УДК 311.218; 311(084); 625.1

П. Г. Королев
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Анализ развития дефектов железнодорожного полотна

Описан алгоритм, применяемый для обработки и анализа результатов, характеризующих величины дефектов железнодорожного полотна. В алгоритме применены правила формирования выборок, пригодных для применения непараметрического рангового критерия. Алгоритм разработан на основе анализа данных, полученных при проведении физического эксперимента. Для проверки алгоритма использованы тестовые наборы данных с известным законом распределения. Рассмотрены особенности применения непараметрического рангового критерия для статистической обработки малых выборок при наличии в них выбросов. Сформулировано правило исключения аномального отсчета из малой выборки. Представлен автоматный граф, описывающий состояния программы, условия переходов между состояниями и выполняемые действия. Рассмотрены диагностические ситуации для различных соотношений изменения меры среднего и меры разброса выборок результатов измерений, характеризующих величину дефектов. Проведенный анализ позволяет идентифицировать ситуации с изменением свойств объекта исследования.

Железнодорожное полотно, измерительная система, малая выборка, ранговый критерий

Обеспечение безопасности на железнодорожном транспорте является важной социально-экономической задачей. Одним из способов снижения аварийности на железной дороге является мониторинг состояния рельсового пути. Информацию о величине дефектов получают от систем динамического мониторинга рельсового пути.

Микромеханические датчики линейных ускорений расположены вблизи от места контакта колесо-рельс на буксах тележки вагона. На раме тележки монтируются бесплатформенная инерциальная навигационная система (БИНС), профилометр, позволяющие измерять шаблон, уровень и кривизну рельсового пути. Для реализации традиционного двухточечного хордового метода измерения просадок рельсовых нитей на базе тележки устанавливается комплект из четырех датчиков линейных перемещений (ДЛП) «букса–тележка» (датчики Б-Т). При этом также реализуется схема определения отклонения от продольного профиля, перекосов. Дефекты поверхности катания рельсов измеряются с помощью микромеханических модулей (ММ), в состав которых входят измерительные каналы угловой скорости и линейных ускорений. ММ устанавливаются на буксах ходовой тележки вагона в непосредственной близости от зоны контакта колесо-рельс [1], [2].

Результат измерения дефекта (возвышение одной рельсовой нити над другой, перекосы на базе ходовой тележки, короткая просадка, ширина рельсовой колеи) в общем случае является случайной величиной, зависящей от многих факторов – сезонного, скорости движения локомотива или путеизмерительного вагона, соседствующих дефектов и т. п., приводящих к появлению дополнительных погрешностей и выбросам. Следовательно, необходимо делать выводы о динамике развития дефекта и принимать организационно-технические решения, основываясь на результатах обработки статистических наблюдений. С другой стороны, в реальных условиях эксплуатации участка обращения локомотивов сложно набрать репрезентативную выборку.

Возникают следующие задачи:

– сформулировать правила формирования достаточных для применения статистических методов выборок;

– разработать алгоритм исследования значимости изменения амплитуды дефекта;

– сформулировать правила удаления выбросов из малых выборок.

Необходим статистический критерий для сравнения математических ожиданий выборок малого объема при условиях, что нет полной уверенности в нормальности закона распределения случайной составляющей погрешности и выборки могут различаться по объему.

Для проверки гипотезы о равенстве математических ожиданий при распределении, отличном от нормального, и неравных объемах выборок $n_1 \neq n_2$ можно использовать критерий Манна и Уитни [3].

Ранговый критерий Манна и Уитни основан на критерии Уилкоксона для независимых выборок. Он является непараметрическим аналогом *t*-критерия для сравнения двух средних значений непрерывных распределений. Минимальный объем выборки – три отсчета, если объем выборки 5 и более отсчетов, вторая выборка может состоять из двух отсчетов. Максимальный объем выборки – 60 отсчетов. В доступных для анализа результатах измерений от бортовой измерительной системы было по 5 отсчетов. *U*-критерий Уилкоксона, Манна и Уитни проверяет гипотезу о принадлежности двух выборок одной и той же генеральной совокупности H_0: $F_1(x) = F_2(x)$. Эта гипотеза включает также равенство значений медиан $\tilde{\mu}_1 = \tilde{\mu}_2$ и равенство средних значений $\mu_1 = \mu_2$. Для вычисления статистики упорядочивается $n_1 + n_2$ значений объединенной выборки по возрастанию значения и все элементы ранжируются. В случае совпадения нескольких значений ранг такого значения следует считать, как среднее арифметическое тех позиций, которые занимают данные значения при их расположении в порядке возрастания. Далее определяется сумма рангов R_1, соответствующая элементам первой выборки, и сумма рангов второй R_2, определяется бóльшая из ранговых сумм T_x. Вычисляется эмпирическое значение

$$U_{\text{эмп}} = n_1 n_2 + \frac{n_x(n_x+1)}{2} - T_x,$$

где n_1 и n_2 – количество значений в выборках 1 и 2; n_x – количество значений в выборке с большей суммой рангов.

При заданном уровне значимости находится табличное значение $U_{\text{кр}}$ и проверяется неравенство $U_{\text{эмп}} > U_{\text{кр}}$. Если оно истинно, то гипотеза H_0 принимается, в противном случае – отвергается, и различие математических ожиданий определяется как существенное [3].

На рис. 1 представлены варианты распределения отсчетов, полученные от бортовой системы вагона – путеизмерителя.

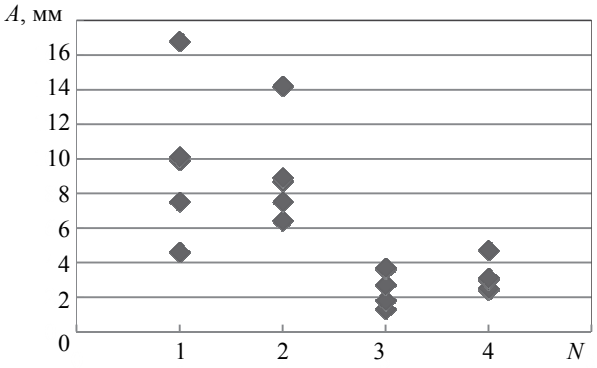

Рис. 1

На рис. 1–3 на горизонтальной оси расположены номера выборок *N*, на вертикальной – величина дефекта *A*. Из рис. 1 видно, что выборки из пяти отсчетов или хорошо сгруппированы, или имеют по одному выбросу. Теоретически, возможно группирование и по два–три отсчета. На рис. 2 и 3 показаны модельные выборки, с помощью которых проводилось исследование критерия Манна и Уитни. Установлено, что критерий устойчив к одиночным и двойным выбросам за пределы области перекрытия отсчетов.

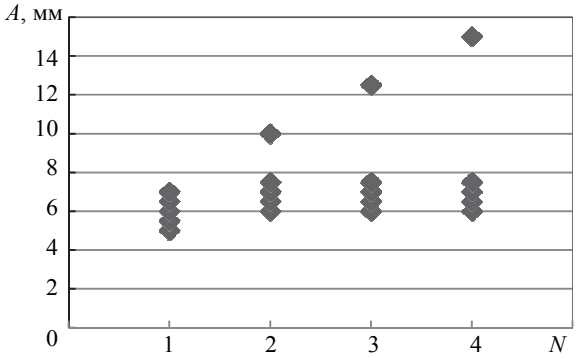

Рис. 2

Тем не менее критерий Манна и Уитни чувствителен к ошибкам второго рода (в частности, при минимальном пересечении выборок).

На рис. 3 представлены выборки из пяти отсчетов. Расхождение выборок № 2 и 1 определено как значимое, выборок № 3 и 1 – зона неопределенности критерия, выборок № 4 и 1 – расхождение определено как незначимое, что с точки зрения автора является ошибкой второго рода, ввиду возможности появления аномальных отсчетов в

системах динамического мониторинга рельсового пути [4], [5].

Критерий Манна и Уитни показывает значимое расхождение выборок при «непересечении» четырех отсчетов из пяти при равноудаленных отсчетах в выборках, значительно отклонившийся отсчет приводит к ошибке. Классическая проверка на выбросы ни при равномерном, ни при нормальном законе распределения не позволяет исключить этот приводящий к ошибке второго рода отсчет.

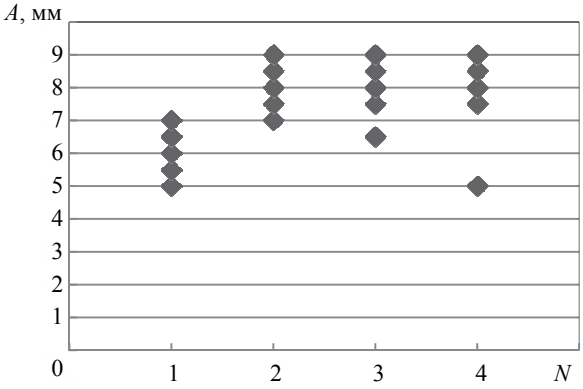

Рис. 3

При проверке отсчета, равного 10, на выброс (рис. 3) получено значение математического ожидания 15.20, среднеквадратического отклонения 3.11, и, исходя из того что результаты были смоделированы и распределены равномерно, этот отсчет выбросом не является. Таким образом, предотвратить ошибку второго рода, классифицировав наиболее удаленный от среднего значения отсчет как выброс, не удалось.

В процессе расчета $U_{эмп}$ множествам значений выборок $X1\{x_1, \ldots, x_{n1}\}$ и $X2\{x_1, \ldots, x_{n2}\}$ ставятся в соответствие множества рангов $R1\{R_1, \ldots, R_{n1}\}$ и $R2\{R_1, \ldots, R_{n2}\}$. Пример приведен в табл. 1. Графическое изображение данной ситуации представлено на рис. 3.

Таблица 1

№	Выборка 1	Ранг 1	Выборка 2	Ранг 2
1	10.0	1.5	10.0	1.5
2	11.0	3.0	15.0	7.0
3	12.0	4.0	16.0	8.0
4	13.0	5.0	17.0	9.0
5	14.0	6.0	18.0	10.0
Сумма	–	19.5	–	35.5

Поскольку даже в малой выборке из-за особенностей объекта исследования возможны аномальные отклонения значений (выбросы), необходимо сформулировать правила удаления выбросов из малой выборки.

Введем следующие термины:

– *базовая выборка* (БВ) – набор отсчетов $n \geq 3$, полученный в начале наблюдений с короткими временными интервалами в предположении, что между проездами не происходило деградации состояния железнодорожного полотна;

– *рабочая выборка* (РВ) – набор отсчетов переменного размера, получаемый в процессе эксплуатации линейного участка пути;

– *опорная рабочая выборка* (ОРВ) – набор отсчетов $n \geq 3$, полученный после формирования базовой выборки в процессе эксплуатации линейного участка пути, расхождение рабочей выборки от опорной рабочей выборки классифицировано как значимое;

– *предшествующая выборка* (ПВ) – при накоплении первой рабочей выборки предшествующей является базовая выборка, а при накоплении последующих выборок – опорная рабочая.

По мере появления групп результатов измерений, которые нельзя объединить с рабочей выборкой, РВ переводится в статус ОРВ и начинается формирование новой рабочей выборки.

Возможны две ситуации, при сравнении рабочей и опорной (или базовой) выборок, когда имеет место выброс:

1. Отсутствие тенденции к нарастанию дефекта (значительное перекрытие выборок, «перемешанные ранги») и выброс в сторону увеличения дефекта. В данном случае критерий Манна и Уитни позволяет принять корректное решение, так как нечувствителен к ошибкам первого рода (рис. 1, выборки 1 и 2).

2. Наличие тенденции к расхождению (отсутствие перекрытия выборок), но один отсчет имеет ранг, пересекающийся с рангами базовой (опорной) выборки. Ситуация представлена на рис. 3, выборки в моменты времени 1, 3, 4. В этом случае выброс оказывает существенное влияние на значение $U_{эмп}$ и возможно принятие неверного решения (ошибка второго рода).

Пример. Имеет место тенденция к развитию дефекта, т. е. математическое ожидание рабочей выборки больше математического ожидания опорной (базовой) выборки $M_{x2} > M_{x1}$, но критерий не показывает значимости расхождений выборок.

Для исключения этого отсчета можно воспользоваться сравнением значений рангов. Упорядочим по неубыванию множества отсчетов и множества рангов и выполним проверку условия:

если

$$\forall\{R_{22}, R_{23}, R_{24}, R_{25}\} > \{R_{11}, R_{12}, R_{13}, R_{14}, R_{15}\} \wedge \\ \wedge \exists R_{21} < \{R_{11}, R_{12}, R_{13}, R_{14}, R_{15}\}, \quad (1)$$

то возникает возможность принятия ошибочного решения, и результат, имеющий ранг R_{21}, следует исключить из выборки.

В общем случае результат, имеющий ранг R_{2i}, следует исключить из выборки, если

$$\forall\{R_{21}, ..., R_{2i}, ..., R_{2N}\} > \{R_{11}, ..., R_{1i}, ..., R_{1M}\} \wedge \\ \wedge \exists R_{2i} < \{R_{11}, ..., R_{1i}, ..., R_{1M}\}. \quad (2)$$

На рис. 4 представлен процесс развития дефекта и накопления информации об этом процессе. На вертикальной оси обозначены амплитуды, характеризующие величину дефекта: A_b – базовое значение; A_1, A_2, A_3 – средние значения величины дефекта; Δt_1, Δt_2, Δt_3 – интервалы времени, на которых накапливались выборки, по которым рассчитаны значения A_i.

Алгоритм. Исходное состояние – нет ни одного результата по i-му дефекту:

1) накопить 3 результата с короткими временными интервалами (желательно несколько проездов подряд);

2) при наличии еще трех результатов применить критерий Манна–Уитни. Если расхождение значимо – зарегистрировать ситуацию как требующую решения, ограничить БВ тремя отсчетами;

3) если расхождение незначимо или находится в зоне неопределенности, проверить отличия рангов – если разошлись не более, чем на один результат, объединить выборки; иначе – новые три отсчета отнести к выборке № 2 (РВ);

4) если в выборке № 1 (БВ) оказалось 6 отсчетов, при наборе выборки № 2 для проверки значимости расхождений достаточно ограничиться двумя отсчетами, при этом критерий классифицирует расхождение как значимое, если выборки полностью не пересекаются;

5) если расхождение незначимо, то накопление РВ продолжается до пяти отсчетов, применяется критерий Манна и Уитни с учетом (1), (2);

6) перевод РВ в состояние ОРВ происходит или при установлении значимости расхождения выборок, или при достижении объема выборки 60 отсчетов, максимального для применяемого критерия.

Представим работу программы обнаружения роста дефекта в виде графа состояний (рис. 5).

Состояния:

1) исходное – накопление БВ;

2) накопление РВ (текущей);

3) проверка значимости расхождения РВ и БВ;

4) проверка значимости расхождения ОРВ и РВ;

5) перевод РВ в состояние ОРВ.

Условия:

У1 – количество отсчетов менее трех;

У2 – расхождение признаков РВ и БВ значимо;

У3 – расхождение признаков ОРВ и РВ значимо;

У4 – количество отсчетов в выборке достигло 60;

У5 – в качестве ОРВ выступает БВ.

Действия:

Д1 – объединение выборок;

Д2 – «закрытие выборки» (отсчеты в нее перестают добавляться);

Д3 – регистрация факта роста дефекта.

Для анализа изменения меры разброса малых выборок применяется критерий Зигеля–Тьюки [6]. В отличие от критерия Манна и Уитни, данный критерий требует нормального распределения выборки.

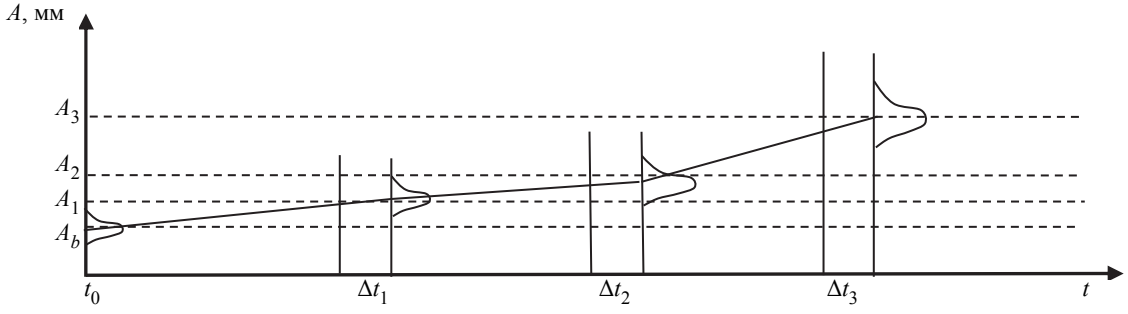

Рис. 4

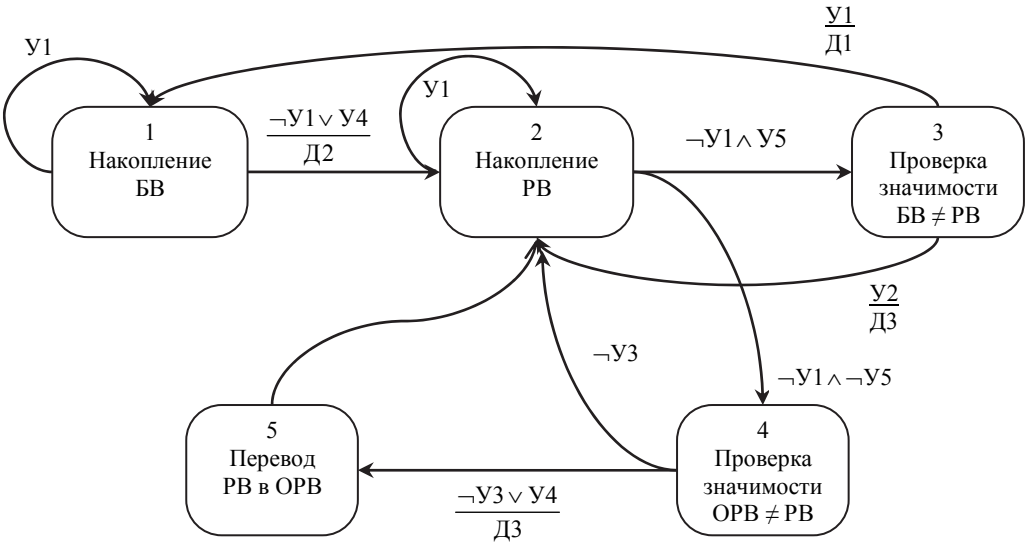

Рис. 5

Таблица 2

Математическое ожидание	Расхождение незначимо (МРН)	Расхождение значимо (рост) (МРЗР)	Расхождение значимо (спад) (МРЗС)
Мера разброса	Расхождение незначимо (РРН)	Расхождение значимо (рост) (РРЗР)	Расхождение значимо (спад) (РРЗС)

В технике часто применяется диагностика состояния по двум параметрам [7], [8]. В табл. 2 применены следующие обозначения: МР – расхождение математического ожидания, РР – расхождение меры разброса, З – значимость, Н – незначимость, Р – рост, С – спад.

При анализе результатов многократных измерений двух выборок возможны ситуации, представленные в табл. 2.

Проведем классификацию ситуаций:

C_1 = МРН ∧ РРН – стабильное состояние;

C_2 = МРН ∧ РРЗР – рост случайной составляющей: возможно влияние сезонного фактора, возможны метрологические отказы;

C_3= МРН ∧ РРЗС – рост случайной составляющей: прекращение действия сезонного фактора;

C_4 = МРЗР ∧ РРН – значимость расхождения математических ожиданий при незначительной случайной составляющей, позволяет уверенно идентифицировать рост дефекта;

C_5= МРЗР ∧ РРЗР – предполагаемый рост дефекта, требуется дополнительная информация;

C_6 = МРЗР ∧ РРЗС – идентификация роста дефекта, снижение случайной составляющей;

C_7 = МРЗС ∧ РРН – признаки устранения дефекта;

C_8 = МРЗС ∧ РРЗР – значительное снижение математического ожидания и рост случайной составляющей позволяет подозревать неисправность измерительного канала и провести техническую диагностику измерительной системы и ее калибровку;

C_9= МРЗС ∧ РРЗСН – признаки устранения дефекта (может использоваться для оценки качества ремонта).

Таким образом, разработан алгоритм, предназначенный для обработки и анализа результатов, характеризующих величины дефектов железнодорожного полотна. Рассмотренные особенности применения непараметрического рангового критерия для обработки малых выборок при наличии в них выбросов, а также диагностические ситуации для различных соотношений изменения меры среднего и меры разброса выборок результатов измерений параметров дефектов, позволят принимать статистически обоснованные решения по состоянию железнодорожного полотна.

СПИСОК ЛИТЕРАТУРЫ

1. Экспериментальные исследования системы динамического мониторинга рельсового пути / П. Г. Королев, Д. Ю. Ларионов, М. Т. Рзиева, М. Н. Шилов // Изв. СПбГЭТУ «ЛЭТИ». 2014. № 10. С. 50–54.

2. MEMS_Based Inertial System for Railway Track Diagnostics / A. M. Boronakhin, L. N. Podgornaya, E. D. Bokhman, N. S. Filipenya, Yu. V. Filatov, R. B. Shalymov, D. Yu. Larionov // Gyroscopy and Navigation. 2011. Vol. 2, № 4. P. 261–268.

3. Лемешко Б. Ю., Помадин С. С. Проверка гипотез о математических ожиданиях и дисперсиях в задачах метрологии и контроля качества при вероятностных законах, отличающихся от нормального // Метрология. 2004. № 3. С. 3–15.

4. Концепция построения системы динамического мониторинга рельсового пути / В. В. Алексеев, П. Г. Королев, Л. Д. Ларионов, М. Н. Шилов // Изв. СПбГЭТУ «ЛЭТИ». 2014. № 10. С. 45–49.

5. Оценка дефектов железнодорожного полотна. Исследование критерия / П. Г. Королев, В. А. Костыря, С. А. Кук, О. А. Микус // XXI Междунар. конф. по мягким вычислениям и измерениям «SCM'2018». СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2018. С. 85–88.

6. Siegel S., Tukey J. W. A nonparametric sum of ranks procedure for relative spread in unpaired samples // J. Amer. Statist. Assoc. 1960. Vol. 55. P. 429–455.

7. Алгоритм идентификации диагностических признаков по параметрам вибрации / В. В. Алексеев, П. Г. Королев, В. С. Коновалова, И. В. Калякин, А. Г. Перкова // XVIII МНК по мягким вычислениям и измерениям: сб. докл. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2015. Т. 2. С. 131–136.

8. Алексеев В. В., Иващенко О. А. Измерение скорости протекания процессов в информационно-измерительных и управляющих системах с целью предупреждения аварийных ситуаций // Приборы. 2013. № 8. С. 32–37.

P. G. Korolev
Saint Petersburg Electrotechnical University «LETI»

RAILWAY DEFECTS EVOLUTION ANALYSIS

An algorithm is described for processing and analyzing the results characterizing the values of defects in a railway track. The algorithm uses rules for the formation of samples suitable for the application of a nonparametric ranking criterion. The algorithm is developed based on the analysis of data obtained during the physical experiment. To test the algorithm, test sets of data with a known distribution law are used. The rule for excluding an anomalous sample from a small sample is formulated. An automatic graph describing the program states, the conditions of state transitions and the actions performed is presented. The peculiarities of using a nonparametric rank criterion for statistical processing of small samples with the presence of emissions are considered. Diagnostic situations for different ratios of the variation of the mean and measure of the scatter of samples of measurement results characterizing the magnitude of defects are considered. The analysis that was carried out makes it possible to identify situations with a change in the properties of the research object.

Railway track; measuring system; small sample; rank test