УДК 674* DOI:10.18324/2077-5415-2020-4-127-133

Разработка математической модели для определения параметров оптимальных режимов процессов отделки древесины лакокрасочными материалами

В.А. Соколова

Санкт-Петербургский государственный лесотехнический университет им. С.М. Кирова,
Институтский пер., 5, Санкт-Петербург, Россия
sokolova_vika@inbox.ru
https://orcid.org/0000-0001-6880-445X
Статья поступила 01.10.2020, принята 23.10.2020

В статье рассматриваются вопросы оптимизации процесса подготовки древесины и ее отделки лакокрасочными материалами. Приведены методики и результаты экспериментов и моделирования по нескольким факторам и по трем основным показателям лакокрасочных покрытий. Однако данная методика является общей для любого количества показателей и факторов. Предлагаемая методика оптимизации использует линейные модели, зависимости между показателями лакокрасочных покрытий и факторами, которые влияют на эти показатели. Это связано с тем, что основной задачей работы является не определение точной зависимости, а оптимальный выбор наиболее значимых факторов, что может быть сделано только с использованием линейных моделей. Только в линейных моделях коэффициенты при факторах определяют их значимость. Значения коэффициентов по модулю определяют их значимость. Предлагаемые математические модели не являются в полной степени линейными. Метод Лассо и метод гребневой регрессии используют регуляризацию для сглаживания или ограничения множества решений, что делает модель полулинейной. Кроме того, представление Бахадура использует пары факторов, что существенно расширяет класс используемых функций, но, тем не менее, в основе всех этих моделей все равно лежат линейные функции. Разработанная методика позволяет с наименьшими затратами и с использованием автоматизированного расчета менять режимы технологий отделки древесины на основе предпочтений с точки зрения значимости различных целевых показателей. Метод парных сравнений Саати позволяет существенно упростить эту процедуру, используя только оценки парных сравнений.

Ключевые слова: алгоритм; модель; метод Лассо; гребневая регрессия; отделка древесины; лакокрасочные материалы.

Development of a mathematical model for determining the parameters of optimal modes of wood finishing processes with paint and varnish materials

V.A. Sokolova

St. Petersburg State Forest Technical University under name of S.M. Kirov; 5, Institutsky per., St. Petersburg, Russia
sokolova_vika@inbox.ru
https://orcid.org/0000-0001-6880-445X
Received 01.10.2020, accepted 23.10.2020

The article deals with the optimization of the process of wood preparation and its finishing with paints and varnishes. Methods and results of experiments and modeling are presented for several factors and for three main indicators of paint and varnish coatings. However, this technique is common for any number of indicators and factors. The proposed optimization technique uses linear models, the relationship between the indicators of paints and varnishes and the factors that affect these indicators. This is due to the fact that the main task of the work is not to determine the exact relationship, but to select the most significant factors, which can only be done when using linear models. Only in linear models, the coefficients with factors determine their significance. The values of the coefficients in modulus determine their significance. The proposed mathematical models are not completely linear. The Lasso and ridge regression methods use regularization to smooth or restrict the set of solutions, making the model semi-linear. In addition, the Bahadur representation uses pairs of factors, which significantly expands the class of functions used, but, nevertheless, all these models are still based on linear functions. The developed methodology allows, at the lowest cost and using automated calculation, to change the modes of wood finishing technologies based on preferences in terms of the significance of various target indicators. The Saaty paired comparison method can significantly simplify this procedure by using only paired comparison estimates.

Keywords: algorithm; model; Lasso method; ridge regression; wood finishing; paint materials.

Введение. Огромное количество методов и алгоритмов отбора факторов основано на предположении, что множество факторов содержит несущественные и избыточные элементы. Грубо говоря, несущественные факторы не несут какой-либо полезной информации, улучшающей соответствующий классификатор. Избыточные факторы содержат информацию, которая уже имеет место в более информативных факторах. Удаление несущественных и избыточных факторов — это одна из целей задачи отбора факторов.

Три основных подхода используется для отбора факторов. Первый подход, называемый фильтрацией, использует статистические свойства факторов и основан на предположении, что эти свойства различны для различных классов. Если некоторый фактор имеет различные распределения вероятностей для двух классов, то методы фильтрации основаны на определении расстояния между распределениями. В качестве одного из наиболее распространенных показателей различия факторов или их распределений вероятностей для двух классов служит критерий Фишера:

$$F(i)=\left|\frac{\mu_i^+-\mu_i^-}{(\sigma_i^+)^2+(\sigma_i^-)^2}\right|. \tag{1}$$

Здесь μ_i^+ и μ_i^- — средние значения *i*-го фактора для классов y = –1 и y = +1; σ_i^+ и σ_i^- — соответствующие среднеквадратические отклонения. Чем больше различие между значениями фактора для двух классов, тем более информативен соответствующий фактор и тем больше значение критерия Фишера. Другими показателями, характеризующими факторы, являются t-статистика, расстояние Кульбака – Лейблера и др.

Второй подход, охватывающий так называемые методы упаковки, во многих случаях обеспечивает более точное решение, чем методы фильтрации, но одновременно требует больших вычислительных ресурсов. Методы упаковки используют метод классификации, например, метод опорных векторов (SVM) для определения, как меняется точность классификации при удалении или добавлении тех или иных факторов. Одним из наиболее популярных методов упаковки является метод рекурсивного удаления факторов (SVM-RFE).

Третий подход использует встроенные методы, которые позволяют отобрать факторы в процессе классификации. Наиболее распространенными встроенными методами являются методы преобразования целевой функции в SVM, например, использование регуляризации в виде единичной нормы $\|\mathbf{w}\|_1=|w_1|+...+|w_m|$ (l_1-SVM) или нулевой нормы $\|\mathbf{w}\|_0=|\{i\ :\ w_i\neq 0\}|$ (l_0-SVM) вместо квадратичной.

Одним из наиболее известных методов в рамках третьего подхода является метод Лассо [1; 2].

Методика исследования. Модификации метода Лассо являются эффективным инструментом для решения задачи отбора значимых факторов и поиска регрессионной модели. Их эффективность была экспериментально доказана многими работами [3–5]. Однако каждое реальное приложение имеет некоторые особенности, учет которых позволил бы улучшить регрессионную модель. Перечислим эти особенности в рассматриваемых задачах.

Прежде всего, наша цель заключается не в том, чтобы найти наилучшую регрессионную модель для имеющейся информации, а выбрать значимый фактор, который влияет на наименьшие значения результатов исследований физико-механических свойств лакокрасочных покрытий. Это не означает, что полностью регрессионная модель не важна для нас. Необходимо комбинировать две цели: отбор значимых факторов, влияющих на результаты исследований, и построение регрессионной модели в целом. Это может быть осуществлено введением весовых штрафных слагаемых специального вида в методе Лассо. Этот вид должен учитывать, прежде всего, минимальные значения результатов исследований физико-механических свойств лакокрасочных покрытий. Меньшие значения результатов исследований образуют большие веса, в то время как большие значения должны быть также учтены и рассмотрены. Отсюда следует, что веса не должны приписываться значениям результатов исследований. Их можно приписывать ожидаемым значениям в соответствии с функцией распределения вероятностей, определенной на значениях совокупности факторов [6].

Значения совокупности факторов, соответствующие каждому значимому фактору, в рассматриваемом приложении образуют двоичный вектор. Зависимость значимых факторов ведет к зависимости соответствующих двоичных векторов, которая может быть оценена.

Вероятности факторов и корреляции неявно оказывают влияние на наименьшие значения результатов исследований физико-механических свойств лакокрасочных покрытий.

Следуя идеям, предложенным в работах [7–12], и развивая их, определим новые веса в штрафном слагаемом в методе Лассо. Основная идея, положенная в основу их определения, состоит в следующем. Мы определяем средний вклад каждого значимого фактора в математическое ожидание значений результатов. Эти вклады являются ничем иным, как весами в адаптивном методе Лассо. Они должны учитывать вероятности факторов, корреляции между значимым фактором и значениями результатов исследований. Следующий вопрос заключается в том, как определить средний вклад каждого значимого фактора. Это может быть сделано следующим образом.

Для каждой пары значимых факторов X_k и X_i вычисляем совместную вероятность $\pi(x_{kj},x_{ij})$ *j*-го экземпляра выборки $i=1,...,p\quad i\neq k$. Это можно сделать, используя представление Бахадура. Представление Бахадура позволяет учесть корреляцию между двоичными переменными. В случае двух двоичных переменных с номерами *k* и *i* совместная вероятность определяется как:

$$\pi(x_k,x_i)=\pi_k^{x_k}(1-\pi_k)^{1-x_k}\cdot\pi_i^{x_i}(1-\pi_i)^{1-x_i}\cdot\ 1+\rho_{ki}u_k u_i\ . \tag{2}$$

Здесь π_k — вероятность *k*-го значимого фактора, т. е. $\pi_k=\Pr\{x_k=1\}$; ρ_{ki} — коэффициент корреляции между *k*-м и *i*-м значимыми факторами, который определяется как $\rho_{ki}=\mathrm{E}\ U_k U_i$, где случайная стандартизованная величина U_k принимает значения u_k таким образом, что:

$$U_k = \frac{X_k - \pi_k}{\sqrt{\pi_k(1-\pi_k)}},\ u_k = \frac{x_k - \pi_k}{\sqrt{\pi_k(1-\pi_k)}}. \qquad (3)$$

Следует отметить, что представление Бахадура может быть записано также для совместных вероятностей трех, четырех и т. д. случайных величин.

Основная идея использования совместных вероятностей заключается в учете корреляций между случайными величинами. Чтобы определить средний вклад *k*-го значимого фактора в математическое ожидание значений результатов исследований, мы рассмотрим все возможные пары значимых факторов, такие, что один из значимых факторов в каждой паре является *k*-м значимым фактором, т. е. нам интересно рассмотреть $p-1$ пар значимых факторов с номерами $(k,1),...,(k,k-1),(k,k+1),...,(k,p)$. Каждая пара определяет ожидаемое значение результатов исследований R_{ki}, соответствующее этой паре *k*-го и i -го значимых факторов следующим образом:

$$R_{ki} = \frac{\sum_{j=1}^{n}\pi(x_{kj}, x_{ij})y_j}{\sum_{j=1}^{n}\pi(x_{kj}, x_{ij})}. \qquad (4)$$

Величина может рассматриваться как вклад *k*-го и i-го значимых факторов в ожидаемое значение результатов исследований.

Тогда вклад *k*-го значимого фактора в ожидаемое значение результатов исследований может быть определен при помощи использования средних вкладов R_{ki}, т. е. он вычисляется как:

$$R_k = \frac{1}{p-1}\sum_{i=1, i\neq k}^{p} R_{ki}. \qquad (5)$$

Таким образом, полученные веса учитывают корреляцию между значимыми факторами, вероятности факторов, двоичный характер данных и тот факт, что наименьшие значения результатов исследований физико-механических свойств лакокрасочных покрытий являются более значимыми по сравнению с другими значениями, так как наша цель — поиск значимых факторов с наибольшим влиянием. В то же время, нет необходимости непосредственно использовать полученные веса и реализовывать алгоритм адаптивного Лассо. Этот алгоритм может быть преобразован в стандартный метод Лассо посредством введения новых значений совокупности факторов [13–16].

Результаты исследования. Построение модели Лассо и гребневой регрессии проводили для образцов древесины с лакокрасочными покрытиями (табл. 1).

Таблица 1. Образцы древесины с лакокрасочными покрытиями

Номер образца	Подложка	Лакокрасочное покрытие
1	Древесина сосны, обработанная ультразвуком	Водно-дисперсионный лакокрасочный материал
2	Древесина сосны, обработанная ультразвуком	Уретановый лакокрасочный материал
3	Древесина осины, обработанная ультразвуком	Водно-дисперсионный лакокрасочный материал
4	Древесина осины, обработанная ультразвуком	Уретановый лакокрасочный материал

В статье представлены результаты исследований образцов № 1 и 3 — древесина сосны и осины с водно-дисперсионными лакокрасочными материалами.

Предлагаемая методика реализована в виде программного комплекса, который включает таблицы в Excel и программные модули на языке программирования R, который является одним из наиболее популярных и эффективных языков для математических и статистических расчетов. Программа модификации алгоритма Бахадура является элементом общего программного пакета оптимального режима автоматизированного процесса технологии отделки древесины.

Используем парные сравнения критериев.

Таблица 2. Парные сравнения целевых показателей по методу Саати

Показатель	Адгезия	Теплостойкость	Водостойкость
Адгезия	1	8	6
Теплостойкость	1/8	1	3
Водостойкость	1/6	1/3	1

Используя парные сравнения критериев, получаем значения весов и нормированных весов целевых критериев (табл. 3).

Таблица 3. Значения весов и нормированных весов целевых критериев

W	Нормирование
3,634241186	0,767213263
0,721124785	0,152234393
0,381571414	0,080552345

Результаты исследований для образца № 1.

В результате использования парных сравнений критериев получены показатели, представленные в табл. 4.

Таблица 4. Определение показателей по методу Саати с использованием матрицы парных сравнений

Y_1	Y_2	Y_3	Весовая композиция
0,667	1,000	0,615	0,713
0,367	0,778	0,769	0,462
0,533	0,500	0,692	0,541
0,967	0,222	0,462	0,813
0,667	0,333	0,462	0,599
1,000	0,222	0,385	0,832
0,367	0,000	0,769	0,343
0,700	0,167	0,000	0,562
0,633	0,778	1,000	0,685
0,667	0,444	0,769	0,641
0,033	0,278	0,462	0,105
0,200	0,000	0,462	0,191
0,000	0,444	1,000	0,148

0,367	0,222	0,538	0,359
0,200	0,056	0,231	0,180
0,333	0,000	0,154	0,268

Модели зависимости адгезии от факторов представлены в табл. 5.

Таблица 5. Модели зависимости адгезии от факторов (образец № 1) по методам Лассо и гребневой регрессии

Показатель «адгезия»	
Метод Лассо	Метод гребневой регрессии
$y_1 = 0{,}849 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4$	$y_1 = 0{,}635 \cdot x_1 + 0{,}251 \cdot x_2 + 0{,}024 \cdot x_3 + 0{,}002 \cdot x_4$

Из табл. 5 следует, что два метода, Лассо и гребневая регрессия, показали, что наиболее значимым фактором является вязкость. Расход лакокрасочного материала также оказывает влияние, но существенно меньше.

Составляем модели корреляций пар факторов, чтобы исследовать, какие факторы совместно существенно оказывают влияние на адгезию:

$$y_1 = \alpha_{12} \cdot x_1 \cdot x_2 + \alpha_{13} \cdot x_1 x_3 + \ldots + \alpha_{34} \cdot x_3 x_4. \quad (6)$$

Используем модификацию алгоритма Бахадура и алгоритм Лассо с парами переменных $X_i \cdot X_j$, $i = 1,\ldots,4$, $j \neq I$, $j = 1,\ldots,\ldots 4$.

Модели зависимости адгезии от пар коррелированных факторов представлены в табл. 6.

Таблица 6. Модели зависимости адгезии от пар коррелированных факторов для образца № 1 по методу Лассо и алгоритму Бахадура

Метод Лассо	Алгоритм Бахадура
$\alpha_{13} = 0{,}681$ $\alpha_{12} = 0{,}294$ все остальные $\alpha_{ij} = 0$	$\alpha_{13} = 0{,}325$ $\alpha_{12} = 0{,}288$ все остальные $\alpha_{ij} = 0$

Из табл. 6 видно, что предлагаемый метод на основе модификации алгоритма Бахадура показывает для пар факторов аналогичные результаты, как и метод Лассо, но метод Лассо более сильно выделяет пару «вязкость – температура сушки», в то время как алгоритм Бахадура практически уравнивает эту пару с парой «вязкость – расход лакокрасочного материала».

Интересно отметить, что обе модели показали, что пара «вязкость – температура сушки» в комбинации дает наиболее сильный эффект, чем по отдельности вязкость и расход лакокрасочного материала, несмотря на то, что последнее в предположении их зависимости более значимо.

Модели зависимости теплостойкости от факторов представлены в табл. 7.

Таблица 7. Модели зависимости теплостойкости от факторов (образец № 1) по методам Лассо и гребневой регрессии

Показатель «теплостойкость»	
Метод Лассо	Метод гребневой регрессии
$y_2 = 0 \cdot x_1 + 0{,}532 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4$	$y_2 = 0{,}058 \cdot x_1 + 0{,}734 \cdot x_2 + 0{,}335 \cdot x_3 + 0 \cdot x_4$

Из табл. 7 следует, что два метода, Лассо и гребневая регрессия, показали, что наиболее значимым фактором является расход лакокрасочного материала. Температура сушки также может влиять, но существенно меньше.

Модели зависимости теплостойкости от пар коррелированных факторов представлены в табл. 8.

Таблица 8. Модели зависимости теплостойкости от пар коррелированных факторов для образца № 1 по методу Лассо и алгоритму Бахадура

Метод Лассо	Алгоритм Бахадура
$\alpha_{23} = 0{,}714$ все остальные $\alpha_{ij} = 0$	$\alpha_{23} = 0{,}541$ все остальные $\alpha_{ij} = 0$

Из табл. 8 видно, что для теплостойкости результаты анализа пар факторов полностью совпадают с анализом отдельных факторов.

Интересно отметить, что обычная регрессия для такой выборки дает неправильные результаты, показывая, что температура сушки и вязкость влияют больше, чем расход лакокрасочного материала, что не соответствует физической интерпретации.

Модели зависимости водостойкости от факторов представлены в табл. 9.

Таблица 9. Модели зависимости водостойкости от факторов (образец № 1) по методам Лассо и гребневой регрессии

Показатель «водостойкость»	
Метод Лассо	Метод гребневой регрессии
$y_3 = 0 \cdot x_1 + 0{,}418 \cdot x_2 + 0{,}301 \cdot x_3 + 0 \cdot x_4$	$y_3 = 0{,}214 \cdot x_1 + 0{,}735 \cdot x_2 + 0{,}704 \cdot x_3 + 0{,}348 \cdot x_4$

Из табл. 9 следует, что два метода, Лассо и гребневая регрессия, показали, что наиболее значимым фактором является расход лакокрасочного материала. Температура сушки также может влиять, но незначительно.

Модели зависимости водостойкости от пар коррелированных факторов представлены в табл. 10.

Таблица 10. Модели зависимости водостойкости от пар коррелированных факторов для образца № 1 по методу Лассо и алгоритму Бахадура

Метод Лассо	Алгоритм Бахадура
$\alpha_{23} = 0{,}618$ $\alpha_{24} = 0{,}101$ все остальные $\alpha_{ij} = 0$	$\alpha_{23} = 0{,}352$ $\alpha_{24} = 0{,}19$ $\alpha_{12} = 0{,}094$ все остальные $\alpha_{ij} = 0$

Из табл. 10 видно, что исследование пар факторов показало аналогичные результаты, но модель алгоритма Бахадура выявила пару «вязкость – расход лакокрасочного материала», также влияющий на водостойкость. Также можно отметить влияние пары «расход лакокрасочного материала – скорость подачи».

Интересно отметить, что, в отличие от теплостойкости, где расход оказывал более сильное влияние, здесь наблюдается примерно одинаковое влияние. Кроме того, гребневая регрессия показала, что и скорость подачи тоже стала одним из значимых факторов, несмотря на то, что его влияние существенно меньше, чем влияние факторов «расход лакокрасочного материала» и «температура сушки».

Строим модель обобщенной характеристики свойств лакокрасочных покрытий с учетом предпочтений производственного процесса. Модели зависимости обобщенной характеристики от факторов представлены в табл. 11.

Таблица 11. Модели зависимости обобщенной характеристики от факторов для образца № 1

Метод Лассо	Метод гребневой регрессии
$y = 0{,}629 \cdot x_1 + 0{,}361 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4$	$y = 0{,}482 \cdot x_1 + 0{,}409 \cdot x_2 + 0{,}118 \cdot x_3 + 0{,}074 \cdot x_4$

Результаты исследований для образца № 3.

В результате использования значения весов, нормированных весов целевых критериев (см. табл. 3), парных сравнений критериев по методу Саати получены композиции характеристик, представленные в табл. 12.

Таблица 12. Определение показателей по методу Саати с использованием матрицы парных сравнений

Y_1	Y_2	Y_3	Весовая композиция
1,000	0,667	0,600	0,917
0,941	0,667	1,000	0,904
0,882	1,000	1,000	0,910
0,824	0,778	0,800	0,815
0,824	0,667	0,400	0,766
0,941	0,778	0,600	0,889
0,588	0,778	0,800	0,634
0,059	0,889	0,600	0,229
0,882	0,556	1,000	0,842
0,941	0,556	0,600	0,855
0,882	0,889	0,600	0,861
0,059	1,000	0,400	0,230
0,588	0,778	0,400	0,602
0,353	0,444	0,200	0,355
0,000	0,111	0,000	0,017
0,000	0,000	0,200	0,016

Модели зависимости адгезии от факторов представлены в табл. 13.

Таблица 13. Модели зависимости адгезии от факторов (образец № 3) по методам Лассо и гребневой регрессии

Показатель «адгезия»	
Метод Лассо	Метод гребневой регрессии
$y_1 = 0{,}38 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4$	$y_1 = 0{,}301 \cdot x_1 + 0{,}214 \cdot x_2 + 0{,}025 \cdot x_3 + 0{,}004 \cdot x_4$

Из табл. 12 следует, что два метода, Лассо и гребневая регрессия, показали, что наиболее значимым фактором является вязкость. Расход лакокрасочного материала также может влиять, но существенно меньше.

Модели зависимости адгезии от пар коррелированных факторов представлены в табл. 14.

Таблица 14. Модели зависимости адгезии от пар коррелированных факторов для образца № 3 по методу Лассо и алгоритму Бахадура

Метод Лассо	Алгоритм Бахадура
$\alpha_{13} = 0{,}306$ $\alpha_{12} = 0{,}251$ все остальные $\alpha_{ij} = 0$	$\alpha_{13} = 0{,}238$ $\alpha_{12} = 0{,}184$ все остальные $\alpha_{ij} = 0$

Из табл. 14 видно, что предлагаемый метод на основе модификации метода Бахадура показывает для пар факторов аналогичные результаты, как и метод Лассо, но метод Лассо более сильно выделяет пару «вязкость – температура сушки», в то время как метод Бахадура практически уравнивает эту пару с парой «вязкость – расход лакокрасочного материала».

Интересно отметить: обе модели показали, что пара «вязкость – температура сушки» в комбинации дает наиболее сильный эффект, чем по отдельности вязкость и расход лакокрасочного материала, несмотря на то, что последнее в предположении их зависимости более значимо.

Модели зависимости теплостойкости от факторов представлены в табл. 15.

Таблица 15. Модели зависимости теплостойкости от факторов (образец № 3) по методам Лассо и гребневой регрессии

Показатель «теплостойкость»	
Метод Лассо	Метод гребневой регрессии
$y_2 = 0 \cdot x_1 + 0{,}387 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4$	$y_2 = 0{,}045 \cdot x_1 + 0{,}526 \cdot x_2 + 0{,}264 \cdot x_3 + 0 \cdot x_4$

Из табл. 15 следует, что два метода, Лассо и гребневая регрессия, показали, что наиболее значимым фактором является расход лакокрасочного материала. Температура сушки также может влиять, но существенно меньше.

Модели зависимости теплостойкости от пар коррелированных факторов представлены в табл. 16.

Таблица 16. Модели зависимости теплостойкости от пар коррелированных факторов для образца № 3 по методу Лассо и алгоритму Бахадура

Метод Лассо	Алгоритм Бахадура
$\alpha_{23} = 0{,}389$ все остальные $\alpha_{ij} = 0$	$\alpha_{23} = 0{,}358$ все остальные $\alpha_{ij} = 0$

Из табл. 16 видно, что для показателя «теплостойкость» результаты анализа пар факторов полностью совпадают с анализом отдельных факторов.

Интересно отметить, что обычная регрессия для такой выборки дает неправильные результаты, показывая, что температура сушки и вязкость влияют больше, чем расход лакокрасочного материала.

Модели зависимости водостойкости от факторов представлены в табл. 17.

Таблица17. Модели зависимости водостойкости от факторов (образец № 3) по методам Лассо и гребневой регрессии

Показатель «водостойкость»	
Метод Лассо	Метод гребневой регрессии
$y_3 = 0 \cdot x_1 + 0{,}398 \cdot x_2 + 0{,}212 \cdot x_3 + 0 \cdot x_4$	$y_3 = 0{,}254 \cdot x_1 + 0{,}586 \cdot x_2 + 0{,}465 \cdot x_3 + 0{,}398 \cdot x_4$

Из табл. 17 следует, что два метода Лассо и гребневая регрессия показали, что наиболее значимым фактором является расход лакокрасочного материала. Температура сушки также может влиять, но существенно меньше.

Модели зависимости водостойкости от пар коррелированных факторов представлены в табл. 18.

Таблица 18. Модели зависимости водостойкости от пар коррелированных факторов для образца № 3 по методу Лассо и алгоритму Бахадура

Метод Лассо	Алгоритм Бахадура
$\alpha_{23} = 0{,}542$ $\alpha_{24} = 0{,}201$ все остальные $\alpha_{ij} = 0$	$\alpha_{23} = 0{,}268$ $\alpha_{24} = 0{,}198$ $\alpha_{12} = 0{,}085$ все остальные $\alpha_{ij} = 0$

Из табл. 18 видно, что исследование пар факторов показало аналогичные результаты, но модель Бахадура выявила пару «вязкость – расход лакокрасочного материала» также влияющий на водостойкость, но значительно меньше, чем «расход – температура» и «расход – скорость подачи». Также можно отметить влияние пары «расход лакокрасочного материала – скорость подачи».

Интересно отметить, что, в отличие от теплостойкости, где расход оказывал более сильное влияние, здесь наблюдается примерно одинаковое влияние. Кроме того, гребневая регрессия показала, что и скорость подачи тоже стала одним из значимых факторов, несмотря на то, что его влияние незначительно меньше, чем факторов «расход лакокрасочного материала» и «температура сушки».

Строим модель обобщенной характеристики свойств лакокрасочных покрытий с учетом предпочтений производственного процесса. Модели зависимости обобщенной характеристики от факторов представлены в табл. 19.

Таблица 19. Модели зависимости обобщенной характеристики от факторов для образца № 3

Метод Лассо	Метод гребневой регрессии
$y = 0{,}312 \cdot x_1 + 0{,}201 \cdot x_2 + + 0 \cdot x_3 + 0 \cdot x_4$	$y = 0{,}254 \cdot x_1 + 0{,}231 \cdot x_2 + + 0{,}112 \cdot x_3 + 0{,}032 \cdot x_4$

В результате исследований образцов № 1 и 3 были получены следующие заключения: два метода, Лассо и гребневой регрессии, показали, что наиболее значимыми факторами являются вязкость и расход лакокрасочного материала. Комбинированная оценка определяет обобщенный режим технологии отделки древесины с учетом всех целевых показателей.

Выводы. Предлагаемая методика оптимизации использует линейные модели, зависимости между показателями лакокрасочных покрытий и факторами, которые влияют на эти показатели. Это связано с тем, что основной задачей работы является не определение точной зависимости, а оптимальный выбор наиболее значимых факторов, а это может быть сделано только с использованием линейных моделей. Только в линейных моделях коэффициенты при факторах определяют их значимость. Значения коэффициентов по модулю определяют их значимость. Предлагаемые математические модели не являются в полной степени линейными. Например, и метод Лассо, и метод гребневой регрессии используют так называемую регуляризацию для сглаживания или ограничения множества решений, что делает модель полулинейной. Кроме того, представление Бахадура использует пары факторов, что существенно расширяет класс используемых функций, но, тем не менее, в основе всех этих моделей все равно лежат линейные функции.

Отметим основные преимущества предложенного метода. Прежде всего, он не требует разработки специальных алгоритмов для решения оптимизационных задач при вычислении вектора регрессионных коэффициентов. Полученная задача решается с использованием стандартного метода Лассо после простых преобразований. Во-вторых, метод является достаточно общим, так как можно изменять веса в соответствии с другими целями. В-третьих, мы рассмотрели не только корреляции между значимыми факторами, но и совместные вероятности, учитывающие эти корреляции. Совместные вероятности являются более информативными по сравнению с коэффициентами корреляции. В-четвертых, упрощенная процедура вычисления совместных вероятностей существенно сокращает время вычислений.

Литература

1. Altidor W., Khoshgoftaar T.M., Van Hulse J., Napolitano A. Ensemble feature ranking methods for data intensive computing applications. In B. Furht and A. Escalante, editors, Handbook of Data Intensive Computing. Springer, New York, 2011. P. 349–376.
2. Buhlmann P., S. van de Geer. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics. Springer, Berlin Heidelberg, 2011.
3. Zeny Z. Feng, Xiaojian Yang, Sanjeena Subedi, Paul D. McNicholas. The LASSO and sparse least squares regression methods for SNP selection in predicting quantitative traits. IEEE/ACM Trans. Comput. Biol. Bioinformatics, 9(2), 2012. P. 629–636.
4. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, New York, 2009.
5. Anhui Huang, Shizhong Xu, Xiaodong Cai. Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC Genetics, 14 (5), 2013. P. 1–14.
6. Kohavi R., John G.H. Wrappers for feature subset selection. Artificial Intelligence, 97 (1–2), 1997. P. 273–324.
7. Lindsay B.G., Yi G.Y., Sun J. Issues and strategies in the selection of composite likelihoods. Statistica Sinica, 21 (1), 2011. P. 71–105.
8. Ogutu J.O., Schulz-Streeck T., Piepho H.-P. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC Proceedings, 6(2), 2012. P. 1–10.
9. Subedi S., Feng Z., Deardon R., Schenkel F.S. SNP selection for predicting a quantitative trait. Journal of Applied Statistics. 2013. № 40 (3). P. 600–613.
10. Tibshirani R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996. № 58(1). P. 267–288.
11. Tipping M. Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research. 2001. № 1. P. 211–244.
12. Tutz G., Ulbricht J. Penalized regression with correlation-based penalty. Statistics and Computing, 19(3), 2009. P. 239–253.
13. Usai M.G., Carta A., Casu S. Alternative strategies for selecting subsets of predicting SNPs by LASSO-LARS procedure. BMC Proceedings, 6 (2), 2012. P. 1–9.

14. Zou H. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association. 2006. № 101 (476). P. 1418–1429.
15. Zou H., Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005. № 67 (2). P. 301–320.
16. Соколова В.А. Модели отбора наиболее значимых факторов при исследовании свойств лакокрасочных покрытий // Науч.-технической конф. ин-та технологических машин и транспорта леса по итогам науч.-исследовательских работ 2019 года: сб. ст. (27 янв. – 6 фев. 2020 г.). СПб.: СПбГЛТУ, 2020. С. 452–458.

References

1. W. Altidor, T.M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. Ensemble feature ranking methods for data intensive computing applications. In B. Furht and A. Escalante, editors, Handbook of Data Intensive Computing, pages 349–376. Springer, New York, 2011.
2. P. Buhlmann and S. van de Geer. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics. Springer, Berlin Heidelberg, 2011.
3. Zeny Z. Feng, Xiaojian Yang, Sanjeena Subedi, and Paul D. McNicholas. The LASSO and sparse least squares regression methods for SNP selection in predicting quantitative traits. IEEE/ACM Trans. Comput. Biol. Bioinformatics, 9 (2): 629–636, 2012.
4. T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer, New York, 2009.
5. Anhui Huang, Shizhong Xu, and Xiaodong Cai. Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC Genetics, 14 (5): 1–14, 2013.
6. R. Kohavi and G.H. John. Wrappers for feature subset selection. Artificial Intelligence, 97 (1-2): 273–324, 1997.
7. B.G. Lindsay, G.Y. Yi, and J. Sun. Issues and strategies in the selection of composite likelihoods. Statistica Sinica, 21 (1): 71–105, 2011.
8. J.O Ogutu, T. Schulz-Streeck, and H.-P. Piepho. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC Proceedings, 6 (2): 1–10, 2012.
9. S. Subedi, Z. Feng, R. Deardon, and F.S. Schenkel. SNP selection for predicting a quantitative trait. Journal of Applied Statistics, 40 (3): 600–613, 2013.
10. R. Tibshirani. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58 (1): 267–288, 1996.
11. M. Tipping. Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research, 1: 211–244, 2001.
12. G. Tutz and J. Ulbricht. Penalized regression with correlation-based penalty. Statistics and Computing, 19 (3): 239–253, 2009.
13. M.G Usai, A. Carta, and S. Casu. Alternative strategies for selecting subsets of predicting SNPs by LASSO-LARS procedure. BMC Proceedings, 6 (2): 1–9, 2012.
14. H. Zou. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association, 101 (476): 1418–1429, 2006.
15. H. Zou and T. Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (2): 301–320, 2005.
16. Sokolova V. A. models of selection of the most significant factors in the study of properties of paint coatings. Collection of articles based on the materials of the scientific and technical conference of the Institute of technological machines and forest transport based on the results of research work 2019 [Electronic resource] / ed. by V. A. Sokolov. SPb: Spbgltu, 2020. P. 452–458.