

ФИЗИКО-ХИМИЧЕСКИЕ ПРОБЛЕМЫ ЗАЩИТЫ МАТЕРИАЛОВ

УДК 620.193.21

ПРИМЕНЕНИЕ АЛГОРИТМА “СЛУЧАЙНЫЙ ЛЕС” ДЛЯ ПРЕДСКАЗАНИЙ КОРРОЗИОННЫХ ПОТЕРЬ АЛЮМИНИЯ ЗА ПЕРВЫЙ ГОД ЭКСПОЗИЦИИ В РАЗЛИЧНЫХ РЕГИОНАХ МИРА

© 2023 г. М. А. Гаврюшина¹, *, А. И. Маршаков¹, Ю. М. Панченко¹

¹Институт физической химии и электрохимии имени А.Н. Фрумкина Российской академии наук,
Москва, 119071 Россия

*e-mail: maleeva.marina@gmail.com

Поступила в редакцию 17.11.2022 г.

После доработки 22.11.2022 г.

Принята к публикации 09.12.2022 г.

С помощью алгоритма “случайный лес” (RF) получены две модели для предсказаний годовых коррозионных потерь (K_1) алюминия в открытой атмосфере в различных регионах мира. Модель RF1 получена с использованием объединенных баз данных международных программ ISO CORRAG и MICAT и испытаний на территории России и предназначена для оценки K_1 в различных типах атмосферы в различных регионах мира. Модель позволяет предсказать K_1 только в континентальных районах мира. Для всех типов атмосфер проведено сравнение точности прогноза K_1 по модели RF1 и функции “доза–ответ” (ФДО), представленной в стандарте ISO 9223. Для континентальных мест сравнение достоверности прогноза дано по модели RF2 и функциям “доза–ответ”, представленной в стандарте ISO 9223 и новой ФДО. Показано, что достоверность предсказаний по обеим моделям RF существенно лучше, чем с использованием функций “доза–ответ”.

Ключевые слова: атмосферная коррозия, машинное обучение, случайный лес, алюминий

DOI: 10.31857/S0044185623700249, **EDN:** SZJHKA

1. ВВЕДЕНИЕ

Коррозионные потери металлов в атмосфере могут варьироваться в больших интервалах в зависимости от агрессивности окружающей среды. По этой причине оправдан интерес к аналитическим и численным моделям, которые позволяют предсказывать массопотери металлов в различных климатических регионах мира и типах атмосферы. Наличие в атмосфере значительного числа агрессивных агентов, многостадийность, нелинейность и взаимное влияние физико-химических процессов, протекающих в тонком слое электролита на поверхности металла, делают задачу создания предиктивных моделей атмосферной коррозии очень трудной. Вместе с тем, для решения инженерных задач, таких как предсказание коррозионной стойкости материала конструкций, срока их службы, выбора средств антикоррозионной защиты, требуется разработка моделей, которые использовали бы минимальный набор параметров атмосферы. В идеале, для предсказания коррозионных потерь должны использоваться параметры, которые определяются на метеорологических станциях и на станциях, следящих за загрязнением атмосферы, на всей территории Земного шара. В настоящее время этому требова-

нию отвечают функции доза–ответ (ФДО), которые позволяют предсказать массопотери металлов за первый год экспозиции (K_1) в зависимости от ограниченного числа климатических и аэрохимических параметров атмосферы. Величины K_1 необходимы для определения коррозионной агрессивности атмосферы [1] и для предсказаний долговременных коррозионных потерь в различных регионах мира без проведения натурных испытаний образцов металлов [2–4].

Модели для предсказания величин K_1 типовых металлов, а именно, низкоуглеродистой стали, цинка, меди и алюминия, в различных регионах мира представлены в международном стандарте (ФДО^С) [1]. Новые ФДО (ФДО^Н) для континентальных районов мира даны в [5]. ФДО^С были получены регрессионным анализом баз данных, которые включали экспериментальные коррозионные годовые потери типовых металлов ($K_1^{\text{эк}}$), метеорологические и аэрохимические параметры мест испытаний по программе ISO CORRAG [6], проекту MICAT [7] и результатам российских исследований, проведенных в приморских и континентальных районах Дальнего Востока. При разработке ФДО^Н для стали и цинка использовались

данные программ ЕСЕ ООН [8] и РФ [9], для меди — данные проекта МІСАТ и программ ЕСЕ ООН и РФ, для алюминия — данные проекта МІСАТ и программы РФ.

Предсказание коррозионных потерь алюминия с помощью ФДО является более сложной задачей по сравнению с предсказанием K_1 других типовых металлов. Атмосферная коррозия алюминия имеет ярко выраженный локальный характер с относительно небольшими потерями массы, а образование питтингов, как стохастический процесс, прогнозируется с меньшей точностью в любых коррозионных средах. Поэтому предсказания K_1 алюминия по стандарту [1] имеют больший допустимый интервал неопределенности, то есть, относительную ошибку в пределах от -50% до $+100\%$, по сравнению с другими типовыми металлами, для которых допустимая ошибка предсказаний K_1 находится в пределах от -30% до $+50\%$ [1]. Однако, сопоставление рассчитанных по ФДО^С значений K_1 алюминия с величинами $K_1^{\text{экc}}$, полученными в континентальных местах испытаний проекта МІСАТ и программы РФ, показало, что в значительном числе мест ошибка предсказаний K_1 больше допустимой ошибки [5]. Предсказания K_1 алюминия по ФДО^Н в континентальных районах являются, как правило, более достоверными [5], однако надо учесть, что ФДО^Н была разработана на основе базы данных вышеуказанных испытательных программ. В приморской атмосфере предсказания K_1 алюминия с использованием ФДО^С также имеют ошибку, значительно превышающую допустимый интервал неопределенности согласно стандарта [1], при этом не удалось разработать новую ФДО, применимую в приморских местах всего мира [10].

В связи с этим, представляется необходимым дальнейший поиск моделей для прогноза атмосферной коррозии алюминия. Для этого можно использовать алгоритм случайного леса (RF) — один из популярных методов машинного обучения [11].

Случайный лес состоит из отдельных деревьев — моделей. Дерево решений — это метод представления решающих правил в иерархической структуре, состоящей из элементов двух типов — узлов и листьев. В узлах находятся решающие правила и производится проверка соответствия примеров этому правилу по какому-либо атрибуту обучающего множества. Число деревьев является гиперпараметром модели.

Алгоритм RF использовался для построения предиктивных моделей атмосферной коррозии малолегированных сталей [12, 13]. Скорости коррозии сталей, предсказанные RF моделью, искусственной нейронной сетью, методами регрессии опорных векторов и логистической регрессии,

были сопоставлены с экспериментальными значениями, полученными в 6 местах экспозиции на территории Китая [12]. Оценка достоверности предсказаний скорости коррозии по таким статистическим показателям, как коэффициент детерминации (R^2), средняя абсолютная процентная ошибка (MAPE) и корень из среднеквадратичной ошибки (RMSE), показала преимущество RF модели [13]. RF модель, построенная на основе базы данных, полученной при экспозиции тринадцати видов низколегированной стали в трех местах Японии в открытой атмосфере и под навесом, также показала более точные предсказания скорости коррозии сталей по сравнению с другими методами машинного обучения [13]. В этом случае достоверность моделей оценивалась по величинам R^2 и средней абсолютной ошибке (MAE). Надо отметить, что RF модель, обученная по данным двух мест экспозиции, показала существенно большую ошибку предсказаний в третьем месте экспозиции, данные которого не использовались для обучения этой модели [13].

Алгоритм RF позволяет определить наиболее значимые параметры атмосферы, влияющие на коррозию металлов [12–14]. Это позволяет уменьшить число параметров во входных наборах, которые используются другими методами машинного обучения. Модель, в которой были объединены RF и алгоритм машинного обучения с учителем, была использована для предсказаний скорости коррозии углеродистой стали в 10 местах на территории Китая и показала высокую точность предсказаний [14]. Вместе с тем, достоверность RF моделей [12–14] не была проверена в различных регионах мира, то есть, в местах испытаний, результаты которых не были использованы при разработке этих моделей. Необходимо отметить, что применение в вышеуказанных работах параметра R^2 может приводить к неправильной оценке достоверности предсказанных величин K_1 ($K_1^{\text{пр}}$) в сравнении с $K_1^{\text{экc}}$ [15].

Целью настоящей работы является разработка RF модели на основании результатов годовых коррозионных испытаний алюминия по программам [6, 7, 9], проведенных в различных регионах мира в местах с различным типом атмосферы, а также сопоставление величин K_1 алюминия, предсказанных по RF модели и функциям доза—ответ [1, 5], с экспериментальными данными.

2. МЕТОДИКА РАБОТЫ

2.1. Базы данных натурных коррозионных испытаний алюминия

Для разработки RF моделей использованы базы данных одногодных экспозиций в каждом месте испытаний по программе ISO CORRAG [6]

Таблица 1. Коды мест испытаний программ, данные которых использованы при разработке RF моделей. * отмечены континентальные места испытаний

Программа	Код мест испытаний
ISO CORRAG [5]	ARG1, ARG2, ARG3, ARG4, AGR5, CND1, CS1, CS2, CS3, D1, E1, E2, E3, E4, F1, F2, F3, F4, F5, F6, F7, F8, JAP1, JAP2, JAP3, N1, N2, N3, N4, N5, N6, S1, S2, S3, SF1, SF2, SF3, UK1, UK4, US1, US3, US6, SU1, SU2, SU3, SU4
MICAT [6]	A1, A2*, A4*, A5, A6*, B2, B3, B4, B5, B6*, B7, B9, CH1, CH2, CH3, CH4, CO1, CO2*, CR1, CR2, CR3, CR4, CU1, CU2, CU3, E1*, E4*, E7, E8*, EC1*, EC3, EC5, M1*, M2*, M4, PE2, PE3, PE4*, PE5*, PE6*, PO1, PO2, PO3, U1*, U2, U3*, U4, U5, V3, V4, V5
RUS [4, 8]	Армань, Апапельхино, Аян, Чумикан, о. Айон, Оха, Охотск, У-Хайрюзово, П-Камч, о. Байдуков, м. Шмидта, Невельск, м. Чаплина, м. Гамов, Владивосток, ДВКС1, ДВКС2, ДВКС3, СКС1, СКС2, СКС3, Никольское, м. Лопатка, ГЦКИ1, ГЦКИ2, ГЦКИ3, Билибино*, Оймякон*, Усть-Омчуг*, Атка*, Сусуман*, Тында*, Ключи*, Алдан*, Победино*, Яковлевка*, Пограничный*, Комсомольск-на-Амуре*

(далее БД ISO), проекту MICAT [7] (далее БД MICAT) и по российским программам [9] (далее БД RUS).

Из БД ISO использованы 258 наборов данных, полученных в 46 местах за разные одногодные испытания, включающих коррозионные потери алюминия, $K_1^{\text{экс}}$ (мкм) и соответствующие этому году среднегодовые значения параметров агрессивности атмосферы: температуры (T , °C) и относительной влажности воздуха (RH , %), концентрации SO_2 в воздухе ($[\text{SO}_2]$, мкг/м³) и скорости осаждения хлоридов Cl^- ($[\text{Cl}^-]$, мг/(м² сут)). Значения RH в отдельных местах приведены в соответствии с [16, 17]. В случае, если в местах экспозиции отсутствовали данные о концентрации SO_2 и скорости осаждения хлоридов, а $K_1^{\text{экс}}$ были небольшие, то были приняты фоновые значения: $[\text{SO}_2] = 1$ мкг/м³ и $[\text{Cl}^-] = 1$ мг/(м² сут).

Из БД MICAT использовано 129 наборов данных, полученных в 52 местах испытаний. В набор данных входят величины $K_1^{\text{экс}}$, T , RH , $[\text{SO}_2]$, $[\text{Cl}^-]$ и среднегодовое количество атмосферных осадков (P_{prec} , мм). В базу данных не включены результаты, полученные в местах испытаний с кодами A3, B1, B10, B11, B12, CO3, E5, M3. Обоснование выбраковки мест испытаний дано в [18].

БД RUS состоит из 38 наборов данных, полученных в 38 местах испытаний. В набор данных входят величины $K_1^{\text{экс}}$, T , RH , $[\text{SO}_2]$, $[\text{Cl}^-]$ и P_{prec} .

Для разработки модели RF1 для всех типов атмосфер и сопоставления ее результатов с прогнозом K_1 , сделанным по модели ФДО^С, объединены БД ISO, БД MICAT и БД RUS в общую базу данных (БД INT), которая включает 425 наборов данных. Коды мест испытаний в соответствии с программами представлены в табл. 1.

Для разработки модели RF2 для континентальных мест и сопоставления ее результатов с предсказаниями K_1 , сделанными по модели ФДО^С и ФДО^Н, объединены данные, полученные в континентальных местах испытаний по проекту MICAT и РФ, и сформирована база данных БД CON, которая включает 52 набора данных. Коды мест испытаний, которые попали в БД CON, отмечены в табл. 1 звездочкой.

В табл. 2 приведены интервалы среднегодовых параметров атмосферы и первогодных коррозионных поражений алюминия, для мест испытаний, включенных в БД INT и БД CON.

2.2. Функции доза–ответ

Для прогнозирования коррозионный потерь алюминия за первый год использованы функции доза-ответ, разработанные для двух температурных интервалов.

ФДО^С для атмосфер, содержащих SO_2 и Cl^- , представлены в виде уравнения (1) [1]:

$$\begin{aligned}
 r_{\text{corr}} = & 0.0042P_d^{0.73} \times \\
 & \times \exp[0.025RH + 0.009(T - 10)] + \\
 & + 0.0018S_d^{0.60} \exp(0.02RH + 0.094T), \\
 & \text{при } T \leq 10^\circ\text{C}, \\
 r_{\text{corr}} = & 0.0042P_d^{0.73} \times \\
 & \times \exp[0.025RH - 0.043(T - 10)] + \\
 & + 0.0018S_d^{0.60} \exp(0.02RH + 0.094T), \\
 & \text{при } T > 10^\circ\text{C},
 \end{aligned} \tag{1}$$

где r_{corr} (мкм/год) – скорость коррозии алюминия за первый год экспозиции; T – среднегодовая температура, °C; RH – среднегодовая относитель-

Таблица 2. Параметры атмосферы и коррозионные потери алюминия, их символы, единицы измерения, интервалы среднегодовых значений для мест испытаний, включенных в БД INT и БД CON

Параметр	Символ	Единицы измерения	Интервал	
			БД INT	БД CON
Температура воздуха	T	°C	от –17.1 до +28.2	от –16.6 до +26.9
Относительная влажность воздуха	RH	%	от 33 до 98	от 33 до 98
Количество атмосферных осадков	$Prec$	мм/год	—	от 17 до 1810
Концентрация диоксида серы	$[SO_2]$	мкг/м ³	от 1 до 214.6	от 1 до 67.2
Скорость выпадения хлоридов	$[Cl^-]$	мг/(м ² сут)	от 1 до 1093	—
Первогодовые коррозионные потери Al	$K_1^{экс}$	мкм	от 0.01 до 5.47	от 0.01 до 0.68

ная влажность воздуха, %; P_d и S_d — среднегодовые выпадения SO_2 и Cl^- соответственно, мг/(м²сут).

ФДО^H для атмосфер, содержащих SO_2 , представлены в виде уравнения (2) [5,18]:

$$\begin{aligned}
 K_1^{np} &= 0.01[SO_2]^{0.67} \times \\
 &\times \exp[0.039RH + 0.032(T - 10) - \\
 &- 0.0001Prec], \quad T \leq 10^\circ C, \\
 K_1^{np} &= 0.01[SO_2]^{0.67} \times \\
 &\times \exp[0.039RH - 0.065(T - 10) - \\
 &- 0.0001Prec], \quad T > 10^\circ C,
 \end{aligned}
 \quad (2)$$

где K_1^{np} (г/м²) — коррозионные массопотери алюминия за первый год экспозиции; $[SO_2]$ — среднегодовая концентрация SO_2 в воздухе, мкг/м³; $Prec$ — среднегодовое количество атмосферных осадков, мм/год.

Для пересчета K_1 , выраженной в г/м² (2), в мкм использована плотность алюминия, равная 2.70 г/см³. Для (1) сделан пересчет скорости осаждения SO_2 (P_d , мг/(м²сут)) в концентрацию SO_2 в воздухе ($[SO_2]$, мкг/м³) по соотношению [1]: $P_d = 0.8[SO_2]$.

2.3. Модель “случайный лес”

В машинном обучении принята следующая терминология: каждый набор данных в БД является объектом, что соответствует местам испытаний. Объект характеризуется признаками (входные данные для модели, то есть параметры атмосферы) и величиной прогноза (выходные данные, то есть, величина K_1).

Реализация алгоритма “случайный лес” проводилась при помощи библиотеки *scikit-learn* [19]. Обучение деревьев проводилось на основе обучающей выборки, которая составляет 70% от всей базы данных. Каждое из деревьев получало на вход свою подвыборку, которая с помощью бутстрапа получалась из исходной обучающей

подвыборки. Бутстрап — один из популярных подходов к построению подвыборок. Он заключается в том, что из обучающей выборки длины L (длина выборки — количество принадлежащих ей объектов) выбирают с возвращением L объектов. При этом новая выборка также будет иметь длину L , но некоторые объекты в ней будут повторяться, а некоторые объекты из исходной выборки в нее не попадут.

Размер выборки был равен размеру обучающей выборки (т.е. часть данных дублировалась). Ветвление производили по случайно выбранным признакам (количество которых является гиперпараметром) до исчерпания данных. Ветвление производилось в согласии с критерием информативности (среднеквадратичной ошибкой) так, чтобы дисперсия значений в листе была минимальной.

В тестовой выборке (30% объектов БД) каждое из деревьев давало величину прогноза на основании признаков каждого объекта из этой выборки. В итоге прогнозом для объекта тестовой выборки становилось среднее значение прогноза по всем деревьям.

Значения глобальных гиперпараметров (число признаков для ветвления и число деревьев в лесу) подбирали с помощью функции *GridSearchCV* [20]: число деревьев в лесу от 50 до 600 с шагом 50, число признаков: от 1 до 5. Лучший набор соответствовал наименьшему значению средней относительной ошибки *MAPE* по пяти тестовым выборкам.

Расчет важности признаков, показывающих влияние каждого из признаков на величину прогноза коррозионных потерь, проводился с помощью библиотеки *scikit-learn* [19].

2.4. Статистические критерии достоверности прогноза

Для оценки достоверности предсказаний моделей использовали следующие статистические критерии

Таблица 3. Гиперпараметры, использованные при построении моделей “случайный лес”

Название модели	База данных	Число объектов		Гиперпараметры	
		обучающая выборка	тестовая выборка	число деревьев	число признаков для ветвления
RF1	БД INT	297	128	50	3
RF2	БД CON	36	16	50	3

(1) Средняя абсолютная процентная ошибка (MAPE):

$$\text{MAPE}(x, y) = \frac{1}{N} \sum_{i=1}^N \frac{|x_i - y_i|}{|x_i|} \times 100, \quad (3)$$

где x_i и y_i – экспериментальное и прогнозное значения K_1 , соответственно, N – количество объектов в БД. Чем меньше MAPE, тем меньше модель ошибается в прогнозе.

(2) Симметричная средняя абсолютная процентная ошибка (SMAPE):

$$\text{SMAPE}(x, y) = \frac{2}{N} \sum_{i=1}^N \frac{|x_i - y_i|}{|x_i| + |y_i|} \times 100. \quad (4)$$

Преимущество SMAPE по сравнению с MAPE в том, что SMAPE учитывает возможную погрешность не только прогноза, но и экспериментального значения.

(3) Обобществленный коэффициент детерминации ($R_{\text{нов}}^2$) [15]

$$R_{\text{нов}}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \frac{(yx)_{\text{cp}}}{(x^2)_{\text{cp}}} x_i)^2}{\sum_{i=1}^n (y_i - x_i)^2}, \quad (5)$$

$$\text{где } (yx)_{\text{cp}} = \frac{1}{n} \sum_{i=1}^n y_i x_i, \quad (x^2)_{\text{cp}} = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Коэффициент $R_{\text{нов}}^2$ показывает, насколько хорошо распределение точек с координатами x_i и y_i описывается функцией $y = x$. Значения $R_{\text{нов}}^2$ изменяются от 0 до 1; при $R_{\text{нов}}^2 = 0$ все точки $(x_i; y_i)$ попадают на биссектрису, то есть, на прямую $y = x$. Увеличение коэффициента $R_{\text{нов}}^2$ показывает, что точки $(x_i; y_i)$ наилучшим образом описываются прямой $y = ax$, где коэффициент $a \neq 1$.

Необходимость использования коэффициента $R_{\text{нов}}^2$ связана с тем, что стандартный коэффициент детерминации (R^2) не подходит для определения достоверности модели путем сравнения прогнозного и истинного значения [15]. Коэффициент R^2 показывает, насколько хорошо линейная модель вида $y = ax + b$ описывает данные в сравнении с

моделью $y = b$, но при $R^2 = 1$ условие $a = 1$ не обязательно выполняется. Так, например, если прогноз будет всегда вдвое больше, чем ожидаемое значение, коэффициент R^2 будет в точности таким же, как и в случае, когда прогноз совпадает с ожидаемым значением.

(4) Процент удовлетворительных значений y (PSV):

$$\text{PSV} = \frac{M}{N} \times 100, \quad (6)$$

где M – число y_i , значения которых находятся в интервале от $0.5x_i$ до $2.0x_i$. На графике с координатами $y = K_1^{\text{пр}}$, $x = K_1^{\text{экс}}$ значения $K_1^{\text{пр}}$ должны находиться между линиями относительных ошибок $K_1^{\text{пр}}$, равных -50% и $+100\%$, соответственно. Этот интервал относительных ошибок прогноза соответствуют интервалу неопределенности расчета годовых коррозионных потерь алюминия по стандарту [1]. Чем ближе PSV к единице, тем большее число $K_1^{\text{пр}}$ лежит между линиями указанных относительных ошибок прогноза, и, следовательно, модель является более достоверной.

3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

3.1. Получение моделей “случайный лес” на объединенных базах данных БД INT и БД CON

Базы данных были случайным образом разделены на обучающую (70% объектов БД) и тестовую (30% объектов БД) выборки данных (табл. 3). Величины подобранных гиперпараметров для БД INT и БД CON представлены в табл. 3. Таким образом, на основе объединенных баз данных получены две модели “случайный лес”: RF1 и RF2. В дальнейшем эти модели будут применяться как к объединенным БД, так и к БД различных программ натурных испытаний.

При обучении моделей были рассчитаны значения важности признаков – параметров атмосферы, которые представлены в табл. 4. Нужно отметить, что этот параметр не характеризует важность признаков для решения задачи, но лишь для настройки конкретной модели: RF1 или RF2. Видно, что в случае RF1 наибольшее влияние на величину коррозионных потерь оказывает относительная влажность воздуха, чуть меньше –

Таблица 4. Важность признаков моделей, полученных на основе алгоритма “случайный лес”

Модель	Значение важности признака				
	[SO ₂]	[Cl ⁻]	<i>T</i>	<i>RH</i>	<i>Prec</i>
RF1	0.240	0.280	0.140	0.340	—
RF2	0.65	—	0.126	0.030	0.194

Таблица 5. Значения статистических критериев достоверности модели RF1 и ФДО^С, полученных на БД INT

Модель	Число объектов	Характер выборки	<i>R</i> _{new} ²	MAPE, %	SMAPE, %	PSV, %
RF1	425	БД INT	0.44	47	32	85
	297	Обучение	0.46	35	23	93
	128	Тест	0.59	76	52	68
ФДО ^С	425	БД INT	0.30	125	70	53
	128	Тест	0.52	101	71	50

скорость осаждения хлоридов и содержание SO₂, и еще меньше — температура воздуха. В случае RF2 величину коррозионных потерь в первую очередь определяет содержание SO₂ в воздухе, в гораздо меньшей степени влияют количество осадков и температура, и еще меньше — относительная влажность воздуха.

3.2. Сравнение достоверности моделей RF1 и ФДО^С для БД INT

Величины *K*₁^{пр} были рассчитаны в соответствии с моделью RF1 и ФДО^С (1), используя БД

INT. Сопоставляя предсказанные величины с экспериментальными значениями *K*₁^{экс} (рис. 1), были рассчитаны показатели достоверности этих моделей (табл. 5). Сплошная линия на рис. 1 отвечает условию *K*₁^{пр} = *K*₁^{экс}.

Модель RF1 была получена на 70% объектов БД INT (обучающая выборка), а ее проверка на оставшихся 30% объектов (тестовая выборка) этой базы данных. ФДО^С была получена, используя наборы данных, большая часть которых входит в БД INT. Поэтому для корректного сравнения достоверности моделей RF1 и ФДО^С, значе-

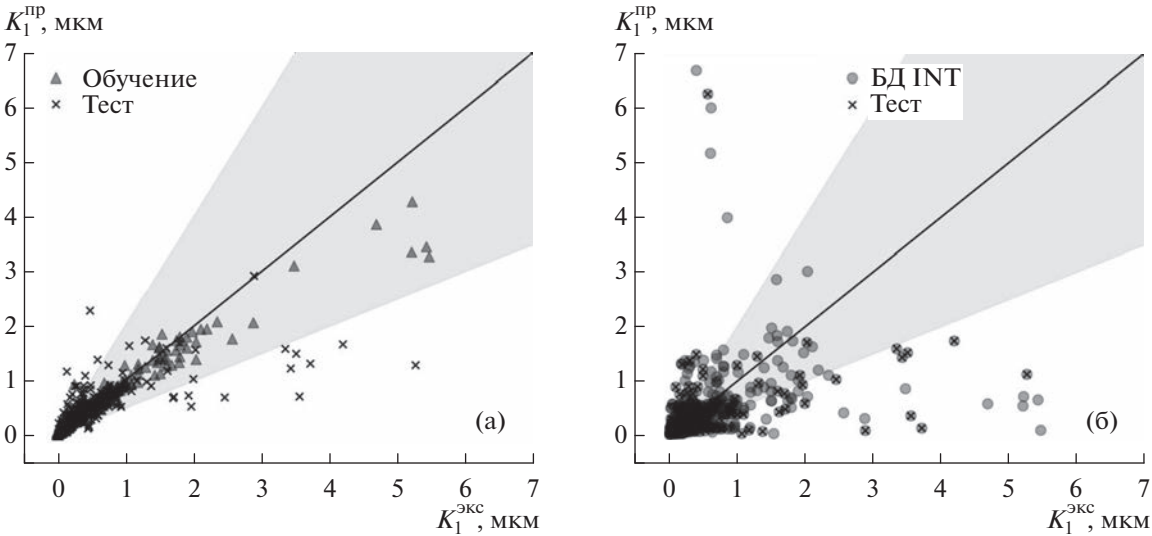


Рис. 1. БД INT. Соответствие между экспериментальными и предсказанными значениями *K*₁: по RF1 (а) и ФДО^С (б). Линия соответствует *K*₁^{пр} = *K*₁^{экс}. Выделенная область показывает относительную ошибку предсказаний в интервале от –50% до + 100%.

ния $K_1^{\text{пр}}$ были рассчитаны по обеим моделям, как для всей базы данных, так и для тестовой выборки. Тестовая выборка, на которой определялись критерии достоверности моделей RF1 и ФДО^С, включала одни и те же объекты.

Полученные результаты (табл. 5) свидетельствуют, что предсказания по модели RF1 для всей БД INT являются более точными, чем для ФДО^С: коэффициенты MAPE и SMAPE имеют меньшие значения, а PSV – большее, величины R_{new}^2 сопоставимы. Применение обеих моделей к тестовым выборкам данных показывает, что RF1 также имеет лучшие значения PSV, MAPE и SMAPE (табл. 5). Величины R_{new}^2 сопоставимы и значительно больше нуля, что можно объяснить несимметричным расположением точек относительно линии $K_1^{\text{экс}} = K_1^{\text{пр}}$ (рис. 1а).

Модель RF1 дает значения PSV = 85 и 68% для всей базы данных и тестовой выборки, соответственно. Это означает, большая часть значений $K_1^{\text{пр}}$ лежит в интервале от $0.5K_1^{\text{экс}}$ до $2.0K_1^{\text{экс}}$. Функция “доза–ответ” может предсказать не более 50–53% значений $K_1^{\text{пр}}$, которые попадают в интервал допустимых ошибок прогноза [1].

3.3. Сравнение достоверности моделей RF2, функций “доза–ответ” ФДО^Н и ФДО^С

Сопоставление рассчитанных по модели RF2 и функциям ФДО^Н и ФДО^С значений годовых коррозионных потерь алюминия в континентальных местах испытаний (БД CON) с соответствующими экспериментальными величинами $K_1^{\text{экс}}$ показано на рис. 2. Значения статистических критериев достоверности всех моделей были рассчитаны как для всей БД CON, так и для 30% тестовой выборки объектов этой базы данных (табл. 6). Как видно, в обоих случаях модель RF2 дает более точный прогноз $K_1^{\text{пр}}$: низкие значения R_{new}^2 , меньшие значения MAPE, SMAPE и большее значение PSV. При использовании RF2 величина R_{new}^2 близка к нулю, то есть, прогноз $K_1^{\text{пр}}$ дает наиболее симметричный разброс точек вокруг линии $K_1^{\text{пр}} = K_1^{\text{экс}}$ как при малых, так и при больших величинах $K_1^{\text{пр}}$. Это означает, что RF2 точнее предсказывает коррозионные потери алюминия, если рассматривать весь диапазон полученных экспериментальных данных (рис. 2а).

Сравнивая достоверность двух функций доза–ответ, можно отметить, что разброс значений $K_1^{\text{пр}}$, рассчитанных по ФДО^Н, более симметричен вокруг линии $K_1^{\text{пр}} = K_1^{\text{экс}}$ по сравнению с разбро-

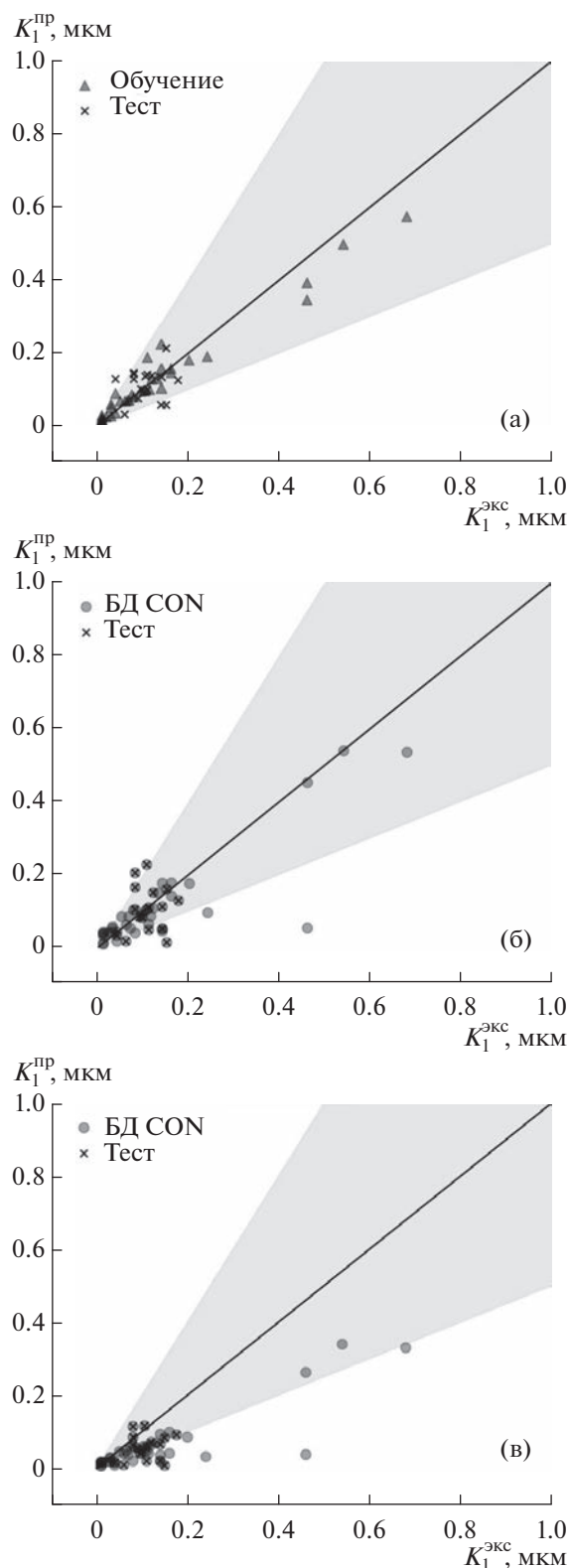


Рис. 2. БД CON. Соответствие между экспериментальными и предсказанными значениями K_1 по RF2 (а), ФДО^Н (б) и ФДО^С (в).

Таблица 6. БД CON. Значения статистических критериев достоверности модели RF2, ФДО^H и ФДО^C

Модель	Число точек	Характер выборки	R_{new}^2	MAPE, %	SMAPE, %	PSV, %
RF2	52	БД CON	0.20	38	30	89
	36	Обучение	0.39	33	25	92
	16	Тест	0.001	49	42	81
ФДО ^H	52	БД CON	0.27	56	48	65
	16	Тест	0.03	51	53	56
ФДО ^C	52	БД CON	0.87	50	69	60
	16	Тест	0.71	51	76	56

Таблица 7. Значения статистических критериев достоверности моделей, примененных к различным базам данных

БД	Модель	Число точек	Признаки	R_{new}^2	MAPE, %	SMAPE, %	PSV, %
ISO	RF1	258	$T, RH, [SO_2], [Cl^-]$	0.0004	44	29	90
	ФДО ^C	258		0.002	156	71	53
MICAT	RF1	129	$T, RH, [SO_2], [Cl^-]$	0.63	57	40	75
	ФДО ^C	129		0.75	85	65	58
	RF2	40	$T, RH, [SO_2], Prec$	0.20	46	36	85
	ФДО ^H	40		0.31	66	56	58
RUS	RF1	38	$T, RH, [SO_2], [Cl^-]$	0.0008	37	25	90
	ФДО ^C	38		0.94	56	83	39
	RF2	12	$T, RH, [SO_2], Prec$	0.15	11	12	100
	ФДО ^H	12		0.0001	24	21	92

сом предсказаний по ФДО^C (рис. 2б и 2в), поэтому R_{new}^2 для тестовой выборки БД CON равен 0.03 и 0.71 для ФДО^H и ФДО^C, соответственно (табл. 2). Величины MAPE и PSV для обоих ФДО примерно равны, а значения SMAPE для ФДО^H меньше как для тестовой выборки, так и для всей БД CON (табл. 2). Следовательно, по совокупности статистических критериев ФДО^H является более достоверной функцией доза—ответ.

3.4. Оценка достоверности моделей “случайный лес” и ФДО на базах данных различных программ натурных испытаний

Программы натурных испытаний типовых металлов ISO CORRAG, MICAT и РФ, результаты которых были объединены в БД INT и БД CON, были проведены в разные годы в различных климатических регионах мира. Естественно, что БД отдельных испытательных программ существенно отличаются, и достоверность предсказаний $K_1^{\text{пр}}$ может быть различна. Необходимо проверить достоверность моделей RF1 и RF2 в случае их применения к БД различных испытательных про-

грамм. Как и в случае объединенных баз данных (п. 3.2 и 3.3), достоверность $K_1^{\text{пр}}$ по моделям RF сравнивали с точностью предсказаний $K_1^{\text{пр}}$, полученных при использовании ФДО^C и ФДО^H. ФДО^C применяли для мест испытаний с любым типом атмосферы, ФДО^H — только для континентальных мест.

БД ISO. В эту БД вошли 258 результатов испытаний с любым типом атмосферы. Для предсказаний K_1 использовали RF1 и ФДО^C. Соответствие фактических и прогнозных значений K_1 представлено на рис. 3. Статистические критерии достоверности моделей даны в табл. 7. Ошибки MAPE и SMAPE для модели RF1 более, чем в 2 раза меньше, чем для ФДО^C. 90% предсказанных моделью RF1 значений $K_1^{\text{пр}}$ укладываются в допустимый интервал ошибок (табл. 7). Величины R_{new}^2 для обеих моделей примерно одинаковы и близки к нулю.

БД MICAT. Статистические критерии были рассчитаны для всей базы данных проекта MICAT (129 объектов) по моделям RF1 и ФДО^C и только для континентальных мест испытаний (40 объек-

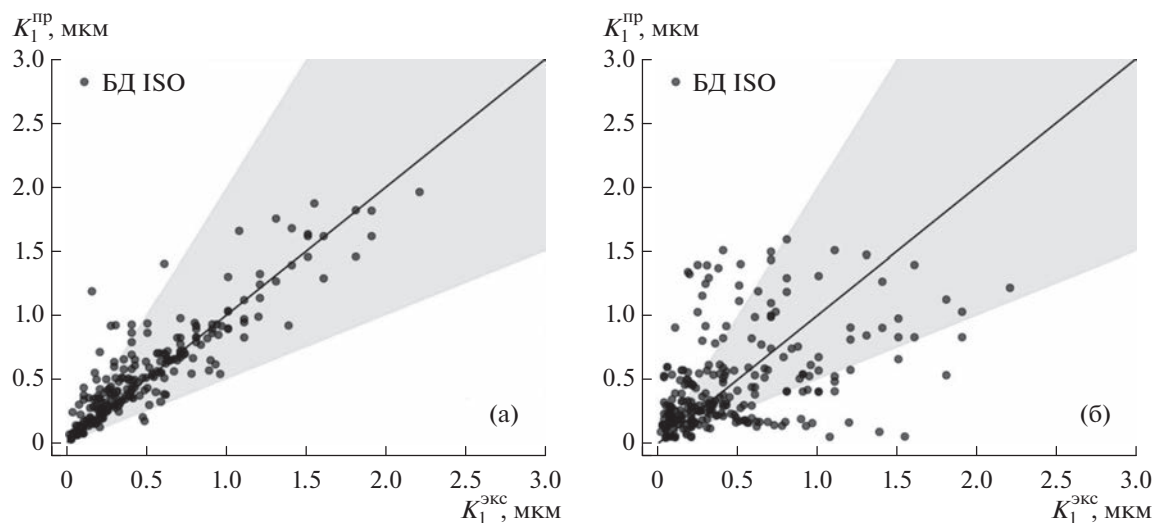


Рис. 3. БД ISO. Соответствие между экспериментальными и предсказанными значениями K_1 по RF1 (а) и ФДО^C (б).

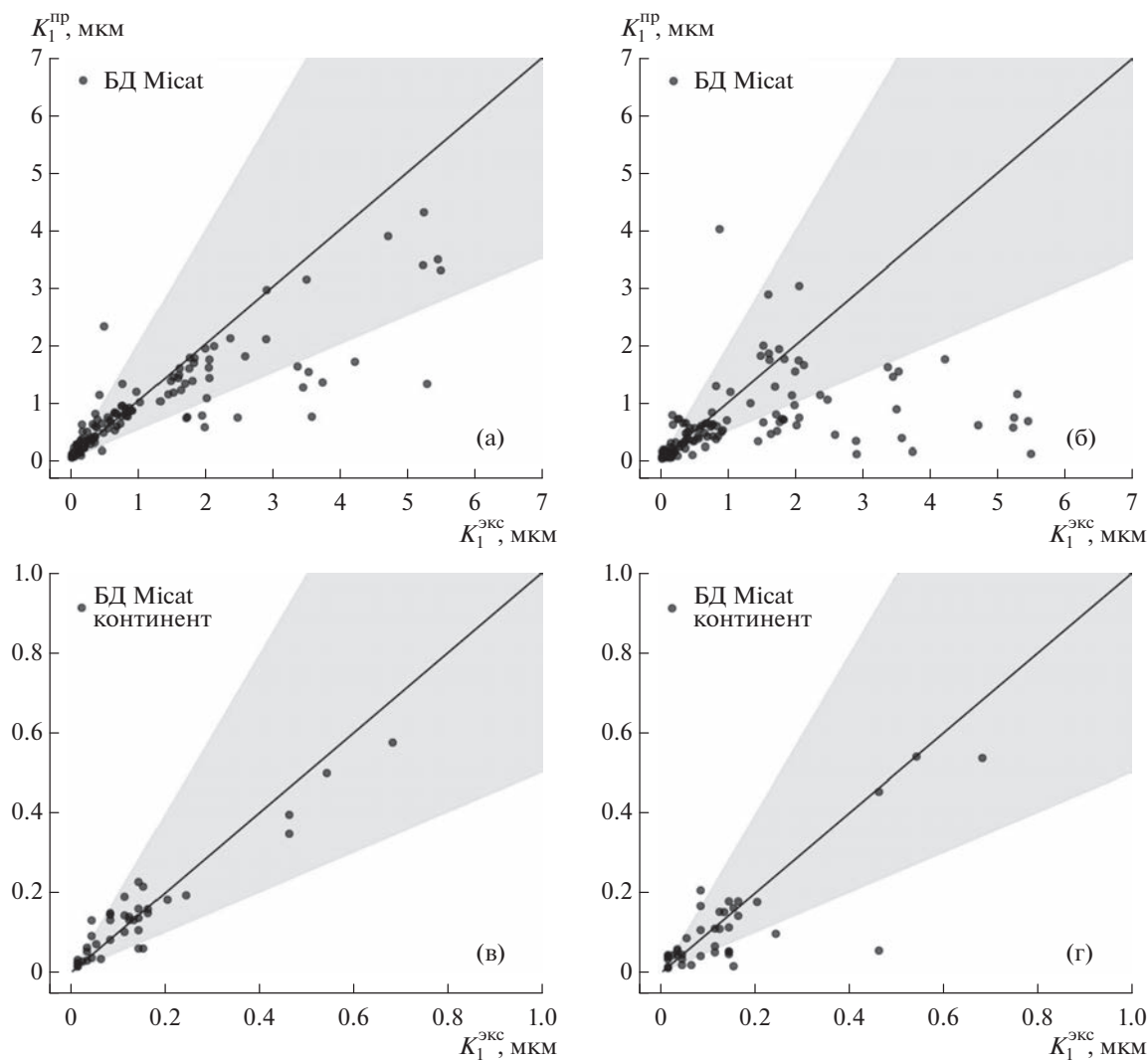


Рис. 4. БД МІСАТ. Соответствие между экспериментальными и предсказанными значениями K_1 по RF1 (а), ФДО^C (б), RF2 (в), ФДО^H (г).

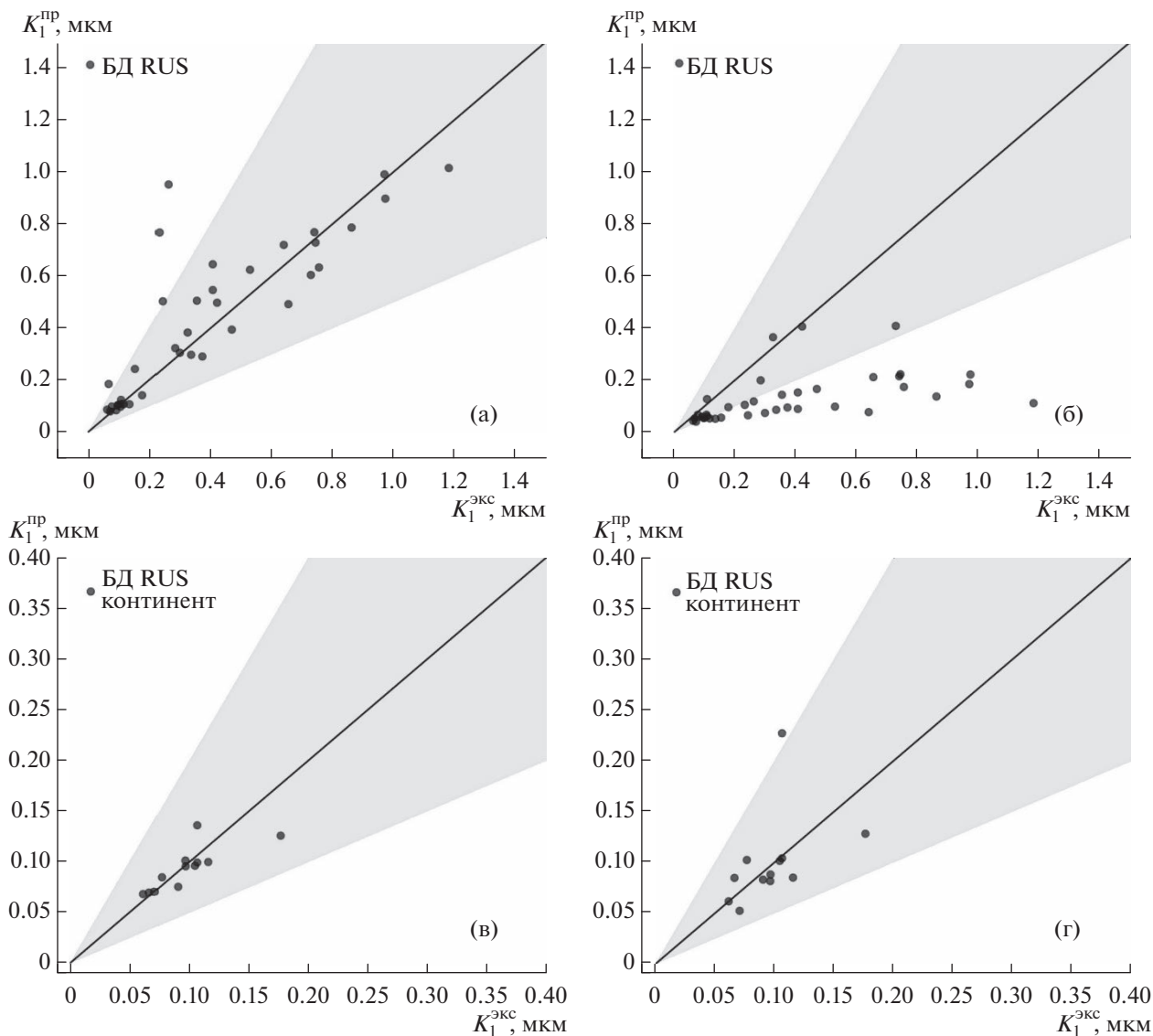


Рис. 5. БД RUS. Соответствие между измеренными и предсказанными значениями по RF1 (а), ФДО^C (б), RF2 (в), ФДО^H (г).

тов) по моделям RF2 и ФДО^H (табл. 7). Соответствие фактических и прогнозных значений K_1 представлено на рис. 4. Все критерии достоверности модели RF1 для БД MICAT несколько хуже, чем для БД ISO (табл. 7). Однако, модель RF1 дает более точные предсказания K_1 в местах испытаний проекта MICAT, чем ФДО^C: ошибка MAPE и SMAPE меньше, а коэффициент PSV — больше. Для континентальной выборки БД MICAT статистические критерии достоверности модели RF2 также лучше, чем для ФДО^H (табл. 7).

БД RUS. Статистические критерии достоверности всех моделей были рассчитаны для полной базы данных (38 объектов) и континентальной выборки (12 объектов). Соответствие фактиче-

ских и прогнозных значений K_1 представлено на рис. 5. Для полной БД RUS точность прогноза K_1 при использовании RF1 значительно выше, чем для ФДО^C (рис. 5а, 5б). Результат прогноза по модели RF1 характеризуют меньшие значения MAPE и SMAPE, чем для ФДО^C (табл. 7). Число предсказанных K_1 , которые имеют допустимую ошибку для обеих моделей отличается значительно: коэффициент PSV = 90 и 39% для RF1 и ФДО^C, соответственно.

Для континентальных объектов БД RUS предсказания K_1 по обеим моделям сопоставимы и достаточно высоки (рис. 5в, 5г), что отражается в показателях достоверности моделей (табл. 7). Так, число предсказанных по ФДО^H значений K_1 ,

которые имеют допустимую ошибку, равно 92%, что является лучшим результатом для функций “доза–ответ” для всех рассмотренных баз данных. Однако, все значения K_1 , рассчитанные по RF2 попадают в интервал допустимых ошибок (PSV = 100%). Величины MAPE и SMAPE при использовании модели RF2 равны 11 и 12%, соответственно, и примерно в 2 раза ниже, чем для модели ФДО^H.

ЗАКЛЮЧЕНИЕ

С помощью алгоритма “случайный лес” получены две модели RF для предсказаний годовых коррозионных потерь (K_1) алюминия в открытой атмосфере в различных регионах мира. Модель RF1 получена на объединенной базе данных, которая включает данные программ ISO CORRAG, MICAT и RUS, и предсказывает величины K_1 по значениям T , RH , $[SO_2]$, $[Cl^-]$. Модель RF2 получена на объединенной базе данных, которая включает данные, полученные в континентальных местах испытаний программ MICAT и RUS, и предсказывает величины K_1 по значениям T , RH , $Prec$, $[SO_2]$.

Достоверность предсказаний моделей RF оценивалась по совокупности статистических критериев: обобщенному коэффициенту детерминации R_{new}^2 , MAPE, симметричной средней относительной ошибке SMAPE и коэффициенту CRV. Коэффициент CRV был предложен в этой работе и показывает долю предсказаний K_1 , относительная ошибка которых не выходит за интервал ошибок, допустимый в соответствии со стандартом [1].

Проведено сравнение точности предсказаний K_1 по моделям RF и двум функциям “доза–ответ” (ФДО): ФДО стандарта [1] для всех типов атмосферы и новой ФДО [5] для не морской атмосферы. Показано, что для объединенных баз данных достоверность обеих моделей RF лучше, чем ФДО.

Сопоставление точности прогнозов коррозионных потерь по моделям RF и ФДО, проведенное на базах данных отдельных испытательных программ, подтвердило, что модели RF дают более точный прогноз.

СПИСОК ЛИТЕРАТУРЫ

1. ISO 9223:2012(E). Corrosion of metals and alloys. Corrosivity of atmospheres. Classification, determina-

tion and estimation, International Standards Organization, Geneva, 2012.

2. ISO 9224:2012(E) Corrosion of metals and alloys. Corrosivity of atmospheres. Guiding values for the corrosivity categories, 2012.
3. Panchenko Yu.M., Marshakov A.I. // Corr. Sci. 2016. V. 109. P. 217.
4. Abramova M.G., Panchenko Y.M., Vetrova E.Y. et al. // Prot. Met. Phys. Chem. Surf. 2021. V. 57. № 7. P. 1272–1282.
5. Panchenko Yu.M., Marshakov A.I., Nikolaeva L.A., Ignonin T.N. // Civil Eng. J. 2020. V. 6. № 8. P. 1503.
6. Knotkova D., Boschek P., Kreislova K. In *Atmospheric Corrosion*, Kirk W.W. and Lawson H.H., Eds., Philadelphia, PA, USA: American Soc. Test. Mater., 1995. P. 38.
7. Morcillo M., In *Atmospheric Corrosion*, Kirk W.W. and Lawson H.H., Eds., Philadelphia, PA, USA: American Soc. Test. Mater. 1995. P. 257.
8. Tidblad J., Kucera V., Mikhailov A.A., Henriksen J., Kreislova K., Yaites T., Stöckle B., Schreiner M. // Water, Air, and Soil Pollution. 2001. V. 130. P. 1457.
9. Панченко Ю.М., Шухаева Л.А., Михайловский Ю.Н. // Защита металлов. 1982. Т. 18. С. 575.
10. Panchenko Yu.M., Marshakov A.I., Nikolaeva L.A., Ignonin T.N. // Corr. Eng. Sci. Tech. 2020. V. 55. № 8. P. 655.
11. Breiman L. // Machine Learning. 2001. V. 45. P. 5.
12. Zhi Y., Fu D., Zhang D., Yang T., Li X. // Metals. 2019. V. 9. № 3. P. 383.
13. Yan L., Diao Y., Gao K. // Materials. 2020. V. 13. № 15. P. 3266.
14. Zhi Y., Jin Z., Lu L., Yang T., Zhou D., Pei Z., Wu D., Fu D., Zhang D., Li X. // Corrosion Science. 2021. V. 178. № 109084.
15. Panchenko Yu.M., Marshakov A.I., Bardin I.V., Shklyayev A.V. // Prot. Metals Phys. Chem. Surf. 2019. V. 55. № 4. P. 753.
16. Mikhailov A.A., Tidblad J., Kucera V. // Prot. Metals. 2004. V. 40. № 6. P. 541.
17. Tidblad J., Kucera V., Mikhailov A.A., Knotkova D. In *Outdoor Atmospheric Corrosion*, Townsend H.E., Eds., West Conshohocken, PA, USA: American Soc. Test. Mater., 2002, p. 73.
18. Panchenko Yu.M., Marshakov A.I., Nikolaeva L.A., Kovtanyuk V.V. // AIMS Materials Sci. 2018. V. 5. № 4. P. 624.
19. Scikit-learn. Machine Learning in Python // <https://scikit-learn.org/stable/index.html>
20. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html