A Framework for Evaluating the Context-Dependent Performance of Intelligent Construction Technologies

1st Guang Yang
Department of Civil Engineering
Tsinghua University
Beijing, China
yangguang@mails.tsinghua.edu.cn

2nd Wen Liu
State Key Laboratory of Hydroscience and Engineering
Tsinghua University
Beijing, China
liuw92@foxmail.com

3rd Jia-Rui Lin
Department of Civil Engineering
Tsinghua University
Beijing, China
lin611@tsinghua.edu.cn

4th Peng Pan*
Department of Civil Engineering, State Key Laboratory of Hydroscience and Engineering, Key Laboratory of Digital Construction and Digital Twin
Tsinghua University
Beijing, China
panpeng@tsinghua.edu.cn

Abstract—**The adoption of Intelligent Construction (IC) technologies, including construction robots, is pivotal for the architecture, engineering, and construction (AEC) industry's transformation, yet their performance often varies significantly across projects—a challenge known as context dependency. Existing evaluation methods, focused on single-case studies, fail to capture this variance, hindering reliable investment decisions. To address this gap, we propose a novel "Expectation-Robustness" framework. The framework first utilizes a multi-criteria benefit evaluation system to quantify a technology's performance in a single project into a standardized Overall Benefit Score (OBS). Then, by treating the OBS scores from multiple projects as a statistical sample, we define two key metrics: (1) Expected Benefit, measured by the mean of OBS, representing the technology's average performance; and (2) Performance Robustness, measured by the inverse of the Coefficient of Variation (CV) of OBS, representing its stability across diverse contexts. We validated the framework using a large-scale dataset of 678 technology application cases from national IC pilot projects in China. An analysis of 10 typical technologies, including various construction robots, revealed distinct performance profiles. Our framework moves beyond simplistic single-point estimates, offering a more scientific basis for stakeholders to make informed decisions on technology investment, research, development, and deployment, thereby promoting the sustainable development of intelligent construction.**

Keywords- Intelligent construction; comprehensive benefit; context dependency; construction robots; empirical research

I. INTRODUCTION

The widespread adoption of Intelligent Construction (IC) technologies, ranging from Building Information Modeling (BIM) and digital twins to various construction robots, is widely recognized as the core driver for the digital transformation of the architecture, engineering, and construction (AEC) industry [1]. These technologies promise to address long-standing industry challenges, such as low productivity, high safety risks, and labor shortages, by enhancing efficiency, quality, and sustainability [2]. Consequently, construction firms are increasingly investing in a diverse portfolio of IC technologies. However, selecting the right technology and justifying the investment remain significant hurdles for decision-makers.

A primary challenge lies in the robust evaluation of a technology's benefits. While numerous studies have demonstrated the positive impacts of specific IC technologies through case studies, their findings are often difficult to generalize. This is due to a critical, yet frequently overlooked, phenomenon: context dependency. The performance and benefits of an IC technology are not inherent properties but are heavily influenced by the specific project environment, including project type, scale, organizational maturity, and team expertise. For instance, a sophisticated robotic system that yields remarkable productivity gains on a large-scale, standardized project may prove inefficient and costly on a smaller, more complex one. This performance variability creates significant uncertainty and risk for potential adopters.

The existing literature on technology evaluation predominantly suffers from two limitations in addressing this issue [3], [4], [5], [6]. First, most studies rely on single-project case analyses or small-scale pilot tests, which, while providing in-depth insights, lack the statistical power to reveal general performance trends and variances. Second, there is a scarcity of evaluation frameworks that systematically incorporate performance volatility into the assessment. Traditional metrics like Return on Investment (ROI) typically provide a single-point estimate of benefits, failing to quantify the risk or robustness of achieving those benefits across different scenarios. This leaves a critical gap: decision-makers lack a tool to distinguish between a versatile "workhorse" technology that performs consistently well and a context-sensitive "specialist" technology that excels only under specific conditions.

To bridge this gap, this paper proposes a novel "Expectation-Robustness" framework for the cross-project evaluation of IC technologies. This framework moves beyond single-point benefit assessment by characterizing each technology using two fundamental dimensions: (1) its Expected Benefit, representing the average performance level across multiple applications, and (2) its Performance Robustness, quantifying the consistency and stability of that performance. By validating this framework with a large-scale empirical dataset from China's national IC pilot projects, this study aims to:

(1) Introduce and formalize a multi-stage and two-dimensional methodology for evaluating and comparing IC technologies.

(2) Empirically classify a set of 10 typical technologies, including construction robots, into distinct performance profiles based on the framework.

(3) Provide actionable insights for industry stakeholders to make more strategic and risk-informed decisions regarding technology adoption, investment, and development.

II. METHODS

To systematically evaluate the context-dependent performance of IC technologies, we developed a multi-stage methodology. First, we established a multi-criteria benefit evaluation system to serve as a standardized measurement tool. Second, we proposed our core theoretical contribution: the "Expectation-Robustness" framework for cross-project assessment. Finally, we collected a large-scale dataset to validate the framework.

A. The Multi-Criteria Benefit Evaluation System

To quantify the multifaceted benefits of an IC technology within a single project, we adopted a comprehensive evaluation system developed in our prior research. This system is built upon a three-tiered hierarchical structure based on Multi-Criteria Decision Making (MCDM) theory [7].

1) Objective Layer: The top level is the Overall Benefit Score (OBS), a normalized index ranging from 0 to 1 that represents the technology's aggregate performance in a project.

2) Criteria Layer: The OBS is decomposed into three core criteria consistent with sustainable development principles: Economic Benefit (E), Social Benefit (S), and Environmental Benefit (V). Through an Analytic Hierarchy Process (AHP) survey involving 35 experts from industry, academia, and government, the weights for these criteria were determined as 0.4223 (E), 0.2390 (S), and 0.3387 (V), respectively.

3) Indicator Layer: The three criteria are underpinned by 22 specific, measurable indicators. These indicators were systematically mined from a large corpus of policy documents, academic literature, and project case studies using Natural Language Processing (NLP) techniques to ensure objectivity and comprehensiveness. The contribution weight of each indicator to its corresponding criterion was also determined via AHP.

The complete system, including all indicators and their weights, is presented in Table I.

TABLE I. THE MULTI-CRITERIA BENEFIT EVALUATION SYSTEM WITH INDICATOR WEIGHTS

No	**Indicators**	**Contribution weight of each indicator to:**		
		Economic Benefit	***Social Benefit***	***Environmental Benefit***
1	Reduce construction costs	0.2823	——	——
2	Reduce operation & maintenance costs	0.2210	——	——
3	Shorten construction duration	0.1614	0.0157	——
4	Improve labor productivity	0.1035	——	——
5	Reduce change orders and rework	0.0912	——	——
6	Enhance resource utilization efficiency	0.0485	——	0.1522
7	Shorten investment payback period	0.0433	——	——
8	Reduce water and electricity consumption	0.0227	——	0.2159
9	Decrease reliance on manual labor	0.0155	0.0814	——
10	Reduce design costs	0.0106	——	——
11	Mitigate safety risks	——	0.2068	——
12	Improve workers' occupational health	——	0.1914	——
13	Enhance building performance	——	0.1077	——
14	Improve construction quality	——	0.0898	——
15	Foster technology development & diffusion	——	0.0855	——
16	Improve design quality	——	0.0775	——
17	Enhance communication & coordination efficiency	——	0.0738	——
18	Improve operation and maintenance efficiency	——	0.0704	——
19	Reduce life-cycle carbon emissions	——	——	0.2504
20	Reduce industrial waste	——	——	0.1302
21	Reduce dust emissions	——	——	0.1296
22	Reduce noise pollution	——	——	0.1217

The final OBS for a technology in a single project is calculated through a bottom-up aggregation process, using the following formula:

$$OBS = \omega_E \cdot S_E + \omega_S \cdot S_S + \omega_V \cdot S_V \qquad (1)$$

where $S_j = \sum_{i=1}^{22}(\omega_{ij} \cdot s_i)$ for $j \in \{E, S, V\}$. In these formulas, s_i is the normalized score of indicator i, ω_{ij} is the contribution weight of indicator i to criterion j, and ω_j is the weight of criterion j.

B. The "Expectation-Robustness" Framework

While the OBS provides a robust measure for a single-project context, it fails to capture the performance variability across different projects. To address this, we propose the "Expectation-Robustness" framework, which models a technology's performance as a random variable. By collecting a set of OBS observations $\{OBS_1, OBS_2, \ldots, OBS_n\}$ from n different projects, we can analyze the statistical properties of its performance distribution.

We define two key metrics to characterize a technology's cross-project performance profile:

1) Expected Benefit: This metric represents the technology's average performance level and is estimated by the sample mean of the OBS scores. A higher mean indicates a greater expected return from adopting the technology.

$$\text{Expected Benefit} = \bar{x}_{OBS} = \frac{1}{n}\sum_{i=1}^{n} OBS_i \qquad (2)$$

2) Performance Robustness: This metric quantifies the consistency and stability of the technology's performance. We define a Robustness Index as the inverse of the Coefficient of Variation (CV) of the OBS scores. A higher Robustness Index signifies that the technology's benefits are less sensitive to contextual factors, indicating lower risk and higher predictability.

$$\text{Robustness Index} = \frac{1}{CV} = \frac{\bar{x}_{OBS}}{\sigma_{OBS}} \qquad (3)$$

where σ_{OBS} is the sample standard deviation of the OBS scores.

By plotting technologies on a two-dimensional coordinate system with Expected Benefit as the x-axis and the Robustness Index as the y-axis, we can classify them into four distinct categories, providing a powerful decision-making tool for stakeholders (Fig. 1). Each quadrant represents a unique technology profile.

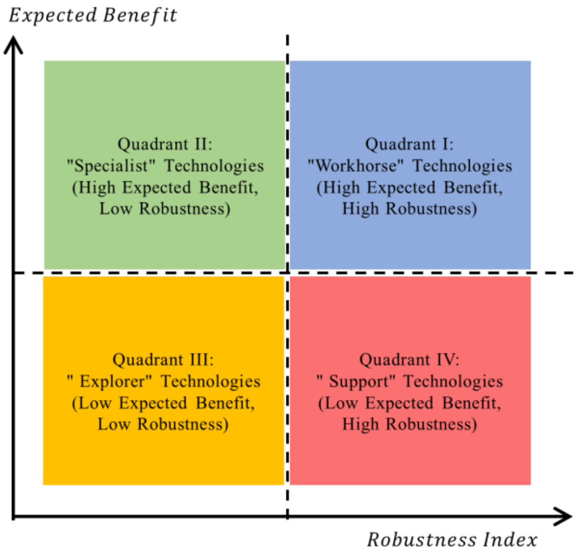

Figure 1. The "Expectation-Robustness" framework for classifying IC technologies.

a) Quadrant I (High Expected Benefit, High Robustness): "Workhorse" technologies. These are highly reliable and valuable technologies that consistently deliver significant benefits across a wide range of project contexts. They represent mature, low-risk, high-reward solutions that are prime candidates for widespread adoption.

b) Quadrant II (High Expected Benefit, Low Robustness): "Specialist" Technologies. These technologies offer substantial potential benefits but are highly sensitive to the project context. Their success is contingent on specific enabling conditions (e.g., project scale, team expertise). They are high-reward but also high-risk, requiring careful feasibility analysis before deployment.

c) Quadrant III (Low Expected Benefit, Low Robustness): "Explorer" Technologies. These technologies exhibit both low average benefits and high performance variability. This profile is often characteristic of immature or emerging technologies that have not yet found their optimal application niche. They represent high-risk options that require further research and development before they are ready for broad implementation.

d) Quadrant IV (Low Expected Benefit, High Robustness): "Support" Technologies. These technologies provide modest but consistent benefits. They are typically low-risk, point solutions that address niche problems. While not transformative, they can be valuable as reliable, incremental improvements, especially in risk-averse environments.

C. *Data Collection and Processing*

To empirically validate our framework, we collected data through a large-scale survey conducted in collaboration with the Ministry of Housing and Urban-Rural Development of China, targeting national IC pilot projects.

1) Data Collection Instrument: A structured questionnaire, the Intelligent Construction Pilot Project Survey Form, was designed to capture detailed information on the application of a single IC technology in a specific project. The form consists of five sections: (1) technology identification, (2) application context, (3) benefit analysis, (4) conclusions and suggestions, and (5) project information. The crucial "Benefit Analysis" section employed a mixed-method approach, requiring respondents to provide both a qualitative rating (a 1-10 "benefit fulfillment" score) and detailed quantitative evidence and descriptions.

2) Data Sample: We received 678 valid survey responses from 21 provinces in China, covering a wide range of project types (e.g., residential, public, industrial) and investment scales. This rich dataset provides a representative snapshot of the current state of IC technology application in China.

3) Data Processing to OBS: A systematic procedure was established to convert the raw survey data into the standardized indicator scores (s_i) required by our evaluation system. The process prioritized quantitative data.

a) For quantitative indicators: We first mapped the narrative and numerical data from the survey's "Benefit Analysis" section to the 22 indicators in our system. For "cost-type" indicators (e.g., construction cost), the score was calculated based on the reduction rate:

$$s_i = \frac{B-A}{B} \tag{4}$$

where B is the baseline value and A is the actual value after technology adoption. For "benefit-type" indicators (e.g., labor productivity), the score was calculated based on the degree of improvement relative to a target:

$$s_i = \frac{A-B}{T-B} \tag{5}$$

where T is the target value.

b) For qualitative indicators: When quantitative data was unavailable or the indicator was inherently qualitative (e.g., improving design quality), a five-level rating scale (Excellent=0.9, Good=0.7, Medium=0.5, Poor=0.3, No Improvement=0.1) was used. The rating was determined by interpreting the qualitative descriptions in the survey, cross-referenced with the respondent's 1-10 fulfillment score.

c) If a technology had no discernible impact on a specific indicator, its score was set to 0. This rigorous data processing pipeline ensures that the raw, heterogeneous field data is transformed into consistent, comparable OBS inputs for our framework.

III. RESULTS

To validate the proposed "Expectation-Robustness" framework, we applied it to the collected dataset. We selected 10 typical IC technologies that had sufficient sample sizes and were representative of different application domains. The analysis reveals significant variations in their cross-project performance profiles, demonstrating the framework's diagnostic power.

For each of the 10 selected technologies, we first calculated the Overall Benefit Score (OBS) for every application case according to the methodology described in Section II. Subsequently, we computed the mean of OBS (Expected Benefit) and the Coefficient of Variation (CV) for each technology's OBS sample. The final evaluation results, including the calculated Robustness Index (1/CV), are summarized in Table II.

TABLE II. CROSS-PROJECT PERFORMANCE EVALUATION RESULTS FOR 10 TYPICAL IC TECHNOLOGIES

Technology	***Expected Benefit (Mean of OBS)***	***Std. Dev. of OBS***	***CV (%)***	***Robustness Index (1/CV)***
BIM Forward Design	0.425	0.169	39.86%	2.51
Smart Production of Precast Concrete (PC) Elements	0.388	0.232	59.75%	1.67
Smart Site Management Platform	0.511	0.104	20.42%	4.90
Smart Tower Crane Technology	0.224	0.056	25.10%	3.98
Smart Construction Hoist Technology	0.503	0.034	6.84%	14.62
Wall-Finishing Construction Robot	0.386	0.133	34.37%	2.91
Floor-Finishing Construction Robot	0.405	0.080	19.72%	5.07
Tile-Laying Robot	0.205	0.143	69.57%	1.44
3D Simulation and Roaming	0.112	0.020	17.68%	5.66
Structural Health Monitoring (SHM)	0.358	0.162	45.38%	2.20

To classify the technologies objectively, we defined quantitative thresholds based on the data distribution: technologies with an Expected Benefit ⩾ 0.3 were considered high-benefit, and those with a Robustness Index ⩾ 4.0 were considered high-robustness. Based on their positions in the framework, the technologies are classified as follows, with insights drawn directly from the analysis of survey responses.

1) "Workhorse" technologies (Expected Benefit ⩾ 0.3, Robustness Index ⩾ 4.0): The Smart Site Management Platform, Smart Construction Hoist, and Floor-Finishing Construction Robot are archetypal workhorse technologies. The Smart Construction Hoist, in particular, exhibits the highest robustness (Index=14.62) along with a very high expected benefit (OBS=0.503). Our survey data shows that these technologies address common project pain points (e.g., site management, vertical transportation, flooring) in a standardized manner, making their benefits both significant and highly repeatable.

2) "Specialist" technologies (Expected Benefit ⩾ 0.3, Robustness Index < 4.0): BIM Forward Design, Smart Production of PC Elements, the Wall-Finishing Robot, and SHM are identified as specialist technologies. They all offer substantial average benefits but have low robustness. For example, qualitative feedback from the surveys consistently indicates that BIM Forward Design's effectiveness is highly dependent on team proficiency and inter-disciplinary collaboration. Similarly, the benefits of Smart Production of PC Elements were reported to be much higher on large-scale projects with a high degree of repetition.

3) "Explorer" technologies (Expected Benefit < 0.3, Robustness Index < 4.0): The Tile-Laying Robot and Smart Tower Crane Technology are classified as explorer technologies. The Tile-Laying Robot, with the lowest robustness in our sample (Index=1.44), exemplifies an early-stage technology. Survey responses often highlighted challenges in adapting to complex site conditions and the high level of debugging required, explaining its unpredictable performance.

4) "Support" technologies (Expected Benefit < 0.3, Robustness Index ⩾ 4.0): Only 3D Simulation and Roaming fits this profile, providing modest but reliable benefits. Respondents frequently reported its value in client communication and virtual mock-ups, which are useful but

have a limited direct impact on core project metrics like cost and schedule.

IV. DISCUSSION AND CONCLUSION

This study set out to address the critical challenge of context dependency in evaluating Intelligent Construction (IC) technologies. By proposing and empirically validating a novel "Expectation-Robustness" framework, we have demonstrated that assessing technologies based on both their average benefit and cross-project stability provides a more nuanced and powerful approach than traditional methods. Our classification of 10 typical technologies into distinct profiles—Workhorse, Specialist, Support, and Explorer—using a large-scale dataset from China offers several key takeaways for both theory and practice.

A central finding from our analysis is that a technology's robustness appears strongly linked to the standardization of its application scenario. The "Workhorse" technologies, like the Smart Construction Hoist and Floor-Finishing Robot, excel in highly structured and repetitive environments, leading to predictable and consistent benefits. In contrast, technologies classified as "Specialist" or "Explorer," such as BIM Forward Design or the Tile-Laying Robot, must contend with more variable conditions—from the "soft" dynamics of team collaboration to the "hard" realities of site imperfections. This insight suggests that the pathway to technological maturity in construction involves not just enhancing a technology's core function, but critically, increasing its adaptability to the industry's inherently unstructured settings.

The theoretical contribution of this work is twofold. First, it introduces a new paradigm for technology evaluation in construction that formally incorporates risk and uncertainty. Moving beyond single-point, deterministic metrics, our framework provides a quantifiable method to capture the performance variability that practitioners intuitively understand but have lacked the tools to measure. Second, this study offers a replicable, data-driven methodology for creating an empirical technology typology, providing a standardized vocabulary for future research on technology diffusion and innovation lifecycles in the AEC industry.

From a practical standpoint, the "Expectation-Robustness" framework serves as a strategic decision-making tool. For technology adopters, it enables a portfolio approach to IC investment, aligning technology choices with risk appetite. For technology developers, it pinpoints clear research and development priorities, shifting the focus from simply maximizing potential benefits to also ensuring performance consistency. For policy makers, it allows for the formulation of more targeted promotion strategies tailored to the maturity profile of different technologies.

While this study provides a robust foundation, its limitations open avenues for future research. The dataset, while extensive, is from Chinese pilot projects; its generalizability warrants further investigation in other international contexts. Furthermore, future work could aim to decouple the sources of performance variance by building regression models that explicitly link specific project characteristics to a technology's OBS. The synergistic effects of technology combinations, intentionally excluded for focus, also remain a rich area for subsequent analysis.

In conclusion, by shifting the evaluation perspective from a single point to a distribution, this research offers a more scientific and actionable basis for navigating the complex landscape of intelligent construction. It empowers stakeholders to make smarter, risk-informed decisions, ultimately fostering a more sustainable and successful technological transformation of the AEC industry.

REFERENCES

[1] J. Yan, Z. Zheng, Y. Zhou, J. Lin, and Y. Deng, “A Review on Current Advances of Intelligent Construction Based on Bibliometric Analysis,” the International Conference on Construction and Real Estate Management 2022, American Society of Civil Engineers, Dec. 2022, pp. 340–355.

[2] M. Štefanič and V. Stankovski, “A review of technologies and applications for smart construction,” *Proceedings of the Institution of Civil Engineers - Civil Engineering*, ICE Publishing, May 2019, pp. 83–87.

[3] H. Gao, Y. Lu, and H. Wu, “Intelligent Construction Assessment for Construction Projects: Toward an Evaluation Framework,” *Proceedings of the 26th International Symposium on Advancement of Construction Management and Real Estate*, H. Guo, D. Fang, W. Lu, and Y. Peng, Eds., Singapore: Springer Nature, 2022, pp. 829–844.

[4] K. Liu, Q. Meng, Q. Kong, and X. Zhang, “Review on the Developments of Structure, Construction Automation, and Monitoring of Intelligent Construction,” *Buildings*, vol. 12, no. 11, Art. no. 11, Nov. 2022.

[5] C. Lu, J. Liu, Y. Liu, and Y. Liu, “Intelligent construction technology of railway engineering in China,” *Frontiers of Engineering Management*, vol. 6, no. 4, pp. 503–516, Dec. 2019.

[6] G. Xu, G. K. Chang, D. Wang, A. G. Correia, and S. Nazarian, “The pioneer of intelligent construction—An overview of the development of intelligent compaction,” *Journal of Road Engineering*, vol. 2, no. 4, pp. 348–356, Dec. 2022.

[7] M. Stojčić, E. K. Zavadskas, D. Pamučar, Ž. Stević, and A. Mardani, “Application of MCDM Methods in Sustainability Engineering: A Literature Review 2008–2018,” *Symmetry*, vol. 11, no. 3, Art. no. 3, Mar. 2019.