Problem Chosen	2025 MCM/ICM Summary Sheet	Team Control Number
A		2522170

Stairway to Wear: Analyzing Stair Behavior with Wear Fields and Convolutional Networks

Summary

Have you ever noticed the **insignificant** stairs when visiting temples or churches, worn down by the steady passage of countless individuals over time? Unassuming marks left by human footsteps endure steadfastly over time. Archaeologists rely on these traces of wear to study human path across historical periods. This paper conceptualizes this wear as a "field" to examine the usage dynamics of stairs.

For task 1, we construct a dataset by collecting data on the wear LiDAR point cloud data of the stairs, material types, human behavior, and other relevant factors. By combining **Archard wear law** and **differential equations**, we develop an integrated model for **reverse** deduction of stair wear analysis and crowd behavior inference. This model allows us to calculate the wear depth, footstep count, and footstep frequency for each wear region.

For task 2 and 3, we define the "**wear field**", a quantifiable representation of wear intensity and direction. A dynamic behavior analysis of this field is then conducted, incorporating fundamental knowledge from field theory. Analyzing its **gradient**, **divergence**, and **curl** to characterize directional preferences of individuals, evaluate crowd distribution density, and identify parallel behaviors. Additionally, this approach helps assess more complex movement patterns, such as turns or lingering.

For task 4 to 8, environmental and topographical factors are introduced as key variables. Building on the conclusions derived from task 1 to 3, a combination of **linear regression, logistic regression, probabilistic statistics, K-means material clustering**, and **Kernel algorithms** are applied to derive the solutions for each problem. Considering the fragmentation and dispersion of traditional methods, **CNN** is introduced to enhance the model's fitting capabilities. This allows for a deeper exploration of the relationships between environmental, topographical factors, and the various issues at hand, thereby improving the accuracy and efficiency of the analysis. Due to CNN's sensitivity to data patterns, **regularization bias** is incorporated in the final model to prevent overfitting, ensuring a more robust and reliable outcome.

Finally, this paper presents a series of methods to predict the annual usage of stairs. Each staircase can be viewed as a "field," much like our lives can be understood as a "**Life Field**" enduring and eternal, similar to the timeless nature of stone stairs.

Keywords: LiDAR Point Cloud; Wear Field; Regression Analysis; CNN; Regularization Bias

Contents

I Introduction

1.1 Problem Background

Stairs in ancient structures, whether made of stone, wood, or other materials, exhibit unique wear patterns over time. These patterns can reveal critical insights about the use and history of the structure. The wear on stairs is influenced by factors like the number of people using the stairs, the frequency of usage, and the direction of travel. Despite their durability, materials like stone are not impervious to wear, and over centuries or millennia, the edges and treads of stairs can undergo significant changes.

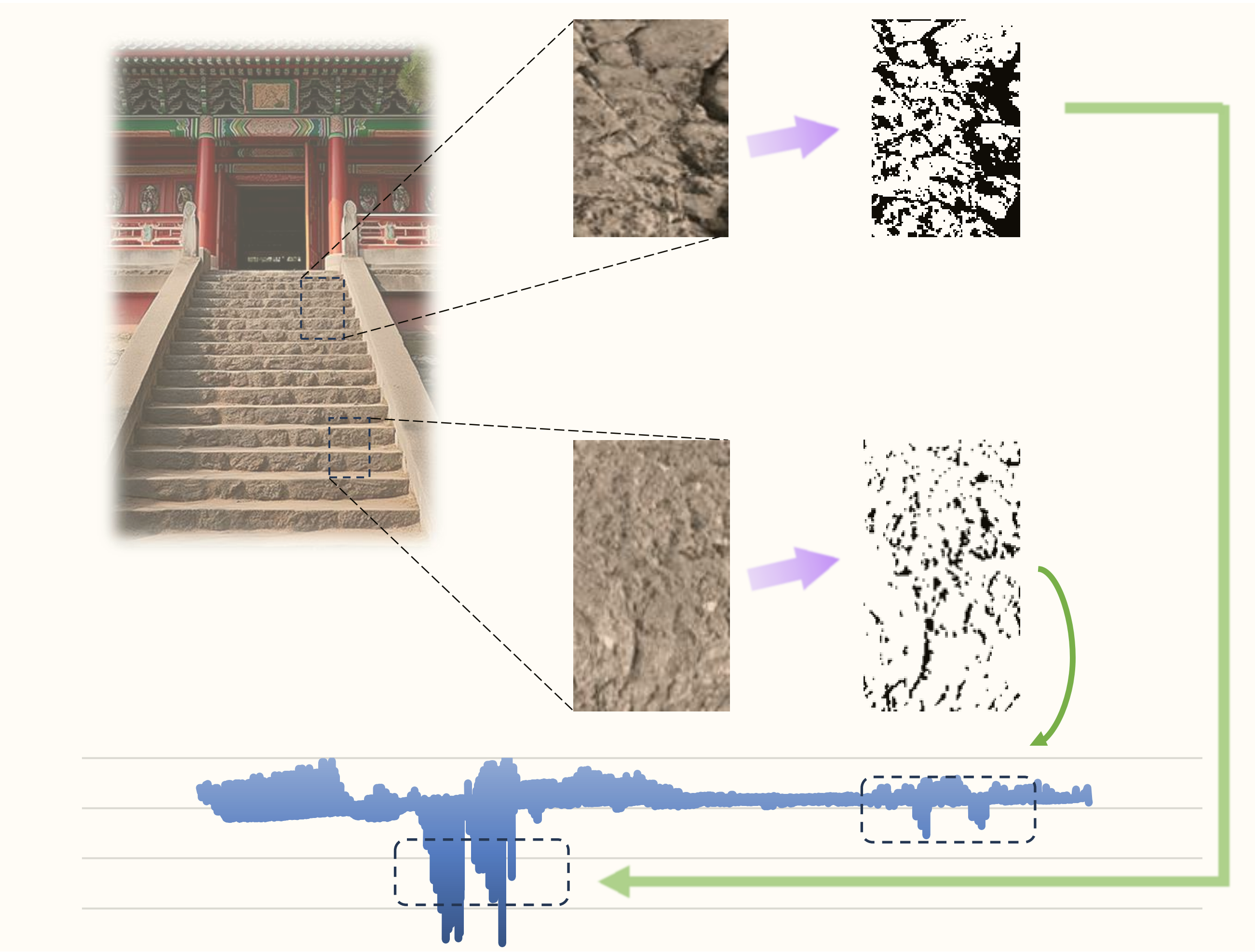


Figure 1: Exhibition of ancient building staircase wear

A rough estimate of how often a staircase is used can be made by examining the wear. This can be reflected in the depth and evenness of wear across the surface. Areas of more pronounced wear may indicate more frequent traffic, while areas of less wear may indicate periods of lesser use. As shown in Figure 1 the upper step is clearly more heavily worn than the lower one. However, by intuition, the lower step may have been weakened by wear due to wind and sand buildup.

In general, in ancient buildings where the gate is located in the center, people tend to converge on the gate from different directions when they enter, unless there are specific religious practices, such as the requirement to walk around the temple in a clockwise direction.

As people walk up the stairs, they leave marks that create wear. Therefore, wear patterns can be used to estimate the frequency of stair usage. Areas with more pronounced wear may indicate higher traffic, as people tend to walk there more often, while areas with less wear may suggest lower usage. Wear can also provide insight into how many people are using the stairs simultaneously.

1.2 Problem Restatement

We are asked to provide guidance on what information archaeologists can determine from a set of worn stairs. The problem involves analyzing the wear patterns on a staircase to draw conclusions about its usage history, particularly in the context of ancient buildings like temples or churches. The goal is to use physical properties of the staircase to infer multiple aspects of its historical usage. Tasks involved:

1. Measure and analyze the wear depth and distribution across the staircase. Look for patterns that could indicate high-traffic areas.
2. Estimate how often the stairs were used and how many people used the stairs simultaneously from wear depth and distribution.
3. Investigate whether the wear is asymmetric and use wear patterns to infer if people used the stairs singly or in pairs.
4. Search for relevant information to compare wear in our models.
5. Use wear patterns to estimate the true age of the staircase and determine the reliability of the given age estimate.
6. Determine if there are any signs that the staircase has been repaired or renovated.
7. Investigate how material properties such as stone or wood have degraded over time and whether the wear is consistent with the time periods during which the stairs were in use.
8. Analyze the number of people using the stairs in a typical day, focusing on whether there is a large number of people using the stairs over a short time or a small number of people using it over a longer time.

1.3 Our Approach

In order to avoid complicated descriptions, intuitively reflect our work process, the flow chart is shown in Figure 2

Firstly, we collect Architectural history, human behavior, Scatter plot, Step size, Material properties and wear function to build up a database, and utilize the ideas of Wear amount, Archard's law and Differential equation to build up a Wear field (WF). The Wear Field (WF) is established by utilizing the ideas of Wear Volume, Archard's Law and Differential Equation.

Second, the data were analyzed through the wear field, using gradients, dispersion and spin in order to find the results of Task1,2,3.

Third, introduce environmental and topographic variables, and use data analysis to apply regression, statistical, and other methods to solve questions 4-8 separately. Then, integrate the results to obtain a comprehensive predictive value.

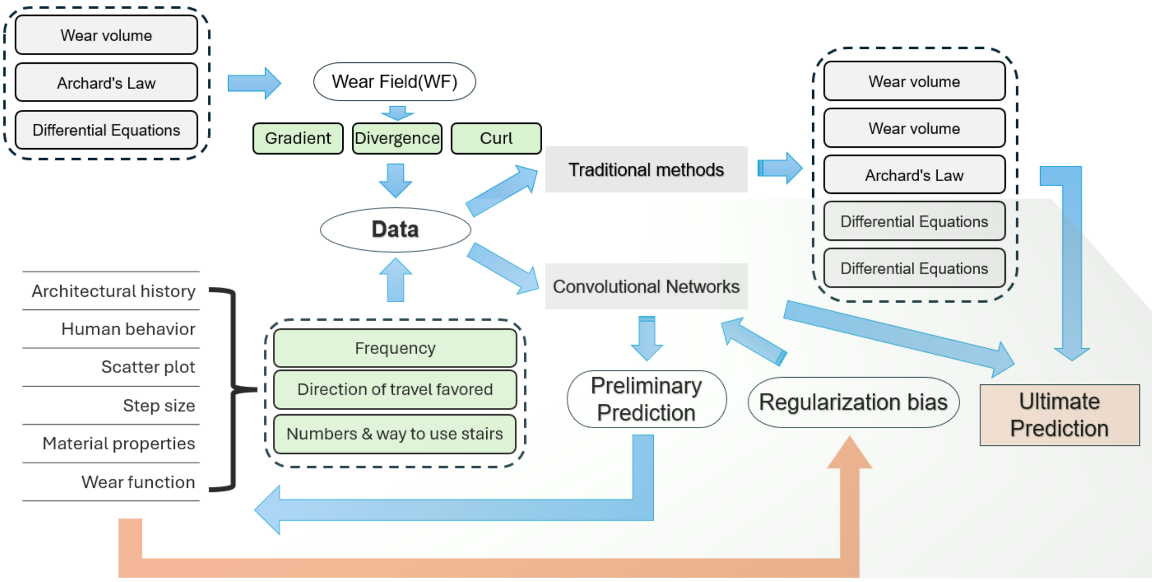


Figure 2: Mind map of our work

Fourth, use a three-layer convolutional neural network to train all the previously collected data, optimizing and adjusting hyperparameters to ultimately obtain another prediction value. To prevent overfitting, regularization techniques are applied to refine and optimize the model, improving prediction accuracy.

Finally, compare and integrate the results from both prediction methods to obtain the final optimized prediction value.

II Assumptions and Justifications

Assumption 1: The wear of the stairs plane is regarded as a flat field (not curved field).
Justification: A section of a step plane at any point wear is considered, there is a magnitude and direction of the amount, then the wear field exists on the step, like the electric field and magnetic field.
Assumption 2: Sample totals using mean substitution for basic human data
Justification: The study of the problem is not a priority and has little impact on the assessment of the overall system, while the data on human beings is heterogeneous and complex, and influenced by many factors, such as culture and religion, so is not considered.
Assumption 3: Repairing gives a new staircase without changing the characteristics of the staircase itself.
Justification: Wood rectorate wooden staircases, the same stone rectorate stone staircases, and as adhesives etc. have less effect on their own properties.
Assumption 4: The direction of travel favored by the people is not affected by the nature of the steps.
Justification: To simplify the calculation, the effect of stair pattern and stair damage on stair use is not considered.

III Notations

Table 1: Notations used in this paper

Symbol	Description	Unit
$\boldsymbol{W}$	Wear volume vector at a point on the step	m^3
L	Slide path vector	m
k	Wear coefficient	/
h	Material hardness parameters	Pa
Q	Energy consumed at a point through wear	J
P	Wear power at a point	W
$\boldsymbol{F}$	Local force at a point during wear	N
$\boldsymbol{v}$	Slide path velocity vector	$m \cdot s^{-1}$
f	Footstep frequency in staircase	t^{-1}
R	Wear coefficient	$m^3 \cdot t^{-1}$
N	Footstep count in staircase	/

IV Task1, 2 and 3: Estimation of Wear Characteristics

The Geghard (Armenian: Գեղարդ) Monastery is in Armenia and is rich in Christian historical and cultural significance. Its central church was built between 1215 and 1283, and in 2000, listed as a UNESCO World Heritage site.

This chapter analyzes one of the granite rock stairs at the monastery as a case study, with the approach outlined in the following diagram.

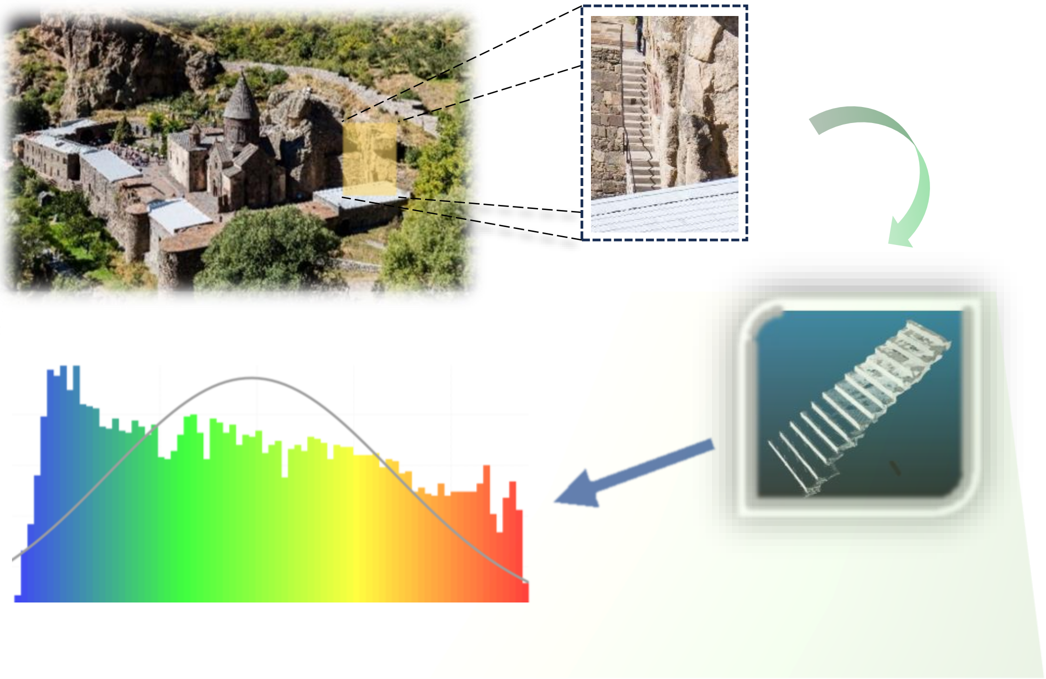


Figure 3: The Geghard Monastery Data Extraction

4.1 Data Searching and Preprocessing

LiDAR point cloud:

In order to obtain data that can be analyzed and to avoid damaging the building proper, we obtained LiDAR maps of the ancient building through Open Heritage's publicly available data. Open Heritage uses a LiDAR method to non-destructively and with low energy consumption obtain surface data of stairs. After extracting the overall data of the staircase in Cloud Compare, cuts were made to obtain data for each piece of the steps. The height distribution maps were generated after the stair surfaces were extracted by cleaning the point cloud data, removing noise and invalid points, and these scatter data were stored in CSV files for subsequent analysis.

Stair geometry data:

Identify the wear regions by extracting the surface curvature, thermograms through the above csv file. Obtain boundary data and segment independent wear regions Obtain height boundary data by processing the point cloud data and using a local fitting method to calculate the surface height variation Finally, by outputting a height map as an approximation of the wear depth. Segment the LiDAR map using a clustering algorithm to extract hardness data for multiple independent wear regions R.

Assuming a section of step wear is continuous. But since the data sought is discrete, the step surface is divided into discrete grid cells using a point cloud meshing method with each grid cell defined as (i, j, k).

Material properties:

Table 2: Wear coefficients of different materials

Material Type	**Wear Coefficient (k)**	**Description**
Stone	0.01~0.05	Marble, sandstone, **high** hardness and excellent wear resistance
Wood	0.07~0.30	Softwood exhibits **rapid** wear; hardwood exhibits **slower** wear
Concrete/Brick	0.02~0.06	**Medium** hardness
Metal	0.005~0.02	**Minimal** wear **resistance** but is **highly susceptibl**e to corrosion
Composite	0.02~0.04	Offers superior wear **resistance** and **stability**

Human behavior and architectural history:

Human behavior is highly complex and random. To simplify the analysis, global human physiological data is obtained from the World Health Organization, and the mean values are calculated with a bias adjustment for reference. Meanwhile, the history of architecture evolves with societal needs, technological advancements, and cultural changes. By examining the construction history and visitation data of buildings, trends in the use and functional changes of different building types throughout history can be analyzed.

Wear function:

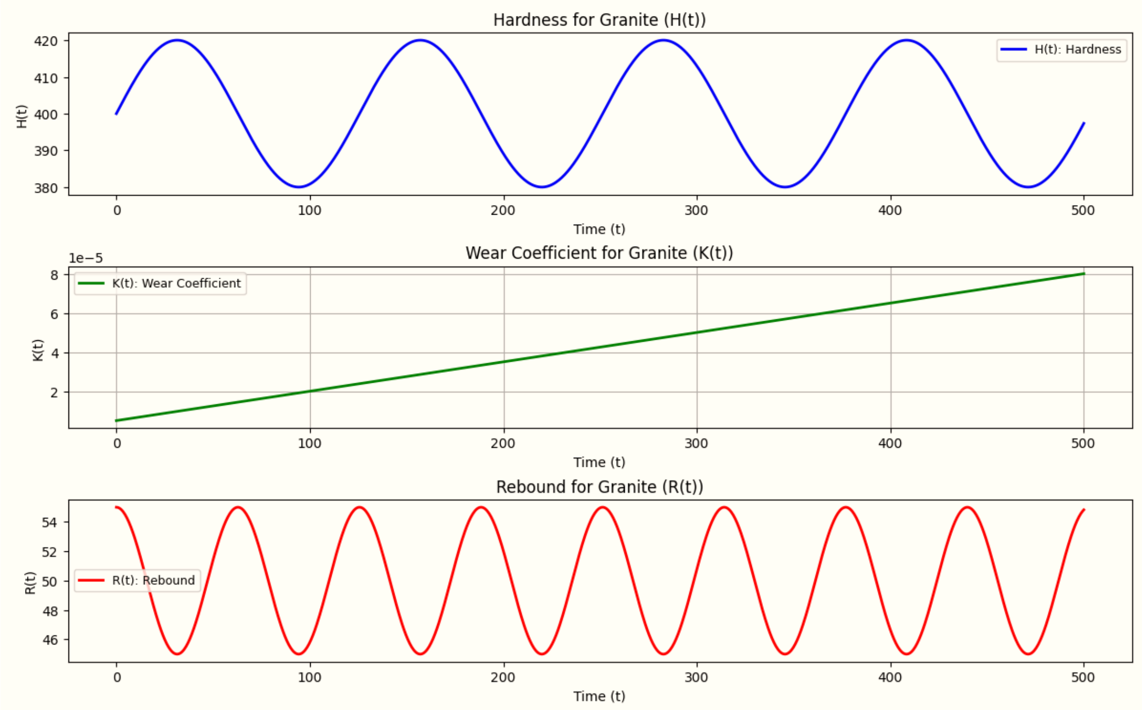


Figure 4: H,W,R **for Granite**

4.2 Vectorized Wear for Establishing Wear Field

Traditional Archard wear law:

$$W=\frac{k\boldsymbol{F}\cdot\boldsymbol{L}}{h}$$

Where W is the wear volume, $\boldsymbol{F}$ is the local force, $\boldsymbol{L}$ is the sliding path vector, k is the wear coefficient, and h is the material hardness parameter.

By discretizing, we get:

$$\Delta W=k\frac{\boldsymbol{F}\cdot\boldsymbol{v}\,\Delta t}{h}$$

Considering the rebound feedback again, then:

$$\frac{\partial W}{\partial t}=k\frac{\boldsymbol{F}\cdot\boldsymbol{v}}{h}+R$$

R is the rebound function, t is the elapsed time.

Then:

$$\frac{W_{n+1}-W_{n-1}}{2\Delta t}=k\frac{F\cdot v}{h}+R$$

Now, setting Wear Field(WF):

The wear distribution of the step is a three-dimensional vector space $\boldsymbol{d}=(x,y,z)$. The wear volume W and its unit normal vector $\boldsymbol{n}$ define the wear field as:

$$\boldsymbol{W}=W\cdot\boldsymbol{n}$$

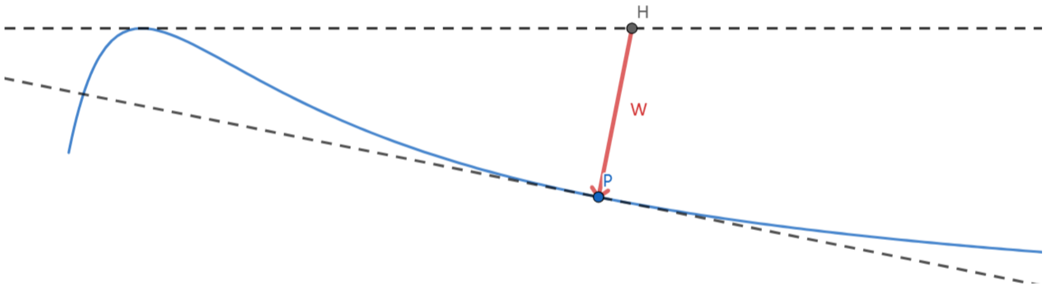


Figure 5: Wear Field Direction Schematic

Wear volume:

$$W=\sqrt{(1-d(x,y))^{2}\left(\left(\frac{\partial d}{\partial x}\right)^{2}+\left(\frac{\partial d}{\partial y}\right)^{2}+\left(\frac{\partial d}{\partial z}\right)^{2}\right)}$$

Normal vector:

$$\boldsymbol{n}=\left[\frac{\partial d}{\partial x},\frac{\partial d}{\partial y},\frac{\partial d}{\partial z}\right]^{T}$$

Then:

$$\boldsymbol{W}=\begin{bmatrix}\dfrac{\partial z}{\partial x}(d(x,y)-d_{\mathrm{m}})\\ \dfrac{\partial z}{\partial y}(d(x,y)-d_{\mathrm{m}})\\ 1-d(x,y)\end{bmatrix}$$

Number of pedals and frequency of use:

In $W=\dfrac{k\boldsymbol{F}\cdot\boldsymbol{L}}{h}$, k, h are material properties, which can be obtained from the above LiDAR diagram, while the stepping frequency $\boldsymbol{N}$ is related to the applied force $\boldsymbol{F}$ and the sliding path $\boldsymbol{L}$.

And $Q=F\cdot\boldsymbol{L}$, $P=F\cdot\boldsymbol{v}$, Therefore, the total footstep frequency N and footstep frequency f:

$$N=\frac{h\cdot\|\boldsymbol{W}\|}{k\cdot Q}\qquad\qquad f=\frac{h\cdot\|\boldsymbol{W}\|}{k\cdot P}$$

Obtain frequency distribution:

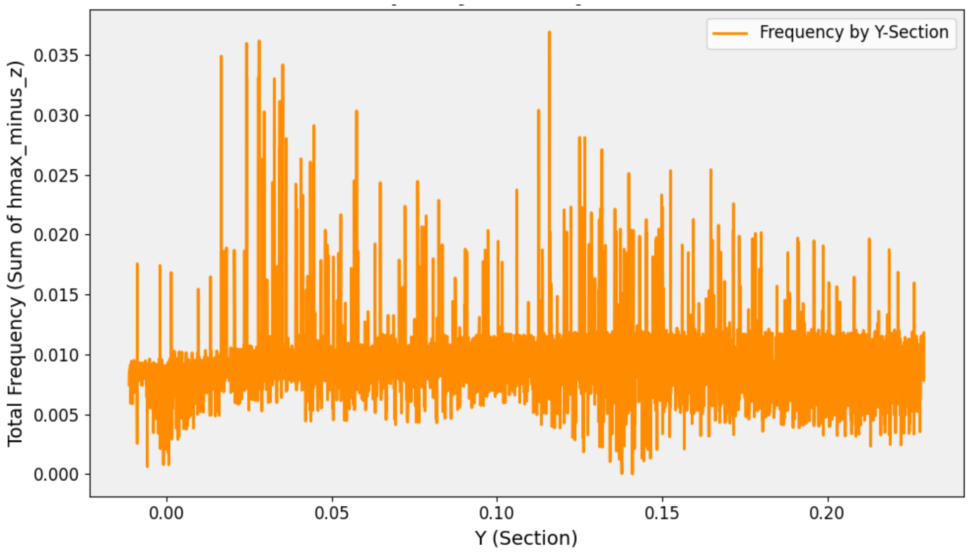


Figure 6: Frequency Curve

4.3 Dynamic Behavioral Analysis

After establishing the wear field, we use field theory to reveal directional preferences through gradient, determine complex behaviors such as circulation and stay, and divergence is used to evaluate the density and parallel behavior of crowd distribution. The integration of these parameters enables us to more comprehensively analyze Frequency, Direction of travel favored and way to use stairs.

Gradient analysis:

The wear gradient of each grid cell is calculated using the finite difference method (FDM):

$$\nabla \boldsymbol{W} = \left(\frac{\partial \boldsymbol{W_x}}{\partial x}, \frac{\partial \boldsymbol{W_y}}{\partial y}, \frac{\partial \boldsymbol{W_z}}{\partial z}\right)$$

Discretization:

$$\nabla \boldsymbol{W} = \left(\frac{W_{i+1,j,k} - W_{i-1,j,k}}{2\Delta x}, \frac{W_{i,j+1,k} - W_{i,j-1,k}}{2\Delta y}, \frac{W_{i,j,k+1} - W_{i,j,k-1}}{2\Delta z}\right)$$

The direction of the gradient indicates the direction of the primary force on wear, and the magnitude indicates the rate of change.

The direction of the gradient points in the direction of the most intense wear. If the gradient is significant along a particular direction, it indicates that the footwear prefers that direction. If the gradient is uniform, it indicates a more random crowd activity.

Figure 7 displays the gradient direction distribution of the staircase section of Geghard Monastery. This distribution is based on eight directions, with 300 data points randomly sampled from approximately 6,000 data points for analysis.

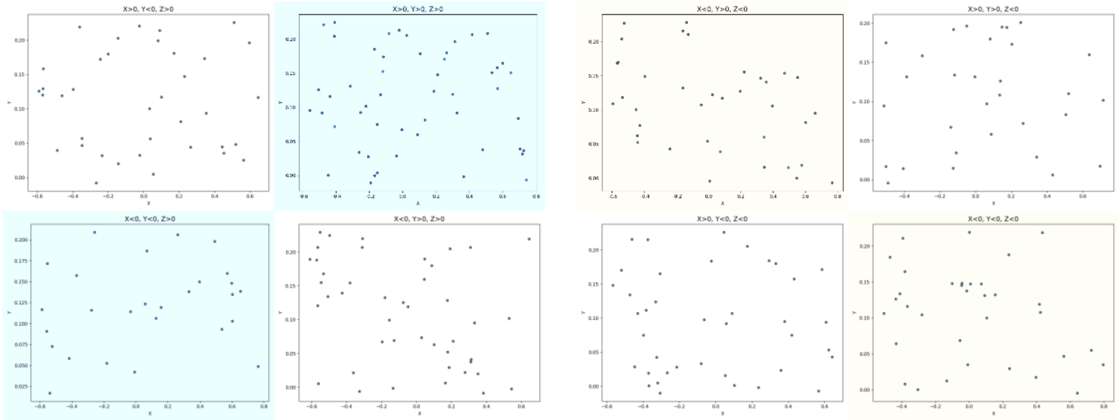
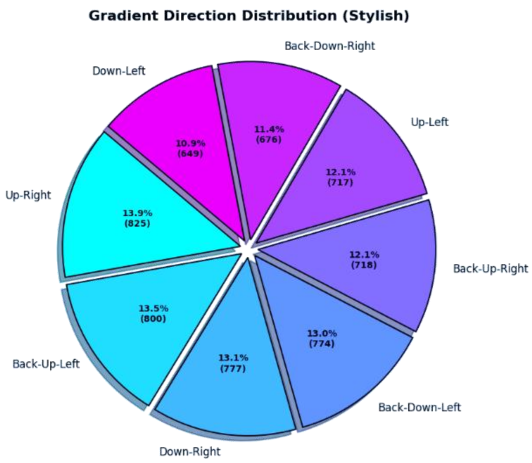


Figure 7: Scatterplot of Gradient Direction Distribution

The analysis indicates that the number of individuals traveling uphill exceeds those traveling downhill, and a greater proportion of individuals tend to favor the right side compared to the left side. This finding is consistent with the real-world scenario, where the right side is characterized by a mountainside and the left side by a railing.

Figure 8 is pie charts depicting the distribution of individuals across eight directional sectors. Given the large dataset and potential factors such as maintenance or renovations, the proportions across the eight directions are approximately equal. Nevertheless, the uphill direction and rightward preference are still predominant.

Figure 9 illustrates a 3D gradient directional plot based on a random sample of 50 data points. The results reveal a significantly higher number of vectors oriented towards the right and a higher density of points located above, compared to those below, thus validating the conclusions.



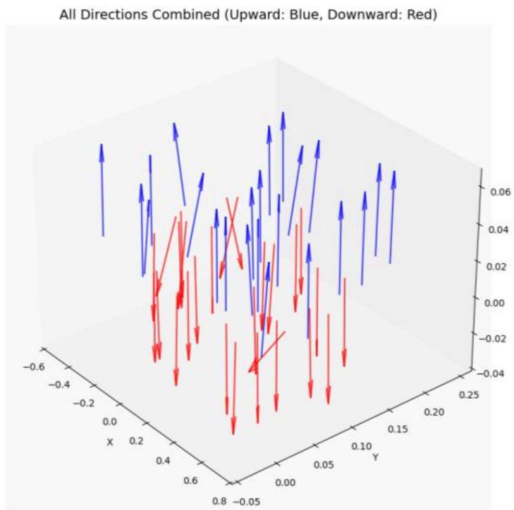


Figure 8: Pie Chart of Gradient Direction **Figure 9: 3D gradient directional plot**

Curl analysis:

$$\nabla \times \boldsymbol{W} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \boldsymbol{W_x} & \boldsymbol{W_y} & \boldsymbol{W_z} \end{vmatrix}$$

If $\nabla \times W \neq 0$, cyclic behaviors such as stopping and turning are present.

If $\nabla \times W = 0$, the footpath is linearly distributed.

As shown in Figure 10, the upward arrows are noticeably longer than the downward arrows, indicating that there are more upward detours than downward ones. This is reasonable, as it reflects the common observation that individuals often ascend several steps before deciding to retrace their path. Furthermore, the arrows pointing from the upper right to the lower left are particularly dominant, suggesting that a greater number of people tend to move from the left side to the right, following the natural contours of the hillside. The diversity in arrow directions further implies that the staircase resembles those found in scenic areas, where people are free to move without a specific destination in mind. This observation is consistent with the real-world context, as the staircase in question is located externally and follows the topography of the mountain.

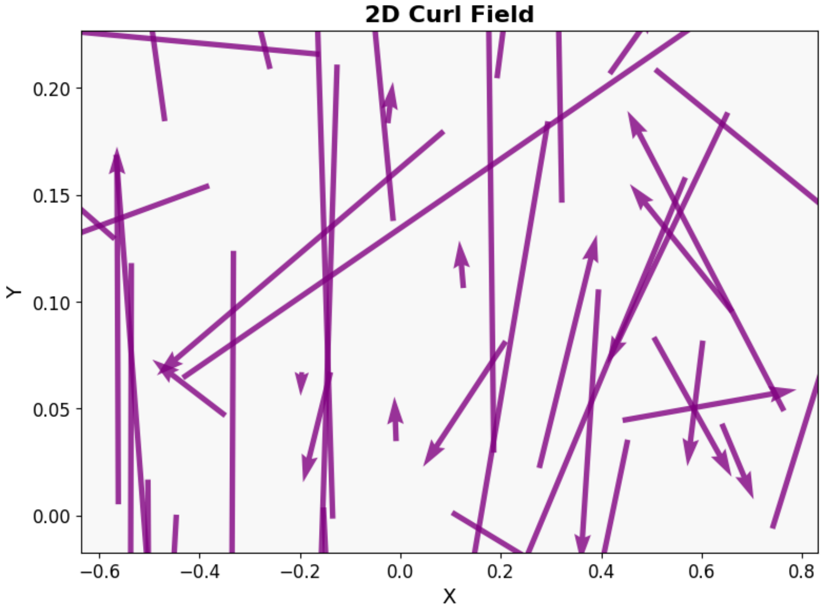


Figure 10: Curl Field Analyze

Divergence analysis:

$$\nabla \cdot \boldsymbol{W} = \frac{\partial \boldsymbol{W_x}}{\partial x} + \frac{\partial \boldsymbol{W_y}}{\partial y} + \frac{\partial \boldsymbol{W_z}}{\partial z}$$

If a church features a large set of stairs descending from the lower center, individuals are more likely to converge towards the entrance from various directions, analogous to the behavior of a negative charge, which attracts movement from all sides. In contrast, during departure, there is a tendency for individuals to disperse away from the entrance, mirroring the dispersal pattern of a positive charge. This analogy highlights the directional flow of movement in response to the spatial configuration of the structure.

Areas of large scatter indicate multiple people walking in parallel resulting in wear spreading; areas of small scatter indicate a single person walking in single file.

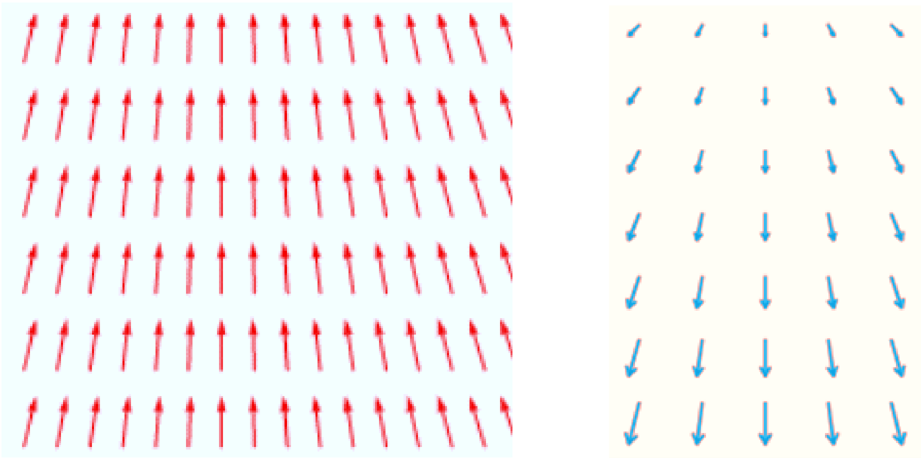


Figure 11: Divergence Field Analyze

Figure 11 is an optimized Divergence Field analyze, illustrating the typical flow of people either moving from various positions at the lower level towards the central upper area or dispersing gradually from the central upper area. This represents the idealized scenario. However, due to the large volume of data, a more detailed analysis, considering the number of stairs, their placement, and the functional role of the building, is necessary. Therefore, for the sake of clarity, up until here, only a simple path analysis will be performed in this paper.

V Task 4 to 8: Regression Analysis

The first step to be taken is to verify that the actual wear and tear is as expected, which requires that the dependent and independent variables be defined firstly.

Dependent variable is the wear volume vector. It’s obtained by physical measurements. The wear data is related to the morphological changes of the stair surface on the point cloud map. Independent variables are applied force, material coefficient, rebound coefficient, climate factors, topographic coefficient.

The selection of building materials and their wear patterns are typically influenced by local geological and climatic conditions. Different geological and climate environments significantly affect the types of building materials used in specific regions. Therefore, this chapter aims to establish a correlation model between material wear patterns and various environmental factors. The model will be based on global geological and climate type distribution maps, incorporating two typical types of wood and stone—granite, marble, pine, and oak. Key parameters will be extracted for estimating the building lifespan, followed by validation and testing.

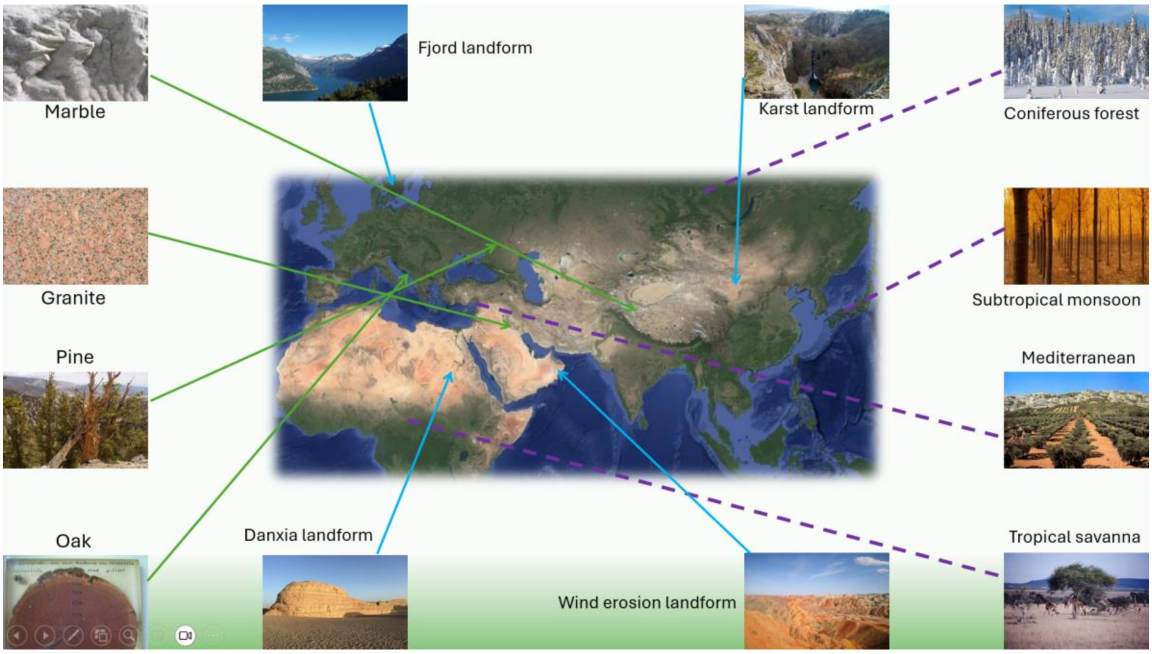


Figure 12: Types of landforms and climates towards common materials

As illustrated, we have gathered data on common climates (such as temperate monsoon climate), landforms (such as karst topography), and material sources (such as granite) from around the world, quantified them, in order to more comprehensively assess the mechanisms of wear.

Chichen Itza, situated in the northern part of the Yucatan Peninsula in the state of Yucatan, Mexico, was constructed by the Maya civilization and is recognized as one of the New Seven Wonders of the World. It was inscribed as a UNESCO World Heritage Site in 1988, standing as a magnificent manifestation of human culture.

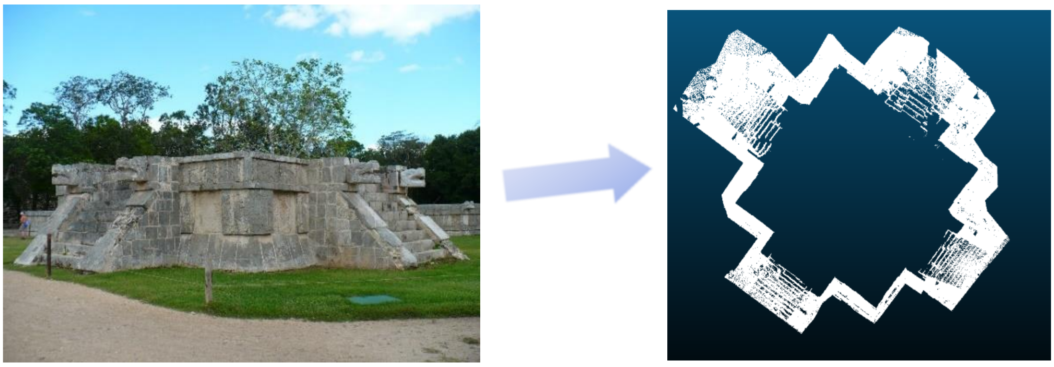
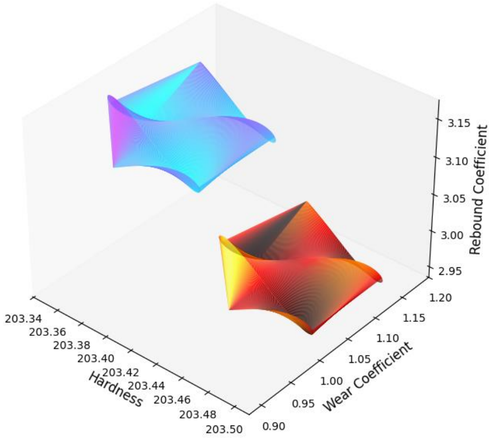


Figure 13 Chichen Itza relics

Taking the four sections of stairs in the building as an example, we collect data on the local stone materials, environmental factors, as well as the applied forces and wear conditions. A random selection of six stair sections is made, with five of the steps forming the initial dataset. The sixth set of data is then trained based on this information. Finally, the trained sixth set of data is compared with the actual data for validation and analysis, ensuring the correctness of Tasks 1, 2, and 3.



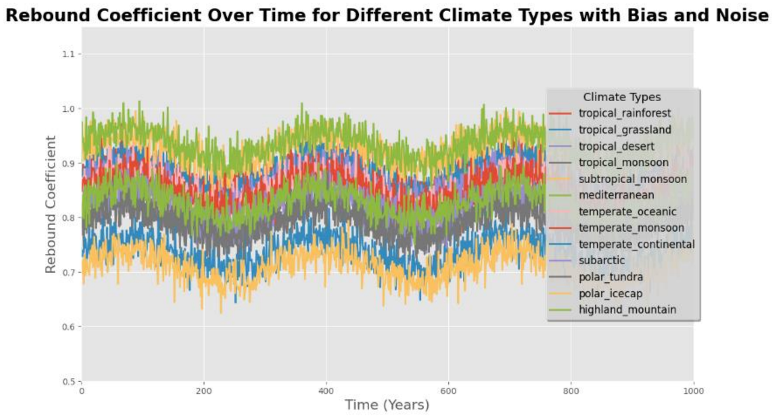


Figure 14: Granite vs. Wood **Figure 15: Rebound Coefficient Over time**

5.1 Task 4: Correlation Checking on Analysis

Build a regression model: Establish the relationship between wear data and independent variables:

$$\boldsymbol{W} = \vec{\alpha} \cdot \mathbf{X} + \epsilon$$

Where $\vec{\alpha} = [\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4]$ is the regression coefficient vector, which indicates the influence of each factor on wear. $\mathbf{X} = [\varphi(k,h,R),\boldsymbol{F},E,L]^{\mathrm{T}}$ denotes the aggregate vector of applied force, hardness, wear coefficient, rebound coefficient, environmental coefficient, material coefficient and geomorphology coefficient. ϵ is the error term.

Figure 16 is a comparison between the predicted and actual data. It is evident that the actual data exhibits a greater degree of dispersion, while the predicted trend closely aligns with the real trend, demonstrating a high level of accuracy in the prediction.

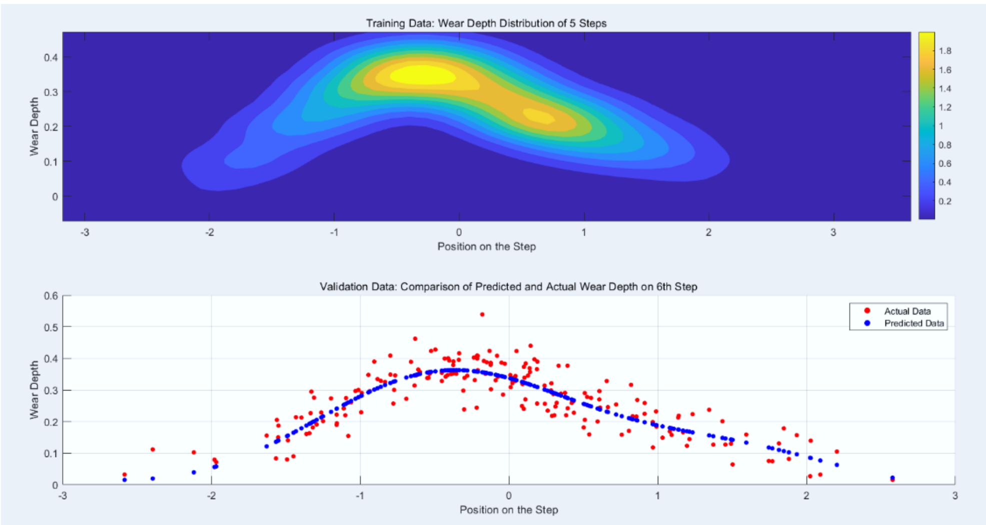


Figure 16: Comparison of prediction and reality

The R^2 value is calculated to characterize the accuracy of the model. As shown in Figure 17, the maximum value of R^2 does not exceed that of (1.4×10^{-5}, 8.0×10^{-20}), confirming that the wear pattern is consistent with the information provided. This supports the reliability of the model's predictions.

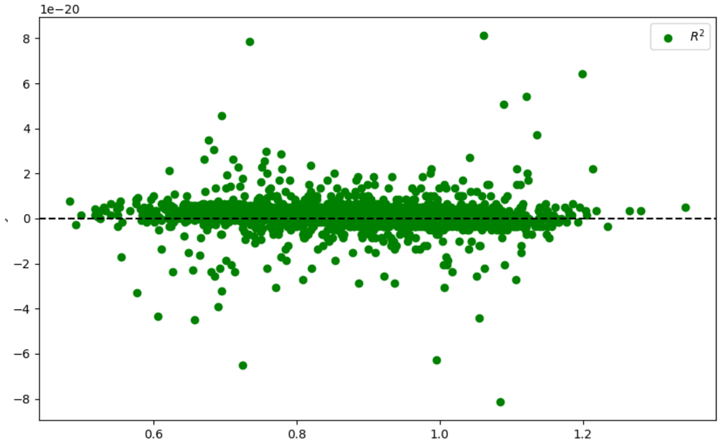


Figure 17: R^2 analysis

5.2 Task 5: Estimation of the Age of the Stairs

To estimate the age of the staircase, a model is constructed using existing data, leveraging the principles of logistic regression, while recognizing that age is a continuous variable, necessitating the application of a linear regression model.

To enhance the model's performance, feature engineering is applied to extract relevant features that improve age prediction. These features include geographic location attributes, wear gradient rotational dispersion, material characteristics, and environmental factors such as climate and geomorphology. Given the significant variation in the magnitude of these features, Z-score normalization is employed to standardize them, ensuring that each feature contributes equally in terms of scale before being input into the model.

To establish the logistic regression model, assuming that wear is the binary classification target, with "mild wear" = 0 and "severe wear" = 1. The logistic regression model can be expressed as:

$$P(y=1|\mathbf{X}) = \frac{1}{1+e^{-(\vec{\beta}\mathbf{x})}}$$

Where $\vec{\beta} = [\beta_1, \beta_2, ..., \beta_n]$ is the regression coefficient vector. $\mathbf{X} = [x_1, x_2, ..., x_n]$ is the input feature. $P(y=1|\mathbf{X})$ is the probability of heavy wear of the target variable.

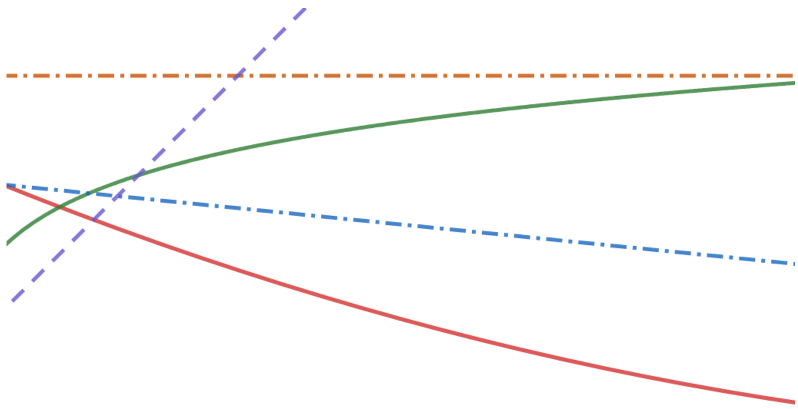


Figure 18: logistics for 5 different locations

The learning process of the regression coefficient $\vec{\beta}$ is accomplished by maximizing the likelihood function to find the optimal value of $\vec{\beta}$ that makes the model the most accurate on the known data.

5.3 Task 6: Forecast of Repair or Renovation

Typically, wear data, in the absence of external interventions such as repairs or renovations, follow a statistical pattern. A common assumption is that these data adhere to a normal distribution.

Assuming that the wear (denoted as "wear") on a staircase without repairs follows a normal distribution, we adjust the mean and variance based on different environmental factors such as climate, material type, etc.

For the wear value on each step, we assume it follows a normal distribution:

$$W \sim \mathcal{N}(\mu, \sigma^2)$$

Where μ is the mean value of wear and σ is the standard deviation of wear. μ and σ are a function of climate, geomorphology and material.

The parameters of these distributions are estimated using the Maximum Likelihood Estimation (MLE) method, which is then used to predict the normal wear range for each step.

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\omega_i - \mu)^2}{2\sigma^2}\right)$$

The optimal values of μ and σ are obtained through the estimation process.

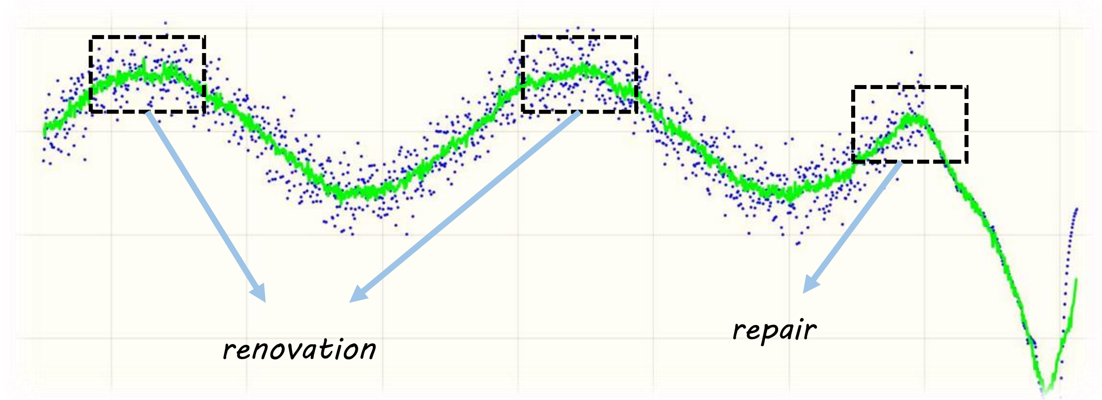


Figure 19: Repair and Renovation analysis

Anomaly Detection

We can use statistical tests to detect anomalous wear values, which may indicate whether a step has been repaired or renovated. The most common method is to use the Z-score (standardized deviation), which is defined as:

$$Z = \frac{W_i - \mu}{\sigma}$$

Where W_i is the wear value of the i-th step. μ is the mean wear for that step's location. σ is the standard deviation of the wear.

5.4 Task 7: Source of Materials

To use the input features for K-means clustering analysis, it is essential to normalize the data before proceeding with the clustering. Similar to the previous steps, the data is standardized to ensure that each feature contributes equally to the clustering process.

The initialization of the cluster centers involves randomly selecting K data points from the dataset to serve as the initial centroids. These initial cluster centers are then iteratively refined during the K-means algorithm to minimize the within-cluster variance.

For each data point X_i, calculate its Euclidean distance d to each cluster center C_i:

$$d(X_i, C_i) = \sqrt{\sum_{j=1}^{n}(x_{ij} - c_{ij})^2}$$

where X_i is the i-th data point, C_k is the center of the k cluster, and X_{ij} is the j-th feature of the i-th data point.

Assign the data point X_i to the nearest cluster center, in other words, X_i belongs to cluster k

IFF:

$$k = \arg \min_{1 \leqslant k \leqslant K} d(X_i, C_k)$$

Update cluster centers: for each cluster k, calculate the mean of all points within that cluster as the new cluster center:

$$C_k = \frac{1}{|S_k|} \sum_{X_i \in S_k} X_i$$

where S_k is the set of data points in cluster k and $|S_k|$ is the number of data points in cluster k.

Repeat steps 2 and 3 until the cluster centers no longer change or the change becomes minimal.

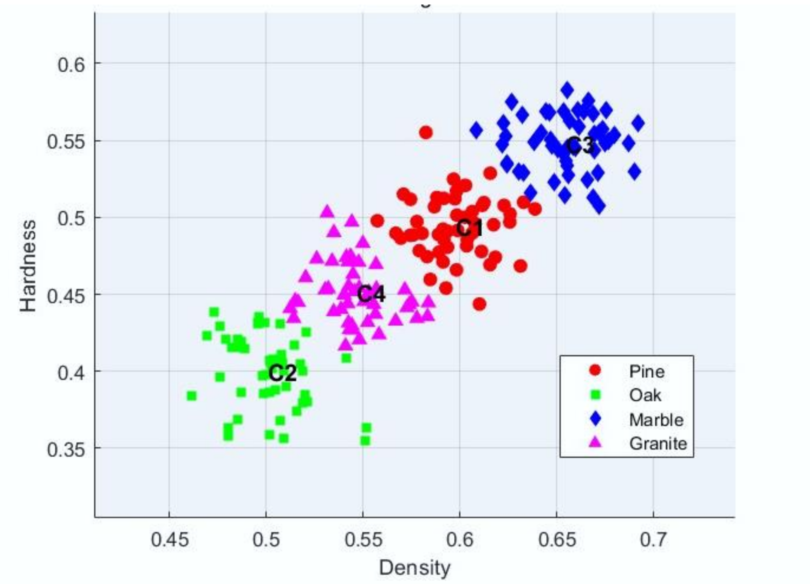


Figure 20: K-means Clustering of Four Materials

Choosing the optimal K is a crucial step in K-means clustering. A commonly used method is the Elbow Method. This involves calculating the Sum of Squared Errors (SSE) for different values of K and plotting the relationship between K and SSE. The optimal value of K is chosen at the point where the SSE curve exhibits an "elbow" indicating a diminishing return in error reduction as K increases. This K corresponds to the best balance between cluster compactness and the number of clusters.

SSE：

$$SSE(K) = \sum_{k=1}^{K} \sum_{X_i \in S_k} d(X_i, C_k)^2$$

After completing the K-means clustering, you will obtain the cluster labels for each step, with each cluster center representing the characteristics of a particular type of material.

By analyzing these cluster centers, you can infer the potential source of the materials. Additionally, you can compare the wear features of the steps with known material properties, such as hardness and wear coefficients for stone or wood, to assess whether the wear aligns with the expected material type. Finally, you can validate the clustering results by comparing them with established archaeological data to confirm the reasonableness and accuracy of the inferred material sources and wear patterns.

5.5 Task 8: Wear Frequency

The wear level is typically related to the frequency of use of the stairs, meaning that the higher the usage frequency, the more severe the wear. Therefore, we can assume that the wear data is proportionally related to the number of users of the stairs. Let's assume that during a given period, the wear level of the stairs is represented by:

$$W \sim I \cdot t \qquad\qquad I = kN$$

$$W = kN \cdot t$$

where W denotes the degree of wear at step location (x,y) at t. I is the intensity of use of the point at t, usually proportional to the number of people using the point. k is a constant of proportionality indicating the contribution of each person to wear and tear

Kernel Density Estimation (KDE) to estimate the number of people: Assuming that the distribution of the position of the steps (x,y) and the degree of wear W is smooth：

$$I(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)$$

where: x_i is the position of the i-th step. N is the total number of data points. h is the bandwidth parameter, which controls the degree of smoothing of the kernel function.

K is the kernel function, and commonly used kernels include Gaussian kernel and so on.

$$N(t) = \int_0^T \int_A I(x)\, dx dt$$

T is a period; A is the surface area of the step.

Location 1 is a densely populated and bustling area, suitable for commercial, transportation, or important facilities. It is located at a famous historical building in the city center. Location 2 is a relatively quiet area, likely suitable for residential or low-density commercial use. It is situated at a lesser-known historical building in a more secluded location.

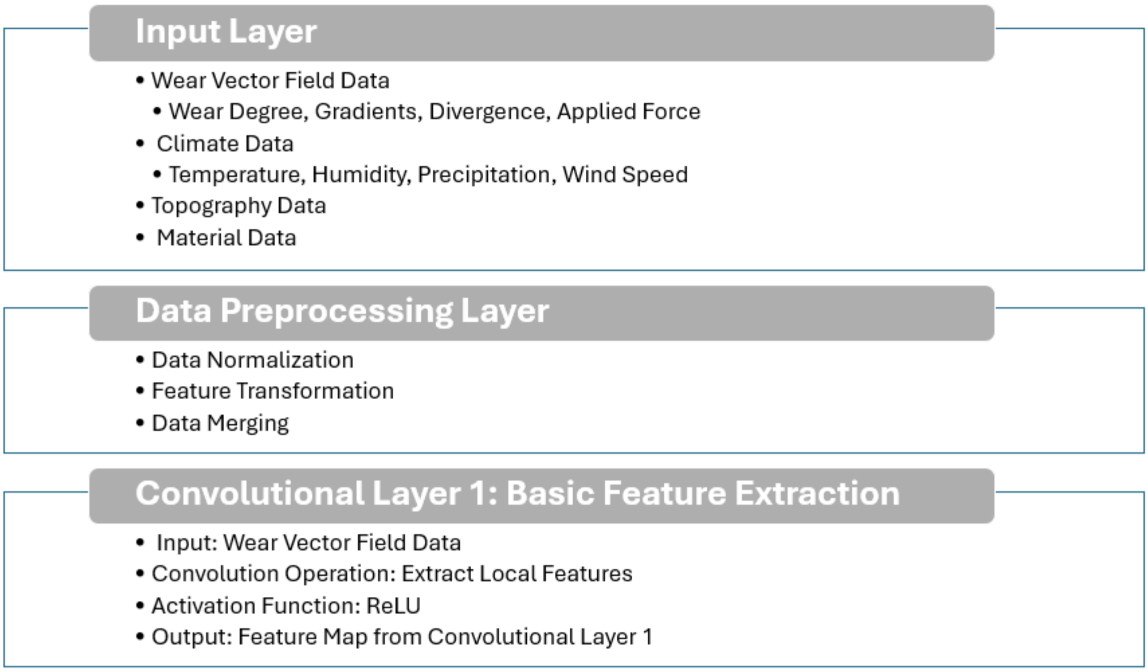
Table 3: Intensity of use

Time	Location 1 Intensity of use	Location 2 Intensity of use
6 a.m.~9 a.m.	965	99
9 a.m.~12 a.m.	759	176
12 a.m.~3 p.m.	574	113
3 p.m.~6 p.m.	1028	312
6 p.m.~9 p.m.	453	241
9 p.m.~12 p.m.	307	58
12 p.m.~3 a.m.	9	6
3 a.m.~6 a.m.	1	0

5.6 Again: Convolutional Neural Networks (CNN)

Due to the fragmented nature of the methods used previously, tasks 4 to 8 essentially control one variable at a time to predict other variables. In this context, we integrate these tasks using Convolutional Neural Networks (CNNs). CNNs are particularly effective at extracting wear patterns from different steps and learning the influence of external factors, such as climate and topography, on wear. Additionally,

CNNs can handle multiple types of input data and enhance the model's generalization capabilities. This is because CNNs feature local connections and weight sharing, which not only maintain low computational complexity when processing large datasets but also effectively capture complex spatial and nonlinear relationships. Therefore, convolutional neural networks provide a powerful tool for staircase wear analysis, enabling a deeper investigation into the causes and patterns of wear.



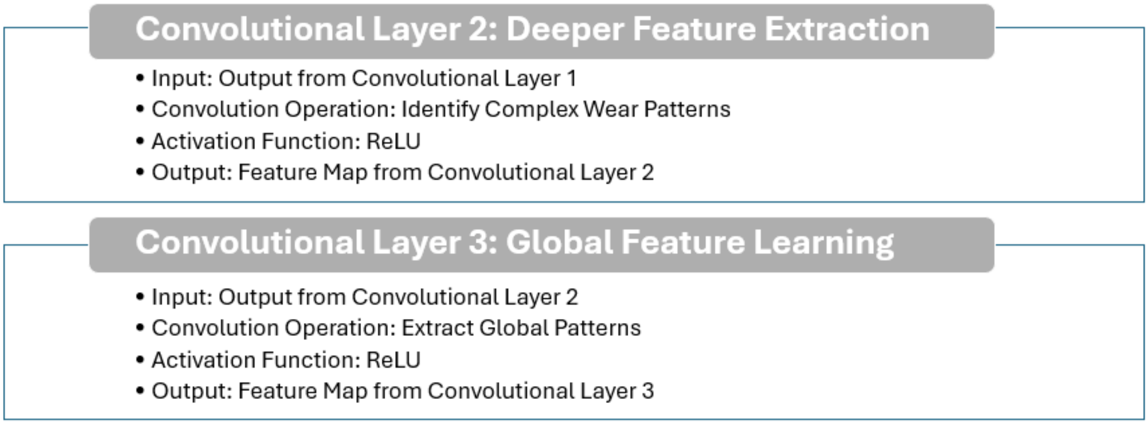


Figure 21: CNN Flowchart

5.7 Regularization Bias

Introducing regularization in Convolutional Neural Networks (CNN) is aimed at preventing overfitting and improving the model's generalization ability, particularly when data is limited, or the model is complex. Regularization techniques add a penalty term to the loss function, which helps constrain the model's complexity and prevent it from fitting too closely to the noise in the training data. This allows the model to better generalize to unseen data, leading to more reliable and robust predictions.

Combine L2 regularization and Dropout regularization, the loss function for the CNN model is:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \lambda \sum W_i^2 + \frac{1}{2} \sum_j \text{Dropout}\left(\text{Layer}_j\right)$$

Where: $\mathcal{L}$ is the original loss function. the first term is the L2 regularization term. The second term represents the Dropout regularization.

VI Model Evaluation

6.1 Sensitivity Analysis and Error Analysis

The blue dots represent the actual data. From the graph, it is clear that using traditional methods, although the accuracy for some individual points is quite high, the overall accuracy is moderate, and the variance is large.

For the CNN without regularization, the accuracy improves significantly, but it remains sensitive to the data, exhibiting some volatility.

After applying regularization to the CNN, although the accuracy for a few individual points slightly decreases, the overall accuracy improves, and the variance becomes smaller. This indicates a more stable and reliable model, with better generalization performance.

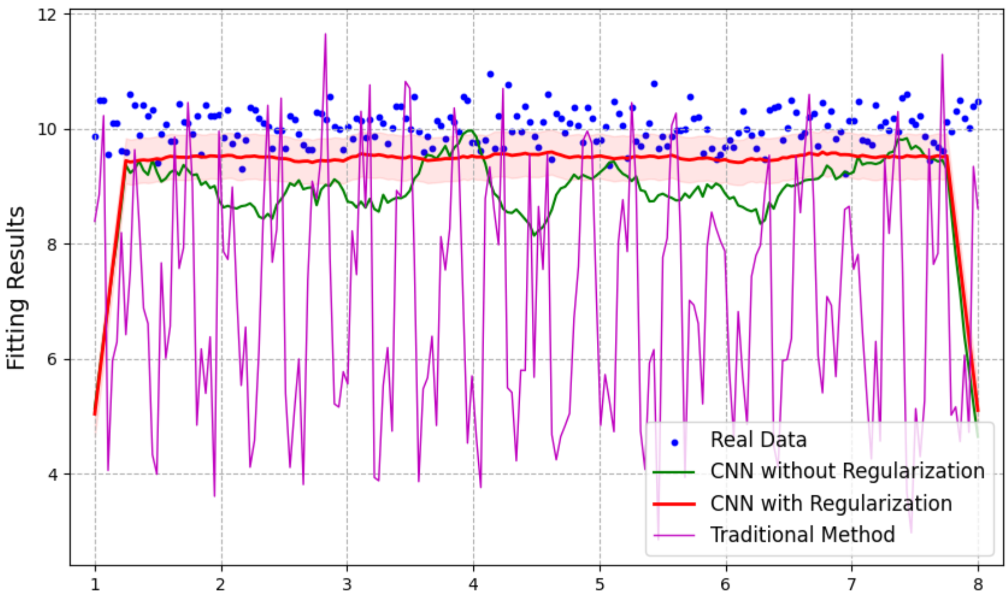


Figure 22: Comparison of Different Modeling Fitting Effects

6.2 Strengths

- **Authentic Data:** The model for wear prediction is based on authentic data sourced from the publicly available heritage site database OpenHeritage3D. The LiDAR point cloud data was captured using the high precision Leica SS1 scanner, employing Time of Flight (ToF) and Terrestrial LiDAR methods.
- **Innovative Use of Field Theory:** We creatively introduced Field Theory to model wear. The wear field, influenced by multiple factors, was vectorized at a high resolution. This provides a novel and intuitive predictive method for studying user behavior patterns, especially in terms of preferred directions of movement and trends under repetitive actions, offering new insights into how users interact with architectural structures.
- **Model Integration:** The model integrates multiple linear regression with convolutional neural network (CNN) prediction methods. This hybrid approach combines the fitting accuracy limits of linear models with the stability and predictive power of CNNs, ensuring that the final model closely aligns with the real-world data. It provides a comprehensive analysis of wear patterns for the staircase under study.

6.3 Weaknesses and Possible Improvement

- **Limitations of Traditional Methods:** Traditional approaches do not account for the interdependencies between different questions, leading to isolated answers. In contrast, Convolutional Neural Networks (CNNs) are black-box models that require large datasets for training, but due to the limitations of scanning techniques, it is difficult to gather sufficient data for training in the first two cases.
- **Fatigue Analysis:** The current model lacks an in-depth analysis of the mechanical properties of materials. Integrating material mechanics, specifically focusing on fatigue accumulation over time due to repetitive or cyclic loading, would enhance the model. By applying theories such as Miner's Rule for fatigue effects and using the linear damage

accumulation theory alongside the Norton-Bailey creep behavior equation, we can more accurately predict how materials deform and fail under prolonged stress, improving the wear prediction and material performance modeling.

VII Conclusion

- **Regular inspections:** Regular inspections of the staircase should focus on wear depth and slope changes, with timely repairs or renovations to extend its lifespan. For stairs made of different materials, proper cleaning and maintenance should be carried out to minimize the impact of climate and topography on wear.
- **Implement crowd control plans:** To optimize visitor flow management, crowd control plans can be implemented during peak periods to reduce the number of people using the stairs simultaneously, thus minimizing wear. Additionally, guiding systems, such as directional signs or warning markers at the entrance, can help balance the flow of visitors going up and down, reducing pressure on specific steps.
- **Choose durable materials**. When constructing or renovating stairs, it is recommended to choose more durable materials, such as reinforced concrete or specially treated wood, to increase the staircase's usage frequency and extend its lifespan. Additionally, during material selection and stair design, the local climate conditions should be considered to ensure that the materials are well-suited to the environment.

References

[1] M. Gurova and C. Bonsall, “Use-wear analyses and provenance determination of pitchstone artefacts: a pilot study from western Scotland,” Journal of Archaeological Science: Reports, vol. 30, p. 102189, 2020. [Online]. Available: https://doi.org/10.1016/j.jasrep.2019.102189

[2] W. J. Stemp, M. Mikhail, and A. J. M. K. Alastair, “Quantifying lithic microwear with load variation on experimental basalt flakes using LSCM and area-scale fractal complexity (Asfc),” Surface Topography: Metrology and Properties, vol. 3, no. 3, p. 034006, 2015.

[3] J. Pijoan-López, J. A. Barceló-Álvarez, I. Clemente-Conte, and A. Vila-Mitjà, “Variabilidad estadística en imágenes digitalizadas de rastros de uso: resultados preliminares,” in Análisis Funcional: su aplicación al estudio de sociedades prehistóricas, I. Clemente, R. Risch, and J. F. Gibaja, Eds. Oxford: BAR Int. Series, 1073, Archeopress, pp. 55–64, 2002.

[4] W. J. Stemp and M. Stemp, “UBM laser profilometry and lithic use-wear analysis: a variable length scale investigation of surface topography,” Journal of Archaeological Sciences, vol. 28, pp. 81–88, 2001.

[5] W. J. Stemp, H. J. Lerner, and E. H. Kristant, “Quantifying microwear on experimental Mistassini quartzite scrapers: preliminary results of exploratory research using LSCM and scale-sensitive fractal analysis,” Scanning, vol. 35, pp. 28–39, 2013, pmid:22688593.

[6] W. J. Stemp, “A review of quantification of lithic use-wear using laser profilometry: a method based on metrology and fractal analysis,” Journal of Archaeological Science, vol. 48, pp. 15–25, 2014.

[7] J. J. Werner, “An experimental investigation of the effects of post-depositional damage on current quantitative use-wear methods,” Journal of Archaeological Science: Reports, vol. 17, pp. 597–604, 2018.

[8] H. Lerner, X. Du, A. Costopoulos, and M. Ostoja-Starzewski, “Lithic raw material physical properties and use-wear accrual,” Journal of Archaeological Science, vol. 34, no. 5, pp. 711–722, 2007.

[9] Wikipedia https://en.wikipedia.org/wiki

[10] Open Heritage 3D https://openheritage3d.org/

Report on Use of AI

1. OpenAI ChatGPT (Dec. 2024 version, ChatGPT-4o)

Query1: Please translate these paragraphs into English as follows.

Output: Sure! Please provide the paragraphs you'd like me to translate into English.

Query2: Organize the logically confused language into logically coherent language.

Output: Could you share the text you'd like me to organize? I can help clarify it and make it more logically coherent.

Query3: Analyze the feasibility of using neural convolutional networks in this modeling.

Output: Using neural convolutional networks (CNNs) in this modeling could help identify wear patterns on stairs by analyzing high-resolution images of the steps, which could then be linked to specific usage patterns. These models might also aid in determining the directionality of wear, traffic frequency, and simultaneous usage, offering insights into the number of people and their movement patterns over time.