



ARTICLE

Performance Evaluation of Multi-Stage Equipment Allocation Chains Based on Interval Cloud Model and Network DEA

Tianqiao Yan, Cenyu Hu and Yabin Wang*

Shijiazhuang Campus of Army Engineering University of PLA, Shijiazhuang, China

*Corresponding Author: Yabin Wang. Email: wangyabin123@163.com

Received: 13 July 2026; Accepted: 04 September 2026; Published: 22 October 2026

ABSTRACT: Performance evaluation of equipment allocation chains is a critical closed loop linking strategy optimization and resilience design. To address the limitations of traditional evaluation methods—namely, the fragmentation of multi-stage transmission mechanisms, the neglect of expert cognitive uncertainty, and the lack of robustness verification—this paper constructs an integrated framework of “cloud model quantification, entropy-weight dimensionality reduction, network data envelopment analysis (DEA) evaluation, and Bootstrap robust ranking”. Specifically, an interval cloud model is employed to transform experts’ fuzzy linguistic evaluations into cognitive consensus boundaries; a stage-wise dimensionality reduction model based on fused entropy weights and game-theoretic combination weighting is designed to output subjective–objective coordinated composite intervals; a polarity-corrected interval serial network DEA (ISN-DEA) model with intermediate linking constraints is constructed to achieve four-stage efficiency decoupling and bottleneck identification from demand planning to performance feedback; and a possibility degree matrix is introduced to produce a complete ranking, whose robustness and convergence validity are jointly verified via Bootstrap resampling, multi-baseline comparison, and sensitivity experiments. Case studies demonstrate that the proposed method can effectively characterize the performance degradation patterns of the chain under different disruption scenarios, differentially identify typical performance states such as the “critical state” and the “system failure state”, precisely locate stage bottlenecks, and output statistically significant strategy rankings, thereby providing quantitative decision support for equipment prepositioning strategies and chain reconfiguration.

KEYWORDS: Chain performance evaluation; interval cloud model; network DEA; entropy-weight dimensionality reduction; bootstrap robustness

1 Introduction

Equipment allocation and supply support, as a core component of the military logistics system, directly bears on the precise execution of equipment support missions and the efficient flow of resources. Against the backdrop of rapidly evolving modern warfare forms and an increasingly complex battlefield environment, scientifically evaluating the operational effectiveness of equipment allocation supply management chains and identifying critical bottlenecks has become an urgent imperative for enhancing equipment support capabilities [1–3]. Particularly empowered by emerging technologies such as artificial intelligence and big data, resilience assessment and performance evolution analysis of military logistics systems have emerged as frontier research directions [4,5].

In the field of supply chain and logistics network efficiency evaluation, data envelopment analysis (DEA) and its extended methods have been widely applied. However, traditional DEA treats decision-making units as “black boxes”, neglecting the internal inter-stage transmission mechanisms [6]. Network DEA, by disentangling the internal structure of multi-stage systems, can more accurately characterize the sources and transmission pathways of efficiency [7–11]. Nevertheless, confronted with the strongly coupled characteristics of demand planning, scheme design, organizational execution, and performance feedback in military equipment allocation, existing network models struggle to effectively decouple the inter-stage performance degradation and cascading effects. Recent fuzzy/interval Network DEA studies—fuzzy fair efficiency [1], fuzzy dynamic network DEA [2], interval type-2 fuzzy super-SBM network DEA [3], multi-period network DEA with feedback [8], and hybrid AHP–entropy dynamic network DEA [9]—have enriched uncertain efficiency evaluation, yet they still rely on per-DMU optimized weights or α -cut arithmetic, lacking interval-cloud cognitive boundaries, common-weight linking constraints with telescoping-consistent stage decoupling, and resampling-based ranking robustness.

Regarding the processing of uncertain information, the cloud model has attracted considerable attention for its ability to simultaneously characterize randomness and fuzziness, and has been applied to evaluation domains such as emergency medical rescue equipment, multi-criteria equipment evaluation, and photovoltaic fire risk [12–14], and weapon equipment combat indices and power-system security risk [15,16]; in equipment support capability assessment, it has been combined with combination weighting and capability-evaluation frameworks [17,18], with group decision-making, extension clouds and the VIKOR method [19–21], and with fuzzy-number-based TOPSIS and resampling-based robustness techniques [22,23] to construct various evaluation frameworks. However, existing equipment support assessments predominantly adopt “black-box” comprehensive measurement, failing to reveal the performance transmission mechanisms across the stages of the allocation chain. More importantly, how to deeply integrate the cognitive boundary characterization capability of the interval cloud model with the stage-decoupling advantage of network DEA so as to resolve the efficiency inversion caused by polarity misalignment of cost-type indicators remains a theoretical bottleneck awaiting breakthrough.

In multi-criteria decision-making and robustness verification for complex systems, methods such as fuzzy TOPSIS are frequently employed in shipbuilding supply chain and equipment support evaluation [24,25]. However, when handling expert cognitive uncertainty under small-sample conditions, existing methods largely rely on point estimates or simple intervals and lack statistical robustness verification based on resampling techniques, making it difficult to effectively simulate the cognitive drift and system coordination mechanisms under stochastic demand [26]. Such “false precision” point-estimate rankings often lack statistical robustness when confronting extreme disruption scenarios.

Building on the above analysis, Table 1 positions the proposed framework against existing cloud model–DEA studies along six dimensions: existing studies mostly handle uncertainty with point estimates or simple intervals under black-box structures and generally lack robustness verification, whereas this paper fuses interval cloud cognitive boundaries, linking constraints, and resampling verification within a serial network, forming a closed loop.

Table 1: Methodological feature comparison between existing cloud-DEA studies and the proposed framework.

Dimension	Existing Cloud-DEA Studies	Proposed Framework
Uncertainty representation	Point estimates/simple intervals, with cognitive boundaries lost	Interval cloud (Ex , En) cognitive consensus boundaries

(Continued)

Table 1 (continued)

Dimension	Existing Cloud-DEA Studies	Proposed Framework
Network structure	Black-box or single-stage	Four-stage series + intermediate linking constraints
Polarity handling	Cost-type indicators mostly left unprocessed	Inverse mapping of cost-type intervals, eliminating efficiency inversion
Dimensionality reduction mechanism	Median/mean collapse	Cloud-feature fused entropy weight + game-theoretic combination
Robustness verification	Usually absent	Bootstrap + scenario expansion drift experiment
Result form	Point ranking	Interval efficiency + stage decoupling + probabilistic ranking

To address the aforementioned issues, this paper proposes a performance evaluation method for equipment allocation supply management chains that fuses the cloud model with network DEA. The main contributions include: (1) constructing a four-stage serial network structural model to reveal the performance transmission mechanisms; (2) introducing the interval cloud model to aggregate expert cognition and constructing composite intervals through a stage-wise fused-entropy/game-theoretic dimensionality reduction that avoids the median collapse of cloud information; (3) establishing a polarity-corrected common-weight ISN-DEA with linking constraints and the telescoping property (Proposition 1) to achieve stage efficiency decoupling; and (4) employing Bootstrap resampling, multi-baseline comparison, and sensitivity experiments to verify the robustness of the ranking. The novelty thus lies in both these specific methodological improvements and their closed-loop integration, which guarantees end-to-end consistency from linguistic evaluation to probabilistic ranking and renders the stage-decoupled interval efficiencies mutually self-consistent and statistically verified, rather than a mere assembly of off-the-shelf components.

2 Construction of the Allocation Chain Task Process and Indicator System

Equipment allocation and supply support is not mere point-to-point material transfer but a complex dynamic network with multi-node synergy and multi-resource coupling; this section deconstructs the chain stage by stage and constructs a life-cycle indicator system.

2.1 Stage Deconstruction and Logical Relationships of the Allocation Support Chain

The equipment allocation chain is abstracted as a serial network of four core stages—demand planning (S_1), scheme design (S_2), organizational execution (S_3), and performance feedback (S_4)—tightly coupled through information and material flows (Fig. 1). S_1 , centered on “precise alignment”, parses requirements and outputs resource prepositioning plans, determining the initial resource occupation of the whole chain; S_2 , the hub under “global optimization”, generates route planning and resource grouping schemes with reserved buffers; S_3 , the core interface confronting disruptions, offsets degradation via real-time monitoring and local reallocation under “dynamic resilience”; S_4 consolidates execution results into standardized rules feeding back to S_1 , driving the transition from reactive response to proactive adaptation. The output of each stage serves as the input of the next, so any node’s loss or delay is amplified through cascading

effects; traditional “black-box” models cannot capture this transmission mechanism, necessitating a network evaluation framework.

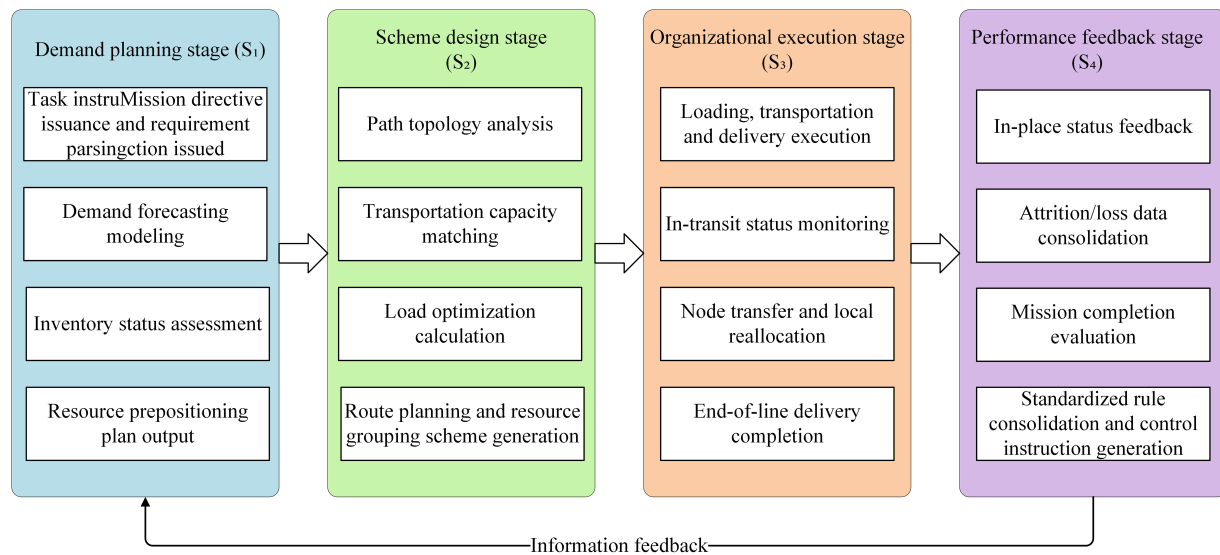


Figure 1: Topology diagram of the four-stage task flow for equipment allocation and supply support.

2.2 Design Logic and Dimensions of the Performance Evaluation Indicator System

To quantify the operational effectiveness of the aforementioned four-stage chain, this study constructs a multi-dimensional performance evaluation indicator system from a life-cycle perspective, grounded in the Supply Chain Operations Reference (SCOR) model and system resilience theory. The indicator selection adheres to the principles of systematicity, measurability, and operational orientation, aiming to comprehensively capture the topological characteristics and dynamic response capabilities across all four stages. The detailed definitions, codes, and selection rationales for these specific indicators are presented in Table 2.

Table 2: Performance evaluation indicator system and definitions for equipment allocation chains.

Operational Stage	Evaluation Indicators	Operational Implications and Selection Rationale
Demand Planning Stage (S_1)	Resource Demand Matching Degree SA_1	Measures the alignment between prepositioned resources and mission directives, reflecting the precision of source planning
	Initial Input Redundancy Rate SA_2	Measures the proportion of excess resources reserved for uncertainty, reflecting cost control capability
	Planning Time Compliance Rate SA_3	Evaluates the timeliness of demand parsing and plan issuance, reflecting the response speed of the command hub

(Continued)

Table 2 (continued)

Operational Stage	Evaluation Indicators	Operational Implications and Selection Rationale
Scheme Design Stage (S ₂)	Path Topology Rationality CC ₁	Evaluates the degree to which planned routes avoid adverse road conditions and risks, reflecting the optimization capability of the core hub
	Transportation Capacity Matching Degree CC ₂	Measures the alignment between allocated transportation capacity and actual demand, reflecting the scientific nature of resource scheduling
	Cross-domain Collaborative Response Time CC ₃	Measures the friction and response delay during multi-agent joint support, reflecting collaborative friction costs
Organizational Execution Stage (S ₃)	Command Transmission Efficiency RC ₁	Measures the losslessness and timeliness of cross-level and cross-departmental information transmission
	Dynamic Routing Switching Success Rate RC ₂	Evaluates the effectiveness of alternative route planning and execution when encountering sudden disruptions
	Resource Integrity Retention Rate RC ₃	Measures the proportion of equipment free from combat damage or environmental degradation during allocation, reflecting execution resilience
Performance Feedback Stage (S ₄)	Mission On-time Achievement Rate RE ₁	Core output indicator, measuring whether equipment delivery time meets the baseline requirement of mission timeliness
	Comprehensive Support Effectiveness Index RE ₂	Reflects the integrity rate and availability of the finally delivered equipment, embodying the ultimate quality of physical displacement
	Experience Reuse Conversion Rate RE ₃	Core feedback indicator, measuring the proportion of current mission data converted into standardized strategic rules

As shown in Table 2, the selection of the 12 indicators further adheres to three criteria: theoretical mapping (each indicator is derived individually from the SCOR model and system resilience theory, covering the full-chain characteristics), stage balance (three indicators are configured for each stage, matching the serial topology and ensuring sufficient information bandwidth for entropy-weight computation and DEA solving), and small-sample feasibility (under the conditions of 4 scenarios and 15 experts, the number of indicators is controlled at 12 to avoid over-parameterization). The expert screening process eliminated highly

redundant or difficult-to-measure candidate indicators, thereby ensuring the independence and measurability of the indicator system. This architecture covers the complete business chain from resource planning to performance feedback, providing structured data input for the subsequent interval cloud aggregation and network DEA decoupling; the polarity delineation and quantification rules are systematically elaborated in [Section 3](#).

3 Construction of the Evaluation Model and Computational Methodology

The equipment allocation chain exhibits pronounced stage-wise transmission characteristics and uncertainty coupling under multiple disruptions. To scientifically quantify the resilience performance of the chain across different scenarios, this paper constructs a quantitative evaluation framework that integrates expert linguistic evaluation, interval cloud aggregation, entropy-weight dimensionality reduction, network DEA decoupling, and possibility-degree ranking. This framework strictly follows the operational flow, ensuring absolute alignment between mathematical computation and engineering semantics.

3.1 Evaluation Framework and Indicator Polarity Delineation

Guided by the operational flow characteristics of the allocation chain, this paper deconstructs the chain into four logical stages: S_1 (Demand Planning) focuses on initial resource matching and constraint control; S_2 (Scheme Design) emphasizes route planning and collaborative mechanisms; S_3 (Organizational Execution) characterizes delivery integrity and dynamic disturbance resistance; and S_4 (Performance Feedback) reflects system recovery and value consolidation. To support the subsequent interval network DEA evaluation, 12 secondary indicators are selected, and their polarity attributes and stage affiliations are strictly delineated. The specific mapping relationships are presented in [Table 3](#).

Table 3: Polarity delineation and attribute mapping of evaluation indicators for allocation chains.

Indicator	Polarity Attribute	Treatment in DEA Model	Indicator	Polarity Attribute	DEA Model Processing Method
Resource Demand Matching Degree SA_1	Benefit-type	Direct output/Positive mapping	Command Transmission Efficiency RC_1	Benefit-type	Intermediate variable
Initial Input Redundancy Rate SA_2	Cost-type	Inverse mapping (High evaluation indicates low input)	Dynamic Routing Switching Success Rate RC_2	Benefit-type	Intermediate variable
Planning Time Compliance Rate SA_3	Benefit-type	Direct output/Positive mapping	Resource Integrity Retention Rate RC_3	Benefit-type	Intermediate variable
Path Topology Rationality CC_1	Benefit-type	Intermediate variable	Mission On-time Achievement Rate RE_1	Benefit-type	Direct final output

(Continued)

Table 3 (continued)

Indicator	Polarity Attribute	Treatment in DEA Model	Indicator	Polarity Attribute	DEA Model Processing Method
Transportation Capacity Matching Degree CC_2	Benefit-type	Intermediate variable	Comprehensive Support Effectiveness Index RE_2	Benefit-type	Direct final output
Cross-domain Collaborative Response Time CC_3	Cost-type	Inverse mapping	Experience Reuse Conversion Rate RE_3	Benefit-type	Direct final output

Upon completing the stage-wise deconstruction of the allocation chain, the screening of evaluation indicators, and the polarity normalization, a complete quantitative evaluation algorithm is further constructed, integrating interval cloud cognitive aggregation, game-theoretic combined entropy-weight dimensionality reduction, interval serial network DEA decoupling, complete ranking, and robustness verification. The overall algorithm adopts a serial computation mode in which the output of each preceding module serves as the input to the subsequent one, sequentially performing uncertainty quantification, stage-wise indicator dimensionality reduction, stage-wise efficiency solving, scenario ranking, and statistical verification, and ultimately yielding evaluation conclusions such as the scenario performance ranking and the identification of chain bottlenecks. The complete computational flow is illustrated in Fig. 2.

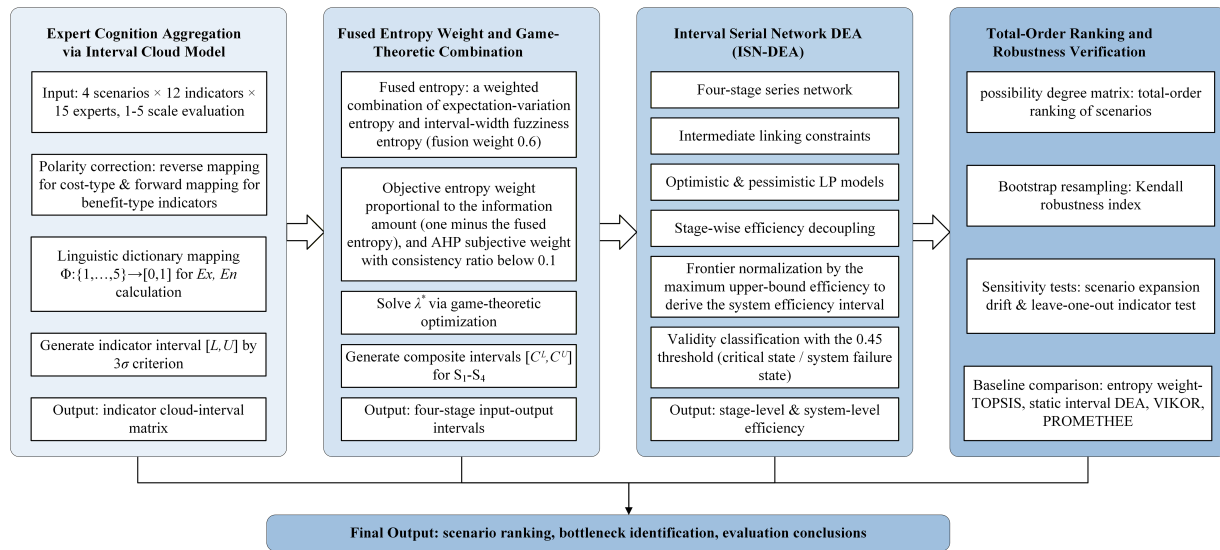


Figure 2: Overall flowchart of the equipment allocation chain performance evaluation algorithm based on the interval cloud model and network DEA.

Experts independently score the 12 indicators of each scenario on a 5-level linguistic scale (1 = extremely poor to 5 = extremely excellent). The scale semantics and polarity mapping rules are detailed in Supplementary Table S1: benefit-type indicators are directly mapped to high output/high capability, whereas cost-type

indicators (SA2, CC3) undergo inverse mapping, converting high evaluations into low physical inputs, thereby precluding, at the source, the efficiency inversion caused by polarity confusion.

3.2 Expert Cognitive Aggregation Based on the Interval Cloud Model

As stated in [Section 1](#), expert linguistic evaluations simultaneously embody fuzziness and randomness. This section presents the specific aggregation mechanism by which the interval cloud model achieves a robust mapping from qualitative semantics to quantitative intervals. Let the expert panel consist of K experts, and let the score given by the k -th expert to indicator m under scenario i be denoted as s_{imk} , which is mapped to a baseline cloud drop $\Phi(s_{imk})$ via the preset dictionary $\Phi: \{1, 2, 3, 4, 5\} \rightarrow [0, 1]$.

The expectation of group cognition Ex_{im} serves as an unbiased estimate of the central tendency:

$$Ex_{im} = \frac{1}{K} \sum_{k=1}^K \Phi(s_{imk}). \quad (1)$$

The entropy of group cognition En_{im} adopts the unbiased sample standard deviation estimate with $K - 1$ degrees of freedom, characterizing the dispersion of expert opinions:

$$En_{im} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\Phi(s_{imk}) - Ex_{im})^2}. \quad (2)$$

The hyper-entropy He , as a second-order measure, is retained as a small positive constant and does not participate in the interval construction. Based on the 3σ criterion, the cognitive interval is constructed and truncated to the physically feasible domain $[0, 1]$:

$$L_{im} = \max(0, Ex_{im} - 3En_{im}), \quad U_{im} = \min(1, Ex_{im} + 3En_{im}). \quad (3)$$

Should the extreme dispersion case $L_{im} > U_{im}$ occur, $L_{im} = U_{im} = Ex_{im}$ is set to guarantee interval validity. This aggregation enables the interval $[L_{im}, U_{im}]$ to simultaneously carry central and dispersion information, avoiding the collapse of linguistic evaluations into a single numerical value; the 3σ truncation naturally suppresses anomalous evaluations, providing structured interval input for [Section 3.3](#).

To facilitate understanding, we first distinguish three types of uncertainty: randomness reflects the fluctuation of the same expert's judgments at different points in time; fuzziness arises from the unclear semantic boundaries of the linguistic scale; and expert divergence refers to the dispersion of different experts' cognitive centers, which is statistically characterized by En . In the interval cloud, Ex captures the consensus center, En absorbs randomness and fuzziness, and the 3σ interval $[L, U]$ constitutes the consensus boundary "acceptable to the vast majority of experts". To illustrate, SA₁ of Scenario A receives unanimous "4" ratings (mapped to 0.8) from all 15 experts, yielding $Ex = 0.800$, $En = 0.000$, and $[L, U] = [0.800, 0.800]$ —the interval degenerates to a single point, reflecting complete consensus; by contrast, RE₂ of Scenario D receives eleven "1" ratings (0.2) and four "2" ratings (0.4), yielding $Ex = 0.253$, $En = 0.092$, and $[L, U] = [0.000, 0.528]$ —a low consensus center with a wide cognitive boundary. The comparison of the two examples demonstrates that the interval cloud can adaptively characterize the continuous spectrum from "complete consensus" to "significant divergence".

3.3 Stage-Wise Dimensionality Reduction and Composite Interval Construction Based on Cloud-Feature Fused Entropy Weights

The indicators at each stage differ in dimension and semantics, and must be reduced to scalar composite intervals before being input into the network DEA. The traditional approach of taking the interval median to compute entropy weights discards the divergence information preserved by the cloud model, causing a break in the “aggregation–dimensionality reduction” pipeline. To this end, this paper constructs a fused entropy from the two dimensions of expectation variation and fuzziness to comprehensively assess the information content of the indicators. Let stage p contain the indicator set M_p , and let the number of scenarios be $n = 4$.

The normalized probability and information entropy based on expectation measure the objective variation of the indicators:

$$p_{im}^{Ex} = \frac{Ex_{im}}{\sum_i Ex_{im} + \varepsilon}, H_m^{Ex} = -\frac{1}{\ln n} \sum_i p_{im}^{Ex} \ln p_{im}^{Ex}. \quad (4)$$

The fuzziness entropy based on the interval width $W_{im} = U_{im} - L_{im}$ (a direct measure of fuzziness) is:

$$p_{im}^{En} = \frac{W_{im}}{\sum_i W_{im} + \varepsilon}, H_m^{En} = -\frac{1}{\ln n} \sum_i p_{im}^{En} \ln p_{im}^{En}. \quad (5)$$

The comprehensive information entropy H_m is obtained by weighted fusion of the two (the adjustment coefficient α is taken as 0.6 to balance objective data variation and cognitive fuzziness)

$$H_m = \alpha H_m^{Ex} + (1 - \alpha) H_m^{En}. \quad (6)$$

Given that the pure objective entropy weight is sensitive to sample perturbation when the number of scenarios is small, and that variation-only weighting may ignore business importance, an AHP subjective weight w_m^{sub} passing the consistency test ($CR < 0.1$) is introduced, and the combination coefficient λ^* is optimized via game theory. The judgment matrices are independently provided by all 15 experts of the evaluation panel and aggregated via the geometric mean; the matrices and CR values are detailed in Supplementary Table S2. Let $w^{(1)} = w^{obj}$, $w^{(2)} = w^{sub}$; the game-equilibrium combination coefficients (α_1, α_2) satisfy the first-order condition linear system:

$$\begin{bmatrix} w^{(1)} w^{(1)} & w^{(1)} w^{(2)} \\ w^{(2)} w^{(1)} & w^{(2)} w^{(2)} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} w^{(1)} w^{(1)} \\ w^{(2)} w^{(2)} \end{bmatrix}. \quad (7)$$

Solving and normalizing yields $\lambda^* = \alpha_1 / (\alpha_1 + \alpha_2)$. The final combined weight $w_m^{(p)}$ and the stage composite intervals are:

$$w_m^{(p)} = \lambda^* w_m^{obj} + (1 - \lambda^*) w_m^{sub}, \quad (8)$$

$$C_{ip}^L = \sum_{m \in p} w_m^{(p)} L_{im}, C_{ip}^U = \sum_{m \in p} w_m^{(p)} U_{im}, \quad (9)$$

where the objective entropy weight is proportional to the information amount $(1 - H_m)$, i.e., $w_m^{obj} = (1 - H_m) / \sum_{m \in p} (1 - H_m)$. This mechanism genuinely incorporates fuzziness into the weight computation and, through the subjective–objective combination, suppresses small-sample weight fluctuations, yielding perturbation-resistant structured interval inputs.

3.4 Interval Common-Weight Series Network DEA Efficiency Decoupling Model

As stated in the introduction, the “black-box” assumption of traditional DEA cannot reveal the sources and paths of the efficiency degradation within the chain. To this end, this paper constructs a series network DEA linear program with intermediate linking constraints, solves for the globally unique common weights under the optimistic and pessimistic bounds, and decomposes the system efficiency stage by stage into the stage efficiencies of S_1 – S_4 .

Network topology and linking constraints. Let the composite interval of stage k in scenario j be $C_{jk} = [CL_{jk}, CU_{jk}]$ ($k = 1, \dots, 4$). In the series structure, the output of stage k is the input of stage $k + 1$; let the linking variable be $Z_{jk} = C_{jk}$, and let the input of stage 1 be the exogenous normalized endowment $C_{j0} = 1$. Let u_k be the common weight of the output of stage k ; from the linking constraints, $v_{k+1} = u_k$, $v_1 = 1$.

Joint linear program. This paper adopts the common-weight paradigm: the same set of stage weights $\{u_k\}_{k=1}^4$ is imposed on all $J = 4$ scenarios, and the optimal weights that keep every stage efficiency of every scenario no greater than 1 are solved in one shot. The optimistic (upper-bound) model is:

$$\max_u \quad u_4, \quad (10a)$$

$$\text{s.t.} \quad u_k CU_{j,k} \leq u_{k-1} CU_{j,k-1}, \quad \forall j = 1, \dots, J, k = 1, \dots, 4, \quad (10b)$$

$$u_k \geq \varepsilon, k = 1, \dots, 4, \quad (10c)$$

where $\varepsilon = 10^{-6}$, and by convention $C_{j,0} \equiv 1$, $u_0 \equiv 1$ (so the $k = 1$ case of Eq. (10b) reads $u_1 CU_{j,1} \leq 1$). This stage-1 constraint bounds $u_1 \leq 1/\max_j CU_{j,1}$; chaining Eq. (10b) over $k = 2, 3, 4$ then gives $u_4 CU_{j,4} \leq u_1 CU_{j,1} \leq 1$ for all j , i.e., $u_4 \leq 1/\max_j CU_{j,4}$, so the maximization of u_4 is bounded and every stage efficiency in Eq. (11) satisfies $E_{jk} \leq 1$. The pessimistic model replaces CU by CL with $v_0 \equiv 1$.

Stage-0 input and weight normalization. The input of stage 1 is the exogenous normalized endowment $C_{j,0} \equiv 1$; the corresponding weights are defined as $u_0 \equiv 1$ (optimistic model) and $v_0 \equiv 1$ (pessimistic model), i.e., the denominator of the stage-1 efficiency degenerates to the constant 1, ensuring that Eq. (11) is meaningful for $k = 1$.

Stage efficiency decoupling. The optimistic and pessimistic stage efficiencies are defined as:

$$E_{jk}^U = \frac{u_k^* CU_{jk}}{u_{k-1}^* CU_{j,k-1}}, \quad E_{jk}^L = \frac{v_k^* CL_{jk}}{v_{k-1}^* CL_{j,k-1}}, \quad k = 1, \dots, 4, \quad (11)$$

where for $k = 1$ the denominator is $u_0 C_{j,0} = 1$ (optimistic) or $v_0 C_{j,0} = 1$ (pessimistic).

Proposition 1 (telescoping property): The product of the four stage efficiencies equals the system efficiency:

$$\prod_{k=1}^4 E_{j,k}^U = u_4^* \cdot CU_{j,4} \equiv \rho_j^U, \quad \prod_{k=1}^4 E_{j,k}^L = v_4^* \cdot CL_{j,4} \equiv \rho_j^L. \quad (12)$$

Proof (taking the optimistic model as an example): Expanding Eq. (11) stage by stage, the intermediate terms cancel pairwise, leaving only $\frac{u_4^* CU_{j,4}}{u_0^* C_{j,0}} = \frac{u_4^* CU_{j,4}}{1} = u_4^* CU_{j,4} = \rho_j^U$; the pessimistic model follows analogously. $\square$

Normalizing by $f = \max_j \tilde{\rho}_j^U$ yields the system efficiency interval:

$$\rho_j^U = \tilde{\rho}_j^U / f, \quad \rho_j^L = \tilde{\rho}_j^L / f. \quad (13)$$

Note 1. Proposition 1 shows that, once the optimal common weights (u^*, v^*) are obtained by numerically solving the linking-constrained series network DEA linear program (10a)–(10c), the end-to-end interval efficiency $[\rho_j^L, \rho_j^U]$ can be exactly recovered as the product of the stage efficiencies (telescoping property); that is, the stage decoupling is strictly self-consistent with the system efficiency.

Note 2 (structural extensibility). For a series structure, each additional stage only requires appending the corresponding intermediate linking constraint, and the telescoping property of Proposition 1 still holds; for parallel branches, introducing branch-level linking variables and flow-conservation constraints yields a general network DEA without changing the solution architecture.

Note 3 (computational complexity). Cloud aggregation is $O(KJM)$; fused entropy weighting and combination weighting are $O(JM) + O(1)$ closed-form; Eqs. (10a)–(10c) have only 4 decision variables (16 constraints for $J = 4$, single solve < 1 ms with the HiGHS solver), and Bootstrap with $B = 1000$ takes about 10 s in total; scaling up only increases the order of the constraints, so the complexity remains controllable, and the main manual cost lies in constructing the AHP judgment matrices.

Envelope and bottleneck identification. Since under the two sets of common weights the relation $E_{jk}^L \leq E_{jk}^U$ does not necessarily hold stage by stage, the envelope is taken to ensure interval validity:

$$\left[\min(E_{jk}^L, E_{jk}^U), \max(E_{jk}^L, E_{jk}^U) \right], \quad (14)$$

and, with the median adopted as the point estimate, the bottleneck stage is defined as

$$k_j^* = \operatorname{argmin}_k \frac{1}{2} (E_{jk}^L + E_{jk}^U). \quad (15)$$

This definition enables the weakest stage of each scenario to be uniquely identified, providing stage-level quantitative evidence for the subsequent mechanism analysis.

3.5 Complete Ranking Based on Possibility Degree Matrix and Bootstrap Robustness Verification

The output of the interval serial network DEA is a set of efficiency intervals, which in itself does not constitute a complete ordering and cannot be directly used for strategy selection; meanwhile, the ranking conclusions obtained under a small-sample expert panel also require statistical robustness verification. To this end, this section first constructs a complete-ranking method based on the possibility degree matrix, and then introduces Bootstrap nonparametric resampling to statistically test the ranking conclusions.

3.5.1 Interval Possibility Degree-Based Complete Ranking

To meet the complete-ranking requirement for interval efficiencies, this paper constructs a superiority measure based on the possibility degree matrix. For any two scenario efficiency intervals $\tilde{\rho}_i, \tilde{\rho}_j$, the possibility degree $P(\tilde{\rho}_i \geq \tilde{\rho}_j)$ is defined as a relative superiority measure over the interval overlap region:

$$P_{ij} = \frac{\max(0, U_i - L_j) - \max(0, L_i - U_j)}{(U_i - L_i) + (U_j - L_j)}. \quad (16)$$

When the denominator is less than 10^{-9} , $P_{ij} = 0.5$ is set. This measure satisfies reflexivity ($P_{ii} = 0.5$) and complementarity ($P_{ij} + P_{ji} = 1$), guaranteeing the logical consistency of pairwise comparisons. The possibility degree matrix $\mathbf{P} \in \mathbb{R}^{4 \times 4}$ is then constructed, and the row sums $R_i = \sum_j P_{ij}$ are computed. The larger R_i is, the

higher the probability that scenario i is relatively superior to the other scenarios; ranking in descending order of R_i yields the baseline comprehensive ranking $\mathbf{r}^{(0)}$.

3.5.2 Bootstrap Resampling-Based Robustness Verification

For a small-sample expert panel ($K \leq 15$), the traditional leave-one-out method is prone to the statistical artifact of an invariant ranking and cannot objectively reflect the impact of cognitive perturbation on the ranking. This paper therefore adopts Bootstrap nonparametric resampling for robustness verification: K experts are drawn with replacement from the original scoring matrix to form a resampling set $\mathbf{D}^{(b)}$, and this is repeated B times, each time independently recomputing the full pipeline to obtain a resampled ranking vector $\mathbf{r}^{(b)}$. The Kendall concordance coefficient is used to quantify the ordinal consistency between the baseline ranking $\mathbf{r}^{(0)}$ and the resampled ranking:

$$W^{(b)} = \frac{1 + \tau(\mathbf{r}^{(0)}, \mathbf{r}^{(b)})}{2}, \quad (17)$$

$$\bar{W} = \frac{1}{B} \sum_{b=1}^B W^{(b)}, \quad (18)$$

where $W^{(b)}$ is the robustness index of the b -th resampling, $\tau(\cdot, \cdot)$ is the Kendall rank correlation coefficient, and $\bar{W}$ is the average robustness index. The closer $\bar{W}$ is to 1, the more stable the ranking conclusion is against expert cognitive perturbation. Meanwhile, the 2.5%–97.5% empirical quantiles of each scenario's ranking are extracted as the 95% confidence interval ($CI_{95\%}$) to quantify the uncertainty boundary of the ranking. By simulating the drift of the cognitive centroid after resampling, this mechanism provides statistical evidence on whether the ranking conclusion is robust, avoiding the “false precision” of point-estimate rankings.

4 Empirical Analysis and Discussion of Results

4.1 Data Sources and Processing of the Expert Evaluation Matrix

To systematically validate the engineering applicability and methodological robustness of the proposed evaluation framework, this paper takes a typical equipment support system as the empirical case and convenes a 15-expert evaluation panel with backgrounds spanning equipment command, operations research optimization, and systems engineering (six frontline commanders, five academic scholars, and four engineering technicians). Using a predefined 5-level linguistic scale (1 = extremely poor to 5 = extremely excellent), the experts independently assess four prototypical allocation disruption scenarios: Scenario A (Routine) encodes normalized support missions under standard operating procedures; Scenario B (Emergency) encodes rapid-response and resource-reallocation missions under sudden disruptions; Scenario C (Collaborative) encodes complex coordination missions entailing multi-departmental, cross-hierarchical linkage; and Scenario D (Mixed) encodes high-uncertainty missions subject to superimposed multiple disruptions. Together, the four scenarios span the complete performance-evolution spectrum of the chain under routine operation and extreme disruption alike, forming an orthogonal perturbation spectrum of “baseline—single-dimensional extreme—organizational complexity—multi-dimensional coupling” that provides the comparative benchmark for the subsequent differentiated mechanism analysis.

A caveat is in order: military equipment allocation data are highly classified, and large-sample objective records of extreme disruption scenarios are unobtainable in peacetime. This paper therefore adopts an expert-judgment-based scenario simulation approach, and introduces Bootstrap resampling together with interval cloud consensus boundaries, so that the robustness of the conclusions under cognitive

perturbation is guaranteed by statistical verification rather than by assumption. The raw judgments constitute a three-dimensional evaluation matrix $\mathbf{S} = [s_{ijk}]_{K \times J \times M}$ ($K = 15$ experts, $J = 4$ scenarios, $M = 12$ indicators); after consistency verification and outlier screening, the effective retention rate reaches 100%. Pursuant to Eqs. (1)–(3), the discrete ratings are mapped into the continuous domain and aggregated into interval cloud parameters; representative aggregates include SA_1 of Scenario A ($Ex = 0.800$, $En = 0.000$, $[L, U] = [0.800, 0.800]$), CC_2 of Scenario A ($Ex = 0.787$, $En = 0.052$, $[L, U] = [0.632, 0.942]$), CC_2 of Scenario C ($Ex = 0.827$, $En = 0.070$, $[L, U] = [0.616, 1.000]$), and RE_2 of Scenario D ($Ex = 0.253$, $En = 0.092$, $[L, U] = [0.000, 0.528]$). The complete-consensus form of SA_1 (Scenario A), whose interval degenerates to a single point, corroborates the stability of the routine process, whereas the low-expectation, high-divergence form of RE_2 (Scenario D) exposes the pronounced uncontrollability of delivery quality under mixed disruption. The juxtaposition of these parameter sets quantifies the cognitive evolution law of “high consensus–low dispersion under normality” vs. “low consensus–high divergence under perturbation”.

For reproducibility, the principal parameter settings are summarized as follows: linguistic mapping $\Phi: \{1, 2, 3, 4, 5\} \rightarrow \{0.2, 0.4, 0.6, 0.8, 1.0\}$; 3σ truncation to $[0, 1]$ (Eq. (3)); fused-entropy adjustment coefficient $\alpha = 0.6$ (Eq. (6)), balancing objective variation and cognitive fuzziness; entropy protection term $\varepsilon = 10^{-9}$ (Eqs. (4) and (5)); LP non-Archimedean bound $\varepsilon = 10^{-6}$ (Eq. (10c)); empirical threshold $\theta = 0.45$, with sensitivity verified over $[0.40, 0.55]$ (Section 4.3.1); game-theoretic combination coefficients λ^* solved per stage from Eq. (7) (1.000, 1.000, 0.690, 0.758 for S_1 – S_4 ; Supplementary Table S2); Bootstrap resampling $B = 1000$; scenario-expansion drift experiment $M = 200$.

4.2 Stage Entropy Weights and Composite Interval Analysis

Applying Eqs. (4)–(9), the stage-wise indicator weights are computed and the composite intervals are generated, as reported in Table 4.

Table 4: Stage entropy weights and composite interval results (partial).

Scenario	Stage	L	U	Weights
Scenario A	S_1 Demand Planning	0.756	0.788	[0.324, 0.064, 0.611]
Scenario B		0.588	0.615	[0.324, 0.064, 0.611]
Scenario C		0.681	0.851	[0.324, 0.064, 0.611]
Scenario D		0.362	0.483	[0.324, 0.064, 0.611]
Scenario A	S_4 Performance Feedback	0.683	0.859	[0.143, 0.150, 0.707]
Scenario B		0.512	0.644	[0.143, 0.150, 0.707]
Scenario C		0.742	0.850	[0.143, 0.150, 0.707]
Scenario D		0.285	0.451	[0.143, 0.150, 0.707]

Beyond the numerical entries of Table 4, the fused entropy weights reveal the core driving dimensions underlying performance differentiation across stages. The S_1 weights concentrate heavily on planning time achievement rate SA_3 (0.611), those of S_2 on path topology rationality CC_1 (0.633), and those of S_4 on experience reuse conversion rate RE_3 (0.707), indicating that, within the variation structure of the present sample, timeliness control, path optimization, and experience consolidation constitute the most decisive dimensions for discriminating scenario quality; by contrast, the S_3 weights are comparatively balanced ([0.384, 0.322, 0.294]), reflecting that organizational execution relies on multi-dimensional coordination. The game-theoretic combination weighting (Supplementary Table S2) effectively injects subjective priors into S_3 and S_4 ($\lambda^* = 0.690$ and 0.758 , respectively), smoothing the extreme concentration of the pure objective

entropy weight at S_4 (where the objective weight of RE_3 reaches 0.852), so that the weighting system balances data-driven variation with operational importance.

The composite interval distributions of the four scenarios, as listed in Table 4, show that, in terms of input-control precision, the lower bound of S_1 for Scenario D is the lowest (0.362), exposing the loss-of-control risk of upstream planning under mixed perturbations; the S_1 interval of Scenario A is the narrowest (width 0.032), evidencing the exceptionally high consensus and stability of resource pre-positioning under the routine process. Regarding output transformation, the S_4 upper bound of Scenario C (0.850) is close to that of A (0.859), whereas that of D reaches only 0.451, revealing the severe blockage that multi-source perturbations impose on the final performance feedback. A further dissection of the interval widths shows that the S_1 widths follow the order $C > D > A > B$: Scenario B exhibits the smallest width yet a comparatively high upper bound—a consensus essentially attained through passive compromise; Scenario C exhibits the largest width, corroborating the principle that “synergy gains come with friction costs”. Such interval-based characterization supplies structured inputs for the subsequent network DEA model and effectively prevents the smoothing-over of expert disagreement inherent in traditional point estimates.

4.3 Interval Network DEA Efficiency Decoupling and Complete Ranking

4.3.1 Interval Network DEA Efficiency Decoupling Results

Applying Eqs. (10)–(15), the efficiency interval of each scenario is computed and normalized, with the results reported in Table 5. The distributions of the normalized efficiency interval bounds and medians are depicted in Fig. 3, in which horizontal gray segments intuitively represent the span of the efficiency interval $[\rho^L, \rho^U]$, colored diamonds mark the efficiency median $\tilde{\rho}^{mid}$, and a vertical dashed line is introduced as the empirical threshold (0.45) for discriminating the “critical state” from the “system failure state”.

Determination rules. Let the empirical threshold be $\theta = 0.45$: if the median efficiency $\geq \theta$ and the lower bound $\geq \theta$, the scenario is determined as being in the “effective operation state”; if the median efficiency $\geq \theta$ but the lower bound $< \theta$, it is determined as being in the “critical state” (the mean is acceptable, yet a worst-case failure risk exists); if the median efficiency $< \theta$, it is determined as being in the “system failure state”. The threshold sensitivity test shows that the comprehensive ranking $C > A > B > D$ remains invariant over the entire interval $[0.40, 0.55]$; the state labels are stable within $[0.40, 0.45]$ (C/A/B critical, D failed), and only when the threshold is raised to 0.50 does B shift from critical to failed, indicating that the threshold mainly affects the discrete labels rather than the continuous ranking.

Table 5: Interval network DEA efficiency decoupling results.

Scenario	Efficiency Lower Bound ρ^L	Efficiency Upper Bound ρ^U	Median Efficiency	Status Classification
Scenario A	0.249	1.000	0.624	Critical state
Scenario B	0.186	0.750	0.468	Critical state
Scenario C	0.270	0.990	0.630	Critical state
Scenario D	0.104	0.526	0.315	System failure state

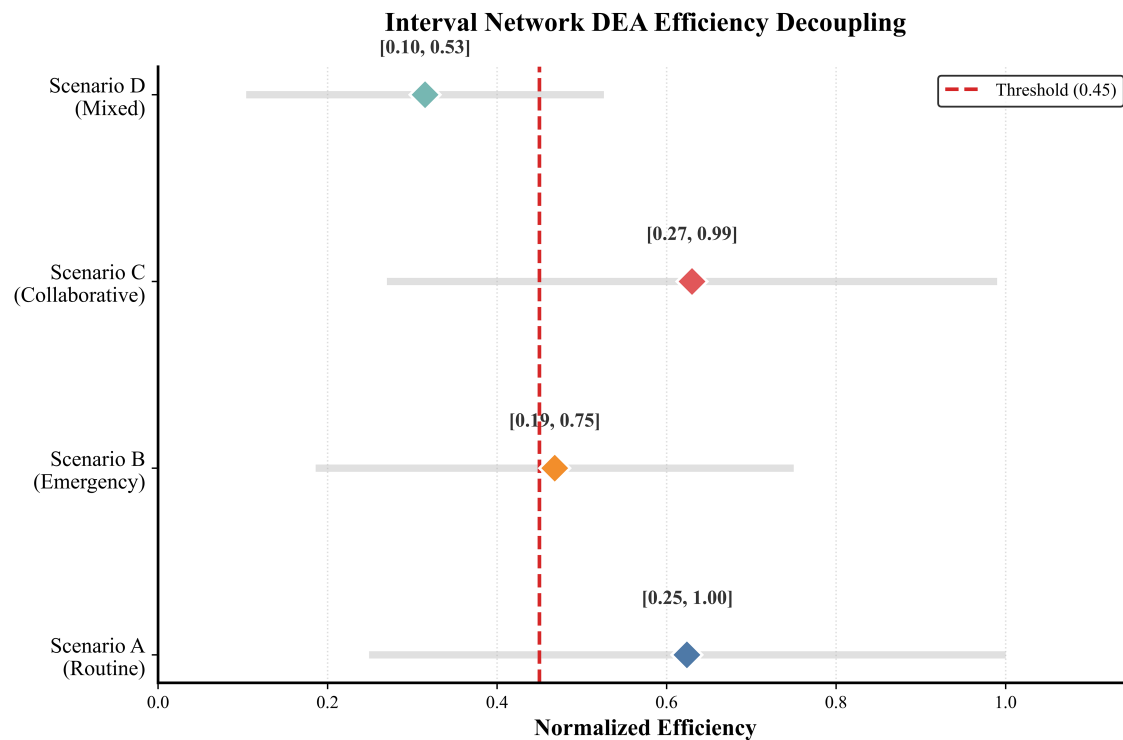


Figure 3: Interval network DEA efficiency decoupling results.

Combined with the geometric morphology of Fig. 3, the efficiency differentiation mechanisms of the four scenarios can be clearly identified. The diamond medians of the collaborative scenario (C) and the routine scenario (A) (0.630 and 0.624) are extremely close, and both straddle the 0.45 threshold; however, their efficiency lower bounds (0.270 for C and 0.249 for A) both fall below the threshold, exhibiting the typical “acceptable mean, worrisome worst-case” profile, and hence both are determined as being in the “critical state”. This indicates that, although the collaborative mechanism and the routine process can sustain satisfactory efficiency under most cognitive realizations, both carry failure risk under extreme cognitive perturbation. The median of the emergency scenario (B) (0.468) lies slightly above the threshold and likewise belongs to the “critical state”. The median (0.315) and upper bound (0.526) of the mixed scenario (D) are both markedly depressed, and it is determined as being in the “system failure state”, revealing the structural blockage that multi-source concurrent perturbations impose on the transformation channel.

The stage-wise efficiency decoupling results are reported in Table 6 (full four-stage results in Supplementary Table S3). Jointly analyzing these results identifies the weak link of each scenario: the bottlenecks of Scenarios A, B, and C all lie at S_4 (median efficiencies of 0.629, 0.597, and 0.615, respectively), indicating that the performance feedback stage is the common weak link constraining their overall efficiency; in contrast, the bottleneck of Scenario D lies at S_3 (median efficiency of only 0.422), revealing that the organizational execution stage suffers severe degradation under mixed perturbation.

Table 6: Stage-wise efficiency decoupling results (bottleneck stages S_3 – S_4).

Scenario	Stage	Lower Bound	Upper Bound	Median	Scenario	Stage	Lower Bound	Upper Bound	Median
Scenario A	S_3	0.911	1.000	0.955	Scenario C	S_3	0.915	1.000	0.957
	S_4	0.294	0.965	0.629		S_4	0.229	1.000	0.615
Scenario B	S_3	0.591	0.967	0.779	Scenario D	S_3	0.197	0.647	0.422
	S_4	0.236	0.958	0.597		S_4	0.831	1.000	0.916

4.3.2 Possibility Degree Matrix and Complete Ranking

Based on Eq. (16), the possibility degree matrix is constructed and its row sums are computed, with the results reported in Table 7. Ranking in descending order of the row sums yields the comprehensive ordering: Scenario C > Scenario A > Scenario B > Scenario D.

Table 7: Possibility degree matrix and complete ranking results.

	Scenario A	Scenario B	Scenario C	Scenario D	Row Sum R_i	Overall Ranking
Scenario A	0.5	0.619	0.496	0.764	2.378	2
Scenario B	0.381	0.5	0.374	0.656	1.911	3
Scenario C	0.504	0.626	0.5	0.776	2.406	1
Scenario D	0.236	0.344	0.224	0.5	1.304	4

Note: P_{ij} denotes the possibility degree that Scenario i is superior to Scenario j .

Table 7 reports the pairwise superiority probabilities (the probability that the row scenario outperforms the column scenario, with the diagonal fixed at 0.5). The possibility degree of C over A is only 0.504—nearly equivalent to a random coin toss—placing the two in a statistically indistinguishable “near-tie” state that reflects a delicate equilibrium between “synergy gain” and “process standardization” under the current indicator system; meanwhile, the first tier’s advantage probabilities over the last-placed Scenario D all exceed 0.7 ($P_{CD} = 0.776$, $P_{AD} = 0.764$), confirming the ranking’s statistical significance. Given the high interval overlap and the small row-sum gap between C and A, the Bootstrap further shows a swap probability of up to 24.4% (confidence intervals both [1.0, 2.0]); such near-tie cases should therefore be interpreted as “statistically indistinguishable, with only a marginal advantage” rather than a strict superiority, and decision-makers should treat the two as the same tier and reserve elastic space in light of the confidence intervals. The value of the framework lies precisely in quantifying this indistinguishability probabilistically—providing decision boundaries rather than fabricating false precision.

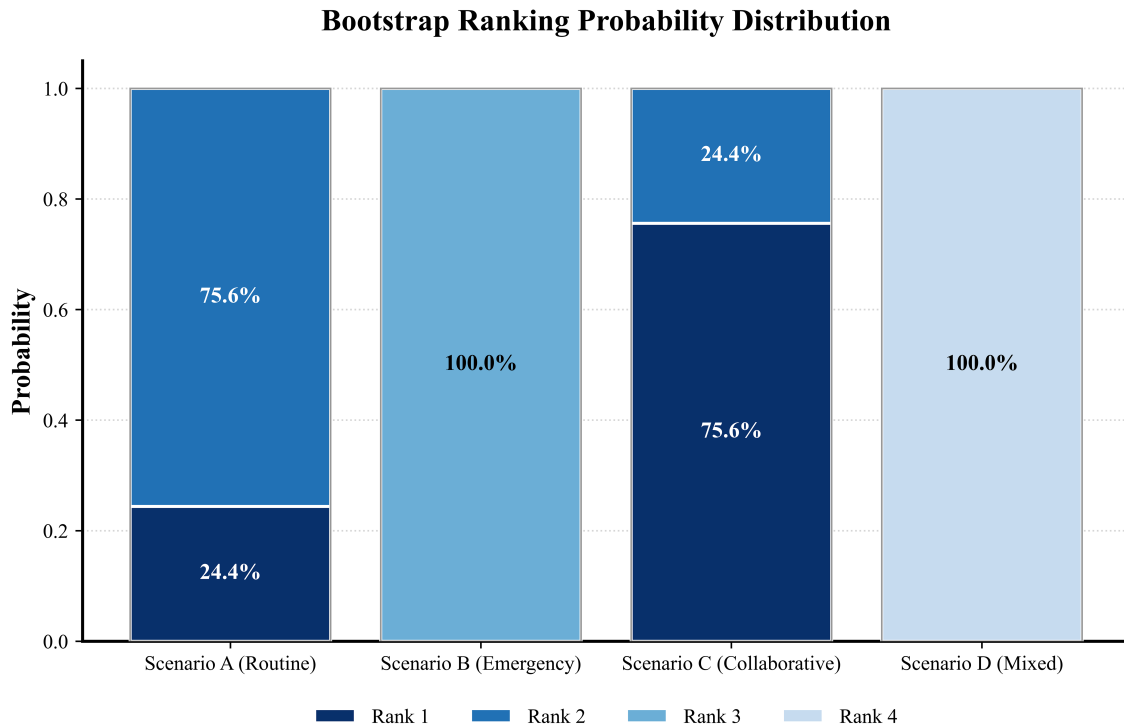
4.4 Bootstrap Robustness Verification

To rigorously verify the statistical robustness of the obtained ranking (C > A > B > D), a nonparametric Bootstrap resampling procedure ($B = 1000$) was executed. Each iteration independently recomputes the full pipeline of “cloud aggregation—entropy-weight dimensionality reduction—interval DEA—possibility degree ranking” and computes the rank consistency according to Eqs. (17) and (18). The ranking probability distributions and 95% confidence intervals of the four scenarios are reported in Table 8.

Table 8: Bootstrap ranking probability distribution and confidence intervals.

Scenario	Baseline Ranking	P (Rank = 1)	P (Rank = 2)	P (Rank = 3)	P (Rank = 4)	CI _{95%}
Scenario A	2	24.4%	75.6%	0	0	[1.0, 2.0]
Scenario B	3	0	0	100%	0	[3.0, 3.0]
Scenario C	1	75.6%	24.4%	0	0	[1.0, 2.0]
Scenario D	4	0	0	0	100%	[4.0, 4.0]

Fig. 4 visualizes the distribution morphology of the ranking probability mass as a percentage stacked bar chart.

**Figure 4:** Scenario ranking probability distributions based on Bootstrap resampling.

In conjunction with Fig. 4, two complementary convergence features can be identified. The first is controlled alternation at the head: C dominates rank 1 with a probability of 75.6% and A dominates rank 2 with a probability of 75.6%, with a swap probability between the two of up to 24.4% (confidence intervals both [1.0, 2.0]), which closes the loop perfectly with the “near-tie” conclusion of $P_{CA} = 0.504$ in Section 4.3.2; decision-makers should therefore treat them as the same first tier. The second is strong consensus at the bottom: B and D are locked at ranks 3 and 4, respectively, with 100% probability. After 1000 iterations, the average robustness index $\bar{W} = 0.960$, at the upper bound; the deficit below 1.0 is precisely contributed by the genuine 24.4% competitive uncertainty at the head, yielding a non-degenerate robustness profile of “strictly locked bottom, controlled alternation at the head”.

To clarify the positioning of this verification strategy, compared with Monte Carlo simulation, which requires imposing prior distributional assumptions, and with deterministic sensitivity analysis, which struggles to capture the joint drift of multiple experts’ cognitions, the Bootstrap in this paper resamples the

observed panel directly and assumption-free; the variability stems from the data itself and is therefore best suited to characterizing epistemic uncertainty under small samples.

4.5 Comparative Validation against Baseline Methods

To systematically examine the methodological characteristics of the proposed framework, it is compared against six baselines reported in Table 9: four mature methods (entropy-weighted TOPSIS, static interval DEA, VIKOR, PROMETHEE) and two closely related paradigms—interval Network DEA (Kao–Liu style, per-scenario weights) and robust (worst-case) Network DEA (pessimistic lower-bound ranking).

Table 9: Comparison of quantitative performance metrics.

Method	Scene Ranking	Kendall's W	Core Methodological Limitation
Entropy-weighted TOPSIS	$C > A > B > D$	0.997	Point-estimate measure; lacks interval uncertainty and stage decoupling
Static Interval DEA	$C > D > A > B$	0.780	End-to-end black-box ratio; ignores polarity correction and network transmission
Interval Network DEA (Kao–Liu)	$C > A > B > D$	0.960	Per-scenario self-favoring weights; cross-scenario efficiencies incomparable
Robust (worst-case) Network DEA	$C > A > B > D$	1.000	Uses only pessimistic lower bound; discards upper-bound and stage-structure information
VIKOR	$C > A > B > D$	1.000*	Point-estimate compromise solution; relies on the utility–regret trade-off and does not retain interval boundaries
PROMETHEE	$C > A > B > D$	1.000*	Point-estimate net-flow ranking; lacks interval uncertainty and stage decoupling
Proposed Framework	$C > A > B > D$	0.960	Polarity correction + linking constraints + interval uncertainty + stage decoupling (comprehensively optimal)

Note: *denotes the Kendall ordinal consistency between that baseline's ranking and the proposed framework (point-estimate baselines, not incorporated into the interval resampling); the remaining values are each method's own Bootstrap robustness.

Notably, the static interval DEA method yielded the ranking $C > D > A > B$, severely misjudging A and B to positions after D (W only 0.780). This starkly exposes that black-box ratio models, lacking stage decoupling and polarity correction, are highly susceptible to being misled by the extreme inputs at S_1 or the localized outputs at S_4 , resulting in distorted efficiency denominators/numerators. Although entropy-weighted TOPSIS, VIKOR, and PROMETHEE perform inconsistently regarding the marginal inferiority of the head C/A, their overall gradients converge with the proposed framework. With $\bar{W} = 0.960$, the proposed framework retains genuine cognitive uncertainty (as opposed to the false precision of 1.000 for point estimates) while achieving the highest ranking stability and structural explanatory power.

To validate the advantage under more closely related paradigms, two additional baselines are tested under the same Bootstrap resampling: an interval Network DEA (Kao–Liu style) with per-scenario weights,

and a robust (worst-case) Network DEA ranking by pessimistic lower bounds. The former's optimistic self-evaluation reorders the head to $A > C$; although its interval-midpoint ranking converges with the proposed framework ($C > A > B > D$, with an identical own Bootstrap robustness $\bar{W} = 0.960$), this stability is achieved at the cost of cross-scenario incomparability and loss of stage decoupling, as self-favoring weights degenerate the ranking to the terminal-stage (S_4) output. The latter yields $C > A > B > D$ with extreme stability ($W = 1.000$) but discards upper-bound and stage-structure information. Neither baseline provides the cross-scenario comparability or stage-level bottleneck identification of the proposed common-weight ISN-DEA.

Scoring each method along four objectively quantifiable dimensions yields the results reported in Table 10: the proposed framework scores 1.0 on all four dimensions, whereas the baseline methods contract on the corresponding dimensions due to point estimation, black-boxing, or the absence of polarity correction; the two closely related paradigms (interval Network DEA and robust Network DEA), although polarity-consistent, still contract on interval completeness and stage decoupling.

Table 10: Four-dimensional capability scores.

Method	Interval Information Completeness	Stage Decoupling	Statistical Robustness	Polarity Consistency
Entropy-weighted TOPSIS	0.0	0.0	0.5	0.5
Static Interval DEA	0.5	0.0	0.5	0.0
Interval Network DEA (Kao–Liu)	0.5	0.0	1.0	1.0
Robust (worst-case) Network DEA	0.0	0.0	0.5	1.0
VIKOR	0.0	0.0	0.0	0.5
PROMETHEE	0.0	0.0	0.0	0.5
Proposed Framework	1.0	1.0	1.0	1.0

Note: Scoring rules—Interval information completeness: outputs efficiency intervals together with width information, scored as 1.0; outputs only simple intervals without width analysis, scored as 0.5; point estimates only, scored as 0.0. Stage decoupling capability: decouples all four stages and locates the bottlenecks, scored as 1.0; partially decoupled, scored as 0.5; end-to-end only, scored as 0.0. Statistical robustness: Bootstrap performed with $\bar{W} \in [0.85, 0.96]$, scored as 1.0; test performed but $\bar{W}$ out of range (too low indicating insufficient robustness, too high indicating near-degenerate point estimation), scored as 0.5; no test, scored as 0.0. Polarity consistency: explicit polarity-correction mechanism, scored as 1.0; relies on manual handling, scored as 0.5; not handled, scored as 0.0.

Combining Tables 9 and 10, three conclusions can be drawn.

First, convergence validity is established: among the seven methods, six—entropy-weighted TOPSIS, VIKOR, PROMETHEE, interval Network DEA, robust Network DEA, and the proposed framework—all converge to $C > A > B > D$, with only static interval DEA misjudging A and B to positions after D owing to its end-to-end black-box ratio. The head C/A ordering falls within the near-tie interval under every method ($P_{CA} = 0.504$, Bootstrap swap probability 24.4%) and is statistically indistinguishable; the overall performance gradient (C/A markedly superior to B/D) remains consistent across all seven methods, demonstrating strong convergence validity.

Second, information completeness constitutes the core advantage of the proposed framework: entropy-weighted TOPSIS and static interval DEA output only point estimates or end-to-end ratios, with the two dimensions of “interval information completeness” and “stage decoupling capability” in Table 10 markedly contracted; the interval Network DEA outputs only simple efficiency intervals without width analysis or stage decoupling (0.5/0.0 in Table 10), and the robust Network DEA retains merely the pessimistic lower

bound (0.0/0.0); the proposed framework, by contrast, additionally outputs efficiency intervals (Table 5) and stage efficiencies (Table 6), uniquely identifies the bottleneck of each scenario (Scenario D at S_3 and the remaining scenarios at S_4), and, based on the 0.45 threshold, delivers differentiated determinations of “critical state/system failure state”, thereby providing stage-level decision evidence for targeted intervention that the baseline methods cannot offer.

Third, robustness should be interpreted in conjunction with information capacity: the point-estimate baselines exhibit (near-) degenerate stability ($\bar{W} = 0.997$ for entropy-weighted TOPSIS; $W^* = 1.000$ for VIKOR and PROMETHEE, where W^* denotes the Kendall ordinal consistency with the proposed framework, marked with * in Table 9); and the robust Network DEA attains $\bar{W} = 1.000$ because ranking by a single lower-bound series is insensitive to resampling—these reflect insensitivity rather than genuine robustness. The interval Network DEA attains $\bar{W} = 0.960$, which lies within the recommended band $[0.85, 0.96]$ and is therefore scored 1.0 on statistical robustness in Table 10; however, this apparent robustness is obtained under per-scenario self-favoring weights that sacrifice cross-scenario comparability and stage decoupling, i.e., its robustness is assessed on self-favoring, scenario-specific weights rather than a common basis. The proposed framework, with $\bar{W} = 0.960$, retains the genuine cognitive-uncertainty boundaries while achieving high robustness. Accordingly, the baselines are comparable on the “statistical robustness” dimension but diverge essentially on information completeness and stage decoupling.

4.6 Weight and Indicator Sensitivity and Scenario Expansion Drift Experiments

The objective weights of the entropy-weight method are determined by the variation structure of the indicators across scenarios; under the small-sample condition ($J = 4$), their values may be rather sensitive to the composition of the scenario set. To systematically examine the stability of the weighting system and the ranking conclusions, this paper designs sensitivity experiments along two complementary dimensions, “scenario expansion” and “indicator removal”: the former examines the drift of the weights with respect to the composition of the scenario set, and the latter examines the robustness of the ranking against single-indicator perturbation.

- (1) Scenario expansion drift experiment: a virtual fifth scenario is constructed via indicator-wise convex combinations of two randomly selected real scenarios (coefficient $t \sim U(0.2, 0.8)$, with the endpoints excluded), and the procedure is independently repeated 200 times. This construction guarantees that the virtual scores remain within the evaluation domain and constitute an “intermediate state” of the real scenarios (extrapolation scenarios would require validation with actual mission data). After each expansion, the pure objective entropy weights and the game-theoretic combination weights are recomputed, respectively, and the weight drift $\|\Delta w\|$ is measured by the L_2 norm. The results are reported in Table 11.

The results indicate that, following scenario expansion, the mean drift of the pure objective entropy weights reaches 0.1187 ± 0.0441 , confirming the pronounced sensitivity of the objective weights under small samples; by contrast, the game-theoretic combination weights suppress the mean drift to 0.1056 ± 0.0399 , an average stability improvement of 11.0%, and in all 200 trials the combination drift is strictly smaller than the objective drift (accounting for 100%), demonstrating that this suppression effect is statistically significant.

- (2) Leave-one-out indicator sensitivity. Beyond weight stability, a leave-one-out procedure is further employed to test the robustness of the ranking conclusions against single-indicator perturbation: each of the 12 indicators is removed in turn, and the composite intervals and efficiency ranking are recomputed. The results show that, except for the resource integrity retention rate RC_3 , the ranking remains $C > A > B > D$ upon removal of any single indicator (ranking-invariance rate 92%, Kendall

τ mean 0.972, minimum 0.667); only when RC_3 is removed do the head scenarios A and C swap to $A > C$ ($\tau = 0.667$). This finding carries a dual implication: on the one hand, it indicates that the indicator system is overall robust to single-indicator perturbation and that the ranking conclusion does not hinge on any specific indicator; on the other hand, it reveals that the relative ordering of A/C is comparatively sensitive to the resource-integrity retention indicator of the S_3 stage, mutually corroborating the near-tie conclusion of $P_{CA} = 0.504$ in Section 4.3.2 and further supporting the recommendation to treat the two as the same tier, while reminding managers to closely monitor equipment integrity retention in the organizational execution stage.

- (3) Cross-domain generalizability. The four-stage series topology and the “interval cloud—entropy weight—network DEA—Bootstrap” pipeline of this paper do not depend on military-domain-specific semantics and thus possess cross-domain generalization potential. In civilian logistics the stages correspond to demand forecasting, route optimization, transport execution, and performance review, and in disaster emergency response and medical supply chains they correspond to demand assessment, resource scheduling, material delivery, and effect feedback (full cross-domain mapping in Supplementary Table S4). When migrating, it suffices to redefine the indicator system and polarity attributes according to the business semantics of the target domain (Tables 2 and 3); the mathematical structure of the linking constraints and the interval solution remains unchanged.

Table 11: Comparison of weight drift under scenario expansion ($M = 200$).

Metric	Mean $\pm$ Standard Deviation
Drift of pure objective entropy weights	0.1187 $\pm$ 0.0441
Drift of game-theoretic combination weights	0.1056 $\pm$ 0.0399
Average stability improvement	11.0%
Proportion of trials with combination drift < objective drift	100%

In summary, combined with the Bootstrap ranking-robustness verification in Section 4.4, the experiments in this section constitute a complementary verification system of “ranking robustness + weight robustness”, jointly underpinning the statistical credibility of the ranking conclusions.

5 Performance Decay Mechanism Analysis and Managerial Implications

5.1 Mechanism Analysis of Scenario Performance Discrepancies

The performance gradient of the four scenarios ($C > A > B > D$) is jointly determined by the link topology and the perturbation-response mechanisms, exhibiting the structural feature of “a clear median gradient with collectively threshold-breaching lower bounds”: the median efficiencies of the top three scenarios (0.630, 0.624, 0.468) all exceed the 0.45 threshold, yet none of the lower bounds (0.270, 0.249, 0.186) crosses it, indicating that the managerial focus should shift from “pursuing median excellence” to “lifting lower-bound resilience”. Specifically, C, through cross-node information sharing, validates that “synergy gains can compensate for coordination costs”; A, by virtue of standardized processes, drives the upper bound to the theoretical optimum, validating that “process normalization can compress planning redundancy”; however, both exhibit large interval widths (0.720, 0.751) with threshold-breaching lower bounds and bottlenecks located at S_4 (see Section 4.3.1), and are therefore both determined as being in the “critical state”. B, owing to the “trading time for space” strategy, suffers systematic degradation of delivery quality and likewise belongs to the “critical state”. D, with its “input–output” transformation channel blocked by multi-source concurrent perturbations and its bottleneck concentrated at S_3 (median only 0.422), is determined as being in the “system

failure state”. At the indicator level, the S_4 weight is highly concentrated on RE_3 (0.707) and the S_1 weight on SA_3 (0.611), mutually corroborating the stage-level evidence that S_4 serves as the common bottleneck.

5.2 Targeted Optimization Strategies and Dynamic Maintenance Mechanisms

Differentiated interventions should be implemented according to the validity determination and the bottleneck stage of each scenario. For the common S_4 bottleneck of the first tier (C and A), resource allocation should follow the principle of parallel reinforcement: jointly strengthening the closed-loop knowledge consolidation and standardized strategy reuse in the performance feedback stage (focusing on lifting RE_3) and raising the efficiency lower bound to hedge against worst-case failure risk; specifically, A should introduce high-precision demand forecasting at S_1 to compress planning-timeliness fluctuation, while C should establish a dynamic resource pool to absorb coordination friction. For the “critical state” of B, the mindset of “emphasizing planning over execution” should be abandoned, and a cross-hierarchical direct command channel should be established at S_2 to mitigate the drag of the “trading time for space” strategy on S_4 delivery quality. For the “system failure state” of D, real-time state monitoring and local resource reconfiguration capabilities at S_3 should be strengthened to unblock the transformation channel, and allocation experience should be consolidated into a standardized strategy library, realizing the transition from passive response to proactive evolution.

In addition, the Bootstrap test ($\overline{W} = 0.960$) confirms that the core ranking possesses strong robustness, and its deficit below 1.0 is precisely contributed by the genuine 24.4% competitive uncertainty at the head. It is recommended to establish a link-performance early-warning system benchmarked against the DEA efficiency-interval lower bound and the 0.45 threshold: when the lower bound approaches or breaches the threshold, resource reconfiguration and plan escalation are automatically triggered; given that the lower bounds of the top three scenarios all breach the threshold, the current warning emphasis should be placed on “lower-bound lifting”. Meanwhile, frontline personnel should be regularly absorbed into the evaluation panel, and weight drift should be corrected through rolling Bootstrap and scenario expansion experiments (Section 4.6), ensuring the long-term adaptation of the model to operational evolution. The above strategies are grounded via a four-layer deployment path: a data layer interfacing the command information system under encrypted storage; an evaluation-collection layer embedding the 5-level scale and automatic polarity mapping (Table 3 and Supplementary Table S1); a computation layer encapsulating Eqs. (1)–(18) into an automated pipeline (Python 3.10, HiGHS; ≈ 10 s per run at this scale); and a display-and-warning layer outputting an interval-efficiency and bottleneck dashboard with the 0.45 warning line.

5.3 Research Limitations

Despite the methodological safeguards above, the following limitations must be objectively acknowledged. First, the empirical data derive primarily from expert judgment rather than objective operational records; the interval cloud, Bootstrap, and combination weighting mitigate but cannot fully eliminate subjective bias. Second, the expert panel size is limited ($K = 15$) and the scenario set is an orthogonal simulation spectrum; the conclusions await external validation with larger samples and real missions, and the framework can accommodate such extension through the natural expansion of K . Third, the 0.45 threshold is an empirical value; threshold sensitivity analysis shows that the ranking is robust to its variation, yet more rigorous calibration methods remain to be explored. Fourth, the empirical study is based solely on four simulated scenarios of a single equipment support system, and data-confidentiality restrictions make it difficult to obtain independent cross-unit operational data, so external validity has not been verified; the scenario-expansion drift and leave-one-out indicator experiments in Section 4.6 have

examined, from an internal-consistency perspective, the stability of the stage weights against the scenario-set composition and the robustness of the ranking against single-indicator removal, and future work will strengthen external validation through independent civilian case collaborations, cross-validation within confidentiality-permitted scopes, and recalibration after panel expansion ($K > 30$).

6 Conclusion

Addressing the performance-evaluation needs of equipment allocation supply management chains, this paper constructs an evaluation framework that deeply integrates the interval cloud model with network DEA. The interval cloud mapping characterizes the fuzziness and randomness of expert cognition; the cloud-feature fused entropy weights together with the game-theoretic combination safeguard the objectivity and operational importance of the dimensionality reduction; the polarity-corrected interval serial network DEA with intermediate linking constraints decomposes the four-stage chain performance and locates the stage bottlenecks; and the possibility degree matrix complete ranking, Bootstrap resampling, and multi-dimensional sensitivity experiments systematically verify the robustness and convergence validity of the ranking. Empirical results demonstrate that the proposed method can effectively characterize the performance degradation patterns of the chain under different perturbation scenarios, differentially identify typical performance states such as the “critical state” and the “system failure state”, and output strategy rankings with statistical significance. The research not only fills the methodological gap in the “process–performance–resilience” evaluation of military logistics chains but also provides a reproducible and extensible theoretical tool for the performance measurement of complex supply chain systems; its four-stage serial topology and evaluation pipeline possess cross-domain generalization potential.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Tianqiao Yan and Yabin Wang; methodology, Tianqiao Yan and Cenyu Hu; software, Tianqiao Yan and Cenyu Hu; validation, Tianqiao Yan, Yabin Wang and Cenyu Hu; formal analysis, Tianqiao Yan and Cenyu Hu; investigation, Tianqiao Yan and Cenyu Hu; resources, Yabin Wang; data curation, Tianqiao Yan and Cenyu Hu; writing—original draft preparation, Tianqiao Yan; writing—review and editing, Yabin Wang, Tianqiao Yan and Cenyu Hu; visualization, Tianqiao Yan; supervision, Yabin Wang; project administration, Yabin Wang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The anonymized $15 \times 4 \times 12$ expert scoring matrix and the full Python implementation of the evaluation pipeline (Eqs. (1)–(18)) are provided as supplementary files, together with Supplementary Tables S1–S4, alongside the online version of this article; any additional materials are available from the corresponding author, Yabin Wang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at <https://www.techscience.com/doi/10.32604/cmc.2026.088982/s1>.

References

1. Tavassoli M. Fuzzy fair efficiency assessment in network data envelopment analysis models for complex system: an application in sustainable supply chain management. *Sustain Comput Inform Syst.* 2025;46(6):101105. doi:10.1016/j.suscom.2025.101105.

2. Tavassoli M. Resilience-sustainability assessment of electricity supply chain under uncertainty: a new fuzzy dynamic network DEA based on rang adjusted measure. *OPSEARCH*. 2024;113(3):96. doi:10.1007/s12597-024-00827-4.
3. Pishdar M, Shakib MD, Antucheviciene J, Vilkonis A. Interval type-2 fuzzy super SBM network DEA for assessing sustainability performance of third-party logistics service providers considering circular economy strategies in the era of Industry 4.0. *Sustainability*. 2021;13(11):6497. doi:10.3390/su13116497.
4. Primadasa R, Kusriani E, Mansur A. Evaluating sustainable supply chain performance in SMEs: an integrated entropy-data envelopment analysis (DEA) and maturity classification framework. *Process Integr Optim Sustain*. 2026;15(2):1379. doi:10.1007/s41660-026-00778-x.
5. Jiang T, Yu C. Analysis and improvement of equipment maintenance management fee allocation efficiency based on DEA theory. In: *Proceedings of the 2021 International Conference on E-Commerce and E-Management (ICECEM)*; 2021 Sep 24–26; Dalian, China. p. 99–104. doi:10.1109/icecem54757.2021.00028.
6. Thomas AV, Mahanty B. Dynamic assessment of control system designs of information shared supply chain network experiencing supplier disruption. *Oper Res*. 2021;21(1):425–51. doi:10.1007/s12351-018-0435-9.
7. Soheilrad S, Govindan K, Mardani A, Zavadskas EK, Nilashi M, Zakuan N. Application of data envelopment analysis models in supply chain management: a systematic review and meta-analysis. *Ann Oper Res*. 2018;271(2):915–69. doi:10.1007/s10479-017-2605-1.
8. Xu YW, Zhang HJ, Cheng K, Zhang ZX, Chen YT. Efficiency measurement in multi-period network DEA model with feedback. *Expert Syst Appl*. 2021;175(4):114815. doi:10.1016/j.eswa.2021.114815.
9. Kalyan S, Bansal P, Kumar S. A novel hybrid AHP-entropy weighted dynamic network DEA framework for comprehensive efficiency assessment. *Expert Syst Appl*. 2026;296(1):129137. doi:10.1016/j.eswa.2025.129137.
10. Wang H, Liu Z, Zhou XY, Li L, Fei S, Wang X. Multi-ship collision risk situation assessment based on finite interval cloud model. *Reliab Eng Syst Saf*. 2026;271(4):112290. doi:10.1016/j.res.2026.112290.
11. Yang M, Zhang H, Ye W. Risk assessment of equipment procurement contract performance based on combination weighting two-dimensional cloud model. *Int J Model Simul Sci Comput*. 2026;17:2550083. doi:10.1142/s1793962325500837.
12. Zhao Y, Luan T, Li X, Wang K, Shi S. Application of Bayesian optimization-based cloud model in fire risk assessment of distributed photovoltaic power plants. *Energy Sources Part A Recovery Util Environ Eff*. 2025;47(2):2551874. doi:10.1080/15567036.2025.2551874.
13. Lu C, Zhao X, Yu M, Gao S, Guo Y, Wang Y. Evaluation of the effectiveness of emergency medical rescue equipment system based on cloud model. *Reliab Eng Syst Saf*. 2026;269(4):112120. doi:10.1016/j.res.2025.112120.
14. Guan J, Liu J, Chen H, Bi W. A multi-criteria decision-making approach for equipment evaluation based on cloud model and VIKOR method. *Int J Adv Comput Sci Appl*. 2024;15(7):1311–21. doi:10.14569/ijacsa.2024.01507127.
15. Fu Z, Zhang T, Tan C, Zhang Y. Weapon equipment operational index evaluation method based on cloud model. *J Phys Conf Ser*. 2023;2650(1):012014. doi:10.1088/1742-6596/2650/1/012014.
16. Peng L, Tu J, Cai S, Chen D. Power distribution Internet of Things security risk evaluation based on combined weighting and cloud model. *Entropy*. 2026;28(4):433. doi:10.3390/e28040433.
17. Chang H, Zhang W, Yin Y. Research progress on logistics support capability evaluation model and its application. In: *Proceedings of the International Conference on Computer Applications and Information Security (ICCAIS 2021)*; 2021 Dec 18–19; Wuhan, China. doi:10.1117/12.2637844.
18. Qahtan S, Alsattar HA, Zaidan AA, Deveci M, Pamucar D, Delen D. Performance assessment of sustainable transportation in the shipping industry using a q-rung orthopair fuzzy rough sets-based decision making methodology. *Expert Syst Appl*. 2023;223(1):119958. doi:10.1016/j.eswa.2023.119958.
19. Wu A, Sun Y, Zhang H, Sun L, Wang X, Li B. Research on resilience evaluation of coal industrial chain and supply chain based on interval type-2F-PT-TOPSIS. *Processes*. 2023;11(2):566. doi:10.3390/pr11020566.
20. Topcu I, Karpak B, Ülengin F, Aktas E. Big data analytics in supply chain management: uncovering emerging trends through a bibliometric network analysis and a systematic literature review. *J Enterp Inf Manag*. 2026;39(3):1166–99. doi:10.1108/jeim-07-2024-0374.

21. Wang NN, Jian M, Fang X. Simulation-based robustness evaluation of coordination under stochastic demand. *Int J Simul Model*. 2025;24(4):730–41. doi:10.2507/ijsimm24-4-co20.
22. Michali M. Bootstrapping in network data envelopment analysis [dissertation]. Birmingham, UK: Aston University; 2022.
23. Shishodia A, Sharma R, Rajesh R, Munim ZH. Supply chain resilience: a review, conceptual framework and future research. *Int J Logist Manag*. 2023;34(4):879–908. doi:10.1108/ijlm-03-2021-0169.
24. Katsaliaki K, Galetsi P, Kumar S. Supply chain disruptions and resilience: a major review and future research agenda. *Ann Oper Res*. 2022;319(1):965–1002. doi:10.1007/s10479-020-03912-1.
25. Negri M, Cagno E, Colicchia C, Sarkis J. Integrating sustainability and resilience in the supply chain: a systematic literature review and a research agenda. *Bus Strategy Environ*. 2021;30(7):2858–86. doi:10.1002/bse.2776.
26. Ma C, Zhang L, You L, Tian W. A review of supply chain resilience: a network modeling perspective. *Appl Sci*. 2025;15(1):265. doi:10.3390/app15010265.