

Е.В. Максимюк, В.С. Микшина

## ПРИМЕНЕНИЕ РЕГРЕССИОННОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ СТЕПЕНИ ВЛИЯНИЯ ОТДЕЛЬНЫХ УЗЛОВ ОБОРУДОВАНИЯ НА ЭФФЕКТИВНОСТЬ РАБОТЫ КОТЕЛЬНЫХ

*В работе предложена новая методика выбора мероприятия при выходе процесса выработки тепловой энергии на газовой котельной из энергоэффективного режима с использованием теории системного анализа. Методика заключается в выявлении степени влияния отдельных узлов оборудования на эффективность работы котельных с помощью множественной линейной регрессионной модели.*

**Регрессионная модель, котельные установки, энергетическая эффективность.**

Для деятельности производственных объектов нефтегазодобывающих компаний, расположенных на буровых, необходима тепловая энергия. Тепловая энергия вырабатывается с помощью котельных установок. Котельные установки (источники теплоснабжения) — сложный комплекс технологически связанных тепловых энергоустановок, расположенных в обособленных производственных зданиях с котлами, водонагревателями и котельно-вспомогательным оборудованием, предназначенных для выработки теплоты. Работа оборудования котельной установки контролируется множеством датчиков, установленных на отдельных узлах.

Под энергетической эффективностью понимают характеристики, отражающие отношение полезного эффекта от использования энергетических ресурсов к затратам энергетических ресурсов, произведенным в целях получения такого эффекта, применительно к продукции, технологическому процессу. К использованным энергетическим ресурсам относят обычно расход топлива (попутного газа), расход электроэнергии, расход воды в котельной установке [1].

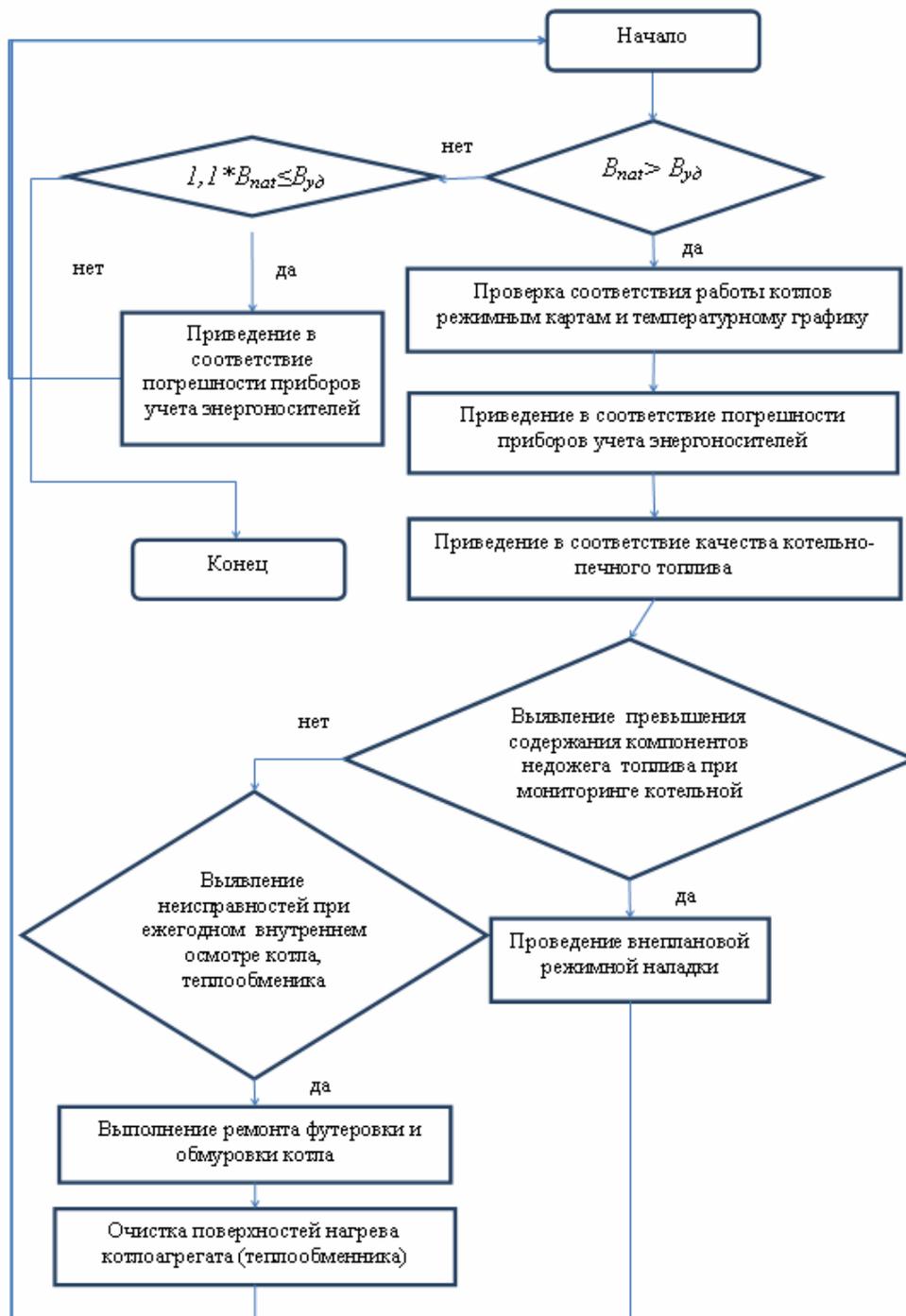
В настоящее время для управления процессами повышения энергоэффективности источников теплоснабжения используют информацию об отклонениях технологических показателей котельных установок от их удельных норм.

Таким контролируемым технологическим параметром является, например, расчет удельного расхода котельно-печного топлива на отпуск тепловой энергии:

$$b_{nat} = \frac{(B_{nat} * 10^3)}{Q_{отпуск\_в\_сеть}}, \quad \text{м}^3 / \text{Гкал} \quad (1)$$

где  $B_{nat}$  — расход натурального топлива (газ — тыс. м<sup>3</sup>, для нефти, мазута и дизельного топлива — тн) согласно показаниям счетчика топлива за учетный период;  $Q_{отпуск\_в\_сеть}$  — количество тепловой энергии, за учетный период отпущенной от источника тепла в тепловую сеть, принимаемое согласно показаниям узла учета тепловой энергии, Гкал.

На рис. 1 представлен алгоритм по выполнению мероприятий в случае превышения фактического удельного расхода топлива на отпуск тепловой энергии над расчетным.



**Рис. 1.** Алгоритм по выполнению мероприятий в случае превышения фактического удельного расхода топлива на отпуск тепловой энергии над расчетным:  $b_{уд}$  — фактический удельный расход котельно-печного топлива на отпуск тепловой энергии;  $b_{уд}^p$  — нормативный удельный расход котельно-печного топлива

Алгоритм принятия решений при превышении фактического расхода газа над нормативным заключается в следующем:

1. Проверяются погрешности приборов учета энергоносителей для приведения их в соответствие с требованиями ГОСТ.

2. Проводится осмотр котлов и теплообменников. В первую очередь проверяется соответствие работы котлов режимным картам и температурному графику. При выявлении превышения содержания компонентов несгоревшего топлива проводится внеплановая режимная наладка. В случае неисправности оборудования, как правило, это отсутствие тяги в дымоходном канале, отложение накипи в теплообменниках, запыленность, загрязненность и забивание сажей горелки, выход из строя расширительного бака, проводятся различные мероприятия: ремонт футеровки и обмуровки котла, очистка поверхностей нагрева теплообменника и т.д.

Такой подход содержит в себе существенный недостаток. В случае с увеличением удельного расхода топлива данный алгоритм предполагает полную проверку всего котельно-печного оборудования. Контроль удельного расхода топлива не позволяет однозначно определить, за счет какого котла или теплообменника возникла проблема. Из-за необходимости проведения осмотра всего котельно-печного оборудования происходит значительная временная задержка. Это приводит к увеличению материальных затрат от продолжительной неэффективной работы оборудования.

Для повышения энергоэффективности работы источников теплоснабжения представляется важным выявление узлов, ответственных за выход контролируемых параметров за границы установленных нормативов, а также степени влияния отдельных узлов оборудования на эффективность работы котельных.

Для решения этой задачи на первом этапе с помощью экспертного анализа определили комплекс факторов, являющихся важнейшими показателями энергоэффективности котельной установки. На втором этапе с помощью корреляционного анализа были установлены зависимые (контролируемые) и независимые факторы. В качестве зависимых факторов выступают расход топлива (газа), расход электроэнергии и выработанная тепловая энергия. И на третьем этапе была получена многомерная линейная регрессионная модель зависимости расхода энергоресурсов от независимых параметров источника теплоснабжения.

В работе [2] представлен перечень независимых параметров, влияющих на расход топлива при выработке тепловой энергии (табл. 1).

Таблица 1

**Перечень факторов, влияющих на расход газа при выработке тепловой энергии котельной**

№ п/п	$X_i$	Наименование фактора
1	$X_1$	Давление воды в прямой трубе
2	$X_2$	Давление воды в обратной трубе
3	$X_3$	Котел водогрейный 1-1 (наработка)
4	$X_4$	Котел водогрейный 1-1 (температура)
5	$X_5$	Котел водогрейный 1-2 (наработка)
6	$X_6$	Котел водогрейный 1-2 (температура)
7	$X_7$	Котел водогрейный 2-1 (наработка)

Окончание табл. 1

№ п/п	$X_i$	Наименование фактора
8	$X_8$	Котел водогрейный 2-1 (температура)
9	$X_9$	Котел водогрейный 2-2 (наработка)
10	$X_{10}$	Котел водогрейный 2-2 (температура)
11	$X_{11}$	Котел водогрейный 3-1 (наработка)
12	$X_{12}$	Котел водогрейный 3-1 (температура)
13	$X_{13}$	Котел водогрейный 3-2 (наработка)
14	$X_{14}$	Котел водогрейный 3-2 (температура)
15	$X_{15}$	Объем воды в прямой трубе
16	$X_{16}$	Объем воды в обратной трубе
17	$X_{17}$	Сетевой насос СН-1 (наработка)
18	$X_{18}$	Сетевой насос СН-2 (наработка)
19	$X_{19}$	Температура воды в обратной трубе
20	$X_{20}$	Температура воды в прямой трубе
21	$X_{21}$	Температура окружающей среды
22	$X_{22}$	Циркуляционный насос ЦН-1 (наработка)
23	$X_{23}$	Циркуляционный насос ЦН-2 (наработка)

В качестве исходного массива экспериментальных данных исследовались показания датчиков котельных установок за период с 01.10.2013 г. по 31.03.2014 г. Рассматриваемая нами котельная оборудована водогрейным котлом «СТГ-Классик», состоящим из двух модулей. Котел работает на газообразном топливе. Номинальная производительность 0,4 МВт. Данные котлы предназначены для отопления и горячего водоснабжения производственных, административных и жилых зданий. Каждый модуль котла оснащен цилиндрическим чугунно-медным теплообменником и горелкой предварительного смешения с наддувом.

В работе [2] рассматривалась математическая модель множественной линейной регрессии вида:

$$y_k = \theta_{1k}x_{1k} + \dots + \theta_{jk}x_{jk} + \dots + \theta_{mk}x_{mk} + \varepsilon_k, \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad k = \overline{1, l}, \quad (2)$$

где  $y_k$  — значение зависимой переменной в  $k$ -м слое наблюдения — расход газа;  $x_{ml}$  — известное значение независимой переменной в  $k$ -м слое  $m$ -й строке наблюдения;  $\theta_{jk}$  — неизвестный коэффициент при  $j$ -ой объясняющей переменной;  $\varepsilon_k$  — случайная составляющая.

Расчеты по линейной регрессионной модели проводились с помощью ППП *Statistica*.

Была проведена проверка адекватности регрессионной модели расхода энергоресурсов по экспериментальным данным, полученным с трех котельных установок [2]. Проверка показала довольно высокий результат достоверности.

В ходе исследования было сделано предположение, что на разных временных интервалах регрессионная модель будет изменяться, а именно будет меняться состав и значимость зависимых переменных. Для проверки этого утверждения были составлены регрессионные модели расхода энергоресурсов (расхода газа) котельной № 1 на выработку тепловой энергии за три временных периода:

1. с 02.01.2014 г. по 21.01.2014 г.
2. с 06.02.2014 г. по 25.02.2014 г.
3. с 05.03.2014 г. по 24.03.2014 г.

Каждый временной период охватывал массив экспериментальных данных, состоящих из 480 измерений по каждому параметру. Измерения проводились каждый час.

В результате расчетов получены три математические модели множественной линейной регрессии зависимости расхода топлива от независимых переменных для котельной.

По данным за период с 02.01.2014 г. по 21.01.2014 г. математическая модель имеет вид:

$$y_1^1 = 0,65x_1 + 2,01x_3 + 0,86x_4 + 6,003x_5 + 4,07x_6 - 0,02x_8 - 0,004x_{10} + 0,003x_{11} + 0,06x_{12} + 0,01x_{14} + 0,3x_{17} - 0,94x_{19} + 0,2x_{20} - 0,24x_{21} - 0,08x_{23}, \quad (3)$$

где  $y_1^1$  — расход топлива на выработку тепловой энергии,  $x_1-x_{21}$  — независимые переменные, описанные в табл. 1.

Полученная регрессионная модель (3) была оценена рядом статистик.

Начнем со статистики  $R^2$  — коэффициента детерминации. Она показывает, насколько условная дисперсия модели отличается от дисперсии реальных значений  $Y$ . Если этот коэффициент близок к 1, то условная дисперсия модели достаточно мала и весьма вероятно, что модель хорошо описывает данные. Если же коэффициент  $R$ -квадрат сильно меньше, например меньше 0.5, то с большой долей уверенности можно сказать, что модель не отражает реальные экспериментальные данные.

$$R^2 = 1 - \frac{\sum(\hat{x}_i - x_i)}{\sum(x_i - \bar{x})^2}, \quad (4)$$

где  $x_i$  — реальные значения  $x$  в каждом наблюдении,  $\hat{x}_i$  — значения, предсказанные моделью,  $\bar{x}$  — среднее по всем реальным значениям  $x_i$ .

Однако у статистики  $R^2$  есть один серьезный недостаток: при увеличении числа предикторов эта статистика может только возрастать. Поэтому может показаться, что модель с большим количеством предикторов лучше, чем модель с меньшим, даже если все новые предикторы никак не влияют на зависимую переменную. Тут можно вспомнить про принцип «бритвы Оккама». Следуя ему, по возможности, стоит избавляться от лишних предикторов в модели, поскольку она становится более простой и понятной. Для этих целей применяется статистика скорректированного  $R^2$ .

$$R_{adj}^2 = 1 - R^2 \frac{k - 1}{k - n - 1} \quad (5)$$

Эта статистика представляет собой  $R^2$ , но со штрафом за большое количество предикторов. Основная идея: если новые независимые переменные дают большой вклад в качество модели, значение этой статистики растет, если нет — то наоборот уменьшается.

Для уравнения (3) получены следующие результаты:

$$- R^2 = 0,78;$$

$$- R_{adj}^2 = 0,76.$$

Как видим из результатов расчета, скорректированный коэффициент детерминации близок к единице, а это значит, что данная модель возможно достоверна и объясняет поведение большего числа точек.

Теперь проверим гипотезу о равенстве нулю всех коэффициентов при предикторах, то есть гипотезу о том, зависит ли вообще величина  $y$  от величин  $x_i$  линейно. Для этого можно использовать следующую статистику, которая, если гипотеза о равенстве нулю всех коэффициентов верна, имеет распределение Фишера с  $n$  и  $k - n - 1$  степенями свободы.

$$F = \frac{R^2 / n}{(1 - R^2) / (k - n - 1)} \quad (6)$$

Соответственно, если значение этой статистики больше критического значения при данном уровне значимости, то нулевая гипотеза отвергается, что означает статистическую значимость регрессии. В противном случае модель признается незначимой.

В рассматриваемом нами уравнении  $F(18,461) = 20,2$  при  $p < 0,00001$ , а  $F_{кр} = 3,229$ . Таким образом,  $F_{расч} > F_{кр}$ , и, соответственно, модель признается значимой.

Стандартная ошибка регрессии рассматривается в качестве меры разброса данных наблюдений от смоделированных значений, она рассчитывается как квадратный корень из несмещенной оценки дисперсии регрессии:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{(n-2)} \left[ \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\left[ \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]}, \quad (7)$$

где  $n$  — общее число наблюдений,  $y_i$  — значения наблюдаемой переменной,  $x_i$  — значения объясняющей переменной,  $\bar{y}$  — среднее значение наблюдаемой переменной по выборке,  $\bar{x}$  — среднее значение объясняющей переменной по выборке,  $\hat{\sigma}^2$  — несмещенная оценка дисперсии регрессии. Чем меньше значение стандартной ошибки регрессии, тем качество модели выше.

Стандартная ошибка оценки регрессии (3):  $\hat{\sigma} = 1,2$ .

Стандартная ошибка оценки регрессии (3) достаточно мала, следовательно, можно сделать вывод, что качество модели достаточно высокое.

В табл. 2 представлен рейтинг факторов, влияющих на расход газа при выработке тепловой энергии котельной, полученный на основе линейной регрессионной модели.

Наиболее значимыми факторами модели (3) для  $Y_1^1$  является «наработка котла 1-2» и «температура котла 1-2», далее идут факторы: «наработка котла 1-1» и «температура воды в обратной трубе». Это дает основание предположить, что наибольший вклад в расход газа вносит работа котла 1-2.

Уравнение расхода топлива на выработку тепловой энергии для котельной № 1 за период с 06.02.2014 г. по 25.02.2014 г. имеет вид:

$$y_2^1 = 0,02x_2 + 3,1x_3 + 0,04x_4 + 0,902x_5 + 3,04x_7 + 0,24x_8 + 0,008x_{11} + 2,05x_{12} + 3,01x_{13} + 0,54x_{15} - 0,68x_{19} + 0,65x_{20} - 0,23x_{21} - 8,78, \quad (8)$$

где  $y_2^1$  — расход топлива,  $x_1-x_{21}$  — независимые переменные, описанные в табл. 1.

Таблица 2

**Рейтинг факторов, влияющих на расход газа при выработке тепловой энергии котельной, полученный на основе линейной регрессионной модели**

№ п/п	$X_i$	Наименование фактора	Значение коэфф. в регрессионной модели	Рейтинг фактора
1	$X_5$	Котел водогрейный 1-2 (наработка)	6,003	1
2	$X_6$	Котел водогрейный 1-2 (температура)	4,07	2
3	$X_3$	Котел водогрейный 1-1 (наработка)	2,01	3
4	$X_{19}$	Температура воды в обратной трубе	-0,94	4
5	$X_4$	Котел водогрейный 1-1 (температура)	0,86	5
6	$X_1$	Давление воды в прямой трубе	0,65	6
7	$X_{17}$	Сетевой насос СН-1 (наработка)	0,3	7
8	$X_{21}$	Температура окружающей среды	-0,24	8
9	$X_{20}$	Температура воды в прямой трубе	0,2	9
10	$X_{23}$	Циркуляционный насос ЦН-2 (наработка)	-0,08	10
11	$X_{12}$	Котел водогрейный 3-1 (температура)	0,06	11
12	$X_8$	Котел водогрейный 2-1 (температура)	-0,02	12
13	$X_{14}$	Котел водогрейный 3-2 (температура)	0,01	13
14	$X_{10}$	Котел водогрейный 2-2 (температура)	-0,004	14
15	$X_{11}$	Котел водогрейный 3-1 (наработка)	0,003	15

Общие результаты оценки регрессионной модели (8):

—  $R^2 = 0,77$ ;

—  $R_{adj}^2 = 0,72$ ;

—  $F(13,466) = 51,5$   $p < 0,00001$ ;  $F_{кр} = 466$ ;

— стандартная ошибка оценки регрессии: 1,3.

Как видим из результатов исследования, скорректированный коэффициент детерминации близок к единице, а это значит, что данная модель достаточно достоверна.

Так как  $F_{расч} > F_{кр}$ , то регрессия статистически значима. Стандартная ошибка оценки регрессии достаточно мала, а следовательно, модель достаточно достоверна.

В табл. 3 представлен рейтинг факторов, влияющих на расход газа при выработке тепловой энергии котельной, полученный на основе линейной регрессионной модели.

Наиболее значимыми факторами модели (8) для  $Y_2^1$  являются «наработки котлов 1-1, 2-1 и 3-2». Значения их коэффициентов в регрессионной модели примерно одинаково. Это означает, что они вносят существенный и равноценный вклад в процесс расхода газа. Далее идут факторы «температура котла 3-1» и «наработка котла 1-2». Эти факторы также вносят существенный вклад в расход топлива.

Множественная регрессия расхода топлива на выработку тепловой энергии на котельной № 1 за период с 05.03.2014 г. по 24.03.2014 г. имеет следующий вид:

$$y_3^1 = 0,32 x_1 + 15,09 x_3 + 0,004 x_4 + 0,006 x_5 + 0,031 x_6 + 0,008 x_7 + 0,001 x_9 + 0,016 x_{11} + 8,005 x_{13} + 0,28 x_{14}, \quad (9)$$

где  $y_3^1$  — расход топлива;  $x_1-x_{14}$  — независимые переменные, описанные в табл. 1.

Таблица 3

**Рейтинг факторов, влияющих на расход газа при выработке тепловой энергии котельной, полученный на основе линейной регрессионной модели**

№ п/п	$X_i$	Наименование фактора	Значение коэфф. в регрессионной модели	Рейтинг фактора
1	$X_3$	Котел водогрейный 1-1 (наработка)	3,1	1
2	$X_7$	Котел водогрейный 2-1 (наработка)	3,04	2
3	$X_{13}$	Котел водогрейный 3-2 (наработка)	3,01	3
4	$X_{12}$	Котел водогрейный 3-1 (температура)	2,05	4
5	$X_5$	Котел водогрейный 1-2 (наработка)	0,902	5
6	$X_{19}$	Температура воды в обратной трубе	-0,68	6
7	$X_{20}$	Температура воды в прямой трубе	0,65	7
8	$X_{15}$	Объем воды в прямой трубе	0,54	8
9	$X_8$	Котел водогрейный 2-1 (температура)	0,24	9
10	$X_{21}$	Температура окружающей среды	-0,23	10
11	$X_4$	Котел водогрейный 1-1 (температура)	0,04	11
12	$X_2$	Давление воды в обратной трубе	0,02	12
13	$X_{11}$	Котел водогрейный 3-1 (наработка)	0,008	13

Общие результаты оценки регрессионной модели:

—  $R^2 = 0,72$ ;

—  $R_{adj}^2 = 0,69$ ;

—  $F(9,470) = 7,43$ ;  $p < 0,00001$ ;

— стандартная ошибка оценки регрессии: 2,4.

Результаты оценки регрессии (9) показывают, что регрессия статистически значима и достоверна.

В таблице 4 представлен рейтинг факторов, влияющих на расход газа при выработке тепловой энергии котельной, полученный на основе линейной регрессионной модели.

Таблица 4

**Рейтинг факторов, влияющих на расход газа при выработке тепловой энергии котельной, полученный на основе линейной регрессионной модели**

№ п/п	$X_i$	Наименование фактора	Значение коэфф. в регрессионной модели	Рейтинг фактора
1	$X_3$	Котел водогрейный 1-1 (наработка)	15,09	1
2	$X_{13}$	Котел водогрейный 3-2 (наработка)	8,005	2
3	$X_1$	Давление воды в прямой трубе	0,32	3
4	$X_{14}$	Котел водогрейный 3-2 (температура)	0,28	4
5	$X_6$	Котел водогрейный 1-2 (температура)	0,031	5
6	$X_{11}$	Котел водогрейный 3-1 (наработка)	0,016	6
7	$X_7$	Котел водогрейный 2-1 (наработка)	0,008	7
8	$X_5$	Котел водогрейный 1-2 (наработка)	0,006	8
9	$X_4$	Котел водогрейный 1-1 (температура)	0,004	9
10	$X_9$	Котел водогрейный 2-2 (наработка)	0,001	10

Наиболее значимым фактором модели (9) для  $u_3^1$  является фактор «наработка котла 1-1», далее с большим отрывом идет «наработка котла 3-2» и «давление воды в прямой трубе». То есть можно сказать, что наибольший вклад в процесс расхода газа вносит котел 1-1.

Из составленных уравнений регрессии работы котельных видно, что наборы значимых независимых переменных за разные даты, но равные промежутки времени отличаются. Отличаются и коэффициенты при переменных. Данную разницу можно объяснить изменениями в состоянии оборудования и режимов процесса выработки тепловой энергии с течением времени. Коэффициенты при независимых переменных показывают, насколько сильно влияние данной переменной. Отметим, что каждая независимая переменная представляет собой измерения датчика определенного параметра с конкретного узла оборудования котельной. Таким образом, узел оборудования, вносящий наибольшее влияние в режим работы всей котельной установки, может быть однозначно идентифицирован.

Итак, результаты математического моделирования процесса получения тепла для принятия решений об обеспечении энергетической эффективности работы котельной могут использоваться следующим образом:

— проводится мониторинг контролируемых показателей, являющихся зависимыми переменными линейной регрессионной модели;

— при превышении какого-либо контрольного параметра над нормативом по 23 независимым переменным, отобранным методами экспертного и корреляционного анализа, строится линейная регрессионная модель на основе экспериментальных данных, полученных на предыдущем периоде (примерно 1000 точек);

— по весовым коэффициентам математической модели определяется, какой из независимых параметров оказывает наибольшее влияние на ситуацию. Так как каждый параметр описывает состояние определенного узла котельной, то в режиме, приближенном к реальному времени, возможно выявить, в каком узле производственного оборудования происходит сбой, и сформировать мероприятие по нормализации его работы.

## ЛИТЕРАТУРА

1. *Абдульманов Р.Р.* Подходы к созданию систем поддержки принятия решения реального времени в области энергоэффективности источников теплоснабжения / Р.Р. Абдульманов, Е.В. Максимюк, В.С. Микшина // Инновационные информационные технологии. Материалы международной научно-практической конференции / Научн. ред. А.Н. Тихонов; Общ. ред. С.У. Увайсов; Отв. ред. И.А. Иванов. — М.: МИЭМ НИУ ВШЭ. — С. 228–231.
2. *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.
3. *Максимюк Е.В.* Прогнозирования зависимых контролируемых параметров работы источников теплоснабжения // Современные проблемы науки и образования. — 2014. — № 6 (приложение «Технические науки»). — С. 22.
4. *Рао С.Р.* Линейные статистические методы и их применения. — Наука, 1968.
5. *Розанов Ю.А.* Теория вероятностей, случайные процессы и математическая статистика. — Наука, 1985.
6. *Боровиков В.П.* STATISTICA, искусство анализа данных на компьютере. — Питер 2001.

7. *Maksimyuk E. V. Mathematical methods and algorithms for decision support to ensure energy efficiency / E. V. Maksimyuk, V. S. Mikshina // Innovative Information Technologies: Materials of the International scientific — practical conference. Part 3 / Ed. Uvaysov S. U. — M.: HSE, 2014. — С. 295–296.*

*Сураутский государственный университет  
for\_mev@mail.ru  
mikshinavs@gmail.com*

*E.V. Maksimyuk, V.S. Mikshina*

*Application of the regression analysis for identification of extent of influence of separate knots of the equipment on efficiency performance of boiler installations*

*In this paper we propose a new method for selecting actions on exit process heat generation of energy-efficient gas boiler mode. The method is using multiple regression for determine the degree of influence of individual pieces of equipment on the efficiency of boiler installations.*

***Regression model, boiler installation, energy efficiency.***