

УДК 519.2, 004.8

**ПОСТРОЕНИЕ ЗАВИСИМОСТИ ТЕМПЕРАТУРЫ НЕПРЕРЫВНО
ЛИТОЙ ЗАГОТОВКИ ОТ ИСТОРИИ ЕЁ НАГРЕВА
DEVELOPMENT OF DEPENDENCE OF CONTINUOUSLY CAST
BILLETS TEMPERATURE FROM ITS HEATING RETROSPECTION**

**Жуков П.И., Глущенко А.И., Фомин А.В.
Zhukov P.I., Glushchenko A.I., Fomin A.V.**

Старооскольский технологический институт им. А.А. Угарова (филиал)
ФГАОУ ВО «Национальный исследовательский технологический
университет «МИСиС»

Sary Oskol Technological Institute n.a. A.A. Ugarov (branch) National
University of Science and Technology "MISIS"

Аннотация: В работе построена зависимость между температурой стальной заготовки на выходе методической печи прокатного цеха металлургического предприятия от температуры в зонах этой печи и времени нахождения заготовки в них. Вид зависимости – модель поверхности отклика 4-го порядка с погрешностью прогноза 15⁰С (1,5%).

Abstract: A relationship between a steel billet temperature after heating in a continuous furnace of a rolling shop of a metallurgical enterprise from a temperature in zones of this furnace and time of billet being in them is developed. This dependence is based on four-order response surface model and allows to obtain the forecast error of about 15⁰С (1,5%).

Ключевые слова: регрессионный анализ, разведочный анализ, нагревательные печи, обобщённая линейная регрессия, полиномиальная регрессия, поверхность отклика.

Keywords: regression analysis, exploration analysis, metallurgy, general linear model, polynomial regression, heating furnaces, response surface.

Введение

В технологической цепочке современного металлургического предприятия немаловажное место занимает этап производства проката. На данном этапе в том или ином виде осуществляется обработка заготовок давлением [1]. Для непрерывно литых заготовок (НЛЗ) она производится при предварительном нагреве заготовки до технологической температуры прокатки с целью снизить коэффициент сопротивления деформации металла [2].

Как правило, нагрев заготовок для дальнейшей прокатки производят в пламенных или электрических печах. В сортопрокатных цехах с начальной температурой проката близкой к 1200°С чаще всего применяют пламенные многозонные методические печи [3].

В настоящий момент существует необходимость в модели, способной описать зависимость между температурами в зонах печи в то время, когда заготовка проходила через них, и непосредственно температурой заготовки при ее попадании на прокатный стан. Данная необходимость связана с постоянным повышением требований к качеству металлопродукции и, как следствие, к необходимости более точного управления технологическим процессом [4]. Сегодня эта проблема решается попытками компьютерного моделирования как нагревательного устройства, так и непосредственно термодинамических процессов нагрева металла на физических моделях [5-6].

В то же время, современные АСУ ТП подобных нагревательных устройств собирают и архивируют большие массивы информации с нижнего (полевого) уровня автоматизации, которые служат в основном для фиксации нормального технологического режима работы системы. Предполагается, что эта информация содержит множество зависимостей, в том числе такие, которые позволят предсказать температуру заготовки на выходе печи еще до того, как она попадет в прокатные клетки.

Основываясь на опыте применения методов статистического анализа к большим объемам данных, полученных с технологических объектов [7-9], предлагается проанализировать эти данные и выявить характерные зависимости для построения модели прогнозирования температуры заготовки на выходе из печи от истории её нагрева в ней. Подобная система будет полезна как персоналу, обеспечивающему контроль качества технологического процесса, так и непосредственно операторам нагревательных устройств.

Поскольку скорость прокатки стана зависит от температуры прокатываемой заготовки (измеренной пирометром непосредственно в группе клеток), возможность прогнозировать температуру позволит регулировать скорость проката до захода металла в стан. Помимо этого, предполагается, что регулирование скорости проката с учетом модели позволит снизить риск «бурёжки» за счёт правильного выбора скорости.

Вышесказанное применимо для металлургических производств, имеющих в составе прокатные цеха, в том числе и для Оскольского электрометаллургического комбината (ОЭМК), на базе которого и предлагается рассмотреть возможность построения подобной модели.

В качестве нагревательного устройства была рассмотрена шестизонная методическая печь нагрева сортопрокатного цеха №1 АО ОЭМК (первая зона расположена над второй, далее идет пара третья-четвертая зона, затем – пятая-шестая). Каждая зона печи снабжена

датчиками температуры. Целевой параметр (температура заготовки) снимается пирометром непосредственно в черновой клетке, это позволяет считать её температурой раската, а не поверхности металла, что делает модель более информативной для принятия решений для соответствующего персонала.

1 Предварительный анализ имеющихся данных

Современные АСУ ТП нагревательных устройств (печи нагрева) для качественного управления тепловыми режимами опрашивают и сохраняют информацию с датчиков температур (термопар) в регулируемых зонах, а также осуществляют фиксацию посадки и выдачи заготовок, формируя таким образом соответствующие наборы данных (табл.1 и табл.2).

Как правило, эти данные слабоструктурированы, однако предполагается, что в результате их обработки возможно получить историю нагрева заготовки для дальнейшего исследования.

Таблица 1 Пример температурной карты

<i>Id</i>	<i>Timestamp</i>	T_1	T_2	T_3	T_4	T_5	T_6
1	21.12.2018	986	979	1211	1208	1184	1184
2	21.12.2018	986	977	1211	1206	1186	1189

Здесь: 1) *Id* – уникальный номер записи; 2) *Timestamp* – временная метка (включает в себя временной интервал с точностью до секунды); 3) T_1 , T_2 , T_3 , T_4 , T_5 и T_6 – температуры в соответствующих зонах.

Таблица 2 Пример данных по заготовкам

<i>Id</i>	<i>Date_ Create</i>	<i>Tel</i>	<i>Act</i>	<i>Plv_ Nr</i>	<i>Party</i>	<i>Unit_ Nr</i>
1	21.12.2018	216	Посад	43116	11	202
17175	19.11.2018	216	Посад	31825	12	104

Здесь: 1) *Id* – уникальный номер записи; 2) *Data_create* – дата создания записи (включая временной интервал); 3) *Tel* – номерной идентификатор состояния; 4) *Act* – маркер состояния; 5) *Plv_Nr* – номер плавки; 6) *Party* – единица прокатки; 7) *Unit_Nr* – номер заготовки плавки.

На основании временных меток в рамках исследования удалось структурировать исходные наборы данных, а также аналитически рассчитать время нахождения каждой заготовки в печи. На основании этого была сформирована сводная таблица данных (табл. 3).

Таблица 3 Фрагмент структурированной выборки

<i>FTZ,</i> мин	<i>STZ,</i> мин	<i>TTZ,</i> мин	$T_1,$ °C	$T_2,$ °C	$T_3,$ °C	$T_4,$ °C	$T_5,$ °C	$T_6,$ °C	$T_{rez},$ °C
120	80	88	978	979	1174	1171	1173	1172	1039
148	80	680	977	979	1171	1169	1172	1171	1043

Здесь: 1) *FTZ* – время, проведенное заготовкой в первой паре зон; 2) *STZ* – время во второй паре зон; 3) *TTZ* – время в третьей паре зон; 4) $T_1, T_2, T_3, T_4, T_5, T_6$ – температуры в соответствующих зонах; 5) T_{rez} – температура заготовки, снятая пирометром. Общий объем выборки составил 7880 заготовок в общей сложности за 62 дня.

Предполагалось, что для построения зависимости температуры заготовки от температур в зонах печи будет использоваться регрессионный анализ. Перед тем, как выбрать план регрессионного анализа, необходимо оценить значимость предполагаемых регрессоров для целевой переменной. А также исследовать данные на наличие мультиколлинарности, которая согласно контекстуальному анализу предметной области может возникнуть между регрессорами, описывающими температуры в парных зонах печи.

Предполагается, что вышеизложенные тесты потребуют расчета коэффициента корреляции, поэтому было решено сначала определить тип распределения регрессоров, так как параметрический критерий Пирсона требует нормальность распределения исследуемых величин [10].

Для аналитической оценки нормальности тестом Шапиро-Уилка [10] исходная выборка является слишком большой (>5000 записей на каждый регрессор). Поэтому было решено оценить нормальность, используя непараметрический метод гистограмм.

Параметры, отвечающие за время, проведенное заготовкой в каждой паре зон, распределены ненормально (рис.1 и рис.2).

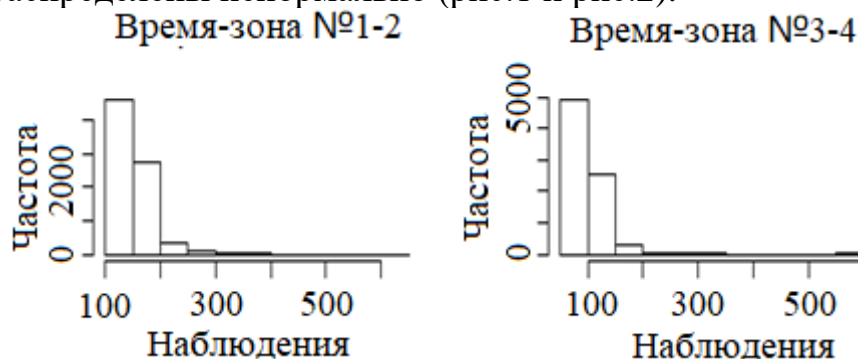


Рис. 1 Распределение параметров *FTZ* и *STZ*

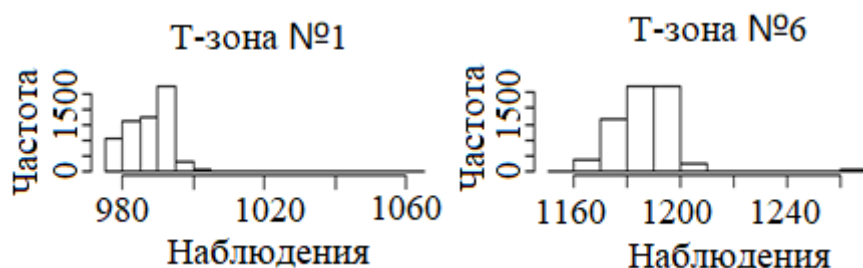


Рис. 2 Распределение *TTZ* и суммы параметров ($TTZ + STZ + FTZ$)

Параметры, отвечающие за температуры в соответствующих зонах, также распределены ненормально, о чем свидетельствуют полученные

гистограммы (рис.3). Остальные параметры температур также распределены ненормально.

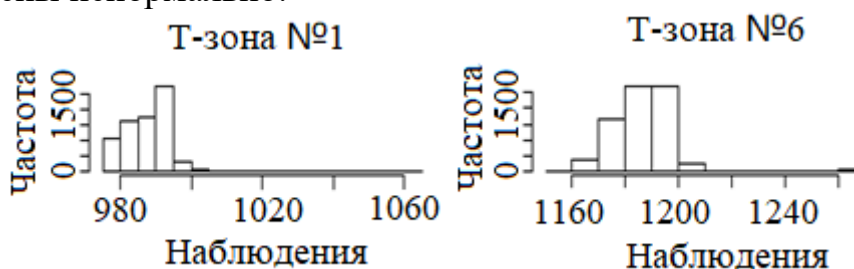


Рис. 3 Распределение параметров температур

Поскольку данные распределены ненормально, то для корреляционного теста был выбран непараметрический критерий Спирмена (1) [11].

$$\rho = 1 - ((6 * \sum d^2) / n * (n^2 - 1)) \quad (1)$$

Здесь: d – разность рангов каждой пары значений, n – размерность выборки. Для оценки уровня значимости соответствующих коэффициентов использовался t -критерий Стюдента (2).

$$t = \rho * \sqrt{n-2} / \sqrt{1-\rho^2} \quad (2)$$

Результат расчета (табл. 4 и табл.5) показал, что все имеющиеся непрерывные переменные имеют слабую корреляционную связь с целевой переменной (<0.3). При этом нулевая гипотеза, оцененная t -статистикой для каждого из коэффициентов корреляции, $H_0: r = 0$ (корреляция отсутствует) отвергается.

Таблица 4 Критерий Спирмена для температурных параметров

Регрессоры	ρ	t –значения
T_1	0.14130114	$4.6897 * 10^{-35}$
T_2	0.20038445	$1.0043 * 10^{-70}$
T_3	0.19202570	$6.9654 * 10^{-65}$
T_4	0.17032720	$5.9658 * 10^{-51}$
T_5	0.25125971	$7.8308 * 10^{-112}$
T_6	0.23508100	$1.3983 * 10^{-97}$

Таблица 5 Критерий Спирмена для временных параметров

Регрессоры	ρ	t –значения
FTZ	0.05821654	$2.3151 * 10^{-6}$
STZ	0.12429026	$3.6174 * 10^{-27}$
TTZ	0.12203112	$3.2576 * 10^{-26}$

Исходя из того, что каждый отдельный регрессор слабо коррелирует с целевой переменной, и одновременно с этим является статистически значимым, было принято решение не исключать ни один из параметров

выборки на данном этапе, предполагая в дальнейшем рассмотрение их взаимодействий.

Поскольку из контекста предметной области известно, что зоны печи регулируются попарно (четные и нечетные зоны), имеется риск полной коллинеарности соответствующих параметров. Тест на мультиколлинеарность проводился при помощи фактора инфляции дисперсий (3) с контрольным порогом $VIF = 5$ [11].

$$VIF_j = 1 / (1 - R_j^2) \quad (3)$$

R – коэффициент линейной корреляции. Результат теста (табл.6) подтвердил контекстуальные предположения о наличии мультиколлинеарности.

Таблица 6 Результаты теста на мультиколлинеарность

Параметр	R	VIF
T_1 (с учетом T_2)	0.9173339	12.09686
T_2 (с учетом T_1)	0.9084988	10.92882
T_3 (с учетом T_4)	0.9859616	71.23302
T_4 (с учетом T_3)	0.985129	67.24513
T_5 (с учетом T_6)	0.8973527	9.7421
T_6 (с учетом T_5)	0.9220852	12.83453

Повторный тест (3) без учета парных параметров температур (табл.7) свидетельствует об отсутствии внутренних линейных зависимостей между остальными параметрами.

Таблица 7 Тест VIF для усредненных температур

Параметр	R	VIF
T_1 (Без учета T_2)	0.528394	2.120414
T_2 (Без учета T_1)	0.4779898	1.915671
T_3 (Без учета T_4)	0.739756	3.842548
T_4 (Без учета T_3)	0.7243226	3.627428
T_5 (Без учета T_6)	0.7188037	3.556235
T_6 (Без учета T_5)	0.7865568	4.685088

Основываясь на данном факте, было решено усреднить значения параметров $T_1, T_2, T_3, T_4, T_5, T_6$, отвечающие за значения температур в парных зонах (№1 и №2; №3 и №4; №5 и №6), получив параметры T_{12}, T_{34}, T_{56} . Вследствие чего исходная выборка приобрела следующий вид (табл.8).

Входными параметрами модели будут являться: 1) FTZ, STZ, TTZ – время, проведенное заготовкой в первой, второй и третьей паре зон соответственно; 2) T_{12}, T_{34}, T_{56} – температуры в соответствующих парах зон. Выходной параметр: T_{rez} – температура раската заготовки.

Таблица 8 Фрагмент полученной выборки

<i>FTZ</i> , мин	<i>STZ</i> , мин	<i>TTZ</i> , мин	<i>T</i> ₁₂ , °C	<i>T</i> ₃₄ , °C	<i>T</i> ₅₆ , °C	<i>T</i> _{рез} , °C
120	80	88	979	1173	1172	1040
148	80	680	976	1172	1173	1042

Основываясь на полученной выборке, было решено промоделировать искомую зависимость, используя различные планы регрессионного анализа.

2. Построение искомой зависимости различными методами

Основываясь на проведенном выше разведочном анализе данных, для первого этапа моделирования был выбран план обобщенной линейной регрессии (4), конечная модель которого зависит от типа распределения исследуемых независимых переменных [12].

$$E(Y) = g^{-1}(X\beta) \quad (4)$$

Y – матрица выходов, где зависимая переменная принадлежит экспоненциальному семейству $E(Y)$; g – функция связи; X – матрица независимых переменных; β – матрица коэффициентов регрессии.

Поскольку исходная выборка содержит большее количество наблюдений для каждого регрессора, то в соответствии с центральной предельной теоремой было решено допустить нормальные распределения всех исследуемых регрессоров и выбрать общие линейные модели с характерной функцией связи (5).

$$g(\mu) = \mu \quad (5)$$

μ – математическое ожидание, а сама функция осуществляет тождественную линейную связь между предикторами и средним значением распределения данного предиктора.

Предварительно исходная выборка была разбита на три части: обучающую (~6000 значений), тестовую (~1500 значений) и валидационную (~380 значений).

Расчет параметров модели здесь и далее осуществлялся средствами языка *R* и среды разработки *RStudio* с соответствующими интегрированными библиотеками. Также для данной модели были взяты только главные эффекты (модель 1-го порядка) (6).

$$f(y_j) = \beta_{0j} + \sum_{i=1}^6 \beta_i * x_{ij} \quad (6)$$

β_i – i -й регрессионный коэффициент; x_{ij} – это j -е значение i -ой независимой переменной (см. табл. 8); y_j – j -е значение зависимой переменной. β_{0j} – свободный член j -го наблюдения; $f(y_j) = y_j$ – линейность выхода, обусловленная (5).

Помимо общей линейной модели, также была смоделирована дробная факторная регрессия (без учета эффектов 2-го порядка) (7).

$$f(y_j) = \sum_{i=1}^6 \beta_i * x_i + \sum_{g=1}^5 \left(\sum_{k=1}^g (\beta_{gk} * x_{6-g} * x_{7-k}) \right) \quad (7)$$

Где: g и k – коэффициенты, позволяющие выбрать только эффекты первого порядка, перебрав их сочетания без повторений. Полученная по результатам расчета аналитическая оценка прогноза рассмотренных выше моделей на валидационной выборке представлена в таблице 9.

Таблица 9 Аналитическая оценка модели

	Ср. ошибка, °C	Макс. ошибка, °C	Мин. ошибка, °C
Модель (6)	22.5	85,5	0.006
Модель (7)	20.1	85	0.01

Было выдвинуто предположение, что среднюю ошибку можно уменьшить, усложнив модель. В результате множественных экспериментов было решено использовать полиномиальную регрессию (8).

$$y_j = \beta_{0j} + \sum_{i=1}^6 \sum_{k=1}^p \beta_{ik} * x_{ij}^k \quad (8)$$

p – степень полинома.

После оценки полученных результатов (табл.10) было решено использовать полином 4-й степени, так как он имеет наименьшую среднюю ошибку при прогнозе. Увеличение степени полинома приводит к росту расхождения с эталонными значениями и, как следствие, к повышению средней ошибки модели на валидационной выборке. Полиномиальная модель 2-го порядка не рассматривалась, так как на стадии экспериментов с дробной факторной регрессией эффекты второго порядка (в том числе и для взаимодействия регрессоров) показали себя хуже, чем первого.

Таблица 10 Численная оценка полиномиальных моделей

Степень полинома	Ср. ошибка, °C	Макс. ошибка, °C	Мин. ошибка, °C
3	19.03	86.14	0.042
4	17.025	70.4	0.042
5	17.99	74.49	0.15
6	18.35	87.33	0.096

В результате экспериментов с полиномиальной моделью (8) была получена модель дробной регрессии поверхности отклика (9) путем итеративного исключения взаимодействий регрессоров.

$$f(y_j) = \beta_{0j} + \sum_{i=1}^6 \sum_{k=1}^p \beta_{ik} * x_{ij}^k + \sum_{m=1}^4 \sum_{k=1}^p \beta_{mk} * (x_{1j}^m * x_{3j}^{(p+1)-k}) \quad (9)$$

$p = 4$ – степень используемого полинома; $m = 4$ – коэффициент, позволяющий определить наивысший порядок эффектов для сочетаний выбранных взаимодействий; x_{1j} – это j -е значение параметра FTZ (см.табл.8); x_{3j} – это j -е значение параметра TTZ (см.табл.8). В качестве регуляризатора использовалась обобщенная регрессия «эластичной сети» (9).

$$\hat{\beta} = \arg \min \left(\sum_{j=1}^n (y_j - f(y_j))^2 + \lambda_1 (\beta)^2 + \lambda_2 |\beta| \right) \quad (10)$$

Где: n – размерность выборки; $\lambda_1 = 0.001$ и $\lambda_2 = 0.5$ – штрафующие коэффициенты; $f(y)$ – регрессионная модель (9); β – параметры этой модели.

Результат аппроксимации полученной модели (9) на обучающей выборке (рис.4) показал достаточно высокую степень соответствия эталонным значениям, однако температуры ниже математического ожидания аппроксимируются плохо. Величина ошибки аппроксимации обучающей выборки составила $\sim 19^{\circ}\text{C}$ в среднем от эталонных значений.

Фрагмент аппроксимации на обучающей выборке

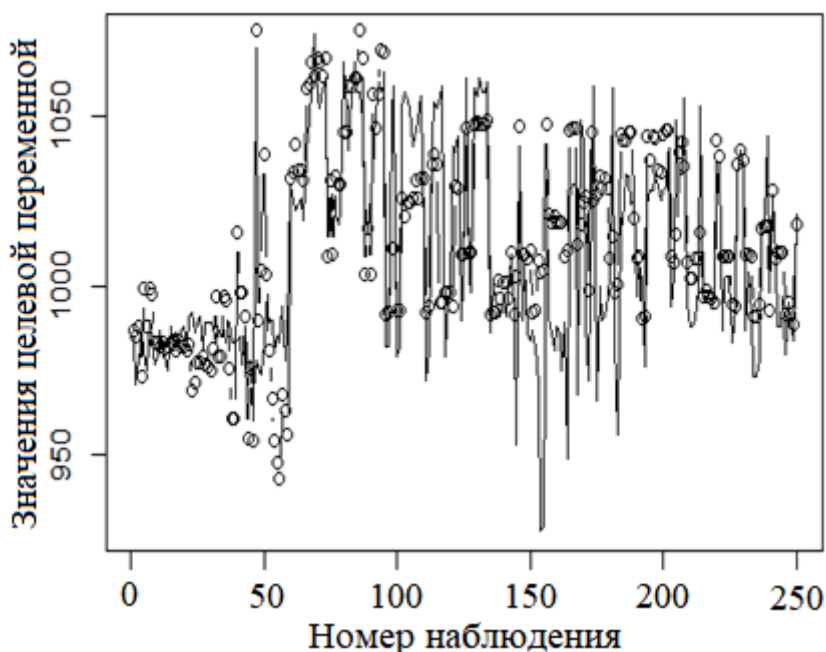


Рис. 4 «←» - График эталонных значений. «o-o» - аппроксимация модели

Точность аппроксимации на тестовом интервале (рис.5) немного выше, чем в пределах всей модели (ошибка $\sim 15^{\circ}\text{C}$ в среднем от эталонных против $\sim 19^{\circ}\text{C}$).

Фрагмент аппроксимации на тестовой выборке

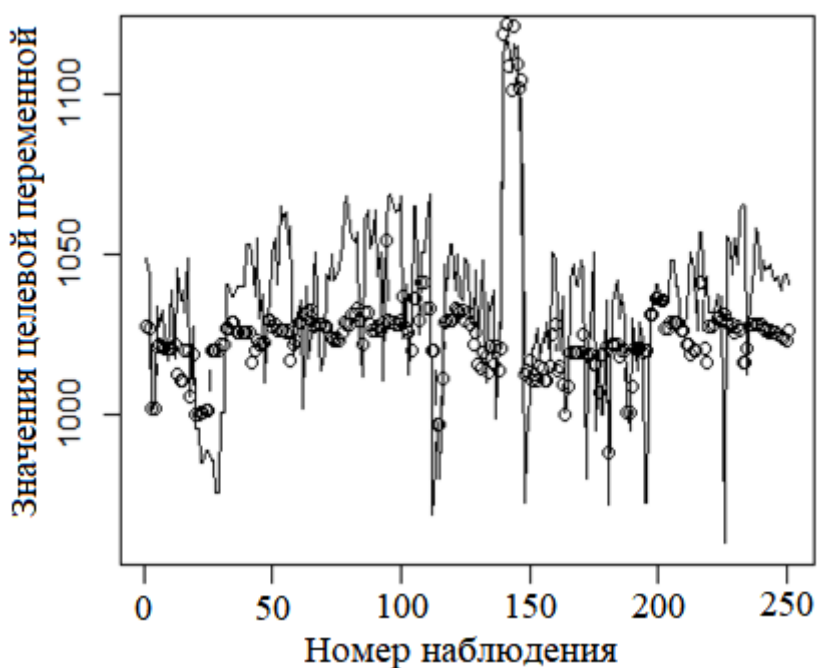


Рис. 5 Аппроксимация на тестовом интервале

Прогноз на валидационной выборке

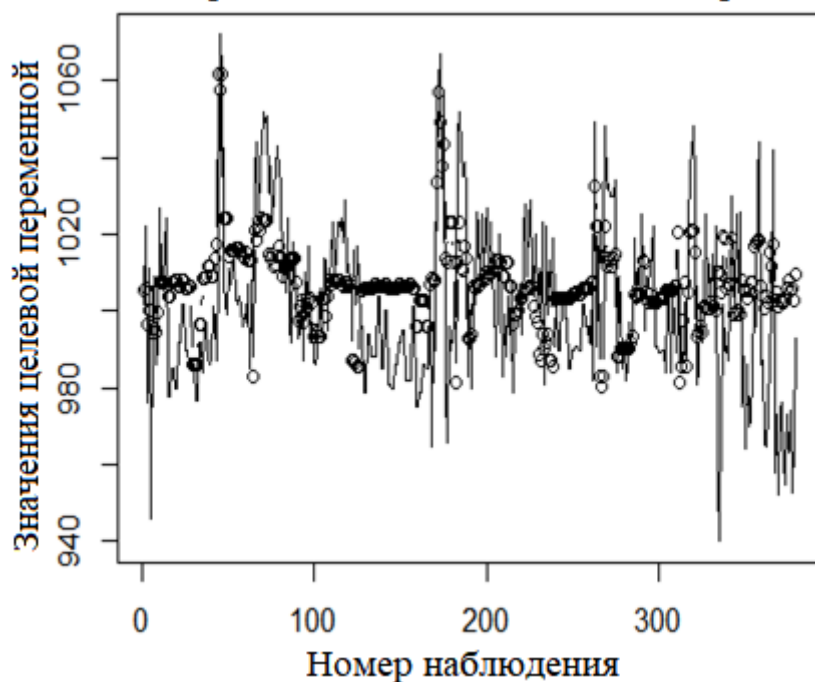


Рис. 6 Апробация на валидационной выборке

В результате апробации (рис.6), полученная модель (9) хорошо предсказывает значения, лежащие в пределах математического ожидания (или выше), при этом в значительной мере игнорируются значения ниже математического ожидания.

Аналитическая оценка результатов апробаций (табл. 11) показывает, что полученная модель (9) имеет преимущество в точности прогноза по сравнению с полиномиальной моделью (8) и общей линейной моделью (6) и моделью множественных взаимодействий регрессоров (7).

На основании результатов сравнения моделей (табл.11), взаимодействия регрессоров, отвечающих за время, проведенное заготовкой в первой и третьей паре зон соответственно, предполагается считать более значимым, чем любые другие взаимодействия регрессоров. Учет слабо значимых взаимодействий в моделях (7) и (8) приводит к уменьшению общей точности прогноза.

Таблица 11. Численное сравнение моделей

	Средняя ошибка, С°	Макс. ошибка, С°	Мин. ошибка, С°
ЛМ	22.5	85.5	0.006
ДФМ	20.1	85	0.01
ПМ	17.025	70.4	0.042
ЭМ	15.968	69.84	0.018

ЛМ – линейная модель (6), ДФМ – модель дробной факторной регрессии (7); ПМ – полиномиальная модель (8), ЭМ – полученная экспериментально модель (9). Численная оценка модели (9) представлена в таблице 12. Здесь $1Q$ – первый квартиль, $3Q$ – третий квартиль.

Таблица 12. Числовые параметры аппроксимации

	Мин.	$1Q$	Меди- ана	Сред- нее	$3Q$	Макс.
Эталон	928	994	1016	1018	1042	1125
Модель	943	1010	1019	1018	1026	1130

Выводы

Полиномиальная модель, полученная в результате множественных экспериментов, показала приемлемую точность для решения проблемы, связанной с прогнозным управлением скоростью проката ($\sim 15^\circ\text{C}$). Предполагается, что невозможность модели обобщать значения целевой переменной, находящиеся ниже математического ожидания, может быть связана с ошибками при предобработке и структуризации данных.

Предполагается, что применение интеллектуальных методов анализа данных (например, искусственных нейронных сетей) позволит добиться увеличения точности прогноза на имеющихся данных. Также предполагается, что использование выборки с более точными значениями регрессоров позволит добиться лучшего результата и при классических регрессионных планах.

Список литературы

1. Пейсахов А.М., Кучер А.М. Металловедение и технология конструкционных материалов. Учебник / А.М. Пейсахов, А.М. Кучер. – СПб.: Изд-во Михайлова В.А., 2003. – 407 с.
2. Кучер А.М. Технология металлов / А.М Кучер – 4-е изд., перераб. и доп. – Л: Машиностроение, Ленингр. отделение, 1987. – 214 с.
3. Шухов Ю. В., Еленев С.А. Холодная штамповка / Ю.В Шухов, С.А Еленев – М: Изд-во Высшая школа, 1977. – 208 с.
4. Сердобинцев, Ю. П. Перспективные направления повышения качества функционирования технологического оборудования: монография / Ю. П. Сердобинцев, О. В. Бурлаченко, А. Г. Схиртладзе. – Старый Оскол: ООО «Тонкие наукоемкие технологии», – 2010. – 412 с.
5. Кухтик М.П., Сердобинцев Ю.П. Математическое моделирование процесса нагрева слябов в методической печи / М.П. Кухтик, Ю.П. Сердобинцев // Известия ВолгГТУ. – 2011. – №13 (86). – С. 80-83.
6. Крючков О.Б., Копасов Е.А., Роненко В.О. Оптимизация нагрева теплотехнически массивных заготовок с использованием физического моделирования / О.Б. Крючков, Е.А Копасов, В.О. Роненко // Известия ВолгГТУ. – 2013. – №6(109). – С. 135-137.
7. Сараев П.В. Последовательный подход к оперативному анализу данных в крупномасштабных производствах. / П.В. Сараев // Управление развитием крупномасштабных систем (mlsd'2011). – М: ИПУ РАН, 2011. – С. 299-301.
8. Кычкин А.В., Бочкарев С.В Интеллектуализация мониторинга и анализа данных о потреблении тепловой энергии на производстве / А.В. Кычкин, С.В. Бочкарев // Энергобезопасность и энергосбережение. – 2017. – № 5. – С. 30-36.
9. Коллерт М. и др. Компьютерная поддержка в производстве полимерных пленок на основе интеллектуального анализа данных / М. Коллерт, М.А. Тетерин, К. Кёниг, Т.Б Чистякова // Математические методы в технике и технологиях – ММТТ. – 2015. – № 3 (73). – С. 20-25.
10. Кобзарь А. И. Прикладная математическая статистика. / А.И Кобзарь // Справочник для инженеров и научных работников. – М.: Физматлит, 2006. – 816 с.
11. Дрейпер Н., Смит Г. Прикладной регрессионный анализ / Н. Дрейпер., Г. Смит. – М: Вильямс, 2016. – 912 с.
12. Corder, G. W. Nonparametric Statistics: A Step-by-Step Approach / G. W. Corder, D. I. Foreman. – New Jersey: Wiley, 2014. – 190 p.