для реализации процесса эксплуатации, так и сам процесс.

В-четвертых, управление такой системой будет эффективно только при переходе на цифровизацию управления, то есть при построении цифрового прообраза энергетического оборудования и учета влияния всех факторов на показатели работы ТЭС на всех стадиях эксплуатации.

Список литературы

- Абрамов С. Ю. Системный анализ, комплексное оценивание, совершенствование учета и управления в системах централизованного теплоснабжения: дисс....канд. техн. наук: 05.13.01 Самара, 2007. 168 с.
- Бахтадзе Н. Н. Интеллектуальные алгоритмы идентификации состояния энергообъектов / Н. Н. Бахтадзе, В. А. Лотоцкий, Е. М. Максимов, Н. Е. Максимова // Информационные технологии и вычислительные системы. 2011. № 3. С. 45–50.
- ГОСТ Р 50779.42–99. Статистические методы. Контрольные карты средних арифметических с предупреждающими границами.
- Кузнецов Р. С. Регрессионный анализ архивной информации тепловычислителей с использованием информационно-аналитической системы СОНА / Р. С. Кузнецов, В. П. Чипулис // Коммерческий учет энергоносителей: материалы 21-й международной научно-практической конференции. СПб.: Борей-Арт, 2005. С. 430–434.
- Максимюк Е. В. Математическое моделирование для поддержки принятия решений в области обеспечения энергетической эффективности / Е. В. Максимюк, В. С. Микшина. // Качество. Инновации. Образование. 2014. № 8. С. 54–63.
- Ефремов И. А., Таран А. С., Филиппова Т. А. Принципы повышения эффективности управления режимами работы электроэнергетических систем // Доклады Томского государственного университета систем управления и радиоэлектроники. 2012. № 1. С. 199–203.
- Боруш О. В. Оценка реального расхода топлива энергоблоками ТЭЦ на основе оптимизационных расчетов их режимов работы / О. В. Боруш, С. В. Зыков и др. // Доклады Академии наук Высшей школы Российской Федерации. Новосибирск: НГТУ, 2013. № 2 (21). С. 26–33.
- Draper N., Smith G. Applied regression analysis. М.: Диалектика, 2007. 911 p. ISBN: 978-5-8459-0963-3.
- Mises, Richard Edler von. Wahrscheinlichkeit Statistik und Wahrheit. М.: URSS, 2009. 253 p. ISBN: 978-5-397-00518-0.
- Grace Joel. Data science from scratch. First principles with Python. Санкт-Петербург: БХВ-Петербург, 2022. 416 с. ISBN: 978-5-9775-6731-2.
- Чимэд О. Сравнительный анализ методов расчета экономичности ТЭЦ / О. Чимэд, О. В. Боруш // Наука. Технологии. Инновации: сборник научных трудов. Новосибирск: Изд-во НГТУ, 2014. С. 39–42.
- Рыжкин В. Я. Тепловые электрические станции: учебник для вузов / Под ред. В. Я. Гирифельда. 3-е изд., перераб. и доп. М.: Энергоатомиздат, 1987. 328 с.

References

- Abramov S. Y. System analysis, integrated assessment, improvement of accounting and management in district heating systems: diss. for the academic degree in technical sciences: 05.13.01 Samara, 2007. 168 p.
- Bakhtadze N. H., Lototsky V. A., Maksimov E. M., Maksimova N. E. Intelligent algorithms for identifying the state of energy facilities. Information Technologies and Computing Systems. 2011. No. 3. Pp. 45–50.
- GOST R 50779.42–99 Statistical methods. Arithmetic mean control cards with warning borders.
- Kuznetsov P. C., Chipulis V. P. Regression analysis of archival information of thermal calculators using the SONA information and analytical system. Commercial accounting of energy resources: proceedings of the 21st International scientific and practical conference. St. Petersburg: Borey-Art, 2005. Pp. 430–434.
- Maksimyuk E. V., Mikshina B. C. Mathematical modeling to support decision-making in the field of energy efficiency. Quality. Innovation. Education. 2014. No. 8. Pp. 54–63.



6. Efremov I. A., Taran A. S., Filippova T. A. Principles of increasing the efficiency of operating modes of electric power systems // Reports of Tomsk State University of Control Systems and Radio Electronics. 2012. No. 1. Pp. 199–203.
7. Borush O. V., Zykov S. V. and others. Assessment of the real fuel consumption by CHP power units based on optimization calculations of their operating modes. Reports of the Academy of Sciences of the Higher School of the Russian Federation. Novosibirsk: NSTU, 2013. № 2 (21). Pp. 26–33.
8. Draper N., Smith G. Applied regression analysis. Moscow: Dialektika, 2007. 911 p. SBN: 978-5-8459-0963-3.
9. Mises, Richard Edler von. Wahrscheinlichkeit Statistik und Wahrheit. Moscow: URSS, 2009. 253 p. ISBN: 978-5-397-00518-0.
10. Grace Joel. Data science from scratch. First principles with Python. St. Petersburg: BHV-Petersburg, 2022. 416 p. ISBN: 978-5-9775-6731-2.
11. Chimed O. A., Borush O. V. Comparative analysis of methods for calculating the efficiency of TETs. Science. Technologies. Innovations: a collection of scientific papers. Novosibirsk, 2014. NSTU Publishing House, 2014. pp. 39–42.
12. Ryzhkin V. Ya. Thermal power plants: textbook for universities. Edited by V. Ya. Girshfeld. 3rd ed., revised and supplemented. M.: Energoatomizdat, 1987. 328 p.

■ Информация об авторах

Ширяев Александр Дмитриевич, ассистент кафедры теплосиловых установок и тепловых двигателей Высшей школы технологии и энергетики Санкт-Петербургского государственного университета промышленных технологий и дизайна, 198095, Россия, Санкт-Петербург, ул. Ивана Черных, д. 4; SPIN-код: 7117–1745, Author ID: 1172930, e-mail: a. d. shiryayev@rambler.ru.

Злобин Владимир Германович, кандидат технических наук, доцент, заведующий кафедрой теплосиловых установок и тепловых двигателей Высшей школы технологии и энергетики Санкт-Петербургского государственного университета промышленных технологий и дизайна, 198095, Россия, Санкт-Петербург, ул. Ивана Черных, д. 4; SPIN-код: 4015–0144, Author ID: 957934; e-mail: zlobin_v@list.ru.

■ Information about the authors

Shiryayev Alexander Dmitrievich, Assistant at the Department of Thermal Power Plants and Thermal Engines, Higher School of Technology and Energy, St. Petersburg State University of Industrial Technologies and Design, 4 Ivana Chernykh St., St. Petersburg, 198095, Russia; SPIN-code: 7117–1745, Author ID: 1172930, e-mail: a. d. shiryayev@rambler.ru.

Zlobin Vladimir Germanovich, PhD in Engineering, Associate Professor, Head of the Department of Thermal Power Plants and Thermal Engines of the Higher School of Technology and Energy, Saint Petersburg State University of Industrial Technologies and Design, 4 Ivan Chernykh Street, St. Petersburg, 198095, Russia; SPIN-code: 4015–0144, Author ID: 957934, e-mail: zlobin_v@list.ru.

