A dynamic framework for contractors to measure the performance of highway maintenance based on grey theory

Xuelian Wu[1], Xiaoli Shi[1*], Yongjian Ke[2], Yuhuan Li[3], Malongzhi Wan[4] & Xiaotian Gong[1]

[1]School of Highway, Chang'an University, Xi'an 710064, Shaanxi, People's Republic of China.

[2]School of Built Environment, University of Technology Sydney, Ultimo 2007, NSW, Australia.

[3]Xi'an Municipal Engineering Design & Research Institute Co., Ltd., Xi'an 710064, Shaanxi, People's Republic of China.

[4]College of Engineering, Qinghai Institute of Technology, Xi'ning 810016, Qinghai, People's Republic of China.

*Email: glxl@gl.chd.edu.cn

Abstract: Highway maintenance performance measurement is the key to implementing performance-based contracting (PBC). However, contractors have struggled with accurately understanding client intent and translating performance goals into measurable indicators. To make PBC successfully applied, this study proposes a dynamic performance measurement framework of highway routine maintenance based on grey theory. A performance indicator system is established based on the organizational structure of highway assets and the bill of quantities in the real world. In addition, a statistical analysis using the data about maintenance cost and highway condition score has been conducted to ascertain the weights and thresholds assigned to various indicators. As a result, the performance indicators, their corresponding weights, and thresholds can be flexibly adjusted to accommodate diverse periods and areas. Then, the grey theory has been employed to evaluate the comprehensive performance level of service within a region. This framework is experimented with using the data of three maintenance areas in Shaanxi Province, China: Qiaogouwan, Huaziping, and Ansai. At the end of the process, comparison and sensitivity analysis are conducted. The results demonstrate the model's reasonability and robustness, providing valuable tools for contractors to assess the performance of highway maintenance in meeting client specifications.

Keywords Highway routine maintenance, Performance measurement, Grey theory, PBC

Maintaining the highway infrastructure at a high condition level with generally limited available funding is challenging for many transportation agencies. Many highway administrators worldwide have implemented PBC to address this challenge, resulting in an average reduction of 21% in indirect costs [1]. The PBC is a contractual approach of tying at least a portion of supplier (i.e. contractor) payment to specific and measurable performance [2–4]. However, at the start of every PBC project, the buyer (i.e. client) is confronted with serious measurement problems and translation issues. The contract fails to indicate precisely the required performance ("cryptic texts"). Because in many working arrangements and projects parties have to deal with 'hidden fuzzy information'. Not all information is explicitly included in the contract but merely implied [5]. Consequently, the accurate translation of fuzzy functional specifications from Performance-Based Contracting (PBC) into detailed technical specifications has become a crucial concern for contractors.

To address this issue, there are relative performance measurement methods and absolute performance measurement methods. Relative performance measurement primarily focuses on evaluating the performance of multiple competitors, and the results are expressed by relative efficiency, such as data envelopment analysis (DEA). Based on the DEA method, Ma (2018) took maintenance cost, AADT, and annual rainfall per lane kilometre as input indicators, some highway condition indexes as output indicators, and established a performance evaluation model [6]. Fallah‑Fini et al. (2010) presented an application of an environmental classification method employed in conjunction with DEA for assessing the relative efficiency of highway maintenance operations in eight counties throughout the state of Virginia [7]. Fallah-Fini et al. (2012) adapted the non-parametric meta-frontier framework to the two-stage boot-strapping technique to develop an analytical approach for evaluating the relative efficiency of two highway maintenance contracting strategies [8]. Different from relative performance measurement methods, the values obtained by absolute performance measurement are more independent. Some researchers have explored deterministic, probabilistic, and fuzzy logic-based approaches in absolute performance evaluation [9]. The deterministic approach typically assigns fixed weights to performance indicators [10,11]. Abu Dabous et al. (2021) introduced an evidential reasoning (ER) theory-based approach to pavement condition evaluation [12]. This is a probabilistic model and provides a comprehensive evaluation of all possible ratings of a pavement section with the probability of each rating, which is an enhanced evaluation compared to a single numeric indicator. A few studies show that the fuzzy approach proves to be one of the best mathematical tools for assessing highway performance [13]. Because uncertainty inherent in pavement condition performance lends itself to applying fuzzy theory, which can effectively quantify such ambiguity and model the associated uncertainties [14]. Sun and Gu (2011) have proposed using fuzzy logic theory and AHP to assess the pavement condition and prioritize highway maintenance projects [15]. Singh et al. (2018) presented two approaches of fuzzy mathematical analysis to conceive strategic planning for the maintenance and rehabilitation of pavements [16]. They were fuzzy AHP and fuzzy weighted average methods. While the concept of fuzzy theory is more well-established, grey theory has gained significant traction in recent years. Grey theory was created in 1982 in China by Ju-Long (1982) and provides a different perspective for capturing uncertainty [17]. Both theories of fuzzy and grey can be considered proponents of uncertainty modelling, but they differ in their underlying principles. Grey theory, an extension of fuzzy theory, is adaptable for both discrete and continuous data. It emphasizes handling small, low-quality datasets, and addressing scenarios where data may be incomplete or redundant [18]. In addition, grey theory is founded upon objective uncertainty, whereas fuzzy theory addresses subjective uncertainty. Yu et al. (2017) developed a new pavement quality index model utilizing grey theory [19]. Sasmal and Ramanjaneyulu (2008) proposed a fuzzy and AHP-based framework for condition ranking of reinforced concrete bridges [9], while Liu et al. (2012) combined fuzzy AHP and grey theory analysis for optimal old bridge retrofit [20]. These experiments prove that grey theory is effective in the field of performance evaluation.

The level of complexity of PBC can range from “simple” to “comprehensive” depending on the number of assets and range of services included [21,22]. The “simple” PBC would cover a single service (e.g., only mowing) and could be awarded for relatively short periods (several months or one year). The “comprehensive” PBC would typically cover all highway assets with the right-of-way and comprise the full range of services needed to manage and maintain the contracted highway corridor [22]. Such services could include routine maintenance, periodic maintenance, traffic accident assistance, etc. For instance, In Republic of Serbia, the World Bank Project included implementing PBC for routine highway maintenance and winter maintenance as well as routine bridge maintenance on about 1200 km of highway network for 3 years [21]. In Argentina, a combined rehabilitation and maintenance contract required the contractor to rehabilitate and subsequently maintain a sub-network of highways under a lump-sum contract for a total period of five years [23]. The type of maintenance works contracted using PBC in Malaysia was either periodic or routine works. In addition to maintenance works, the government practiced PBC in works related to the repair and overhaul of complex equipment [24]. In Netherlands, Rijkswaterstaat (RWS)—the Dutch public agency responsible for the design, construction, management, and maintenance of the main infrastructure facilities—employs integral contracts which means within a single contract and are contracted out to a single (relatively large) main contractor [5]. Since the first maintenance PBC was piloted in British Columbia in 1988, many PBCs have been implemented around the world. However, implementation of PBC is still a challenge for many developing countries because of resource and skill limitations, corruption, and poor management systems [25]. Therefore, a simple pilot trial project of PBC is needed before it is fully introduced to measure the feasibility, capability, cost, and quality of work. Compared with other maintenance projects, the characteristics of routine maintenance are small scale, large dispersion, and short working time. In this way, if the contracts were cancelled due to the contractor’s failure to meet the required performance, the projects can then be stopped immediately, and without much impact on the operating environment of highway. As a result, routine maintenance is very suitable as a simple pilot trial project of PBC.

Despite various suggestions in the literature concerning performance evaluation, majority of performance measurement methods have been devised from the client’s standpoint to assess the contractor's performance. However, the contractor's self-evaluation is also very important. The implementation of the payment principle within the PBC incentivizes the contractors to undertake measures to enhance the condition of the highway. Therefore, it is necessary to establish a maintenance performance evaluation indicator system from the bottom (maintenance items) to the top (maintenance projects) for contractors. In addition, to achieve the level of service (LOS) specified by the client, it is more appropriate for the contractor to use absolute measurement methods for self-evaluation. Besides, the grey theory is a very potential method for measuring maintenance performance. On the other hand, highway deterioration and maintenance are ongoing processes. Dynamic phenomena embedded in highway systems where the uncontrollable factors of the environment (e.g. climate condition) and operational conditions (e.g. traffic load), as well as controllable factors (e.g. maintenance policy), affect the highway condition [7]. In such a setting, the measurement method of maintenance performance should also be dynamically adjustable.

In this context, this study aims to establish a dynamic framework for routine maintenance performance measurement from contractors' perspective. The rest of this paper is organized as follows. The next section reviews the research content included in the performance measurement. In the section "Research methodology and model development", the performance indicator system is established based on the analysis hierarchical principle, taking into account the organizational structure of highway assets, the indicator weights and thresholds of the LOS have been calculated by the graph of cumulative frequency distribution of highway condition scores. Then, the dynamic performance measurement method is established by grey theory. This framework helps contractors improve highway maintenance performance to meet client's requirements and continually improve the conditions of the highways.

Literature review

A complete performance measurement framework should include the performance indicator, weight of the performance indicator, level of service, and performance measurement method [26–28]. First, the performance indicator is the basis of applying PBC by the maintenance contractor [29]. Second, the problems of appropriate choice of weights cannot be ignored as managers may disagree with the choice of weights [30]. Moreover, the weight has a great influence on the calculation result. Third, service of level is a criterion to judge whether the measurement results meet the minimum acceptable level. Lastly, the performance measurement method is to multiply a series of selected indicators by their corresponding weights to arrive at a final value that can be used to measure maintenance performance. This paper adopts the grey theory as a performance measurement method, so the following are summarized from these first three parts.

Performance indicator

For every type of maintenance and operation, performance indicators should be clearly defined and objectively measurable to avoid ambiguity and risk disputes. Performance indicators in this study are the technical indicators translated from the functional indicators given by the clients, so it is mainly used to evaluate the maintenance quality. Part of the performance indicators are shown in Table 1. Most of them are taken from highway condition indexes [26,28,29,31] or a combination of them [7,32–34]. For example, Liu et al. (2006) established an evaluation indicator system for highway subgrade based on slope, retaining wall, drainage facilities, and shoulder [35]. Ma et al. (2012) designed the bridge evaluation indicator system, and the bridge structural integrity and road–bridge adaptability were taken into account [36]. Xiao et al. (2023) proposed a post-evaluation indicator system of utility model maintenance technology from five aspects engineering technology, management technology, economic benefits, social environment, and sustainable development [16]. Orugbo et al. (2015) developed the strength, weakness, opportunity, and threat indicators which are specific, measurable, attainable, realistic, and time-sensitive to measure the performance of the different variants of road maintenance outsourcing (e.g. Operations and Maintenance contracts, Build Operate and Transfer, Based Road Contracts and Operations, and Traditional Road Maintenance contracts) and road maintenance work [30]. Haas et al. (2008, 2009) proposed performance indicators for bridges, namely the remaining life in years and safety considerations [37,38]. Xie (2022) established two long-term pavement maintenance performance indicators, namely the sustainable level and safety rate [39]. Calculation of both indicators utilized the area under the performance curve. Lu et al. (2000) suggested evaluating the maintenance performance of highways by using a good rate of pavement, structures (e.g. bridges and culverts), and facilities along highway

[40]. In some studies, performance indicators are often expressed in LOS represented by rating scales corresponding to the asset's condition. For example, Fallah - Fini et al (2010) designed three levels: Not severe, severe, and very severe [7]. Contractors engaged in multiple types of projects necessitate a comprehensive performance indicator. A typical approach is to combine multiple indicators describing highway condition characteristics into a common performance indicator, thereby capturing the holistic state of highway [41]. The methodology used to calculate this comprehensive indicator is described in the section titled "Performance measurement method based on grey theory".

Author	Indicator	Description
Abu et al (2017) [26], Ozbek et al (2011) [28], Alyami et al (2013) [29]	International roughness index, rutting depth, alligator cracking extent, surface rating, pavement condition	Key condition-based performance indexes are chosen.
Ozbek et al (2011) [28]	Rating on a number scale	Indicators from National Bridge Inventory program
Xiao et al (2023) [31]	Short-term benefit and service life	The two indicators are calculated based on key condition-based performance indexes.
Xiao et al (2023) [31]	Preliminary process, technical scheme, road condition inspection and evaluation, and operation and maintenance	Expert score.
Fallah-Fini et al (2015) [32]	Critical condition index (CCI)	The CCI represents the frequency and severity of different types of distress that exist on a road.
Ma, Y. F. et al (2012) [36]	Highway bridge quality	The indicator is calculated based on key indexes of national bridge inventory program.
Shah et al (2014) [42]	Road condition index, traffic volume factor, and dedicated factor	Key condition-based performance indexes and environmental factors are chosen.
Xie (2022) [39]	Sustainable level and rate of safety	The two indicators are calculated based on the area enclosed by the performance curve and the axes.
Fallah - Fini et al (2010) [7]	Not-severe, severe, and very-severe.	Severity represents depth and openness of a crack.
Lu et al (2000) [40]	The good rate of pavement, structures, and facilities	Expert score.

Table 1. Maintenance performance indicators.

Performance indicator weight

When it comes to evaluating an object according to different indicators, indicator weights are essential. The commonly used indicator weighting methods are the subjective weighting method, objective weighting method, and subjective and objective combination weighting method. As a subjective weighting method, the weights can be determined through expert analysis [36]. Zheng et al. (2019) used the analytic hierarchy process (AHP) method to obtain the weights. While expert scoring can provide insight, it often lacks the requisite objectivity. Therefore, it is imperative to seek out a more impartial methodology for ascertaining indicator weights [43]. Lu (2010) provided a way to obtain the indicator weights by the weighted average of the score obtained through collected data and expert scores [44]. Shervin Zakeri (2023) studied some examples of the subjective weighting methods and objective weighting methods in materials selection [45]. For subjective weighting methods, AHP and best-worst method (BWM) methods have been used mostly [45]. AHP provides a systematic, structured approach to decision-making by organizing complex problems into hierarchical frameworks. On the other hand, BWM employs a more straightforward methodology, whereby decision-makers select the best and worst criteria from a given set of options. Additionally, the entropy method is a widely used objective approach for computing criteria weights in material selection applications. In the domain of performance evaluation for PBC, it may be feasible to derive weights directly from the data itself to further analyze the characteristics of the existing information. It aligns with the principle of "letting the data speak for itself" [46] and makes the objective method more acceptable.

Performance LOS

LOS have been divided by thresholds in this study. Commonly, the thresholds are subjectively defined based on the experience and judgment of the inspector or expert [12,47]. This method can misrepresent the obtained evaluation result and cause wrong intervention decisions. Ram B.

Kulkarni (2010) determined the threshold by multiplying the total potential score (linked to sample inspection) by the performance target set for asset item [28]. Mehmet E. Ozbek Alsharqawi (2018) developed a robust method based on k-means clustering technique to solve this issue. Several clusters were generated during the clustering process. The condition thresholds were achieved by calculating the mean between two consecutive clusters [48]. Xie (2022) established an interest game model and subsequently determined the range of thresholds [39]. However, these methods are not suitable for dealing with the thresholds between multiple service levels. On this basis, there is a need for a quantitative methodology wherein thresholds can be systematically considered and dynamically adjusted as the underlying data is updated.

Research methodology and model development

As illustrated in Fig. 1, there are four main contents of the framework: First, the maintenance performance indicators have been designed with the application of analysis hierarchical principle, taking into account the structural relationships that exist between various maintenance items. Weights and thresholds of performance indicators are the statistical analysis results of actual maintenance and inspection data. Specifically, indicator weights are determined based on the ratio of actual routine maintenance cost, while the thresholds are determined by the cumulative frequency distribution of the highway condition scores. The scores corresponding to points of obvious jump are taken as the threshold of performance LOS. Then, the grey theory is used for performance measurement. Finally, the proposed method is validated using a practical project as an example. Comparison analysis is made between the new method and existing evaluation methods, and sensitivity analysis is used to measure the robustness of the proposed new method itself.

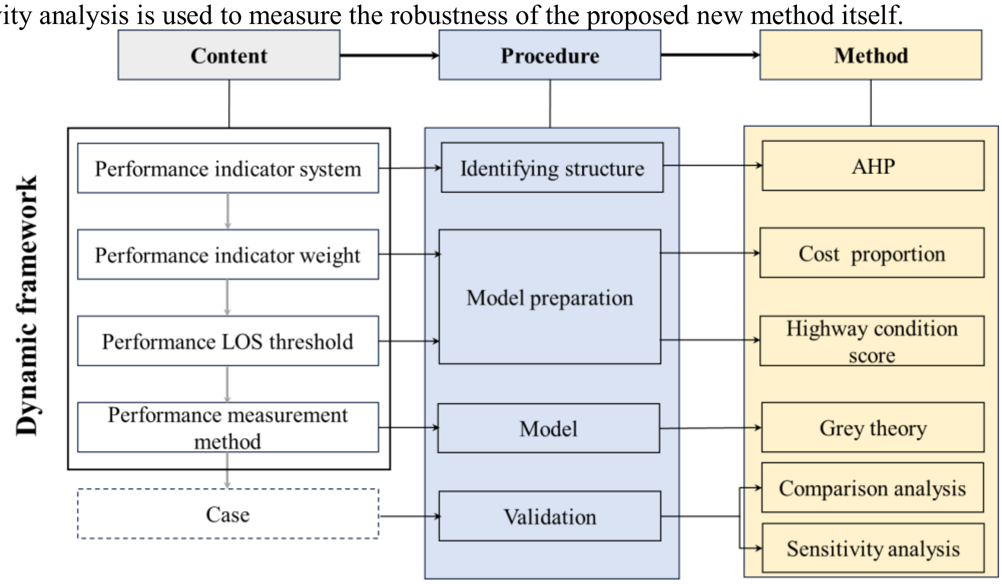

Figure 1. The overview of the study.

Performance indicator system

The indicator system starts from the overall objective level of the evaluation object and then expands step by step. At last, the four-layer indicator system can be established as shown in Fig. 2. Highway maintenance performance achieved by the objective level, five types of highway asset comprise first-level indicators, the routine cleaning and routine repair of each highway asset comprise ten second-level indicators, and the third-level indicators are determined according to the items of each routine maintenance during the analysis period. This will change depending on the project the contractor is working on. The objective level indicator is set as W, and composed of first level indicators U_i (i.e. $U=\{U_1, U_2, \ldots, U_5\}$), the first level indicators are composed of second level indicators U_{ij} (i.e. $U_i=\{U_{i1}, U_{i2}\}$), and the second level indicators are composed of third level indicators U_{ijk} (i.e. $U_{ijk}=\{U_{ij1}, U_{ij2}, \ldots, U_{ijn}\}$).

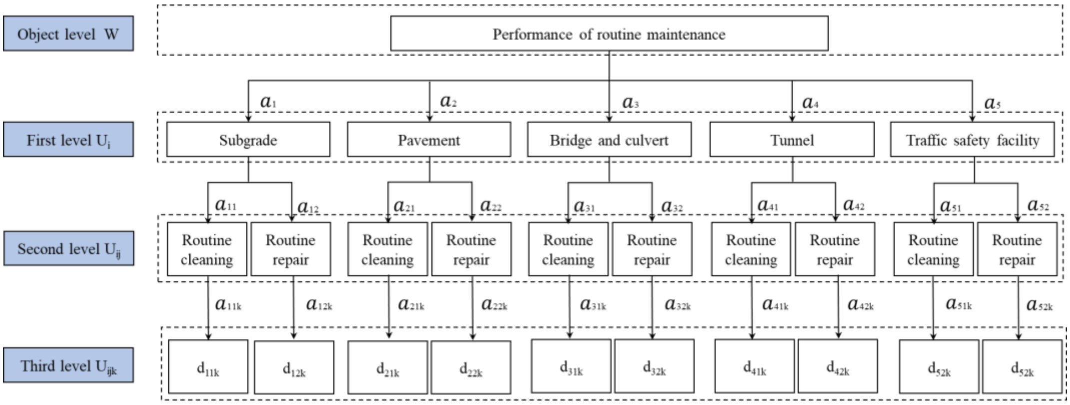

Figure 2. Performance indicator system of routine maintenance performance.

In contrast to establishing the indicator system, the calculation step of the indicator value is from the bottom to the top. According to the highway condition inspection data, the score of each third-level indicator is obtained. d_{ijk}^{p} is the score of d_{ijk} at P_{th} highway condition inspection, and m times inspections are performed. Then, the third-level indicators evaluation matrix U_{ijk} (i.e. $U_{ijk}=(d_{ijk}^{1},d_{ijk}^{2},\ldots,d_{ijk}^{m})$) can be developed.

a_{ijk} in Fig. 2 is the weight of U_{ijk}. A (i.e. $A=(a_1,a_2,...,a_5)$), A_i (i.e. $A_i=(a_{i1},a_{i2})$), and A_{ij} (i.e. $A_{ij}=(a_{ij1},a_{ij2},...,a_{ijn})$) are the weight vectors of the first, second, and third-level indicators respectively. Finally, the W of the object level is calculated step by step according to a_{ijk} and U_{ijk}.

Computing weights of performance indicator

Due to the maintenance performance level achieved by the contractor with various maintenance cost inputs, the indicator weight can be determined according to the ratio of actual routine maintenance cost of each item. Because the ratio of the pay item cost to the total project cost can indicate the complexity and importance of a project [49]. Beyond this, the improvement of performance such as quality, safety, and efficiency can also be reflected by costs. This weighting method can dynamically adjust the weights according to the difference of each expense each year, and ensure the objectivity of the weights. For example, the ratio of the cost of the third-level indicators to the routine cleaning or routine maintenance cost of the second level-indicators respectively, the ratio of routine cleaning cost and routine repair cost in the second-level indicators to the cost of each asset in the first-level indicators, and the ratio of the cost of each asset in the first-level indicators to the total routine maintenance cost. These ratios are taken as weights of indicators at all levels. The calculations for each indicator weight are shown in the following equations:

$$a_i=\frac{\mathrm{Cos}\,t_i}{\mathrm{Cos}\,t} \tag{1}$$

$$a_{ij}=\frac{\mathrm{Cos}\,t_{ij}}{\mathrm{Cos}\,t_i} \tag{2}$$

$$a_{ijk}=\frac{\mathrm{Cos}\,t_{ijk}}{\mathrm{Cos}\,t_{ij}} \tag{3}$$

$\mathrm{Cos}\,t$, $\mathrm{Cos}\,t_i$, $\mathrm{Cos}\,t_{ij}$, and $\mathrm{Cos}\,t_{ijk}$ represent the total routine maintenance cost, the maintenance cost of asset i, the routine cleaning cost $\mathrm{Cos}\,t_{i1}$ or routine repair cost $\mathrm{Cos}\,t_{i2}$, and the maintenance cost of item ijk, respectively.

Performance measurement method based on grey theory

This study employs grey clustering analysis, a technique grounded in grey theory, for the dynamic measurement of performance. The grey clustering approach relies on the construction of grey correlation matrix or the establishment of Albino functions corresponding to select observation indicators, which are then used to categorize the performance into distinct groups [50]. In this work, we apply the grey numbers white weight function method, as this method helps to describe the "preference" degree of grey numbers for different values in the value interval using the Albino function.

In this section, a set of g grey classes (i.e. C={C_1, C_2, …, C_g}), and a set of m objects, according to sample values x_{ijk}, are defined. They represent g LOS, m inspection results, and the probability value the U_{ijk} in this performance measurement method, respectively. The value of the objective level indicator W can be calculated in Eq. (4). The Eqs. (5) to (8) are carried into Eq. (4) for calculation.

$$W = B\lambda^T \tag{4}$$

$$B = A \cdot R \tag{5}$$

$$B_i = A_i \cdot R_i \tag{6}$$

$$B_{ij} = A_{ij} \cdot R_{ij} \tag{7}$$

$$R_{ij} = \begin{bmatrix} r_{ij1} \\ r_{ij2} \\ \cdots \\ r_{ijk} \end{bmatrix} = \begin{bmatrix} r_{i11}\ r_{i12} \cdots r_{i1k} \\ r_{i21}\ r_{i22} \cdots r_{i2k} \\ \cdots \\ r_{in1}\ r_{in2}\ \cdots r_{ink} \end{bmatrix} \tag{8}$$

where R (i.e. $R = (B_1, B_2, B_3, B_4, B_5)$) is a matrix composed of measurement value vector B_i of U_i, R_i (i.e. $R_i = (B_{i1}, B_{i2})$) is a matrix composed of measurement value vector B_{ij} of U_{ij}, R_{ij} is the grey measurement value matrix of the second-level indicator composed of the grey measurement value vector of the third-level indicator, $r_{ijk} = (\mathrm{r}_{ijk}^1, \mathrm{r}_{ijk}^2, \ldots, \mathrm{r}_{ijk}^g)$ is the value vector of U_{ijk} belongs to the different grey type, as shown in Eq. (9):

$$r_{ijk}^q = \frac{x_{ijk}^q}{x_{ijk}}, \ q = 1, 2, \ldots, g \tag{9}$$

where x_{ijk}^q is the probability of U_{ijk} belonging to the q$_{\text{th}}$ grey classes, the x_{ijk} is probability of U_{ijk}. The steps of the method of calculating x_{ijk}^q and x_{ijk} can be presented as follows:

Step 1: The maximum score of highway condition is set as 10, and different grey classes are divided according to the inspected results of the highway condition score. As shown in Fig. 3, plot the cumulative frequency distribution of highway condition score. The threshold of the score interval is the horizontal coordinate of the point where urgent jumps occur. These are "a" and "c" in Fig. 3.

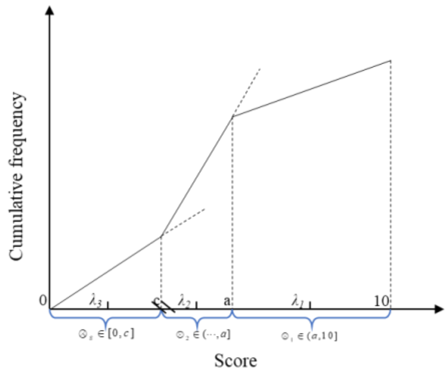

Figure 3. Highway condition score accumulation curve.

Step 2: The grey class interval can be obtained according to the threshold calculated in Step 1. Set the mean of each grey class interval as the Albino function threshold λ={λ_1, λ_2, …, λ_g}, as shown in Table 2.

Grey class/ LOS	Class 1	Class 2	. . .	Class *g*
Score	*a*	*b*	. . .	*c*
λ	λ_1	λ_2	. . .	λ_g
Interval	$Ä_1$ Î $(a,10]$	$Ä_2$ Î (×××;a]	. . .	$Ä_g$ Î $[0,c]$

Table 2. The grey class and threshold.

Step 3: The Albino functions are established according to the grey class number, as shown in Fig. 4. Then, the correspondence rules of the functions $F(X)=\{f_1(x), f_2(x), \ldots, f_g(x)\}$ are obtained by Eqs. (10) to (14).

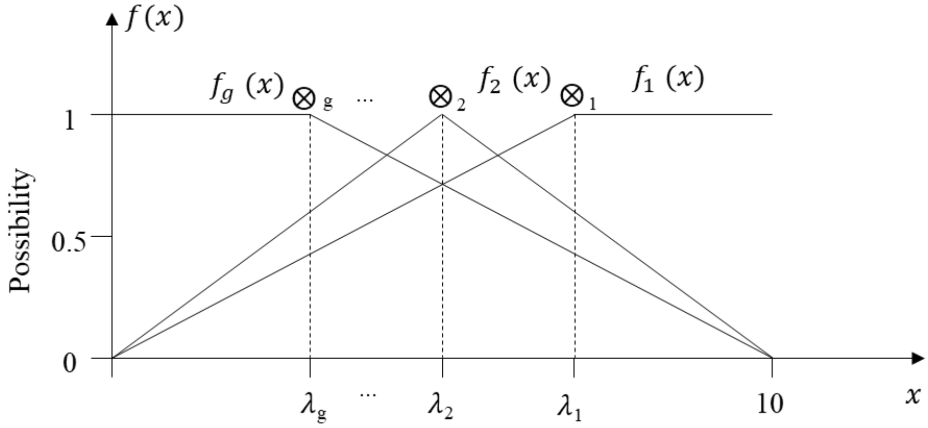

Figure 4. Albino function.

If q equal to 1,

$$f_q(x)=\begin{cases}\dfrac{1}{\lambda_q}x & 0\le x\le\lambda_q\\ 1 & \lambda_q<x\le 10\end{cases} \tag{10}$$

If q equal to g,

$$f_q(x)=\begin{cases}\dfrac{1}{\lambda_q}x & 0\le x<\lambda_q\\ \dfrac{10-x}{10-\lambda_q} & \lambda_q\le x<10\end{cases} \tag{11}$$

Others,

$$f_q(x)=\begin{cases}1 & 0\le x<\lambda_q\\ \dfrac{10-x}{10-\lambda_q} & \lambda_q\le x<10\end{cases} \tag{12}$$

Step 3: x_{ijk}^q and x_{ijk} can be obtained by Eqs. (13) and (14).

$$x_{ijk}^q=\sum_{p=1}^{m} f_q(\mathrm{d}_{ijk}^p) \tag{13}$$

$$x_{ijk}=\sum_{p=1}^{m}\sum_{q=1}^{g} f_q(\mathrm{d}_{ijk}^p)=\sum_{q=1}^{g} x_{ijk}^q \tag{14}$$

Application example

Data collection and preparation

To demonstrate the practicability of the proposed framework, a case study was conducted. The project involved the routine maintenance performance evaluation of three maintenance areas (i.e.

Qiaogouwan, Huaziping, and Ansai in Shaanxi Province, China) by contractors. Bill of Quantity tables for fiscal year 2016 were provided by Shaanxi Transportation Holding Group (STHG), which showed the costs and quantity of routine maintenance items. In addition, the tables of highway condition scores are also provided, which contain the scores of various highway maintenance items. The score is the result of inspections and can be scored out of 10. The deduction standard and deduction limit are specified for each maintenance item. The final score is the full score minus the deduction limit or the actual deduction.

Routine maintenance performance indicator system development

According to the routine maintenance content, the maintenance indicator system was determined. In 2016, STHG implemented routine cleaning and routine repair including assets of subgrade, pavement, bridge & culvert, and traffic safety facility, and their costs are shown in Table 3, respectively. Table 4-9 describes the items of cleaning and repair. Tables 3 to 9 constitute STHG 2016's maintenance performance indicator system. The weights (i.e. Ratio) corresponding to these indicators are calculated by Eqs. (1) to (3).

Objective level	Cost (10,000 CNY)	First level	Cost (10,000 CNY)	Ratio (%)	Second level	Cost (10,000 CNY)	Ratio (%)
Total cost of routine maintenance（U）	638.19	Subgrade（U_1）	61.49	0.10	Cleaning（U_{11}）	26.65	0.43
					Repair（U_{12}）	34.84	0.57
		Pavement（U_2）	330.14	0.52	Cleaning（U_{21}）	261.11	0.79
					Repair（U_{22}）	69.03	0.21
		Bridge & culvert（U_3）	69.62	0.11	Cleaning（U_{31}）	35.05	0.50
					Repair（U_{32}）	34.577	0.50
		Tunnel（U_4）	24.43	0.04	Cleaning（U_{41}）	11.35	0.46
					Repair（U_{42}）	13.08	0.54
		Traffic safety facility（U_5）	152.50	0.24	Cleaning（U_{51}）	43.14	0.28
					Repair（U_{52}）	109.36	0.72

Table 3. Objective performance indicator system.

First level indicator	Second level indicator	Third level indicator	Cost (CNY)	Ratio (%)
Subgrade（U_1）	Cleaning（U_{11}）	Slope U_{111}	46061.24	0.172
		Drainage system U_{112}	218669.83	0.821
		Evaporation pool, cooling pool U_{113}	1759.94	0.007
	Repair（U_{12}）	Slope surface U_{121}	1001.37	0.003
		Rubble masonry U_{122}	64045.78	0.181
		Ashlar masonry U_{123}	72947.15	0.206
		Concrete structure U_{124}	60452.98	0.171
		Pointing and plastering U_{125}	102937.71	0.291
		Road shoulder U_{126}	26681.59	0.075
		Garbage removal U_{127}	25951.52	0.073

Table 4. Subgrade performance indicator system.

First level indicator	Second level indicator	Third level indicator	Cost (CNY)	Ratio (%)
Pavement（U_2）	Cleaning	Corridor U_{211}	2137323.85	0.851

	(U_{21})	Ramp U_{212}	116480.51	0.046
		Connector U_{213}	68668.66	0.027
		Toll plaza U_{214}	180337.92	0.076
	Repair (U_{22})	Crack U_{221}	295776.89	0.45
		Pothole U_{222}	309038.03	0.47
		Water barrier U_{223}	52948.26	0.08

Table 5. Pavement performance indicator system.

First level indicator	Second level indicator	Third level indicator	Cost (CNY)	Ratio (%)
Bridge & culvert (U_3)	Cleaning (U_{31})	Culvert dredging U_{311}	141317.50	0.445
		Abutment U_{312}	176166.94	0.555
	Repair (U_{32})	Guardrail painting U_{321}	6665.96	0.020
		Handrail painting U_{322}	14672.00	0.044
		Expansion joint U_{323}	130945.90	0.389
		Rubber strip replacement U_{324}	141174.81	0.418
		Concrete structure U_{325}	43314.48	0.129

Table 6. Bridge & culvert performance indicator system.

First level indicator	Second level indicator	Third level indicator	Cost (CNY)	Ratio (%)
Tunnel (U_4)	Cleaning (U_{41})	Tunnel wall U_{411}	31660.43	0.811
		Tunnel traffic safety facility U_{412}	7395.28	0.189
	Repair (U_{42})	Aluminium-plastic panel U_{421}	22323.38	0.171
		Structural painting U_{422}	108436.74	0.829

Table 7. Tunnel performance indicator system.

First level indicator	Second level indicator	Third level indicator	Cost (CNY)	Ratio (%)
Traffic safety facility (U_5)	Cleaning (U_{51})	Guardrail U_{511}	300687.97	0.648
		Anti-glare board U_{512}	91003.30	0.196
		Sign U_{513}	40357.32	0.087
		Toll booth canopy U_{514}	11786.40	0.025
		Toll booth U_{515}	7131.33	0.015
		Sound barrier U_{516}	13410.79	0.029
	Repair (U_{52})	Bumper board U_{521}	417.78	0.001
		Anti-collision barrel U_{522}	8668.62	0.010
		Sign board U_{523}	82617.76	0.100
		Sign film U_{524}	3292.88	0.004
		Hundred-meter pile (brand) U_{525}	7158.40	0.009
		Traffic barrier U_{526}	5394.50	0.007
		Isolation pier painting U_{527}	6592.19	0.008

		Folding guardrail U_{528}	5994.69	0.007
		Barrier fence U_{529}	84119.86	0.102
		Barbed wire (without column) $U_{52(10)}$	1476.00	0.002
		Barbed wire (with column) $U_{52(11)}$	157328.13	0.190
		Anti-glare board correction $U_{52(12)}$	56386.07	0.068
		Anti-glare board replacement $U_{52(13)}$	132706.91	0.161
		Anti-glare board bracket $U_{52(14)}$	10539.84	0.013
		Anti-glare board painting $U_{52(15)}$	23047.00	0.028
		Sound barrier $U_{52(16)}$	4281.47	0.005
		Aluminium gusset board $U_{52(17)}$	14334.36	0.017
		Toll booth painting $U_{52(18)}$	204230.99	0.246
		Wall painting $U_{52(19)}$	17801.78	0.022

Table 8. Traffic safety facility performance indicator system.

As shown in Fig. 5, the routine maintenance cost of STHG in 2016 demonstrates that pavement accounts for the largest proportion of routine maintenance costs, with a corresponding maximum weight. Conversely, the ratio of routine maintenance cost for the tunnel is the lowest, with the smallest corresponding weight. This observation aligns with the reality that, within the same maintenance system, the quantity of pavement assets typically exceeds that of tunnel assets, resulting in a greater need for maintenance of the former [51]. Thus, pavement maintenance performance directly affects the comprehensive evaluation of the contractor’s working ability. It is reasonable that the pavement indicator has the highest weight in the comprehensive maintenance performance evaluation. Similarly, As shown in Fig. 6, U211 accounted for the highest ratio of costs in pavement cleaning and was correspondingly given the highest weight.

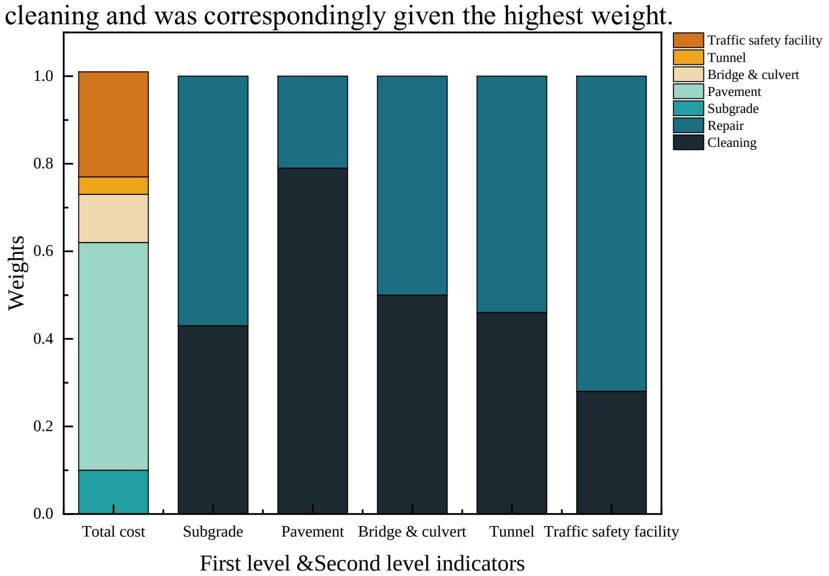

Figure 5. Weights of first level &second level indicators.

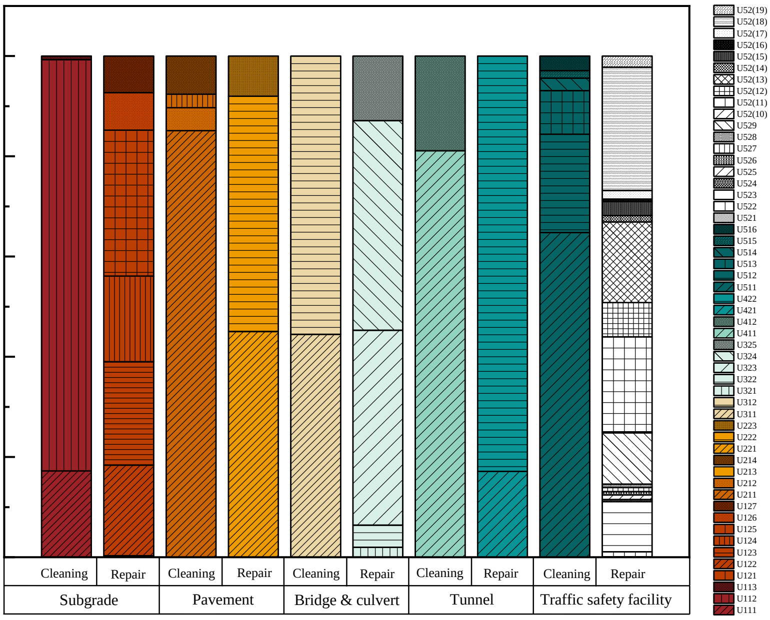

Figure 6. Weights of third level indicators.

Levels and thresholds of routine maintenance performance determination

In 2016, the number of routine maintenance items was 53, and the number of inspections was 11. As a result, the number of ratings for every area was 583 times. The frequency and cumulative frequency of various scores of the third-level indicators in the three areas were summarized, and the results are shown in Fig. 7. It can be found in Fig. 7(a), Fig. 7(b), and Fig. 7(c) that most of the highway asset condition inspection scores are in the range of 9 to 10. The scores of Qiaogouwan, Huaziping and Ansai were 90.1%, 90.2%, and 88.9% between 9 and 10, respectively. The scores of Qiaogouwan, Huaziping, and Ansai less than 5 are maintained within 3.4%, 1.6%, and 2.6%. The routine maintenance work was effectively carried out in these three areas. To better observe the urgent jump point in cumulative frequency for other scores, the 10 score is excluded due to its significantly higher occurrence rat. Finally, the result is shown in Fig. 7(d).

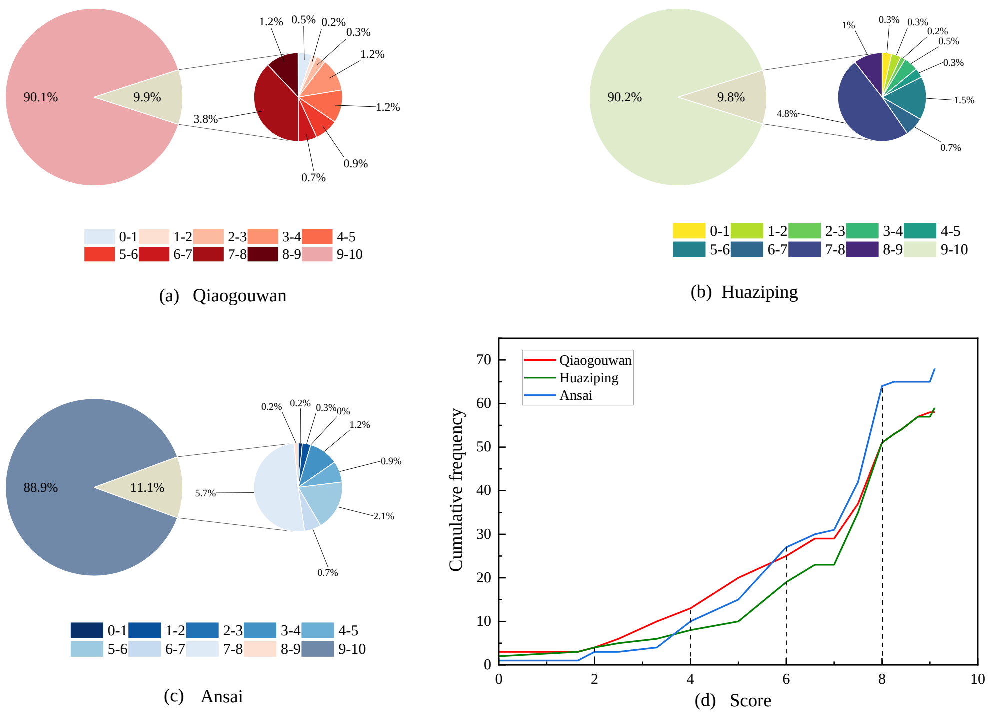

Figure 7. The score frequency statistics of routine maintenance items.

Based on Fig. 7(d), there are urgent jumps at about 2,4,6 and 8 scores. Accordingly, the LOS has been divided into five: excellent, good, poor, critical, and failed. The LOS and Albino function thresholds (λ) are shown in Table 9.

LOS	Excellent	Good	Poor	Critical	Failed
Score	8	6	4	2	0
λ	9	7	5	3	1
Interval	$\otimes_1 \in (8,10]$	$\otimes_2 \in (6,8]$	$\otimes_3 \in (4,6]$	$\otimes_4 \in (2,4]$	$\otimes_5 \in [0,2]$

Table 9. The LOS and threshold.

Albino function establishment

By using Eqs. (10) to (12), and the thresholds in Table 9, the Albino functions can be obtained and shown in Fig. 8, and the specific expression is shown in the following equations.

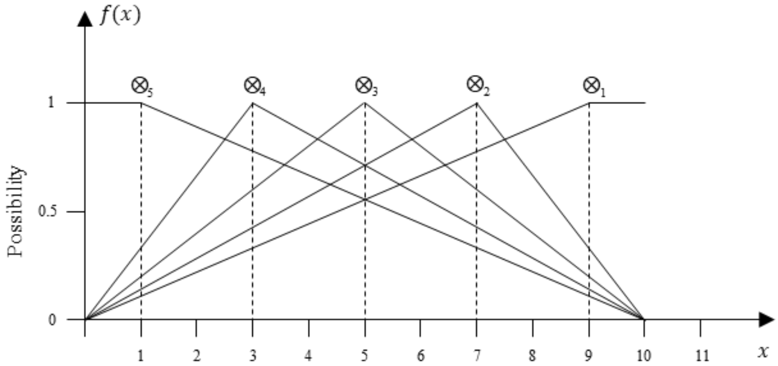

Figure 8. Albino function.

$$f_1(x)=\begin{cases}1 & 9< x\le 10\\ \dfrac{1}{9}x & 0\le x\le 9\end{cases} \qquad (15)$$

$$f_2(x)=\begin{cases}\dfrac{1}{7}x & 0\le x< 7\\ \dfrac{10-x}{10-7}=\dfrac{10-x}{3} & 7\le x\le 10\end{cases} \qquad (16)$$

$$f_3(x)=\begin{cases}\dfrac{1}{5}x & 0\le x< 5\\ \dfrac{10-x}{10-5}=\dfrac{10-x}{5} & 5\le x\le 10\end{cases} \qquad (17)$$

$$f_4(x)=\begin{cases}\dfrac{1}{3}x & 0\le x< 3\\ \dfrac{10-x}{10-3}=\dfrac{10-x}{7} & 3\le x\le 10\end{cases} \qquad (18)$$

$$f_5(x)=\begin{cases}1 & 0\le x< 1\\ \dfrac{10-x}{10-1}=\dfrac{10-x}{9} & 1\le x\le 10\end{cases} \qquad (19)$$

Routine maintenance performance evaluation

According to the Albino functions, the grey measurement values were calculated. Taking the subgrade slope indicator U_{111} of routine cleaning in Qiaogouwan as an example, the values of 11 times different inspections are substituted into 5 functions for calculation. The results are shown in Table 10. At last, the value vector r_{111} of U_{111} is $(0.800, 0.064, 0.060, 0.043, 0.033)$.

LOS	r_{111}												Value
	10	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**6**	**10**	**Σ**	
1	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(10)$	$f_1(6)$	$f_1(10)$	10.667	0.800
2	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(10)$	$f_2(6)$	$f_2(10)$	0.857	0.064
3	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(10)$	$f_3(6)$	$f_3(10)$	0.800	0.060
4	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(10)$	$f_4(6)$	$f_4(10)$	0.571	0.043
5	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(10)$	$f_5(6)$	$f_5(10)$	0.444	0.033

Table 10. The value vector r_{111} of U_{111}.

The other grey measurement value vectors of third-level indicators of Qiaogouwan were also calculated. Then, the second-level indicator measurement value matrix R_{11} is composed of value vectors of third-level indicators. As shown in flowing:

$$R_{11} = \begin{bmatrix} r_{111} \\ r_{112} \\ r_{113} \end{bmatrix} = \begin{bmatrix} 0.800 & 0.064 & 0.060 & 0.043 & 0.033 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix}$$

Based on the weight of each indicator in Table 3 and Table 4, the weight vector $A = (0.10,\ 0.52,\ 0.11,\ 0.04,\ 0.24)$, $A_1 = (0.43, 0.57)$, and $A_{11} = (0.172,\ 0.821, 0.007)$. Using Eqs (4) to (8), these can be calculated: $B_{11} = A_{11}R_{11} = (0.966, 0.011, 0.010, 0.007, 0.006)$, $B_1 = A_1R_1 = (0.948, 0.020, 0.014, 0.008, 0.008)$, and $B = AR = (0.685, 0.094, 0.089, 0.077, 0.065)$. Afterward, the grey evaluation value of the objective level W is:

$$W = B\lambda^T = (0.685, 0.094, 0.089, 0.077, 0.065)\begin{pmatrix} 9 \\ 7 \\ 5 \\ 3 \\ 1 \end{pmatrix} = 7.564$$

$$f_1(W) = 0.840 \quad f_2(W) = 0.812 \quad f_3(W) = 0.487 \quad f_4(W) = 0.348 \quad f_5(W) = 0.271$$

These values show that the probability of being excellent level in routine maintenance performance in Qiaogouwan is 84%. The same steps can be applied in Huaziping and Ansai. At last, in Huaziping: $f_1(W) = 0.869$ $f_2(W) = 0.727$ $f_3(W) = 0.436$ $f_4(W) = 0.312$ $f_5(W) = 0.242$, and in Ansai: $f_1(W) = 0.834$ $f_2(W) = 0.830$ $f_3(W) = 0.498$ $f_4(W) = 0.356$ $f_5(W) = 0.277$.

Based on the measurement results, it is evident that the possibility of reaching an excellent level is highest in the three areas. However, it is important to acknowledge that the possibility of achieving a good level in Qiaogouwan and Ansai is also considerably high, nearly on par with the excellent level. This suggests that the judgment of an excellent level is unstable, and the contractor should enhance their maintenance techniques to ensure consistent excellence. Referring to Fig. 7, the highway condition scores with the highest frequency of 9-10 are observed in Huaziping, followed by Qiaogouwan, and the lowest in Ansai. This comparison aligns with the probability of the three regions achieving an excellent level, leading to the same conclusion.

Validation

The comparison analysis was initially employed to validate the grey theory evaluation method for assessing maintenance performance LOS. This involved comparing the evaluation results against inspection data and expert scores to ascertain its feasibility. Subsequently, the thresholds of the grey theory method were adjusted, and the method's sensitivity was analyzed to confirm its robustness.

Comparison analysis

To verify the results, three project staff were invited to assign scores to the performance LOS of maintenance items in Tables 3 to 8. The scoring scale ranged from 0 to 10, with scores falling into the following classifications: 0-2 denoted as Failed, 2-4 as Critical, 4-6 as Poor, 6-8 as Good, and 8-10 as Excellent. Average scores were calculated, and the corresponding results are shown in Table 11. A comparison was made between performance LOS calculated based on inspection data, expert scores, and grey theory. Table 11 illustrates that the performance LOS for all three areas falls within the excellent level. While the evaluation results derived from inspection data or expert scores are fixed values, the evaluation results obtained through grey theory are represented by a grey number [18]. This grey number signifies that the true value fluctuates around the value. The probability of achieving an excellent performance LOS closely aligns with the percentage derived from the inspection data and expert scores. This observation suggests the feasibility of employing grey theory to evaluate maintenance performance LOS, with the calculated performance LOS probability by grey theory demonstrating similarity to the actual performance LOS probability distribution.

Area	Performance LOS	Inspection	Expert 1	Expert 2	Expert 3	Grey theory
Qiaogouwan	Average/Comprehensive score	8.985	7.95	8.532	8.995	7.564
	Percentage/Possibility of excellent level	77.7%	50.9%	63.2%	90%	84%
Huaziping	Average/Comprehensive score	9.110	8.105	8.582	9.036	7.819
	Percentage/Possibility of excellent level	77.3%	55.5%	61.8%	90.5%	86.9%
Ansai	Average/Comprehensive score	8.834	7.973	8.536	8.955	7.509
	Percentage/Possibility of excellent level	70.5%	55.5%	62.3%	91.4%	83.4%

Table 11. Performance level comparison between inspection, expert, and grey theory

Sensitivity analysis

To study the influence of the threshold on the measurement results, sensitivity analysis was carried out. Based on the above findings, there are 8, 8.5, and 9 three kinds of thresholds λ between good level and excellent level were set, respectively. In this section, all the input data are the same as the sources mentioned above except for the values of λ. The results are given in Fig. 9. It shows that the results of both 9 and 9.5 as the thresholds for measurement are consistent in these three areas and all of them are excellent. This suggests that taking 9 as the threshold between a good level and an excellent level can ensure the stability of measurement results. On the contrary, if 8.5 is the threshold, the maintenance performance of Huaziping is excellent, while both Ansai and Qiaogouwan are good. It implies that the change of λ value is sensitive to the results of measurement. Additionally, this phenomenon supports the previous discussion that both Ansai and Qiaogouwan need to improve their maintenance performance. In summary, it can be found that this method of measuring routine maintenance performance is reasonable.

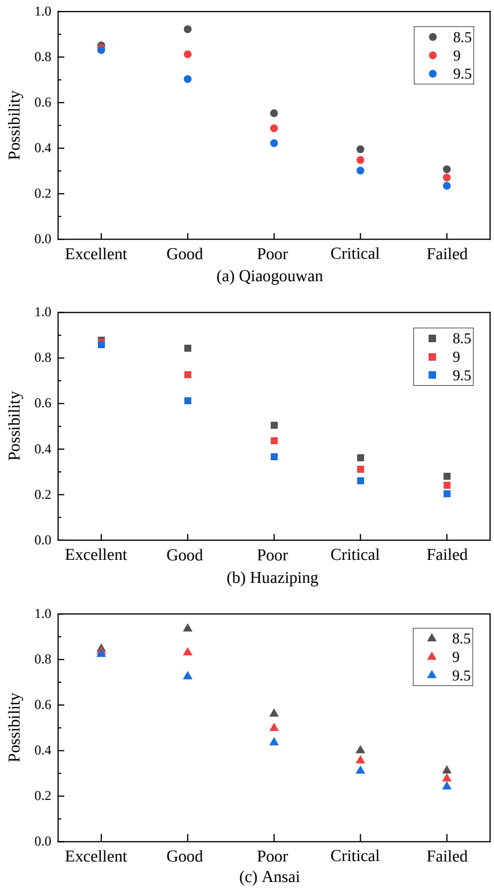

Figure 9. Sensitivity analysis results for the measurement model.

Conclusions

To help contractors improve the maintenance performance of highway routine maintenance and ensure that client requirements are met, this study adopted the grey theory to build a dynamic performance measurement model. Afterward, the 11 times inspection data in Qiaogouwan, Huaziping, and Ansai were taken as examples to carry out empirical analysis. Finally, through a comparison analysis and sensitivity analysis, the rationality of employing this method for performance evaluation was verified.

The contributions can be summarized as follows. First, this study provides an innovative and comprehensive framework to assess routine maintenance performance by systematically integrating items of all routine maintenance occurring in the maintenance area. In addition, the weights of performance indicators are determined by the ratio of maintenance costs corresponding to each indicator. In particular, the thresholds of performance LOS are adopted by the graph of cumulative frequency distribution of highway condition scores. Besides, the grey theory is used to estimate the probabilities that the maintenance performance of an area reaches different levels based on the performance indicator system and inspection data. In summary, these quantitative methods overcome the subjective arbitrariness in previous studies. Due to the bill of quantities varying from year and region, the measurement framework allows for dynamic adjustments. Consequently, it facilitates the continuous enhancement of highway conditions. Moreover, this study provides a feasible measurement method for contractors intending to implement PBC, which contributes to the development of PBC and saves maintenance costs.

Some limitations of this research and potential research directions should be acknowledged. Firstly, this paper does not establish specific thresholds for the LOS of each maintenance item. In future studies, emphasis should be placed on refining the third-level indicators by introducing fourth-level indicators, which might encompass aspects such as structural and functional integrity, defect severity, repair time limits, and other relevant factors. Therefore, it is crucial to pay attention to collecting refined data and further enhancing the performance measurement system.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Straub, A. Cost savings from performance-based maintenance contracting. *Int. J. Strateg. Prop. Manag.* **13**, 205–217 (2009).
2. Liinamaa, J. *et al.* Performance-based and functional contracting in value-based solution selling. *Ind. Mark. Manag.* **59**, 37–49 (2016).
3. Selviaridis, K. & Wynstra, F. Performance-based contracting: a literature review and future research directions. *Int. J. Prod. Res.* **53**, 3505–3540 (2015).
4. Shrestha, K. & Shrestha, P. P. Framework development of performance-based striping maintenance contracts. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* **12**, 04520014 (2020).
5. Gelderman, Semeijn, & Vries. Contracting for road maintenance in the Netherlands—the downside of performance-based contracting. *Infrastructures* **4**, 41 (2019).
6. Ma, Y. L. Evaluation of highway pavement maintenance effect and study of effective road maintenance strategies in Guangdong Province based on DEA. (South China University of Technology, 2018).
7. Fallah-Fini, S., Rahmandad, H., Triantis, K. & De La Garza, J. M. Optimizing highway maintenance operations: dynamic considerations. *Syst. Dyn. Rev.* **26**, 216–238 (2010).
8. Fallah-Fini, S., Triantis, K., De La Garza, J. M. & Seaver, W. L. Measuring the efficiency of highway maintenance contracting strategies: A bootstrapped non-parametric meta-frontier approach. *Eur. J. Oper. Res.* **219**, 134–145 (2012).
9. Sasmal, S. & Ramanjaneyulu, K. Condition evaluation of existing reinforced concrete bridges using fuzzy based analytic hierarchy approach. *Expert Syst. Appl.* **35**, 1430–1443 (2008).
10. Tatari, O., Sargand, S. M., Masada, T. & Tarawneh, B. Neural Network Approach to Condition Assessment of Highway Culverts: Case Study in Ohio. *J. Infrastruct. Syst.* **19**, 409–414 (2013).
11. Aktan, A. E. *et al.* Condition Assessment for Bridge Management. *J. Infrastruct. Syst.* **2**, 108–117 (1996).
12. Abu Dabous, S., Zeiada, W., Al-Ruzouq, R., Hamad, K. & Al-Khayyat, G. Distress-based evidential reasoning method for pavement infrastructure condition assessment and rating. *Int. J. Pavement Eng.* **22**, 455–466 (2021).

13. Su, H. *et al.* Research on safety condition assessment methodology for single tower steel box girder suspension bridges over the sea based on improved AHP-fuzzy comprehensive evaluation. *Sci. Rep.* **14**, 12079 (2024).
14. Ouma, Y. O., Opudo, J. & Nyambenya, S. Comparison of fuzzy AHP and fuzzy TOPSIS for road pavement maintenance prioritization: Methodological exposition and case study. *Adv. Civ. Eng.* **2015**, 1–17 (2015).
15. Sun, L. & Gu, W. Pavement condition assessment using fuzzy logic theory and analytic hierarchy process. *J. Transp. Eng.* **137**, 648–655 (2011).
16. Singh, A. P., Sharma, A., Mishra, R., Wagle, M. & Sarkar, A. K. Pavement condition assessment using soft computing techniques. *Int. J. Pavement Res. Technol.* **11**, 564–581 (2018).
17. Ju-Long, D. Control problems of grey systems. *Syst. Control Lett.* **1**, 288–294 (1982).
18. Khuman, A. S. The similarities and divergences between grey and fuzzy theory. *Expert Syst. Appl.* **186**, 115812 (2021).
19. Yu, J., Zhang, X. & Xiong, C. A methodology for evaluating micro-surfacing treatment on asphalt pavement based on grey system models and grey rational degree theory. *Constr. Build. Mater.* **150**, 214–226 (2017).
20. Tan, Y., Zhang, Z., Wang, H. & Zhou, S. Gray relation analysis for optimal selection of bridge reinforcement scheme based on Fuzzy-AHP weights. *Math. Probl. Eng.* **2021**, 1–8 (2021).
21. Duran, A. Comparison of performance-based and traditional road maintenance contracts. *Zb. Rad. Građev. Fak.* **37**, 555–562 (2021).
22. Stankevich, N., Qureshi, N. & Queiroz, C. Performance-based contracting for preservation and improvement of road assets. (2005).
23. Silva, M., Marcela. Performance-based road rehabilitation and maintenance contracts (CREMA) in Argentina. (2011).
24. Anwar, M. F. *et al.* Implementation of performance based contracting in Malaysia. **6**, (2016).
25. Sultana, M., Rahman, A. & Chowdhury, S. A review of performance based maintenance of road infrastructure by contracting. *Int. J. Product. Perform. Manag.* **62**, 276–292 (2013).
26. Abu, S. S., Osman, H. & Hosny, O. Optimal maintenance and rehabilitation policies for performance-based road maintenance contracts. *J. Perform. Constr. Facil.* **31**, 04016075 (2017).

27. Fuller, J., Brown, C. J. & Crowley, R. Performance-based maintenance contracting in Florida: Evaluation by surveys, statistics, and content analysis. *J. Constr. Eng. Manag.* **144**, 05017021 (2018).

28. Ozbek, M. E. & De La Garza, J. M. Comprehensive evaluation of Virginia department of transportation's experience with its first performance-based road-maintenance contract. *J. Transp. Eng.* **137**, 845–854 (2011).

29. Alyami, Z. & Tighe, S. L. Development of maintenance and rehabilitation program: Pavement assets under performance-based contracts. *Transp. Res. Rec. J. Transp. Res. Board* **2361**, 1–10 (2013).

30. Rouse, P. Towards a general managerial framework for performance measurement: A comprehensive highway maintenance application. **8**, 127–149 (1997).

31. Xiao, Y. & Zhang, Y. F. Construction of highway asphalt pavement maintenance comprehensive post-evaluation system. *E3S Web Conf.* **439**, 01001 (2023).

32. Fallah-Fini, S., Triantis, K., Rahmandad, H. & De La Garza, J. M. Measuring dynamic efficiency of highway maintenance operations. *Omega* **50**, 18–28 (2015).

33. Setianingsih, A. I., Sangaji, S. & Setyawan, A. Road maintenance and rehabilitation program using functional and structural assessment. *IOP Conf. Ser. Mater. Sci. Eng.* **176**, 012030 (2017).

34. Yuan, J. B. & Zhu, W. X. Research on the content system of post-evaluation of highway construction projects. *highway* 92-95+125 (2001).

35. Liu, F. J., Lu, J., Zhang, G. Q., & Xiang, Q. J. Research on evaluation method of expressway subgrade maintenance. *J. Highw. Transp. Res. Dev.* 39–42 (2006).

36. Ma, Y. F., Jian, J. L., Zhang, W. B., & Zhao, J. G. Comprehensive model for highway bridge maintenance quality evaluation in Jiangsu province, China. *Transp. Res. Rec. J. Transp. Res. Board* **2292**, 150–159 (2012).

37. Haas, R., Tighe, S. & Yeaman, J. Long term warranty provisions for sustained preservation of pavement networks. (2008).

38. Haas, R., Felio, G., Lounis, Z. & Cowe Falls, L. Measurable performance indicators for roads Canadian and international practic. 1–22 (2009).

39. Xie, Y. C. Research on asphalt pavement maintenance mode based on performance appraisal.

(Southwest University, 2022).

40. Lu, Y. F., Ren, H.X., Li, Fu, Feng, Y. G., & Ning, X. W. Talking about the evaluation index of highway maintenance. *Technol. Econ. Areas Commun.* 11–12 (2000) doi:10.19348/j.cnki.issn1008-5696.2000.02.006.
41. Marcelino, P., Lurdes Antunes, M. D. & Fortunato, E. Comprehensive performance indicators for road pavement condition assessment. *Struct. Infrastruct. Eng.* **14**, 1433–1445 (2018).
42. Shah, Y. U., Jain, S. S. & Parida, M. Evaluation of prioritization methods for effective pavement maintenance of urban roads. *Int. J. Pavement Eng.* **15**, 238–250 (2014).
43. Zheng, X. Y., Easa, S. M., Yang, Z. X., Ji, T. & Jiang, Z. L. Life-cycle sustainability assessment of pavement maintenance alternatives: Methodology and case study. *J. Clean. Prod.* **213**, 659–672 (2019).
44. Lu, J. *Evaluation and Prediction Technology of Comprehensive Maintenance of High Grade Highway*. (2010).
45. Zakeri, S., Chatterjee, P., Konstantas, D. & Ecer, F. A decision analysis model for material selection using simple ranking process. *Sci. Rep.* **13**, 8631 (2023).
46. Gutman, A. J. & Goldmeier, J. *Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning*. (John Wiley & Sons, 2021).
47. Kulkarni, R. B., Golabi, K., Finn, F. N., Johnson, R. & Consultants, W.-C. A systematic procedure for the development of maintenance of severice. *Transp. Res. Rec.*
48. Alsharqawi, M. Integrated decision support methodology for bridge deck management under performance-based contracting. (2018).
49. Qiao, Y., Fricker, J. D. & Labi, S. Quantifying the similarity between different project types based on their pay item compositions: Application to bundling. *J. Constr. Eng. Manag.* **145**, 04019053 (2019).
50. Liu, K., Shen, X. L., Tan, Z. F. & Guo, W. Y. Grey clustering analysis method for overseas energy project investment risk decision. *Syst. Eng. Procedia* **3**, 55–62 (2012).
51. Wu, X., Shi, X., Li, Y. & Gong, X. Estimation of annual routine maintenance cost for highway tunnels. *Adv. Civ. Eng.* **2022**, 1–14 (2022).

Acknowledgments
This research was supported by the Natural Science Basic Research Program of Shaanxi Province, No.2022 JM-307, China Scholarship Council (202306560107), and China Scholarship Council (202206560047). The authors sincerely acknowledge the data support from Shaanxi Transportation Holding Group.

Author contributions
Concepts and design: Xiaoli Shi. Acquisition, analysis, or interpretation of the data: Xuelian Wu, Yuhuan Li, and Xiaotian Gong. Drafting of the manuscript: Xuelian Wu and Yuhuan Li. Critical revision of the manuscript for important intellectual content: Xiaoli Shi, Yongjian Ke, Malongzhi Wan, and Yuhuan Li.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to X.S.