УДК 681.53 DOI: 10.18324/2077-5415-2018-3-57-61

Корреляционно-регрессионный анализ технологических параметров

Т.А. Григорьева[a], В.Н. Толубаев[b]

Братский государственный университет, ул. Макаренко 40, Братск, Россия
[a]umubrgu@mail.ru, [b]tolubaevvladimir@gmail.com
[a]https://orcid.org/0000-0002-5361-6832, [b]https://orcid.org/0000-0002-7438-2254
Статья поступила 20.03.2018, принята 20.04.2018

В работе исследуются динамические свойства сушильной части пресспата на основе корреляционно-регрессионного анализа данных, полученных в результате технологического процесса. Одним из основных требований, предъявляемых к продукции целлюлозного производства, является влажность готового целлюлозного полотна. Этот параметр в работе принимается в качестве результативного показателя. Факторными параметрами приняли давление пара в цилиндрах сушильной группы, температуру сушильного воздуха и уровень конденсата в сепараторе. Многофакторный корреляционно-регрессионный анализ позволяет построить четырехмерную регрессионную модель исследуемых параметров, оценить меру влияния каждого из включенных в модель факторов на исследуемый результативный показатель, а также с определенной степенью точности найти теоретическое значение результирующего показателя при любых возможных сочетаниях факторов. В работе проанализированы характеристики тесноты связи между зависимой и независимыми переменными: парные, частные и множественные коэффициенты корреляции и детерминации. Определена взаимосвязь факторных признаков при устранении влияния результативного признака. На основе полученных значений парных коэффициентов корреляции найдены значения совокупных коэффициентов множественной корреляции и детерминации. Проведена проверка значимости уравнения множественной регрессии по статистическим критериям Фишера и Стьюдента. Анализ исходных данных позволяет установить наличие причинно-следственной связи между исследуемыми параметрами. Данный метод находит широкое применение в различных областях науки и техники, в том числе в автоматизации технологических процессов.

Ключевые слова: корреляция; регрессионный анализ; линейная математическая модель; корреляционно-регрессионный анализ; многофакторная регрессионная модель; парные коэффициенты корреляции; частные коэффициенты корреляции; сушильная часть пресспата.

Correlation-regression analysis of technological parameters

T.A. Grigorieva[a], V.N. Tolubaev[b]

Bratsk State University; 40, Makarenko St., Bratsk, Russia
[a]umubrgu@mail.ru, [b]tolubaevvladimir@gmail.com
[a]https://orcid.org/0000-0002-5361-6832, [b]https://orcid.org/0000-0002-7438-2254
Received 20.03.2018, accepted 20.04.2018

In the paper, the dynamic properties of the drying part of the press are studied on the basis of correlation-regression analysis of data obtained as a result of the technological process. One of the basic requirements for pulp production is the moisture of the finished cellulose web. This parameter is taken as an effective indicator. Factor parameters were the steam pressure in the cylinders of the drying group, the temperature of the drying air, and the condensate level in the separator. Multifactor correlation-regression analysis makes it possible to construct a four-dimensional regression model of the investigated parameters, to estimate the measure of influence of each of the factors included in the model on the investigated effective indicator, and also to find with a certain degree of accuracy the theoretical value of the resulting indicator for any possible combinations of factors. In the paper, the characteristics of the tightness of the relationship between the dependent and independent variables were analyzed: paired, partial and multiplicative correlation and determination coefficients. The interrelation of factor attributes is determined at elimination of influence of effective attribute. Based on the obtained values of paired correlation coefficients, the values of the combined coefficients of multiple correlation and determination were found. The validity of the multiple regression equation was tested by the statistical criteria of Fisher and Student. Analysis of the initial data makes it possible to establish the presence of a causal relationship between the parameters studied. This method finds wide application in various fields of science and technology, including the automation of technological processes.

Keywords: correlation; regression analysis; linear mathematical model; correlation-regression analysis; multifactorial regression model; paired correlation coefficients; partial correlation coefficients; drying part of the press.

Введение

Целью данной работы является исследование динамических свойств сушильной части пресспата на основе многофакторного корреляционно-регрессионного анализа [1; 2].

В рамках этой работы был выполнен анализ данных с датчиков, построена регрессионная трехфакторная модель, определены и проанализированы характеристики тесноты связи между зависимой и независимой переменными: парные, частные и множественные коэффициенты корреляции и детерминации, проверена адекватность данной модели [3–7].

Постановка и решение задачи. В качестве исследуемых признаков выступают данные с датчиков, установленных в сушильной части пресспата целлюлозного производства: x_1 — давление пара в цилиндрах сушильной группы; x_2 — температура сушильного воздуха; x_3 — уровень конденсата в сепараторе; y — влажность целлюлозного полотна [8; 9].

Анализ исходных данных позволяет установить наличие причинно-следственной связи между исследуемыми параметрами.

Одним из основных требований, предъявляемых к продукции целлюлозного производства, является влажность готового целлюлозного полотна. Этот параметр в работе принимается в качестве результативного показателя. Факторами, которые могут оказывать влияние на результативный признак, являются:

– температура воздуха в сушильной камере (^oC);

– давление пара в сушильных цилиндрах, соприкасающихся с полотном (*кПа*);

– уровень конденсата в сепараторах (*м*).

Изменение факторных и результативного показателей во времени показано на рис. 1–4.

На основе экспериментальных данных была построена четырехмерная регрессионная модель в линейной форме:

$$y_{x_1x_2x_3} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 ,$$

где $y_{x_1x_2x_3}$ — расчетное значение зависимой переменной (результативного признака); x_1, x_2, x_3 — независимые переменные (факторные признаки); a_0, a_1, a_2, a_3 — параметры модели.

Коэффициенты a_0, a_1, a_2, a_3 определены путем решения системы нормальных уравнений:

$$\begin{cases} a_0 n + a_1\sum x_1 + a_2\sum x_2 + a_3\sum x_3 = \sum y \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + a_3\sum x_1x_3 = \sum yx_1 \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 + a_3\sum x_2x_3 = \sum yx_2 \\ a_0\sum x_3 + a_1\sum x_1x_3 + a_2\sum x_2x_3 + a_3\sum x_3^2 = \sum yx_3 \end{cases},$$

где n — объем выборки анализируемых показателей.

С учетом исходных данных составлена система нормальных уравнений:

$$\begin{cases} 21a_0 + 6538{,}6a_1 + 2544{,}2a_2 + 36{,}6a_3 = 92{,}5 \\ 6538{,}6a_0 + 2036131a_1 + 791812{,}1a_2 + 11363{,}4a_3 = 28759{,}2 \\ 2544{,}2a_0 + 791812{,}1a_1 + 348532{,}5a_2 + 5746{,}07a_3 = 14231{,}9 \\ 36{,}6a_0 + 11363{,}4a_1 + 5746{,}07a_2 + 115{,}1a_3 = 263{,}58 \end{cases}$$

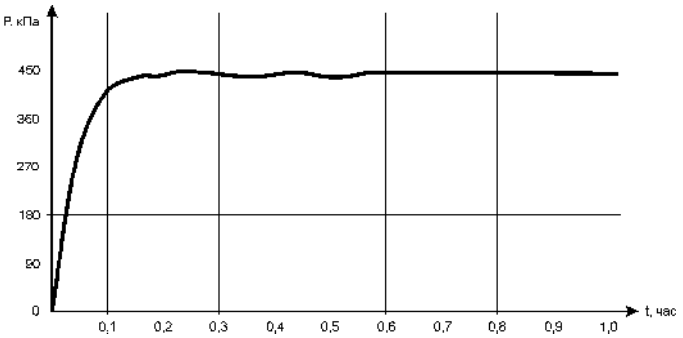

Рис. 1. Динамика исследуемых показателей: давление пара (x_1)

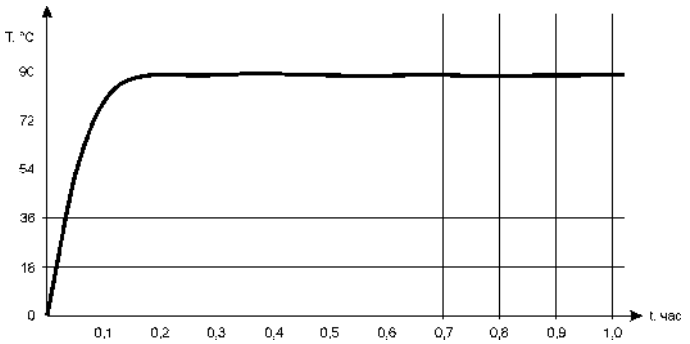

Рис. 2. Динамика исследуемых показателей: температура воздуха (x_2)

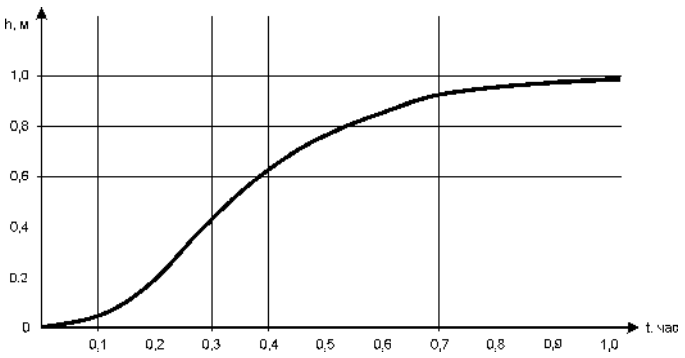

Рис. 3. Динамика исследуемых показателей: уровень конденсата (x_3)

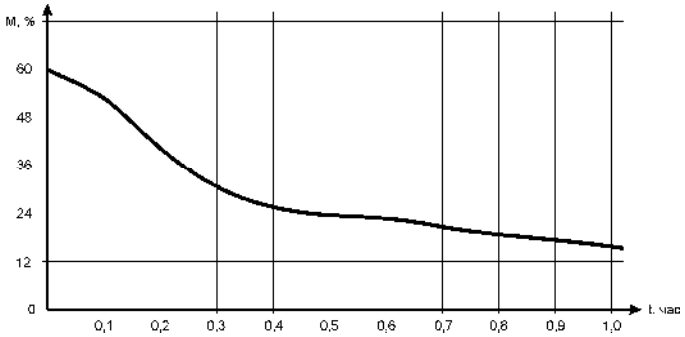

Рис. 4. Динамика исследуемых показателей: влажность полотна (y)

Для решения данной системы уравнений использована программа расчета «Решение СЛАУ», в которой применяется метод Зейделя (последовательные приближения) или метод Крамера.

При решении данной системы уравнений получены следующие результаты:

$a_0 = –139,12;$

$a_1 = –2,45;$

$a_2 = –10;$

$a_3 = –47,44.$

На основе решения данной системы уравнений получаем регрессионную модель вида:

$$y_{x_1x_2x_3} = –139,12 – 2,45x_1 – 10x_2 – 47,44x_3 .$$

В работе были проанализированы характеристики тесноты связи между зависимой и независимыми переменными: парные, частные и множественные коэффициенты корреляции и детерминации.

Для измерения тесноты связи между двумя из рассматриваемых показателей (без учета их взаимодействия с другими показателями) применяются парные коэффициенты корреляции:

$$r_{yx_i} = \frac{\overline{x_1 y} - \overline{x_1}\,\overline{y}}{\sigma_{x_i}\sigma_y} ,$$

где r_{yx_i} — парный коэффициент корреляции; σ_{x_i}, σ_y — среднеквадратические отклонения анализируемых показателей.

Результаты расчетов парных коэффициентов корреляции исследуемого объекта приведены в табл. 1.

Таблица 1

Парные коэффициенты корреляции

	y	x_1	x_2	x_3
x_1	–0,69	1	0,98	0,56
x_2	–0,65	0,98	1	0,51
x_3	–0,97	0,56	0,51	1

Отрицательные значения коэффициентов свидетельствуют о наличии обратной связи между результативным и факторными показателями, т. е. влажность полотна снижается при повышении давления пара, температуры воздуха или уровня конденсата.

Однако в реальных условия все переменные, как правило, взаимосвязаны. Теснота этой связи определяется частными коэффициентами корреляции, которые характеризуют степень и влияние одного из аргументов на функцию при условии, что остальные независимые переменные закреплены на постоянном уровне [10–14].

Изучение парных и частных коэффициентов корреляции позволяет отобрать наиболее существенные, значимые факторы.

Частный коэффициент корреляции первого порядка между признаками y и x_i при исключении влияния признака x_j вычислим по формуле:

$$r_{yx_i(x_j)} = \frac{r_{yx_i} - r_{yx_j} r_{x_i x_j}}{\sqrt{(1 - r_{yx_j}^2)(1 - r_{x_i x_j}^2)}} .$$

Также рассчитаем взаимосвязь факторных признаков при устранении влияния результативного признака:

$$r_{yx_i(y)} = \frac{r_{x_i x_j} - r_{yx_i} r_{x_i x_j}}{\sqrt{(1 - r_{yx_i}^2)(1 - r_{yx_j}^2)}} .$$

Значения частных коэффициентов корреляции приведены в табл. 2 и 3.

Таблица 2

Частные коэффициенты корреляции между признаками $y – x_n$ и $y – x_{n+1}$

	r_{y_1}	r_{y_2}	r_{y_3}	искл.
r_{y_1}	1	0,74	–0,97	x_1
r_{y_2}	–0,77	1	–0,96	x_2
r_{y_3}	–0,75	–0,71	1	x_3

Таблица 3

Частные коэффициенты корреляции между признаками $x_n – x_{n+1}$

	r_{x_1}	r_{x_2}	r_{x_3}	искл.
r_{x_1}	1	0,049	0,64	y
r_{x_2}	0,049	1	0,015	y
r_{x_3}	0,64	0,015	1	y

По данным табл. 2 и 3 можно сделать вывод, что все факторы влияют на результирующий показатель, но даже при исключении результативного показателя связь между факторными признаками присутствует. Например, чем выше давление пара в цилиндрах, тем выше уровень конденсата в сепараторе, $r_{x_1x_3} = 0,64$, что соответствует реальным процессам в рассматриваемом объекте. В то же время, практически отсутствует связь между факторными признаками $r_{x_1x_2} = 0,049$ и $r_{x_2x_3} = 0,015$, т. е. температура сушильного воздуха имеет слабую связь с давлением пара и уровнем конденсата. Это наблюдение также подтверждается на практике.

Показателем тесноты связи, устанавливаемой между результативным и факторными признаками, является совокупный коэффициент множественной корреляции.

На основе полученных значений парных коэффициентов корреляции были получены значения совокупных коэффициентов множественной корреляции и детерминации.

Совокупный коэффициент множественной корреляции $R_{yx_1x_2x_3}$ отображает одновременное влияние факторных признаков на результативный. Чем ближе его значение к 1, тем выше интенсивность корреляционной связи. Значение совокупного коэффициента множественной корреляции для двухфакторных связей равно:

$$R_{yx_ix_j} = \sqrt{\frac{r_{yx_i}^2 + r_{yx_j}^2 - 2 r_{yx_i} r_{yx_j} r_{x_i x_j}}{1 - r_{x_i x_j}^2}} .$$

Так как в рассматриваемой модели учитываются три факторных показателя, был вычислен совокупный

коэффициент множественной корреляции для трехфакторной связи:

$$R_{yx_1x_2x_3} = \frac{\sum R_{yx_ix_j}}{3} = 0{,}95 .$$

Совокупный коэффициент детерминации R^2 показывает, какая доля вариации изучаемого показательного признака объясняется влиянием факторных признаков, включенных в уравнение множественной регрессии:

$$R^2 = R^2_{yx_1x_2x_3} = 0{,}89 .$$

На основе полученных результатов был сделан вывод о том, что вариация значения влажности полотна на 89 % обуславливается тремя анализируемыми факторами.

Это значит, что выбранные факторы существенно влияют на влажность полотна. Такая связь свидетельствует о целесообразности построения трехфакторной регрессионной модели в виде линейного уравнения регрессии, однако показатели могут оказаться подвержены действию случайных факторов. Поэтому только после проверки адекватности оно может быть пригодно для применения.

Была проведена проверка значимости уравнения множественной регрессии по статистическим критериям Фишера и Стьюдента. Вычисление данных критериев показало, что рассматриваемое уравнение множественной регрессии является адекватным.

Табличное значение *F*-критерия при доверительной вероятности 0,95 (уровень значимости α = 0,05) и при числах степеней свободы $k_1 = 2$ и $k_2 = 17$ составляет $F_{табл} = 4{,}76$. Так как $F_{расч} > F_{табл}$ более чем в 4 раза, данное уравнение регрессии следует признать адекватным и пригодным для практического применения.

Табличное значение *t*-критерия при доверительной вероятности 0,95 (α = 0,05) при $k = 17$ степенях свободы составляет 2,11, соответственно, все величины следует признать значимыми, существенными.

Таким образом, результатом проведенного трехфакторного корреляционно-регрессионного анализа стало получение линейной модели:

$$y_{x_1x_2x_3} = -139{,}12 - 2{,}45x_1 - 10x_2 - 47{,}44x_3 ,$$

описывающей связь технологических параметров сушильной камеры пресспата.

Выводы

Исходя из сказанного, можно сделать соответствующие выводы и осуществить практические мероприятия, направленные на оптимизацию технологического процесса и улучшение качества готовой продукции [15–18].

Так, например, $a_1 = 2{,}45$ свидетельствует о том, что с увеличением давления на 1 кПа следует ожидать уменьшения влажности полотна на 2,45 %. Повышение же температуры на 1 °С может привести к снижению влажности на 10 %, а увеличение уровня конденсата на 1 м приведет к снижению влажности на 47,4 %.

Следующим шагом для автоматизации процесса сушки целлюлозы является выбор контроллера для многоконтурной системы автоматического регулирования [19; 20].

Литература

1. Григорьева Т.А., Шуманский Э.К. Многофакторный корреляционно-регрессионный анализ технологических параметров сушки целлюлозы // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2016. Т. 2. С. 134-138.

2. Григорьева Т.А., Толубаев В.Н. Управление динамическими свойствами сушильной установки // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2015. Т. 2. С. 52-54.

3. Дойников А.Н., Григорьева Т.А. Методика формирования модели многосвязной системы для адаптивного управления качеством переходных процессов с использованием регуляторов на смежных станциях. Братск, 2004. Рукоп. деп. в ВИНИТИ 06.08.2004. № 1367-В2004.

4. Дойников А.Н., Крумин О.К., Григорьева Т.А. Методика и алгоритм адаптивной стабилизации многосвязной электроэнергетической системы // Вестн. Иркут. гос. техн. ун-та. 2004. Т. 2, № 18. С. 31-36.

5. . Григорьева Т.А., Короткая А.С. Управление динамическими свойствами парового котла БКЗ 75/39// Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2015. Т. 2. С. 64-67.

6. Григорьева Т.А., Толубаев В.Н. Система сбора и представления информации об учете энергоресурсов // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2006. Т. 2. С. 73-75.

7. Григорьева Т.А., Патрусова А.М. Моделирование динамических процессов в печи обжига металла // Наука и современность. 2014. № 28. С. 167-171.

8. Григорьева Т.А., Толубаев В.Н. Выбор преобразователей температуры в современных системах автоматического регулирования // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2015 Т. 1. С. 150-153.

9. Григорьева Т.А., Толубаев В.Н. Выбор преобразователей давления в современных системах автоматического регулирования// Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2015. Т. 2. С. 54-58.

10. Дойников, А. Н., Григорьева Т.А. Анализ динамических свойств и синтез моделей электроэнергетических систем по режимным частотным характеристикам // Информационные технологии и проблемы математического моделирования сложных систем: сб. науч. тр. Иркутск, 2005. С. 91-102.

11. Григорьева Т.А. Параметрическая идентификация электроэнергетических систем для управления собственными динамическими свойствами: автореф. дис. … канд. техн. наук. Братск, 2005.

12. Григорьева Т.А. Параметрическая идентификация электроэнергетических систем для управления собственными динамическими свойствами: дис. … канд. техн. наук. Братск, 2005.

13. Григорьева Т.А. Управление динамическими свойствами в теплоэнергетических системах // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2010. Т. 1. С. 54-58.

14. Григорьева Т.А., Патрусова А.М. Проблемы управления динамическими свойствами промышленных объектов // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2013. Т. 2. С. 210-213.

15. Патрусова А.М. Алгоритмы идентификации и диагностики аналоговых промышленных объектов: дис. ... канд. техн. наук. Братск, 2003. 125 с.

16. Патрусова А.М. Метод вторичной идентификации линейных динамических объектов // Информатика и системы управления. 2001. № 2 (02). С. 96-99.

17. Патрусова А.М., Колтыгин Д.С. Анализ и разработка прикладных методов идентификации промышленных объектов. Братск, 2003. Рукоп. деп. в ВИНИТИ 02.04.2003. № 583-В2003.

18. Григорьева Т.А., Толубаев В.Н. Автоматизация технологических процессов и производств: Братск: Изд-во БрГУ, 2017. 107 с.

19. Григорьева Т.А., Толубаев В.Н. Выбор программируемых контроллеров в современном производстве // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2014. Т. 1. С. 269-271.

20. Григорьева Т.А., Толубаев В.Н. Автоматизированные системы управления на базе ПТК «ТЕКОН» // Труды Братского государственного университета. Сер. Естественные и инженерные науки. 2014. Т. 1. С. 271-274.

References

1. Grigor'eva T.A., Shumanskij Eh.K. Multifactorial correlation-regression analysis of technological parameters of cellulose drying // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2016. T. 2. P. 134-138.

2. Grigor'eva T.A., Tolubaev V.N. Control of the dynamic properties of the drying plant // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2015. T. 2. P. 52-54.

3. Dojnikov A.N., Grigor'eva T.A. The method of forming a model of a multiply connected system for adaptive control of the quality of transient processes using regulators at adjacent stations. Bratsk, 2004. Rukop. dep. v VINITI 06.08.2004. № 1367-V2004.

4. Dojnikov A.N., Krumin O.K., Grigor'eva T.A. Metodika and algorithm of adaptive stabilization of multiply connected electric power system // Vestnik of Irkutsk State Technical University. 2004. T. 2, № 18. P. 31-36.

5. Grigor'eva T.A., Korotkaya A.S. Control of the dynamic properties of the steam boiler BKZ 75/39 // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2015. T. 2. P. 64-67.

6. Grigor'eva T.A., Tolubaev V.N. System for collecting and presenting information on the accounting of energy resources // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2006. T. 2. P. 73-75.

7. Grigor'eva T.A., Patrusova A.M. Modeling of dynamic processes in a metal firing furnace // Nauka i sovremennost'. 2014. № 28. P. 167-171.

8. Grigor'eva T.A., Tolubaev V.N. Choice of temperature converters in modern automatic control systems // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2015 T. 1. P. 150-153.

9. Grigor'eva T.A., Tolubaev V.N. Choice of pressure transducers in modern automatic control systems // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2015. T. 2. P. 54-58.

10. Dojnikov, A. N., Grigor'eva T.A.The analysis of dynamic properties and synthesis of models of electric power systems by the regime frequency characteristics// Informacionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnyh sistem: sb. nauch. tr. Irkutsk, 2005. P. 91-102.

11. Grigor'eva T.A. Parametric identification of electric power systems for managing their own dynamic properties: avtoref. dis. … kand. tekhn. nauk. Bratsk, 2005.

12. Grigor'eva T.A. Parametric identification of electric power systems for managing their own dynamic properties: dis. … kand. tekhn. nauk. Bratsk, 2005.

13. Grigor'eva T.A. Management of dynamic properties in heat-and-power systems // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2010. T. 1. P. 54-58.

14. Grigor'eva T.A., Patrusova A.M. Problems of controlling the dynamic properties of industrial facilities // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2013. T. 2. P. 210-213.

15. Patrusova A.M. Algorithms for identification and diagnostics of analog industrial objects: dis. ... kand. tekhn. nauk. Bratsk, 2003. 125p.

16. Patrusova A.M. Method of secondary identification of linear dynamic objects // Information Science and Control Systems. 2001. № 2 (02). P. 96-99.

17. Patrusova A.M., Koltygin D.S. Analysis and development of applied methods for identifying industrial objects. Bratsk, 2003. Rukop. dep. v VINITI 02.04.2003. № 583-V2003.

18. Grigor'eva T.A., Tolubaev V.N. Automation of technological . Bratsk: Izd-vo BrGU, 2017. 107 p.

19. Grigor'eva T.A., Tolubaev V.N. Choice of programmable controllers in modern production // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2014. T. 1. P. 269-271.

20. Grigor'eva T.A., Tolubaev V.N. Automated control systems on the basis of PTK "TEKON" // Trudy Bratskogo gosudarstvennogo universiteta. Ser. Estestvennye i inzhenernye nauki. 2014. T. 1. P. 271-274.