World Journal of Engineering Research

Volume 3, Issue 6, 2025

Published by Upubscience Publisher

World Journal of Engineering Research
Print ISSN: 2959-9865 Online ISSN: 2959-9873
Email: info@upubscience.com
Website: http://www.upubscience.com/

Table of Content

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3064

LLM-GUIDED ADAPTIVE STEP-SIZE ZEROING NEURAL NETWORK FOR ROBOTIC MANIPULATOR TRAJECTORY TRACKING

GuangYu Long, ZhuoSong Fu*
College of Computer Science and Engineering, Jishou University, Jishou 416000, Hunan, China.
**Corresponding Author: ZhuoSong Fu*

Abstract: Zeroing Neural Network (ZNN) is widely used for robotic trajectory tracking due to their rapid convergence and strong error attenuation properties. In discrete implementations, however, the choice of step-size critically influences tracking accuracy, numerical stability, and computational cost. Conventional variable-step strategies typically rely on fixed heuristics or manually tuned rules, limiting their adaptability in dynamic task conditions. To address this limitation, this paper introduces a Large Language Model (LLM)-based adaptive step-size mechanism for discrete ZNN in manipulator trajectory tracking. The LLM receives natural-language instructions together with essential system-state descriptions and outputs step-size adjustments that guide the ZNN update. This enables intuitive human-robot interaction and allows the controller to flexibly shift between high-precision tracking and low-computation execution without modifying the underlying ZNN formulation. Results show that the proposed method improves tracking accuracy when finer steps are selected, reduces computational load when coarser steps suffice, and maintains high semantic consistency in interpreting step-size-related instructions. These findings demonstrate the potential of integrating LLM reasoning into step-size regulation to enhance the flexibility and interpretability of discrete ZNN-based robotic tracking.
Keywords: Zeroing neural network; Variable step-size; Large Language Model; Robotic manipulator

1 INTRODUCTION

Zeroing Neural Network (ZNN) has been extensively studied as an effective computational framework for solving time-varying problems such as inverse kinematics and trajectory tracking of robotic manipulators [1-3]. By embedding the solution of a dynamic equation into an exponentially convergent neural error system, ZNN provides fast response, strong disturbance rejection, and stable tracking properties [4-5]. When implemented in discrete time, however, the performance of ZNN strongly depends on the choice of the numerical step size τ. A small step size yields higher tracking accuracy but significantly increases computational load, while a large step size improves computational efficiency at the cost of reduced precision or even numerical instability [6-7]. Designing an adaptive step-size strategy that balances accuracy and efficiency has therefore become an important issue in discrete ZNN research.

Existing approaches to variable step-size selection typically rely on fixed analytical rules, derivative-based indicators, error thresholds, or predefined schedules [8-10]. Although these methods can adjust τ according to certain system metrics, they lack flexibility, semantic interpretability, and the ability to incorporate human intentions. In practica applications, high-level objectives such as “make the tracking more accurate” or “run faster even if precision decreases” using natural language are more intuitive than manually tuning numerical parameters. Conventional adaptive algorithms cannot directly interpret such natural-language instructions, nor can they generalize across linguistically varied expressions that imply similar operational adjustments.

Recent advances in Large Language Models (LLMs) have shown remarkable capability in semantic understanding, instruction following, and reasoning under loosely specified objectives [11-13]. These properties make LLMs promising candidates for assisting numerical algorithms that require flexible or context-dependent parameter regulation. While LLMs have been applied to robot task planning, code generation, and decision-making [14], their potential as interpreters for numerical step-size modulation in ZNN solvers has not yet been investigated. To the best of our knowledge, no prior work has explored integrating LLM-driven semantic guidance with discrete ZNN to achieve human-interpretable and adaptive step-size control for trajectory tracking tasks.

To bridge this gap, this paper proposes an LLM-guided step-size adjustment mechanism for discrete ZNN manipulator trajectory tracking. The LLM receives natural-language instructions together with key system-state information and outputs step-size τ. This design allows the controller to dynamically shift between high-accuracy tracking and low-cost computation while maintaining an intuitive interaction interface. More importantly, the method introduces semantic consistency into numerical regulation: instructions that imply similar intentions yield similar step-size decisions, even across diverse linguistic expressions.

The main contributions of this work are summarized as follows.

1) Introduce an LLM-guided step-size adjustment mechanism that interprets natural-language instructions and system-state descriptions to dynamically regulate the step-size τ of discrete ZNN without modifying its computational formulation.

2) Develop a variable step-size discrete ZNN scheme for robotic manipulator trajectory tracking that flexibly transitions between high-precision updates and low-computation operation based on LLM-generated decisions.
3) Validate the effectiveness and robustness of the LLM-driven step-size adjustments through experiments on robotic manipulator trajectory tracking.

2 PROBLEM AND METHOD

This section presents the discrete ZNN formulation used for trajectory tracking, the inverse kinematics mapping of the robotic manipulator, and the proposed LLM-guided step-size adjustment mechanism. To avoid ambiguity, the parameter τ is consistently referred to as the step-size, which also determines the real-time update interval of the controller.

2.1 Zeroing Neural Network for Inverse Kinematics

For a robotic manipulator with joint configuration $\mathbf{q}(t) \in R^n$ and end-effector pose $\mathbf{x}(t) \in R^m$, the forward kinematics is expressed as

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{q}(t)). \tag{1}$$

Given a desired trajectory $\mathbf{x}_{\mathrm{d}}(t)$, the tracking error is defined as

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{q}(t)) - \mathbf{x}_{\mathrm{d}}(t). \tag{2}$$

The continuous-time ZNN imposes an exponentially stable error dynamics:

$$\dot{\mathbf{e}}(t) = -\alpha \mathbf{e}(t), \tag{3}$$

where $\alpha > 0$ is a constant decay coefficient. Differentiating $\mathbf{e}(t)$ with respect to time gives

$$\dot{\mathbf{e}}(t) = J(\mathbf{q}(t))\dot{\mathbf{q}}(t) - \dot{\mathbf{x}}_{\mathrm{d}}(t), \tag{4}$$

where $J(\mathbf{q}(t))$ is the manipulator geometric Jacobian. Combining the above equations leads to the ZNN update law:

$$J(\mathbf{q}(t))\dot{\mathbf{q}}(t) = \dot{\mathbf{x}}_{\mathrm{d}}(t) - \alpha \mathbf{e}(t), \tag{5}$$

and the joint velocity can be obtained via

$$\dot{\mathbf{q}}(t) = J^{\dagger}(\mathbf{q}(t))(\dot{\mathbf{x}}_{\mathrm{d}}(t) - \alpha \mathbf{e}(t)) \tag{6}$$

where $J^{\dagger}(\mathbf{q}(t))$ denotes the Moore–Penrose pseudoinverse.

2.2 Discrete ZNN Update with Step Size τ

For real-time implementation, the continuous dynamics are discretized using a forward-Euler scheme. At discrete time index k, the update rule becomes

$$\mathbf{q}_{k+1} \doteq \mathbf{q}_k + \tau_k \dot{\mathbf{q}}_k \tag{7}$$

where τ_k is the step-size applied at iteration k.
Substituting the ZNN expression of $\dot{\mathbf{q}}_k$, the discrete update becomes

$$\mathbf{q}_{k+1} = \mathbf{q}_k + \tau_k J^{\dagger}(\mathbf{q}_k)(\dot{\mathbf{x}}_{\mathrm{d},k} - \alpha \mathbf{e}_k). \tag{8}$$

A smaller τ_k increases update frequency and improves tracking accuracy, while a larger τ_k gives faster execution but reduce precision. Therefore, selecting an appropriate τ_k plays a critical role in discrete ZNN performance.

2.3 Step-Size Decision Based on System State Description

At iteration k, the system constructs a compact description of the current motion state, including: 1) the discrete time index k, or elapsed time t_k; 2) the current tracking error norm $\|\mathbf{e}_k\|$; and 3) the currently applied step-size τ_k. This information is formatted as a short text string and used as part of the input to the LLM and enables the step-size decision to depend on the real-time tracking condition.

2.4 Natural-Language-Guided Step-Size Modulation

In addition to the state description, the user provides a natural-language instruction such as “make the tracking more accurate” or “update less frequently”. The LLM processes 1) the human instruction and 2) the system-state description, and outputs a recommended step size:

$$\tau_k^{\mathrm{LLM}} = g_{\mathrm{LLM}}(\mathrm{instruction}, \mathrm{statedescriptioin}). \tag{9}$$

The output is constrained to a safe range:

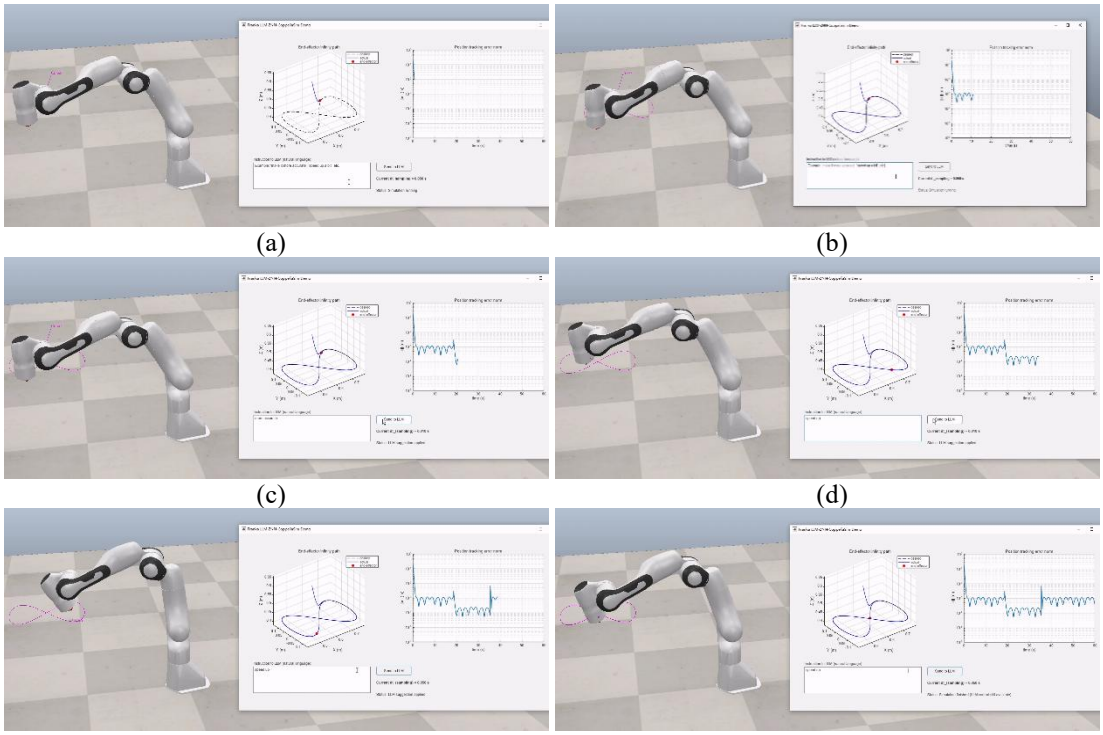


Figure 1 Real-Time Path Tracking under the LLM-Guided Step-Size Adjustment Mechanism

$$\tau_{\min} \le \tau_k^{\mathrm{LLM}} \le \tau_{\max}, \tag{10}$$

ensuring numerical stability and preventing overly aggressive updates. On the basis of Eq. (10), the final step-size used in ZNN iteration is

$$\tau_k^* = \min(\max(\tau_k^{\mathrm{LLM}}, \tau_{\min}), \tau_{\max}). \tag{11}$$

This formulation allows: 1) intuitive human control through natural language; 2) semantically consistent adjustment of τ; adaptation to varying error levels and trajectory conditions. Notably, the ZNN computational structure remains unchanged. Only the numerical step size is modulated.

2.5 Overall Procedure

The algorithm for iteration k to change the step-size can be summarized as

1) Compute tracking error $\mathbf{e}_k$ and derivative $\dot{\mathbf{x}}_{\mathrm{d},k}$;
2) Assemble the state description including t_k, τ_k, and $\|\mathbf{e}_k\|$;
3) Query the LLM using the natural-language instruction and the state description.
4) Obtain the recommended step-size τ_k^*;
5) Perform the ZNN update: $\mathbf{q}_{k+1} = \mathbf{q}_k + \tau_k^* J^{\dagger}(\mathbf{q}_k)(\dot{\mathbf{x}}_{\mathrm{d},k} - \alpha \mathbf{e}_k)$.

3 EXPERIMENTS

To evaluate the proposed LLM-guided step-size adjustment mechanism, a series of trajectory-tracking experiments are conducted using a Franka Emika Panda manipulator simulated in CoppeliaSim [15]. The robot follows a planar infinity-like path, while the discrete ZNN solver computes joint updates in real time. MATLAB serves as the controller environment, generates ZNN-based velocity updates, issues step-size decisions, and communicates with CoppeliaSim through the Remote API.

Three experiments are designed.

1) **Experiment 1** illustrates how the LLM modulates the step-size τ in response to human natural-language instructions and shows the resulting trajectory-tracking behavior.
2) **Experiment 2** compares fixed-step ZNN solvers with the LLM-guided variable-step scheme under identical conditions.
3) **Experiment 3** evaluates the robustness and semantic consistency of LLM-generated step-sizes using multiple natural-language expressions.

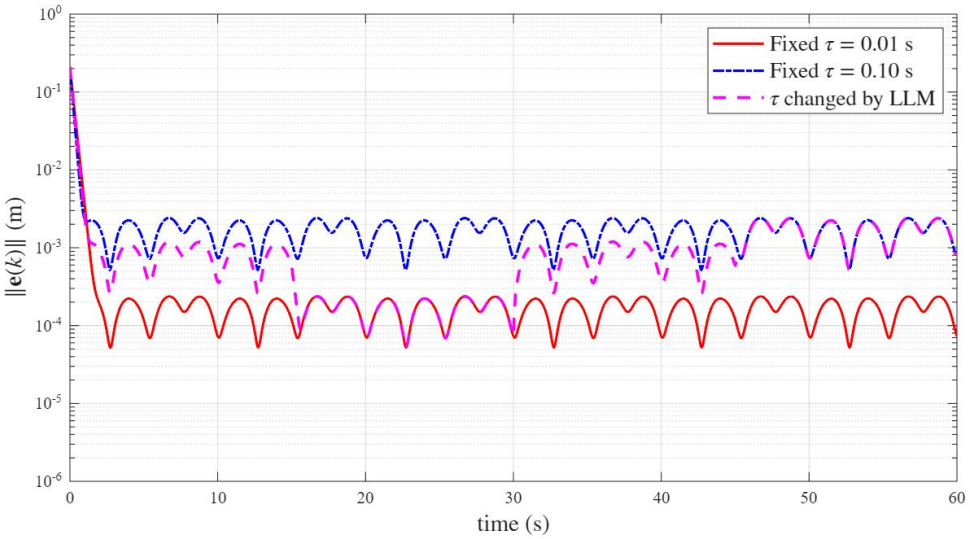


Figure 2 Tracking-Error Comparison between Fixed Step-Sizes and the LLM-Guided Step-Size

All experiments use the same ZNN formulation, robot model, and infinity-like path. The results confirm that the proposed mechanism adjusts τ appropriately, improves tracking precision when instructed to be "more accurate", and increases update efficiency when instructed to "speed up", while maintaining stable tracking behavior.

3.1 Experiment 1: Real-Time Step-Size Modulation via Natural-Language Instructions

Experiment 1 investigates how the LLM adjusts the step size in real time during robotic trajectory tracking. The user provides natural-language commands such as "speed up" or "make it more accurate", and the LLM outputs a suitable step-size based on both the instruction and the current system state. The manipulator begins tracking the infinity-like path with an initial step-size $\tau = 0.05$ s. Six representative snapshots from the experiment appear in Figure 1(a)-(f). These frames show the robot motion in CoppeliaSim, the desired and actual end-effector trajectories, and the corresponding error curves.
1) **Experiment begins**. At the start of the experiment, the manipulator moves from its initial configuration toward the infinity-like path. The tracking error is relatively high, as the end-effector has not yet reached the desired path.
2) **Error stabilizes under the current step size**. After several iterations, the ZNN solver drives the robot onto the desired path, and the error norm stabilizes at a characteristic level determined by the initial step-size. The tracking motion becomes smooth and periodic.
3) **The first natural-language instruction is issued**. When the user enters a command (e.g., "more accurate"), the LLM generates a new step-size. The change in τ modifies the update frequency of the ZNN solver, and the error curve responds immediately. A smaller step size produces a sharp decrease.
4) **Error stabilizes under the new step size**. After the initial transient, the tracking error settles into a new steady pattern that corresponds to the updated step-size. A reduced τ yields smaller error errors.
5) **The second natural-language instruction is issued**. The user inputs another natural-language instruction. The LLM interprets the new intent and updates τ. The tracking error immediately reflects this change, showing a rise when the LLM decreases or increases the step-size.
6) **Experiment concludes**. Near the end of the trajectory, the robot continues tracking the desired path using the most recently selected step-size. The tracking behavior remains stable, and the error remains bounded, demonstrating that LLM-guided step-size adjustment mechanism does not compromise the stability of the discrete ZNN solver.

3.2 Experiment 2: Comparison Between Fixed and LLM-Guided Step-Sizes

Experiment 2 evaluates the tracking performance of the proposed LLM-guided step-size adjustment mechanism by comparing it with two fixed-step ZNN baselines. All three methods operate on the same infinity-like path and use identical ZNN parameters except for the step-size τ. The goal is to assess whether the proposed mechanism produces behavior consistent with the numerical characteristics of the underlying solver and whether its performance falls

between the fine-step and coarse-step baselines, as theoretically expected. Three tracking-error curves are plotted in Figure 2.

1) The fixed-step baselines reveal the expected behavior of the discrete ZNN solver. The solid line ($\tau = 0.01$ s) shows the smallest steady-state error and the smoothest oscillation pattern. The error decreases rapidly at the beginning of the experiment and then remains in a narrow band around 10^{-4} m. This reflects the high numerical resolution achieved with a small step-size.

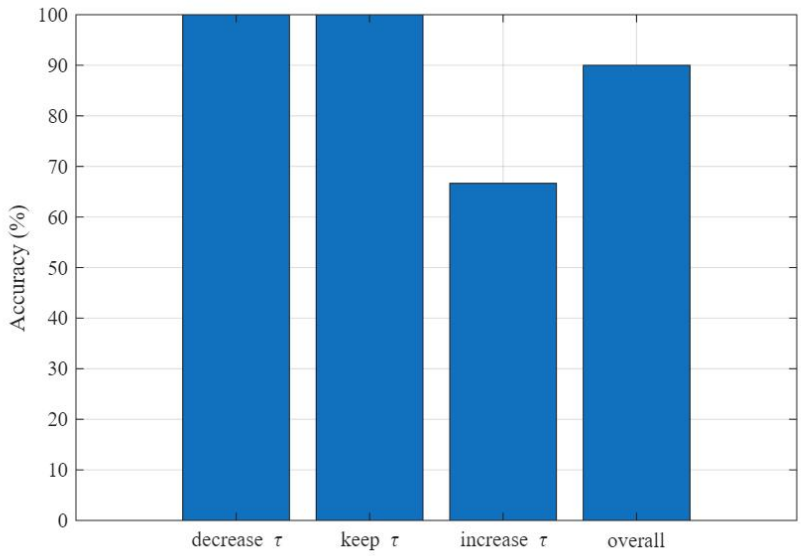


Figure 3 Accuracy of Instruction Interpretation

2) The dash-dotted line (($\tau = 0.10$ s) shows a noticeably larger steady-state error, remaining around 10^{-3} m. The oscillations are more pronounced and occur in a wider band, which is consistent with the reduced update frequency of the solver. Although the error remains bounded and stable, this configuration provides visibly lower precision.

3) The dashed line, corresponding to the LLM-guided step-size adjustment mechanism, oscillates between the behaviors of the two baselines. When the mechanism selects a smaller step-size, the error decreases and approaches the pattern of the solid line. When it selects a larger step-size, the error increases temporarily and moves toward the pattern of the dash-dotted line. After each change in step-size, the error settles into a predictable and stable range.

Overall, the dashed line remains entirely between the solid and dash-dotted lines, demonstrating that the LLM-guided step-size adjustment mechanism produces performance commensurate with the numerical properties of the selected step -sizes. In addition, the transitions between different steady-state levels occur smoothly. There is no sign of divergence, instability, or abrupt irregularities. This confirms that the mechanism respects the stability characteristics of the discrete ZNN solver while enabling adaptive behavior.

3.3 Experiment 3: Semantic Consistency

Experiment 3 evaluates the semantic consistency of the LLM-guided step-size adjustment mechanism when processing diverse natural-language expressions that convey similar intentions. The goal is to determine whether the mechanism reliably maps different verbal formulations to the correct type of step-size action: 1) decrease τ; 2) keep τ; and increase τ. To conduct the test, 100 generated instruction phrases are prepared for each of the three action categories. For example:

1) "Make it more accurate", "reduce the step size" and "increase update frequency" denote instructions that decrease τ.
2) "Maintain the current setting", "leave it unchanged" and "keep going as is" denote instructions that keep τ.
3) "Speed up", "use a larger step size" and "reduce update frequency" denote instructions that increase τ.

Each instruction is independently fed to the mechanism, together with a neutral system-state description, and the resulting step-size decision is labeled as correct or incorrect. Figure 3 shows the accuracy across the three action types and the overall performance.

The results indicate that the mechanism demonstrates high semantic reliability. For instructions that decrease τ, the accuracy reaches 100%, meaning every tested phrase is interpreted correctly. For instructions that keep τ, the accuracy also reaches 100%, showing that the mechanism recognizes neutrality consistently. For instructions that increase τ, the accuracy is slightly lower at 66.7%, indicating that upward step-size adjustments are somewhat harder for the mechanism to infer, possibly due to greater linguistic diversity in how users express speeding-up intentions. The overall accuracy reaches 90%, confirming strong robustness across all categories.

Across all tests, the mechanism does not produce unstable or extreme step-size recommendations, showing that the semantic interpretation remains well-behaved even when minor misclassifications occur.

4 CONCLUSION

This paper has presented an LLM-guided step-size adjustment mechanism for discrete ZNN-based trajectory tracking of robotic manipulators. By allowing the step-size τ to be determined through natural-language instructions and system-state descriptions, the proposed framework has enabled intuitive human–robot interaction without modifying the underlying ZNN formulation. The method has maintained the numerical properties of the solver while introducing semantic interpretability and adaptability into the control process. A series of experiments has been conducted to validate the effectiveness and robustness of the mechanism. The results have shown that integrating LLM reasoning into numerical step-size selection offers a promising direction for adaptive, human-interpretable robotic control. The framework has established a foundation for future extensions, such as richer multimodal instructions, more advanced ZNN formulations, and real-robot experiments.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research was supported by the Research Foundation of Jishou University under Grant JD24004.

REFERENCES

[1] Fu Z, Zhang Y, Tan N. Gradient-feedback Zhang neural network for unconstrained time-variant convex optimization and robot manipulator application. IEEE Transactions on Industrial Informatics, 2023, 19(10): 10489 -10500.
[2] Fu Z, Zhang Y, Li W. Solving future nonlinear equation system via ZNN and novel general ILR3S formula with multitype manipulator applications. IEEE Transactions on Industrial Electronics, 2024, 71(10): 12623-12633.
[3] Lai G, Lin C, Ouyang Y, et al. Finite-time uncalibrated visual servoing for robotic manipulators based on model-free zeroing neural networks. IEEE Transactions on Automation Science and Engineering, 2025, 22: 24347-24361.
[4] Li J, Qu L, Zhu Y, et al. A novel zeroing neural network for time-varying matrix pseudoinversion in the presence of linear noises. Tsinghua Science and Technology, 2025, 30(5): 1911-1926.
[5] Liao B, Han L, Cao X, et al. Double integral-enhanced Zeroing neural network with linear noise rejection for time-varying matrix inverse. CAAI Transactions on Intelligence Technology. 2024, 9(1): 197–210.
[6] Guo J, Zhang Y. Stepsize interval confirmation of general four-step DTZN algorithm illustrated with future quadratic programming and tracking control of manipulators. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(3): 1662-1670.
[7] Qiu B, Guo J, Li X, et al. New discretized zeroing neural network models for solving future system of bounded inequalities and nonlinear equations aided with general explicit linear four-Step rule. IEEE Transactions on Industrial Informatics, 2021, 17(8): 5164-5174.
[8] Park J, Park C, Park P. A variable step-size MSD-based diffusion LMS algorithm with nnoisy input and quantization link noise. IEEE Wireless Communications Letters, 2025, 14(11): 3734-3738.
[9] Liu J. An improved arithmetic optimization algorithm with multi-strategy variable step size for optimizing engineering problems. IEEE Access, 2025, 13: 156566-156599.
[10] Shen L, Jiang D, Zhu H. Control study for suppressing rotor vibration of the six-pole hybrid magnetic bearing system based on no-parameter variable-step-size least-mean-square algorithm. IEEE Transactions on Power Electronics, 2025, 40(8): 10672-10681.
[11] Hongying Z, Javed A, Abdullah M, et al. Large language models with contrastive decoding algorithm for hallucination mitigation in low-resource languages. CAAI Transactions on Intelligence Technology. 2025, 10: 1104-1117.
[12] Duan J, Li W, Bai Q, Nguyen M, et al. LLM-botguard: A novel framework for detecting LLM-driven bots with mixture of experts and graph neural networks. IEEE Transactions on Computational Social Systems, 2025, 12(5): 3488-3500.
[13] Chen W, Li G, Li M, et al. LLM-enabled incremental learning framework for hand exoskeleton control. IEEE Transactions on Automation Science and Engineering, 2025, 22: 2617-2626.
[14] Fakhoury S, Naik A, Sakkas G, et al. LLM-based test-driven interactive code generation: User study and empirical evaluation. IEEE Transactions on Software Engineering, 2024, 50(9): 2254-2268.
[15] Ferro M, Mirante A, Ficuciello F, et al. A Coppeliasim dynamic simulator for the Da Vinci research kit. IEEE Robotics and Automation Letters, 2023, 8(1): 129-136.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3065

DYNAMIC CHARACTERISTICS OF GEARBOX HOUSING VIBRATION ON THE BEIJING-SHANGHAI LINE

Qi Wang, JiMin Zhang*, Kun He
School of Transportation, Tongji University, Shanghai 201804, China.
**Corresponding Author: JiMin Zhang*

Abstract: Gearbox housing vibration critically impacts the operational safety and reliability of high-speed trains. However, existing research mainly relies on the track spectrum of Wuhan-Guangzhou Line, lacking analysis tailored to the unique characteristics and curve negotiation conditions of Beijing-Shanghai Line. To address this gap, this study establishes an experimentally validated vehicle-track coupled dynamics model, conducting multi-condition simulations to explore gearbox vibration correlations with operating parameters, track conditions, and wheel polygon excitation under Beijing-Shanghai Line scenarios. Key findings show that as speed increases, vertical (Z) and lateral (Y) vibration responses of three measurement points converge. Low speeds are dominated by 11th-order wheel polygon effects, while high speeds are influenced more by the 23rd-order polygon. Upper gearbox points exhibit prominent Y-direction vibration, and lower points show significant Z-direction vibration. Beyond 300 km/h, speed becomes the dominant factor, with polygon order's influence on vibration acceleration RMS values diminishing. Gearbox vibration intensifies at 275 km/h under 23rd-order polygon excitation, and vibration severity increases with wheel polygon amplitude. Curve negotiation analysis reveals that smaller radii induce more severe vibration: Z-direction vibration decreases monotonically with increasing radius. Y-direction vibration behaves differently by speed—increasing with radius at 250 km/h but decreasing at other speeds. Local peaks occur at 9,000 m radius (275/325 km/h) and 8,000 m radius (350 km/h) due to enhanced wheel-rail excitation.
Keywords: Gearbox housing vibration; Beijing-Shanghai Line; Wheel polygon; Curve radius

1 INTRODUCTION

Under complex operating conditions, high-speed train gearboxes experience excessive vibration and dynamic stress due to multi-source excitations (including traction system dynamic loads, wheel-rail coupled vibrations) and high-frequency impacts induced by track irregularities/wheel polygonalization, accelerating fatigue damage and threatening operational safety. To address this issue, researchers have conducted a series of studies: Parey et al. established multi-degree-of-freedom dynamic models for gear transmission systems [1-2]; Carbonelli et al. revealed the inherent modal shapes of gearbox housings through finite element analysis [3]; Kahraman et al. investigated vibration response characteristics under external excitations [4]. Zhang et al. identified stress distribution patterns under external excitations via dynamic stress testing [5], while Wang et al. quantified the impacts of track irregularities and wheel polygonalization on dynamic stresses, emphasizing speed, track conditions, and wheel tread profiles as critical factors [6-7]. Notably, Huang et al. demonstrated that time-varying meshing stiffness and transmission errors significantly amplify vertical vibrations, with spectral characteristics dominated by gear meshing frequencies, highlighting the importance of internal excitations [8]. Hu et al. revealed through multi-domain analysis that coupled effects of wheel-rail excitation and casting defects reduce fatigue strength, proposing that optimized casting processes could prevent resonance between excitation frequencies and structural natural frequencies [9]. Zhang et al. developed finite element modal analysis methods to identify local resonance and stress concentration points caused by external excitations, based on gearbox crack failure databases [10]. Currently, most studies adopt the Wuhan-Guangzhou (Wuguang) track spectrum, primarily because the Wuguang high-speed railway, as one of China's early-built mainlines, had its track spectrum data publicly available earlier, providing a more mature research foundation. Early simulations and tests were predominantly based on Wuguang data, and subsequent studies have continued using this spectrum to maintain comparative consistency. However, with the sustained development of China's economy, the operational frequency of the Beijing-Shanghai High-Speed Railway has continuously increased in practical application, making it the busiest rail line in the country. Designed for a speed of 380 km/h, which is higher than the Wuhan-Guangzhou line's 350 km/h, the track spectrum of the Beijing-Shanghai line more accurately reflects wheel-rail excitation characteristics under ultra-high-speed conditions. Therefore, this paper establishes a vehicle-track coupled dynamics model incorporating wheel polygonal wear, aiming to systematically investigate the vibration characteristics of the high-speed train gearbox under Beijing-Shanghai line conditions.

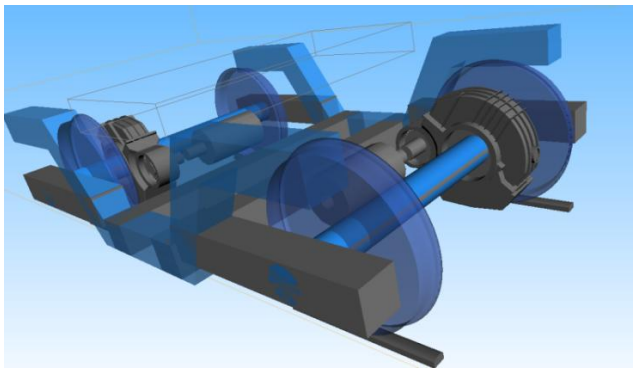
2 DYNAMIC MODEL CONSTRUCTION

Based on the given vehicle parameters, this study establishes a full vehicle dynamics model. Given the significant influence of the transmission system on the gearbox vibration characteristics, key components such as the motor, drive shaft, pinion, driven shaft, gear(parameters detailed in Table 1), and gearbox housing are incorporated in detail into the model. The gear meshing interactions are characterized using the F225 force element in SIMPACK, as illustrated in

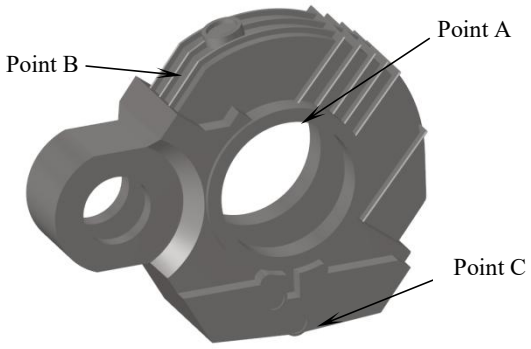
Figure 1(a). To further investigate the local dynamic response of the gearbox, three key points are selected as observation points on its flexible body: the gear bearing hole seat (Point A), the gear tooth observation hole (Point B), and the oil level observation hole (Point C), with their locations shown in Figure 1(b) [11-12].

Table 1 Gear Parameters

	Pinion	Gear
Teeth	35	85
Tooth width(mm)	66	65
Profile shift	0.225	0.024
Helix angle(deg)	18	-18
Module	6	
Normal pressure angle(deg)	20	
Center distance(mm)	380	
Stiffness Ratio	0.8	
Young's Modulus(Pa)	2.1e11	
Poisson's Ratio	0.3	
Damping Coefficient(N•s/m)	5000	



(a) Dynamic model construction



(b) Observation points

Figure 1 Dynamic Model Construction and Observation Points

3 KINEMATIC THEORY AND EXCITATION

3.1 Kinematic Theory

For the matrix form of the motion differential equations of the vehicle system:

$$[M]\{\ddot{x}\}+[C]\{\dot{x}\}+[K]\{x\}=\{F\} \tag{1}$$

Based on the gear transmission system dynamics theory, a coupled bending-torsion-axial-lateral vibration dynamic model of helical gears is established. The model incorporates nonlinear factors such as time-varying mesh stiffness, transmission error, and backlash of helical gears, while tooth surface friction is neglected.The meshing force $F_{\text{meshi}}(i=x,y,z)$ acting on the large gear can be expressed as [13]:

$$\begin{bmatrix} F_{\text{mesh}x} \\ F_{\text{mesh}y} \\ F_{\text{mesh}} \end{bmatrix} = \begin{bmatrix} F_{\text{m}}\sin\alpha\cos\beta \\ F_{\text{m}}\sin\beta \\ F_{\text{m}}\cos\alpha\cos\beta \end{bmatrix} \tag{2}$$

$$F_{\text{m}} = k_{\text{m}}\left(\delta + e_0\right) + c_{\text{m}}\dot{\delta} \tag{3}$$

In the equation, $k_{\text{m}}, c_{\text{m}}$ represent the time-varying mesh stiffness and damping of the helical gears, respectively. δ denotes the dynamic transmission error. e_0 represents the transmission error. $x_i, y_i, z_i (i=p,w)$ are the translational displacements of the pinion and wheel, respectively. β is the helix angle. α is the pressure angle. F_m is the normal force acting on the helical gear tooth surface.

3.2 The Wheel Polygonalization Excitation Function

Considering the functional relationships between wheel polygon radius deviation, order number, out-of-roundness, and rotation angle, these parameters are modeled in the form of harmonic functions to derive the mathematical expression for wheel polygonization [14].

$$\Delta R(\beta) = CA\sin(n(\beta+\beta_0)) \qquad \beta \in (0, 2\pi] \tag{4}$$

Where ΔR denotes the wheel radius deviation; A represents the amplitude; n indicates the polygonal order of the wheel; C stands for the out-of-roundness; β corresponds to the wheel rotation angle; and β_0 signifies an additional relative rotation and offset.By configuring the relevant settings in Simpack, the corresponding simulations can be performed.

4 MODEL VALIADATION

Based on field measurement data of gearbox housing vibrations on the Wuhan-Guangzhou Line(train speed: 300 km/h), a comparative analysis was conducted between simulations and operational conditions regarding the lateral and vertical vibration accelerations and power spectral density (PSD) of the gearbox. As shown in Figure 2, the experimental test results showed that the lateral and vertical vibration accelerations of gearbox housing fluctuated within ranges of approximately -70 to 60 m/s² and -124.58 to 136.47 m/s², respectively, while the simulation results showed ranges of approximately -68.08 to 58.88 m/s² and -117.69 to 64.85 m/s², respectively. The main vibration range and maximum amplitude of both vertical and lateral vibration accelerations were similar between simulation and experiment. In the frequency domain, as illustrated in Figure 3, both simulations and measurements exhibited consistent trends, though the measured spectra displayed more complex components. These discrepancies primarily originated from three factors: the simulated track excitation spectrum failed to fully account for short-wave irregularities induced by rail weld joints and fasteners; the actual track conditions deteriorated after prolonged service; and the rigid-body simplification of components in the simulation model. These limitations collectively led to lower vibration responses in simulations compared to field data. Nevertheless, this comparative study validates the accuracy of the proposed vehicle-track transmission system dynamics model, thereby providing a robust foundation for further theoretical investigations.

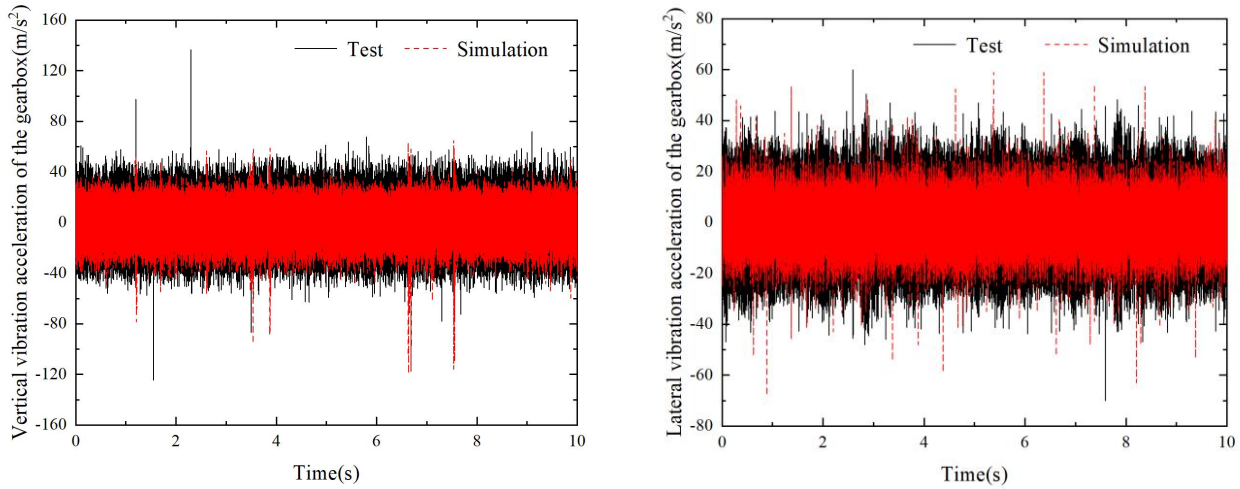


Figure 2 Gearbox Vibration Acceleration

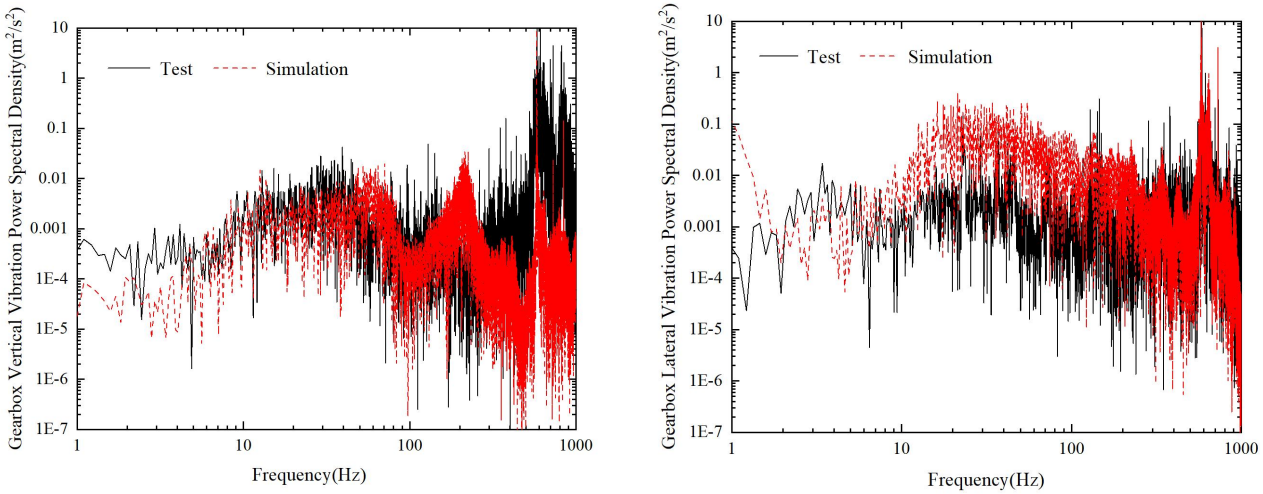


Figure 3 PSD of Gearbox Vibration Acceleration

5 GEARBOX HOUSING VIBRATION ANALYSIS

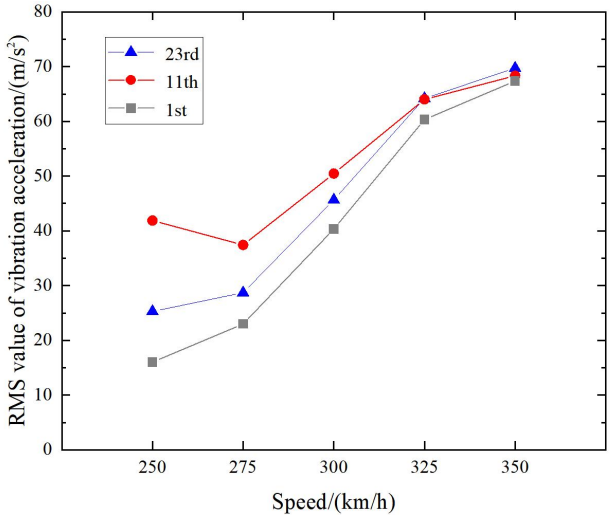
5.1 Effect of Wheel Polygon Excitation on Gearbox Housing Vibration Characteristics

Wheel-rail excitation is the primary source of gearbox housing vibration. Wheel polygonalization can significantly exacerbate the impact loads within the wheel-rail system, thereby influencing the dynamic response of the gearbox. Therefore, a thorough investigation into its impact is necessary.

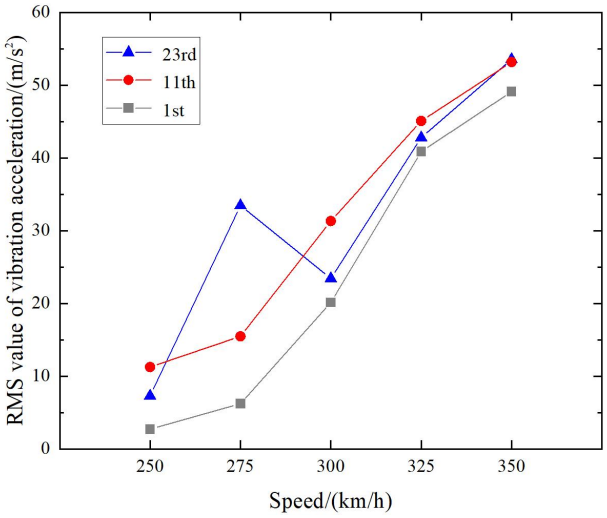
5.1.1 Effect of wheel polygon order on gearbox housing vibration characteristics

The polygon order is a key parameter characterizing wheel polygonal features. To investigate its influence, this study compares and analyzes the vibration acceleration amplitudes at three measurement points on the gearbox housing under combined conditions of five train speeds (250, 275, 300, 325, and 350 km/h) and three polygon orders. Based on vibration test data of a certain EMU gearbox and wheel polygon statistics, the polygon orders selected for this study include the 1st, 11th, and 23rd orders [15-16].

As shown in Figure 4, the vibration trends in the Z-direction are generally consistent across the three observation points, but the amplitude at Point C is higher than at the other two points. In the low-speed range (e.g., 250 km/h), the 11th order polygon induces the most pronounced vibration. As the speed increases, the vibration levels under the three orders gradually converge, with the 23rd order exhibiting a relatively more prominent influence. A notable phenomenon is observed at 250 km/h, where the vibration caused by the 11th order is substantially greater than that of the other orders. This is attributed to the stronger vertical vibrations of both the pinion and gear under this specific order condition. When the speed exceeds 300 km/h, the influence of the polygon order on the root mean square (RMS) values of vibration acceleration at each measurement point diminishes, with speed becoming the dominant factor. In the Y-direction, the vibration amplitudes follow the order Point B > Point C > Point A, yet the overall trends remain relatively consistent. Specifically, the vibrations under the 11th and 1st orders show a general increasing trend with speed, while the 23rd order exhibits a distinct peak at 275 km/h. This peak occurs because the wheel-rail excitation frequency at this speed (approximately 643 Hz) coincides with the natural frequency of the gearbox, leading to resonance.

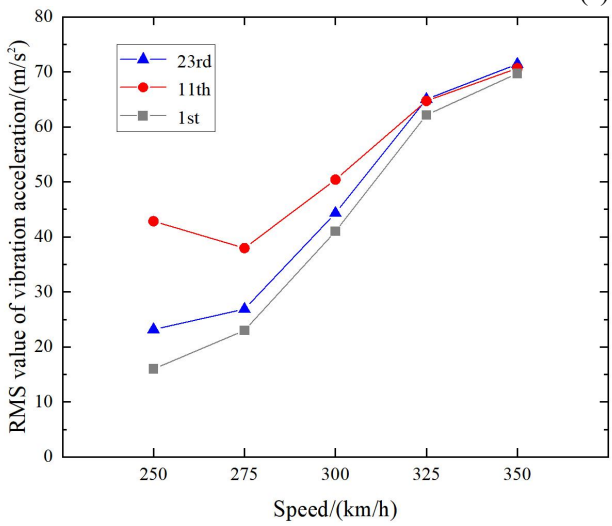


(i) z-direction

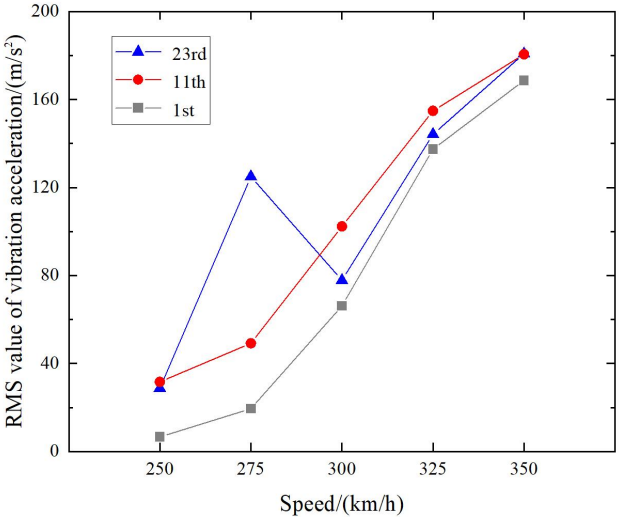


(ii) y-direction

(a) Point A

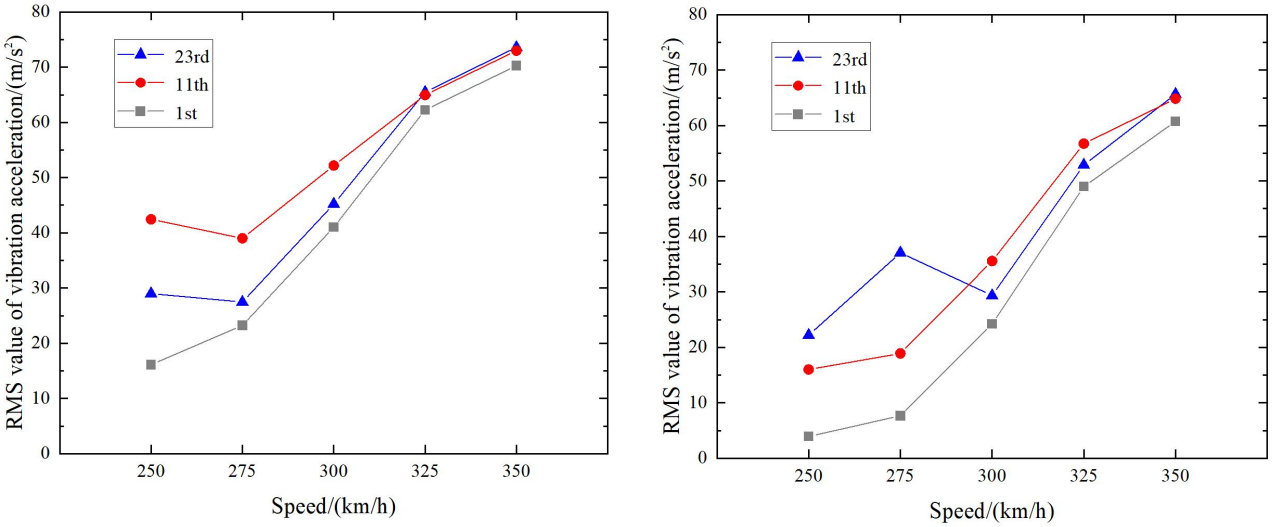


(i) z-direction



(ii) y-direction

(b) Point B



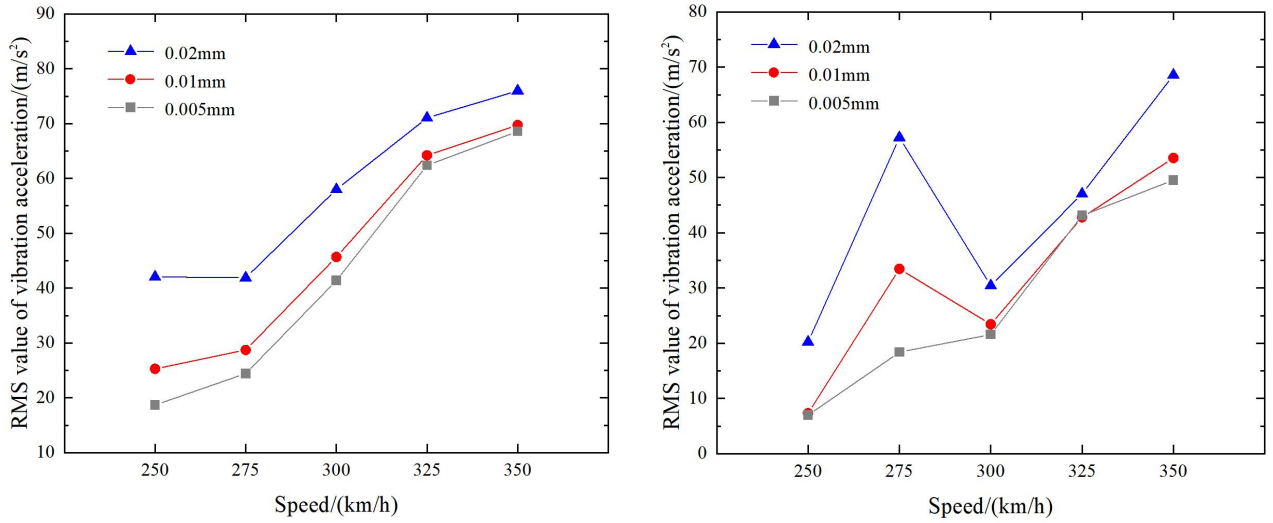
(i) z-direction (ii) y-direction

(c) Point C

Figure 4 Effect of Wheel Polygon Order on the RMS Value of Gearbox Vibration Acceleration at Different Speeds

5.1.2 Effect of wheel polygon amplitude on gearbox housing vibration characteristics

Under high-speed conditions, vibrations intensify, with the 23rd-order wheel polygon causing the most significant impact. Therefore, this study sets the polygon order to 23. According to literature, the amplitude limit for wheel polygons of orders 18–24 is typically below 0.020 mm. Simulations were conducted with amplitudes of 0.005 mm, 0.010 mm, and 0.020 mm, at speeds ranging from 250 to 350 km/h, to analyze the effect of amplitude on gearbox housing vibration.As shown in Figure 5, at a given speed, the RMS values of vibration acceleration are highest at an amplitude of 0.020 mm, with overall trends remaining consistent. Comparative simulations for the 11th-order polygon yielded similar results, which are not detailed here. Among the three measurement points, Point B shows the most intense vibration response overall, requiring special attention in fatigue analysis.



(i) z-direction (ii) y-direction

(a) Point A

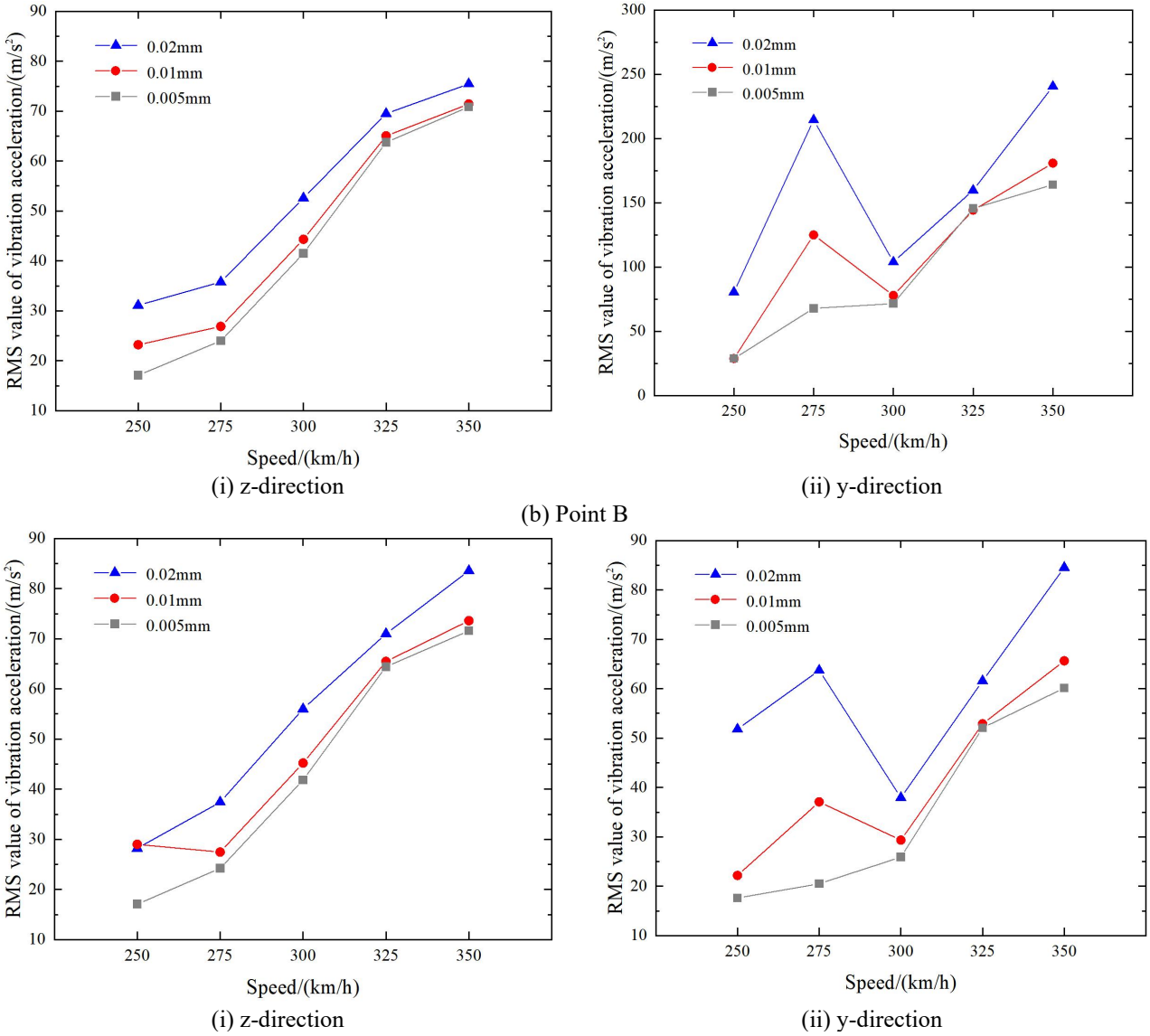


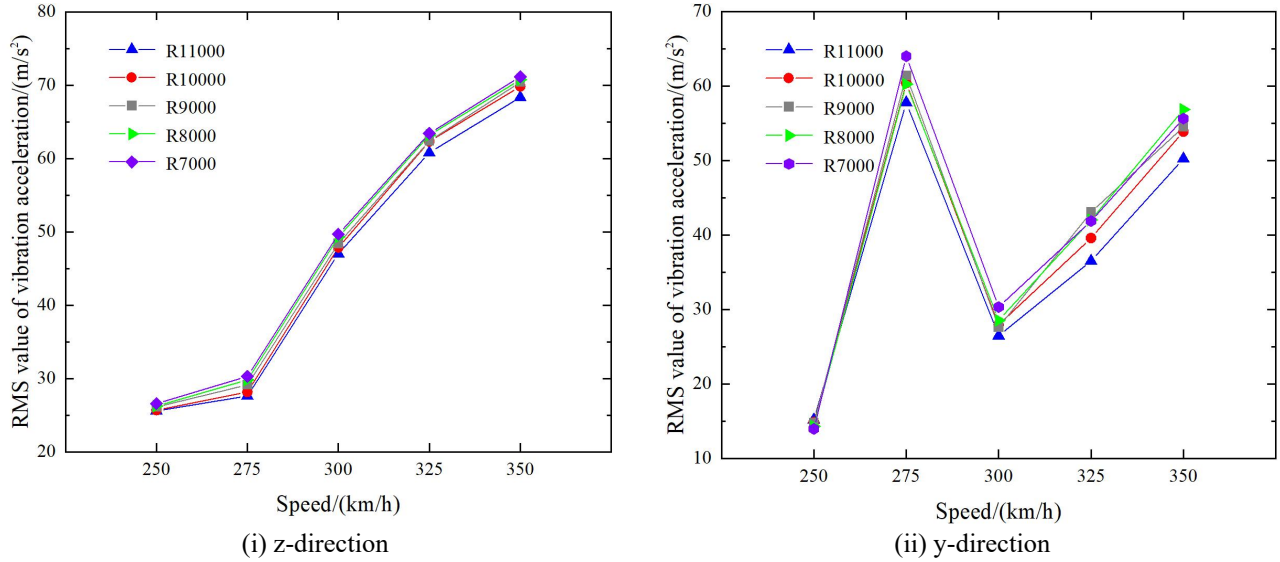
Figure 5 Effect of Wheel Polygon Amplitude on the RMS Value of Gearbox Housing Vibration Acceleration at Different Speeds

5.2 Effect of Curve Radius on Gearbox Housing Vibration Characteristics

Curve negotiation performance is a key indicator for evaluating the operational quality of electric multiple units (EMUs), and the vibration characteristics of the gearbox under curve negotiation conditions are also a major research focus. Taking the Beijing-Shanghai High-Speed Railway, a trunk line designed for a speed of 380 km/h, as an example, the minimum curve radius standard for its alignment is 7,000 meters. To systematically investigate the influence of curve radius on gearbox housing vibration, this study selects a range of five typical curve radii for detailed simulation analysis: 7,000 m, 8,000 m, 9,000 m, 10,000 m, and 11,000 m. The operational speed is set within the common high-speed range of 250–350 km/h. Track parameters, including superelevation and transition curve lengths, are configured according to real-world design specifications and operational scenarios to ensure the simulations accurately reflect practical dynamic interactions.

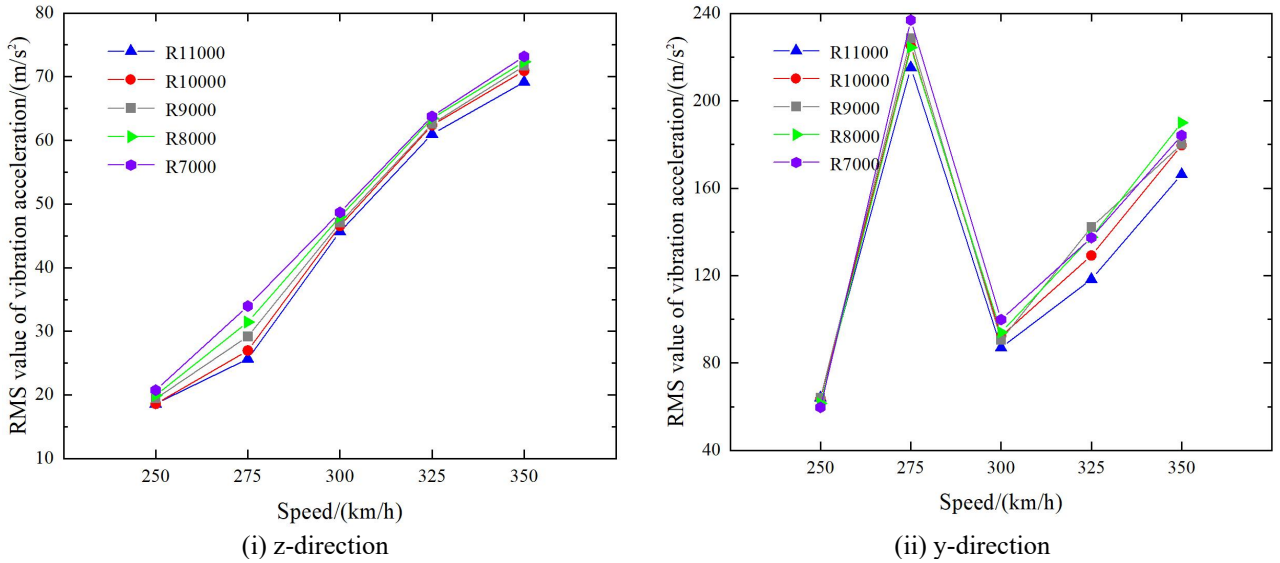
The results indicate a clear relationship between curve radius and vibration intensity: smaller curve radii generally induce more intense gearbox vibrations. In the vertical (z) direction, vibration amplitude consistently decreases as the curve radius increases, showing a clear monotonic trend. However, the lateral (y-direction) vibration exhibits a more complex, speed-dependent behavior. At 250 km/h, y-direction vibration amplitude actually increases with larger curve radii, contrary to the overall trend. Under other speeds, it decreases as the radius increases. Notable local vibration peaks are observed at specific speed-radius combinations: at 275 km/h and 325 km/h, a local peak occurs at a curve radius of 9,000 meters; at 350 km/h, a local peak appears at 8,000 meters.Spectrum analysis reveals that, under these specific conditions, the combination of speed and curve radius intensifies wheel-rail excitation, which in turn amplifies gearbox vibration. Furthermore, comparing the vibration response across the three measurement points (A, B, C), the y-direction patterns are consistent with those observed under straight-line operation. In the z-direction, the location of

maximum vibration shifts with speed: it is initially at point A under lower speeds, but gradually transitions to point C as the operating speed increases.



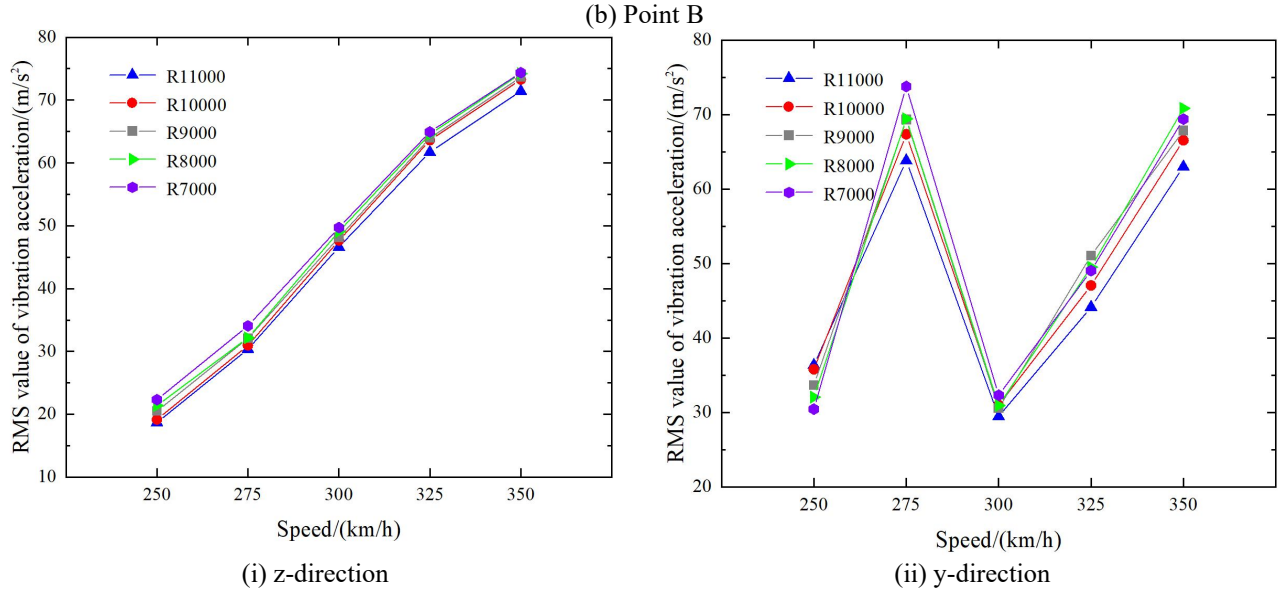
(i) z-direction (ii) y-direction

(a) Point A



(i) z-direction (ii) y-direction

(b) Point B



(i) z-direction (ii) y-direction

(c) Point C

Figure 6 Effect of Curve Radius on the RMS Value of Gearbox Housing Vibration Acceleration at Different Speeds

6 CONCLUSION

This study investigates the correlation between gearbox housing vibration characteristics and train operating parameters, track conditions, and wheel polygon excitation under the Beijing-Shanghai High-Speed Railway scenario by establishing a vehicle dynamics model, combined with experimental data validation and multi-condition simulation analysis. The main findings are summarized as follows:

(1) As the operating speed increases, the vibration responses at the three measurement points in the Z (vertical) and Y (lateral) directions gradually converge. In the low-speed range, the 11th-order polygon has a more pronounced impact on vibrations, whereas under high-speed conditions, the influence of the 23rd-order polygon becomes more significant. Additionally, the Y-direction vibration response is more prominent at the upper measurement point of the gearbox (e.g., Point B), while the Z-direction vibration is more noticeable at the lower measurement point (e.g., Point C). These two types of locations should be regarded as key areas of focus. When the speed exceeds 300 km/h, the influence of the polygon order on the root mean square (RMS) values of vibration acceleration at each measurement point diminishes, and speed becomes the dominant factor affecting vibration characteristics. Notably, under the 23rd-order polygon condition, gearbox vibration intensifies at 275 km/h. Furthermore, as the amplitude of the wheel polygon increases, the vibration intensity of the gearbox exhibits an increasing trend.

(2) When the train passes through curved sections, a smaller curve radius leads to more severe gearbox vibrations. Specifically, the Z-direction vibration decreases monotonically as the curve radius increases, while the Y-direction vibration exhibits differentiated patterns depending on the operating speed. At 250 km/h, the Y-direction vibration increases with larger curve radii, whereas under other speed conditions, it generally decreases with increasing curve radii. It is worth noting that under the 275 km/h and 325 km/h conditions, a local peak in Y-direction vibration occurs at a curve radius of 9,000 meters. In contrast, under the 350 km/h condition, the local peak appears at a curve radius of 8,000 meters.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is supported by the Fundamental Research Project of China Academy of Railway Sciences Corporation Limited (2022Y1303).

REFERENCES

[1] Parey A, Badaoui E M, Guillet F, et al. Dynamic modelling of spur gear pair and application of empirical mode decomposition-based statistical analysis for early detection of localized tooth defect. Journal of Sound and Vibration, 2005, 294(3): 547-561.

[2] Pandya Y, Parey A. Crack behavior in a high contact ratio spur gear tooth and its effect on mesh stiffness. Engineering Failure Analysis, 2013, 34: 69-78.

[3] Carbonelli A, Rigaud E, Perret-Liaudet J. Vibro-acoustic analysis of geared systems-predicting and controlling the whining noise. Automotive NVH Technology, 2016, 10: 63-79.

[4] Kahraman A, Singh R. Non-linear dynamics of a geared rotor- bearing system with multiple clearances. Journal of Sound and Vibration, 1991, 144(3): 469-506.

[5] Zhang B, Tan C A, Lin J. Gearbox fault diagnosis of high-speed railway train. Engineering Failure Analysis, 2016, 66: 407-420.

[6] Wang W J, Yan R G, Li G Q, et al. Dynamic characteristics of Chinese standard EMU gearbox housing. Journal of Vibration Engineering, 2019, 32(3): 534-539.

[7] Wang W J, Li G Q, Han J C, et al. Influence rule of dynamic stress of high-speed train gearbox housing. Journal of Traffic and Transportation Engineering, 2019, 19(1): 85-95.

[8] Huang G, Zhou N, Zhang W. Effect of internal dynamic excitation of the traction system on the dynamic behavior of a high-speed train. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2016, 230(8): 1899-1907.

[9] Hu W, Liu Z, Liu D, et al. Fatigue failure analysis of high speed train gearbox housings. Engineering Failure Analysis, 2017, 73: 57-71.

[10] Zhang B, Tan C A, Lin J. Gearbox fault diagnosis of high-speed railway train. Engineering Failure Analysis, 2016, 66: 407-420.

[11] Yue Zhou. Dynamic behavior of high-Speed train helical gear transmission systems. Beijing: Beijing Jiaotong University, 2024.

[12] Zhiwei Wang. Dynamic modeling and coupled vibration analysis of gear and bearing Systems in high-speed trains under service conditions. Chengdu: Southwest Jiaotong University, 2019.

[13] Liu Y, Chen Z, Wang K, et al. Dynamic characteristics analysis of gear transmission and its support bearings of high-speed train on the curve. Vehicle System Dynamics, 2024, 62(3): 623-650.

[14] Yang J H. Study on vibration behavior and prevention measures of polygon wheels of EMUs. Chengdu: Southwest Jiaotong University, 2022.
[15] Zou H Y. Study on vibration characteristics and fatigue strength of high-speed train gearbox considering out-of-round wheel. Chengdu: Southwest Jiaotong University, 2018.
[16] Wang Y. Study on influence of out-of-round railway wheels on wheel/rail interaction force. Chengdu: Southwest Jiaotong University, 2016.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3066

CROP PLANTING STRATEGIES BASED ON PARTICLE SWARM OPTIMIZATION ALGORITHM

YanZhuo Wu, GuangWu Ao*
School of Computer Science and Software Engineering, University of Science and Technology, Anshan 114051, Liaoning, China.
**Corresponding Author: GuangWu Ao*

Abstract: In order to solve the problems of efficient utilization of limited cultivated land resources and sustainable agricultural development in the North China Plain, this study aims to maximize total profits from 2024-2030, and combines field inspection data to build a planting strategy optimization model. First, the 2023 crop yield per mu and planting area data are processed through the VLOOKUP function to determine the expected future sales volume; the lowest sales price is extracted and profits are calculated based on planting costs. After Python data cleaning and standardization, SPSSPRO is used to visually analyze the profit and income ratio of different plots (flat dry land, terraced fields, etc.) to screen high-quality crops. For the two scenarios of "total waste of excess sales" and "price reduction of excess sales", corresponding objective functions are constructed, constraints such as plot planting cycle, continuous cropping taboos, and bean crop rotation are introduced, and particle swarm optimization (PSO) is used to solve the optimal solution through Python. The results show that the model can effectively balance resource utilization, cost control and ecological constraints, and provide quantitative support for scientific planting decisions in the North China Plain.
Keywords: Particle swarm optimization; Crop planting strategy optimization; North China Plain; Objective function; Cultivated land resource utilization

1 INTRODUCTION

Agriculture is one of the most important industries in the world; it is my country's largest industry; and it is the foundation of national economic development [1]. Agriculture plays a vital role in food supply, economic growth and social stability. Since the 40 years of reform and opening up, the structural adjustment of planting and agricultural industries has made great contributions to ensuring my country's grain production and food supply, and has promoted my country's great achievements in the development of modern agriculture. However, the core and foundation of agriculture is planting. Therefore, the adjustment of the structure of the planting industry is very critical to the adjustment of the agricultural structure. As the times advance and technology develop, people are increasingly pursuing scientific planting. In exploration and practice, people have found that optimizing and adjusting the crop planting structure through local actual conditions and local actual crop planting statistics can improve the adaptability of crops and make crops more suitable for the local environment. Therefore, formulating planting plans, improving and updating them can not only meet people's daily food needs, but also bring good economic benefits to achieve sustainable agricultural development [2].

The north and south regions of the North China Plain span large latitudes, and there are obvious differences in temperature, precipitation and other conditions between north and south. Therefore, people's production activities will be affected accordingly, resulting in obvious differences in key phenological periods of crops [3]. We must take a long-term perspective on crop planting issues, not affect the subsequent growth of crops, and not damage the soil environment because of immediate benefits. Planting strategies should follow the growth laws of crops. Just as the same crop or crops of the same family cannot be continuously planted on a piece of land. In order to achieve sustainable development, keep the land healthy and suitable for crop growth, the land can be used to plant beans at least once within three years. crop. Therefore, there are many constraints in the process of formulating a reasonable planting strategy. There are many influencing factors to judge whether it is suitable for planting, and the data statistical process is complex. Therefore, it is necessary to establish a reliable model and then obtain a reasonable planting plan. It is conducive to achieving efficient and sustainable development of agriculture under the rigid constraints of the market and resources and the strategic requirements of the new food security concept, and embarking on the path of agricultural modernization with product safety, resource conservation, and environmental friendliness [4-6].

In view of the complexity of planting scenarios and the diversity of constraints in the North China Plain, and the lack of systematic data processing and intelligent algorithm support for existing planting schemes, making it difficult to achieve the optimal balance between resource utilization and profits, the research purposes of this paper are as follows: First, based on the field planting data in 2023, the expected crop sales volume is determined through the VLOOKUP function, the lowest sales price is screened, and data cleaning and standardization are completed based on planting costs to clarify the distribution characteristics of crop profits in different plots; The second is to build a functional model with the goal of maximizing total profit from 2024-2030 for the two market scenarios of "all waste of excess sales" and "sales at reduced prices for excess sales", and incorporate the plot planting cycle, continuous cropping taboos, ecological and

production constraints such as bean rotation; The third is to use particle swarm optimization algorithm to solve the optimal solution through Python programming, and finally form a scientific planting strategy that adapts the characteristics of the North China Plain land and takes into account economic benefits and ecological sustainability, providing quantitative basis for optimizing local agricultural planting structure and efficient utilization of resources. Decision basis.

2 MATERIALS AND METHODS

2.1 Data

The research data were obtained through local field investigations and statistical yearbooks.

2.1.1 Expected sales volume

First, use the vlookup function to correspond the yield per mu in the 2023 planting situation table, multiply the yield per mu by the planting area to obtain the 2023 production data, and use the 2023 production as the total sales volume as the expected sales volume in the future.

2.1.2 Standardized treatment

In order to ensure accurate data analysis and simple data cleaning, python was used to type convert the crop number, convert non-numeric values into NAN, and then replace NAN with-1 to identify invalid tags. Finally, the invalid numbers were filtered. At the same time, the data is standardized to prevent the scope from being too large and causing the proportion of data to decrease and affect analysis.

2.2 Research methods

2.2.1 Research and analysis

If you want to obtain the optimal planting strategy in the future, you first need to reasonably process the expected sales volume, sales unit price, and planting cost data in 2023. Secondly, you need to formulate plans to reduce costs and reduce waste of land resources. Therefore, you need to add a small number of realistic variables. Consider (Such as the total output of crops, market sales, price fluctuations and other factors), set the objective function according to two different processing methods of remaining products, generate code according to the objective function, and finally use the optimization model to obtain the results.

2.2.2 Particle Swarm Optimization (PSO)

Particle swarm algorithm is a method of finding an optimal solution through iteration in multi-dimensional space. Particle swarm algorithm randomly gives the velocity and position of particles in the solution space, and initializes the number of particles, maximum number of iterations, inertia weights, individual learning factors and social learning factors [7]. Then evaluate the fitness of particles according to the objective function, find out the historical best position (pbest) and global best position (gbest) of particles, and then iterate, and continue to evaluate particles using the objective function. If the particle's new fitness is better than the historical fitness, update the position to the historical best position (pbest), compare the best positions of all particles again, and find a global best position (gbest) until a termination condition is encountered [8].

The iteration ends. Each potential solution is called a particle. Each particle remembers the historical best position (pbest) and shares a global best position (gbest). Each particle adjusts based on these two positions to obtain the optimal solution.

Applicable conditions are as follows:

Condition 1: Particle swarm optimization solves continuous problems, that is, decision variables can be taken within a continuous range.

Condition 2: The objective function is computable in order to find the optimal solution based on fitness.

Condition 3: Progress during the search process constantly adjusts the PSO parameters so that it does not converge prematurely due to obtaining a local optimal solution.

Condition 4: Since particle swarm optimization can store past information to predict future information, the objective function or constraint conditions can change over time.

The velocity of particle a is represented by $V_a = (v_a1, v_a2, ..., v_aD)$. For each iteration, its d-th dimension ($1 \le d \le D$) changes according to the following equation:

$$V_{ad}^{(t+1)} = wv_{ad}^{(t)} + c_1 r_1 \left(p_{ad} - x_{ad}^{(t)}\right) + c_1 r_2 \left(g_d - x_{ad}^{(t)}\right) \quad (1)$$

$$X_{ad}^{(t+1)} = X_{ad}^{(t)} + V_{ad}^{(t+1)} \quad (2)$$

Where w is the inertia weight, c1 and c2 are both acceleration constants, and rand () and Rand () are random numbers varying in the range of [0,1] respectively. The superscript (t) denotes the state at the t-th iteration, while (t+1) indicates the updated state after the current iteration. The original notation vid is corrected to $V_{ad}^{(t+1)}$ (updated velocity), and V_{ad} is corrected to $V_{ad}^{(t)}$ (current velocity).

3 RESULTS

3.1 Crop Profit Data Processing and Visualization Analysis

In 2023, the yield per mu of relevant data will be multiplied by the sales unit price to obtain the minimum sales amount and the maximum sales amount per mu. Then, the two sales amounts will be subtracted from the planting cost to obtain the two profits per mu. The maximum profit will be divided by the cost to obtain the maximum profit ratio. Using the pivot table, using the two profits and income ratios as values, and crop names and plots as columns to obtain new tabular data, import it into spsspro visualization, and screen out crops with poor profits., narrowing the scope of the data. The relevant results are shown in Figures 1-6.

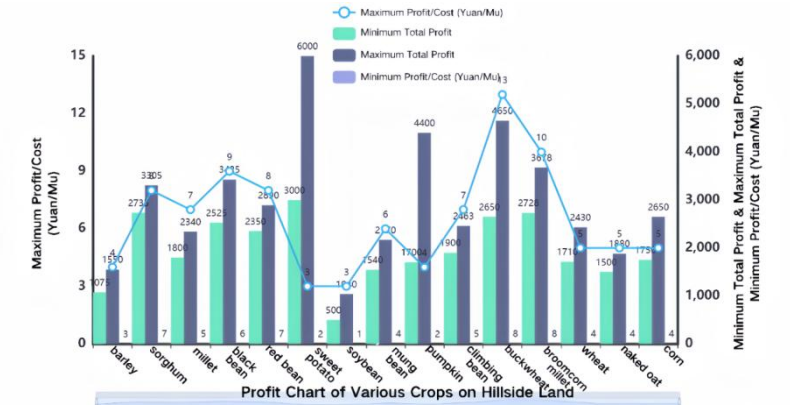


Figure 1 Profit Map of Various Crops on Hillside Land

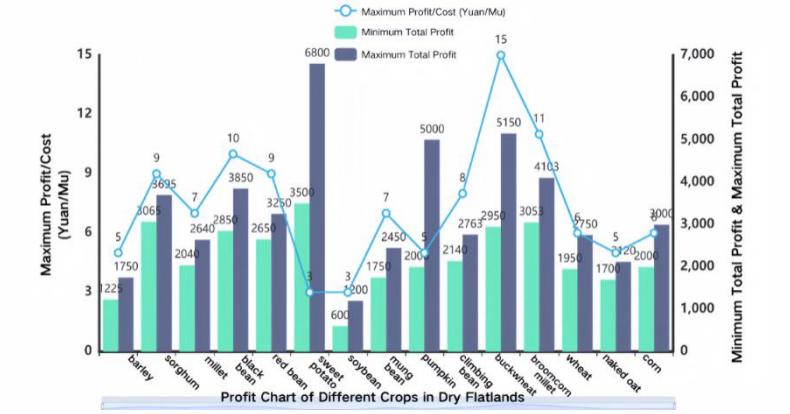


Figure 2 Profit Map of Different Crops in Dry Flatlands

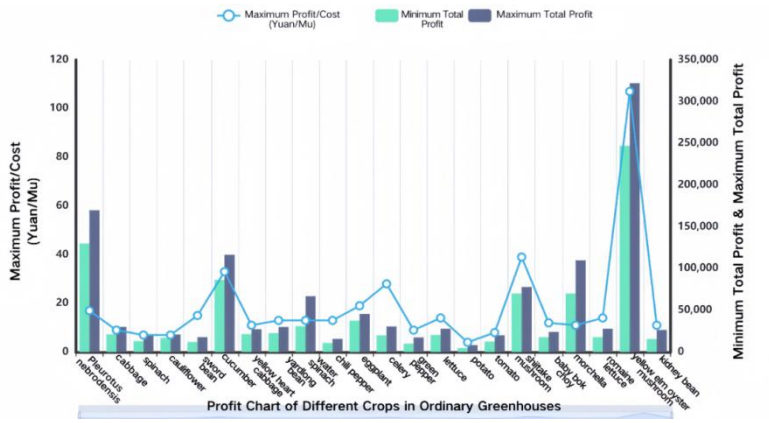


Figure 3 Profit Map of Different Crops in Ordinary Greenhouses

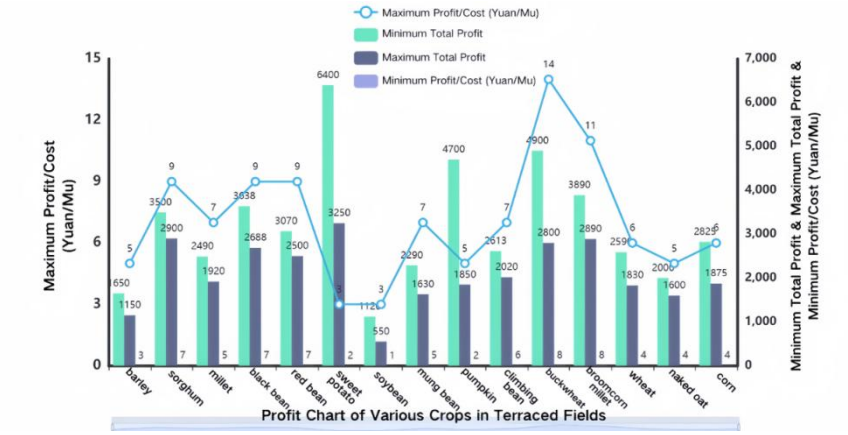


Figure 4 Profit Map of Each Crop in Terraced Fields

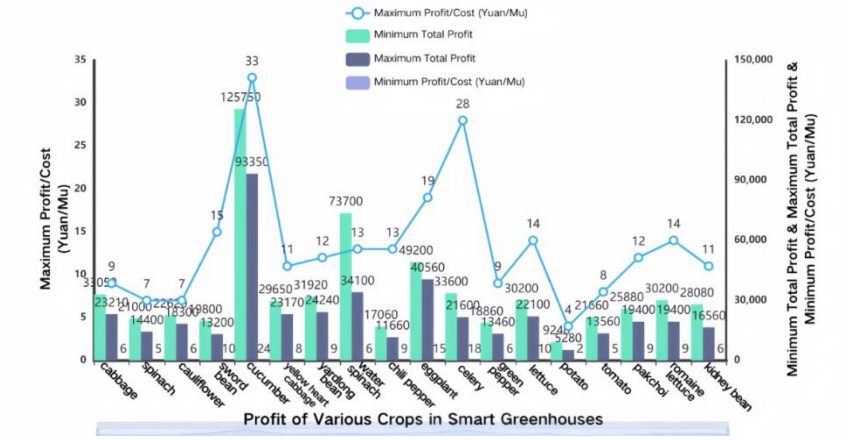


Figure 5 Profits of Various Crops in Smart Greenhouses

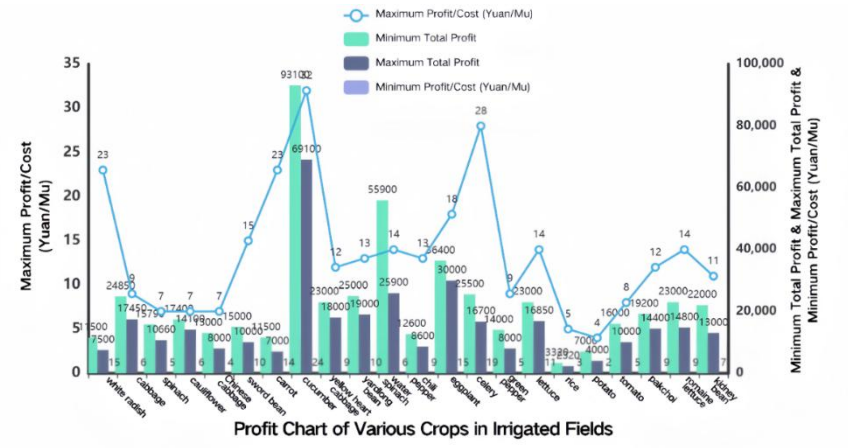


Figure 6 Profit Chart of Each Crop in Irrigated Land

Based on the statistical data from 2023 and adhering to a consistent analytical framework, this study conducted an analysis of the planting profits of different crops in various environments, including hillside land, dry flat land, ordinary greenhouses, terraced fields, smart greenhouses, and irrigated land. First, we calculated the product of crop yield per mu (a Chinese unit of area, approximately 0.067 hectares) and the selling price in each environment to determine the minimum and maximum sales revenue per mu. After deducting the costs, we derived the minimum and maximum profits as well as the maximum profit ratio. Subsequently, we used pivot tables to consolidate the data and generate new tables, which were then imported into SPSSPRO software to create profit charts.

From these charts, it is evident that there are extremely significant differences in crop profits across different environments. For instance, in smart greenhouses, the maximum total profit per mu for knife beans (a type of legume) reached 125,750 yuan, while in irrigated land, the maximum total profit per mu for carrots reached 93,100 yuan. However, some crops yielded relatively low profits. These analytical results provide crucial data references for planting management in various environments, facilitating the rational selection of crops, optimization of planting structures, maximization of planting benefits, and promotion of efficient and precise agricultural development.

3.2 Construction of Total Profit Objective Function Based on Different Sales Scenarios

In order to obtain the optimal planting strategy, the total profit from 2024-2030 is used as the objective function. Because there are many data variables, the profit of each crop in different plots during the year is first calculated. Ask for the profit of all crops for one year, and finally ask for the total profit for seven years. According to the two different conditions in the title, the objective function is constructed as follows:

Objective function 1:

$$\text{Max: } Z_1=\sum_{r=0}^{7}\sum_{f=1}^{2}\sum_{i}\left(p_i\cdot\frac{\min\left(\sum_v x_{i,v,f,r}\cdot q_i,D_i\right)}{q_i}-c_i\cdot\sum_v x_{i,v,f,r}\right) \tag{3}$$

Objective function 2:

$$\text{Max: } Z_2=\sum_{r=0}^{7}\sum_{f=1}^{2}\sum_{i}\left(p_i\cdot\frac{\min\left(\sum_h x_{i,v,f,r}\cdot q_i,D_i\right)}{q_i}+p_i'\cdot\frac{\max\left(\sum_h x_{i,v,f,r}\cdot q_i-D_i,0\right)}{q_i}-c_i\cdot\sum_b x_{i,v,f,r}\right) \tag{4}$$

Among them, Z_1 and Z_2 represent the maximum profits, p_i represents the selling price per mu of the i-th crop, q_i is the kilogram yield per mu of the i-th crop, D_i represents the expected kilogram sales of the i-th crop, and c_i is the planting cost per mu of the i-th crop.

The decision variable is: $x_{i,b,f,r}$, which represents the area (unit: mu) of the i-th crop planted on the h-th plot or greenhouse in the f-th quarter in the r-th year.

3.3 The Implementation of Particle Swarm Optimization Algorithm with Integrated Planting Constraints

Add constraints: Among them, crop type M_i: represents the type of crop i (such as grain, grain (bean), vegetable, vegetable (bean), edible fungus). Plot type L_h: Represents the type of plot h (such as flat dry land, terraced fields, sloping land, irrigated land, ordinary greenhouses, smart greenhouses).

(1) For "ordinary greenhouses" (L_v=L_5), one season of vegetables and one season of edible fungi are planted each year, and the area of the vegetable and edible fungi planting does not exceed the total area of all plot types.

$$\sum_{i\in\{i\mid M_i=M_3\}}x_{i,v,1,r}+\sum_{i\in\{i\mid M_i=M_5\}}x_{i,v,2,r}\le A_v,\ \forall v \text{ with } L_v=L_5,r\in\{0,1,2,3,4,5,6,7\} \tag{5}$$

(2) For "Smart greenhouses" (L_v=L_6), two vegetable planting cycles can be achieved each year:

$$\sum_{i\in\{i\mid M_i=M_3\}}x_{i,v,f,r}\le A_v,\ \forall v \text{ with } L_v=L_6,f\in\{1,2\},r\in\{0,1,2,3,4,5,6,7\} \tag{6}$$

(3) The same crop cannot be continuously planted on the same plot of land.

$$x_{i,v,f,r}\cdot x_{i,v,f,r+1}=0,\ \forall i,v,f\in\{1,2\},r\in\{0,1,2,3,4,5,6,7\} \tag{7}$$

(4) Dry land, terraced fields and sloping land complete one crop cycle each year:

$$\sum_{f=1}^{2}\sum_{i}x_{i,v,k,r}\le A_v,\ \forall v \text{ with } L_v\in\{L_1,L_2,L_3\},\forall r \tag{8}$$

(5) The irrigated land can be grown with water straw in a single season or with rice straw crops in two seasons each year:

$$\begin{gathered}x_{i,v,1,r}+x_{i,v,2,r}\le A_v,\ \forall v \text{ with } L_v=L_4,\forall i \text{ with } M_i=M_1,i=C_{16},\forall r\\ \sum_{i\in\{i\mid M_i\in\{M_3,M_4\}\}}x_{i,v,2,r}\le A_v\cdot b_{v,r},\ \forall v \text{ with } L_v=L_4,\forall r\end{gathered} \tag{9}$$

Finally, the solution matrix is obtained through Python using particle swarm optimization algorithm.

4 CONCLUSIONS

Focusing on the goals of efficient utilization of cultivated land resources and sustainable agricultural development in the North China Plain, this study forms a complete set of optimization plans for crop planting strategies through data processing, model construction and intelligent algorithm solution: first based on 2023 field data, VLOOKUP function, data standardization and visualization analysis to clarify the distribution of crop profits and the range of high-quality crops; An objective function is then constructed for two residual product processing scenarios, combined with constraints such as plot characteristics (such as "vegetable + edible fungus" rotation in ordinary greenhouses, double-season vegetable planting in smart greenhouses), continuous cropping taboos, and rotation requirements, and accurately solved with the help of particle swarm optimization algorithm. The optimal planting plan from 2024 to 2030 effectively achieves a balance between maximizing profits and minimizing resource waste, while taking into account soil health and ecological sustainability. In the future, the research dimension can be further expanded to include variables such as the impact of climate change on crop yields, dynamic fluctuations in market prices, and adaptability of new crop varieties, optimize the inertia weights and learning factor parameters of the particle swarm optimization algorithm, and improve the adaptability of the model to complex agricultural scenarios.; GIS geospatial analysis technology can also be combined to refine planting strategies at different latitudes and climate zones, promote the popularization and application of research results to larger areas, and provide more comprehensive technical support for modern agricultural precision planting.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Liu Tao. Discussion on countermeasures to optimize crop planting. Agricultural Technology and Information, 2021(22): 73-74.
[2] Huang Malan. The impact of rural labor transfer and rising prices on changes in crop planting structure. Huazhong Agricultural University, 2019.
[3] Xue Bingyu, Wang Cong, Hu Qiong, et al. Analysis of spatio-temporal variation characteristics and influencing factors of key phenological periods of main crops in Hubei Province. Journal of Ecology, 2025(18): 1-16.
[4] Xia Wen, Shi Yufei, Zou Shuai, et al. Optimization of multi-terrain and multi-crop planting strategies based on particle swarm algorithm. Smart Agriculture Guide, 2025, 5(17): 49-53.
[5] Wang Xin. Restrictions on the promotion of corn planting in saline-alkali land in Cangzhou and countermeasures. Agricultural Technology and Equipment, 2025(08): 34-36.
[6] Wang Lin, Guo Yaxin. Research on crop planting strategies based on linear programming model. Journal of Henan Institute of Technology, 2025, 33(04): 27-30.
[7] Mohankumar A, Duraisamy T, Packkirisamy V. Optimizing cold spray process parameters for AA2024/Al2O3 coatings to minimize wear loss via response surface methodology and particle swarm optimization. Journal of Adhesion Science and Technology, 2025, 39(17): 2686-2709.
[8] Liu Z, Hu Y, Fang Z, et al. Improved prediction model for daily PM2.5 concentrations with particle swarm optimization and BP neural network. Scientific Reports, 2025, 15(1): 32050-32050.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3067

DYNAMIC MODELING AND ANALYSIS OF IN-WHEEL SUSPENDED ELECTRIC WHEEL SYSTEM

Peng Zhang*, Lei Jiang, ShangQing Liu, ShengShuo Yan, JingXin Bai, Ping Zhang
School of Mechanical Electronic and Information Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China.
**Corresponding Author: Peng Zhang*

Abstract: The introduction of in-wheel motors increases the unsprung mass of electric vehicles, exacerbates vibrations when the vehicle is subjected to road excitation, and affects ride comfort and stability. The in-wheel suspended electric wheel has become an important technical means to improve the above-mentioned problems of in-wheel motors. This paper conducts a vertical dynamic characteristic analysis of two electric wheel configurations. A 1/4 vehicle model was established based on the Matlab/Simulink platform, and the vibration response characteristics of the two electric wheels under different road excitations were obtained. Experimental tests were performed on the vibration characteristics of the two electric wheels using a swept-sine road excitation. The experimental results were in good agreement with the simulation results, verifying the reliability and accuracy of the simulation model.
Keywords: In-wheel motor; Electric wheels; Internal suspension; Unsprung mass

1 INTRODUCTION

In-wheel motor (IWM) drive technology, which simplifies the powertrain system and enhances chassis layout flexibility, has become a key area in pure electric vehicle (EV) development. However, the integration of IWM significantly increases the unsprung mass of EV, leading to aggravated vertical vibrations under road excitation, which severely compromises ride comfort and handling stability [1]. This “negative unsprung mass effect” constitutes a major bottleneck hindering the widespread application of IWM drive technology. To address this issue, extensive research has been conducted on optimizing electric wheel configurations. The in-wheel suspended electric wheel, which integrates the motor into the wheel as a dynamic vibration absorber (DVA) via dedicated spring-damper mechanisms, effectively reduces the proportion of unsprung mass, offering a viable path to mitigate vibration problems. Current research has preliminarily validated the vibration improvement potential of suspended configurations. For instance, Rojas et al. compared the effects of passive, semi-active/active suspensions, and motor suspension devices on wheel dynamic load and body vibration acceleration, concluding that motor suspension significantly improves body vibration acceleration [2]. Chen et al. introduced a multifunctional integrated electric wheel structure with in-wheel suspension, which avoids the increase in unsprung mass [3]. Zhao et al. proposed a motor stator suspension scheme that transferred the motor stator mass to the sprung mass, resulting in body vertical vibration acceleration peaks comparable to those of centrally driven EV [4]. Li investigated the influence laws of the mass of mounting components in the dynamic vibration absorber and the mass of the motor as a part of the unsprung mass on the vertical dynamic performance of the vehicle, and obtained the optimal mass ranges of the two which were beneficial to the improvement of ride comfort [5]. Meng et al. studied a new type of electric wheel integrating a flexible transmission mechanism and a vibration damping mechanism. In this electric wheel, power transmission and vertical vibration were no longer coupled with each other, and could be designed or controlled separately, thus achieving favorable longitudinal and vertical dynamic characteristics of the vehicle [6]. However, existing studies often focus on single road conditions or localized performance optimization. Aiming to improve the vertical vibration performance of IWM-driven vehicles, this paper compares the vibration response characteristics of two electric wheel configurations under different road excitation and validates the simulation results through swept-sine excitation bench tests.

2 DYNAMIC MODELING OF THE ELECTRIC WHEEL SUSPENSION SYSTEM

2.1 Mathematical Model of the Electric Wheel

2.1.1 Mathematical model of the fixed in-wheel electric wheel

In most IWM-driven vehicles, the motor stator is fixed to the axle or steering knuckle arm through bearings, while the rotor is fixedly connected to the wheel rim. This configuration is termed the fixed in-wheel drive. The quarter-car model for the fixed in-wheel electric wheel suspension system is shown in Figure 1.

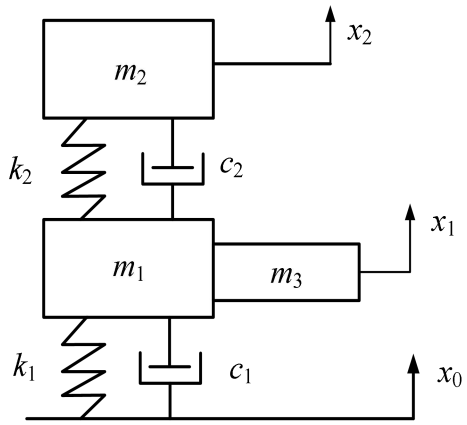


Figure 1 Quarter-Car Model of the Fixed In-Wheel Electric Wheel Suspension System

The vertical vibration differential equations for the fixed in-wheel vehicle are:

$$\begin{cases} (m_1+m_3)\ddot{x}_1 + c_1(\dot{x}_1 - \dot{x}_0) + c_2(\dot{x}_1 - \dot{x}_2) + k_1(x_1 - x_0) + k_2(x_1 - x_2) = 0 \\ m_2\ddot{x}_2 + c_2(\dot{x}_2 - \dot{x}_1) + k_2(x_2 - x_1) = 0 \end{cases} \tag{1}$$

where m_1 is the unsprung mass excluding the IWM system; m_2 is the vehicle sprung mass; m_3 is the IWM system mass; k_1 and c_1 are the tire stiffness and damping; k_2 and c_2 are the suspension stiffness and damping; x_0、x_1、x_2 are the road excitation, wheel vertical displacement, and body vertical displacement, respectively.

2.1.2 Mathematical model of the in-wheel suspended electric wheel

The in-wheel suspended electric wheel connects the motor to the unsprung mass via dedicated springs and dampers, effectively utilizing the motor itself as a DVA. This structure significantly reduces the unsprung mass, making it comparable to conventional EV. Furthermore, by introducing dampers, the DVA effect is achieved without increasing the total vehicle mass, offering superior performance compared to traditional fixed in-wheel drive EV. The quarter-car model of the electric wheel suspension system based on the DVA is shown in Figure 2.

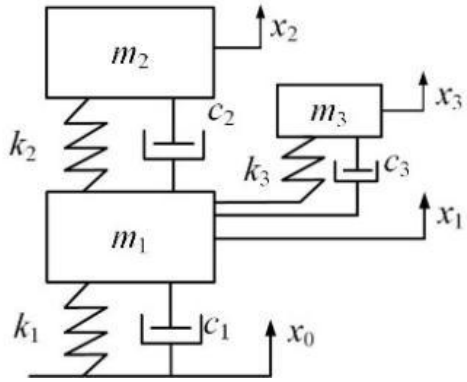


Figure 2 Quarter-Car Model of the In-Wheel Suspended Electric Wheel Suspension System

The vertical vibration dynamic equations for the in-wheel suspended vehicle are：

$$\begin{cases} m_1\ddot{x}_1 = k_1(x_0 - x_1) + c_1(\dot{x}_0 - \dot{x}_1) + k_2(x_2 - x_1) + c_2(\dot{x}_2 - \dot{x}_1) + k_3(x_3 - x_1) + c_3(\dot{x}_3 - \dot{x}_1) \\ m_2\ddot{x}_2 = k_2(x_1 - x_2) + c_2(\dot{x}_1 - \dot{x}_2) \\ m_3\ddot{x}_3 = k_3(x_1 - x_3) + c_3(\dot{x}_1 - \dot{x}_3) \end{cases} \tag{2}$$

where m_1、m_2、m_3 represent the unsprung mass, quarter-car sprung mass, and motor mass, respectively; k_1、k_2、k_3 represent the tire vertical stiffness, vehicle suspension stiffness, and motor DVA system stiffness, respectively; c_1、c_2、c_3 represent the tire vertical damping, vehicle suspension damping, and motor DVA system damping, respectively. x_0 denotes the road excitation, x_1、x_2、x_3 denote the vertical displacements of the respective masses.

2.2 Road Excitation Models

2.2.1 Random road

The random road model based on spectral analysis is a classical approach. Roads of different roughness levels are characterized by power spectral density (PSD) functions as input signals. The PSD of the road vertical velocity is calculated as follows:

$$G_q(f)=4p^2G_q(n_0)n_0^2u \tag{3}$$

where $G_q(n_0)$ is the road PSD value at the reference spatial frequency n_0=0.1m^{-1}. This paper selects a Class E road surface with a roughness coefficient $G_q(n_0)$ =4.1×10^{-3}m^3. The vehicle speed is u=20km/h.

2.2.2 Bump road

The bump road simulates obstacles such as speed bumps. The cross-section of a speed bump can be approximated by the cosine function over the interval [−π/2, π/2]. It is mathematically expressed as:

$$X_r = \begin{cases} \frac{1}{2}A\left[1+\cos(2\pi\frac{L-L_0}{\lambda}-\pi)\right] & L_0 \le L \le L_0+\lambda \\ 0 & \text{else} \end{cases} \tag{4}$$

where X_r is the vertical displacement (m), L is the vehicle travel distance (m), the bump height is A=0.05m, the bump length is λ=0.25m, and the distance from the starting point to the bump is L_0=5m.

2.2.3 Swept-sine road

Swept-sine excitation is commonly used in vibration studies. Since the natural frequencies of most vehicle bodies, seats, and wheels range from 0.7 Hz to 15 Hz [7], suspension research often employs swept-sine surfaces with linearly varying frequency [8]. It uses vibration spectra from typical road conditions to generate test excitation signals, simulating vibrations on different road surfaces.

The swept-sine road profile is calculated as:

$$X_r(L)=A\sin(\Omega L) \tag{5}$$

Where A is the road amplitude, A=0.1m; L is the vehicle travel distance (m); Ω is the angular spatial frequency of the sine road (rad/m). The vehicle speed increases from 10 m/s to 15 m/s with an acceleration of 1 m/s².

2.3 Performance Evaluation Metrics

To evaluate the vehicle's vertical dynamic characteristics, four metrics are selected: wheel vertical dynamic load, motor vertical vibration acceleration, body vertical acceleration, and body suspension dynamic deflection. The wheel vertical dynamic load represents the adhesion effect between the tire and the road. As the vertical dynamic load of the wheel decreases, the traction performance of the vehicle also decreases accordingly. The motor vertical vibration acceleration reflects the motor's vertical vibration amplitude. The body vertical acceleration indicates the vibration level experienced by the vehicle and its occupants. The body suspension dynamic deflection relates to the spatial requirements of the structure; a larger dynamic deflection increases the likelihood of suspension-contact with limit stops, implying reduced comfort. The calculation formulas are as follows:

$$\begin{cases} F_r = k_1(x_1 - x_0) + c_1(\dot{x}_1 - \dot{x}_0) \\ a_1 = \ddot{x}_2 \\ y = x_3 - x_1 \\ a_2 = \ddot{x}_3 \end{cases} \tag{6}$$

Where Fr is the wheel vertical dynamic load, a_1 is the motor vertical acceleration, y is the body suspension deflection, a_2 is the body vertical acceleration.

To reflect the vertical characteristics of the vehicle or motor at each frequency within the studied band, the relative amplitude-frequency characteristics of each metric are used for evaluation, the calculation formula is as follows:

$$\begin{cases} J_1 = \left|\dfrac{\widehat{F_r}}{G\widehat{x_0}}\right| \\ J_2 = \left|\dfrac{\widehat{\ddot{x}_2}}{\widehat{\dot{x}_0}}\right| \\ J_3 = \left|\dfrac{\widehat{x_1} - \widehat{x_3}}{\widehat{\dot{x}_0}}\right| \\ J_4 = \left|\dfrac{\widehat{\ddot{x}_1}}{\widehat{\dot{x}_0}}\right| \end{cases} \tag{7}$$

where G is the total weight of the quarter-car; J_1、J_2、J_3、J_4 represent the relative amplitude-frequency characteristics of wheel dynamic load, motor vertical acceleration, body suspension dynamic deflection, and body vertical acceleration under road excitation, respectively. "^" denotes the Laplace transform.

3 SIMULATION RESULTS AND ANALYSIS

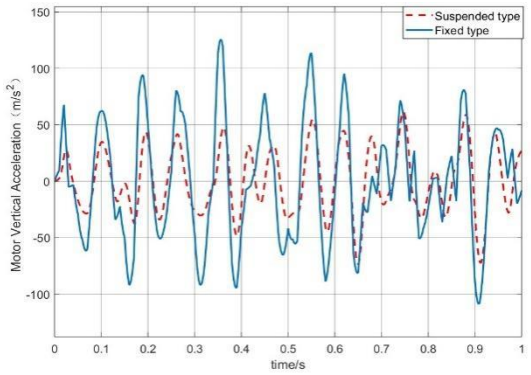
The specific parameters for the in-wheel suspended and fixed configurations are shown in Table 1. A corresponding quarter-car model was established in Matlab/Simulink. Road excitations were applied to the model to simulate vibration response characteristics under different conditions.

Table 1 Simulation Parameters

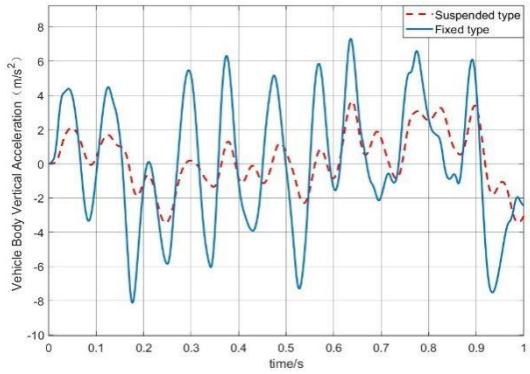
Parameter Symbol	In-Wheel Suspended	Fixed	Parameter Symbol	In-Wheel Suspended	Fixed
m_1(kg)	40	40	k_2(N/m)	32000	32000
m_2(kg)	340	340	c_2(N·s/m)	1496	1496
m_3(kg)	30	30	k_3(N/m)	41000	
k_1(N/m)	360000	360000	c_3(N·s/m)	1000	
c_1(N·s/m)	50	50			

3.1 Simulation Analysis Under Random Road Excitation

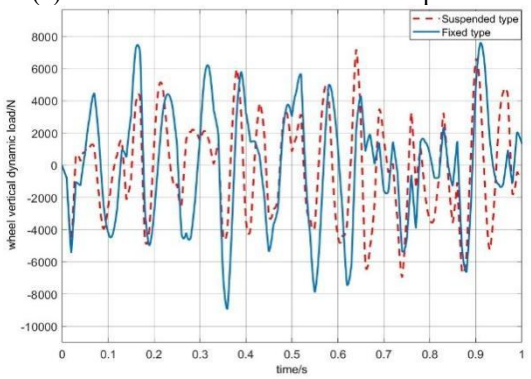
Setting the random road excitation, the time-domain comparisons of various vehicle performance metrics are shown in Figure 3.



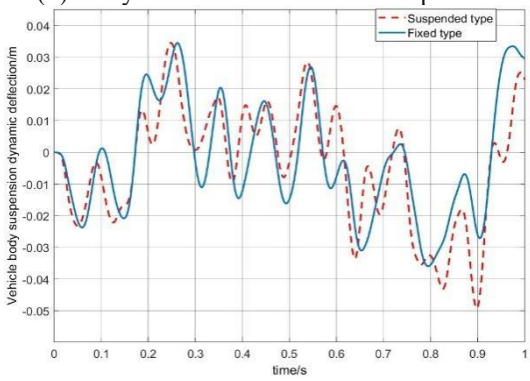
(a) Motor vertical acceleration comparison



(b) Body vertical acceleration comparison



(c) Wheel dynamic load comparison



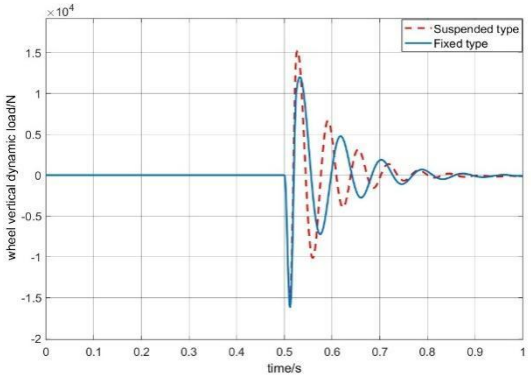
(d) Body suspension dynamic deflection comparison

Figure 3 Time-Domain Comparison of Performance Metrics under Random Road Excitation

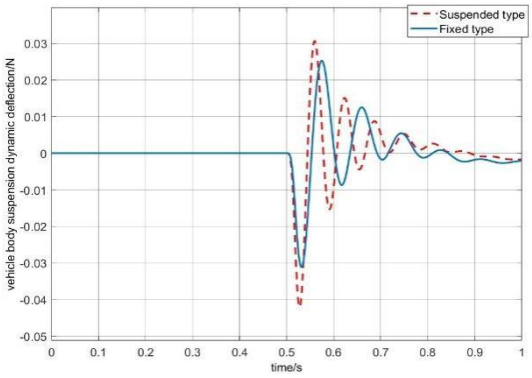
As shown in Figure 3, the in-wheel suspended configuration effectively reduces the vertical vibration of the IWM. Compared to the fixed configuration, the in-wheel suspended motor reduces body vibration levels, thereby improving passenger comfort. The amplitude of the wheel vertical dynamic load for the in-wheel suspended configuration is significantly smaller than that of the fixed configuration, indicating better adhesion between the tire and road, and consequently better traction performance. The body suspension dynamic deflection of the in-wheel suspended configuration is similar to that of the fixed configuration. And it is even larger than the fixed configuration's deflection between 0.8s and 0.9s. This suggests that the increased unsprung mass due to placing the motor inside the wheel is not fully mitigated by suspending the motor, and the issue of potential suspension-contact with limit stops persists.

3.2 Simulation Analysis Under Bump Road Excitation

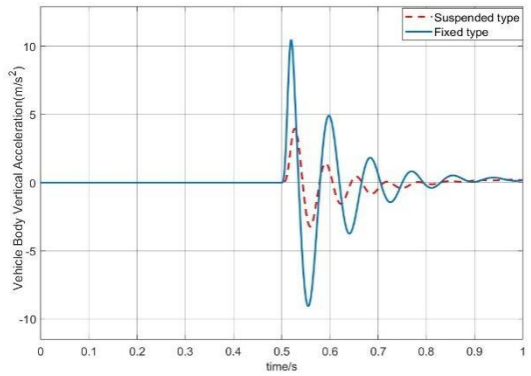
Setting the bump road excitation, the time-domain comparisons of various performance metrics are shown in Figure 4.



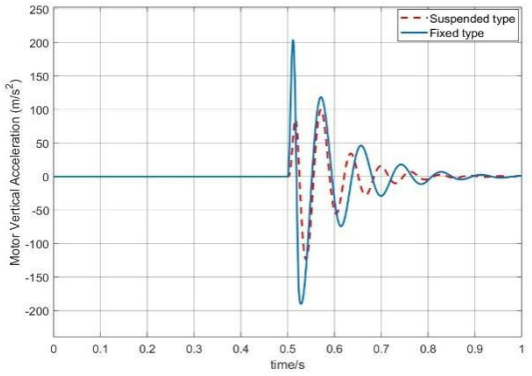
(a) Wheel dynamic load comparison



(b) Body suspension dynamic deflection comparison



(c) Body vertical acceleration comparison



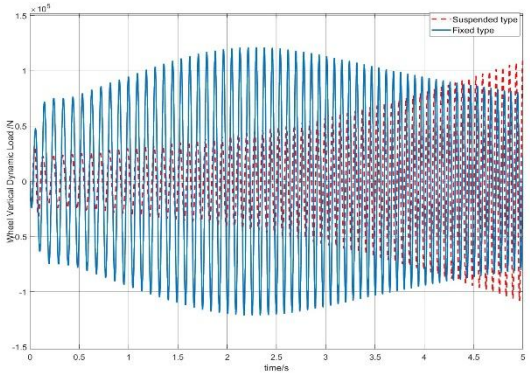
(d) Motor vertical acceleration comparison

Figure 4 Time-Domain Comparison of Performance Metrics under Bump Road Excitation

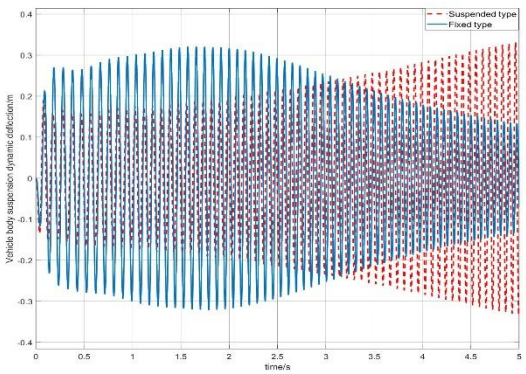
As shown in Figure 4, the wheel dynamic load and body suspension dynamic deflection of the in-wheel suspended electric vehicle are slightly larger than those of the fixed electric vehicle, but the difference is minimal when traversing a bump road. This indicates that suspending the in-wheel motor doesn't significantly enhance tire-road adhesion or reduce the likelihood of suspension collision on bumpy roads. However, the body vertical acceleration and motor vertical acceleration of the in-wheel suspended electric vehicle are smaller than those of the fixed configuration, suggesting that passengers experience lower vibration amplitudes in the suspended configuration, meaning that the comfort level has been improved.

3.3 Simulation Analysis Under Swept-Sine Excitation

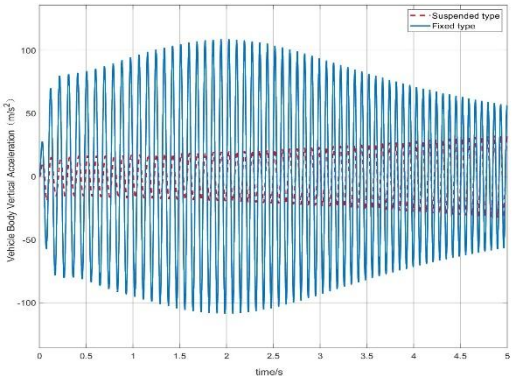
Setting the swept-sine road excitation, the time-domain comparisons of various performance metrics are shown in Figure 5.



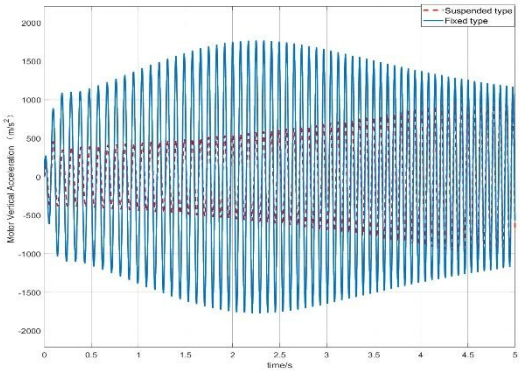
(a) Wheel dynamic load comparison



(b) Body suspension dynamic deflection comparison



(c) Body vertical acceleration comparison



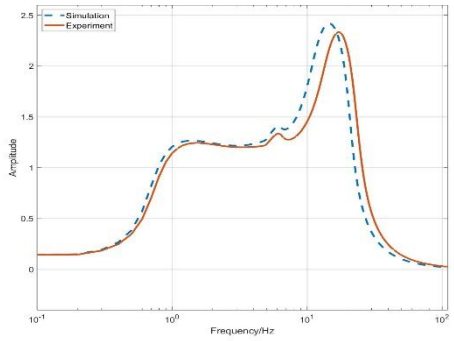
(d) Motor vertical acceleration comparison

Figure 5 Time-Domain Comparison of Performance Metrics under Swept-Sine Road Excitation

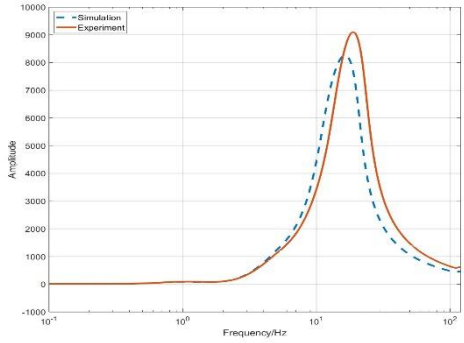
As shown in Figure 5, the wheel dynamic load and body suspension dynamic deflection are related to vehicle speed. On a fixed-frequency sine road, the in-wheel suspended configuration exhibits larger wheel dynamic load and body suspension deflection than the fixed configuration after vehicle speeds exceed 14 m/s and 13 m/s, respectively. This indicates that at lower speeds, the in-wheel suspended configuration can effectively reduce wheel dynamic load and body suspension deflection, enhancing road-holding ability and reducing the likelihood of suspension collision. However, this improvement diminishes and may even become worse than the fixed configuration at higher speeds. In contrast, the improvement in body vertical acceleration and motor vertical acceleration for the in-wheel suspended configuration doesn't deteriorate with increasing speed.

4 EXPERIMENTAL VALIDATION AND COMPARATIVE ANALYSIS

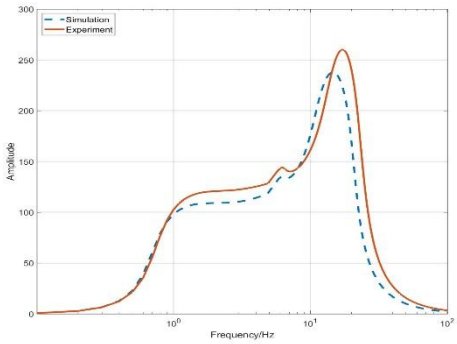
The complexity and strong randomness of actual road conditions pose significant challenges for quantitative analysis. To validate the effectiveness of the dynamic model for the in-wheel suspended electric wheel, this experiment and simulation comparison selected swept-sine road excitation as the unified condition. This excitation covers the main vibration frequency bands encountered in actual vehicle operation through continuously varying frequency and maintains consistent parameters with the simulation phase, comprehensively reflecting the system's vertical dynamic response at different frequencies. The amplitude-frequency diagrams of acceleration outputs from various parts for both test and simulation are shown in Figure 6.



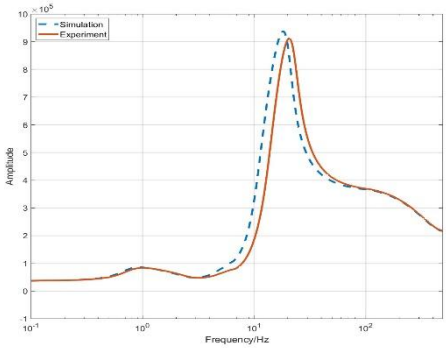
(a) The suspension dynamic deflection



(b) The motor vertical acceleration



(c) The body vertical acceleration



(d) The wheel dynamic load

Figure 6 Comparison of Frequency Response Curves between Test and Simulation for Various Metrics under Swept-Sine Road Excitation

Analysis of the amplitude-frequency characteristic curves above reveals that for the in-wheel suspended configuration, the overall fluctuation amplitudes of the experimental and simulated amplitude-frequency curves for wheel, motor, and body vertical acceleration are relatively close. At lower frequencies, the experimental and simulation results are nearly identical. As speed increases, deviations gradually appear between experiment and simulation. This is due to the difficulty in accurately determining the stiffness and damping coefficients of the body suspension and motor suspension springs/dampers in the actual experiment, leading to discrepancies. However, the results are generally in good agreement, indicating the correctness of the simulation model.

5 CONCLUSIONS

Mathematical models based on mass-spring-damper systems were established for both fixed and in-wheel suspended electric wheel systems. Wheel vertical dynamic load, motor vertical vibration acceleration, body vertical acceleration, and body suspension deflection were identified as vehicle performance evaluation metrics. Random road, bump road, and swept-sine road models were constructed.

The dynamic characteristics were analyzed based on time-domain plots of the four metrics under random, bump, and swept-sine road excitations. It was concluded that the in-wheel suspended configuration improves vehicle vertical vibration performance and ride comfort compared to the fixed configuration. However, the suspended configuration's ability to improve wheel dynamic load and body suspension deflection is influenced by vehicle speed, while its improvement in body and motor vertical acceleration is not affected by speed.

Using swept-sine excitation as input, amplitude-frequency characteristic curves for wheel, motor, and body vertical acceleration were obtained experimentally. And the amplitude- frequency characteristic curves from road excitation to wheel dynamic load, body suspension deflection, and the aforementioned accelerations were derived. Comparison between experimental and simulation results further verified the reliability of the simulation model.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This study was supported by National Training Program of Innovation and Entrepreneurship for Undergraduates (Project No.202513038) and Fundamental Research Funds for the Central Universities.

REFERENCES

[1] Ning GB, Wan G. Research status of the influence of wheel-side drive system on vehicle vertical performance. Automobile Technology, 2007, 3: 21-25.
[2] Rojas AER, Niederkofler H, Willberger J. Comfort and safety enhancement of passenger vehicles with in-wheel motors. Society of Automotive Engineers, 2010: 1-12.
[3] Chen C, Cheng Y, Meng F. Optimum design of a novel in-wheel suspension of the electric wheel. IEEE 3rd International Conference on Green Energy and Applications, 2019: 3-10.
[4] Zhao YE, Zhang JW, Han X. Design and research on dynamic vibration reduction mechanism of in-wheel motor independent drive electric vehicle. Mechanical Science and Technology, 2008, 3: 3-4.
[5] Li G. Research on ride comfort of mounting configuration for in-wheel motor driven electric vehicles. Qingdao: Shandong University of Science and Technology, 2020.
[6] Meng L, Qin Y, Zou Y, et al. A new type of hub-driven electric wheel and its vertical vibration analysis. Proceedings of 2019 SAE-China Congress & Exhibition, 2019, 2: 1-2.
[7] Wei W. Design and control strategy of electromagnetic active suspension. Shenyang: Shenyang University of Technology, 2020.
[8] Nie SD. Research on comprehensive performance of semi-active suspension considering fixed point characteristics. Changchun: Jilin University, 2018.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3068

INORGANIC COMPOSITE STONE-LIKE CAST-IN-PLACE TECHNOLOGY AND ENGINEERING APPLICATION PERFORMANCE EVALUATION FOR URBAN RENEWAL

LianDi Zhao*, XianWei Wang
Shandong Jiaotong University, Jinan 250357, Shandong, China.
**Corresponding Author: LianDi Zhao*

Abstract: To address the demand for green, adaptable building materials in urban renewal's existing building renovation and public space transformation, and solve traditional stone-like materials' defects (long construction period, poor adaptability, low solid waste utilization), this study optimizes inorganic composite stone-like cast-in-place technology and evaluates its performance. Industrial solid wastes (steel slag, recycled construction aggregates) serve as main raw materials; composite cementitious systems and functional admixtures are developed, with key process parameters determined via orthogonal experiments. Results show that at 72% solid waste replacement rate, 0.28 water-binder ratio, 0.4% retarder content and 20–25℃ pouring temperature, the material's 28-day compressive/flexural strength reaches 105/9.8 MPa, setting time 4–6 h, meeting on-site construction requirements. It exhibits excellent durability, with 8.2% strength loss after 50 freeze-thaw cycles and grade-4+ stain resistance. Engineering verification confirms 40% higher construction efficiency, 35% lower cost and 1.2 t/100 m² carbon emission reduction versus traditional dry-hanging. This scheme offers green, efficient technical support for urban renewal, with significant engineering and environmental value.
Keywords: Urban renewal; Inorganic composite stone-like material; Cast-in-place technology; Performance evaluation; Solid waste utilization; Construction adaptability

1 INTRODUCTION

1.1 Research Background

With the acceleration of urban renewal in China, the renovation of existing buildings, transformation of public spaces, and restoration of historical blocks have become important components of urban development[1]. Architectural decoration materials for urban renewal are required to meet multiple demands such as green environmental protection, high performance, fast construction, and cultural compatibility[2]. Natural stone has been widely used in urban decoration due to its excellent decorative effect, but it faces problems such as limited resources, high mining costs, and difficulty in on-site adjustment, which cannot adapt to the complex and changeable construction conditions of urban renewal[3].

Traditional artificial stone construction mainly adopts dry-hanging or precast installation processes, which have disadvantages such as long construction period, high installation cost, and poor adaptability to on-site size changes[4]. In urban renewal projects, especially in old community renovation and historical block restoration, there are often constraints such as narrow construction space, limited construction time, and strict environmental protection requirements, which make traditional processes difficult to implement[5]. At the same time, China produces more than 3 billion tons of industrial solid waste every year, and the low utilization rate of solid waste such as construction waste and steel slag has become a bottleneck restricting green urban development[6]. The development of on-site pouring imitation stone materials with high solid waste utilization rate, excellent construction adaptability, and reliable performance has become an urgent need to promote the green transformation of urban renewal projects.

Therefore, this study focuses on the core needs of urban renewal, optimizes the on-site pouring process of solid waste-based inorganic composite imitation stone, evaluates its comprehensive performance in engineering applications, and provides technical support for the efficient and green development of urban renewal projects.

1.2 Research Status at Home and Abroad

Foreign research on cast-in-place decorative materials started early. European countries have developed polymer-modified cement-based cast-in-place materials for urban landscape renovation, with good decorative effects but high cost and limited solid waste utilization[7]. Japanese scholars have studied the application of solid waste aggregates in cast-in-place concrete, improving the environmental performance of materials, but the mechanical properties and durability need to be further improved[8]. American researchers have focused on the construction adaptability of cast-in-place materials, developing fast-curing admixtures to shorten the construction period, but lack targeted research on the special needs of urban renewal such as historical style integration[9].

Domestic research on cast-in-place imitation stone has gradually increased in recent years. Li Wei et al. prepared cast-in-place imitation marble using fly ash and cement[10], which has good workability but low strength and poor durability. Wang Hong et al. optimized the mix ratio of cast-in-place artificial stone[11], improving mechanical properties but ignoring the adaptability to complex construction environments in urban renewal. Zhang Xiaoming et al. studied the application of cast-in-place imitation stone in old community renovation[12], but the research on construction process optimization and long-term performance evaluation is insufficient. At present, domestic research still has shortcomings such as inconsistent performance of cast-in-place materials, poor adaptability to urban renewal scenarios, and incomplete performance evaluation systems, which restrict the large-scale application of cast-in-place imitation stone in urban renewal[13-15].

2 EXPERIMENTAL MATERIALS AND METHODS

2.1 Experimental Materials

2.1.1 Cementitious materials

P.O42.5 ordinary Portland cement (compressive strength 48.2MPa at 28d, specific surface area 340m²/kg); S95 grade blast furnace slag powder (specific surface area 410m²/kg, activity index 96% at 28d); Grade I fly ash (specific surface area 380m²/kg, loss on ignition 4.2%), all meeting the requirements of GB/T 18046-2017 and GB/T 1596-2017.

2.1.2 Solid waste aggregates

Steel slag: Provided by a steel plant in Hebei, with particle size 0-16mm, crushing index 11.5%, water absorption rate 2.1% after magnetic separation and aging treatment; construction waste recycled aggregate: crushed from waste concrete in urban demolition projects, particle size 0-10mm, apparent density 2420kg/m³, water absorption rate 3.3%; tailings sand: provided by a mining area in Shandong, particle size 0.15-5mm, SiO_2 content 68%, meeting the requirements of GB/T 25177-2010.

2.1.3 Functional additives

Composite retarder: Self-developed, composed of sodium gluconate, citric acid, and borax, with retarding time 4-6h, solid content 95%; polycarboxylate superplasticizer: water reduction rate ≥30%, solid content 40%; composite crack inhibitor: composed of calcium oxide expansion agent and polyvinyl alcohol fiber, fiber length 6mm; titanium dioxide: rutile type, whiteness ≥98%, used for color adjustment.

2.1.4 Other materials

Quartz sand: 40-80 mesh, fineness modulus 2.7; tap water, meeting the requirements of JGJ 63-2006.

2.2 Experimental Methods

2.2.1 Mix ratio and process parameter design

First, carry out single-factor experiments to study the effects of solid waste replacement rate (60%-80%), water-binder ratio (0.25-0.32), composite retarder dosage (0.2%-0.6%), and pouring temperature (15-30℃) on the workability (slump, setting time) and mechanical properties of cast-in-place imitation stone. Then, design an $L16(4^5)$ orthogonal test with 5 factors and 4 levels, taking 28d compressive strength, flexural strength, and construction adaptability as evaluation indicators to optimize the mix ratio and process parameters.

2.2.2 Sample preparation and construction simulation

Raw material pretreatment: Dry solid waste aggregates at 105℃ for 24h to control the moisture content ≤0.5%; sieve to ensure the particle size distribution meets the design requirements.

Mixing: Mix cementitious materials, aggregates, and dry additives for 30s, then add water and liquid additives, and stir for 90s to form a uniform mixture with slump 120-150mm.

Casting and molding: Simulate on-site construction conditions, cast the mixture into molds of 100mm×100mm×100mm (mechanical properties) and 500mm×500mm×50mm (decorative effect), vibrate with a plate vibrator (frequency 50Hz) for 30s to eliminate bubbles.

Curing: Adopt standard curing (temperature 20±2℃, relative humidity ≥95%) and on-site natural curing respectively, and test performance at 7d, 28d, and 90d.

2.2.3 Performance testing methods

Mechanical properties: Refer to GB/T 50081-2019, use a universal testing machine to test compressive strength (loading rate 0.5MPa/s) and flexural strength (loading rate 0.05MPa/s).

Workability and construction adaptability: Test slump according to GB/T 50080-2016; record initial and final setting time using Vicat apparatus; evaluate fluidity, mold filling ability, and surface forming quality through on-site construction simulation.

Durability: Conduct freeze-thaw cycle test (50 cycles, GB/T 50082-2009), water absorption test (GB/T 17671-1999), and stain resistance test (JC/T 2604-2021, test with coffee, soy sauce, engine oil, etc.).

Environmental performance: Test carbon emission coefficient according to GB/T 38596-2020; detect leaching of heavy metals (Pb, Cr, Cd) according to GB 5085.3-2007.

Decorative effect: Evaluate surface smoothness (GB/T 13891-2008), glossiness (80-90GU), and color uniformity ($\Delta E \leq 2$) using a gloss meter and colorimeter.

3 RESULTS AND ANALYSIS

3.1 Optimization of Mix Ratio and Process Parameters

3.1.1 Single-factor experiment results

• Effect of solid waste replacement rate: As shown in Figure 1, with the increase of solid waste replacement rate, the strength of cast-in-place imitation stone first increases and then decreases. When the replacement rate is 72%, the 28d compressive strength reaches 98MPa, and flexural strength 9.2MPa. Excessive replacement rate (> 75%) leads to poor interface bonding and reduced strength.

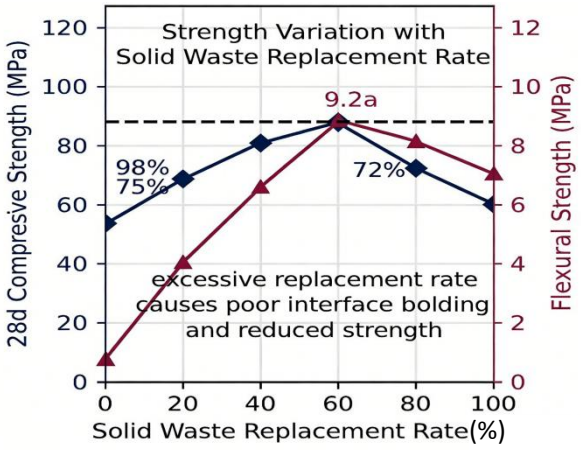


Figure 1 Strength Variation with Solid Waste Replacement Rate

• Effect of water-binder ratio: As shown in Figure 2, the optimal water-binder ratio is 0.28. When the ratio is too low (< 0.26), the workability is poor and mold filling is difficult; when too high (> 0.30), the strength decreases significantly, and surface shrinkage cracks easily occur.

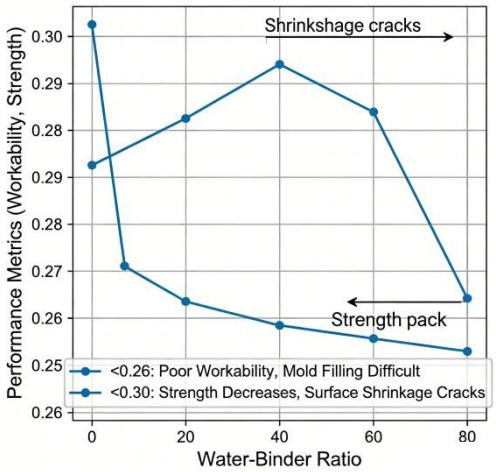


Figure 2 Optimal Water-Binder Ratio Analysis

• Effect of composite retarder dosage: As shown in Figure 3, when the dosage is 0.4%, the setting time is 4.5h, which matches the on-site construction window. Excessive dosage (> 0.5%) leads to strength loss, while insufficient dosage (< 0.3%) results in too fast setting and difficult construction.

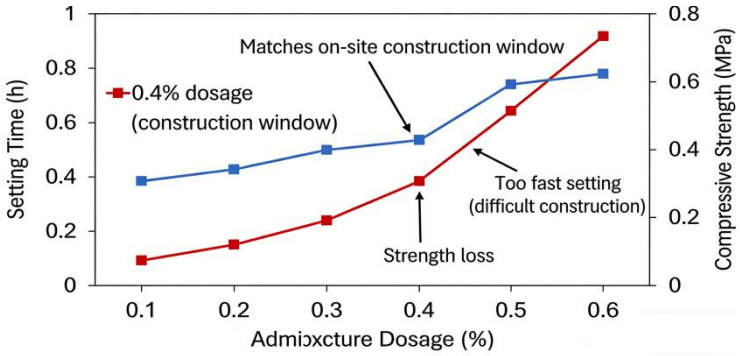


Figure 3 Effect of Admixiture Dosage on Setting Time and Strength

• Effect of pouring temperature: As shown in Figure 4, the optimal pouring temperature is 20-25℃. When the temperature is lower than 15℃, the setting time is prolonged to more than 8h; when higher than 30℃, the hydration heat is too high, leading to internal cracks.

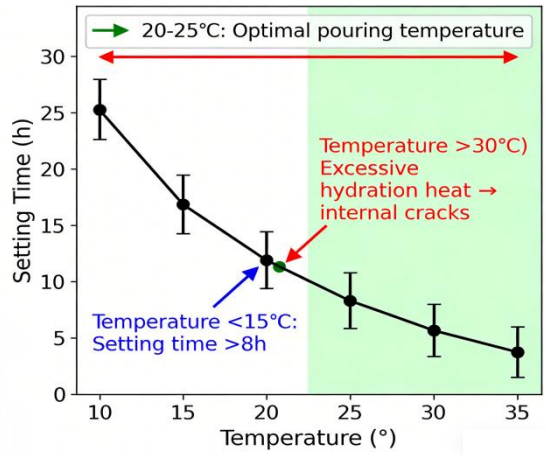


Figure 4 20-25°C: Optimal Pouring Temperature

3.1.2 Orthogonal test results

Based on single-factor experiments, the orthogonal test factors and levels are shown in Table 1, and the test results are shown in Table 2. Through range analysis, the primary and secondary order of factors affecting compressive strength is: water-binder ratio (B) > solid waste replacement rate (A) > retarder dosage (C) > pouring temperature (D) > curing method (E). The optimal combination is A3B2C3D2E1, i.e., solid waste replacement rate 72%, water-binder ratio 0.28, retarder dosage 0.4%, pouring temperature 22℃, standard curing. Under this combination, the 28d compressive strength is 105MPa, flexural strength 9.8MPa, slump 135mm, and setting time 4.2h, which fully meets the requirements of urban renewal construction.

Table 1 Factors and Levels of Orthogonal Test

Factor	Level 1	Level 2	Level 3	Level 4
Solid waste replacement rate (A, %)	65	70	72	75
Water-binder ratio (B)	0.26	0.28	0.30	0.32
Retarder dosage (C, %)	0.2	0.3	0.4	0.5
Pouring temperature (D, ℃)	15	22	25	30
Curing method (E)	Standard curing	Natural curing	Steam curing	Spray curing

Table 2 Orthogonal Test Results (Partial)

Test number	A	B	C	D	E	Compressive strength (MPa)	Flexural strength (MPa)	Slump (mm)	Setting time (h)
1	1	1	1	1	1	85	8.2	110	3.8
8	2	2	3	2	1	102	9.6	130	4.3
12	3	2	3	2	1	105	9.8	135	4.2
16	4	3	4	3	2	92	8.9	140	5.1

3.2 Comprehensive Performance Evaluation

3.2.1 Mechanical properties

The mechanical properties of the optimized cast-in-place imitation stone are shown in Table 3. Compared with traditional precast imitation stone and natural marble, the cast-in-place material has higher compressive strength and flexural strength, and the 90d strength continues to increase by 8.5%, showing good late strength development potential.

The bonding strength with the base layer is 1.8MPa, which is higher than the requirement of 1.5MPa in JC/T 2604-2021, ensuring the stability of the decorative layer in urban renewal projects.

Table 3 Comparison of Mechanical Properties

Material	28d Compressive strength (MPa)	28d Flexural strength (MPa)	90d Compressive strength (MPa)	Bonding strength (MPa)
Cast-in-place inorganic composite imitation stone	105	9.8	114	1.8
Traditional precast imitation stone	85	7.6	88	-
Natural marble	42	4.5	43	-
GB/JC Standard requirement	≥60	≥6.0	-	≥1.5

3.2.2 Workability and construction adaptability

The optimized cast-in-place imitation stone has slump 120-150mm, good fluidity and mold filling ability, and can adapt to complex shapes such as curved surfaces and special-shaped components in urban renewal. The initial setting time is 3.5-4.5h, and final setting time 5.5-6.5h, which allows sufficient time for on-site adjustment and surface finishing. The construction thickness can be adjusted between 30-80mm, and the material can be directly poured on the existing base layer (concrete, masonry, etc.) without complex pretreatment, significantly improving construction efficiency.

3.2.3 Durability

As shown in Table 4, after 50 freeze-thaw cycles, the mass loss rate of the cast-in-place imitation stone is 0.9%, and strength loss rate 8.2%, which is far lower than the standard limit (mass loss rate ≤3%, strength loss rate ≤15%). The water absorption rate is 2.3%, and the stain resistance grade is Level 4 (stain removal rate ≥90%), which can resist common pollutants in urban environments. The carbonation depth is 3.2mm after 28d, showing good carbonization resistance, adapting to the complex service environment of urban public spaces.

Table 4 Durability Test Results

Test item	Test condition	Test result	Standard limit
Freeze-thaw resistance	50 cycles (-20℃~20℃)	Mass loss 0.9%, strength loss 8.2%	Mass loss ≤3%, strength loss ≤15%
Water absorption	24h immersion	2.3%	≤5%
Stain resistance	Coffee, soy sauce, engine oil	Grade 4 (removal rate ≥90%)	≥Grade 3
Carbonation resistance	28d carbonation	Depth 3.2mm	≤5mm

3.2.4 Environmental performance

The carbon emission coefficient of the cast-in-place imitation stone is 286kg CO_2/m^3, which is 42% lower than that of natural marble (493kg CO_2/m^3) and 25% lower than that of traditional precast imitation stone (381kg CO_2/m^3). The leaching concentrations of Pb, Cr, and Cd are all lower than the limit values in GB 5085.3-2007, and the solid waste utilization rate reaches 72%, realizing the resource utilization of industrial solid waste and meeting the green development requirements of urban renewal.

3.3 Microscopic Mechanism Analysis

3.3.1 Microstructure Analysis (SEM)

As shown in Figure 6, the internal structure of the cast-in-place imitation stone is dense, with few pores and cracks. The solid waste aggregates are closely bonded with the cementitious matrix, and the interface transition zone is narrow (< 20μm). The composite admixtures promote the hydration reaction to generate a large amount of C-S-H gel, which fills the internal pores and improves the compactness. The polyvinyl alcohol fibers are uniformly distributed in the matrix, playing a role in bridging cracks and improving toughness.

3.3.2 Phase Composition Analysis (XRD)

As shown in Figure 7, the main crystalline phases of the cast-in-place imitation stone are C-S-H gel, $Ca(OH)_2$, AFt, and quartz. The addition of composite retarder inhibits the rapid growth of $Ca(OH)_2$ crystals, promoting the formation of dense C-S-H gel. The active components in solid waste aggregates (such as SiO_2, Al_2O_3) react with $Ca(OH)_2$ to generate additional C-S-H gel through pozzolanic reaction, further enhancing the interface bonding strength and material compactness.

4 ENGINEERING APPLICATION VERIFICATION

4.1 Overview of Demonstration Projects

The optimized cast-in-place inorganic composite imitation stone was applied to two typical urban renewal projects:

1. Renovation project of an old community in Tianjin: The project involves the renovation of 8 residential buildings with a construction area of 32,000m². The cast-in-place imitation stone is used for the decoration of the lobby floor, corridor walls, and community square pavement, with a total application area of 4,800m².
2. Restoration project of a historical block in Jinan: The project aims to restore the traditional style of the block, with the cast-in-place imitation stone used for the restoration of street pavement, shop facades, and landscape components, with an application area of 3,200m².

4.2 Construction Process and Quality Control

The on-site construction process of cast-in-place imitation stone is as follows:
1. Base layer treatment: Clean the base layer, repair cracks and uneven parts, and apply an interface agent to improve bonding performance.
2. Formwork installation: Install formwork according to the design shape and thickness, and ensure the flatness and firmness of the formwork.
3. Material mixing and pouring: Mix the materials on site according to the optimized mix ratio, and pour them into the formwork, followed by vibration and leveling.
4. Surface finishing: After initial setting, carry out surface polishing and texture treatment to simulate the texture of natural stone.
5. Curing: Adopt spray curing for 7d to ensure the hydration reaction is sufficient.
During construction, the key quality control points include: controlling the mixing ratio of materials, ensuring the pouring temperature is within 20-25℃, and strengthening the vibration to eliminate internal bubbles. The construction quality is tested by on-site sampling, and the qualified rate of compressive strength and surface quality reaches 100%.

4.3 Application Effect Evaluation

4.3.1 Construction efficiency and cost

Compared with the traditional dry-hanging natural stone process, the cast-in-place process reduces the construction procedures such as cutting, transportation, and installation of precast components. The construction efficiency is improved by 40%, and the construction period is shortened by 30%. The comprehensive cost (material + construction) is 185 yuan/m², which is 35% lower than that of natural marble (285 yuan/m²) and 20% lower than that of traditional precast imitation stone (231 yuan/m²), achieving significant economic benefits.

4.3.2 Use effect and user evaluation

After 18 months of operation, the on-site inspection results show that the cast-in-place imitation stone decorative layer is intact without cracks, peeling, or discoloration. The compressive strength of the pavement and wall materials is still above 100MPa, and the stain resistance and durability are good. A questionnaire survey was conducted among 200 residents and merchants in the project area, and 92% of the respondents were satisfied with the decorative effect, 88% were satisfied with the construction efficiency, and 95% recognized the environmental protection performance of the material.

4.3.3 Scenario adaptability analysis

The cast-in-place imitation stone shows good adaptability in different urban renewal scenarios:
• Old community renovation: It can be directly poured on the existing base layer, adapting to the narrow construction space and short construction period requirements.
• Historical block restoration: The color and texture can be adjusted according to the historical style, realizing the integration of modern materials and traditional culture.
• Public space transformation: It has good wear resistance and durability, adapting to the high traffic volume of public spaces such as squares and corridors.

5 CONCLUSION

1. A cast-in-place inorganic composite imitation stone material suitable for urban renewal was developed, with solid waste replacement rate up to 72%. The optimal mix ratio and process parameters are: water-binder ratio 0.28, composite retarder dosage 0.4%, pouring temperature 20-25℃, and standard curing.
2. The prepared cast-in-place imitation stone has excellent comprehensive performance: 28d compressive strength 105MPa, flexural strength 9.8MPa, good workability and construction adaptability, and meets the requirements of complex construction conditions in urban renewal. It has good durability (freeze-thaw resistance, stain resistance) and environmental performance (low carbon emission, high solid waste utilization rate).
3. The strength formation mechanism is that the composite admixtures promote the hydration reaction to generate dense C-S-H gel, fill internal pores, and improve the interface bonding strength between solid waste aggregates and cementitious matrix. The polyvinyl alcohol fibers play a role in toughening and crack resistance, and the pozzolanic reaction of solid waste active components further enhances the material performance.
4. Engineering application verification shows that the cast-in-place process has the advantages of high construction efficiency, low cost, and good scenario adaptability. It can be widely used in old community renovation, historical block

restoration, and public space transformation, providing a green and efficient technical solution for urban renewal projects.

In the future, further research can be carried out on the functional modification of cast-in-place imitation stone (such as waterproof, anti-slip, and thermal insulation) and the expansion of applicable solid waste types, to better meet the diverse needs of urban renewal and promote the sustainable development of the construction industry.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Ministry of Housing and Urban-Rural Development. Guiding Opinions on Accelerating Urban Renewal. Beijing: Ministry of Housing and Urban-Rural Development, 2021.
[2] Li Y, Wang J. Research on the application of green building materials in urban renewal projects. Journal of Building Materials, 2022, 25(3): 456-462.
[3] China Stone Industry Association. Development Status and Trends of China's Stone Industry in Urban Renewal. Beijing: China Stone Industry Association, 2022.
[4] Zhang L, Liu H. Problems and countermeasures of traditional artificial stone construction in urban renewal. Construction Technology, 2021, 52(8): 98-101.
[5] Wang C, Zhang Y. Discussion on the characteristics and material requirements of urban renewal projects. Urban Development Studies, 2020, 27(6): 34-39.
[6] Ministry of Ecology and Environment. Statistical Communique on National Ecological Environment Status 2022. Beijing: Ministry of Ecology and Environment, 2023.
[7] Rossi F, Martelli A. Development and application of polymer-modified cementitious cast-in-place materials for urban renewal. Construction and Building Materials, 2019, 210: 678-685.
[8] Tanaka Y, Suzuki K. Utilization of steel slag aggregate in cast-in-place decorative concrete for urban landscape. Journal of Materials in Civil Engineering, 2020, 32(7): 04020185.
[9] Smith J, Johnson M. Fast-curing cast-in-place artificial stone for urban renovation projects. Cement and Concrete Composites, 2021, 119: 103987.
[10] Li W, Zhang S. Preparation of fly ash-based cast-in-place imitation marble and its performance. New Building Materials, 2019, 46(5): 78-81.
[11] Wang H, Li J. Mix ratio optimization of cast-in-place artificial stone for building decoration. Journal of Wuhan University of Technology (Materials Science Edition), 2020, 35(2): 345-350.
[12] Zhang X M, Liu Y. Application of cast-in-place imitation stone in old community renovation engineering. Construction Science and Technology, 2021(12): 89-91.
[13] Liu Z, Chen W. Research progress of cast-in-place decorative materials for urban renewal. Materials Review, 2022, 36(S1): 289-293.
[14] GB/T 50081-2019, Standard for Test Methods of Mechanical Properties on Ordinary Concrete. Beijing: China Architecture and Building Press, 2019.
[15] JC/T 2604-2021, Inorganic Artificial Stone. Beijing: China Building Materials Industry Press, 2021.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3069

CURRENT STATUS OF NEAR-SURFACE OZONE POLLUTION RESEARCH IN CHINA

Xin Lin
Qinghai Institute of Technology, Xining 810016, Qinghai, China.

Abstract: In recent years, with the effective control of fine particulate matter (PM2.5) pollution in China, the problem of near-surface ozone (O3) pollution has become increasingly prominent, becoming one of the important factors restricting the continuous improvement and attainment of ambient air quality standards. This article systematically reviews the latest progress in research on near-surface ozone pollution in China. Studies show that since the 13th Five-Year Plan, the overall O3 concentration in China has shown a trend of first rising and then plateauing, exhibiting significant regional differences and seasonal high incidence in key urban clusters such as the Beijing-Tianjin-Hebei region and the Yangtze River Delta. In terms of formation mechanism, the nonlinear relationship between ozone formation and its precursors, nitrogen oxides (NOx) and volatile organic compounds (VOCs), has always been the core of research: most urban central areas are in VOCs control zones, while regional background areas and some suburban areas are mostly NOx control zones or transition zones, and their control sensitivity dynamically evolves with the progress of pollution reduction . Meteorological conditions, especially extreme high temperatures and strong solar radiation, have a significant promoting effect on ozone formation and accumulation, and the "heat wave-ozone" compound pollution is gradually becoming a new prominent problem. In terms of prevention and control strategies, the synergistic control of ozone and PM2.5 has risen to an important strategic direction. It is urgent to implement differentiated NOx and VOCs emission reduction measures based on precise sensitivity analysis, and to strengthen regional joint prevention and control mechanisms. Future research should focus on the synergistic governance of compound pollution, refined source spectrum analysis and efficient control of VOCs, the microscopic processes of ozone formation, and the systematic upgrading of forecasting and early warning technologies.
Keywords: Ozone pollution; Research review; PM2.5 synergistic control; VOCs; NOx; Control sensitivity; Regional transport

1 INTRODUCTION

In recent years, with the continuous deepening of China's air pollution prevention and control work, the concentration of traditional pollutants represented by fine particulate matter (PM2.5) has decreased significantly, and the overall ambient air quality has shown an improving trend. However, while the PM2.5 level has continued to decrease, the problem of near-ground ozone (O3) pollution has become increasingly prominent and has gradually evolved into one of the primary pollutants affecting China's air quality standards[1]. This change indicates that China's air pollution prevention and control is moving from a single pollution control stage dominated by PM2.5 to a new stage of coordinated prevention and control of PM2.5 and O3, which puts forward higher requirements for scientific research and policy formulation.

Ground-level ozone is a typical secondary pollutant. It is not emitted directly, but is generated under strong solar radiation conditions by a series of complex photochemical reactions of precursors such as nitrogen oxides (NOx) and volatile organic compounds (VOCs) [2]. Ozone has extremely strong oxidizing properties and is significantly harmful to human health, ecosystems and crop production. Long-term exposure to high concentrations of ozone can induce or aggravate respiratory diseases, reduce lung function, and even cause irreversible damage [3]. At the same time, ozone can also inhibit plant photosynthesis, leading to reduced crop yields and damage to the structure and function of ecosystems [4].

Given the complex causes and widespread impact of ozone pollution, a systematic review of research progress on near-surface ozone pollution in my country is of significant theoretical and practical value for deepening scientific understanding and supporting precise prevention and control. This paper systematically reviews research findings on the spatiotemporal distribution characteristics, formation mechanisms, and influencing factors of ozone pollution in my country, as well as collaborative control and prevention strategies. It also provides an outlook on future research priorities, aiming to offer scientific references for ozone pollution prevention and control in my country.

2 STATUS AND SPATIOTEMPORAL CHARACTERISTICS OF OZONE POLLUTION IN CHINA

2.1 Overall Trend

Since the implementation of the Action Plan for Air Pollution Prevention and Control in 2013, my country’s ambient air quality monitoring network has been continuously improved, laying a solid data foundation for ozone pollution research.

Numerous studies have shown that during the 13th Five-Year Plan period (2016-2020), my country's ozone pollution showed an overall change characteristic of "first rising and then stabilizing" [5].

During the period from 2015 to 2018, ozone concentrations in most cities across the country generally increased. This is partly due to the significant decrease in PM2.5, which weakened the extinction effect of aerosols and enhanced near-surface solar radiation, thereby promoting the photochemical generation of ozone[6]. On the other hand, in VOCs control areas, the initial stage of NOx emission reduction may trigger a "seesaw effect", that is, the decrease in NOx concentration leads to an increase in ozone levels[7].

During the period from 2018 to 2020, as ozone pollution prevention and control measures were gradually strengthened, especially in the spring and summer of 2020, ozone concentrations across the country showed a general downward trend for the first time [8], indicating that the emission reduction measures for precursors began to show results. However, overall, ozone concentrations remained relatively high, and some key areas continued to exceed the standards.

Since 2021, ozone pollution in my country has remained severe against the backdrop of frequent extreme weather conditions. For example, in the summer of 2022, southern China experienced an unprecedented extreme high-temperature process, which led to a significant rebound in ozone concentrations across the country[9]. High temperatures and strong radiation significantly enhance the intensity of photochemical reactions, causing the risk of "heat wave-ozone" compound pollution to continue to rise[10]. The latest ecological and environmental status bulletin shows that ozone remains one of the main pollutants causing air quality to exceed standards in many places[11].

2.2 Regional Distribution Characteristics

Near-ground ozone pollution in China has significant spatial heterogeneity, with high-value areas mainly concentrated in economically developed, densely populated, and high-emission-intensity urban clusters [12].

The Beijing-Tianjin-Hebei region and surrounding areas have long been in the high ozone pollution zone in the country, with ozone concentrations in most cities in the region exceeding the national secondary standard limit for many consecutive years [13]. Ozone pollution in this region is not only significantly affected by local emissions, but also by the superimposed effect of regional transport.

The Yangtze River Delta region also faces a prominent ozone pollution problem. Although ozone levels in some cities have improved in recent years (e.g., ozone concentration in Shanghai has dropped to a low level in recent years[14]), ozone is still an important factor restricting the achievement of regional air quality standards. Regional transport plays an important role in ozone pollution in the Yangtze River Delta, and the transport of precursors from upstream areas has a significant impact on ozone levels in downstream cities[15].

The Pearl River Delta region is one of the earliest regions in my country to carry out research and control of ozone pollution, and its ozone pollution has obvious seasonal characteristics. Related studies have shown that most areas in this region are in NOx control areas, which is significantly different from the characteristics of the core urban areas of Beijing-Tianjin-Hebei and Yangtze River Delta, which are mainly controlled by VOCs[16].

In addition, ozone pollution problems in areas such as the Fenwei Plain, the Chengdu-Chongqing urban agglomeration and the middle reaches of the Yangtze River have become increasingly prominent, and their formation mechanism and control pathways have gradually become new research hotspots[17].

2.3 Seasonal and Diurnal Variation Characteristics

The near-surface ozone concentration in my country exhibits a significant seasonal variation, typically reaching its highest value in late spring and early summer (April to September), while it is at its lowest point during winter [18]. This pattern is mainly controlled by seasonal variations in solar radiation intensity, air temperature, and atmospheric chemical reaction conditions.

On a diurnal scale, ozone exhibits typical photochemical characteristics, with its concentration typically peaking in the afternoon (13:00–16:00) and gradually decreasing at night [19]. At night, ozone is mainly consumed through titration with NO and dry deposition. In urban centers, where NO emissions are higher, the ozone concentration decreases more significantly at night, while in suburban areas and background stations, it tends to maintain a higher background level.

3 THE FORMATION MECHANISM AND INFLUENCING FACTORS OF OZONE POLLUTION

3.1 Precursor Control Sensitivity Analysis

The formation of near-ground ozone (O3) is a highly nonlinear photochemical process, and its formation efficiency mainly depends on the relative abundance of precursors nitrogen oxides (NOx) and volatile organic compounds (VOCs) and the reaction environment. Clarifying the response characteristics of ozone formation to different precursors, i.e., the ozone formation control sensitivity, is the key foundation for formulating scientific and effective emission reduction strategies [20].

Theoretically, the relationship between ozone formation and precursor concentration can usually be characterized by empirical kinetic model curves (EKMA curves), and the study area can be divided into VOCs control area, NOx control area and transition area [21]. In the VOCs control area, ozone formation is limited by the VOCs level, and prioritizing the reduction of VOCs emissions can effectively reduce ozone concentration; in the NOx control area, ozone formation

is more sensitive to NOx, and NOx emission reduction has a significant inhibitory effect; while in the transition area, ozone formation is affected by both NOx and VOCs. In recent years, indicators based on satellite remote sensing observations (such as the HCHO/NO2 ratio) have been widely used in the rapid identification and regional comparison studies of ozone formation sensitivity[22].

Existing research shows that there are significant spatial differences in the sensitivity of ozone generation control among different urban agglomerations and regions in China, and these differences exhibit dynamic evolution characteristics as pollution prevention and control progresses [23] . The core urban areas of the Beijing-Tianjin-Hebei urban agglomeration are mostly VOCs-controlled, while the suburbs and regional scales gradually shift to transition zones or NOx control zones; the Yangtze River Delta urban agglomeration is mainly VOCs-controlled, but there are significant local differences; the Pearl River Delta region is mostly in NOx control zones, which may be closely related to the higher background concentration of ozone and the earlier implementation of NOx emission reduction measures [24–27].

Based on this, the study further focused on the reactivity analysis of key VOC species. Different types of VOCs (such as olefins, aromatic hydrocarbons and aldehydes and ketones) have significantly different contributions to ozone formation [28]. Through methods such as maximum incremental reactivity (MIR), the study identified some highly reactive VOC components in industrial solvent use and traffic emissions as important factors driving ozone formation , providing a scientific basis for implementing differentiated and precise VOC emission reduction.

3.2 Impact of Meteorological Conditions

Meteorological conditions play a crucial regulatory role in the generation, accumulation and transport of ozone, and their influence is mainly reflected in two aspects: photochemical reaction rate and pollutant diffusion conditions[29].

Solar radiation is the direct energy source for photochemical reactions, and the intensity of ultraviolet radiation is usually significantly positively correlated with the concentration of ozone near the ground. Rising temperatures can not only accelerate the photochemical reaction kinetics, but also significantly enhance the emission of biogenic VOCs (such as isoprene), thereby promoting ozone formation [30]. The mechanism of relative humidity affecting ozone is more complex: low humidity conditions are usually conducive to ozone accumulation, while high humidity conditions may inhibit ozone formation by affecting free radical chemical processes. At the same time, stable weather conditions (low wind speed, high air pressure) are not conducive to the diffusion of pollutants, which can easily cause ozone and its precursors to accumulate in the near-surface layer and induce regional ozone pollution events [31].

Against the backdrop of climate change, the frequency and intensity of extreme high-temperature events continue to increase, making the "heat wave-ozone" compound pollution problem increasingly prominent [32]. Studies have found that under high-temperature heat wave conditions, enhanced vertical mixing processes and increased emissions of biogenic VOCs will change the relative ratio of NOx to VOCs, thereby affecting ozone formation sensitivity [33]. Related results indicate that conventional emission reduction measures alone are insufficient to effectively suppress the rising trend of ozone under high-temperature conditions, suggesting that future ozone control strategies urgently need to incorporate climate change adaptability and forward-looking considerations .

3.3 Regional Transmission and Source Analysis

Ozone pollution has significant regional and transport characteristics. The ozone level in a single city or local area is not only controlled by local emissions, but also by the transport of ozone and its precursors from upwind areas [34].

Previous studies have used backward trajectory analysis, source-sink apportionment, and adjoint models to quantitatively assess the role of regional transport in ozone pollution formation. The results show that ozone pollution in the Yangtze River Delta region is affected to some extent by transport from northern Jiangsu and Anhui and surrounding areas, while the Beijing-Tianjin-Hebei region exhibits obvious zonal transport characteristics, with emissions from upwind cities making a significant contribution to the ozone peak in downstream cities [35,36].

In terms of source apportionment technology, advanced numerical simulation tools such as adjoint models can efficiently quantify the sensitivity of ozone to different precursor species, emitting industries and geographical sources, providing strong support for regional joint prevention and control and emission reduction priority ranking[37]. In addition, observation-based models (OBM) and receptor models are also widely used in qualitative and quantitative research on ozone sources.

4 RESEARCH PROGRESS ON OZONE POLLUTION PREVENTION AND CONTROL STRATEGIES AND SYNERGISTIC CONTROL

4.1 The Necessity and Challenges of Synergistic Control of Ozone and PM2.5

Ozone and PM2.5 are two core issues facing my country’s current air pollution prevention and control. The two have a complex coupling relationship in terms of generation mechanism and control path, mainly reflected in the shared precursors (NOx and VOCs) and the regulatory role of atmospheric oxidation capacity on the generation of secondary pollutants [38].

From the perspective of coordinated control, NOx and VOCs are not only key precursors to ozone formation, but also important sources of nitrates and secondary organic aerosols in PM2.5. Implementing coordinated emission reduction strategies can help achieve simultaneous reduction of multiple pollutants and maximize environmental and health benefits [39]. In addition, as an important indicator of atmospheric oxidation capacity, changes in ozone concentration

will affect the efficiency of SO2, NOx and other pollutants in their conversion to particulate matter, thereby further affecting PM2.5 levels.

However, due to the strong nonlinearity of ozone formation, coordinated control also faces practical challenges such as the "seesaw effect" [40]. In VOCs control areas, simply reducing NOx emissions may weaken the titration effect of NO on ozone, leading to a short-term increase in ozone concentration. Therefore, coordinated control strategies must be based on refined sensitivity analysis, and the emission reduction ratios of NOx and VOCs should be determined according to local conditions to avoid neglecting one aspect for the other.

4.2 VOCs Emission Control Research

Given that ozone generation in the central areas of most cities in China is mainly controlled by VOCs, the refined management of VOCs has become one of the core links in ozone pollution prevention and control[41].

Existing research has continuously improved the VOCs emission inventory system, focusing on enhancing the characterization accuracy of various types of emissions, including industrial sources, mobile sources, solvent use, and biological sources [42]. In recent years, atypical emission sources and fugitive emissions have gradually attracted attention, providing new research directions for improving the reliability of emission inventories.

In terms of control technology and management, the focus of industrial source control is on the optimized application of end technologies such as adsorption, catalytic combustion and regenerative thermal oxidation (RTO), as well as the source substitution of low VOCs raw and auxiliary materials[43]. In terms of mobile sources, emission intensity is continuously reduced by raising emission standards, promoting new energy vehicles and strengthening the supervision of oil and gas recovery. For bio-source VOCs, research mainly focuses on the emission change characteristics under the background of climate change and their parameterized expression in models.

4.3 Regional joint prevention and control mechanism

Given the significant regional transport characteristics of ozone pollution, its prevention and control work urgently needs to break through administrative boundary restrictions and implement regional joint prevention and control[44].

At present, key regions such as Beijing-Tianjin-Hebei and the Yangtze River Delta have established relatively mature collaborative prevention and control mechanisms. Model simulation and scenario analysis results show that cross-regional collaborative reduction of NOx and VOCs emissions is an effective way to reduce regional ozone peaks and pollution frequency[45].

Future research will focus more on scientific zoning based on pollution characteristics and transport paths, identify key source areas that contribute the most to ozone in downstream cities through high-resolution simulation, and formulate differentiated emission reduction plans accordingly to achieve precise and coordinated governance of multiple cities and industries in the region [46].

5 CONCLUSION AND OUTLOOK

5.1 Research Summary

Over the past decade, my country has made significant progress in research on near-surface ozone pollution. Overall, the research findings are mainly reflected in the following aspects: First, the spatiotemporal evolution characteristics of ozone pollution nationwide and in key regions have been systematically clarified, identifying ozone as a key pollutant affecting air quality standards. Second, through extensive sensitivity analysis studies, the understanding of ozone formation mechanisms has been deepened, revealing significant differences in control sensitivity between urban centers and regional background areas. Third, a theoretical framework for the synergistic control of ozone and PM2.5 has been gradually established, providing scientific support for addressing the complex problems of multiple pollutants. Fourth, the application of new technologies such as accompanying models and satellite remote sensing has significantly improved the scientific rigor of ozone source analysis and control decisions.

5.2 Future Research Directions

Despite significant research progress, ozone pollution control in China still faces long-term and complex challenges. Future research urgently needs to deepen in the following areas: First, strengthen research on the coordinated governance of ozone, PM2.5, and climate change, enhancing the systematicness and resilience of emission reduction strategies; second, further refine VOCs source profile analysis and develop efficient and economical emission reduction and substitution technologies; third, explore the role of atmospheric oxidation capacity and atypical ozone formation pathways under China's environmental conditions at the microscopic level; fourth, promote the development of high-resolution forecasting, early warning, and intelligent source tracing technologies to provide technical support for precise prevention and control and emergency response; and fifth, strengthen the quantitative assessment of ozone's health and ecological effects to provide a scientific basis for evaluating pollution control effectiveness and optimizing standards.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] China National Environmental Monitoring Centre. 2023 China Ecological Environment Status Bulletin, 2024. Available: https://www.mee.gov.cn/hjzl/sthjzk/zghjzkgb/202406/P020240604551536165161.pdf.

[2] CMA. Ozone, Ozone Pollution and Pollution Control, 2010. Available: https://www.cma.gov.cn/kppd/kppdqxyr/kppdjsqx/201211/t20121128_192962.html.

[3] Xinhua Net. From little-known to becoming the primary pollutant affecting summer air quality, beware of ozone, 2020. Available: http://www.xinhuanet.com/politics/2020-06/09/c_1126094036.htm

[4] Zhou, Derong. A Study on the Sensitivity to Prevention and Control and Emission Reduction Scenarios of Typical Ozone Pollution Processes in Eastern China, 2023. Available: http://dqkxxb.ijournals.cn/dqkxxb/article/html/20230503?st=search.

[5] Ye Shen. Spatiotemporal variation characteristics and influencing factors of ozone concentration in China's three major urban agglomerations from 2015 to 2020, 2023. Available: http://www.sssampling.cn/down/2023%20%E5%8F%B6%E6%B7%B1%20-%202015%E2%80%942020%E5%B9%B4%E4%B8%AD%E5%9B%BD%E4%B8%89.pdf.

[6] A review of the causes and influencing factors of atmospheric O3 pollution in China. Environmental Science and Engineering Research, 2022. Available: https://www.hjkxyj.org.cn/cn/article/pdf/preview/10.13198/j.issn.1001-6929.2022.09.01.pdf.

[7] MEE. Analysis of Ozone Pollution Situation and Suggestions for Prevention and Control in China, 2020. Available: https://www.caep.org.cn/yclm/zghjghyzjzs/zghjghyzjzs_21956/202008/t20200813_793622.shtml

[8] Song Xichen. Peking University Air Quality Report VIII: Regional Pollution in “3+95” Cities from 2013 to 2020, 2020. Available: https://songxichen.com/index.php/Article/index/id/26.

[9] Analysis of Near-Surface Ozone Pollution and Meteorological Causes in China under Historical Extreme High Temperatures in Summer 2022. Atmospheric Science, 2023. Available: https://www.iapjournals.ac.cn/dqkx/cn/article/pdf/preview/10.3878/j.issn.1006-9895.2302.22211.pdf.

[10] Nanjing University. Professor Huang Xin and Associate Professor Li Mengmeng have made significant progress in the study of heat wave-ozone complex pollution in China under climate change, 2024. Available: https://as.nju.edu.cn/91/18/c11323a692504/page.htm.

[11] China National Environmental Monitoring Centre. 2024 Bulletin on the State of China's Ecological Environment, 2025. Available: https://www.cnemc.cn/jcbg/zghjzkgb/202506/P020250611406628093675.pdf.

[12] Liu C, Lian X，Huang J. Resesarch Review on the Statio- lefmporal Distribution of Osoe Pullation and Iis CGases in China 1.Joumal of Arid Mefevomdogy, 2020, 38(3): 355-361.

[13] CCTV. Report: National PM2.5 Concentration Slightly Decreased Last Year, but Improvement Was Not Satisfactory in Key Areas, 2023. Available: https://news.cctv.com/2023/10/27/ARTIdKi4HhO9n3Vs5VhNE4Mb231027.shtml?spm=C94212.P4YnMod9m2uD.ENPMkWvfnaiV.177.

[14] Shanghai Municipal People's Government. Shanghai Ecological Environment Status Bulletin Released: AQI Excellent Rate Reaches 88.5%, Ozone Improvement Most Significant, 2025. Available: https://www.shanghai.gov.cn/nw4411/20250604/4e3fb65e206b4ad1929be5d969cd6437.html.

[15] Gong K J, Li L, Shi Z H, et al. Quantitative analysis of the impact of pollutant transport on ozone levels in Nanjing within the Yangtze River Delta Region. Journal of Atmospheric Sciences, 2023, 46(5): 703-712.

[16] CAS. STOTEN: Evaluation of the Optimal O3 Control Strategy in the Pearl River Delta Based on Spatial Sensitivity, 2023. Available: https://iap.cas.cn/gb/xwdt/kyjz/202309/t20230928_6888759.html.

[17] Wu K, Kang P, Yu L, et al. Pollution status and spatio-temporal variations of ozone in China during 2015-2016. Acta Scientiae Circumstantiae, 2018, 38(6): 2179-2190.

[18] A study on the spatiotemporal variation of ozone concentration in Chinese cities from 2015 to 2016. Journal of Environmental Sciences, 2018. Available: https://html.rhhz.net/hjkxxb/html/20171119003.htm.

[19] Urban Environment Institute. Advances by the Urban Environment Institute in atmospheric oxidation capacity and ozone pollution mechanisms, 2022. Available: https://shb.cas.cn/kydt2024/kjjz2024/202202/t20220228_6378489.html.

[20] Scientific and precise prevention and control of ozone pollution. China Environment News, 2020. Available: https://www.cfej.net/hbyq/rdpx/202008/t20200804_792628.shtml.

[21] Lou Si jia, Zhu Bin, Liao Hong. Impacts of O precursor on surface O, concentration over China. Transactions of Atmospheric Sciences, 2010, 33(4): 451-459.

[22] ZHUO J L, ZHU S X, LONG Z, et al. A satellite-based method and application for identifying high ozone production area. Journal of Environmental Engineering Technology, 2022, 12(6): 2039-2048.

[23] CAS. SB: Spatiotemporal distribution of ozone formation sensitivity and the impact of “dual-carbon” targets on it, 2024. Available: https://iap.cas.cn/gb/xwdt/kyjz/202401/t20240103_6950998.html.

[24] SU Xingtao, FENG Jing, AN Hao, et al. Trends Analysis of Fine Particulate Matter and Ozone Pollution in Typical Cities in the Beijing–Tianjin–Hebei Region during 2015–2021. Chinese Journal of Atmospheric Sciences (in Chinese), 2023, 47(5): 1641-1653.

[25] Cui Y R, Suo N, Wang L, et al. Study of ozone generation and photochemical regimes in the urban atmosphere of Qinhuangdao. Acta Scientiae Circumstantiae, 2020, 40(11): 4105-4112.

[26] Jiang M Q, Li J, Wang X, Zhang L. O3 pollution causes and key VOCs activity analysis in typical urban agglomerations in China. Environmental Science, 2018, 39(10): 4145-4154.
[27] Sheng Y W, Zhu Y, Tao J, et al. Source contribution analysis of ambient ozone and cost-benefit assessment of control scenarios in a typical ozone-polluted city. Acta Scientiae Circumstantiae, 2017, 37(9): 3306-3315.
[28] Sources of volatile organic compounds and their role in the formation of secondary air pollutants. Peking University, 2018. Available: http://www.research.pku.edu.cn/UploadFiles/2018111171630605.pdf.
[29] Cao T W, Wu K, Kang P, et al. Study on ozone pollution characteristics and meteorological cause of Chengdu-Chongqing urban agglomeration. Acta Scientiae Circumstantiae, 2018, 38(4): 1275-1284.
[30] Nanjing University. Professor Huang Xin's research group published an article in Nature Cities revealing the formation mechanism of urban high-temperature-ozone compound pollution, 2025. Available: https://as.nju.edu.cn/f7/32/c11323a784178/page.htm.
[31] Chang L Y, Xu J M, Qu Y H, et al. Study on objective synoptic classification on ozone pollution in Shanghai. Acta Scientiae Circumstantiae, 2019, 39(1): 169-179.
[32] Mengmeng Li, Xin Huang, Dan Yan, et al. Coping with the concurrent heatwaves and ozone extremes in China under a warming climate, Science Bulletin, 2024.
[33] Zhou X, Li M, Huang X. et al. Urban meteorology–chemistry coupling in compound heat–ozone extremes. Nat Cities, 2025, 2: 847-856.
[34] H Liu, M Zhang, X Han. A review of surface ozone source apportionment in China, Atmospheric and Oceanic Science Letters, 2020, 13(5).
[35] Gong K J,Li L,Shi Z H,et al.,2023.Quantitative analysis of the impact of pollutant transport on ozone levels in Nanjing within the Yangtze River Delta Region.Journal of Atmospheric Sciences, 2023, 46(5): 703-712.
[36] Cui M R, Bai L Y, Feng J Z, et al. Analysis of temporal and spatial variations of ozone coupling with dynamics of meteorological factors in the Beijing-Tianjin-Tangshan region. Acta Scientiae Circumstantiae, 2021, 41(2): 373-385.
[37] Professor Zongmei Fu's group at SUSTech used an adjoint model to refine the analysis of ground-based ozone sources. Southern University of Science and Technology, 2021. Available: https://osrp.sustech.edu.cn/news/707.html.
[38] Li H, Peng L, Bi F, et al. Research on coordinated control strategies for PM and ozone pollution in China. Environmental Science Research, 2019, 32(10): 1763-1778.
[39] Li H, Bao J M, Bi F, et al. Synergistic Control of PM2.5 and Ozone Pollution: Challenges and Responses. China Environment News, 2021. Available: http://www.chinaeol.net/zyzx/sjhjzz/zzlm/fmgs/202102/P020210204383551919228.pdf.
[40] Dong Z X, Ding D, Jiang Y Q, et al. Responses of PM2.5 and ozone to precursor emission reduction and meteorological change and their policy implications. Research of Environmental Sciences, 2023, 36(2): 223-236.
[41] MEE. Technical Guidelines for Satellite Remote Sensing Identification and Source Tracing of High-Value Areas of Atmospheric Volatile Organic Compounds, 2025. Available: https://www.mee.gov.cn/xxgk2018/xxgk/xxgk06/202511/W020251124619593423304.pdf.
[42] Sources of volatile organic compounds and their role in the formation of secondary air pollutants. Peking University, 2018. Available: http://www.research.pku.edu.cn/UploadFiles/2018111171630605.pdf.
[43] Blue Book on Atmospheric Ozone Pollution Prevention and Control in China. Science Press, 2023. Available: https://book.sciencereading.cn/shop/book/Booksimple/show.do?id=B14768BCCC7E4ACE7E063010B0A0A1E8F000.
[44] Li H, Bao J M, Bi F, et al. Synergistic Control of PM2.5 and Ozone Pollution: Challenges and Responses. China Environment News, 2021. Available: http://www.chinaeol.net/zyzx/sjhjzz/zzlm/fmgs/202102/P020210204383551919228.pdf.
[45] Zhou D R, Liu Y, Gao J, et al.Assessment of ozone sensitivity and emission reduction scenarios in typical pollution processes in eastern China.Transactions on Atmospheric Science, 2023, 46(5): 667-678.
[46] Wang R, Wang L, Sun J, et al. Maximizing ozone control by spatial sensitivity-oriented mitigation strategy in the Pearl River Delta region, China. Science of the Total Environment, 2023, 905: 166987.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3070

SOME PROPERTIES OF HYPERBOLIC TRIGONOMETRIC FUNCTIONS

HaoLin Fu
School of Mathematics and Statistics, Hubei University of Education, Wuhan 430205, Hubei, China.

Abstract: This paper investigates some fundamental properties of hyperbolic trigonometric functions on the hyperbolic number plane. Hyperbolic numbers form a commutative ring with zero divisors, generated by two real numbers via the Cartesian basis. Within this algebraic framework, this paper systematically establish the system of hyperbolic trigonometric functions. By fully leveraging the decomposition properties of hyperbolic numbers and their associated trigonometric functions, this work effectively overcomes the mathematical difficulties arising from the presence of zero divisors in the hyperbolic number ring. On this basis, this paper for the first time in the context of a hyperbolic number ring containing zero divisors, systematically derived and rigorously proved the addition theorems for the hyperbolic sine, cosine, tangent, and cotangent functions. On this basis, we have successfully derived and rigorously proved the addition theorems for the hyperbolic sine, cosine, tangent, and cotangent functions, establishing a complete system of angle addition formulas and laying a solid foundation for the theory of hyperbolic functions. The addition theorems established in this research will provide an important theoretical foundation for the further development of hyperbolic analysis in function theory, while simultaneously injecting new research momentum into the further study of the properties and expansions of hyperbolic trigonometric series. These theoretical achievements are expected to play significant roles in the study of hyperbolic differential equations, geometric analysis, and related physical problems.
Keywords: Hyperbolic numbers; Hyperbolic trigonometric functions; Identity transformation; Addition theorems

1 INTRODUCTION

Hyperbolic numbers, as a two-dimensional extension of the real number field, provide an algebraic framework with unique properties such as zero divisors. Ferhat Kuruz and Ali Dagdeviren studied matrices with hyperbolic number entries, establishing foundational properties of hyperbolic matrices [1]. Iskender Ozturk and Mustafa Ozdemir systematically investigated affine transformations in the hyperbolic number plane [2], revealing their invariants and geometric structure. Cayo D. and Nara P. applied hyperbolic geometry to surface determination via short curves [3], while Rachid A. et al. designed capstone courses for mathematics majors using complex and hyperbolic numbers [4]. Yuksel S. explored dual hyperbolic generalized Fibonacci numbers, further enriching the algebraic theory of hyperbolic systems [5].

In the complex domain, trigonometric functions play a central role in analysis, and analogous developments are sought in the hyperbolic setting. Chunli Li and Wenchang Chu computed several classes of definite integrals involving hyperbolic and trigonometric functions [6], providing precise analytical expressions. Medvegyev P. revisited the construction of elementary trigonometric functions [7], offering insights into their foundational structure. Youssef A. and Elhoucien E. derived cosine and sine addition laws with an automorphism [8], extending classical identities to broader algebraic contexts. Hassan A. and Mehdi D. proposed diffusive representations for fractional derivatives using sine and cosine functions [9], linking functional equations to fractional calculus. Ebanks B. studied the cosine-sine functional equation on semigroups [10], establishing general solutions under algebraic constraints.

Despite these advances, a systematic theory of hyperbolic trigonometric functions—particularly their addition theorems—remains underdeveloped. Guarin Garcia Julian derived transformations between trigonometric and hyperbolic functions based on physical models of cable statics [11]. Alibrahim Hamzah Ali and Das Saptarshi bridged p-special functions between generalized hyperbolic and trigonometric families [12], obtaining several connecting identities. Chunli Li and Wenchang Chu further evaluated improper integrals involving powers of inverse trigonometric and hyperbolic functions [13]. Stojiljkovic V. et al. established sharp bounds for trigonometric and hyperbolic functions with applications to fractional calculus [14]. Bagul J. et al. derived polynomial-exponential bounds for certain trigonometric and hyperbolic functions [15], contributing to quantitative estimates in analysis. However, a rigorous and systematic treatment of addition formulas for hyperbolic trigonometric functions within the hyperbolic number plane is still lacking, which motivates the present work.

2 THE BASIC THEORY OF HYPERBOLIC NUMBERS

Hyperbolic numbers are an extension of the real numbers, defined as numbers of the form:

$$\mathbb{H}=\mathbb{R}[h]=\{x_0+x_1h;\ x_0,\ x_1\in\mathbb{R},\ h^2=1,\ h\neq\mathbb{R}\}. \tag{1}$$

Let any two hyperbolic numbers be denoted as $h_1 = x_0 + x_1 h$ and $h_2 = y_0 + y_1 h$ (where $x_0, x_1, y_0, y_1 \in \mathbb{R}$, and $h^2 = 1,\ h \notin \mathbb{R}$), their definitions of equality, addition, and multiplication are as follows:

$$h_1 = h_2 \Leftrightarrow x_0 = y_0 \text{ and } x_1 = y_1, \tag{2}$$

$$h_1 + h_2 = (x_0 + y_0) + (x_1 + y_1)h, \tag{3}$$

$$h_1 h_2 = (x_0 y_0 + x_1 y_1) + (x_1 y_0 + x_0 y_1)h. \tag{4}$$

To simplify the power operation and function expansion of hyperbolic numbers, a special representation form called the idempotent basis can be introduced. Define the idempotent basis elements as: $h_+ = \frac{1}{2}(1+h)$ and $h_- = \frac{1}{2}(1-h)$. Meanwhile, for any hyperbolic number $h = x_0 + x_1 h$, difine $x_+ = x_0 + x_1$ and $x_- = x_0 - x_1$. At this point, the hyperbolic number can be rewritten as a linear combination of the idempotent basis:

$$h = x_+ h_+ + x_- h_-. \tag{5}$$

The idempotent basis has the following three key properties:

$$h = x_+ h_+ + x_- h_-, \tag{6}$$

$$h_+ h_- = 0. \tag{7}$$

The basis elements $\{h_+, h_-\}$ form another set of basis for $\mathbb{H}$, and every hyperbolic number has the unique idempotent representation.

Now given a hyperbolic number: $\delta = x_0 + x_1 h$, on the hyperbolic plane, the partial order is defined as follows it can be transformed via a coordinate change into: $\delta = \alpha h_+ + \beta h_-$. And $\alpha = x_0 + x_1,\ \beta = x_0 - x_1$.

For any hyperbolic number delta, the following holds:

$$\delta = \alpha h_+ + \beta h_-,\ \ \alpha, \beta \in \mathbb{R}, \tag{8}$$

$$\delta_1 \preceq \delta_2 \Leftrightarrow \delta_2 - \delta_1 \succeq 0, \tag{9}$$

$$\begin{cases} \delta_1 = \alpha_1 h_+ + \beta_1 h_- \\ \delta_2 = \alpha_2 h_+ + \beta_2 h_- \end{cases} \Rightarrow (\alpha_2 - \alpha_1)h_+ + (\beta_2 - \beta_1)h_- \pm 0 \Leftrightarrow \begin{cases} \alpha_2 \geq \alpha_1 \\ \beta_2 \geq \beta_1 \end{cases}, \tag{10}$$

$$\delta_1 \preceq \delta_2 \Leftrightarrow \delta_2 - \delta_1 \succeq 0 \Leftrightarrow \delta_1 - \delta_2 \in \mathbb{H}^+, \tag{11}$$

$$\mathbb{H}^+ = \{\delta = \alpha h_+ + \beta h_-; \alpha \geq 0, \beta \geq 0\}. \tag{12}$$

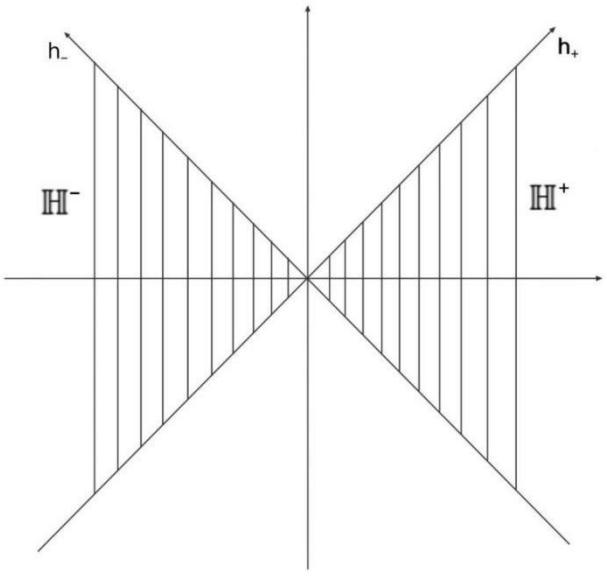


Figure 1 The Positive And Negative Hyperbolic Numbers

In Figure 1 illustrates the fundamental partition of the hyperbolic number plane into four sectors based on the signs of the idempotent components h_+ and h_-. The first and third quadrants, which are shaded, represent the sets of positive and negative hyperbolic numbers, respectively. This structure is fundamental to understanding the partial order and the

behavior of functions defined on hyperbolic numbers, as it directly stems from the existence of zero divisors in the algebra.

Define a hyperbolic interval as follows:

$$[\delta,\omega]_{\mathbb{H}}=\{\xi\in\mathbb{H}:\delta\preceq\xi\preceq\omega\}. \tag{13}$$

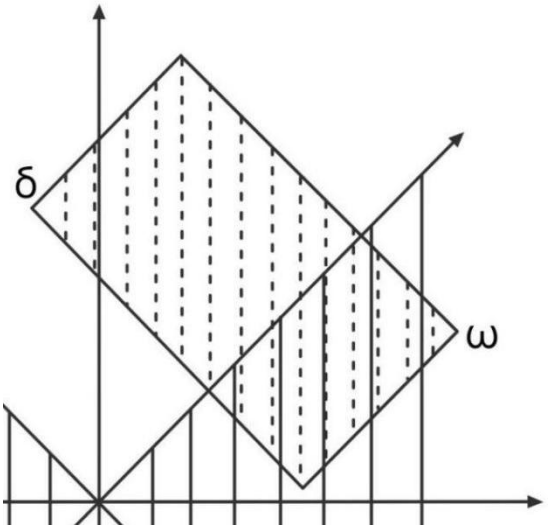


Figure 2 The Hyperbolic Interval $[\delta,\omega]_{\mathbb{H}}$.

In Figure 2 depicts a hyperbolic interval on the hyperbolic number plane. Geometrically, this interval represents a rectangular region defined by the partial order inherent to hyperbolic numbers, bounded by the idempotent components δ and ω. This structure generalizes the concept of a real-number interval to two dimensions within the hyperbolic plane. The hyperbolic interval is fundamental for defining domains of functions and convergence regions in hyperbolic analysis, providing a crucial geometric framework for subsequent theoretical developments.

The modulus of a hyperbolic number is defined as follows:

$$\left|\delta\right|_{\mathbb{H}}=\left|\alpha h_{+}+\beta h_{-}\right|_{\mathbb{H}}=\left|\alpha\right|h_{+}+\left|\beta\right|h_{-}\in\mathbb{H}^{+}. \tag{14}$$

3 DEFINITION AND PROPERTIES OF HYPERBOLIC TRIGONOMETRIC FUNCTIONS

3.1 Definition of Hyperbolic Trigonometric Functions

The computation of powers for hyperbolic numbers can be efficiently performed using idempotent bases. Specifically, for every $\mathrm{h}\in\mathbb{H}$ and $n\in\mathbb{R}$, the following relation can be obtained according to formula (6) and formula (7):

$$h^{n}=(x_{+}h_{+}+x_{-}h_{-})^{n}=(x_{+})^{n}(h_{+})^{n}+(x_{-})^{n}(h_{-})^{n}=(x_{+})^{n}h_{+}+(x_{-})^{n}h_{-}, \tag{15}$$

Formula (15) provides the idempotent basis representation for the power computation of hyperbolic numbers, laying the foundation for subsequent function definitions.

Similar to zero divisor factorization, if a hyperbolic function f is differentiable on the hyperbolic plane, then f admits a zero divisor factorization. Moreover, this operation can be formally expressed as a functional:

$$f(h)=f(x_{+}h_{+}+x_{-}h_{-})=f(x_{+})h_{+}+f(x_{-})h_{-}. \tag{16}$$

Equation (16) gives the idempotent decomposition form of a function acting on a hyperbolic number, which transforms the functional operation on the domain of hyperbolic numbers into the functional operations on its two idempotent components.

Example 1. If $f(h)=\delta=\sin(h)$ and $h=x_{0}+x_{1}h$ then, It has:

$$f(h)=\delta=\sin\left(h\right)=\sin(x_{0}+x_{1}h)=\sin(x_{+}h_{+}+x_{-}h_{-})=\sin(x_{+})h_{+}+\sin(x_{-})h_{-}, \tag{17}$$

Taking the hyperbolic interval $[-\pi,\pi]_{\mathbb{H}}$ as the domain, The graph of the hyperbolic trigonometric function is shown below:

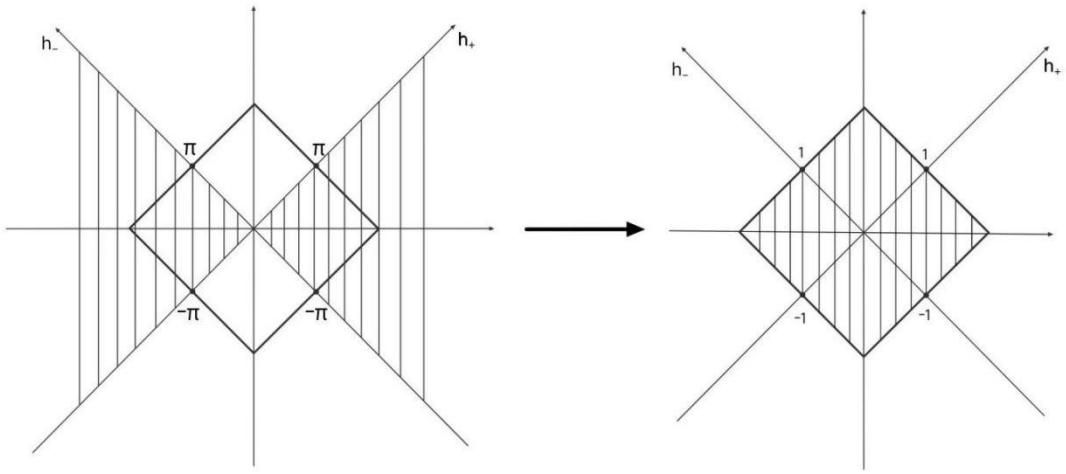


Figure 3 The Graph of The Hyperbolic Trigonometric Function

In Figure 3 presents the four-dimensional hyperbolic sine function through two two-dimensional cross-sectional views, illustrating the graph of the hyperbolic trigonometric function $f(h)$ over the hyperbolic interval $[-\pi,\pi]_{\mathbb{H}}$. These plots depict the variation patterns of the hyperbolic sine function within the hyperbolic number plane, revealing its unique oscillatory and growth characteristics. Geometrically, the visualization demonstrates how these functions evolve along the idempotent components, highlighting the intrinsic interplay between the algebraic structure of hyperbolic numbers and their geometric representation.

Let $\delta = x_0 + x_1 h = \alpha h_+ + \beta h_- \in \mathbb{H}$, then it follows that:

$$\begin{aligned} \sin(\delta) &= \sin(\alpha h_+ + \beta h_-) = \sin(\alpha) h_+ + \sin(\beta) h_-, \\ \cos(\delta) &= \cos(\alpha h_+ + \beta h_-) = \cos(\alpha) h_+ + \cos(\beta) h_-. \end{aligned} \tag{18}$$

By analogy with the complex domain, the hyperbolic tangent and cotangent functions are defined as follows:

If $\delta \in \mathbb{H}$, the hyperbolic tangent function is defined if and only if $\cos\delta$ is invertible, which requires that neither of values α and β equals $\frac{\pi}{2} + k\pi$ for any integer k. By computation, it follows that:

$$\tan\delta = \frac{\sin\delta}{\cos\delta} = \tan(\alpha) h_+ + \tan(\beta) h_-. \tag{19}$$

Similarly, if $\delta \in \mathbb{H}$, the hyperbolic cotangent function is defined if and only if $\sin\delta$ is invertible, which requires that neither of values α and β equals $k\pi$ for any integer k. By computation, it follows that

$$\cot\delta = \frac{\cos\delta}{\sin\delta} = \cot(\alpha) h_+ + \cot(\beta) h_-. \tag{20}$$

3.2 Study of The Properties of Hyperbolic Trigonometric Functions

Theorem 3.2.1: In the hyperbolic number plane $\mathbb{H}$, for any hyperbolic numbers $\delta_1, \delta_2 \in \mathbb{H}$, the hyperbolic sine function satisfies the addition formula:

$$\sin(\delta_1 + \delta_2) = \sin\delta_1 \cos\delta_2 + \cos\delta_1 \sin\delta_2. \tag{21}$$

Proof: Let $\forall \delta_1, \delta_2 \in \mathbb{H}$ .Within this framework, By the standard canonical representation, hyperbolic numbers are expressed as follows: $\delta_1 = \alpha_1 h_+ + \beta_1 h_- = x_0 + x_1 h,\ \delta_2 = \alpha_2 h_+ + \beta_2 h_- = y_0 + y_1 h.$ where the $\alpha_1, \alpha_2, \beta_1, \beta_2$ are real coefficients, and h_+, h_- denote the basis elements of $\mathbb{H}$.

First, calculate the sum $\delta_1 + \delta_2$:

$$\delta_1 + \delta_2 = (\alpha_1 h_+ + \beta_1 h_-) + (\alpha_2 h_+ + \beta_2 h_-) = (\alpha_1 + \alpha_2) h_+ + (\beta_1 + \beta_2) h_-. \tag{22}$$

From which we obtain:

$$\begin{aligned} \sin(\delta_1 + \delta_2) &= \sin((\alpha_1 + \alpha_2) h_+ + (\beta_1 + \beta_2) h_-) \\ &= \sin((\alpha_1 + \alpha_2) h_+) + \sin((\beta_1 + \beta_2) h_-). \end{aligned} \tag{23}$$

Apply the addition formula for the sine function in real analysis to each component, it has:

$$\begin{aligned}\sin(\delta_1+\delta_2)&=\sin((\alpha_1+\alpha_2)h_+ +(\beta_1+\beta_2)h_-)\\ &=(\sin(\alpha_1)\cos(\alpha_2)+\cos(\alpha_1)\sin(\alpha_2))h_+ +(\sin(\beta_1)\cos(\beta_2)+\cos(\beta_1)\ \sin(\beta_2\ ))h_-.\end{aligned} \tag{24}$$

It can changed:

$$\begin{aligned}\sin(\delta_1+\delta_2)&=(\sin(\alpha_1)\cos(\alpha_2)+\cos(\alpha_1)\sin(\alpha_2))\,h_+ +\\ &\quad(\sin(\beta_1)\cos(\beta_2)+\cos(\beta_1)\sin(\beta_2))\,h_-,\end{aligned} \tag{25}$$

$$\begin{aligned}\sin(\delta_1+\delta_2)&=(\sin(\alpha_1)\cos(\alpha_2)h_+ +\sin(\beta_1)\cos(\beta_2)h_-)+\\ &\quad(\cos(\alpha_1)\sin(\alpha_2)h_+ +\cos(\beta_1)\sin(\beta_2)h_-),\end{aligned} \tag{26}$$

$$\begin{aligned}\sin(\delta_1+\delta_2)&=(\sin(\alpha_1)h_+ +\sin(\beta_1)h_-)(\cos(\alpha_2)h_+ +\cos(\beta_2)h_-)+\\ &\quad(\cos(\alpha_1)h_+ +\cos(\beta_1)h_-)(\sin(\alpha_2)h_+ +\sin(\beta_2)h_-).\end{aligned} \tag{27}$$

Therefore, equation (21) is established.

Building upon the established angle addition formula for the hyperbolic sine function, now extend this analysis to the cosine function.

Theorem 3.2.2: In the hyperbolic number plane $\mathbb{H}$, for any hyperbolic numbers $\delta_1,\delta_2\in\mathbb{H}$, the hyperbolic cosine function satisfies the addition formula:

$$\cos(\delta_1+\delta_2)=\cos\delta_1\cos\delta_2-\sin\delta_1\sin\delta_2. \tag{28}$$

Proof: Let $\forall\delta_1,\delta_2\in\mathbb{H}$. By the canonical representation of hyperbolic numbers, this paper express them in as follows $\delta_1=\alpha_1h_+ +\beta_1h_-=x_0+x_1h,\ \delta_2=\alpha_2h_+ +\beta_2h_-=y_0+y_1h.$ where the are real coefficients, and denote the basis elements of $\mathbb{H}$.

From equation (22), it can be seen that:

$$\begin{aligned}\cos(\delta_1+\delta_2)&=\cos((\alpha_1+\alpha_2)h_+ +(\beta_1+\beta_2)h_-)\\ &=\cos((\alpha_1+\alpha_2)h_+)+\cos((\beta_1+\beta_2)h_-).\end{aligned} \tag{29}$$

Apply the addition formula for the cosine function in real analysis to each component, it has:

$$\begin{aligned}\cos(\delta_1+\delta_2)&=\cos((\alpha_1+\alpha_2)h_+ +(\beta_1+\beta_2)h_-)\\ &=(\cos(\alpha_1)\cos(\alpha_2)-\sin(\alpha_1)\sin(\alpha_2))h_+ +(\cos(\beta_1)\cos(\beta_2)-\sin(\beta_1)\ \sin(\beta_2\ ))h_-.\end{aligned} \tag{30}$$

It can changed:

$$\begin{aligned}\cos(\delta_1+\delta_2)&=(\cos(\alpha_1)\cos(\alpha_2)-\sin(\alpha_1)\sin(\alpha_2))\,h_+ +\\ &\quad(\cos(\beta_1)\cos(\beta_2)-\sin(\beta_1)\sin(\beta_2))\,h_-,\end{aligned} \tag{31}$$

$$\begin{aligned}\cos(\delta_1+\delta_2)&=(\cos(\alpha_1)\cos(\alpha_2)h_+ +\cos(\beta_1)\cos(\beta_2)h_-)-\\ &\quad(\sin(\alpha_1)\sin(\alpha_2)h_+ +\sin(\beta_1)\sin(\beta_2)h_-),\end{aligned} \tag{32}$$

$$\begin{aligned}\cos(\delta_1+\delta_2)&=(\cos(\alpha_1)h_+ +\cos(\beta_1)h_-)(\cos(\alpha_2)h_+ +\cos(\beta_2)h_-)-\\ &\quad(\sin(\alpha_1)h_+ +\sin(\beta_1)h_-)(\sin(\alpha_2)h_+ +\sin(\beta_2)h_-).\end{aligned} \tag{33}$$

Therefore, equation (28) is established.

Based on the results above, following the successful derivation of the angle addition formulas for both sine and cosine functions in the hyperbolic number plane, now complete this trigonometric foundation by examining the tangent function.

Theorem 3.2.3: Let $\delta_1,\delta_2\in\mathbb{H}$, $\delta_i=\alpha_ih_+ +\beta_ih_-\in\mathbb{H}$, and neither of values α and β equals $\frac{\pi}{2}+k\pi$.

When $\cos\delta_1\cos\delta_2-\sin\delta_1\sin\delta_2$ is invertible, it has:

$$\tan(\delta_1+\delta_2)=\frac{\tan\delta_1+\tan\delta_2}{1-(\tan\delta_1\tan\delta_2)}. \tag{34}$$

But when $\cos\delta_1\cos\delta_2-\sin\delta_1\sin\delta_2$ is invertible, it has:

$$\tan(\delta_1+\delta_2)=\frac{\tan\alpha_1+\tan\alpha_2}{1-(\tan\alpha_1\tan\alpha_2)}h_+ +\frac{\tan\beta_1+\tan\beta_2}{1-(\tan\beta_1\tan\beta_2)}h_-. \tag{35}$$

Proof: $\forall \delta_1, \delta_2 \in \mathbb{H}$, when $\cos\delta_1 \cos\delta_2 - \sin\delta_1 \sin\delta_2$ is invertible. By the definition of the hyperbolic tangent function, $\tan\delta = \dfrac{\sin\delta}{\cos\delta}$. Thus, $\tan(\delta_1 + \delta_2) = \dfrac{\sin(\delta_1 + \delta_2)}{\cos(\delta_1 + \delta_2)}$.

From Theorem 3.2.1 and Theorem 3.2.2, it follows that:

$$\tan(\delta_1 + \delta_2) = \frac{\sin(\delta_1 + \delta_2)}{\cos(\delta_1 + \delta_2)} = \frac{\sin(\delta_1)\cos(\delta_2) + \cos(\delta_1)\sin(\delta_2)}{\cos(\delta_1)\cos(\delta_2) - \sin(\delta_1)\sin(\delta_2)}. \tag{36}$$

Divide both the numerator and the denominator by $\cos\delta_1 \cos\delta_2$, it simplify to obtain (34).

When $\cos\delta_1 \cos\delta_2 - \sin\delta_1 \sin\delta_2$ is not invertible. Let $\forall \delta_1 = \alpha_1 h_+ + \beta_1 h_-$ and $\forall \delta_2 = \alpha_2 h_+ + \beta_2 h_-$, express $\delta_1 + \delta_2$ in terms of the idempotent basis :

$$\delta_1 + \delta_2 = (\alpha_1 + \alpha_2) h_+ + (\beta_1 + \beta_2) h_-. \tag{37}$$

From which we obtain:

$$\begin{aligned} \tan(\delta_1 + \delta_2) &= \tan((\alpha_1 + \alpha_2) h_+ + (\beta_1 + \beta_2) h_-) \\ &= \tan((\alpha_1 + \alpha_2) h_+) + \tan((\beta_1 + \beta_2) h_-). \end{aligned} \tag{38}$$

Apply the addition formula for the tangent function in real analysis to each component, it has:

$$\tan(\alpha_1 + \alpha_2) = \frac{\sin(\alpha_1 + \alpha_2)}{\cos(\alpha_1 + \alpha_2)} = \frac{\tan\alpha_1 + \tan\alpha_2}{1 - \tan\alpha_1 \tan\alpha_2}, \tag{39}$$

$$\tan(\beta_1 + \beta_2) = \frac{\sin(\beta_1 + \beta_2)}{\cos(\beta_1 + \beta_2)} = \frac{\tan\beta_1 + \tan\beta_2}{1 - \tan\beta_1 \tan\beta_2}. \tag{40}$$

It can changed:

$$\begin{aligned} \tan(\delta_1 + \delta_2) &= \tan((\alpha_1 + \alpha_2) h_+ + (\beta_1 + \beta_2) h_-) \\ &= \tan(\alpha_1 + \alpha_2) h_+ + \tan(\beta_1 + \beta_2) h_-, \end{aligned} \tag{41}$$

$$\begin{aligned} \tan(\delta_1 + \delta_2) &= \frac{\sin(\alpha_1 + \alpha_2)}{\cos(\alpha_1 + \alpha_2)} h_+ + \frac{\sin(\beta_1 + \beta_2)}{\cos(\beta_1 + \beta_2)} h_- \\ &= \frac{\tan(\alpha_1) + \tan(\alpha_2)}{1 - \tan(\alpha_1)\tan(\alpha_2)} h_+ + \frac{\tan(\beta_1) + \tan(\beta_2)}{1 - \tan(\beta_1)\tan(\beta_2)} h_-. \end{aligned} \tag{42}$$

Therefore, equation (35) is established.

The hyperbolic cotangent is the reciprocal of the hyperbolic tangent. Building on the angle addition formula for tanh, now derive the corresponding formula for coth. This establishes the following theorem.

Theorem 3.2.4: Let $\delta_1, \delta_2 \in \mathbb{H}$, $\delta_i = \alpha_i h_+ + \beta_i h_- \in \mathbb{H}$, and neither of values $\alpha = \sum \alpha_i$ and $\beta = \sum \beta_i$ equals $\dfrac{\pi}{2} + k\pi$. When $\sin\delta_1 \cos\delta_2 + \cos\delta_1 \sin\delta_2$ is invertible , it has:

$$\cot(\delta_1 + \delta_2) = \frac{\cot\delta_1 \cot\delta_2 - 1}{\cot\delta_1 + \cot\delta_2}. \tag{43}$$

But when $\sin\delta_1 \cos\delta_2 + \cos\delta_1 \sin\delta_2$ is invertible , it has:

$$\cot(\delta_1 + \delta_2) = \frac{\cot\alpha_1 \cot\alpha_2 - 1}{\cot\alpha_1 + \cot\alpha_2} h_+ + \frac{\cot\beta_1 \cot\beta_2 - 1}{\cot\beta_1 + \cot\beta_2} h_-. \tag{44}$$

Proof: $\forall \delta_1, \delta_2 \in \mathbb{H}$, when $\sin\delta_1 \cos\delta_2 + \cos\delta_1 \sin\delta_2$ is invertible. By the definition of the hyperbolic cotangent function, $\cot\delta = \dfrac{\cos\delta}{\sin\delta}$. Thus, $\cot(\delta_1 + \delta_2) = \dfrac{\cos(\delta_1 + \delta_2)}{\sin(\delta_1 + \delta_2)}$.

From Theorem 3.2.1 and Theorem 3.2.2, it follows that:

$$\cot(\delta_1 + \delta_2) = \frac{\cos(\delta_1 + \delta_2)}{\sin(\delta_1 + \delta_2)} = \frac{\cos(\delta_1)\cos(\delta_2) - \sin(\delta_1)\sin(\delta_2)}{\sin(\delta_1)\cos(\delta_2) + \cos(\delta_1)\sin(\delta_2)}. \tag{45}$$

Divide both the numerator and the denominator by $\sin\delta_1 \sin\delta_2$, it simplify to obtain (43).

When $\sin\delta_1\cos\delta_2+\cos\delta_1\sin\delta_2$ is not invertible. Let $\forall\delta_1=\alpha_1h_++\beta_1h_-$ and $\forall\delta_2=\alpha_2h_++\beta_2h_-$, express $\delta_1+\delta_2$ in terms of the idempotent basis :

$$\delta_1+\delta_2=(\alpha_1+\alpha_2)h_++(\beta_1+\beta_2)h_-. \tag{46}$$

From which we obtain:

$$\begin{aligned}\cot(\delta_1+\delta_2)&=\cot((\alpha_1+\alpha_2)h_++(\beta_1+\beta_2)h_-)\\&=\cot((\alpha_1+\alpha_2)h_+)+\cot((\beta_1+\beta_2)h_-).\end{aligned} \tag{47}$$

Apply the addition formula for the cotangent function in real analysis to each component, it has:

$$\cot(\alpha_1+\alpha_2)=\frac{\cos(\alpha_1+\alpha_2)}{\sin(\alpha_1+\alpha_2)}=\frac{\cot\alpha_1\cot\alpha_2-1}{\cot\alpha_1+\cot\alpha_2}, \tag{48}$$

$$\cot(\beta_1+\beta_2)=\frac{\cos(\beta_1+\beta_2)}{\sin(\beta_1+\beta_2)}=\frac{\cot\beta_1\cot\beta_2-1}{\cot\beta_1+\cot\beta_2}. \tag{49}$$

It can changed:

$$\begin{aligned}\cot(\delta_1+\delta_2)&=\cot((\alpha_1+\alpha_2)h_++(\beta_1+\beta_2)h_-)\\&=\cot(\alpha_1+\alpha_2)h_++\cot(\beta_1+\beta_2)h_-,\end{aligned} \tag{50}$$

$$\begin{aligned}\cot(\delta_1+\delta_2)&=\frac{\cos(\alpha_1+\alpha_2)}{\sin(\alpha_1+\alpha_2)}h_++\frac{\cos(\beta_1+\beta_2)}{\sin(\beta_1+\beta_2)}h_-\\&=\frac{\cot\alpha_1\cot\alpha_2-1}{\cot\alpha_1+\cot\alpha_2}h_++\frac{\cot\beta_1\cot\beta_2-1}{\cot\beta_1+\cot\beta_2}h_-.\end{aligned} \tag{51}$$

Therefore, equation (44) is established.

4 CONCLUSIONS

This paper has conducted a systematic investigation into the fundamental properties of hyperbolic trigonometric functions defined on the hyperbolic number plane. The core achievement lies in the rigorous establishment of the addition theorems for the four primary hyperbolic trigonometric functions: sine, cosine, tangent, and cotangent, as detailed in Theorems 3.2.1 through Theorems 3.2.4. The successful derivation of these formulas relied critically on leveraging the unique decomposition properties afforded by the idempotent basis of hyperbolic numbers, which effectively circumvented the analytical challenges posed by the presence of zero divisors in the underlying algebra. The proven addition theorems constitute a fundamental and complete set of identities for hyperbolic trigonometric functions, significantly enriching their theoretical framework. These results not only solidify the theoretical foundation of hyperbolic function theory but also provide indispensable analytical tools. They are poised to facilitate and stimulate further research in hyperbolic analysis, particularly in the study of hyperbolic differential equations, integral transforms, and the convergence properties of hyperbolic trigonometric series. Ultimately, this work opens several promising avenues for future research. The established properties are expected to serve as a cornerstone for more advanced analytical developments within the hyperbolic number system and to find meaningful applications in related physical and geometric contexts where the structure of hyperbolic numbers proves natural and powerful.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Ferhat K, Ali D. Matrices with Hyperbolic Number Entries. Turkish Journal of Mathematics, 2022, 14(2): 306-313.
[2] Iskender O, Mustafa O. Affine transformations of hyperbolic number plane. Boletin de la Sociedad Matematica Mexicana, 2022, 28(3): 61.
[3] Cayo D, Nara P. Determining surfaces by short curves and applications. arXiv, 2402.18676, 2024.
[4] Rachid A, Curtis K, Barbara S. Capstone Studies for Math Majors via Complex and Hyperbolic Numbers. Primus: Problems Resources and Issues in Mathematics Undergraduate Studies, 2023, 33(6): 602-621.
[5] Yuksel S. On dual hyperbolic generalized Fibonacci numbers. Indian Journal of Pure and Applied Mathematics, 2021, 52(1): 1-17.
[6] Chunli Li, Wenchang Chu. Four Classes of Definite Integrals about Hyperbolic and Trigonometric Functions. Computational Mathematics and Mathematical Physics, 2023, 63(7): 1199-1217.
[7] Medvegyev P. On the construction of the elementary trigonometric functions. Pure Mathematics and Applications, 2022, 30(3): 54-64.

[8] Youssef A, Elhoucien E. Cosine and Sine Addition and Subtraction Law with an Automorphism. Annales Mathematicae Silesianae, 2024, 38(2): 155-176.
[9] Hassan A, Mehdi D. The sine and cosine diffusive representations for the Caputo fractional derivative. Applied Numerical Mathematics, 2024, 204: 265-290.
[10] Ebanks B. The Cosine-Sine Functional Equation on Semigroups. Annales Mathematicae Silesianae, 2021, 36(1): 30-52.
[11] Julian G. Transformation of Trigonometric Functions into Hyperbolic Functions Based on Cable Statics. Applied Sciences, 2025, 15(5): 2647-2647.
[12] Ali A, Saptarshi D. Bridging the p -Special Functions between the Generalized Hyperbolic and Trigonometric Families. Mathematics, 2024, 12(8): 1242.
[13] Chunli Li, Wenchang Chu. Improper Integrals Involving Powers of Inverse Trigonometric and Hyperbolic Functions. Mathematics, 2022, 10(16): 2980-2980.
[14] Stojiljkovic V, Radojevic S, Cetin E, et al. Sharp Bounds for Trigonometric and Hyperbolic Functions with Application to Fractional Calculus. Symmetry, 2022, 14(6): 1260-1260.
[15] Bagul J, Dhaigude M, Kostic M, et al. Polynomial-Exponential Bounds for Some Trigonometric and Hyperbolic Functions, Axioms, 2021, 10(4): 308-308.

World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3071

MULTI-AGENT COORDINATION AND RESOURCE ALLOCATION OPTIMIZATION STRATEGIES FOR SMOKE SCREEN DEPLOYMENT AGAINST DYNAMIC TARGETS

TianJian Zhong
School of Automation, Central South University, Changsha 410083, Hunan, China.

Abstract: This paper investigates optimal defense strategies for countering single or multiple high-speed incoming missiles by deploying smoke screen decoys through coordinated operations of multiple unmanned aerial vehicles (UAVs) in complex dynamic battlefield environments. The research focuses on the challenging problem of multi-agent cooperative defense and large-scale resource allocation. First, for a three-UAV, single-decoys cooperative countermeasure strategy, a joint optimization model incorporating parameters such as UAV flight direction, velocity, deployment timing, and detonation timing is constructed. A differential evolution algorithm is employed to solve this high-dimensional combinatorial optimization problem. Through iterative differential evolution, optimal parameter configurations for the three UAVs are obtained, achieving effective shielding against missile M1. Subsequently, the scenario is expanded to the most complex five-UAV, three-missile integrated defense problem. This problem involves multiple UAVs, multiple missiles, multiple missile deployments, and resource allocation, constituting a typical large-scale combinatorial optimization problem. This paper innovatively proposes a two-layer hybrid optimization framework combining the Hungarian algorithm and genetic algorithm. The upper layer utilizes the Hungarian algorithm to determine the optimal task allocation between missiles and UAVs, minimizing the initial distance cost to ensure optimal spatial separation. The lower layer, with fixed allocations, employs a genetic algorithm to refine the deployment parameters for each UAV. The final optimized solution successfully achieves comprehensive interference against three missiles, demonstrating the dual-layer model's capability to obtain near-global-optimal multi-objective defense strategies while ensuring real-time feasibility.
Keywords: Spatial coordinate system; Kinematic model; Genetic algorithm

1 INTRODUCTION

Smoke-based jamming technology has found extensive application in modern military operations. Its core value lies in generating smoke clouds to create a protective screen in front of targets, effectively obscuring enemy missile detection systems and safeguarding genuine objectives. In practical applications, existing long-endurance UAVs equipped with smoke countermeasure munitions must instantly adjust flight states and deployment strategies (e.g., flight direction, velocity, release point, detonation point) upon detecting missiles to maximize the effective shielding duration of smoke clouds over the true target[1-2]. This challenge raises two core technical questions: First, how to precisely calculate the effective shielding duration of a single smoke decoy against a dynamic real target, given known motion states and initial parameters of the missile, UAV, and smoke decoy. Second, how to determine the single-resource jamming strategy that maximizes the real target's complete shielding duration by optimizing strategic parameters such as the UAV's flight direction, speed, and the smoke decoy's deployment and detonation timing. Previous studies addressing the masking effects of complex dynamic targets often relied excessively on simplified geometric models or coarse time steps, failing to provide precise calculations meeting high-precision engineering requirements[3]. The primary innovations of this research are: First, establishing spatial dynamic models for missiles, UAVs, and smoke grenades, laying the foundation for precise calculations. Second, it achieves high-precision determination of smoke-induced complete concealment by discretely sampling the target's cylindrical surface and applying geometric discriminants for line-sphere intersections. Most critically, addressing the multivariate nonlinear optimization problem of maximizing concealment duration, this section introduces a genetic algorithm for global optimization. This algorithm logically integrates complex kinematic models and geometric criteria without simplification, thereby identifying the optimal single-resource deployment strategy. The research approach comprises two steps: First, constructing spatial and geometric determination models to precisely calculate concealment duration under given parameters; Second, using this duration as the objective function, the genetic algorithm searches the strategy parameter space for the optimal combination[4-5].

2 DIFFERENTIAL EVOLUTION OPTIMIZATION FOR MULTI-UAV COOPERATIVE DEFENSE

2.1 Model Establishment

This part requires using three UAVs (FY1, FY2, FY3), each deploying one smoke interference bomb, to interfere with missile M1. This section will detail how to establish a mathematical model to solve this problem.
Model Constraints

Deployment time constraint: The time interval between smoke interference bombs deployed by each UAV must be at least 1 second:

$$t_{j(k+1),p} \geq t_{jk,p} + 1, \forall j, k \tag{1}$$

Detonation time constraint: The detonation time shall not be earlier than 1 second after deployment:

$$t_{jk,b} \geq t_{jk,p} + 1, \forall j, k \tag{2}$$

Speed and direction constraints: The UAV's speed and direction angle must be within the allowable range:

$$v_j \in [70,140], \forall j \tag{3}$$

$$\theta_j \in [0,2\pi), \forall j \tag{4}$$

Resource allocation constraint: Each UAV can deploy at most one smoke interference bomb:

$$\sum_{m} x_{jkm} \leq 1, \forall j, k \tag{5}$$

where x_{jkm} is the smoke bomb allocation indicator variable, $x_{jkm} \in \{0,1\}$, indicating whether smoke bomb k is used to interfere with missile m.

Smoke bomb allocation constraint: Each smoke bomb can only be allocated to one target:

$$\sum_{k} x_{jkm} \leq 1, \forall j, m \tag{6}$$

Objective Function

Maximize the total effective shielding time:

$$\max \sum_{j=1}^{3} \sum_{k=1}^{K_j} T_{c,jk} \tag{7}$$

where $K_j = 1$ is the number of smoke bombs deployed by UAV j, and $T_{c,jk}$ is the effective shielding time of the k-th smoke bomb deployed by the j-th UAV.

2.2 Solution Model

The Differential Evolution (DE) algorithm is used to solve this optimization problem. The differential evolution algorithm is a population-based optimization algorithm suitable for global optimization problems in continuous spaces[6]. It gradually approaches the optimal solution through mutation, crossover, and selection operations during iterations. Randomly generate a set of UAV flight speeds, direction angles, deployment times, and detonation times as the initial population. Each individual consists of four parameters: flight speed v_j, direction angle θ_j, deployment time $t_{d,j}$, and detonation time $t_{b,j}$. Calculate the total effective shielding time for each individual. The fitness function is the total effective shielding time, with the goal of maximizing this value. Perform mutation operation on each individual to generate a mutation vector. The mutation vector is generated based on the difference between the current individual and two other randomly selected individuals[7-8]. Perform crossover operation between the mutation vector and the current individual to generate a trial vector. The crossover operation randomly selects corresponding parameters from the mutation vector and the current individual for combination.

The algorithm compares the fitness of the trial vector and the current individual. If the fitness of the trial vector is better than that of the current individual, the trial vector replaces the current individual[9-10]. Repeat steps 3-5 until the preset number of iterations is reached or the convergence condition is satisfied. During the solution process, the UAV flight parameters are first discretized to adapt to the search space of the differential evolution algorithm. The specific parameter ranges are as follows:

Flight speed: $v_j \in [70,140]$ m/s

Flight direction angle: $\theta_j \in [0,2\pi)$

Deployment time: $t_{j,p} \geq 0$

Detonation time: $t_{j,b} \geq t_{j,p} + 1$

Through iterative optimization of the differential evolution algorithm, three sets of solutions meeting the conditions are found, each corresponding to the optimal flight parameters of one UAV. The specific solutions are as follows:

Solution 1: $i = 44, j = 118, q = 5, s = 7$, Result = 4.1 seconds

Solution 2: $i = 43, j = 80, q = 12, s = 7$, Result = 3.35 seconds

Solution 3: $i = 49, j = 98, q = 8, s = 6$, Result = 0.72 seconds

2.3 Analysis of Problem 4 Results

The differential evolution algorithm calculates three sets of solutions meeting the conditions, each corresponding to the flight parameters and deployment strategy of one UAV. These solutions maximize the effective shielding time for missile M1 while satisfying all constraints. The specific results are as follows: Flight direction angle: 252.10°; Flight speed: 118 m/s; Deployment point coordinates: (11818.67, 838.55, 1400) m; Detonation point coordinates: (11564.82, 52.53, 1159.9) m; Effective shielding time: 4.1 seconds; Flight direction angle: 246.37°; Flight speed: 80 m/s; Deployment point coordinates: (11615.23, 520.48, 1400) m; Detonation point coordinates: (11390.79, 7.43, 1159.9) m;

Effective shielding time: 3.35 seconds; Flight direction angle: 280.75°; Flight speed: 98 m/s; Deployment point coordinates: (12146.23, 629.76, 1400) m; Detonation point coordinates: (12255.89, 52.08, 1223.6) m; Effective shielding time: 0.72 seconds.

From the results, it can be seen that the deployment strategy of FY1 provides the longest effective shielding time of 4.1 seconds, indicating that under the given conditions, the smoke bomb can more effectively cover the flight path of missile M1. FY2 also performs well, providing a shielding time of 3.35 seconds. However, the shielding time of FY3 is relatively short, only 0.72 seconds, which may be due to the poor selection of its deployment and detonation points, resulting in the smoke cloud failing to provide sufficient shielding during the critical phase when the missile approaches the real target. Through visualization results, we can see the flight trajectory of each UAV, the deployment and detonation positions of the smoke bombs, and the effective shielding effect on missile M1. The graphs show that the smoke bombs of FY1 and FY2 successfully form effective shielding when the missile approaches the decoy target, while the smoke bomb of FY3 also forms shielding but is less effective than the previous two. The strategy of collaborative deployment of smoke interference bombs by three UAVs is successfully obtained through the differential evolution algorithm to maximize the effective shielding time for missile M1. The results show that by optimizing the flight parameters and deployment strategy of UAVs, continuous and effective interference can be provided during the missile's approach phase.

3 A TWO-LAYER OPTIMIZATION FRAMEWORK FOR MULTI-AIRCRAFT MULTI-WEAPON MISSION ALLOCATION

3.1 Model Establishment

This part requires using 5 UAVs (each with a maximum of 3 smoke bombs) to simultaneously interfere with 3 missiles (M1, M2, M3). Compared with the previous four problems, this problem introduces multiple UAVs, multiple missiles, multiple bomb deployments, and resource allocation, making it a typical large-scale combinatorial optimization problem.

On the basis of the single-UAV/multi-UAV modeling in Problems 3 and 4, we establish a two-layer structure of "UAV-missile allocation" and "multi-bomb deployment optimization": the upper layer matches 3 missiles with 5 UAVs (ensuring optimal spatial distance); the lower layer optimizes the deployment parameters (course, speed, deployment time, detonation delay) of each UAV; the overall goal is to maximize the total shielding time of the three missiles when approaching the real target.

Decision Variables

$\{\theta_i, v_i, t_{i,j,p}, t_{i,j,b}, x_{j,k,m}\}$, where $x_{j,k,m} \in \{0,1\}$ is the assignment variable, indicating whether the j-th smoke bomb of UAV i is used to interfere with missile m.

Constraints

Deployment time interval: Smoke bombs continuously deployed by the same UAV must satisfy $\Delta t \geq 1$ second.

Speed/direction constraints: $70 \leq v_i \leq 140$, $\theta_i \in [0,2\pi)$.

Resource constraint: Each UAV can carry at most 3 bombs.

Allocation constraint: Each bomb can only be assigned to one missile.

Objective Function

$$\max \mathcal{L} = \sum_{m=1}^{3} \sum_{i=1}^{5} \sum_{j=1}^{3} x_{j,k,m} \cdot T_{i,j,m} \tag{8}$$

where $T_{i,j,m}$ represents the effective shielding time of the j-th smoke bomb of UAV i on missile m.

This model inherits the geometric shielding determination method (missile-real target-cloud cluster relationship) from Problems 1 to 4. However, due to the superposition of multiple targets and multiple resources, the problem exhibits exponential complexity and requires heuristic intelligent algorithms for solution.

3.2 Model Solution

We adopt a two-layer hybrid optimization framework of "Hungarian algorithm + Genetic Algorithm (GA)". The upper layer (task allocation) uses the Hungarian algorithm to solve the optimal matching problem between 3 missiles and 5 UAVs. The cost matrix takes the minimum distance from the initial position of the UAV to the missile's flight path to ensure that UAVs are preferentially allocated to missiles that are easier to interfere with spatially. For the remaining 2 UAVs, the "nearest path distance" criterion is used for auxiliary allocation. The lower layer (deployment parameter optimization) calls the genetic algorithm GA to optimize the deployment plan of a single UAV under fixed task allocation.

Chromosome Encoding

$[\theta, v, q_1, s_1, q_2, s_2, q_3, s_3]$, a total of 8 continuous variables, corresponding to the course, speed, and deployment time and delay of 3 bombs respectively. Feasibility checks are also performed; individuals with a deployment interval of less than 1 second are directly judged as invalid, and their fitness is set to zero.

Fitness Function

Call eval_cover_time() to return the total shielding time.
Genetic Operators
Roulette selection, single-point crossover, and multi-point mutation with a probability of 0.25.
Algorithm Parameters
Population size: 100, number of iterations: 100, crossover probability: 0.7.

3.3 Result Analysis

After optimization, the task allocation is as follows: FY1→M1, FY2→M2, FY3→M3, FY4→M2, FY5→M3. For each UAV, the optimal flight and deployment parameters are obtained through optimization (sorted by fitness).
The operation results show that after genetic algorithm optimization, all 5 UAVs (FY1－FY5) obtain clear flight direction angles, flight speeds, as well as the deployment coordinates, detonation coordinates, and corresponding effective shielding times of 3 smoke bombs.
FY1−FY2−FY5 (allocated to M1, M2, M3): The overall shielding times are approximately 9.29 s, 4.06 s, and 2.28 s respectively, indicating that the three UAVs allocated in the early stage play a major protective role.
FY3 and FY4 provide supplementary protection for M3 and M2 respectively, increasing the effective shielding time by 1.58 s and 3.17 s.
From the single-UAV results, in most cases, the first interference bomb contributes the longest shielding time (e.g., the first bomb of FY1 has an effective duration of up to 4.28 s), while the third bomb often has a shielding time close to 0 s, indicating that the available window shortens as the missile gradually approaches the target.

4 CONCLUSIONS

This section successfully establishes an accurate calculation model for effective smoke screen duration and utilizes a genetic algorithm to optimize the single-resource interference strategy. For duration estimation, we constructed a spatial scenario involving missiles, UAVs, smoke grenades, and real targets, establishing dynamic models for each entity. Through discrete sampling of the target surface and geometric criteria for segment-sphere intersections, we precisely calculated the complete effective concealment duration of 1.39 seconds under a given deployment strategy. For strategy optimization, we constructed an optimization model with UAV flight parameters and smoke grenade timing parameters as decision variables to maximize effective concealment duration. This model employs a genetic algorithm for global optimization, ultimately achieving a maximum effective concealment duration of 4.51 seconds, significantly enhancing single-resource deployment performance.
This study retains certain limitations in modeling and solution processes. Model construction incorporates simplifying assumptions, such as assuming all entities move without resistance. Additionally, we assumed that after detonation, the smoke-screen interference projectile maintains a constant spherical shape with unchanging radius throughout its effective interference duration, disregarding other factors that could distort its shape. Furthermore, the true target was modeled as a regular cylinder, neglecting the possibility of irregular structures encountered in practice. Future research should focus on enhancing the model's engineering applicability and accuracy. Exploring the incorporation of environmental drag terms into the dynamic model and developing more sophisticated smoke cloud shape models could improve prediction accuracy. To address the limitation where genetic algorithms may converge to similar parameters in later iterations, making it difficult to find solutions superior to existing ones, future work could investigate combining other optimization algorithms or refining inertial weights. This would further enhance the global optimization capability for finding optimal strategies in complex, multi-variable scenarios.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Song Qinxiang, Zhan Tingbian, Hu Yaping, et al. Study on the Smoke Screen Shielding Characteristics of Modified Fly Ash. Pyrotechnics, 2025: 1-6.
[2] Wang Shouxing, Wang Fei, Bao Xiaoyun, et al. Preparation and Interference Performance Study of Multi-Spectral Smoke Screen Materials. China Building Materials Science and Technology, 2025, 34(04): 81-85.
[3] Xie Xinghua, Qian Qingyu, Xie Qiang, et al. Formulation Design and Far-Infrared Interference Performance of Combustion-Type Smoke Agent Based on Organic Solid Waste. Journal of Henan University of Urban Construction, 2025, 34(04): 71-77.
[4] Ding Jialin. Research on Smoke Interference Effectiveness Evaluation Technology Based on Image Correlation. Academy of Military Sciences, 2025.
[5] Guo Aiqiang, Gao Xinbao. Research and Development Trends of Smoke-Generating Ammunition Interference Effectiveness Evaluation System. Journal of Ordnance Engineering, 2025, 46(01): 38-45.
[6] Xie Yue, Li Ming, Hou Yali, et al. Equivalent Modeling Method for Laser Fuse Anti-Smoke Interference in Concentration-Layered Scenarios. Journal of Detection and Control, 2024, 46(04): 39-45.

[7] Liang Jianxing, Chen Qingliang, Liu Hu, et al. A Measurement and Calculation Method for Smoke Curtain Coverage Area in Passive Jammer Testing. Optics and Optoelectronics Technology, 2024, 22(03): 57-62.
[8] Ding Jialin, Chen Chunsheng, Li Qingwei, et al. Evaluation Indicators and Calculation Methods for Smoke Screen Interference Effectiveness. Infrared, 2024, 45(03): 29-39.
[9] Chen X, Hu Y, Gu Y, et al. Technique based on the grayscale value for evaluating the shielding performance of infrared smokescreen. Optical Engineering, 2024, 63(3): 034107.
[10] Chen L, Li X, Wang X, et al. Research on evaluation methods for smoke screen shielding and interference effects. Progress in Lasers and Optoelectronics, 2023, 60(22): 41-50.