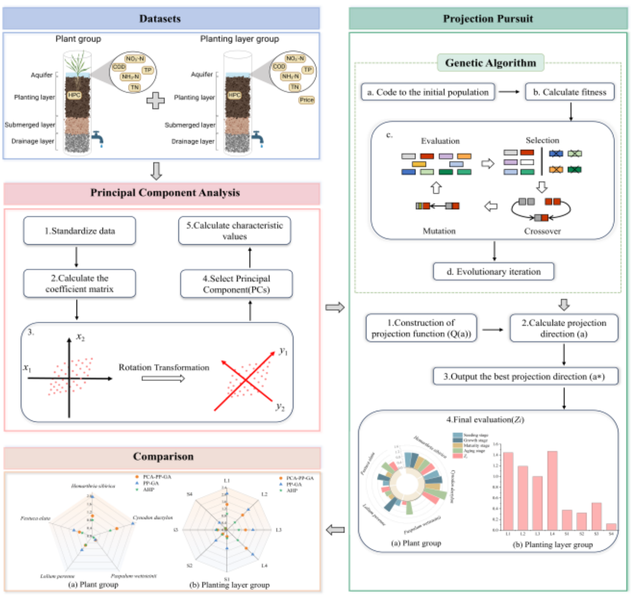
Global NEST Journal, Vol 27, No X, pp 1-8

Multi-objective evaluation of bioretention systems based on principal component analysis-projection pursuit model

Zhen Liu[1,2], Yiwei Fu[1,2], Weibo Qiang[3], Fangfang Ji[4], Qiming Cheng[1,2], Jingsong Liao[5], Sifei Wu[5], Bing Zhao[5], Sixu Hu[5] and Yao Chen[1,2*]
[1]School of River and Ocean Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[2]Key Laboratory of Hydraulic and Waterway Engineering of the Ministry of Education, Chongqing Jiaotong University, Chongqing 400074, China
[3]Central & Southern China Municipal Engineering Design and Research Institute Co., Ltd., Wuhan 430010, China
[4]T. Y. Lin International (China), Chongqing 401121, China
[5]Chongqing Shipping Construction Development (Group) Company, Ltd, Chongqing 400013, China
Received: 08/11/2024, Accepted: 27/02/2025, Available online: 13/03/2025
*to whom all correspondence should be addressed: e-mail: chenyao@cqjtu.edu.cn
https://doi.org/10.30955/gnj.06977

Graphical abstract



Abstract

Hydraulic infiltration and decontamination performance of bioretention systems are influenced by plant species and planting layers. However, traditional evaluation methods are limited by their high subjectivity and inability to capture complex relationships among multidimensional data accurately. This study developed a coupled PCA-PP-GA model, integrating principal component analysis (PCA) for dimensionality reduction, projection pursuit (PP) for comprehensive evaluation, and genetic algorithm (GA) for optimization. Through analyzing Pearson correlation coefficients and principal component loadings, which showed strong multicollinearity and significant weights, the rationality of employing PCA for dimensionality reduction was validated. In evaluating five plant species (*Cynodon dactylon*, *Hemarthria sibirica*, *Paspalum wettsteinii*, *Lolium perenne*, and *Festuca elata*) across growth stages and different planting layers for stormwater runoff control, results indicated that *Cynodon dactylon* exhibited the highest score of 1.22, and L4 planting layer composition (10.0% loamy sandy soil + 90.0% fine sand) scored 1.47. Furthermore, compared to the analytic hierarchy process (AHP) and traditional projection pursuit model (PP-GA), the data-driven PCA-PP-GA offers a more comprehensive consideration of both cost and pollutant removal efficiency, demonstrating advantages in reducing subjective bias and enhancing information screening efficiency. This study provides a reference for evaluating the effectiveness and implementation of ecological engineering in stormwater runoff control.

Keywords: bioretention systems, evaluation, genetic algorithm, principal component analysis, projection pursuit

1. Introduction

In recent years, ecological engineering (e.g., bioretention systems, slope engineering, and grass swales) has been pivotal in the prevention and control of urban flooding and water pollution (Muhammad *et al.*, 2024). Among these measures, bioretention systems stand out as a crucial strategy for enhancing water quality (Li *et al.*, 2021; Vijayaraghavan *et al.*, 2021). They could effectively reduce pollutants in stormwater runoff through multiple actions, including plant uptake, filler filtration, and microbial degradation, thereby acting as a protective barrier for river ecosystems (Mehmood *et al.*, 2021). As an essential component of bioretention systems, plants and planting layers can directly affect their pollutant removal efficiency (Xu *et al.*, 2019; De-yong *et al.*, 2020; Liu *et al.*, 2020). However, the variability in pollutant removal capacity and hydraulic performance among different plants and planting layers introduces a multidimensional and non-linear relationship between evaluation indicators, rendering common analytical methods inadequate for accurate evaluation (Liu *et al.*, 2018; Wang, Zhang & Li, 2019; Aziz *et al.*, 2021; Mehmood *et al.*, 2021). Consequently, how to scientifically and reasonably evaluate the control effect of plants and planting layers on stormwater runoff has become an important problem yet to be solved in ecological engineering.

Nowadays, there is no standardized approach, either domestically or internationally, for evaluating the effectiveness of ecological engineering in managing rainwater runoff (Muhammad *et al.*, 2022). Among them, the evaluation of bioretention systems typically adopts methods such as the analytic hierarchy process (AHP) and fuzzy comprehensive evaluation. Mei et al. integrated fuzzy set theory with an improved analytic hierarchy process (IAHP) to develop a fuzzy synthetic evaluation model to evaluate the comprehensive removal capability and infiltration rate of sand media in different bioretention systems. The findings indicated that the media with a sand content of 96% performed the best (Mei *et al.*, 2018). Xu et al. constructed an evaluation methodology for environmental pollution management systems, utilizing 5 primary indicators and 26 secondary indicators, combined with the Delphi method to collect experts' opinions and feature vector analysis by AHP, and finally determined the importance ranking of secondary indicators (Xu, Ling & Jin, 2017). Jia also used the same approach for evaluating ecological suitability (Jia & Zhessakov, 2021). The above methods provide significant flexibility in evaluation and analysis, making them adaptable to diverse evaluation scenarios. Nevertheless, since they rely on expert subjective judgment, the evaluation results of these methods are prone to subjectivity and ambiguity, potentially leading to large errors.

The ability of PP to reveal the intrinsic structure of data has garnered widespread attention within the domain of the ecological environment. PP mitigates the interference of empowerment with outcomes and ensures the objectivity of the results by searching for projections that maximize non-Gaussianity (Wentzell *et al.*, 2015). In evaluating the rationality of water resource utilization across 31 provinces, Wu applied the evaluation results from the PP model to yield projections that more closely align with the actual target value (Wu, 2019). Similarly, Yang et al. demonstrated in their evaluation of the groundwater resources and environmental carrying capacity that PP can more accurately reflect the actual situation of the data (Yang *et al.*, 2010). However, Zhang found that the PP model may encounter challenges such as noise interference, computational instability, and a large number of multi-objective evaluations during the iterative adjustment of projection directions (Zhang, Ye & Wang, 2023). Therefore, optimizing the computational process of PP projection could be key to improving the accuracy of the evaluation results.

Given that dimensionality reduction techniques can effectively reduce data noise and improve computational efficiency, researchers have explored the integration of PP with methods including independent component analysis (ICA) and PCA. ICA separates independent signal sources, providing a clearer projection direction for PP in cases of signal interference, thereby mitigating the impact of noise. However, Seonjoo Lee et al. pointed out in their study that ICA is sensitive to initial conditions and algorithm parameters, which may fluctuate with changes in data distribution, leading to different outcomes in multi-objective evaluation (Lee *et al.*, 2015). In contrast, PCA, as a classical dimensionality reduction technique, can provide consistent key features by compressing large datasets when dealing with the same dataset, thus optimizing the indicator system, reducing the number of multi-objective evaluations, and providing a more robust projection foundation for the PP model (Pereira *et al.*, 2017; Ferde *et al.*, 2021). When employing the combined PCA-PP model, a critical aspect is the determination of the optimal projection direction.

This process involves solving nonlinear problems, which generally cannot be tackled using straightforward mathematical methods. Owing to their efficiency in automatically optimizing complex systems, intelligent algorithms such as GA, particle swarm optimization (PSO), and simulated annealing (SA) have been extensively utilized for selecting the projection direction (Garg, 2016; Xu *et al.*, 2022). Among these, GA has demonstrated a superior capability in reliably identifying the globally optimal solution (Jafari *et al.*, 2020). The PCA-PP combined model has been preliminarily applied in the domains of cybersecurity and data mining, effectively improving the accuracy of evaluation and prediction (Wen & Chen, 2012), but its application within the ecological environment, especially in evaluating the effectiveness of bioretention system plants and planting layers for stormwater runoff management, remains unexplored.

To this end, this study proposed a PCA-PP model optimized through GA for the multi-objective evaluation of the effectiveness of stormwater runoff management in bioretention systems, focusing on the role of plants and planting layers. Initially, PCA was used to perform dimensionality reduction on the raw data, including hydraulic performance and pollutant removal rates. This process extracted several comprehensive indicators to reduce the computational complexity inherent in traditional models. Subsequently, GA was used to solve the constructed multi-objective comprehensive evaluation model (PCA-PP). The evaluation outcomes were then compared and analyzed against those obtained from AHP and PP-GA models to verify the reliability of the proposed method. This novel approach not only advances the theoretical foundation of multi-objective evaluation but also provides a scalable framework for optimizing ecological engineering solutions, offering a scientific foundation for the plants and planting layers selection.

2. Methods

2.1. Data sources

The data and planting layer ratios utilized in this study are derived from the previous experiments conducted by the research group. Detailed information on the specific groupings and evaluation indicators can be found in Table 1. The experiments in the plant group were designed to investigate the effects of plants at various growth stages (seedling, growth, maturity, and aging) on pollutant removal and soil hydraulic properties in stormwater runoff (Muhammad *et al.*, 2019). And the experiments in the planting layer group, which were conducted in the absence of plants, aimed to explore the impact of planting layer structural ratios on the efficiency of pollutant removal in stormwater runoff.

Table 1. Evaluation indicator system

Groups	Research subject		Evaluation indicators	Sample size
Plant group	*Cynodon dactylon*, *Hemarthria sibirica*, *Paspalum wettsteinii*, *Lolium perenne*, *Festuca elata*		TN, NH_4^+–N, TP, NO3⁻–N, COD, HPC	80
Planting layer group	L1	9.8%Ls+88.2%S+2.0% Ve	TN, NH_4^+–N, TP, NO3⁻–N, COD, HPC, Price	72
	L2	9.8% Ls +88.2%S+2.0% Bi		
	L3	9.8%Ls +88.2%S+2.0% Pe		
	L4	10.0% Ls +90.0% Fs		
	S1	18.4% Ss+73.6%S+8.0% Ve		
	S2	18.4% Ss +73.6%S+8.0% Bi		
	S3	18.4% Ss +73.6%S+8.0% Pe		
	S4	20.0% Ss +80.0% Fs		

Based on the identification of sandy soil (Ss) and loamy sand (Ls) as the predominant soil types in Chongqing's urban area, commonly used fine sand (Fs), vermiculite (Ve), biochar (Bi), and perlite (Pe) were selected as amendments. In accordance with soil particle size gradation guidelines for bioretention facilities, Ss and Ls were graded to obtain different planting layer ratio combinations.

The evaluation indicators include common stormwater runoff pollution indicators –total nitrogen (TN), nitrate nitrogen (NO_3^--N), ammonia nitrogen (NH_4^+-N), total phosphorus (TP), and chemical oxygen demand (COD)– alongside the hydraulic permeability coefficient (HPC). Additionally, the price of the planting layer is a key factor influencing the economic feasibility of ecological restoration projects. Therefore, it is necessary to incorporate price into the evaluation framework to ensure the validity of the study.

2.2. Model construction

2.2.1. Principal component analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique that performs an orthogonal transformation on the original data, mapping it onto a new coordinate system. By using a few mutually independent variables to capture as much variance in the data as possible, thus most of the original information is retained while eliminating irrelevant noise. **Figure 1** illustrates the geometric procedure of PCA. In this transformation, the initial variables x_1 and x_2 are linearly combined to generate new variables y_1 and y_2. A sample with greater dispersion along y_1 has a large variance contribution. It can be extracted as the first principal component (PC1), while a sample with less dispersion along the y_2 has a small variance contribution and can be used as the second principal component (PC2). By following this process, the dimensionality of the data can be reduced. Typically, cumulative variance denotes the aggregated proportion of variance explained by the initial principal components, reflecting the extent of information retained from the original dataset (Yuan *et al.*, 2022). This study selects the number of principal components, *m*, based on the principle that the cumulative variance contribution reaches 85%.

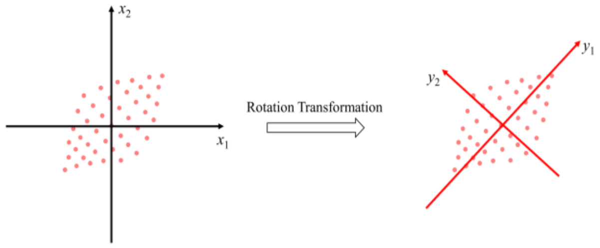


Figure 1. Geometrical interpretation of PCA

In conclusion, assuming that there are *n* variables in the original data, if the Pearson correlation analysis reveals strong correlations between the variables, PCA will reduce the dimensionality by transforming the data into *m* principal components (*m*<*n*). The reduced components y_i (i=1,2, ⋯, m) can be represented by Equation (1).

$$\begin{cases} y_1 = u_{11}x_1 + u_{12}x_2 + \cdots + u_{1n}x_n \\ y_2 = u_{21}x_1 + u_{22}x_2 + \cdots + u_{2n}x_n \\ \vdots \\ y_m = u_{m1}x_1 + u_{m2}x_2 + \cdots + u_{mn}x_n \end{cases} \quad (1)$$

Where u_{mn} is the loading coefficient, representing the weight and contribution of each original variable in the corresponding principal component. The higher the absolute value of the loading coefficient, the greater the explanatory power of the variable in the corresponding principal component.

2.2.2. Projection pursuit

Projection Pursuit (PP) is a statistical method that identifies an optimal projection direction to highlight specific patterns or features in the given data. Based on the dimensionality-reduced data obtained from PCA, this study employs PP to construct the projection index function Q(a). By employing the intelligent algorithm to optimize the projection direction, the final evaluation results are obtained.

(1) Normalization processing

To reduce the errors caused by the scale of different indicators in the calculation, this study used the min-max normalization (Equation (2)) to adjust the reduced dimensionality of the data, so that the values are uniformly distributed between 0 and 1 while maintaining the relative relationships in the data.

$$S_{ij}^{*}=\frac{S_{ij}-S_{j,min}}{S_{j,max}-S_{j,min}} \quad (2)$$

Where i is the number of samples in each group of experiments; j is the number of sample indicators; S_{ij} is the j^{th} indicator of the i^{th} sample; $S_{j,\,max}$ is the maximum value of the j^{th} indicator of all the samples; $S_{j,\,min}$ represents the minimum value of the j^{th} indicator of all the samples; and S_{ij}^{*} is the result of normalization of the j^{th} indicator of the i^{th} sample.

(2) Model construction

The core of PP lies in determining the projection direction a=a (a_1, a_2, ⋯, a_m), formulating the projection index function Q(a), and solving for the optimal projection direction to obtain the one-dimensional evaluation value Z_i (Equation (3)). To ensure that Z_i demonstrates both local clustering and overall dispersion, Q(a) can be represented by both the standard deviation S_z and the local density D_z (Huang & Lu, 2014). The model equation is shown as follows.

$$Z_i=\sum_{j=1}^{p}a(j)\cdot S_{ij}^{*},\ i=1,2,\cdots,n \quad (3)$$

$$S_z=\left[\sum_{i=1}^{n}\left(Z_i-E_z\right)^2(n-1)^{-1}\right]^{\frac{1}{2}} \quad (4)$$

$$D_z=\sum_{i=1}^{n}\sum_{k=1}^{n}(R-r_{ik})\cdot U\left(R-r_{ik}\right) \quad (5)$$

$$\max Q(a)=S_z\cdot D_z \quad (6)$$

$$s.t.\sum_{j=1}^{p}a(j)^2=1 \quad (7)$$

Where E_z is the projected mean of the sequence {Z_i|I = 1, 2, …, n}; R denotes the window radius of the local density in the projection distribution, typically set as 0.1S_Z; r_{ik} = |Z_i−Z_k| refers to the distance between samples; U(R−r_{ik}) represents the Heaviside step function.

(3) Solving for the optimal projection direction

In the PP model, solving for the optimal projection direction constitutes a nonlinear optimization problem, typically necessitating the application of advanced computational algorithms. (Zhu & Chen, 2017; Huo *et al.*, 2023). Among the available optimization techniques, GA stands out for its robust global search capabilities, adaptability, and efficiency in handling complex, high-dimensional problems (Li & Fu, 2022). It mimics biological evolution by iteratively applying operations such as selection, crossover, and mutation to improve a population of candidate solutions. These operations not only maintain diversity within the solution space but also enable GA to effectively explore large and complex search spaces, gradually converging toward the optimal or near-optimal solution.

2.2.3. Comprehensive evaluation

After solving for the optimal projection direction, combining the normalized data, and substituting it into Equation (3), the final evaluation results Z_i of each group of data, followed by ranking from best to worst. To uniformly evaluate the hydraulic permeability and pollutant removal capabilities of the plants, the Z_i from the four stages are weighted at 25% each.

The workflow of PCA-PP-GA is illustrated in detail in Figure S1.

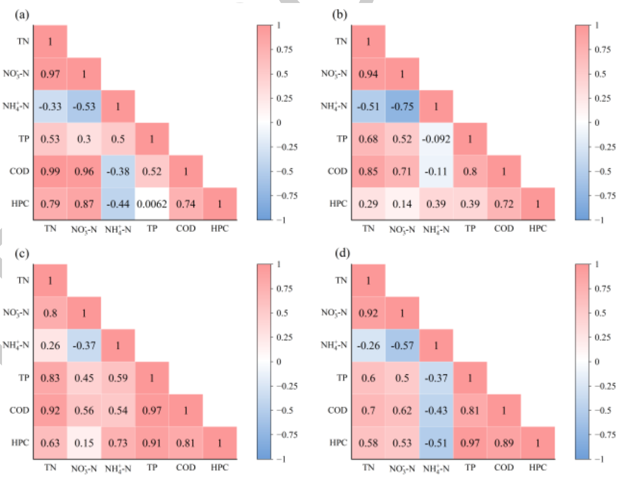


Figure 2. Correlation coefficient heatmap of the plant group: (a) Seedling Stage, (b) Growth Stage, (c) Maturity Stage, and (d) Aging Stage

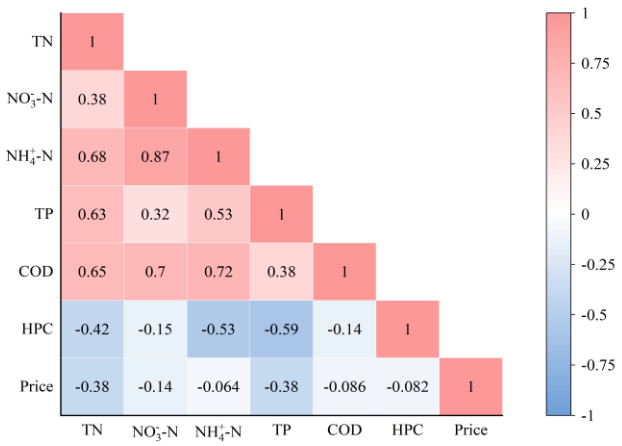


Figure 3. Correlation coefficient heatmap of the planting layer group

3. Results and discussion

3.1. Results of principal component analysis

By calculating the Pearson correlation coefficients of the plant group and planting layer group, the correlation analyses of each evaluation indicator were carried out separately. As illustrated in **Figure 2**, TN and NO_3^-−N exhibit a high correlation across different stages (r=0.97, 0.94, 0.80, 0.92) during the removal of pollutants from stormwater runoff by plants. However, the correlation is low (r= 0.38) in the planting layer shown in **Figure 3**. Additionally, within the plant group, HPC is predominantly positively correlated with other indicators. In contrast, HPC generally shows significant negative correlations in the planting layer group. These observations suggested that, in the absence of plants, the soil permeability in the bioretention system increased, resulting in accelerated water flow and diminished contact time between pollutants and soil particles, thereby reducing the effectiveness of stormwater runoff purification. At the plant maturity stage (**Figure 2(c)**), the correlation of TP with TN and COD (r = 0.83 and 0.97), respectively, significantly increases compared to the earlier two stages ((**Figure 2(a)** and **Figure 2(b)**). It indicated that plant decomposition activity may have intensified, indirectly enhancing pollutant removal rates. This aligned with Liu et al.'s findings on the effects of Phragmites at different stages in constructed wetlands (Liu *et al.*, 2018). In the correlation analysis of the planting layer group (**Figure 3**), the coefficient of NO_3^-−N and NH_4^+−N is 0.87, indicating a synergistic trend. In summary, both the plant group and the planting layer group showed high correlations among evaluation indicators, proving that there was a strong linear dependence between multiple variables. Therefore, PCA can effectively extract representative principal components, thus simplifying the data dimensions.

By analyzing the correlation coefficients, potential multicollinearity between the variables was preliminarily identified, forming the foundation for the PCA dimensionality reduction. After calculating the covariance matrix and performing eigen decomposition, the contribution rate of each group of data was obtained (**Figure 4**). Following the criterion that the cumulative variance contribution rate exceeds 85%, the corresponding principal components of the plant group and planting layer group were selected to maximize the reflection of the original information (Yuan *et al.*, 2023). As seen in **Figure 4(a)**, the evaluation indicators for the seedling, growth, and maturity stages of the plants were reduced from six to two dimensions (PC1 and PC2), with cumulative variance contribution rates of 93.54%, 89.60%, and 96.22%, respectively. **Figures 4(a)**-**(b)** showed that the data for the aging stage and the planting layer group were reduced to three dimensions (PC1, PC2, and PC3), with cumulative variance contribution rates of 97.0% and 86.53%, respectively.

The principal component factor loading matrix was obtained by calculating the weights of the original variables in the new coordinates (Table S1). Based on the data presented in Table S1, the loadings of PC1 on TP at maturity is 0.982, and on TN is 0.907. This result is consistent with the correlation coefficient between TP and TN (r=0.83) depicted in **Figure 2(c)**. The high correlation between TP and TN during the maturity stage, coupled with their substantial loadings on PC1, underscored the significant impact of these variables on ecosystem functions.

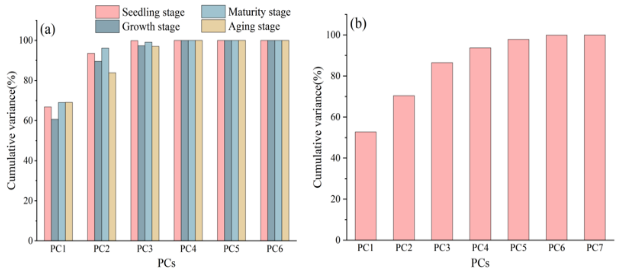


Figure 4. Cumulative contribution rate: (a) Plant group (b) Planting layer group

The above results showed that the PCA-based dimensionality reduction method simplified the data dimensions for pollutant removal performance and hydraulic permeability of plants and planting layers in bioretention systems.

3.2. Results of the principal component analysis-projection pursuit model

Based on the above PCA dimensionality reduction results, this study further developed the PCA-PP model to evaluate plants and planting layers within bioretention systems. The best projection coefficients were determined by solving the model using GA, as shown in Table S2, the comprehensive scores of the plant group (*Hemarthria sibirica*, *Cynodon dactylon*, *Paspalum wettsteinii*, *Lolium perenne*, and *Festuca elata*) and the planting layer group (L1, L2, L3, L4, and S1, S2, S3, S4) were obtained (**Figure 5**). In **Figure 5(a)**, *Cynodon dactylon* had the most outstanding performance in removing stormwater runoff pollutants, whereas *Paspalum wettsteinii* performed the worst, with a score (Z_i) difference of 0.97 between the two species. And analyzing the ratings of the plants at each growth stage, it could be seen that *Paspalum wettsteinii*, *Lolium perenne*, and *Festuca elata* were dominant only at a single stage, achieving their highest rating of merely 1.11. This value was significantly lower than the peak rating of *Cynodon dactylon*, which reached 1.58 during the aging stage. The ratings for *Paspalum wettsteinii*, *Lolium perenne*, and *Festuca elata* in other stages were generally below 0.5, contributing to their lower overall rankings. Conversely, *Cynodon dactylon* and *Hemarthria* consistently maintained high scores across all stages, with values exceeding 0.95. This may be attributed to the well-developed root systems and growth rates of both plants, which enhanced the hydraulic infiltration properties and pollutant filtration capacity of the soil. Compared to *Hemarthria sibirica*, *Cynodon dactylon* was able to maintain higher physiological activity during the aging period, therefore it was rated the highest among the five plants.

In addition, the scores of each planting layer are illustrated in **Figure 5(b)**, revealing that L4 performed the best in removing stormwater runoff pollutants, while S4 performed the worst. By analyzing the planting layer ratios (**Table 1**), it was evident that the scores of the planting layers with Ss added (Z_{S1}、Z_{S2}、Z_{S3} 和 Z_{S4}=0.37、0.32、0.51 and 0.12) were generally lower than those of the planting layers with Ls (Z_{L1}、Z_{L2}、Z_{L3} 和 Z_{L4}=1.44、1.29、1.0 and 1.47). This difference may be due to the silt and clay particles in Ls, which enhanced the soil's adsorption capacity. However, the single-grain structure of Ss was not conducive to pollutant adsorption and treatment. Moreover, in the Ss combinations, the substrate layers with added Ve (S1), Bi (S2), and Pe (S3) had higher scores than S4, which had no additives. In contrast, the same additives did not produce similar effects in the Ls combinations (L1, L2, L3, and L4), possibly because the additives disrupted the structural balance of Ls.

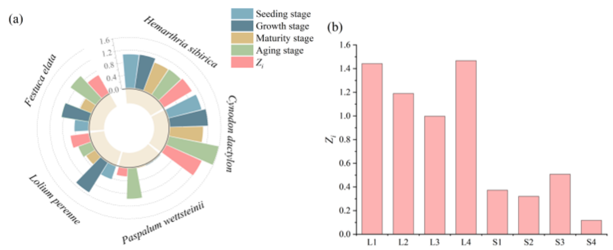


Figure 5. Evaluation results: (a) the plant group and (b) the planting layer group

Based on the results of this study, *Cynodon dactylon* was recommended as the plant, and L4 as the planting layer, for use in bioretention systems.

3.3. Comparison of evaluation methods

To verify the validity of the models, this study conducted a comparative analysis of the AHP and the PP-GA model in evaluating the hydraulic infiltration and pollutant removal performance of various plants and planting layers within the bioretention system. AHP is a decision-making method that helps evaluate complex problems by breaking them down into a set of criteria. By comparing the importance of each criterion in a pairwise comparison, it calculates the relative weight of each factor to help prioritize decisions. The judgment matrix derived from AHP is presented in Table S3, and the optimal projection coefficients obtained from the PP-GA model are detailed in Figure S2. By combining the results of both methods, the final performance ranking was derived (**Figure 6**).

In the ranking of the plant group results, the ranking obtained using the method proposed in this study is consistent with that of the PP-GA model, specifically following the order *Cynodon dactylo*＞*Hemarthria sibirica*＞*Festuca elata*＞*Lolium perenne*＞*Paspalum wettsteinii*. However, when evaluated using AHP, the ranking between *Festuca elata* and *Lolium perenne*, differed from the other methods. Upon examining the AHP calculation process, it was found that the weight of HPC was merely 0.05 (Table S3). This indicated that the AHP model mainly focused on pollutant removal performance during the weight allocation process, while insufficient consideration was given to hydraulic permeability performance, resulting in an underestimation of the HPC weight, which is a key indicator. In comparison, the PP model in the remaining two methods could analyze the characteristics of the data structure of the evaluation object and automatically adjust the weights of the indicators, thus providing more realistic results.

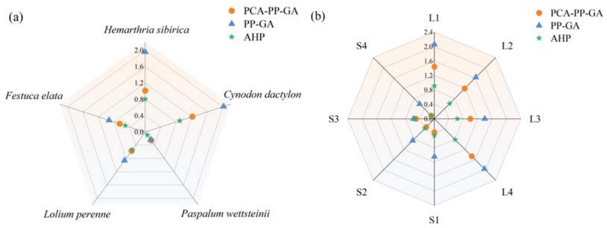


Figure 6. Evaluation results of three methods: (a) Plant group and (b) Planting layer group

When analyzing the evaluation results of different planting layers (**Figure 6**), it can be seen the ranking of PCA-PP-GA was L4 > L1 > L2 > L3 > S3 > S1 > S2 > S4, whereas both the AHP and the traditional PP-GA model identified L1 as the optimal planting layer ratio scheme. This discrepancy arose because, during PCA in this study, the data dimensionality was reduced and the focus was placed on important variables, taking the impact of price into account. For instance, the weight of the price indicator in PC3 was as high as 0.962, while in AHP it was only 0.0412 (Table S3). Therefore, PCA-PP-GA chose the more cost-effective L4 when the two planting layers had similar pollutant removal effects, and AHP chose L1 with less consideration of cost in assigning values. Furthermore, unlike AHP, which used consistent weights at all stages (Table S3), the PP-GA model could adjust the weights of different indicators based on the actual data characteristics of each variable at the four stages (Figure S2), thereby providing a more objective evaluation of the various plant growth periods. Nevertheless, the conventional PP-GA relied on the direct projection of the original data, and there existed a significant strong correlation between COD and TN, NO_3^--N, NH_4^+-N and other indicators (**Figure 2**). This could result in this model's derived projection coefficients not being accurate enough when faced with complex or multidimensional data, thereby causing a loss of information. The comparison of the evaluation methods is summarized in Table S4.

In conclusion, the PCA-PP-GA model not only reduced information redundancy and enhanced accuracy through dimensionality reduction but also achieved a better balance between performance and economic factors. It was more appropriate for bioretention system designs under economic constraints.

4. Conclusion

Aiming at the problem of multi-objective evaluation of the effectiveness of bioretention systems on stormwater runoff management, a PCA-PP model combined with GA was proposed in this study. By applying PCA to optimize the indicators of the raw data, the processing of complex multi-dimensional data was simplified. Meanwhile, the GA avoided the difficulties of traditional methods, such as parameter determination and the risk of falling into local optima, achieving more reasonable indicator weights. The evaluation results showed that *Cynodon dactylon* exhibited superior performance within the bioretention system, while the L4 configuration (10.0% Ls + 90.0% S) was identified as the optimal planting layer. Compared with AHP, the ranking obtained through PCA-PP was based on data-driven analysis, reducing the interference of subjective judgment and thus capturing the importance of cost-effectiveness for practical applications. In comparison to the PP-GA model, this study initially employed PCA to transform the raw data into principal components (e.g., PC1, PC2) with higher explanatory power, effectively mitigating the interference of irrelevant or secondary information and improving the accuracy of the multi-objective evaluation method. These results provide actionable guidance for the practical design and optimization of bioretention systems, particularly in enhancing cost-effective stormwater management strategies. By prioritizing these configurations, urban planners and ecological engineers could improve pollutant removal and hydraulic performance, contributing to the broader goals of sustainable urban development.

5. Limitations and prospects

Despite the promising potential of the PCA-PP-GA model in evaluating the performance of bioretention systems, this study primarily relies on a limited dataset for model construction and validation, as well as specific plant species and planting layer ratio combinations. This limitation introduces potential biases, restricting the generalizability of the results to other geographical regions or climatic conditions.

Future researches are recommended to focus on expanding the dataset by incorporating field data from various climates, soil types, and pollution characteristics to validate the applicability and robustness of the findings. Additionally, in addressing the economic and logistical constraints of real-world applications, exploring local material alternatives and modular construction methods would be valuable. It is suggested that prioritizing the use of local materials in small-scale experiments could help ensure scalability and adaptability to different urban environments.

Funding

This study was supported by the Natural Science Foundation of Chongqing Municipal Science and Technology Commission (cstc2020jcyj-msxmX0716), Key Laboratory of Hydraulic and Waterway Engineering of the Ministry of Education, Chongqing Jiaotong University (SLK2021A09), Technology Innovation and Application Development Project of Chongqing (CSTB2022TIAD-KPX0200).

References

Aziz P., Muhammad N., Intisar A., Abid M.A., Din M.I., Yaseen M., Kousar R., Aamir A., Quratulain and Ejaz R. (2021) Constituents and antibacterial activity of leaf essential oil of plectranthus scutellarioides. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*. **155**, 1247–1252.

De-yong W., Jun-long H., Jing W. and Qingjun Z. (2020) Study on the selection of vegetation for ecological restoration of residual soil slopes in south china. *E3S Web of Conferences*. **198**, 4039.

Ferde M., Costa V.C., Mantovaneli R., Wyatt N.L.P., Rocha P. de A., Brandão G.P., de Souza J.R., Gimenes A.C.W., Costa F.S., da Silva E.G.P. and Carneiro M.T.W.D. (2021) Chemical characterization of the soils from black pepper (piper nigrum L.) cultivation using principal component analysis (PCA) and kohonen self-organizing map (KSOM). *Journal of Soils and Sediments*. **21**, 3098–3106.

Garg H. (2016) A hybrid PSO-GA algorithm for constrained optimization problems. *Applied Mathematics and Computation*. **274**, 292–305.

Huang H. and Lu J. (2014) Identification of river water pollution characteristics based on projection pursuit and factor analysis. *Environmental Earth Sciences*. **72**, 3409–3417.

Huo Z., Zha X., Chu Y., Lu M. and Zhang S. (2023) Research on river water quality evaluation based on the GA-PP and improved fuzzy model. *Water Science and Technology*. **88**, 2160–2173.

Jafari A., Khalili T., Babaei E. and Bidram A. (2020) A hybrid optimization technique using exchange market and genetic algorithms. *IEEE Access*. **8**, 2417–2427.

Jia Q. and Zhessakov A. (2021) Study on ecological evaluation of urban land based on GIS and RS technology. *Arabian Journal of Geosciences*. **14**, 261.

Lee S., Caffo B.S., Lakshmanan B. and Pham D.L. (2015) Evaluating model misspecification in independent component analysis. *Journal of Statistical Computation and Simulation*.

Li M. and Fu Y. (2022) Prediction of Supply Chain Financial Credit Risk Based on PCA-GA-SVM Model. *Sustainability*. **14**, 16376.

Li Y., Zhang Y., Yu H., Han Y. and Zuo J. (2021) Enhancing nitrate removal from urban stormwater in an inverted bioretention system. *Ecological engineering*. **170**, 106315.

Liu X., Zhang Y., Li X., Fu C., Shi T. and Yan P. (2018) Effects of influent nitrogen loads on nitrogen and COD removal in horizontal subsurface flow constructed wetlands during different growth periods of *phragmites australis*. *Science of the Total Environment*. **635**, 1360–1366.

Liu Y.F., Dunkerley D., López-Vicente M., Shi Z.H. and Wu G.L. (2020) Trade-off between surface runoff and soil erosion during the implementation of ecological restoration programs in semiarid regions: a meta-analysis. *Science of the Total Environment*. **712**, 136477.

Mehmood T., Gaurav G., Liu C., Klemeš J., Usman M., Bokhari A. and Lu J. (2021) A review on plant-microbial interactions, functions, mechanisms and emerging trends in bioretention system to improve multi-contaminated stormwater treatment. *Journal of Environmental Management*. **292**, 113108.

Mei Y., Gao L., Zhou H., Wei K.-H., Cui N.-Q. and Chang C.-C. (2018) Ranking media for multi-pollutant removal efficiency in bioretention. *Water Science and Technology*. **77**, 2023–2035.

Muhammad N., Guo D., Zhang Y., Intisar A., Subhani Q., Qadir M.A. and Cui H. (2019) Online clean-up setup for the determination of non-fluorescent acidic pharmaceutical drugs in complex biological samples. *Journal of Chromatography B*. **1126–1127**, 121708.

Muhammad N., Hussain I., Ali A., Noureen L., He Q., Subhani Q., Khan N.A., Cui H. and Zhu Y. (2024) Ion chromatography: a comprehensive review of sample preparation methods for analysis of halogens and allied nonmetals in critically challenging inorganic matrices. *Journal of Chromatography A*. **1734**, 465311.

Muhammad N., Hussian I., Ali A., Hussain T., Intisar A., Ul Haq I., Subhani Q., Hedar M., Zhong J.-L., Asif M., Guo D., Cui H. and Zhu Y. (2022) A comprehensive review of liquid chromatography hyphenated to post-column photoinduced fluorescence detection system for determination of analytes. *Arabian Journal of Chemistry*. **15**, 104091.

Pereira J.F.Q., Silva C.S., Braz A., Pimentel M.F., Honorato R.S., Pasquini C. and Wentzell P.D. (2017) Projection pursuit and PCA associated with near and middle infrared hyperspectral images to investigate forensic cases of fraudulent documents. *Microchemical Journal*. **130**, 412–419.

Vijayaraghavan K., Biswal B.K., Adam M.G., Soh S.H., Tsen-Tieng D.L., Davis A.P., Chew S.H., Tan P.Y., Babovic V. and Balasubramanian R. (2021) Bioretention systems for stormwater management: recent advances and future prospects. *Journal of Environmental Management*. **292**, 112766.

Wang R., Zhang X. and Li M.-H. (2019) Predicting bioretention pollutant removal efficiency with design features: a data-driven approach. *Journal of Environmental Management* **242**, 403–414.

Wen B. and Chen G. (2012) Principal component analysis of network security data based on projection pursuit. In *Network Computing and Information Security*. 380–387.

Wentzell P.D., Hou S., Silva C.S., Wicks C.C. and Pimentel M.F. (2015) Procrustes rotation as a diagnostic tool for projection pursuit analysis. *Analytica Chimica Acta*. **877**, 51–63.

Wu J. (2019) Evaluation and prediction of water resources utilization structure based on projection pursuit method. *IOP Conference Series: Earth and Environmental Science*. **300**, 22164.

Xu B., Ling L. and Jin H. (2017) Research on evaluation of third-party governance operation services for environmental pollution. IOP Conference Series: Earth and Environmental Science. **94**, 12079.

Xu D., Liu D., Liu D., Fu Q., Huang Y., Li M. and Li T. (2022) New method for diagnosing resilience of agricultural soil-water resource composite system: projection pursuit model modified by sparrow search algorithm. *Journal of Hydrology* **610**, 127814.

Xu R., Li X., Yang W., Jiang C. and Rabiei M. (2019) Use of local plants for ecological restoration and slope stability: a possible application in Yan'an, loess plateau, china. *Geomatics Natural Hazards Risk*. **10**, 2106–2128.

Yang, L., Qu, J., Chen, N., Xu, C. and Li, Z. (2010) Assessment of groundwater resources and environment carrying capacity based on coupled model of PSO and projection pursuit. In *2010 4th International Conference on Bioinformatics and Biomedical Engineering*. pp.1–5.

Yuan H., Cao Z., Xiong L., Li H. and Wang Y. (2022) A machine learning method for engineering risk identification of goaf. *Water*. **14**, 4075.

Yuan H., Ji S., Liu G., Xiong L., Li H., Cao Z. and Xia Z. (2023) Investigation on intelligent early warning of rock burst disasters using the PCA-PSO-ELM model. *Applied Science*. **13**, 8796.

Zhang C., Ye J. and Wang X. (2023) A computational perspective on projection pursuit in high dimensions: feasible or infeasible feature extraction. *International Statistical Review*. **91**, 140–161.

Zhu Z. and Chen X. (2017) Evaluating the Effects of Low Impact Development Practices on Urban Flooding under Different Rainfall Intensities. *Water*. **9**, 548.