Modeling Staircase Wear Using Logistic Growth and Clustering Analysis

Yiming Yao[1, *, #], Qihao Tang[2]

[1] Haide College, Ocean University of China, Qingdao, China

[2] Dundee international institute, Central South University, Changsha, China

*Corresponding author: 17363911985@163.com

#These authors contributed equally.

Abstract

This study develops a novel framework for assessing staircase wear in high-traffic environments by integrating logistic growth modeling with clustering-based crowd behavior analysis. Traditional visual inspection methods often fail to provide quantitative wear assessments, necessitating data-driven approaches to evaluate deterioration patterns and predict service life. The proposed two-phase logistic model distinguishes between natural degradation (unrepaired state) and post-repair recovery, incorporating adjusted growth rates and maximum wear capacities to reflect maintenance impacts. Clustering techniques analyze pedestrian flow patterns, revealing an annual usage frequency of 75.002 million people and identifying predominant upward movement tendencies (82.4% side-by-side walking). Spatial wear analysis demonstrates concentrated degradation on step outer edges and right sides, correlating with biomechanical movement patterns. Results demonstrate the model's effectiveness in quantifying wear progression, offering actionable insights for maintenance planning and stair design optimization. The hybrid approach bridges theoretical wear mechanics with real-world usage dynamics, providing a scalable solution for infrastructure management that could be extended to other high-traffic structures.

Keywords

Staircase Wear, Logistic Model, Crowd Clustering.

1. Introduction

Staircases in high-traffic environments experience significant wear that impacts their structural integrity and safety. Traditional maintenance approaches relying on visual inspections often fail to provide accurate, quantitative assessments of deterioration patterns. As urban infrastructure ages and pedestrian volumes increase, there is growing need for more precise methods to evaluate wear progression and predict service life. Current understanding of staircase deterioration lacks comprehensive analysis of how mechanical wear interacts with actual usage patterns under real-world conditions. This gap highlights the importance of developing data-driven approaches that can objectively quantify wear processes while accounting for complex human behavior factors in built environments.

The current research primarily focuses on three methodological approaches: regression analysis, time series modeling, and cluster analysis, which have demonstrated extensive applications across various engineering and scientific domains. In the field of regression analysis, multiple linear regression and discriminant classification models were employed to optimize monitoring strategies for water and sediment in the Yellow River [1]. Logistic regression models have shown effectiveness in diverse applications, including quality assessment of traditional Chinese herbs [8] and medical risk prediction for postoperative complications [9]. Additionally, multiple linear regression has proven valuable in geological fault analysis and extension prediction [10]. Time series analysis has been successfully implemented in infrastructure monitoring, particularly for subway tunnel settlement and deformation tracking [4]. Recent advancements have incorporated multi-source data integration to enhance time series forecasting accuracy [3]. Cluster analysis techniques have demonstrated utility in power system optimization through line loss modeling [5] and in material classification for archaeological studies [6]. The logistic growth model has also been effectively applied in transportation planning for vehicle ownership forecasting [2], while theoretical developments in random matrix theory continue to contribute to mathematical applications [7].

The main contributions of the article can be summarized as follows: 1) The study develops an integrated modeling framework combining logistic growth theory with clustering analysis to quantify stair wear dynamics, providing a novel approach for infrastructure condition assessment. 2) The research introduces a two-phase wear model that distinguishes between natural degradation and post-repair recovery patterns, enabling more accurate prediction of staircase service life under different maintenance scenarios. 3) The application of crowd behavior analysis through clustering techniques reveals distinct usage patterns, including a quantified annual usage frequency of 75.002 million people and identification of predominant upward movement tendencies. 4) The developed model demonstrates practical utility by identifying specific wear concentration areas (outer edges and right sides) and quantifying pedestrian flow characteristics, offering actionable insights for maintenance planning and stair design optimization.

The structure of this paper is organized as follows: The first section presents Related Theories, introducing the fundamental concepts of logistic growth models for wear progression analysis and clustering algorithms for crowd behavior pattern recognition, along with their respective applications in infrastructure monitoring. The second section describes Experiments, detailing the development of a two-phase wear model (unrepaired and repaired states), the integration of time series analysis for trend detection, and the implementation of clustering techniques to analyze pedestrian flow characteristics. The third section provides Results and Analysis, presenting quantitative findings including the annual stair usage frequency of 75.002 million people, spatial wear patterns showing predominant outer-edge and right-side degradation, and behavioral statistics revealing 82.4% side-by-side walking patterns. The final section contains Conclusions, summarizing the model's effectiveness in infrastructure assessment, discussing practical applications for maintenance planning and design optimization, and suggesting potential extensions to other high-traffic structures.

2. Related Theories

The logistic growth model is a concept used to describe how populations or systems expand over time when resources are limited. Initially, growth is rapid and nearly exponential, as conditions are ideal and resources are abundant. However, as the population increases, competition for resources intensifies, slowing the growth rate. Eventually, the system stabilizes at a maximum sustainable level, known as the carrying capacity. This S-shaped curve is widely applicable in fields like biology (e.g., animal populations), economics (e.g., market saturation), and epidemiology (e.g., the spread of diseases). Unlike unrestricted exponential growth, the logistic model captures the realistic constraints of real-world environments.

Clustering algorithms are unsupervised learning techniques that organize data points into meaningful groups based on their similarities. These algorithms identify patterns by grouping data such that points in the same cluster are more alike to each other than to those in other clusters. Popular methods include K-means, which divides data into a predefined number of clusters by iteratively refining their centers, and hierarchical clustering, which builds a tree-like structure of nested clusters. Another approach, DBSCAN, focuses on density, forming clusters where data points are closely packed together while marking outliers in sparse regions. These techniques are essential in applications like customer segmentation, image recognition, and anomaly detection, helping uncover hidden structures in large datasets without prior labeling.

3. Experiments

In order to address both the reliability of the age estimation of stairs and the repair or renovation status of stairs, this paper constructed a comprehensive model that combines a logistic growth model and time series analysis. This joint model not only simulates the wear and tear process of stairs, but also captures trend changes, seasonal fluctuations, and the effects of external events in the data to more accurately infer the service life and repair history of stairs. When analyzing the stair wear process, our study has divided the process of wear into two main stages:

No Repair Phase: In the case where the staircase has not been repaired or refurbished, a classical Logistic growth model is used to describe the wear process:

$$\frac{dN}{dt} = \gamma N\left(1 - \frac{N}{K}\right) \tag{1}$$

By solving this differential equation, this model can derive an expression for $N(t)$:

$$N(t) = \frac{K\dfrac{N_0}{K - N_0}e^{\gamma t}}{1 + \dfrac{N_0}{K - N_0}e^{\gamma t}} \tag{2}$$

where:

- $N(t)$ indicates the level of wear on the stairs at time t.
- γ indicates the rate of growth of wear and tear.
- K indicates the maximum wear value of the staircase, indicating the limit at which the staircase can no longer be worn.
- t indicates time.

Repair phase: As the staircase undergoes repair or refurbishment, the parameters γ and K of the model are adjusted according to the state of the staircase after repair. Repair may result in a decrease in the wear rate γ_{new} and an increase in the maximum wear value K_{new}, thus reflecting the restoration of wear and the extension of the service life of the staircase after repair:

$$\frac{dN}{dt} = \gamma_{new} N\left(1 - \frac{N}{K_{new}}\right) \tag{3}$$

where γ_{new} denotes the rate of wear after repair and K_{new} denotes the maximum wear value of the repaired staircase.

Clustering is the process of discovering natural groupings within data by identifying similarities between observations. The core idea is that data points within the same cluster should share common characteristics, while those in different clusters should be distinct. Unlike supervised learning, clustering doesn't rely on predefined labels-instead, it autonomously detects patterns based on the inherent structure of the data. The steps are shown in Figure 1:

Figure 1. Mind map for cluster analysis

Different clustering approaches emphasize different aspects of similarity. Some methods, like K-means, focus on minimizing the distance between points within clusters, effectively grouping data into spherical shapes. Others, such as spectral clustering, detect complex structures by analyzing connectivity in a graph representation of the data. Density-based techniques, like DBSCAN, excel at finding irregularly shaped clusters by identifying regions where points are tightly packed. The choice of algorithm depends on the data's nature and the desired interpretation, making clustering a versatile tool for exploratory data analysis across fields like biology, marketing, and social network analysis.

4. Results and Analysis

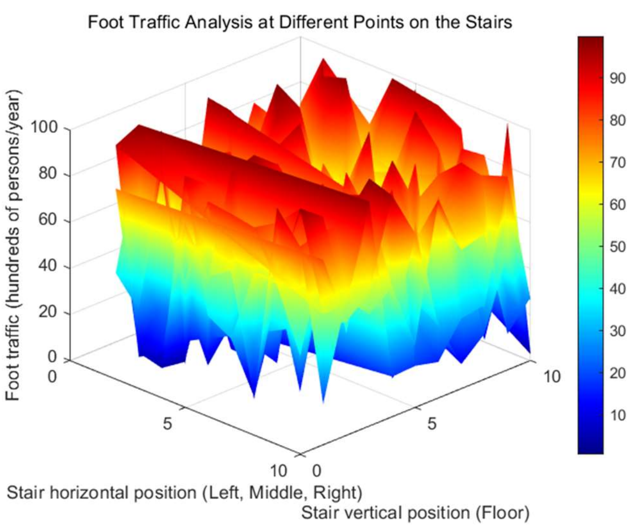

Figure 2. Step Wear Heat Map

Through fieldwork, this research measures the flow of people over a period of time, and then calculates the actual total wear of the steps through their own physical properties and specific wear and tear, and then calculates the actual total wear and tear based on the crowd-driven model, solving for the conversion factor, and then generalizing the model, allowing us to measure and calculate the actual total wear and tear of the steps to calculate the flow of people, and ultimately solving for the frequency of use of the stairs. Finally, our study calculated the frequency of stair use to be 75.002 million people/year which is shown in Figure 2.

By categorizing the crowd to calculate the actual total wear driven by the crowd and measuring the wear distribution of the ladder, this study can make Figure 3:

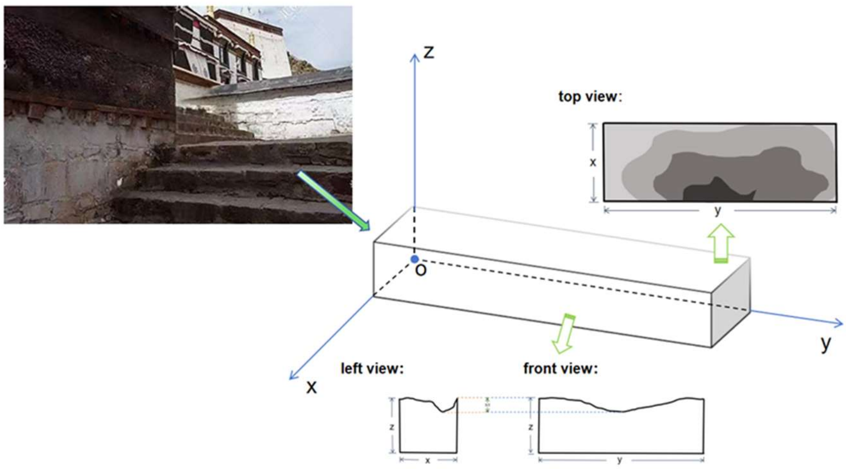

Figure 3. Schematic detail of stair wear

It can be seen that the wear is more severe on the outside of the step and the right side of the step. Firstly, according to the crowd-driven wear model, this study can conclude that the wear on the outside of the step is more severe because this staircase is more suitable for upward movement because the crowd-driven wear model suggests that the center of gravity of people is more biased towards the outside of the step during upward movement, and is more biased towards the inner measurement during downward movement.

Under the condition that this staircase is more suitable for upward movement, our paper and model observe the comparison of wear on the left and right sides of the staircase, and find that the wear on the right side of the staircase is more severe, so this paper can roughly conclude that there are more people traveling up and down on the right side of the staircase than on the left side of the staircase .So this paper can conclude that people tend to use this staircase more for upward traveling

Assuming that the staircase is a bi-directional staircase, during the specific operation of the staircase, people tend to choose to avoid the staircase in consideration of the presence of people in the opposite direction, so in solving the problem, this paper do not analyze the case of walking on the left side of the staircase when going up as well as the case of walking on the left side of the staircase when going down. First, our models analyze each case of side-by-side walking according to the population model, and conclude that side-by-side walking has the same effect on the wear and tear of the left and right staircases and is equal to the wear and tear produced for the middle, so this paper can get the equation:

$$
\begin{cases}
A_1 \cdot 0.25 \cdot K_1 + C_2 \cdot K_2 + A_2 \cdot 0.25 \cdot K_2 = W_1 \\
A_1 \cdot 0.5 \cdot K_1 + B_1 \cdot K_1 + B_2 \cdot K_2 + A_2 \cdot 0.5 \cdot K_2 = W_2 \\
A_1 \cdot 0.25 \cdot K_1 + C_1 \cdot K_2 + A_2 \cdot 0.25 \cdot K_2 = W_3 \\
W_1 \cdot \left(\dfrac{A_1}{P_1}\right) + W_2 \cdot \left(\dfrac{P_2}{A_2}\right) + K_1 \cdot \left(\dfrac{A_1 + A_2}{K_2}\right) = C_2 \cdot A_2 \cdot \left(\dfrac{W_1}{C_1}\right) + P_1 \cdot A_2 \cdot \left(A_1 + A_2\right) \\
A_1 + B_1 + C_1 = P_1 \\
A_2 + B_2 + C_2 = P_2
\end{cases} \tag{4}
$$

From this system of equations it is easy to see that there are 6 variables and 6 systems of equations, and our research can then calculate the values of each variable. With the crowd model and the stair wear model, combined with these equations this paper can calculate the percentage of the total crowd that shown in Table 1.

Table 1. Percentage of total population

	People walking side by side	People walking in single file
Percentage of total population	0.82436	0.17564

5. Conclusion

This study demonstrates that combining logistic growth modeling with clustering analysis effectively quantifies stair wear dynamics and usage patterns. The logistic model accurately captures wear progression under both unrepaired and repaired conditions, showing how maintenance interventions alter degradation rates and extend service life. Field measurements reveal an annual usage frequency of 75.002 million people, with wear patterns indicating predominant upward movement and side-by-side walking behavior. The model's ability to integrate physical wear mechanics with crowd behavior provides practical insights for maintenance planning and stair design optimization. These findings highlight the value of hybrid modeling approaches for infrastructure management, offering a framework that could be extended to other high-traffic structures. The study underscores how theoretical models combined with data-driven techniques can address real-world engineering challenges.

References

[1] GENG Yinghao, ZHOU Xiaodong, QI Junlu, et al. Research on optimization strategy of Yellow River water and sand monitoring based on multiple linear regression and discriminant classification model[J]. Science and Technology Innovation,2025,(04):60-63.

[2] Wang P. Forecast of private car ownership in Harbin city based on logistic model[J]. Heilongjiang Transportation Science and Technology,2024,47(08):153-157.DOI:10.16402/j.cnki.issn1008-3383.2024.08.013.

[3] Moying Li. A time series analysis model based on multi-source data[J]. Information Technology,2025,(01):112-118+125.DOI:10.13274/j.cnki.hdzj.2025.01.018.

[4] HER YC, GUO GI X, LIU Han. Application of time series analysis model in monitoring settlement and deformation of subway tunnels[J]. Surveying, Mapping and Spatial Geographic Information,2022,45(09):238-240.

[5] Zhengang Zhang,Yong Yang,Gaofei Yan. Research on line loss model based on cluster analysis[C]//Thirty-third Beijing-Tianjin-Hebei-Jin-Meng-Lu-North China Six Provinces and Cities (Regions) Electrical Engineering (Electric Power) Society Proceedings. Handan Power Supply

Bureau of Hebei Power Grid Limited Liability Company;,2023:5.DOI:10.26914/c.cnkihy.2023.040778.

[6] Shao Guangming,Xia Xianqi,Yin Hejie. Research on ancient glass classification prediction model based on CART and cluster analysis[J]. Journal of Tonghua Normal College,2024,45(02):31-35.DOI:10.13877/j.cnki.cn22-1284.2024.02.005.

[7] Guangyi Zou. Marginal spectral distribution and graphical expansion of random banded matrices [D]. University of Science and Technology of China, 2024.

[8] Dong Shuangtao, Xu Lixia, Gao Jianping, et al. Prediction of quality grade of Baiqian based on one-measure-multi-rating scale combined with chemometrics and logistic regression model[J]. Chinese herbal medicine,2025,56(02):656-666.

[9] HAO Guangzhi, SUN Linlin, ZHANG Bingying, et al. Construction of a risk prediction model for postoperative hydrocephalus after cerebral hemorrhage based on LASSO-Logistic regression[J/OL]. Chinese Journal of Neurosurgical Disease Research,2025,(01):13-19[2025-01-27].http://kns.cnki.net/kcms/detail/61.1384.R.20250114.1443.006.html.

[10] Zhou P. Research on the prediction of extension length of small faults under multiple linear regression model and evaluation of effect[J]. Inner Mongolia Coal Economy,2024,(22):1-4.DOI:10.13487/j.cnki.imce.026082.