Evaluating Matrix Factorization Techniques for Thematic Mapping of Wilderness Walkability Using Multiple GPX Datasets

Ljiljana Šerić[1], Boris Draško[2], Antonia Ivanda[1]

[1] University of Split Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia - (ljiljana.seric, antonia.senta.00)@fesb.hr
[2] University of Mostar, Mostar, Bosnia and Herzegovina - borisdrasko@gmial.com

Keywords: Matrix Factorization, Wilderness Walkability, Sparse Data,GPX Trail Data, Thematic Mapping.

Abstract

Quantitative thematic mapping of walkability in wilderness areas is challenging due to sparse and unreliable data. Unlike urban walkability, which depends on built infrastructure, wilderness walkability is influenced by natural terrain features such as slope, surface stability, and vegetation density. This study leverages 1,620 GPX trail datasets from Croatia to infer walkability by analyzing movement speed across spatial cells. To extract latent walkability patterns, we apply matrix factorization techniques, including Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), Stochastic Gradient Descent (SGD), Alternating Least Squares (ALS), and Fast Independent Component Analysis (FastICA). Evaluation results indicate that NMF and Truncated SVD yield the most accurate and interpretable walkability maps. These findings highlight the potential of matrix factorization for mapping hidden variables in geospatial studies and suggest applications in related fields such as fire risk assessment.

1. Introduction

Quantitative thematic mapping is widely used in meteorology (e.g., weather maps), geology (e.g., topographic maps), and environmental science (e.g., pollution distribution). However, mapping continuous quantitative data is challenging, especially when data is sparse or unreliable. Sparse data arises from uneven measurement distribution, while unreliable data stems from inconsistencies or errors, such as subjective self-reports or satellite-derived estimates. These challenges intensify when mapping hidden variables requiring indirect proxies (Ewing and Handy, 2009), introducing further uncertainty. Advanced techniques for integrating multiple datasets and statistical methods like interpolation or machine learning are essential for improving accuracy. Mapping walkability and fire risk in wilderness areas is particularly difficult due to the hidden nature of these variables and data limitations. Wilderness walkability depends on factors such as trail connectivity, slope, surface quality, and accessibility, which are difficult to measure comprehensively. Desktop analyses often miss real-world obstacles like debris or vegetation overgrowth. Similarly, fire risk is influenced by vegetation dryness, wind patterns, topography, and human activity—complex interactions that are hard to quantify due to sparse sensor coverage and environmental variability. Both require indirect proxies, such as fire behavior data or trail condition audits, which introduce inaccuracies. Addressing these challenges demands advanced mapping techniques that integrate multiple data sources, including crowdsourced trail reviews, IoT sensors, and remote sensing data. This study focuses on thematic mapping of walkability. Unlike urban walkability(Horak et al., 2022), which is linked to built infrastructure, wilderness walkability depends on natural terrain features such as slope, surface stability, vegetation density, and trail connectivity. Measuring these factors directly is difficult due to terrain heterogeneity and dynamics. For instance, steep inclines and loose surfaces can impede movement, while dense undergrowth or debris can block trails entirely. Walkability can be assessed using GPX trail data by calculating walking speed along trails, providing an objective measure of terrain difficulty. GPX files contain time-stamped geographic coordinates that allow speed calculations based on distance and time. However, individual differences in fitness, experience, and preferences introduce subjectivity when expressing walkability as walking speed. One hiker may struggle on rocky trails, while another navigates them with ease. Aggregating data from multiple users helps mitigate these biases, capturing broader patterns and providing a more accurate walkability representation. Walking speed alone is insufficient for defining inherent walkability, so we propose matrix factorization as a technique for revealing latent walkability values. Using multiple GPX trails, we evaluate different matrix factorization methods for thematic mapping.

2. Materials and methods

This section will described the methodology of the research starting from data collection and preprocessing. Processed data forms a dataset used in evaluation of matrix factorization techniques. Evaluation framework is described in the methods section.

2.1 Data

Main data source is a collection of GPX trails. The trails are obtained from personal contacts and friends. We collected 1,620 GPX trails from users across Croatia, including mountain rescue teams, hikers, runners, dog walkers, and casual users. To ensure anonymity, each GPX file was assigned a unique user ID without any personal information. However, since no personal information was collected, we did not take into account possibility that more than one trail was recorded by the same user. Each trail contains a list of geographic coordinates and timestamps recorded during a continuous session. After parsing the files, variations of precision of data records were noticed that were attributed to variations in recording instruments. More precise instruments record position of the user more frequently and accurately. However, all instrumets' recordings were taken into consideration in this study. Movement speed was calculated by comparing time and location of neighboring points expressed in Croatian Terrestrial Reference System ('epsg:3765'). This reference system perserves the distances and rectangular grid of coordinates thus ensuring homogeneity of the data. After filtering out outliers, we obtained 1,795,663 valid segments, each described by location, time, user ID, and speed. Figure 1 shows the location of study area and plot of all trails in red.

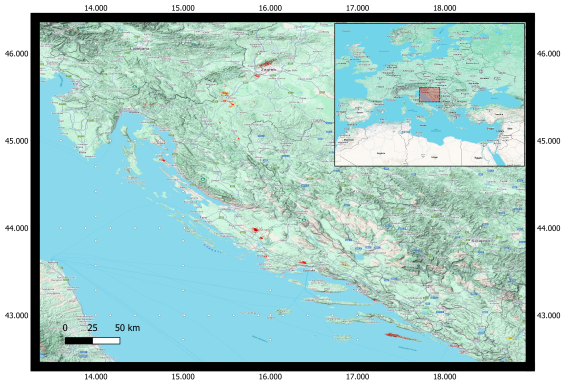

Figure 1. Study area.

To address previously mentioned inconsistencies, segments were grouped into 100-meter spatial cells per user. This was trivially performed by adding one more descriptor of the cell calculated by rounding the coordinates of two border points. Each descriptor represent a cell in the 100m grid of the study area. Direction of movement was ignored. Finally, for each cell and user the **median** movement speed per user-cell combination was computed, resulting in 127,478 user-cell speed descriptions. Median is statistical value representing a measure of central tendency thus is more appropriate in this problem than mean value that is more sensible from outliers and noise data. The final dataset was structured as a 1,609 × 24,349 sparse matrix, where rows represent users and columns represent terrain cells, with values indicating median walking speed of the user. Median walking speed is utilized as a users objective rating of difficulty to walk the cell of the terrain.

2.2 Methods

Matrix factorization is a process of braking down a matrix into simpler matrix, or factors. The aim is to find factors whose product make the original matrix. It's widely used in fields like linear algebra, computer science, and especially machine learning and recommendation systems(Koren et al., 2009). In this study, we utilize matrix factorization techniques to decompose a user-cell median speed of walking into two matrices - user latent representation and cell latent representations. The process of factorization is shown in figure 2.

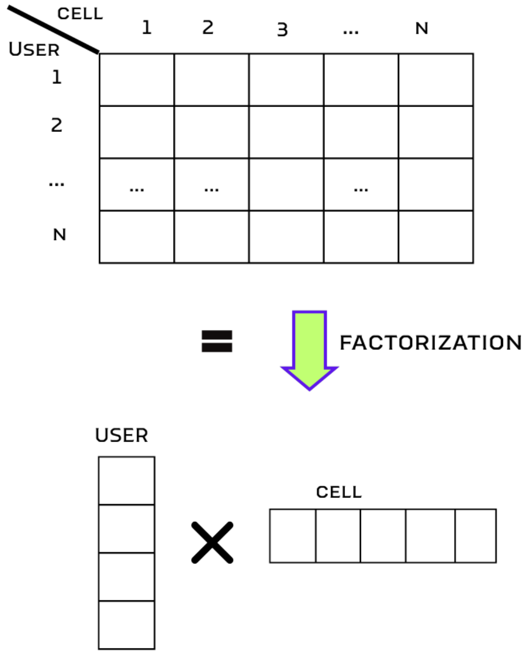

Figure 2. Factorization of a matrix into user and cell latent values.

When factorizing user-item rating matrices, various techniques uncover latent features and improve predictions ((Khalitov et al., 2021) (Du et al., 2023)). In this study we focused on six techniques found in literature. While there are more techniques proposed, we focus on these due to their popularity and ease of implementation.

NMF (Non-negative Matrix Factorization) (Lee and Seung, 1999) This method has been applied because it has the ability to decompose data into non-negative factors, which is especially applicable in analyzing data such as movement speed, distance and time that naturaly cannot be negative in its essence. Unlike non-linear methods NMF ensures that results remain interpretable. In context of our research, NMF enables identification of latent patterns between users and locations, giving results that can be directly interpreted as components that contribute to certain movement patterns. As NMF minimizes reconstruction error while preserving interpretability, it is very suitable for reconstruction of thematic maps that visualize spatial variations of speed in intuitive and explainable way. For example, certain latent component can match specific type of movement like faster movement on flat surface while other can match slow movement in hilly terrain. All this makes NMF useful for space classification and mapping, beacause it produces components that semantically make sense and are directly related to physical or behavioral characteristics of users and locations. Also, NMF not only helps in reconstructing original matrix, but it contributes to discovering latent structures that can be used for clustering, visualization and predictive analysis in spatial-temporal data.

SVD (Singular Value Decomposition) (Golub and Reinsch, 1970). SVD is used as a basic method for reducing dimensionality and identifying global structures in data. Despite the fact that it does not restrict factors non-negative values, SVD is a strong tool because of its mathematical robustness and its ability to find orthogonal latent dimensions that capture greatest variance within a matrix. In our research, this method enables the compression of complex spatio-temporal patterns into low-dimensional latent space that can be used for efficient reconstruction and classification of movement segments. It is also suitable for comparison with other algorithms as it ensures optimal matrix approximation using matrix norms.

Truncated SVD (Falini, 2022). Truncated SVD was used for dimensionality reduction and compression of user-terrain data into fundamental latent components. Unlike the complete SVD decomposition, this sub-method keeps only the first few most significant singular values (only one in our case), thus achieving simpler and faster processing of a large and sparse matrix. Although factors can have negative values, Truncated SVD enables efficient matrix reconstruction and clearly separates dominant data patterns. In this research, the method serves as a reference for comparing to other techniques.

SGD (Stochastic Gradient Descent) (Abdelbar et al., 2018) . SGD has been applied because of its flexibility and control over learning of latent factors of users and locations. Unlike SVD, SGD enables iterative optimization of user and location factors through directly minimizing error function. It minimizes MSE between known and prediction values, which is mathematical equivalent of Frobenius norm minimization. As the factors are updated stochastically in small steps, model can learn behavioral patterns adaptively, including more complex local variations that might not be detected by deterministic methods. This makes SGD suitable for processing highly variable data, such as different movement conditions in the field. In context of our research, where we analyze movement patterns between points in space, SGD enables us to train the model even when data is sparse and irregular, which is common in real-world GPS trajectory segments. It also enables fine optimization of parameters such as learning rate, regularization, epoch number. This is useful for experimenting with user behaviour in different geographical contexts.

ALS (Alternating Least Squares) (Hastie et al., 2014). ALS is used because of its robustness, numerical stability and efficiency in processing large and sparse matrix typical for data derived from GPS segments. Unlike stochastic methods, ALS uses deterministic approach in which user and location latent factors are alternatively optimized by solving systems of linear equations, keeping one set fixed while updating the other. This procedure enables convergence toward a locally optimal solution even when the data is unevenly distributed. In this research, ALS is useful because it can extract latent movement patterns without the need for gradient methods, which results in faster and more stable optimization. It also reduces overfitting risk and improves model generalization ability. Through minimizing the squared error between observed and predicted values, ALS enables the construction of a low-dimensional latent space that accurately captures the dynamics of movement in real-world spatial environments.

Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000). ICA has been used in order to explore the possibility of discovering statistically independent latent factors that contribute to movement patterns, that might have not been detected using linear and ortoghonal methods such as SVD or NMF. ICA is based on the assumption that different movement patterns emerge from mutually independent factors such as terrain type, user behavioral habits, or environmental conditions. In this research we used FastICA algorithm, which is numerically optimized and efficient ICA implementation method from scikit-learn library. It enables fast decomposition of user-location matrix through maximizing statistical independence of latent components, that enables pattern discovery not necessarily related to dominant variations. This approach adds to recognizing hidden behaviour dimensions that further explain spatial dynamics of users and improves results interpretability.

2.3 Evaluation strategy

The evaluation was conducted by performing an experimental matrix factorization and comparing the results with the original values, with each other, and with geospatial features of the corresponding terrain. All factorization techniques were benchmarked using the same dataset.

Using Python, scikit-learn, and custom implementations, we factorized a user–cell matrix populated with median movement speeds into two factor matrices. The source code for the experiment is available on **GitHub** (Seric, 2025).

Each method was used to extract **a single latent factor** for every user and every cell. As a result, the original user–item walking performance matrix was decomposed into two matrices: (1) a user–latent-factor matrix and (2) a cell–latent-factor matrix, as illustrated in Equation 1.

$$S_{m\times n} \approx U_{m\times r} \times C_{r\times n} \tag{1}$$

S is a sparse matrix representing the median walking speed of users in individual cells. U denotes the latent factor representation of users, and C the latent factor representation of cells. Here, m is the number of distinct users in the dataset, and n is the number of spatial cells. The variable r represents the number of latent factors considered; in our case, we set $r = 1$.

The extracted latent factors were expected to preserve most of the variability present in the original data, allowing for its reconstruction as the product of the two latent factor matrices, as shown in Equation 2.

$$S_r = U_{m\times 1} \times C_{1\times n} \tag{2}$$

The ability of each factorization technique to extract meaningful latent features was evaluated using two complementary strategies. First, performance was assessed by calculating the Root Mean Squared Error (RMSE) between the reconstructed values and the original observed values. The reconstructed values were obtained as the product of the user latent factor matrix U and the cell latent factor matrix C:

$$RMSE = \sqrt{\frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \left(S_{ij} - (UC)_{ij}\right)^2} \tag{3}$$

Here, Ω denotes the set of index pairs (i, j) for which the original value S_{ij} is observed (i.e., not missing), $|\Omega|$ is the number of such observed entries, and $(UC)_{ij}$ represents the reconstructed value at position (i, j).

A lower RMSE indicates that a greater proportion of the data's variability is captured by the first latent factor, suggesting that the factorization technique more effectively extracts underlying latent features. However, it is important to note that a low RMSE does not necessarily imply that the extracted factor represents walkability.

The latent factor associated with each cell is interpreted as its inferred walkability. By assigning these inferred values to their corresponding spatial locations, we generate *walkability maps*. These maps are inherently sparse, containing values only at locations where observations are available. Nevertheless, they serve as a foundation for statistical analysis and the construction of a walkability model.

Walkability maps produced by different factorization techniques were compared against satellite imagery, topographic data, and land cover classifications using statistical tools available in GRASS GIS (GRASS Development Team, 2024).

Covariance in raster data is a statistical measure that quantifies how the values of two raster layers (gridded spatial datasets) vary together across corresponding pixels.

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{4}$$

The covariance between pairs of raster layers was calculated as shown in Equation 4, quantifying the relationship between walkability inferred by different techniques and terrain features such as height, slope, and land cover.These covariance values provide insight into how well each technique captures terrain-dependent variations in walkability.

3. Results

The factorization algorithms described in Section 2.2 were implemented in Python using the scikit-learn library along with custom code. For each technique, a single latent value was extracted per user and per cell, resulting in the latent matrices U and C as defined in Equation 1. The original values were reconstructed using Equation 2, and RMSE values were computed as described in Equation 3. The RMSE results summarized in Table 1 highlight substantial differences in reconstruction performance among the evaluated matrix factorization methods. Lower RMSE values indicate more accurate reconstruction of the original data from the latent factor matrices.

Among all methods, SGD (Stochastic Gradient Descent) achieved the lowest RMSE of 2.250, indicating its superior ability to capture the underlying structure in the data using a single latent factor. This suggests that SGD is particularly well-suited for sparse, noise-prone geospatial datasets like the one used in this study, likely due to its flexibility and capacity for fine-tuned optimization.

ALS (Alternating Least Squares) also performed well, with an RMSE of 4.968, outperforming traditional decomposition techniques such as SVD (8.1788) and FastICA (8.1801). This demonstrates the advantage of optimization-based approaches over closed-form decompositions when modeling latent geospatial relationships, especially under constraints such as sparsity.

Interestingly, both NMF and TruncatedSVD achieved identical RMSE scores (7.0092), which are significantly higher than those of SGD and ALS but lower than standard SVD and FastICA. The similarity in performance between NMF and TruncatedSVD may reflect their shared reliance on linear structure, although NMF's non-negativity constraint typically alters the factorization outcome.

Overall, these results emphasize that while classical methods like SVD and ICA provide useful baseline performance, optimization-driven techniques such as SGD and ALS offer more accurate reconstructions in this context, particularly when extracting a single latent factor to represent walkability.

To further compare the errors of values reconstructed from obtained factors compared to observed values we plotted the error plot of each technique shown in figure 3. The error plots

Method	RMSE
NMF	7.0092
TruncatedSVD	7.0092
SVD	8.1788
FastICA	8.1801
ALS	4.968
SGD	2.250

Table 1. RMSE results for various matrix factorization methods on the dataset.

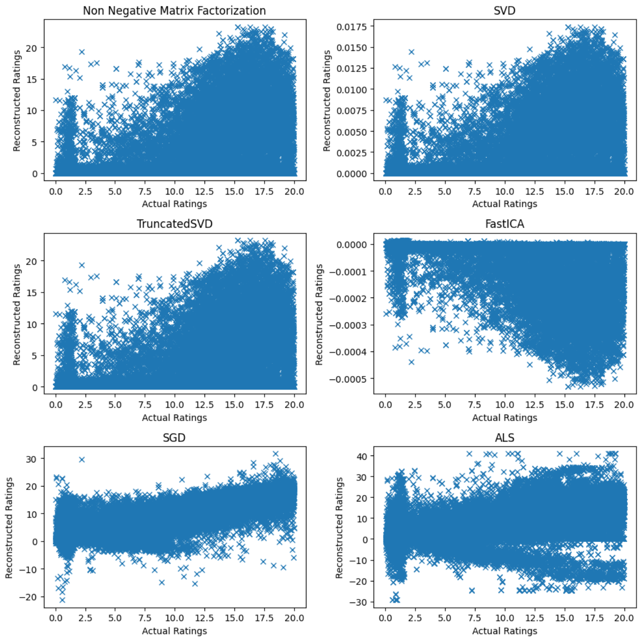

Figure 3. Error plot of comparing observed and reconstructed median speed of walking .

do not exhibit a consistent trend across the different factorization methods, indicating variability in their reconstruction performance. However, it is noteworthy that NMF, SVD, and Truncated SVD successfully maintain non-negative reconstructed values, which aligns better with the physical nature of the data—median walking speed cannot be negative. This property is particularly expected from NMF, due to its inherent non-negativity constraint, and somewhat preserved by the structure of truncated linear decompositions like SVD.

In contrast, FastICA produces reconstructed values that are entirely negative, which is likely a consequence of its internal centering and whitening operations. Since FastICA attempts to maximize statistical independence rather than minimize reconstruction error, and tends to subtract the mean during preprocessing, the resulting components may not align well with the original scale or sign of the data.

SGD and ALS, on the other hand, yield reconstructions containing both positive and negative values. This outcome reflects the unconstrained nature of their optimization, where minimizing the RMSE does not necessarily ensure the physical interpretability of the output values.

Overall, reconstructions obtained from factorization using only a single latent component do not closely match the observed mean walking speed. This is expected, as capturing the full variability and complexity of user–cell interactions likely requires a richer latent representation. Interestingly, the first four techniques—NMF, SVD, Truncated SVD, and FastICA—exhibit a similar error plot shape, suggesting a shared underlying structure in their factorization approach, despite differences in constraints or objectives.

Next, we investigate the walkability values inferred by the different factorization techniques. As an initial step, we conducted pairwise comparisons of the walkability scores produced by each method. Interestingly, despite the differences in underlying algorithms and assumptions, most techniques yield highly correlated walkability estimates, as evidenced by the strong linear trends observed in the pairwise scatter plots. This suggests that these methods, though methodologically distinct, are capturing a similar underlying spatial pattern.

However, walkability values derived using ALS deviate significantly from those produced by the other techniques. The divergence is apparent both in the magnitude and distribution of values, indicating that ALS may be capturing a different aspect of the data or is more sensitive to certain patterns in user–cell interactions. This inconsistency highlights the importance of carefully selecting factorization methods, especially when interpreting latent components in terms of real-world phenomena like walkability.

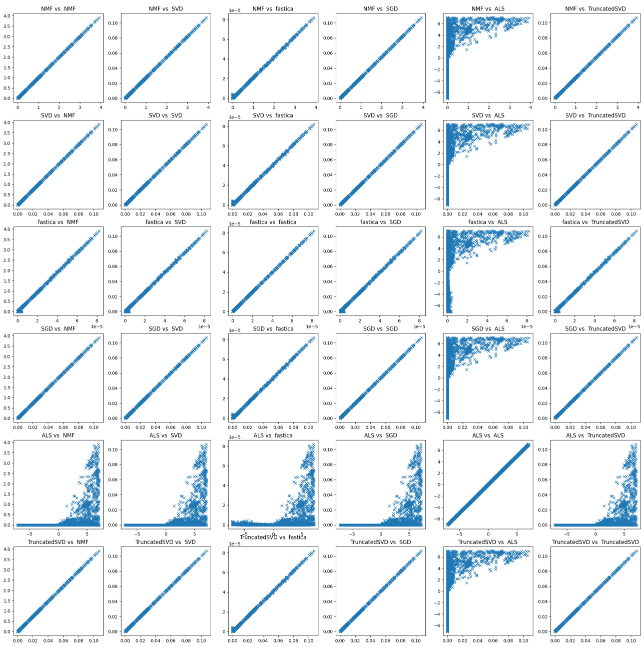

Figure 4. Pairwise plot of walkability values for pairs of techniques.

Finally, the walkability values inferred from matrix factorization were compared with geospatial features that are intuitively related to walkability. Specifically, we examined correlations between walkability and the following datasets:

- **Digital Elevation Model (DEM30)**: A 30-meter resolution elevation dataset obtained from the Copernicus Earth Observation Service (European Environment Agency, 2019).
- **Slope**: Derived using the `r.slope` module in GRASS GIS, based on the DEM30.
- **Land Cover**: CORINE Land Cover codes obtained from the Copernicus Land Monitoring Service (European Environment Agency, 2018).

Walkability maps were generated in GeoTIFF format, assigning the extracted walkability values to cells for which observations were available. All other cells were assigned a value of 0. These maps represent latent walkability factors derived from each technique.

To ensure consistency across datasets, all geospatial layers were converted to raster format and resampled to a common spatial resolution of 100 meters. Additionally, the geospatial datasets were masked to retain only the cells for which walkability values were available; all other cells were set to 0 to avoid introducing bias into the analysis.

The `r.covar` module of GRASS GIS was used to compute the covariance matrix between walkability and the geospatial features. The resulting covariance values are presented in Table 2.

	Height	Slope	Land Cover
Height	0.046590	0.036881	1.200949
Slope	0.036881	0.073646	1.489132
Land Cover	1.200949	1.489132	53.360328
NMF	0.000000	0.000001	0.000009
SVD	-0.000000	-0.000000	-0.000000
TruncSVD	0.000000	0.000000	0.000000
FastICA	0.000000	0.000000	0.000000
SGD	-0.002367	-0.001601	-0.054017
ALS	0.000248	0.001165	0.053605

Table 2. Covariance matrix of masked raster layers (truncated to geospatial features).

As expected, the geospatial layers themselves show moderate to high covariance, particularly between Slope and Land Cover (1.4891), and between Height and Land Cover (1.2009), suggesting that land cover types are closely associated with elevation gradients and slope values in the study area.

In contrast, the walkability values inferred by matrix factorization techniques show very low covariance with geospatial features. Most techniques, including NMF, SVD, Truncated SVD, and FastICA, exhibit near-zero covariance with all three geospatial layers. This indicates that the single latent component extracted by these methods does not align strongly with terrain elevation, slope, or land cover. Notably, all values are either exactly zero or negligible, reflecting minimal spatial correspondence.

Two techniques stand out slightly:

SGD shows negative covariance with all three geospatial features, especially with Land Cover (-0.0540), which may suggest that its inferred walkability factor is inversely related to land use intensity or urbanization.

ALS yields slightly positive covariance, particularly with Land Cover (0.0536), indicating some alignment with spatial variation in land cover types, although the relationship remains weak.

Overall, these findings suggest that the single latent component extracted by most factorization techniques captures patterns that are largely orthogonal to the physical terrain features. This could be due to the limited expressiveness of single-component models, or because the latent factors are more reflective of behavioral or infrastructural patterns not directly represented by elevation, slope, or land cover.

These results support the conclusion that while matrix factorization can extract meaningful latent structures from movement data, interpreting these structures as geospatial phenomena like "walkability" requires caution. More complex models (e.g., multi-component factorizations) or the inclusion of domain knowledge may be necessary to better align latent features with known geographic correlates.

4. Conclusion

Among the evaluated matrix factorization techniques, Stochastic Gradient Descent (SGD) achieved the lowest RMSE (2.25), indicating that it best reconstructed the original movement data. This suggests that SGD is most effective at capturing variability in the observed user–cell matrix using only a single latent factor. Its optimization-driven nature allows it to minimize reconstruction error efficiently, making it a strong choice for data-driven prediction tasks.

However, SGD showed negative covariance with geospatial features, suggesting that its latent factors do not align well with intuitive walkability determinants such as elevation, slope, or land cover.

Alternating Least Squares (ALS), while producing a higher RMSE (4.97), yielded the strongest positive covariance with geospatial variables—particularly with land cover (0.0536). This implies that the latent factors extracted by ALS are more geospatially meaningful, potentially reflecting real-world influences on walkability, even if they are less precise in reconstructing the original data.

Together, these results illustrate a trade-off: SGD excels in numerical accuracy, while ALS offers better alignment with environmental factors, making it more suitable when interpretability and spatial correlation are desired. This distinction can guide method selection depending on whether the goal is prediction accuracy or geospatial explainability.

Although the extracted latent factors do not exhibit strong covariance with core geospatial features such as elevation, slope, or land cover, they still reveal distinct spatial patterns of walkability. These patterns highlight areas where walking conditions are inferred to be more or less favorable based on movement behavior alone. As such, these latent factors offer a valuable, data-driven basis for further analysis, even if they do not align neatly with traditional geospatial indicators.

To enhance spatial generalization, future work could focus on extrapolating these latent factors across unsampled regions using auxiliary datasets such as land cover maps, topographic indices, and other environmental attributes. Integrating these latent walkability patterns with external data could also support the development of predictive models for assessing walkability in heterogeneous or remote landscapes. Beyond walkability, this methodology may be transferable to other domains—such as fire risk assessment—where latent structure in sparse observational data could inform susceptibility mapping and targeted interventions.

References

Abdelbar, A. M., Elnabarawy, I., Salama, K. M., Wunsch, D. C., 2018. Matrix factorization based collaborative filtering with resilient stochastic gradient descent. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–7.

Du, K.-L., Swamy, M., Wang, Z.-Q., Mow, W. H., 2023. Matrix factorization techniques in machine learning, signal processing, and statistics. *Mathematics*, 11(12), 2674.

European Environment Agency, 2018. Corine land cover (clc) 2018. https://land.copernicus.eu/en/products/corine-land-cover/clc2018 (15 April 2025). Copernicus Land Monitoring Service, European Environment Agency.

European Environment Agency, 2019. Copernicus digital elevation model (dem) - glo-30. https://browser.dataspace.copernicus.eu/ (15 April 2025). Accessed via the Copernicus Open Access Hub.

Ewing, R., Handy, S., 2009. Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design*, 14(1), 65–84.

Falini, A., 2022. A review on the selection criteria for the truncated SVD in Data Science applications. *Journal of Computational Mathematics and Data Science*, 5, 100064.

Golub, G. H., Reinsch, C., 1970. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5), 403–420. https://doi.org/10.1007/BF02163027.

GRASS Development Team, 2024. Geographic Resources Analysis Support System (GRASS GIS) Software, Version 8.4. Open Source Geospatial Foundation, USA.

Hastie, T., Mazumder, R., Lee, J., Zadeh, R., 2014. Matrix completion and low-rank svd via fast alternating least squares. arXiv preprint.

Horak, J., Kukuliac, P., Maresova, P., Orlikova, L., Kolodziej, O., 2022. Spatial Pattern of the Walkability Index, Walk Score and Walk Score Modification for Elderly. *ISPRS International Journal of Geo-Information*, 11(5).

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4), 411–430.

Khalitov, R., Yu, T., Cheng, L., Yang, Z., 2021. Sparse factorization of large square matrices. *arXiv preprint arXiv:2109.08184*.

Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37.

Lee, D. D., Seung, H. S., 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401, 788–791.

Seric, L., 2025. Python notebook for evaluation of factorization performance on walkability . https://github.com/ljiljana44/Walkability_Factorization .